Redbooks
ibm.com/redbooks

Implementing an IBM InfoSphere BigInsights Cluster using Linux on Power

Dino Quintero
Esteban Arias Navarro
Pablo Barquero Garro
Rodrigo Ceron Ferreira de Castro
Luis Carlos Cruz Huertas
Peng Jiang
Franz Friedrich Liebinger Portela
Peter McCullagh
Ichsan Mulia Permata
Joanna Wong
John Wright

Analytics

Power Systems

IBM
Redbooks

International Technical Support Organization

Implementing an IBM InfoSphere BigInsights Cluster using Linux on Power

June 2015

SG24-8248-00

Note: Before using this information and the product it supports, read the information in “Notices” on page vii.

First Edition (June 2015)

This edition applies to Linux Red Hat 6.5, IBM InfoSphere BigInsights Version 3.0, IBM Spectrum Scale (formerly GPFS) Version 3.5.0-17, IBM Platform Platform Cluster Manager Version 4.2, xCAT 2.8.5, and IBM Installation Toolkit Version 5.6.

Contents

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

AIX®
BigInsights™
DataStage®
DB2®
developerWorks®
Easy Tier®
Global Business Services®
GPFS™
IBM®
IBM Elastic Storage™
IBM Spectrum™
IBM Watson™
InfoSphere®
LSF®
OpenPower™
POWER®
Power Systems™
Power Systems Software™
POWER7®
POWER8™
PowerLinux™
PowerVM®
PureData®
PureFlex®
PureSystems®
Redbooks®
Redbooks (logo) ®
Symphony®
System i®
Tivoli®

The following terms are trademarks of other companies:

SoftLayer, and SoftLayer device are trademarks or registered trademarks of SoftLayer, Inc., an IBM Company.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

UNIX is a registered trademark of The Open Group in the United States and other countries.

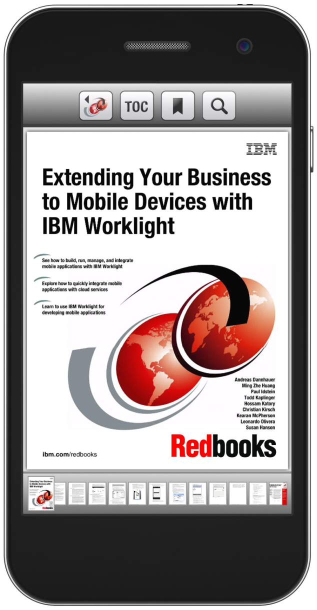
Other company, product, or service names may be trademarks or service marks of others.

Find and read thousands of IBM Redbooks publications

- ► Search, bookmark, save and organize favorites
- ► Get up-to-the-minute Redbooks news and announcements
- ► Link to the latest Redbooks blogs and videos

Get the latest version of the Redbooks Mobile App

◄ iOS **Download Now** Android ►



IBM

Extending Your Business to Mobile Devices with IBM Worklight

See how to build, run, manage, and integrate mobile applications with IBM Worklight

Explore how to quickly integrate mobile applications with cloud services

Learn to use IBM Worklight for developing mobile applications

Andreas Dannhauer
Ming Zhe Huang
Paul Idstein
Todd Kaplinger
Hossam Katory
Christian Kirsch
Kearan McPherson
Leonardo Olivera
Susan Hanson

ibm.com/redbooks

Redbooks

Promote your business in an IBM Redbooks publication

Place a Sponsorship Promotion in an IBM® Redbooks® publication, featuring your business or solution with a link to your web site.

Qualified IBM Business Partners may place a full page promotion in the most popular Redbooks publications. Imagine the power of being seen by users who download millions of Redbooks publications each year!

ibm.com/Redbooks
About Redbooks → Business Partner Programs

THIS PAGE INTENTIONALLY LEFT BLANK

Preface

This IBM® Redbooks® publication demonstrates and documents how to implement and manage an IBM PowerLinux™ cluster for big data focusing on hardware management, operating systems provisioning, application provisioning, cluster readiness check, hardware, operating system, IBM InfoSphere® BigInsights™, IBM Platform Symphony®, IBM Spectrum™ Scale (formerly IBM GPFS™), applications monitoring, and performance tuning. This publication shows that IBM PowerLinux clustering solutions (hardware and software) deliver significant value to clients that need cost-effective, highly scalable, and robust solutions for big data and analytics workloads.

This book documents and addresses topics on how to use IBM Platform Cluster Manager to manage PowerLinux BigData data clusters through IBM InfoSphere BigInsights, Spectrum Scale, and Platform Symphony. This book documents how to set up and manage a big data cluster on PowerLinux servers to customize application and programming solutions, and to tune applications to use IBM hardware architectures. This document uses the architectural technologies and the software solutions that are available from IBM to help solve challenging technical and business problems.

This book is targeted at technical professionals (consultants, technical support staff, IT Architects, and IT Specialists) that are responsible for delivering cost-effective Linux on IBM Power Systems™ solutions that help uncover insights among client's data so they can act to optimize business results, product development, and scientific discoveries.

Authors

This book was produced by a team of specialists from around the world working at the International Technical Support Organization, Poughkeepsie Center.

Dino Quintero is a technical Project Leader and an IT Generalist with the International Technical Support Organization (ITSO) in Poughkeepsie, NY. His areas of expertise include enterprise continuous availability planning and implementation, enterprise systems management, virtualization, and clustering solutions. He is an Open Group Master Certified IT Specialist - Server Systems. He holds a master’s degree in Computing Information Systems, and a Bachelor of Science degree in Computer Science from Marist College.

Esteban Arias Navarro is a Software Engineer with an interest in cloud computing. He has a strong background in VMware private clouds as an administrator and implementation consultant. He also is an Openstack administrator and developer. He works as an IBM Tivoli® System Automation SME for advanced cloud solutions for customers within IBM SoftLayer®, with a focus on networking, storage, and compute services.

Pablo Barquero Garro is a Certified Experienced Architect working for GTS Global Solutions in Costa Rica. He has more than 15 years of experience with information technology in different fields, such as programming, information systems, artificial intelligence, and cloud management systems, and database and middleware administration, integration, and support. Pablo has been working on cloud-related projects since 2011, which includes architecture, development, evangelization, adoption and enablement, and legacy infrastructure migration and integration to private, public, or hybrid clouds by using closed or open standards cloud technologies.

Rodrigo Ceron Ferreira de Castro is a Master Inventor and Consultant at IBM Lab Services and Training Latin America, in Brazil. He has 14 years of experience in the UNIX/Linux area and over 10 years working at IBM, where he received eight intellectual property patents in multiple areas. He graduated with honors in Computer Engineering from the University of Campinas (UNICAMP) and holds an IEEE CSDA credential. He is also an IBM Certified Expert IT Specialist. His areas of expertise include IBM Power Systems high availability, performance, cloud, and analytics, including SAP HANA (IBM POWER® and x86 systems). He has also published papers about Operations Research in IBM developerWorks®.

Luis Carlos Cruz Huertas is a Transition and Transformation Solution Architect with IBM GTS Delivery. During his two-plus years at IBM, he has held research positions in the big data analytics, mobility, and cloud areas. He also has held several positions in the Midrange and Storage Technical Solution Architecture areas. Luis comes from GBM, which is a strategic IBM Alliance company in Latin America, where he holds positions in strategy, IBM Tivoli Architecture, project management, and management positions. He primarily works in Tivoli Service Management capabilities, management systems, data warehouse infrastructure, information integration, database administration, performance management, and database development technology. He has been a prominent speaker at industry events, such as IBM Edge and customer briefings, and is a frequent contributor to industry articles, analyst research, and other publications.

Peng Jiang is an Advisory Software Engineer at the STG lab in Beijing, China. He has worked with IBM since 2008. He is a Team Leader of IBM Spectrum Scale (GPFS), supports solutions such High Performance Computing, cloud, and analytics. He has extensive experience with the infrastructure and software stacks of the solutions. He is certified in PMP, IBM AIX®, and Linux, and holds a master's degree in Software Engineering from Beijing University of Aeronautics & Astronautics.

Franz Friedrich Liebinger Portela is a Datacenter Relocation Architect with over 15 years of IT experience in serving Fortune 500 and government clients in the Americas region as an IT Solution Integrator and Cloud Solution Architect. Franz is an IBM Certified Expert Architect with a focus on public, private, and hybrid cloud implementation, and migrations from legacy or virtual environments to cloud-based solutions.

Peter McCullagh is a big data consultant from the UK and has been working with InfoSphere BigInsights since 2011. Peter holds a degree in Chemistry from the University of Bristol. He is a member of the IBM UKI Software Group (SWG) Services big data team and works with several IBM products, including InfoSphere BigInsights, Watson Explorer, InfoSphere Streams, IBM DB2®, and IBM Smart Analytics System.

Ichsan Mulia Permata is a Client Technical Specialist at IBM Indonesia. He has expertise in IBM Power Systems with a focus on AIX and Linux operating systems. He works with IBM Power Systems Software™ as well, such as IBM PowerVM®, PowerVC, and Spectrum Scale. He champions PowerLinux initiatives, proof of concept, and ISV porting in Indonesia. He is involved in building the Cloud Demo Center and leads the big data Demo Center at IBM Indonesia. He recently was given an Outstanding Technical Achievement Award. Ichsan holds a bachelor's degree in Electrical Engineering from Bandung Institute of Technology.

Joanna Wong is an Executive IT Specialist for the IBM STG Worldwide Client Centers, focusing on IBM Platform Computing solutions. She has experience in HPC application optimization and large systems scalability performance on the x86-64 and IBM POWER architectures. Joanna also has industry experience in engagements with Oracle database server and Enterprise Application Integration. Joanna has an Artium Baccalaureatus degree in Physics from Princeton University, and a Master of Science degree and a doctorate degree in Theoretical Physics from Cornell University. She also has a Master of Business Administration degree from the Walter Haas School of Business at the University of California at Berkeley.

John Wright is an IBM Power Systems, AIX, and PureSystems® Specialist for IBM GTS in the UK. He has over 13 years of experience in IT across many industries, including the automotive, financial, and service sectors. He holds certifications from IBM, Oracle, and SUSE. His areas of expertise are virtualization, converged systems, high availability, and data center migrations.

Thanks to the following people for their contributions to this project:

Richard Conway, Don Brennan, Dave Bennin
International Technical Support Organization, Poughkeepsie Center

Piyush Chaudhary, Gina L. King
IBM US

Mehdi Bozzo-Rey
IBM Canada

Guangcheng Li, Junhui Bu, Peng Jia, Xinyu Pan, Lei Ai, Xiaoming Bao, Rong Li, Yu Song
IBM China

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

- Use the online **Contact us** review Redbooks form found at:

  ibm.com/redbooks

- Send your comments in an email to:

  redbooks@us.ibm.com

- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

1

Introduction to the solution

This book provides information about how to deploy an IBM InfoSphere BigInsights cluster on IBM Power Systems servers running Linux. It also provides information about how to manage the clustered solution, including automated deployment by using IBM Platform Cluster Manager.

This chapter covers the following topics:

- InfoSphere BigInsights
- Linux on IBM Power Systems
- IBM Platform Symphony
- IBM Spectrum Scale-FPO (formerly GPFS-FPO)
- IBM Platform Cluster Manager

After reading this chapter, you will understand the advantages of the complete solution that uses IBM pieces for the Hadoop framework (IBM Platform Symphony and IBM Spectrum Scale, formerly GPFS), Linux on IBM POWER, and Platform Cluster Manager.

1.1 InfoSphere BigInsights

InfoSphere BigInsights is a software solution package from IBM that is based on open source Hadoop. The solution contains all the open source components that you expect from a Hadoop deployment, and comes with other IBM technologies to maximize the value of the deployment. Additional information about Hadoop can be found in *Implementing IBM InfoSphere BigInsights on IBM System x*, SG24-8077. The various components that are installed as part of InfoSphere BigInsights are shown in Figure 1-1.

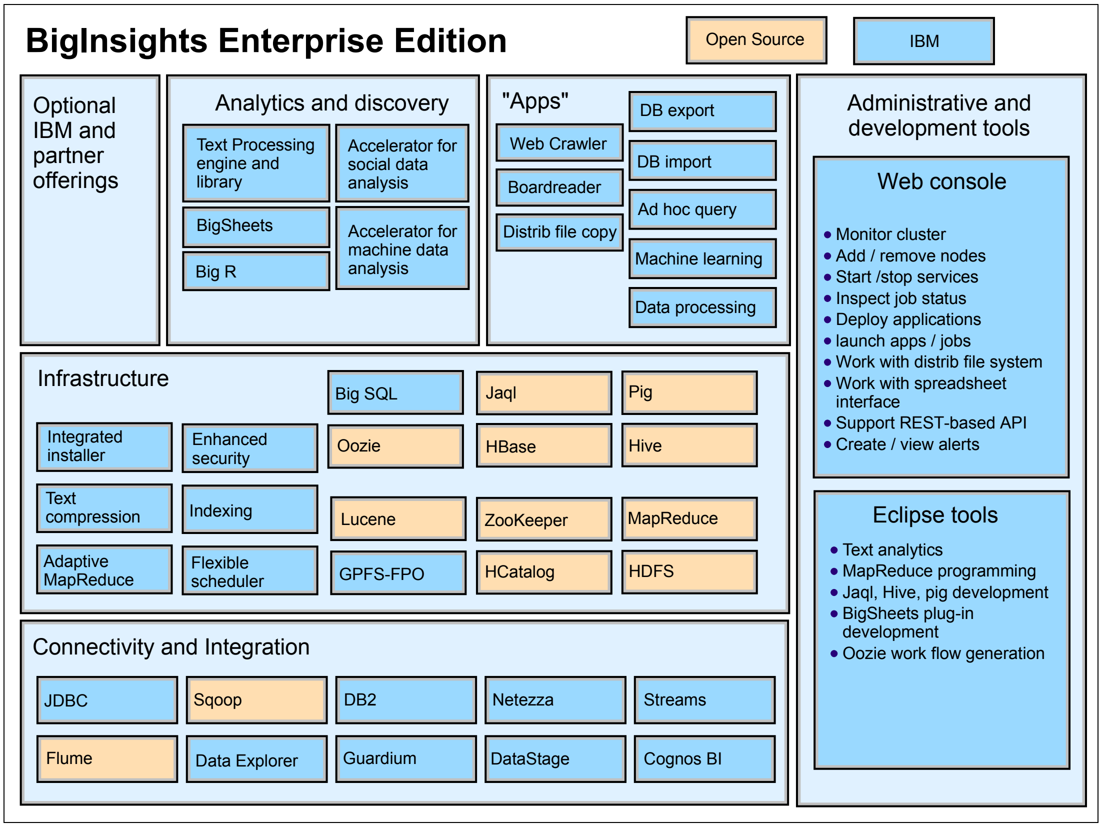


Figure 1-1 InfoSphere BigInsights components - yellow are open source and blue are added by IBM

This suite of components covers the full spectrum of what you might want to do with your big data cluster. Here, you have import tools, a file system for storage, databases to apply structure to the data, and a powerful framework for analytics on a massive scale. This is a complete framework that gathers insight from your data. More important is that InfoSphere BigInsights is a flexible platform that grows with your analytics requirements, which means that adding nodes and scaling out is a simple process. Also, this technology can be used as a powerful cloud solution.

InfoSphere BigInsights is the Hadoop offering in the IBM BigData portfolio and is built to complement other technologies that are built for big data. As such, it integrates seamlessly with IBM InfoSphere Streams, IBM Watson™ Data Explorer, IBM InfoSphere Master Data Management, and data storage solutions, such as IBM DB2.

1.1.1 The advantages of InfoSphere BigInsights

IBM added components to provide tangible benefits over the open source version of Hadoop. Some of these benefits are listed in this section.

Integrated installation

The InfoSphere BigInsights installer provides an easy to use graphical user interface (GUI) to set up your environment. This installer decreases the amount of time that is required to start getting value from your cluster because you install only one product, not all the individual components. Installing many components individually is not only time-consuming, but issues can arise around which version of each component works with all the other components you have installed. With InfoSphere BigInsights, you know that all of the components are chosen to work in harmony.

InfoSphere BigInsights deployments are cluster deployments on many nodes (a minimum of four). The installer provides a single point from which you install all nodes. The components that you want to install on each node are configured in the GUI before the installation. The installer can even be made rack-aware and install across multiple racks.

The installer makes the installation simple. Many of the settings that you might want to use are already set in the installer to allow clusters to be built quickly. That said, if you want to configure settings at installation time, you can.

After the cluster is built, the installer has an integrated verification, which means that when you finish the installation, you know that you have a healthy cluster ready to start running analytics.

BigSQL

Businesses across the world have been using Structured Query Language (SQL) to analyze their data for decades, primarily because it is simple and powerful. This situation has created departments with strong SQL skills, and BigSQL uses those skills to maximize the value that you can pull from your InfoSphere BigInsights environment.

BigSQL uses SQL to produce a view of your existing data. There is no new proprietary storage format. The table definitions are shared with Hive, but unlike working with Hive directly, you can use your usual SQL queries. In fact, the same SQL that you use on your data warehouse runs with few or no modifications, including the use of subqueries and all standard join operations.

In addition to working with the data that you have stored in Hive, BigSQL can federate to other data stores, which facilitates joining across databases. Because the query optimizer understands the capabilities of the external systems, it can push as much work as possible to each system to process.

BigSheets

BigSheets is a spreadsheet-like, web-based application that allows dynamic analysis of data. With BigSheets, users can work with smaller subsets of the data to ensure that they are performing high-value transformations before performing those transformations on the whole cluster. This situation keeps the workload on the system down while providing valuable insight.

Flexibility of the underlying file system

InfoSphere BigInsights, like traditional Hadoop, can run on the Hadoop distributed file system (HDFS), but InfoSphere BigInsights is not limited to running on HDFS. Another supported option is IBM Spectrum Scale-FPO (formerly GPFS - File Placement Optimizer). The benefits of Spectrum Scale-FPO over HDFS are explained in 1.4, “IBM Spectrum Scale-FPO (formerly GPFS-FPO)” on page 10.

Integration with existing architectures

Postinstallation, the first thing that you probably do is start importing data into your InfoSphere BigInsights environment. InfoSphere BigInsights easily connects to IBM InfoSphere DataStage®. InfoSphere BigInsights has both a standard JDBC connector to provide connectivity to existing systems, such as Microsoft SQL Server, and Oracle. InfoSphere BigInsights also contains several custom high-speed connectors to other IBM products, including DB2 PureData® System for Operational Analytics, DB2 PureData for Analytics, IBM InfoSphere Warehouse, and DB2.

BigR

R is an open source language that is used for statistical analysis and creating graphical displays of data. Although InfoSphere BigInsights does not install R itself, it does provide BigR. If you decide to install and use R, BigR is a collection of functions that integrate with R and remove the complexity of converting these jobs into MapReduce. The result is that your BigR jobs scale with your cluster.

Support

Although InfoSphere BigInsights is based on open source software, it is an IBM product, so it is packaged with the enterprise-level support that you expect for IBM software. Help can be reached by phone 24 hours a day, or you can have onsite help from IBM Global Business Services®.

1.1.2 What practical problems can be solved with InfoSphere BigInsights

InfoSphere BigInsights is a low-cost file system with a job scheduling framework that is attached. It is an ideal environment for running batch jobs of analytics on large data sets.

The big data tenets: Storage

InfoSphere BigInsights is often compared to a data reservoir. There is a large body of stationary data and you can control what goes in and what comes out. The “Vs” of big data are volume, variety, velocity, and veracity . InfoSphere BigInsights addresses two quite well: volume and variety.

Volume

InfoSphere BigInsights has Hadoop at its core, so it scales well. It can handle petabytes of data and thousands of nodes. These features open two significant uses for InfoSphere BigInsights:

- After data has passed its most useful stage, it is often moved from the data warehouse and into low-cost archive storage, for example, tape. Although this data is still technically accessible, in reality it is no longer accessible, which potentially means that insight that might be gleamed from this data is lost. Having this data stored and accessible in InfoSphere BigInsights means that this insight can be revealed to the business for a competitive advantage.

- You can also use InfoSphere BigInsights high-volume capabilities to collect multiple data sources into one place. This capability can be useful if you are using InfoSphere BigInsights as a staging area to gather, cleanse, and join large sets of data before exporting the data into your data warehouse. This process takes the workload of your data warehouse and reduces the amount of storage.

Variety

Sometimes you cannot join sources together in your data warehouse because it is difficult to store structured, semi-structured, and unstructured data in the same place. With InfoSphere BigInsights, unstructured data and structured data can exist together. This situation allows businesses to explore relationships that they previously could not explore. New relationships can lead to new insights, which in turn can lead to a competitive advantage.

Analytics

There is a range of analytic functions that you can perform on data in an InfoSphere BigInsights cluster. Which types you decide to use depends on the data, what you want to extract from it, and how you want to interact with the data.

MapReduce

MapReduce was the original type of workload that was designed for Hadoop. MapReduce breaks up the job into segments that are called *maps* that can be passed to data nodes, and then these multiple maps are reduced to one answer that is returned to the user. You can use this distributed approach to working with data to analyze high volumes efficiently and maximize the capabilities of your nodes.

BigSQL

BigSQL is an additional IBM component that allows you to access data that is stored in your InfoSphere BigInsights cluster by using SQL. BigSQL does not use MapReduce; instead, it uses an advanced database engine to propagate queries across the cluster. This process makes the data in your Hadoop environment more accessible than ever before.

MapReduce and BigSQL answer different challenges, and they are only two of the many components that make up InfoSphere BigInsights. Together, they make up a series of tools that you can use to overcome the challenges of big data and help revolutionize the way that you use your data.

1.2 Linux on IBM Power Systems

Linux on IBM Power Systems offers a highly performant and highly reliable infrastructure for your big data environment. IBM POWER8™ is the first processor that is designed for big data, and it can handle large volumes of both structured and unstructured data. POWER8 provides continuous data load, parallel processing, massive memory bandwidth, and massive I/O bandwidth to support compute-intensive or data-intensive big data workloads.

1.2.1 IBM Power Systems

This section provides information about the IBM POWER8.

POWER8 processor

The POWER8 processor is manufactured by IBM 22 nm Silicon-On-Insulator (SOI) technology. Each chip is 650 square millimeters and contains 4.2 billion transistors. The POWER8 processor chip, which is shown in Figure 1-2, contains 12 cores. Each core has its own 512 KB L2 cache and all cores share a 96 MB L3 embedded dynamic random access memory (eDRAM) cache, two memory controllers, PCIe Gen3 I/O controllers, and an interconnection system that connects all components within the chip. The interconnect also extends through module and system board technology to other POWER8 processors, DDR3 memory, and various I/O devices. The number of memory controllers, memory buffer chips, PCIe lanes, and cores that are available for use depends on the POWER8 system.

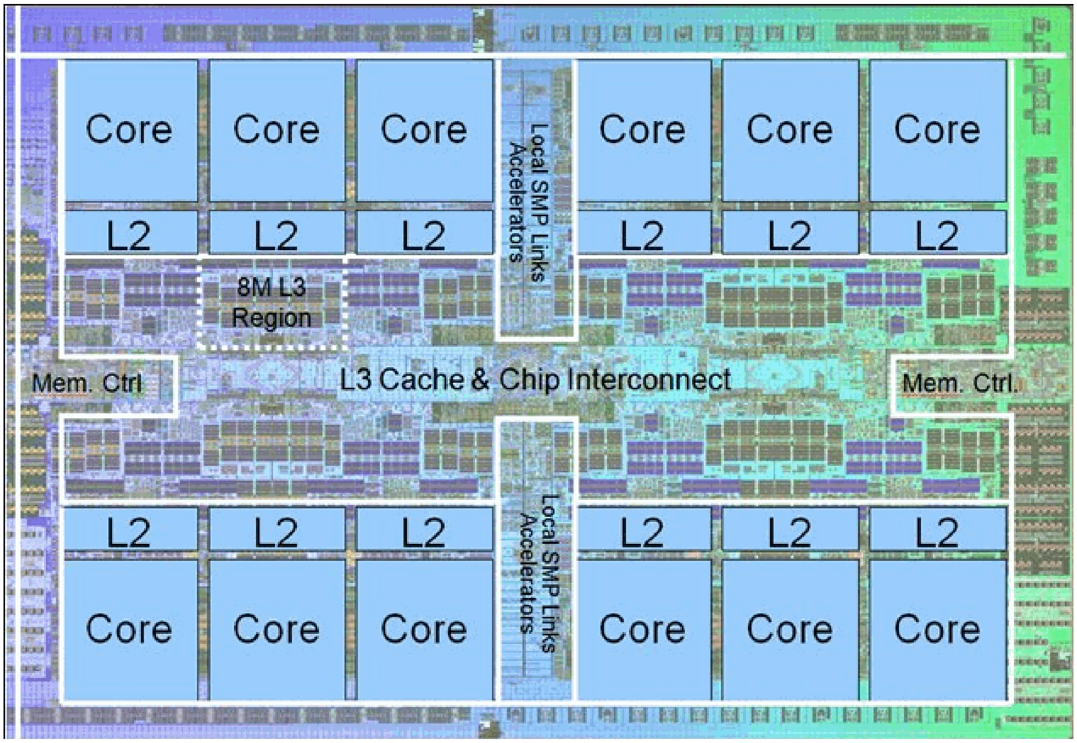


Figure 1-2 IBM POWER8 processor architecture

Each core is a 64-bit implementation of the IBM Power ISA V2.07 and has the following features:

- Multi-thread design that supports up to an eight-way SMT.
- 32 KB, eight-way set-associative L1 i-cache.
- 64 KB, eight-way set-associative L1 d-cache.
- 72-entry ERAT for effective-to-real address translation for instructions (fully associative).
- 48-entry primary ERAT for effective-to-real address translation for data (fully associative).
- Aggressive branch prediction that uses local and global prediction tables with a selector table to choose the best predictor.
- 16-entry link stack.
- 256-entry count cache.
- Aggressive out-of-order execution.
- Two symmetric fixed-point execution units.

- Two symmetric load/store units and two load units, all four of which can also run simple fixed-point instructions.
- An integrated, multipipeline vector-scalar floating point unit that supports up to eight flops per cycle (four double precision or eight single precision) and that runs the following Scalar and Single Instruction Multiple Data (SIMD)-type instructions:
  - The Vector Multimedia Extension (VMX) instruction set
  - The Vector Scalar Extension (VSX) instruction set
- On-chip encryption.
- Hardware data prefetching with 16 independent data streams and software control.
- Hardware decimal floating point (DFP) capability.

The POWER8 processor is designed for system offerings from single-socket blades to multi-socket enterprise servers. It incorporates a triple-scope broadcast coherence protocol over local and global symmetric multiprocessor (SMP) links to provide superior scaling attributes.

POWER8 and Simultaneous Multi-Threading

The Simultaneous Multi-Threading (SMT) technology allows applications to use the same processor core through multiple threads by performing an intelligent processor register allocation to each of the threads. The processor registers that are not in use by a running machine instruction can be used by another machine instruction, which is how to achieve parallel use of the processor.

The POWER8 chip is enhanced and can work with eight simultaneous threads on a core. This is a 2x improvement over its predecessor, the IBM POWER7® chip.

POWER8 cache architecture and memory buffer chip

The POWER8 architecture uses a Non-Uniform Cache Architecture (NUCA) cache policy, which benefits data-intensive workloads. Each processor core is associated with a L3 cache. However, the NUCA cache policy provides the capability for a processor core to access an L3 cache that is associated with another core processor. It provides benefits when one job must access an L3 cache larger than the associated L3 core cache on which it is running. Figure 1-3 shows this architecture.

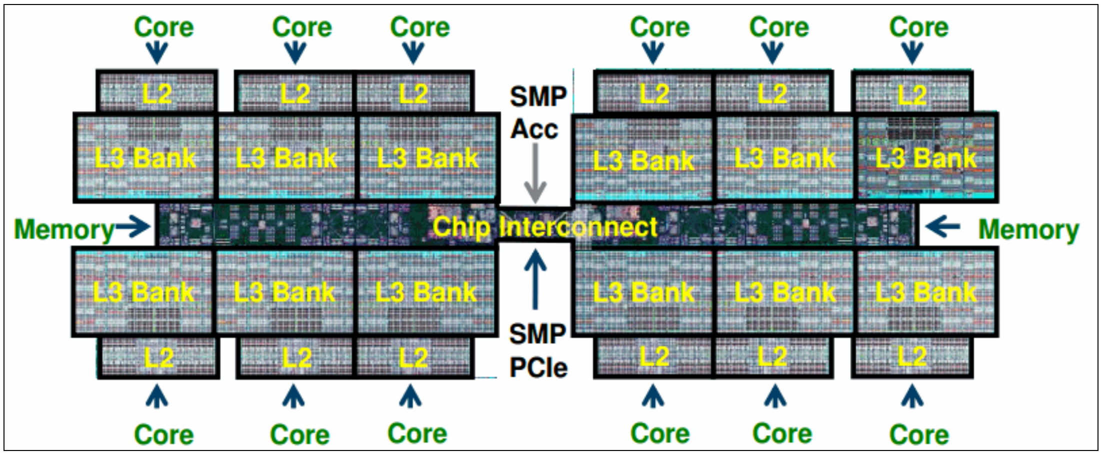


Figure 1-3 POWER8 non-uniform cache architecture

The POWER8 processor connects to memory through the memory buffer chip. The memory buffer chip has scheduling logic to achieve a lower latency between the memory chip and the processor chip. It has a 16-MB cache that performs as an L4 cache. Links between the POWER8 processor and the memory buffer chip run at 9.6 GBps each, with a 40-ns latency. Figure 1-4 shows this concept.

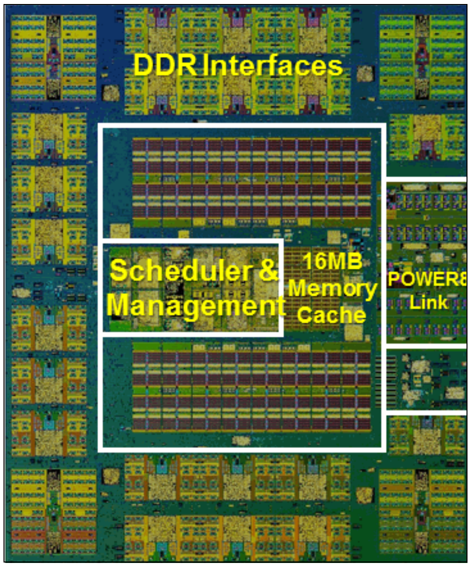


Figure 1-4 The memory buffer chip

Power Systems PCIe Gen3 and Coherent Accelerator Processor Interface technology

The new generation of Power Systems use PCIe Gen3 and Coherent Accelerator Processor Interface (CAPI) technology. It increases I/O bandwidth to boost data transfer from disk to memory. CAPI sits directly on the POWER8 chip and works with the same memory addresses that the processor uses, so it reduces the latency of the typical I/O model. With CAPI, this additional hardware accelerator can access memory directly without operating system and device driver impact. Some hardware accelerators that are projected to use the CAPI technology are FPGAs, flash cards, and GPUs.

Power Systems Easy Tier Internal Disk

The new generation of Power Systems supports IBM Easy Tier® (Automatic Tiering) in its internal disks. This technology moves hot data from SAS disk to SSD automatically. It brings benefits to big data environments, which need SAS disks for capacity, but also need SSDs to have low latency data access.

1.2.2 Linux for big data

Linux is the fastest growing operating system at the time of writing. It is multi-platform, follows open standards, is adopted by multiple industries, has its development performed by thousands of contributors around the globe, and most importantly, has enterprise support available through partners such as Red Hat, SUSE, and Canonical. Also, IBM is committed to the development of technologies that run on Linux, and having its big data stack run on it, along with a one-billion-dollar investment in Linux, announced in 2013, demonstrates that it is a *de facto* standard for the world of big data.

Linux has been able to run on IBM Power Systems servers for a long time. If you have POSIX-compliant software that runs on Linux, chances are that it can run on multiple architectures. Moreover, support for little-endian Linux distributions to run on the POWER8 chip was announced and more development efforts are under way, which means that you can more easily run existing x86 Linux applications on POWER8 servers. Also, with the competitive pricing of POWER8 Linux servers such as the S822L, opting to run MapReduce based applications on POWER8 provides you with all the features and advantages of the POWER hardware: reliability, availability, scalability, security, and performance.

In summary, the combination of POWER8 and Linux for big data workloads ensures that you have an optimized and resilient hardware technology in hand for data processing, along with all of the open standards that are provided by the operating system.

> **Note:** The IBM Watson technology runs on top of Linux to efficiently process large amounts of structured and unstructured data.

1.3 IBM Platform Symphony

IBM Platform Symphony is a resource scheduler for grid environments. It works with grid-enabled applications and can provide high resource utilization rates along with low latency for certain types of jobs.

Platform Symphony can be used in a InfoSphere BigInsights environment as a job scheduler for MapReduce tasks. Platform Symphony can replace the open source Hadoop scheduler in a MapReduce based framework and provide advantages such as the following ones:

- Better performance by providing lower latency for certain MapReduce based jobs.
- Dynamic resource management that is based on slot allocation according to job priority and server thresholds.
- A fair-share scheduling scheme with 10.000 priority levels for jobs of a application.
- A complete set of management tools for providing reports, job tracking, and alerting.
- Reliability by providing a redundant architecture for MapReduce jobs in terms of name nodes (in case the Hadoop file system is in use), job trackers, and task trackers.
- Support for rolling upgrades, hence maximizing uptime of your applications.
- Open, so it is compatible with multiple APIs and languages such as Hive, Pig, Java, and others. Also, it is compatible with both HDFS and IBM Spectrum Scale (formerly GPFS).

Also, using Platform Symphony as a scheduler for an InfoSphere BigInsights environment is a choice that you can make while installing InfoSphere BigInsights, as Platform Symphony comes pre-packaged with InfoSphere BigInsights and is handled by the installer in an integrated fashion. As a bundled offer with InfoSphere BigInsights, you benefit from extensive hours of integration tests between Platform Symphony and InfoSphere BigInsights.

For more information about using Platform Symphony within a MapReduce framework, including reference architectures, see *IBM Technical Computing Clouds*, SG24-8144.

1.4 IBM Spectrum Scale-FPO (formerly GPFS-FPO)

IBM Spectrum Scale-FPO stands for Spectrum Scale *File Placement Optimizer*. It is a feature that you can enable in Spectrum Scale at the time of the creation of a file system, and allows the file system to store and use data locality properties. For more information about how Spectrum Scale-FPO works, see “Spectrum Scale File Placement Optimizer topology” on page 26.

Big data workloads that are based on the Hadoop framework must use data locality to process efficiently data by not having to transfer large chunks of data around the network. Therefore, a file system that works with data locality is a must for this kind of workload. The Hadoop file system was designed for this task, but it lacks some features that are addressed by Spectrum Scale-FPO:

- POSIX compliance
- A redundant file system architecture for metadata processing
- A general use file system

Spectrum Scale-FPO provides these features and differentiates itself from the HDFS, therefore making it a good choice for running big data workloads. InfoSphere BigInsights can work with Spectrum Scale-FPO and has an integrated installer that handles the installation and configuration of Spectrum Scale-FPO.

High availability of file system metadata

When InfoSphere BigInsights is configured to use Spectrum Scale-FPO instead of HDFS, you remove the need for a NameNode. HDFS requires that a NameNode stores the metadata of where all the data is physically on the disks, but with Spectrum Scale-FPO this metadata is distributed across multiple nodes in the cluster. This arrangement removes the single point of failure from traditional architectures that use the HDFS.

Note: InfoSphere BigInsights has a built-in failover option for the NameNode when using HDFS.

POSIX compliant

Spectrum Scale is a POSIX-compliant file system, which means that it can talk to the operating system kernel and vice versa through native APIs as through it were an ordinary file system from the operating system kernel point of view. So, you do not need to load and unload your data into your Hadoop-based applications for processing and checking the results, which is a situation that forces you to duplicate data and spend time copying the data in and the results out of a non-POSIX compliant file system.

You can perform ordinary file system operations from within your operating system shell. Commands such as copy, move, delete, or data visualization can be started natively as you can do in Linux or AIX.

Because of its POSIX behavior, Spectrum Scale can be used as a general-purpose file system. You can choose to use it for big data processing, data sharing across multiple nodes, better performance because of its parallelism, or data high availability. Whatever your need is, Spectrum Scale can meet it natively.

Highly scalable, enterprise-ready file system

Spectrum Scale was designed with high availability (HA) in mind. It can provide HA across every layer with which it works: metadata handling and storing, access to disks, data replication if your storage arrays cannot provide it (which is useful when working with local disks on a JBOD-like approach), and file system cluster management.

Spectrum Scale is a cluster that handles data, which means that all of the complexity of organizing and performing cluster management is done by Spectrum Scale. You do not need to know what goes on at that level; you simply use it. Spectrum Scale nodes must perform multiple tasks to provide a file system cluster arrangement, such as cluster management, file system management, metadata management, token management[1], and I/O management. The good news is that all nodes can take on those roles, so a node failure should never bring down the file system itself, unless the cluster was poorly designed and cannot provide HA.

A Spectrum Scale key feature is that it can access data in parallel, therefore increasing the amount of data that it can process. This task is accomplished by having data that is spread over all the available disks in a file system, and by working with a data locking mechanism that works at the file system block level as opposed to a file level. This setup provides parallelism that results in an enhanced I/O throughput.

Moreover, Spectrum Scale has features that provide you with standard enterprise required features that you need in your environment:

- Snapshots.
- Backup and restoring of data.
- Data replication (up to three copies). This feature can be used for data protection or, in a Spectrum Scale-FPO environment, increase data locality alternatives by allowing more than one node to be eligible to run a job.
- Asynchronous caching.
- Cluster security mechanisms for file system management and operations.
- Information lifecycle management (ILM).
- Storage pools implementations.

All of the features that are mentioned in this section makes Spectrum Scale a candidate for a general use file system. Coupled with the data locality feature that is provided by FPO, it is a strong candidate for use with big data workloads.

For more information about IBM Spectrum Scale, see *Implementing the IBM General Parallel File System (GPFS) in a Cross Platform Environment*, SG24-7844 and *IBM Spectrum Scale (formerly GPFS)*, SG24-8254.

1.5 IBM Platform Cluster Manager

Managing many systems is a time-consuming, error-prone, and tedious task when done manually. In the past, before the era of virtualization, systems management did not take up much time from IT personnel because of the small number of systems that they managed for a solution. With the advent of virtualization, the increasing processing power of servers, and the requirements to process large amounts of data, it is inevitable that you face a situation in which you must manage a large server farm within your company.

[1] A token is the IBM Spectrum Scale concept of a data locking mechanism.

IBM Platform Cluster Manager is a cluster management software that can perform bare-metal or virtualized[2] systems deployment, and also can create cluster configurations on the deployed systems. Imagine that you want to deploy multiple, independent InfoSphere BigInsights clusters. If you did that manually, not only do you have much work, but both installations might end up being slightly different because of the lack of automation. Also, imagine that you had multiple InfoSphere BigInsights clusters, but some of them had peaks while others were idler in a certain period. Would you like to be able to reassign some compute nodes from one InfoSphere BigInsights cluster to another based on your business needs and make better use of your computation power? With Platform Cluster Manager, you can.

Platform Cluster Manager can be used to manage POWER servers hardware, install a particular operating system image onto them, and use a few of these servers to create a InfoSphere BigInsights cluster. This is what the authors have done in this book, and we share our experience with you. At the time of writing, there were a few IBM Power Systems 7R2 servers available for use, and we used them in a bare-metal fashion with no I/O virtualization because this would most probably be what customers would do in the field if they opted to manage many servers with Platform Cluster Manager.

Here is a list of other Platform Cluster Manager advantages that you can leverage when you use it to manage your clusters:

- Elimination of cluster silos: You can grow and shrink your clusters on demand as you want, or through on cluster load monitoring.
- Multi-tenancy: Clusters are independent from one another, and isolated.
- Automation: You can create customization rules for operating system deployment and cluster software installation, such as for InfoSphere BigInsights.
- Self-service: After cluster creation rules are published, any authorized user can create, grow, or shrink a cluster based on the provided cluster template rules and resource utilization quotas.
- Support for multiple cluster software: In addition to InfoSphere BigInsights, you can use Platform Cluster Manager to manage other clusters, such as IBM Platform Symphony or IBM Platform LSF® clusters. You can use Platform Cluster Manager to manage any cluster configuration by scripting the cluster software installation, adding and removing nodes from a cluster, and performing cluster software updates.
- There is support for bare-metal or hypervisor-based (virtualization) deployments.

[2] Supported hypervisors only, such as PowerKVM and PowerVM on IBM Power Systems servers.

2

Reference architecture

This chapter provides information about some of the possible reference architectures that you might use for deploying an IBM InfoSphere BigInsights cluster. This book refers to architectures on top of IBM Power System servers.

Both the hardware architecture and the software components that are involved from the point of view of deploying and managing an InfoSphere BigInsights cluster with IBM cluster management software are described here.

This chapter illustrates how to integrate the following components to build a robust, easy-to-manage, and performant InfoSphere BigInsights environment:

- IBM Power Systems servers: Virtualization and bare metal approaches
- Data storage alternatives: Internal and external
- Linux
- The InfoSphere BigInsights software stack: IBM Spectrum Scale, IBM Platform Symphony, Hadoop, and big data workloads
- IBM Platform Cluster Manager - Advanced Edition

You can create the environments by using all or part of the architecture, according to your goals, environment size, and so on.

In addition to these topics, this chapter describes the following topics:

- The big data environment reference architecture
- Hardware architecture for InfoSphere BigInsights clusters on POWER
- Software architecture for InfoSphere BigInsights clusters

2.1 The big data environment reference architecture

Making the most of your big data environment requires that you include data from many different sources. Data sources vary by source and structure. Some are text-based documents, such as logs from web applications and data feeds from social-networking applications, some are spreadsheets, and others can include transactional data from relational databases and data warehouses. Integration of these data sources forms the foundation of federated analytics. Data in these various forms tends to be large and continues to grow at an unprecedented rate. The speed at which data is analyzed and turned into business intelligence is critical to enterprises.

Figure 2-1 shows the components of a big data environment. This environment includes core Hadoop open source components (orange) along with IBM products (blue) that seamlessly extend or replace Hadoop components to make the solution more enterprise-ready.

The components that are applicable to your deployment depend on the data sources that you plan to integrate. Consider your current plans and future needs when making decisions for initial deployment so that the infrastructure can easily grow to support your business as it changes.

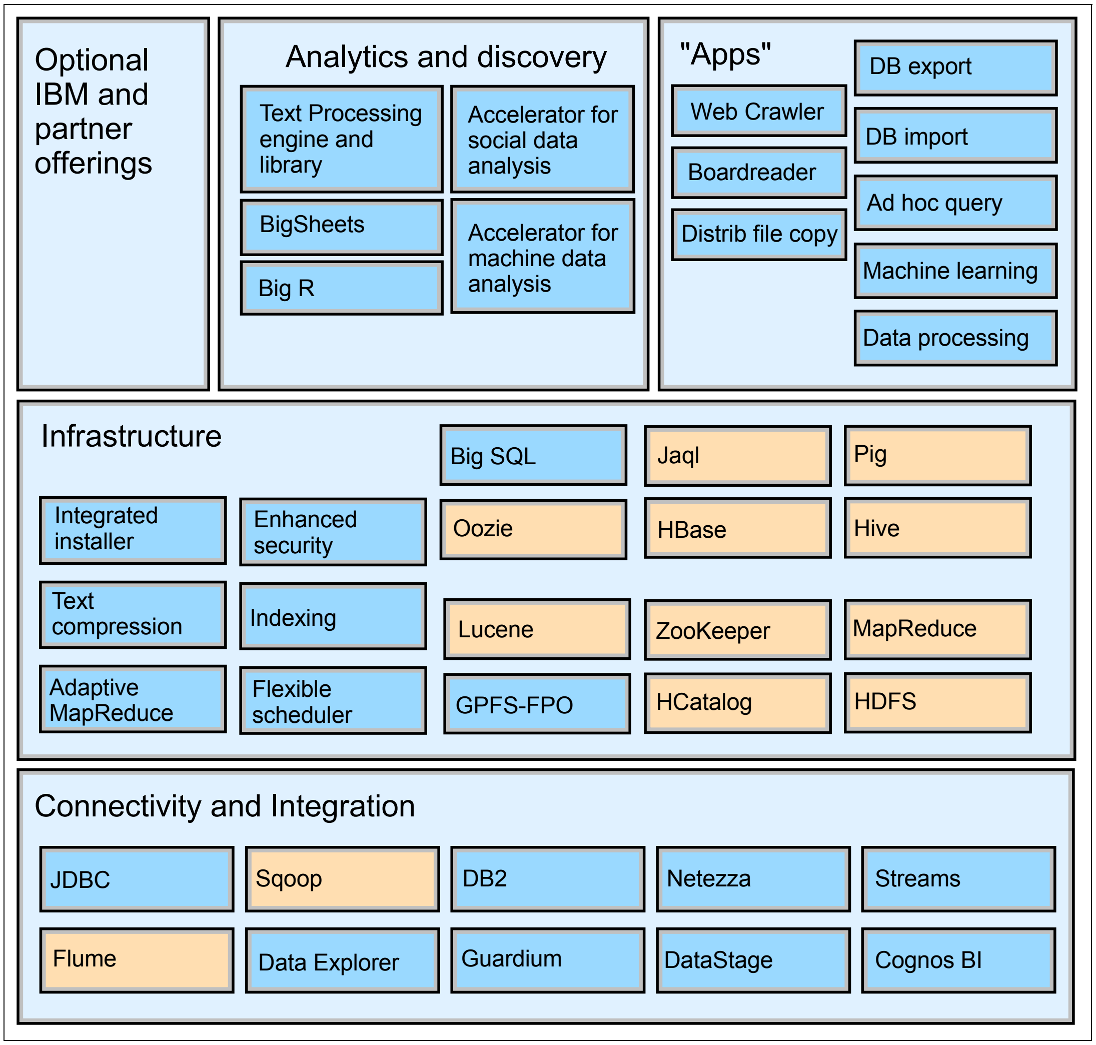


Figure 2-1 The components of a big data environment

2.2 Hardware architecture for InfoSphere BigInsights clusters on POWER

This section provides you with insights about IBM Power Systems hardware architectures that you might select to run your InfoSphere BigInsights environment on. This book shows that there are multiple alternatives that are available, which depend on your environment size and your budget.

One aspect to consider is which hardware architecture is correct for your environment. The answer to this question depends on whether you own any POWER hardware, what kind of hardware it is, how and whether you should virtualize it, and so on. Consider the following questions:

- Do you use existing servers or new servers?
- Do you use high-end Power Systems servers or entry-level ones?
- Do you use Linux-only Power Systems servers or general Power Systems servers?
- Do you use virtualization or bare-metal servers?
- Do you deploy on a cloud environment or a Technical Computing Clouds environment?

Existing hardware or new hardware

Using new or existing hardware might sound like a question that relates only to budgetary concerns. However, consider that a balanced cluster environment for deploying big data solutions tends to be symmetric in terms of hardware configuration.

If you analyze multiple vendors' solutions for big data, especially the so called *black-box* solutions, each specific part of the cluster hardware in the solutions have the same role to achieve symmetry. A symmetric solution has consistent performance results because of correct load balancing, easier maintenance, and high availability. You can build a big data cluster by using any piece of hardware that you want, and aggregating any kind of hardware that you want. The hardware can be from different vendors, have different processor types, or have multiple disks subsystems with different I/O rates. The Hadoop framework works in this manner. However, if you have the opportunity to choose your environment architecture, consider the overall performance gains that a symmetric design offers you.

If you think about one of the software pieces that compose a InfoSphere BigInsights solution, for example, Spectrum Scale (formerly GPFS), you conclude that creating a symmetric environment ensures maximum performance because of the way that Spectrum Scale handles parallel I/O onto the disks. If the hardware is unbalanced in terms of number of disks per node, disks size, disks type, or network throughput per node, this asymmetry might create a performance penalty for its parallel algorithms.

So, focus on creating a symmetric hardware architecture for your InfoSphere BigInsights cluster on IBM Power Systems, whether you plan to reuse your existing servers or acquire new ones.

High-end IBM Power Systems servers or entry-level servers

There are a few points to consider when you select a Power Systems model on which to deploy your environment. You can leverage all of the advantages of having a large server, such as systems consolidation, reduced floor space, and energy savings. However, it is likely that you want to virtualize these servers because some big data workloads might not scale up to the high-end server capacity in terms of processor and memory.

Conversely, entry-level servers can be leveraged as bare-metal nodes without the use of virtualization because they have less capacity in terms of resources. Although you can remove the extra work of virtualization management, you lose in terms of consolidation, energy costs, and floor space.

If you are planning on deploying big data services that can be managed within a cloud environment, it makes sense to use high-end servers. You can manage logical partitions (LPARs) on your servers to create resource boundaries so that your application workload can fully leverage what it is given and still ensure consolidation of your hardware. Moreover, you can create a multi-tenant environment in the sense that cloud software, such as the IBM Cloud Manager with Openstack, handles cloud management and multi-tenancy for you. However, achieving symmetry might be more challenging in this case.

Entry-level servers can be managed in a more Platform Computing way, where you have a clearer distinction between nodes roles, such as management nodes, and computing or data nodes. There is no physical hardware overlap among these roles. With this architecture type, it is easier to achieve symmetry of the hardware architecture. Also, software such as Platform Cluster Manager - Advanced Edition can be leveraged to automate the deployment of a InfoSphere BigInsights cluster on top of the nodes.

The last scenario is a predefined black-box solution, such as the POSH architecture for InfoSphere BigInsights on Power Systems, as described in 2.2.2, “The POSHv2 architecture” on page 21. This is a built-to-the-task hardware architecture that IBM has available for running InfoSphere BigInsights on Power Systems. If your goal is performance and seamless integration of software and hardware, this is a good option for you.

In addition to the POSHv2 reference architecture, the more recent IBM Data Engine for Analytics can also be leveraged as a hardware architecture for running InfoSphere BigInsights on Power Systems.

Linux-only Power Systems servers or general Power Systems servers

Since the launch of the POWER7 hardware, IBM has made available some server models that run only Linux (as opposed to also running AIX and IBM System i®). IBM has the following Power Systems server models that run only Linux:

- Power S812L (POWER8)
- Power S822L (POWER8)
- PowerLinux 7R1 (POWER7)
- PowerLinux 7R2 (POWER7)
- PowerLinux 7R4 (POWER7)

The fundamental advantage of this hardware is that because it runs only Linux, it is cheaper than the genera-purpose Power Systems server models. If all you want to run are Linux-based workloads, such as InfoSphere BigInsights, then this is price-competitive hardware, even compared to other hardware vendor’s x86-based servers.

These server models fit more into the entry- or mid-end servers than high-end ones. So, they might benefit more from the bare-metal approach than the virtualized one. However, if you must virtualize them, you can. The POWER8 models also benefit from PowerKVM virtualization. The S812L or S822L combined with PowerKVM-based virtualization provides the preferred price-performance in terms of virtualizing a Power Systems server for Linux environments only.

Using virtualization or using bare metal servers

This consideration depends on your goals. Here are some important points to consider:

- Solution symmetry
- The automated deployment tools that are available at your disposal
- The size of your servers (entry-level or high-end).

You might want to consider a mixed architecture in terms of virtualization. For example, you might want to virtualize two or three management nodes with two LPARs each, which host management services for InfoSphere BigInsights, with all of the data nodes used as bare-metal nodes. This configuration better uses the resources on the management nodes while also ensuring a proper high availability and better consolidation levels for those nodes.

The next sections outline a few hardware architectures that can be used to deploy a InfoSphere BigInsights environment on Power Systems servers. This list is not restrictive, but you should have the aspects that have been described in this chapter when deciding on a hardware architecture for your use.

2.2.1 General architecture

An InfoSphere BigInsights cluster is composed of management nodes and data nodes. Management nodes host services such as the InfoSphere BigInsights console, Oozie, BigSQL, catalog, ZooKeeper, HBase, Hive, and Platform Symphony. Data nodes are the ones that perform work based on the workloads that are running on the cluster.

The nodes are interconnected through Internet Protocol networks. Therefore, a well-designed network architecture is important and plays a central role in cluster performance.

As a preferred practice for a InfoSphere BigInsights cluster, define at least three networks:

- A data network
- A public network
- A user administrative network

Another aspect of an InfoSphere BigInsights hardware architecture is where the data is stored. Because of the way that the MapReduce framework works, jobs are scheduled on nodes where data is found locally to minimize network transfers of massive amounts of data. Therefore, the typical architecture uses disks that are assigned to only a single node. This is called a *shared-nothing* architecture. File systems must be aware of this architecture to achieve the goals of MapReduce workloads. The Hadoop file system does that, and so does Spectrum Scale through its File Placement Optimizer feature. This book focuses on using Spectrum Scale as the file system to build the architecture design. For more information about possible disk layouts, see “Disks layout” on page 20.

Taking a step further into the architecture, you can add nodes that ease the management of hardware in the InfoSphere BigInsights cluster. Imagine that you want to either add or remove a node from an existing InfoSphere BigInsights cluster, or you want to create multiple, independent InfoSphere BigInsights clusters. Doing so manually is a time-consuming task. Cluster management software eases this task.

This book uses IBM Platform Cluster Manager - Advanced Edition to provision your InfoSphere BigInsights nodes. Platform Cluster Manager - Advanced Edition can perform bare-metal provisioning and apply cluster templates during this process so that you can conveniently deploy a whole cluster, along with management and data nodes, in an automated fashion.

You can add a system management node to your overall InfoSphere BigInsights server farm to install Platform Cluster Manager - Advanced Edition for managing your hardware. Doing so changes the network layout of your network environment some because you must have Platform Cluster Manager - Advanced Edition communicated with the FSP ports of your Power Systems hardware. Also, Platform Cluster Manager uses a provisioning network to do systems deployment. You might choose to use your administrative network for doing so, or use a fourth network to isolate the traffic for systems provisioning. A detailed explanation of these network pieces is available in “Networking” on page 18.

Networking

InfoSphere BigInsights uses three networks: administrative, public, and data.

The *administrative network* is used for accessing the nodes to perform administrative tasks, such as verify logs, start or stop services, and perform maintenance. Administrators use it to SSH into the nodes, or access them through virtual network computing (VNC). This network can be as simple as a 1-Gb Ethernet port that might or might not have high availability through the form of Ethernet bonding.

Based on your environment requirements, the administration network can be segregated onto separate VLANs or subnets, and is directly connected to your company’s administrative network through a firewall to prevent non-IT management personnel from reaching the IT servers.

The *public network* is the gateway to the applications and services that are provided by the InfoSphere BigInsights cluster itself. It can be though of as the public face of your corporate network. It is the network that you use to access the InfoSphere BigInsights web portal and perform your big data work. Although all cluster nodes can be connected to this network, the management nodes are the only ones with configurable services, such as an HTTP server service, on it. The reason why you connect all of your nodes to the public network is to prevent cabling rework if you are working with a dynamic environment, for example, when managing multiple clusters through Platform Cluster Manager - Advanced Edition.

The *data network* is a private, fast interconnect network for the cluster nodes that is used to move data among nodes, and move data in to or out of the Hadoop file system for processing. It can be built with 10-Gb Ethernet adapters, InfiniBand adapters, or any other technology that provides high throughput and low latency network data transfers.

An InfoSphere BigInsights cluster can connect to the corporate data network by using one or more *edge nodes*. These edge nodes provide a layer between your InfoSphere BigInsights cluster and your data network. You can use these nodes to import data into your cluster. They can be other Power Systems servers running Linux, or any other server type at all. If you think of a large InfoSphere BigInsights cluster, each rack can have an edge node, although this is not mandatory.

Figure 2-2 on page 19 shows how the networking architecture looks like in an environment that follows the guidelines in this chapter.

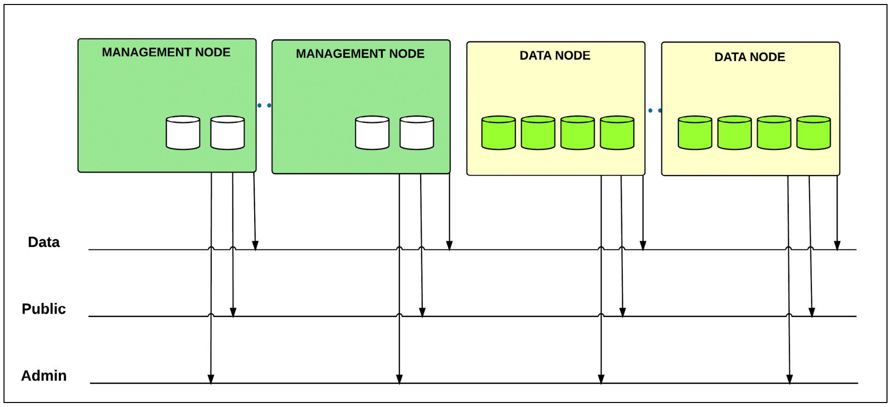


Figure 2-2 High-level InfoSphere BigInsights cluster architecture: - nodes, networks, and disks

Notice that you can deploy the general architecture that is described in this section in any kind of environment, whether that environment is bare-metal nodes, LPARs onto larger servers, or even a cloud environment.

If you plan to use Platform Cluster Manager - Advanced Edition to perform cluster management, then add two more networks to the hardware architecture: The provisioning and FSP networks.

The *FSP* network connects Platform Cluster Manager - Advanced Edition to the FSP port of your Power Systems hardware in the same fashion as an HMC is connected to those ports. In fact, the Power Systems hardware has two FSP ports through which it communicates with the external world for hardware management. If you are using an HMC to manage your hardware, it uses the primary HMC port on the server. So, in this case, you can use the secondary HMC port to allow Platform Cluster Manager to manage the hardware as well.

The *provisioning network* is used by Platform Cluster Manager - Advanced Edition to transfer operating system installation images onto the hardware, and perform preinstallation and postinstallation scripts for deployment customization.

Figure 2-3 shows a complete network architecture of an InfoSphere BigInsights environment that is managed by Platform Cluster Manager - Advanced Edition. This scenario implements a provisioning network that is different from the administrative network. Because of the low traffic of an administrative network, and because provisioning traffic happens at particular points only, these two networks can be the same one.

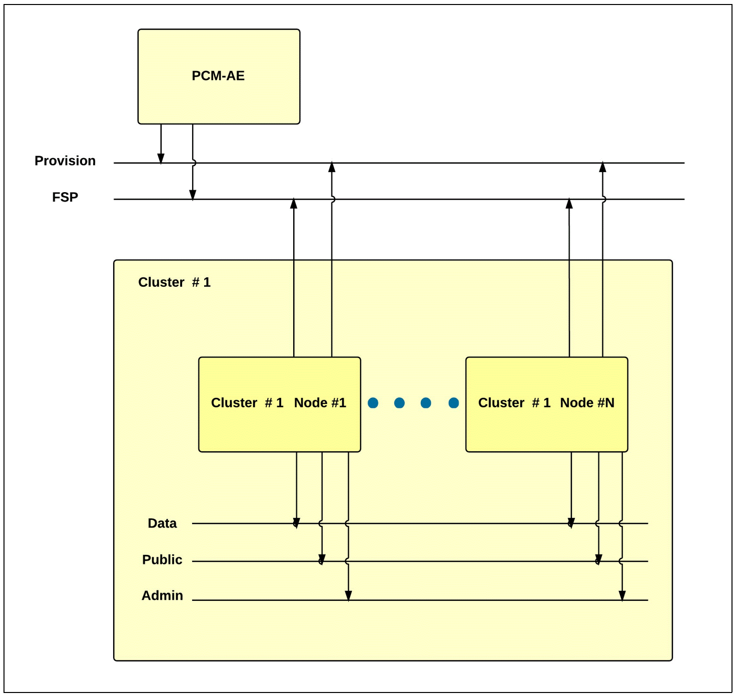


Figure 2-3 An InfoSphere BigInsights cluster under provisioning by Platform Cluster Manager - Advanced Edition - network diagram

Disks layout

The simplest shared-nothing disk layout that can be used with MapReduce workloads is using internal disks in the cluster nodes. It is usually the cheapest scenario. However, you can still achieve a shared-nothing environment by using disks that are outside of the machine, either on storage expansions or storage devices.

Scenarios that work with an external storage device can leverage high availability in terms of disk access while also ensuring a shared-nothing layout. This task is accomplished by assigning the disks to two nodes simultaneously and having Spectrum Scale assign primary and secondary Network Shared Drive (NSD) servers in an alternated fashion. If the primary disk server node for a storage disk fails, the secondary node can still serve the disk, and serves only the disk if the primary node fails. This task is controlled by a Spectrum Scale FPO failure groups configuration.

One point to consider is that Hadoop-based technologies process large amounts of data that are found locally on a server to reduce I/O transfers over the network and leverage fast I/O. Suppose that your environment is composed of 100 nodes, each with access to 10 disks, for a total of 1000 disks. Can a single SAN unit and its SAN topology provide enough bandwidth to feed I/O to all of these 1000 disks with a performance that is as good as though each of the 100 nodes were accessing 10 internal disks each? For Hadoop workloads, using a SAN architecture without properly considering I/O performance is not suggested.

2.2.2 The POSHv2 architecture

To provide a high-performing hardware architecture for Hadoop workloads that are based on Power Systems servers, IBM defined the POSHv2 architecture. This architecture is flexible in terms of disk layout, and is available in various building block sizes, as described in “POSHv2 building blocks” on page 22. So, you design your own InfoSphere BigInsights cluster hardware architecture based on a given POSHv2 POD building block size. POSHv2 even provides some predefined configurations, such as for low-cost proof-of-concept environments, general-purpose landing zones or data lakes, or powerful NoSQL or complex analytics environments.

The POSHv2 hardware architecture is composed of the following components:

- An HMC
- Management nodes
- Data nodes
- Edge nodes
- A cluster management node

The POSHv2 architecture was designed with one goal in mind: provide customer value. It provides customers with the following capabilities:

- Flexibility and extensibility.

  There is no single one architecture size. POSHv2 can meet your sizing needs, whether you plan to build a small or large Hadoop environment. Also, if you choose to start small, you can either grow your environment by extending vertically (by using bigger POSHv2 building blocks) or horizontally (by adding more of your building block size).

- Best-in-class hardware.

  The IBM Power Systems hardware has been in the market for many years, and is known for its performance, high availability through component redundancy, robustness, and reliability. Moreover, and especially for Hadoop workloads, the fact that there are Power Systems server models that are targeted to run only Linux makes it a compelling choice over other x86-based server models. In essence, you have all of the Power Systems servers advantages at prices that compete with x86 servers.

- Dense storage subsystem.

  Either by choosing to use large internal disks, attaching an expansion I/O drawer, or connecting to a DS3700 storage unit, the POSHv2 architecture configurations can provide you with as much disk space as you need for your Hadoop workload.

- Scalability.

  Choose to use whether the S812L or S822L server models, which scale up to 12 or 24 cores per node respectively, and up to 512 GB or1024 GB of memory respectively.

- Multi-tenancy.

  The architecture supports multitenancy, which is achieved by using a cluster management node that runs Platform Cluster Manager - Advanced Edition. You can create a consistent and easy-to-grow environment by using the POSHv2 building blocks and deploy multiple, independent Hadoop environments within your server farm.

- Uses modern technology running on top of the hardware.

  All of this powerful hardware is driven by Linux, Spectrum Scale, and InfoSphere BigInsights.

Figure 2-4 shows a high-level overview of the POSHv2 solution.

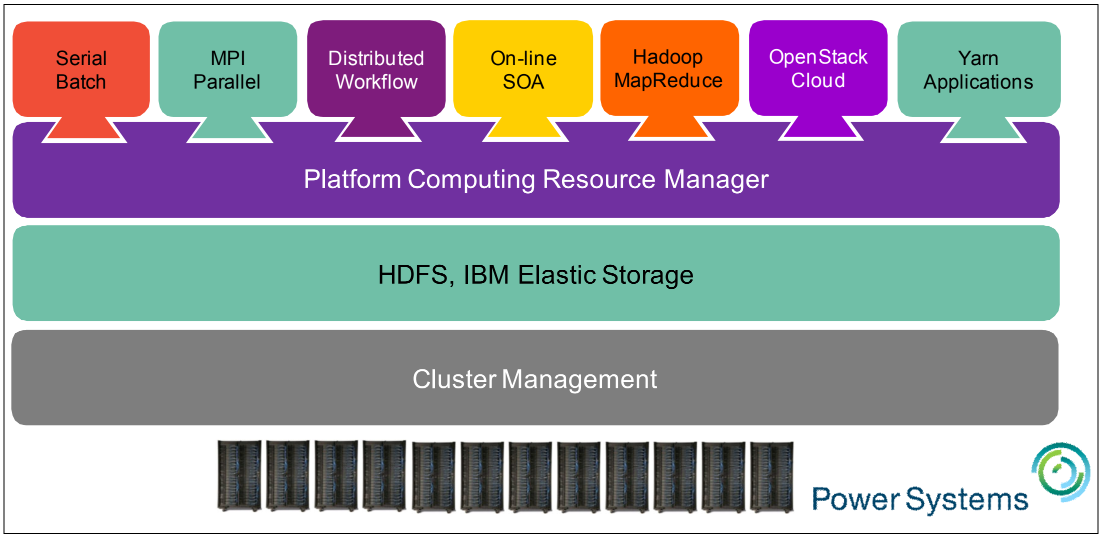


Figure 2-4 The POSHv2 solution for Hadoop workloads

POSHv2 building blocks

You can use the flexibility of the POSHv2 architecture to create customized hardware environments that are based on your sizing needs. The building blocks options are shown in Figure 2-5.

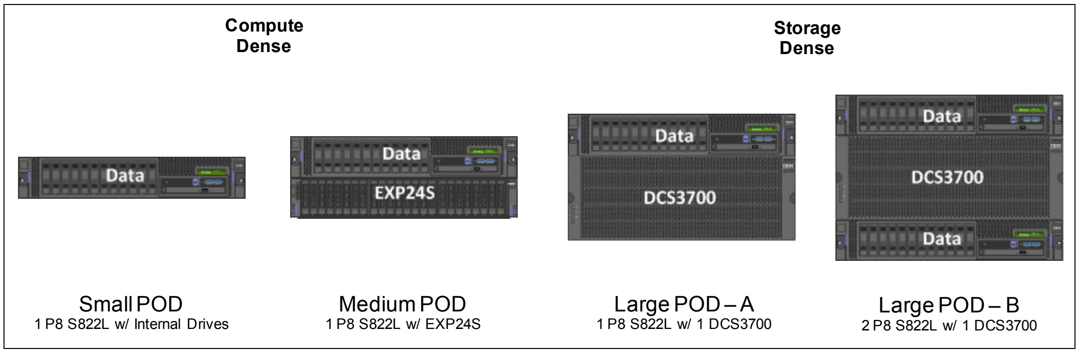


Figure 2-5 POSHv2 building blocks

Here are the available sizes:

- Small POD: One POWER8 S22L with internal drives
- Medium POD: One POWER8 S822L with internal drives and an expansion I/O unit (EXP24S)
- Large POD - A: One POWER8 S822L with internal drives and one DS3700 unit
- Large POD - B: Two POWER8 S822L systems with internal drives and one DS3700 unit

Regarding the large POD - B block, a DS37000 unit is shared between two POWER8 nodes. In an InfoSphere BigInsights with a Spectrum Scale-based configuration, this does not mean that both nodes see and serve I/O onto all of the disks at the same time, but can have half of the disks be primarily served by one node and have the other node as a backup. The other half of the disks is served with alternated primary and backup roles. This layout applies to what was briefly mentioned in “Disks layout” on page 20.

So, when using these building blocks and clustering them together, you can create cluster environments that can easily grow larger in the future. Figure 2-6 shows racks that are built with the Small POD, Medium POD, and Large POD - A building blocks. The HMC is not shown in the figure. You might use an existing one or attach it to your POSHv2 rack solution.

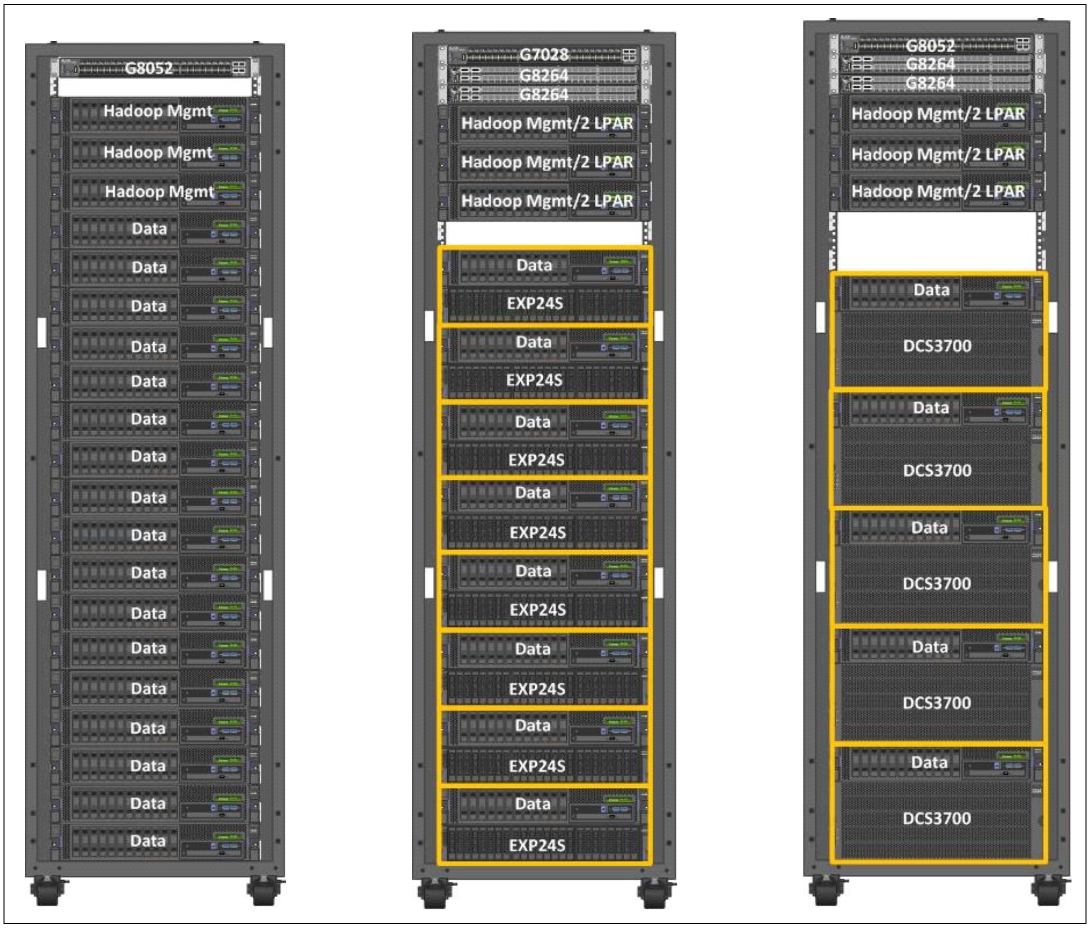


Figure 2-6   POSHv2-built rack solutions

The POSHv2 management nodes layout for InfoSphere BigInsights

The POSHv2 architecture suggests a particular management node configuration for InfoSphere BigInsights to provide high availability to the services that run within them. Basically, each one of the three physical Power Systems servers that are dedicated to running management services, as shown in Figure 2-6 on page 23, are partitioned into two LPARs. So, a total of six LPARs are available to run management services. With that layout in mind, the distribution of services can be done by following the design that is shown in Figure 2-7.

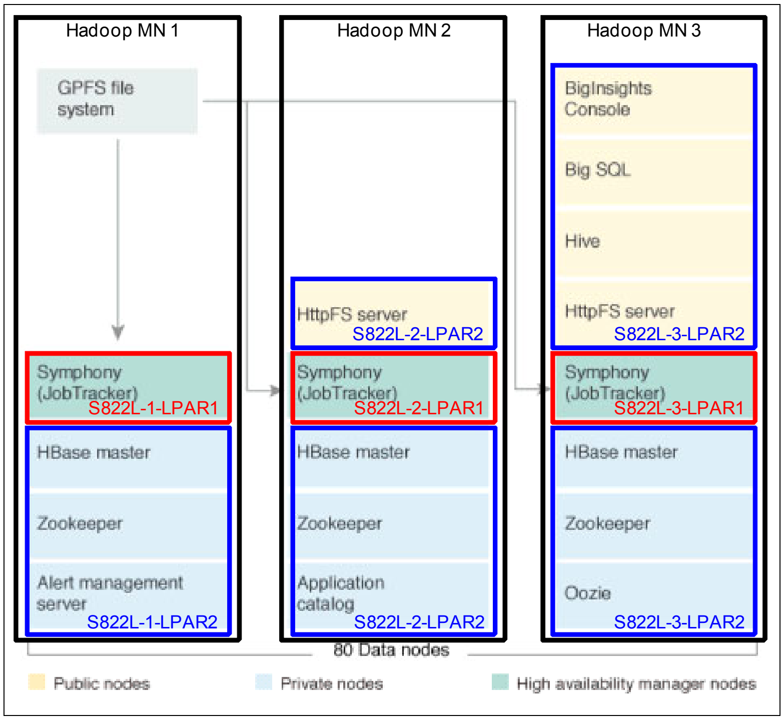


Figure 2-7 Six LPAR management node high availability for InfoSphere BigInsights

A degree of high availability is required when you design your InfoSphere BigInsights cluster to use the adaptive MapReduce scheduler on top of Spectrum Scale as your distributed file system. The Spectrum Scale cluster requires three system pool machines that share disks for the high availability manager component, which provides high availability for adaptive MapReduce. These machines are represented by the red LPARs in Figure 2-7. Also, at a minimum, the ZooKeeper service must have a dedicated local disk to provide high I/O transactions of ZooKeeper snapshots.

The three dedicated nodes for the high availability management components include the Job Tracker and the Spectrum Scale HA_POOL. The HA_POOL is used to store the Job Tracker and task tracker logs. Each of the HA nodes holds one of the three replications of the logs. The JobTracker runs on the first HA node, and the other two nodes run in standby mode. During an HA event, the Job Tracker is migrated to the HA standby node. A second HA event results in the migration of the JobTracker to the third HA standby node.

ZooKeeper is intended to be run as an ensemble. Because it maintains a quorum, there should be an odd number of nodes. POSHv2 specifies three management nodes, which provide a sufficient failover configuration.

The Web Console node acts as the edge node, and includes the HttpFS and BigSQL components, which must be on an edge node.

POSHv2 networking architecture

The POSHv2 network architecture follows the architecture that is described in “Networking” on page 18. The options that are available for the switches are the following models:

- IBM RackSwitch G8052:
  - 48x 1 Gb RJ45 ports, plus four SFP+ uplinks for 1 or 10 Gb connectivity
  - Provides a 1.18-ms latency
- IBM RackSwitch G8264:
  - 48x 1 or 10 Gb SFP ports, plus four QSFP+ uplinks for 10 or 40 Gb connectivity
  - Provides an 880-ns latency
- IBM RackSwitch G8316:
  - 416x QSFP+ ports
  - Provides an 880-ns latency

2.3 Software architecture for InfoSphere BigInsights clusters

The InfoSphere BigInsights solution is composed of multiple underlying software components. You can visualize the InfoSphere BigInsights software architecture as the integration of some software components that leverage the hardware architecture that is described in 2.2, “Hardware architecture for InfoSphere BigInsights clusters on POWER” on page 15 to digest efficiently massive amounts of data. Figure 2-8 shows the software components.

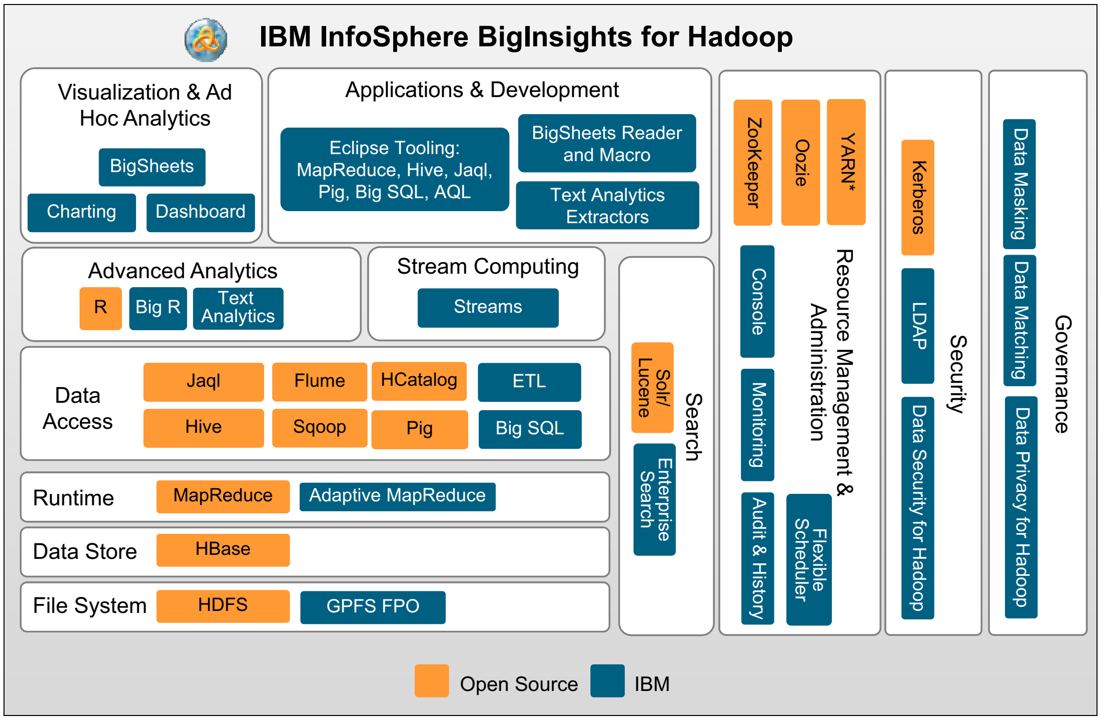


Figure 2-8 InfoSphere BigInsights software components

The solution in Figure 2-8 on page 25 is built on some open source software pieces. IBM, however, has components that can be used to replace some of these open source layers and provide enhanced functions, for example, Spectrum Scale (replacing HDFS) and Platform Symphony for adaptive MapReduce scheduling. In addition, the *Visualization and Ad Hoc Analytics* software components, and the *Application and Development* ones that are shown in Figure 2-8 on page 25, are designed and provided by IBM so that you can gain insights into your large amounts of data and take advantage of it in the business or data processing challenges you face.

These components are the basis for building an InfoSphere BigInsights cluster. This solution can be deployed on the hardware architectures that have been described in 2.2, "Hardware architecture for InfoSphere BigInsights clusters on POWER" on page 15. Likewise, what is shown in Figure 2-8 on page 25 can be considered a general software architecture for an InfoSphere BigInsights cluster.

This software stack runs on PowerLinux. For InfoSphere BigInsights on Power Systems, the only supported version at the time of writing is Red Hat Enterprise Linux 6.5. For more information about operating system requirements for InfoSphere BigInsights on Power Systems, see "Supported Linux distributions and levels" on page 32.

InfoSphere BigInsights and Spectrum Scale: An integrated approach

InfoSphere BigInsights V3.0 has an installation process that integrates the installation of Spectrum Scale version V3.5.0-17. If you do not have a pre-existing Spectrum Scale environment, there is no need to install and configure one manually. In fact, it is easier to let InfoSphere BigInsights performs these tasks for you because the proper file system configuration and tuning is performed by the installer.

If you do have an existing Spectrum Scale environment that you want to use with your new InfoSphere BigInsights environment, or if you want to use a different version of Spectrum Scale than the one that is packaged with the product (if it is supported), then you can opt to install and configure Spectrum Scale manually, and later direct your InfoSphere BigInsights installation to use this existing file system. In this case, you must understand how the Spectrum Scale-FPO cluster topology works. There is an overview in "Spectrum Scale File Placement Optimizer topology" on page 26.

From an architectural point of view, the use of Spectrum Scale does not change anything in terms of the number of cluster nodes. It replaces one file system software for another one, as shown in Figure 2-8 on page 25. The advantages of using Spectrum Scale over native HDFS are described in 1.4, "IBM Spectrum Scale-FPO (formerly GPFS-FPO)" on page 10.

Spectrum Scale File Placement Optimizer topology

This section provides a sample topology to reference when you plan to install and configure Spectrum Scale manually for your InfoSphere BigInsights cluster.

Figure 2-9 on page 27 illustrates a Spectrum Scale File Placement Optimizer (FPO) topology with three racks that each contain four nodes. Two nodes are in the top half and bottom half of each rack. Each rack corresponds to the rack that you specify for each node when you create your InfoSphere BigInsights cluster. All nodes use the same hardware except for the type of disk drive that is used for data storage. In a typical Spectrum Scale FPO cluster, solid-state drives (SSDs) are recommended for metadata, although you can use hard disk drives (HDDs) throughout your cluster.

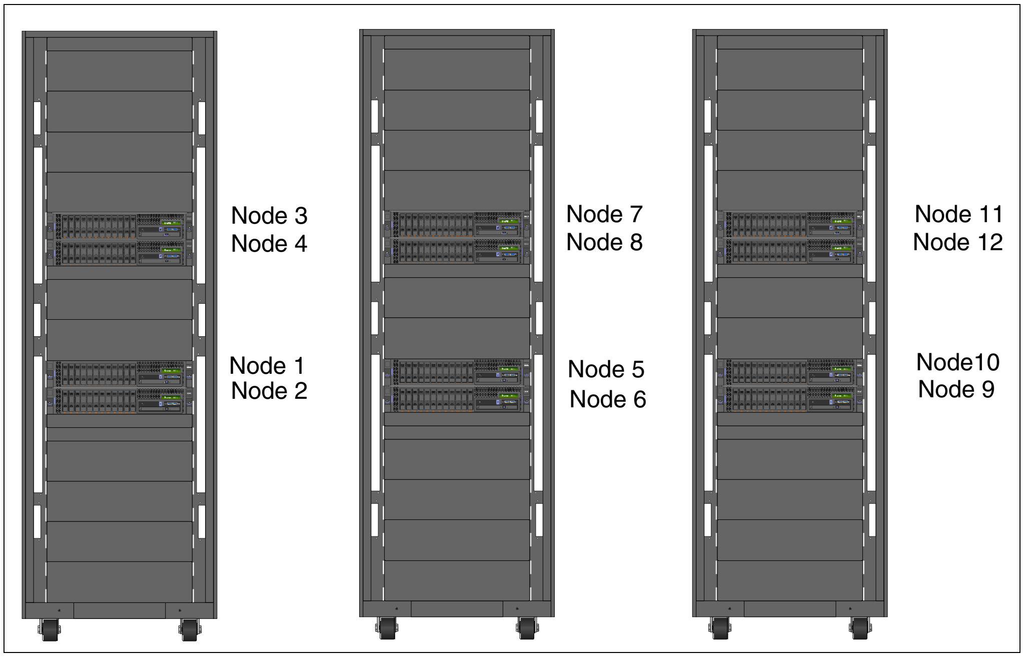


Figure 2-9 The topology of a Spectrum Scale-FPO cluster for use with InfoSphere BigInsights

One node on each rack has disks that are dedicated for metadata. To have three replicas for metadata, a minimum of four failure groups are necessary to maintain the replication level if there is single-node failure. Maintaining the maximum replication factor (three replicas) for metadata at any time is important to ensure the maximum availability of the cluster.

For clusters that can tolerate less-replicated data if there is a single-node failure, three failure groups might be sufficient. For larger clusters, you typically configure more than three failure groups for data.

Figure 2-9 illustrates two failure groups per rack, which combine for six total failure groups for this cluster. Each failure group contains two nodes.

Table 2-1 shows the failure groups for each of the racks in Figure 2-9. The cluster is organized with metadata and data nodes in four different failure groups. The failure groups correspond to the rack and rack position in which they are located.

Table 2-1 Spectrum Scale-FPO failure groups and failure group vectors

Failure group	Rack number	Rack location	Node number	Failure group vector
FG1: (1,0)	1	Bottom	Node 1 Node 2	1,0,1 1,0,2
FG2: (1,1)	1	Top	Node 3 Node 4	1,1,1 1,1,2
FG3: (2,0)	2	Bottom	Node 5 Node 6	2,0,1 2,0,2
FG4: (2,1)	2	Top	Node 7 Node 8	2,1,1 2,1,2

Failure group	Rack number	Rack location	Node number	Failure group vector
FG5: (3,0)	3	Bottom	Node 10 Node 11	3,0,1 3,0,2
FG6: (3,1)	3	Top	Node 12 Node 13	3,1,1 3,1,2

A three-number notation, which is known as the *failure group topology vector*, is used to denote disks that are part of a data storage pool that is configured for the Spectrum Scale FPO. The first number is the rack number: rack 1, rack 2, and rack 3. The second number denotes whether the node is at the bottom (0) or at the top (1) half of the rack. Finally, the third number denotes the position of the node within the specified half of the rack. For example, in Figure 2-9 on page 27, the notation 2,1,1 refers to the second rack, top half, position 1, which refers to node 7. All nodes that use disks in a certain half of a rack belong to the same failure group. For example, the two nodes with disks that use the failure group topology vectors (2,1,0) and (2,1,1), are in the (2,1) failure group. In Figure 2-9 on page 27, the failure groups for the data disks are: (1,0), (1,1), (2,0), (2,1), (3,0), (3,1).

InfoSphere BigInsights and Platform Symphony: An integrated approach

By default, InfoSphere BigInsights installs and implements the open source Apache MapReduce runtime scheduler during its installation. You might, however, opt to replace this scheduler with IBM Platform Symphony Adaptive MapReduce. The advantages of doing so are described in 1.3, “IBM Platform Symphony” on page 9.

In terms of architecture, using the Platform Symphony Adaptive MapReduce runtime scheduler from Platform Symphony might require a change in your architecture: Your cluster must have at least three nodes to ensure high availability of this component. From a software architecture overview, it fills the *runtime scheduler* layer of the overall InfoSphere BigInsights architect, as shown in Figure 2-8 on page 25.

Hadoop components overview

The Hadoop project is composed of three parts: HDFS, the Hadoop MapReduce model, and Hadoop Common. The goal of Hadoop is to use commodity servers in a large cluster to provide high I/O performance with minimal cost.

Because the scenarios in this book are not going to be using HDFS (because of the Spectrum Scale-FPO functions that are part of the implementation), this book goes in to more detail about Spectrum Scale-FPO. However, you can also use HDFS as an alternative to Spectrum Scale. Implementing HDFS or Spectrum Scale has other effects on the environment, as the two file systems are different, so your installation process and your backup/recovery process must be redesigned depending on which component you use.

MapReduce

MapReduce is the heart of Hadoop. The term MapReduce refers to two separate distinct tasks that Hadoop programs perform. The first is the map function, which takes a set of data and processes it to capture interesting aspects of the data.

The reduce function takes the output from map jobs and combines it into a larger set of output to provide insights into the data. In a Hadoop cluster, jobs are managed by a component that is called the JobTracker, which communicates with the name node to create multiple map-and-reduce tasks based on data locality. The tasks are managed by TaskTracker agents that run continually on the cluster nodes and report back the status to the JobTracker.

Hadoop Common Components are a set of libraries that support various Hadoop subprojects. These components are intended to simplify the complex task of creating MapReduce applications. Because HDFS is not POSIX-compliant, which means that common Linux or UNIX tools and APIs cannot be used, it is necessary to use the **`hdfs dfs <args>`** (or in Hadoop Version1.x, **`Hadoop dfs <args>`**) file system shell command interface and Hadoop file system APIs. Flume is another component that is used for importing data into HDFS. The MapReduce application development environment includes several components to reduce the time and skill that is needed for application development. These components include Hive, Pig, Oozie, ZooKeeper, HBase, and Avro as examples.

2.3.1 Cluster management software

Section 2.2, "Hardware architecture for InfoSphere BigInsights clusters on POWER" on page 15 presented a general hardware architecture for a pure InfoSphere BigInsights cluster and later added nodes to provide cluster management through Platform Cluster Manager - Advanced Edition. This section describes the advantages of such a solution and its overall software architecture.

Using Platform Cluster Manager - Advanced Edition to manage an environment of clusters provides the following advantages:

- Management of multi-tenancy environments

  You can create multiple, isolated clusters within your server farm.

- Support for deploying multiple products

  This book describes a scenario that implements Platform Cluster Manager - Advanced Edition to deploy an InfoSphere BigInsights cluster, but this book might also be used to automate deployment of other solutions, such as ones based on IBM Symphony, IBM LSF, GridEngine, PBS Pro, and open source Hadoop.

- On-demand and self-service provisioning

  You can create clusters definitions and use them to deploy automatically the cluster nodes. A person with little or no cluster setup knowledge can then quickly deploy a cluster environment.

- Increased server consolidation

  By being able to grow or shrink dynamically a cluster environment, you minimize the amount of idle resources because of the creation of siloed clusters.

Figure 2-10 shows how the software pieces integrate into a Platform Cluster Manager - Advanced Edition managed cluster for InfoSphere BigInsights.

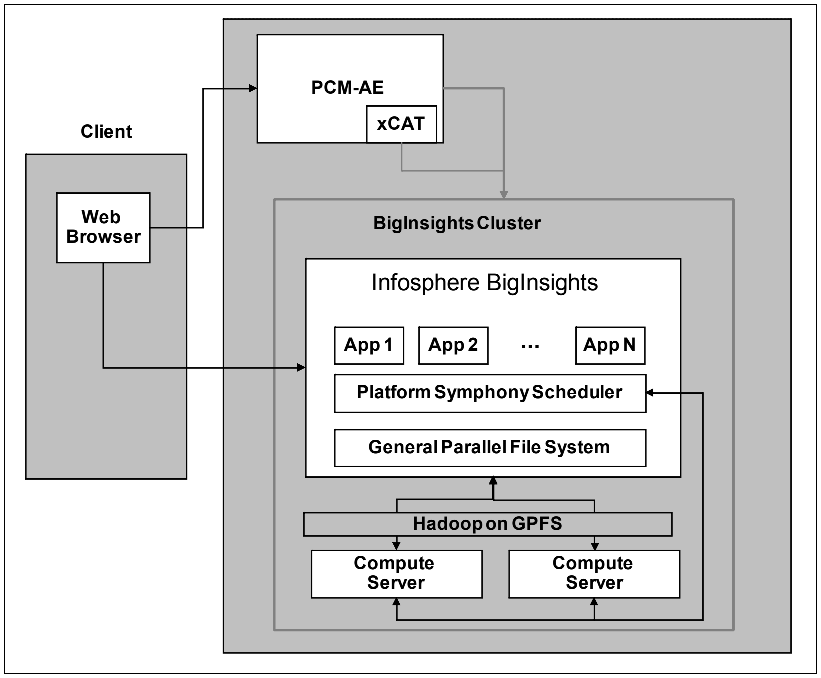


Figure 2-10 Platform Cluster Manager - Advanced Edition managed cluster for InfoSphere BigInsights

One of the parts that integrate the software architecture of a Platform Cluster Manager - Advanced Edition based solution is xCAT. Past versions of Platform Cluster Manager - Advanced Edition integrated xCAT into the whole solution as an add-on, external software component. With Platform Cluster Manager - Advanced Edition Version 4.2, xCAT comes integrated within the solution.

During Platform Cluster Manager - Advanced Edition Version 4.2 installation, xCAT also is installed on the node. As described in 3.3, "Platform Cluster Manager - Advanced Edition" on page 73, some of the xCAT configuration is automatically performed during this step. You can, however, leverage xCAT commands to further configure or reconfigure your cluster provisioning environment. As of today, some features still must be configured through xCAT:

- Establishing a connection to the server's FSP port for hardware operations management
- Creating a hardware profile with the correct serial connection settings for connecting to the LPAR console.
- Optionally, using the Platform Cluster Manager environment as a DHCP server to the FSP hardware management network.

All of the setup details that you must go through to create a working Platform Cluster Manager - Advanced Edition environment are described in 3.3.2, "Performing the installation" on page 77.

If you are running a version of Platform Cluster Manager - Advanced Edition that is older than Version 4.2, you must install and configure an xCAT environment separately.

3

Installation

Now that you understand the pieces of hardware and software that are used to build an IBM InfoSphere BigInsights environment on Linux, it is time to learn how to install each piece of it.

This chapter is organized into two conceptual approaches:

- Deployment for small environments
- Deployment for large environments by using IBM Platform Cluster Manager - Advanced Edition

Read the first part of this chapter if your goal is to deploy quickly and simply a single InfoSphere BigInsights cluster. Read the second part if you plan to manage multiple InfoSphere BigInsights clusters within your server farm.

This chapter covers the following topics:

- Linux on Power Systems
- InfoSphere BigInsights
- Platform Cluster Manager - Advanced Edition

3.1 Linux on Power Systems

There are a few ways to install an operating system onto a server or logical partition (LPAR) on Power Systems servers. As you are aware, these alternatives mainly depend on the operating system (OS) that you intend to install because these alternatives are provided by the OS.

Linux can be installed on IBM Power Systems servers by using DVD media, or it can be installed over the network if you set up a second environment to serve the content over the network. The latter approach is called a *network installation*.

The methodology that you choose to install your environment is up to you. If you are planning to create a small InfoSphere BigInsights environment with the minimum recommended number of nodes, you might decide to install the operating system from DVD media. If you plan to install a larger environment, you might consider performing a first installation and then cloning this first system to the other nodes over network installations. If you are planning to create a large, flexible, or multi-cluster environment, consider using Platform Cluster Manager - Advanced Edition as the cluster management tool.

The section “Installing Linux by using the IBM Installation Toolkit” on page 33 is intended for users who are not familiar installing Linux on Power Systems servers. This method is the easiest way to perform a single installation, and that section briefly covers it because the scope of this book is not the Linux installation. Although it is the simplest method to install Linux on Power Systems, use the methodology that you are most familiar with while following the prerequisites to install InfoSphere BigInsights, which are described in 3.1.2, “Operating system prerequisites setup for InfoSphere BigInsights” on page 39.

3.1.1 Operating system installation

This section outlines planning aspects about which Linux distribution to use and how to install it in your Power Systems server environment.

Supported Linux distributions and levels

At the time of writing, InfoSphere BigInsights is supported only on Red Hat Enterprise Linux (RHEL) 6 update 3 and above on IBM POWER7 processor-based servers. On POWER8, the minimum required operating system level is RHEL 6 update 5. In any of these scenarios, only Big Endian architecture is supported, as summarized in Table 3-1.

Table 3-1 Supported Linux versions for running InfoSphere BigInsights on Power Systems

Hardware architecture	Supported versions	Endianness
POWER7	RHEL 6 update 3 RHEL 6 update 4 RHEL 6 update 5	Big Endian Big Endian Big Endian
POWER8	RHEL 6 update 5	Big Endian

For an updated list of supported hardware and Linux distributions, see the InfoSphere BigInsights website in the IBM Knowledge Center:

http://ibm.co/1DlnmVy

Installing Linux by using the IBM Installation Toolkit

If you are not familiar with Linux installations in general or not familiar with doing so on Power Systems servers, the IBM Installation Toolkit is publicly available to help you.

The IBM Installation Toolkit eases the installation of Linux on Power Systems by providing simplified installation wizards that can install any of the two supported Linux distributions on POWER: Red Hat Enterprise Linux and SUSE Linux Enterprise Server. Also, the IBM Installation Toolkit easily delivers IBM add-on packages to integrate fully a Linux environment with POWER hardware. Such add-on packages are IBM Java, RSCT packages for Dynamic Logical Partition Operations, IBM Performance Manager Monitor, and so on. For a comprehensive list of the product features, add-on packages, users guide, and so on, go to the following website:

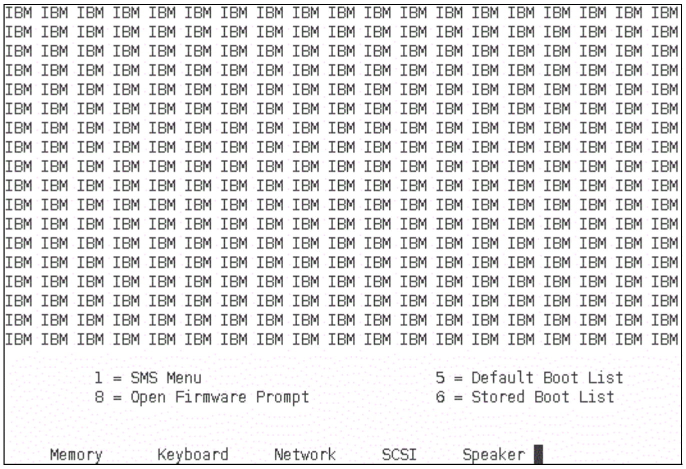
https://www-304.ibm.com/webapp/set2/sas/f/lopdiags/installtools/home.html

Even if you do not plan to use the IBM Installation Toolkit to install your servers, it might serve you as a rescue bootable media that can help you perform preinstallation tasks, such as setting up a hardware RAID array for your operating system disks.

Booting from the IBM Installation Toolkit media

The IBM Installation Toolkit media can be booted from the network, but this section covers only the DVD boot method for simplicity.

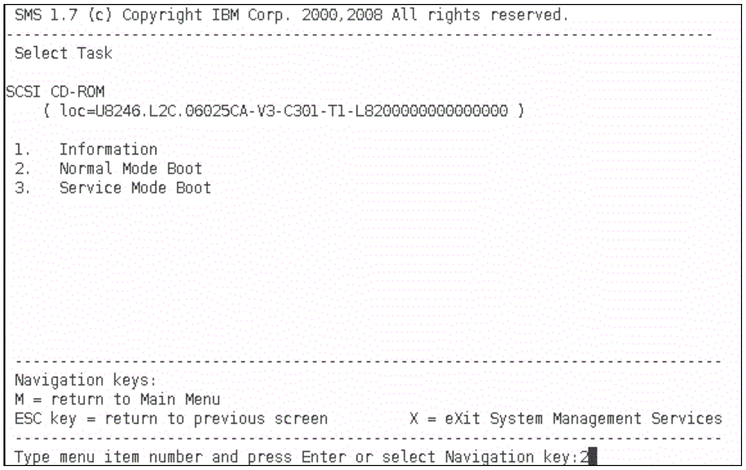
After you turn on your LPAR, or your server if you are using bare metal without any virtualization at all (a preferred practice when using IBM OpenPower™ hardware as data / compute nodes), go to the SMS menu and select the proper options to boot from the DVD drive by pressing 1 when the SMS menu opens, as shown in Figure 3-1.



```
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM
IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM IBM

         1 = SMS Menu                          5 = Default Boot List
         8 = Open Firmware Prompt              6 = Stored Boot List


     Memory      Keyboard     Network     SCSI     Speaker
```

Figure 3-1 The SMS screen on IBM Power Systems

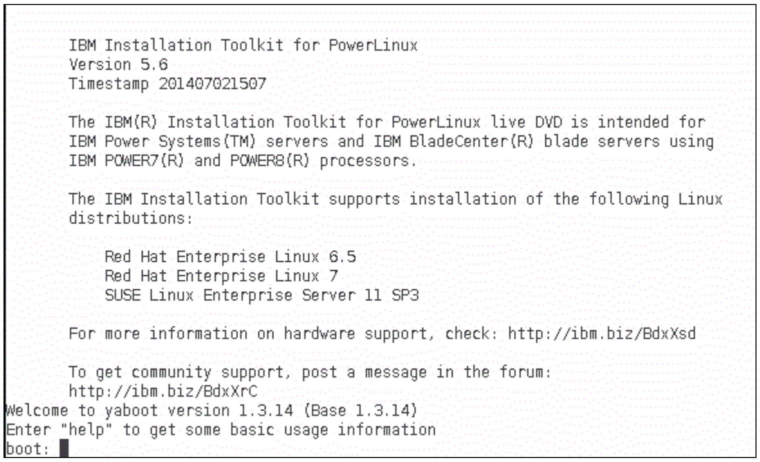
Select options 5 (Select Boot Options) → 1 (Select Install/Boot Device), then select the DVD as the boot device and perform a normal boot. Figure 3-2 shows where you should be within the SMS menu just before booting off DVD media.



```
 SMS 1.7 (c) Copyright IBM Corp. 2000,2008 All rights reserved.
-------------------------------------------------------------------------------
 Select Task

SCSI CD-ROM
    ( loc=U8246.L2C.06025CA-V3-C301-T1-L8200000000000000 )

 1.   Information
 2.   Normal Mode Boot
 3.   Service Mode Boot










 -------------------------------------------------------------------------------
 Navigation keys:
 M = return to Main Menu
 ESC key = return to previous screen         X = eXit System Management Services
 -------------------------------------------------------------------------------
 Type menu item number and press Enter or select Navigation key:2
```

Figure 3-2 Within SMS - about to boot from the DVD

After the system starts to boot, the IBM Installation Toolkit bootloader screen opens, as shown in Figure 3-3.

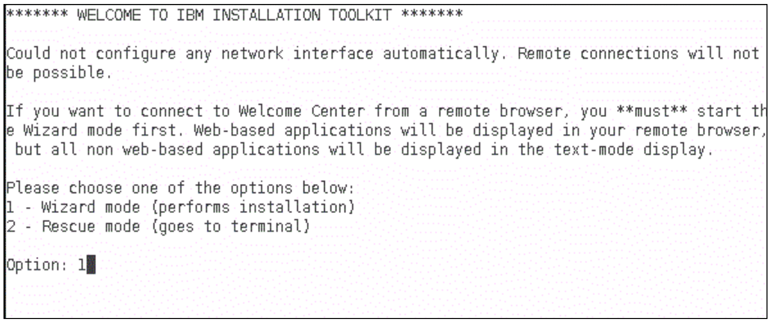


```

       IBM Installation Toolkit for PowerLinux
       Version 5.6
       Timestamp 201407021507

       The IBM(R) Installation Toolkit for PowerLinux live DVD is intended for
       IBM Power Systems(TM) servers and IBM BladeCenter(R) blade servers using
       IBM POWER7(R) and POWER8(R) processors.

       The IBM Installation Toolkit supports installation of the following Linux
       distributions:

           Red Hat Enterprise Linux 6.5
           Red Hat Enterprise Linux 7
           SUSE Linux Enterprise Server 11 SP3

       For more information on hardware support, check: http://ibm.biz/BdxXsd

       To get community support, post a message in the forum:
       http://ibm.biz/BdxXrC
Welcome to yaboot version 1.3.14 (Base 1.3.14)
Enter "help" to get some basic usage information
boot:
```

Figure 3-3 IBM Installation Toolkit bootloader screen

Now, all you need to do is let the system boot. No actions are required until you reach the screen that is shown in Figure 3-4 on page 35. The IBM Installation Toolkit can be used in two ways:

- Wizard mode: Used for setting up and performing a Linux installation on POWER.
- Rescue mode: Used as an in-memory Linux RAM disk to rescue the system or perform pre-installation activities, such as formatting disks for RAID use and creating RAID arrays with disks.



```
******* WELCOME TO IBM INSTALLATION TOOLKIT *******

Could not configure any network interface automatically. Remote connections will not
be possible.

If you want to connect to Welcome Center from a remote browser, you **must** start th
e Wizard mode first. Web-based applications will be displayed in your remote browser,
 but all non web-based applications will be displayed in the text-mode display.

Please choose one of the options below:
1 - Wizard mode (performs installation)
2 - Rescue mode (goes to terminal)

Option: 1
```

Figure 3-4 IBM Installation Toolkit - wizard or rescue modes selection

IBM Power Systems provides you with some of the most reliable hardware in terms of availability. This reliability is achieved by having redundant hardware components and a redundant virtualization architecture.

If you do not use a Power Systems server as a bare metal server, that is, you either virtualize it with PowerVM or PowerKVM, the hardware redundancy and virtualization architecture redundancy exists within your virtualization layer. For example, if your LPARs disks come from an external storage unit, these disks are probably protected by a RAID array in the storage itself, so you do not need to create another data protection layer at the LPAR level. In this case, you might proceed to install Linux into your LPAR and choose option 1, as shown in Figure 3-4, to get into the *wizard mode*.

However, if you intend to use your server as a bare metal server or in a full-server single LPAR mode, using its internal disks only, protect your operating system installation by using a RAID 10 array that uses two of your internal disks. In this case, you should select option 2 to get into *rescue mode*, as shown in Figure 3-4. The following section shows how to set up RAID with internal disks in your environment.

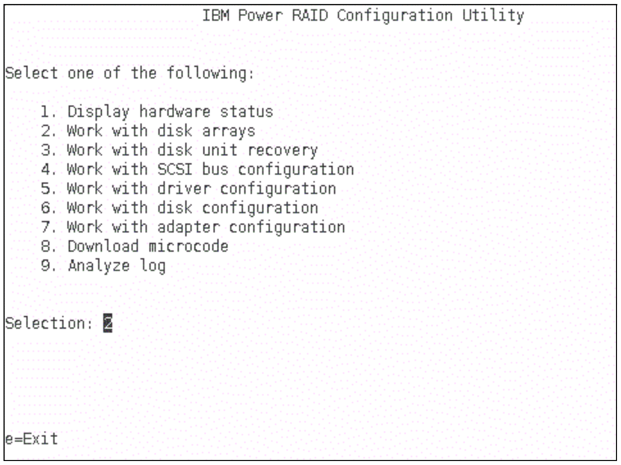
Creating RAID arrays with internal disks

To create RAID arrays with the internal disks of your Power Systems server, run the **iprconfig**[1] command. The **iprconfig** command is a text-mode, but menu-driven, tool that you use to format disks for either RAID or JBOD[2] use. Your server factory disk setup has all of them configured as JBOD disks. So, to use a couple of them as RAID disks, you must use the **iprconfig** command to format them for RAID use first.

[1] IBM Power RAID Configuration Utility.

[2] Just a Bunch Of Disks, commonly referred to as non-RAID disks.

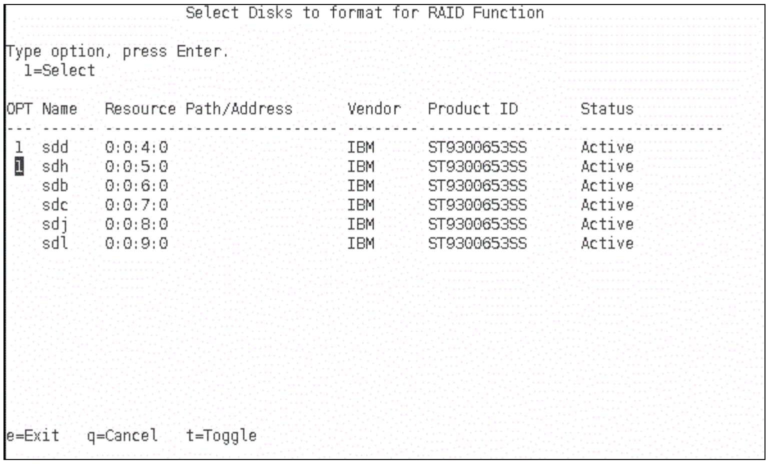
Run **iprconfig** on your shell so that its menu opens, as shown in Figure 3-5.



```
                      IBM Power RAID Configuration Utility


Select one of the following:

    1. Display hardware status
    2. Work with disk arrays
    3. Work with disk unit recovery
    4. Work with SCSI bus configuration
    5. Work with driver configuration
    6. Work with disk configuration
    7. Work with adapter configuration
    8. Download microcode
    9. Analyze log


Selection: 2





e=Exit
```

Figure 3-5 IBM Power RAID Configuration Utility menu

After you are in the **iprconfig** menu, select 2. Work with disks arrays → 5. Format device For RAID function. Now, you are presented with a set of all of the internal disks that are available on your Power Systems server, as shown in Figure 3-6.



```
                    Select Disks to format for RAID Function

Type option, press Enter.
  1=Select

OPT Name   Resource Path/Address      Vendor   Product ID       Status
--- ------ ------------------------- -------- ---------------- ----------------
 1  sdd    0:0:4:0                    IBM      ST9300653SS      Active
 1  sdh    0:0:5:0                    IBM      ST9300653SS      Active
    sdb    0:0:6:0                    IBM      ST9300653SS      Active
    sdc    0:0:7:0                    IBM      ST9300653SS      Active
    sdj    0:0:8:0                    IBM      ST9300653SS      Active
    sdl    0:0:9:0                    IBM      ST9300653SS      Active









e=Exit   q=Cancel   t=Toggle
```

Figure 3-6 .iprconfig - RAID formatting disk selection

Select the first two disks to create your RAID array. The first two disks are not necessarily the ones with the lowest device numbering (which is alphabetical order in this case, that is, `sda`, `sdb`, `sdc`, and so on), but the ones with the lowest resource address numbering. This is a preferred practice so that you always pick up the disks that are labelled as 1 and 2 from the physical disks bay of your Power Systems server. In Figure 3-6, the first two disks are selected with 1, as required by the tool. Press the Enter key and follow the **iprconfig** guidance to start formatting your disks.

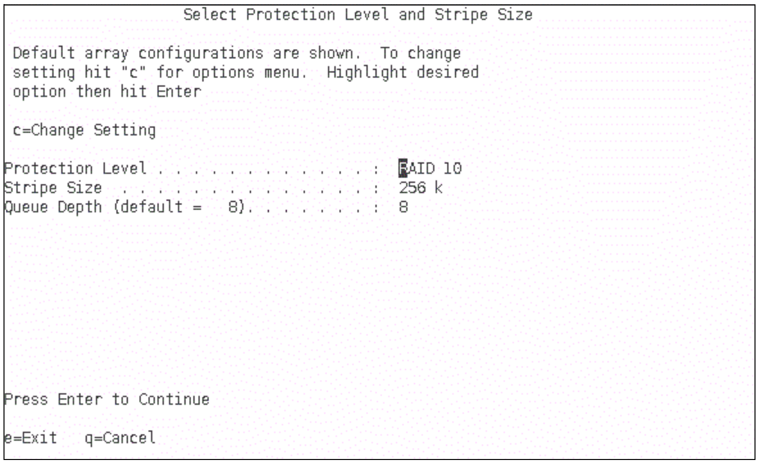
> **Note:** Although disks formatting happens in parallel, it might take a while to complete. In our lab machines, it took 25 - 45 minutes.

After formatting completes, these disks description changes from Physical Disk to Advanced Function Disk in the Option 6. Work with disk configuration menu entry (see Figure 3-5).

Now, you must create the RAID array. Go back to the first menu screen of the **iprconfig** utility (Figure 3-5 on page 36) and select 2. Work with disk arrays → 2. Create a disk array. Then, select the adapter under which your formatted disks are found. In an entry-level Power Systems server box such as OpenPower, you see two adapters for accessing the internal disks. Selecting either one at this step is fine because both can see your disks.

The next screen shows the disks that are available to use for RAID functions. Only the disks you formatted as RAID disks appear on this list. Select the two disks in to create your array and proceed through the wizard.

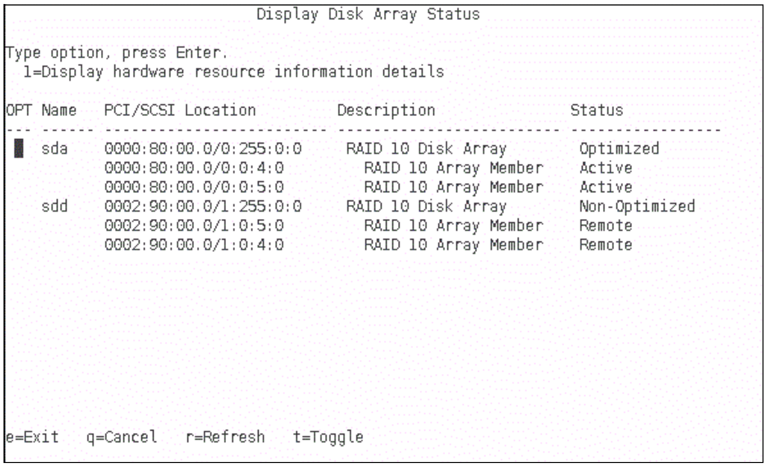
Now, you are given the choice to change the RAID type, stripe size, and queue depth of the disk array, as shown in Figure 3-7. As there are two disks, and you want to protect the Linux installation from any single disk failure, select the RAID 10 configuration, which provides data redundancy through disk mirroring.



```
                    Select Protection Level and Stripe Size

 Default array configurations are shown.  To change
 setting hit "c" for options menu.  Highlight desired
 option then hit Enter

 c=Change Setting

Protection Level . . . . . . . . . . . . :  RAID 10
Stripe Size  . . . . . . . . . . . . . . :  256 k
Queue Depth (default =   8). . . . . . . :  8









Press Enter to Continue

e=Exit   q=Cancel
```

Figure 3-7 Configure the array type and characteristics

Proceed to the next steps in the array creation wizard. Confirm your choices and create your array. After it is complete, you can verify that your RAID 10 disk is operational by selecting 2. Work with disk arrays → 1. Display disk array status. You see an output similar to Figure 3-8.



```
                            Display Disk Array Status

Type option, press Enter.
  1=Display hardware resource information details

OPT Name   PCI/SCSI Location         Description              Status
--- ------ ------------------------- ------------------------ ----------------
    sda    0000:80:00.0/0:255:0:0     RAID 10 Disk Array       Optimized
           0000:80:00.0/0:0:4:0         RAID 10 Array Member    Active
           0000:80:00.0/0:0:5:0         RAID 10 Array Member    Active
    sdd    0002:90:00.0/1:255:0:0     RAID 10 Disk Array       Non-Optimized
           0002:90:00.0/1:0:5:0         RAID 10 Array Member    Remote
           0002:90:00.0/1:0:4:0         RAID 10 Array Member    Remote










e=Exit   q=Cancel   r=Refresh   t=Toggle
```

Figure 3-8 RAID 10 disk array

In Figure 3-8 on page 37, the test machines are IBM Power7R2 boxes with six internal disks and two internal disk controllers. Because there are two controllers, you see two device names for the RAID 10 disk array, that is, `sda` and `sdd`. Both controllers display two disks, and you can tell that these are the same disks by looking at the disk hardware address location (0:4:0 and 0:5:0; the location prefix refers to the controller hardware address, which is naturally different). Linux deals with this situation later on by creating a *multipathed* disk device. So, whenever the operating system refers to this disk array later on, it does so by using a multipathed device name (for example, `mpathe`) as opposed to a disk name (for example, `sda`).

> **Note:** It is important to remember that the RAID10 array is, in this example, on disks `sda` and `sdf` because as you proceed with the IBM Installation Toolkit, this information is needed in one of the installation wizard steps.

Starting a Linux installation with the IBM Installation Toolkit

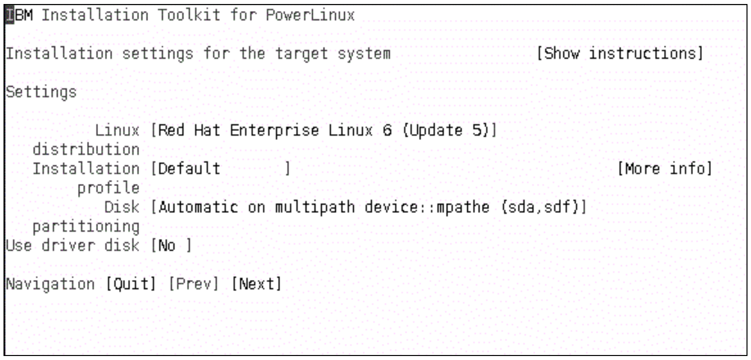
After you are sure that your installation disk is either protected by an internal RAID array as explained in “Creating RAID arrays with internal disks” on page 35, or it is protected by other means (for example, an external storage disk with RAID, or a virtual disk with some underlying redundancy), you can start a Linux installation by using the *wizard mode* of the IBM Installation Toolkit, as described in Figure 3-4 on page 35.

The wizard tool is straightforward and has instructions and help available throughout all of its steps. For more information about this wizard, go to the following website:

http://ibm.biz/BdxXrC

Note the disk partitioning screen, which is shown in Figure 3-9. Select the disk (or disk list in the case of multipathing) that corresponds to your protected installation disk. In these tests, use the simplest disk partitioning scenario, which is automatic partitioning in the multipathed RAID 10 disk. If you must create elaborate partitioning schemes, see the *IBM Installation Toolkit Users Guide* at the following website:

https://www-304.ibm.com/webapp/set2/sas/f/lopdiags/installtools/home.html



```
IBM Installation Toolkit for PowerLinux

Installation settings for the target system             [Show instructions]

Settings

          Linux [Red Hat Enterprise Linux 6 (Update 5)]
   distribution
   Installation [Default       ]                                  [More info]
        profile
           Disk [Automatic on multipath device::mpathe (sda,sdf)]
   partitioning
Use driver disk [No ]

Navigation [Quit] [Prev] [Next]
```

Figure 3-9 IBM Installation Toolkit disk selection screen

After the installation wizard finishes, the system restarts and starts to install Linux. Because this scenario uses RHEL as the Linux version for the Power Systems environment, Anaconda opens and retrieves the kickstart information that is generated by the IBM Installation Toolkit to install the system. After it is finished, your Linux environment is ready to use.

After all of the Power Systems servers you planned to use for your InfoSphere BigInsights installation have Linux up and running, it is time to set them up according to InfoSphere BigInsights prerequisites. This task is described in 3.1.2, “Operating system prerequisites setup for InfoSphere BigInsights” on page 39.

3.1.2 Operating system prerequisites setup for InfoSphere BigInsights

Before you attempt to install InfoSphere BigInsights in a Linux server, you must check that all the prerequisites are met. These prerequisites include available disk space, installed software, some system configuration in `/etc`, and so on. This section explains what these prerequisites are and helps you get organized before you start your InfoSphere BigInsights installation.

Security and authentication

InfoSphere BigInsights needs an administrator user ID to run. There are a few alternatives that you can use depending on your environment's security policies. From a high-level perspective, and not specific to its installation only, InfoSphere BigInsights works with both local and central user management. For example, if you have an LDAP database, you might choose to have your InfoSphere BigInsights environment configured to use it.

From a low-level perspective, and specific to the installation steps for InfoSphere BigInsights, you can have the administrator user be the following users:

- The root user
- A non-root user with sudo privileges, with two alternatives:
  - A non-root user that can gain root privileges on the node where the installation is done, and then uses the root user account to access all other nodes.
  - A non-root user that can gain root privileges on all of the cluster nodes.

It is important to know which type of user you plan to use for installing the software, as it affects the network prerequisites that you must set up before the installation. For example, if you choose to use root for installing the software, a passwordless SSH authentication mechanism for root is required to all nodes, including the node where the installation is done. If you choose a non-root user that can gain root privileges and from there use root to access other nodes, you also need this SSH passwordless setup for the root user among the cluster nodes. The last option, a non-root user account that can gain root privileges on all nodes, requires that the user exists on all nodes, that its password is the same on all of them, and that **`sudo`** is properly configured on all nodes to grant the user privileges.

In the test environment, install InfoSphere BigInisights as root and configure passwordless SSH authentication among the nodes for the root user.

Users and groups

Unless you have a pre-existing environment on top of which you plan to install InfoSphere BigInsights, or you manage users and groups through an LDAP central management server, you do not need to take actions for setting up users and groups because the InfoSphere BigInsights installer handles these tasks for you.

If you cannot have InfoSphere BigInsights set up these items in your environment, make sure that the biadmin user and group exist, and that the group has the proper group ID of 123. Then, set a password for the biadmin user.

It is important to make the password consistent across your cluster nodes, especially if you plan to install InfoSphere BigInsights with biadmin as a non-root user and not use SSH exchange keys as explained in “Security and authentication” on page 39. Example 3-1 shows how to accomplish these steps locally on a given server.

Example 3-1 Create the biadmin user and group and set up a password

```
# groupadd -g 123 biadmin
# useradd -g biadmin -u 123 admin
# passwd biadmin
```

> **Note:** The password for the biadmin user must be fewer than 31 characters in length.

As a second step, if you need to set up manually the biadmin user and group, add the biadmin user to the `suders` file by completing the following steps, which are also shown in Example 3-2:

1. Comment out this line to run **`sudo`** without a terminal. Add a # at the beginning of the line.
2. Locate the line `#%wheel ALL=(ALL) NOPASSWD: ALL` and replace it with the one that is shown in Example 3-2.

Example 3-2 Adding the biadmin user to the /etc/sudoers file

```
vi /etc/sudoers
[... snpip...]
# Defaults requiretty 1
[... snip...]
biadmin ALL=(ALL) NOPASSWD: ALL 2
```

The last step is to check that the biadmin user has **`bash`** set up as its shell interpreter. This information is in the `/etc/passwd` file. Check that the setup looks like the lines that are shown in Example 3-3.

Example 3-3 Bash shell interpreter for user biadmin

```
# grep biadmin /etc/passwd
biadmin:x:200:123:/home/biadmin:/bin/bash
```

File system and disk setup prerequisites

A minimum amount of free disk space is required to install InfoSphere BigInsights. Table 3-2 shows the disk space values. Ensure that the cluster nodes meet these requirements.

Table 3-2 File system space requirements for InfoSphere BigInsights

Directory	Minimum free space
`/`	10 GB
`/tmp`	5 GB

`/$BIGINSIGHTS_HOME`
`(default is /opt/ibm`
	15 GB

`/$BIGINSIGHTS_VAR`
`(default is /var/ibm`
	5 GB

`/home/$USER_HOME`
`(default is biadmin)`
	5 GB

You must ensure that your disks are referred to in /etc/fstab by UUID[3] as opposed to the device name to prevent a mismatch of a disk device and its mount point if one of the disks becomes temporarily unavailable and the system is rebooted. Using a disk's UUID is a preferred practice and is the default method that RHEL uses after a default installation. Your /etc/fstab file should look like Example 3-4.

Example 3-4 Example of a proper /etc/fstab file that uses a disk's UUID

```
# /etc/fstab
# Created by anaconda on Mon Sep 15 15:15:14 2014
UUID=5b585a08-9ebd-460f-8a0c-60ede9452146 /        ext4 defaults        1 1
UUID=e167b3de-ee5e-4818-a46c-4ddbc51bfa2e /boot    ext4 defaults        1 2
UUID=5e64e32c-2ea9-4dad-9928-2d8a8d8d89d4 swap     swap defaults        0 0
UUID=3634a8b6-1170-49f1-a852-a509e7e7fbf7 swap     swap defaults        0 0
tmpfs                   /dev/shm                tmpfs   defaults        0 0
devpts                  /dev/pts                devpts  gid=5,mode=620  0 0
sysfs                   /sys                    sysfs   defaults        0 0
proc                    /proc                   proc    defaults        0 0
/dev/bigpfs      /gpfs gpfs       rw,nomtime,relatime,dev=bigpfs,noauto 0 0
```

If your /etc/fstab file does not refer to disks by UUID, run the **blkid** command to obtain the UUIDs and replace your disk devices definitions with them, as shown in Example 3-5.

Example 3-5 Obtaining a disk's UUID

```
[root@mgmt02 ~]# blkid
/dev/mapper/mpathep2: UUID="e167b3de-ee5e-4818-a46c-4ddbc51bfa2e" TYPE="ext4"
/dev/mapper/mpathep3: UUID="3634a8b6-1170-49f1-a852-a509e7e7fbf7" TYPE="swap"
/dev/mapper/mpathep4: UUID="5b585a08-9ebd-460f-8a0c-60ede9452146" TYPE="ext4"
/dev/loop0: LABEL="CDROM" TYPE="iso9660"
```

Networking prerequisites

Regarding a cluster solution, communication among the cluster nodes is importance. To ensure InfoSphere BigInsights works correctly, complete the following steps:

1. Set up passwordless SSH authentication among the nodes.

   This setup is required if you do not have a non-root user on all cluster nodes that is granted root privileges in its sudoers configuration file. Example 3-6 shows how to perform this operation. Perform it for the root user. The InfoSphere BigInisights installation program performs this step for the biadmin user if it does not exist; otherwise, perform key exchange for this user as well.

   Example 3-6 Create and exchange SSH keys for a user

   ```
   #Run this command on each node for each user for which you want to set up keys.
   ssh-keygen -t rsa

   #Then, run the following command from each node to every other node, including
   #the node itself.
   ssh-copy-id -i ~/.ssh/id_rsa.pub <user>@<cluster_node_N>
   ```

[3] Universally Unique Identifier.

2. Define IP addresses in the /etc/hosts file.

   The IP addresses and the hosts' fully qualified names, followed by their short names, must be defined within the /etc/hosts file of every cluster node, as shown in Example 3-7.

   Example 3-7 /etc/hosts file nodes definitions

   ```
   [root@mgmt02 ~]# cat /etc/hosts
   127.0.0.1 localhost
   ## data network
   192.168.0.2     mgmt01.test.ibm.com      mgmt01
   192.168.0.3     mgmt02.test.ibm.com      mgmt02
   192.168.0.4     mgmt03.test.ibm.com      mgmt03
   192.168.0.5     mgmt04.test.ibm.com      mgmt04
   192.168.0.6     data01.test.ibm.com      data01
   192.168.0.7     data02.test.ibm.com      data02
   192.168.0.8     data03.test.ibm.com      data03
   ```

3. Disable the firewall (**iptables**).

   The firewall must be disabled on all nodes. This is accomplished by running the commands that are shown in Example 3-8.

   Example 3-8 Disable the iptables

   ```
   # service iptables save
   # service iptables stop
   # chkconfig iptables off
   ```

4. Disable IPv6.

   IPv6 must be disabled on all nodes. This is done by editing the disable-ipv6.conf and the network system configuration files and appending the information that is shown in Example 3-9. You must restart your servers after performing these changes.

   Example 3-9 Disable IPv6 on RHEL

   ```
   vi /etc/modprobe.d/disable-ipv6.conf
      install ipv6 /bin/true

   vi /etc/sysconfig/network
      NETWORKING=yes
      NETWORKING_IPV6=no

   echo “net.ipv6.conf.all.disable_ipv6 = 1” >> /etc/sysctl.conf
   ```

5. Increase the network operation timeout value (optional).

   If you are using a slow network to perform the installation, you might want to increase the timeout value for the installer operations. This is an InfoSphere BigInsights installation setting in the installer/hdm/conf/hdm.properties file found within your InfoSphere BigInsights software extract directory. Under the section DEPLOY_OPTIONS, change the value of the default.exec.timeout property to 600000 (10 minutes). Example 3-10 illustrates this step.

   Example 3-10 Increase the execution timeout for the InfoSphere BigInsights installer

   ```
   vi <BigInsights_installer_extract_dir>/installer/hdm/conf/hdm.properties
   [... snip ...]
   # [optional]
   ```

```
# When there is an internal parameter that is used to control the timeout
#period when hdm #runs a command or a shell script executable file, increasing
#this value might help when hdm runs into timeout errors under some bad network
#conditions; otherwise, keeping this value at 5 minutes is reasonable because
#some operations or scripts have several minutes of execution time.
# A value <=0 means no timeout at all.
 default.exec.timeout=600000
```

> **Note:** This step assumes that you have extracted your InfoSphere BigInsights installation file into the cluster node from where you start the installation.

6. Synchronize the system clock with a Network Time Protocol (NTP) server.

   There is no secret about synchronizing the cluster nodes clock with an NTP server. You probably use your local, private NTP server address, in which case you must add only the line `server <ntp_server> iburst` to your `/etc/ntp.conf` file.

   However, it is a prerequisite that the cluster nodes have time synchronization at the time of installation. Therefore, if you do not have an internal NTP server available or cannot connect to a public one over the internet, then you must configure an internal, private NTP server. You can use your first cluster node to act as an NTP server, and synchronize the rest of the nodes with it. Example 3-11 shows a bare minimal functional configuration for a node to act as an NTP server.

   Example 3-11 Configuration for a node to act as an NTP server

   ```
   # vi /etc/ntp.conf

   driftfile /var/lib/ntp/drift
   disable auth
   restrict 127.0.0.1
   server  127.127.1.0
   fudge   127.127.1.0 stratum 10
   ```

   Verify that the clocks are synchronized on all nodes by running the **ntpstat** command.

Kernel and system configuration

Check and tune the kernel parameters in the `/etc/sysctl.conf`, according to Example 3-12.

Example 3-12 Tune some of the kernel parameters

```
# vi /etc/sysctl.conf
[...]
kernel.pid_max = 4194303
[...]
net.ipv4.ip_local_port_range = 1024 64000
[...]
```

Installed software prerequisites

Some prerequisite software must be installed on all of the cluster nodes, and some on the installation cluster node in particular. You might not have all of the required packages installed at the time of the operating system installation, so set up a **yum** repository before you start the InfoSphere BigInsights installation.

You might want to create a repository in your local network that is exported over NFS, or point to your RHEL6 installation DVD media. Either works, but a network **yum** repository is more convenient because InfoSphere BigInsights can request package installation through the **yum** command dynamically during the installation. Some guidance about how to create a local **yum** repository and make it available through **yum** in your local network is at the following website:

http://red.ht/1CHzQX6

The following software packages must be installed in the cluster node from where you start the installation:

- expect
- numactl
- ksh

You can check whether these packages are already installed by running the **rpm -qa | grep <pacakge_name>** command, and install any of the missing packages.

Here are the packages that must be installed in all cluster nodes:

- compat-libstdc++-33
- gcc-c++
- imake
- kernel-devel
- kernel headers
- libstdc++
- redhatlsb
- vacpp.rte

Operating system configuration prerequisites

A few of the operating system's settings should be changed before installing InfoSphere BigInsights. The installer can change these settings during installation run time, but it is a preferred practice to do so before the installation so that the installer pre-check does not show warning or error messages when it runs. The following list summarizes these configuration changes:

1. *Minimum* values for some parameters within the ulimits configuration.

   You can configure these values *either* globally 1, to a group 2, or to a user 3. Doing it globally also affects all other users and groups for which there is no specific configuration in the file. Either way, the numbers in Example 3-13 are the bare minimum that is required for a InfoSphere BigInsights environment. Increase the ulimits number according to your workload.

   Example 3-13 Configure /etc/security/limits.conf

   ```
   # vi /etc/security/ulimits.conf
   [... snip ...]
   * hard nofile 16384 1
   * soft nofile 16384 1
   * hard noproc 10240 1
   * soft noproc 10240 1
   [... snip...]
   @biadmin hard nofile 16384 2
   @biadmin soft nofile 16384 2
   @biadmin hard noproc 10240 2
   @biadmin soft noproc 10240 2
   [... snip ...]
   biadmin hard nofile 16384 3
   ```

```
biadmin soft nofile 16384 3
biadmin hard noproc 10240 3
biadmin soft noproc 10240 3
```

2. The root user’s `umask` should be set to `022` during the InfoSphere BigInsights installation process.

   Usually, system administrators do not have an issue with a system umask of `022`, which is the default for Linux installations. However, if you work with tougher security constraints, you change this value in your environment when you install InfoSphere BigInsights. If you are not sure or if your environment is highly secured, simply proceed as follows and later remember to undo it after InfoSphere BigInsights has been installed. On all nodes, open the `.bashrc` file for the root user and append the `umask` line, as shown in Example 3-14.

   Example 3-14 Set up umask for the root user

   ```
   # vi ~/.bashrc
   [... snip...]
   umask 022
   ```

3.2 InfoSphere BigInsights

Check that your environment meets all of the prerequisites that are listed in 3.1.2, “Operating system prerequisites setup for InfoSphere BigInsights” on page 39 before proceeding with the InfoSphere BigInsights installation.

There are multiple scenarios for an InfoSphere BigInsights installation, for example, by using Spectrum Scale or HDFS. This section covers the installation flow of the InfoSphere BigInsights software stack. In this example, the environment uses Spectrum Scale as the file system and Adaptive MapReduce as the runtime scheduler. Both are installed and configured by the installer.

If you plan to customize your Spectrum Scale setup, configure it before the installation of InfoSphere BigInsights. For information about how to perform a standard Spectrum Scale configuration, see *IBM Spectrum Scale (formerly GPFS)*, SG24-8254.

3.2.1 Installation

InfoSphere BigInsights can be installed in one of two ways:

- By using the installation wizard
- By using the silent installer and a response file

The installation wizard guides you through the configuration of InfoSphere BigInsights. A response file is generated at the end of the wizard. This response file can then be used for a GUI or silent installation of InfoSphere BigInsights.

Start the installation wizard web server from the extracted InfoSphere BigInsights directory, as shown in Example 3-15.

Example 3-15 Start the InfoSphere BigInsights installation web server

```
[root@mgmt01 ~]# cd /tmp/biginsights-3.0.0.1-Linux-ppc64-b20140711_1547/

[root@mgmt01 biginsights-3.0.0.1-SNAPSHOT-Linux-ppc64-b20140711_1547]# ls
artifacts        installer          install.properties  licenses        start.sh
fullinstall.xml  installer-console  _jvm                silent-install

[root@mgmt01 biginsights-3.0.0.1-Linux-ppc64-b20140711_1547]# ./start.sh
```

Although the InfoSphere BigInsights installation is run on a single node, the process itself is done on all nodes that you have specified during the configuration process. There is no need to run the installation process on each node.

> **Note:** A prerequisite for a successful silent installation of InfoSphere BigInsights is that this web server is stopped after you generate the response file. Run the following command from the extracted installation directory:
>
> `# ./start.sh shutdown`

For more information about the wizard, see “Installation wizard” on page 47.

> **Note:** The InfoSphere BigInsights installer listens on port *8300*.

The second method, a silent installation that uses a response file, assumes that you have built a response file for the installation either by using the installation wizard or by customizing some of the examples that are found in the `silent-install` (needs to be *`cmd/directory/font`*) directory. You should either start with the sample XML files or you can run the installation wizard, as explained previously, to generate a response file.

Generating a response file by using the installation wizard is the suggested method because of the following reasons:

- You do not need to understand the internals of the XML response file.
- You can easily restart the installation process if you must perform any troubleshooting during the installation.

After the response file is created, it can be used as an argument for the **`silent-install.sh`** script, as shown in Example 3-16. The `fullinstall.xml` file is the response file that is generated by the installation wizard.

Example 3-16 Start a InfoSphere BigInsights silent installation

```
[root@mgmt01 biginsights-3.0.0.1-Linux-ppc64-b20140711_1547]#
./silent-install/silent-install.sh fullinstall.xml
```

For more information about how to use the installation wizard to generate a response file, see Figure 3-28 on page 63.

Installation wizard

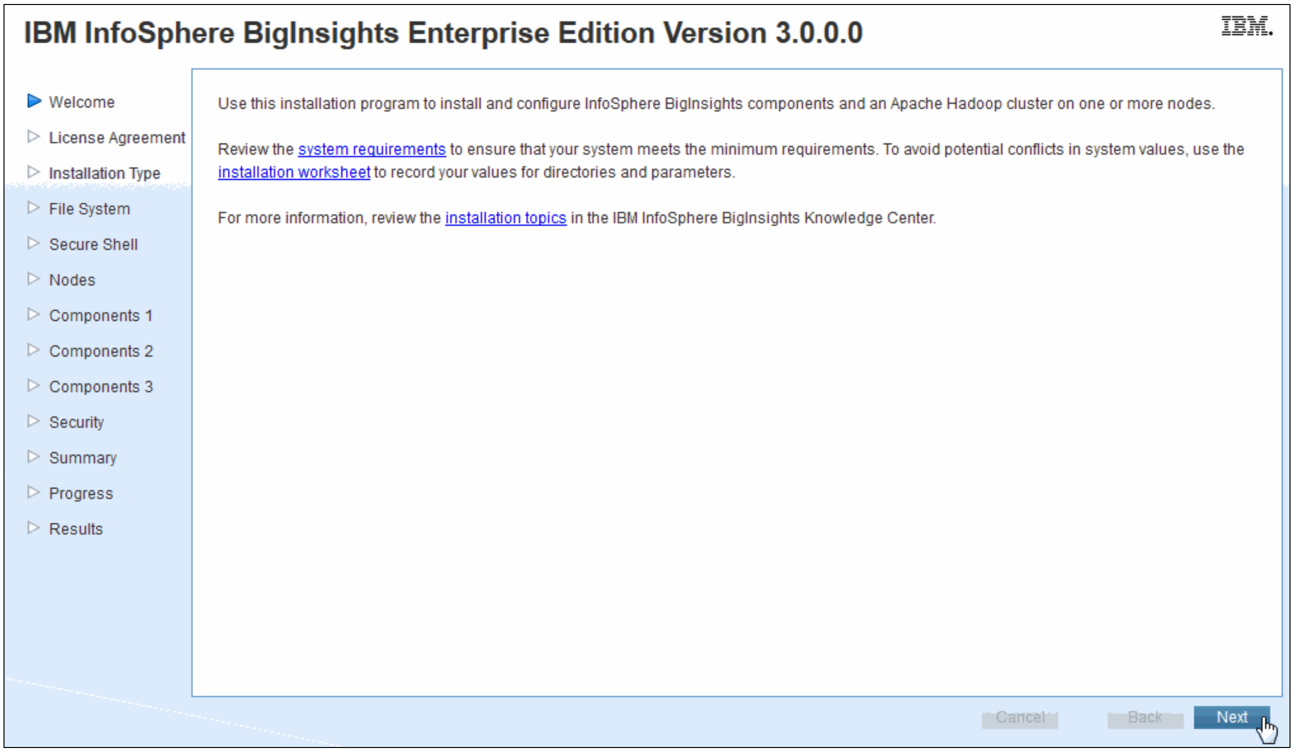
The installation wizard is used to configure the InfoSphere BigInsights cluster. This process determines the distribution of services between the management nodes, the configuration of those services, the number of data nodes, the Spectrum Scale/HDFS configuration, and the security settings. The POSHv2 architecture can be used as a reference when deciding on the distribution of services across the management nodes, as described in “The POSHv2 management nodes layout for InfoSphere BigInsights” on page 24.

The MapReduce scheduler can be configured to use either Apache MapReduce or Adaptive MapReduce (also known as IBM Platform Symphony MapReduce). This choice is made through the `install.properties` file in the installation folder. The contents of the file, and how to choose the scheduler, are shown in Example 3-17.

Example 3-17 Enable the adaptive MapReduce algorithm

```
[root@mgmt01 biginsights-3.0.0.1-Linux-ppc64-b20140711_1547]# vi
install.properties

#-----------------------------------------------------------
# OCO Source Materials
# (C) Copyright IBM Corp. 2013, 2014
# The source code for this program is not published or
# otherwise divested of its trade secrets, irrespective of
# what has been deposited with the US Copyright Office.
#-----------------------------------------------------------
# set AdaptiveMR.Enable to true if you want to install AdaptiveMR instead of
# Apache MapReduce
AdaptiveMR.Enable=true
```

Figure 3-10 shows the Welcome window.



IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

Use this installation program to install and configure InfoSphere BigInsights components and an Apache Hadoop cluster on one or more nodes.

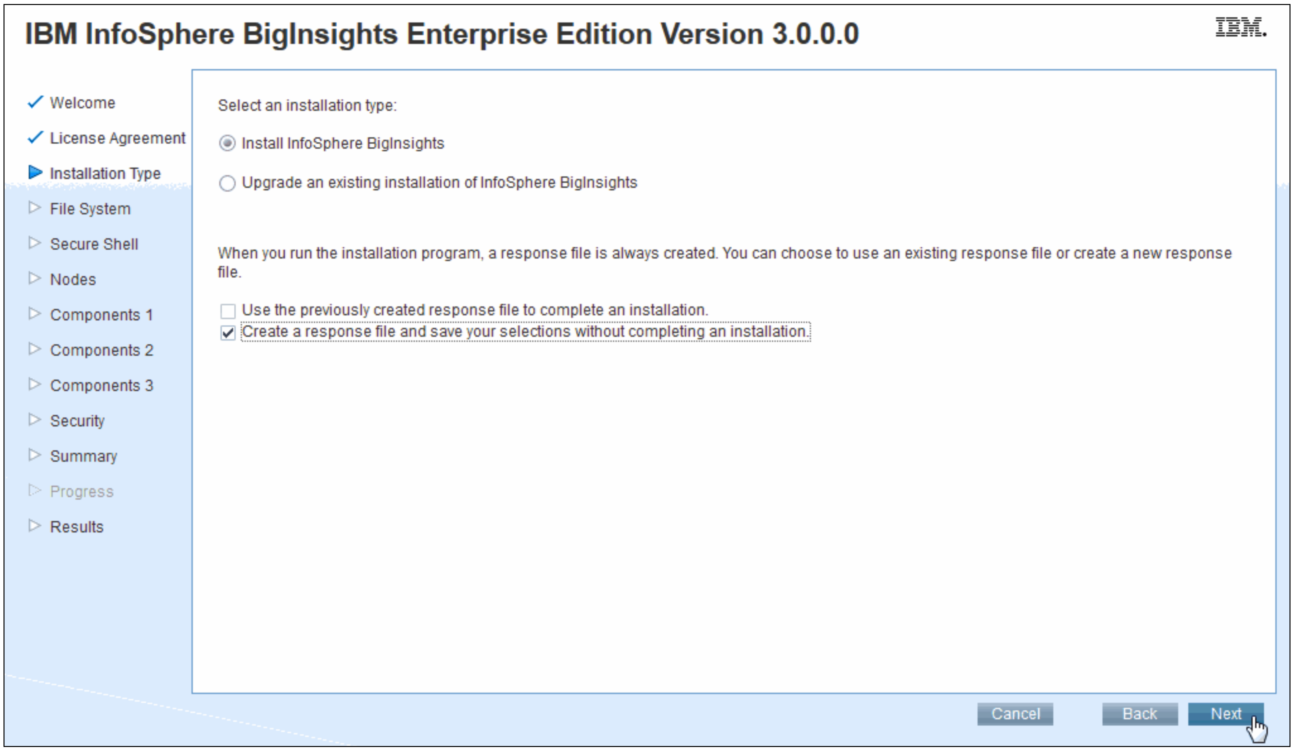
Review the system requirements to ensure that your system meets the minimum requirements. To avoid potential conflicts in system values, use the installation worksheet to record your values for directories and parameters.

For more information, review the installation topics in the IBM InfoSphere BigInsights Knowledge Center.

Figure 3-10 Welcome window

Click **Next** to open the License Agreement window. Accept the license agreement to open the next window.

Figure 3-11 shows the installation type window. Choose whether to perform an installation or an upgrade and whether to use a previously generated response file or create one.



IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

Select an installation type:

● Install InfoSphere BigInsights

○ Upgrade an existing installation of InfoSphere BigInsights

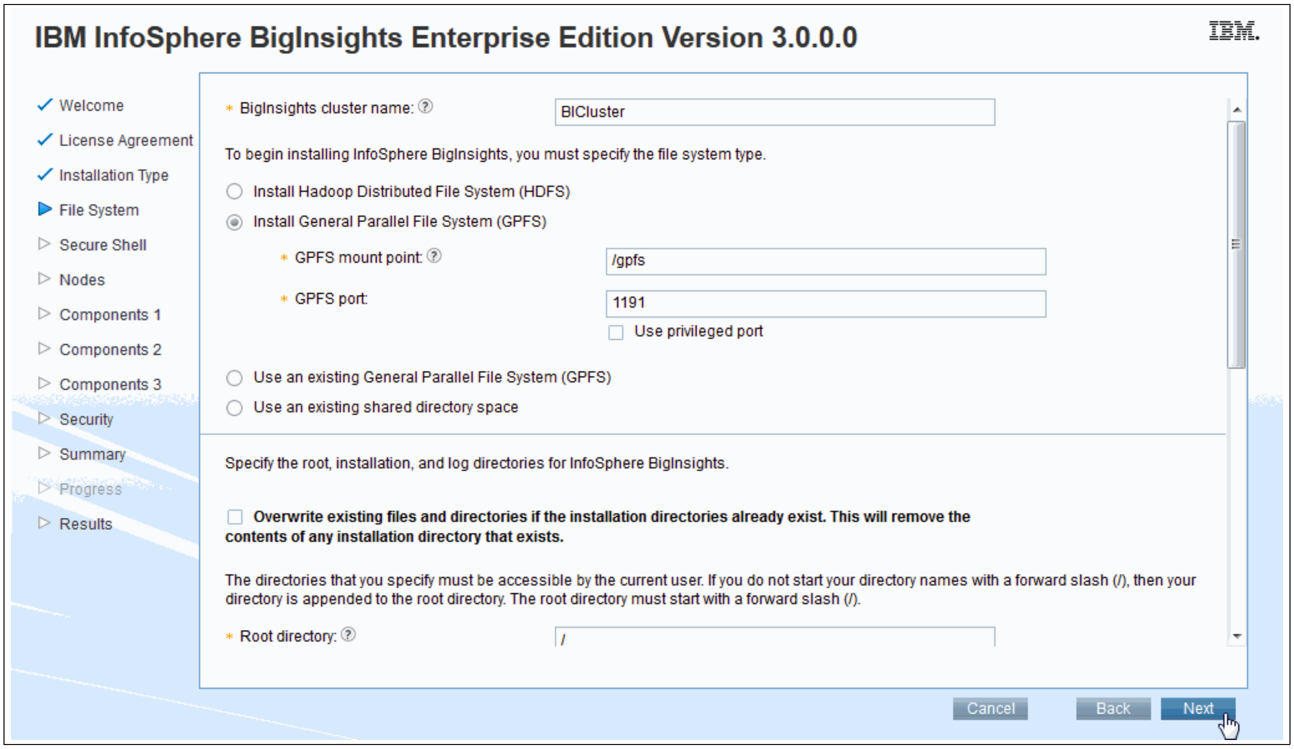
When you run the installation program, a response file is always created. You can choose to use an existing response file or create a new response file.

□ Use the previously created response file to complete an installation.

☑ Create a response file and save your selections without completing an installation.

Figure 3-11 The installation type window

Create an HDFS or Spectrum Scale file system, choose an existing Spectrum Scale file system, or use an existing share directory space, as shown in Figure 3-12. If the **Overwrite existing files and directories if the installation directories already exists. This will remove the contents of any installation directory that exists** check box is selected, the installation can be rerun by running the same command.



IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

* BigInsights cluster name: BICluster

To begin installing InfoSphere BigInsights, you must specify the file system type.

○ Install Hadoop Distributed File System (HDFS)

● Install General Parallel File System (GPFS)

* GPFS mount point: /gpfs

* GPFS port: 1191

□ Use privileged port

○ Use an existing General Parallel File System (GPFS)

○ Use an existing shared directory space

Specify the root, installation, and log directories for InfoSphere BigInsights.

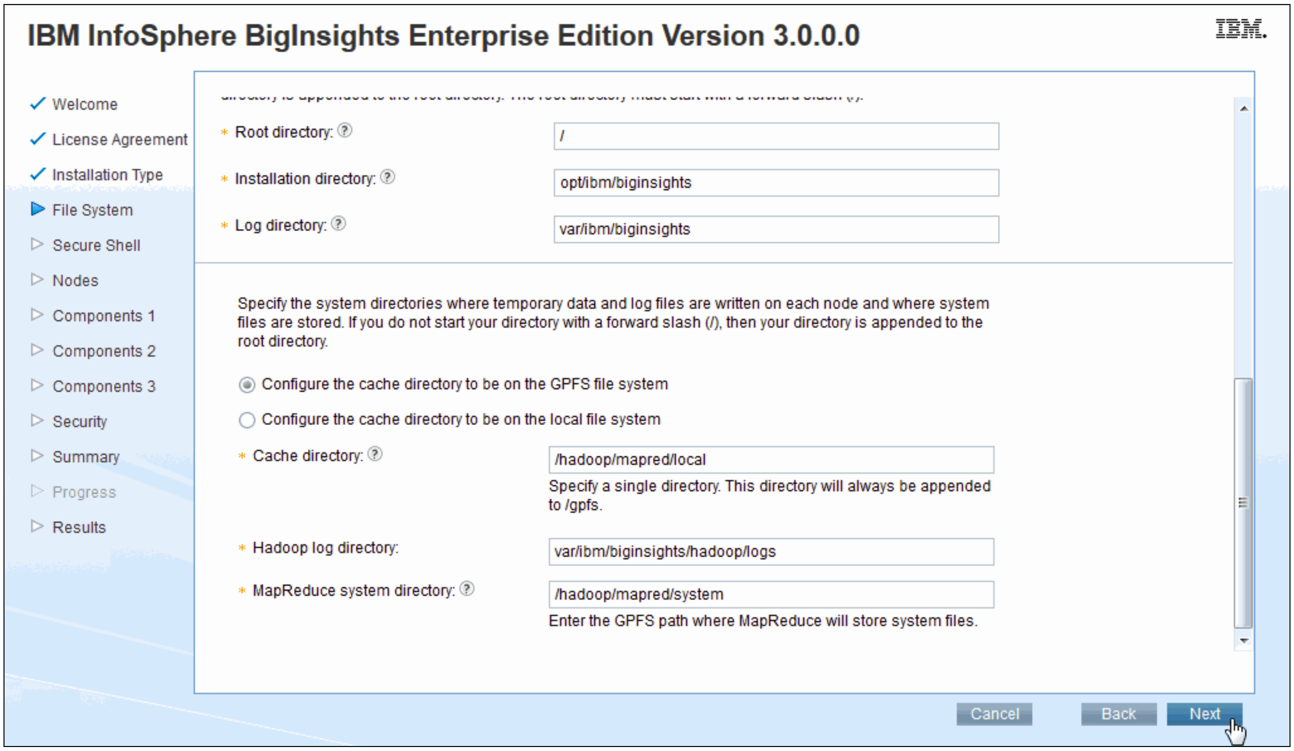
□ **Overwrite existing files and directories if the installation directories already exist. This will remove the contents of any installation directory that exists.**

The directories that you specify must be accessible by the current user. If you do not start your directory names with a forward slash (/), then your directory is appended to the root directory. The root directory must start with a forward slash (/).

* Root directory: /

Figure 3-12 Select and configure the file system

Figure 3-13 on page 51 shows the bottom part of this step on the wizard. This is where you provide information such as the installation directory, logs directory, and caching directory.



IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

directory is appended to the root directory. The root directory must start with a forward slash (/).

* Root directory: /

* Installation directory: opt/ibm/biginsights

* Log directory: var/ibm/biginsights

Specify the system directories where temporary data and log files are written on each node and where system files are stored. If you do not start your directory with a forward slash (/), then your directory is appended to the root directory.

● Configure the cache directory to be on the GPFS file system

○ Configure the cache directory to be on the local file system

* Cache directory: /hadoop/mapred/local

Specify a single directory. This directory will always be appended to /gpfs.

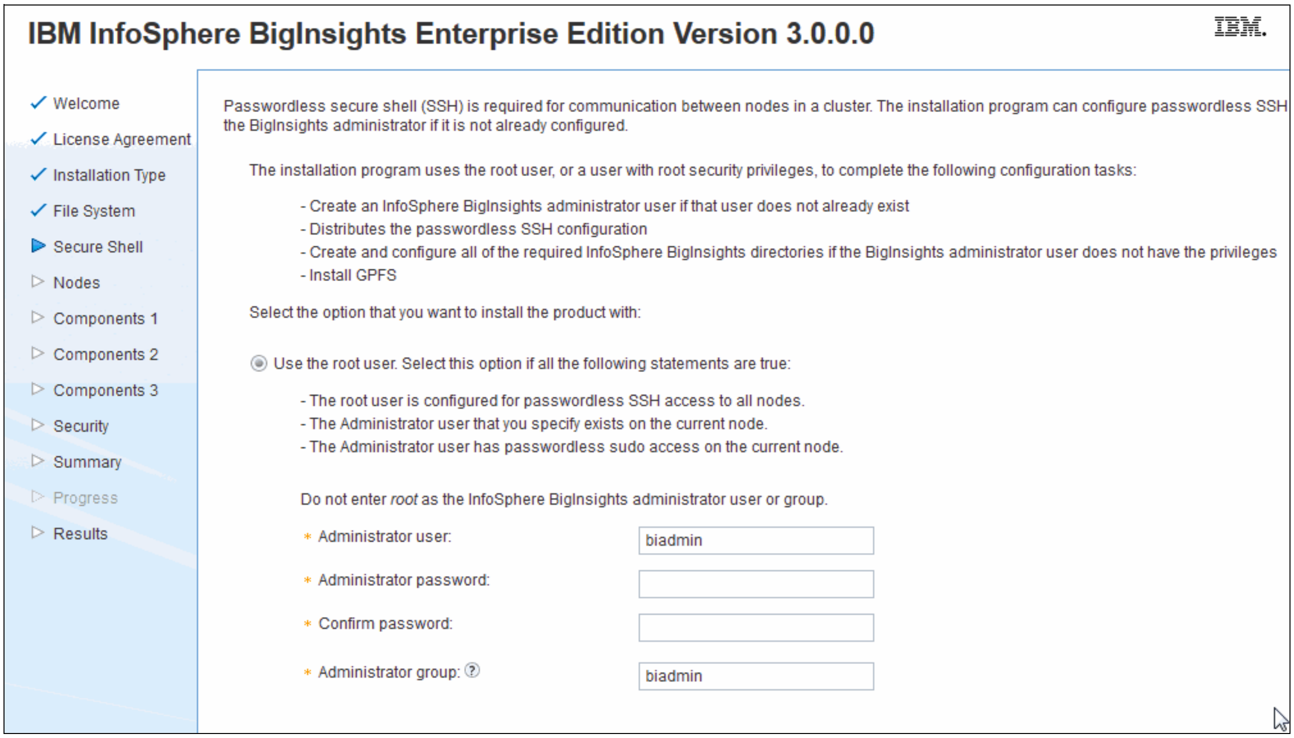
* Hadoop log directory: var/ibm/biginsights/hadoop/logs

* MapReduce system directory: /hadoop/mapred/system

Enter the GPFS path where MapReduce will store system files.

Figure 3-13 Configure the cache directory

Passwordless SSH is required for communication between nodes in the cluster. The installer can configure this feature for all users, including root, but it is preferable that the root user is configured beforehand. Figure 3-14 shows the Secure Shell setup window.



IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

Passwordless secure shell (SSH) is required for communication between nodes in a cluster. The installation program can configure passwordless SSH the BigInsights administrator if it is not already configured.

The installation program uses the root user, or a user with root security privileges, to complete the following configuration tasks:

- Create an InfoSphere BigInsights administrator user if that user does not already exist
- Distributes the passwordless SSH configuration
- Create and configure all of the required InfoSphere BigInsights directories if the BigInsights administrator user does not have the privileges
- Install GPFS

Select the option that you want to install the product with:

● Use the root user. Select this option if all the following statements are true:

- The root user is configured for passwordless SSH access to all nodes.
- The Administrator user that you specify exists on the current node.
- The Administrator user has passwordless sudo access on the current node.

Do not enter *root* as the InfoSphere BigInsights administrator user or group.

* Administrator user: biadmin

* Administrator password:

* Confirm password:

* Administrator group: biadmin

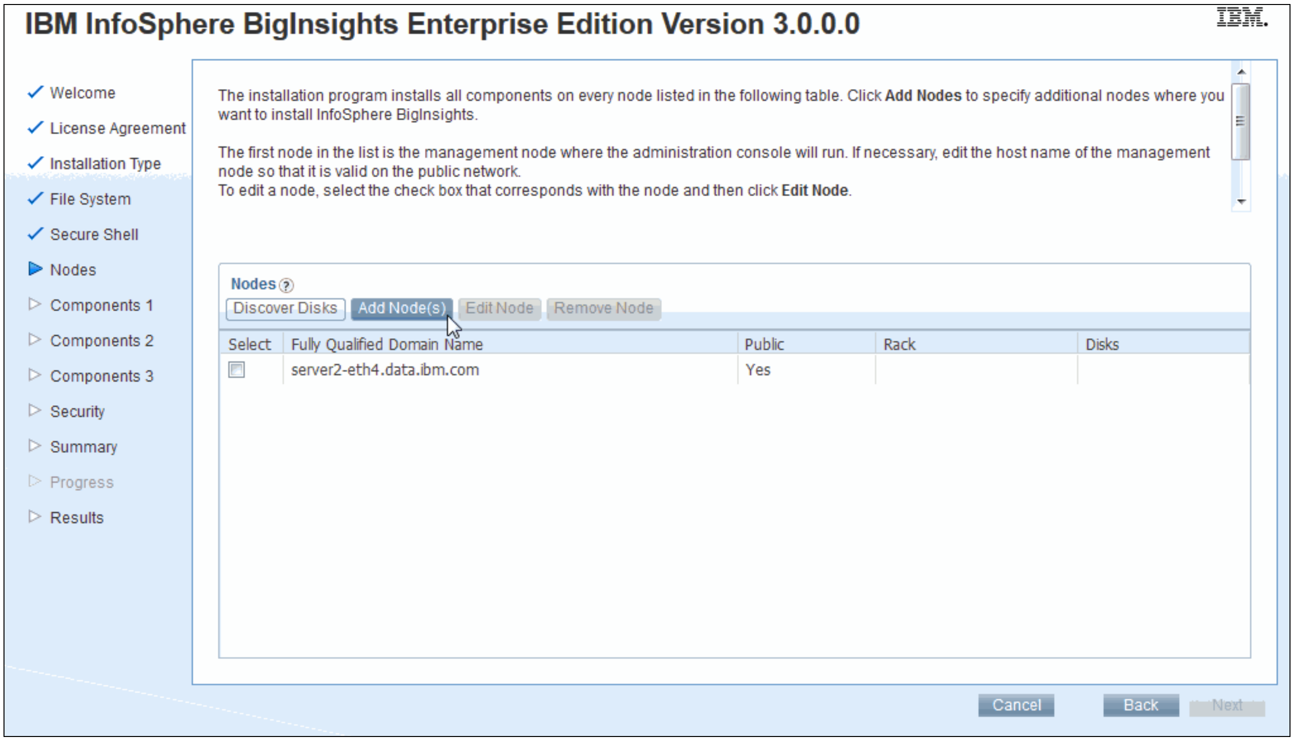
Figure 3-14 Secure Shell configuration

Up to this point, the installation is the same for a non-HA or HA installation.

Non-HA installation

The minimum number of servers for a non-HA installation is four, that is, one management node and three compute nodes.

Figure 3-15 shows how to select nodes to be included in the cluster.



IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

The installation program installs all components on every node listed in the following table. Click **Add Nodes** to specify additional nodes where you want to install InfoSphere BigInsights.

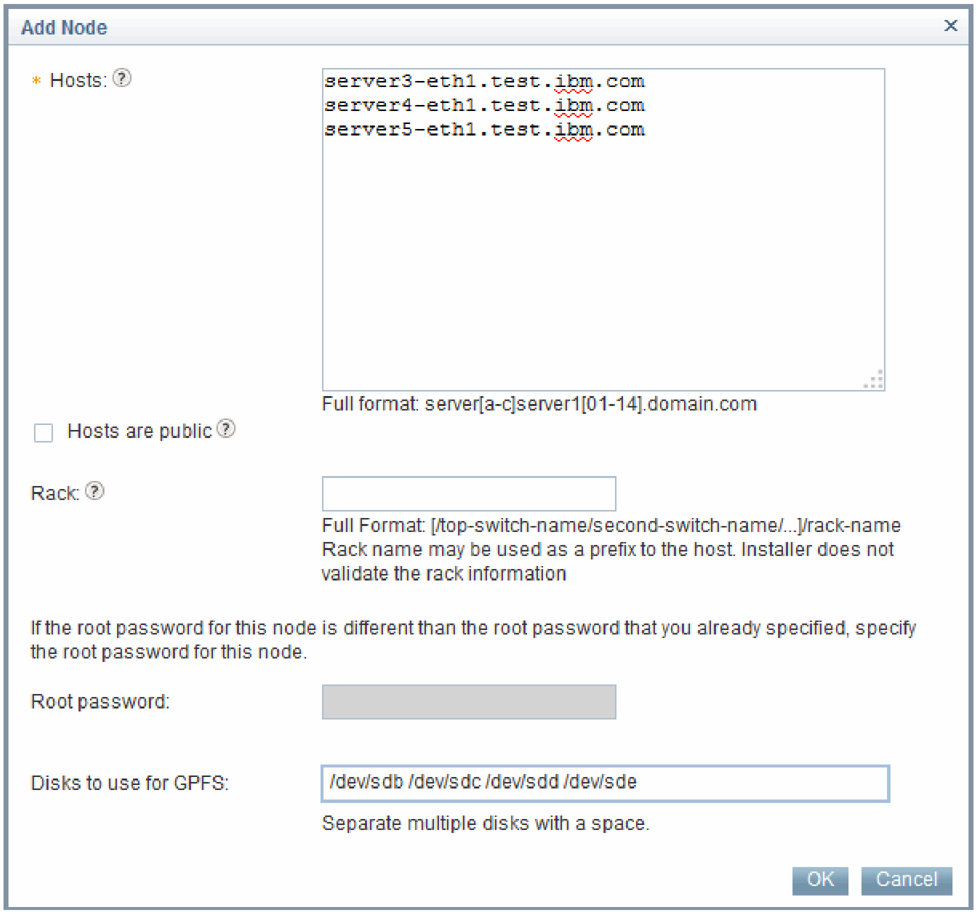
The first node in the list is the management node where the administration console will run. If necessary, edit the host name of the management node so that it is valid on the public network.
To edit a node, select the check box that corresponds with the node and then click **Edit Node**.

Select	Fully Qualified Domain Name	Public	Rack	Disks
□	server2-eth4.data.ibm.com	Yes		

Figure 3-15 Add nodes

In Figure 3-15, `server2-eth4.data.ibm.com` is the server on which the installation wizard web server is running.

Figure 3-16 on page 53 shows the addition of the three compute nodes and the disks on these nodes that are used for Spectrum Scale. Here, it is possible to detail which rack these servers are in and whether these nodes are publicly accessible.



* Hosts: ?

```
server3-eth1.test.ibm.com
server4-eth1.test.ibm.com
server5-eth1.test.ibm.com
```

Full format: server[a-c]server1[01-14].domain.com

□ Hosts are public ?

Rack: ?

Full Format: [/top-switch-name/second-switch-name/...]/rack-name
Rack name may be used as a prefix to the host. Installer does not validate the rack information

If the root password for this node is different than the root password that you already specified, specify the root password for this node.

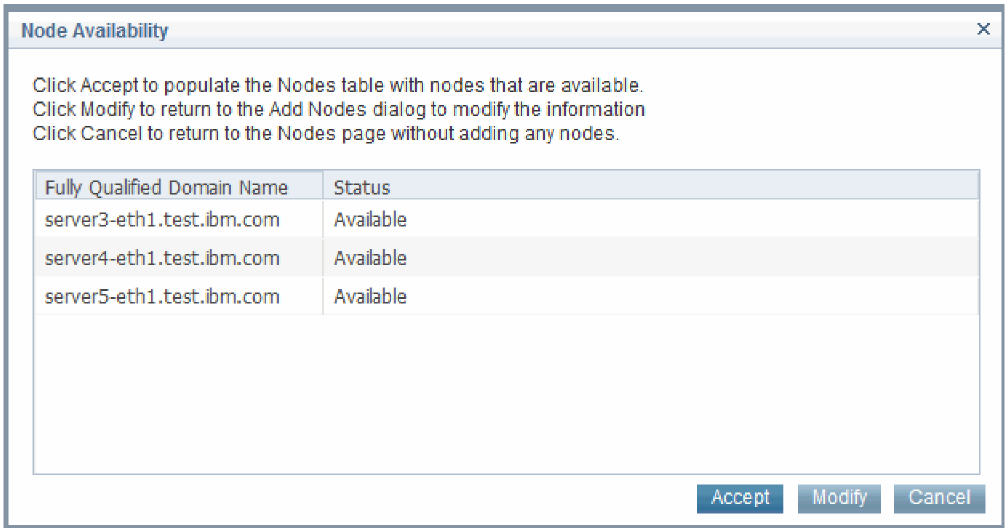
Root password:

Disks to use for GPFS: /dev/sdb /dev/sdc /dev/sdd /dev/sde

Separate multiple disks with a space.

Figure 3-16 Three compute nodes added to the cluster and their disks added to Spectrum Scale

Figure 3-17 shows that the nodes that are selected are available for use. The installer prompts you for confirmation that they nodes should be added to the cluster configuration.

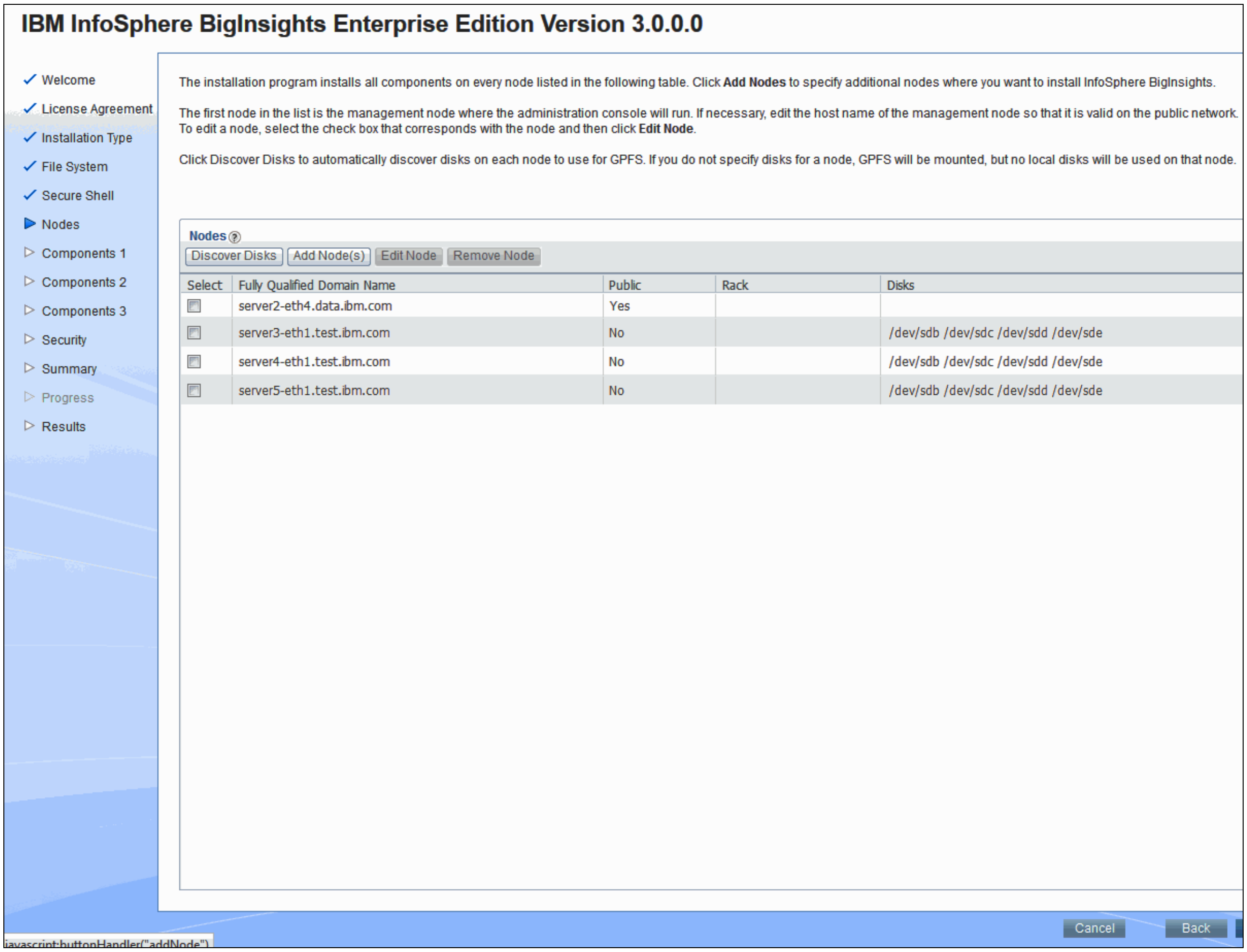


Click Accept to populate the Nodes table with nodes that are available.
Click Modify to return to the Add Nodes dialog to modify the information
Click Cancel to return to the Nodes page without adding any nodes.

Fully Qualified Domain Name	Status
server3-eth1.test.ibm.com	Available
server4-eth1.test.ibm.com	Available
server5-eth1.test.ibm.com	Available

Figure 3-17 Confirm node availability

Figure 3-18 shows a summary of the nodes in the cluster and which disks those nodes are donating to the Spectrum Scale configuration.



IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

The installation program installs all components on every node listed in the following table. Click **Add Nodes** to specify additional nodes where you want to install InfoSphere BigInsights.

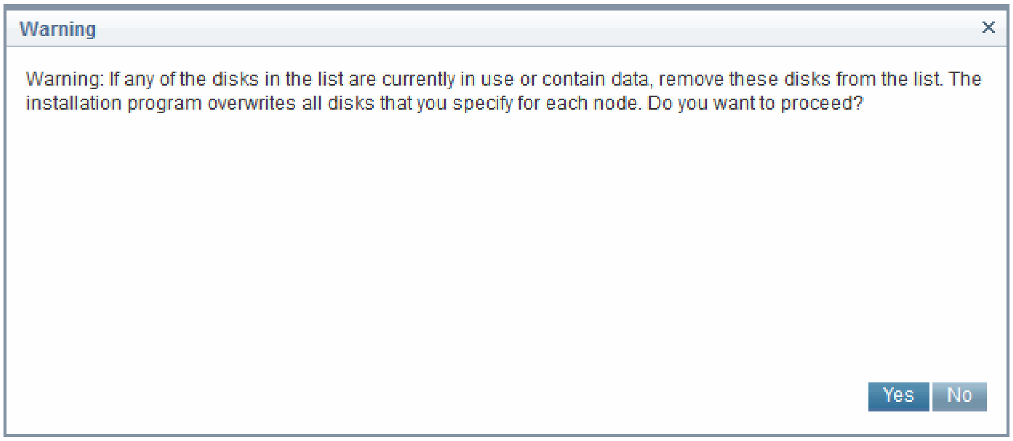
The first node in the list is the management node where the administration console will run. If necessary, edit the host name of the management node so that it is valid on the public network. To edit a node, select the check box that corresponds with the node and then click **Edit Node**.

Click Discover Disks to automatically discover disks on each node to use for GPFS. If you do not specify disks for a node, GPFS will be mounted, but no local disks will be used on that node.

Select	Fully Qualified Domain Name	Public	Rack	Disks
□	server2-eth4.data.ibm.com	Yes		
□	server3-eth1.test.ibm.com	No		/dev/sdb /dev/sdc /dev/sdd /dev/sde
□	server4-eth1.test.ibm.com	No		/dev/sdb /dev/sdc /dev/sdd /dev/sde
□	server5-eth1.test.ibm.com	No		/dev/sdb /dev/sdc /dev/sdd /dev/sde

Figure 3-18 Node addition summary

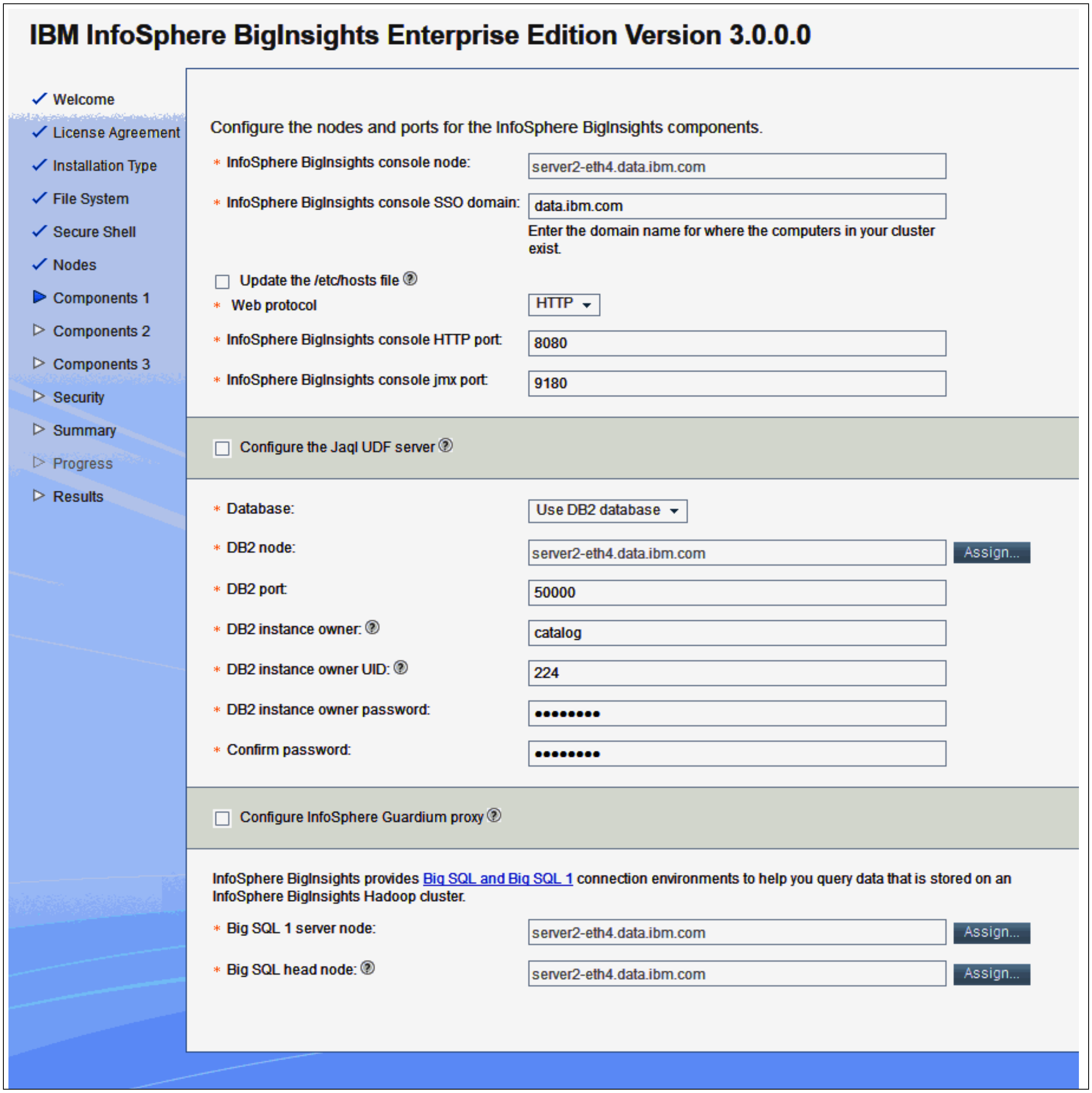
A window with a warning about data loss as a result of the installation process opens. Because this is a fresh installation, click **Yes**, as shown in Figure 3-19.



Warning: If any of the disks in the list are currently in use or contain data, remove these disks from the list. The installation program overwrites all disks that you specify for each node. Do you want to proceed?

Figure 3-19 Data loss warning

Figure 3-20 shows the Components 1 section. From this window, you can configure the nodes and ports for InfoSphere BigInsights, and the user names and passwords for application users for DB2 and BigSQL.



IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

Configure the nodes and ports for the InfoSphere BigInsights components.

* InfoSphere BigInsights console node: server2-eth4.data.ibm.com

* InfoSphere BigInsights console SSO domain: data.ibm.com

Enter the domain name for where the computers in your cluster exist.

□ Update the /etc/hosts file

* Web protocol: HTTP

* InfoSphere BigInsights console HTTP port: 8080

* InfoSphere BigInsights console jmx port: 9180

□ Configure the Jaql UDF server

* Database: Use DB2 database

* DB2 node: server2-eth4.data.ibm.com [Assign...]

* DB2 port: 50000

* DB2 instance owner: catalog

* DB2 instance owner UID: 224

* DB2 instance owner password: ••••••••

* Confirm password: ••••••••

□ Configure InfoSphere Guardium proxy

InfoSphere BigInsights provides Big SQL and Big SQL 1 connection environments to help you query data that is stored on an InfoSphere BigInsights Hadoop cluster.

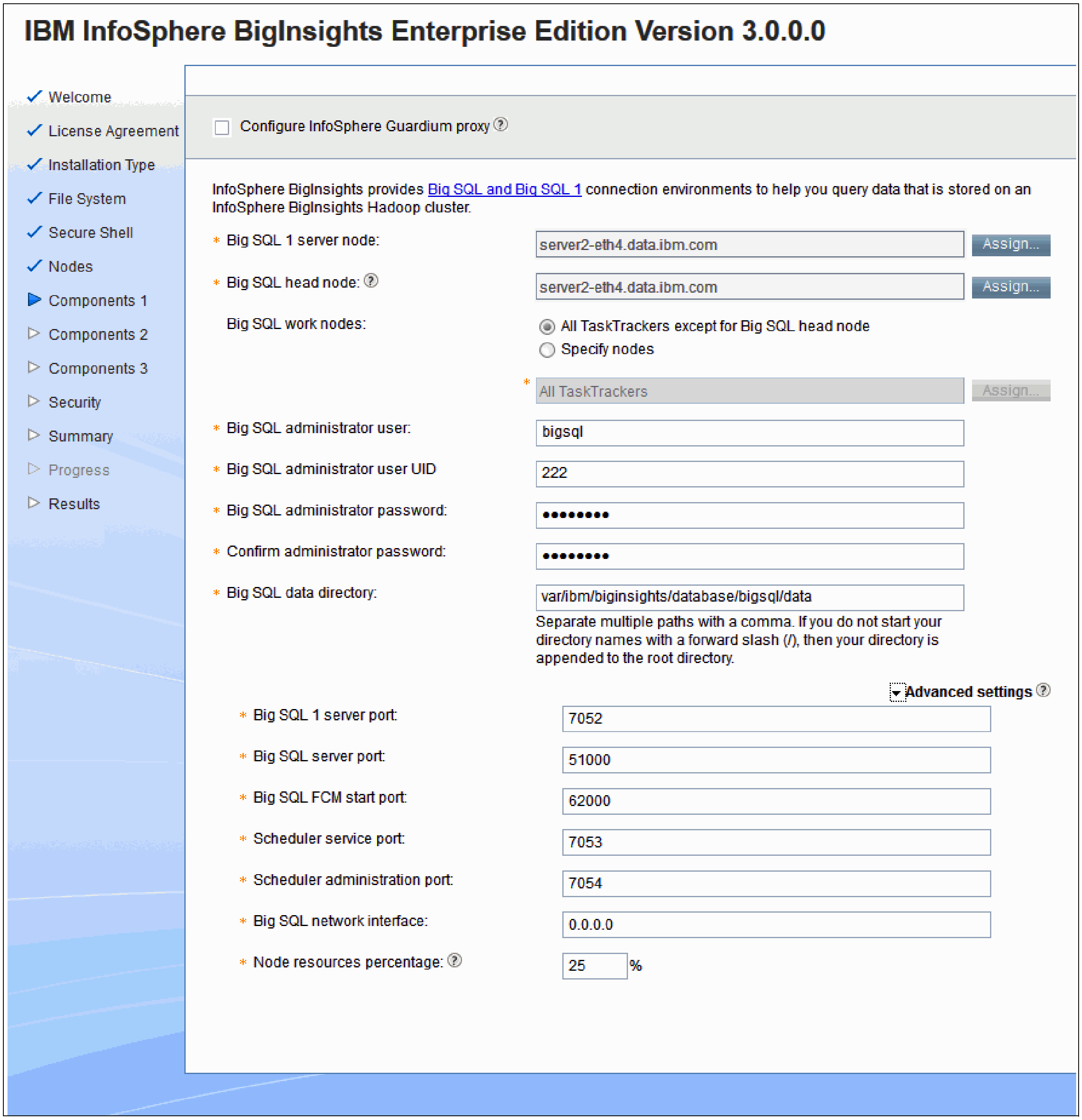
* Big SQL 1 server node: server2-eth4.data.ibm.com [Assign...]

* Big SQL head node: server2-eth4.data.ibm.com [Assign...]

Figure 3-20 Configure the nodes and ports for the InfoSphere BigInsights components

Figure 3-21 shows the advanced settings for the configuration of the BigSQL 1 and BigSQL environments.

The BigSQL services are constrained to use 25 - 75% of the server resources. If the box is dedicated to the BigSQL service, you can use higher percentages, but if the box is shared with other services, then more than about 25% risks crashing the box.



□ Configure InfoSphere Guardium proxy ⓘ

InfoSphere BigInsights provides Big SQL and Big SQL 1 connection environments to help you query data that is stored on an InfoSphere BigInsights Hadoop cluster.

* Big SQL 1 server node:	server2-eth4.data.ibm.com
* Big SQL head node: ⓘ	server2-eth4.data.ibm.com
Big SQL work nodes:	● All TaskTrackers except for Big SQL head node ○ Specify nodes
*	All TaskTrackers
* Big SQL administrator user:	bigsql
* Big SQL administrator user UID	222
* Big SQL administrator password:	••••••••
* Confirm administrator password:	••••••••
* Big SQL data directory:	var/ibm/biginsights/database/bigsql/data

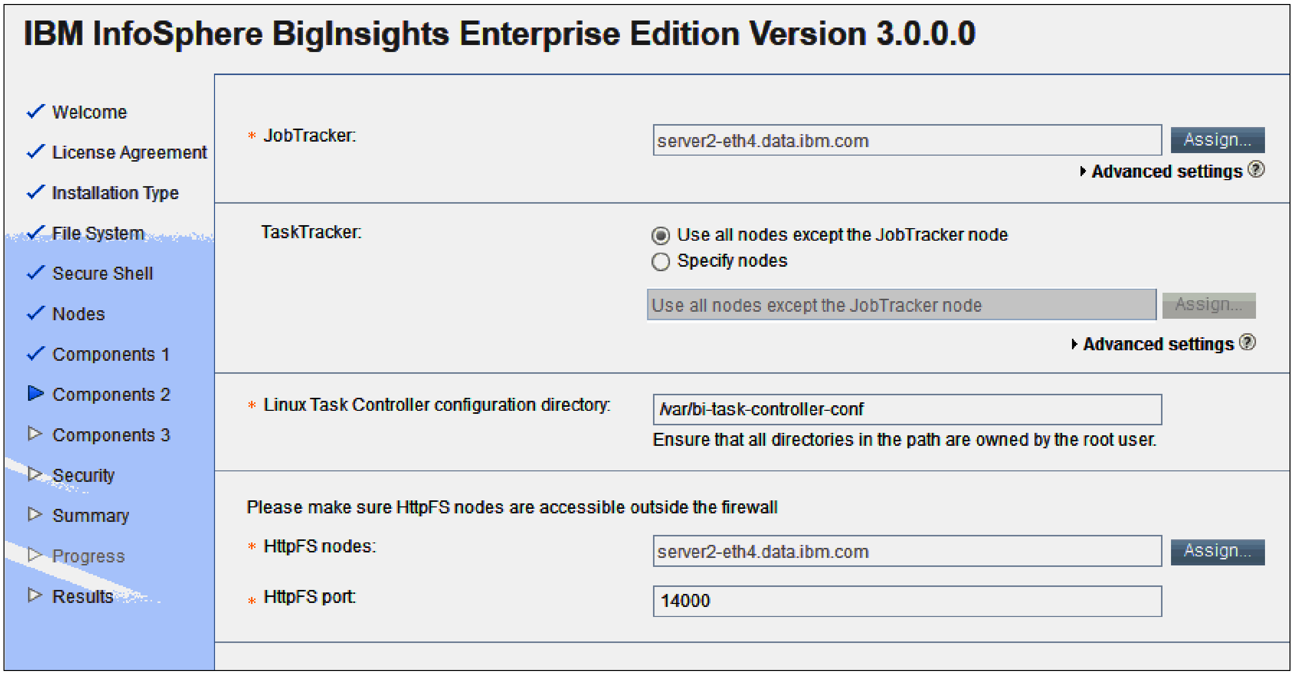
Separate multiple paths with a comma. If you do not start your directory names with a forward slash (/), then your directory is appended to the root directory.

Advanced settings ⓘ

* Big SQL 1 server port:	7052
* Big SQL server port:	51000
* Big SQL FCM start port:	62000
* Scheduler service port:	7053
* Scheduler administration port:	7054
* Big SQL network interface:	0.0.0.0
* Node resources percentage: ⓘ	25 %

Figure 3-21 Advanced settings

Configure the JobTracker, TaskTracker, Linux Task Controller Directory, and HttpFS settings, as shown in Figure 3-22.



IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

* JobTracker: server2-eth4.data.ibm.com [Assign...]

Advanced settings

TaskTracker:
● Use all nodes except the JobTracker node
○ Specify nodes

Use all nodes except the JobTracker node [Assign...]

Advanced settings

* Linux Task Controller configuration directory: /var/bi-task-controller-conf

Ensure that all directories in the path are owned by the root user.

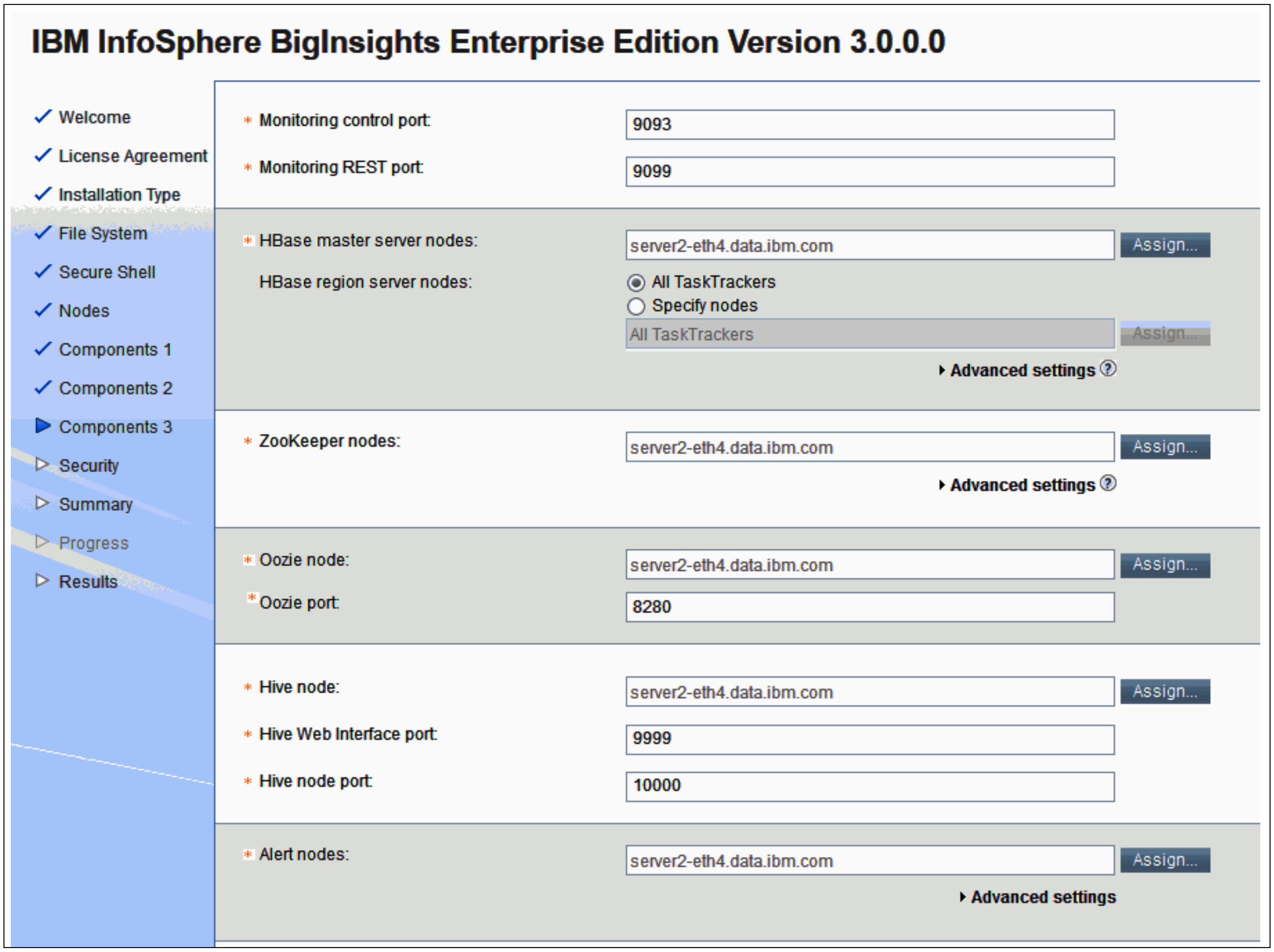
Please make sure HttpFS nodes are accessible outside the firewall

* HttpFS nodes: server2-eth4.data.ibm.com [Assign...]

* HttpFS port: 14000

Figure 3-22 Configure the components 2 section of the installation wizard

Figure 3-23 shows the configuration of Monitoring, HBase, ZooKeeper, Oozie, Hive, and the Alert nodes.



IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

* Monitoring control port: 9093

* Monitoring REST port: 9099

* HBase master server nodes: server2-eth4.data.ibm.com Assign...

HBase region server nodes: ● All TaskTrackers ○ Specify nodes

All TaskTrackers Assign...

▸ Advanced settings

* ZooKeeper nodes: server2-eth4.data.ibm.com Assign...

▸ Advanced settings

* Oozie node: server2-eth4.data.ibm.com Assign...

* Oozie port: 8280

* Hive node: server2-eth4.data.ibm.com Assign...

* Hive Web Interface port: 9999

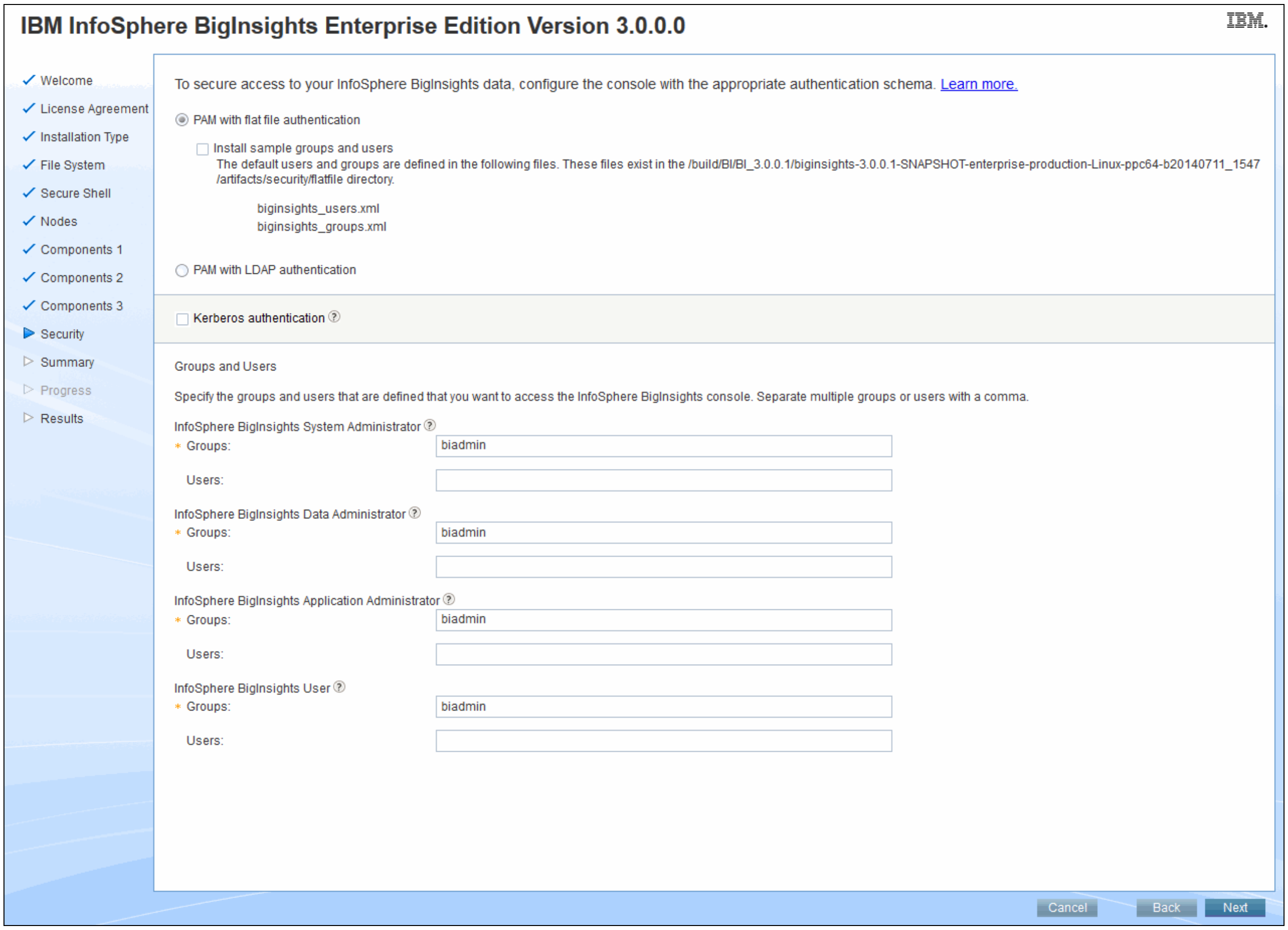
* Hive node port: 10000

* Alert nodes: server2-eth4.data.ibm.com Assign...

▸ Advanced settings

Figure 3-23 Configuration options for component 3

Configure the authentication method for users, as shown in Figure 3-24.



To secure access to your InfoSphere BigInsights data, configure the console with the appropriate authentication schema. Learn more.

● PAM with flat file authentication

□ Install sample groups and users
The default users and groups are defined in the following files. These files exist in the /build/BI/BI_3.0.0.1/biginsights-3.0.0.1-SNAPSHOT-enterprise-production-Linux-ppc64-b20140711_1547 /artifacts/security/flatfile directory.

biginsights_users.xml
biginsights_groups.xml

○ PAM with LDAP authentication

□ Kerberos authentication ?

Groups and Users

Specify the groups and users that are defined that you want to access the InfoSphere BigInsights console. Separate multiple groups or users with a comma.

InfoSphere BigInsights System Administrator ?
* Groups: biadmin
Users:

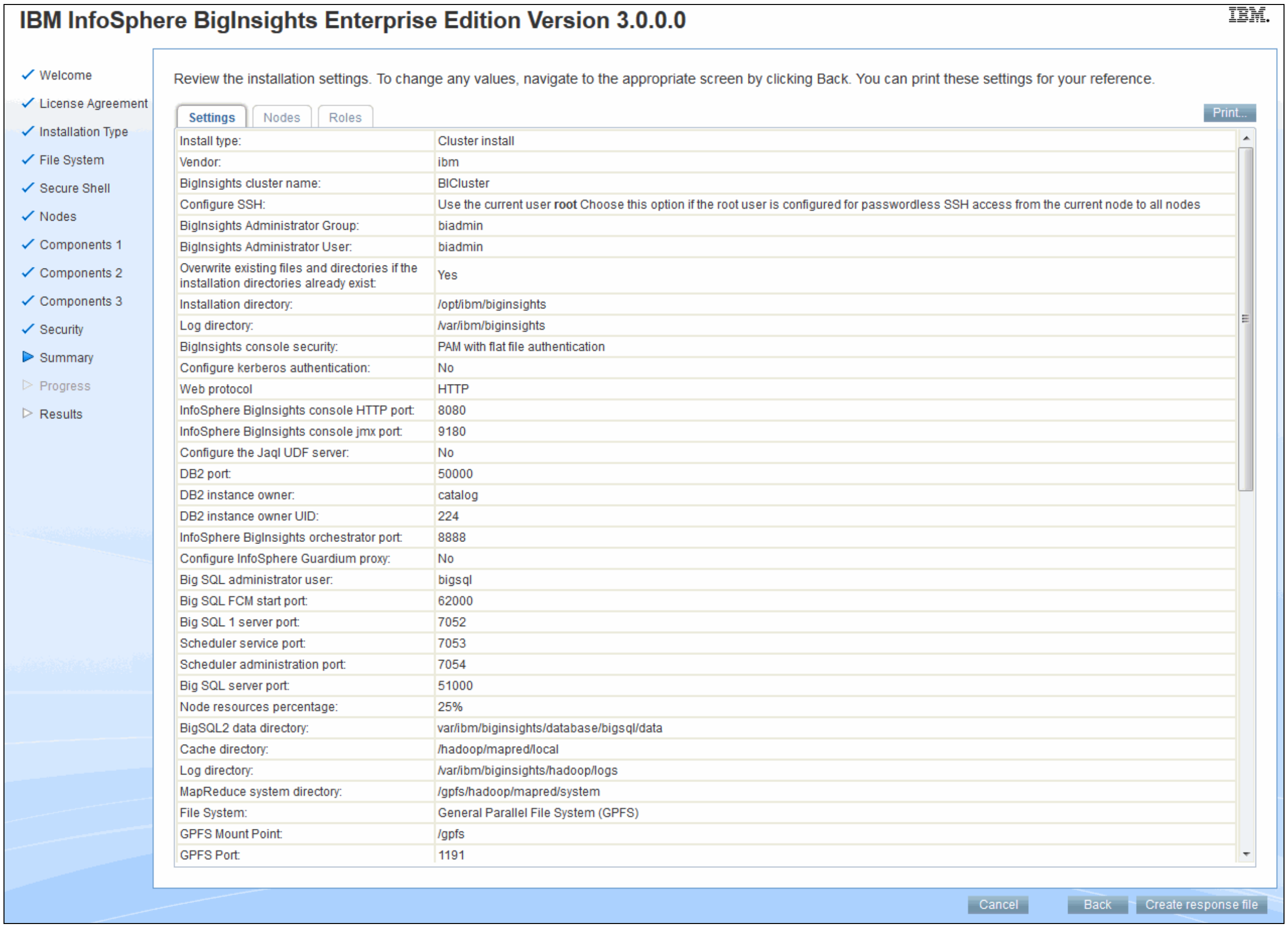
InfoSphere BigInsights Data Administrator ?
* Groups: biadmin
Users:

InfoSphere BigInsights Application Administrator ?
* Groups: biadmin
Users:

InfoSphere BigInsights User ?
* Groups: biadmin
Users:

Figure 3-24 User authentication

Figure 3-25 shows the summary of the installation settings.

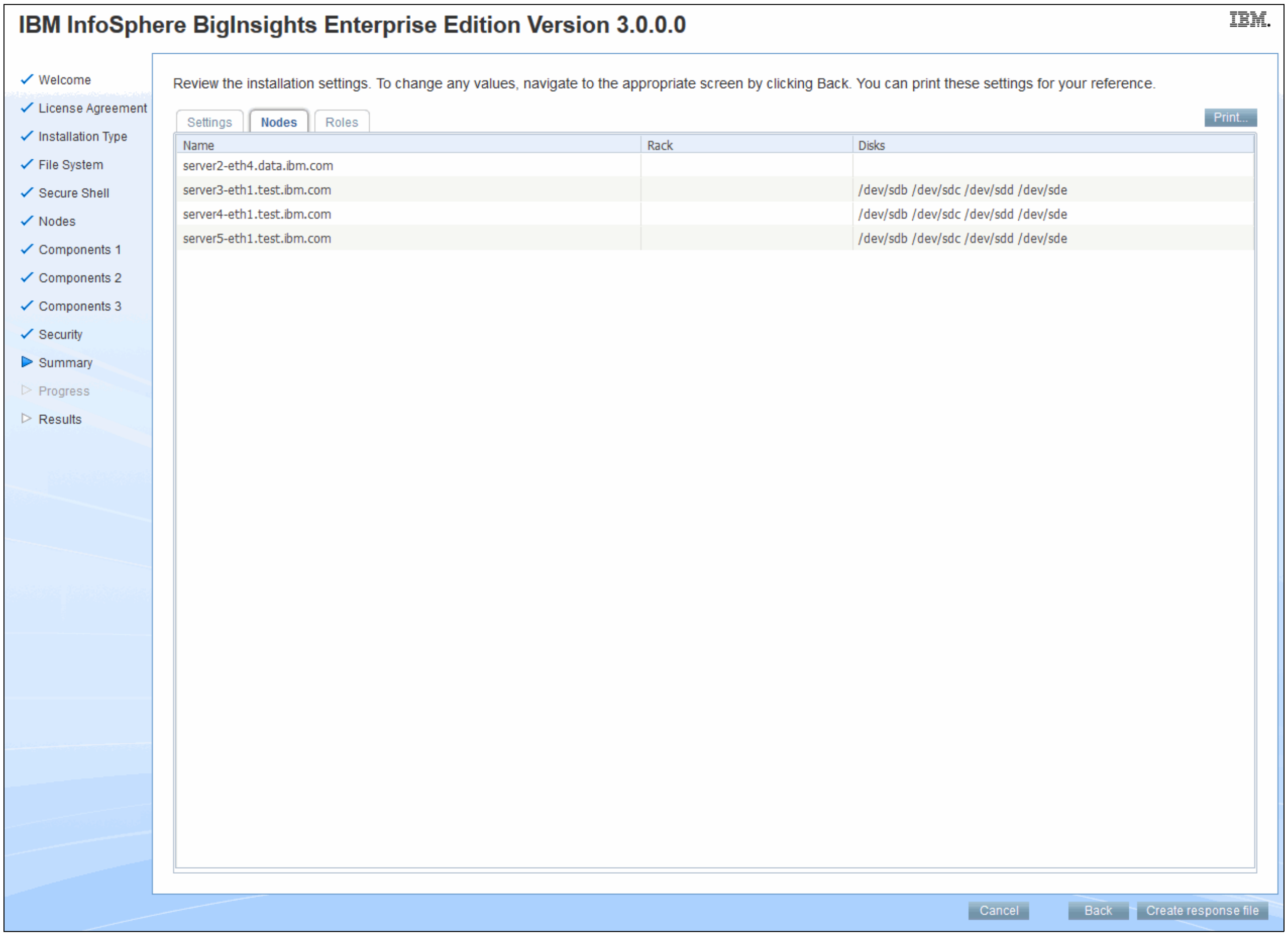


Review the installation settings. To change any values, navigate to the appropriate screen by clicking Back. You can print these settings for your reference.

Install type:	Cluster install
Vendor:	ibm
BigInsights cluster name:	BICluster
Configure SSH:	Use the current user **root** Choose this option if the root user is configured for passwordless SSH access from the current node to all nodes
BigInsights Administrator Group:	biadmin
BigInsights Administrator User:	biadmin
Overwrite existing files and directories if the installation directories already exist:	Yes
Installation directory:	/opt/ibm/biginsights
Log directory:	/var/ibm/biginsights
BigInsights console security:	PAM with flat file authentication
Configure kerberos authentication:	No
Web protocol	HTTP
InfoSphere BigInsights console HTTP port:	8080
InfoSphere BigInsights console jmx port:	9180
Configure the Jaql UDF server:	No
DB2 port:	50000
DB2 instance owner:	catalog
DB2 instance owner UID:	224
InfoSphere BigInsights orchestrator port:	8888
Configure InfoSphere Guardium proxy:	No
Big SQL administrator user:	bigsql
Big SQL FCM start port:	62000
Big SQL 1 server port:	7052
Scheduler service port:	7053
Scheduler administration port:	7054
Big SQL server port:	51000
Node resources percentage:	25%
BigSQL2 data directory:	var/ibm/biginsights/database/bigsql/data
Cache directory:	/hadoop/mapred/local
Log directory:	/var/ibm/biginsights/hadoop/logs
MapReduce system directory:	/gpfs/hadoop/mapred/system
File System:	General Parallel File System (GPFS)
GPFS Mount Point:	/gpfs
GPFS Port:	1191

Figure 3-25 Summary of the installation settings

Figure 3-26 shows the summary of the node configuration.



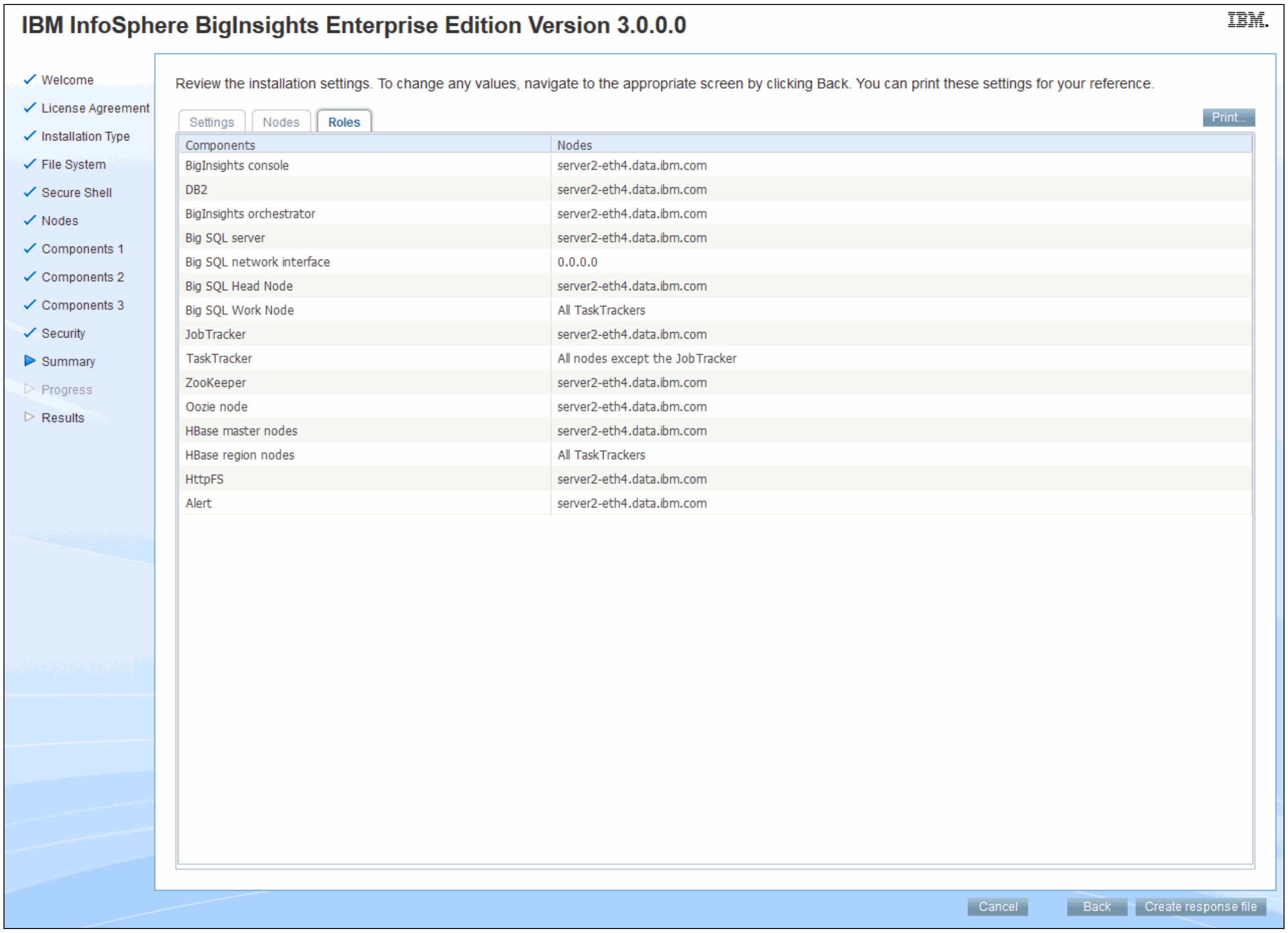
IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

Review the installation settings. To change any values, navigate to the appropriate screen by clicking Back. You can print these settings for your reference.

Name	Rack	Disks
server2-eth4.data.ibm.com		
server3-eth1.test.ibm.com		/dev/sdb /dev/sdc /dev/sdd /dev/sde
server4-eth1.test.ibm.com		/dev/sdb /dev/sdc /dev/sdd /dev/sde
server5-eth1.test.ibm.com		/dev/sdb /dev/sdc /dev/sdd /dev/sde

Figure 3-26 Summary of the node configuration

Figure 3-27 shows the distribution of the roles across the nodes.



Review the installation settings. To change any values, navigate to the appropriate screen by clicking Back. You can print these settings for your reference.

Components	Nodes
BigInsights console	server2-eth4.data.ibm.com
DB2	server2-eth4.data.ibm.com
BigInsights orchestrator	server2-eth4.data.ibm.com
Big SQL server	server2-eth4.data.ibm.com
Big SQL network interface	0.0.0.0
Big SQL Head Node	server2-eth4.data.ibm.com
Big SQL Work Node	All TaskTrackers
JobTracker	server2-eth4.data.ibm.com
TaskTracker	All nodes except the JobTracker
ZooKeeper	server2-eth4.data.ibm.com
Oozie node	server2-eth4.data.ibm.com
HBase master nodes	server2-eth4.data.ibm.com
HBase region nodes	All TaskTrackers
HttpFS	server2-eth4.data.ibm.com
Alert	server2-eth4.data.ibm.com

Figure 3-27   Summary of the distribution of services across nodes

The response file can now be generated by clicking **Create response file**. The file is called `fullinstall.xml` and is saved in the extracted InfoSphere BigInsights directory.

Now, you can restart the installer and instead of generating a new response file, you can choose to use a previously generated one. Taking this action guides you through a sequence of installation windows, at the end of which you can choose to perform the installation. Alternatively, this response file can be used as a parameter to the **`silent-installer.sh`** script, as shown in Example 3-18.

Example 3-18   Silent installation by using a previously generated file as input

```
[root@mgmt01 biginsights-3.0.0.1-Linux-ppc64-b20140711_1547]#
./silent-install/silent-install.sh fullinstall.xml
```

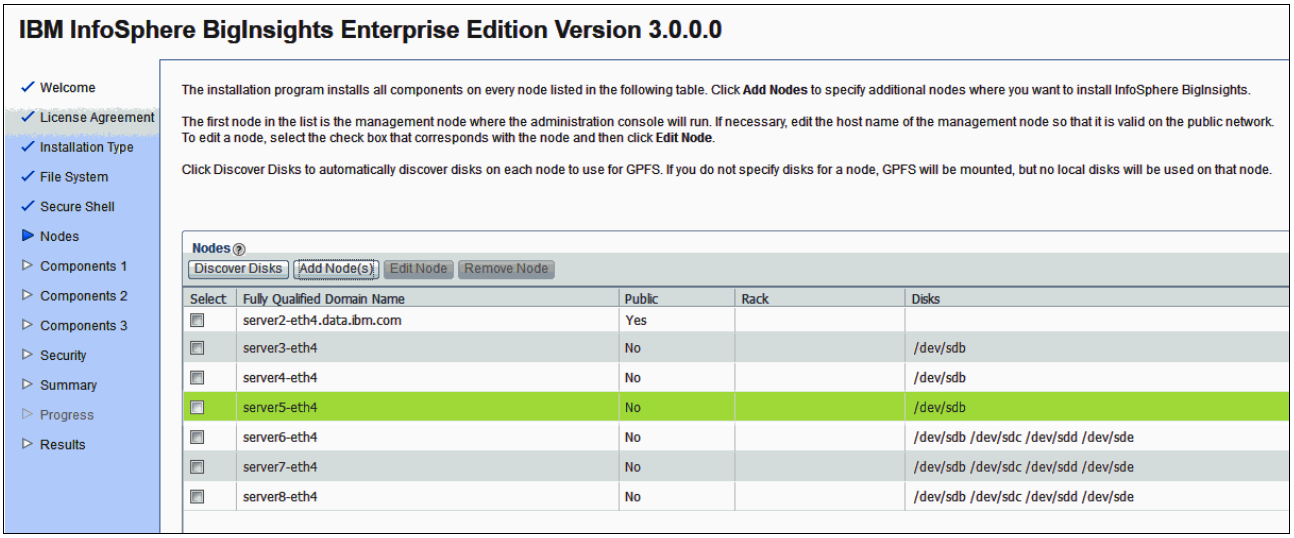
HA installation

The minimum number of nodes for the HA installation of InfoSphere BigInsights is seven: four management nodes and three compute nodes. Three out of the four management nodes are used for the JobTracker. The fourth management node runs all other services (including BigSQL). However, this configuration is suboptimal. Ideally, a fifth management node is available and solely dedicated to BigSQL.

This scenario shows the seven-node configuration.

Note: To configure HA, you *must* enable Adaptive MapReduce.

Figure 3-28 shows all seven servers and their disk requirements. The first node is a management node and runs all the services except the JobTracker. The next three nodes are purely for the HA Job Tracker. The last three nodes are compute nodes. Note their different storage requirements.



IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

The installation program installs all components on every node listed in the following table. Click **Add Nodes** to specify additional nodes where you want to install InfoSphere BigInsights.

The first node in the list is the management node where the administration console will run. If necessary, edit the host name of the management node so that it is valid on the public network. To edit a node, select the check box that corresponds with the node and then click **Edit Node**.

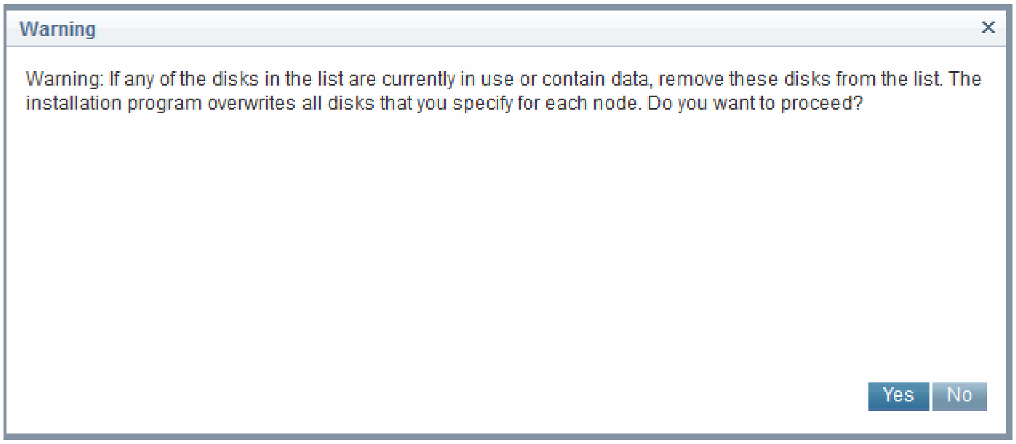
Click Discover Disks to automatically discover disks on each node to use for GPFS. If you do not specify disks for a node, GPFS will be mounted, but no local disks will be used on that node.

Nodes

Select	Fully Qualified Domain Name	Public	Rack	Disks
□	server2-eth4.data.ibm.com	Yes		
□	server3-eth4	No		/dev/sdb
□	server4-eth4	No		/dev/sdb
□	server5-eth4	No		/dev/sdb
□	server6-eth4	No		/dev/sdb /dev/sdc /dev/sdd /dev/sde
□	server7-eth4	No		/dev/sdb /dev/sdc /dev/sdd /dev/sde
□	server8-eth4	No		/dev/sdb /dev/sdc /dev/sdd /dev/sde

Figure 3-28 Server and disk selection window

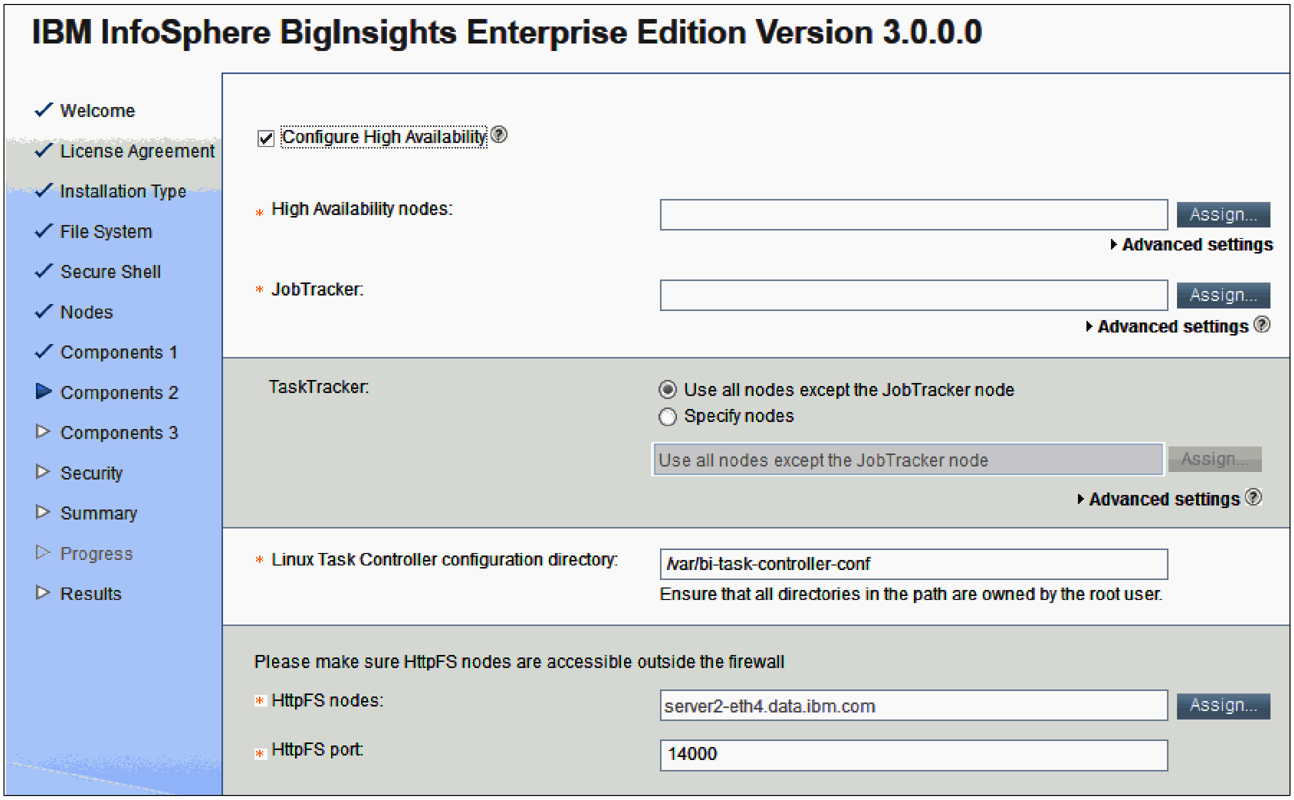
Figure 3-29 shows the data loss warning.



Warning: If any of the disks in the list are currently in use or contain data, remove these disks from the list. The installation program overwrites all disks that you specify for each node. Do you want to proceed?

Figure 3-29 Warning about data on disks

Select **Configure High Availability**, as shown in Figure 3-30.



☑ Configure High Availability

* High Availability nodes:

* JobTracker:

TaskTracker: ● Use all nodes except the JobTracker node ○ Specify nodes

Use all nodes except the JobTracker node

* Linux Task Controller configuration directory: /var/bi-task-controller-conf

Ensure that all directories in the path are owned by the root user.

Please make sure HttpFS nodes are accessible outside the firewall

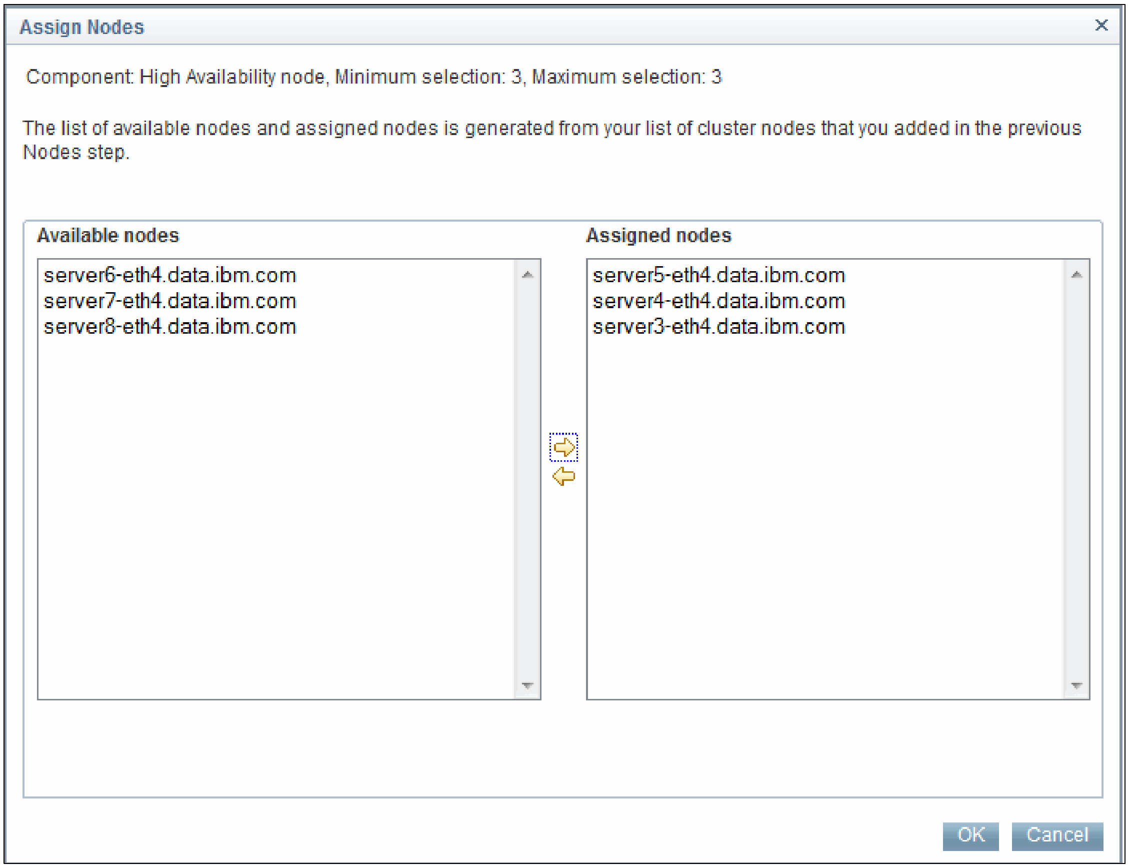
* HttpFS nodes: server2-eth4.data.ibm.com

* HttpFS port: 14000

Figure 3-30 Configure high availability

Select the nodes to be used for HA, as shown in Figure 3-31.

.



Assign Nodes

Component: High Availability node, Minimum selection: 3, Maximum selection: 3

The list of available nodes and assigned nodes is generated from your list of cluster nodes that you added in the previous Nodes step.

Available nodes

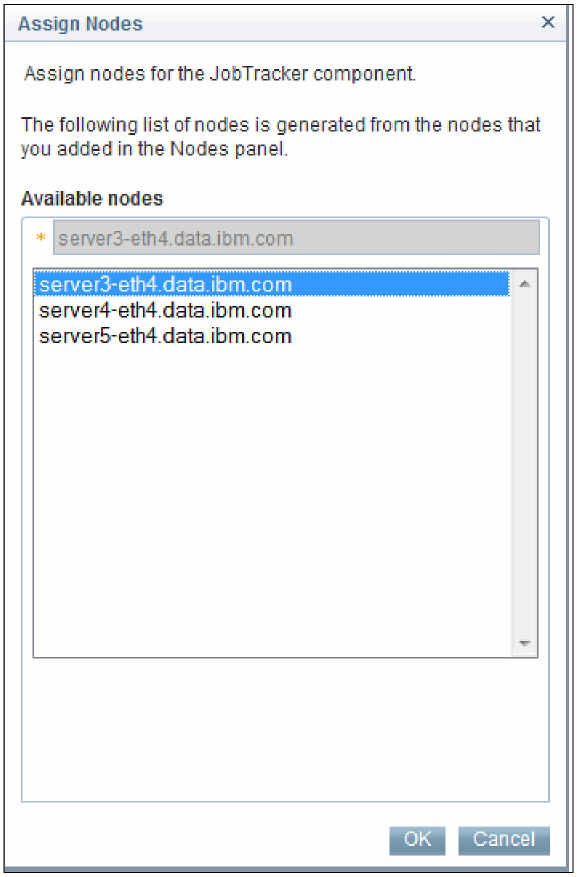
server6-eth4.data.ibm.com
server7-eth4.data.ibm.com
server8-eth4.data.ibm.com

Assigned nodes

server5-eth4.data.ibm.com
server4-eth4.data.ibm.com
server3-eth4.data.ibm.com

Figure 3-31 Select HA nodes

Assign the JobTracker component, as shown in Figure 3-32.



Assign nodes for the JobTracker component.

The following list of nodes is generated from the nodes that you added in the Nodes panel.

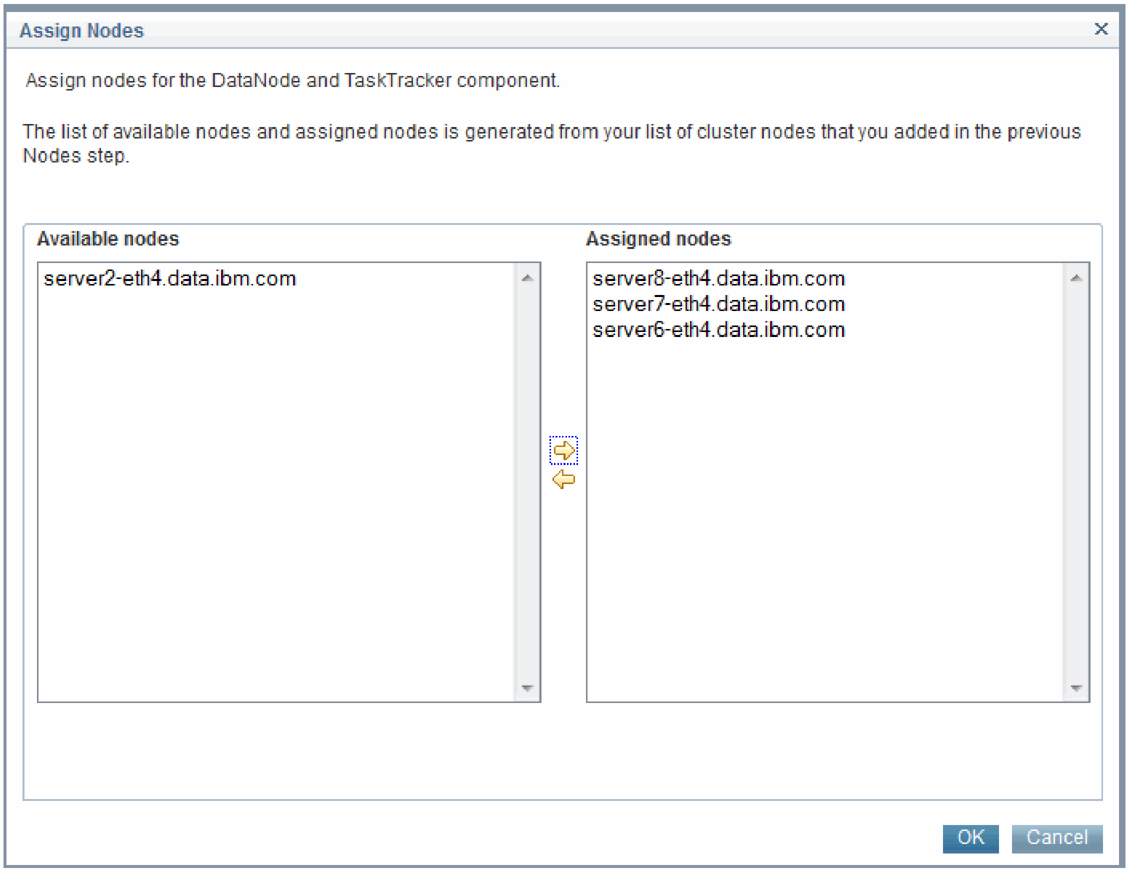
Available nodes

* server3-eth4.data.ibm.com

server3-eth4.data.ibm.com
server4-eth4.data.ibm.com
server5-eth4.data.ibm.com

Figure 3-32 Assign the JobTracker

Figure 3-33 shows how to assign the data and TaskTracker nodes.



Assign nodes for the DataNode and TaskTracker component.

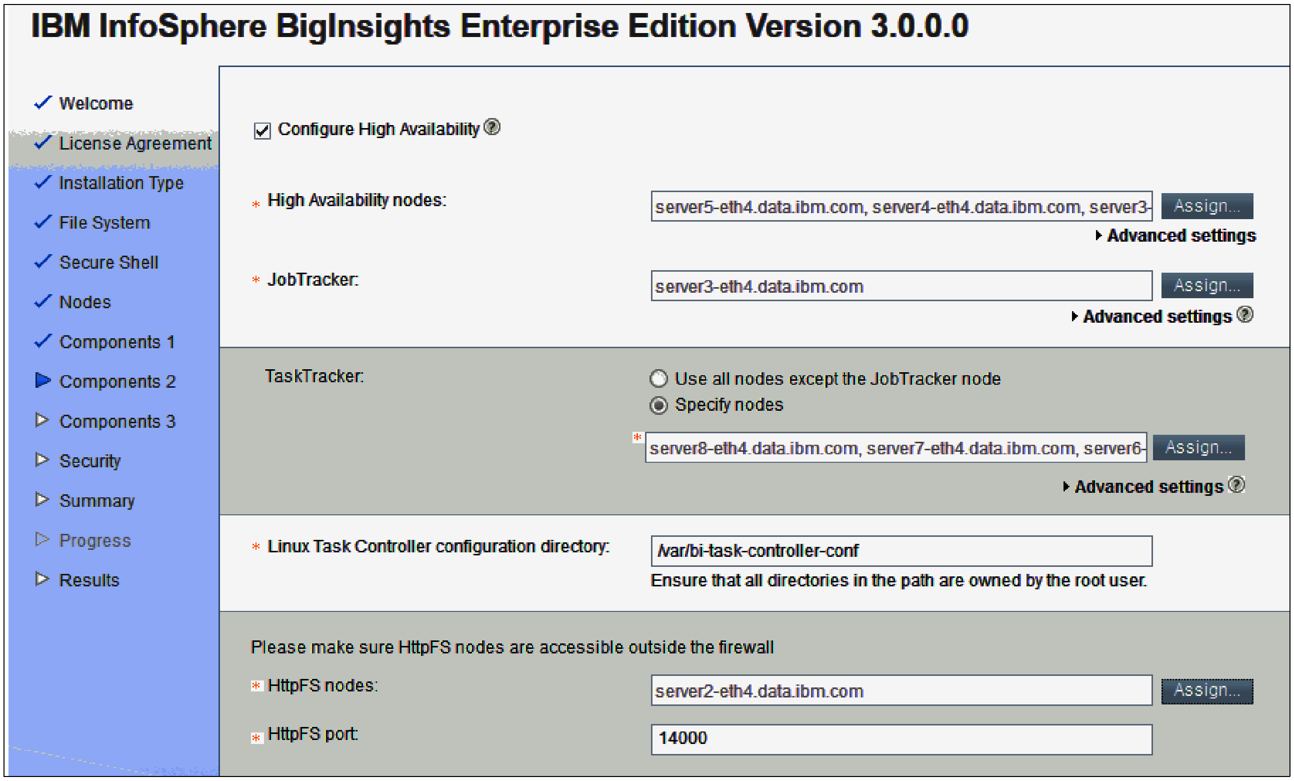
The list of available nodes and assigned nodes is generated from your list of cluster nodes that you added in the previous Nodes step.

server2-eth4.data.ibm.com

server8-eth4.data.ibm.com
server7-eth4.data.ibm.com
server6-eth4.data.ibm.com

Figure 3-33 Data and TaskTracker node selection window

Figure 3-34 shows the window after the selections are made.



IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

☑ Configure High Availability

* High Availability nodes: server5-eth4.data.ibm.com, server4-eth4.data.ibm.com, server3- Assign...

Advanced settings

* JobTracker: server3-eth4.data.ibm.com Assign...

Advanced settings

TaskTracker:

○ Use all nodes except the JobTracker node

● Specify nodes

* server8-eth4.data.ibm.com, server7-eth4.data.ibm.com, server6- Assign...

Advanced settings

* Linux Task Controller configuration directory: /var/bi-task-controller-conf

Ensure that all directories in the path are owned by the root user.

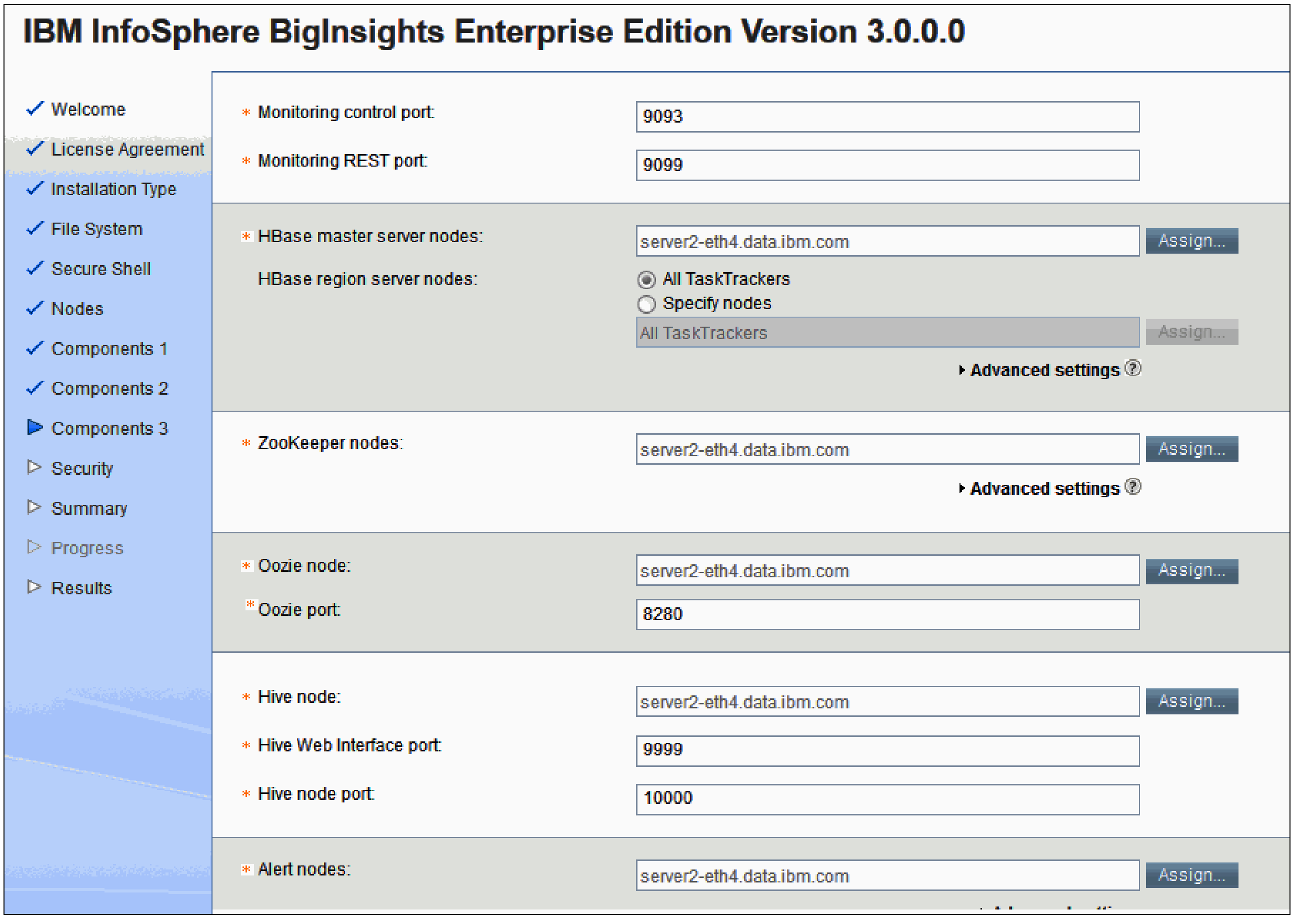
Please make sure HttpFS nodes are accessible outside the firewall

* HttpFS nodes: server2-eth4.data.ibm.com Assign...

* HttpFS port: 14000

Figure 3-34 The high availability nodes section is now populated with the earlier selections

Figure 3-35 shows the configuration of Monitoring, HBase, ZooKeeper, Oozie, Hive, and the Alert nodes.



IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

* Monitoring control port: 9093

* Monitoring REST port: 9099

* HBase master server nodes: server2-eth4.data.ibm.com

HBase region server nodes: ● All TaskTrackers ○ Specify nodes

All TaskTrackers

* ZooKeeper nodes: server2-eth4.data.ibm.com

* Oozie node: server2-eth4.data.ibm.com

* Oozie port: 8280

* Hive node: server2-eth4.data.ibm.com

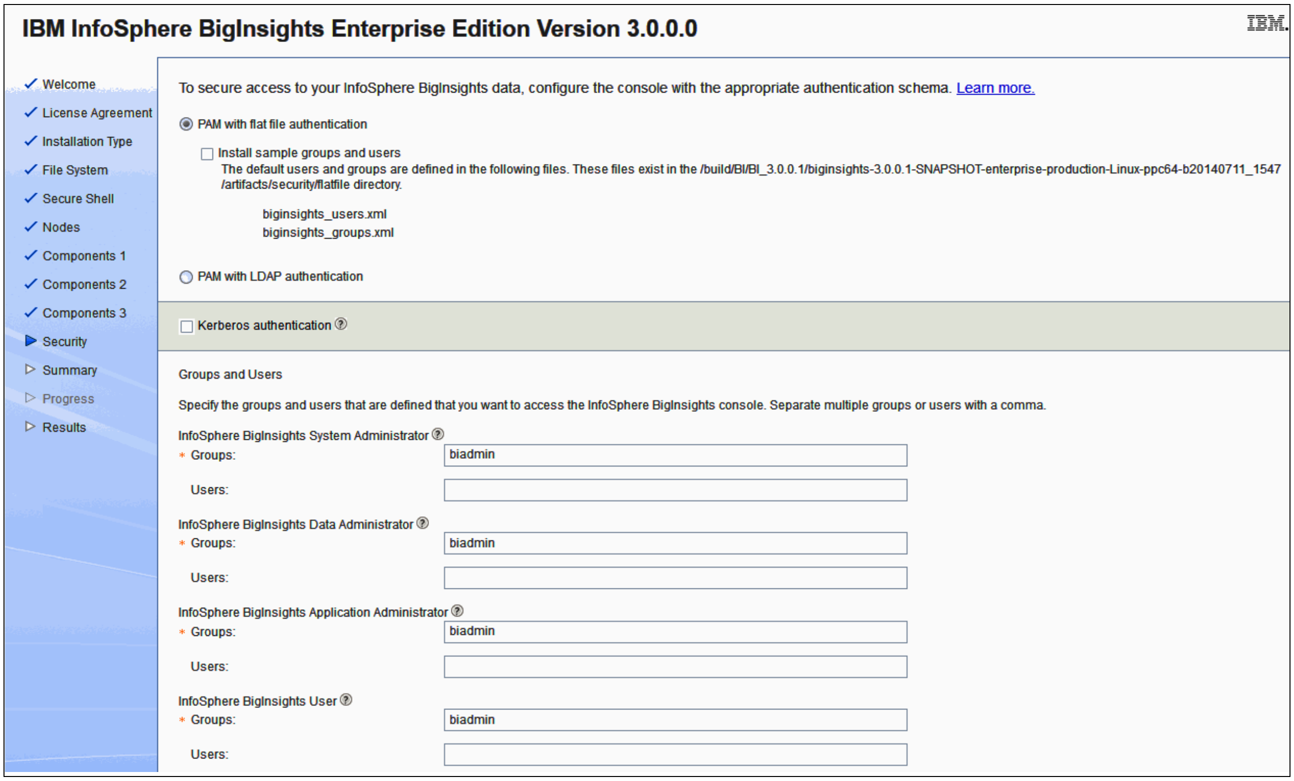
* Hive Web Interface port: 9999

* Hive node port: 10000

* Alert nodes: server2-eth4.data.ibm.com

Figure 3-35 Configuration options for Component 3

Configure the authentication method for users, as shown in Figure 3-36.



IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

To secure access to your InfoSphere BigInsights data, configure the console with the appropriate authentication schema. Learn more.

● PAM with flat file authentication

□ Install sample groups and users
The default users and groups are defined in the following files. These files exist in the /build/BI/BI_3.0.0.1/biginsights-3.0.0.1-SNAPSHOT-enterprise-production-Linux-ppc64-b20140711_1547 /artifacts/security/flatfile directory.

biginsights_users.xml
biginsights_groups.xml

○ PAM with LDAP authentication

□ Kerberos authentication ?

Groups and Users

Specify the groups and users that are defined that you want to access the InfoSphere BigInsights console. Separate multiple groups or users with a comma.

InfoSphere BigInsights System Administrator ?
* Groups: biadmin
Users:

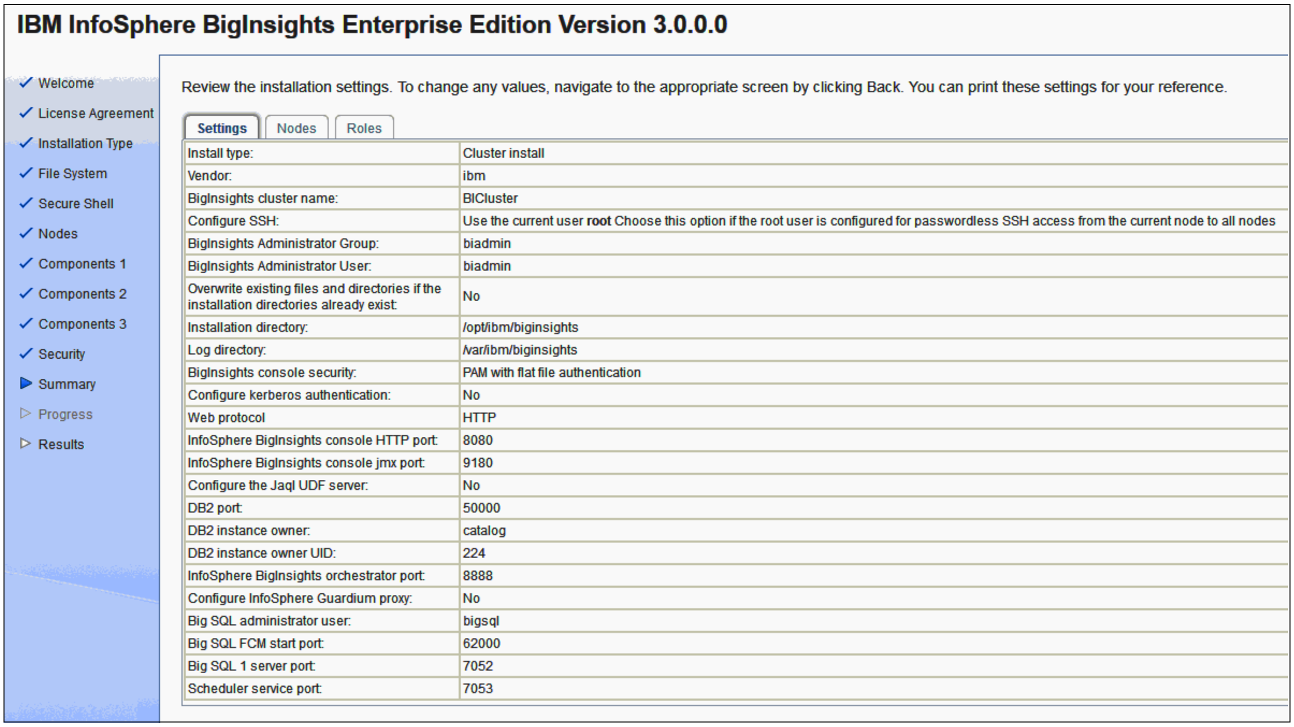
InfoSphere BigInsights Data Administrator ?
* Groups: biadmin
Users:

InfoSphere BigInsights Application Administrator ?
* Groups: biadmin
Users:

InfoSphere BigInsights User ?
* Groups: biadmin
Users:

Figure 3-36 User authentication

Figure 3-37 shows the summary of the installation settings.

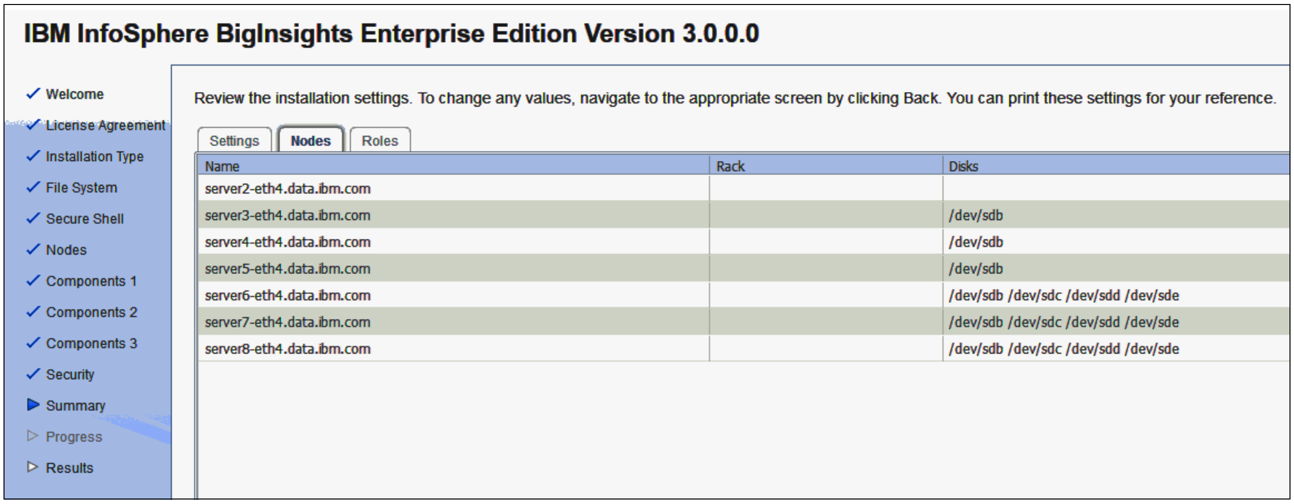

IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

Review the installation settings. To change any values, navigate to the appropriate screen by clicking Back. You can print these settings for your reference.

Install type:	Cluster install
Vendor:	ibm
BigInsights cluster name:	BICluster
Configure SSH:	Use the current user **root** Choose this option if the root user is configured for passwordless SSH access from the current node to all nodes
BigInsights Administrator Group:	biadmin
BigInsights Administrator User:	biadmin
Overwrite existing files and directories if the installation directories already exist:	No
Installation directory:	/opt/ibm/biginsights
Log directory:	/var/ibm/biginsights
BigInsights console security:	PAM with flat file authentication
Configure kerberos authentication:	No
Web protocol	HTTP
InfoSphere BigInsights console HTTP port:	8080
InfoSphere BigInsights console jmx port:	9180
Configure the Jaql UDF server:	No
DB2 port:	50000
DB2 instance owner:	catalog
DB2 instance owner UID:	224
InfoSphere BigInsights orchestrator port:	8888
Configure InfoSphere Guardium proxy:	No
Big SQL administrator user:	bigsql
Big SQL FCM start port:	62000
Big SQL 1 server port:	7052
Scheduler service port:	7053

Figure 3-37 Summary of the installation settings

Figure 3-38 shows the summary of the node configuration.

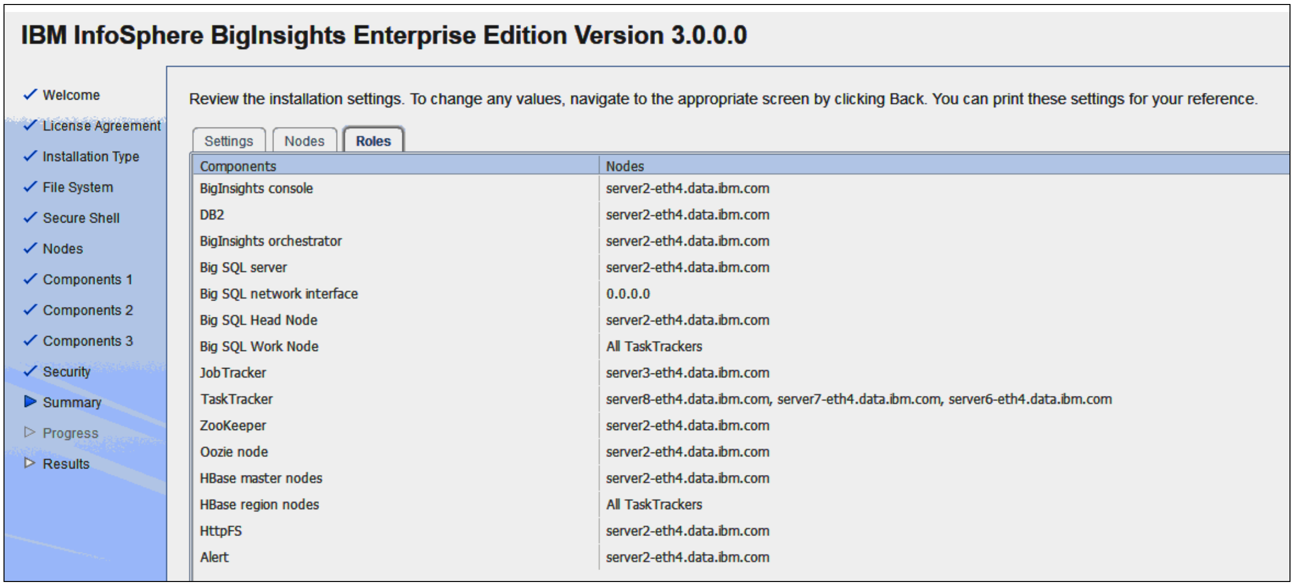

IBM InfoSphere BigInsights Enterprise Edition Version 3.0.0.0

Review the installation settings. To change any values, navigate to the appropriate screen by clicking Back. You can print these settings for your reference.

Name	Rack	Disks
server2-eth4.data.ibm.com		
server3-eth4.data.ibm.com		/dev/sdb
server4-eth4.data.ibm.com		/dev/sdb
server5-eth4.data.ibm.com		/dev/sdb
server6-eth4.data.ibm.com		/dev/sdb /dev/sdc /dev/sdd /dev/sde
server7-eth4.data.ibm.com		/dev/sdb /dev/sdc /dev/sdd /dev/sde
server8-eth4.data.ibm.com		/dev/sdb /dev/sdc /dev/sdd /dev/sde

Figure 3-38 Summary of the node configuration

Figure 3-39 shows the distribution of the roles across the nodes.

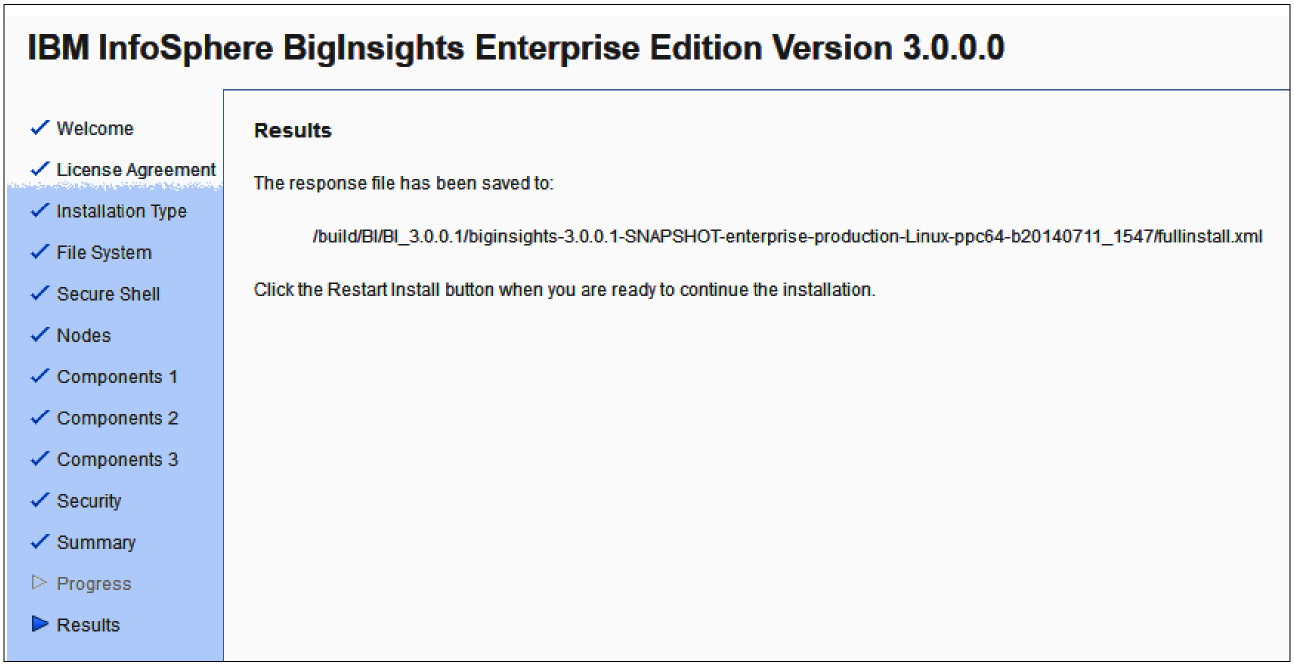


Review the installation settings. To change any values, navigate to the appropriate screen by clicking Back. You can print these settings for your reference.

Components	Nodes
BigInsights console	server2-eth4.data.ibm.com
DB2	server2-eth4.data.ibm.com
BigInsights orchestrator	server2-eth4.data.ibm.com
Big SQL server	server2-eth4.data.ibm.com
Big SQL network interface	0.0.0.0
Big SQL Head Node	server2-eth4.data.ibm.com
Big SQL Work Node	All TaskTrackers
JobTracker	server3-eth4.data.ibm.com
TaskTracker	server8-eth4.data.ibm.com, server7-eth4.data.ibm.com, server6-eth4.data.ibm.com
ZooKeeper	server2-eth4.data.ibm.com
Oozie node	server2-eth4.data.ibm.com
HBase master nodes	server2-eth4.data.ibm.com
HBase region nodes	All TaskTrackers
HttpFS	server2-eth4.data.ibm.com
Alert	server2-eth4.data.ibm.com

Figure 3-39 Summary of the distribution of services across nodes

Figure 3-40 shows the location of the response file.



Results

The response file has been saved to:

/build/BI/BI_3.0.0.1/biginsights-3.0.0.1-SNAPSHOT-enterprise-production-Linux-ppc64-b20140711_1547/fullinstall.xml

Click the Restart Install button when you are ready to continue the installation.

Figure 3-40 Location of the response file

3.3 Platform Cluster Manager - Advanced Edition

Platform Cluster Manager - Advanced Edition can be used to automate operating system installation and the prerequisites setup for InfoSphere BigInsights in all nodes, as described in 3.1.2, “Operating system prerequisites setup for InfoSphere BigInsights” on page 39, and then use these nodes to install InfoSphere BigInsights, as described in 3.2, “InfoSphere BigInsights” on page 45. Also, Platform Cluster Manager - Advanced Edition makes it possible to automate an InfoSphere BigInsights installation in a cluster of nodes by using a *cluster template*.

This section describes how to install and set up Platform Cluster Manager - Advanced Edition, automate the operating system installation, and perform the prerequisites setup for InfoSphere BigInsights. To automate an InfoSphere BigInsights installation by using a cluster template, see 6.2, “Managing InfoSphere BigInsights cluster nodes within Platform Cluster Manager - Advanced Edition” on page 139.

3.3.1 Planning a system configuration

This section gives you an overview of the hardware, software, and networking requirements for a Platform Cluster Manager - Advanced Edition installation. Platform Cluster Manager - Advanced Edition can be installed into an LPAR, which becomes the Platform Cluster Manager - Advanced Edition master node, or management node.

Hardware requirements

Here are the minimum hardware requirements for the management node:

- 100 GB of free disk space
- 4 GB of memory (RAM)
- At least one statically configured Ethernet interface

Note: If you use IBM PureFlex® System, the management node must be a node that is not in the IBM Flex Chassis.

Here are the minimum requirements for compute nodes for a stateful package-based installation (or performing a standard operating system installation on top of which you can install InfoSphere BigInsights):

- 1 GB of physical memory (RAM)
- 40 GB of free disk space
- One statically configured Ethernet interface

Here are the minimum requirements for a compute node stateless image-based installation (or bootable rescue disk):

- 4 GB of physical memory (RAM)
- One statically configured Ethernet interface

Networking requirements

At a minimum, you must set up the physical network configuration for a public network, a provisioning network, and an FSP network within the Platform Cluster Manager - Advanced Edition cluster environment.

More optional network interfaces can be configured based on your needs:

- An extra Ethernet interface for application networking
- An extra Ethernet interface for other public networks
- More interconnects for high-performance message passing, such as 10 Gb or InfiniBand
- More interconnects for high-performance application data, such as 10 Gb or InfiniBand

Because the idea is to deploy an InfoSphere BigInsights cluster within the Platform Cluster Manager - Advanced Edition cluster in the lab environment, four networks are defined: provisioning, public, FSP, and data. For more information about this architecture, see “Networking” on page 18.

Software requirements

You have two choices for the operating system that you use with your Platform Cluster Manager - Advanced Edition management node:

- Red Hat Enterprise Linux 6.5 ppc64
- SUSE Linux Enterprise Server (SLES) 11.3 ppc64

The version that is used in this book is RHEL 6.5 ppc64.

Operating system prerequisites

Before you install Platform Cluster Manager - Advanced Edition on your management node, ensure that following conditions are met:

1. Decide on a partitioning layout. Here is the preferred partitioning layout:
   - Ensure that the `/opt` partition has a least 4 GB.
   - Ensure that the `/var` partition has at least 40 GB.
   - Ensure that the `/install` partition has at least 40 GB.
2. Configure at least one static network interface.
3. Use a fully qualified domain name (FQDN) for the management node.
4. The `/home` directory must be writable.

   If the /home directory is mounted by AutoFS, you must first disable the AutoFS configuration by running the following commands:

   ```
   # chkconfig autofs off
   # service autofs stop
   ```

   To make the `/home` directory writable, run the following commands as root:

   ```
   # chmod u+w /home
   # ls -al / |grep home
   ```
5. The package `openais-devel` must be removed manually if it is already installed.
6. Before you install Platform Cluster Manager - Advanced Edition on the management node, check that the shadow passwords authentication method is enabled. Run **`setup`** in a shell and make sure that the **Use Shadow Passwords** check box is selected under the Authentication Configuration menu item.
7. Ensure that IPv6 is enabled for remote power and console management. Do not disable IPv6 during the operating system installation. To enable IPv6, complete the following steps:

   For RHEL: If the `disable-ipv6.conf` file exists in the `/etc/modprobe.d` directory, comment out the following line to disable IPv6:

   ```
   install ipv6 /bin/true
   ```

8. The `70-persistent-net.rules` file is created under `/etc/udev/rules.d` to make the names persistent across restarts.
9. Before installing Platform Cluster Manager - Advanced Edition, you must stop the NetworkManager service. Run the following command:

   ```
   /etc/init.d/NetworkManager stop
   ```

10. Disable SELinux by completing the following steps:
    a. On the management node, edit the /etc/selinux/config file to set `SELINUX=disabled`.
    a. Restart the management node.
11. Ensure that the traditional naming scheme *ethN* is used for the Ethernet interface. If you have a system that does not use the traditional naming scheme ethN, you must revert to the traditional naming scheme ethN by completing the following steps:
    a. Rename all `ifcfg-ethN` and `ifcfg-p*` configuration files and modify the contents of the files accordingly. The content of these files is distribution-specific (for details, see `/usr/share/doc/initscripts-version`). For example, the `ifcfg-ethN` files in RHEL 6.x contain a `DEVICE=` field that is assigned with the ethN name. Modify it to suit the new naming scheme, such as `DEVICE=eth0`.
    b. Comment the *HWADDR* variable in the `ifcfg-eth*` files if it is present because it is not possible to predict here which of the network devices is named eth0, eth1, and so on.
    c. Restart the system.
    d. Log in to check the correctness of the ethN interfaces.

To configure a high availability environment for the Platform Cluster Manager - Advanced Edition management node, go to this website:

http://www-01.ibm.com/support/knowledgecenter/SSDV85_4.2.0/pcm_welcome.html

Planning your network configuration

As described in 2.2.1, “General architecture” on page 17, Platform Cluster Manager - Advanced Edition needs at least three networks: public, provisioning, and FSP.

The public network is the one that you use to access the Platform Cluster Manager - Advanced Edition management node. You can have the same public network with the cluster nodes, or you can differentiate between these two networks. If you have a multi-cluster environment with different users, each cluster should have its own distinct public network.

The provision network is used by the Platform Cluster Manager - Advanced Edition management node to communicate with the cluster nodes to perform operating system installation, synchronize files, and perform preinstallation and postinstallation scripts, monitoring, and message passing. You can use a high-speed 10 Gb network or InfiniBand to have higher throughput and lower latency for large cluster environments. The FSP network is used by Platform Cluster Manager to have control over Power Systems hardware for powering on and off the LPARs, powering on and off the servers itself, and creating LPAR configurations.

The following information is required to set up and configure your Platform Cluster Manager - Advanced Edition network:

- Provisioning network information:
  - Network subnet
  - Network domain name
  - Static IP address range

- Public network information:
  - Network subnet
  - Network domain name
  - Static IP address range
- FSP network information:
  - Network subnet
  - Network domain name
  - Static IP address range
- Management node information:
  - Node name (use a fully qualified domain name with a public domain suffix, for example: `management.domain.com`)
  - Static IP address and subnet mask for public network
  - Static IP address and subnet mask for provision network
  - Default gateway address
  - External DNS server IP address

You can use the information in Table 3-3 while installing your Platform Cluster Manager - Advanced Edition environment. The values that are outlined here are the ones that were used in our lab installation.

Table 3-3 Network configuration details

Network	Provisioning	Public	FSP
Domain	`cluster.ibm.com`	test.ibm.com	`fspcluster.ibm.com`
Management interface	`eth0`	`eth1`	`eth2`
Gateway	`<xcatmaster>`	`N/A`	`<xcatmaster>`
Subnet mask	`255.255.255.0`	`255.255.255.0`	`255.255.255.0`

Static range	`192.168.0.101-`
	`192.168.0.199`
		`10.0.0.101-`
		`10.0.0.199`
			`192.168.1.102-`
			`192.168.1.199`

Dynamic range	`192.168.0.201-`
	`192.168.0.254`
		`10.0.0.201-`
		`10.0.0.254`
			`192.168.1.201-`
			`192.168.1.254`

Static range increment	`1`	`1`	`1`

3.3.2 Performing the installation

This section provides a walkthrough for a Platform Cluster Manager - Advanced Edition installation according to the specifications that are outlined in the previous sections.

Preparing for a Platform Cluster Manager - Advanced Edition installation

There are some prerequisite steps and information that you should determine before performing a Platform Cluster Manager - Advanced Edition installation:

1. Download the Platform Cluster Manager ISO and entitlement file from IBM Entitlement Software Downloads. You can contact an IBM Representative to help you procure the software. Then, copy these files into your management node.
2. Summarize the Platform Cluster Manager - Advanced Edition installation settings information. Table 3-4 shows the values that were used in our lab tests. Table 3-4 provides a column for you to enter the values that are used during your installation.

Table 3-4 Platform Cluster Manager - Advanced Edition installation information

No	Option	Lab values	Your values
1.	Platform Cluster Manager ISO path.	`/root/pcmsource/pcm-4.2-140913-193148.ppc64.iso`	
2.	Product entitlement path.	`/root/pcmsource/pcm_adv_entitlement.dat`	
3.	Select a mount point for the depot (/install) directory.	`/`	
4.	Select from where to install the operating system.	ISO image or mount point	
5.	Operating system ISO image path for compute node installation.	`/mnt/RHEL65ppc64.iso`	
6.	Specify a provision network interface.	`eth0`	
7.	IP address range that is used for provisioning the compute nodes.	`192.168.0.102` - `192.168.0.200`	
8.	Temporary IP address range that is used for provisioning compute nodes by node discovery. This range cannot overlap the range that is specified for the provisioning compute nodes.	`192.168.0.200` - `192.168.0.254`	
9.	Specify a public network interface.	`eth1`	
10.	Domain name for the provision network.	`cluster.ibm.com`	
11.	Domain name for the public network	`test.ibm.com`	
12.1	DNS server.	N/A	

No	Option	Lab values	Your values
13.1	NTP server.	`10.0.0.22`	
14.1	Platform Cluster Manager database administrator password.	`pcmdbadm`	
15.1	Root account password for all compute nodes.	`Cluster`	

Installation: A step-by-step approach

Use the information in Table 3-4 on page 77 for your installation settings. The values inside [ ] throughout the installation wizard mean the default value if you do not specify any value.

There are four steps for the process of installing Platform Cluster Manager - Advanced Edition:

1. License agreement.
2. Management node pre-check.
3. Specify installation settings.
4. Installation.

Installation procedure

Complete the following steps:

1. Run the installer as the root user.
2. Go to the directory where you have placed the Platform Cluster Manager - Advanced Edition ISO, then mount the Platform Cluster Manager installation media to your installation path:

```
# cd /root/pcmsource/pcmiso
# mkdir /root/pcmsource/pcmiso
# mount -o loop pcm-4.2-140913-193148.ppc64.iso pcmiso
```

3. Start the Platform Cluster Manager - Advanced Edition installer:

```
[root@pcm pcmiso]# ./pcm-installer
   Preparing to install 'pcm-installer'...[  OK  ]
```

4. Enter the path of your Platform Cluster Manager - Advanced Edition entitlement file:

```
Enter the path to the product entitlement file:
/root/pcmsource/pcm_adv_entitlement.dat
Parsing the product entitlement file...[  OK  ]
   =================================================================

Welcome to the IBM Platform Cluster Manager - Advanced Edition 4.2 Installation

=================================================================
The complete IBM Platform Cluster Manager - Advanced Edition 4.2 installation
includes the following:

   1. License Agreement
   2. Management node pre-checking
   3. Specify installation settings
   4. Installation
Press ENTER to continue the installation or CTRL-C to quit the installation.
```

5. Accept the license agreement and continue:

```
   ================================================================

   Step 1 of 4: License Agreement

   ================================================================

   International Program License Agreement

Part 1 - General Terms

BY DOWNLOADING, INSTALLING, COPYING, ACCESSING, CLICKING ON
AN "ACCEPT" BUTTON, OR OTHERWISE USING THE PROGRAM,
LICENSEE AGREES TO THE TERMS OF THIS AGREEMENT. IF YOU ARE
ACCEPTING THESE TERMS ON BEHALF OF LICENSEE, YOU REPRESENT
AND WARRANT THAT YOU HAVE FULL AUTHORITY TO BIND LICENSEE
TO THESE TERMS. IF YOU DO NOT AGREE TO THESE TERMS,

* DO NOT DOWNLOAD, INSTALL, COPY, ACCESS, CLICK ON AN
"ACCEPT" BUTTON, OR USE THE PROGRAM; AND

* PROMPTLY RETURN THE UNUSED MEDIA, DOCUMENTATION, AND

Press Enter to continue viewing the license agreement, or
enter "1" to accept the agreement, "2" to decline it, "3"
to print it, "4" to read non-IBM terms, or "99" to go back
to the previous screen.
1
```

6. The management node pre-checking automatically starts:

```
   ================================================================

   Step 2 of 4: Management node pre-checking

   ================================================================

   Checking hardware architecture...[  OK  ]

   Checking OS compatibility...[  OK  ]

   Checking free memory...[  OK  ]

   Checking if SELinux is disabled...[  OK  ]

   Checking if Auto Update is disabled...[  OK  ]

   Checking if NetworkManager is disabled...[  OK  ]

   Checking if PostgreSQL is disabled...[  OK  ]

   Checking for DNS service...[  OK  ]

   Checking for DHCP service...[  OK  ]

   Checking for available ports...[  OK  ]
   Checking management node name...[  OK  ]

   Checking static NIC...[  OK  ]

   Probing DNS settings...[OK]

   Probing language and locale settings...[  OK  ]

   Checking home directory (/home) ...[  OK  ]

   Checking mount point for depot (/install) directory...[  OK  ]

   Checking required free disk space for opt directory...[  OK  ]

   Checking required free disk space for var directory...[  OK  ]
```

7. Select the Custom Installation option:

```
   ================================================================

   Step 3 of 4: Specify installation settings

   ================================================================

Select the installation method from the following options:

   1) Quick Installation

   2) Custom Installation

   Enter your selection [1]:  2
```

8. Select a mount point for the depot (`/install`) directory. The depot (`/install`) directory stores installation files for Platform Cluster Manager - Advanced Edition. The Platform Cluster Manager management node checks for the required disk space.

```
Select a mount point for the depot (/install) directory from the following
options:

   1) Mount point: '/' Free space: '493 GB'

   Enter your selection [1]: 1
```

9. Select the location of the operating system image that you want to use for the compute node operating system installation:

```
The OS version must be the same as the OS version on the management node.

   From the following options, select where to install the OS from:

   1) CD/DVD drive

   2) ISO image or mount point

   Enter your selection [1]:  2
Enter the path to the first ISO image or mount point:  /mnt/IBMIT_RHEL65
```

10. Select a network interface for the provisioning network. This step shows your available networks.

```
Select a network interface for the provisioning network from the following
options:

  1) Interface: eth0, IP: 192.168.0.30, Netmask: 255.255.255.0

  2) Interface: eth1, IP: 10.0.0.30, Netmask: 255.255.255.0

  3) Interface: eth2, IP: 192.168.1.30, Netmask: 255.255.255.0

  Enter your selection [1]:  1
```

11. Enter the IP address range that is used for provisioning compute nodes:

```
Enter IP address range used for provisioning compute nodes
[192.168.0.3-192.168.0.200]: 192.168.0.101-192.168.0.199
```

12. Choose whether to provision compute nodes automatically with the node discovery method:

```
Do you want to provision compute nodes with node discovery? (Y/N) [Y]: Y
```

13. Enter a node discovery IP address range to be used for provisioning compute nodes by node discovery. The node discovery IP address range is a temporary IP address range that is used to provision automatically nodes by using the auto node discovery method. This range cannot overlap the range that is specified for the provisioning compute nodes.

```
Enter a temporary IP address range to be used for provisioning compute nodes
   by node discovery. This range cannot overlap the range specified for the
   provisioning compute nodes. [192.168.0.201-192.168.0.254]:
```

14. Select that interface that is used by the management node to connect to the public network, or choose `It is not connected to the public network` if it is not connected to public network.

    If your management node is connected to public network, you can optionally enable the following setting:

    a. Enable Platform Cluster Manager specific rules for the management node firewall that is connected to the public interface.

    b. Enable NAT forwarding on the management node for all compute nodes.

```
The management node is connected to the public network by:

  1) Interface: eth1, IP: 10.0.0.30, Netmask: 255.255.255.0

  2) Interface: eth2, IP: 192.168.1.30, Netmask: 255.255.255.0

  3) It is not connected to the public network

  Enter your selection [1]:  1

Enable Platform Cluster Manager specific rules for the management node firewall
to the public interface? (Y/N) [Y]:

Enable NAT forwarding on the management node for all compute nodes? (Y/N) [Y]:
```

15. Enable an FSP network that uses the default provisioning template. In rack-mounted Power Systems servers, use the FSP network on the management node to communicate with the Power Systems hardware, but not for provisioning the operating system, so do not enable this option. Define the FSP network after installation as described in “Configuring the DHCP service for the FSP network” on page 86.

> **Note:** Do not enable an FSP network that uses the default provisioning template when managing rack-mounted Power Systems servers with Platform Cluster Manager - Advanced Edition.

```
Enable an FSP network that uses the default provisioning template (Y/N) [N]:

   For rack-mounted Power servers, do not enable this option. N
```

16. Enter a domain name for the provisioning network:

```
Enter a domain name for the provisioning network [private.dns.zone]:
cluster.ibm.com
```

17. Set and specify the domain name for the public network:

```
Set a domain name for the public network (Y/N) [Y]:
Enter a domain name for the public network [example.com]:  test.ibm.com
```

18. Enter the IP addresses of your extra name server if you have one:

```
Enter the IP addresses of extra name servers that are separated by commas
[N/A]:
```

19. Set the NTP server. This test uses a local lab NTP server.

```
Enter NTP server [pool.ntp.org]:  10.0.0.22
Synchronizing management node with the time server...[  OK  ]
```

20. Export the home directory on the management node and use it for all compute nodes:

```
Do you want to export the home directory on the management node

   and use it for all compute nodes? (Y/N) [Y]:
```

21. Enter the Platform Cluster Manager database administrator password and root account password for all compute nodes:

```
Do you want to change the root password for compute nodes and the

   default password for the Platform Cluster Manager database? (Y/N) [Y]:

   Enter the Platform Cluster Manager database administrator password
[pcmdbadm]: ********

   Enter the password again: ********

   Enter the root account password for all compute nodes [Cluster]: ********

   Enter the password again: ********
```

22. A summary of your selected installation settings is displayed. To change any of these settings, press ‘99’ to reselect the settings or press '1' to begin the installation.

```
   ================================================================

    Platform Cluster Manager Installation Summary
```

```
================================================================
You have selected the following installation settings:

Provision network domain:                cluster.ibm.com
Provision network interface:             eth0, 192.168.0.0/255.255.255.0
Public network domain:                   test.ibm.com
Public network interface:                eth1, 10.0.0.0/255.255.255.0
Depot (/install) directory mount point:  /
OS media:                                /mnt/IBMIT_RHEL65
Network Interface:                       eth0
eth0 IP address range for compute nodes: 192.168.0.102-192.168.0.200
eth0 IP address range for node discovery:192.168.0.201-192.168.0.254
Enable firewall:                         Yes
Enable NAT forwarding:                   Yes
NTP server:                              10.0.0.22
Name servers:                            127.0.0.1
Database administrator password:         ************
Compute node root password:              ************
Export home directory:                   Yes
================================================================


Note: To copy the OS from the OS DVD, you must insert the first
OS DVD into the DVD drive before beginning the installation.

To modify any of the above settings, press "99" to go back
to "Step 3: Specify installation settings", or press "1"
to begin the installation.
 1
```

23. After the installation completes, you must run the **`source /opt/pcm/bin/pcmenv.sh`** command to configure the needed environment variables for your current session. This is not required for new login sessions.

```
The Platform Cluster Manager installation is complete.


   Installation log can be found here: /opt/pcm/log/pcm-installer.log.

Run the 'source /opt/pcm/bin/pcmenv.sh' command to configure environment
variables for this session. This is not required for new login sessions.

To get started with IBM Platform Cluster Manager - Advanced Edition 4.2, using
your web browser, you can access the Web portal at http://10.0.0.30:8080 with
the 'root' user on the management node.

# source /opt/pcm/bin/pcmenv.sh
```

24. Log in with the root user account and password on the management node. You can point your web browser to the Platform Cluster Manager - Advanced Edition’s portal to log in:

`http://hostname:8080` or `http://IPaddress:8080`.

Note: After the installation completes, Platform Cluster Manager - Advanced Edition can be accessed in your browser on port 8080 of your management node’s IP address.

Verifying the installation

Check that you have successfully installed Platform Cluster Manager after you finish your installation. You can check the installation log file, which includes details and results about your Platform Cluster Manager installation, in `/opt/pcm/log/pcm-installer.log`.

To verify that your installation is working correctly, log on to the management node as the root user and complete the following steps:

1. Source the Platform Cluster Manager environment variables:

```
[root@pcm ~]# source /opt/pcm/bin/pcmenv.sh
```

2. Check that the PostgreSQL database server is running:

```
[root@pcm ~]# service postgresql status
 (pid  2623) is running...
```

3. Check that the Platform Cluster Manager services are running:

```
[root@pcm ~]# service xcatd status
xCAT service is running
[root@pcm ~]# service pcm status

Cluster name           : PCM
EGO master host name   : pcm
EGO master version     : 1.2.10
SERVICE  STATE    ALLOC CONSUMER RGROUP RESOURCE SLOTS SEQ_NO INST_STATE ACTI
RULE-EN* STARTED  14    /Manage* Manag* pcm      1     1      RUN        38
PCMD     STARTED  13    /Manage* Manag* pcm      1     1      RUN        37
PTC      STARTED  9     /Manage* Manag* pcm      1     1      RUN        33
PLC      STARTED  10    /Manage* Manag* pcm      1     1      RUN        34
WEBGUI   STARTED  15    /Manage* Manag* pcm      1     1      RUN        39
PURGER   STARTED  11    /Manage* Manag* pcm      1     1      RUN        35
ACTIVEMQ STARTED  12    /Manage* Manag* pcm      1     1      RUN        36
```

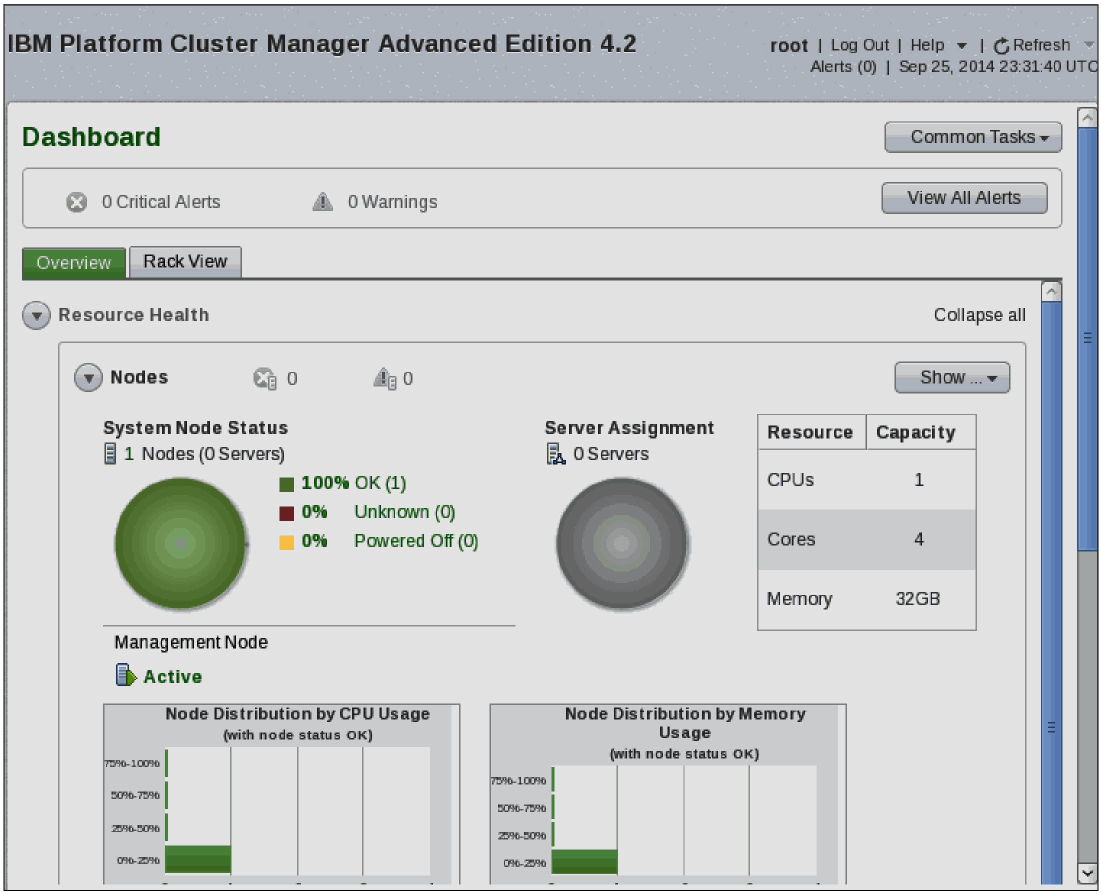
4. Log in to the web portal:

   a. Open a supported web browser. For a list of supported web browsers, see the *Release Notes*.

   b. Go to `http://`*`mgmtnode-IP`*`:8080`, where *`mgmtnode-IP`* is the real management node IP address. If you are connected to a public network, you can also go to `http://mgmtnode-hostname:8080`, where *`mgmtnode-hostname`* is the real management node host name.

   c. Log in as the root user. The root user has administrative privileges and maps to the operating system root user.

   d. After you log in, the Resource Dashboard is displayed in the web portal, as shown in Figure 3-41.



Resource	Capacity
CPUs	1
Cores	4
Memory	32GB

Figure 3-41 Platform Cluster Manager - Advanced Edition dashboard

3.3.3 Managing and configuring the IBM Power Systems nodes

After the installation, proceed to managing and configuring your Power Systems servers within Platform Cluster Manager - Advanced Edition so that Platform Cluster Manager - Advanced Edition can discover and manage theme as cluster nodes. Platform Cluster Manager - Advanced Edition manages Power Systems servers through an FSP port (the HMC1 or HMC2 port). If you have an HMC, you can use the HMC1 port for connecting your server to the HMC and the HMC2 port for Platform Cluster Manager - Advanced Edition to gain access to one of the Power Systems server's FSP.

Platform Cluster Manager - Advanced Edition uses the SLP protocol to discover hardware components on the FSP network. Before attempting a discovery, make sure that you validate the following items:

- Check that the Platform Cluster Manager - Advanced Edition management node and the central electrical complex FSP (HMC port1 or HMC port2 of the managed Power Systems servers) Ethernet connections are physically connected to LAN Ethernet switches and can therefore communicate.
- The central electrical complex FSP ports are 1-Gb ports, so you must ensure that they are connected to 1-Gb ports on the LAN Ethernet switches.
- Check that the SLP protocol is supported on the LAN Ethernet switches by enabling IGMP-Snooping.
- Configure dhcpd with dynamic range to give dynamic DHCP IP addresses to the hardware components so that they can respond to SLP broadcasts.

Configuring the DHCP service for the FSP network

During a Platform Cluster Manager - Advanced Edition installation for rack-mounted Power Systems servers, dhcpd is configured only to listen for requests on the provisioning network, so you should enable dhcpd on the FSP network as well. Prepare your FSP configuration detail and enable dhcpd for the FSP network by completing the following steps:

1. Create an IP pool for your FSP network by using the information that is provided in Table 3-3 on page 76. The values in bold should match your configuration.

```
[root@pcm ~]# mkdef -t network -o FSP mgtifname=eth2 domain=cluster.ibm.com
gateway="<xcatmaster>" net=192.168.1.0 mask=255.255.255.0
staticrange=192.168.1.102-192.168.1.200
dynamicrange=192.168.1.201-192.168.1.254  staticrangeincrement=1
```

> **Tip:** You can run the **chdef** command if you must change the values after you run the **mkdef** command. Also, you can list the parameters by running the **lsdef** command.
>
> ```
> # chdef -t network -o FSP staticrangeincrement=2
> # lsdef -t network -o FSP
> ```

2. Add your FSP interface to the dhcpd configuration.

   First, check the DHCP interface configuration:

```
[root@pcm ~]# lsdef -t site clustersite |grep dhcpinterface
    dhcpinterfaces=eth0
```

   In the test environment, *eth0* is used for the provisioning network. Now, add the FSP interface to be managed by dhcpd as well. In our tests, this is *eth2*, which is the interface that is connected to the FSP network.

```
[root@pcm ~]# chdef -t site clustersite dhcpinterfaces="eth0 eth2"
1 object definitions have been created or modified.
[root@pcm ~]# lsdef -t site clustersite |grep dhcpinterfaces
    dhcpinterfaces=eth0 eth2
```

3. Update the dhcpd configuration and restart the DHCP service:

```
[root@pcm ~]# makedhcp -n

Renamed existing dhcp configuration file to  /etc/dhcp/dhcpd.conf.xcatbak
```

   You can check the DHCP configuration file in `/etc/dhcp/dhcpd.conf` to ensure that it contains the correct configuration, and then restart the dhcpd service:

```
[root@pcm ~]# service dhcpd restart
```

```
Shutting down dhcpd:                                        [  OK  ]
Starting dhcpd:                                             [  OK  ]
```

4. Verify whether DHCP is working on your FSP interface:

   ```
   [root@pcm ~]# detect_dhcpd -i eth2

   ++++++++++++++++++++++++++++++++++
   There are 1 servers reply the dhcp discover.
       Server:192.168.1.30 assign IP [192.168.1.209] to you. The next server is
   [192.168.1.30]!
   ++++++++++++++++++++++++++++++++++
   ```

   You should have only one DHCP server on a given network. If you get a reply from two or more dhcp servers, you should choose which one that you want to use and shut down the dhcp service on the other servers.

Discovering and managing the hardware

After the dhcp daemon is enabled on the FSP network, discover and manage Power Systems server hardware by completing the following steps:

1. Discover hardware:

   a. Check for SLP responses from your Platform Cluster Manager - Advanced Edition management node by running the **lsslp** command. Make sure that you get a response from all of the servers that you want to add to your cluster. If you do not get a response, check your physical connection, dhcp configuration, and IGMP-snooping configuration in the network’s configuration.

      ```
      [root@pcm ~]# lsslp
      Sending SLP request on interfaces: 192.168.1.30,10.0.0.30,192.168.0.30 ...
      Received 9 responses.
      Sending SLP request on interfaces: 192.168.1.30,10.0.0.30,192.168.0.30 ...
      Received 0 responses.
      Sending SLP request on interfaces: 192.168.1.30,10.0.0.30,192.168.0.30 ...
      Received 0 responses.
      Sending SLP request on interfaces: 192.168.1.30,10.0.0.30,192.168.0.30 ...
      Received 0 responses.
      4 requests with 9 responses.  Now processing responses.  This will take 0-1
      minutes...
      device  type-model  serial-number  side  ip-addresses   hostname
      fsp     8246-L2T    06061BA        A-0   172.17.0.11    172.17.0.11
      fsp     8246-L2C    06025DA        A-0   172.17.0.12    172.17.0.12
      fsp     8246-L2T    06061EA        A-0   172.17.0.4     172.17.0.4
      fsp     8246-L2T    060670A        A-0   172.17.0.5     172.17.0.5
      fsp     8246-L2T    06061DA        A-0   172.17.0.6     172.17.0.6
      fsp     8246-L2T    060671A        A-0   172.17.0.7     172.17.0.7
      fsp     8246-L2T    06066FA        A-0   172.17.0.8     172.17.0.8
      fsp     8246-L2T    1008CBA        A-0   172.17.0.9     172.17.0.9
      fsp     8246-L2T    060671A        A-1   192.168.1.201  192.168.1.201
      fsp     8246-L2T    060670A        A-1   192.168.1.202  192.168.1.202
      fsp     8246-L2T    06061EA        A-1   192.168.1.203  192.168.1.203
      fsp     8246-L2T    06066FA        A-1   192.168.1.204  192.168.1.204
      fsp     8246-L2T    06061DA        A-1   192.168.1.205  192.168.1.205
      fsp     8246-L2C    06025DA        A-1   192.168.1.206  192.168.1.206
      fsp     8246-L2T    06061BA        A-1   192.168.1.207  192.168.1.207
      fsp     8246-L2T    1008CBA        A-1   192.168.1.208  192.168.1.208
      cec     8246-L2C    06025DA                             Server-8246-L2C-SN06025DA
      cec     8246-L2T    06061BA                             Server-8246-L2T-SN06061BA
      cec     8246-L2T    06061DA                             Server-8246-L2T-SN06061DA
      cec     8246-L2T    06061EA                             Server-8246-L2T-SN06061EA
      ```

```
cec     8246-L2T    06066FA                                   Server-8246-L2T-SN06066FA
cec     8246-L2T    060670A                                   Server-8246-L2T-SN060670A
cec     8246-L2T    060671A                                   Server-8246-L2T-SN060671A
cec     8246-L2T    1008CBA                                   Server-8246-L2T-SN1008CBA
```

b. Create definition file and define the central electrical complex for the cluster:

```
[root@pcm ~]# lsslp -z -s CEC > mycecs
[root@pcm ~]# cat mycecs | mkdef -z
Warning: The node name
'Server-8246-L2C-SN06025DA,Server-8246-L2T-SN06061EA,Server-8246-L2T-SN06066
FA,Server-8246-L2T-SN06061DA,Server-8246-L2T-SN1008CBA,Server-8246-L2T-SN060
671A,Server-8246-L2T-SN06061BA,Server-8246-L2T-SN060670A' contains capital
letters which may not be resolved correctly by the dns server.
24 object definitions have been created or modified.
[root@pcm ~]# nodels cec
Server-8246-L2C-SN06025DA
Server-8246-L2T-SN06061BA
Server-8246-L2T-SN06061DA
Server-8246-L2T-SN06061EA
Server-8246-L2T-SN06066FA
Server-8246-L2T-SN060670A
Server-8246-L2T-SN060671A
Server-8246-L2T-SN1008CBA
```

2. Initialize the hardware connection:

```
[root@pcm ~]# mkhwconn cec -t
[root@pcm ~]# lshwconn cec
Server-8246-L2C-SN06025DA: 172.17.0.12: LINE DOWN
Server-8246-L2C-SN06025DA:
sp=primary,ipadd=192.168.1.206,alt_ipadd=unavailable,state=LINE UP
Server-8246-L2T-SN06061DA: 172.17.0.6: LINE DOWN
Server-8246-L2T-SN06061DA:
sp=primary,ipadd=192.168.1.205,alt_ipadd=unavailable,state=LINE UP
Server-8246-L2T-SN1008CBA: 172.17.0.9: LINE DOWN
Server-8246-L2T-SN1008CBA:
sp=primary,ipadd=192.168.1.208,alt_ipadd=unavailable,state=LINE UP
Server-8246-L2T-SN060671A: 172.17.0.7: LINE DOWN
Server-8246-L2T-SN060671A:
sp=primary,ipadd=192.168.1.201,alt_ipadd=unavailable,state=LINE UP
Server-8246-L2T-SN06061BA: 172.17.0.11: LINE DOWN
Server-8246-L2T-SN06061BA:
sp=primary,ipadd=192.168.1.207,alt_ipadd=unavailable,state=LINE UP
Server-8246-L2T-SN06061EA: 172.17.0.4: LINE DOWN
Server-8246-L2T-SN06061EA:
sp=primary,ipadd=192.168.1.203,alt_ipadd=unavailable,state=LINE UP
Server-8246-L2T-SN060670A: 172.17.0.5: LINE DOWN
Server-8246-L2T-SN060670A:
sp=primary,ipadd=192.168.1.202,alt_ipadd=unavailable,state=LINE UP
Server-8246-L2T-SN06066FA: 172.17.0.8: LINE DOWN
Server-8246-L2T-SN06066FA:
sp=primary,ipadd=192.168.1.204,alt_ipadd=unavailable,state=LINE UP
```

In the previous results, each server shows an SLP response for two HMC ports. Subnet 172.17.0.0 refers to HMC port1 for HMC management, and subnet 192.168.1.0 refers to HMC port2, which is connected to the FSP network on Platform Cluster Manager - Advanced Edition.

> **Hint:** If the **`lshwconn` cec** command displays `state=LINE UP`, the hardware connection is fine as is. If the command displays `state=CEC AUTHENTICATION FAILED` or `state=PASSWORDS REQUIRED CHANGE`, you must change the password to connect to the CEC. To change the password, run the following command:
>
> ```
> # rspconfig cec *_passwd=old_password,new_password
> ```
>
> *old_password* is the old password and *new_password* is the new password. By default, the old_password value is empty.

3. Verify that the hardware control setup is correct for the CECs:

```
[root@pcm ~]# rpower cec state
   Server-8246-L2C-SN06025DA:operating
   Server-8246-L2T-SN06061BA:operating
   Server-8246-L2T-SN06061DA:operating
   Server-8246-L2T-SN06061EA:operating
   Server-8246-L2T-SN06066FA:operating
   Server-8246-L2T-SN060670A:operating
   Server-8246-L2T-SN060671A:operating
   Server-8246-L2T-SN1008CBA:operating
```

Configuring the console on demand

In a Power Systems servers cluster environment, the consoles are opened by the *fsp-api* through the FSP network. If the `consoleondemand` attribute is set to `no`, all the consoles of the cluster open immediately. For a large environment, this setting affects the performance of the Platform Cluster Manager - Advanced Edition management node. When the `consoleondemand` attribute is set to `yes`, conserver connects to and logs the console output only when the user opens the console by running the **`rcons`** command. The default value of the `consoleondemand` attribute on Linux is `no`, so change it to `yes`.

```
[root@pcm ~]# lsdef -t site clustersite |grep console
    consoleondemand=no
[root@pcm ~]# chdef -t site clustersite consoleondemand=yes
1 object definitions have been created or modified.
[root@pcm ~]# lsdef -t site clustersite |grep console
    consoleondemand=yes
[root@pcm ~]# makeconservercf
```

3.3.4 Provisioning templates

In Platform Cluster Manager - Advanced Edition, you can provision a node or several nodes at a time by using a provisioning template. A provisioning template defines characteristics for provisioning nodes, including a hardware profile, an image profile, and network profile to use. To create a convenient provisioning template, you must create all of the other profiles first.

Creating a hardware profile

In Platform Cluster Manager - Advanced Edition, there are some predefined hardware profiles that are available, including a hardware profile for Flex Systems and one for rack-mounted servers. This section shows how to create a hardware profile. Create one named `IBM_System_p_CEC_9600` because on the OpenPower 7R2 servers, the serial console speed is 9600, which is a different serial port speed than the default predefined profile for rack-mounted servers. The default hardware profile `IBM_System_p_CEC` uses a serial console speed of 115200.

To create a hardware profile, complete the following steps:

1. You can see the existing hardware profiles by running the **tabdump nodehm** command:

```
[root@pcm ~]# tabdump nodehm
#node,power,mgt,cons,termserver,termport,conserver,serialport,serialspeed,seria
lflow,getmac,cmdmapping,consoleondemand,comments,disable
"__HardwareProfile_IBM_Flex_System_p",,"fsp",,,,,,,,,"/opt/pcm/etc/hwmgt/mappin
gs/HWCmdMapping_flex_p.xml",,,
"__HardwareProfile_IBM_System_p_CEC",,"fsp","fsp",,,,,,,,"/opt/pcm/etc/hwmgt/ma
ppings/HWCmdMapping_rackmount_p.xml",,,
"__Chassis_IBM_Flex_chassis",,"blade",,,,,,,,,,,,
"192.168.1.205",,"fsp",,,,,,,,,,,,
"Server-8246-L2C-SN06025DA",,"fsp",,,,,,,,,,,,
"192.168.1.204",,"fsp",,,,,,,,,,,,
"172.17.0.9",,"fsp",,,,,,,,,,,,
"192.168.1.203",,"fsp",,,,,,,,,,,,
"192.168.1.207",,"fsp",,,,,,,,,,,,
"172.17.0.8",,"fsp",,,,,,,,,,,,
"192.168.1.208",,"fsp",,,,,,,,,,,,
"192.168.1.202",,"fsp",,,,,,,,,,,,
"Server-8246-L2T-SN06061EA",,"fsp",,,,,,,,,,,,
"172.17.0.6",,"fsp",,,,,,,,,,,,
"Server-8246-L2T-SN06066FA",,"fsp",,,,,,,,,,,,
"Server-8246-L2T-SN06061DA",,"fsp",,,,,,,,,,,,
"Server-8246-L2T-SN1008CBA",,"fsp",,,,,,,,,,,,
"192.168.1.206",,"fsp",,,,,,,,,,,,
"Server-8246-L2T-SN060671A",,"fsp",,,,,,,,,,,,
"172.17.0.5",,"fsp",,,,,,,,,,,,
"Server-8246-L2T-SN06061BA",,"fsp",,,,,,,,,,,,
"172.17.0.12",,"fsp",,,,,,,,,,,,
"192.168.1.201",,"fsp",,,,,,,,,,,,
"Server-8246-L2T-SN060670A",,"fsp",,,,,,,,,,,,
"172.17.0.7",,"fsp",,,,,,,,,,,,
"172.17.0.4",,"fsp",,,,,,,,,,,,
"172.17.0.11",,"fsp",,,,,,,,,,,,
```

   You can see that the path of `cmdmapping` for rack-mounted servers is `/opt/pcm/etc/hwmgt/mappings/HWCmdMapping_rackmount_p.xml`. Use the same `cmdmapping` file for the new hardware profile because these are 7R2 nodes; change only the serial connection speed.

2. Add a hardware profile by using the same `cmdmapping` file:

```
[root@pcm ~]# pcmaddhwprofile name=IBM_System_p_CEC_9600 type=fsp
cmdmapping=/opt/pcm/etc/hwmgt/mappings/HWCmdMapping_rackmount_p.xml

Hardware profile IBM_System_p_CEC_9600 added.
```

3. Change the attribute of the new hardware profile to what is suitable for your machine. Note the change of the `hwtype` parameter to `lpar`. This change is made because the Power Systems server nodes, as bare metal systems with a single LPAR, uses all of the server resources. After you make changes, verify them by running the **lsdef** command again.

```
[root@pcm ~]# lsdef -t group -o __HardwareProfile_IBM_System_p_CEC_9600

Object name: __HardwareProfile_IBM_System_p_CEC_9600
    cmdmapping=/opt/pcm/etc/hwmgt/mappings/HWCmdMapping_rackmount_p.xml
    grouptype=static
```

```
        hwtype=blade
        members=
        mgt=fsp

[root@pcm ~]# chdef -t group -o __HardwareProfile_IBM_System_p_CEC_9600
hwtype=lpar cons=fsp serialspeed=9600 usercomment="7R1,7R2 baudrate=9600"

1 object definitions have been created or modified.

[root@pcm ~]# lsdef -t group -o __HardwareProfile_IBM_System_p_CEC_9600

Object name: __HardwareProfile_IBM_System_p_CEC_9600
        cmdmapping=/opt/pcm/etc/hwmgt/mappings/HWCmdMapping_rackmount_p.xml
        cons=fsp
        grouptype=static
        hwtype=lpar
        members=
        mgt=fsp
        serialspeed=9600
        usercomment=7R1,7R2 baudrate=9600
```

4. Run the **plcclient.sh** command to make the hardware profile available from the web portal:

```
[root@pcm ~]# plcclient.sh -d pcmhardwareprofileloader
Loaders start successfully.
```

Now, you have a new hardware profile for rack-mounted servers that uses a baud rate of 9600 for its serial speed connection, and you can see it in the Platform Cluster Manager - Advanced Edition web portal. Use this information when defining a provisioning template in "Creating a provisioning template" on page 101.

Creating an image profile

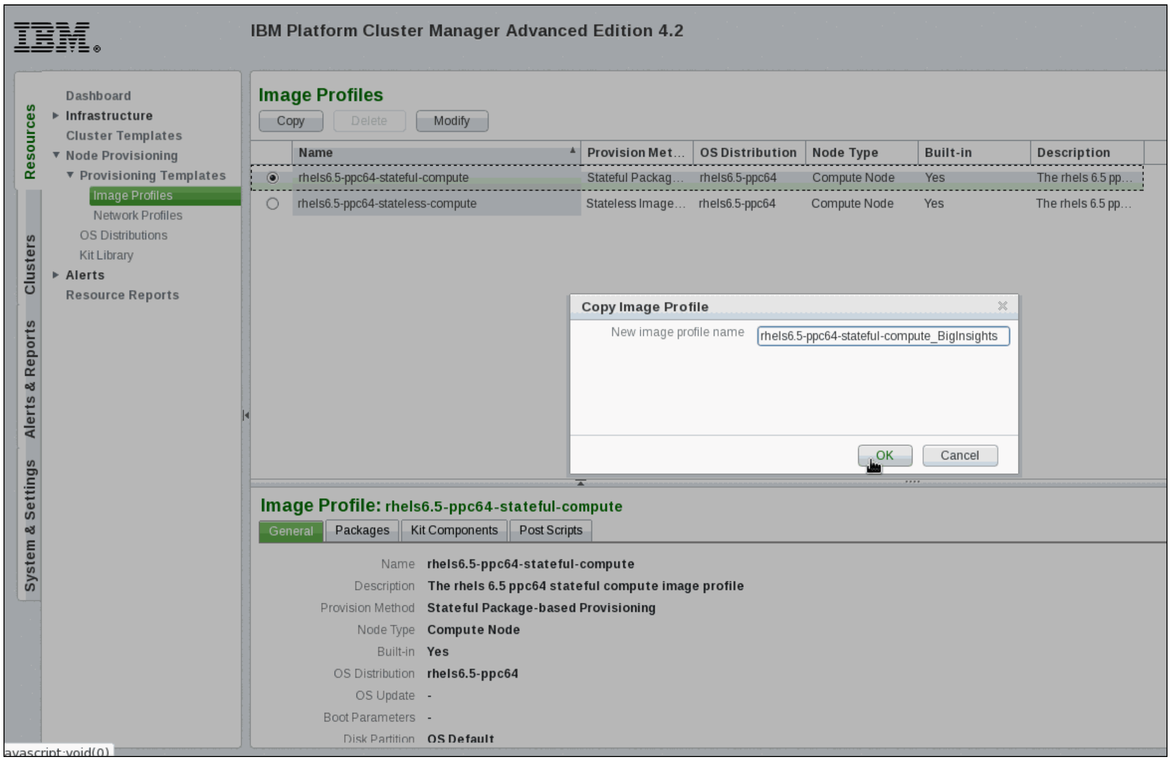
Image profiles represent a logical definition of what is provisioned on compute nodes. An image profile includes the operating system version to be used, other software (kits) with their configuration scripts, postinstallation and post-boot scripts, and custom packages and kernel modules.

Image profiles are used to provision and update the compute nodes with all the definition within them.

Note: An image profile is associated to a compute node. Any modification of the image profile that is used by a node makes the compute node out of synchronization. You can synchronize the compute node and apply changes to all associated nodes after performing modifications to an existing profile.

You can add an image profile by completing the following steps in the Platform Cluster Manager - Advanced Edition web portal:

1. To add an image profile, copy an existing default image profile, and then modify it. Log on to the Platform Cluster Manager - Advanced Edition web portal, go to the **Resources** tab, and click **Provisioning** → **Provisioning Templates** → **Image Profile**. Select the radio button for the stateful images, and click **Copy**. These steps are shown in Figure 3-42.



	Name	Provision Met…	OS Distribution	Node Type	Built-in	Description
●	rhels6.5-ppc64-stateful-compute	Stateful Packag…	rhels6.5-ppc64	Compute Node	Yes	The rhels 6.5 pp…
○	rhels6.5-ppc64-stateless-compute	Stateless Image…	rhels6.5-ppc64	Compute Node	Yes	The rhels 6.5 pp…

Copy Image Profile

New image profile name: rhels6.5-ppc64-stateful-compute_BigInsights

Image Profile: rhels6.5-ppc64-stateful-compute

Name	rhels6.5-ppc64-stateful-compute
Description	The rhels 6.5 ppc64 stateful compute image profile
Provision Method	Stateful Package-based Provisioning
Node Type	Compute Node
Built-in	Yes
OS Distribution	rhels6.5-ppc64
OS Update	-
Boot Parameters	-
Disk Partition	OS Default

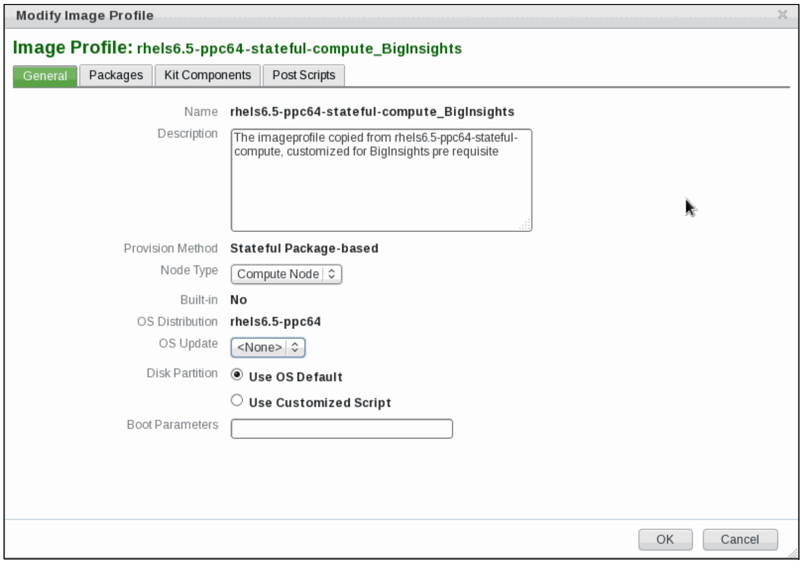
Figure 3-42 Copy an image profile

> **Note:** Stateful provisioning loads the operating system to persistent storage, such as a local disk, iSCSI, or a SAN device. Changes that are made to the operating system are persistent across node restarts. This process is used for standard node installation.
>
> Stateless provisioning loads the operating system to memory and does not need a disk partition definition for the provisioning. Changes that are made to the operating system while it is running are not persistent across node restarts. These changes are often used as rescue systems. You can use diskless provisioning through RAM-root or compressed RAM-root. The RAM-root method is using an uncompressed file system in the operating system, while the compressed RAM-Root is using compressed file system in the operating system. Using the compressed RAM-root can reduce the amount of memory that is needed for the provisioning.

2. Modify your image profile. Select the radio button on your image profile and then click **Modify**. You see the image profile configuration, as shown in Figure 3-43. Here is a description of this configuration:

   a. The General tab.

      In this tab, you can add a description of your image profile, use customized disk partition tables, and specify boot parameters.



Modify Image Profile

Image Profile: rhels6.5-ppc64-stateful-compute_BigInsights

Name rhels6.5-ppc64-stateful-compute_BigInsights

Description The imageprofile copied from rhels6.5-ppc64-stateful-compute, customized for BigInsights pre requisite

Provision Method Stateful Package-based

Node Type Compute Node

Built-in No

OS Distribution rhels6.5-ppc64

OS Update <None>

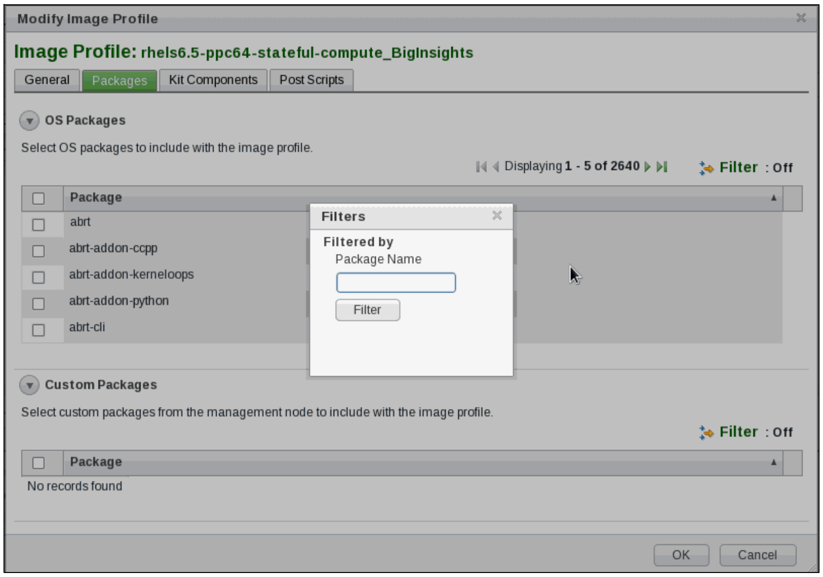
Disk Partition ● Use OS Default ○ Use Customized Script

Boot Parameters

Figure 3-43 Customize image profiles

   b. The Packages tab.

      In this tab, you can choose software packages to include in the image for deployment. Place the InfoSphere BigInsights package prerequisite here. For a list of packages, see "Installed software prerequisites" on page 43. Figure 3-44 exemplifies this step.



Modify Image Profile

Image Profile: rhels6.5-ppc64-stateful-compute_BigInsights

OS Packages

Select OS packages to include with the image profile.

Displaying 1 - 5 of 2640

Filter : Off

□	Package
□	abrt
□	abrt-addon-ccpp
□	abrt-addon-kerneloops
□	abrt-addon-python
□	abrt-cli

Filters

Filtered by

Package Name

Custom Packages

Select custom packages from the management node to include with the image profile.

Filter : Off

□	Package

No records found

Figure 3-44 Add packages to an image profile for custom deployment

c. The Kit Components tab.

   In this tab, you can select which component kits to use with your image. Platform Cluster Manager - Advanced Edition comes with a kit that can be used for node monitoring. Leave this component selected.

d. The Post Scripts tab.

   In this tab, you can add postinstallation scripts, which are run once after the installation but before the first boot, and post boot scripts, which are run every time the nodes restart. You can put a script to set up all of the InfoSphere BigInsights prerequisites, as described in 3.1.2, “Operating system prerequisites setup for InfoSphere BigInsights” on page 39, as postinstallation scripts. Figure 3-45 shows this step.

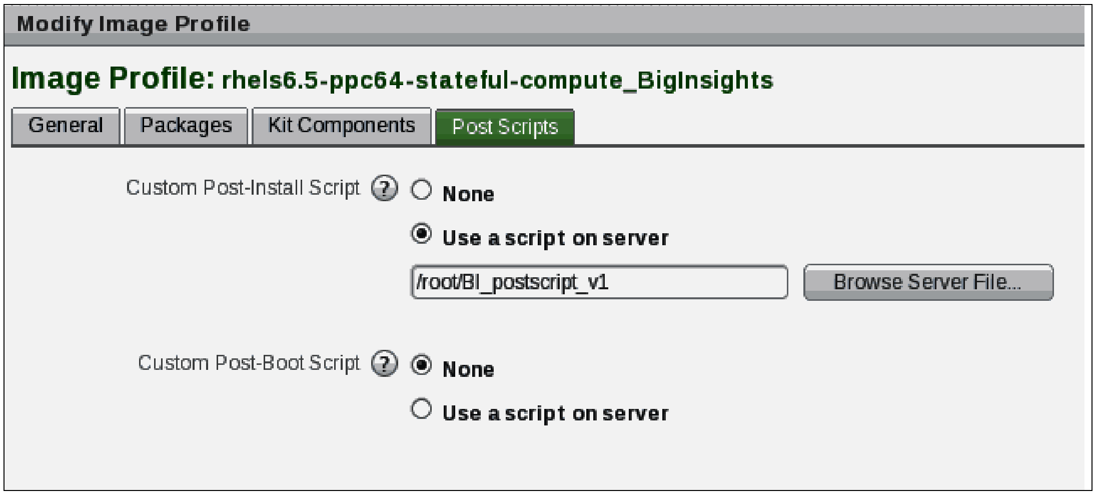


Image Profile: rhels6.5-ppc64-stateful-compute_BigInsights

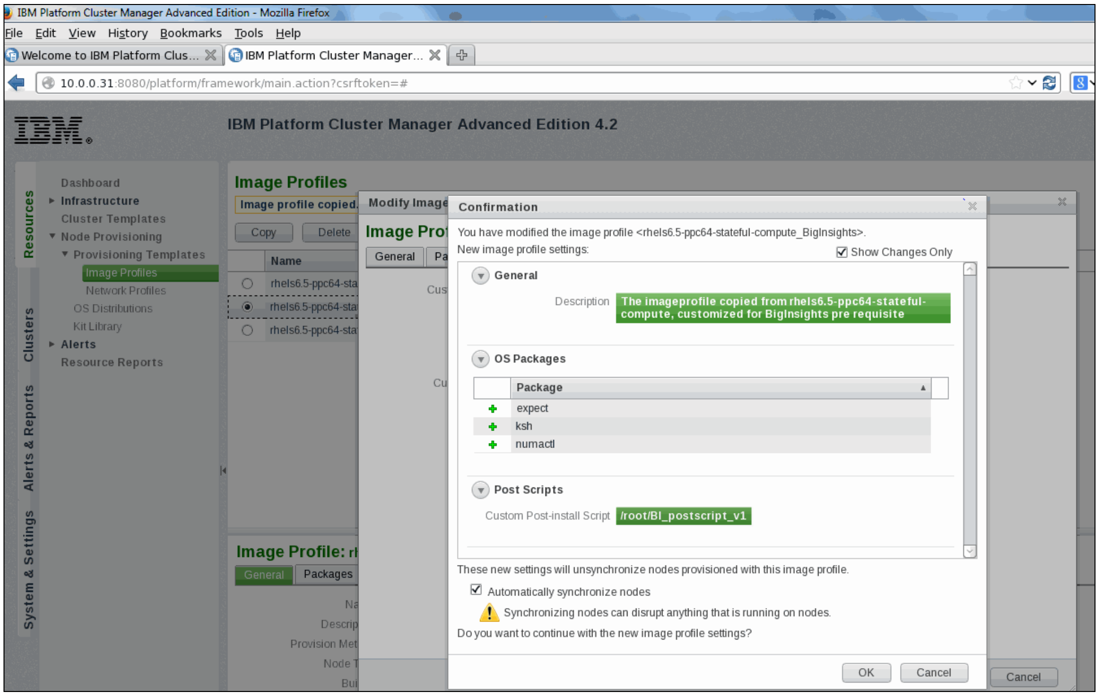
Custom Post-Install Script: ○ None ● Use a script on server

/root/BI_postscript_v1

Custom Post-Boot Script: ● None ○ Use a script on server

Figure 3-45 Add postinstallation or post-boot scripts to an image profile

3. Verify your modification and confirm, as shown in Figure 3-46.


IBM Platform Cluster Manager Advanced Edition - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Welcome to IBM Platform Clus...
IBM Platform Cluster Manager...
10.0.0.31:8080/platform/framework/main.action?csrftoken=#
IBM
IBM Platform Cluster Manager Advanced Edition 4.2
Resources
Clusters
Alerts & Reports
System & Settings
Dashboard
Infrastructure
Cluster Templates
Node Provisioning
Provisioning Templates
Image Profiles
Network Profiles
OS Distributions
Kit Library
Alerts
Resource Reports
Image Profiles
Copy
Delete
Name
Modify Image
General
Packages
OK
Cancel

Confirmation

You have modified the image profile <rhels6.5-ppc64-stateful-compute_BigInsights>.

New image profile settings: ☑ Show Changes Only

General

Description: The imageprofile copied from rhels6.5-ppc64-stateful-compute, customized for BigInsights pre requisite

OS Packages

	Package
✚	expect
✚	ksh
✚	numactl

Post Scripts

Custom Post-install Script: /root/BI_postscript_v1

These new settings will unsynchronize nodes provisioned with this image profile.

☑ Automatically synchronize nodes

⚠ Synchronizing nodes can disrupt anything that is running on nodes.

Do you want to continue with the new image profile settings?

Figure 3-46 Confirm changes to an image profile

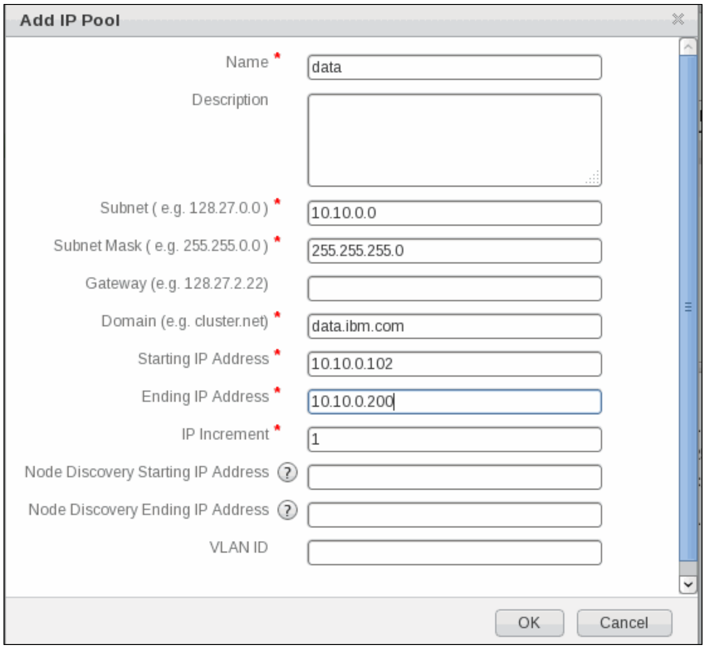
Creating a network profile

Network profiles define the network configuration for the compute nodes. You can define networks that are used by the compute nodes in such a profile. However, networks that can be added in a network profile are based on IP pools. IP pools represent a subnet that includes an IP address range, a subnet, a subnet mask, and a gateway. If you want to configure multiple network interfaces on the compute nodes for application use, create an IP pool for it first.

Adding IP pools

To add IP pools, go to the **Resources** tab and click **Infrastructure** → **Networks** → **IP Pools** → **Add**. Figure 3-47 shows the addition of IP pools for the InfoSphere BigInsights data network, which uses the 10-Gb network adapter.

This is where you use the information from Table 3-3 on page 76. Make sure that you have IP pools for the provisioning, FSP, public, and data networks. The provisioning and FSP ones should be there because they were added during the installation steps.

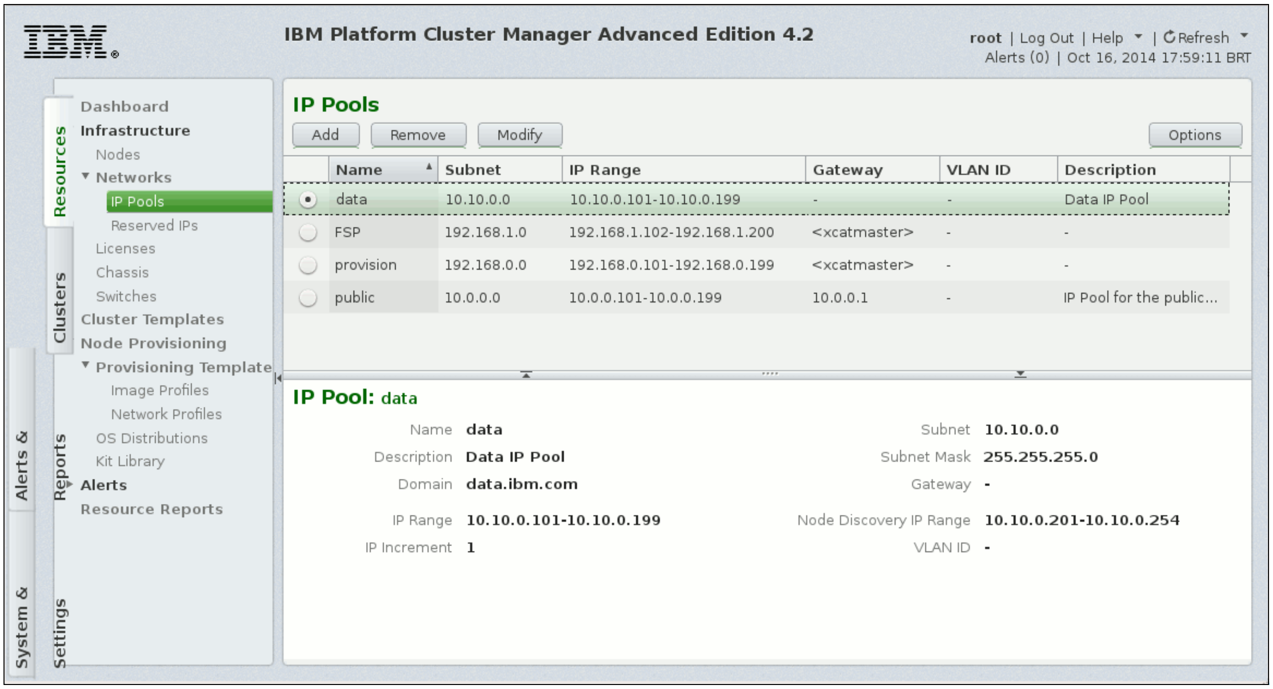


Add IP Pool

Name *	data
Description	
Subnet (e.g. 128.27.0.0) *	10.10.0.0
Subnet Mask (e.g. 255.255.0.0) *	255.255.255.0
Gateway (e.g. 128.27.2.22)	
Domain (e.g. cluster.net) *	data.ibm.com
Starting IP Address *	10.10.0.102
Ending IP Address *	10.10.0.200
IP Increment *	1
Node Discovery Starting IP Address	
Node Discovery Ending IP Address	
VLAN ID	

Figure 3-47 Add an IP pool

Figure 3-48 shows all of the required IP pools for the InfoSphere BigInsights environment.



IP Pools

	Name	Subnet	IP Range	Gateway	VLAN ID	Description
●	data	10.10.0.0	10.10.0.101-10.10.0.199	-	-	Data IP Pool
○	FSP	192.168.1.0	192.168.1.102-192.168.1.200	<xcatmaster>	-	-
○	provision	192.168.0.0	192.168.0.101-192.168.0.199	<xcatmaster>	-	-
○	public	10.0.0.0	10.0.0.101-10.0.0.199	10.0.0.1	-	IP Pool for the public...

IP Pool: data

Name	data	Subnet	10.10.0.0
Description	Data IP Pool	Subnet Mask	255.255.255.0
Domain	data.ibm.com	Gateway	-
IP Range	10.10.0.101-10.10.0.199	Node Discovery IP Range	10.10.0.201-10.10.0.254
IP Increment	1	VLAN ID	-

Figure 3-48 IP pools for automating the deployment of an InfoSphere BigInsights cluster

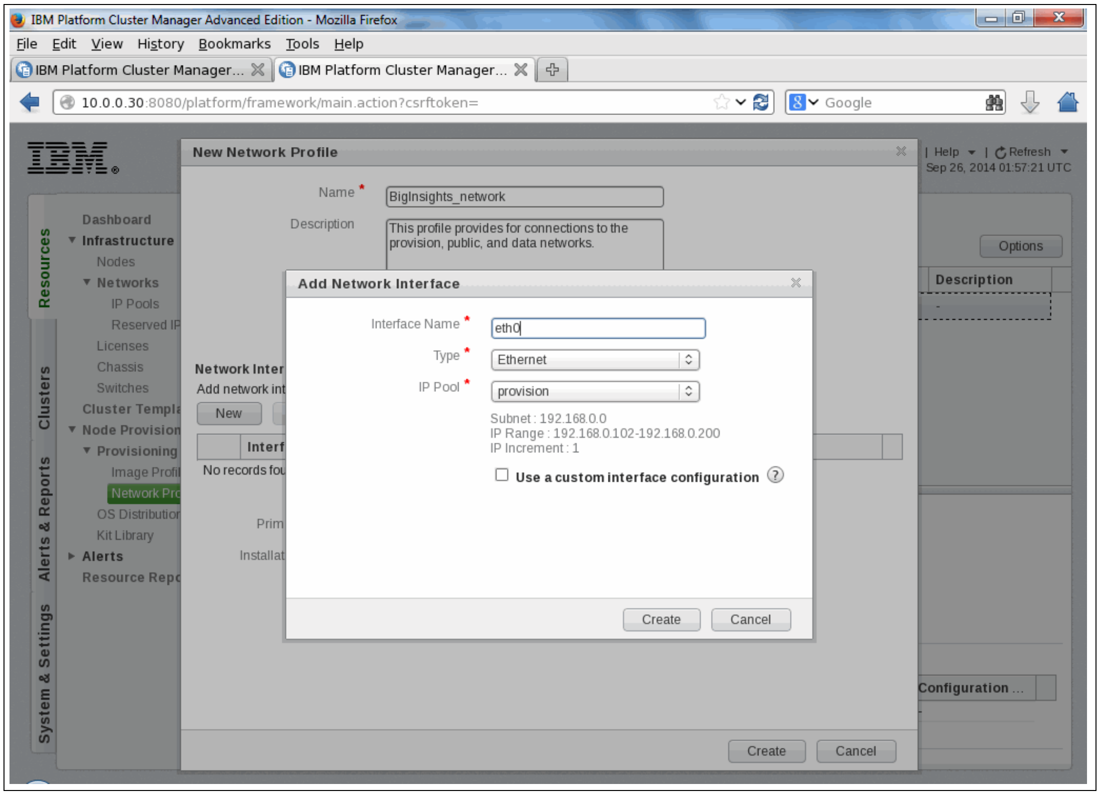
Creating a network profile

To create a network profile, go to the **Resource** tab, click **Node Provisioning** → **Provisioning Template** → **Network Profile** → **New**, and complete the following steps:

1. Choose a network profile name and description that represents the network profile.

2. Add network interfaces to be configured in the compute nodes. For the InfoSphere BigInsights cluster, configure three network interfaces for the compute nodes:

   a. A network interface for the provision network (*eth0* in the test lab). This interface is used for provisioning an image, monitoring, and updating the nodes, as shown in Figure 3-49. Choose the appropriate IP pool for the provisioning network that you created in “Adding IP pools” on page 95.



New Network Profile

Name * BigInsights_network

Description: This profile provides for connections to the provision, public, and data networks.

Add Network Interface

Interface Name * eth0

Type * Ethernet

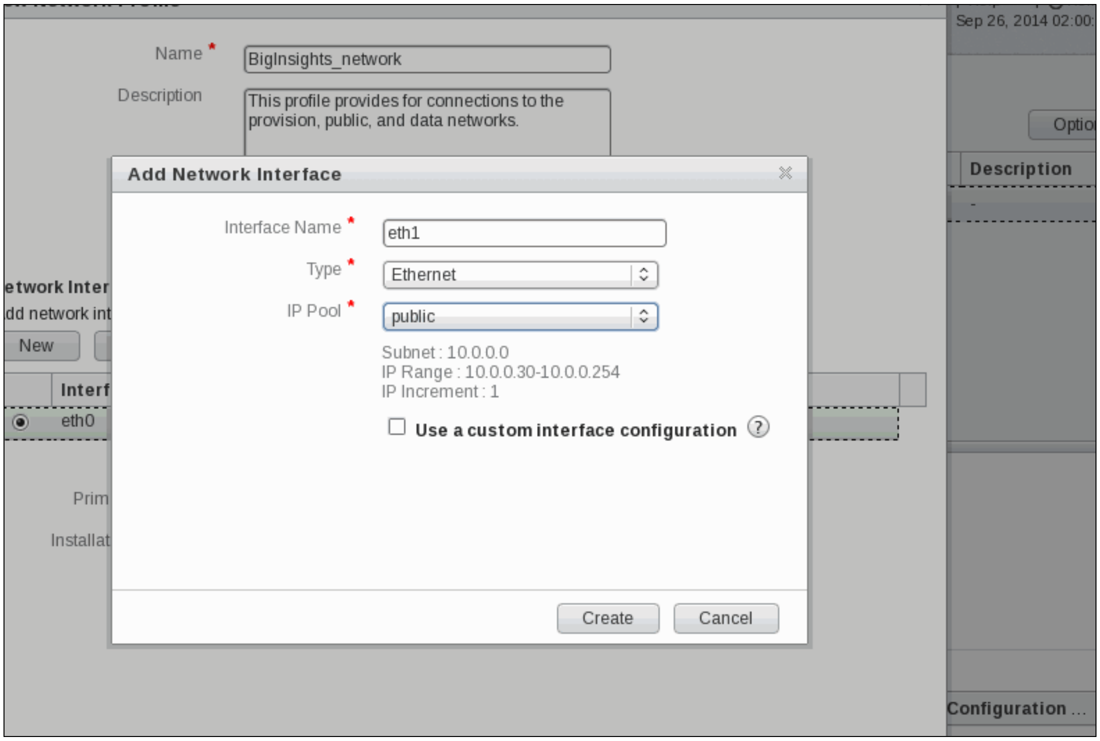
IP Pool * provision

Subnet : 192.168.0.0
IP Range : 192.168.0.102-192.168.0.200
IP Increment : 1

□ Use a custom interface configuration ?

Figure 3-49 Add a network interface to a network profile

b. A network interface for the public network. This interface is connected to your company's public network. InfoSphere BigInsights users reach the cluster through this interface and its IP address. Adding the public network is shown in Figure 3-50. In our lab environment, *eth1* is the interface for the public network, and you must select the proper IP pool for it.



Name * BigInsights_network

Description: This profile provides for connections to the provision, public, and data networks.

Add Network Interface

Interface Name * eth1

Type * Ethernet

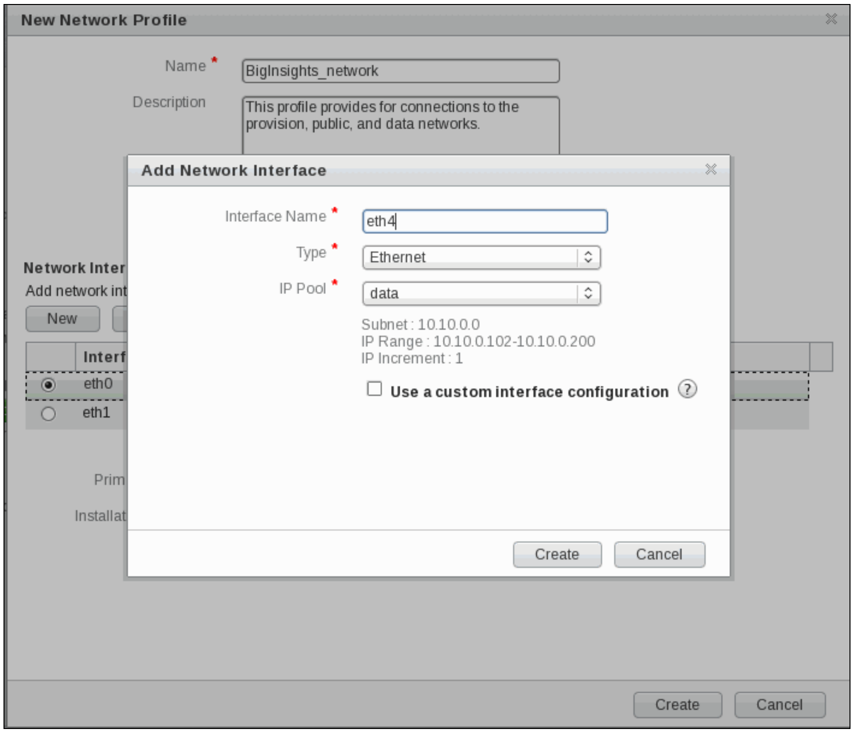
IP Pool * public

Subnet : 10.0.0.0
IP Range : 10.0.0.30-10.0.0.254
IP Increment : 1

☐ Use a custom interface configuration

Figure 3-50 Add a public network to a network profile for InfoSphere BigInsights

c. A network interface for the application network. This interface is used for InfoSphere BigInsights installation and administration. It is the InfoSphere BigInsights main working interface. In this environment, *eth4* is the data network interface that maps to a 10-Gb interface on the servers. Figure 3-51 shows this step. Again, select the appropriate IP pool for this network.



New Network Profile

Name * BigInsights_network

Description This profile provides for connections to the provision, public, and data networks.

Add Network Interface

Interface Name * eth4

Type * Ethernet

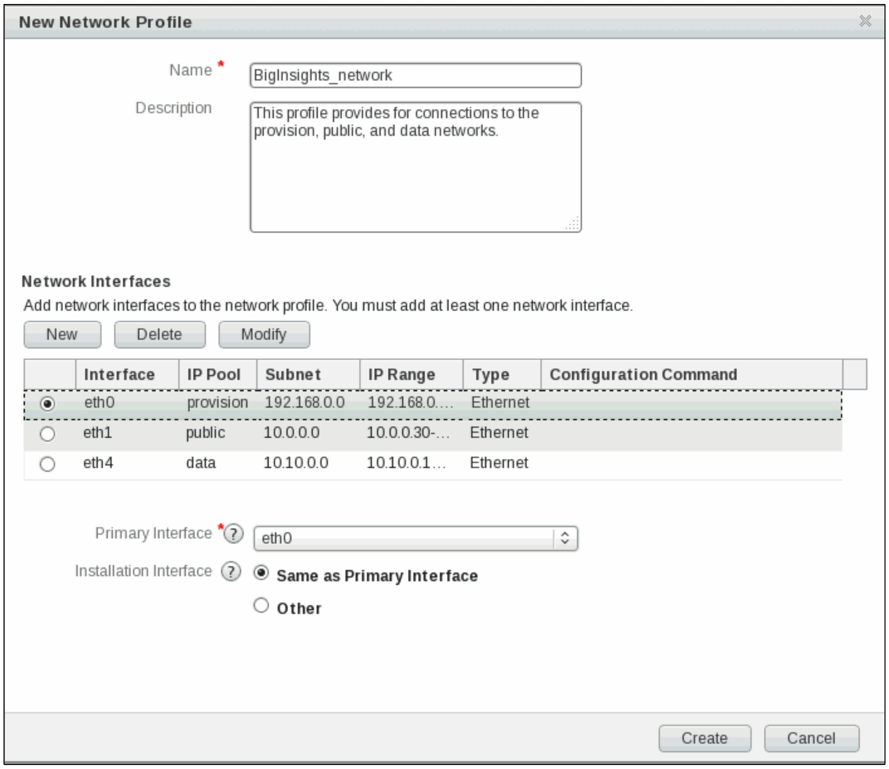
IP Pool * data

Subnet : 10.10.0.0
IP Range : 10.10.0.102-10.10.0.200
IP Increment : 1

☐ Use a custom interface configuration

Figure 3-51 Add a data network to a network profile for InfoSphere BigInsights

3. After all networks are mapped to the network template, choose the primary interface and the installation interface for the profile. A *primary interface* is used for monitoring and updating nodes. An *installation interface* is used for installing and provisioning a node only. You can make the primary and installation interfaces the same one; in this case, they are the provisioning interface. Figure 3-52 shows this step.



New Network Profile

Name *: BigInsights_network

Description: This profile provides for connections to the provision, public, and data networks.

Network Interfaces

Add network interfaces to the network profile. You must add at least one network interface.

	Interface	IP Pool	Subnet	IP Range	Type	Configuration Command
●	eth0	provision	192.168.0.0	192.168.0.…	Ethernet	
○	eth1	public	10.0.0.0	10.0.0.30-…	Ethernet	
○	eth4	data	10.10.0.0	10.10.0.1…	Ethernet	

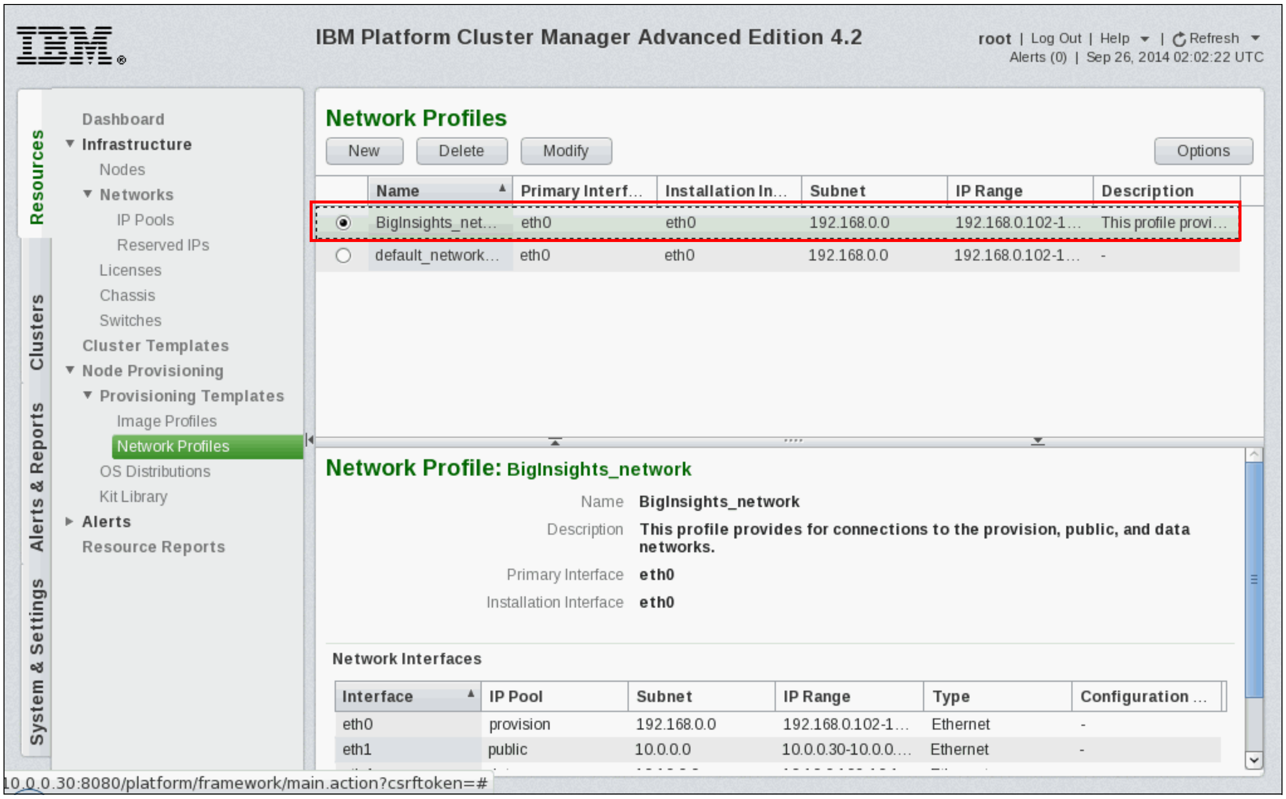
Primary Interface *: eth0

Installation Interface: ● Same as Primary Interface ○ Other

Figure 3-52 Finish the creation of a network profile

4. Verify your configuration and click **Create**.

5. Validate that your network template is created. It should be listed under the network profile section of the Platform Cluster Manager - Advanced Edition interface, as shown in Figure 3-53 on page 101.



Network Profiles

	Name	Primary Interf…	Installation In…	Subnet	IP Range	Description
●	BigInsights_net…	eth0	eth0	192.168.0.0	192.168.0.102-1…	This profile provi…
○	default_network…	eth0	eth0	192.168.0.0	192.168.0.102-1…	-

Network Profile: BigInsights_network

Name BigInsights_network

Description This profile provides for connections to the provision, public, and data networks.

Primary Interface eth0

Installation Interface eth0

Network Interfaces

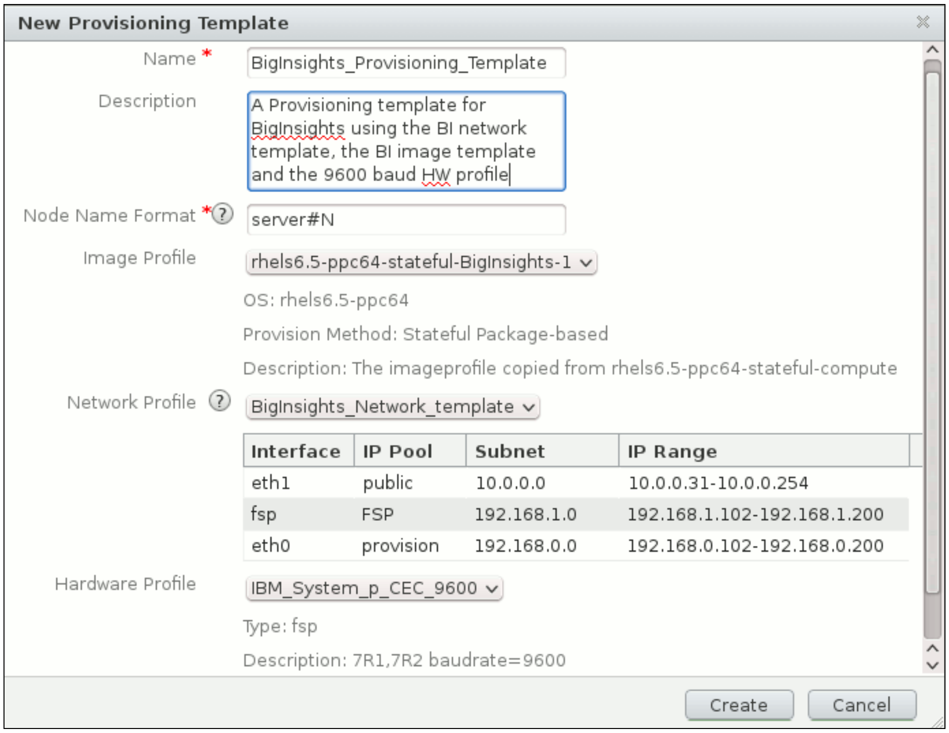
Interface	IP Pool	Subnet	IP Range	Type	Configuration …
eth0	provision	192.168.0.0	192.168.0.102-1…	Ethernet	-
eth1	public	10.0.0.0	10.0.0.30-10.0.0.…	Ethernet	-

Figure 3-53 A network profile has been created

Creating a provisioning template

This is the last step after you define a hardware profile, a network profile (with its associated IP pools), and an image profile. Check that the profiles that you intend to use with your provisioning template suit your needs. The template that is used in this book works for creating an InfoSphere BigInsights cluster on rack-mounted Power Systems servers.

To create a provisioning template, go to the **Resources** tab and click **Node Provisioning** → **Provisioning Templates** → **New**. You are prompted to enter the provisioning template information, as shown in Figure 3-54.



New Provisioning Template

Name * BigInsights_Provisioning_Template

Description: A Provisioning template for BigInsights using the BI network template, the BI image template and the 9600 baud HW profile

Node Name Format * server#N

Image Profile: rhels6.5-ppc64-stateful-BigInsights-1

OS: rhels6.5-ppc64

Provision Method: Stateful Package-based

Description: The imageprofile copied from rhels6.5-ppc64-stateful-compute

Network Profile: BigInsights_Network_template

Interface	IP Pool	Subnet	IP Range
eth1	public	10.0.0.0	10.0.0.31-10.0.0.254
fsp	FSP	192.168.1.0	192.168.1.102-192.168.1.200
eth0	provision	192.168.0.0	192.168.0.102-192.168.0.200

Hardware Profile: IBM_System_p_CEC_9600

Type: fsp

Description: 7R1,7R2 baudrate=9600

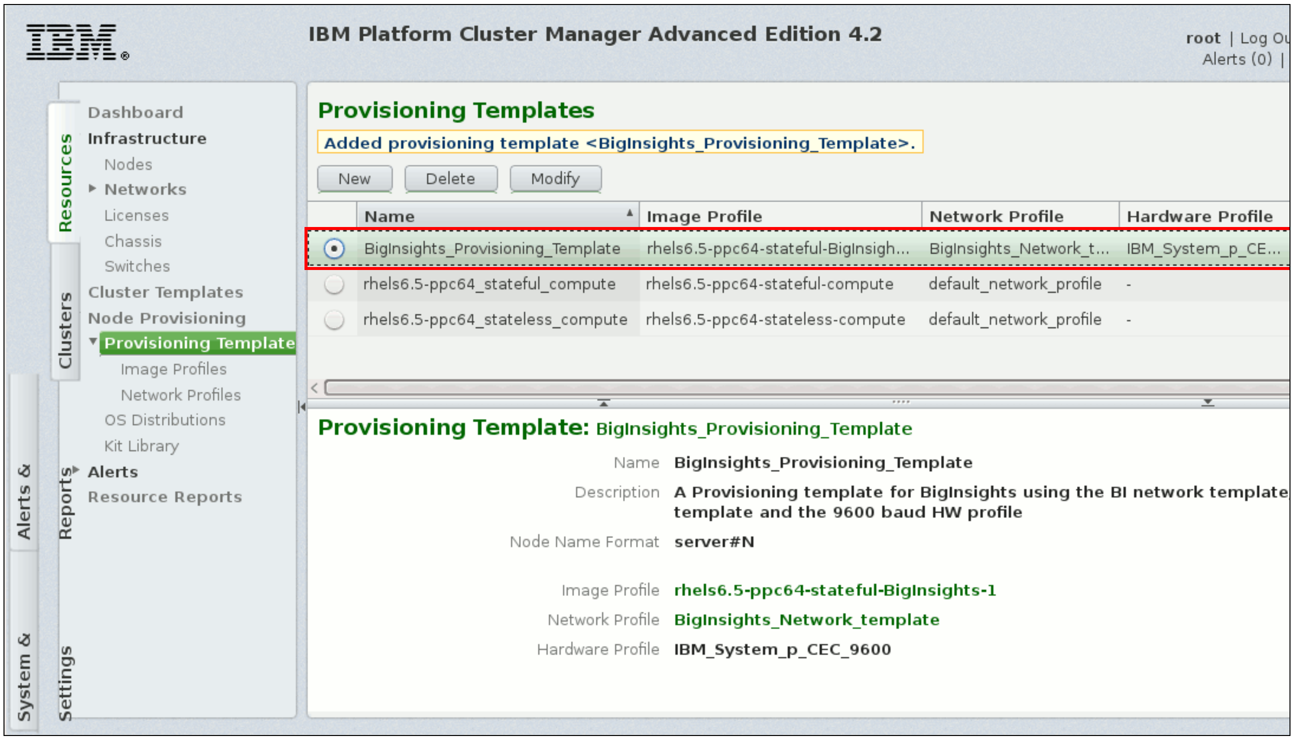
Figure 3-54 Create a provisioning template for an InfoSphere BigInsights cluster on Power Systems

Here are the items that you must enter:

- The template’s name: Assign a meaningful name.
- A description for the template.
- A node name format.

  When you add your Power Systems servers as Platform Cluster Manager - Advanced Edition managed nodes, you can define the node names yourself during the node addition, or you can have Platform Cluster Manager - Advanced Edition assign one for you. This option tells Platform Cluster Manager - Advanced Edition the name format if it assigns a name for you. In this example, the node names are *server1*, *server2*, *server3*, and so on.
- An image profile: Select the InfoSphere BigInsights image profile that you created in “Creating an image profile” on page 91.
- A network profile: Select the InfoSphere BigInsights network profile that you created in “Creating a network profile” on page 95.
- A hardware profile: Select the hardware profile that you created in “Creating a hardware profile” on page 89.

Verify your settings and create your provisioning template. After you are done, you can verify that a new template is available for your use in Platform Cluster Manager - Advanced Edition's node provisioning templates section, as shown in Figure 3-55.



IBM Platform Cluster Manager Advanced Edition 4.2

root | Log Ou
Alerts (0) |

Provisioning Templates

Added provisioning template <BigInsights_Provisioning_Template>.

	Name	Image Profile	Network Profile	Hardware Profile
●	BigInsights_Provisioning_Template	rhels6.5-ppc64-stateful-BigInsigh...	BigInsights_Network_t...	IBM_System_p_CE...
○	rhels6.5-ppc64_stateful_compute	rhels6.5-ppc64-stateful-compute	default_network_profile	-
○	rhels6.5-ppc64_stateless_compute	rhels6.5-ppc64-stateless-compute	default_network_profile	-

Provisioning Template: BigInsights_Provisioning_Template

Name	BigInsights_Provisioning_Template
Description	A Provisioning template for BigInsights using the BI network template template and the 9600 baud HW profile
Node Name Format	server#N
Image Profile	rhels6.5-ppc64-stateful-BigInsights-1
Network Profile	BigInsights_Network_template
Hardware Profile	IBM_System_p_CEC_9600

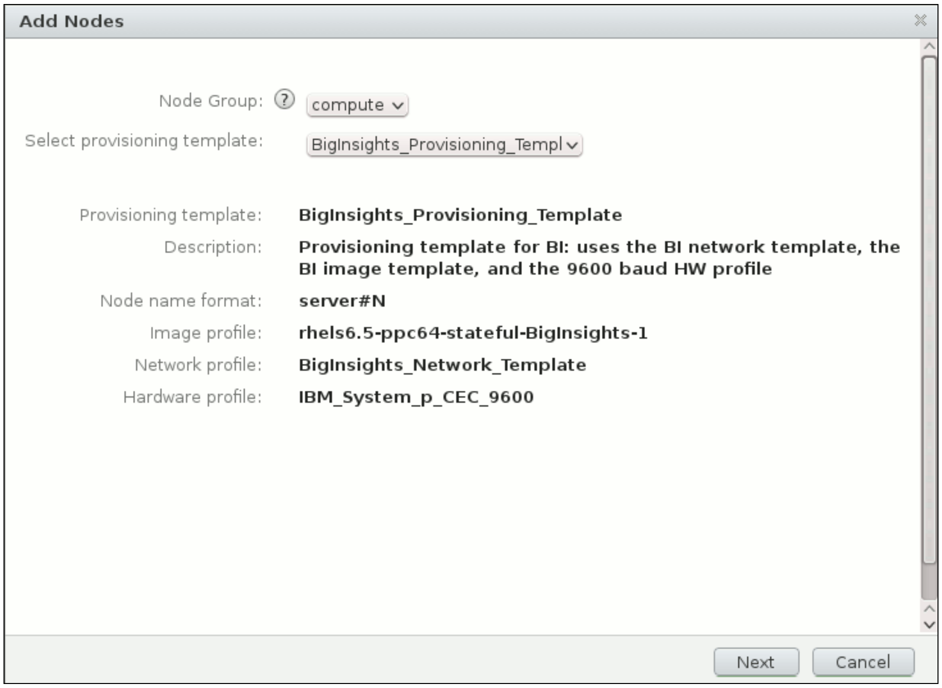
Figure 3-55 Finish the creation of a provisioning template

Now, you can add the Power Systems servers to be managed by your Platform Cluster Manager - Advanced Edition environment, and deploy them by using an InfoSphere BigInsights provisioning template.

3.3.5 Adding nodes to Platform Cluster Manager - Advanced Edition

As a prerequisite to adding nodes to Platform Cluster Manager - Advanced Edition, complete the steps that are shown in "Discovering and managing the hardware" on page 87. Also, having a working provisioning template simplifies this step, so it is a preferred practice to also complete the steps in 3.3.4, "Provisioning templates" on page 89.

To add a Power Systems server to Platform Cluster Manager - Advanced Edition, log on to it and select the **Resources** tab. Click **Infrastructure** → **Nodes** and then click **Add**. You are prompted to select some properties, as shown in Figure 3-56.



Node Group: compute

Select provisioning template: BigInsights_Provisioning_Templ

Provisioning template: BigInsights_Provisioning_Template

Description: Provisioning template for BI: uses the BI network template, the BI image template, and the 9600 baud HW profile

Node name format: server#N

Image profile: rhels6.5-ppc64-stateful-BigInsights-1

Network profile: BigInsights_Network_Template

Hardware profile: IBM_System_p_CEC_9600

Figure 3-56 Add a node to a Platform Cluster Manager - Advanced Edition environment

Here are the properties that you must determine:

- Node group: Use the *compute* node group for your nodes.
- Provisioning template: Select the InfoSphere BigInsights provisioning template that was created earlier for this purpose. Note that in Figure 3-56 that all of the information that is contained in the provisioning template appears in the window and cannot be changed.

Because this scenario uses the provisioning template that was defined earlier, there is no need to set manually anything else. Simply proceed as shown in the wizard.

The second and last step for adding a node is to provide a *node information file*, which contains the information for the nodes to be added. The contents of this file, applied to the test lab, is shown in Example 3-19.

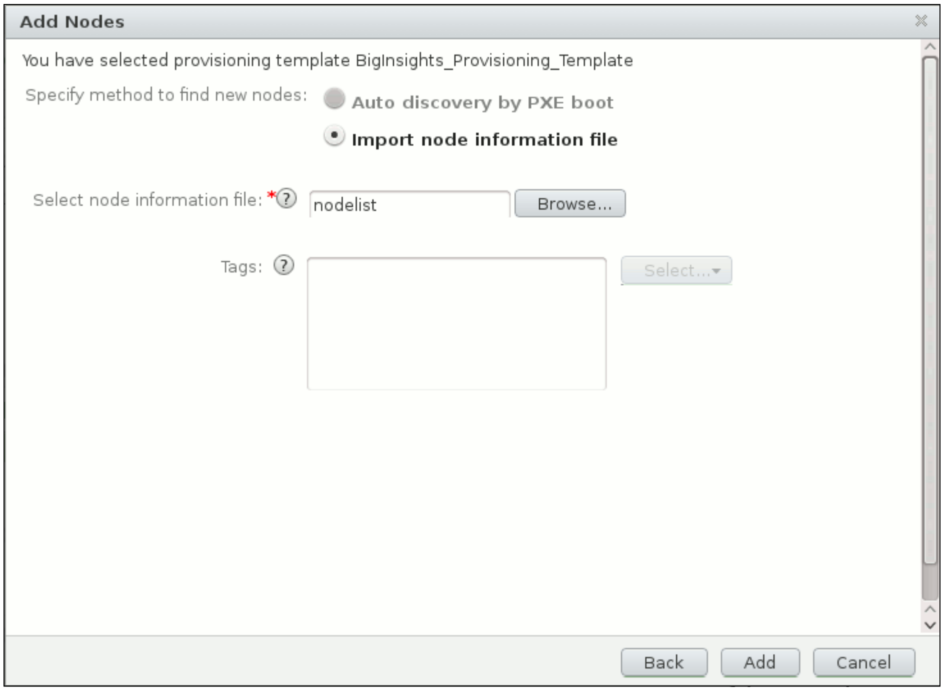
Example 3-19 Node information file for adding nodes to Platform Cluster Manager - Advanced Edition

```
#node definition file
server2:
cec=Server-8246-L2C-SN06025DA

server3:
cec=Server-8246-L2T-SN060670A

server4:
cec=Server-8246-L2T-SN06066FA

server5:
cec=Server-8246-L2T-SN06061BA
```

```
server6:
cec=Server-8246-L2T-SN060671A

server7:
cec=Server-8246-L2T-SN06061EA

server8:
cec=Server-8246-L2T-SN06061DA

server9:
cec=Server-8246-L2T-SN1008CBA
```

Example 3-19 on page 104 contains the CEC information for each server to be added. You can add multiple servers simultaneously. Also, it shows a *serverN* tag for each one of the CECs. Although the provisioning template provides a node name, as explained in “Creating a provisioning template” on page 101, ensure that the server names within Platform Cluster Manager - Advanced Edition match the physical servers tags in your data center. So, provide this information when adding the servers.

Figure 3-57 shows the step of pointing out the information for adding the eight servers.



Add Nodes

You have selected provisioning template BigInsights_Provisioning_Template

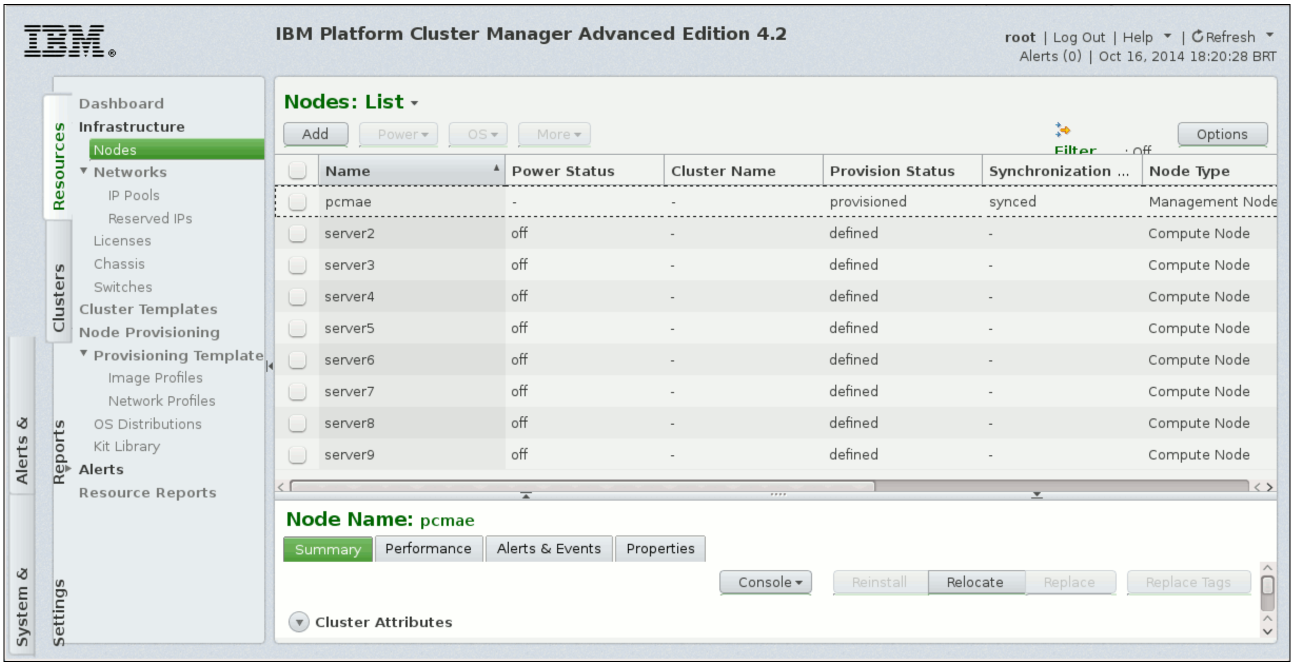
Specify method to find new nodes: ○ Auto discovery by PXE boot
● Import node information file

Select node information file: * nodelist

Tags:

Figure 3-57 Point out to a node information file

Proceed with the wizard. The following windows show the progress of node addition and a summary of successful and failed nodes. After the process is complete, validate that your servers were added. You can find them by clicking **Resources** → **Infrastructure** → **Nodes** in the Platform Cluster Manager - Advanced Edition web portal, as shown in Figure 3-58.



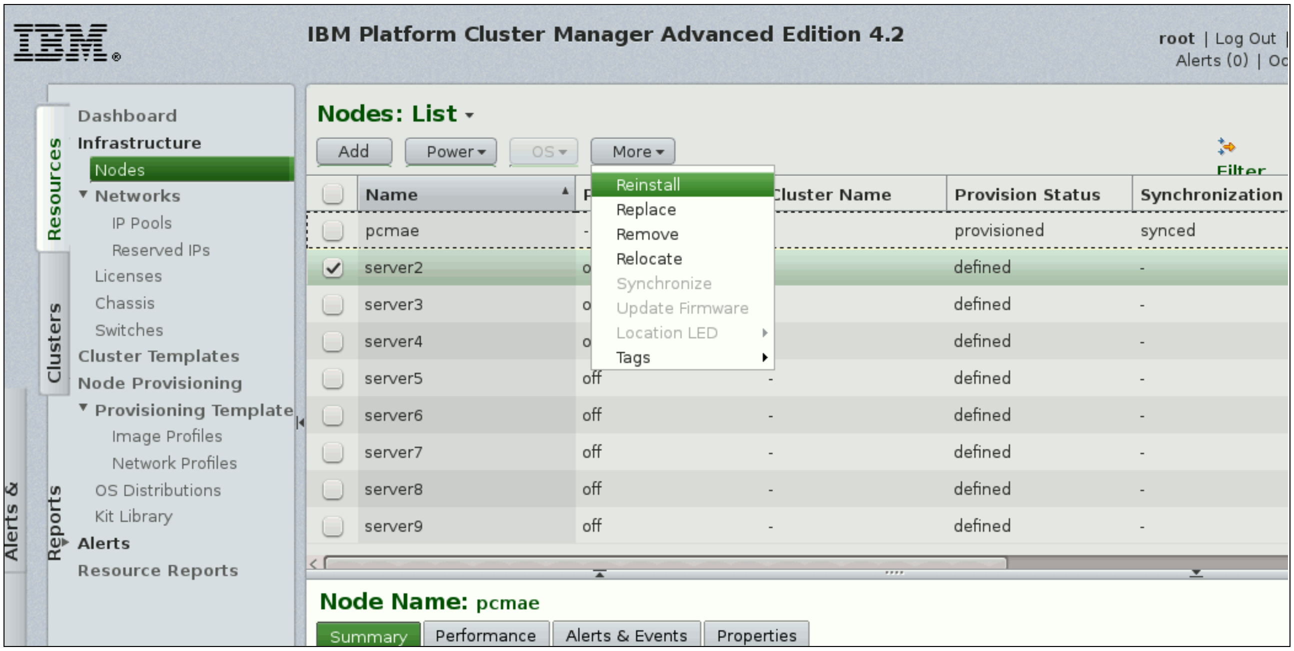
Name	Power Status	Cluster Name	Provision Status	Synchronization ...	Node Type
pcmae	-	-	provisioned	synced	Management Node
server2	off	-	defined	-	Compute Node
server3	off	-	defined	-	Compute Node
server4	off	-	defined	-	Compute Node
server5	off	-	defined	-	Compute Node
server6	off	-	defined	-	Compute Node
server7	off	-	defined	-	Compute Node
server8	off	-	defined	-	Compute Node
server9	off	-	defined	-	Compute Node

Figure 3-58 Verify node addition

Verifying the deployment of a Platform Cluster Manager - Advanced Edition node

After you add a Platform Cluster Manager - Advanced Edition node for the first time after setting up a Platform Cluster Manager - Advanced Edition environment and creating a provisioning template, it is a preferred practice to attempt to provision a node to test your new environment. Pick one of your nodes and install it.

To install the node, select it and click **More**, and then select **Reinstall**. This step is shown in Figure 3-59 on page 107.



	Name	P	Cluster Name	Provision Status	Synchronization
□	pcmae	-		provisioned	synced
☑	server2	o		defined	-
□	server3	o		defined	-
□	server4	o		defined	-
□	server5	off	-	defined	-
□	server6	off	-	defined	-
□	server7	off	-	defined	-
□	server8	off	-	defined	-
□	server9	off	-	defined	-

Node Name: pcmae

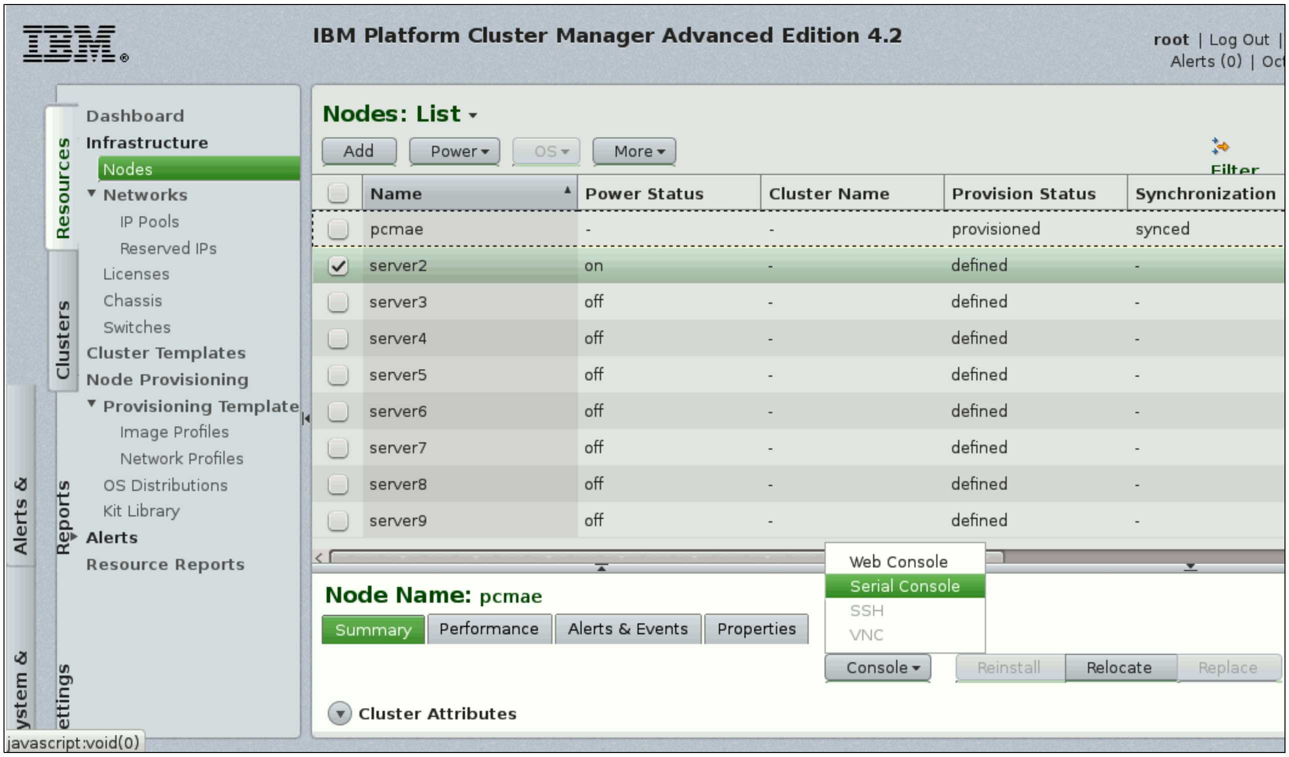
Figure 3-59 Test a newly created provisioning template

Use a *new provisioning template* and select the InfoSphere BigInsights template that you created for this task. Subsequent deployments on the same node can remember your choice of provisioning template, so you can then opt to use the *existing provisioning template* option.

The provisioning process starts. You can verify its progress in a few ways:

- View the status codes on the HMC for the LPAR on the target server. It should cycle the codes for booting and transferring a TFTP image until it displays information for a booted Linux kernel. This method is simple and does not provide you with much information about what is occurring. Also, you must have your nodes managed by an HMC in addition to being managed by the Platform Cluster Manager - Advanced Edition FSP network.
- Open the serial console on the HMC for the target server. Although valid, this is not the preferred approach. The reason for using this approach is because Platform Cluster Manager uses the FSP network for sending TFTP boot commands to the LPAR, which is the same interface the HCM uses for opening a serial console. Instead, use the next method.

- Use the serial connection that is provided by Platform Cluster Manager - Advanced Edition for the node. To access this approach, ensure that you select the node for which you want to open the console, and click **Console**, and select the **Serial Console** entry, as shown in Figure 3-60. Here is where the 9600-baud rate that you configured in “Creating a hardware profile” on page 89 is used.



☐	Name	Power Status	Cluster Name	Provision Status	Synchronization
☐	pcmae	-	-	provisioned	synced
☑	server2	on	-	defined	-
☐	server3	off	-	defined	-
☐	server4	off	-	defined	-
☐	server5	off	-	defined	-
☐	server6	off	-	defined	-
☐	server7	off	-	defined	-
☐	server8	off	-	defined	-
☐	server9	off	-	defined	-

Figure 3-60 Open the Platform Cluster Manager - Advanced Edition serial console connection to the target node

Wait until the provisioning process is complete. Platform Cluster Manager - Advanced Edition applies any postinstallation and boot scripts that your provisioning template determines and then restarts the target node after the installation completes. Be patient because if you try to log in and use the system too early, you might not see the results of your scripts. After the whole process completes, it is a good idea to validate that all of the network interfaces that you defined in your network profile are configured.

Note: Some hardware configurations might require some extra post-scripting work to set up networking correctly. For example, the Mellanox cards need a special driver that must be installed in the system before Linux can recognize and list those cards as *ethX* interfaces. There are customization post-scripts for this hardware in Appendix B, “Scripts” on page 187.

4

Design considerations

This chapter provides design considerations for an IBM InfoSphere BigInsights infrastructure and its different options. This chapter covers key aspects to consider for improving a big data and analytics infrastructure.

At first, many products from various sources appear to meet big data and analytics requirements, but the fact is that this is not the full story. You are better off implementing solutions that have better features in them. So, in addition to processing and storage capacity, look for products with intuitive management that require less manual intervention and results in less administrative time.

This chapter covers the following topics:

- Important factors for sizing an InfoSphere BigInsights cluster
- IBM Spectrum Scale (formerly GPFS) considerations
- High availability considerations
- Throughput and bandwidth considerations
- Data volumes considerations
- Security, user authentication, and edge nodes
- Impact of use cases in design

4.1 Important factors for sizing an InfoSphere BigInsights cluster

An infrastructure to support big data must meet critical business requirements, such as grow and retain customers, transform financial processes and optimize operations, reduce fraud and manage risk, improve IT economics, and create business models.

For each data entry point, there are certain infrastructure requirements and design points, of which speed, access, and availability are the most important of all, but you must consider others. One key consideration is that there are certain design points that are better enabled by specific hardware and software infrastructure capabilities.

Some deployments use scale-out capabilities, and others use scale-up capability solutions:

1. Scaling out: Scale horizontal, which means adding more nodes to a system. The key concept of scale-out is to obtain distributed computing from hundreds of small computing systems that exceed the computing power of a traditional single processor type.
2. Scale up: Scale vertical, which means adding computing resources to a single-node system. Scaling up can include adding processors or memory to a single computing node. This vertical scale can be used to enable virtualization technology more efficiently.

   When choosing a scale-out or scale-up server infrastructure, consider the complexity and breadth of the analytic workloads. This area is where IBM provides an integrated, high-performance infrastructure that can include core servers, storage, networking, and systems software technologies.

4.1.1 Scalability

An InfoSphere BigInsights architecture is linearly scalable. When the maximum capacity of the existing infrastructure is reached, the cluster can be horizontally scaled-out by adding more data nodes and, if necessary, management nodes. As the capacity of the existing racks is reached, new racks can be added to the cluster.

Note: Some workloads might not scale linearly. For these cases, you might consider scale-up capabilities in your nodes.

When you design a new InfoSphere BigInsights architecture implementation, future horizontal scale-out is a key consideration in the initial design. You must consider the two key aspects of networking and management. Both of these aspects are critical to cluster operation and become more complex as the cluster infrastructure grows.

The networking model that is described in 4.5, “Throughput and bandwidth considerations” on page 116 is designed to provide robust network interconnection of racks within the cluster. As more racks are added, the predefined networking topology remains balanced and symmetrical. If there are plans to scale the cluster beyond one rack, initially design the cluster with multiple racks, even if the initial number of nodes might fit within one rack. Starting with multiple racks enforces proper network topology and prevents future reconfiguration and hardware changes. Also, as the number of nodes within the cluster increases, many of the tasks of managing the cluster also increase, such as updating node firmware or operating systems.

Building a cluster management framework as part of the initial design and proactively considering the challenges of managing a large cluster pays off in the end.

Platform Cluster Manager - Advanced Edition or Extreme Cloud Administration Toolkit (xCAT), an open source project that IBM supports, are scalable distributed computing management and provisioning tools that provide a unified interface for hardware control, discovery, and operating system deployment. In contrast to the command-line scripting environment that is provided by xCAT, Platform Cluster Manager - Advanced Edition provides a robust and easy to use GUI-based tool that accelerates time to value for deploying, managing, and monitoring a clustered hardware infrastructure. Within the InfoSphere BigInsights architecture, the IBM Power Systems server FSP and the cluster management network provide an out-of-band management framework that management tools, such as Platform Cluster Manager or xCAT, can use to facilitate or automate the management of cluster nodes.

Proactive planning for future scale-out and the development of a cluster management framework as part of the initial cluster design provides a foundation for future growth that minimizes hardware reconfigurations and cluster management issues as the cluster grows.

4.1.2 Availability

When you implement an InfoSphere BigInsights cluster in a Power Systems server, consider the availability requirements as part of the final hardware and software configuration.

Edge nodes: This calculation does not consider edge nodes. Based on the client's choice of an edge node, proportions can vary. Every two 1U edge nodes displace one data node, and every one 2U edge node displaces one data node.

Typically, Hadoop is considered a highly reliable solution. Hadoop and InfoSphere BigInsights preferred practices provide protection against data loss. Generally, failures can be managed without causing an outage. Redundancy can be added to make a cluster even more reliable, considering both hardware and software redundancies.

As shown in Figure 4-1, IBM Power Systems is highly virtualized, based on the IBM next generation POWER8 multi-core processors. Power Systems are designed around the first open server infrastructure that brings together the computing power, memory bandwidth, and I/O optimized for high-volume data processing that supports operating systems, such as Linux, AIX, and IBM i.

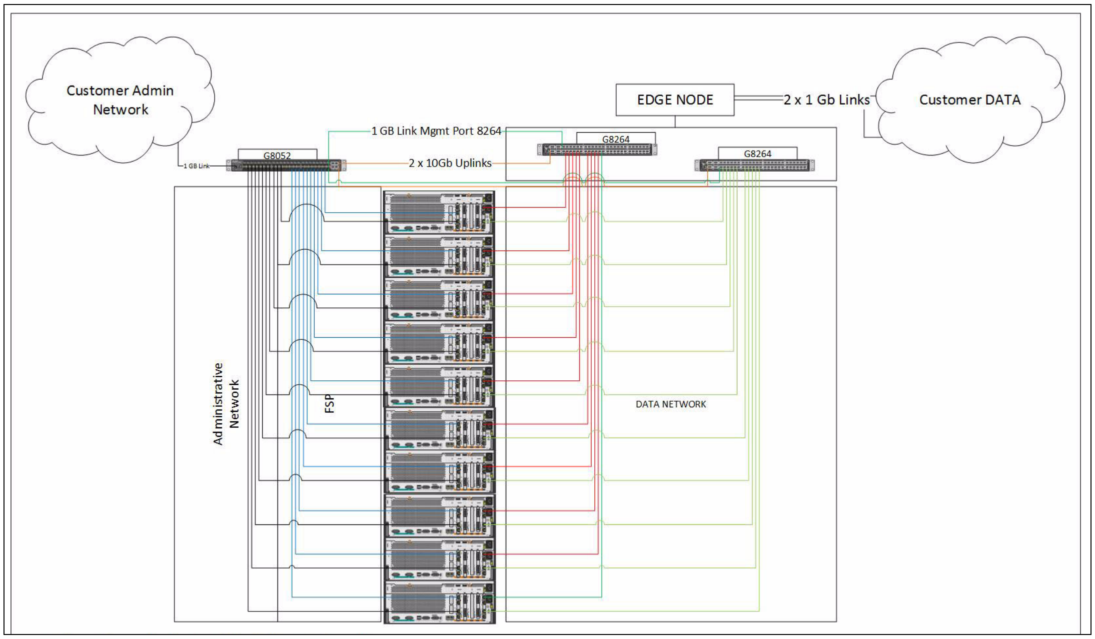


Figure 4-1 Network high availability - redundant configuration

There are multiple deployment styles for business analytics, including server nodes for business analytics, grids, and scalable server nodes to analyze data that is stored in data warehouses.

The proximity of data that is stored near or on these systems enhances overall throughput, as do fast server I/O links connecting to external storage network switches. For these capabilities, the Power platform has driven innovative big data analytics solutions, such as IBM Watson, which is now commercially available for clients looking to take their analytics to the next level by using cognitive computing.

4.2 Customizing the predefined configurations

The predefined configuration provides a baseline configuration for an InfoSphere BigInsights cluster, and provides modifications for an InfoSphere BigInsights cluster running HBase. The predefined configurations represent a baseline configuration that can be implemented *as is* or modified based on specific client requirements such as lower cost, improved performance, and increased reliability, as shown in Figure 4-2 on page 113.

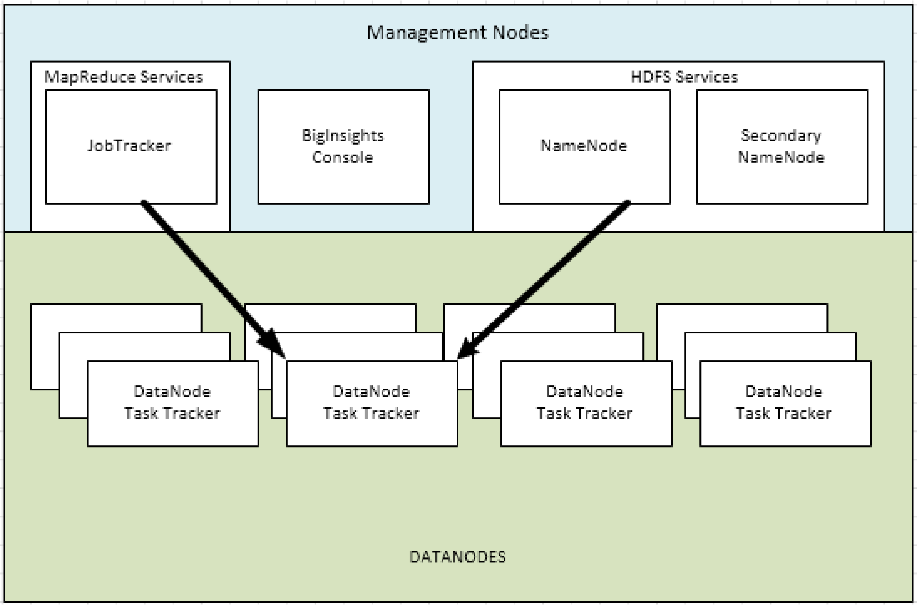


Figure 4-2 IBM BigInsights component model

When you consider modifying the predefined configuration, you must understand key aspects of how the cluster will be used. In terms of data, you must understand the current and future total data to be managed, the size of a typical data set, and whether access to the data will be uniform or skewed. In terms of ingest, you must understand the volume of data to be ingested and ingest patterns, such as regular cycles over specific periods and bursts in ingest.

Also, consider the data access and processing characteristics of common jobs and whether query-like frameworks, such as IBM BigSQL, are used.

When designing an InfoSphere BigInsights cluster infrastructure, conduct the necessary testing and proof of concepts against representative data and workloads to ensure that the proposed design achieves the necessary success criteria. The following sections provide information about customizing the predefined configuration. When considering customizations to the predefined configuration, work with a systems architect who is experienced in designing InfoSphere BigInsights cluster infrastructures.

4.3 IBM Spectrum Scale (formerly GPFS) considerations

Spectrum Scale File Placement Optimizer (Spectrum Scale-FPO) is a high-performance, cost-effective storage methodology that started as a clustered file system and has evolved into much more than a file system. Today, Spectrum Scale is a full-featured set of file management tools, including advance storage virtualization, integrated high availability, automated tiered storage management, and the performance to manage effectively large quantities of file-based data.

Spectrum Scale supports various application workloads and is effective in large and demanding environments. Spectrum Scale is installed in clusters, and supports big data, analytics, gene sequencing, digital media, and scalable file serving. All indications are that InfoSphere BigInsights might bring more unstructured and file-based data into the application.

Figure 4-3 shows the architecture of Spectrum Scale-FPO.

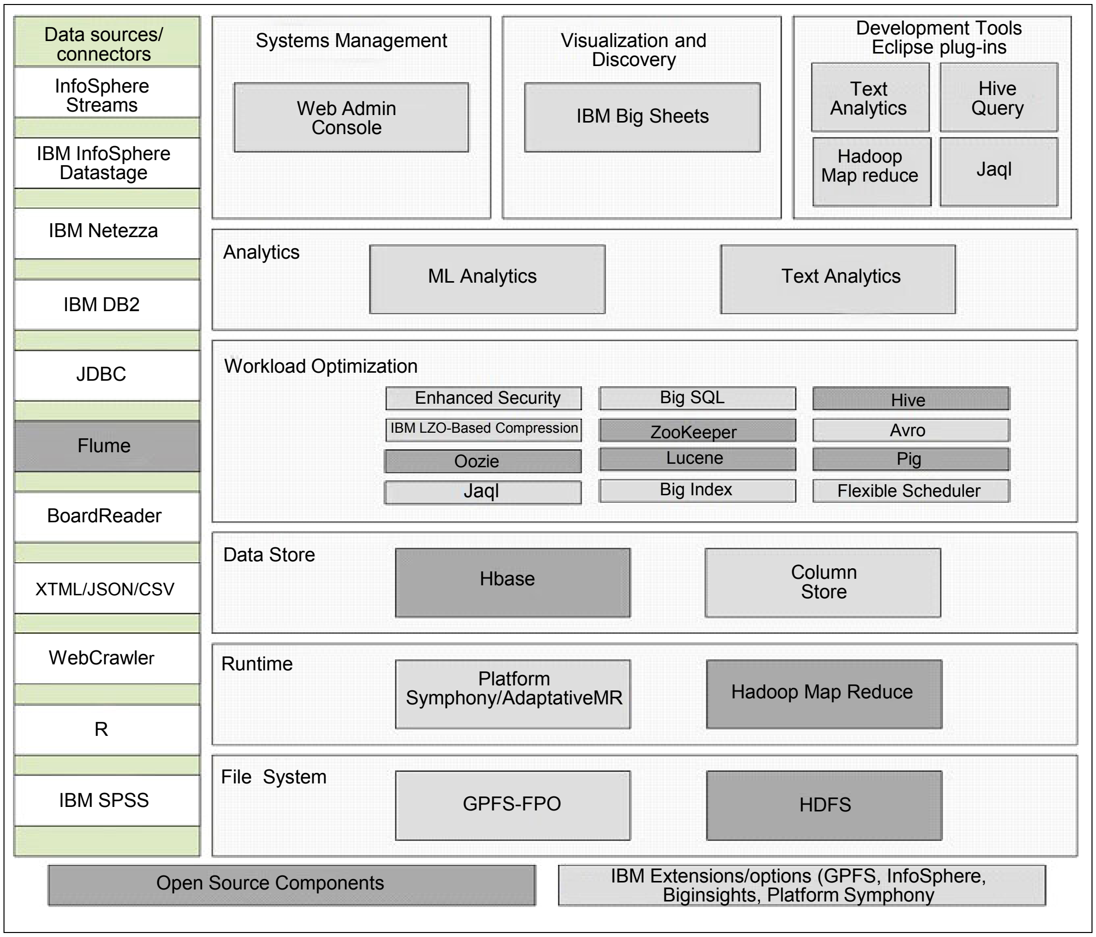


Figure 4-3 IBM Spectrum Scale-FPO

For high-performance computing environments, IBM Spectrum Scale offers a distributed, scalable, reliable, and single namespace file system. Spectrum Scale-FPO (File Placement Optimizer) is based on a shared-nothing architecture so that each node on the file system can function independently and be self-sufficient within the cluster. Typically, Spectrum Scale-FPO can be a substitute for HDFS, removing the need for the HDFS NameNode, Secondary NameNode, and DataNode services.

However, in performance-sensitive environments, placing Spectrum Scale metadata on higher-speed drives might improve performance of the Spectrum Scale file system.

Spectrum Scale-FPO has significant and beneficial architectural differences from HDFS. HDFS is a file system that is based on Java that runs on top of the operating system file system and is not POSIX-compliant. Spectrum Scale-FPO is a POSIX-compliant, kernel-level file system that provides Hadoop with a single namespace, distributed file system with performance, manageability, and reliability advantages over HDFS.

As a kernel-level file system, Spectrum Scale is free from the impact that is incurred by HDFS as a secondary file system, running within a JVM on top of the operating systems' file system. As a POSIX-compliant file system, files that are stored in Spectrum Scale-FPO are visible to authorized users and applications by using standard file access/management commands and APIs. An authorized user can list, copy, move, or delete files in Spectrum Scale-FPO by using traditional operating system file management commands without logging in to Hadoop.

Additionally, Spectrum Scale-FPO has significant advantages over HDFS for backup and replication. Spectrum Scale-FPO provides point-in-time snapshot backup and off-site replication capabilities that enhance cluster backup and replication capabilities.

When using Spectrum Scale-FPO instead of HDFS as the cluster file system, the HDFS NameNode and Secondary NameNode daemons are not required on cluster management nodes, and the HDFS DataNode daemon is not required on cluster data nodes. Equivalent tasks are performed by Spectrum Scale in a distributed way across all nodes in the cluster, including data ones. From an infrastructure design perspective, including Spectrum Scale-FPO can reduce the number of management nodes that are required.

Because Spectrum Scale-FPO distributes metadata across the cluster, no dedicated name service is needed. Management nodes within the InfoSphere BigInsights predefined configuration or InfoSphere BigInsights HBase predefined configuration that are dedicated to running the HDFS NameNode or Secondary NameNode services can be eliminated from the design. The reduced number of required management nodes can provide sufficient space to allow for more data nodes within a rack.

For more information about implementing IBM Spectrum Scale-FPO in an InfoSphere BigInsights solution, see *Deploying a big data solution using IBM Spectrum Scale-FPO*, found at:

http://ibm.co/1NBnGTj

4.4 High availability considerations

This section provides information about high availability considerations.

4.4.1 Designing for high availability

Designing for high availability entails assessing potential failure points and planning so that potential failure points do not impact the operation of the cluster. Whenever you address enhanced high availability, you must understand and consider the trade-offs between the cost of outage and the cost of adding redundant hardware components.

Within an InfoSphere BigInsights cluster, several single points of failure exist: A typical Hadoop HDFS is implemented with a single NameNode service instance. A couple of options exist to address this issue. InfoSphere BigInsights V3.0 supports an active/standby redundant NameNode configuration as an alternative to the standard NameNode/Secondary NameNode configuration.

4.5 Throughput and bandwidth considerations

This section describes networking considerations, including throughput and bandwidth.

4.5.1 The data network

The data network is a private 10 GbE cluster data interconnect among data nodes that are used for data access, moving data across nodes within the cluster, and ingesting data into HDFS. The InfoSphere BigInsights cluster typically connects to the client's corporate data network by using one or more edge nodes. These edge nodes can be IBM Power 740 or an IBM Power 750 servers, other Power Systems servers, or other client-specified servers. Edge nodes act as interface nodes between the InfoSphere BigInsights cluster and the outside client environment (for example, data ingested from a corporate network into a cluster). Not every rack has an edge node connection to a client network. Data can be ingested into the cluster through edge nodes or through parallel ingest.

4.5.2 Administrative/management network

The administrative/management network is a 1 GbE network that is used for in-band operating system administration and out-of-band hardware management. In-band administrative services such as Secure Shell (SSH) or virtual network computing (VNC) that run on the host operating system allow administration of cluster nodes. Out-of-band management uses the Power Systems server Hardware Management Console (HMC) connection for hardware management.

The HMC ports on the Power Systems server allow for hardware management of the cluster nodes, which is a requirement for systems deployment on them. Based on client requirements, the administration and management links can be segregated into separate virtual LANs (VLANs) or subnets. The administrative/management network is typically connected directly to the client’s administrative network.

4.6 Data volumes considerations

Data volumes are an inherent part of a big data workload and solution, and one of the biggest impacts in terms of design considerations. For data volumes, you must consider not only the original data but the intermediate data that the MapReduce process produces to accomplish its purpose.

Data volumes impact the space to be allocated for storing the data itself and the networking design that enables the replication and the data transference among the different nodes in charge of the different stages of the Hadoop job, as shown in Figure 4-4 on page 117.

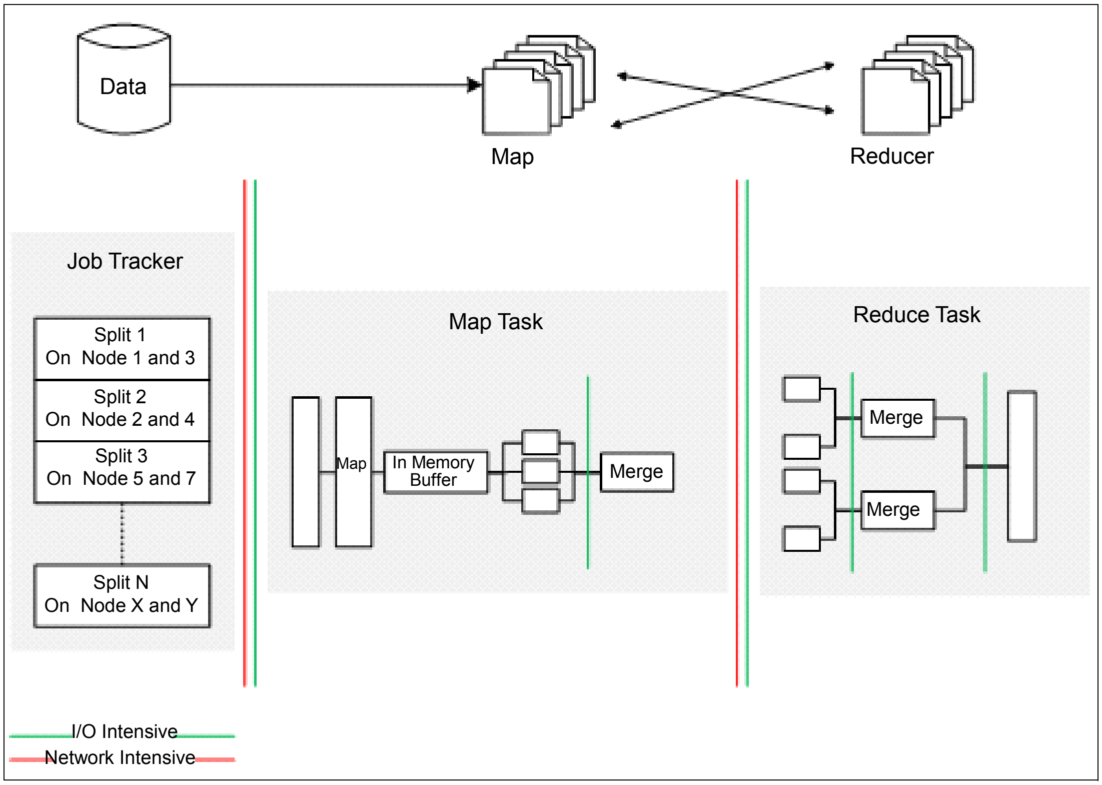


Figure 4-4 Hadoop stages and impact

Besides the networking impact on big data volumes, the main design considerations are related to Spectrum Scale components:

- Block size: Spectrum Scale supports sizes of 16 KB - 16 MB, and defaults to 256 KB. This configuration determines the ratio for any I/O operation because represents the largest unit for a single I/O operation.
- Chunk size: On an FPO configuration, a chunk is a grouping of blocks, which defines the minimal size into which a file is divided. Each chunk is then distributed among the cluster. The chunk size is determined multiplying the *block group factor* by the *block size*. The block group factor is a value 1 - 1024.
- Storage pool: This item is basically a collection of disks in Spectrum Scale that allows you to group storage devices according to your requirements and the characteristics of the data being handled.
- Replication: This item must be properly designed and considered because it impacts not only storage space but I/O and networking on the entire solution. The replication can be set up either granularly at file level or at the entire file system level.

Tuning according to these characteristics maximizes the efficiency and effectiveness of the usage of your resources in terms of how the data is allocated among the cluster, and how efficient are the I/O and network operations within the large volumes of data.

4.7 Security, user authentication, and edge nodes

From a security and privacy perspective, big data is different from other traditional data, and requires a different approach, although many of the existing methodologies and preferred practices can easily be extended to support the big data paradigm.

This section begins by defining data privacy because it is a core concept. The International Association of Privacy Professionals (IAPP) defines data or information privacy as "The claim of individuals, groups, or institutions to determine for themselves when, how and to what extent, information about them is communicated to others."[1]

Privacy laws throughout the world are a patchwork of laws and regulations. Most contain a shared set of principles that are known as Fair Information Practice Principles (FIPPs) that govern notice, choice and consent, collection, use, access, disposal, and program management of information.

From this definition, you can start to see the complexity that is involved in maintaining this privacy within a big data scenario, and even though on the surface big data appears to have similar risks and exposures as traditional data, there are several key areas where it is different and that you must take into account when designing this type of environment. More data translates into higher risk of exposure if there is a breach.

New types of data are uncovering new privacy implications, with few privacy laws or guidelines to protect the information. Examples include the connected home and digital meters for monitoring electricity usage (eMeters), cell phone beacons that broadcast physical location, health devices, such as medical, fitness and lifestyle trackers, and telematics data that tracks automobile locations.

There is an increased security risk of revealing privacy data and data that exists under compliance regulations, such as the Health Insurance Portability and Accountability Act (HIPAA), Payment Card Industry Data Security Standard (PCI DSS), and the Sarbanes-Oxley Act (SOX), and critical internal company information, such as intellectual capital (software programs, algorithms, and so on).

As a technology, many areas of Hadoop are still evolving, which might create more exposures to the new data. For example, data linkage and combined sensitive data that combines multiple data sources creates an unanticipated sensitive data exposure, often with a lack of awareness on your part.

These are all issues that you must consider to ensure that your big data solution is implemented in a way that avoids encountering these pitfalls. A good example is the anonymization of production data. Much of the value of big data is in uncovering patterns that do not require identification of the individual. As a result, organizations are increasingly using anonymization and de-identification to remove individual identifiers while still gaining utility from the data.

[1] Source: https://privacyassociation.org/resources/glossary

To ensure that you have a secure environment and that there is proper data privacy, start by applying proper data governance to the data (structured or not), and a security framework that provides a set of preferred practices, operational standards, and controls to guide your organization in assessing vulnerabilities, planning and implementing secure computing, which reduces the risk of data and systems intrusion. There are several security frameworks in use today, such as Control Objectives for Information and Related Technology (COBIT), NIST, and International Standards Organization (ISO 27000). Of the three, ISO 27000 is the only framework that contains a standard by which companies can be certified. It is also multinational in nature. You must verify with your organization as to which security framework you are using, and have a clear understanding of what data privacy means.

You have many principles that come from common sense and common practice in the realm of structured data. Their origins can be traced to good data management practices that were developed decades earlier. They are an extension of those practices, which often provide the foundation for many regulatory requirements, and are extended here to include the principles that are necessary for big data.

This list is a good starting point if your organization does not have a governance program/structure.

Data governance practices must allow for nimble responses to changes in technology, customer needs, and internal processes, including the following ways:

- Responding to emergent technology
- Considering how rules and bureaucratic controls might reduce productivity
- Standardizing where doing so does not significantly impact the work flow

Here are some items to consider:

- Be compliant: You must fulfill the obligations of external regulations from international, national, regional, and local governments in concert with policies and regulations that might exist within the organization.
- Manage quality: Because information and data is the core of your business, the quality of the content of that information is paramount to your continued success.
  - For big data, the data must be fit for purpose; context might need to be hypothesized for evaluation.
  - Quality does not imply cleansing activities, which might mask the results.
- Map your information: Understanding your complete business and the flow of information across all processes enables you to succeed with your first two principles. This requires the capture and recording of data at rest and data in motion throughout the organization.
  - Provenance and lineage are still, if not more important, in big data.
  - What is the source? This is a crucial question to validate analytics results as fit for purpose.
  - Where is it going or has gone? This is a crucial question for ongoing system maintenance and support efforts.
  - Enables audit reporting, which is considered an essential regulatory compliance function.

- For unstructured data, it is especially important to manage meaning: Understanding the language that you use and managing it actively reduces ambiguity, redundancy, and inconsistency, which relates directly to the quality of information.
  - Big data might not have a logical data model that is provided, so, as early as possible, any structured data should be mapped to the enterprise model.
  - Big data still has context, so modeling becomes increasingly important to creating knowledge and understanding.
  - Determine the degree of ambiguity that you can tolerate in reporting and analytics. Measure and monitor this ambiguity as a metric.
  - Plan for meaning to change over time. The definitions evolve over time and the enterprise must plan to manage the shifting meaning.
- Manage classification: It is critical for the business/steward to classify the overall source and the contents within as soon as it is brought in by its owner to support information lifecycle management, access control, and regulatory compliance.
  - Public versus private.
  - Retention period/schedule.
  - Security level.
  - Applicable regulatory controls (for example, PII, PCI, and illegal content).
  - Non-destructive testing to make assessment/classification in a staging zone (for data at rest) or threat assessment/function in stream (for data in motion).
  - If you do not have positive control over the incoming stream (for example, a Twitter feed is uncontrolled), then you must have components to monitor and classify what crosses the boundary. This is needed to exclude any data that does not meet the acceptance criteria.
- Protect information: Protection of data quality and access is essential to the ability to maintain the trust of your customers and their customers.
  - Your information protection must not be compromised for the sake of expediency, convenience, or deadlines (for example, simple data exploration). If conditions warrant them, exceptions should be decided by the appropriate management team and documented.
  - Accept as a fact the difficulty in predicting how new risks manifest themselves.
  - Protect not just what you bring in, but what you join/link it to, and what you derive. Your customers will fault you for failing to protect them from malicious links.
  - The enterprise must formulate the strategy to deal with more data, longer retention periods, more data subject to experimentation, and less process around it, all while trying to derive more value over longer periods.

4.7.1 Security preferred practices in non-relational data stores

Most security features that are built into relational databases are not present in big data platforms. Many of the third-party security tools do not work with big data because these non-relational data management platforms are architecturally different, so you must rethink how you approach database security.

Big data is composed of a cluster of servers, each with a slice of the stored data. Applications do not speak with a single node in this cluster; instead, they communicate with hundreds or even thousands of nodes. With the multi-node architecture, the sheer scale, variety, and velocity of data, the capabilities of traditional security products are underpowered to secure fully the systems, as they either have trouble scaling (row-level encryption, masking, and packet analysis) or they simply do not work (content filtering and query monitoring) with many NoSQL clusters.

It is important to understand your data and workflows to ensure that you can control your risk and secure the environment to the extent that is required.

One common way to enforce security is to place the entire cluster onto its own network and tightly control logical access through firewalls and access controls. This setup has the benefit of being easy to deploy and maintain, but it provides no security for the data itself, and so it is fragile because it does not help you ensure that the users that have been able to get through the firewall are not misusing, tampering with, or viewing the data in the cluster. This is a low-cost approach and works well if your data does not require a high level of security.

Another approach is to use extra tools for cryptographically enforced access control and secure communication, which might include SSL/TLS (secure communication), Kerberos for node authentication, transparent encryption for data-at-rest security, identity and authorization (groups, roles), and so on.

This second approach is more difficult, and often more expensive, requiring the setup of several security functions that are targeted at specific risks to the database infrastructure. However, it does secure the cluster from attackers in a much better way than the first approach, and therefore must be evaluated when deciding the level of security that your data requires.

4.7.2 Securing IBM Spectrum Scale

Currently, a prerequisite of any IBM Spectrum Scale installation is that all the nodes in the cluster must be able to communicate with each other by using the root user ID, and that this is done without a password to facilitate automated tasks. But this configuration represents a security exposure because security preferred practices forbid that the root user can log in remotely. As part of the hardening process of the environment, you must allow the Spectrum Scale cluster nodes to continue communicating with each other by using the root ID and at the same time block this type of connection to other users and nodes. This task is done by making some changes that can be applied to the Secure Shell (SSH) configuration file on each one of the Spectrum Scale nodes so that remote login operations with the root user are established only from the Spectrum Scale nodes. So, by implementing a conditional configuration of the **`PermitRootLogin`** parameter through the **`Match`** directive in SSH, the risk of allowing remote root login connections from non-authorized or non Spectrum Scale cluster servers or networks is mitigated, and access is granted only to those who strictly need it.

To get more details about securing IBM Spectrum Scale, see the following website:

http://www.ibm.com/developerworks/aix/library/au-aix-modifying-ssh-configuration/

4.7.3 Securing data storage and transaction logs

The storage and transaction logs must be protected for integrity and privacy, but also for establishing completeness, with an emphasis on availability.

Endpoint input validation and filtering

Strong authentication can be achieved through X.509v3 certificates, and potentially by using a SAFE bridge in lieu of general PKI. Consider the following items:

- Manage long-term requirements: Policies and standards are the mechanism by which management communicates their long-range business requirements. These are essential to an effective governance program.
  - Projects that impact the manner in which data is captured, stored, or moved must refer to all applicable policies and standards. These policies and standards must be treated as equals to the business requirements.
  - If you are initiating a big data project, you might need a big data policy (or revisit existing policies) to see what applies and what unique exposures/value opportunities are created. Revisit the rules because big data is changing the game.
  - Factor in that big data implies less capacity for human intervention and identify new approaches for responding.
  - You must build in interruption processes to avoid runaway processes and monitor for such events. Teams must plan for feedback mechanisms to control the rate of change to prevent becoming overwhelmed by a runaway process or decision engine.
- Control third-party content: Third-party data plays an expanding role in big data. There are three types (defined below), and governance controls must be adequate for the circumstances. The controls must consider applicable regulations for the operating geographic regions; therefore, you must understand and manage these obligations.
  - Outsourced delivery (proxy for you): Contracts must reflect your policies, as you are still responsible for the content of the data (for example, traceability).
  - Providers of data (publishing or selling to you): What is your responsibility (for example, a bad action that is based on bad data) versus that of the third-party provider (wrong emergency response)? The responsibility belongs with the third party, depending on the terms and conditions.
  - Subscribers of data (buying from you): The responsibility belongs with the deliverer, depending on the terms and conditions.
- Remember, your people are the means to quality, security, and governance/compliance in your data:
  - Continuously generate awareness as a critical task for the governance program. As an example, many companies offer periodic training about the protection of personal information.
  - Audit and measure the organization against these principles routinely and include the results in performance monitoring for individuals and departments.

4.8 Impact of use cases in design

To address the analysis of the impact in the design according to the workload, and because big data is meant to handle a wide variety of workloads, many of these workloads are not yet defined. You must clarify the characteristics of the workloads and the weight of each on the different components of a big data solution.

4.8.1 Workload characteristics definition

An analytic workload exhibits one or more of the following characteristics, each of which elevates a given workload's degree of difficulty, and has its level of impact on the different components of the big data solution:

Data volume
Refers to the cardinality and size of data to be handled by the specific workload to be applied on this big data solution.

Data model complexity
Refers to the complexity of the data that is handled, structured or unstructured. Either related to the number of objects or the data variety itself.

Variable and unpredictable traversal paths, patterns, and frequencies
Refers to the variance of the data, how difficult is to find, identify or compute relationships or patterns. Required to accomplish the workload purpose.

Set-oriented processing and bulk operations
Refers to the ability to handle multiple rows of data as a set at once, which grows almost exponentially, impacting the complexity of the workload either live or in batch offline operations, such as ETLs.

Multi-step, multi-touch analysis algorithms
Refers to the need of several operations or iterations on the same set of data to properly get a valid result to cover the purpose of the workload.

Complex computation
Refers to the complexity of the algorithms that are performed to accomplish the defined workload purpose. This characteristic is influenced and impacted by the previous workload characteristics, which increments the workload and impacts the design.

Temporary or intermediate staging of data
Refers to the need to keep intermediate results to fulfill and accomplish the workload purpose. This scenario is common on workloads that require complex computations on a complex data model with multi-step characteristics. This characteristic is important to the amount of intermediate data that is generated and how efficiently this intermediate data can be accessed to generate the final result.

Change isolation/data stability implications
Refers to the impact of the changes on the working data sets, either intended and normal data changes, or errors and failures on the infrastructure, on the overall workload algorithm and purpose.

Table 4-1 shows the relationship among the workload characteristics and its impact on the solution components.

Table 4-1 Workload characteristics impact

Workload characteristic	Network	Storage/Size	Storage/Redundancy	Storage/Tiering
Data volume	High	High	High	High
Data model complexity	Medium	High	High	Medium
Variable and unpredictable traversal paths, patterns, and frequencies	Medium	Medium	Medium	Medium
Set-oriented processing and bulk operations	Medium	Medium	Medium	Medium
Multi-step, multi-touch analysis algorithms	Medium	Medium	Medium	Medium
Complex computation	Medium	Low	Low	Medium
Temporary or intermediate staging of data	Medium	High	Medium	High
Change isolation/data stability implications	High	Medium	High	High

As shown in Table 4-1, the relationship between the workload characteristics and the impact on the different components is what drives the different design considerations for the big data solution.

Note: Data volume is the main characteristic that impacts the components in a big data solution, which must be reflected in the detailed design that is described in 4.3, “IBM Spectrum Scale (formerly GPFS) considerations” on page 113.

5

Solution customization

This chapter provides details about how to customize solutions, and use these solutions for clients environments.

This chapter covers the following topics:

- IBM Elastic Storage Server
- IBM Platform Symphony MapReduce
- File system: Spectrum Scale / HDFS (architectural changes when not using Spectrum Scale)
- InfoSphere BigInsights high availability
- Security: InfoSphere BigInsights user authentication

5.1 IBM Elastic Storage Server

IBM Spectrum Scale File Placement Optimizer (Spectrum Scale-FPO) is deployed on commodity hardware, including internal storage-rich servers. A fact of such commodity hardware is that its failure rate is higher than enterprise system servers. Node failure also results in all disks inside these nodes to be out of service. In addition, such a commodity hardware deployment model is inflexible regarding the growth of compute resources and storage resources.

IBM Elastic Storage™ Server (ESS) addresses these enterprise-level requirements. The ESS is a big data storage system that combines IBM Power Systems servers, storage enclosures, and disks along with Spectrum Scale and its native RAID technology. Therefore, ESS provides analytic and technical computing storage and data services for elastic storage workloads. ESS scales in a building block approach where adding more storage servers adds to the overall capacity, bandwidth, and performance all within a single name space. The ESS modern, declustered RAID technology is designed to recover from multiple disk failures in minutes, versus hours and days as in older technology. 8+2 and 8+3 RAID protection and platter-to-client data protection are included with ESS.

To integrate ESS and InfoSphere BigInsights, create a Spectrum Scale file system that is based on ESS first, and deploy it on the file system with an option to *install on an existing file system* during installation.

5.2 IBM Platform Symphony MapReduce

To enable Platform Symphony Adaptive MapReduce, see “Installation wizard” on page 47. If you do not want to enable Platform Symphony, you do not need to act. The high availability of JobTracker services requires Platform Symphony Adaptive MapReduce to be enabled.

Platform Symphony MapReduce sets the number of slots by default to the number of effective/logical processors, including hyper threading/SMT. For POWER8 servers that support up to eight threads per core (SMT8), the default number of slots might be too high if the node does not have enough disks to sustain the I/O requests. For example, if the server has 16 physical cores and 24 disks, you might consider a test performance of the hardware running at SMT4 mode (64 effective processors) instead of SMT8 mode (128 effective processors) to avoid an I/O bottleneck.

For a POWER8 server with more than 40 slots per node, the default setting for shuffle service (MRSS) in Platform Symphony MapReduce is conservative. You can get better performance for Platform Symphony MapReduce applications by increasing both `PMR_MRSS_WORKING_THREADS` and `PMR_MRSS_DATA_WRITE_WORKING_THREADS_NUMBER` to the number of slots per node. These parameters are defined in the `$EGO_TOP/eservice/esc/conf/services/mrss.xml` file. Example 5-1 shows the settings on nodes with number of slots=64.

Example 5-1 Customize MRSS parameters in Platform Symphony MapReduce

```
<ego:EnvironmentVariable
name="PMR_MRSS_WORKING_THREADS_NUMBER">64</ego:EnvironmentVariable>
<ego:EnvironmentVariable
name="PMR_MRSS_DATA_WRITE_WORKING_THREADS_NUMBER">64</ego:EnvironmentVariable>
```

For the change to be effective, stop and start the MRSS service as the InfoSphere BigInsights administrator. See Example 5-2.

Example 5-2 Update the MRSS service

```
egosh service stop MRSS
egosh service start MRSS
```

5.3 File system: Spectrum Scale / HDFS (architectural changes when not using Spectrum Scale)

The selection and installation of the underlying file system should be done when the rest of the Hadoop cluster is installed. The installation of Spectrum Scale with InfoSphere BigInsights is described in Appendix A, “Integration and configuration for IBM Spectrum Scale, Hadoop, and IBM Platform Symphony” on page 157. The installation of HDFS is also done during the InfoSphere BigInsights installation.

If you are going to install an HDFS file system, there are a few differences in the prerequisites:

- The cluster requires a minimum of one extra management server for the NameNode. However, it is highly recommended that you have two for high availability.
- As with Spectrum Scale, all of the disks should be available in a JBOD fashion. However, with HDFS, the disks should be formatted with ext4 and mounted on the node.

Then, during the installation process, you can select to install HDFS instead of Spectrum Scale by using the web-based tool.

5.4 InfoSphere BigInsights high availability

The level of high availability that can be implemented depends upon the type of file system that you choose to implement in your solution.

5.4.1 IBM Spectrum Scale

With a Spectrum Scale installation, you can configure high availability for the following services:

- Platform Symphony Adaptive MapReduce (JobTracker)
- Zookeeper
- HBase

Note: High availability is not required for the NameNode in a Spectrum Scale installation because a Spectrum Scale and InfoSphere BigInsights installation does not use a NameNode.

Platform Symphony MapReduce

To enable the installation of Platform Symphony MapReduce, you must edit the `$EXTRACTION_DIR/install.properties` file and change the `AdaptiveMR.Enable` property from `false` to `true`. You must do this task before you start the InfoSphere BigInsights installer. Three *dedicated* management nodes are required to set up high availability for the JobTracker. These nodes cannot host any other management services. The nodes also require access to a shared disk. Ideally, there should be a local disk on each of the three nodes that you plan to use. At the Add Nodes page in the installation web user interface, ensure that you are adding disks for the management nodes to use for Spectrum Scale. Then, when you reach the Components 2 page, select the check box for high availability and add all three nodes. You can then elect which node to be the primary JobTracker by specifying it in the JobTracker box.

Zookeeper and HBase

If multiple Zookeeper or HBase management servers are required, then these servers should be added in their respective component sections of the installation web user interface. No further action is required.

5.4.2 HDFS

With an HDFS installation, you can configure high availability for the following services:

- NameNode
- Zookeeper
- HBase
- JobTracker
- Platform Symphony Adaptive MapReduce (JobTracker)

> **Note:** Electing to use Platform Symphony MapReduce replaces the need for an Apache MapReduce JobTracker.

NameNode

To have high availability of the NameNode, use a second management server. Although it is a preferred practice that this node hosts no other services, this is not a requirement. Activate the web-based user interface and when you reach the Components 2 window, add multiple host names as NameNodes to configure high availability for this node. If you have Platform Symphony MapReduce turned on, select the **Configure High availability** box at the upper left and add your NameNode FQDN and the IP address that you want to use for the NameNode. Then, on the same page, add multiple host names as NameNodes to configure high availability for this node.

Zookeeper and HBase

If multiple Zookeeper or HBase management servers are required, then they should be added in their respective component sections of the installation web user interface. No further action is required.

JobTracker

To enable high availability of the JobTracker when you reach the Component 2 window of the installation, check the **Configure High Availability** option at the upper left. This option provides the option to enter “Quorum Journal Manager Nodes”. You must enter three nodes to judge quorum on where the JobTracker should exist. Then, under the JobTracker tab, add multiple JobTrackers.

Platform Symphony MapReduce

To enable the installation of Platform Symphony MapReduce, you must edit the $EXTRACTION_DIR/install.properties file and change the AdaptiveMR.Enable property from false to true. You must do this before you start the InfoSphere BigInsights installer. When you reach the Component 2 window of the installation, select the **Configure High Availability** option at the upper left. Enabling this option provides the option to enter five additional options. The first two relate to HA of the NameNode. The final three configure Platform Symphony MapReduce HA. The *High Availability Nodes* are the nodes that you want to be available for the Platform Symphony MapReduce service. These nodes should not be running any other services. The NFS server information and mount point are required because this is where InfoSphere BigInsights stores a file with details about the jobs running so that, if the JobTracker must fail over, job progress is not lost.

5.5 Security: InfoSphere BigInsights user authentication

IBM InfoSphere BigInsights V3.0.0.1 supports two authentication options: Flat file security and LDAP security. Both of these options use the Pluggable Authentication Module (PAM) to pass authentication to the operating system.

5.5.1 Using flat file security

InfoSphere BigInsights with flat file security has two files in the $BIGINSIGHTS_HOME/console/conf/security directory: biginsights_user and biginsights_group. If you are going to use this form of authentication, then every user that you want to allow to sign in to the web console should be added to both of these files. The biginsights_user file contains the user name and password. It is worth noting that the password is stored in human-readable format. The biginsights_groups file maps each user (or group) to a InfoSphere BigInsights role. Information about the various InfoSphere BigInsights roles can be found at the following website:

http://ibm.co/1IRaMxz

5.5.2 Using LDAP security

InfoSphere BigInsights with LDAP security has only the $BIGINSIGHTS_HOME/conf/install.xml file. The groups that are associated with each user, which are retrieved from LDAP, should be entered into this file to map the group to a user role in the security section, as shown in Example 5-3. Multiple users or groups should be separated by a comma.

Example 5-3 InfoSphere BigInsights and LDAP security file map

```
<security>
        <authentication>ldap</authentication>
        <biginsightssystemadministrator>
            <group>biadmin</group>
            <user>biadmin</user>
        </biginsightssystemadministrator>
        <biginsightsdataadministrator>
            <group>group1,group2</group>
            <user>user1,user2</user>
        </biginsightsdataadministrator>
        <biginsightsapplicationadministrator>
            <group>biadmin</group>
        </biginsightsapplicationadministrator>
        <biginsightsuser>
            <group>biadmin</group>
        </biginsightsuser>
```

When the installation completes, it creates a `jpam` file in `/etc/pam.d`. This file is the link between the InfoSphere BigInsights console and the operating system authentication. To use LDAP authentication, the operating system should be configured to authenticate against the LDAP server. Then, by using the PAM file, the application passes the authentication to the LDAP server.

6

Cluster management

This chapter provides information about how to perform cluster operations within IBM Platform Cluster Manager Advanced Edition, which includes the following operations:

- Adding a node into Platform Cluster Manager - Advanced Edition
- Removing an existing node from Platform Cluster Manager - Advanced Edition
- Monitoring nodes with Platform Cluster Manager - Advanced Edition

In addition, this chapter provides information about how to grow and shrink an existing InfoSphere BigInsights cluster that is managed by Platform Cluster Manager - Advanced Edition.

Finally, this chapter provides an overview of how node failures are handled within an InfoSphere BigInsights cluster within a Platform Cluster Manager - Advanced Edition environment.

This chapter covers the following topics:

- Managing nodes in a Platform Cluster Manager - Advanced Edition environment
- Managing InfoSphere BigInsights cluster nodes within Platform Cluster Manager - Advanced Edition

6.1 Managing nodes in a Platform Cluster Manager - Advanced Edition environment

As described in 2.3.1, “Cluster management software” on page 29, you can use Platform Cluster Manager - Advanced Edition to automate the deployment of an InfoSphere BigInsights cluster in your environment. To do so, you must know how to import nodes into Platform Cluster Manager - Advanced Edition, and also how to remove them. Platform Cluster Manager - Advanced Edition is a cluster management software that provides flexibility for managing servers, in particular servers that are intended to create the cluster. The next sections show you how to perform fundamental cluster management.

6.1.1 Adding nodes to a Platform Cluster Manager - Advanced Edition environment

The first action to take when adding a server into a Platform Cluster Manager - Advanced Edition environment is to make sure that it can be seen through the Flexible Service Processor (FSP) network that you set up and configured while installing Platform Cluster Manager - Advanced Edition itself. For more information about the FSP network and how to set it up with Platform Cluster Manager - Advanced Edition, see “Networking” on page 18 and 3.3, “Platform Cluster Manager - Advanced Edition” on page 73.

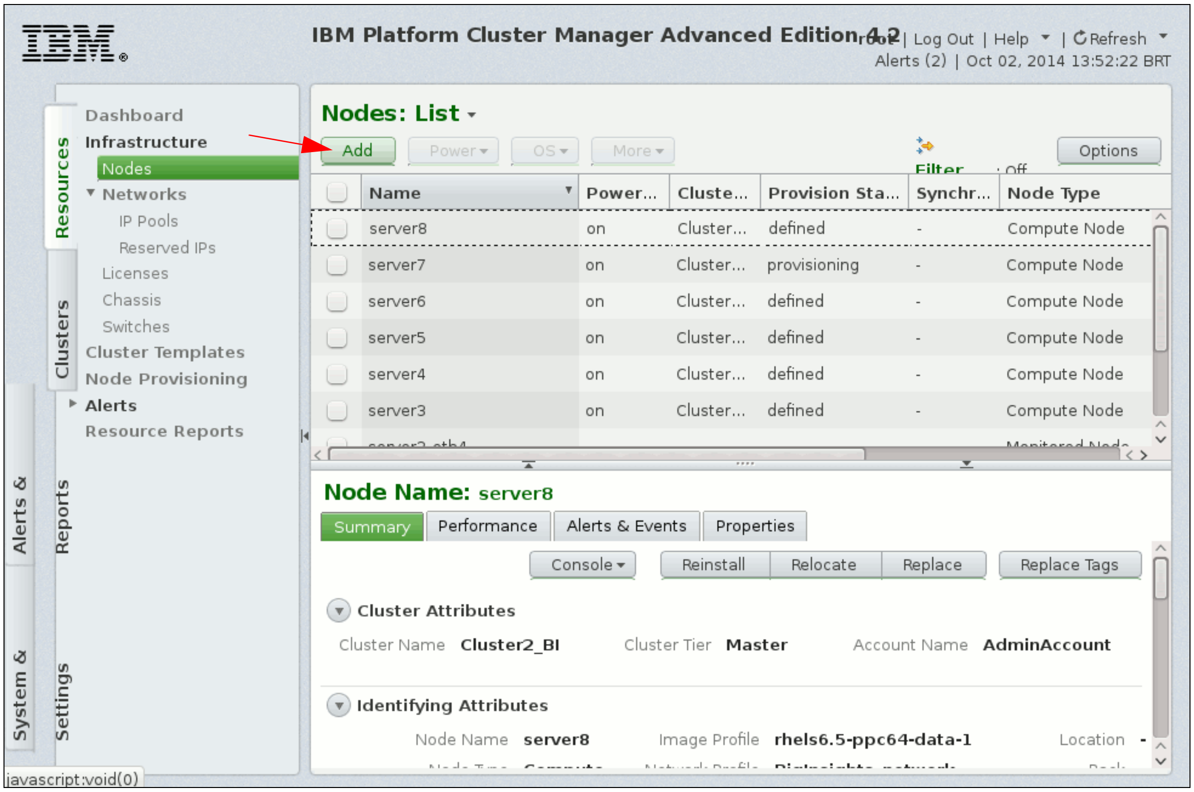
When you connect a new server to the FSP network, the Flexible Service Processor on the Power Systems hardware makes a DHCP request that is answered by Platform Cluster Manager - Advanced Edition. Platform Cluster Manager - Advanced Edition cannot manage the new hardware until it can communicate over the FSP network with the new node. With a proper DHCP setup in the FSP network, after a short time Platform Cluster Manager - Advanced Edition can communicate with the new node. If the communication does not happen, then start troubleshooting your FSP network connections or DHCP setup.

Run the **lshwconn** command, as shown in Example 6-1, to check whether your new server can communicate with Platform Cluster Manager - Advanced Edition. In **1**, it still cannot communicate with the hardware because of failed authentication. The communication uses the server hardware ASMI password, which is similar to what the Hardware Management Console (HMC) uses to establish a connection with the Power Systems servers. In **2**, reset the password. You must provide the old password and the new one. The default old password is an empty value. In **3**, the connection is fully established and displays LINE UP.

Example 6-1 Running the lshwconn command

```
root@pcmha ~]# lshwconn cec 1
Server-SN06066FA: 172.17.0.8: LINE DOWN
Server-SN06066FA: sp=primary,ipadd=192.168.1.204,alt_ipadd=unavailable,state=CEC
AUTHENTICATION FAILED

[root@pcmha ~]# rspconfig cec *_passwd=old_passwd,new_passwd 2
Server-SN06066FA: 172.17.0.8: LINE DOWN
Server-SN06066FA: 192.168.1.204: RESET REQUESTED BY USER

root@pcmha ~]# lshwconn cec 3
Server-SN06066FA: 172.17.0.8: LINE DOWN
Server-SN06066FA: sp=primary,ipadd=192.168.1.204,alt_ipadd=unavailable,state=LINE
UP
```

After your new server shows up as LINE UP, you can then import it into Platform Cluster Manager - Advanced Edition. Log in to the Platform Cluster Manager - Advanced Edition web interface and, in the left side menu, click **Resources** → **Infrastructure** → **Nodes**, and then click **Add** in the main area, as shown in Figure 6-1.



	Name	Power...	Cluste...	Provision Sta...	Synchr...	Node Type
☐	server8	on	Cluster...	defined	-	Compute Node
☐	server7	on	Cluster...	provisioning	-	Compute Node
☐	server6	on	Cluster...	defined	-	Compute Node
☐	server5	on	Cluster...	defined	-	Compute Node
☐	server4	on	Cluster...	defined	-	Compute Node
☐	server3	on	Cluster...	defined	-	Compute Node

Node Name: server8

Cluster Attributes

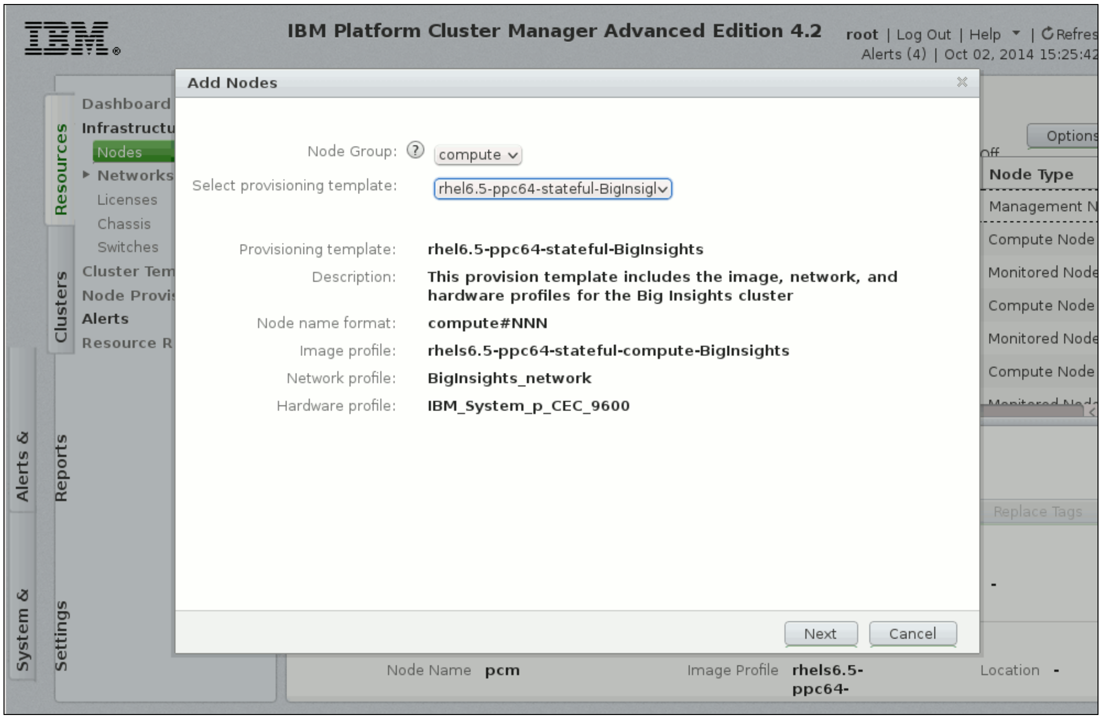
Cluster Name **Cluster2_BI** Cluster Tier **Master** Account Name **AdminAccount**

Identifying Attributes

Node Name **server8** Image Profile **rhels6.5-ppc64-data-1** Location -

Figure 6-1 Add a node to a Platform Cluster Manager - Advanced Edition environment

The next step is to select a provisioning template to use with this new node. For information about how to create a provisioning template, see 3.3.4, “Provisioning templates” on page 89. In Figure 6-2, select an existing InfoSphere BigInsights provisioning template from the test environment. You can choose to specify provisioning properties; if so, you must provide a node name format, and select image, network, and hardware profiles. One suggestion is to have the provisioning template ready to use with your Platform Cluster Manager - Advanced Edition environment because it facilitates node management and standardizes your environment.



Node Group: compute

Select provisioning template: rhel6.5-ppc64-stateful-BigInsigl

Provisioning template: rhel6.5-ppc64-stateful-BigInsights

Description: This provision template includes the image, network, and hardware profiles for the Big Insights cluster

Node name format: compute#NNN

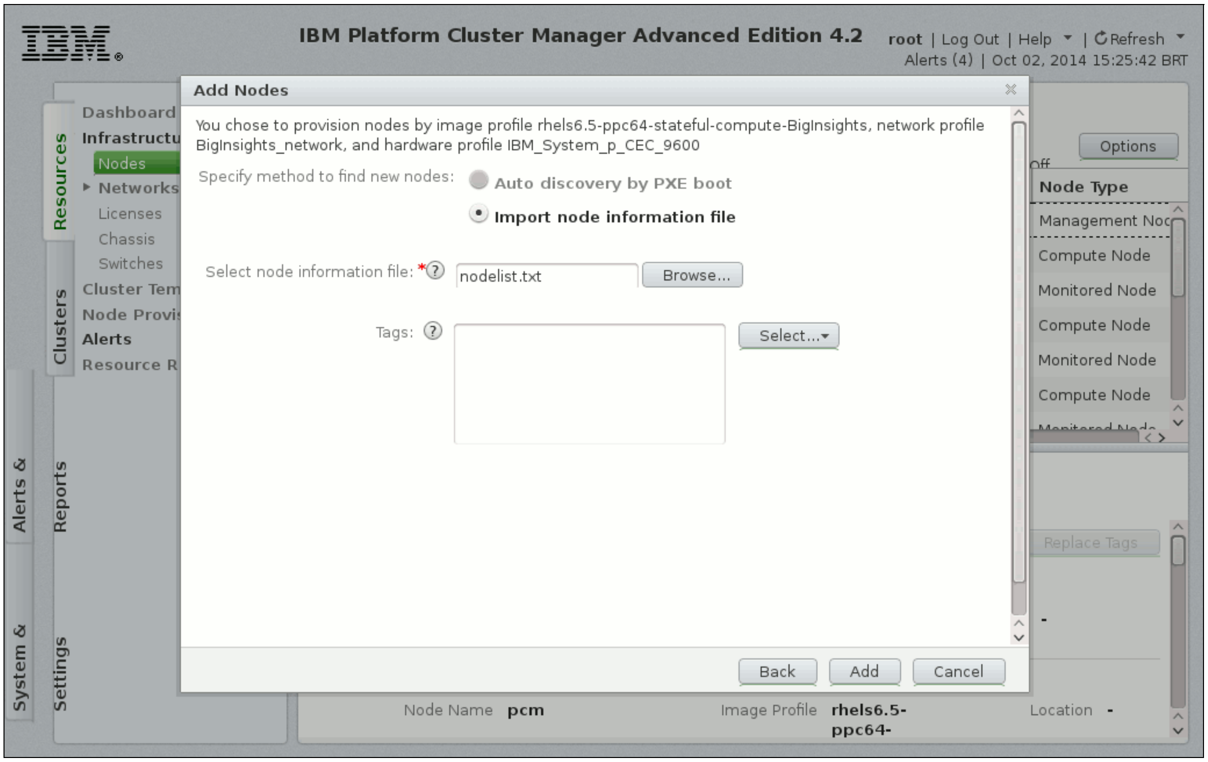
Image profile: rhels6.5-ppc64-stateful-compute-BigInsights

Network profile: BigInsights_network

Hardware profile: IBM_System_p_CEC_9600

Figure 6-2 Select a provisioning template to add a node to Platform Cluster Manager - Advanced Edition

The next and final step is to provide a node information file to import the node, as shown in Figure 6-3 on page 135.



Add Nodes

You chose to provision nodes by image profile rhels6.5-ppc64-stateful-compute-BigInsights, network profile BigInsights_network, and hardware profile IBM_System_p_CEC_9600

Specify method to find new nodes: ○ Auto discovery by PXE boot ● Import node information file

Select node information file: * nodelist.txt

Tags:

Figure 6-3 Specify a node information file to add a node

The contents of the node information file is composed of the host name and the CEC name, as shown in Example 6-2.

Example 6-2 Contents of a node information file

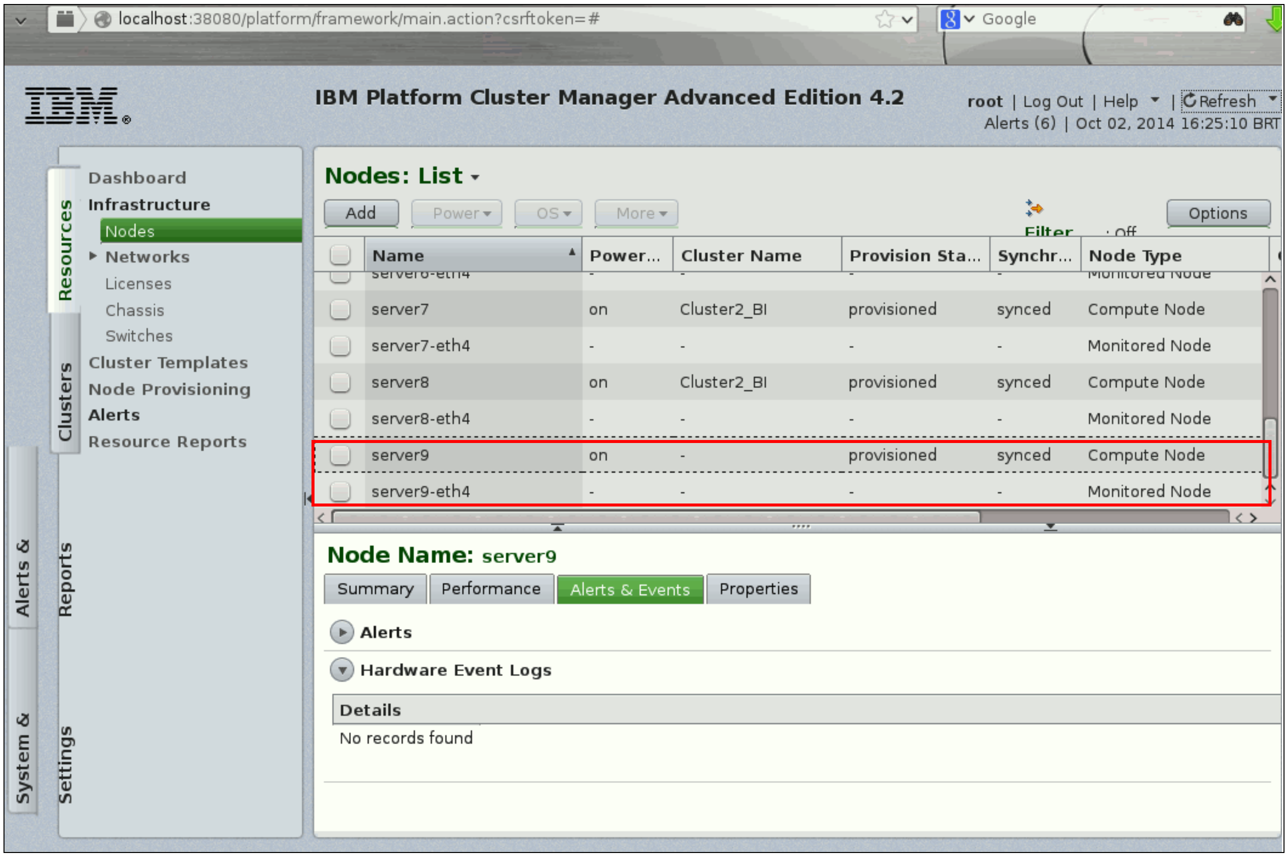
```
#node definition file
server9:
cec=Server-SN1008CBA
```

In Example 6-2, you must specify a name for the new server `server9`. If you do not want to name the node, you can simply use the `__host__` keyword, and Platform Cluster Manager - Advanced Edition names the server according to your provisioning template node name format, as shown in Figure 6-2 on page 134.

Click **Finish** in the wizard and wait for the node addition process to complete. After it completes, you can see an entry for your newly added node in the Resources tab of Platform Cluster Manager - Advanced Edition by clicking **Infrastructure** → **Nodes**.

For the InfoSphere BigInsight installation, the intra-node data communication is over the IP interface of the host name of the node. In the test cluster, the node cannot perform a performance network boot over the high-performance interconnect, that is, the 10 Gb Ethernet. The host name of the node must be changed to be on the 10 Gb Ethernet after the node is provisioned. Initially, only one entry (over the provisioning network) appears in the list with the node type specified during the add node operation. In this scenario, it is the *compute node*.

After you deploy a system into this node, change the host name to the 10 Gb Ethernet, and restart the node, a *monitored node* entry (over the 10 Gb network) appears, as shown in Figure 6-4. The alternative approach is to define the multi-homed environment in the Platform Cluster Manager - Advanced Edition Enterprise Grid Orchestrator (EGO) configuration, as described in 6.2.1, “Creating a cluster template” on page 141. When you use the multi-homed configuration, only a single entry for each node appears in the list.



Name	Power...	Cluster Name	Provision Sta...	Synchr...	Node Type
server6-eth4	-	-	-	-	Monitored Node
server7	on	Cluster2_BI	provisioned	synced	Compute Node
server7-eth4	-	-	-	-	Monitored Node
server8	on	Cluster2_BI	provisioned	synced	Compute Node
server8-eth4	-	-	-	-	Monitored Node
server9	on	-	provisioned	synced	Compute Node
server9-eth4	-	-	-	-	Monitored Node

Node Name: server9

Details

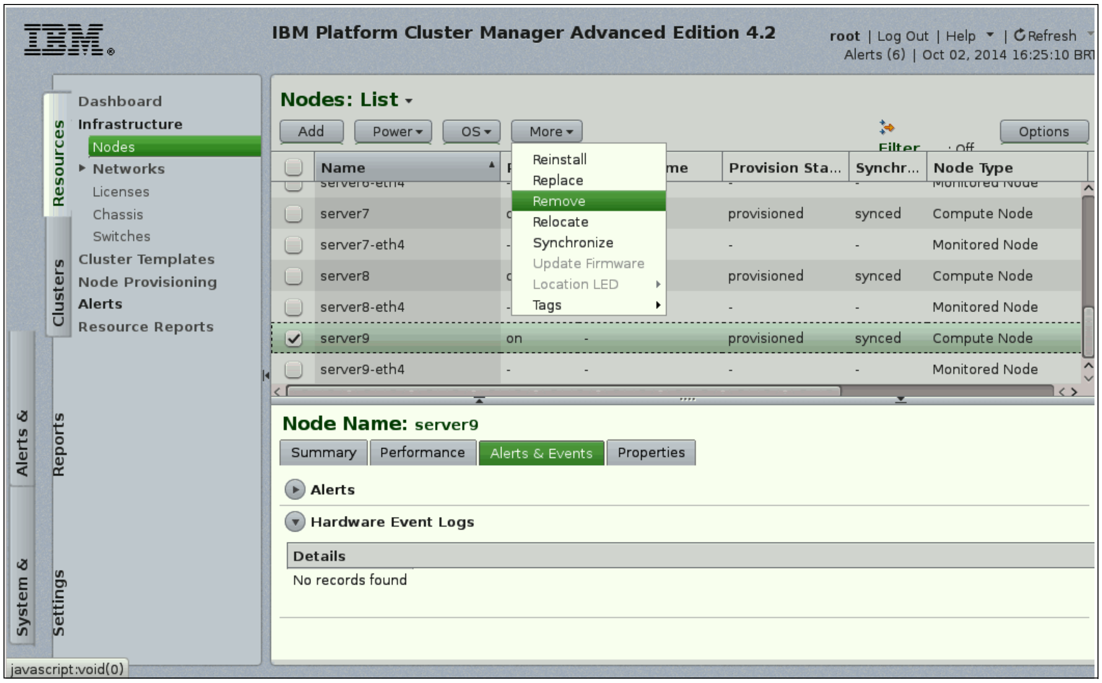
No records found

Figure 6-4 Verify the add node operation

6.1.2 Removing nodes from a Platform Cluster Manager - Advanced Edition environment (including the monitored node entry)

To remove a node from a Platform Cluster Manager - Advanced Edition environment, first remove it from any existing Platform Cluster Manager - Advanced Edition-managed cluster of which it might be part. After the node is free, then proceed as follows.

In the left menu of the Platform Cluster Manager - Advanced Edition interface, click **Resources** → **Infrastructure** → **Node**. Then, select the node that you want to remove. Your node might have two entries: one for the node itself and another listed as a *monitored node*. The next section explains how to deal with the monitored node entry. After you select the node for removal, click **More**, and then click **Remove**, as shown in Figure 6-5 on page 137.



Name	P		me	Provision Sta...	Synchr...	Node Type
server6-eth4	-			-	-	Monitored Node
server7	c			provisioned	synced	Compute Node
server7-eth4	-			-	-	Monitored Node
server8	c			provisioned	synced	Compute Node
server8-eth4	-			-	-	Monitored Node
☑ server9	on	-		provisioned	synced	Compute Node
server9-eth4	-	-		-	-	Monitored Node

Node Name: server9

Details
No records found

Figure 6-5 Remove a node form Platform Cluster Manager - Advanced Edition

After the node is removed, the monitored node entry remains active. To remove that node, complete the following steps:

1. Log in to the Platform Cluster Manager - Advanced Edition node by using SSH and source the require profile.
2. Log on as the cluster admin.
3. Check the status of the monitor agent.
4. If the monitor agent status is other than *unavailable*, close it.
5. Remove the agent resource.

Example 6-3 demonstrates how to perform these actions.

Example 6-3 Remove a monitored node entry from Platform Cluster Manager - Advanced Edition

```
# . /opt/pcm/ego/profile.platform 1
# egosh user logon -u Admin -x Admin 2
# egosh resource list 3
# egosh resource close node-resource-name 4
# egosh resource remove node-resource-name 5
```

6.1.3 Monitoring a Platform Cluster Manager - Advanced Edition managed node

Platform Cluster Manager - Advanced Edition can integrate basic monitoring features by using agents on its deployed systems. You can manually install these agents or specify their installation through the operating system image profile. The default operating system image profile comes with these agents set up, so if you create your operating system image profile by using the default one as a starting point, you do not need to take any further action.

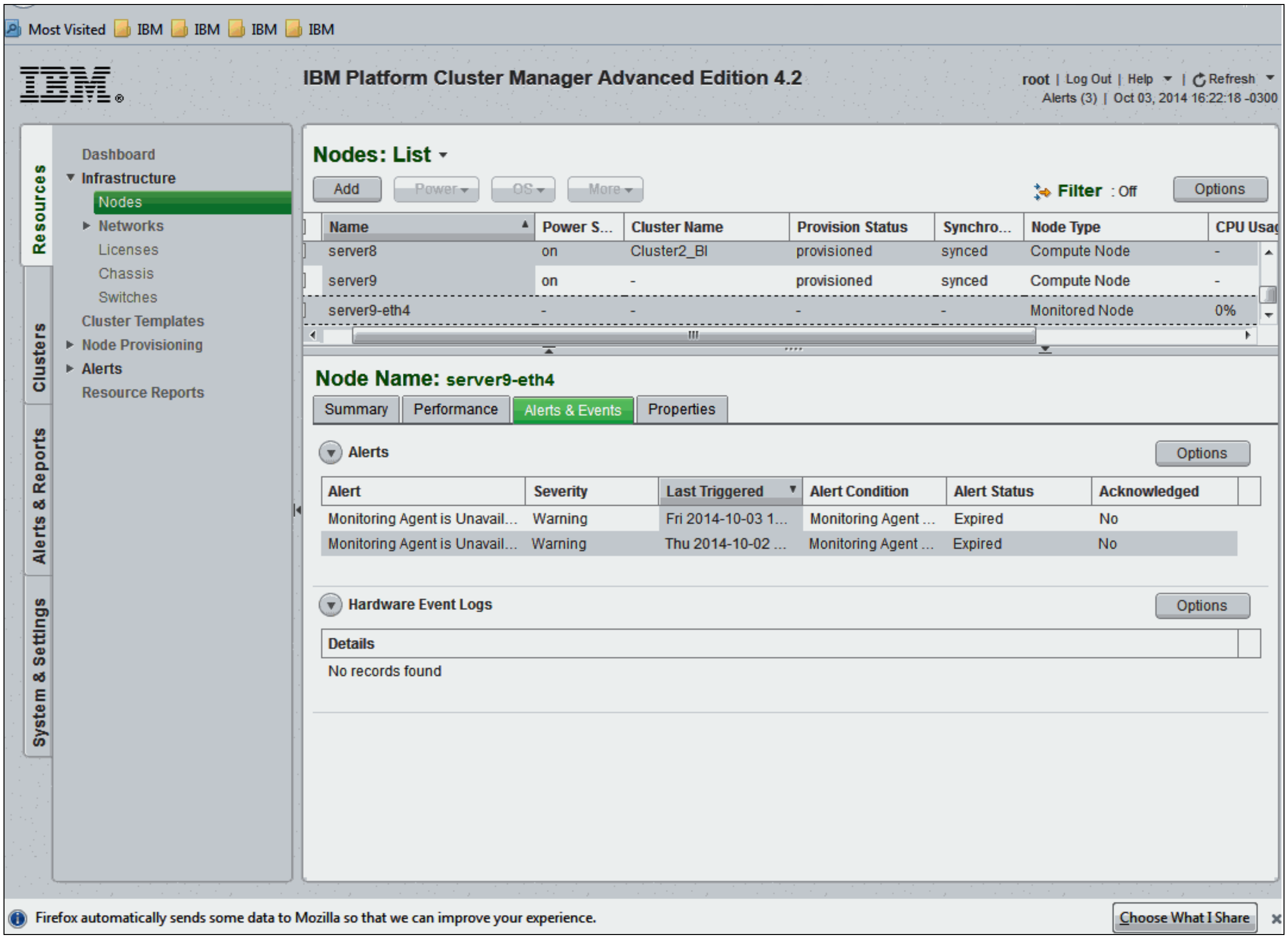
To set up manually node monitoring, review the following high-level overview of the steps:

1. Create a user and group named *pcmadmin*.
2. Install the required RPM monitoring package pcm-ego.
3. Export the required environment variables.
4. Join the monitored node to the Platform Cluster Manager environment.
5. Start the monitoring agent on the node.

For a comprehensive set of instructions about how to perform manually the monitoring actions, see *IBM Platform Cluster Manager, Version 4.2 Administration Guide*, found at:

`http://www-01.ibm.com/support/knowledgecenter/SSDV85_4.2.0/pcm_welcome.html`

Figure 6-6 shows information that you can get from the Platform Cluster Manager - Advanced Edition dashboard regarding node alerts and hardware event logs. This information is displayed by selecting the corresponding monitored node from the node list and then viewing the Alerts & Events tab.



Name	Power S...	Cluster Name	Provision Status	Synchro...	Node Type	CPU Usag
server8	on	Cluster2_BI	provisioned	synced	Compute Node	-
server9	on	-	provisioned	synced	Compute Node	-
server9-eth4	-	-	-	-	Monitored Node	0%

Node Name: server9-eth4

Alerts

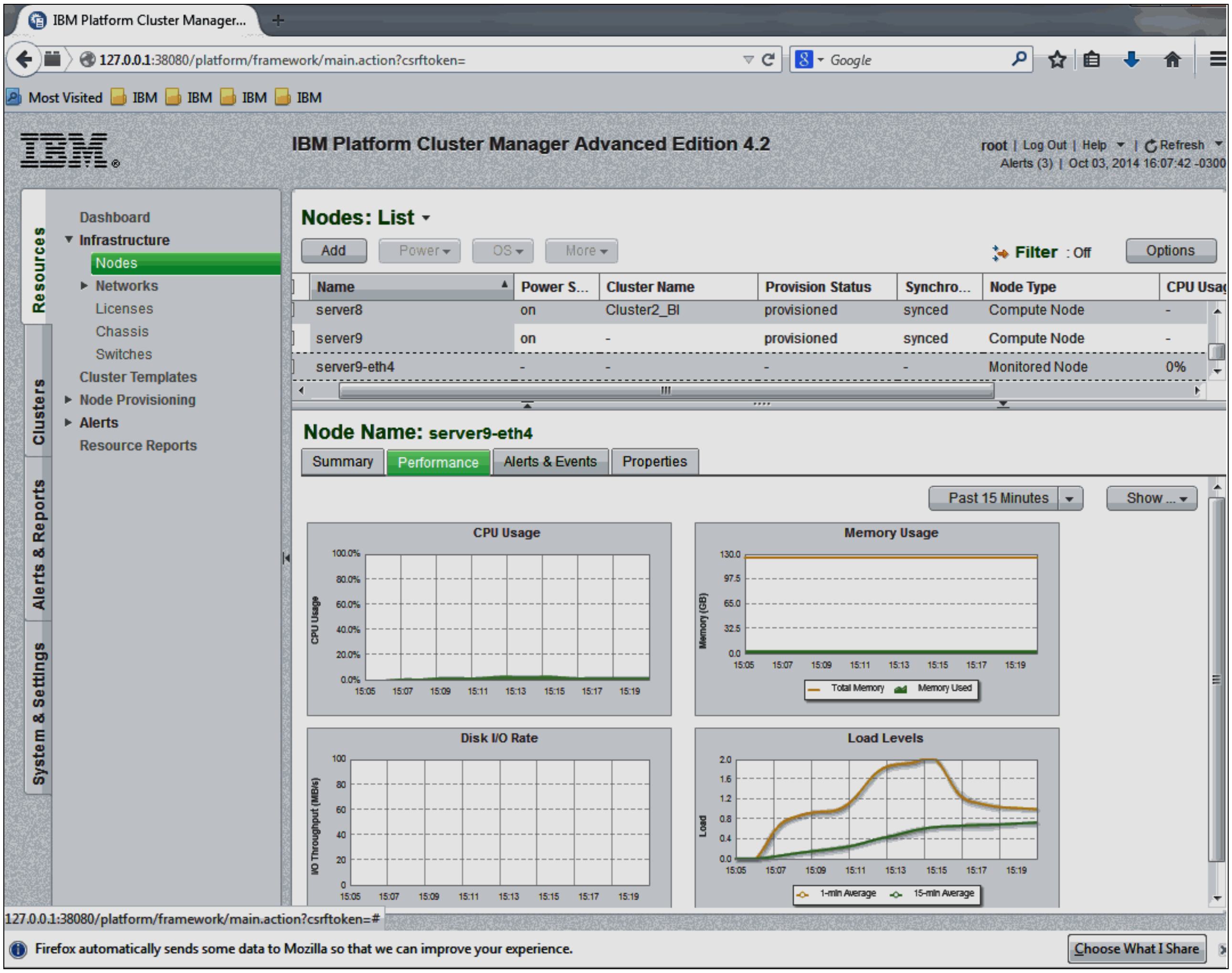
Alert	Severity	Last Triggered	Alert Condition	Alert Status	Acknowledged
Monitoring Agent is Unavail...	Warning	Fri 2014-10-03 1...	Monitoring Agent ...	Expired	No
Monitoring Agent is Unavail...	Warning	Thu 2014-10-02 ...	Monitoring Agent ...	Expired	No

Hardware Event Logs

Details
No records found

Figure 6-6 Alerts and events for a monitored node within Platform Cluster Manager - Advanced Edition

Also, Platform Cluster Manager - Advanced Edition gathers some metrics through its monitoring agent. You can view these metric under the **Performance** tab for a given monitored node. You can customize which charts you want to have in this view, such as processor and memory usage, I/O throughput, system load, and many more, as shown in Figure 6-7 on page 139.



Name	Power S...	Cluster Name	Provision Status	Synchro...	Node Type	CPU Usa
server8	on	Cluster2_BI	provisioned	synced	Compute Node	-
server9	on	-	provisioned	synced	Compute Node	-
server9-eth4	-	-	-	-	Monitored Node	0%

Node Name: server9-eth4

Figure 6-7 Performance metrics for a monitored node within Platform Cluster Manager - Advanced Edition

6.2 Managing InfoSphere BigInsights cluster nodes within Platform Cluster Manager - Advanced Edition

Before you set up Platform Cluster Manager - Advanced Edition to deploy a working cluster, such as an InfoSphere BigInsights cluster, first make sure that Platform Cluster Manager - Advanced Edition can at least deploy a working operating system environment by using a provisioning template, as described in 3.3.4, “Provisioning templates” on page 89. After you complete this process, the next step is to define a cluster template.

A cluster template is an entity that holds the information that is required for installing software across multiple nodes in the cluster. For example, an InfoSphere BigInsights cluster template holds information about how to perform the installation of InfoSphere BigInsights in a given node. Cluster templates are composed of the following components:

- An image profile
- A network profile
- Custom scripts for cluster software installation

Recall that an image profile has some postinstallation or post-boot custom scripts to run upon the operating system deployment. These scripts are, however, related to operating system tuning, not to cluster software customization. In an InfoSphere BigInsights cluster template, the *image profile* scripts deal with tuning the operating system and software installation to meet the prerequisites for an InfoSphere BigInsights installation, but in the test scenario the *cluster template* scripts deal with starting the InfoSphere BigInsights installer after all nodes have the operating system running and ready for an InfoSphere BigInsights installation.

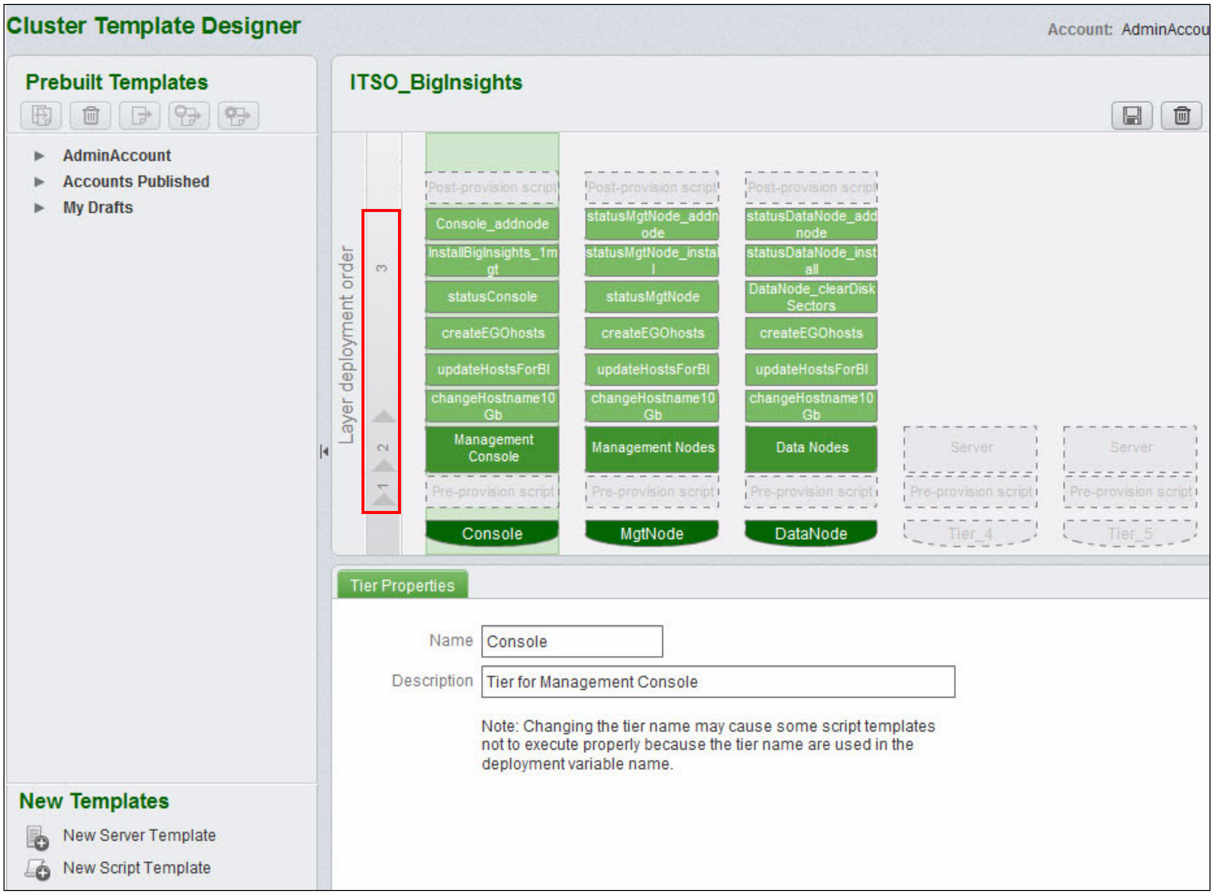
The image profile in the InfoSphere BigInsights cluster template is modified from the stateful image that is built by the xCAT in Platform Cluster Manager - Advanced Edition with the following changes:

- The operating system packages for the InfoSphere BigInsight installation that are added to the image profile include the following ones:
  - `device-mapper-multipath`
  - `device-mapper-multipath-libs`
  - `expect`
  - `gcc-gfortran`
  - `ksh`
  - `libgcc`
  - `numactl`
  - `pam.ppc`
  - `pciutils`
  - `pciutils-libs`
  - `tcl`
  - `tcsh`
  - `tk`
- The image profile in Platform Cluster Manager - Advanced Edition can add one customized postinstallation script and one customized post-boot script. The postinstallation script is run after the operating system is installed and before the node is restarted. The post-boot script is run on the node after the node that is provisioned with the image profile is restarted. The postinstallation script of the image profile in the InfoSphere BigInsights cluster template is defined to run a number of subscripts that modify the operating system environment to meet the prerequisites for the InfoSphere BigInsights installation. These scripts include the following functions:
  - Installation of the Mellanox OpenFabrics Enterprise Distribution (MLNX_OFED 2.3-1.0.1) package for the Mellanox 10 Gb HCA
  - Disabling IPv6
  - Disabling firewall
  - Disabling **`tty sudo`**
  - Changing kernel parameters in sysctl.conf

The InfoSphere BigInsights installation requires that the first entry for each host in the `/etc/hosts` file be the fully qualified name (FQN). When a new node is added in the Platform Cluster Manager console, xCAT updates the `/etc/hosts` file with the entry for the new node; however, the short name (without the domain) is the first entry. The `/etc/hosts` file can be set up to be synchronized across the cluster nodes. However, the `/etc/hosts` file must be modified to change the first entry to FQN for all hosts. A script file can be set to run whenever the `/etc/hosts` file on a node is being updated by using the option **“**`EXECUTE:`**”** `Definition` in synclist that is supported in xCAT 2.9, which is bundled with Platform Cluster Manager V4.2.

6.2.1 Creating a cluster template

To create a cluster template, in the left menu of Platform Cluster Manager - Advanced Edition, click **Resources** → **Infrastructure** → **Cluster Templates**, then click **Add**. Figure 6-8 shows the main interface window where you create a cluster template: The Cluster Template Designer.



Name: Console

Description: Tier for Management Console

Note: Changing the tier name may cause some script templates not to execute properly because the tier name are used in the deployment variable name.

Figure 6-8 Platform Cluster Manager - Advanced Edition - Cluster Template Designer

You now must define some key aspects of your cluster template. Click the corresponding item on the interface to make changes, or drag the server and script template skeletons from the lower left of the template designer to create them.

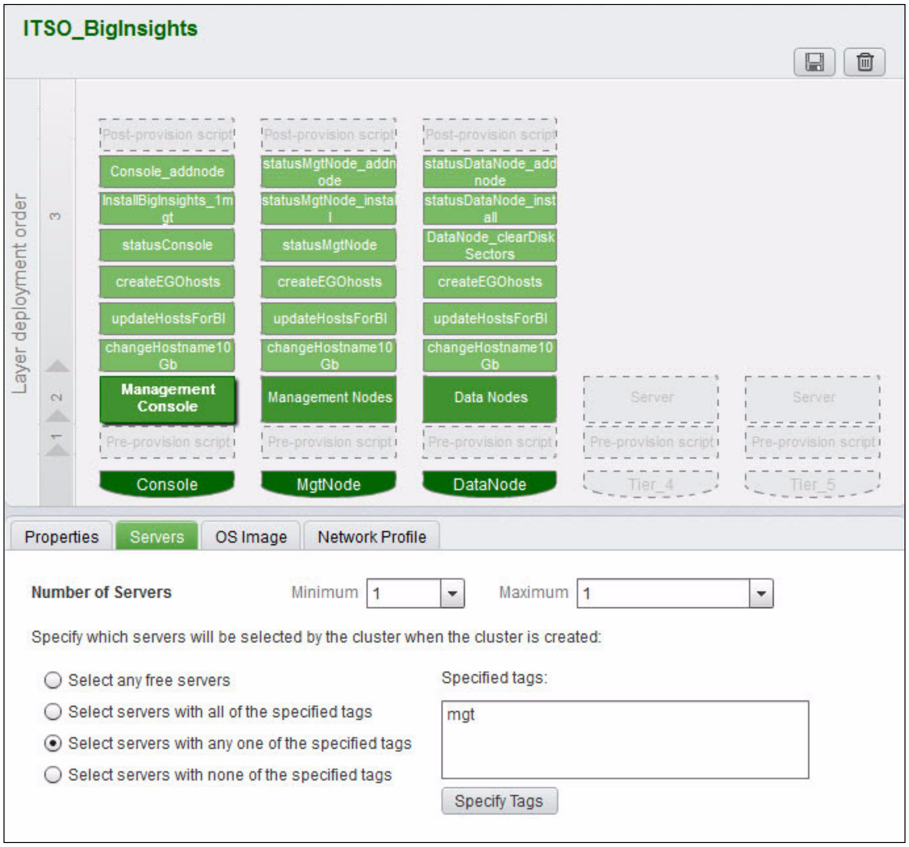
- The number of tiers for your cluster

  In the InfoSphere BigInsights cluster template, define three tiers: The *console* tier for the console node, the *management node* tier, and the *data node* tier. These three tiers are created to differentiate the three types of nodes in terms of script tuning for a particular node type, and to make it easier to manage an InfoSphere BigInsights installation within Platform Cluster Manager - Advanced Edition. Because you can trigger an InfoSphere BigInsights full cluster installation by starting it from the console node, call the installer only one time on that node, as shown in Figure 6-8 on page 141.

  Notice in Figure 6-8 on page 141 that each tier in the nodes layout is composed of three layers. There is a layer deployment order, as shown by the highlighted rectangle around it. The three layers are pre-provisioning scripts, a server template, and the post-provisioning scripts. No pre-provisioning scripts are used in the examples, but you must create and tune the server template and post-provisioning scripts sections.

- The name and properties for the server templates for each tier

  Here, you define the image and network templates to use with that particular tier node type. Also, you can customize which servers, and how many, you want to use with that particular tier. For example, you might want one hardware type to run management nodes and the console node, and another hardware type to run compute nodes. You can place tags on your nodes and then have Platform Cluster Manager select the appropriate hardware for each server type of your cluster template, as shown in Figure 6-9 on page 143. Select the InfoSphere BigInsights image template and your InfoSphere BigInsights network template to complete the OS Image and Network profile tabs. Use the template that is created in “Creating an image profile” on page 91 and “Creating a network profile” on page 95.



Number of Servers Minimum 1 Maximum 1

Specify which servers will be selected by the cluster when the cluster is created:

○ Select any free servers
○ Select servers with all of the specified tags
● Select servers with any one of the specified tags
○ Select servers with none of the specified tags

Specified tags:

mgt

Figure 6-9 Define the properties for a server template within a cluster template

In this test environment, here are the limits for each tier:

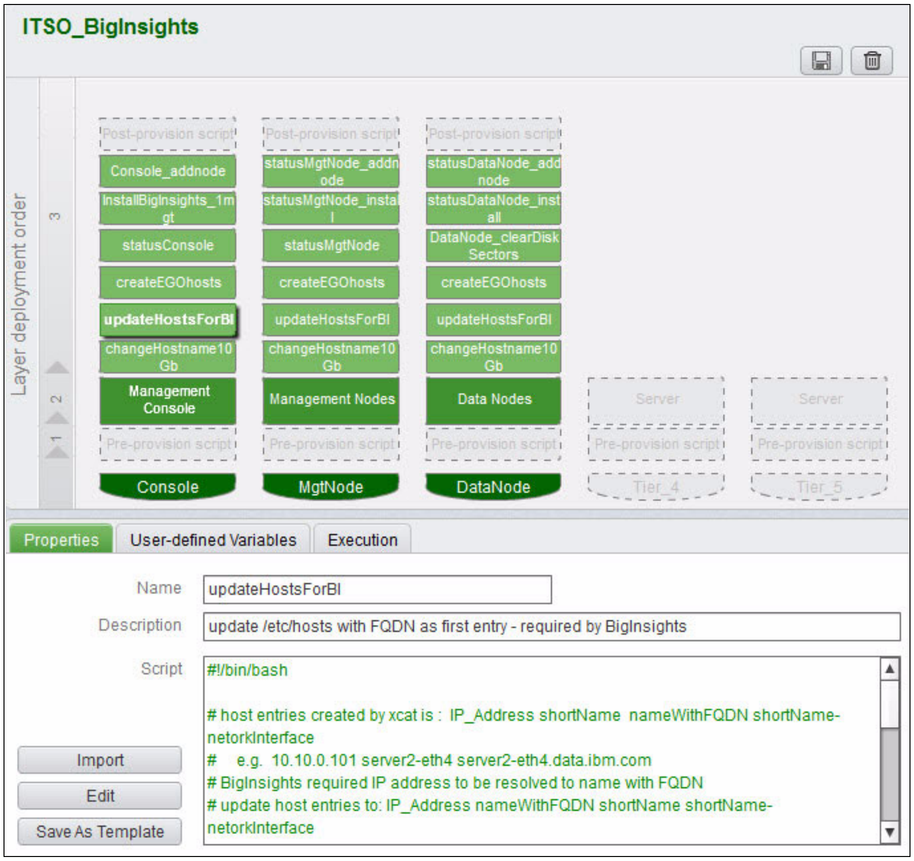
– Console node: Exactly one node.

– Management nodes: Minimum of zero (if you want to place all services in the console node). Maximum of 10.

– Compute nodes: Minimum of 1, maximum of unlimited.

- The post-provisioning scripts section

  In this section, input your custom scripts for the cluster software installation. A post-provisioning script is composed of a name, a description, and the script itself, as shown in Figure 6-10.



Name: updateHostsForBI

Description: update /etc/hosts with FQDN as first entry - required by BigInsights

Script:

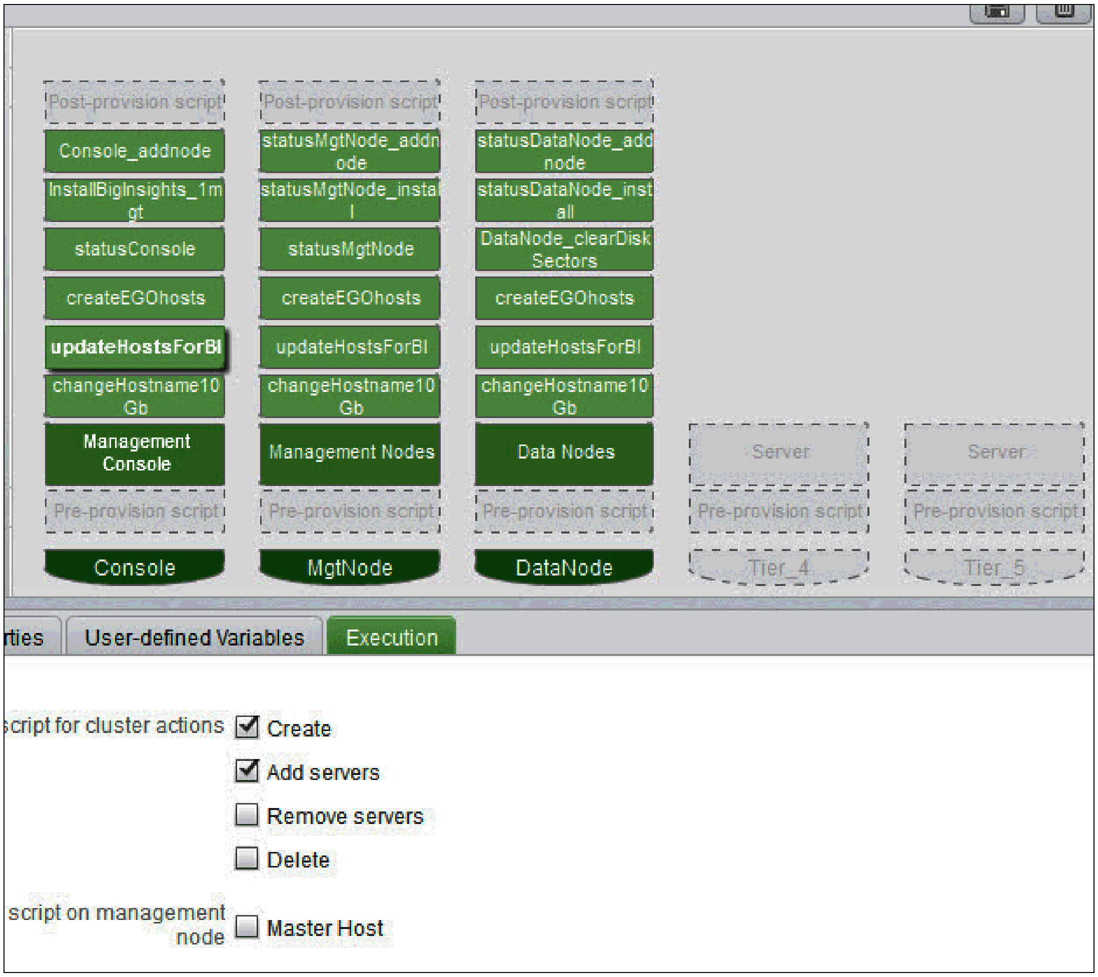
```
#!/bin/bash

# host entries created by xcat is :  IP_Address shortName  nameWithFQDN shortName-netorkInterface
#    e.g.  10.10.0.101 server2-eth4 server2-eth4.data.ibm.com
# BigInsights required IP address to be resolved to name with FQDN
# update host entries to: IP_Address nameWithFQDN shortName shortName-netorkInterface
```

Figure 6-10 A post provisioning script in a cluster template

> **Note:** The scenario uses custom scripts to adjust the host name and `/etc/hosts` entries because the 10 Gb Ethernet Mellanox card cannot perform network boots, so you cannot use it for provisioning the network for the nodes. Because Platform Cluster Manager - Advanced Edition assigns the host name of the nodes to the provisioning interface, scripts must be written to associate the Mellanox interface with the host name. This is what the `ChangeHostname_10Gb` and the `updateHostsforBI` post-provisioning scripts do.

A post-provisioning script that is called `status<nodetype>` is added, as shown in Figure 6-10, to ensure synchronization among the cluster nodes. The Platform Cluster Manager - Advanced Edition cluster deployment already ensures node synchronization, but added an extra layer to make sure that the InfoSphere BigInsights installation did not start before making changes to `/etc/hosts` because of the Mellanox issue.

Notice that one of the properties you can configure when defining a post-provisioning script is the Execution tab, as shown in Figure 6-11. You can select to run a given script for different cluster operations, such as create, add, or remove servers. In the particular case for the `updateHostsforBI` script, run it for the cluster creation and add server operations.



cript for cluster actions ☑ Create

☑ Add servers

□ Remove servers

□ Delete

script on management node □ Master Host

Figure 6-11 Define when to run the custom post-provision scripts

Hosts with more than one IP address are called multi-homed hosts. Because hosts in Platform Cluster Manager are identified by name, the host name must be configured so that all the IP addresses for the multi-homed host are resolved to the same name and are identified as a single host in Platform Cluster Manager. The `/opt/pcm/ego/kernel/conf/hosts` file is created so that for each host entry in `/etc/hosts`, the addresses for both the 1 Gb Ethernet and the 10 Gb Ethernet have the same name. For more information, see the section “Hosts with multiple addresses” in the IBM Platform LSF document *Administering Platform LSF*, found at:

http://publibfp.dhe.ibm.com/epubs/pdf/c2753023.pdf

The multi-homed host file is created by the `createEGOhosts` postscript, as shown in Example B-3 on page 189.

After you finish editing all of the information that you need for the InfoSphere BigInsights cluster deployment, save the cluster template, which makes it available for use in the Cluster Templates area of Platform Cluster Manager - Advanced Edition.

Note: All of the scripts that were used in the test environment are available in Appendix B, “Scripts” on page 187.

6.2.2 Creating a cluster from a cluster template

After you have created a cluster template, you can deploy a cluster by clicking, in the Platform Cluster Manager - Advanced Edition interface, **Clusters** → **Create**, as shown in Figure 6-12.

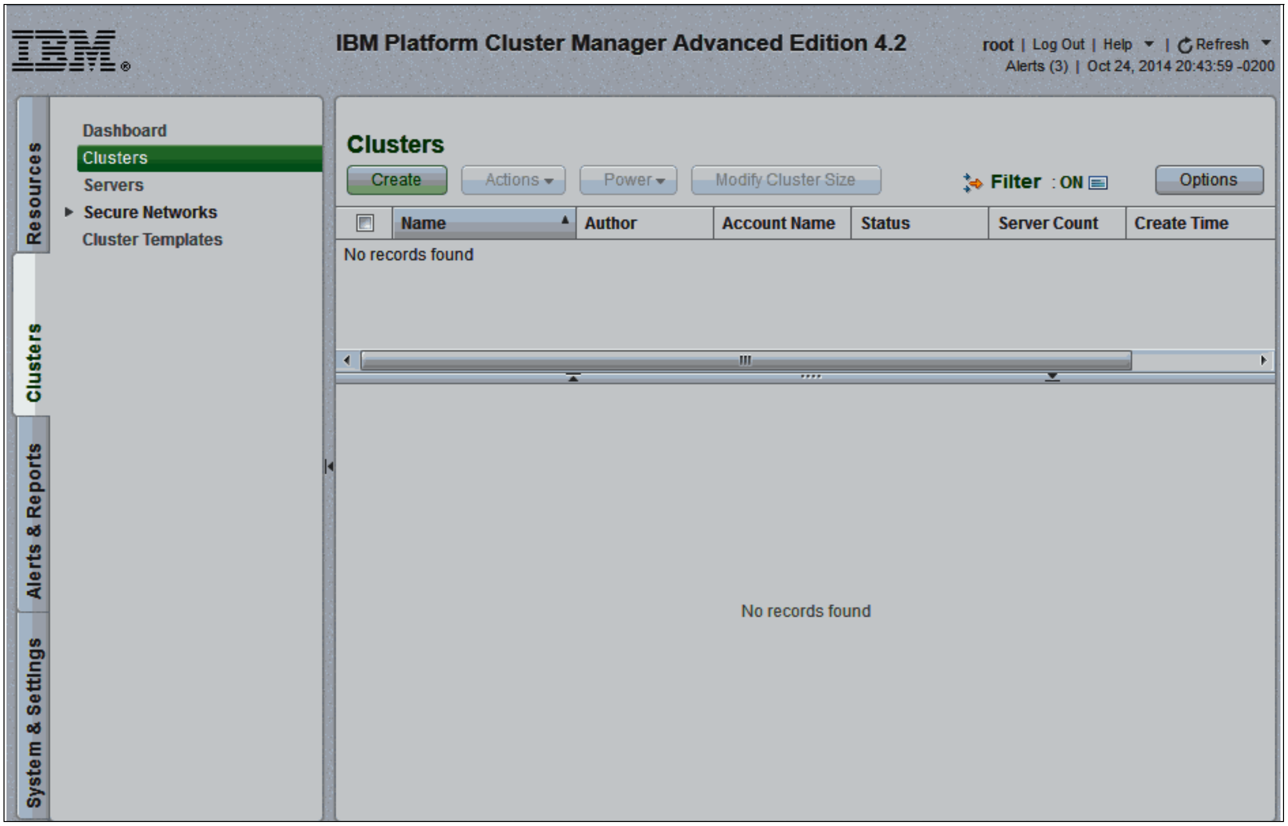
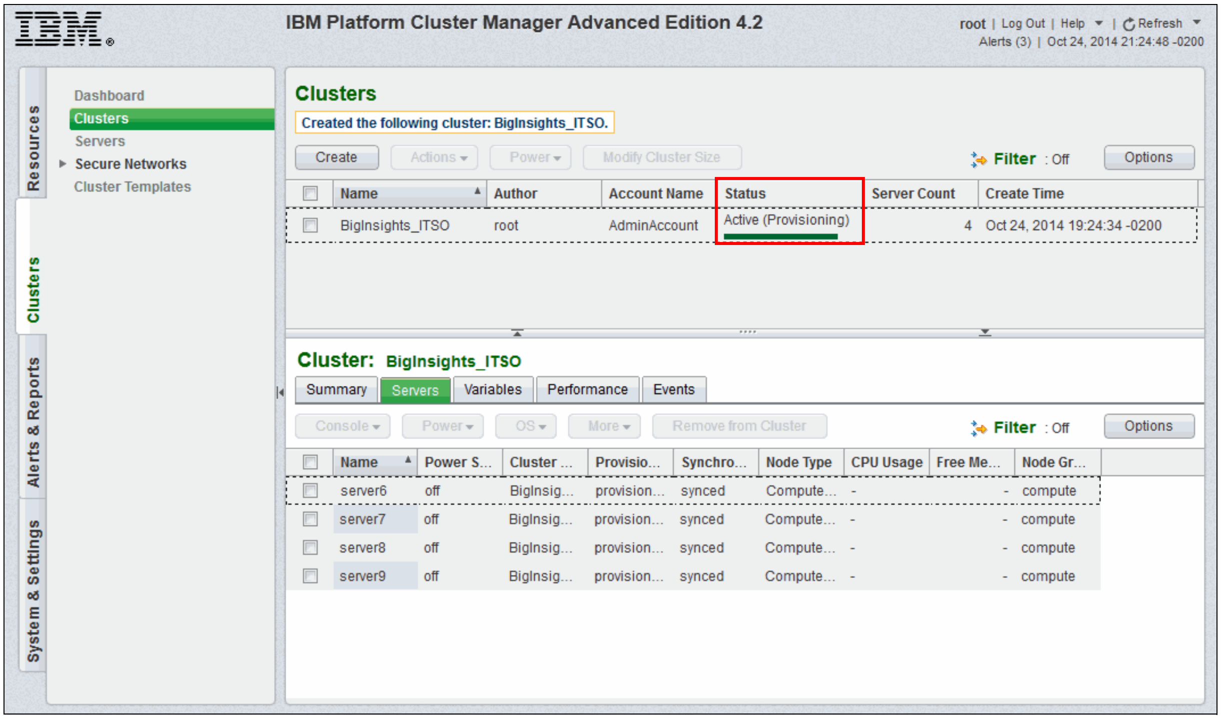


Figure 6-12 Create a cluster

Next, select the cluster template that you defined earlier and enter the information for your new InfoSphere BigInsights cluster, as shown in Figure 6-13 on page 147. Provide the following information:

- Cluster name and description
- The number of management nodes
- The number of data nodes



Created the following cluster: BigInsights_ITSO.

□	Name	Author	Account Name	Status	Server Count	Create Time
□	BigInsights_ITSO	root	AdminAccount	Active (Provisioning)	4	Oct 24, 2014 19:24:34 -0200

□	Name	Power S...	Cluster ...	Provisio...	Synchro...	Node Type	CPU Usage	Free Me...	Node Gr...
□	server6	off	BigInsig...	provision...	synced	Compute...	-	-	compute
□	server7	off	BigInsig...	provision...	synced	Compute...	-	-	compute
□	server8	off	BigInsig...	provision...	synced	Compute...	-	-	compute
□	server9	off	BigInsig...	provision...	synced	Compute...	-	-	compute

Figure 6-13 Provision an InfoSphere BigInsights Cluster with Platform Cluster Manager - Advanced Edition

Then, click **Create**. The cluster is put into the Active (Provisioning) state, as shown in Figure 6-13. Then, choose to create a cluster with the console node, no extra management node, and three data nodes, totaling four nodes.

Wait for the cluster to reach a provisioned state, and then verify that your InfoSphere BigInsights console is available through the IP address that is assigned to the public interface in the console node. You can verify the IP addresses that are assigned to your cluster nodes by checking the per-node information, which you can access by clicking **Resources** → **Nodes** in the Platform Cluster Manager - Advanced Edition interface.

While you wait for the cluster provisioning, you can use the Events tab, as shown in Figure 6-13, to gain access to the events logging of your cluster creation.

7

Tuning

This chapter presents performance tuning hints for your big data environment. This chapter does not go into depth about how to tune your hardware and storage device, or describe network configuration tuning in detail.

This chapter covers the following topics:

- Tuning IBM InfoSphere BigInsights
- Tuning IBM Spectrum Scale (formerly GPFS)
- Tuning the Platform Symphony MapReduce framework

7.1 Tuning IBM InfoSphere BigInsights

Much of how you go about tuning your InfoSphere BigInsights environment depends on your workloads, so this section describes a few of the parameters that you might consider configuring manually.

7.1.1 Tuning at the operating system level

There are many parameters that you might want to alter in the Linux kernel. You might consider changing the scheduler that is used by the operating system. For example, if you are often doing MapReduce jobs and much sequential I/O, you might want to change the scheduler from completely fair to deadline. This change allows faster read access to the data, which can help speed up your jobs. This is configured in the `/sys/block/sd(x)/queue/scheduler` file.

If you are often reading from your disks, then any parameters you can change to improve read speed, read ahead buffers, and similar things, might improve your performance. If you are using HDFS (Spectrum Scale tuning is considered in more detail in 7.2, "Tuning IBM Spectrum Scale (formerly GPFS)" on page 153), then you might also consider some of your `ext4` settings in your JBOD configured disks. Again, the goal here is to improve read speed. Altering `dir_index` can speed up lookups in large directories, and using an extent block scheme can improve the throughput of the file system access, especially with larger files.

Red Hat also has some memory configuration settings that can have an impact on the performance of your InfoSphere BigInsights cluster. `vm.swappiness` is a configuration setting that controls how proactively memory pages are swapped. Lowering this setting decreases how frequently the kernel swaps pages and can offer performance improvements. This parameter is best changed in small increments. `vm.min_free_kbytes` is another parameter worth highlighting. When this value is increased, the kernel starts to reclaim memory earlier, which can be useful in some situations with InfoSphere BigInsights. InfoSphere BigInsights often has this setting at around 2% of the total memory available, so if you are using Spectrum Scale, it should be set to around 6% of the total memory on the server.

7.1.2 Tuning Hadoop

The speed of Hadoop MapReduce jobs depends on many variables, but there are a few key areas that you can look at in an attempt to speed up the jobs. Initially, the job is submitted and the scheduler breaks it down into some maps. These map tasks are then distributed across all the available slots in the cluster. The data from these jobs is then shuffled on to the disk and then a number of reduce tasks are carried out before returning the answer.

If the number of map tasks and the number of available slots aligns poorly, then you can expect suboptimal performance. The number of map and reduce slots on each tasktracker is configured in the `mapred-site.xml` file and is set by default to the outcome of a formula. If your machine has hyperthreading or many processors, then you might find that the formula provides a high value, so reducing it can improve performance. If you configure this value, be careful because a high value can make your cluster unstable, but a low value wastes resources. If you change this value, you can change the Hadoop MapReduce JVM size to accommodate the change. A reasonable value for the `mapred.chid.java.opts -Xmx` setting is 75% of your available memory divided by the maximum number of map and reduce tasks.

The sort parameters also must be edited if the `mapred.chid.java.opts -Xmx` configuration is changed. The `io.sort.factor` setting controls how many threads are joined, and the `io.sort.mb` setting controls the size in megabytes of the buffer to use when sorting the output. These should be tuned with the value set for `mapred.chid.java.opts -Xmx`, or your sort time increases.

Now that the number of slots your system is set correctly, configure the block size of the data in your storage system. This is significant because the number of maps your job creates is a function of the size of the files you are trying to read and the block size. Each block of data has its own mapper. If you have a small block size, it generates a superfluous number of mappers, which incurs a processing impact. Also, a large a block size increases the risk of concurrency issues and the delays that are associated with it.

Compression is another element to take into account when configuring your cluster for performance. If compression is implemented, you sacrifice processing for gains in storage; this can be counter-productive if your processor is already stretched. Nevertheless, if you decide to compress the output of your map tasks, this can speed up your shuffle phase (as there is less to move). If you alter the compression settings, there are several parameters that you can edit in the `mapred-site.xml` file. To specify a specific compression codec, alter `mapred.map.output.compression.codec` to point to the codec of your choice. To compress the output of your maptasks (to attempt to speed up your shuffle phase), change the `false` value of `mapred.compress.map.output` to `true`. Finally, to compress the final output of your Hadoop MapReduce job, set `mapred.output.compress` to `true`.

7.1.3 Tuning BigSQL

BigSQL uses existing IBM expertise from the relational database world to make the data on your Hadoop cluster accessible by using the most popular query language, SQL. This means that you are already using SQL rewrite technology and a database optimizer. There are a few parameters that you might consider when customizing your system to get the absolute most out it.

When BigSQL reads the data from HDFS or Spectrum Scale, it is using an I/O engine to either read the data by using C++ or Java (depending on the format of the data it is trying to read). The settings for these I/O engines and for the BigSQL scheduler can be found in `$BIGSQL_HOME/conf/bigsql-conf.xml` file.

> **Note:** Generally, the reader settings do not need to be modified. This information is provided for advanced tuning only.

`dfsio.num_scanner_threads` is a parameter that controls the number of scanner threads that can be created by the system. This is set to `0` by default, which lets the system decide how many threads to create. This can be specified per query by the user or changed here. `dfsio.mem_limit` is a limit on the amount of memory that can be used by the C++ I/O engines. By default, this is controlled by BigSQL, but you can specify a specific amount of memory to use. `dfsio.num_threads_per_disk`, a parameter that controls threads per disk for C++ I/O engines, is set by default to allow the system to decide. The system then creates five threads per disk, which is generally sufficient to use all of the capacity of the disks. If you have high-performance disk subsystem, you might want to increase this value. For Java I/O engines, you might want to increase the value of `bigsql.java.io.tp.size` from the default of eight if you have a high-performance device. This parameter can also be set on a per query basis.

Your InfoSphere BigInsights instance was configured at installation time to allow only the BigSQL component to use a certain amount of the resources on the management node on which it is. Postinstallation, you might discover that you must assign more resources or you want to throttle the amount that BigSQL is allowed to use to facilitate other functions. By default, this value is set to 25%. Most of the memory that BigSQL uses is for sort space, so if you are finding this memory is running low, this can be considered for adjustment. If you alter this value postinstallation, it is also important to alter `mapred-conf.xml` to match it. BigSQL is good at determining which functions have what proportion of available memory. If you have a relatively fixed workload, then turning on the self-tuning memory manager can be a useful way to allocate memory for buffer pools and sort space. Then, after a few workloads and after the allocations have stabilized, you can turn off this setting and be confident that your cluster has been optimally tuned for your workload.

In addition to the customizable parameters of your environment, there are methods that you can adopt with your database that might help performance. When loading data, the default settings create four map tasks. If you are loading much data, you can increase the `num.map.tasks` by running the following command:

```
overwrite WITH LOAD PROPERTIES (‘num.map.tasks’=’100’);
```

Often, setting the number of map tasks equal to the number of BigSQL worker nodes can provide better performance. Also, consider whether you can get better performance by importing files into the distributed file system before loading, or loading the files directly from the source. The number of map tasks to load from a local file system is equal to the number of files, which means that a low number of large files loads much faster than if you import them into the distributed file system first.

BigSQL and Hive both support table partitioning by data values. This approach stores data partitions as different files across your file system so that during queries the system reads only the files that are relevant. This approach can speed up many queries, although it might not be suitable for all situations.

As you might expect, statistics are important for the query optimizer. So, it is important to keep these statistics as up to date and accurate as possible to give the optimizer a chance to be as efficient as possible.

Finally, you should be aware of the data types you are using. BigSQL works with 32 K pages, so it performs more efficiently when table definitions permit rows to fit inside these 32 K pages. Often, you find that meeting this definition is not a problem, but if your data is using STRING data types, you might find that you are exceeding the page size and performance drops off because STRING maps to VARCHAR(32,672). If you have STRING fields, change these fields to VARCHARs that fit the data.

7.2 Tuning IBM Spectrum Scale (formerly GPFS)

This section describes tuning parameters that you might want to consider configuring to improve Spectrum Scale performance for your application.

7.2.1 Tuning at the operating system level

There are a few parameters that you might consider while tuning the operating system. In some configurations, it is possible to encounter memory exhaustion symptoms when free memory should in fact be available. Setting `vm.min_free_kbytes` to a higher value (the Linux **`sysctl`** utility can be used for this purpose) in the order of magnitude of 5 - 6% of the total amount of physical memory should help you avoid such a situation. For better performance of data traffic between nodes, tune the TCP window settings by enabling TCP window scaling, which increases the TCP buffer limit and the backlog size. Add the following lines to the `/etc/sysctl.conf` file:

```
# set min free memory, replace 1024 with 6% of the total amount of physical memory
vm.min_free_kbytes= 1024
# enable tcp window scaling
net.ipv4.tcp_window_scaling = 1
# increase Linux TCP buffer limits
net.core.rmem_max = 8388608
net.core.wmem_max = 8388608
# increase default and maximum Linux TCP buffer sizes
net.ipv4.tcp_rmem = 4096 262144 8388608
net.ipv4.tcp_wmem = 4096 262144 8388608
# increase max backlog to avoid dropped packets
net.core.netdev_max_backlog=2500
```

7.2.2 Tuning the Spectrum Scale daemon (formerly the GPFS daemon)

The parameters to consider while tuning the Spectrum Scale daemon are listed in this section.

The **`pagepool`** parameter determines the size of the Spectrum Scale file data block cache. Unlike local file systems that use the operating system page cache to cache file data, Spectrum Scale allocates its own cache, which is called the *pagepool*. The Spectrum Scale pagepool is used to cache user file data and file system metadata. Along with file data, the pagepool supplies memory for various types of buffers, such as prefetch and write behind. Usually, you should set the value to 25% of the physical memory. For example, to change the **`pagepool`** value to 48G, run **`mmchconfig pagepool=48G`**. Then, shut down and restart the Spectrum Scale daemon by running **`mmshutdown -a`** and **`mmstartup -a`** respectively.

The **`maxMBpS`** parameter is an indicator of the maximum amount of I/O in megabytes that can be submitted by Spectrum Scale per second in or out of a single node (it is not a hard limit). The **`maxMBpS`** value is used to calculate how much I/O can effectively be done for Spectrum Scale prefetch for readers and Spectrum Scale write-behind for writers. The **`maxMBpS`** value should be adjusted for nodes that are direct-attached to storage or internal disks, and set it to aggregate the throughput of the attached disks. For example, to change the **`maxMBpS`** value to 500 (MB/s), run **`mmchconfig maxMBpS=500 -i`**.

The **maxFilesToCache** parameter should be large enough to handle the number of concurrently open files plus allow caching of recently used files. **maxStatCache** can be set higher on user-interactive nodes and smaller on dedicated compute-nodes. **maxFilesToCache** and **maxStatCache** are indirectly affected by the **distributedTokenServer** parameter because distributing the tokens across multiple token servers might allow keeping more tokens than if a file system has only one token server. For example, to change the **maxFilesToCache** value to 100000, run **mmchconfig maxFilesToCache=100000**. Then, shutdown and restart the Spectrum Scale daemon with the commands **mmshutdown -a** and **mmstartup -a** respectively.

The **prefetchThreads** parameter controls the maximum threads that are dedicated to prefetching data for sequential file reads or to handle sequential write-behind. Prefetch threads should have twice the number of disks that are available to the node. The default value of 72 should work well in most configurations.

The **worker1Threads** parameter controls the maximum number of concurrent file operations at any one instant, primarily for random read and write operations that cannot be prefetched. For big data applications, increase this setting to 72 by running **mmchconfig worker1Threads=72 -i**.

7.2.3 Configuring the Spectrum Scale file system

With the MapReduce process, shuffle/sort data is passed from mappers to reducers by writing the data to a local file system. A preferred practice is to store local MapReduce data on an ext3/ext4 file system. MapReduce jobs end if they need more shuffle file space than what is available. Hence, reserve 25% total disk space for local file system as a rule of thumb because disks can be partitioned with the first partition that is used by Spectrum Scale, and the second by ext3/ext4.

Here are some performance hints for when you create or configure a file system. For a file system based on FPO, to enable the policy to read replicas from local disks, run **mmchconfig readReplicaPolicy=local –i**. The replication for both data and metadata is normally set to three replicas. Some temporary files and logs files can be set to one replica to improve performance. Quotas cannot be activated on this file system. An inode size of 4096 and 256 K meta block size are preferred for typical MapReduce metadata sizes. A data block size of 1 M and a block group factor size of 128 are optimal values for MapReduce data sizes. The **-S** and **-E** options for the **mmcrfs** command help improve performance for *mtime* and *atime* updates. For a file system that is based on the IBM Elastic Storage Server (ESS), there is no data locality, and you can reduce the replication number to 2, or even 1, because Spectrum Scale Native RAID inside the ESS helps protect data.

7.3 Tuning the Platform Symphony MapReduce framework

As with Apache MapReduce, the number of slots is probably the first thing you think of when tuning your environment. However, with Platform Symphony, the way you calculate the number of slots is slightly different. Apache MapReduce has more map slots than reduce slots, usually a ratio of two to one. Platform Symphony has no distinction between map or reduce slots; there are simply slots. This makes it important that you do not over populate your nodes with slots, so consider the amount of memory you have available and the amount of memory you are allocating to each slot before you decide on the number of slots. By default, Platform Symphony sets the number of slots equal to the number of processors on the data node. If you want to change it, open the `$BIGINSIGHTS_HOME/hadoop-conf/components/HAManager/conf/ResourceGroups.xml` file and change the line `<ResourceGroup ResourceGroupName=”ComputeHosts” available slots =”32”>` to the number of slots that you wan to create. Then, run **`syncconf.sh`**. You also can specify the ideal slot ratio to run map tasks and reduce tasks or prioritize critical applications with application profiles.

Considerations for the sort and buffer settings are similar to those that are described in 7.1.2, “Tuning Hadoop” on page 150. In addition, Platform Symphony uses the `pmr.ondemand.2nd.sort.mb` parameter to specify whether the second map sort buffer is allocated for the job always or only when required. This parameter helps determine memory allocation for sorting files (specified as `io.sort.mb`) and the maximum heap size for a map task (specified as the **`-Xmx`** JVM option). Another parameter, `pmr.reduce.f2f.factor`, configures the concurrent file-to-file merge during the reduce phase of a MapReduce job. Instead of a single thread, multiple merger threads decompress and merge the map output, enabling each merge thread to work on a subset of segments or files to speed up the file-to-file merge. This task is done both during the intermediate and final merges. A valid value is a range of numbers greater than 1 (exclusive) but lower than or equal to 2 (inclusive). For example, to add the following properties, open the `pmr-site.xml` file in `$BIGINSIGHTS_HOME/HAManager/data/soam/mapreduce/conf/` and add them:

```
<property>
  <name>pmr.ondemand.2nd.sort.mb</name>
  <value>true</value>
</property>

<property>
  <name>pmr.reducer.f2f.factor</name>
  <value>1.5</value>
</property>
```

A

Integration and configuration for IBM Spectrum Scale, Hadoop, and IBM Platform Symphony

This appendix describes how open source Hadoop can be integrated with Spectrum Scale. This appendix shows a full installation run, which can aid in environment configuration.

This appendix covers the following topics:

- Test cluster description
- Configuration of the IBM Spectrum Scale File Placement Optimizer Hadoop Connector
- Installing and configuring IBM Java and Apache Hadoop
- Running a Hadoop MapReduce job
- Installing and configuring IBM Platform Symphony V7.1
- Running Hadoop MapReduce jobs on Platform Symphony
- Installation and configuration of IBM Spectrum Scale (formerly GPFS)

Test cluster description

In the test cluster, there are four logical partitions (LPARs) on an IBM Power 740 with 10g Ethernet interconnections (node1, node2, node3, and node4), Red Hat Enterprise Linux (RHEL) 6.5, a regular user ‘Jeff’ for submitting MapReduce jobs, and a configured passwordless SSH for ‘root’ and ‘Jeff’. IBM Spectrum Scale V4.1.0-4, Hadoop 2.4.1, and IBM Platform Symphony V7.1 are installed and configured in the cluster. node1 is the master host, and node2, node3, and node4 are compute hosts.

Installation and configuration of IBM Spectrum Scale (formerly GPFS)

This section covers Spectrum Scale V4.1 TL1 on Red Hat Enterprise Linux 6.5.

To perform the installation and configuration, complete the following steps:

1. Install the following RPMs on all nodes in the cluster by running the following commands:

```
rpm -ivh gpfs.base-4.1.0-4.ppc64.rpm
rpm -ivh gpfs.gpl-4.1.0-4.noarch.rpm
rpm -ivh gpfs.msg.en_US-4.1.0-4.noarch.rpm
rpm -ivh gpfs.docs-4.1.0-4.noarch.rpm
rpm -ivh gpfs.ext-4.1.0-4.ppc64.rpm
```

2. Set the environment variables and compile the code on all nodes in the cluster by running the following commands:

```
cd /usr/lpp/mmfs/src
export SHARKCLONEROOT=/usr/lpp/mmfs/src
make Autoconfig
make World
make InstallImages
```

3. Create the `/tmp/nodelist` file and add the following lines to it:

```
node1.test.ibm.com:quorum
node2.test.ibm.com:quorum-manager
node3.test.ibm.com:quorum-manager
node4.test.ibm.com
```

4. Edit the Spectrum Scale configuration file and modify to look like the following lines:

```
[root@node1 ~] cat ./configfile
readReplicaPolicy local
minMissedPingTimeout 60
leaseRecoveryWait 65
maxMBpS 400
nsdMinWorkerThreads 48
nsdInlineWriteMax 1M
nsdSmallThreadRatio 2
nsdMaxWorkerThreads 360
nsdThreadsPerDisk 30
nsdThreadsPerQueue 10
disableInodeUpdateOnFdatasync yes
restripeOnDiskFailure yes
unmountOnDiskFail meta
maxFilesToCache 100000
```

```
maxStatCache 10000
prefetchAggressivenessWrite 0
prefetchAggressivenessRead 2
syncBuffsPerIteration 1
enableRepWriteStream 0
forceLogWriteOnFdatasync no
dataDiskCacheProtectionMethod 2
pagepool 4G
```

5. Create the Spectrum Scale cluster by running the following command:

   ```
   [root@node1 ~]# mmcrcluster -N nodelist --ccr-disable  -p node1.test.ibm.com -s 
   node3.test.ibm.com  -r /usr/bin/ssh -R /usr/bin/scp -C testCluster -A -c 
   configfile
   mmcrcluster: Performing preliminary node verification ...
   mmcrcluster: Processing quorum and other critical nodes ...
   mmcrcluster: Processing the rest of the nodes ...
   mmcrcluster: Finalizing the cluster data structures ...
   mmcrcluster: Processing user configuration file configfile
   mmcrcluster: Command successfully completed
   mmcrcluster: Warning: Not all nodes have proper GPFS license designations.
       Use the mmchlicense command to designate licenses as needed.
   mmcrcluster: Propagating the cluster configuration data to all
     affected nodes.  This is an asynchronous process.
   ```

6. Accept the license for all nodes by running the following command:

   ```
   [root@node1 ~]# mmchlicense server --accept -N all
   The following nodes will be designated as possessing GPFS server licenses:
           node1.test.ibm.com
           node2.test.ibm.com
           node3.test.ibm.com
           node4.test.ibm.com
   mmchlicense: Command successfully completed
   mmchlicense: Propagating the cluster configuration data to all
     affected nodes. This is an asynchronous process.
   ```

7. Start the Spectrum Scale (formerly GPFS daemons) daemons and check the state of the cluster by running the following commands:

   ```
   [root@node1 ~]# mmstartup -a
   Sat Sep 27 19:37:18 EDT 2014: mmstartup: Starting GPFS ...
   [root@node1 ~]# mmgetstate -a
   Node number  Node name        GPFS state
   ------------------------------------------
          1      node1            arbitrating
          2      node2            arbitrating
          3      node3            arbitrating
          4      node4            arbitrating

   #If the GPFS state does not change from arbitrating, look into the firewall 
   #configuration (if there is one) and ensure that all nodes in the cluster can 
   #communicate on port 1191. If the communication between the nodes is OK, the 
   #GPFS state should change to active.
   [root@node1 ~]# mmgetstate -a
   Node number  Node name        GPFS state
   ------------------------------------------
          1      node1            active
          2      node2            active
   ```

```
        3      node3            active
        4      node4            active
```

8. List the possible NSD devices across all nodes by running the following command:

```
[root@node1 ~]# mmdsh -N all 'cat /proc/partitions|grep dm'
node4.test.ibm.com:   253        0  876609536 dm-0
node4.test.ibm.com:   253        1  876609536 dm-1
node4.test.ibm.com:   253        2  876609536 dm-2
node4.test.ibm.com:   253        3  876609536 dm-3
node4.test.ibm.com:   253        4  876609536 dm-4
node2.test.ibm.com:   253        0  876609536 dm-0
node2.test.ibm.com:   253        1  876609536 dm-1
node2.test.ibm.com:   253        2  876609536 dm-2
node2.test.ibm.com:   253        3  876609536 dm-3
node2.test.ibm.com:   253        4  876609536 dm-4
node1.test.ibm.com:   253        0  876609536 dm-0
node1.test.ibm.com:   253        1  876609536 dm-1
node1.test.ibm.com:   253        2  876609536 dm-2
node1.test.ibm.com:   253        3  876609536 dm-3
node1.test.ibm.com:   253        4  876609536 dm-4
node3.test.ibm.com:   253        0  876609536 dm-0
node3.test.ibm.com:   253        1  876609536 dm-1
node3.test.ibm.com:   253        2  876609536 dm-2
node3.test.ibm.com:   253        3  876609536 dm-3
node3.test.ibm.com:   253        4  876609536 dm-4
```

9. Create the NSD stanza file by running the following command:

```
[root@node1 ~]# cat nsdstanza.gpfs1
%pool:
        pool=system
        blockSize=256K
        layoutMap=cluster
        usage=metadataOnly
        allowWriteAffinity=no
%pool:
        pool=sp1
        blockSize=1M
        layoutMap=cluster
        usage=dataOnly
        allowWriteAffinity=yes

        blockGroupFactor=1024
        writeAffinityDepth=1
#This is NSD stanza for Node node1
%nsd:
        device=/dev/dm-2
        servers=node1
        nsd=META_node1_DISK1
        usage=metadataOnly
        failureGroup=1101
        pool=system

%nsd:
        device=/dev/dm-0
        servers=node1
        nsd=DATA_node1_DISK2
        usage=dataOnly
        failureGroup=1,1,101
        pool=sp1

%nsd:
        device=/dev/dm-3
        servers=node1
        nsd=DATA_node1_DISK3
        usage=dataOnly
        failureGroup=1,1,101
        pool=sp1

%nsd:
        device=/dev/dm-1
        servers=node1
        nsd=DATA_node1_DISK4
        usage=dataOnly
        failureGroup=1,1,101
        pool=sp1

%nsd:
        device=/dev/dm-4
        servers=node1
        nsd=DATA_node1_DISK5
        usage=dataOnly
        failureGroup=1,1,101
        pool=sp1
......(similar node stanza section for each node)
```

10. Create the NSD by running the following command:

```
[root@node1 ~]# mmcrnsd -F nsdstanza.gpfs1 -v no
mmcrnsd: Processing disk dm-2
mmcrnsd: Processing disk dm-0
mmcrnsd: Processing disk dm-3
mmcrnsd: Processing disk dm-1
mmcrnsd: Processing disk dm-4
mmcrnsd: Processing disk dm-1
mmcrnsd: Processing disk dm-0
mmcrnsd: Processing disk dm-2
mmcrnsd: Processing disk dm-3
mmcrnsd: Processing disk dm-4
mmcrnsd: Processing disk dm-1
mmcrnsd: Processing disk dm-0
mmcrnsd: Processing disk dm-3
mmcrnsd: Processing disk dm-4
mmcrnsd: Processing disk dm-2
mmcrnsd: Processing disk dm-0
mmcrnsd: Processing disk dm-1
mmcrnsd: Processing disk dm-2
mmcrnsd: Processing disk dm-3
mmcrnsd: Processing disk dm-4
mmcrnsd: Propagating the cluster configuration data to all affected nodes.
This is an asynchronous process.

#List the NSD:
```

```
[root@node1 ~]# mmlsnsd -X

 Disk name    NSD volume ID      Device         Devtype  Node name
Remarks
---------------------------------------------------------------------------------
-------
 DATA_node1_DISK2 1433AC10542770EE  /dev/dm-0    dmm    node1.test.ibm.com
server node
 DATA_node1_DISK3 1433AC10542770EF  /dev/dm-3    dmm    node1.test.ibm.com
server node
 DATA_node1_DISK4 1433AC10542770F0  /dev/dm-1    dmm    node1.test.ibm.com
server node
 DATA_node1_DISK5 1433AC10542770F1  /dev/dm-4    dmm    node1.test.ibm.com
server node
 DATA_node2_DISK2 1434AC10542770F7  /dev/dm-0    dmm    node2.test.ibm.com
server node
 DATA_node2_DISK3 1434AC10542770F9  /dev/dm-2    dmm    node2.test.ibm.com
server node
 DATA_node2_DISK4 1434AC10542770FC  /dev/dm-3    dmm    node2.test.ibm.com
server node
 DATA_node2_DISK5 1434AC10542770FF  /dev/dm-4    dmm    node2.test.ibm.com
server node
 DATA_node3_DISK2 1435AC1054277105  /dev/dm-0    dmm    node3.test.ibm.com
server node
 DATA_node3_DISK3 1435AC1054277108  /dev/dm-3    dmm    node3.test.ibm.com
server node
 DATA_node3_DISK4 1435AC105427710B  /dev/dm-4    dmm    node3.test.ibm.com
server node
 DATA_node3_DISK5 1435AC105427710E  /dev/dm-2    dmm    node3.test.ibm.com
server node
 DATA_node4_DISK2 1436AC1054277113  /dev/dm-1    dmm    node4.test.ibm.com
server node
 DATA_node4_DISK3 1436AC1054277116  /dev/dm-2    dmm    node4.test.ibm.com
server node
 DATA_node4_DISK4 1436AC1054277119  /dev/dm-3    dmm    node4.test.ibm.com
server node
 DATA_node4_DISK5 1436AC105427711C  /dev/dm-4    dmm    node4.test.ibm.com
server node
 META_node1_DISK1 1433AC10542770ED  /dev/dm-2    dmm    node1.test.ibm.com
server node
 META_node2_DISK1 1434AC10542770F4  /dev/dm-1    dmm    node2.test.ibm.com
server node
 META_node3_DISK1 1435AC1054277102  /dev/dm-1    dmm    node3.test.ibm.com
server node
 META_node4_DISK1 1436AC1054277111  /dev/dm-0    dmm    node4.test.ibm.com
server node
```

11. Create the Spectrum Scale file system `/gpfs1` by running the following command:

```
[root@node1 ~]# mmcrfs /gpfs1 /dev/gpfs1 -F ./nsdstanza.gpfs1 -n 4 -R 3 -M 3 -r
3 -m 3 -i 4k -A yes -v no
The following disks of gpfs1 will be formatted on node node2.test.ibm.com:
    META_node1_DISK1: size 856064 MB
    DATA_node1_DISK2: size 856064 MB
    DATA_node1_DISK3: size 856064 MB
    DATA_node1_DISK4: size 856064 MB
    DATA_node1_DISK5: size 856064 MB
    META_node2_DISK1: size 856064 MB
    DATA_node2_DISK2: size 856064 MB
    DATA_node2_DISK3: size 856064 MB
    DATA_node2_DISK4: size 856064 MB
    DATA_node2_DISK5: size 856064 MB
    META_node3_DISK1: size 856064 MB
    DATA_node3_DISK2: size 856064 MB
    DATA_node3_DISK3: size 856064 MB
    DATA_node3_DISK4: size 856064 MB
    DATA_node3_DISK5: size 856064 MB
    META_node4_DISK1: size 856064 MB
    DATA_node4_DISK2: size 856064 MB
    DATA_node4_DISK3: size 856064 MB
    DATA_node4_DISK4: size 856064 MB
    DATA_node4_DISK5: size 856064 MB
Formatting file system ...
Disks up to size 7.2 TB can be added to storage pool system.
Disks up to size 7.2 TB can be added to storage pool sp1.
Creating Inode File
  27 % complete on Sat Sep 27 22:33:39 2014
  59 % complete on Sat Sep 27 22:33:44 2014
  89 % complete on Sat Sep 27 22:33:49 2014
 100 % complete on Sat Sep 27 22:33:51 2014
Creating Allocation Maps
Creating Log Files
Clearing Inode Allocation Map
Clearing Block Allocation Map
Formatting Allocation Map for storage pool system
Formatting Allocation Map for storage pool sp1
Completed creation of file system /dev/gpfs1.
mmcrfs: Propagating the cluster configuration data to all affected nodes.  This
is an asynchronous process.

#Mount the file system:
[root@node1 ~]# mmmount gpfs1 -a
```

12.Configure the file placement policy by running the following commands:

```
[root@node1 ~] cat /tmp/policy.fpo
 /* placement rules */
RULE 'default' SET POOL 'sp1'
[root@node1 base]# mmchpolicy gpfs1 /tmp/policy.fpo
Validated policy `policy.fpo': Parsed 1 policy rules.
Policy `policy.fpo' installed and broadcast to all nodes
```

Now, write a test file from node1 and verify that it can be seen from node2, node3, and node4.

Configuration of the IBM Spectrum Scale File Placement Optimizer Hadoop Connector

The IBM Spectrum Scale File Placement Optimizer (Spectrum Scale-FPO) Hadoop Connector allows Hadoop users to access data from a Spectrum Scale-FPO file system.

To configure the Spectrum Scale connector on all FPO nodes, complete the following steps:

1. Run the following commands in node1:

```
[root@node1 ~] cp /usr/lpp/mmfs/fpo/gpfs-connector-daemon
/usr/lpp/mmfs/fpo/gpfs-callback_start_connector_daemon.sh
/usr/lpp/mmfs/fpo/gpfs-callback_stop_connector_daemon.sh /var/mmfs/etc/
```

2. Register and start the Spectrum Scale connector by running the following command:

```
[root@node1 ~]# /usr/lpp/mmfs/fpo/hadoop-2.4.0/install_script/gpfs-callbacks.sh
--add
```

3. Restart the Spectrum Scale daemons by running the following commands:

```
[root@node1 ~]# mmshutdown -a
Tue Nov  4 12:30:24 EST 2014: mmshutdown: Starting force unmount of GPFS file
systems
Tue Nov  4 12:30:29 EST 2014: mmshutdown: Shutting down GPFS daemons
node4.test.ibm.com:  Shutting down!
node3.test.ibm.com:  Shutting down!
node1.test.ibm.com:  Shutting down!
node2.test.ibm.com:  Shutting down!
node4.test.ibm.com:  'shutdown' command about to kill process 24749
node4.test.ibm.com:  Unloading modules from
/lib/modules/2.6.32-431.el6.ppc64/extra
node2.test.ibm.com:  'shutdown' command about to kill process 29307
node2.test.ibm.com:  Unloading modules from
/lib/modules/2.6.32-431.el6.ppc64/extra
node3.test.ibm.com:  'shutdown' command about to kill process 31367
node3.test.ibm.com:  Unloading modules from
/lib/modules/2.6.32-431.el6.ppc64/extra
node1.test.ibm.com:  'shutdown' command about to kill process 26915
node1.test.ibm.com:  Unloading modules from
/lib/modules/2.6.32-431.el6.ppc64/extra
node4.test.ibm.com:  Unloading module mmfs26
node2.test.ibm.com:  Unloading module mmfs26
node3.test.ibm.com:  Unloading module mmfs26
node1.test.ibm.com:  Unloading module mmfs26
node3.test.ibm.com:  Unloading module mmfslinux
node4.test.ibm.com:  Unloading module mmfslinux
node2.test.ibm.com:  Unloading module mmfslinux
node1.test.ibm.com:  Unloading module mmfslinux
Tue Nov  4 12:30:38 EST 2014: mmshutdown: Finished

[root@node1 ~]# mmstartup -a
Tue Nov  4 12:32:01 EST 2014: mmstartup: Starting GPFS ...
[root@node1 ~]# mmgetstate -a

 Node number  Node name        GPFS state
------------------------------------------
       1      node1            active
       2      node2            active
       3      node3            active
       4      node4            active
[root@node1 ~]# mmmount gpfs1 -a
Tue Nov  4 12:32:29 EST 2014: mmmount: Mounting file systems ...
[root@node1 ~]# mmlsmount gpfs1 -L
File system gpfs1 is mounted on 4 nodes:
  172.16.20.52    node2
  172.16.20.53    node3
  172.16.20.51    node1
  172.16.20.54    node4
```

4. Perform a health check by running the following command:

```
[root@node1 ~]# ps -elf |grep gpfs-connector-daemon
4 S root     49717     1  0  60 -20 -   118 skb_re 12:32 ?        00:00:00
/var/mmfs/etc/gpfs-connector-daemon
```

Installing and configuring IBM Java and Apache Hadoop

Install and configure IBM Java for Power Systems. IBM Java can be downloaded from this website:

http://www.ibm.com/developerworks/java/jdk/linux/download.html

At the website, click **Java 7 for 64-bit IBM Power** → **Login with an IBM ID** and answer the questions. Download the package to the cluster and install it as follows:

```
[root@node1 ~] #./ibm-java-sdk-7.1-1.1-ppc64-archive.bin
Preparing to install... Extracting the JRE from the installer archive... Unpacking
the JRE... Extracting the installation resources from the installer archive...
Configuring the installer for this system's environment...
Launching installer...
 Graphical installers are not supported by the VM. The console mode will be used
instead...
===============================================================================
Choose Locale... ----------------
 1- Bahasa Indonesia 2- Catal? 3- Deutsch ->4- English 5- Espa?ol 6- Fran?ais 7-
Italiano 8- Portugu?s (Brasil)
CHOOSE LOCALE BY NUMBER:
===============================================================================
IBM 64-bit SDK for Linux, Java Technology Edition, Version 7.1(created with
InstallAnywhere)
----------------------------------------------------------------------------------
Preparing CONSOLE Mode Installation...

===============================================================================
License Agreement ------------------
Installation and Use of IBM 64-bit SDK for Linux, Java Technology Edition, Version
7.1 Requires Acceptance of the Following License Agreement:
International License Agreement for Non-Warranted Programs
Part 1 - General Terms
BY DOWNLOADING, INSTALLING, COPYING, ACCESSING, CLICKING ON AN "ACCEPT" BUTTON, OR
OTHERWISE USING THE PROGRAM, LICENSEE AGREES TO THE TERMS OF THIS AGREEMENT. IF
YOU ARE ACCEPTING THESE TERMS ON BEHALF OF LICENSEE, YOU REPRESENT AND WARRANT
```

```
THAT YOU HAVE FULL AUTHORITY TO BIND LICENSEE TO THESE TERMS. IF YOU DO NOT AGREE
TO THESE TERMS,
* DO NOT DOWNLOAD, INSTALL, COPY, ACCESS, CLICK ON AN "ACCEPT" BUTTON, OR USE THE
PROGRAM; AND
* PROMPTLY RETURN THE UNUSED MEDIA AND DOCUMENTATION TO THE PARTY FROM WHOM IT WAS
OBTAINED FOR A REFUND OF THE AMOUNT PAID. IF THE PROGRAM WAS DOWNLOADED, DESTROY
ALL COPIES OF THE PROGRAM.
1. Definitions
"Authorized Use" - the specified level at which Licensee is authorized to execute
or run the Program. That level may be measured by number of users, millions of
service units ("MSUs"), Processor Value Units ("PVUs"), or other level of use
specified by IBM.
PRESS <ENTER> TO CONTINUE:
UNTIL
Licensee may not transfer TPC to another party except as a transfer accompanying
the Program. TPC may contain a disabling device that will prevent it from being
used after the evaluation period ends. Licensee will not tamper with this
disabling device or the TPC. Licensee should take precautions to avoid any loss of
data that might result when the TPC can no longer be used.
L/N: L-EWOD-99YA4J
PRESS <ENTER> TO CONTINUE:
D/N: L-EWOD-99YA4J P/N: L-EWOD-99YA4J
DO YOU ACCEPT THE TERMS OF THIS LICENSE AGREEMENT? (Y/N):Y
 ===============================================================================
Introduction ------------
InstallAnywhere will guide you through the installation of IBM 64-bit SDK for
Linux, Java Technology Edition, Version 7.1.
It is strongly recommended that you quit all programs before continuing with this
installation.
Respond to each prompt to proceed to the next step in the installation. If you
want to change something on a previous step, type 'back'.
You may cancel this installation at any time by typing 'quit'.
PRESS <ENTER> TO CONTINUE:

===============================================================================
Choose Install Folder ---------------------
Where would you like to install?
 Default Install Folder: /root/ibm-java-ppc64-71
ENTER AN ABSOLUTE PATH, OR PRESS <ENTER> TO ACCEPT THE DEFAULT : /opt/ibm/java
INSTALL FOLDER IS: /opt/ibm/java IS THIS CORRECT? (Y/N): Y

===============================================================================
Pre-Installation Summary ------------------------
Review the Following Before Continuing:
Product Name: IBM 64-bit SDK for Linux, Java Technology Edition, Version 7.1
Install Folder: /opt/ibm/java
Link Folder: /root
Disk Space Information (for Installation Target): Required: 251,426,941 bytes
Available: 50,788,216,832 bytes
PRESS <ENTER> TO CONTINUE:
===============================================================================
Installing... -------------
 [==================|==================|==================|==================]
[------------------|------------------|------------------|------------------]
```

```
===============================================================================
Installation Complete ---------------------
Congratulations. IBM 64-bit SDK for Linux, Java Technology Edition, Version 7.1
has been successfully installed to:
 /opt/ibm/java
PRESS <ENTER> TO EXIT THE INSTALLER:
```

Now, configure Hadoop by completing the following steps:

1. Complete the following steps on all nodes in the cluster:
   a. Extract the Hadoop installation files and correct the 32-/64-bit problem:

   ```
   [Jeff@node1 ~]$ tar -xzf hadoop-2.4.1.tar.gz
   ```

   > **Note:** If the Hadoop distribution is downloaded from the Hadoop official link, rebuild it because although the Hadoop native library is for a 64-bit machine, the distribution is built on a 32-bit machine, so there is potential risk of running a 32-bit native library on a 64-bit machine. For more information about this topic, see the following website:
   >
   > http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/NativeLibraries.html

   Build the 64-bit Hadoop native library by running the following commands:

   ```
   [Jeff@node1 ~]$ file libhadoop.so.1.0.0
   libhadoop.so.1.0.0: ELF 64-bit MSB shared object, 64-bit PowerPC or cisco
   7500, version 1 (SYSV), dynamically linked, not stripped
   [Jeff@node1 ~]# cp libhadoop.so.1.0.0
   /home/Jeff/hadoop-2.4.1/lib/native/libhadoop.so.1.0.0'
   [Jeff@node1 ~]# chown Jeff:Jeff
   /home/Jeff/hadoop-2.4.1/lib/native/libhadoop.so.1.0.0
   ```

   b. Append the following lines to ~/.bashrc:

   ```
   export JAVA_HOME=/opt/ibm/java
   export HADOOP_PREFIX=/home/Jeff/hadoop-2.4.1
   export PATH=.:$JAVA_HOME/bin/:$HADOOP_PREFIX/bin:$HADOOP_PREFIX/sbin:$PATH
   ```

   c. Source ~/.bashrc by running the following command:

   ```
   [Jeff@node1 ~]$ source .bashrc
   ```

   d. Link a library on each node. The following example shows the process for node1:

   ```
   [root@node1 ~]#  export HADOOP_HOME=/home/Jeff/hadoop-2.4.1
   [root@node1 ~]#  ln -sf /usr/lpp/mmfs/fpo/hadoop-2.4.0/hadoop-2.4.0-gpfs.jar
   $HADOOP_HOME/share/hadoop/common
   [root@node1 ~]# ln -sf /usr/lpp/mmfs/fpo/hadoop-2.4.0/libgpfshadoop.64.so
   $HADOOP_HOME/lib/native/libgpfshadoop.so
   [root@node1 ~]# ln -sf /usr/lpp/mmfs/lib/libgpfs.so
   $HADOOP_HOME/lib/native/libgpfs.so
   ```

2. Edit the Hadoop configuration files:
   a. Edit core-site.xml as follows:

   ```
   [Jeff@node1 ~]# vim  /home/Jeff/hadoop-2.4.1/etc/hadoop/core-site.xml
   <!-- Put site-specific property overrides in this file. -->

   <configuration>
      <property>
   ```

```
        <name>fs.defaultFS</name>
        <value>gpfs:///</value>
    </property>
    <property>
        <name>fs.gpfs.impl</name>
        <value>org.apache.hadoop.fs.gpfs.GeneralParallelFileSystem</value>
    </property>
    <property>
        <name>fs.AbstractFileSystem.gpfs.impl</name>
        <value>org.apache.hadoop.fs.gpfs.GeneralParallelFs</value>
    </property>
    <property>
        <name>gpfs.default.fpopool.bgf</name>
        <value>1024</value>
    </property>
    <property>
        <name>gpfs.mount.dir</name>
        <value>/gpfs1</value>
    </property>
    <property>
        <name>gpfs.supergroup</name>
        <value>Jeff</value>
    </property>
</configuration>
```

b. Edit `mapred-site.xml` as follows:

```
[Jeff@node1 ~]$ cp
/home/Jeff/hadoop-2.4.1/etc/hadoop/mapred-site.xml.template
/home/Jeff/hadoop-2.4.1/etc/hadoop/mapred-site.xml
[Jeff@node1 ~]$ vim /home/Jeff/hadoop-2.4.1/etc/hadoop/mapred-site.xml

<!-- Put site-specific property overrides in this file. -->

<configuration>
    <property>
        <name>mapreduce.framework.name</name>
        <value>yarn</value>
    </property>
    <property>
        <name>mapreduce.jobhistory.address</name>
        <value>node1:10020</value>
    </property>
    <property>
        <name>mapreduce.jobhistory.webapp.address</name>
        <value>node1:19888</value>
    </property>
</configuration>
```

c. Add subordinate nodes to the `slaves` file by running the following command:

```
[Jeff@node1 ~]$ cd hadoop-2.4.1/etc/hadoop/
[Jeff@node1 hadoop]$ cat slaves
node2
node3
node4
```

d. Append the following two lines to the bottom of the `yarn-env.sh` file and correct the `JAVA_HOME` environment variable:

```
[Jeff@node1 `]$ vim /home/Jeff/hadoop-2.4.1/etc/hadoop/yarn-env.sh
     export JAVA_HOME=/opt/ibm/java
...output cut...
     export HADOOP_COMMON_LIB_NATIVE_DIR=${HADOOP_PREFIX}/lib/native
     export HADOOP_OPTS="-Djava.library.path=$HADOOP_PREFIX/lib/native"
```

e. Edit the `hdfs-site.xml` file as follows:

```
[Jeff@node1 ~]$ vim /home/Jeff/hadoop-2.4.1/etc/hadoop/hdfs-site.xml
<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>dfs.replication</name>
<value>3</value>
</property>
</configuration>
```

f. Edit the `yarn-site.xml` file as follows:

```
...output cut...
<configuration>

<!-- Site specific YARN configuration properties -->
    <property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
    </property>
    <property>
        <name>yarn.resourcemanager.address</name>
        <value>node1:8032</value>
    </property>
    <property>
        <name>yarn.resourcemanager.scheduler.address</name>
        <value>node1:8030</value>
    </property>
    <property>
        <name>yarn.resourcemanager.resource-tracker.address</name>
        <value>node1:8031</value>
    </property>
</configuration>
```

3. Check whether Hadoop connects to Spectrum Scale file systems successfully by running the following command:

```
[Jeff@node1 ~]$ hadoop fs -ls /
14/11/04 13:54:41 INFO Configuration.deprecation: dfs.block.size is deprecated.
Instead, use dfs.blocksize
14/11/04 13:54:41 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus       src=gpfs:/   dst=null
perm=null
14/11/04 13:54:41 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus       src=gpfs:/   dst=null
perm=null
14/11/04 13:54:41 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=listStatus  src=gpfs:/  dst=null perm=null
```

```
Found 12 items
drwxr-xr-x   - root root       4096 2014-10-31 15:39
/QuasiMonteCarlo_1414784357285_2044788281
-rw-r--r--   3 root root        292 2014-10-31 15:39
/api.node1.test.ibm.com.log
drwxrwxrwx   - Jeff Jeff       4096 2014-09-30 08:53 /hadoop
drwxr-xr-x   - root root       4096 2014-11-04 11:14 /in
drwxr-xr-x   - root root       4096 2014-11-04 11:14 /out
drwxrwxrwx   - root root       4096 2014-09-27 22:35 /symphony
drwxr-xr-x   - Jeff Jeff       4096 2014-09-30 14:41 /test
drwxr-xr-x   - Jeff Jeff       4096 2014-09-30 16:31 /test2
drwxr-xr-x   - Jeff Jeff       4096 2014-10-08 03:16 /test3
-rw-r--r--   3 Jeff Jeff          0 2014-10-08 06:05 /testfile
drwx------   - Jeff Jeff       4096 2014-09-30 13:43 /tmp
drwxrwxrwx   - Jeff Jeff       4096 2014-10-08 07:13 /user
```

4. Start yarn and check the node status by running the following command:

```
[Jeff@node1 ~]$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to
/home/Jeff/hadoop-2.4.1/logs/yarn-Jeff-resourcemanager-node1.test.ibm.com.out
node4: starting nodemanager, logging to
/home/Jeff/hadoop-2.4.1/logs/yarn-Jeff-nodemanager-node4.test.ibm.com.out
node2: starting nodemanager, logging to
/home/Jeff/hadoop-2.4.1/logs/yarn-Jeff-nodemanager-node2.test.ibm.com.out
node3: starting nodemanager, logging to
/home/Jeff/hadoop-2.4.1/logs/yarn-Jeff-nodemanager-node3.test.ibm.com.out

[Jeff@node1 ~]$  yarn node -list all
14/11/04 15:52:03 INFO client.RMProxy: Connecting to ResourceManager at
node1/172.16.20.51:8032
Total Nodes:3
         Node-Id             Node-State Node-Http-Address
Number-of-Running-Containers
    node2:32993             RUNNING      node2:8042
0
    node4:43742             RUNNING      node4:8042
0
    node3:52273             RUNNING      node3:8042
0
```

Running a Hadoop MapReduce job

You can run a Hadoop MapReduce job by completing the following steps:

1. Define the `LD-LIBRARY_PATH` to `$HADOOP_HOME/lib/native` by running the following commands:

```
[Jeff@node1 ~]$ export HADOOP_HOME=/home/Jeff/hadoop-2.4.1
[Jeff@node1 ~]$ export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:$HADOOP_HOME/lib/native
```

2. Generate the input files to the Spectrum Scale file system by running the following commands:

```
[Jeff@node1 ~]$ mkdir /gpfs1/input /gpfs1/output
[Jeff@node1 ~]$ vim /gpfs1/input/words
```

```
[Jeff@node1 ~]$ cat /gpfs1/input/words
Implementing an IBM InfoSphere BigInsight Cluster using Linux on Power
Implementing an IBM InfoSphere BigInsight Cluser
Implementing an IBM InfoSphere
```

3. Run a Hadoop MapReduce job by running the following command. Here is an example of *word count.*

```
[Jeff@node1 ~]$ hadoop jar
$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.4.1.jar
wordcount /input /output/wc_out_1
14/11/04 16:21:04 INFO Configuration.deprecation: dfs.block.size is deprecated.
Instead, use dfs.blocksize
14/11/04 16:21:04 INFO client.RMProxy: Connecting to ResourceManager at
node1/172.16.20.51:8032
14/11/04 16:21:04 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff/.staging   dst=null        perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=false
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013  dst=null
perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff/.staging   dst=null        perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff    dst=null        perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus       src=/tmp/hadoop-yarn/staging
dst=null        perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff      ip=null cmd=getFileStatus      src=/tmp/hadoop-yarn dst=null
perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus       src=/tmp     dst=null
perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus       src=/dst=null        perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=mkdirs
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013  dst=null
perm=rwxrwxrwx
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=setPermission
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013  dst=null
perm=rwx------
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.jar
dst=null        perm=rw-rw-rw-
14/11/04 16:21:05 WARN gpfs.GeneralParallelFileSystem: replication out of range
(1..3): 10; replication set to 3
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=setReplication
```

```
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.jar
dst=null        perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=setPermission
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.jar
dst=null        perm=rw-r--r--
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff      ip=null cmd=getFileStatus      src=gpfs:/input     dst=null
perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff      ip=null cmd=getFileStatus      src=gpfs:/input     dst=null
perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=listStatus  src=gpfs:/input      dst=null
perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus       src=gpfs:/input/words
dst=null        perm=null
14/11/04 16:21:05 INFO input.FileInputFormat: Total input paths to process : 1
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.split
dst=null        perm=rw-rw-rw-
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=setPermission
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.split
dst=null        perm=rw-r--r--
14/11/04 16:21:05 WARN gpfs.GeneralParallelFileSystem: replication out of range
(1..3): 10; replication set to 3
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=setReplication
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.split
dst=null        perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.splitmeta
info  dst=null        perm=rw-rw-rw-
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=setPermission
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.splitmeta
info  dst=null        perm=rw-r--r--
14/11/04 16:21:05 INFO mapreduce.JobSubmitter: number of splits:1
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.xml
dst=null        perm=rw-rw-rw-
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=setPermission
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.xml
dst=null        perm=rw-r--r--
14/11/04 16:21:05 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1412772835374_0013
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=resolveLink
```

```
src=gpfs:/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013
dst=null        perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.xml
dst=null        perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=resolveLink
src=gpfs:/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.xml
dst=null        perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.jar
dst=null        perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=resolveLink
src=gpfs:/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.jar
dst=null        perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.split
dst=null        perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=resolveLink
src=gpfs:/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.spli
t     dst=null        perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.splitmeta
info  dst=null        perm=null
14/11/04 16:21:05 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=resolveLink
src=gpfs:/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0013/job.spli
tmetainfo     dst=null        perm=null
14/11/04 16:21:05 INFO impl.YarnClientImpl: Submitted application
application_1412772835374_0013
14/11/04 16:21:05 INFO mapreduce.Job: The url to track the job:
http://node1.test.ibm.com:8088/proxy/application_1412772835374_0013/
14/11/04 16:21:05 INFO mapreduce.Job: Running job: job_1412772835374_0013
14/11/04 16:21:12 INFO mapreduce.Job: Job job_1412772835374_0013 running in
uber mode : false
14/11/04 16:21:12 INFO mapreduce.Job:  map 0% reduce 0%
14/11/04 16:21:18 INFO mapreduce.Job:  map 100% reduce 0%
14/11/04 16:21:24 INFO mapreduce.Job:  map 100% reduce 100%
14/11/04 16:21:24 INFO mapreduce.Job: Job job_1412772835374_0013 completed
successfully
14/11/04 16:21:24 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=150
                FILE: Number of bytes written=186749
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                GPFS: Number of bytes read=234
                GPFS: Number of bytes written=100
                GPFS: Number of read operations=2
                GPFS: Number of large read operations=0
                GPFS: Number of write operations=3
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Rack-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=3108
                Total time spent by all reduces in occupied slots (ms)=3632
                Total time spent by all map tasks (ms)=3108
                Total time spent by all reduce tasks (ms)=3632
                Total vcore-seconds taken by all map tasks=3108
                Total vcore-seconds taken by all reduce tasks=3632
                Total megabyte-seconds taken by all map tasks=3182592
                Total megabyte-seconds taken by all reduce tasks=3719168
        Map-Reduce Framework
                Map input records=3
                Map output records=20
                Map output bytes=231
                Map output materialized bytes=150
                Input split bytes=82
                Combine input records=20
                Combine output records=11
                Reduce input groups=11
                Reduce shuffle bytes=150
                Reduce input records=11
                Reduce output records=11
                Spilled Records=22
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=210
                CPU time spent (ms)=1660
                Physical memory (bytes) snapshot=292290560
                Virtual memory (bytes) snapshot=2374303744
                Total committed heap usage (bytes)=164233216
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=152
        File Output Format Counters
                Bytes Written=100
```

Checking the results:

```
[Jeff@node1 ~]$ cat /gpfs1/output/wc_out_1/part-r-00000
BigInsight      2
Cluster  1
IBM     3
Implementing    3
InfoSphere      3
```

```
Linux   1
Power   1
an      3
on      1
using   1
```

Here is an example of a Hadoop MapReduce job pi:

```
[Jeff@node1 ~]$ hadoop jar
hadoop-2.4.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.4.1.jar pi 10
10
Number of Maps  = 10
Samples per Map = 10
14/11/04 16:14:10 INFO Configuration.deprecation: dfs.block.size is deprecated.
Instead, use dfs.blocksize
14/11/04 16:14:10 INFO GeneralParallelFileSystem.audit: allowed=false
ugi=Jeff        ip=null cmd=getFileStatus
src=QuasiMonteCarlo_1415135649352_2088722269/in      dst=null        perm=null
14/11/04 16:14:10 INFO GeneralParallelFileSystem.audit: allowed=false
ugi=Jeff        ip=null cmd=getFileStatus
src=QuasiMonteCarlo_1415135649352_2088722269 dst=null        perm=null
14/11/04 16:14:10 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=mkdirs
src=QuasiMonteCarlo_1415135649352_2088722269/in      dst=null
perm=rwxrwxrwx
14/11/04 16:14:10 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=QuasiMonteCarlo_1415135649352_2088722269/in/part0        dst=null
perm=rw-rw-rw-
Wrote input for Map #0
14/11/04 16:14:10 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=QuasiMonteCarlo_1415135649352_2088722269/in/part1        dst=null
perm=rw-rw-rw-
Wrote input for Map #1
14/11/04 16:14:10 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=QuasiMonteCarlo_1415135649352_2088722269/in/part2        dst=null
perm=rw-rw-rw-
Wrote input for Map #2
14/11/04 16:14:10 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=QuasiMonteCarlo_1415135649352_2088722269/in/part3        dst=null
perm=rw-rw-rw-
Wrote input for Map #3
14/11/04 16:14:10 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=QuasiMonteCarlo_1415135649352_2088722269/in/part4        dst=null
perm=rw-rw-rw-
Wrote input for Map #4
14/11/04 16:14:10 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=QuasiMonteCarlo_1415135649352_2088722269/in/part5        dst=null
perm=rw-rw-rw-
Wrote input for Map #5
```

```
14/11/04 16:14:10 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=QuasiMonteCarlo_1415135649352_2088722269/in/part6        dst=null
perm=rw-rw-rw-
Wrote input for Map #6
14/11/04 16:14:10 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=QuasiMonteCarlo_1415135649352_2088722269/in/part7        dst=null
perm=rw-rw-rw-
Wrote input for Map #7
14/11/04 16:14:10 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=QuasiMonteCarlo_1415135649352_2088722269/in/part8        dst=null
perm=rw-rw-rw-
Wrote input for Map #8
14/11/04 16:14:10 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=QuasiMonteCarlo_1415135649352_2088722269/in/part9        dst=null
perm=rw-rw-rw-
Wrote input for Map #9
Starting Job
14/11/04 16:14:10 INFO client.RMProxy: Connecting to ResourceManager at
node1/172.16.20.51:8032
14/11/04 16:14:10 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff/.staging   dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=false
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012  dst=null
perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff/.staging   dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff    dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus       src=/tmp/hadoop-yarn/staging
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff      ip=null cmd=getFileStatus     src=/tmp/hadoop-yarn dst=null
perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus       src=/tmp     dst=null
perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus       src=/dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=mkdirs
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012  dst=null
perm=rwxrwxrwx
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=setPermission
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012  dst=null
perm=rwx------
```

```
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.jar
dst=null        perm=rw-rw-rw-
14/11/04 16:14:11 WARN gpfs.GeneralParallelFileSystem: replication out of range
(1..3): 10; replication set to 3
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=setReplication
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.jar
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=setPermission
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.jar
dst=null        perm=rw-r--r--
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=gpfs:/user/Jeff/QuasiMonteCarlo_1415135649352_2088722269/in    dst=null
perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=gpfs:/user/Jeff/QuasiMonteCarlo_1415135649352_2088722269/in    dst=null
perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=listStatus
src=gpfs:/user/Jeff/QuasiMonteCarlo_1415135649352_2088722269/in    dst=null
perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=gpfs:/user/Jeff/QuasiMonteCarlo_1415135649352_2088722269/in/part5
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=gpfs:/user/Jeff/QuasiMonteCarlo_1415135649352_2088722269/in/part8
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=gpfs:/user/Jeff/QuasiMonteCarlo_1415135649352_2088722269/in/part2
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=gpfs:/user/Jeff/QuasiMonteCarlo_1415135649352_2088722269/in/part9
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=gpfs:/user/Jeff/QuasiMonteCarlo_1415135649352_2088722269/in/part0
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=gpfs:/user/Jeff/QuasiMonteCarlo_1415135649352_2088722269/in/part3
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=gpfs:/user/Jeff/QuasiMonteCarlo_1415135649352_2088722269/in/part6
dst=null        perm=null
```

```
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=gpfs:/user/Jeff/QuasiMonteCarlo_1415135649352_2088722269/in/part1
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=gpfs:/user/Jeff/QuasiMonteCarlo_1415135649352_2088722269/in/part4
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=gpfs:/user/Jeff/QuasiMonteCarlo_1415135649352_2088722269/in/part7
dst=null        perm=null
14/11/04 16:14:11 INFO input.FileInputFormat: Total input paths to process : 10
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.split
dst=null        perm=rw-rw-rw-
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=setPermission
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.split
dst=null        perm=rw-r--r--
14/11/04 16:14:11 WARN gpfs.GeneralParallelFileSystem: replication out of range
(1..3): 10; replication set to 3
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=setReplication
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.split
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.splitmeta
info  dst=null        perm=rw-rw-rw-
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=setPermission
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.splitmeta
info  dst=null        perm=rw-r--r--
14/11/04 16:14:11 INFO mapreduce.JobSubmitter: number of splits:10
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=create
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.xml
dst=null        perm=rw-rw-rw-
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=setPermission
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.xml
dst=null        perm=rw-r--r--
14/11/04 16:14:11 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1412772835374_0012
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=resolveLink
src=gpfs:/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.xml
dst=null        perm=null
```

```
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=resolveLink
src=gpfs:/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.xml
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.jar
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=resolveLink
src=gpfs:/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.jar
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.split
dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=resolveLink
src=gpfs:/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.spli
t     dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.splitmeta
info  dst=null        perm=null
14/11/04 16:14:11 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=resolveLink
src=gpfs:/tmp/hadoop-yarn/staging/Jeff/.staging/job_1412772835374_0012/job.spli
tmetainfo     dst=null        perm=null
14/11/04 16:14:12 INFO impl.YarnClientImpl: Submitted application
application_1412772835374_0012
14/11/04 16:14:12 INFO mapreduce.Job: The url to track the job:
http://node1.test.ibm.com:8088/proxy/application_1412772835374_0012/
14/11/04 16:14:12 INFO mapreduce.Job: Running job: job_1412772835374_0012
14/11/04 16:14:19 INFO mapreduce.Job: Job job_1412772835374_0012 running in
uber mode : false
14/11/04 16:14:19 INFO mapreduce.Job:  map 0% reduce 0%
14/11/04 16:14:25 INFO mapreduce.Job:  map 10% reduce 0%
14/11/04 16:14:26 INFO mapreduce.Job:  map 20% reduce 0%
14/11/04 16:14:27 INFO mapreduce.Job:  map 30% reduce 0%
14/11/04 16:14:29 INFO mapreduce.Job:  map 60% reduce 0%
14/11/04 16:14:33 INFO mapreduce.Job:  map 70% reduce 0%
14/11/04 16:14:34 INFO mapreduce.Job:  map 80% reduce 0%
14/11/04 16:14:35 INFO mapreduce.Job:  map 100% reduce 0%
14/11/04 16:14:36 INFO mapreduce.Job:  map 100% reduce 100%
14/11/04 16:14:36 INFO mapreduce.Job: Job job_1412772835374_0012 completed
successfully
14/11/04 16:14:36 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=226
                FILE: Number of bytes written=1029765
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                GPFS: Number of bytes read=8330
                GPFS: Number of bytes written=215
                GPFS: Number of read operations=30
                GPFS: Number of large read operations=0
                GPFS: Number of write operations=4
        Job Counters
                Launched map tasks=10
                Launched reduce tasks=1
                Rack-local map tasks=10
                Total time spent by all maps in occupied slots (ms)=47021
                Total time spent by all reduces in occupied slots (ms)=8221
                Total time spent by all map tasks (ms)=47021
                Total time spent by all reduce tasks (ms)=8221
                Total vcore-seconds taken by all map tasks=47021
                Total vcore-seconds taken by all reduce tasks=8221
                Total megabyte-seconds taken by all map tasks=48149504
                Total megabyte-seconds taken by all reduce tasks=8418304
        Map-Reduce Framework
                Map input records=10
                Map output records=20
                Map output bytes=180
                Map output materialized bytes=280
                Input split bytes=1300
                Combine input records=0
                Combine output records=0
                Reduce input groups=2
                Reduce shuffle bytes=280
                Reduce input records=20
                Reduce output records=0
                Spilled Records=40
                Shuffled Maps =10
                Failed Shuffles=0
                Merged Map outputs=10
                GC time elapsed (ms)=1463
                CPU time spent (ms)=9940
                Physical memory (bytes) snapshot=2111176704
                Virtual memory (bytes) snapshot=13086294016
                Total committed heap usage (bytes)=1571160064
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=1180
        File Output Format Counters
                Bytes Written=97
Job Finished in 26.389 seconds
14/11/04 16:14:36 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=getFileStatus
src=gpfs:/user/Jeff/QuasiMonteCarlo_1415135649352_2088722269/out/reduce-out
dst=null        perm=null
14/11/04 16:14:36 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=open
```

```
src=gpfs:/user/Jeff/QuasiMonteCarlo_1415135649352_2088722269/out/reduce-out
dst=null        perm=null
14/11/04 16:14:36 INFO GeneralParallelFileSystem.audit: allowed=true
ugi=Jeff        ip=null cmd=delete
src=QuasiMonteCarlo_1415135649352_2088722269 dst=null         perm=null
Estimated value of Pi is 3.20000000000000000000
```

Installing and configuring IBM Platform Symphony V7.1

To install and configure IBM Platform Symphony V7.1, run the following steps on each node in the cluster:

1. Set the environmental variables on each node in the cluster by running the following commands:

   ```
   [root@node1 ~]$ export HADOOP_HOME=/home/Jeff/hadoop-2.4.1;export
   HADOOP_VERSION=2_4_x;export JAVA_HOME=/opt/ibm/java;export
   CLUSTERADMIN=root;export CLUSTERNAME=testCluster;export BASEPORT=10020;export
   DERBY_DB_HOST=node1;export SIMPLIFIEDWEM=N

   [root@node1 ~]$ /path/to/symSetup7.1.0_lnx26-lib23-x64.bin --prefix
   /opt/ibm/platformsymphony/ --quiet
   ```

2. Append `EGO_RSH=`”`ssh -q`” to the “`/opt/ibm/platformsymphony/kernel/conf/ego.conf`” file of each node.

3. Run the following commands on the master node (node1):

   ```
   [root@node1 ~]# source /opt/ibm/platformsymphony/profile.platform
   [root@node1 ~]# egoconfig join node1 -f
   [root@node1 ~]# egoconfig setentitlement ./platform_sym_adv_entitlement.dat
   ```

4. Run the following commands on every other node (this example shows node2):

   ```
   [root@node2 ~]# source /opt/ibm/platformsymphony/profile.platform
   [root@node2 ~]# egoconfig join node1 -f
   ```

5. Run the following command on the master node (node1):

   ```
   [root@node1 ~]# egosh ego start all
   ```

6. Verify the symphony installation by running the following commands:

   ```
   [root@node1 ~]# source /opt/ibm/platformsymphony/profile.platform
   [root@node1 ~]# soamlogon -u Admin -x Admin
   Logged on successfully
   [root@node1 ~]# egosh resource list
   NAME     status       mem    swp    tmp   ut    it    pg   r1m  r15s  r15m  ls
   node1.t* ok           11G  1023M   143G   3%     1     0   0.1    10   0.1   1
   node2.t* ok           12G  1023M   162G   2%    37     0     0   3.2     0   1
   node3.t* ok           12G  1023M   162G   2%    36     0   0.1     3     0   1
   node4.t* ok           12G  1014M   162G   3%    36     0     0     3     0   1
   [root@node1 ~]# egosh service list
   SERVICE  STATE    ALLOC CONSUMER RGROUP RESOURCE SLOTS SEQ_NO INST_STATE ACTI
   GPFSmon* DEFINED        /Manage* Manag*
   purger   STARTED  2     /Manage* Manag* node1.t* 1     1      RUN        1
   derbydb  STARTED  3     /Manage* Manag* node1.t* 1     1      RUN        2
   WEBGUI   STARTED  4     /Manage* Manag* node1.t* 1     1      RUN        3
   plc      STARTED  5     /Manage* Manag* node1.t* 1     1      RUN        12
   USSD     DEFINED        /Cluste* Inter*
   ```

```
RS       STARTED  6     /Manage* Manag* node1.t* 1     1      RUN        5
Seconda* DEFINED        /HDFS/S*
MRSS     STARTED  7     /Comput* MapRe* node2.t* 1     2      RUN        13
                                        node1.t* 1     1      RUN        6
                                        node4.t* 1     4      RUN        15
                                        node3.t* 1     3      RUN        14
RSA      STARTED  8     /Cluste* Inter* node2.t* 1     3      RUN        17
                                        node1.t* 1     1      RUN        7
                                        node4.t* 1     4      RUN        18
                                        node3.t* 1     2      RUN        16
WebServ* STARTED  11    /Manage* Manag* node1.t* 1     1      RUN        10
Service* STARTED  9     /Manage* Manag* node1.t* 1     1      RUN        8
NameNode DEFINED        /HDFS/N*
DataNode DEFINED        /HDFS/D*
SD       STARTED  10    /Manage* Manag* node1.t* 1     1      RUN        9

[root@node1 ~]# soamview app
APPLICATION                   STATUS    SSM HOST      SSM PID CONSUMER
MapReduce7.1                  enabled   node1.test.i* 31489   /MapReduceConsum*
symping7.1                    enabled   -             -       /SymTesting/Symp*
symexec7.1                    disabled  -             -       /SymExec/SymExec*

[root@node1 ~]# egosh rg
NAME                  HOSTS         SLOTS         FREE          ALLOCATED
DataNodeRG            4             4             4             0
SecondaryNodeRG       0             0             0             0
NameNodeRG            0             0             0             0
MapReduceInternalReso 4             8             4             4
InternalResourceGroup 4             40            36            4
ComputeHosts          4             8             8             0
ManagementHosts       1             16            6             10
```

Running Hadoop MapReduce jobs on Platform Symphony

To run Hadoop MapReduce jobs on Platform Symphony, complete the following steps:

1. Configure Spectrum Scale-FPO for Platform Symphony by running the following command:

   ```
   [root@node1 ~]# mmdsh -N all ln -sf 
   /usr/lpp/mmfs/fpo/hadoop-2.4.0/hadoop-2.4.0-gpfs.jar 
   /opt/ibm/platformsymphony/soam/mapreduce/7.1/linux2.6-glibc2.5-ppc64/lib/hadoop
   -2.4.x/
   ```

2. Configure the environment for the Platform Symphony administrator by running the following command:

   ```
   [Jeff@node1 ~]$ cat .bash_profile
   # .bash_profile

   # Get the aliases and functions
   if [ -f ~/.bashrc ]; then
           . ~/.bashrc
   fi

   # User specific environment and startup programs
   ```

```
set -o vi
PATH=$PATH:$HOME/bin:/usr/lpp/mmfs/bin
export PATH
export CLUSTERADMIN=Jeff
HADOOP_PREFIX=/home/Jeff/hadoop-2.4.1
PATH=$PATH:.:/opt/ibm/java/bin/:/home/Jeff/hadoop-2.4.1/bin:/home/Jeff/hadoop-2
.4.0/sbin:/usr/lib64/qt-3.3/bin:/usr/local/bin:/bin:/usr/bin:/usr/local/sbin:/u
sr/sbin:/sbin:/home/Jeff/bin
JAVA_HOME=/opt/ibm/java
source /opt/ibm/platformsymphony/profile.platform
soamlogon -u Admin -x Admin
```

3. Submit Hadoop MapReduce jobs. The following job is an example of pi:

```
[Jeff@node1 ~]$ mrsh jar
/home/Jeff/hadoop-2.4.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.4.1.
jar pi 10 10
You are using Hadoop API with 2.4.x version.
Number of Maps  = 10
Samples per Map = 10
14/11/04 21:55:52 GMT INFO Configuration.deprecation: dfs.block.size is
deprecated. Instead, use dfs.blocksize
14/11/04 21:55:52 GMT INFO GeneralParallelFileSystem.audit: allowed=false
ugi=root        ip=null cmd=getFileStatus
src=QuasiMonteCarlo_1415138150695_638312006/in   dst=null        perm=null
14/11/04 21:55:52 GMT INFO GeneralParallelFileSystem.audit: allowed=false
ugi=root        ip=null cmd=getFileStatus
src=QuasiMonteCarlo_1415138150695_638312006      dst=null        perm=null
14/11/04 21:55:52 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=mkdirs
src=QuasiMonteCarlo_1415138150695_638312006/in       dst=null
perm=rwxrwxrwx
14/11/04 21:55:52 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=create
src=QuasiMonteCarlo_1415138150695_638312006/in/part0 dst=null
perm=rw-rw-rw-
Wrote input for Map #0
14/11/04 21:55:52 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=create
src=QuasiMonteCarlo_1415138150695_638312006/in/part1 dst=null
perm=rw-rw-rw-
Wrote input for Map #1
14/11/04 21:55:52 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=create
src=QuasiMonteCarlo_1415138150695_638312006/in/part2 dst=null
perm=rw-rw-rw-
Wrote input for Map #2
14/11/04 21:55:52 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=create
src=QuasiMonteCarlo_1415138150695_638312006/in/part3 dst=null
perm=rw-rw-rw-
Wrote input for Map #3
14/11/04 21:55:52 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=create
src=QuasiMonteCarlo_1415138150695_638312006/in/part4 dst=null
perm=rw-rw-rw-
```

```
Wrote input for Map #4
14/11/04 21:55:52 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=create
src=QuasiMonteCarlo_1415138150695_638312006/in/part5 dst=null
perm=rw-rw-rw-
Wrote input for Map #5
14/11/04 21:55:52 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=create
src=QuasiMonteCarlo_1415138150695_638312006/in/part6 dst=null
perm=rw-rw-rw-
Wrote input for Map #6
14/11/04 21:55:52 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=create
src=QuasiMonteCarlo_1415138150695_638312006/in/part7 dst=null
perm=rw-rw-rw-
Wrote input for Map #7
14/11/04 21:55:52 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=create
src=QuasiMonteCarlo_1415138150695_638312006/in/part8 dst=null
perm=rw-rw-rw-
Wrote input for Map #8
14/11/04 21:55:52 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=create
src=QuasiMonteCarlo_1415138150695_638312006/in/part9 dst=null
perm=rw-rw-rw-
Wrote input for Map #9
Starting Job
14/11/04 21:55:53 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=getFileStatus
src=gpfs:/user/root/QuasiMonteCarlo_1415138150695_638312006/in   dst=null
perm=null
14/11/04 21:55:53 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=getFileStatus
src=gpfs:/user/root/QuasiMonteCarlo_1415138150695_638312006/in   dst=null
perm=null
14/11/04 21:55:53 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=listStatus
src=gpfs:/user/root/QuasiMonteCarlo_1415138150695_638312006/in     dst=null
perm=null
14/11/04 21:55:53 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=getFileStatus
src=gpfs:/user/root/QuasiMonteCarlo_1415138150695_638312006/in/part5
dst=null        perm=null
14/11/04 21:55:53 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=getFileStatus
src=gpfs:/user/root/QuasiMonteCarlo_1415138150695_638312006/in/part8
dst=null        perm=null
14/11/04 21:55:53 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=getFileStatus
src=gpfs:/user/root/QuasiMonteCarlo_1415138150695_638312006/in/part2
dst=null        perm=null
14/11/04 21:55:53 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=getFileStatus
src=gpfs:/user/root/QuasiMonteCarlo_1415138150695_638312006/in/part9
dst=null        perm=null
```

```
14/11/04 21:55:53 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=getFileStatus
src=gpfs:/user/root/QuasiMonteCarlo_1415138150695_638312006/in/part0
dst=null        perm=null
14/11/04 21:55:53 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=getFileStatus
src=gpfs:/user/root/QuasiMonteCarlo_1415138150695_638312006/in/part3
dst=null        perm=null
14/11/04 21:55:53 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=getFileStatus
src=gpfs:/user/root/QuasiMonteCarlo_1415138150695_638312006/in/part6
dst=null        perm=null
14/11/04 21:55:53 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=getFileStatus
src=gpfs:/user/root/QuasiMonteCarlo_1415138150695_638312006/in/part1
dst=null        perm=null
14/11/04 21:55:53 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=getFileStatus
src=gpfs:/user/root/QuasiMonteCarlo_1415138150695_638312006/in/part4
dst=null        perm=null
14/11/04 21:55:53 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=getFileStatus
src=gpfs:/user/root/QuasiMonteCarlo_1415138150695_638312006/in/part7
dst=null        perm=null
14/11/04 21:55:53 GMT INFO input.FileInputFormat: Total input paths to process
: 10
14/11/04 21:55:53 GMT INFO internal.MRJobSubmitter: Connected to
JobTracker(SSM)
14/11/04 21:55:53 GMT INFO Configuration.deprecation:
mapred.output.key.comparator.class is deprecated. Instead, use
mapreduce.job.output.key.comparator.class
14/11/04 21:55:53 GMT INFO Configuration.deprecation:
mapred.compress.map.output is deprecated. Instead, use
mapreduce.map.output.compress
14/11/04 21:55:53 GMT INFO internal.MRJobSubmitter: Job <QuasiMonteCarlo>
submitted, job id <8>
14/11/04 21:55:53 GMT INFO internal.MRJobSubmitter: Job will not verify
intermediate data integrity using checksum.
14/11/04 21:55:53 GMT INFO mapreduce.Job: Running job: job_ssm_0008
14/11/04 21:56:06 GMT INFO mapreduce.Job: Job job_ssm_0008 running in uber mode
: false
14/11/04 21:56:06 GMT INFO mapreduce.Job:  map 0% reduce 0%
14/11/04 21:56:12 GMT INFO mapreduce.Job:  map 10% reduce 0%
14/11/04 21:56:13 GMT INFO mapreduce.Job:  map 100% reduce 0%
14/11/04 21:56:21 GMT INFO mapreduce.Job:  map 100% reduce 100%
14/11/04 21:56:21 GMT INFO mapreduce.Job: Job job_ssm_0008 completed
successfully
14/11/04 21:56:23 GMT INFO mapreduce.Job: Counters: 28
        Map-Reduce Framework
                Map input records=10
                Map output records=20
                Map output bytes=180
                Input split bytes=810
                Combine input records=0
                Combine output records=0
                Reduce input groups=2
                Reduce shuffle bytes=240
                Reduce input records=20
                Reduce output records=0
                Spilled Records=20
                Shuffled Maps =10
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=221
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=0
                FILE: Number of large read operations=0
                FILE: Number of read operations=0
                FILE: Number of write operations=0
                GPFS: Number of bytes read=1180
                GPFS: Number of bytes written=215
                GPFS: Number of large read operations=0
                GPFS: Number of read operations=20
                GPFS: Number of write operations=4
        Shuffle Errors
                CONNECTION=0
                IO_ERROR=0
                WRONG_PATH=0
Job Finished in 30.696 seconds
14/11/04 21:56:23 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=getFileStatus
src=gpfs:/user/root/QuasiMonteCarlo_1415138150695_638312006/out/reduce-out
dst=null        perm=null
14/11/04 21:56:23 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=open
src=gpfs:/user/root/QuasiMonteCarlo_1415138150695_638312006/out/reduce-out
dst=null        perm=null
14/11/04 21:56:23 GMT INFO GeneralParallelFileSystem.audit: allowed=true
ugi=root        ip=null cmd=delete
src=QuasiMonteCarlo_1415138150695_638312006  dst=null        perm=null
Estimated value of Pi is 3.20000000000000000000
```

B

Scripts

This appendix provides the scripts that are used to build the image profile and the cluster template in IBM Platform Cluster Manager to create the IBM InfoSphere BigInsights cluster. These scripts are included *as is* because there is no support that is provided by IBM for them.

This appendix covers the following topics:

- Postscripts that are defined in the InfoSphere BigInsights cluster template
- Postscripts that are defined in the image profile

Postscripts that are defined in the InfoSphere BigInsights cluster template

The following scripts must run on all nodes to set up the InfoSphere BigInsights cluster. Complete the following steps:

1. Change the host name of the node to the host name of the 10 Gb Ethernet interface, as shown in Example B-1.

Example B-1 changeHostname10Gb

```
#!/bin/bash
# BigInsights cluster is defined using hostname of servers.
# Change hostname to 10Gb interface -- first mlx4_core device configured

#echo "start with hostname:  `hostname`" >> $LOG

DOMAIN=data.ibm.com

IF=`grep -A1 mlx4_core /etc/udev/rules.d/70-persistent-net.rules  | grep "NAME=" |
awk -F "NAME="  '{print $2}' | sed -e "s/\"//g" | sort -n | head -1`

grep -q "${IF}.${DOMAIN}" /etc/sysconfig/network
if [ $? -ne 0 ] ; then

sed -i "/HOSTNAME=/ s/$/-${IF}.${DOMAIN}/" /etc/sysconfig/network

# update hostname in current session

hostname10Gb=`grep HOSTNAME /etc/sysconfig/network | awk -F "=" '{print \$2}'`
hostname ${hostname10Gb}

fi
```

2. Update the host entries that are created by xCAT to make the fully qualified domain name the first entry, which is required by InfoSphere BigInsights, as shown in Example B-2.

Example B-2 UpdateHostsForBI

```
#!/bin/bash

# host entries created by xcat is :  IP_Address shortName  nameWithFQDN
shortName-netorkInterface
#     e.g.  10.10.0.101 server2-eth4 server2-eth4.data.ibm.com
# BigInsights required IP address to be resolved to name with FQDN
# update host entries to: IP_Address nameWithFQDN shortName
shortName-netorkInterface
#     e.g.  10.10.0.101 server2-eth4.data.ibm.com server2-eth4

WDIR=/tmp
FN=/etc/hosts

grep -v localhost ${FN} | grep -v pcmae > ${WDIR}/tmphosts.$$
grep -v pcmae ${FN} | grep localhost > ${WDIR}/hosts.$$
grep -v localhost ${FN} | grep pcmae >> ${WDIR}/hosts.$$
```

```
awk -F " " '{print $1 "   " $3 "   " $2 "   " $4 "   " $5 "  " }'
${WDIR}/tmphosts.$$ >> ${WDIR}/hosts.$$

mv -f ${WDIR}/hosts.$$ $FN
rm -f ${WDIR}/tmphosts.$$
```

3. Create the Platform Enterprise Grid Orchestrator (EGO) host file for a multi-homed host, as shown in Example B-3.

Example B-3   createEGOhosts

```
#!/bin/bash

#======= Global variables =======

PCM="pcmae"

DATE=`date +%Y%m%d-%H%M`

# Log File
export LOG_FILE="/tmp/log_createEGOhosts.$DATE"

#------------------------------------------------------------------
# Name: LOG
# Synopsis: LOG "$message"
# Description:
#       Record message into log file.
#------------------------------------------------------------------
LOG ()
{
        echo `date` "$1" | tee -a $LOG_FILE
#        echo `date`  "$1" >> $LOG_FILE
        sync
}

#==================================================================
#  MAIN
#==================================================================

  LOG "create EGO hosts for multi-homed server"

  FN=/opt/pcm/ego/kernel/conf/hosts

AWK=/tmp/awk.$$
cat << EOF > $AWK
#!/bin/awk -f
{
  print \$1 "  " "$PCM" "  "  \$2
}
EOF
  grep -v "^#" /etc/hosts | grep "${PCM}"  |  awk -f ${AWK} > ${FN}
cat << EOF > $AWK
#!/bin/awk -f
{
  print \$1 "  " "$EXECUTION_SERVER" "  "  \$2
}
```

```
EOF
  grep -v "^#" /etc/hosts | grep "${EXECUTION_SERVER}"  |  awk -f ${AWK} >> ${FN}

rm -f ${AWK}

source /opt/pcm/ego/kernel/conf/profile.ego
egosh user logon -u Admin -x Admin
egosh ego restart
```

The scripts from the console tier in the cluster template are run on the InfoSphere BigInsights console node. Complete the following steps:

1. Run the script that is shown in Example B-4 in the console node to install InfoSphere BigInsights in the cluster with one management node and N-number of data nodes.

Example B-4 installBigInsights_1mgt

```
#!/bin/bash
# installBigsights_1mgt:  install BigInsights on cluster configuration
#                         with 1 management node
#-------------------------------------------------------------------------------
#  System-defined variables used:
#   Console_eth0_NIC_HOSTNAMES     MgtNode_eth0_NIC_HOSTNAMES
DataNode_eth0_NIC_HOSTNAMES
#  User-defined variables used:
#   SOURCE_URL=http://192.168.0.31
#   BI_INPUT_VAR -  file to define
#          BI_PACKAGENAME    BI_FULLINSTALL_XML  DISKS   AdaptiveMR
#          CLUSTERADMGRP  CLUSTERADMGID
#          CLUSTERADMIN   CLUSTERADMID  CLUSTERADMPWD
#          BIGSQL   BIGSQLID   BIGSQLPWD

# Global variables

DATE=`date +%Y%m%d-%H%M`

# Log File
export LOG_FILE="/tmp/logall.BI_install.$DATE"
export LOGFILE="log.BI_install.$DATE"

# Directory
export DESTINATION_DIR="/tmp/dest.$$"
export BUILD_DIR="/tmp/build.$$"

export CONSOLE_HOSTNAME="${Console_eth0_NIC_HOSTNAMES}"
export MGT_HOSTNAMES="${MgtNode_eth0_NIC_HOSTNAMES}"
export DATA_HOSTNAMES="${DataNode_eth0_NIC_HOSTNAMES}"

#=================================================================
# Description
# Installation of BigInsights 3.0.0.1
#
#=================================================================

#------------------------------------------------------------------
# Name: LOG
# Synopsis: LOG "$message"
# Description:
#       Record message into log file.
#------------------------------------------------------------------
LOG ()
{
        echo `date` "$1" | tee -a $LOG_FILE
#        echo `date`  "$1" >> $LOG_FILE
        sync
}

#--------------------------------------------------------------------------
#
#  fetchPackage -
#     Download installation package from Management Server.
#
#---------------------------------------------------------------------------
fetchPackage()
{
    if [ $# != 2 ]; then
        LOG "Usage: fetchPackage <source URL path> <File to download>"
        exit 1
    fi

    # Set the URL path
    _url_path="$1"
    _package="$2"

    LOG "Downloading packages...${_package} "

    # fetch package
    logVar=`wget $_url_path/${_package} 2>&1`
    if [ "$?" != "0" ] ; then
        echo $logVar 1>&2
        LOG "Failed to fetch package < $_package > from $_url_path."
        return 1
    fi

    LOG "All packages were downloaded successfully."
    return 0
}

#-----------------------------------------
# Name: create_node_lists
# Description:
#    Create node lists
#-----------------------------------------
create_node_lists ()
{
  LOG " Create Cluster node lists "

  NODELIST_MASTER_eth0=`echo $CONSOLE_HOSTNAME $MGT_HOSTNAMES | sed "s/;/ /g"`
  NODELIST_SLAVES_eth0=`echo $DATA_HOSTNAMES | sed "s/;/ /g"`
  NODELIST_ALL_eth0=`echo "$NODELIST_MASTER_eth0 $NODELIST_SLAVES_eth0"`
  NODELIST_MASTER=`echo $NODELIST_MASTER_eth0 | sed "s/\.provision/-eth4\.data/g"`
  NODELIST_SLAVES=`echo $NODELIST_SLAVES_eth0 | sed "s/\.provision/-eth4\.data/g"`
  NODELIST_ALL=`echo $NODELIST_ALL_eth0| sed "s/\.provision/-eth4\.data/g"`

# Check number of master, current version only for 1 master
  tt=`echo $NODELIST_MASTER  | wc | awk '{print $2}' `
  if [ $tt != "1" ] ; then
    LOG "number of cluster masters ($tt) > 1 ; this script is for 1 master "
    exit 99
  fi

  export MASTER_IP_ADDRESS=`/usr/bin/host $NODELIST_MASTER | awk '{print $4}' `
  LOG " Master hostname <$NODELIST_MASTER> IP <$MASTER_IP_ADDRESS> "

  LOG " NODELIST_ALL ($NODELIST_ALL) "
  LOG " NODELIST_MASTER ($NODELIST_MASTER) "
  LOG " NODELIST_SLAVES ($NODELIST_SLAVES) "

}

#---------------------------------------
# Name: clear_disk_sectors
# Description:
#  Clear at least 2 sectors of the disk, just in case NSD volume ID left from 
previous GPFS instllation
#---------------------------------------

clear_disk_sectors()
{

 LOG " Clear disk sectors on Data nodes"

# Generate script to clear disk sectors
  if [[ -f /tmp/clear_disk_sectors.sh ]] ; then
    /bin/rm /tmp/clear_disk_sectors.sh
  fi

cat << EOF >> /tmp/clear_disk_sectors.sh
  for DEV in ${DISKS[@]}; do
     /bin/dd if=/dev/zero of=\$DEV obs=512 count=2000
  done
EOF
  /bin/chmod a+rx /tmp/clear_disk_sectors.sh

# Run clear_disk_sectors.sh on DATA nodes
  mymaster=`hostname -s`
  for NODE in ${NODELIST_SLAVES[@]}; do
     if [ "`/usr/bin/ssh $NODE hostname -s`" != "$mymaster" ] ;  then
       /usr/bin/scp -p /tmp/clear_disk_sectors.sh 
$NODE:/tmp/clear_disk_sectors.sh
       /usr/bin/ssh $NODE "/tmp/clear_disk_sectors.sh"
     fi
   done
}
```

```
#----------------------------------------
#
# updateConfigFile
#   Update node and GPFS disk entry in config file
#
#---------------------------------------
updateConfigFile()
{
   LOG " Update Master  node in config file "
   for NODE in $NODELIST_MASTER ; do
     sed -i -e "s|__MGMT__|${NODE}|" $CONFILE_NEW
   done

   LOG " Add Data node in config file "
   for NODE in $NODELIST_SLAVES ; do
## includes rack
##    cat $CONFILE_NEW | sed -e "s:-- NODELIST-TOKEN -->:-- NODELIST-TOKEN -->\n
<node>\n            <name-or-ip>$NODE</name-or-ip>\n
<password>{xor}</password>\n             <rack>$rack_number</rack>\n
<hdfs-data-directory/>\n            <gpfs-node-designation/>\n
<gpfs-admin-node/>\n        <gpfs-rawdisk-list>$DISKS</gpfs-rawdisk-list>\n
<gpfs-datapool-disk-list/>\n            <bigsql-data-directory/>\n
<node-type>private</node-type>\n            </node>:g" > $CONFILE
    cat $CONFILE_NEW | sed -e "s:-- NODELIST-TOKEN -->:-- NODELIST-TOKEN -->\n
<node>\n            <name-or-ip>$NODE</name-or-ip>\n
<password>{xor}</password>\n             <rack/>\n
<hdfs-data-directory/>\n            <gpfs-node-designation/>\n
<gpfs-admin-node/>\n        <gpfs-rawdisk-list>$DISKS</gpfs-rawdisk-list>\n
<gpfs-datapool-disk-list/>\n            <bigsql-data-directory/>\n
<node-type>private</node-type>\n            </node>:g" > $CONFILE
            if [ -s $CONFILE ]; then
                /bin/cp -f $CONFILE $CONFILE_NEW
                LOG "INFO:  Added:  $NODE"
            else
                LOG "ERROR:  Failed to update nodelist with node:  $NODE"
                exit 99
            fi
        done
}

#------------------------------------------------------------------
# Name: create_user
# Synopsis: create_user
# Description:
#       Create BigInsigths admin, group, ssh key and sudo
#------------------------------------------------------------------
create_user ()
{
   mygroup=$1
   mygid=$2
   myuser=$3
   myuid=$4
   mypasswd=$5
   mymaster=`hostname -s`
```

```
#   LOG " Create BI user:  group < $mygroup > user  < $myuser > "

# Generate script to create group and user
   if [[ -f /tmp/create_user.sh ]] ; then
      /bin/rm /tmp/create_user.sh
   fi

cat << EOF >> /tmp/create_user.sh
   /usr/bin/id $myuser
   if [ \$? != 0 ] ; then
     /usr/bin/getent group | grep -q $mygroup
     if [ \$? != 0 ] ; then
       /usr/sbin/groupadd -g $mygid $mygroup
     fi
     /usr/sbin/useradd -m -d /home/$myuser -g $mygid -u $myuid $myuser
     /bin/echo -e "$myuser:$mypasswd" | /usr/sbin/chpasswd
   fi
EOF
   /bin/chmod a+rx /tmp/create_user.sh

   for NODE in ${NODELIST_ALL[@]}; do
     if [ "`/usr/bin/ssh $NODE hostname -s`" != "$mymaster" ] ;  then
       scp -p /tmp/create_user.sh  $NODE:/tmp/create_user.sh
     fi
   done

   if [[ -f /tmp/create_user_key.sh ]] ; then
      /bin/rm /tmp/create_user_key.sh
   fi

# Generate script to create ssh keys
cat << EOF >> /tmp/create_user_key.sh
   /usr/bin/id $myuser
   if [ \$? == 0 ] ; then
     cd /home/$myuser
     su - $myuser  -c "/bin/rm -rf ~/.ssh"
     su - $myuser  -c "/usr/bin/ssh-keygen -t rsa -f ~/.ssh/id_rsa -N \"\" "
     su - $myuser  -c "/bin/cp .ssh/id_rsa.pub .ssh/authorized_keys"
     su - $myuser  -c "/bin/chmod 600 .ssh/id_rsa* "
     su - $myuser  -c "/bin/chmod 600 .ssh/authorized_keys "
     echo "LogLevel QUIET" > /tmp/mysshconfig
     echo "StrictHostKeyChecking no" > /tmp/mysshconfig
     chmod a+rw /tmp/mysshconfig
     su - $myuser  -c "/bin/cp /tmp/mysshconfig .ssh/config"
     su - $myuser  -c "/bin/chmod 644 .ssh/config"
     rm  /tmp/mysshconfig
   fi
EOF
   /bin/chmod a+x /tmp/create_user_key.sh

# Create group and user
   /tmp/create_user.sh
   /tmp/create_user_key.sh

   for NODE in ${NODELIST_ALL[@]}; do
```

```
     if [ "`/usr/bin/ssh $NODE hostname -s`" != "$mymaster" ] ;  then
       LOG " Create $myuser on node < $NODE > "
       /usr/bin/ssh $NODE "/tmp/create_user.sh"
       LOG  " Copy ssh key from master on node < $NODE >  "
       /usr/bin/scp -pr /home/$myuser/.ssh $NODE:/home/$myuser/
       /usr/bin/ssh $NODE "chown -R $myuser.$mygroup /home/$myuser"
     fi
   done
  /bin/rm  -f  /tmp/create_user_key.sh

   for NODE in ${NODELIST_ALL[@]}; do
       /usr/bin/ssh $NODE "/bin/rm -f /tmp/create_user.sh"
   done
  /bin/rm -f  /tmp/create_user_key.sh

}

#----------------------------------------
#
# runInstallation()
#   Run Installation
#
#---------------------------------------
runInstallation()
{
    LOG " Install BigInsigths 3.0.0.1 in silent mode "

    sed -i -e "/AdaptiveMR.Enable/ s/=.*$/=$AdaptiveMR/"
./$BIinstall/install.properties

    # Run the silent install
    LOG " ./$BIinstall/silent-install/silent-install.sh $1 "
    ./$BIinstall/silent-install/silent-install.sh $1  2>&1 | tee -a $LOGFILE

    if [ "$?" != "0" ] ; then
        LOG "ERROR: BigInsights 3.0.0.1 installation returned $?."
        exit
    fi
}

#===================================================================
#  MAIN
#===================================================================


    if [ "$PCM_LAYER_ACTION" == "CREATE" ] ; then
      LOG "Current action is CREATE"
    fi

    # fetch BigInsights packages

    if [ -d $DESTINATION_DIR ]; then
        rm -rf $DESTINATION_DIR
    fi
    mkdir -p  $DESTINATION_DIR
    cd $DESTINATION_DIR

    fetchPackage "$SOURCE_URL" "$BI_INPUT_VAR"
    source $BI_INPUT_VAR

    fetchPackage "$SOURCE_URL" "$BI_FULLINSTALL_XML"
    fetchPackage "$SOURCE_URL" "$BI_PACKAGENAME"

    if [ ! -d $BUILD_DIR ]; then
      mkdir -p $BUILD_DIR
    fi
##    untarPackage "../$BI_PACKAGENAME" "$BUILD_DIR"
   untarPackage "$BI_PACKAGENAME" "$BUILD_DIR"

    cp -p $BI_FULLINSTALL_XML fullinstall.xml.new
    CONFILE="fullinstall.xml"
    CONFILE_NEW="fullinstall.xml.new"
    create_node_lists

# synchronize clock update
    clockmaster=`grep -v "^#" /etc/ntp.conf | grep server | head -1 |  awk '{print 
$2}'`
    for NODE in ${NODELIST_ALL[@]} ; do
      /usr/bin/ssh $NODE "hostname; /usr/bin/ssh `hostname` \"hostname; 
/sbin/service ntpd stop; /usr/sbin/ntpdate -s $clockmaster ; /sbin/service ntpd 
start\" "
    done

    updateConfigFile

##  Clear disk sectors on data nodes
##  clear_disk_sectors
    create_user "$CLUSTERADMGRP" "$CLUSTERADMGID" "$CLUSTERADMIN" "$CLUSTERADMID" 
"$CLUSTERADMPWD"
    create_user "$CLUSTERADMGRP" "$CLUSTERADMGID" "$BIGSQL" "$BIGSQLID" 
"$BIGSQLPWD"

    mv -f fullinstall.xml $BUILD_DIR/

    cd $BUILD_DIR
    export BIinstall=`ls -d biginsights-3.0*SNAPSHOT-enterprise-*`
    runInstallation $BUILD_DIR/fullinstall.xml
```

2. Add a node to the InfoSphere BigInsights cluster by using the `addnode.sh` script, as shown in Example B-5.

Example B-5 Console_addnode

```
#!/bin/bash

#======= Global variables =======

DATE=`date +%Y%m%d-%H%M`

# Log File
```

```
export LOG_FILE="/tmp/log_addnode.$DATE"

# Directory
export DESTINATION_DIR="/tmp/dest.$$"

#------------------------------------------------------------------
# Name: LOG
# Synopsis: LOG "$message"
# Description:
#       Record message into log file.
#------------------------------------------------------------------
LOG ()
{
        echo `date` "$1" | tee -a $LOG_FILE
#        echo `date`  "$1" >> $LOG_FILE
        sync
}
#--------------------------------------------------------------------------
#
#  fetchPackage -
#     Download installation package from Management Server.
#
#---------------------------------------------------------------------------
fetchPackage()
{
    if [ $# != 2 ]; then
        LOG "Usage: fetchPackage <source URL path> <File to download>"
        exit 1
    fi

    # Set the URL path
    _url_path="$1"
    _package="$2"

    LOG "Downloading packages...${_package} "

    # fetch package
    logVar=`wget $_url_path/${_package} 2>&1`
    if [ "$?" != "0" ] ; then
        echo $logVar 1>&2
        LOG "Failed to fetch package < $_package > from $_url_path."
        return 1
    fi

    LOG "All packages were downloaded successfully."
    return 0
}

#------------------------------------------------------------------
# Name: addnode_create_user
# Synopsis: addnode_create_user
# Description:
#       Create BigInsigths admin, group, ssh key and sudo
#------------------------------------------------------------------
addnode_create_user ()
```

```
{
   mygroup=$1
   mygid=$2
   myuser=$3
   myuid=$4
   mypasswd=$5
   mymaster=`hostname -s`

#   LOG " Create BI user:  group < $mygroup > user  < $myuser > "

# Generate script to create group and user
   if [[ -f /tmp/addnode_create_user.sh ]] ; then
      /bin/rm /tmp/addnode_create_user.sh
   fi

cat << EOF >> /tmp/addnode_create_user.sh
   /usr/bin/id $myuser
   if [ \$? != 0 ] ; then
     /usr/bin/getent group | grep -q $mygroup
     if [ \$? != 0 ] ; then
       /usr/sbin/groupadd -g $mygid $mygroup
     fi
     /usr/sbin/useradd -m -d /home/$myuser -g $mygid -u $myuid $myuser
     /bin/echo -e "$myuser:$mypasswd" | /usr/sbin/chpasswd
   fi
EOF
   /bin/chmod a+rx /tmp/addnode_create_user.sh

   for NODE in ${addnode_NODELIST[@]}; do
     scp -p /tmp/addnode_create_user.sh  $NODE:/tmp/addnode_create_user.sh
   done

# Create group and user remote

   for NODE in ${addnode_NODELIST[@]}; do
       LOG " Create $myuser on node < $NODE > "
       /usr/bin/ssh $NODE "/tmp/addnode_create_user.sh"
       LOG  " Copy ssh key from master on node < $NODE >  "
       /usr/bin/scp -pr /home/$myuser/.ssh $NODE:/home/$myuser/
       /usr/bin/ssh $NODE "chown -R $myuser.$mygroup /home/$myuser"
   done

   for NODE in ${addnode_NODELIST[@]}; do
       /usr/bin/ssh $NODE "/bin/rm -f /tmp/addnode_create_user.sh"
   done
  /bin/rm -f  /tmp/addnode_create_user.sh

}

#===================================================================
#  MAIN
#===================================================================

#=== Capture environment variables ==
/bin/env
```

```
if [ "$RUN_ADDNODE" == "Y" ] ; then

    if [ "$PCM_LAYER_ACTION" == "ADDSERVERS" ] ; then
      LOG "Current action is $PCM_LAYER_ACTION on `hostname`"
   else
      LOG "Current action is $PCM_LAYER_ACTION.  Only run addnode.sh for 
ADDSERVERS.    Exiting"
     exit 99
    fi

    # fetch BigInsights variable settings

    if [ -d $DESTINATION_DIR ]; then
        rm -rf $DESTINATION_DIR
    fi
    mkdir -p  $DESTINATION_DIR
    chmod 777 $DESTINATION_DIR
    cd $DESTINATION_DIR

    fetchPackage "$SOURCE_URL" "$BI_INPUT_VAR"
    source $BI_INPUT_VAR

    TMPLOG=${DESTINATION_DIR}/log_addnode.`hostname -s`.$$
    TMP=${DESTINATION_DIR}/sh_addnode.`hostname -s`.$$

mypassword=$ROOT
mydisks=`echo $DISKS | sed "s/ /:/g"`
addlist=`echo $DataNode_OP_SERVER_HOSTNAME_LIST | sed "s/;/ /g"`
addhostlist=""
for N in ${addlist[@]}
do
  addhostlist=`echo -n ${addhostlist} \`ssh $N hostname\``
done
addhostlist=`echo ${addhostlist}`

datanode_ADDNODES=${addhostlist}
datanode_ADDNODES_LIST=`echo $datanode_ADDNODES | sed "s/ /;/g"`

##-- create user biadmin and bigsql
addnode_NODELIST=$datanode_ADDNODES
    addnode_create_user "$CLUSTERADMGRP" "$CLUSTERADMGID" "$CLUSTERADMIN" 
"$CLUSTERADMID" "$CLUSTERADMPWD"
    addnode_create_user "$CLUSTERADMGRP" "$CLUSTERADMGID" "$BIGSQL" "$BIGSQLID" 
"$BIGSQLPWD"
##--

CMDgpfs="./addnode.sh gpfs"
for N in ${datanode_ADDNODES[@]} ; do
 CMDgpfs=`echo -n ${CMDgpfs} ${N},$mypassword,,$mydisks `
done
CMDgpfs=`echo ${CMDgpfs}`

CMDhadoop="./addnode.sh hadoop $datanode_ADDNODES"
CMDbigsql="./addnode.sh bigsql $datanode_ADDNODES"
```

```
cat << EOF > $TMP
#! /bin/bash
if [ ! -x /opt/ibm/biginsights/conf/biginsights-env.sh ] ; then
  echo "\$0: Exiting..problem with BigInsights environment"
  exit
fi
source /opt/ibm/biginsights/conf/biginsights-env.sh
cd \${BIGINSIGHTS_HOME}/bin
${CMDgpfs}
${CMDhadoop}
${CMDbigsql}
./status.sh
EOF
chmod a+x $TMP

myhostname=`hostname`
clockmaster=`grep -v "^#" /etc/ntp.conf | grep server | head -1 |  awk '{print 
$2}'`
  for N in ${datanode_ADDNODES[@]} ; do
   /usr/bin/ssh ${N} "/bin/hostname; /usr/bin/ssh $myhostname \"/bin/hostname\""
  done
  for N in ${datanode_ADDNODES[@]} ; do
   /usr/bin/ssh ${N} "/sbin/service ntpd stop; /usr/sbin/ntpdate -s $clockmaster ; 
/sbin/service ntpd start"
  done

  /bin/su -c $TMP biadmin  1> ${TMPLOG}  2>&1
  sleep 10

#     cd ; rm -rf $DESTINATION_DIR

   LOG "Console:  `hostname -s`    `date +%Y%m%d-%H%M`"

fi
```

The following scripts (Example B-6 and Example B-7 on page 202) are run on the data nodes to create the InfoSphere BigInsights cluster:

To clear at least two sectors of the disk to create the Spectrum Scale Network Shared Disk (NDS), run the script that is shown in Example B-6.

Example B-6   DataNode_clearDiskSectors

```
#!/bin/bash

#======= Global variables =======

DATE=`date +%Y%m%d-%H%M`

# Log File
export LOG_FILE="/tmp/log_clearDiskSectors.$DATE"

# Directory
export DESTINATION_DIR="/tmp/dest.$$"
```

```
#------------------------------------------------------------------
# Name: LOG
# Synopsis: LOG "$message"
# Description:
#       Record message into log file.
#------------------------------------------------------------------
LOG ()
{
        echo `date` "$1" | tee -a $LOG_FILE
#        echo `date`  "$1" >> $LOG_FILE
        sync
}
#--------------------------------------------------------------------------
#
#  fetchPackage -
#     Download installation package from Management Server.
#
#---------------------------------------------------------------------------
fetchPackage()
{
    if [ $# != 2 ]; then
        LOG "Usage: fetchPackage <source URL path> <File to download>"
        exit 1
    fi

    # Set the URL path
    _url_path="$1"
    _package="$2"

    LOG "Downloading packages...${_package} "

    # fetch package
    logVar=`wget $_url_path/${_package} 2>&1`
    if [ "$?" != "0" ] ; then
        echo $logVar 1>&2
        LOG "Failed to fetch package < $_package > from $_url_path."
        return 1
    fi

    LOG "All packages were downloaded successfully."
    return 0
}
#----------------------------------------
# Name: DataNode_clear_disk_sectors
# Description:
#  Clear at least 2 sectors of the disk, just in case NSD volume ID left from
previous GPFS instllation
#----------------------------------------
DataNode_clear_disk_sectors ()
{

 LOG " Clear disk sectors on Data nodes"

  for DEV in ${DISKS[@]}; do
     /bin/dd if=/dev/zero of=$DEV obs=512 count=2000
  done
}

#==================================================================
#  MAIN
#==================================================================

    if [ "$PCM_LAYER_ACTION" == "CREATE" ] ; then
      LOG "Current action is $PCM_LAYER_ACTION"
   else
      LOG "Current action is $PCM_LAYER_ACTION..  exiting"
     exit 99
    fi

    # fetch BigInsights variable settings

    if [ -d $DESTINATION_DIR ]; then
        rm -rf $DESTINATION_DIR
    fi
    mkdir -p  $DESTINATION_DIR
    cd $DESTINATION_DIR

    fetchPackage "$SOURCE_URL" "$BI_INPUT_VAR"
    source $BI_INPUT_VAR


    rpm -qa | grep -q gpfs
    if [ $? -ne 0 ] ; then
     echo "Clear disk sectors only if GPFS is not installed -- when creating and 
deleting data nodes"
      DataNode_clear_disk_sectors
    fi

    cd ; rm -rf $DESTINATION_DIR

   LOG  "DataNode:  `hostname -s`    `date +%Y%m%d-%H%M`"
```

An example of the postscript to ensure synchronization (`status<nodetype>`) can be found in Example B-7.

Example B-7   statusConsole

```
#/bin/bash

sleep 10
echo "Console:  `hostname -s`    `date +%Y%m%d-%H%M`"
```

Postscripts that are defined in the image profile

The InfoSphere BigInsights installation instructions include a list of tasks to be completed before you start the installation. The InfoSphere BigInsights installation runs the preinstallation utility **bi-prechecker.sh** to ensure that the operating system environment is ready. Tasks are completed in the postscript that runs after the completion of the diskfull installation of the operating system and before the restart.

The postscript is added to the definition of the image profile by completing the following steps:

1. Install Mellanox OFED 2.3-1.0.1, which is the version for the operating system RHEL 6.5, as shown in Example B-8.

Example B-8 mlnxofed_2.3-1.0.1-install

```
#!/bin/bash
# add to postscripts that should be run after diskfull installation

logger -t mlnx_ofed  -p local4.info  "started $0 on `hostname`"

# pkglist needed for MLNX_OFED_LINUX are added in pkglist.cfm
#INCLUDE: /opt/xcat/share/xcat/ib/netboot/rh/ib.rhels6.ppc64.pkglist
#pciutils-libs
#pciutils
#tcl
#tk
#tcsh
#libgcc.ppc
#gcc-gfortran

#yum -y install pciutils-libs pciutils tcl tk tcsh libgcc.ppc gcc-gfortran

if [ -z ${installroot} ] ; then
  installroot=/mnt
fi

ofeddir=${installroot}/otherpkgs/mellanox
ofediso=${ofeddir}/MLNX_OFED_LINUX-2.3-1.0.1-rhel6.5-ppc64.iso

mkdir -p /tmp/ofed
/bin/mount -o loop ${ofediso} /tmp/ofed

cd /tmp/ofed
echo "y" | perl -x ./mlnxofedinstall --without-32bit --force

service openibd restart
ANS=`/usr/bin/ibv_devinfo | grep -B1  PORT_ACTIVE  | grep "port:"`

if [ "$ANS" == "0" ] ; then
 echo "`hostname`: openibd started successfully"
else
 echo "`hostname`:  openibd failed to start"
 exit 1
fi

cd
```

```
/bin/umount /tmp/ofed
rm -f /tmp/ofed
```

2. Update the operating system prerequisites, as shown in Example B-9:
   a. SELinux must be disabled.
   b. Increase the system default `ulimit` of `nofile` (maximum number of open files) and `nproc` (maximum number of processes).
   c. Update the `umask` of root to `022`.

Example B-9   system_prereq

```
#!/bin/bash

logger -t BIpost  -p local4.info  "started $0 on `hostname`"

#start of test
if [ 1 == 1 ] ; then
#-- SELinux must be disabled (permissive mode is not sufficient)

/usr/sbin/sestatus  | grep -q disabled
RC=$?
if [ $RC -ne 0 ] ; then
  echo "SELinux must be disabled. Editing /etc/selinux/config prior rebooting "
  sed -i -e "/SELINUX=/ s/=.*$/=disabled/" /etc/selinux/config
#  cat /etc/selinux/config
fi

#Not to export home from PCM: tabch key=pcm_export_home site.value=False

#-- Update nproc in /etc/security/limits.d/90-nproc.conf and
/etc/security/limits.conf
#-- Update nofile in /etc/security/limits.conf

maxnproc=65536
maxnofile=65536
NPROC=`ulimit -u`
NOFILE=`ulimit -n`

#-- Edit nproc in /etc/security/limits.d/90-nproc.conf
WDIR=/tmp
FN=/etc/security/limits.d/90-nproc.conf
mkdir -p ${WDIR}

grep -v ^# ${FN}  | grep -q nproc
RC=$?
if [ $RC -eq 0 ] ; then
   sed -i -e "/^#/! s/ nproc.*/ nproc   $maxnproc/" ${FN}

else

cat << EOF > ${WDIR}/tmp_add.$$
*   soft  nproc  $maxnproc
*   hard  nproc  $maxnproc
root   soft  nproc  503689
root   hard  nproc  503689
```

```
biadmin  soft  nproc  $maxnproc
biadmin  hard  nproc  $maxnproc
@biadmin  soft  nproc  $maxnproc
@biadmin  hard  nproc  $maxnproc
EOF

cat ${FN} ${WDIR}/tmp_add.$$ > $WDIR/tmp_file.$$
mv -f ${WDIR}/tmp_file.$$ ${FN}
rm -f ${WDIR}/tmp_add.$$
fi

#-- Edit nproc and nofile limits in /etc/limits.conf
WDIR=/tmp
FN=/etc/security/limits.conf

grep -v ^# ${FN}  | grep -q nproc
RC=$?
if [ $RC -eq 0 ] ; then
   sed -i -e "/^#/! s/ nproc.*/ nproc   $maxnproc/" ${FN}

else

cat << EOF > ${WDIR}/tmp_add.$$
*   soft  nproc  $maxnproc
*   hard  nproc  $maxnproc
root   soft  nproc  503689
root   hard  nproc  503689
biadmin  soft  nproc  $maxnproc
biadmin  hard  nproc  $maxnproc
@biadmin  soft  nproc  $maxnproc
@biadmin  hard  nproc  $maxnproc
EOF

cat ${FN} ${WDIR}/tmp_add.$$ > $WDIR/tmp_file.$$
mv -f ${WDIR}/tmp_file.$$ ${FN}
rm -f ${WDIR}/tmp_add.$$

fi


grep -v ^# ${FN} | grep -q nofile
RC=$?
if [ $RC -eq 0 ] ; then
   sed -i -e "/^#/! s/ nofile.*/ nofile   $maxnofile/" ${FN}

else

cat << EOF > ${WDIR}/tmp_add.$$
*   soft  nofile  $maxnofile
*   hard  nofile  $maxnofile
root   soft  nofile  $maxnofile
root   hard  nofile  $maxnofile
biadmin  soft  nofile  $maxnofile
biadmin  hard  nofile  $maxnofile
@biadmin  soft  nofile  $maxnofile
```

```
@biadmin  hard  nofile  $maxnofile
EOF

cat ${FN} ${WDIR}/tmp_add.$$ > $WDIR/tmp_file.$$
mv -f ${WDIR}/tmp_file.$$ ${FN}
rm -f ${WDIR}/tmp_add.$$

fi

#-- Set umask 022 for root

FN=/root/.bashrc
grep -v ^# ${FN} | grep -q umask
RC=$?
if [ $RC -eq 0 ] ; then
   sed -i -e "/umask/ s/^.*$/umask 022/" ${FN}
else
   echo "# BigInsight prerequisites" >> ${FN}
   echo "umask 022"  >> ${FN}
fi
echo "---${FN}----"
cat ${FN}

#end of test
fi
```

3. Disable IPV6, as shown in Example B-10.

Example B-10   disable_ipv6

```
#!/bin/bash

FN=/etc/sysctl.conf
grep -v ^# ${FN} | grep -q kernel.pid_max
RC=$?
if [ $RC -eq 0 ] ; then
   RESULT=`grep = ${FN} | grep kernel.pid_max | cut -d= -f2 | tr -d " " `
   if [ ${RESULT} -ne 4194303 ] ; then
     sed -i -e "/kernel.pid_max/ s/^.*$/kernel.pid_max = 4194303/" ${FN}
   fi
else
## NOTE: cannot include any comment with kernel.pid_max, net.ipv4.ip_local_port_
range or else the bi-prechecker.sh script will fail: grep did not parse comments
"
   echo "" >> ${FN}
##   echo "# BigInsights prerequisite:  kernel.pid_max = 4194303" >> ${FN}
##   echo "#                            net.ipv4.ip_local_port_range = 1024  640
00 >> ${FN}
   echo "# Add BigInsights prerequisites" >> ${FN}
   echo "kernel.pid_max = 4194303" >> ${FN}
   echo "net.ipv4.ip_local_port_range = 1024  64000" >> ${FN}
fi
```

4. Disable the firewall in the system, as shown in Example B-11 on page 207.

Example B-11 disable_fw

```
#!/bin/bash

##### this script disables the fw on the system

service iptables save
service iptables stop
chkconfig iptables off

exit 0
```

5. Disable *tty* for **sudo** users, as shown in Example B-12.

Example B-12 disable_sudoer_tty

```
#!/bin/bash

FN=/etc/sudoers
grep -v ^# ${FN}  | grep -q " requiretty"
RC=$?
if [ $RC -eq 0 ] ; then
   sed -i -e "/^.*Defaults.*requiretty/ s/ requiretty/ !requiretty/" ${FN}
fi
```

6. Update the system kernel parameters, as shown in Example B-13.

Example B-13 setup_sysctl-conf

```
#!/bin/bash

FN=/etc/sysctl.conf
grep -v ^# ${FN} | grep -q kernel.pid_max
RC=$?
if [ $RC -eq 0 ] ; then
   RESULT=`grep = ${FN} | grep kernel.pid_max | cut -d= -f2 | tr -d " " `
   if [ ${RESULT} -ne 4194303 ] ; then
     sed -i -e "/kernel.pid_max/ s/^.*$/kernel.pid_max = 4194303/" ${FN}
   fi
else
## NOTE: cannot include any comment with kernel.pid_max,
net.ipv4.ip_local_port_range or else the bi-prechecker.sh script will fail: grep
did not parse comments"
   echo "" >> ${FN}
##   echo "# BigInsights prerequisite:  kernel.pid_max = 4194303" >> ${FN}
##   echo "#                            net.ipv4.ip_local_port_range = 1024  64000
>> ${FN}
   echo "# Add BigInsights prerequisites" >> ${FN}
   echo "kernel.pid_max = 4194303" >> ${FN}
   echo "net.ipv4.ip_local_port_range = 1024  64000" >> ${FN}
fi
```

C

BigData Enablement and Administration Toolkit introduction

This appendix describes the BigData Enablement and Administration Toolkit (BEAT), which is an IBM toolkit to help enable a big data solution on IBM Power Systems.

This appendix introduces the functions of the toolkit and shows the key windows that are used to set up an IBM BigData solution by using a web-based GUI wizard.

This appendix covers the following topics:

- BigData Enablement and Administration Toolkit overview
- Big data solution deployment with BigData Enablement and Administration Toolkit

BigData Enablement and Administration Toolkit overview

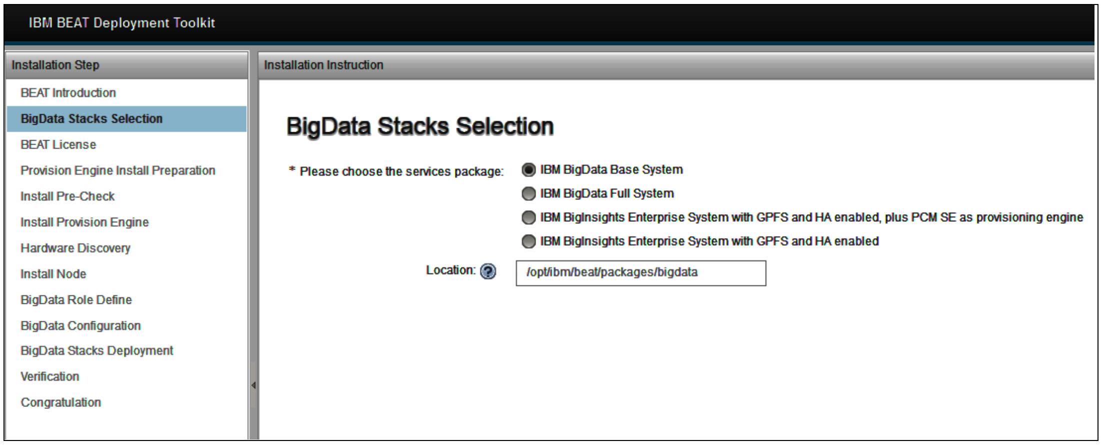
BEAT is targeted at administrators to help them quickly enable a big data solution in their own environment. The tool integrates the Extreme Cloud Administration Toolkit (xCAT) to deploy and configure a cluster of IBM Power Systems servers running IBM InfoSphere BigInsights, IBM Spectrum Scale, IBM Platform Symphony, and open source HBase and Hive.

BEAT provides the following functions:

- Deployment foundation stacks:
  - Quickly enable big data on POWER.
  - Simple deployment steps with predefined templates.
  - A deployment configuration adviser.
  - Node discovery.
- Big data stack integration:
  - Big data stack auditing.
  - Big data stack monitoring and management in a unified web portal.
  - Troubleshooting.

Big data solution deployment with BigData Enablement and Administration Toolkit

A step-by-step web-based GUI wizard helps with the entire deployment process. Figure C-1 shows the BigData Stacks Selection menu.



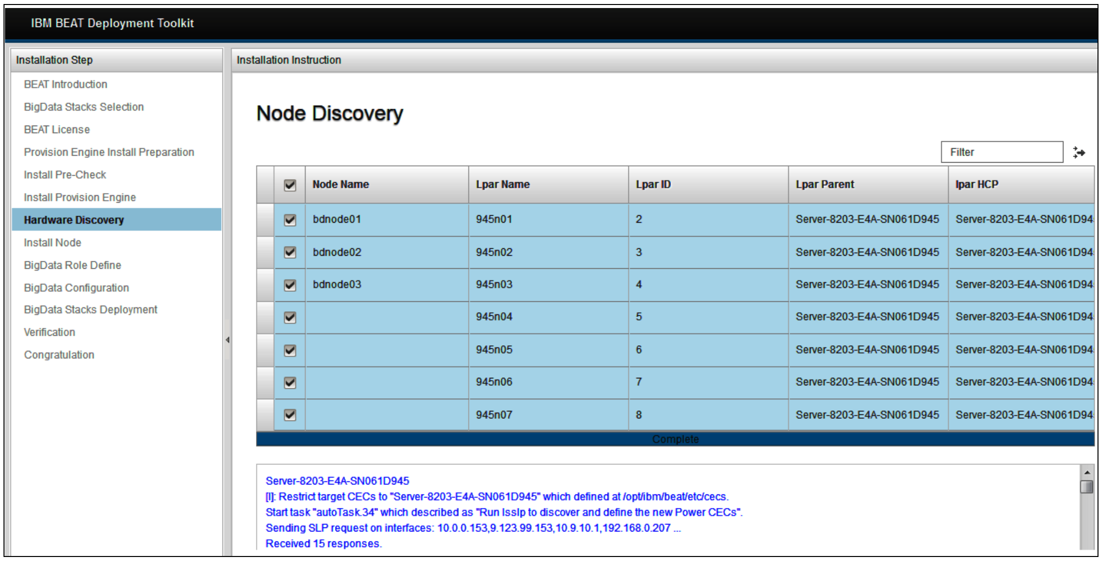
BigData Stacks Selection

* Please choose the services package:
● IBM BigData Base System
○ IBM BigData Full System
○ IBM BigInsights Enterprise System with GPFS and HA enabled, plus PCM SE as provisioning engine
○ IBM BigInsights Enterprise System with GPFS and HA enabled

Location: /opt/ibm/beat/packages/bigdata

Figure C-1 BigData Stacks Selection window

Figure C-2 on page 211 shows the node discovery window.



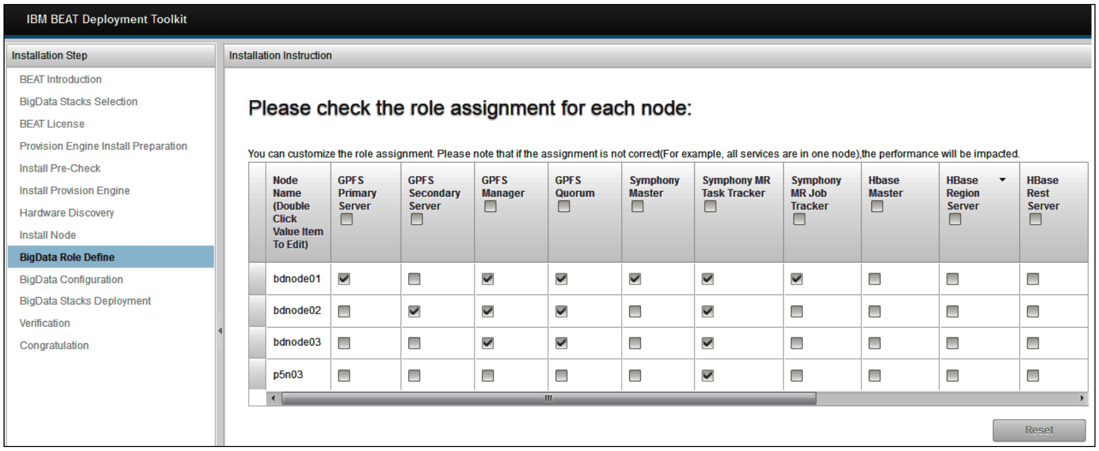
Node Discovery

☑	Node Name	Lpar Name	Lpar ID	Lpar Parent	Ipar HCP
☑	bdnode01	945n01	2	Server-8203-E4A-SN061D945	Server-8203-E4A-SN061D94
☑	bdnode02	945n02	3	Server-8203-E4A-SN061D945	Server-8203-E4A-SN061D94
☑	bdnode03	945n03	4	Server-8203-E4A-SN061D945	Server-8203-E4A-SN061D94
☑		945n04	5	Server-8203-E4A-SN061D945	Server-8203-E4A-SN061D94
☑		945n05	6	Server-8203-E4A-SN061D945	Server-8203-E4A-SN061D94
☑		945n06	7	Server-8203-E4A-SN061D945	Server-8203-E4A-SN061D94
☑		945n07	8	Server-8203-E4A-SN061D945	Server-8203-E4A-SN061D94

```
Server-8203-E4A-SN061D945
[I]: Restrict target CECs to "Server-8203-E4A-SN061D945" which defined at /opt/ibm/beat/etc/cecs.
Start task "autoTask.34" which described as "Run lsslp to discover and define the new Power CECs".
Sending SLP request on interfaces: 10.0.0.153,9.123.99.153,10.9.10.1,192.168.0.207 ...
Received 15 responses.
```

Figure C-2 Node discovery

Figure C-3 shows the role assignment selection window for each node.



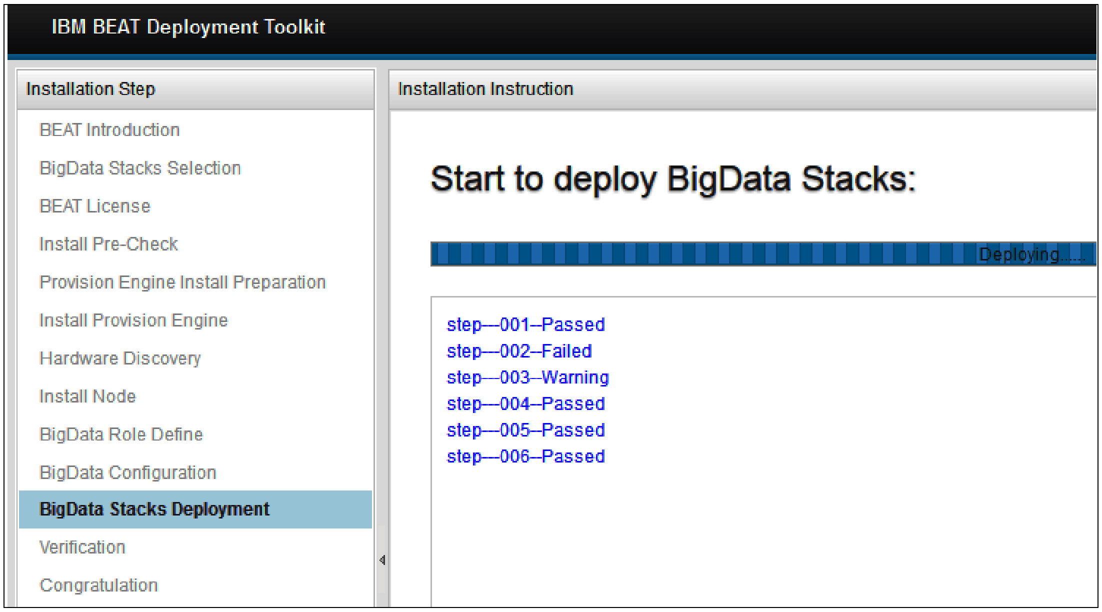
Please check the role assignment for each node:

You can customize the role assignment. Please note that if the assignment is not correct(For example, all services are in one node),the performance will be impacted.

Node Name (Double Click Value Item To Edit)	GPFS Primary Server ☐	GPFS Secondary Server ☐	GPFS Manager ☐	GPFS Quorum ☐	Symphony Master ☐	Symphony MR Task Tracker ☐	Symphony MR Job Tracker ☐	Hbase Master ☐	HBase Region Server ☐	HBase Rest Server ☐
bdnode01	☑	☐	☑	☑	☑	☑	☑	☐	☐	☐
bdnode02	☐	☑	☑	☑	☐	☑	☐	☐	☐	☐
bdnode03	☐	☐	☑	☑	☐	☑	☐	☐	☐	☐
p5n03	☐	☐	☐	☐	☐	☑	☐	☐	☐	☐

Figure C-3 Role assignment for each node

Figure C-4 shows the installation of BigData stacks.

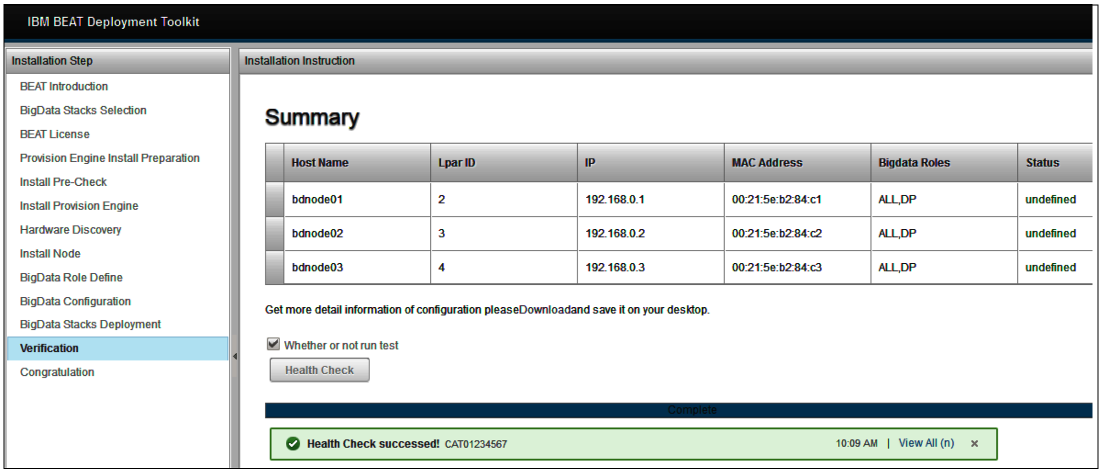


Start to deploy BigData Stacks:

step—001–Passed
step—002–Failed
step—003–Warning
step—004–Passed
step—005–Passed
step—006–Passed

Figure C-4 Installation of BigData stacks

Figure C-5 shows the installation check menu summary.



Summary

Host Name	Lpar ID	IP	MAC Address	Bigdata Roles	Status
bdnode01	2	192.168.0.1	00:21:5e:b2:84:c1	ALL,DP	undefined
bdnode02	3	192.168.0.2	00:21:5e:b2:84:c2	ALL,DP	undefined
bdnode03	4	192.168.0.3	00:21:5e:b2:84:c3	ALL,DP	undefined

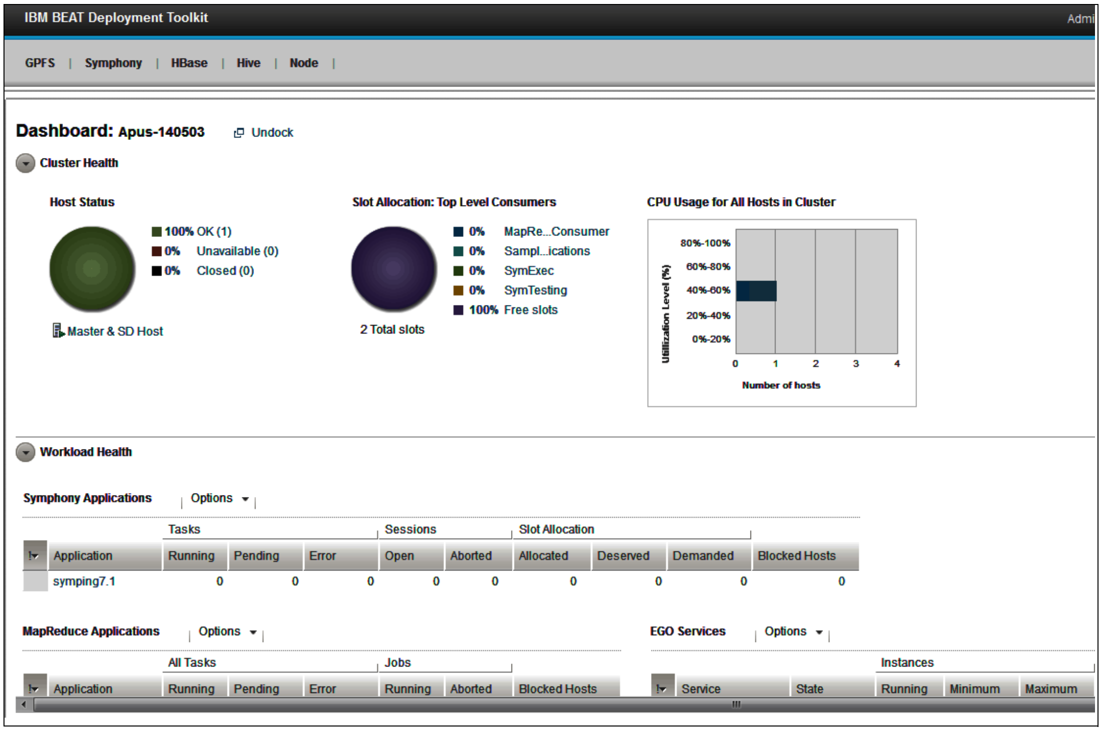
Get more detail information of configuration pleaseDownloadand save it on your desktop.

☑ Whether or not run test

Health Check successed! CAT01234567

Figure C-5 Summary check for the installation

BEAT also provides a dashboard web GUI for monitoring and management, as shown in Figure C-6 on page 213.



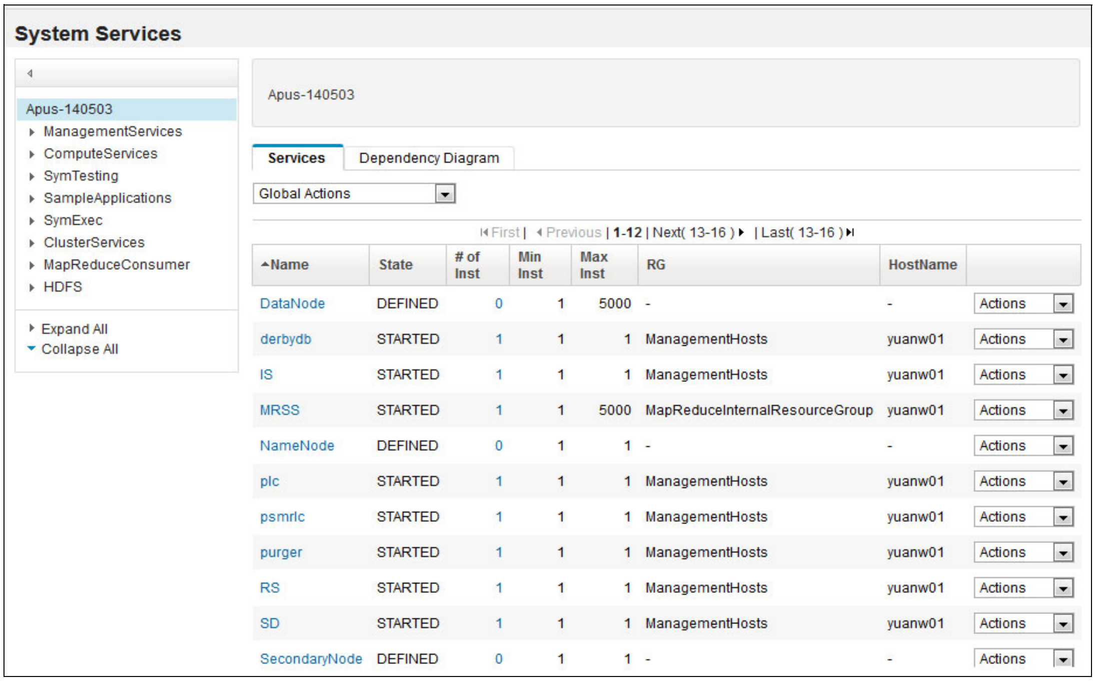
	Tasks			Sessions		Slot Allocation			
Application	Running	Pending	Error	Open	Aborted	Allocated	Deserved	Demanded	Blocked Hosts
symping7.1	0	0	0	0	0	0	0	0	0

	All Tasks			Jobs		
Application	Running	Pending	Error	Running	Aborted	Blocked Hosts

		Instances		
Service	State	Running	Minimum	Maximum

Figure C-6 Dashboard

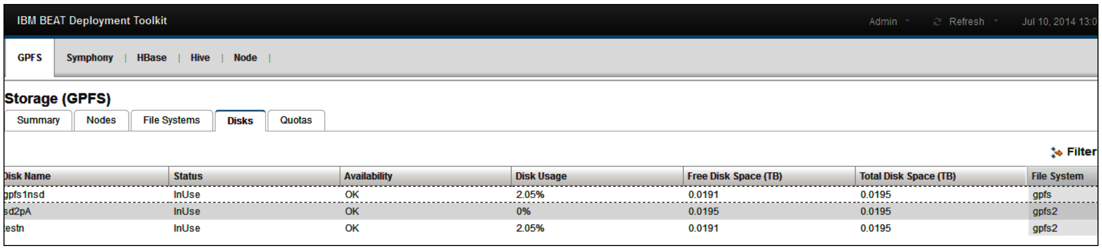
Figure C-7 shows the System Services monitoring window.



Name	State	# of Inst	Min Inst	Max Inst	RG	HostName	
DataNode	DEFINED	0	1	5000	-	-	Actions
derbydb	STARTED	1	1	1	ManagementHosts	yuanw01	Actions
IS	STARTED	1	1	1	ManagementHosts	yuanw01	Actions
MRSS	STARTED	1	1	5000	MapReduceInternalResourceGroup	yuanw01	Actions
NameNode	DEFINED	0	1	1	-	-	Actions
plc	STARTED	1	1	1	ManagementHosts	yuanw01	Actions
psmrlc	STARTED	1	1	1	ManagementHosts	yuanw01	Actions
purger	STARTED	1	1	1	ManagementHosts	yuanw01	Actions
RS	STARTED	1	1	1	ManagementHosts	yuanw01	Actions
SD	STARTED	1	1	1	ManagementHosts	yuanw01	Actions
SecondaryNode	DEFINED	0	1	1	-	-	Actions

Figure C-7 System services monitoring

Figure C-8 shows the storage monitoring dashboard.



Storage (GPFS)

Disk Name	Status	Availability	Disk Usage	Free Disk Space (TB)	Total Disk Space (TB)	File System
gpfs1nsd	InUse	OK	2.05%	0.0191	0.0195	gpfs
sd2pA	InUse	OK	0%	0.0195	0.0195	gpfs2
estn	InUse	OK	2.05%	0.0191	0.0195	gpfs2

Figure C-8 Storage monitoring dashboard

BEAT has been implemented at many customer sites by IBM Lab Services. Currently, BEAT has the following features under development:

- Monitoring the cluster by using a mobile device. Apps in Android or iOS easily connect with a big data cluster that is managed by BEAT, and can retrieve the cluster’s running status.
- Diagnosing bottlenecks from hardware and software applications.
- Providing performance tuning guidance with the configuration adviser.
- Supporting high availability for the management nodes of each service layer.
- Deployment configuration adviser.
- Easy upgrade or turn on/off the applications and services.

Related publications

The publications that are listed in this section are considered suitable for a more detailed description of the topics that are covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Some publications referenced in this list might be available in softcopy only.

- *IBM Spectrum Scale (formerly GPFS)*, SG24-8254
- *IBM Technical Computing Clouds*, SG24-8144
- *Implementing the IBM General Parallel File System (GPFS) in a Cross Platform Environment*, SG24-7844
- *Implementing IBM InfoSphere BigInsights on IBM System x*, SG24-8077

You can search for, view, download, or order these documents and other Redbooks, Redpapers, Web Docs, draft and additional materials, at the following website:

ibm.com/redbooks

Online resources

These websites are also relevant as further information sources:

- Deploying a big data solution by using IBM Spectrum Scale:

  http://ibm.co/1NBnGTj
- IBM InfoSphere BigInsights:

  http://ibm.co/1DlnmVy
- Configuring Yum repositories:

  http://red.ht/1CHzQX6
- IBM Platform Cluster Manager V4.2 documentation:

  http://www-01.ibm.com/support/knowledgecenter/SSDV85_4.2.0/pcm_welcome.html

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Business Services

ibm.com/services

Redbooks

Implementing an IBM InfoSphere BigInsights Cluster using Linux on Power

(0.2”spine)
0.17”<->0.473”
90<->249 pages

SG24-8248-00

ISBN 0738440744

Printed in U.S.A.