Redbooks

ibm.com/redbooks

Implementing an Optimized Analytics Solution on IBM Power Systems

Dino Quintero

Kanako Harada

Reinaldo Tetsuo Katahira

Antonio Moreira de Oliveira Neto

Robert Simon

Brian Yaeger

Analytics

Power Systems

IBM

Redbooks

International Technical Support Organization

Implementing an Optimized Analytics Solution on IBM Power Systems

June 2016

SG24-8291-00

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

First Edition (June 2016)

This edition applies to the following products:
- Version IBM Open Platform for Apache Hadoop 4.1.0.0
- IBM BigInsights 4.1.0.1
- Apache Ambari 2.1.0.0
- IBM Spectrum Scale Advanced Edition 4.1.1.3
- Installation Manager 1.8.3
- IBM Cognos Business Intelligence Server 10.2.2 Fix Pack (FP) 2
- IBM Cognos Framework Manager 10.2.2 FP2
- IBM Cognos Cube Designer 10.2.2 FP2
- Collaboration and Deployment Service 7.0.0.0
- Deployment Manager 7.0.0.0
- IBM SPSS Analytical Decision Management 17.0
- IBM SPSS Modeler Server 17.0
- DB2 with BLU Acceleration 10.5 FP6
- IBM WebSphere 8.5.5.0
- OpenLDAP 2.4.39
- Magento 1.9.2.1
- Apache 2.2.31
- MySQL 5.5.41-MariaDB

Contents

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

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, MD-NC119, Armonk, NY 10504-1785, US

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 jurisdictions 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 provide in any way it believes appropriate without incurring any obligation to you.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

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.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

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 actual people or business enterprises 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. The sample programs are provided “AS IS”, without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

AIX®
AIX/ESA®
BigInsights®
BLU Acceleration®
Bluemix®
Cognos®
DB2®
Global Business Services®
GPFS™
IBM®
IBM BLU®
IBM Elastic Storage™
IBM Spectrum™
IBM Spectrum Scale™
IBM Watson™
Informix®
InfoSphere®
Insight™
LSF®
Passport Advantage®
POWER®
Power Systems™
POWER7®
POWER7+™
POWER8®
Redbooks®
Redbooks (logo)  ®
SPSS®
Symphony®
WebSphere®
z/OS®

The following terms are trademarks of other companies:

Netezza, and N logo are trademarks or registered trademarks of IBM International Group B.V., an IBM Company.

is a trademark or registered trademark of Ustream, Inc., an IBM Company.

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

Microsoft, Windows, 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 personalized notifications of new content
- Link to the latest Redbooks blogs and videos

Get the latest version of the Redbooks Mobile App

iOS **Download Now** Android

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 addresses topics to use the virtualization strengths of the IBM POWER8® platform to solve clients' system resource utilization challenges and maximize systems' throughput and capacity.

This book addresses performance tuning topics that will help answer clients' complex analytic workload requirements, help maximize systems' resources, and provide expert-level documentation to transfer the how-to-skills to the worldwide teams.

This book strengthens the position of IBM Analytics and Big Data solutions with a well-defined and documented deployment model within a POWER8 virtualized environment, offering clients a planned foundation for security, scaling, capacity, resilience, and optimization for analytics workloads.

This book is targeted toward technical professionals (analytics consultants, technical support staff, IT Architects, and IT Specialists) who are responsible for providing analytics solutions and support on IBM Power Systems™.

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 Complex Solutions Project Leader and an IBM Level 3 Certified Senior IT Specialist with the ITSO in Poughkeepsie, New York. His areas of knowledge include enterprise continuous availability, enterprise systems management, system virtualization, technical computing, and clustering solutions. He is an Open Group Distinguished IT Specialist. Dino holds a Master of Computing Information Systems degree and a Bachelor of Science degree in Computer Science from Marist College.

Kanako Harada is a Certified Customer Support Analyst in Japan. She has worked for IBM for seven years. She has 10 years experience in the business analytics and optimization field. During her career, she was certified on IBM Statistical Package for the Social Sciences (SPSS®) and on IBM Cognos® products (Business Intelligence and Financial Planning Management). Kanako holds a degree from Waseda University in Japan. She worked on many integration projects as a Project Manager, and also as an experienced Software Analyst to deliver real-time financial intelligence information for several international organizations.

Reinaldo Tetsuo Katahira is an Information Architect, working for the Client Innovation Center under IBM Global Business Services® (GBS) in Brazil. He is a Thought Leader Certified IT Specialist in Data Management by IBM, and he has DB2®, Oracle, and MS SQL Server certifications. Reinaldo holds a degree in Computer Engineering from the Escola Politecnica da Universidade de Sao Paulo. He has 18 years of experience in database management systems (DBMS) and over 10 years of experience working at IBM, where he filed patent applications that relate to optimization of relational DBMS.

Antonio Moreira de Oliveira Neto is a Client Technical Specialist, working in IBM Systems Hardware in Brazil for pre-sales technical advisory and architecture designs. He has seven years of experience with IBM. His technical expertise areas include IBM Power Systems, Server Virtualization, Infrastructure Architecture, Hybrid Cloud, and Analytics Architectures. He also has extensive experience designing architecture for highly virtualized and highly available environments. He holds a degree in Computer Engineering from Instituto Tecnológico de Aeronáutica (Sao Paulo, Brazil). He is an Expert Certified IT Specialist.

Robert Simon is a Senior Software Engineer for IBM Spectrum™ Scale. With IBM since 1987, Robert works in Software Support (Level 2) for VM, IBM AIX/ESA®, PSSP (HPC), IBM General Parallel File System (GPFS™), and GPFS Native RAID.

Brian Yaeger is an Advisory Software Engineer in the United States. He has eight years of experience in the field of software technical support at IBM and before that two years of experience as a computer technician. He holds an Associates degree in Computer Science from SUNY Dutchess, a Bachelors degree in Computer Science from Plattsburgh State University, and he is actively pursuing a Masters degree in Information Systems Management at Marist College (anticipated graduation 2016). His areas of expertise include clustered computing with Spectrum Scale on Linux.

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

Richard Conway
International Technical Support Organization, Poughkeepsie Center

Cesar Maciel, Paul Hannan, Vlaunir Jimenez da Silva
IBM US

Fabio Martins and Pedro Henrique dos Santos Principeza
IBM Brazil

Dong Joo Song and Hajime Kawada
IBM Japan

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

Achieving significant, measurable business value and insight from big data and analytics technologies requires organizations to put an infrastructure foundation in place that supports the rapidly growing volume, variety, and velocity of data.

IBM Power Systems helps to pave the way by delivering innovation and optimization for analytics databases and applications to accelerate the delivery of insights and speed up decision making, offering our clients a sustainable competitive advantage.

This chapter describes the following topics:

- Why IBM POWER8
- POWER8 processor technology
- POWER8 scale-out servers

1.1 Why IBM POWER8

IBM offers a broad set of analytics capabilities built on the proven foundation of a single platform, IBM Power Systems. Power Systems is an open, secure, and flexible platform that is designed for big data. It has massive input/output (I/O) bandwidth to deliver analytics in real time, and it can provide the necessary capabilities to handle the varying analytics initiatives of each business.

In 2014, IBM announced POWER8, the first microprocessor that was designed for big data and analytics. POWER8 offers numerous advantages for big data and analytics solutions: processing capability, memory capacity and bandwidth, cache workspace, and the ability to move information in and out of the system at the required rapid speeds.

1.2 POWER8 processor technology

A few of the distinguishing capabilities of POWER8 are listed:

- Parallelism is the capability to process more concurrent queries in parallel faster and scale easily to support a growing number of users who need reports, or to perform ad hoc analytics.
- Increased memory bandwidth to move large volumes of data to memory faster to accelerate time to result.
- Four-level cache design in every processor. A robust cache design helps with handling large volumes of data for better response times.
- Faster I/O to ingest, move, and access large volumes of data for various data sources so that analytics results are available faster.
- Acceleration that is enabled by Coherent Accelerator Processor Interface (CAPI) technology, through which graphics processing units (GPUs), flash memory, networking, and field programmable gate arrays (FPGAs) connect directly to the processor, which helps to improve performance, reduce latency, and result in more workload for the dollar.

1.3 POWER8 scale-out servers

With these advantages, POWER8 delivers the levels of performance that you need to make decisions in real time, helping you to capitalize on the currency of data by finding business insights faster and more efficiently. Power Systems are designed for big data, from operational to computational to business and cognitive solutions. They are optimized for performance and can scale to support growing workloads. See Figure 1-1.

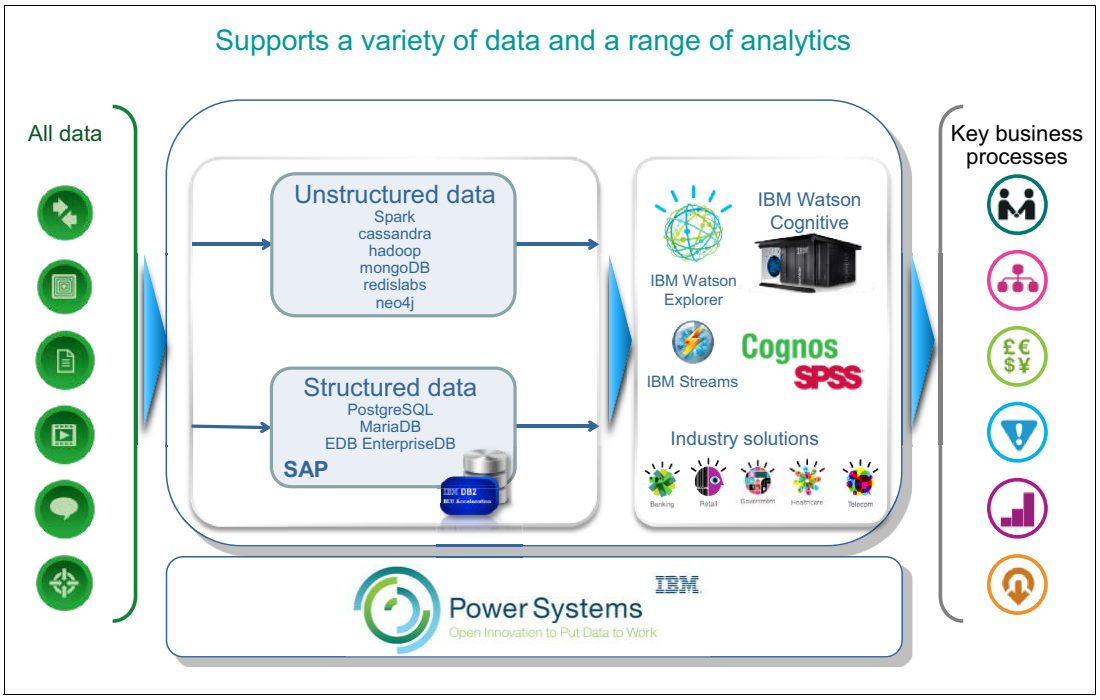


Figure 1-1 IBM Power Systems big data and analytics solutions

For more information about IBM POWER8, see Chapter 3, “IBM POWER8 for analytics workloads” on page 43.

2

Solution reference architecture

This chapter describes possible reference architectures that you can use when you deploy big data and analytics solutions, including the integration of its components. This chapter refers to architectures that are implemented on IBM Power Systems servers, and it focuses on a solution that implements the following solutions:

- Big data exploration
- Relational database with acceleration for analytics
- Analytical decision management
- Reporting and dashboarding on top of many sources of data (structured and unstructured)

Both hardware architecture and software components from the point of view of deploying the solution are described in this chapter. This chapter also illustrates how the components integrate to build a robust and ease-to-manage analytics solution.

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

This chapter also describes the following topics:

- Hardware architecture, including virtualization and management
- Software architecture for applications
- Data store approach
- IBM BigInsights® for Apache Hadoop cluster management deployment

The information is organized into the following areas:

- Big data and analytics general architectures
- Hardware reference architecture
- Software reference architecture
- Solution reference architecture

2.1 Big data and analytics general architectures

With the continuous growth of data and its variety, big data and analytics disciplines and architectures are extensively studied and improved. To gain effective insights from data, it is necessary to understand and explore the many sources, how to store them effectively, and how to discover and visualize it.

Data sources vary by source and structure. The traditional sources are transactional database applications that contain structured data. New sources with data in an unstructured format are explored more often. Certain sources of data are text-based documents, such as logs from web applications and data feeds from social-networking applications. The integration of these data sources forms the foundation for federated analytics. Correlating and analyzing these varieties of data is essential to gain valuable insights. The speed to get these insights is critical to the enterprise.

Figure 2-1 shows a logical view of a high-level big data, data storage, analysis, discovery, and visualization reference architecture.

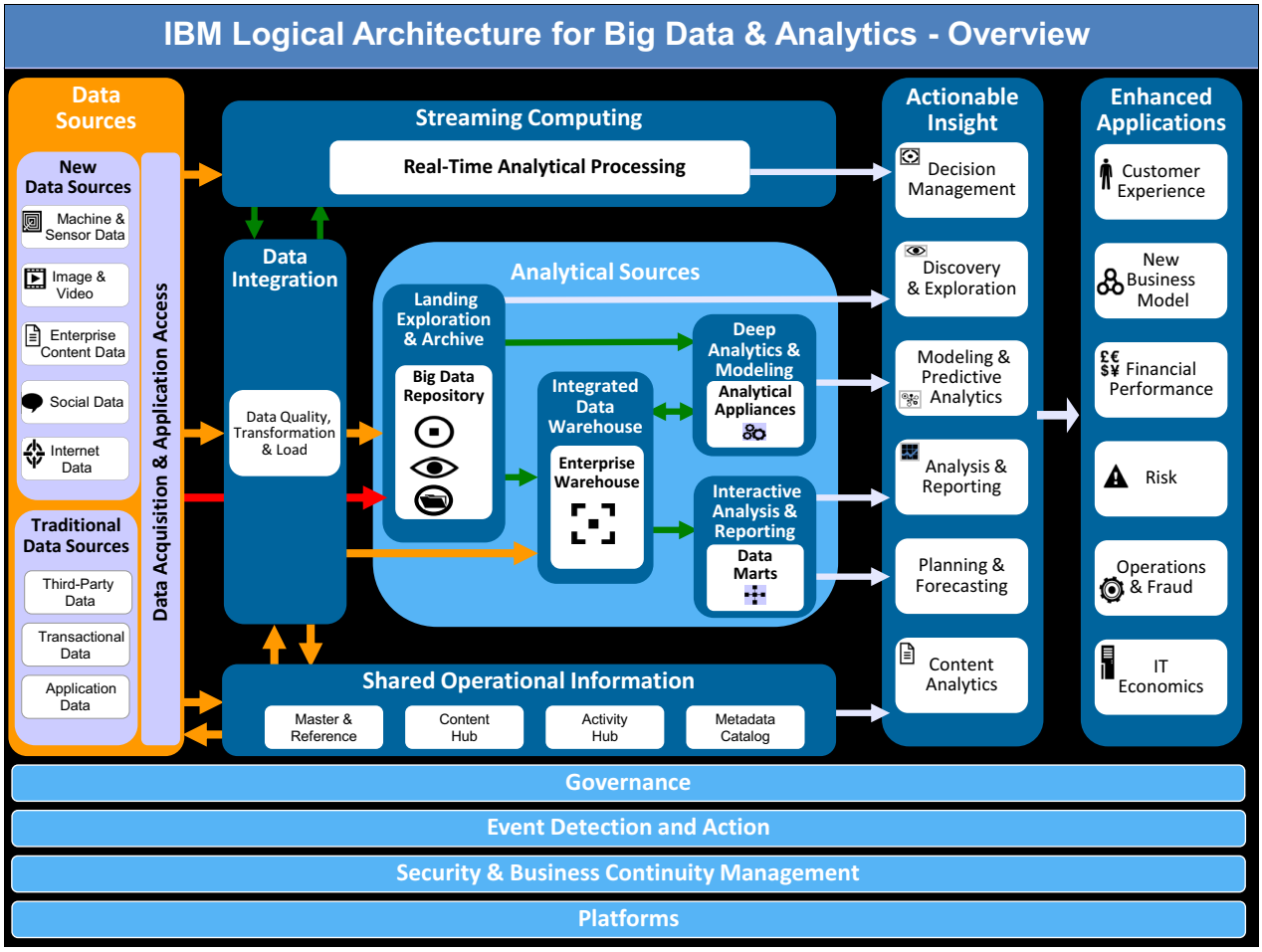


Figure 2-1 IBM big data and analytics reference architecture

From a big data perspective, Hadoop clusters are intensively used as data repositories for landing and exploring many kinds of structured and unstructured data, and discovering ways of gaining insights from it. IBM BigInsights for Apache Hadoop is the IBM platform for managing and analyzing persistent big data. It is built on top of the IBM Open Data Platform for Apache Hadoop, which consists of entirely Apache Hadoop open source components, such as Apache Ambari, HDFS, Flume, Hive, and ZooKeeper (Figure 2-2 shows the components for this solution).

The components that apply to your deployment depend on the data sources that you plan to integrate. Consider your current plans and future needs when you decide about the initial deployment so that the infrastructure can easily grow to support your business as your business changes.

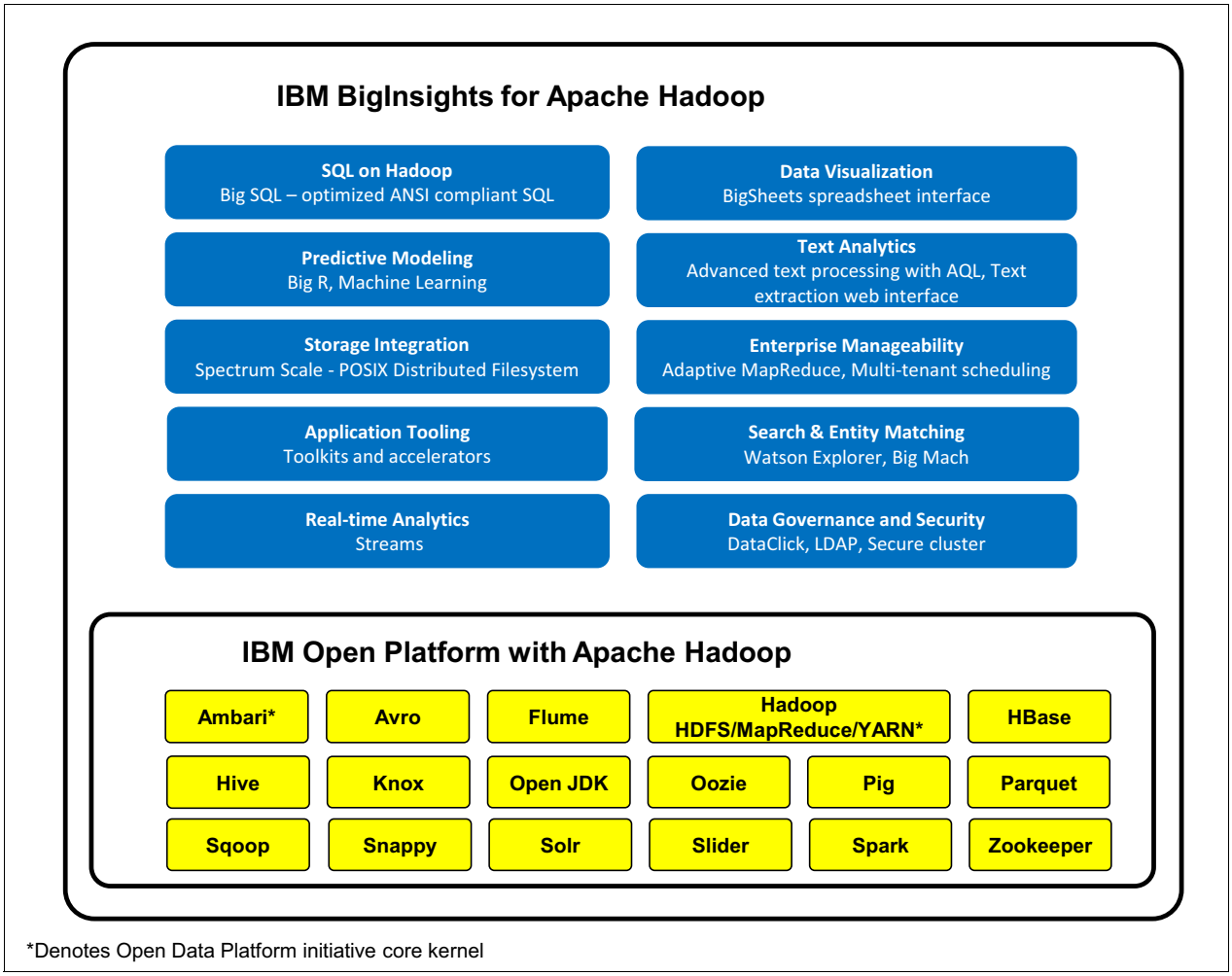


Figure 2-2 IBM BigInsights for Apache Hadoop major components

The data that is stored in Hadoop clusters can provide many kinds of actionable insights. Different sorts of analyses can be performed from this data. For example, based on past events and recorded transactions, you can use reporting and analysis tools, such as IBM Cognos, to help you gain insight and help to answer questions, such as “Why did it happen?” You can also use modeling and predictive analysis tools, such as IBM Statistical Package for the Social Sciences (SPSS) to help you search in the data for answers to questions, such as “What happened?”

Modeling and predictive analysis tools help with the decision-making processes.

In this way, analytics applications can use big data environments to help businesses in many areas, such as customer experience, fraud detection, and IT economics.

2.2 Hardware reference architecture

Selecting a hardware architecture to support the deployment of a big data and analytics solution requires an understanding of the relevant components and how they affect many aspects, such as performance, reliability, availability and serviceability (RAS), costs, and management. This publication shows a few available alternatives and how these components can help with the deployment of an efficient environment. Additionally, this book shows the deployment of a big data and analytics solution scenario that shows many of the reference architecture components.

One aspect to consider is which hardware architecture is correct for your environment. The answer to this question depends on whether you own any IBM Power Systems hardware, the type of hardware, whether you virtualize the hardware, or how you virtualize the hardware. To start this process, consider the following questions:

- Do you use existing servers or new servers?
- Do you use high-end Power Systems servers or entry-level Power Systems servers?
- Do you use only Linux Power Systems servers or generic Power Systems servers?
- Do you use virtualization or bare-metal servers?
- Do you deploy the solution in a cloud environment or in a technical computing cloud environment?

2.2.1 IBM Power Systems servers

IBM Power Systems servers can provide high-processing capacity, high bandwidth, and highly reliable hardware. However, you must consider a few points and make several decisions when you select the hardware for a big data and analytics environment deployment.

Two basic types of IBM Power Systems servers, enterprise and scale-out, exist:

- Enterprise servers provide enough capacity to grow the environment by using the same hardware (vertically). They offer advantages, such as systems consolidation, reduced floor space, and energy savings. You can use these servers to deploy applications that benefit from a scale-up architecture in terms of processor and memory, and virtualization features.
- Scale-out servers are based on the principle that the application will grow in capacity by adding new servers to the environment (horizontally). You can use these servers to deploy big data clustered workloads that do not benefit from a scale-up architecture in terms of memory and processor or from server virtualization (deployed in a bare-metal node architecture).

If you plan to deploy analytics services that can be managed within a virtualized environment, it makes sense to use high-end servers. You can manage logical partitions (LPARs) in your servers to create resource boundaries so that your application workload can fully use what it is given and still ensure consolidation of your hardware. Moreover, you can create a multitenant (cloud-style) environment by using the IBM Cloud Manager with Openstack, which handles the cloud management and multitenancy for you. However, achieving symmetry might be more challenging in this case.

Scale-out servers can be managed by using IBM Platform Computing solutions where you have a clearer distinction between node roles, such as management nodes, and computing or data nodes. No physical hardware overlap exists among these roles. With this architecture type, it is easier to achieve symmetry of the hardware architecture. Also, You can use software, such as the IBM Platform Cluster Manager - Advanced Edition, to automate the deployment of an IBM Open Data Platform for Apache Hadoop on the nodes.

Additionally, several predefined big data solutions can be deployed, such as IBM Data Engine for Analytics and IBM POSHv2, to take advantage of the IBM Open Data Platform for Apache Hadoop and IBM BigInsights for Apache Hadoop workloads.

Another consideration about the server is about using virtualization or bare metal servers. Consider the following important points:

- Scale-out or scale-up components
- Solution symmetry
- Automated, available deployment tools
- Size of the servers (scale-out or enterprise)

Perhaps, you are considering a mixed implementation in terms of virtualization. For example, you are thinking about virtualizing applications, such as Cognos, SPSS, and DB2 that traditionally scale up, and you choose not to deploy them as clustered nodes. In a BigInsights for Apache Hadoop cluster, you can implement the same nodes for the management nodes and use bare-metal nodes for the data nodes because they are traditionally scale-out components.

You also need to keep in mind that you can be restricted to use a specific operating system due to application support. For example, BigInsights for Apache Hadoop supports Linux platforms only. Since the launch of the IBM POWER7® hardware, IBM made available server models that run only Linux (as opposed to also running IBM AIX® and IBM i). At the time that this publication was written, IBM has the following available Power Systems server models that run only Linux:

- Power S812L (POWER8)
- Power S822L (POWER8)
- Power S824L (POWER8)

These server models fit more into the entry-level or mid-level server category than the high-end server category. So, the servers can benefit more from the bare-metal approach than the virtualized approach. However, if you must virtualize them, you can without any challenges. The POWER8 server models also benefit from PowerKVM virtualization. The S812L or S822L server model that is combined with PowerKVM-based virtualization provides a good option for price-performance in terms of virtualizing a Power Systems server for Linux environments only.

For more information about the IBM Power Systems server models, see the IBM Power Systems Quick Reference Guide at the following website:

https://ibm.biz/Bd4yQU

2.2.2 General architecture for BigInsights for Apache Hadoop

An BigInsights for Apache Hadoop cluster consists of management nodes and data nodes. Management nodes host the following services:

- Ambari
- Oozie
- Big SQL
- Catalog
- ZooKeeper
- HBase
- Hive
- IBM Platform Symphony®

Data nodes are the nodes that perform work based on the workloads that are running in the cluster. These nodes are interconnected through Internet Protocol networks. Therefore, a well-designed network architecture is important, and it plays a central role in cluster performance. As a preferred practice for a BigInsights for Apache Hadoop cluster, define at least three networks:

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

Another aspect of a BigInsights for Apache Hadoop 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 local to minimize network transfers of massive amounts of data. Therefore, the typical architecture uses disks that are assigned to only a single node, which 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 supports the *shared-nothing* architecture. Also, Spectrum Scale supports the *shared-nothing* architecture through its File Placement Optimizer feature. This book focuses on the use of Spectrum Scale as the file system to build the architecture design, working with File Placement Optimization and also approaching it with an IBM Elastic Storage™ Server perspective. For more information about possible disk layouts, see 2.2.5, “Data storage” on page 14.

Going further into the architecture, you can add nodes that ease the management of the hardware in the BigInsights for Apache Hadoop cluster. Imagine that you want to either add or remove a node from an existing BigInsights for Apache Hadoop cluster, or you want to create multiple, independent BigInsights for Apache Hadoop clusters. Performing these tasks manually is a time-consuming task. Cluster management software eases this task.

The scenarios and examples that are implemented in this book used IBM Platform Cluster Manager - Advanced Edition to provision the BigInsights for Apache Hadoop 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, with management and data nodes, in an automated fashion. The advantages for this configuration are described in 2.3.6, “Cluster management” on page 34.

You can add a system management node to your overall BigInsights for Apache Hadoop server farm to install Platform Cluster Manager - Advanced Edition for managing your hardware. This addition changes the network layout of your network environment because you set Platform Cluster Manager - Advanced Edition to communicate with the flexible service processor (FSP) ports of your Power Systems hardware. Also, Platform Cluster Manager uses a provisioning network to deploy systems. You can use your administrative network or a fourth network to isolate the traffic for systems provisioning. A detailed explanation of the networking components is available in 2.2.4, “Networking” on page 12.

2.2.3 General architecture for analytics applications

Many analytic applications can be implemented in a big data and analytics environment. This publication focuses on several of these applications:

- Reporting and analysis with the IBM Cognos Business Intelligence, which allows the visualization of data from different sources, such as BigInsights for Apache Hadoop and IBM DB2
- Decision Management with IBM SPSS Analytical Decision Management so that you can set up a “business rules engine” for optimizing transactional decisions and consistently maximize outcomes
- In-memory acceleration for transactional databases with DB2 BLU to maximize performance and efficiency when you analyze transactional data from an online analytical processing (OLAP) approach

These workloads provide flexibility for how you might deploy them. The workloads can easily be deployed as LPARs by using a scale-up hardware architecture, benefitting from virtualization and system consolidation.

DB2 BLU specifically accelerates analytic workloads by working with columnar tables and placing the minimum memory. A scale-up architecture might provide a better memory bandwidth, increasing memory throughput, and better memory RAS, which helps prevent risks of data loss.

In a virtualized environment deployment, you can use the Hardware Management Console (HMC) and other cloud management software, such as IBM PowerVC and IBM Cloud Manager with Openstack, as the infrastructure management components.

In a deployment that uses individual nodes for each application or dedicated LPAR in scale-out systems, Platform Cluster Manager - Advanced Edition can also be used for the deployment and management of those nodes, similar to BigInsights for Apache Hadoop nodes.

Additionally, in a production environment, we suggest that you deploy a high-availability environment to minimize unplanned downtime and minimize the effect of hardware or software failures.

For data storage, consider the use of a storage area network (SAN) for images, repositories, and data. Take advantage of the benefits of storage consolidation, live partition mobility, and so on. Software-defined storage is also an approach that can be used for additional benefits. For more information, see 2.2.5, “Data storage” on page 14.

2.2.4 Networking

BigInsights for Apache Hadoop 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 use Secure Shell (SSH) to get into the nodes or to access them through virtual network computing (VNC). This network can be as simple as a 1-Gb Ethernet port with high availability through Ethernet bonding.

Based on your environment's requirements, the administration network can be segregated into separate virtual local area networks (VLANs) or subnets. It 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 BigInsights for Apache Hadoop cluster. Think of the public network as the public face of your corporate network. It is the network that you use to access the Ambari web interface or the BigInsights home web portal and perform your big data work. Although all cluster nodes can be connected to this network, the management nodes are the only nodes with configurable services, such as an HTTP server service, on them. 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 you manage multiple clusters through Platform Cluster Manager - Advanced Edition.

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

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

Applications, such as Cognos Business Intelligence and SPSS Modeler, can use the edge nodes to connect to the cluster and use the capabilities of the Big SQL component on BigInsights for Apache to connect to those clusters. For more information about this integration, see Chapter 4, "Scenario: How to implement the solution components" on page 49.

Figure 2-3 shows how the networking architecture looks in an environment that follows the guidelines in this chapter.

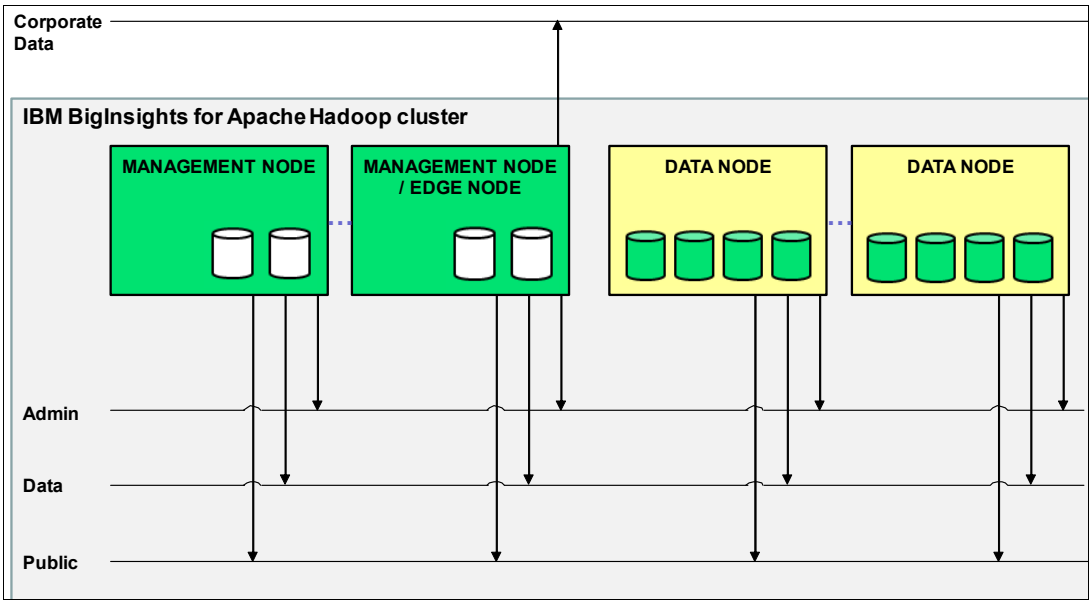


Figure 2-3 High-level BigInsights for Apache Hadoop cluster architecture: nodes, networks, and disks

You can deploy the general architecture that is described in this chapter in any kind of environment, whether the environment is bare-metal nodes, logical partitions (LPARs) in larger servers, or even a cloud environment.

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

The *service* network is used for the hardware-level management functions, such as power-cycling the nodes in the cluster, hardware status monitoring, firmware configuration, and hardware console access. The *service* 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 use 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, also.

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

Figure 2-4 shows a complete network architecture of a BigInsights for Apache Hadoop environment that is managed by Platform Cluster Manager - Advanced Edition. This scenario implements a provisioning network that differs 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 network.

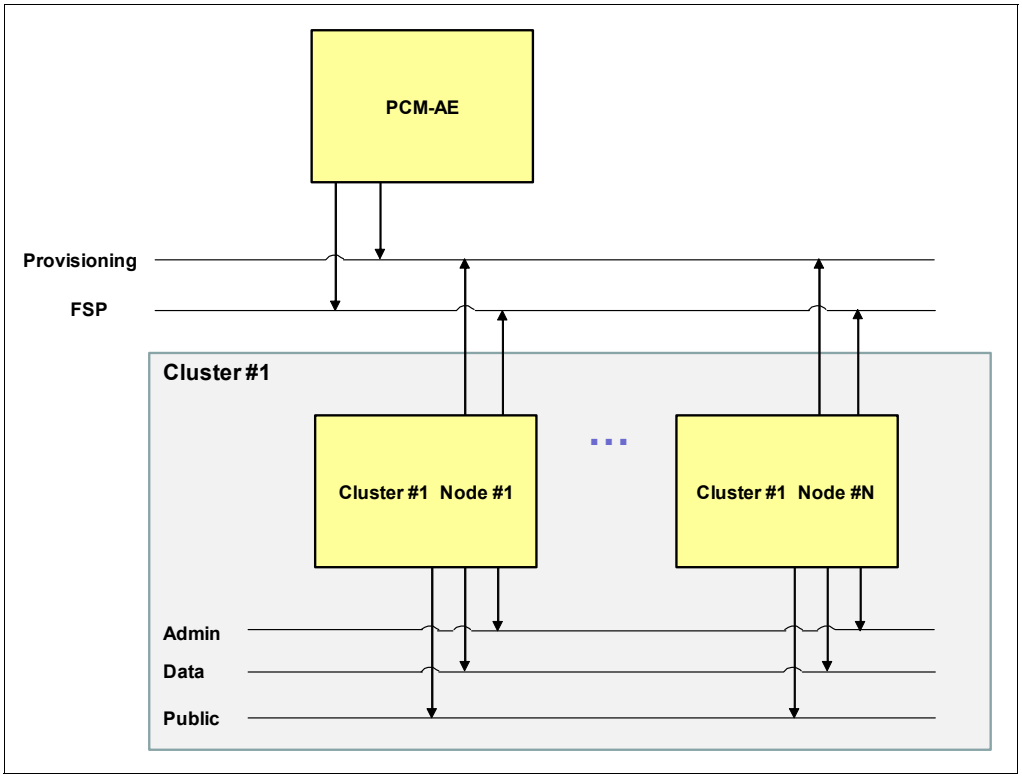


Figure 2-4 BigInsights for Apache Hadoop cluster network diagram

Similar to big data clusters, other analytic applications can have their own network configuration requirements. The applications that are the focus in this book use a traditional network arrangement. The applications basically need access to the corporate data network and the administrative network (to perform administrative tasks). You might need access to other networks if your infrastructure has special requirements, for example, a backup network.

2.2.5 Data storage

When you implement a big data and analytics solution, you must consider many options for data storage, for example, internal disks and external disks.

When you deploy applications, for example, DB2, SPSS, and Cognos, we suggest that you use external storage benefits, such as flexibility, performance, and management capabilities. A software-defined storage approach can be considered by using IBM Spectrum Scale™. The storage solution can be deployed through an Elastic Storage Server implementation. This implementation can combine Power Systems servers, storage enclosures, and disks with Spectrum Scale and its Native RAID technology to provide analytic and technical computing storage and data services for analytic workloads.

When you plan for analytics application storage subsystems, consider the requirements for different applications, including bandwidth and operational throughput. For example, DB2 and Hadoop workloads behave differently and their profiles of data access differ, which might require different disk layout and capabilities.

When you deploy a Hadoop cluster, the simplest shared-nothing disk layout that can be used with MapReduce workloads is the use of internal disks in the cluster nodes. It is usually the most cost-effective scenario. However, you can still achieve a shared-nothing environment by using disks that are external to the machines, either on storage expansion units or storage devices.

Scenarios that work with an external storage device can use high availability in terms of disk access and also ensure a shared-nothing layout. This task is accomplished by assigning the disks to two nodes simultaneously and by using Spectrum Scale to 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 it serves the disk only if the primary node fails. This task is controlled by a Spectrum Scale File Placement Optimizer (FPO) failure groups configuration. For more information about how this technology works, see 2.3.5, “Spectrum Scale and File Placement Optimizer” on page 33.

Consider that Hadoop-based technologies process large amounts of data that is local on a server to reduce I/O transfers over the network and to use fast I/O. Assume that your environment consists of 100 nodes, each with access to 10 disks, for a total of 1,000 disks. Can a single SAN unit and its SAN topology provide enough bandwidth to feed I/O to all 1,000 disks with performance that is as good as though each of the 100 nodes accessed 10 internal disks? For Hadoop workloads, we do not recommend that you use a SAN architecture without considering I/O performance.

2.2.6 IBM Data Engine for Analytics reference architecture

The IBM Data Engine for Analytics - Power Systems Edition (IDEA) provides an expertly designed, tightly integrated solution for running big data workloads. Consider choosing this solution for your analytics environment.

IDEA consists of a hardware and software implementation. It uses IBM Power Systems, which are managed by Platform Cluster Manager Advanced Edition to deploy a big data cluster. Standard open source MapReduce applications are enabled through the inclusion of IBM Open Platform for Apache Hadoop. Additional added value analytics are available through the optional inclusion of BigInsights Data Scientist or BigInsights Data Analyst. This optimized configuration enables users to become productive quickly.

The IDEA architecture was designed to provide client value through the following capabilities:

- Integrated complete cluster solution

  It has the necessary hardware and software components for implementing a BigInsights for Apache Hadoop cluster and start developing applications on top of it.

- Best-in-class hardware

  The IBM Power Systems hardware is known for its performance, high availability through component redundancy, robustness, and reliability. Moreover, and especially for Hadoop workloads, because certain Power Systems server models are targeted to run Linux only, it is 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.

- Innovative storage

  Implement the IBM Elastic Storage Server, which provides the cluster storage solution, with scalable Portable Operating System Interface (POSIX)-compliant storage to house both structured and unstructured data. The Elastic Storage Server is built by using IBM Spectrum Scale, which is based on the same General Parallel File System (GPFS) technology that solved the challenges of managing large data sets in High Performance Computing (HPC) environments for over two decades.

- Flexibility for storage and computing sizing

  Use the shared storage approach, by using Elastic Storage Server, to tailor configuration ratios between CPUs and storage capacity on the data nodes, according to the characteristics of the cluster usage. Also, you can scale them independently. Clusters that support more storage-intensive applications can require a lower CPU-to-storage capacity and clusters that support more CPU-intensive applications can require a higher CPU-to-storage ratio.

- Multitenancy

  The architecture supports multitenancy, which is achieved by using Platform Symphony Advanced Edition to configure a share of resources between groups of users with guaranteed service level agreements (SLAs).

- Ease of deployment

  The full solution is assembled and installed at an IBM delivery center before delivery with all included software preinstalled. Onsite services personnel integrate the solution into the client data center. The solution includes Platform Cluster Manager Advanced Edition to simplify deployment and monitoring of the cluster.

IDEA is built by using IBM POWER8 systems and uses all of the benefits, such as CPU performance, high memory bandwidth, and high RAS capabilities. All of the servers in the cluster are configured with Red Hat Enterprise Cluster.

The hardware architecture consists of the following components:

- An HMC

  The HMC's main function is to manage the system management node.

- A system management node

  Platform Cluster Manager Advanced Edition is installed on a system management node. It is used to provision and monitor the nodes that make up the cluster. The system management node is also used as a repository for operating systems and software images for initial installation and updates. Usually, one node for each system is sufficient, but, if the size of the cluster is large, or if high availability is important, two or more system management nodes are required. IBM Power S812L is used for this node.

- Analytics node: Management nodes

  Management nodes are used for the management services in the BigInsights for Apache Hadoop cluster. Management nodes are typically distributed across three to six nodes, depending on the services that will run and whether high availability is required. Two management nodes can run in LPARs in a single IBM Power S822L server.

- Analytics node: Data nodes

  These servers store the data in the distributed environment. In a default configuration, each node is configured in a Power S822L server in a full partition configuration. If Big SQL is included in the solution through the inclusion of the optional BigInsights Data Scientist or Analyst added value packages, the data nodes must be configured with two LPARs for each server instead of a single LPAR for each server.

- Analytics node: Edge nodes

  An edge node is an optional node type that acts as a gateway between the BigInsights for Apache Hadoop cluster and the external environment as a path to load and unload data. These nodes are configured as management nodes with additional connections to external network. Edge nodes can run as either one or two LPARs for each IBM Power S822L server.

- Elastic Storage Server

  The Elastic Storage Server (ESS) is an integrated shared storage solution that provides the Hadoop Distributed File System (HDFS)-compatible file system through Spectrum Scale. The solution consists of two Power S822L servers, each of which runs a single LPAR, and with two different disk enclosure types. The smaller enclosure has a 24-disk capacity and the larger enclosure has a 60-disk capacity. Four models are available, which use the smaller enclosure type. ESS 5146-GS1 uses one enclosure. ESS 5146-GS2 uses two enclosures. ESS 5146-GS4 uses four enclosures. ESS 5146-GS6 uses six enclosures. Three models are available, which use the larger enclosure type. ESS 5146-GL2 uses two enclosures. ESS 5146-GL4 uses four enclosures. ESS 5146-GL6 uses six enclosures.

Figure 2-5 shows an example of a preconfigured IDEA cluster.

System Management Node – S812L

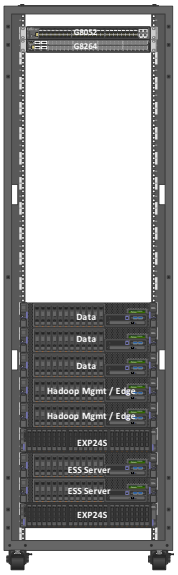
✓10x 3.425GHz cores
✓32GB DRAM
✓2x 1.2TB 10K SAS HDD (OS)
✓1x dual-port 10GbE NIC (data + mgmt)
✓1x 4-port 1GbE NIC (mgmt)

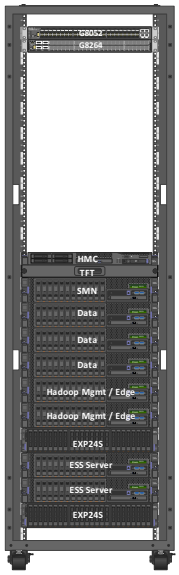
Hadoop Management Node – S822L

✓2 LPARs with split backplane
✓24x 3.02GHz cores, 256GB DRAM
✓8x 1.2TB 10K SAS HDD (OS + data)
✓2x dual-port 10GbE NIC (data + mgmt)

Analytics Node – 5146-S22

✓Standard: 24x 3.02GHz cores, 256GB DRAM, up to 10x 1.2TB 10K SAS HDD, 1x dual-port 10GbE NIC (data + mgmt)
✓Performance: 24x 3.02GHz cores, 512GB RAM, up to 10x 775GB SSD, 1x dual-port 40GbE NIC (data + mgmt)





	Without HMC/TFT/SMN	With HMC/TFT/SMN
Available Filesystem Capacity	38.6 TB	38.6 TB
Number of Data Nodes	3	3
Hadoop Mgmt LPARs or Edge Nodes	4	4
System Mgmt Node	0	1

Figure 2-5 IDEA infrastructure components example

IDEA is configured with three networks: management, service, and data. The management network provides functions for both administrative and provision networks, as described in 2.2.4, “Networking” on page 12.

Both management and service networks use a 1Gb Ethernet top-of-rack (TOR) switch. The service network VLAN requires one connection to the system management node, one connection to each server for out-of-band FSP hardware management, one connection to each network switch for out-of-band switch management, and two connections to each Elastic Storage Server storage enclosure for out-of-band storage management. The management network VLAN requires one connection to the system management node and one connection for each analytics node and storage server.

The data network VLAN requires one or two connections to the system management node, one or two connections to each analytics node, and a variable number of physical links for each storage server, which is determined by balancing aggregate network bandwidth to the aggregate storage bandwidth. The data network can use one of the following high-performance switch options:

- 10 Gb Ethernet top-of-rack switch (Mellanox SX1410, SX1400, or SX1036)
- 40 Gb Ethernet top-of-rack switch (Mellanox SX1710)
- InfiniBand Fourteen Data Rate (FDR) (56 Gbps) top-of-rack switch (Mellanox SX6036)

Figure 2-6 shows the IDEA networks.

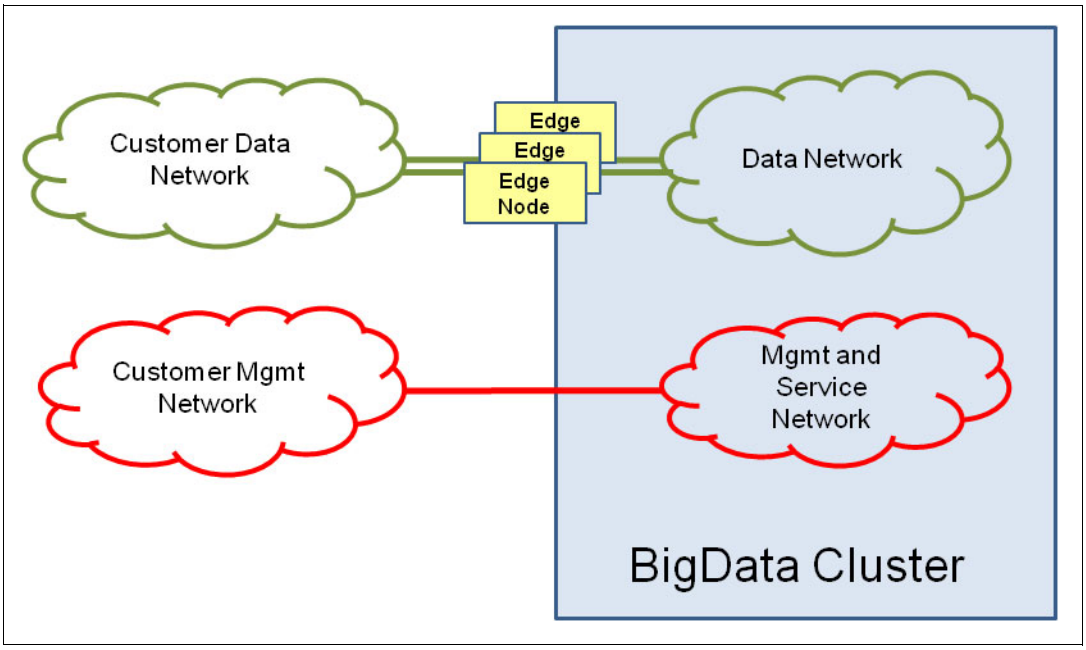


Figure 2-6 Networks that are used in IDEA

2.3 Software reference architecture

A big data and analytics environment can support many processes, as described in 2.1, “Big data and analytics general architectures” on page 6. The deployments of these environments consist of a set of software components, each of which offers benefits to your enterprise.

This section focuses on a set of software components to help solve business problems and manage environments:

- IBM BigInsights for Apache Hadoop for data storage and analysis
- Cognos Business Intelligence for reporting and dashboarding
- SPSS for predictive analytics
- DB2 with BLU acceleration for transactional data and OLAP operations

- Spectrum Scale for data storage and parallel access
- Platform Cluster Manager - Advanced Edition for infrastructure management in a cluster

2.3.1 IBM BigInsights for Apache Hadoop and IBM Open Platform with Apache Hadoop clusters

IBM BigInsights for Apache Hadoop is a software platform for discovering, analyzing, and visualizing data from disparate sources. The solution is used to help process and analyze the volume, variety, and velocity of data that continually enters organizations every day. BigInsights is a collection of added value services that can be installed on top of the IBM Open Platform with Apache Hadoop, which is the open Hadoop foundation.

By combining these technologies, BigInsights for Apache Hadoop extends the Hadoop open source framework with enterprise-grade security, governance, availability, integration into existing data stores, tools that simplify developer productivity, and more.

Hadoop is a computing environment that is built on top of a distributed, clustered file system that is designed specifically for large-scale data operations. Hadoop is designed to scan through large data sets to produce its results through a highly scalable, distributed batch processing system. Hadoop consists of two main components: a file system, which is known as the Hadoop Distributed File System (HDFS), and a programming paradigm, which is known as Hadoop MapReduce. To develop applications for Hadoop and interact with HDFS, you use additional technologies and programming languages, such as Pig, Hive, Flume, and many others.

Figure 2-2 on page 7 shows the software components on the BigInsights for Apache Hadoop architecture. Figure 2-7 shows how these components are packaged from a licensing perspective. IBM BigInsights Analyst, IBM BigInsights Data Scientist, and IBM BigInsights Enterprise Management extend the Hadoop open source framework by adding value packages according to what will be implemented in your big data cluster environment.

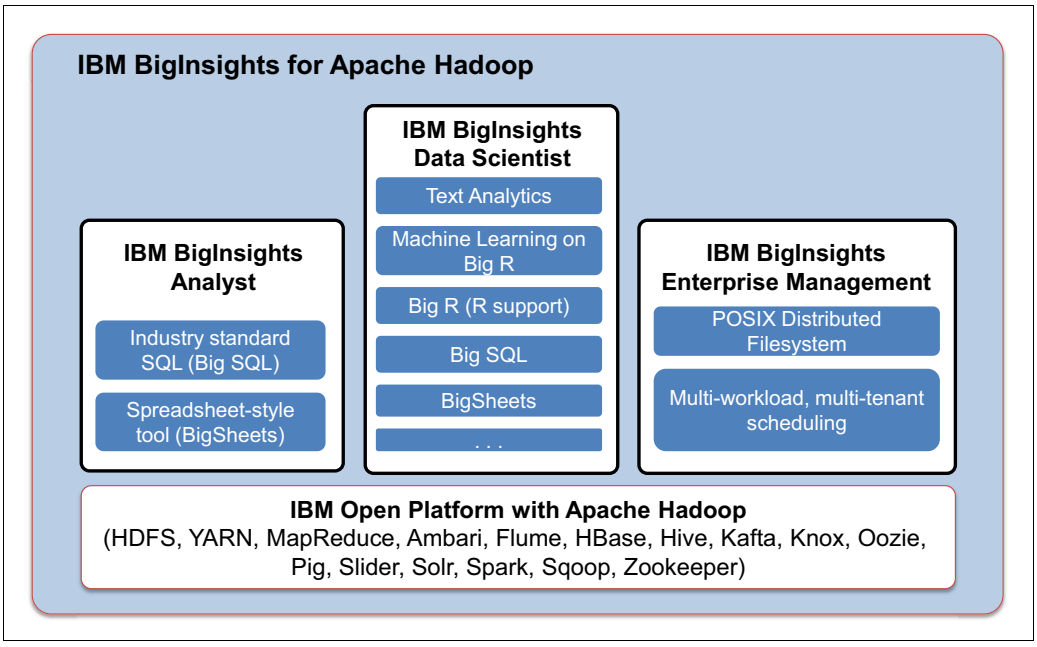


Figure 2-7 IBM BigInsights for Apache Hadoop software components

These components are the basis for building a BigInsights for Apache Hadoop cluster. This solution can be deployed on the hardware architectures that are described in 2.2, “Hardware reference architecture” on page 8.

This software stack runs on Linux on Power. For BigInsights for Apache Hadoop on Power Systems, the only supported version at the time that this publication was written is Red Hat Enterprise Linux 7.1 little endian. For more information about operating system requirements, see 4.2.5, “Installing the BigInsights value-add packages” on page 73.

Hadoop Distributed File System (HDFS)

Open source Hadoop traditional deployments use the Hadoop Distributed File System to store and share data across the many nodes in a cluster. Data is broken into smaller pieces that are called *blocks*, then they are distributed throughout the nodes in the cluster. This process also includes copying the blocks to increase the fault tolerance of the cluster. It is common to have a total of three copies of the data in HDFS deployments.

Am HDFS implementation has two major components:

- DataNode

  Each HDFS cluster has a number of DataNodes, with one DataNode for each node in the cluster. DataNodes manage the storage that is attached to the nodes on which they run. When a file is split into blocks, the blocks are stored in a set of DataNodes that are spread throughout the cluster. DataNodes are responsible for serving read and write requests from the clients on the file system, and they also handle block creation, deletion, and replication.

- NameNode

  An HDFS cluster supports NameNodes. An active NameNode and a standby NameNode are a common setup for high availability. The NameNode regulates the access to files by clients, and it tracks all data files in HDFS. The NameNode determines the mapping of blocks to DataNodes, and handles operations, such as opening, closing, and renaming files and directories. All of the information for the NameNode is stored in memory, which provides quick response times when you add storage or read requests.

  The NameNode is the repository for all HDFS metadata, and the user data never flows through the NameNode. A typical HDFS deployment has a dedicated computer that runs only the NameNode because the NameNode stores metadata in memory. If the computer that runs the NameNode fails, the metadata for the entire cluster is lost, so this server is typically more robust than other servers in the cluster.

IBM Open Platform for Apache Hadoop can be deployed by using an HDFS implementation. BigInsights for Apache Hadoop extends these capabilities by providing Spectrum Scale with File Placement Optimization technology as an alternative that brings many advantages over HDFS. This approach and its benefits are described in 2.2.5, “Data storage” on page 14. It is also possible to use an existing Spectrum Scale installation with a new IBM BigInsights for Apache Hadoop new implementation. 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 file system software.

The Elastic Storage Server, which is described in 2.2.5, “Data storage” on page 14, is based on Spectrum Scale technology. Also, it can be used in these deployments, as a dedicated installation, storing only the BigInsights for Apache Hadoop data, or as a shared installation, where other Spectrum Scale file systems might be created to store data from other applications, such as Cognos and DB2.

MapReduce and YARN

MapReduce is a programming paradigm where applications are divided into self-contained units of work, and each of them can run on any node in the cluster. In a Hadoop cluster, a MapReduce program is known as a *job*. A job is run by being broken down into pieces that are known as *tasks*. These tasks are scheduled to run on the nodes in the cluster where the data exists.

In IBM Open Platform with Apache Hadoop, the MapReduce framework, MapReduce v2, runs as a YARN workload framework. The benefits of this new approach are that resource management is separated from workload management, and MapReduce applications can coexist with other types of workloads, such as Spark or Slider.

MapReduce v2 jobs are executed by YARN in the Hadoop cluster. The YARN ResourceManager creates a MapReduce ApplicationMaster container, which requests additional containers for mapper and reducer tasks. The ApplicationMaster communicates with the NameNode to determine where all of the data that is required for the job exists across the cluster. It attempts to schedule tasks on the cluster where the data is stored, rather than sending data across the network to complete a task. The YARN framework and the HDFS typically exist on the same set of nodes, which enables the ResourceManager program to schedule tasks on nodes where the data is stored.

The reduce task is always completed after the map task. A MapReduce job splits the input data set into independent chunks that are processed by map tasks, which run in parallel. These bits, which are known as *tuples*, are key and value pairs. The reduce task takes the output from the map task as input, and it combines the tuples into a smaller set of tuples. Each MapReduce ApplicationMaster monitors its created tasks. If a task fails to complete, the ApplicationMaster will reschedule that task on another node in the cluster.

This distribution of work enables map tasks and reduce tasks to run on smaller subsets of larger data sets. Ultimately, this distribution provides maximum scalability. The MapReduce framework also maximizes parallelism by manipulating data that is stored across multiple clusters. MapReduce applications do not have to be written in Java, although most MapReduce programs that run natively under Hadoop are written in Java.

Integrating BigInsights for Apache Hadoop and IBM Platform Symphony

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

Platform Symphony can be used in a BigInsights for Apache Hadoop environment as a job scheduler for MapReduce tasks. Platform Symphony can replace the open source Hadoop scheduler in a framework that is based on MapReduce. It can provide the following advantages:

- 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 the jobs of an 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, maximizing the uptime of your applications.
- Open so that it is compatible with multiple application programming languages (APIs) and languages, such as Hive, Pig, and Java. Also, it is compatible with both HDFS and Spectrum Scale.

IBM value add package: Big SQL

One of the most valuable features that is added by IBM BigInsights for Apache Hadoop is IBM Big SQL. Big SQL is a software layer that allows users and applications to query the Hadoop cluster by using familiar SQL statements.

It is a massively parallel processing (MPP) SQL engine that deploys directly on the physical Hadoop Distributed File System (HDFS) or Spectrum Scale cluster. This SQL engine pushes processing down to the same nodes that hold the data. Big SQL uses a low-latency parallel execution infrastructure that accesses Hadoop data natively for reading and writing.

Big SQL consists of two services: *head* and *worker*. Queries are received by the head nodes, which push them to the data nodes to process and return the results. Deployments have head nodes, which are installed on the Hadoop management nodes, and worker nodes, which are installed on two head nodes (a primary and a secondary).

Big SQL uses the Hive database catalog (HCatalog) for table definitions, location, storage format, and the encoding of input files. This Big SQL catalog is on the head node. If the data is defined in the Hive Metastore and accessible in the Hadoop cluster, Big SQL can get to it. Big SQL stores part of the metadata from the Hive catalog locally for ease of access and to facilitate query execution.

Big SQL uses the IBM Data Server Client drivers. This driver package uses the same standards-compliant Java Database Connectivity (JDBC), Java Combined Client (JCC), Open Database Connectivity (ODBC), call level interface (CLI), and .NET drivers that are used in other IBM software products, such as DB2 for Linux, UNIX, and Windows, IBM DB2 for z/OS®, and IBM Informix® database software. Because the same driver is shared across these platforms, other languages that already use these drivers, such as Ruby, Perl, Python, and PHP Hypertext Preprocessor (PHP), can interact with Big SQL with no additional custom configuration or drivers. Therefore, applications can interact between traditional database management systems (DBMSs) or data warehouse systems and Big SQL.

BigSheets

BigSheets is a spreadsheet-like, web-based application that allows the 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 they perform those transformations on the whole cluster. This situation keeps the workload on the system down and provides valuable insight.

BigR

R is an open source language that is used for statistical analysis and creating graphical displays of data. Although IBM BigInsights for Apache Hadoop does not install R itself, it does provide BigR. 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.

2.3.2 DB2 with BLU Acceleration

IBM BLU® Acceleration® is one of the most significant advances in technology in DB2 and in the database market in general. Available with the IBM DB2 10.5 release, BLU Acceleration delivers performance improvements for analytic applications and reporting by using dynamic in-memory optimized columnar technologies.

Although BLU Acceleration is an important new technology in DB2, BLU Acceleration is built directly into the DB2 kernel. BLU Acceleration is not only an extra feature. It is a part of DB2, and every component of DB2 is aware of BLU Acceleration. BLU Acceleration still uses the same storage unit of pages, the same buffer pools, and the same backup and recovery mechanisms.

For more information about BLU Acceleration, see *Architecting and Deploying DB2 with BLU Acceleration*, SG24-8212*,* at the following website:

`http://www.redbooks.ibm.com/abstracts/sg248212.html`

Typical experiences of using DB2 with BLU Acceleration show good approaches and results:

- Helps you to achieve performance improvements of about 10x to 20x
- Helps increase storage savings versus decompressed data of about 5x to 20x

The performance and response time of IT systems, especially business intelligence systems, when you run reports are always a source of concern. No matter what is done, these systems' performance and response times can always be improved.

Simple to implement and use

Keep it simple is a strong, almost mandatory concept in BLU Acceleration. Because of its implementation, maintenance, and daily usability, the following keep-it-simple characteristics are present:

- One setting is necessary to optimize the DB2 system for BLU Acceleration. It is only necessary to set one database variable, when the database is used for analytic workloads, and optimize for optimal analytics performance.
- No additional workload and maintenance: No indexes, multi-dimensional clustering (MDC), statistics views, manual reorganization, or RUNSTATS (these last two tasks are automated).
- All of the features are built into the DB2 kernel: SQL, language interfaces, administration, reusing the DB2 process model, storage concepts, and utilities.
- Simple table creation and conversion.

Column store

The most basic and prominent feature of BLU Acceleration is the column-organized table type. Column-organized tables store each column on a separate set of pages on disk, reducing the necessary I/O for processing queries that are loaded into memory from disk. The following features helped generate savings in tests:

- Minimal I/O: By only performing I/O in the columns and values that match the query, and by reducing the working set of pages during the query progression
- Work that is performed directly in columns: By working on individual columns for predicate evaluations, joins, scans, and so on, and not materializing rows until necessary to build the result set

- Improved memory density and extreme compression: By keeping columnar data compressed in memory and by packing more data values into a small amount of memory or disk
- Cache efficiency: By packing data into CPU cache-friendly structures

By being able to store both row-organized and column-organized tables in the same database, users can implement BLU Acceleration even in database environments where mixed online transaction processing (OLTP) and online analytical processing (OLAP) workloads are required. Again, BLU Acceleration is built into the DB2 engine. The SQL, optimizer, utilities, and other components are fully aware of both row-organized and column-organized tables at the same time.

Data skipping

Data skipping avoids the unnecessary processing of irrelevant data, further reducing the I/O that is required to complete a query.

Automatic detection of large sections of data that does not qualify for a query can be ignored. Data skipping is used for data in memory (buffer pool) and on disk and helps to significantly reduce I/O, memory, and CPU consumption.

BLU Acceleration performs data skipping in the following way. As data is loaded into column-organized tables, BLU Acceleration tracks the minimum and maximum values on ranges of rows in metadata objects that are called the *synopsis tables*. These metadata objects (or synopsis tables) are dynamically managed and updated by the DB2 engine without intervention from the DBA.

When a query is run, BLU Acceleration looks up the synopsis tables for ranges of data that contains the value that matches the query. It effectively avoids the blocks of data values that do not satisfy the query, and it skips straight to the portions of data that matches the query. The net effect is that only necessary data is read or loaded into system memory, which in turn provides a dramatic increase in the speed of the query execution because much of the unnecessary scanning is avoided.

Extreme or adaptive compression

The column data is compressed with actionable compression, which preserves order so that the data can be used without decompression, resulting in storage and CPU savings and a higher density of useful data that is held in memory.

This benefit is possible because of the following features:

- Massive compression with approximate Huffman encoding, considering that the more frequent the value, the fewer bits it takes.
- Encoded values that are packed into bits matching the register width of the CPU.
- Late materialization, which is the ability to operate on the data while it is still compressed. Predicates and joins work directly on encoded values (actionable compression).

In addition to column-level compression, BLU Acceleration also uses page-level compression when appropriate to help further to compress the data based on the local clustering of values on individual data pages.

Because BLU Acceleration can handle query predicates without decoding the values, more data can be packed in the processor cache and buffer pools, which results in less disk I/O, better use of memory, and more effective use of the processor. Therefore, query performance is better and storage utilization is also reduced.

Deep hardware exploitation

BLU Acceleration optimizes the entire access to the hardware and its usage, seeking every opportunity (memory, CPU, and I/O). BLU Acceleration is designed to fully use all of the computing resources that are provisioned to the DB2 server by using Single Instruction Multiple Data (SIMD)-capable CPUs.

SIMD instructions are low-level specific CPU instructions. DB2 can use a SIMD instruction to get results from multiple data elements (perform equality predicate processing, for example) if they are in the same register. DB2 has deep processor and memory usage, including AIX Workload Management (WLM) characteristics, within its own workload policies.

Considering the processor usage, DB2 offers these functions:

- Deep usage of simultaneous multithreading (SMT)
- Key IBM POWER® value proposition with the ability to dispatch many threads
- Decimal arithmetic that is performed directly on the DECFLOAT accelerator

DB2 with BLU Acceleration has special algorithms that automatically take advantage of the built-in parallelism in the processors if SIMD-enabled hardware is available. The algorithms are another feature in BLU Acceleration that allows the use of special hardware instructions that work on multiple data elements with a single instruction.

Core-friendly parallelism

BLU Acceleration is a dynamic in-memory technology. It efficiently uses the number of processor cores in the current system, allowing queries to be processed by using multi-core parallelism and scale across processor sockets. You maximize the processing from processor caches and minimize the latencies from reading from memory and, last, from disk.

Core-friendly parallelism consists of comprehensive algorithms that are designed to carefully place and align data that is likely to be revisited into the processor cache lines to maximize the hit rate to the processor cache, increasing the effectiveness of cache lines.

Parallel vector processing with multi-core parallelism, single instruction, and multiple data parallelism helps to improve performance and use available CPU resources better.

Several physical attributes of the server are listed:

- Queries on BLU tables are automatically parallelized.
- The power of multiple CPU cores is used fully.
- CPU cache efficiency is maximized to optimize cache lines.

Optimal memory caching

DB2 automatically adapts the way that it operates based on the organization of the table (row-organized or column-organized) that is being accessed.

BLU Acceleration uses several attributes to optimize the memory cache:

- New algorithms cache effectively in RAM (buffer pool).
- Data can be larger than RAM. You do no need to ensure that all data fits in memory.
- BLU Acceleration separates caching algorithms for BLU/OLTP data:
  - BLU: Scan-friendly caching that minimizes I/O
  - OLTP: Least recently used (LRU)-based page cleaning that reclaims buffer pool space without regard to future I/O.

BLU Acceleration includes a set of big data-aware algorithms for cleaning out memory pools that are more advanced than the typical LRU algorithms that are associated with traditional row-organized database technologies. These BLU Acceleration algorithms are designed from the bottom up to detect data patterns that are likely to be revisited, and to hold those pages in the buffer pool. These algorithms work with DB2 traditional row-based algorithms.

2.3.3 Predictive Analytics with SPSS

The ability to predict outcomes with a reasonable degree of confidence is of great strategic importance in current businesses. To address this need, the Statistical Package for the Social Sciences (SPSS) is available as part of the IBM Business Analytics solution portfolio.

IBM SPSS consists of a comprehensive set of tooling that uses data for decision-making. It is a solid product with over 40 years of presence in the market. It can be used to provide statistics, create predictive models, and deploy all of these analyses in your business.

The capability to predict is an advantage in your return on investment (ROI). You might have a good sense of your business or eventually several pre-built rules to help with decision making. However, it is not until you use analytics in decision management that you can, sustainably, choose the best path to follow, for every point of impact of your business.

Predictive analytics can capture insights from historic data patterns and provide you with evidence from this data. Predictive analytics is flexible in its modeling to understand changes in trends, and it can analyze massive amounts of structured and unstructured data. Also, it can promptly avail the information that is learned from the insights to everyone who needs access to it. It increases ROI because business areas are not analyzed independently.

Cross-departmental analyses are performed. These analyses do not leave out any business relationships that were uncovered or considered too small. As a result, every aspect of the business is optimized with these analyses.

IBM SPSS software is organized in product families:

- Statistics provides evidence that is based on data and verifies hypotheses.
- Modeling works out accurate predictions to aid decision making.
- Deployment helps you act upon the impact points in your operations.

The following sections describe an overview and highlight the products from each of these families.

Statistics

This family is represented by the IBM SPSS Statistics software suite. It is based on sophisticated mathematics to validate hypotheses and assumptions. It is widely used by government, commercial, and academic institutions to solve business and research problems. You can use it to test an opinion on a new product, predict the acceptance of ideas, experiment with allocation within a supply chain, or test the efficiency of a medical treatment. This software uses data to back up (or not) your theories. With this backup, you are more confident when you decide.

Modeling

This family is represented by the IBM SPSS Modeler software. The previous family, statistics, is used to test hypotheses. With the SPSS Modeler software, you can create business models to predict future outcomes. It uncovers hard-to-identify relationships (within structured and unstructured data) that seem unrelated at first. You can predict the future and understand what happens based on what happened before. This capability is useful to prevent customer churn, for example, and to help people consistently make decisions. Another benefit of this software is that it can also explain the factors that drive future outcomes. You can use it to mitigate risks and take advantage of opportunities.

With the SPSS Modeler, you can create models in an intuitive and quick fashion, without programming. The SPSS Modeler includes the advanced, interactive visualization of models. Multiple techniques can be included within a model, and the results are easy to understand and communicate to your staff.

Deployment

This family is represented by the following software:

- IBM SPSS Decision Management
- IBM SPSS Collaboration and Deployment Services

The first software, SPSS Decision Management, is intended to automate and optimize small decisions that are made in day-to-day business operations in real time. It combines predictive analytics with business rules. The models are created in an easy-to-use interface with which the business user interacts without the specialized help of an analyst, statistician, or data miner. This more independent process allows people at any level of the organization to create automated models for making small decisions, helping to optimize every aspect of the overall business operation.

The second software, SPSS Collaboration and Deployment Services, enables widespread use and deployment of predictive analytics. It provides a centralized, secure, and auditable placeholder of analytical assets and advanced capabilities for management and control of predictive analytic processes, and sophisticated mechanisms to deliver the results of these assets to users.

IBM SPSS Collaboration and Deployment Services architecture

Figure 2-8 illustrates the typical architecture of an SPSS Collaboration and Deployment Services deployment, which consists of these components:

- The central SPSS Collaboration and Deployment Services repository
- The database server
- Execution servers
- Client servers that access the SPSS Collaboration and Deployment Services repository

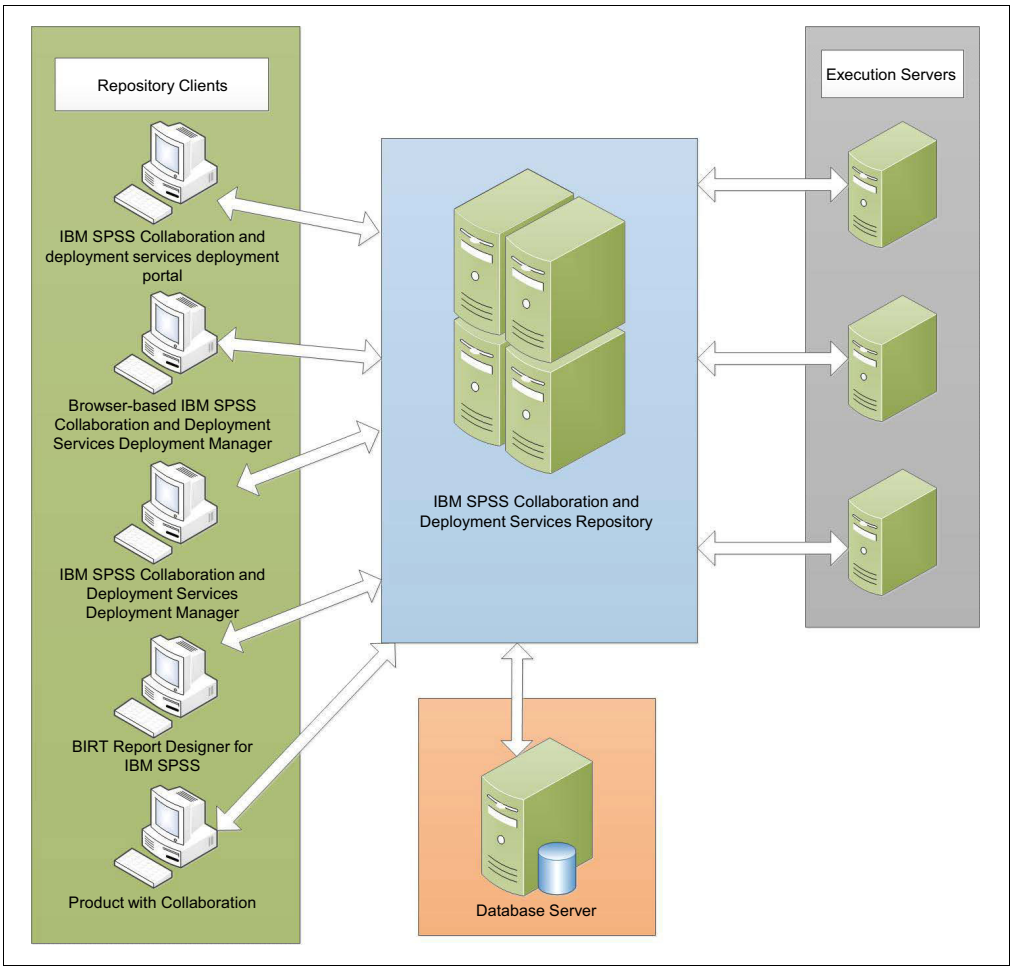


Figure 2-8 SPSS Collaboration and Deployment Services architecture

In this architecture, all of the analytics assets are stored on the SPSS Collaboration and Deployment Services repository. Clients can access these assets through web services or by using specialized client tooling for communications.

The requests that are performed by these clients are sent to execution servers, which perform all of the work on top of the analytics data. The results are then stored in the SPSS Collaboration and Deployment Services, and they can be accessed by the requesting clients.

The following list describes the components that are presented in the architecture of Figure 2-8 on page 28:

- IBM SPSS Collaboration and Deployment Services Repository

  This component is used for collecting and storing analytical assets. It includes models and data at a centralized location.

- IBM SPSS Collaboration and Deployment Services Deployment Manager

  This component is responsible for creating, executing, and automating the analytical task. It includes updating the model that is stored in the repository by users.

- IBM SPSS Collaboration and Deployment Services Deployment Portal

  This web browser-based thin-client interface accesses the IBM SPSS Collaboration and Deployment Services Repository, runs analyses, and views output.

- BIRT Report Designer for IBM SPSS

  Ad hoc reports against relational and file-based data sources can be created by using BIRT Report Designer for IBM SPSS.

- IBM SPSS Collaboration and Deployment Services Enterprise View Driver

  Use this component to access IBM SPSS Collaboration and Deployment Services Enterprise View objects that are stored in the repository, including IBM SPSS Statistics and third-party applications.

- Browser-based Deployment Manager

  This component is used by the SPSS administrator to perform, tune, and update system management tasks.

2.3.4 Reporting insights with IBM Cognos Business Intelligence

Organizations are constantly under pressure to understand and react quickly to new information. In addition, the complexity and volumes of data for all aspects of the environments in which organizations operate are increasing. Markets, regulatory environments, customer and supplier data, competitive information, and internal operational information all affect how data is viewed and interpreted. It is imperative for organizations to react correctly, dynamically, and in a timely fashion to answer key business questions and to outperform the competition.

From business intelligence to financial performance and strategy management to analytics applications, Cognos software can provide what your organization needs to become top-performing and analytics-driven. With products for the individual, workgroup, department, mid-sized business, and large enterprise, Cognos software is designed to help everyone in your organization make decisions that achieve better business outcomes for now and in the future.

Cognos Business Intelligence features

Cognos Business Intelligence provides the following features:

- Reports

  Cognos Business Intelligence software helps ensure that users are equipped with the reports they need to make fact-based decisions in a system that is simpler, faster, and easier to manage. From professional report authors who design one-to-many reports for the enterprise, to business users who need to create their own ad hoc queries or customize existing reports, Cognos Business Intelligence reporting capabilities fit the needs of users throughout your organization.

- Analysis

  With the analytics capabilities of Cognos Business Intelligence software, users can explore information and different perspectives easily and intuitively to ensure that they are making the correct decisions. General business users can easily view, assemble, and analyze the information that is required to make better decisions. Additionally, business and financial analysts can take advantage of more advanced, predictive, or what-if analysis capabilities.

- Scorecards

  Scorecards enable your organization to capture corporate strategy and communicate that strategy at the operational level. Executives and managers can define quantifiable goals and targets and track performance for business units, operating subsidiaries, and geographic regions to quickly identify the areas that need attention.

- Dashboards

  With dashboards, users can access, interact, and personalize content in a way that supports the unique way that they make decisions. Security-rich access to historic, current, and projected data means that users can quickly move from insight to action.

- Statistics

  Statistics capabilities help you incorporate statistical results with core business reporting, reducing the time that it takes to analyze data and prepare business presentations that are based on that analysis.

- Mobile business intelligence

  Mobile business intelligence capabilities enable your mobile workforce to interact with information in new ways by delivering relevant business intelligence wherever the workers are. Users interact with trusted business intelligence through a rich and visual experience, whether offline or online. The flexible platform ensures that mobile decision making is simple, reliable, and safe.

- Real-time monitoring

  Real-time monitoring capabilities provide your employees on the leading edge with a rich view of operational KPIs[1] and measures while they occur to support up-to-the-moment decision making.

[1] Key performance indicators (KPIs) are used by companies to better evaluate their current level of business success and to help plan for the future.

- Collaboration

  Collaboration capabilities help individuals, key stakeholders, workgroups, and teams align their strategic objectives, build stronger relationships, learn from history, and use resources for important decision making effectively.

- Planning and budgets

  Get the right information to the right people in the form they need it to plan, budget, and forecast. Planning and budgeting capabilities in the IBM Cognos Business Intelligence software support a wide range of requirements, from high-performance, on-demand customer and profitability analysis and flexible modeling to enterprise contribution for a broad range of users.

Cognos Business Intelligence components overview

IBM Cognos Business Intelligence is an integrated business intelligence suite that provides a wide range of functionality to help you understand the data of your organization. Everyone in your organization can use Cognos Business Intelligence to view or create business reports, analyze data, and monitor events and metrics so that they can make effective business decisions.

Cognos Business Intelligence integrates the following business intelligence activities in one web-based solution, as shown in Table 2-1.

Table 2-1 IBM Cognos Business Intelligence list of components

Activity	Component
Publishing, managing, and viewing content	IBM Cognos Connection
Interactive workspaces	IBM Cognos Business Insight™
Simple reporting and data exploration	IBM Cognos Business Insight Advanced
Ad hoc querying	IBM Cognos Query Studio
Managed reporting	IBM Cognos Report Studio
Event management and alerting	IBM Cognos Event Studio
Scorecarding and metrics	IBM Cognos Metric Studio
Analyzing your business	IBM Cognos Analysis Studio
Working with IBM Cognos Business Intelligence content in Microsoft Office	IBM Cognos for Microsoft Office

Before you use IBM Cognos Business Intelligence, you must understand how each of the components that make up the IBM Cognos Business Intelligence user interfaces can help you perform your job.

Cognos transaction flow

You can explore the transaction flow through the Cognos environment, which is primarily driven by HTTP requests. The IBM Cognos environment consists of three components: the gateway module, the authentication and authorization module, and the report execution model, as shown in Figure 2-9.

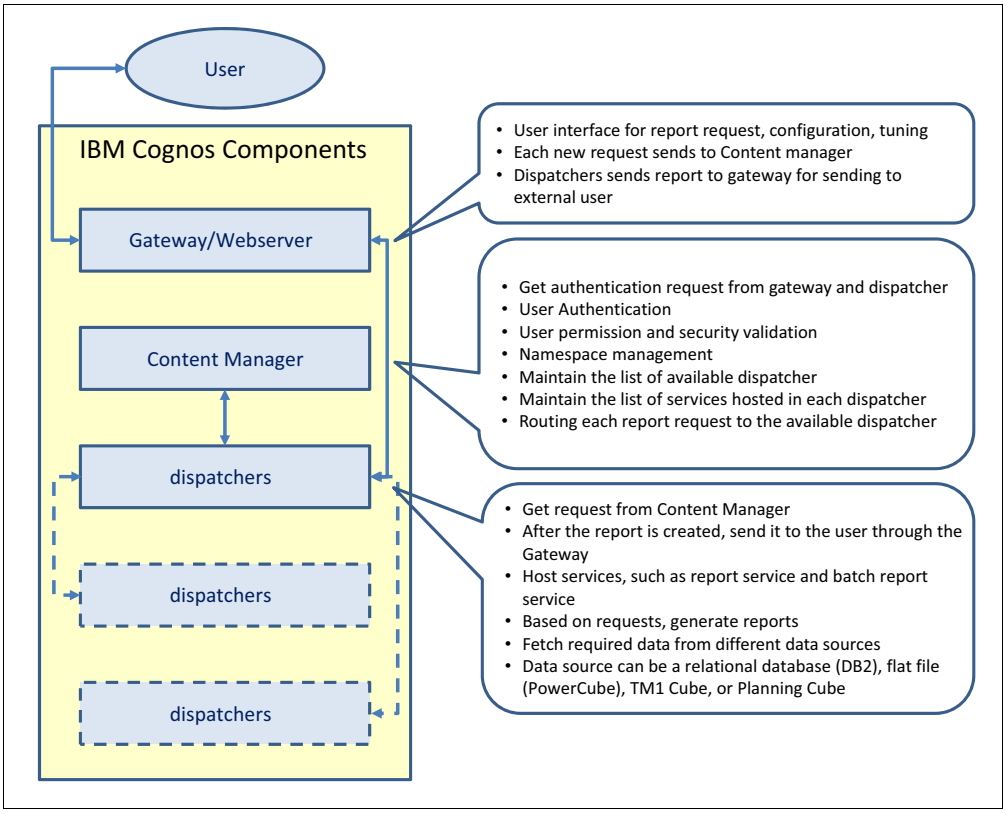


Figure 2-9 Cognos transaction and workflow diagram

The three primary components in a Cognos deployment are listed:

- Cognos Gateway

  The Cognos Gateway is the primary user interface and is accessed through the Cognos Connection and portal URL. User authentication and namespace validation happen at this layer. However, internally, the Gateway communicates with the Content Manager through the Dispatcher for authentication and validation. After successful authentication, a request is sent to the Content Manager for further processing. After the report execution completes, it reaches the Gateway and it is sent to the client.

- Cognos Content Manager

  In an IBM Cognos environment, the main controlling unit is the Content Manager. It performs several important functions. Every Cognos environment consists of only one primary Content Manager. However, a secondary Content Manager can be configured. The secondary Content Manager takes over as primary only when the primary Content Manager is unavailable. Even in a busy environment, all requests and transactions are handled by the single primary Content Manager.

- Application tier and dispatcher

  The dispatcher is the main working thread in the Cognos environment that is used to generate reports that are based on user requests. Each dispatcher hosts several services, including the Presentation Service, which is used in report generation. The dispatcher fetches the required data from the different data sources, and it renders this data in the report based on the specification. Each dispatcher creates multiple business intelligence business processes, and each process handles one request at a time.

2.3.5 Spectrum Scale and File Placement Optimizer

IBM Spectrum Scale is software-defined storage for high-performance, large-scale workloads on-premises or in the cloud. Built on the award-winning IBM General Parallel File System (GPFS), this scale-out storage solution provides file, object, and integrated data analytics for the following areas:

- Compute clusters (technical computing)
- Big data and analytics
- Hadoop Distributed File System (HDFS)
- Private cloud
- Content repositories
- File placement optimization

Spectrum Scale File Placement Optimizer (Spectrum Scale-FPO) is a high-performance, cost-effective storage methodology that started as a clustered file system and evolved into more than a file system. Today, Spectrum Scale is a full-featured set of file management tools, including advanced storage virtualization, integrated high availability, and automated tiered storage management, and offers the performance to effectively manage large quantities of file-based data.

Spectrum Scale supports various application workloads, and it is effective in large and demanding environments. Spectrum Scale is installed in clusters, and it supports big data, analytics, gene sequencing, digital media, and scalable file serving. All indications are that BigInsights for Apache Hadoop might bring more unstructured and file-based data into the application.

For high-performance computing environments, IBM Spectrum Scale offers a distributed, scalable, reliable, and single namespace file system. Spectrum Scale-FPO 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 the 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 it is not Portable Operating System Interface (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 unaffected by the impact that is incurred by HDFS as a secondary file system, running within a Java virtual machine (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 you use 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 the data nodes. From an infrastructure design perspective, including Spectrum Scale-FPO can reduce the number of required management nodes.

Because Spectrum Scale-FPO distributes metadata across the cluster, no dedicated name service is needed. Management nodes within the BigInsights for Apache Hadoop predefined configuration or 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 more data nodes within a rack.

For more information about implementing IBM Spectrum Scale-FPO in an InfoSphere® BigInsights solution, see the *Deploying a big data solution using IBM Spectrum Scale-FPO* white paper at the following website:

http://ibm.co/1NBnGTj

2.3.6 Cluster management

When you manage the Power System nodes in your big data and analytics environment, you can choose from alternatives, such as the Hardware Management Console (HMC) or Platform Cluster Manager - Advanced Edition.

The HMC is the traditional way to manage Power Systems servers. It provides a graphical user interface (GUI) to perform management tasks, such as configuring hardware alerts and creating and configuring LPARs.

You can also use IBM Platform Cluster Manager - Advanced Edition to manage an environment of clusters. It provides the following benefits:

- Management of multitenancy 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 BigInsights for Apache Hadoop cluster, but this book might also be used to automate the deployment of other solutions, such as solutions that are based on IBM Symphony, IBM Load Sharing Facility (LSF®), GridEngine, PBS Pro, and open source Hadoop.

- On-demand and self-service provisioning

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

- Increased server consolidation

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

One of the parts that integrate the software architecture of a Platform Cluster Manager - Advanced Edition 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 the Platform Cluster Manager - Advanced Edition Version 4.2 installation, xCAT is installed on the node. Part of the xCAT configuration is automatically performed during this step. However, you can use xCAT commands to further configure or reconfigure your cluster provisioning environment. As of the writing of this book, several 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
- Optional: Using the Platform Cluster Manager environment as a Dynamic Host Configuration Protocol (DHCP) server to the FSP hardware management network.

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.

2.4 Solution reference architecture

This section describes a big data and analytics solution deployment scenario and its reference architecture to illustrate possible implementations on an enterprise business operation.

2.4.1 Solution scenario overall topology

In this publication, big data and analytics tools were used to implement a business scenario, which can be used in many companies, such as retail companies, that sell products to their customers by using a webstore or e-commerce solution.

In this scenario, an executive from a fictional company can see a Cognos Business Intelligence dashboard that consolidates sales information and results for the company and combine that information with customer social media sentimental analysis to produce a marketing campaign to boost the company's sales. The dashboard contains the following information:

- Revenue trend by product category
- Revenue by product and year
- Gross profit by country
- Revenue social media sentiment by product

The Cognos Business Intelligence dashboard pulls revenue, product categories, and gross profit data from a DB2 with BLU Acceleration database, which might be the company's transactional system or even a data warehouse that might be fed from it.

The revenue social media sentiment is stored on a BigInsights for Apache Hadoop cluster, and the Big SQL technology is used to query this data and display the results on the dashboard. Before this data is stored, it can be captured by using tools, such as IBM Streams or the IBM Bluemix® with Insights for Twitter service.

Figure 2-10 shows how Cognos Business Intelligence consolidates the data into a single dashboard.

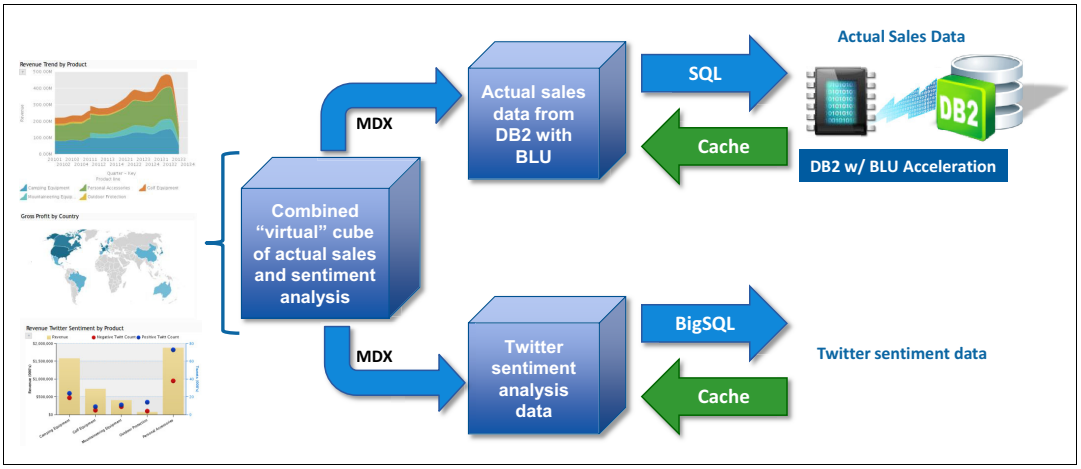


Figure 2-10 Cognos Business Intelligence consolidates data from different sources

By using this information, the company executive can check for products to target that need more sales based on decreased revenue year-to-year and that have positive social media feedback so that the company can define a campaign and a target audience for it.

Based on the executive decision and goals, a company's researcher first creates input data that contains an individual's demographic data (gender and geographic region), banking payment status, and sentiment polarity. This profile will be an input for the SPSS Modeler and Analytical Decision Manager to try to find the best scenario that will predict the most effective expected profit.

After those steps, SPSS Analytical Decision Management for Customer Interactions identifies the best offer for each customer and provides recommendations for offers to deliver to customers through a call center or an email. The offer will be available on a webstore when the customer logs in to the webstore.

Figure 2-11 illustrates the campaign solution process and data flow.

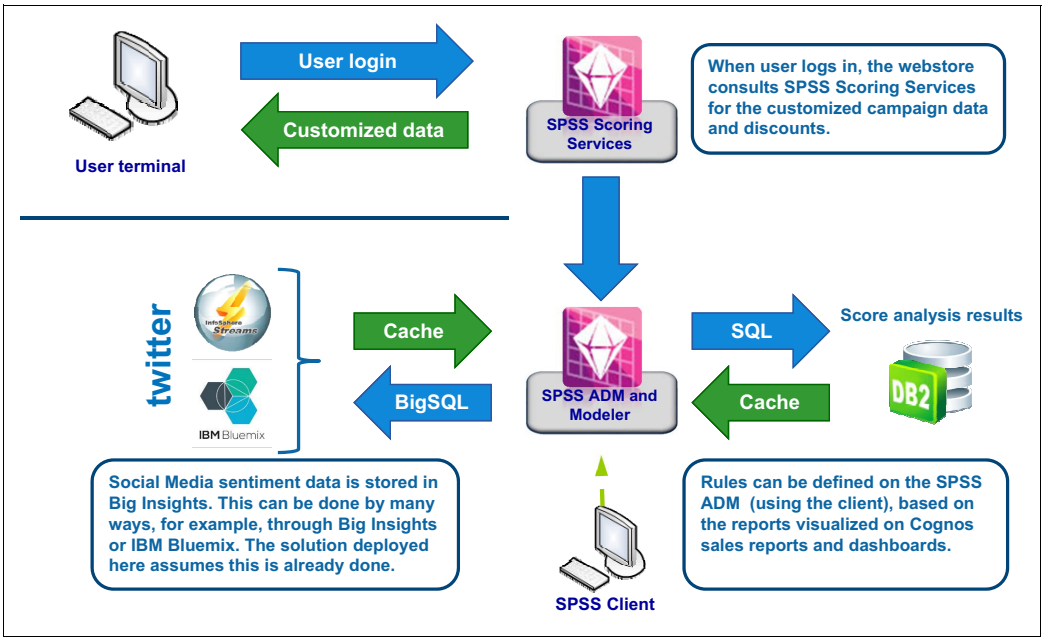


Figure 2-11 Campaign solution process and data flow

2.4.2 Solution scenario architecture

This solution used IBM Power Systems and IBM Data Engine for Analytics capabilities and benefits. To demonstrate IBM Power Systems flexibility, the implementation used AIX and Red Hat Enterprise Linux (RHEL) distributions to install the software components.

The following software component operating systems were installed to implement this solution:

- IBM DB2 with BLU Acceleration on an RHEL 7.1 little endian edition node.
- IBM Open Platform for Apache Hadoop on RHEL 7.1 little endian edition, through an IDEA cluster implementation, which consists of two management nodes (also used as edge nodes) and three data nodes.
- IBM SPSS Collaboration and Deployment Services, IBM SPSS Modeler, and IBM SPSS Analytical Decision Manager on an AIX 7.1 node.
- IBM Cognos Business Intelligence on an RHEL 7.1 big endian edition node.

Figure 2-12 shows the logical architecture for this solution scenario implementation.

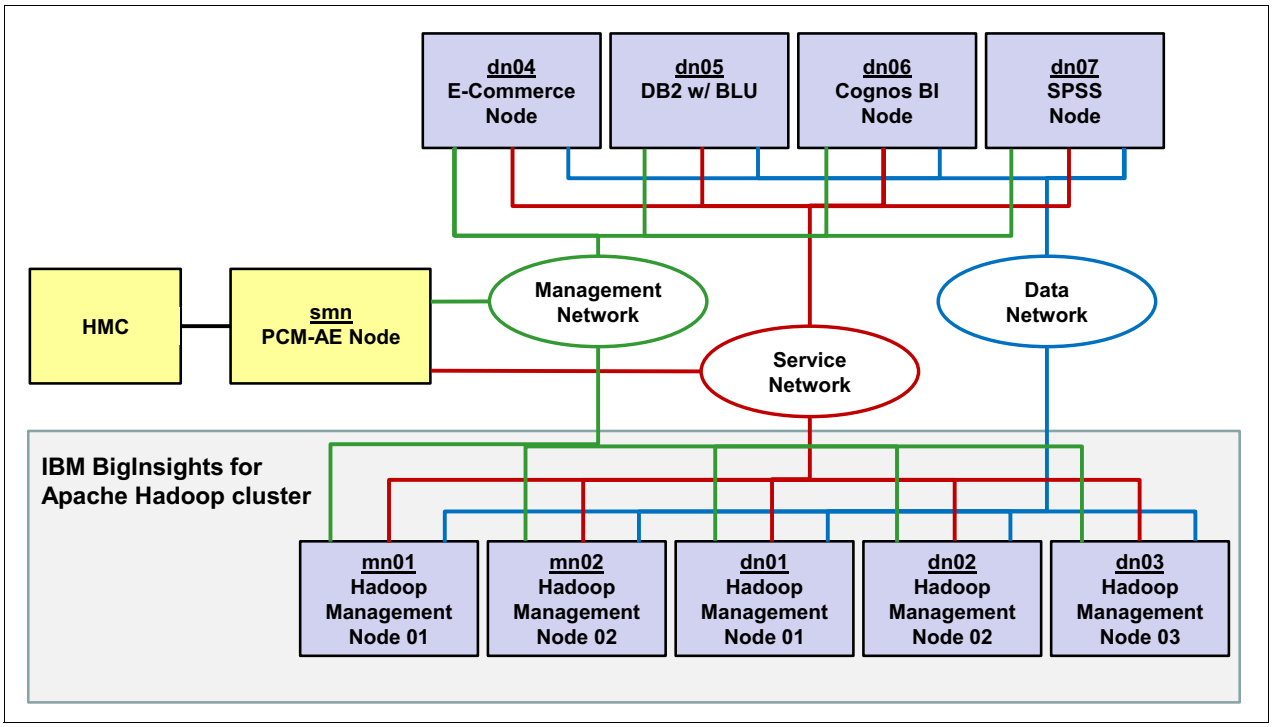


Figure 2-12 Solution scenario architecture

In this implementation, the HMC was used to manage only the System Management Node, which is used to deploy and manage the other nodes, connecting to the flexible service processors (FSPs) from the other nodes through the service network.

The management network has two major purposes: to provision the IBM Open Data Platform for Apache Hadoop, DB2, Cognos Business Intelligence, and SPSS nodes, as described in 2.2.4, “Networking” on page 12, and to perform administrative tasks.

All of the nodes, except the System Management Node, used the data network to send and receive application data, for example, when the Cognos Business Intelligence queries revenue data from the DB2 database or sentimental analysis data from the Open Data Platform for Apache Hadoop (connecting to the Big SQL Head).

The IBM Open Data Platform consists of two management nodes and three data nodes.

Figure 2-13 shows the hardware components that are used to implement the solution scenario that was described earlier in this section.

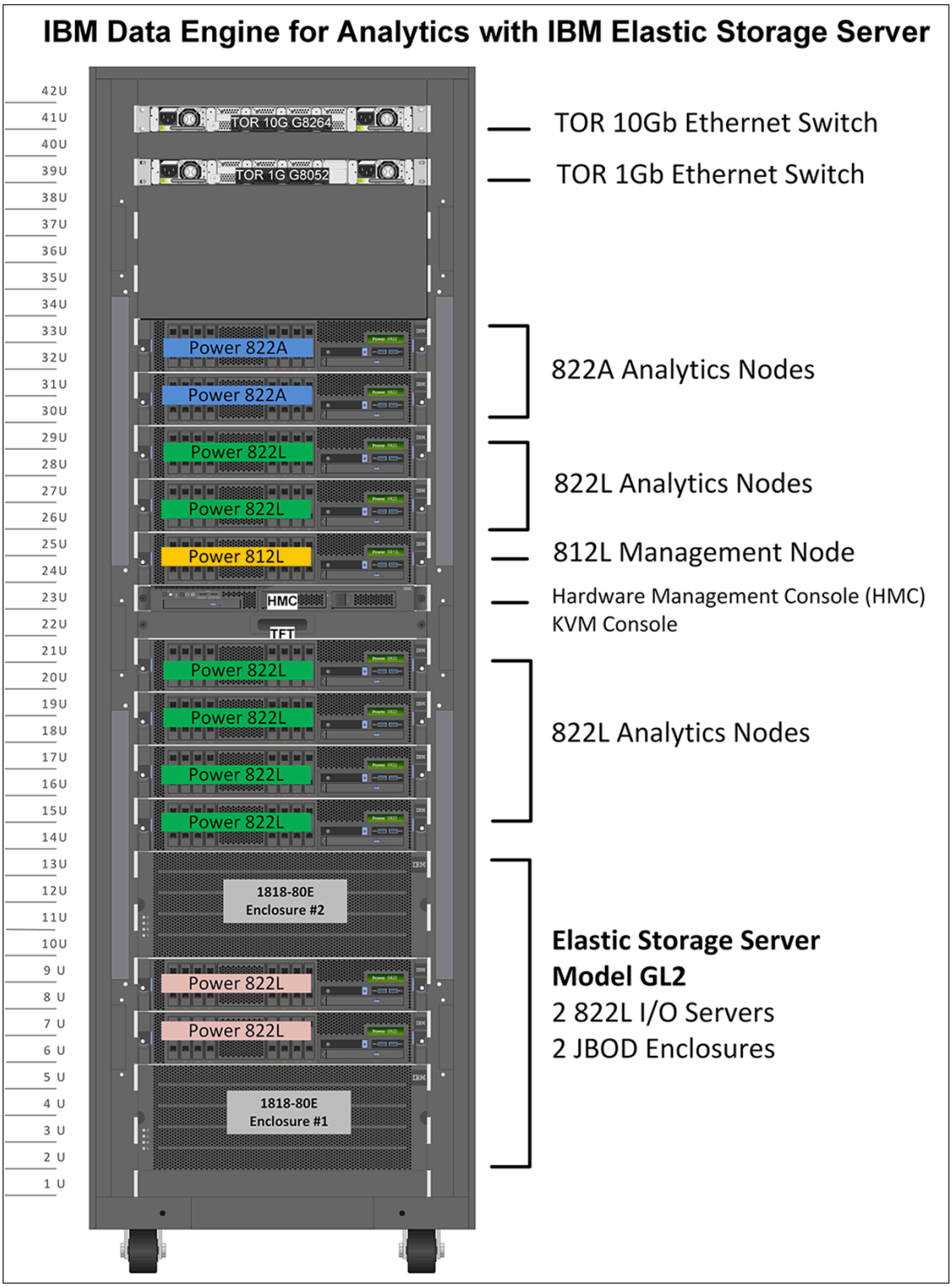


Figure 2-13 Hardware components that are used in the solution scenario implementation

For the IBM Data Engine for Analytics Components, the following configuration was used:

- Elastic Storage Server Model GL2 for storing data.
- Two Hadoop Management Node LPARs, each of which uses half of the resources of a Power S822L server. This server has the following configuration:
  - 24 x 3.02 GHz cores
  - 256 GB memory
  - 4 - 12 x 1.8 TB 10K 2.5-inch serial-attached SCSI (SAS) hard disk drives (HDDs)
  - Split backplane
  - 2 x 4-port 1-Gigabit Ethernet adapter
  - 2 x dual-port 10-Gigabit Ethernet adapter
  - 2 x partitions with equal resources on a server
- Three analytic nodes, which are used as Hadoop data nodes, with each node as a full partition on a Power S822L with the following configuration:
  - 24 x 3.02 GHz cores
  - 256 GB Memory
  - 2 x 4-port 1-Gigabit Ethernet adapter
  - 2 x dual-port 10-Gigabit Ethernet adapter
- One System Management Node (that uses Platform Cluster Manager - Advanced Edition administration capabilities) as a full partition on Power S812L with the following configuration:
  - 10 x 3.425 GHz cores
  - 32 GB memory
  - 2 x 1.8 TB 10K 2.5-inch SAS HDDs
  - 2 x 4-port 1-Gigabit Ethernet adapter
  - 2 x dual-port 10-Gigabit Ethernet adapter

For the other analytic applications, the following configuration was used:

- One Power S822 server for the SPSS node, with the following configuration:
  - 24 x 3.02 GHz cores
  - 256 GB Memory
  - 4 x 1.8 TB 10K 2.5-inch SAS HDDs
  - 2 x 4-port 1-Gigabit Ethernet adapter
  - 2 x dual-port 10-Gigabit Ethernet adapter
- Four Power S822L servers (one for each of the remaining nodes) with the following configuration:
  - 24 x 3.02 GHz cores
  - 256 GB Memory
  - 4 x 1.8 TB 10K 2.5-inch SAS HDDs
  - 2 x 4-port 1-Gigabit Ethernet adapter
  - 2 x dual-port 10-Gigabit Ethernet adapter

For LAN networking, two switches were used:

- IBM System Networking RackSwitch G8052

  1 GbE top-of-rack switch for the management and service networks

- IBM System Networking RackSwitch G8264

  1 GbE top-of-rack switch for the data networks

This configuration reflects the hardware that was available for this solution scenario deployment. The sizing for the server can vary according to the workload that the implementation will support and any special network requirements.

For more information about the installation of the software and hardware components, see Chapter 4, “Scenario: How to implement the solution components” on page 49. The details about the software component integration are in Chapter 5, “Scenario: Integration of the components for the solution” on page 133.

3

IBM POWER8 for analytics workloads

This chapter discusses how IBM POWER8 can help clients with their analytic workload requirements, and it describes the advantages that this solution delivers to help clients manage and understand data to gain competitive advantages.

This chapter contains the following topics:

- Value proposition
- Advantages

3.1 Value proposition

A high-performance infrastructure that is scalable for varying workloads, highly available, and optimized for price performance is critical for getting the most out of the analytics platform and deriving timely insight. It helps infuse analytics seamlessly to drive better business outcomes and allows companies to build on what they already have while they add new capabilities as their needs grow.

IBM Solution for Analytics - Power Systems Edition delivers a solution for business intelligence and predictive analytics. It is flexible and integrated, and it provides options to pre-load and configure one or more IBM analytics applications with data warehouse acceleration on a POWER8 processor-based server.

These analytics tools, with many other tools in the IBM analytics suite, help turn big data into actionable insights to help companies address customer retention and growth, IT costs, management of security risks, counter fraud techniques, optimization of IT operations, and digital innovation opportunities.

3.1.1 Data processing capacity (more data per second)

IBM Power Systems help clients accelerate results for analytics applications. Imagine the possibilities that fast, efficient analytics can open for you. How many more questions can you ask if you can get answers immediately? To improve your business' time to insight, you need an analytics infrastructure that is easy to set up, that is flexible and scalable, and that offers powerful analytics applications.

Helping you make the best decision for your business means understanding what is happening, why it is happening, what can happen in the future, and what you need to do fast and efficiently. Therefore, having the data processing capacity, accelerating the knowledge with this infrastructure, and helping you to deliver faster business outcomes with predictive analytics can provide you with the solution to meet your constant business needs.

3.1.2 More density for same workload

To provide an infrastructure for big data and analytics, the IBM hardware and software teams work together to optimize the entire solution stack for data and analytics. The suite of capabilities that are built on IBM POWER8 can help you develop a thorough plan for addressing your big data and analytics needs.

By integrating data environments across your business, building on a proven infrastructure as IBM Power Systems, using storage solutions that maximize data efficiency, and taking advantage of the latest and best data management and analytics tools, you can accelerate your insights while you reduce costs. As a foundation of the IBM comprehensive big data and analytics portfolio, Power Systems continues to be a key driver in pushing your business forward in the race toward digital transformation with predictive analytics.

3.1.3 More resiliency for the data

With an infrastructure, such as the IBM POWER8 that is tuned for big data and analytics, you can build analytics into operational processes. This POWER8 infrastructure can scale effortlessly while it enables highly secure, shared access, resiliency of the information, and generate insights that are needed to make better decisions.

The POWER8 infrastructure helps to provide continuous availability to address more users, more concurrency, changing demand, and resiliency of the data.

For example, the IBM Data Engine for Analytics provides a customizable, pre-integrated infrastructure solution that brings together the innovative capabilities of IBM Power Systems (massively parallel compute capacity for a fast time to get the result), IBM Elastic Storage Server (software defined storage for fast access to petabytes of data), and IBM Platform Computing (intelligent cluster automation and management for ease of deployment and operations).

This solution can be combined for faster time to value with big data and analytics software, such as IBM InfoSphere BigInsights for Hadoop for analytics workloads, IBM InfoSphere Streams for high ingest streaming analytics, and IBM Watson™ Explorer analytical components for advanced natural language processing of unstructured data.

3.2 Advantages

The suite of capabilities that are built on IBM POWER8 can help businesses develop a thorough plan for addressing their big data and analytics needs. IBM offers a broad set of analytics capabilities built on the proven foundation of a single platform, IBM Power Systems. Power Systems is an open, secure, and flexible platform that is designed for big data. It has massive input/output (I/O) bandwidth to deliver analytics in real time, and it can provide the capabilities that are needed to handle the varying analytics initiatives that each business requires.

3.2.1 POWER8 memory bandwidth

Transactional memory (TM) is a shared-memory synchronization construct that allows process-threads to perform sequences of storage operations that appear to be atomic to other process-threads and applications. This capability allows optimistic execution as a means to take advantage of the inherent parallelism that is found in the latest generation of Power Systems.

One of the main uses of TM is the speed-up of lock-based programs by using the speculative execution of lock-based critical sections (CSs) without first acquiring a lock. This function allows applications that were not carefully tuned for performance to take advantage of the benefits of fine-grain locking. The transactional programming model also provides productivity gains when you develop lock-based shared memory programs.

Applications can also use TM to checkpoint and restore the architectural state, independently of the atomic storage access guarantees that are provided by TM.

3.2.2 POWER8 I/O bandwidth

The POWER8 processor benefits from the next generation Peripheral Component Interconnect Express (PCIe) Gen3 and shorter hardware paths to the adapters. The POWER8 chip design includes on-chip PCIe buses, which means fewer hardware delays and lower latency to PCIe slots. This design results in lower latency for networking and storage protocols, reducing latency by over 1 µs from the previous I/O hub-based POWER7 Systems.

The PCIe Gen3 x16 bus increases the PCIe bus peak bandwidth to 112 Gbps (for a single x16 bus), which is about four times the bandwidth in previous POWER PCIe Gen2 x8 slots in POWER7 and IBM POWER7+™ processor-based systems. Each processor module or socket provides two or four PCIe buses, depending on the system model. Several buses are x16 and others are x8. Depending on the system model, several buses can connect to PCIe I/O hub chips on the system board.

These PCIe slots are for slower-speed adapters so they can share a single higher-speed bus. In addition, for systems that support I/O drawers, the PCIe buses in the I/O drawer all use a PCIe I/O hub. For applications that are sensitive to latency or those applications that require high bandwidth or high message rates, the adapters that are used by the application need to use, when possible, the PCIe slots in the chip where the application runs.

In all cases, bandwidth must be the same no matter where in a POWER8 processor-based system an adapter is plugged. However, for latency, small increases are possible if the adapter is in a different socket, node, or central processor complex drawer.

As a rule, high-speed and low-latency adapters must be placed in the direct PCIe slots on each socket. The PCIe slots behind I/O hubs and in I/O drawers must be used for lower bandwidth and non-latency sensitive adapters. Many adapters are restricted to certain slots. Consult a Power Systems PCI Adapter Placement Guide for a specific adapter.

3.2.3 POWER8 performance

The POWER8 processor contains many new and important performance features, such as support for eight hardware threads in each core and support for transactional memory. The POWER8 processor is a strict superset of the IBM POWER7+ processor, and so all of the performance features of the POWER7+ processor, such as multiple page sizes, also appear in the POWER8 processor.

For more information about POWER performance, see the publication*Performance Optimization and Tuning Techniques for IBM Power Systems Processors Including IBM POWER8*, SG24-8171, at the following website:

http://www.redbooks.ibm.com/abstracts/sg248171.html

3.2.4 IBM Spectrum Scale advantages

IBM Spectrum Scale is software-defined storage for high-performance, large-scale workloads on-premises or in the cloud. Built on the award-winning IBM General Parallel File System (GPFS), this scale-out storage solution provides file, object, and integrated data analytics for the following areas:

- Compute clusters (technical computing)
- Big data and analytics
- Hadoop Distributed File System (HDFS)
- Private cloud
- Content repositories
- File Placement Optimization (FPO)

IBM Data Engine for Analytics is a customizable, pre-integrated infrastructure solution that brings together the innovative capabilities of IBM Power Systems (massively parallel compute capacity for fast time to result), IBM Elastic Storage Server (software defined storage for fast access to petabytes of data), and IBM Platform Computing (intelligent cluster automation and management for ease of deployment and operations).

For faster time to value, big data and analytics software, such as IBM Open Platform with Apache Hadoop, IBM InfoSphere Streams for high ingest streaming analytics, and IBM Watson Explorer analytical components for advanced natural language processing of unstructured data, can be preinstalled and configured before shipment to the client site.

For additional information about IBM Spectrum Scale and the feature that gives analytics implementations a push, see 2.3.5, “Spectrum Scale and File Placement Optimizer” on page 33.

For more information about IBM Spectrum Scale, see the *IBM Spectrum Scale (formerly GPFS)*, SG24-8254, at the following website:

http://www.redbooks.ibm.com/abstracts/sg248254.html

4

Scenario: How to implement the solution components

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

This chapter provides instructions to install and configure the IBM BigInsights added value services, DB2 BLU, Statistical Package for the Social Sciences (SPSS), and Cognos software.

This chapter covers the following topics:

- Basic infrastructure requirements
- Using Ambari to deploy BigInsights with Spectrum Scale
- DB2 with BLU Acceleration to store structured data
- SPSS Analytical Decision Management
- Cognos for Dashboarding

4.1 Basic infrastructure requirements

For this book, we selected the IBM Open Data Platform on the IBM Data Engine for Analytics (IDEA) solution for our implementation. This solution uses IBM Platform Cluster Manager – Advanced Edition (PCM AE) V4.2.1, with its Extreme Cluster Administration Toolkit (xCAT) function to deploy a cluster of IBM POWER8 servers.

The factory installation process results in a ready-to-run environment for analytic workloads that uses the IBM Open Platform edition of BigInsights V4.1, with IBM Spectrum Scale and IBM Platform Symphony, which comes from the Enterprise Manager component. The IDEA solution is based on the IBM POWER8 S812L and S822L servers, including an Elastic Storage Server (ESS) building block.

Therefore, for this book, we do not cover the hardware setup, PCM installation, or operating system deployment. However, expanding on the solution, we provide instruction to install the IBM BigInsights value-add services, DB2 BLU, SPSS, and Cognos. From the hardware perspective, we added additional POWER8 S822A nodes to support the analytics and e-commerce software. The steps that are detailed in this section can also be applied to other POWER8 environments. For more information, see Chapter 2, “Solution reference architecture” on page 5.

Throughout this chapter, we reference the following hosts. as shown in Table 4-1.

Table 4-1 Hosts that are referenced throughout this chapter

Host name	Hardware	Operating system	Software
smn	Power 812L	Red Hat Enterprise License (RHEL) 7.1	ESS and PCM
mn01	Power 822L logical partition (LPAR)	RHEL 7.1le	IBM Open Platform Manager and Ambari
mn02	Power 822L LPAR	RHEL 7.1le	IBM Open Platform Manager
dn01	Power 822L	RHEL 7.1le	IBM Open Platform Worker
dn02	Power 822L	RHEL 7.1le	IBM Open Platform Worker
dn03	Power 822L	RHEL 7.1le	IBM Open Platform Worker
dn04	Power 822L	RHEL 7.1le	E-commerce
dn05	Power 822L	RHEL 7.1le	DB2
dn06	Power 822A	RHEL 7.1	Cognos
dn07	Power 822A	AIX 7.1	IBM WebSphere® and IBM SPSS
ess01	Power 822L	RHEL 7.1	ESS
ess02	Power 822L	RHEL 7.1	ESS

4.2 Using Ambari to deploy BigInsights with Spectrum Scale

This section covers the complete installation of the IBM Open Platform Edition of BigInsights Version 4.1 by using Ambari. If you use a pre-existing installation of the IBM Open Platform software, such as an IDEA solution, skip to 4.2.5, “Installing the BigInsights value-add packages” on page 73 for information about installing the BigInsights value-add packages.

4.2.1 Understanding supported deployment approaches, including Spectrum Scale with Ambari

Multiple supported deployments for Spectrum Scale with Ambari need to be considered. This section covers an installation of IBM Open Platform Edition on a new cluster (option 1) that uses Spectrum Scale File Placement Optimizer (FPO) technology as an alternative to the factory-installed configuration that is provided with the IDEA solution (option 2) on which the remainder of this section is based. Follow these steps:

1. Create a Spectrum Scale cluster.

   Use Ambari to create a new Spectrum Scale cluster by using FPO technology[1]:

   – Create and configure the new Spectrum Scale cluster, including the designation of the manager and quorum node roles.

   – Set basic Spectrum Scale configuration parameters for your environment, for example, pagepool, maxFilesToCache, maxStatCache, and worker1Threads.

   – Create Network Shared Disks (NSDs) and the file system by using basic or advanced configuration methods.

   – Install Open Platform components and configure the Spectrum Scale Hadoop Connector.

2. Add new BigInsights nodes to an existing Spectrum Scale cluster.

   Use Ambari to add new nodes into an existing File Placement Optimizer (FPO) or Elastic Storage Server (ESS) cluster. The installer will perform these tasks:

   – Install Spectrum Scale on the new nodes.

   – Add the new nodes to the existing cluster Spectrum Scale cluster.

   – Ambari will not create any NSDs or file systems. If this cluster is an existing FPO cluster, new NSDs will need to be added manually.

   – Install Open Platform components and configure the Spectrum Scale Hadoop Connector.

3. Add BigInsights to existing Spectrum Scale cluster nodes.

   Use Ambari to deploy BigInsights on a pre-existing cluster. In this case, Spectrum Scale is installed in advance and configured on all nodes by using FPO or ESS technology. The Spectrum Scale Hadoop Connector might not be configured yet.

   The installer will install Open Platform components and configure the Spectrum Scale Hadoop Connector.

[1] GPFS Release for Enterprise Manager from http://ibm.co/1RYHeG1

4.2.2 Download software

This section provides details about how to download the software. Follow these steps:

1. We suggest that you create a mirror of the IBM hosted repository on a machine within your enterprise network. With this approach, you instruct Ambari to use that local repository rather than the repository that is hosted in the IBM cloud. This approach is the preferred approach when internet access is restricted, or when you use Spectrum Scale, which requires local repositories. For additional repository approaches, see the following website:

   http://ibm.co/1nmkEuz

2. For this deployment example, we took advantage of a local http server that was already set up by xCAT in our PCM system management node (smn) to use as a local mirror yum repository.

   Your http server must contain a directory for each of the following repositories, as shown in Figure 4-1.

```
[root@smn ~]# ls /install/repos/
Ambari  GPFS  IOP  IOP-UTILS
```

Figure 4-1 Repository directories

3. Download the following tar archives for the IBM Open Platform repository by using wget:
   - Ambari:

     https://ibm.biz/Bd4SHi
   - IOP:

     https://ibm.biz/Bd4SHK
   - IOP-UTILS:

     https://ibm.biz/Bd4SHG

4. Extract the three repository archives in the `repos` directory on the smn node, as shown in Figure 4-2.

```
cd /install/repos
tar xzvf <path to downloaded tar archives>
```

Figure 4-2 Extracting the repository tar archives

5. Obtain the base installation package files for IBM Spectrum Scale Advanced 4.1.1 Linux POWER8, which can be downloaded from the IBM Passport Advantage® website. Check with your IBM marketing representative or your support team for the URL of this website.
6. Place all of the Spectrum Scale 4.1.1 packages in your GPFS `repo` directory. Ensure that you remove the gpfs.hadoop-connector rpms and obtain the latest connector package. Download it from this website:

   http://ibm.co/1RYItoF

7. You also need to obtain the gpfs-ambari integration package `gpfs.ambari-iop_4.1-1.noarch.bin` from the http://ibm.co/1RYItoF website and place it in `/tmp/` on the management node where you plan to deploy Ambari-server. However, this package will not be a part of the repository for GPFS.

8. Create the local repository metadata, as shown in Figure 4-3.

```
[root@smn ~]# cd /install/repos/GPFS/rhel/7/ppc64le/4.1.1/
[root@smn 4.1.1]# ls
gpfs.base-4.1.1-0.ppc64le.rpm    gpfs.gpl-4.1.1-0.noarch.rpm
gpfs.crypto-4.1.1-0.ppc64le.rpm  gpfs.gskit-8.0.50-40.ppc64le.rpm
gpfs.docs-4.1.1-0.noarch.rpm     gpfs.hadoop-connector-2.7.0-2.ppc64le.rpm
gpfs.ext-4.1.1-0.ppc64le.rpm     gpfs.msg.en_US-4.1.1-0.noarch.rpm
[root@smn 4.1.1]# createrepo .
```

Figure 4-3 Creating the local GPFS repository metadata

9. Test your local repository by browsing the web directory: `http://smn/install/repos`.

4.2.3 Set up and install the Ambari server

Follow these steps:

1. On the management node where you plan to deploy Ambari-server, configure access to the new repository, as shown in Figure 4-4.

```
[root@mn01-dat ~]# cd /etc/yum.repos.d
[root@mn01-dat yum.repos.d]# vi ambari.repo
[root@mn01-dat yum.repos.d]# cat ambari.repo
[BI_AMBARI-2.1.0]
name=ambari-2.1.0
baseurl=http://smn/install/repos/Ambari/rhel/7/ppc64le/2.1.x/GA/2.1/
enabled=1
gpgcheck=0
```

Figure 4-4 Configure access to the Ambari repository

2. Install the Ambari server, as shown in Figure 4-5.

```
[root@mn01-dat ~]# yum -y install ambari-server
...
Running transaction
 Installing : postgresql-libs-9.2.7-1.ael7b.ppc64le
1/4
 Installing : postgresql-9.2.7-1.ael7b.ppc64le
2/4
 Installing : postgresql-server-9.2.7-1.ael7b.ppc64le
3/4
 Installing : ambari-server-2.1.0_IBM-4.ppc64le
4/4
 Verifying  : ambari-server-2.1.0_IBM-4.ppc64le
1/4
 Verifying  : postgresql-libs-9.2.7-1.ael7b.ppc64le
2/4
 Verifying  : postgresql-server-9.2.7-1.ael7b.ppc64le
3/4
 Verifying  : postgresql-9.2.7-1.ael7b.ppc64le
4/4

Installed:
  ambari-server.ppc64le 0:2.1.0_IBM-4

Dependency Installed:
  postgresql.ppc64le 0:9.2.7-1.ael7b  postgresql-libs.ppc64le 0:9.2.7-1.ael7b
postgresql-server.ppc64le 0:9.2.7-1.ael7b

Complete!
[root@mn01-dat ~]# rpm -qa ambari-server
ambari-server-2.1.0_IBM-4.ppc64le
```

Figure 4-5 Installing Ambari

3. Execute the gpfs-ambari integration package, as shown in Figure 4-6.

```
[root@mn01-dat ~]# ./tmp/gpfs.ambari-iop_4.1-1.noarch.bin
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 AGREEMEN
T. 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 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, D
ESTROY ALL COPIES OF THE PROGRAM.
...
Do you agree to the above license terms? [yes or no]
yes
Unpacking...
Done
Installing...
Preparing... ################################# [100%]
Updating / installing...
1:gpfs.ambari-iop_4.1-0 ################################# [100%]
```

Figure 4-6 Installing GPFS and Ambari integration

Important: Do not execute the gpfs-ambari integration package from `/root/` because it can introduce problems.

4. Update the value of the openjdk1.8.url and openjdk1.7.url in `/etc/ambari-server/conf/ambari.properties` to point to your local repository, as shown in Figure 4-7.

```
[root@mn01-dat ~]# cat /etc/ambari-server/conf/ambari.properties | grep
openjdk1.[78].url
openjdk1.8.url=http://smn/install/repos/IOP-UTILS/rhel/7/ppc64le/1.1/openjdk/jd
k-1.8.0.tar.gz
openjdk1.7.url=http://smn/install/repos/IOP-UTILS/rhel/7/ppc64le/1.1/openjdk/jd
k-1.7.0.tar.gz
```

Figure 4-7 Updating the Ambari openjdk repository configuration

5. Update your Ambari configuration to point to the new repositories, as shown in Figure 4-8 and in Figure 4-9.

```
[root@mn01-dat ~]# cd
/var/lib/ambari-server/resources/stacks/BigInsights/4.1.SpectrumScale/repos/
[root@mn01-dat repos]# vi repoinfo.xml
```

Figure 4-8 Updating Ambari repository configuration (part 1 of 2)

```
<reposinfo>
  <mainrepoid>IOP-4.1-Spectrum_Scale</mainrepoid>
  <os family="redhat7">
    <repo>

<baseurl>http://smn/install/repos/IOP/rhel/7/ppc64le/4.1.x/GA/4.1.0.0/</baseurl
>
      <repoid>IOP-4.1-mirror</repoid>
      <reponame>IOP</reponame>
    </repo>
    <repo>
      <baseurl>http://smn/install/repos/IOP-UTILS/rhel/7/ppc64le/1.1/</baseurl>
      <repoid>IOP-UTILS-1.1-mirror</repoid>
      <reponame>IOP-UTILS</reponame>
    </repo>
    <repo>
      <baseurl>http://smn/install/repos/GPFS/rhel/7/ppc64le/4.1.1/</baseurl>
      <repoid>GPFS-4.1.1</repoid>
      <reponame>GPFS</reponame>
    </repo>
</os>
</reposinfo>
```

Figure 4-9 Updating Ambari repository configuration (part 2 of 2)

6. Update `params.py` in the Ambari server to fix the Spark History Service Permission issue. By default, spark_eventlog_dir_mode is 01777, which will cause a permission issue when you start the Spark History Service. This issue might be fixed in the future. However, in the meantime, you must change spark_eventlog_dir_mode to 0777 (Figure 4-10).

```
[root@mn01-dat ~]# vi
/var/lib/ambari-server/resources/stacks/BigInsights/4.1/services/SPARK/package/
scripts/params.py
...
70 spark_hdfs_user_dir = format("/user/{spark_user}")
71 spark_hdfs_user_mode = 0755
72 spark_eventlog_dir_mode = 0777
73 spark_jar_hdfs_dir = "/iop/apps/4.1.0.0/spark/jars"
74 spark_jar_hdfs_dir_mode = 0755
75 spark_jar_file_mode = 0444
76 spark_jar_src_dir = "/usr/iop/current/spark-client/lib"
77 spark_jar_src_file = "spark-assembly.jar"
```

Figure 4-10 Spark history service permission workaround

7. Run the Ambari server setup, as shown in Figure 4-11.

```
[root@mn01-dat ~]# ambari-server setup
Using python  /usr/bin/python2.7
Setup ambari-server
Checking SELinux...
SELinux status is 'disabled'
Customize user account for ambari-server daemon [y/n] (n)? n
Adjusting ambari-server permissions and ownership...
Checking firewall status...
Redirecting to /bin/systemctl status  iptables.service

Checking JDK...
[1] OpenJDK 1.8.0
[2] OpenJDK 1.7.0 (deprecated)
[3] Custom JDK
================================================================================
Enter choice (1): 1
Downloading JDK from
http://smn/install/repos/IOP-UTILS/rhel/7/ppc64le/1.1/openjdk/jdk-1.8.0.tar.gz
to /var/lib/ambari-server/resources/jdk-1.8.0.tar.gz
jdk-1.8.0.tar.gz... 100% (48.3 MB of 48.3 MB)
Successfully downloaded JDK distribution to
/var/lib/ambari-server/resources/jdk-1.8.0.tar.gz
Installing JDK to /usr/jdk64/
Successfully installed JDK to /usr/jdk64/
Completing setup...
Configuring database...
Enter advanced database configuration [y/n] (n)? n
Configuring database...
Default properties detected. Using built-in database.
Configuring ambari database...
Checking PostgreSQL...
Running initdb: This may take upto a minute.
Initializing database ... OK


About to start PostgreSQL
Configuring local database...
Connecting to local database...done.
Configuring PostgreSQL...
Restarting PostgreSQL
Extracting system views...
ambari-admin-2.1.0_IBM_4.jar
......
Adjusting ambari-server permissions and ownership...
Ambari Server 'setup' completed successfully.
```

Figure 4-11 Ambari server setup

8. Start the Ambari server, as shown in Figure 4-12.

```
[root@mn01-dat ~]# ambari-server start
Using python  /usr/bin/python2.7
Starting ambari-server
Ambari Server running with administrator privileges.
Organizing resource files at /var/lib/ambari-server/resources...
Server PID at: /var/run/ambari-server/ambari-server.pid
Server out at: /var/log/ambari-server/ambari-server.out
Server log at: /var/log/ambari-server/ambari-server.log
Waiting for server start...................
Ambari Server 'start' completed successfully.
```

Figure 4-12 Ambari server startup

4.2.4 Deploying the IBM Open Platform edition of BigInsights

This section describes how to deploy the IBM Open Platform edition of BigInsights. Follow these steps:

1. Before the installation, ensure that you configure password-less root access from ambari-server node to all other nodes. This capability is required for Spectrum Scale administration. If you used PCM or xCAT to deploy your nodes, this configuration is performed automatically.
2. You can now connect to the default port 8080 on the server that is running ambari-server through your web browser: `http://mn01-dat:8080/`. The default account username/password is `admin/admin`.

Tip: The client port can be overridden by setting client.api.port in `/etc/ambariserver/conf/ambari.properties`.

Changes require that you restart the server with the command **`ambari-server restart`**.

3. To launch the Ambari Install Wizard, click **Launch Install Wizard**, as shown in Figure 4-13.

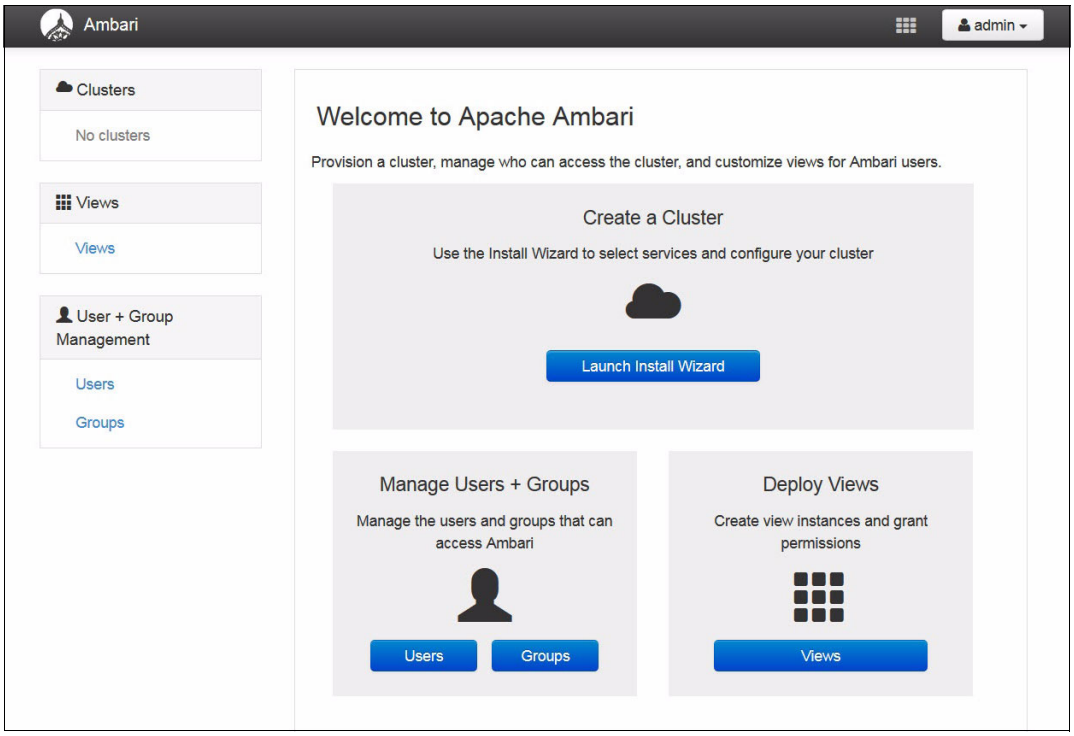


Figure 4-13 Ambari cluster installation wizard: Pre-cluster view

4. Name your cluster, as shown in Figure 4-14.

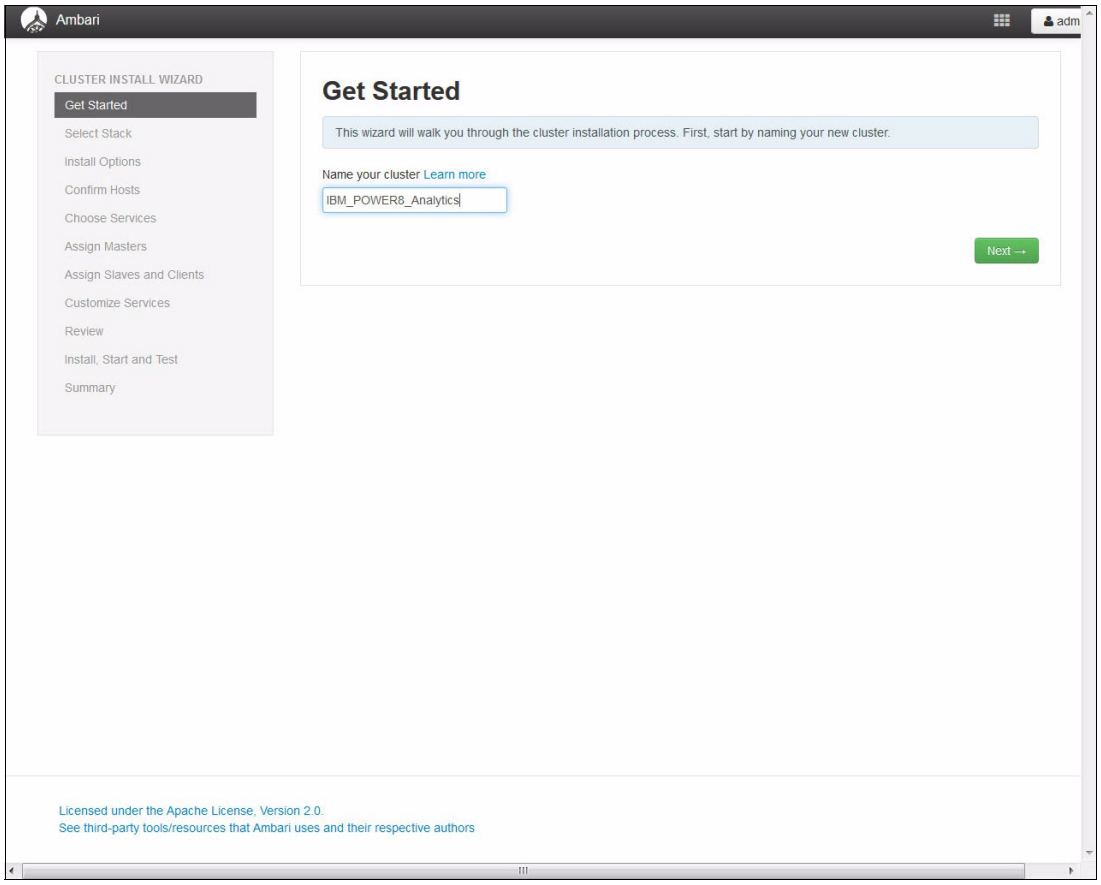


Figure 4-14 Ambari cluster installation wizard: Naming your cluster

5. Select the BigInsights 4.1.SpectrumScale stack by clicking **BigInsights 4.1.SpectrumScale** to pull in the Spectrum Scale services and configuration and the modified configurations for other services to use GPFS instead of HDFS. Click **Advanced Repository Options** to verify and change your repositories, if necessary (Figure 4-15).

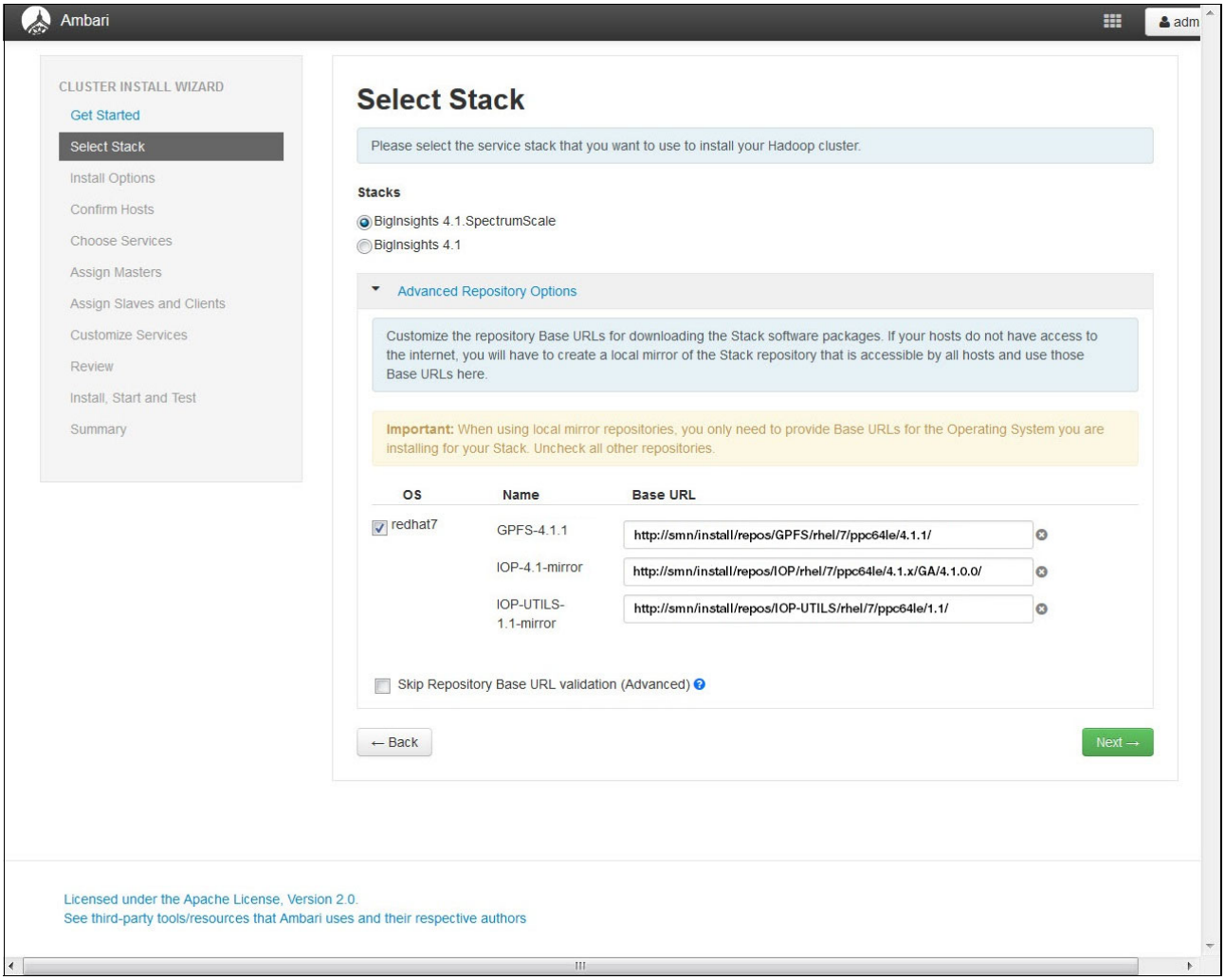


Figure 4-15 Ambari cluster installation wizard: Stack selection

6. Enter information about the nodes in the cluster, as shown in Figure 4-16.

Note: For simplicity, the following windows show only two management nodes (`julia` and `emma`), and they do not show any data nodes.

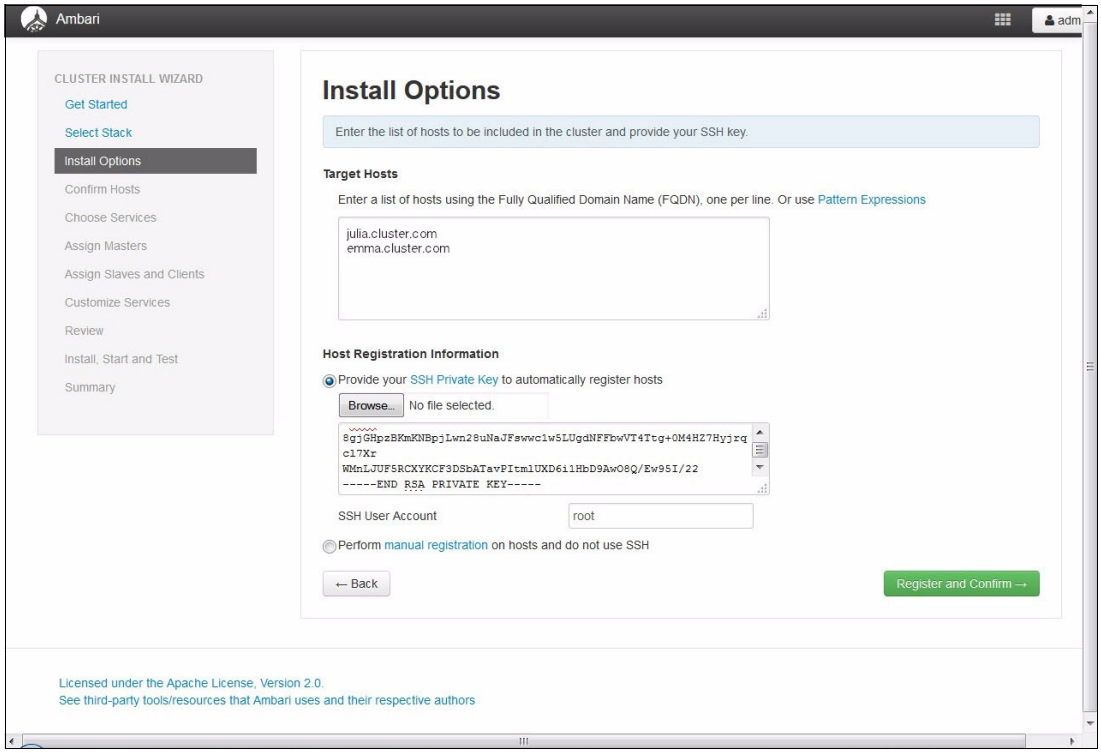


Figure 4-16 Ambari cluster installation wizard: Install options

7. Click **Register and Confirm**. Ambari installs its agent service on each node and runs several basic checks, as shown in Figure 4-17 and in Figure 4-18 on page 64.

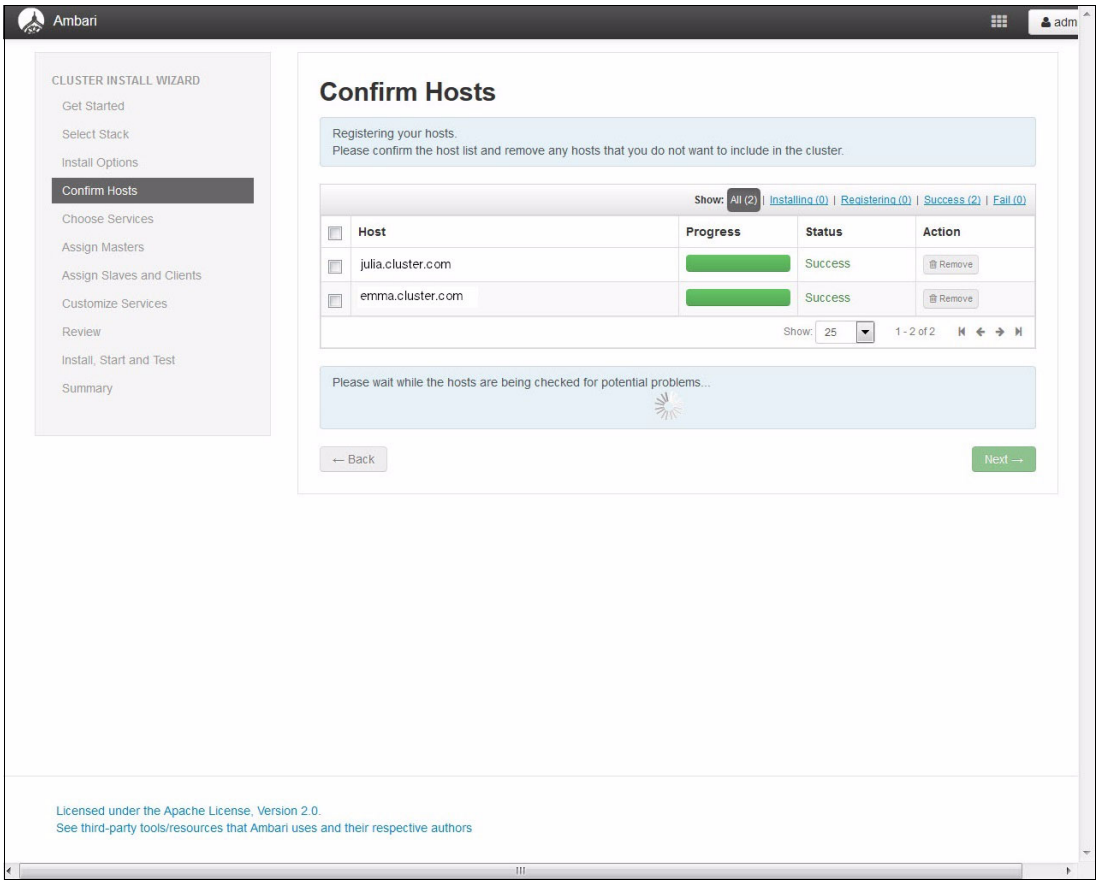


Figure 4-17 Ambari cluster installation wizard: Host confirmation

8. Figure 4-18 shows the Host Checks window.

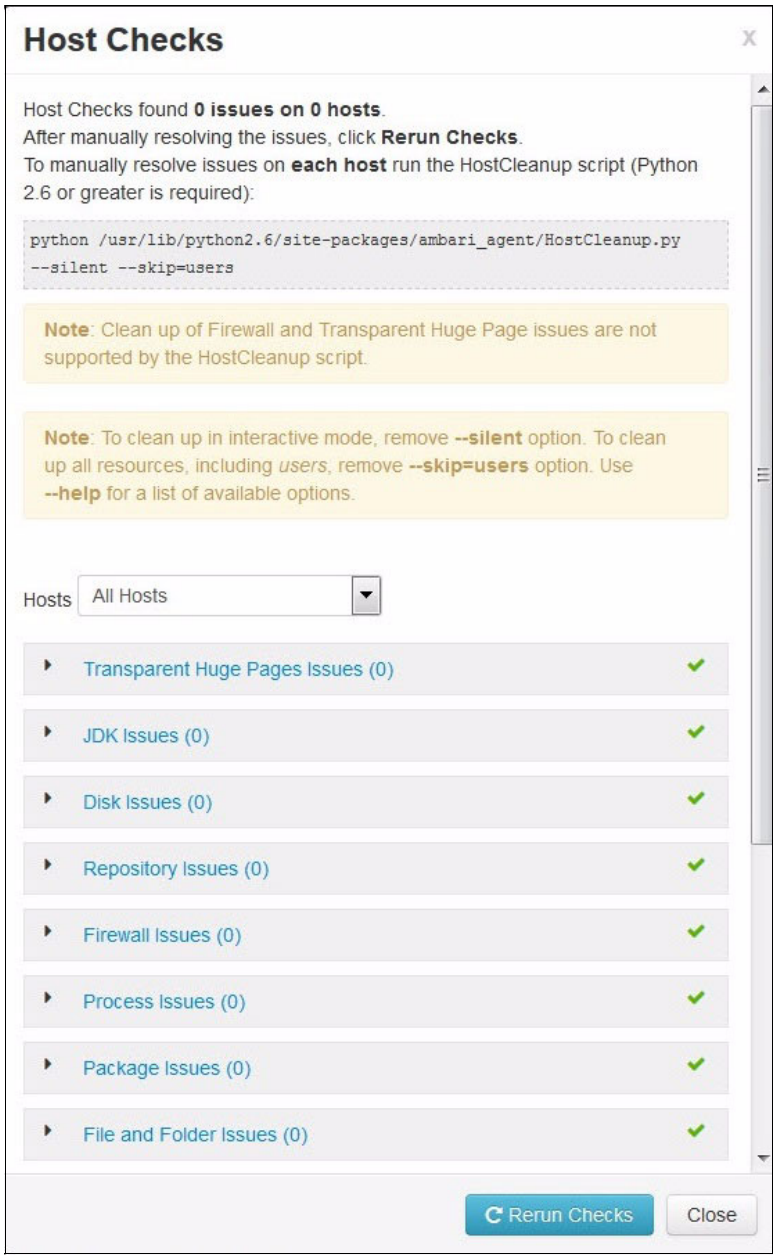


Figure 4-18 Ambari cluster installation wizard: Host Checks

9. Choose **Services**, as shown in Figure 4-19 on page 66.
10. Assign node roles. For more information about recommendations for two, four, six, or eight management server systems, see Appendix B, “Planning Ambari node roles” on page 267.

Important: ResourceManager, Symphony Master, Spark History Server, and Spark Thrift Server must on the same node.

11. The Spectrum Scale Master Node functions as the node where commands that affect the entire cluster will run. For example, when Spectrum Scale is first installed and an FPO cluster is first created, the commands are all executed on the Spectrum Scale Master Node. On the Spectrum Scale Nodes, after the rpms is installed, defer to the Master Node for adding nodes to the cluster, deploying the Hadoop connector, and so on.

    As another example, if the configuration changes after the cluster is deployed, the Spectrum Scale Master Node executes the commands to reconfigure the cluster and if necessary, to restart Spectrum Scale on all nodes. The term “*Master*” is used here only to follow the convention that is used by the other Hadoop services. The Spectrum Scale Master Node has no special role in the Spectrum Scale cluster itself (other than as one of the quorum nodes).

    Assign Slaves and Clients. Spectrum Scale has a node role, GPFS Node, which must be deployed to every node (including the Spectrum Scale Master Node).

Note: Reference `DeployBigInsights4.1_SpectrumScale_with_Ambari 2.1_v0.8.1.pdf` at the following website:

`http://ibm.co/1RYHeG1`

12. Figure 4-19 shows how to assign node master roles.

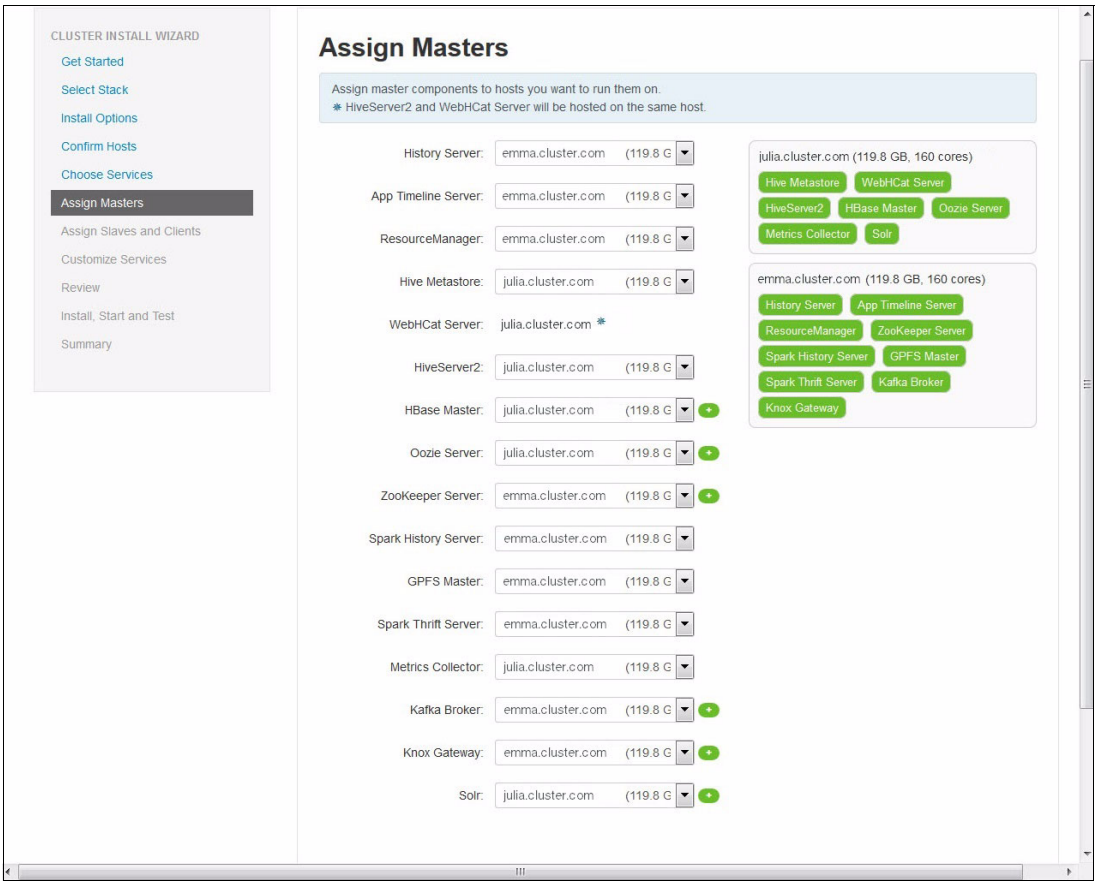


Figure 4-19 Ambari cluster installation wizard: Assign node master roles

13. Figure 4-20 shows how to assign slave and client roles.

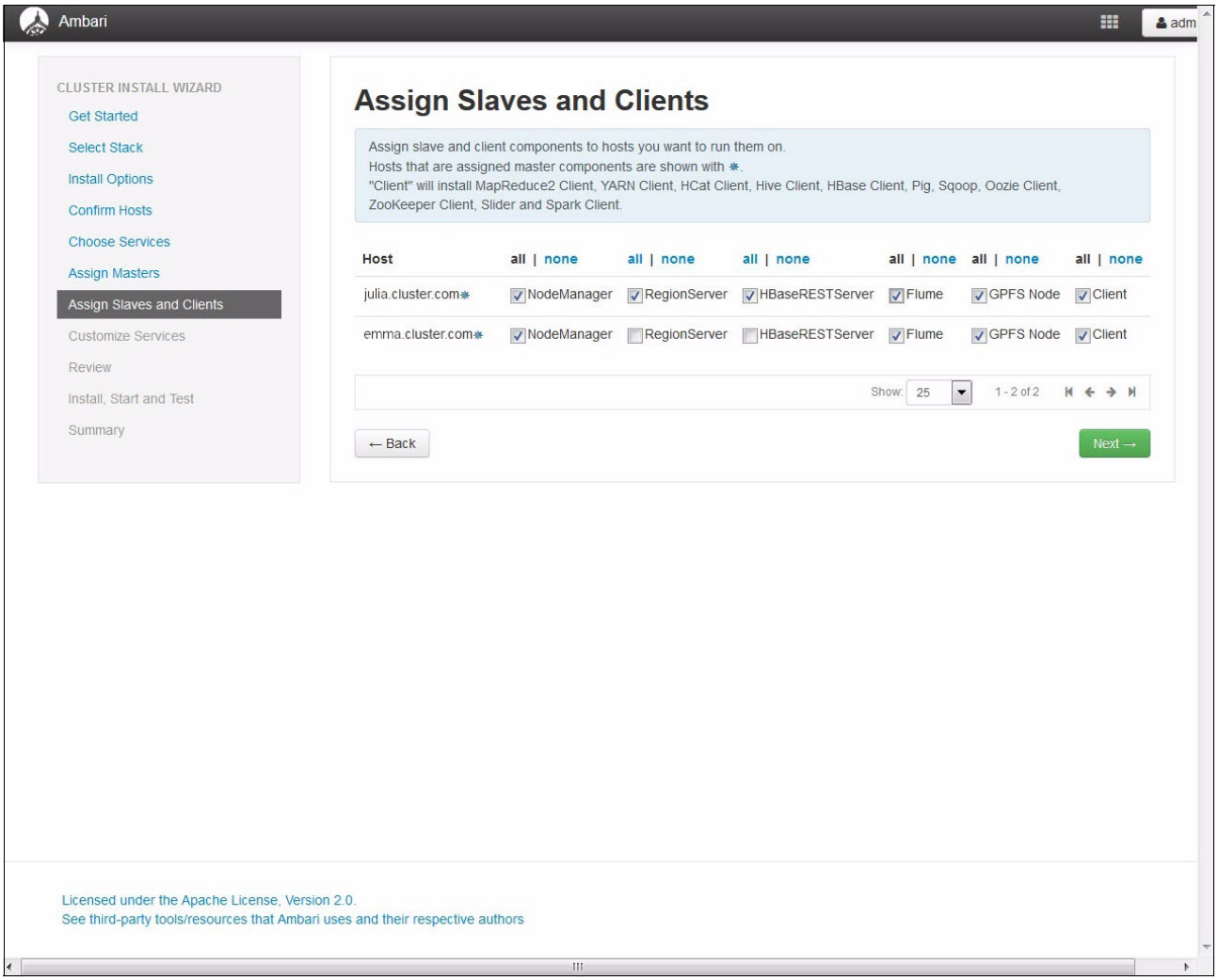


Figure 4-20 Ambari cluster installation wizard: Assign node slave and client roles

14. Customize services. Click each service tab that is flagged with a red circle and enter the required database passwords. Optional: Review the configuration for each service, as shown in Figure 4-21.

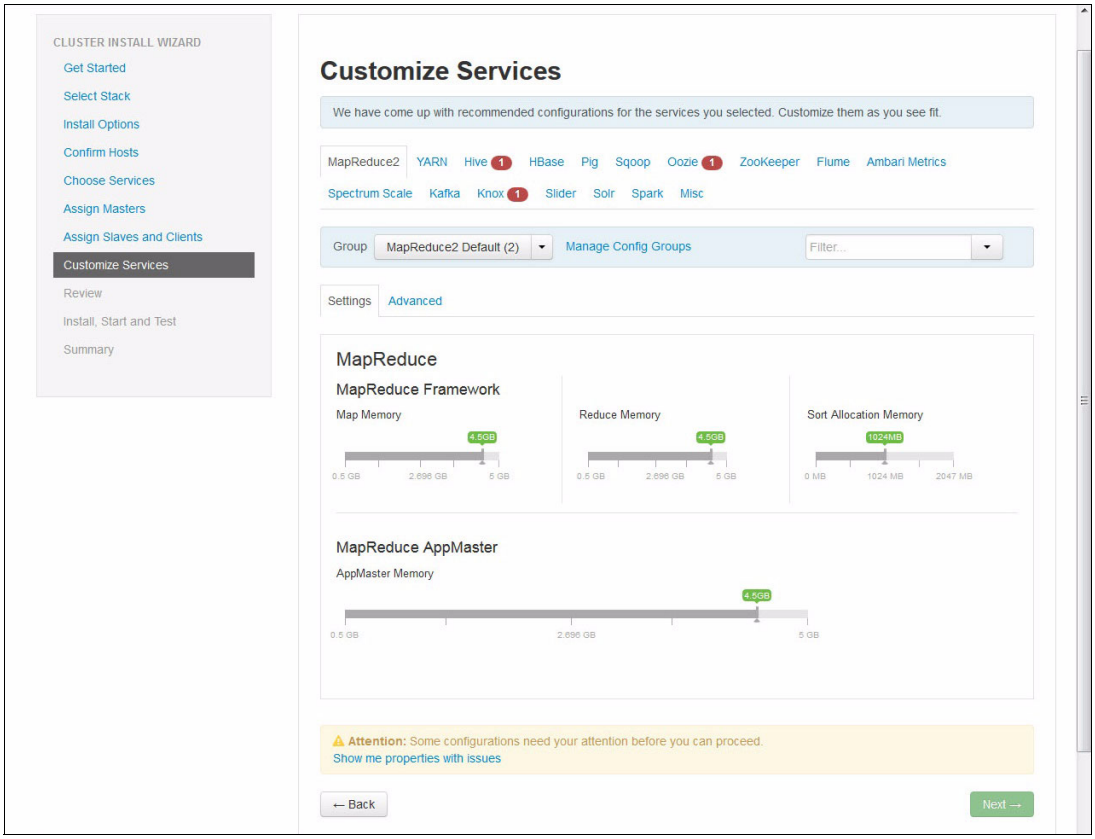


Figure 4-21 Ambari cluster installation wizard: Customize Services

15. Configure the Spectrum Scale stack. Spectrum Scale has its own tab on the Customize Services page. Within that tab, two tabs, Standard and Advanced configuration, appear. In the Standard tab, the user can adjust parameters through either the slider bars or drop-down menus. The Advanced tab contains the parameters that do not need to be changed frequently. When you create a new cluster, all fields on both pages are populated with initial values that are taken from the recommendations in the following white paper:

    http://ibm.co/1UInJQ2

16. The parameters that are followed by a lock icon must not be changed after deployment (for example, Cluster name, RemoteShell, /Filesystem Name, Max Data Replicas/). Before you start the deployment, double-check all important parameters.

17. Two types of NSD files are supported for file system creation. The preferred approach is to use the simple format. However, the standard Spectrum Scale NSD file format is also accepted. If you choose to use the simple NSD file format, Ambari will choose the correct metadata and data ratio for you. If possible, Ambari will also create partitions on several disks for Hadoop intermediate data, which can improve Hadoop performance. If you choose to use the standard Spectrum Scale NSD file format, you need to take all responsibility for the arrangement of all storage space.

18. Due to a limitation in the Ambari framework, the NSD file must be placed on the Ambari server under `/var/lib/ambari-server/resources/gpfs_nsd`. See the `gpfs_nsd.sample` for an example of the simple format.

19. Review and begin the deployment, as shown in Figure 4-22.

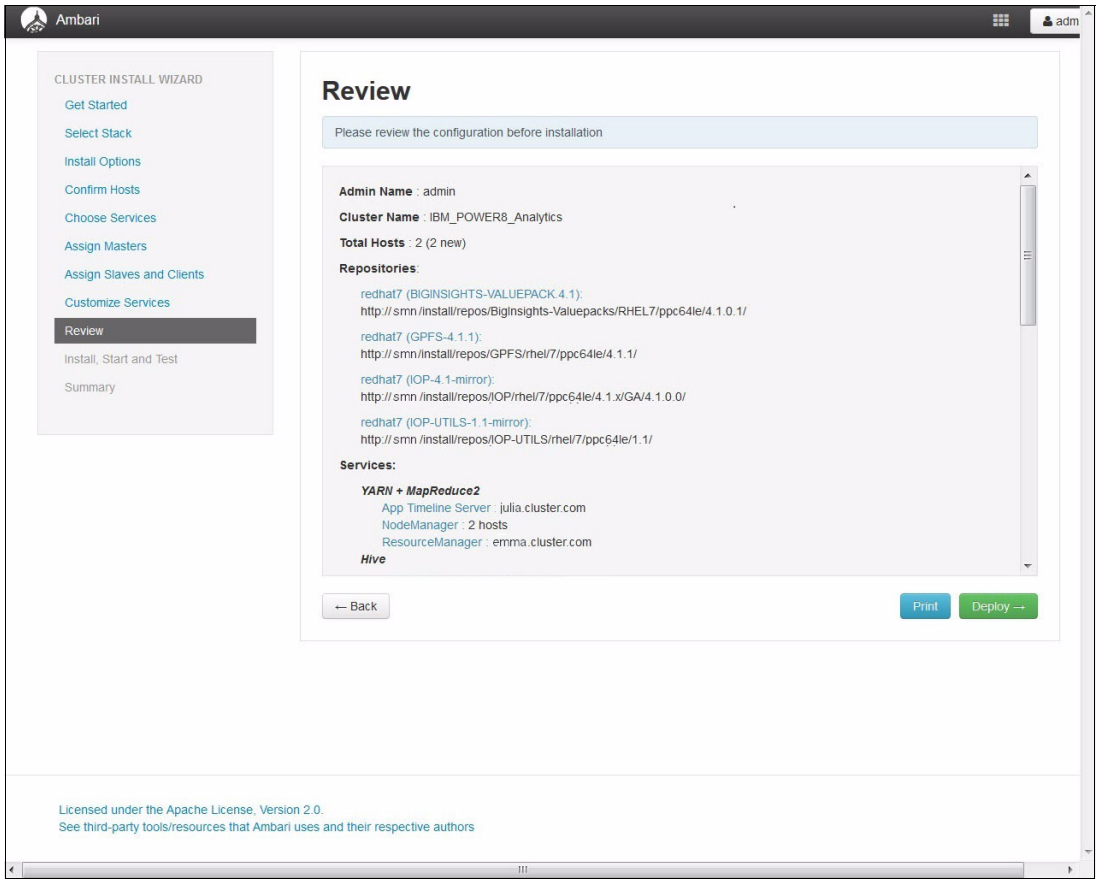


Figure 4-22 Ambari cluster installation wizard: Review deployment

20. Figure 4-23 shows the Install, Start and Test window.

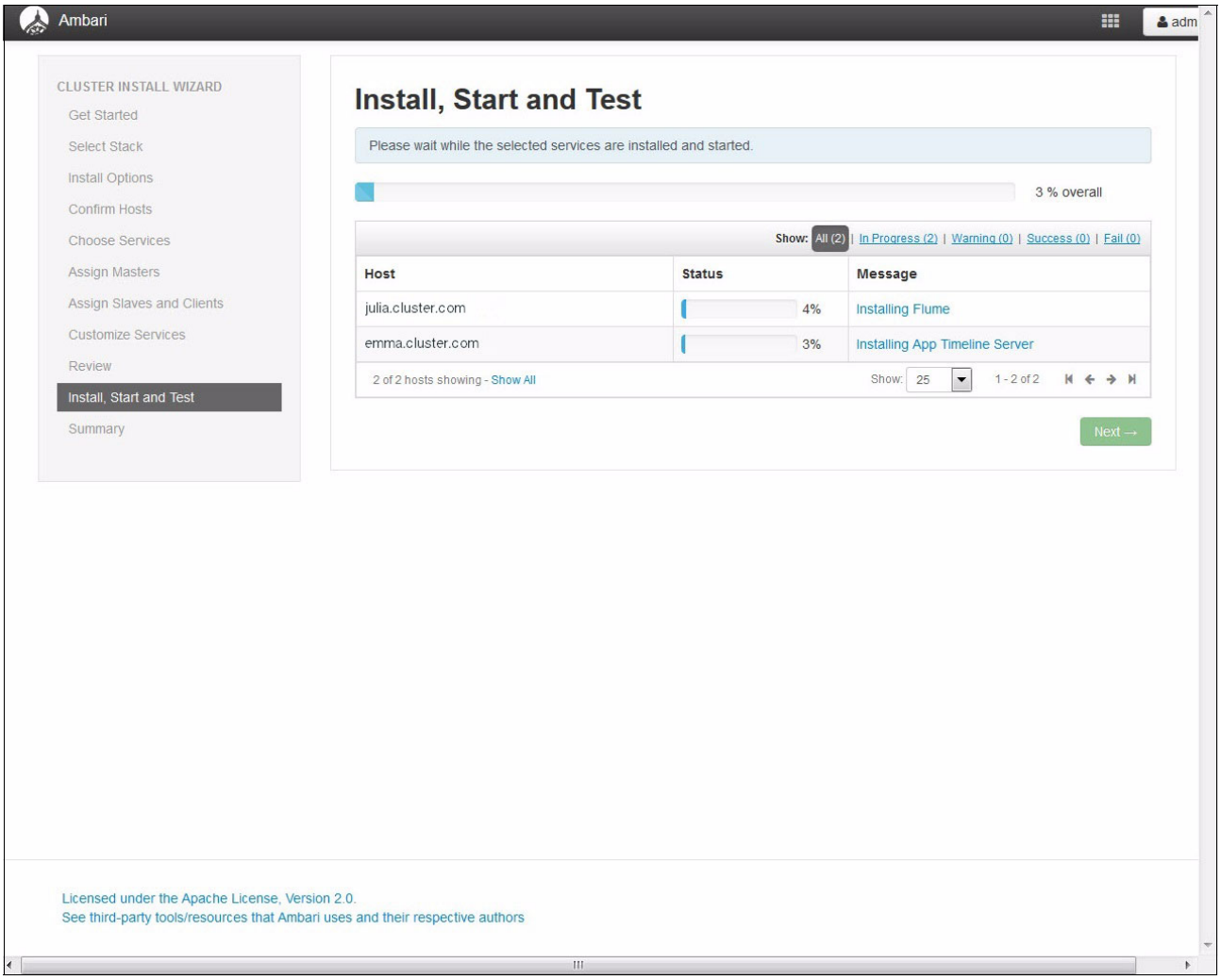


Figure 4-23 Ambari cluster installation wizard: Install and test

21. After the deployment completes, you might need to restart several services, as shown in Figure 4-24.

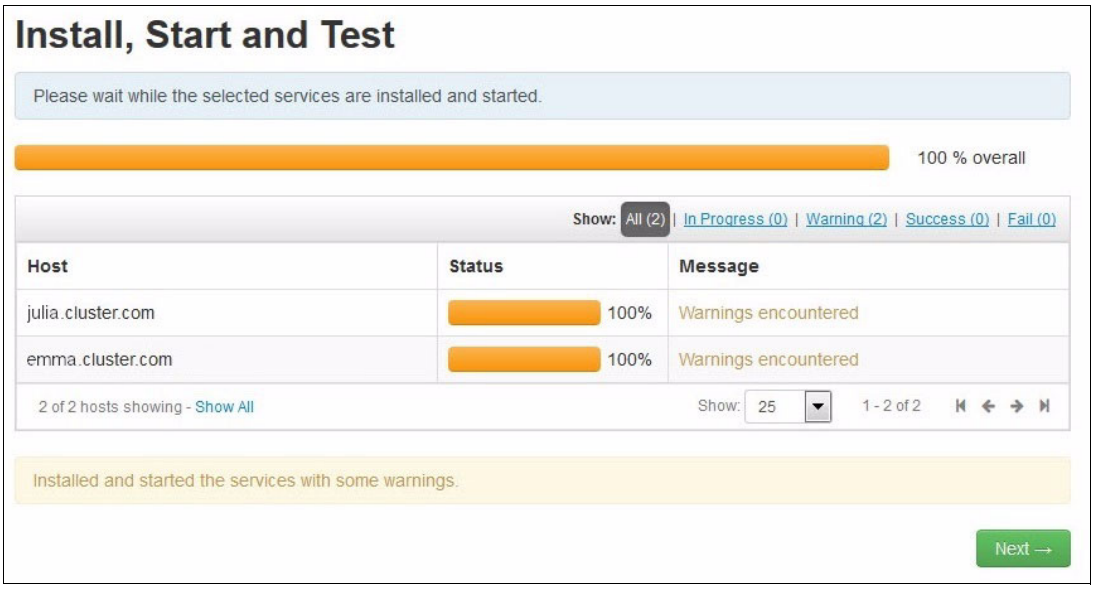


Figure 4-24 Ambari cluster installation wizard: Install warnings

22. Due to limitations in the Ambari stack, you must install the update to Spectrum Scale 4.1.1.3 manually.

Note: For more information about supported versions, see the following website:

`https://ibm.biz/Bd4ayt`

23. Extract the Spectrum Scale update rpm packages to `/install/repos/GPFS/rhel/7/ppc64le/4.1.1.3/` on the smn node.

24. Ensure that you delete any gpfs.hadoop-connector rpms because the latest rpm is installed with the base version repository. Then, create the yum repository metadata, as shown in Figure 4-25.

```
[root@smn 4.1.1.3]# ls
gpfs.base-4.1.1-3.ppc64.update.rpm    gpfs.ext-4.1.1-3.ppc64.update.rpm
gpfs.msg.en_US-4.1.1-3.noarch.rpm
gpfs.crypto-4.1.1-3.ppc64.update.rpm  gpfs.gpl-4.1.1-3.noarch.rpm
gpfs.docs-4.1.1-3.noarch.rpm          gpfs.gskit-8.0.50-47.ppc64.rpm
[root@smn 4.1.1.3]# createrepo .
```

Figure 4-25 Spectrum Scale update rpms

25. Create the yum configuration file, as shown in Figure 4-26.

```
[root@smn 4.1.1.3]# vi /etc/yum.repos.d/GPFS_PTF.repo
[root@smn 4.1.1.3]# cat /etc/yum.repos.d/GPFS_PTF.repo
[GPFS-4.1.1.3]
name=GPFS-4.1.1.3
baseurl=http://smn/install/repos/GPFS/rhel/7/ppc64le/4.1.1.3
path=/
enabled=1
gpgcheck=0
```

Figure 4-26 Spectrum Scale update yum configuration file

26. Copy the yum configuration file to all nodes, as shown in Figure 4-27.

```
[root@smn 4.1.1.3]# xdcp __Managed /etc/yum.repos.d/GPFS_PTF.repo
/etc/yum.repos.d/GPFS_PTF.repo
```

Figure 4-27 Spectrum Scale yum repository distribution

27. Stop all services that use the Ambari graphical user interface (GUI), as shown in Figure 4-28.

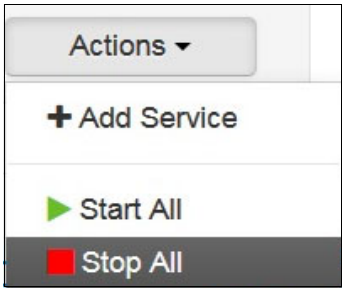


Figure 4-28 Ambari Actions: Stop All services

28. Ensure that Spectrum Scale was shut down.

29. Proceed to install the updated packages, as shown in Figure 4-29.

```
[root@smn ~]# ssh julia /usr/lpp/mmfs/bin/mmgetstate -a

 Node number  Node name        GPFS state
------------------------------------------
       1      julia            down
       2      emma             down
[root@smn ~]# xdsh __Managed "yum -y erase gpfs.gpl gpfs.docs gpfs.gskit
gpfs.msg.en_US gpfs.crypto gpfs.ext"
[root@smn ~]# xdsh __Managed "yum -y install \
gpfs.base gpfs.ext gpfs.crypto gpfs.gpl gpfs.docs gpfs.gskit gpfs.msg.en_US"
[root@smn ~]# xdsh __Managed "rpm -qa gpfs.base"
emma: gpfs.base-4.1.1-3.ppc64le
julia: gpfs.base-4.1.1-3.ppc64le
[root@smn ~]# xdsh __Managed "/usr/lpp/mmfs/bin/mmbuildgpl"
...
[root@smn ~]# xdsh __Managed "/usr/lpp/mmfs/bin/mmstartup"
...
[root@smn ~]# ssh julia /usr/lpp/mmfs/bin/mmgetstate -a

 Node number  Node name        GPFS state
------------------------------------------
       1      julia            active
       2      emma             active
```

Figure 4-29 Spectrum Scale update installation

Tip: If Spectrum Scale is still active, you need to run the **mmshutdown** command. Consider a rolling upgrade if you cannot shut down the entire cluster at one time. For more information, see the *GPFS FPO Cluster Maintenance Guide* on the Big Data Best practices page of the IBM Spectrum Scale Wiki:

http://ibm.co/1RZmRbX

30. Start all services in the Ambari GUI.

4.2.5 Installing the BigInsights value-add packages

This section provides the steps to install the BigInsights value-add packages. Follow the steps:

1. Plan ahead. Before you add the servers, check the suggested layout:

   http://ibm.co/1SX2igv

2. Download IBM BigInsights Analyst 4.1.0.1 for Linux on Power (`bana-1.1.0.0.el7.ppc64le.bin`).
3. Download IBM BigInsights for Apache Hadoop 4.1.0.1 for Linux on Power (`bah-1.1.0.0.el7.ppc64le.bin`).
4. Ensure that the eAssemby files are executable (**`chmod a+x *.bin`**). Execute the programs from a machine with access to the internet. Choose the **OFFLINE** installation type to download `BI-ANA-RHEL7.tar.gz` and `BI-ANA-RHELIOP.tar.gz`.

5. Extract the contents of both archives into /install/repos/BigInsights-Valuepacks/RHEL7/ppc64le/4.1.0.1/ on the smn repository node, and run **createrepo**, as shown in Figure 4-30.

```
[root@smn ~]# cd /install/repos/BigInsights-Valuepacks/RHEL7/ppc64le/4.1.0.1
[root@smn 2.1]# createrepo .
[root@smn 4.1.0.1]# ls
BI-Analyst-IOP-1.1.0.1-4.1.el7.ppc64le.rpm   bigsheets-distrib-5.11.2.rpm
repodata
BI-Apache-Hadoop-IOP-1.1.0.1-4.1.el7.ppc64le.rpm
bigsql-dist_4_1_0_0-5.28.1-Linux-ppc64le.rpm     text-analytics-runtime-4.6.rpm
BigR-4.3.0.5.rpm
bigsql-samples_4_1_0_0-5.28.1.rpm
text-analytics-web-tooling-3.4.rpm
BigR-BigSQL1-3.4.rpm
db2luw-linuxppc64le-10.6.0.3-s150918-db2rpm.rpm  web-ui-framework-2.7.rpm
BigR-Jaql-3.2.0.1.rpm
dsm-1.1.1.1-N20150908_1239.noarch.rpm
BigR-SystemML-5.4.0.2.rpm                         jsqsh-4.4-Linux-amd64.rpm
```

Figure 4-30 BigInsights-Valuepacks repository metadata

6. On the Ambari manager node, set up yum to point to the new repository, as shown in Figure 4-31.

```
[root@mn01-dat ~]# cat /etc/yum.repos.d/BIGINSIGHTS-VALUEPACK.4.1.repo
[BIGINSIGHTS-VALUEPACK.4.1]
name=BIGINSIGHTS-VALUEPACK.4.1
baseurl=http://smn/install/repos/BigInsights-Valuepacks/RHEL7/ppc64le/4.1.0.1/

path=/
enabled=1
```

Figure 4-31 BigInsights-Valuepacks yum configuration file

7. Install the enablement rpms on the Ambari server, as shown in Figure 4-32.

```
yum -y install BI-Apache-Hadoop-IOP-2.13.1-IOP-4_1
yum -y install BI-Analyst-IOP-2.13.1-IOP-4_1
```

Figure 4-32 BigInsights-Valuepacks enablement rpm installation

8. Update your Ambari configuration to point to the new repositories, as shown in Figure 4-33 and in Figure 4-34.

```
   cd
/var/lib/ambari-server/resources/stacks/BigInsights/4.1.SpectrumScale/repos/
   vi repoinfo.xml
```

Figure 4-33 Ambari server repository configuration update (part 1 of 2)

```
<reposinfo>
  <mainrepoid>IOP-4.1-Spectrum_Scale</mainrepoid>
  <os family="redhat7">
    <repo>

<baseurl>http://smn/install/repos/IOP/rhel/7/ppc64le/4.1.x/GA/4.1.0.0/</baseurl
>
      <repoid>IOP-4.1-mirror</repoid>
      <reponame>IOP</reponame>
    </repo>
    <repo>
      <baseurl>http://smn/install/repos/IOP-UTILS/rhel/7/ppc64le/1.1/</baseurl>
      <repoid>IOP-UTILS-1.1-mirror</repoid>
      <reponame>IOP-UTILS</reponame>
    </repo>
    <repo>
      <baseurl>http://smn/install/repos/GPFS/rhel/7/ppc64le/4.1.1/</baseurl>
      <repoid>GPFS-4.1.1</repoid>
      <reponame>GPFS</reponame>
    </repo>
    <repo>
      <baseurl>http://smn/install/repos/ValueAdds/</baseurl>
      <repoid>BigInsights-ValueAdds-IOP-4.1-mirror</repoid>
      <reponame>BI-ValueAdds-IOP-4.1-mirror</reponame>
    </repo>
  </os>
</reposinfo>
```

Figure 4-34 Ambari server repository configuration update (part 2 of 2)

Tip: Ambari uses a relational database management system (rdbms) to store its data. In this project, the default postgresSQL was used. The entire `repoinfo.xml` file is ignored by the ambari post-installation if a single repository is invalid. Ensure that you remove or update any invalid entries. The file also does not get updated by the ambari-server if the file was changed dynamically at installation time. If you experience any problems when you load the `repoinfo.xml` file, check `/var/log/ambari-server/ambari-server.log` for details. For example, an error that is similar to the following example might appear:

“`AmbariManagementControllerImpl:3583 - Could not access base url . http://192.168.9.3/repos/IOP-UTILS/RHEL6/x86_64/1.1 . Network is unreachable`”

9. After you update the `repoinfo.xml` file, restart Ambari, as shown in Figure 4-35.

```
ambari-server restart
```

Figure 4-35 Ambari server restart

10. From the Ambari website, click **Admin → Stack and Versions**. Click the **Versions** tab, and click the **Edit Repositories** icon, as shown in Figure 4-36.

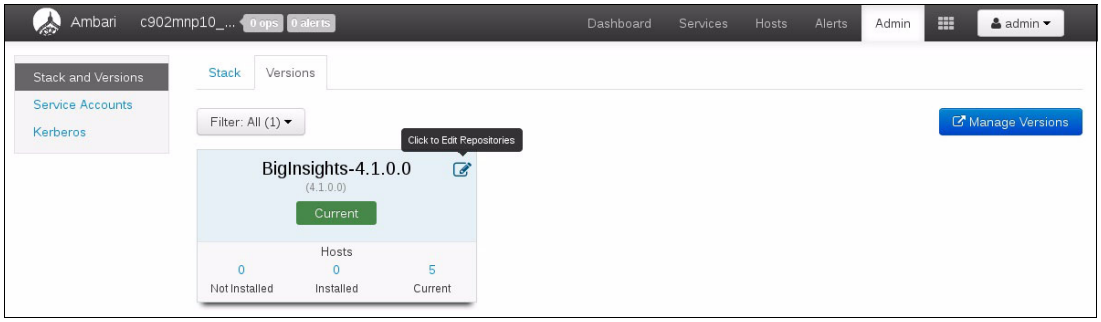


Figure 4-36 Ambari Stack and Versions

11. Ensure that the repositories that you updated in `repoinfo.xml` are displayed, as shown in Figure 4-37.

Figure 4-37 Ambari stack repositories

12. After you install the rpms, you can run the pre-installation check for Big SQL outside of Ambari before you add the services (Figure 4-38).

```
/var/lib/ambari-server/resources/stacks/BigInsights/4.1/services/BIGSQL/package
/scripts/bigsql-precheck.sh -M PRE_ADD_HOST -V -u bigsql
```

Figure 4-38 BigInsights pre-installation check

13. For more information, see the following website:

    http://ibm.co/1TDj4jV

Note: The script is also executed by Ambari when you add the services.

14. Several pre-check issues are common:
    - `Error on sudoers`. You need to comment the line: `Defaults require tty` (on all nodes).
    - `FAIL Hosts file check`: You might see this error from a pre-installation check that is performed by Big SQL. If you followed the IDEA runbook to set up your cluster, your `/etc/hosts` file is probably defined in a way that causes **`hostname -s`** and **`hostname -f`** to return the same short *hostname*.
15. To fix this error, use the **`-l`** (lowercase `L`) option with the **`makehosts`** command on your PCM node to rebuild the `/etc/hosts` file, as shown in Figure 4-39.

```
makehosts __Managed,lpar_ess,switch,smn -l
updatenode __Managed,lpar_ess -F
```

Figure 4-39 xCAT hosts configuration

16. The relevant entries of the corrected `/etc/hosts` file entries are shown in Figure 4-40.

```
172.16.1.10 julia.cluster.com julia
172.16.1.11 emma.cluster.com emma
[root@julia ~]# hostname -s
julia
[root@julia ~]# hostname -f
julia.cluster.com
```

Figure 4-40 /etc/hosts name order

17. Confirm that Hive metastore connectivity exists from the node where Big SQL will be installed, even if Big SQL will be on the same node with Hive. You can test this connectivity by opening the Hive shell from the command line and running a simple command. Perform the steps that are shown in Figure 4-41.

```
Authenticate to hive:

su hive

Open the HIVE shell by typing the following command from the command line:

hive

Run a command such as the following command that displays tables:

hive> show tables;
```

Figure 4-41 Hive metastore connectivity test

18. Perform any remaining steps in the planning guide at the following website:

    https://ibm.biz/Bd4SbT

19. In the Ambari web interface, click **Actions** → **Add Service** (Figure 4-42).

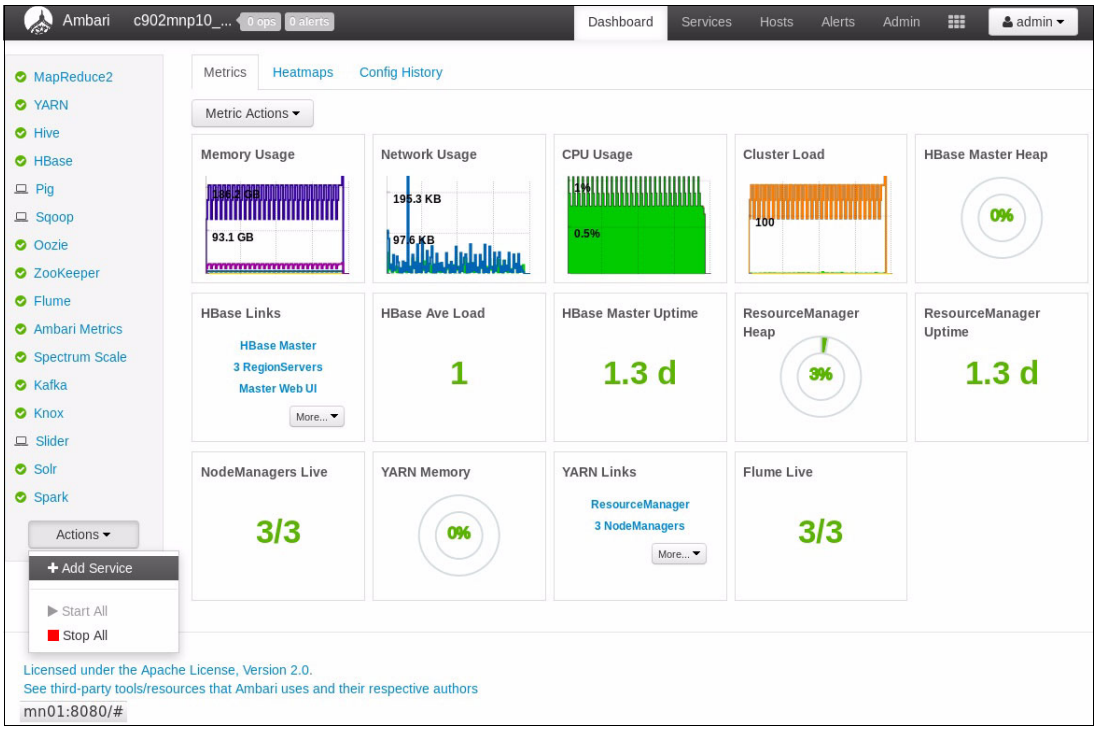


Figure 4-42 Ambari Add Service Actions menu

20. Click **Add Service Wizard** → **Choose Services** → **BigInsights - Data Server Manager** → **Big Insights Big SQL** → **Big Insights Home**. Figure 4-43 shows the Ambari Add Service selection menu.

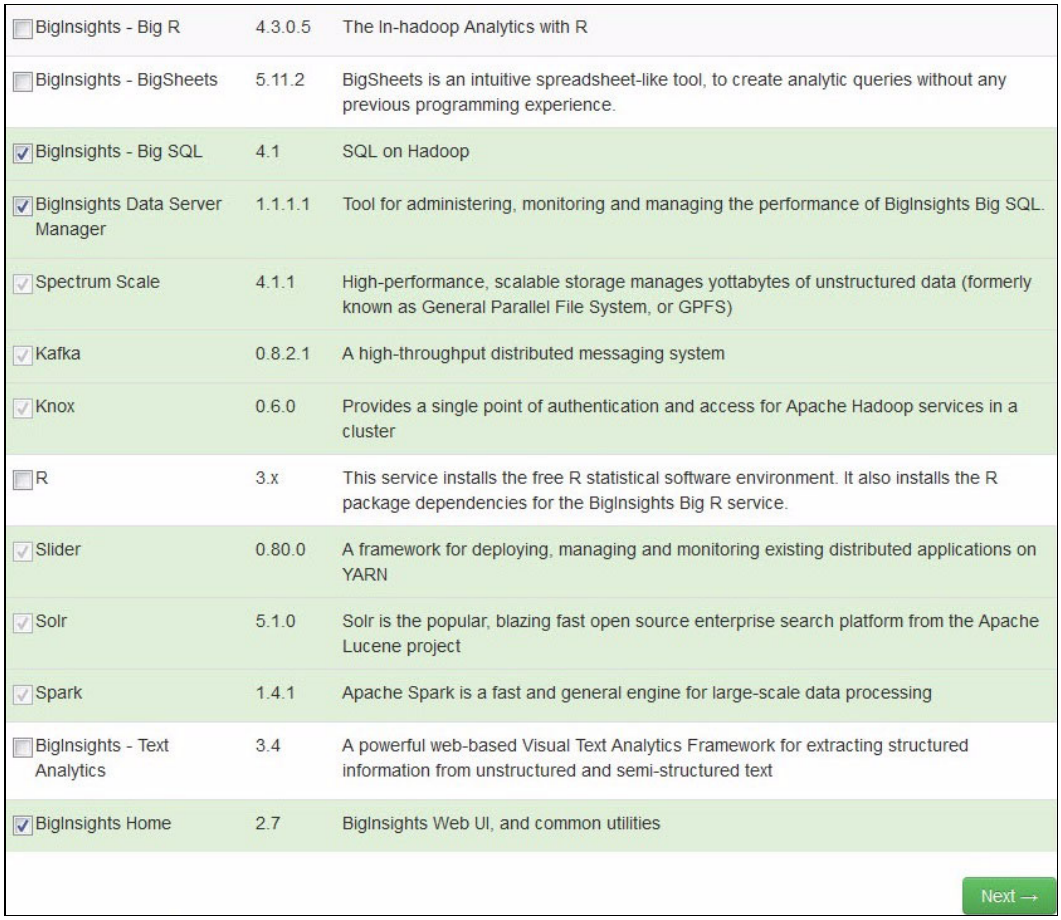


Figure 4-43 Ambari Add Service: Selection

21. Assign the master components to the hosts on which you want to run them, as shown in Figure 4-44.

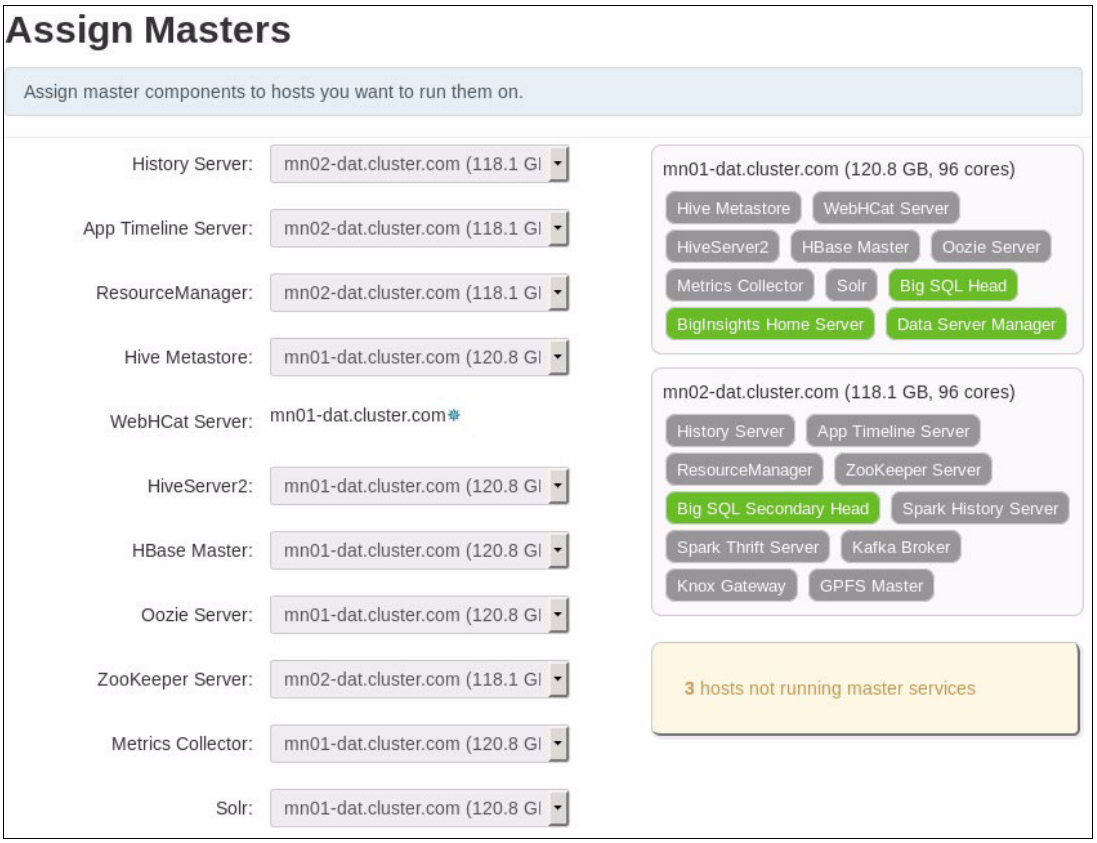


Figure 4-44 Ambari Add Service: Assign Masters

Important: When you choose your Big SQL Head and Secondary Head, you cannot run Big SQL Worker and Head services on the same node. In this case, we selected management nodes mn01 and mn02.

22. Figure 4-45 shows the Ambari window to assign slaves and clients.

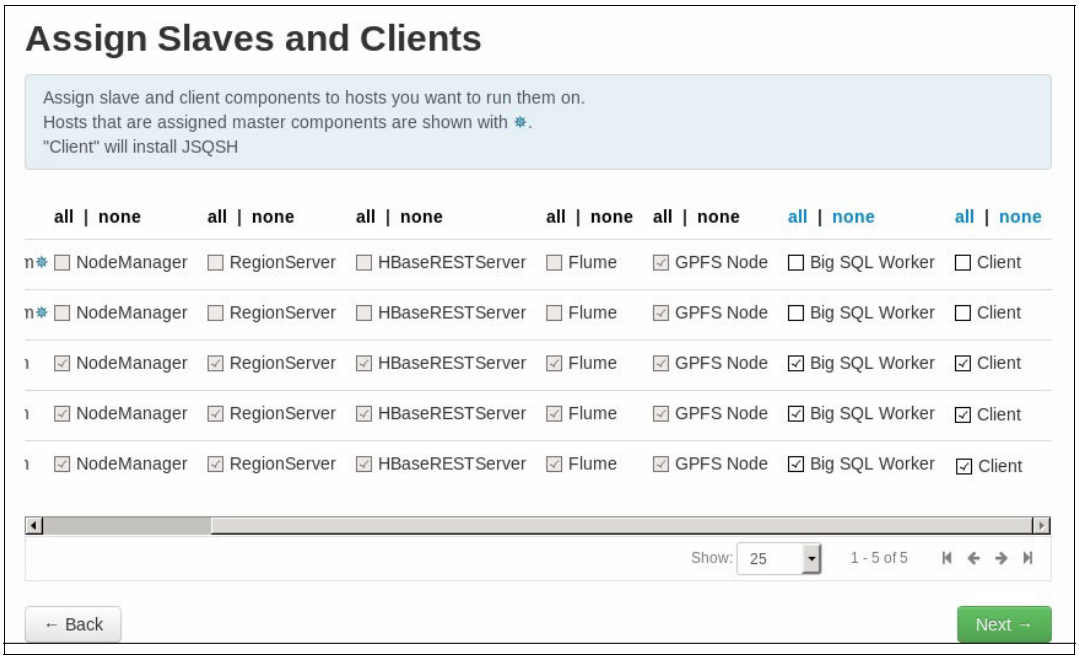


Figure 4-45 Ambari Add Service: Assign Slaves and Clients

Tip: Place the Big SQL workers on the data nodes only. In this example, our data nodes are dn01, dn02, and dn03.

23. Customize the services. In the Customize Services step for BigInsights Big SQL, we defined the password for the bigsql user as `cluster`.

24. For Data Server Manager, for the dsm_admin_user field, we entered `admin`. Type a Knox user name to become the administrator for Data Server Manager. For more information about the configuration steps, see the Installing the BigInsights - Data Server Manager topic in the IBM Knowledge Center:

    http://ibm.co/20900d7

25. Click each service that is flagged in red and complete the required fields, as shown in Figure 4-46.

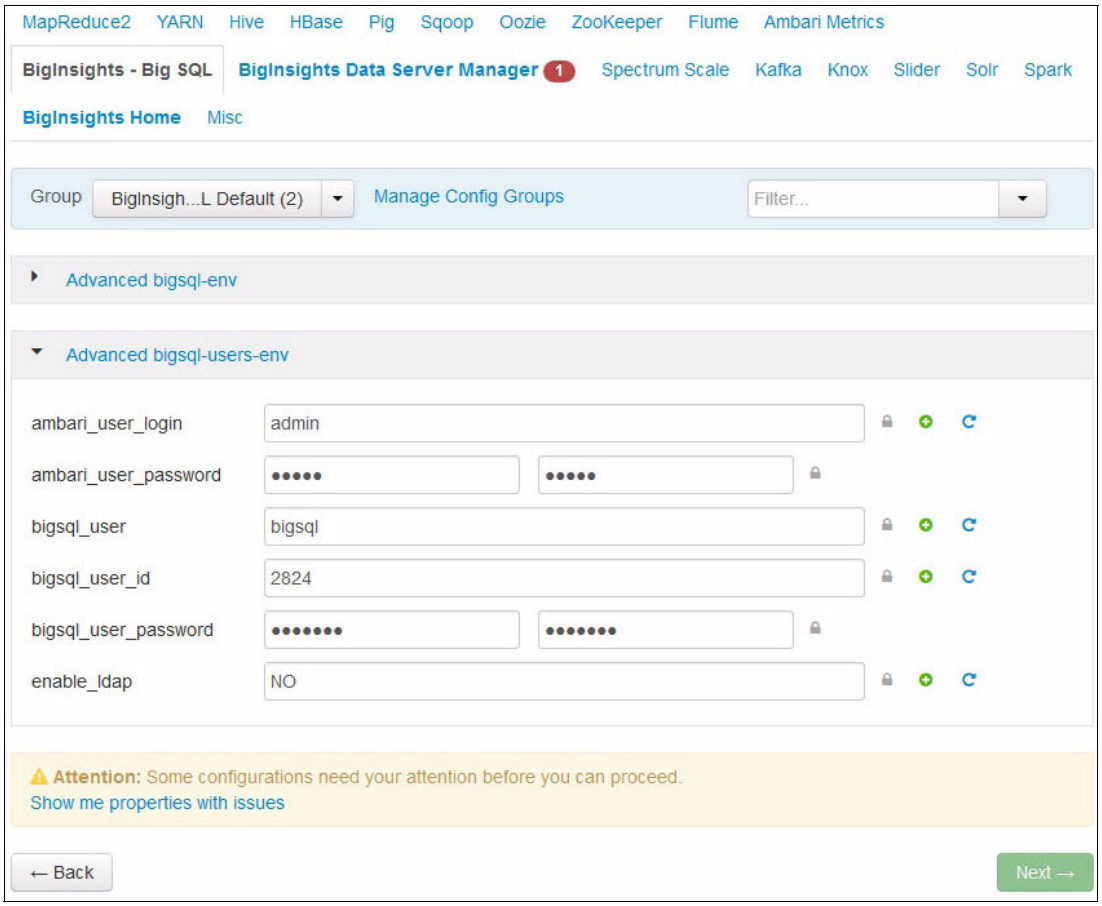


Figure 4-46 Ambari Add Service: Customize Big SQL

26. Figure 4-47 shows the Ambari Add Service window to customize the BigInsights Data Server Manager.

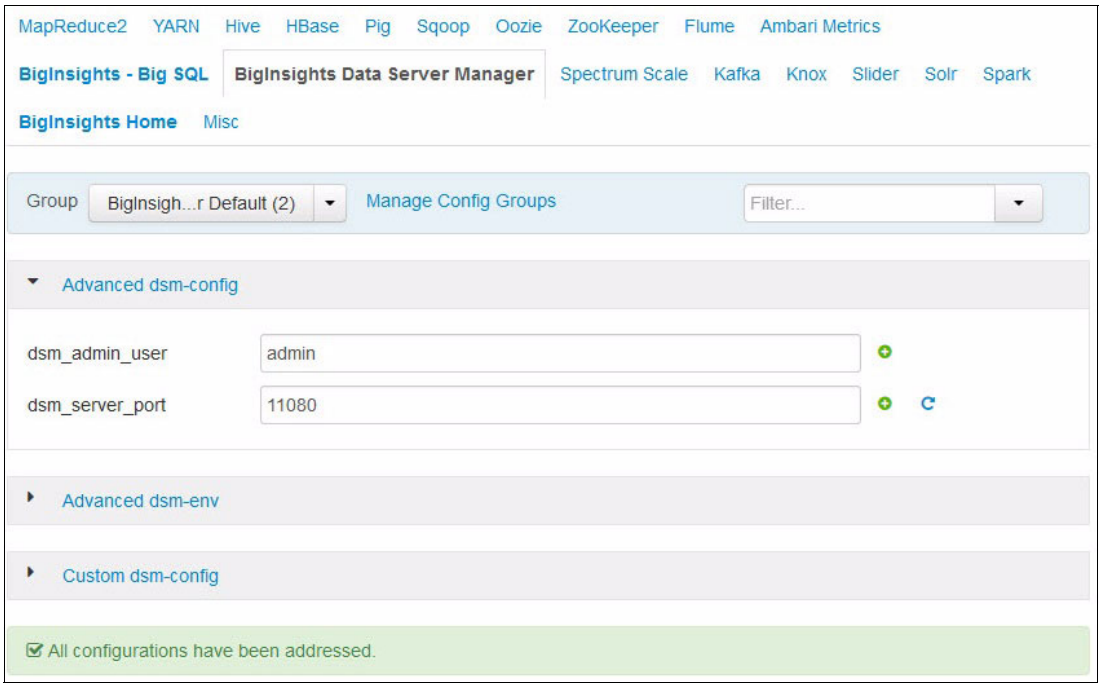


Figure 4-47 Ambari Add Service: Customize BigInsights Data Server Manager

27. Review and deploy, as shown in Figure 4-48.

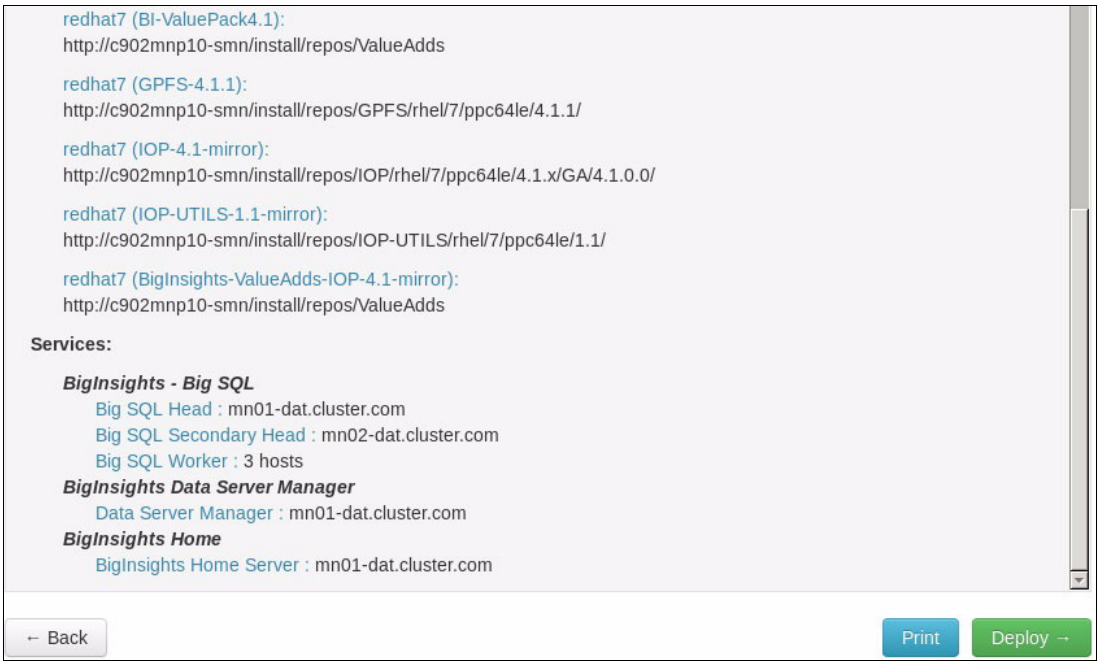


Figure 4-48 Review and deploy

28. Figure 4-49 shows the Ambari Install, Start and Test window.

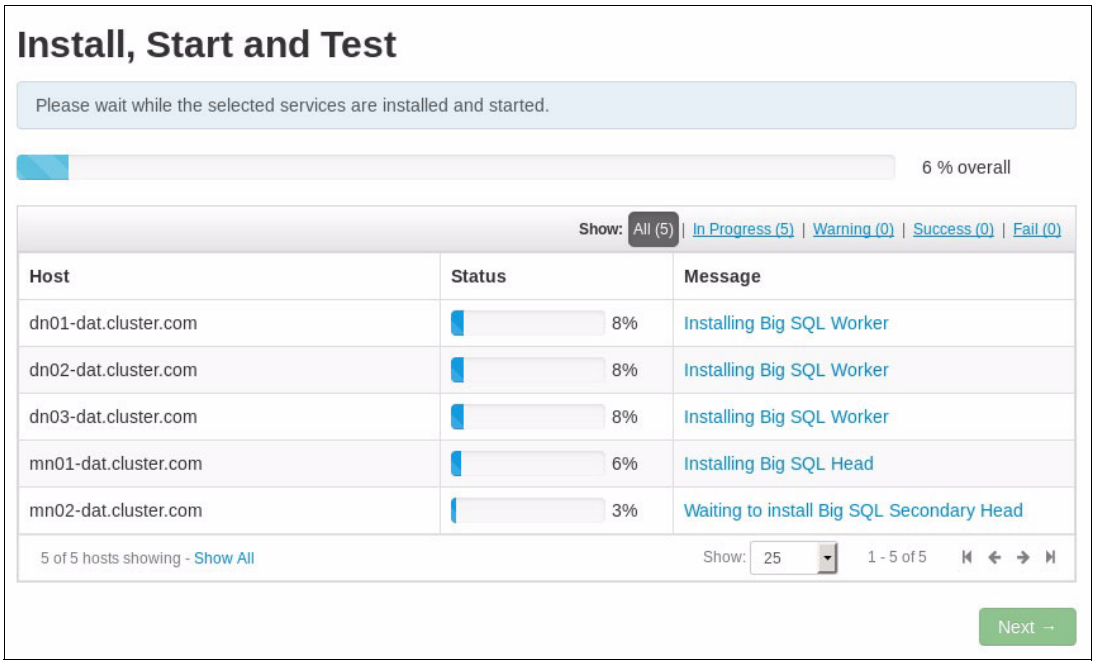


Figure 4-49 Ambari Add Service: Install and Test initial view

29. Figure 4-50 shows the Ambari Add Service window to check the installation and service.

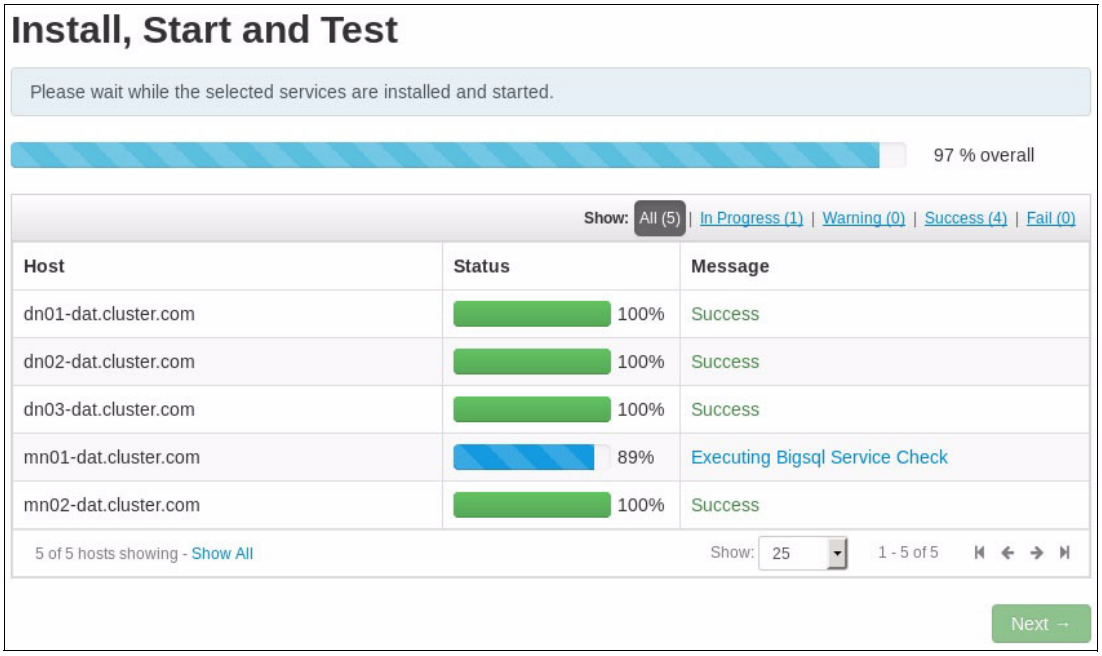


Figure 4-50 Ambari Add Service: Install and test service check

30. During the installation process, if any errors occur, you can review the logs from Ambari, correct the problem, and retry the installation. For example, if you forgot to comment out the require tty line in the `/etc/sudoers` file from all of the nodes, you will see the `error on tty sudo permission`.

    Also, you might see warnings in the logs that include a warning about the need to restart several affected services.

31. You must run **`knox_setup.sh`** to enable Knox for the value-add services. Follow the directions as explained in the "`Enabling Knox for value-add service`" section of the IBM BigInsights manual at the following website:

    `https://ibm.biz/Bd4ay5`

    If you fail to run the script, you will not find the updated JAR files, which are required by dsm, in the `/usr/iop/4.1.0.0/knox/lib/` directory.

    As part of the process, the script will restart Ambari and Knox.

32. You must start the Lightweight Directory Access Protocol (LDAP) for authentication. You can use the Knox Demo LDAP for the authentication, but you must start it before you access the BigInsights home url. The default authentication is `guest/guest-password`.

33. Before you continue with the BigInsights Big SQL configuration, you likely need to reapply a patch to fix a bug in the Ambari web GUI. This bug prevents you from saving core-site changes. See Figure 4-51.

```
/var/lib/ambari-server/resources/scripts/gpfs_core_site_patch.sh
```

Figure 4-51 Ambari web user interface (UI) patch

34. Reload the Ambari web interface on the browser.

35. From the Ambari dashboard, restart the HDFS, YARN, MapReduce2, and Big SQL services.

36. Restart the Knox Service. Start the Knox Demo LDAP service if you did not configure your own LDAP.

37. Restart the BigInsights Home services.

38. Follow the remaining steps from the document at the following website:

    `https://ibm.biz/Bd4ayW`

39. Access the Knox Gateway service by using the following URL:

    `https://mn02:8443/gateway/default/BigInsightsWeb/index.html`

4.3 DB2 with BLU Acceleration to store structured data

Structured data implies that data elements are stored according to a predefined data model. Sets of entities, tables, and files that are organized into attributes, fields, columns, and lines with predefined data types are examples of data models. A *data type* is the predefined type, length, and format of stored data. For example, a timestamp format might be represented as *YYYY-MM-DD HH:mm:SS*, and an instance of that data type is `2015-11-05 11:00:00`. Every instance of one entity, table, or file is considered a new record, row, or line.

A *spreadsheet*, which is an example of structured data, is a set of tables where every cell is an intersection of a column and a row. A *variant data type* is also considered a data type based on its definition, which is the most common data type of spreadsheet cells.

A *relational database*, which is another example of structured data, consists of a set of tables that are organized into rows according to the columns' predefined data types. In addition, relational databases enforce relationship constraints between tables to establish the consistency of data across the database.

Unstructured data has no predefined data model. However, it can be scanned and analyzed to provide the required data. Text data, for example, a digital copy of a contract, is considered unstructured data because the established date of the contract is not necessarily described in a predefined field or format and it can be scanned and found throughout that file.

You might ask yourself about something in between structured and unstructured data. *Semi-structured data* is unstructured data that is combined with metadata that provides tags and instructions for the position, format, length, or type of a specific data element to be addressed within the unstructured data source.

In this context, IBM DB2 with BLU Acceleration stores structured data correctly and speeds up the analytic workloads of your organization. It delivers unparalleled performance improvements for analytic applications and reporting by using dynamic in-memory optimized columnar technologies. Although the industry is abuzz with discussions about in-memory columnar data processing, BLU Acceleration offers so much more. It delivers significant improvements in database compression but it does not require you to have all your data in memory.

DB2 with BLU Acceleration includes several features that work together to make it a significant advancement in technology. We refer to these features as the BLU Acceleration Seven Big Ideas:

- Simplicity and ease of use
- Column store
- Adaptive compression
- Parallel vector processing
- Core-friendly parallelism
- Scan-friendly memory caching
- Data skipping

BLU Acceleration is simple and easy to use. The required effort to deploy and maintain a BLU Acceleration environment is minimal. Advanced technologies, such as columnar compression, parallel vector processing, core-friendly parallelism, scan-friendly memory caching, and data skipping are all used by DB2 automatically without database administrators (DBAs) explicitly deploying auxiliary structures for it to work. It is in the DB2 engine's nature to process queries by using these technologies.

At the center of BLU Acceleration is the column-organized table store. It is combined with actionable compression that operates on a column and page level to save storage space. The column organization eliminates the need for creating and maintaining secondary indexes and aggregates. In DB2 10.5, both column-organized and traditional row-organized tables can coexist in the same database. For optimal performance, run analytical queries against tables that are all column-organized.

For users who intend to convert existing tables to facilitate their analytic processing needs in a mixed workload environment, we suggest that you choose the use of BLU Acceleration only on those tables that are used purely for analytics. The **db2convert** utility converts row-organized tables to column-organized tables, while source tables remain accessible online.

With our client base that uses BLU Acceleration, clients experienced an average, conservatively, of approximately 10 times compression on their analytics databases, without any complex tuning. In terms of performance, their queries ran 35 - 73 times faster on average (and several queries are even faster).

For example, workloads with the following characteristics benefit most from BLU Acceleration:

- Analytical, data mart workloads
- Queries that involve grouping, aggregation, range scans, and joins
- Queries that access only a subset of the columns in a table
- Star or dimensional schemas
- SAP Business Warehouse application workloads

4.3.1 DB2 system requirements

The BLU Acceleration feature in DB2 10.5 is supported on AIX and Linux on Power (Table 4-2). It uses the same DB2 10.5 minimum operating system requirements.

For recent information about DB2 system requirements, consult the general documentation at the following website:

`http://www.ibm.com/support/docview.wss?uid=swg27038033#105`

Table 4-2 Suggestions for IBM DB2 with BLU Acceleration for IBM Power Systems

Operating system	Minimum required version	Suggested hardware
Linux little endian	► Red Hat Enterprise Linux (RHEL) Server 7.1 ► SUSE Linux Enterprise Server 12	IBM POWER8 or later
AIX	► AIX 6.1 Technology Level (TL) 7 ► AIX 7.1 TL1	IBM POWER8 or later

For more information about sizing system resources, see *Best practices: Optimizing analytic workloads using DB2 10.5 with BLU Acceleration* at the following website:

`http://bit.ly/1s5HCX5`

Note: For demonstration, we refer to the GOSALES Cognos sample database, which is in the IBM Cognos samples installation. This database can be stored in less than 10 GB, but we suggest at least 10 GB of disk space for the DB2 BLU database to fully reproduce our demonstration environment.

4.3.2 DB2 license requirements and functionality

DB2 for Linux, UNIX, and Windows Version 10.5 is available in multiple product editions. Each edition includes a different feature number and provides the functionality that we describe in this section.

In terms of required license entitlements, the BLU Acceleration feature includes the following DB2 10.5 editions for production environments:

- Advanced Enterprise Server Edition (AESE)
- Advanced Workgroup Server Edition (AWSE)

Non-production environments can use the following DB2 10.5 edition that also entitles the use of BLU Acceleration:

- Developer Edition (DE)

The license files for DB2 Version 10.5 ship separately for convenience so that you can download the license file in less time due to its small size. You need to download the license activation key from Passport Advantage and then install it.

Contact the Passport Advantage eCustomer Care team for assistance if you encounter problems. See the following website:

https://ibm.biz/BdF2Na

Ensure that you download the corresponding activation key part number for your IBM DB2 edition. Also, download this part number, CN30CML, IBM DB2 BLU Acceleration In-Memory Offering - Quick Start and Activation 10.5.0.5 for Linux, UNIX, and Windows, which includes a `db2baf.lic` file that is required for the BLU Acceleration activation.

4.3.3 IBM DB2 with BLU Acceleration deployment

At the time of writing this publication, IBM DB2 version 10.5.0.5 was compatible with Red Hat Enterprise (RHEL) Server 7.1 on Power System (little endian). Therefore, in addition to all of the unique benefits of IBM DB2 with BLU Acceleration on Power Systems, your existing databases do not need to be migrated to a big endian environment.

For the next set of instructions, we assume the use of the same configuration as before for the demonstration.

Install xlC package

We suggest that you configure a yum repository by using packages from the installation CD. Therefore, you can smoothly install and resolve dependencies for xlC package by using the command that is shown in Figure 4-52.

```
yum install libxlc.ppc64le
```

Figure 4-52 Using yum install libxlc for RHEL on Power little endian

However, if you are not configuring a yum repository, you can also download the xlc package. Ensure that you download a ppc64le extension, which is the correct package for little endian. Use the following URL to download the package:

https://ibm.biz/Bd4aMF

Also, you can use the **rpm** command as an alternative installation method but you must resolve the dependencies manually (Figure 4-53).

```
rpm -ivh libxlc-13.1.2.0-150526a.ppc64le.rpm
```

Figure 4-53 The rpm install libxlc for RHEL on Power little endian

Optionally install the libaio package

We suggest that you use the **yum** command to install the libaio package.

See Example 4-1.

Example 4-1 The yum install libaio output for RHEL on Power Systems little endian

```
[root@dn05-dat server]# yum install libaio
Loaded plugins: product-id, subscription-manager
This system is not registered to Red Hat Subscription Management. You can use subscription-manager to register.
GPFS-4.1.1.1                                                                                                | 2.9 kB  00:00:00
local-rhels7.1-ppc64le-ppc64le                                                                              | 4.1 kB  00:00:00
Resolving Dependencies
--> Running transaction check
---> Package libaio.ppc64le 0:0.3.109-12.ael7b will be installed
--> Finished Dependency Resolution

Dependencies Resolved

================================================================================================================================
 Package              Arch                 Version                       Repository                                    Size
================================================================================================================================
Installing:
 libaio               ppc64le              0.3.109-12.ael7b              local-rhels7.1-ppc64le-ppc64le                24 k

Transaction Summary
================================================================================================================================
Install  1 Package
Total download size: 24 k
Installed size: 158 k
Is this ok [y/d/N]: y
Downloading packages:
libaio-0.3.109-12.ael7b.ppc64le.rpm                                                                         |  24 kB  00:00:00
Running transaction check
Running transaction test
Transaction test succeeded
Running transaction
Warning: RPMDB altered outside of yum.
  Installing : libaio-0.3.109-12.ael7b.ppc64le                                                                 1/1
  Verifying  : libaio-0.3.109-12.ael7b.ppc64le                                                                 1/1

Installed:
  libaio.ppc64le 0:0.3.109-12.ael7b

Complete!
```

Download the DB2 installation package

To install the DB2 with BLU Acceleration trial software on your own platforms, download the 90-day no-charge trial software from the Get DB2 with BLU Acceleration website:

http://www.ibmbluhub.com/get-blu/

For our environment, we use Advanced Enterprise Server Edition (AESE).

Note: We suggest that you download the latest fix pack for your own platform at Download DB2 Fix Packs by version for DB2 for Linux, UNIX, and Windows at the following website:

https://ibm.biz/Bd4aMX

Initial setup

Follow these steps:

1. Ensure that you enabled X11 forwarding on both the client side and the server side if you perform this installation remotely.

> **Note:** Check whether your network throughput is configured optimally because rendering the X application remotely over Secure Shell (SSH) might slow down due to network latency.

2. To start the setup, log in as the root credential. Go to the directory where you extracted the DB2 installation package and run **`db2setup`**, which is an X application. Alternatively, you can use **`db2_install`**, which is a console mode setup.
3. For **`db2setup`**, on the Welcome window (Figure 4-54), click **Install a Product** and then on **DB2 Version 10.5 Fix Pack 5 Workgroup**. For Enterprise and Advanced Editions, click **Install New**.

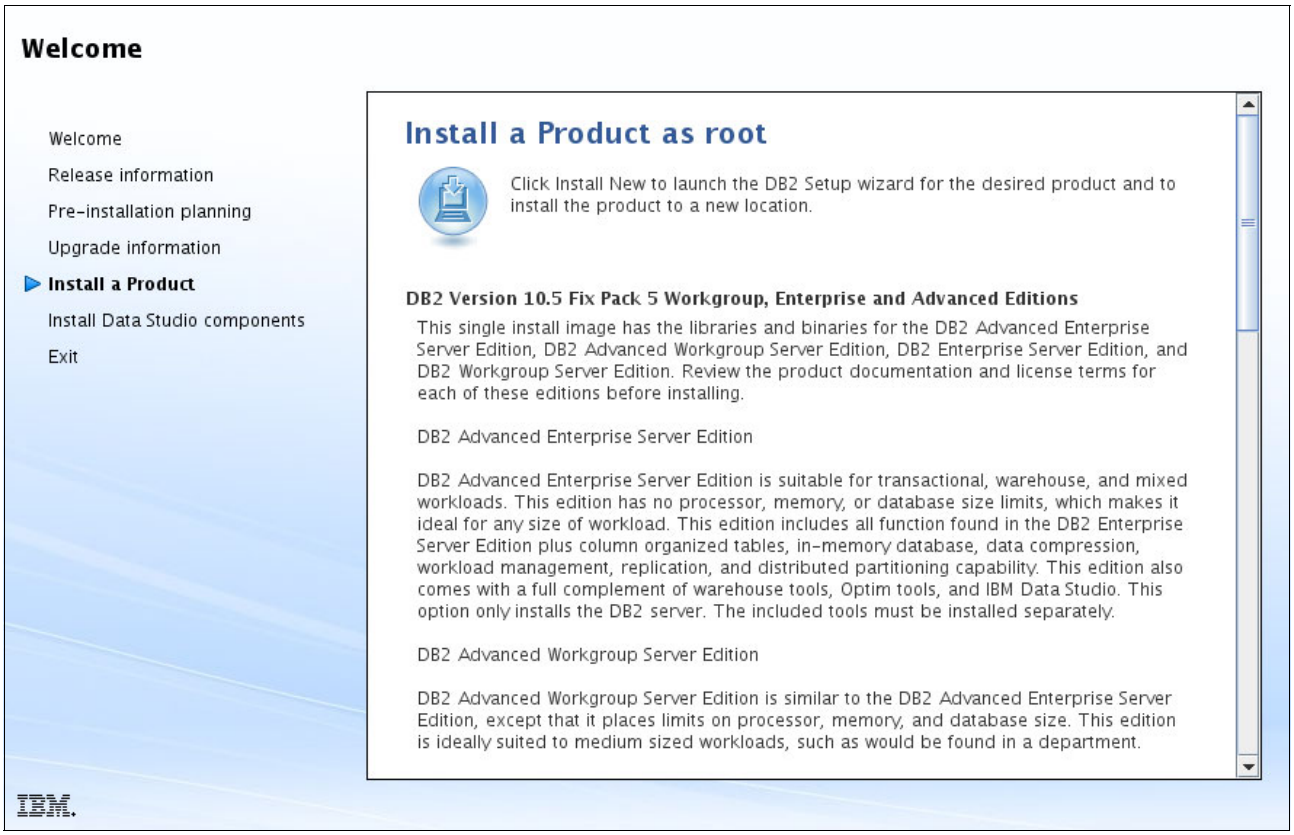


Figure 4-54 Install a product as root for DB2 Version 10.5 Fix Pack 5 Workgroup, Enterprise, and Advanced Editions

4. The following options are shown on the left side of Figure 4-55 on page 91:
   a. On the first setup window, select **1. Introduction**. On the Welcome to DB2 Setup Wizard window, click **Next**.
   b. Click **2. Software License Agreement**. Read the agreement carefully and check **I accept the terms** on the license agreement window. Click **Next**.
   c. Select **3. Installation type**. We encourage you to use the typical installation and click **Next**, but you can customize for your environment.

d. Select **4. Installation action**. Keep the default setting because it creates a response file `/root/db2server.rsp`, which can be reviewed later or used to reinstall on the same or another environment. Click **Next**.

e. Click **5. Installation directory**. We suggest that you keep the default directory, which is `/opt/ibm/db2/V10.5`, but you can change the installation directory to your own preference. Click **Next**.

f. Click **6. DAS user**. We will not use or refer to any DB2 Administration Server feature in this book. We advise that you check **Create the DAS user later**. Click **Next**.

g. Click **7. Instance setup**. Check **Create a DB2 instance**, which is the environment on which you will store data and run applications. You must have an instance to use this product. Click **Next**.

h. Click **8. Partitioning options**. For this demonstration, we do not use multiple partitions. We suggest that you click **Single partition instance**. If you still want or need to use a multiple partition instance, you must have a Database Partitioning Feature license, also. Click **Next**.

i. Click **9. Instance-owning user**. Create a DB2 instance owner with the default user name, which is `db2inst1` under the `db2iadm1` group (Figure 4-55) or change it according to your environment's or organization's standards. Click **Next**.

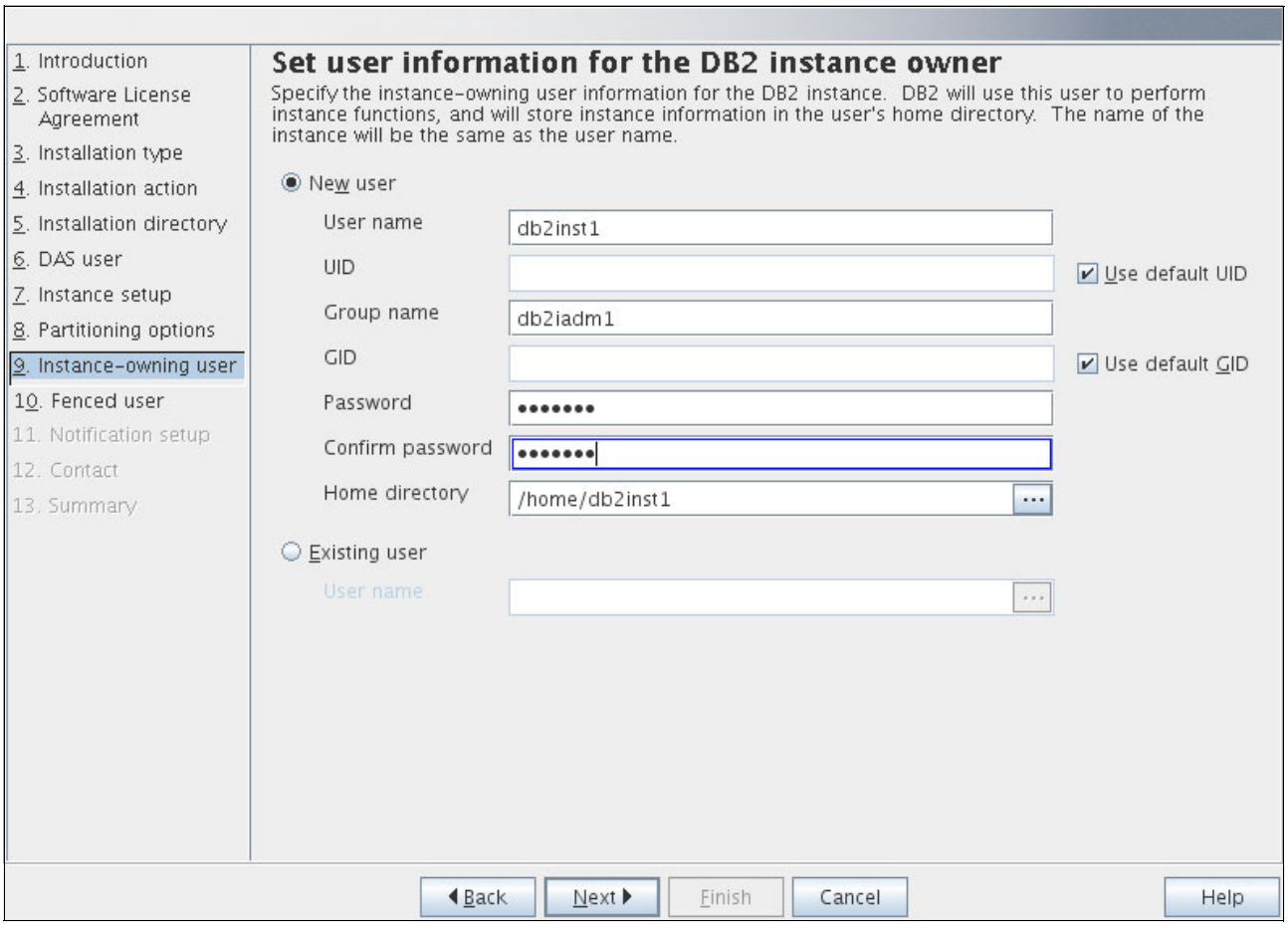


Figure 4-55 Default DB2 instance owner setup

j. Click **10. Fenced user**. Use the default setting because a new fenced user will be created. Fenced user-defined procedures (UDFs) and stored procedures will execute under this user and group (Figure 4-56). Or, change it according to your environment's or organization's standards. Click **Next**.

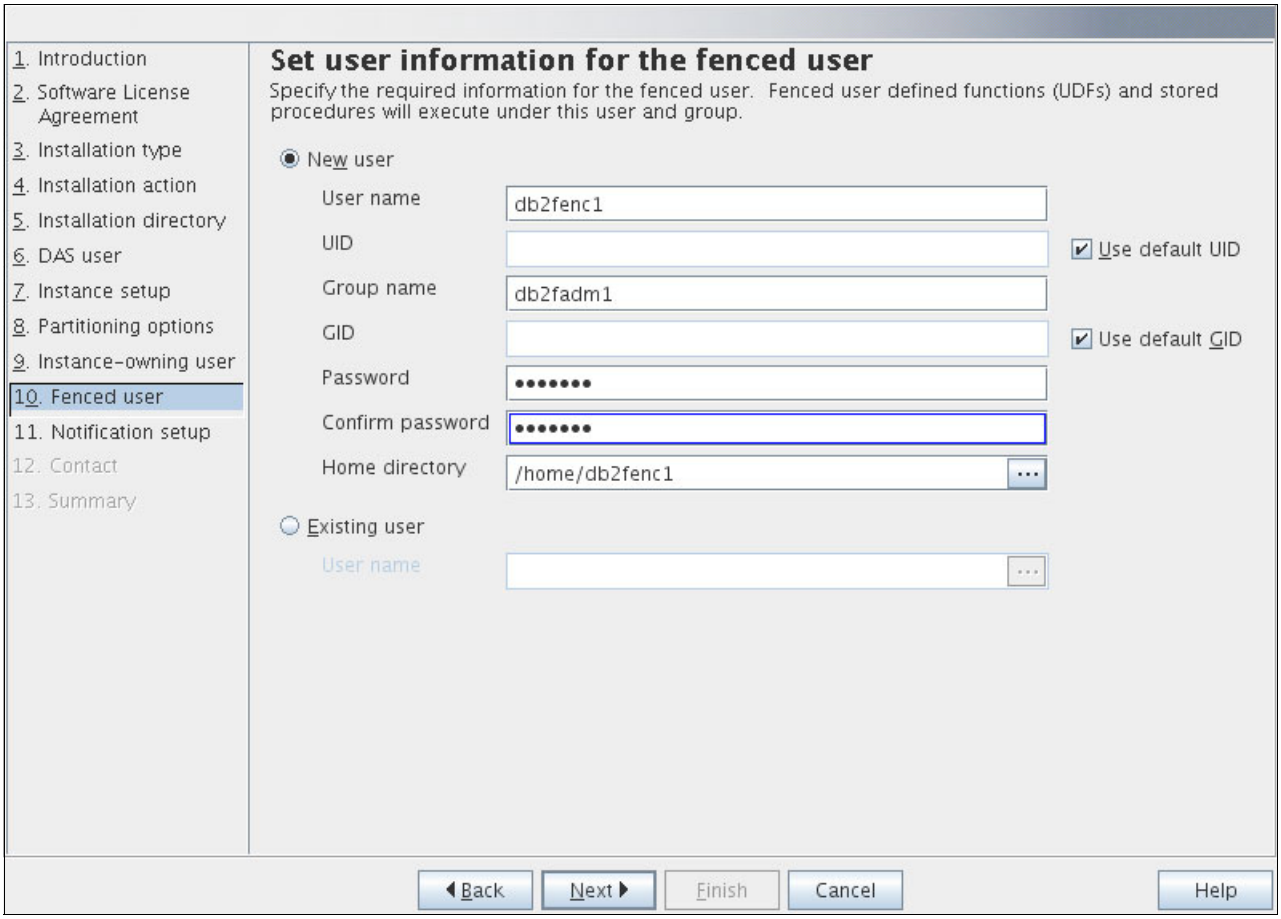


Figure 4-56 Default DB2 fenced user setup

k. Click **11. Notification setup**. Select **Do not set up your DB2 server to send notification** this time, but you can set up your DB2 server to automatically send email or pager notifications to alert administrators when a database needs attention. Use your environment's or corporate Simple Mail Transfer Protocol (SMTP) server in this latter case. Click **Next**.

l. Click **12. Summary**. Review your settings and click **Finish** to start the installation.

Note: As a preferred practice, install the latest fix pack (FP). At the time of writing this publication, the latest fix pack is FP6. This fix pack can be installed from the previously downloaded and then extracted directory, specifying the path where the DB2 database product was installed, *<db2_base_install_path>*:

```
./installFixPack -b <db2_base_install_path>
```

For example, this command installs the fix pack:

```
./installFixPack -b /opt/ibm/db2/V10.5
```

Activate DB2 with BLU Acceleration by using the db2baf.lic license file

After the installation, you are required to activate DB2 BLU Acceleration with the correct DB2 server edition license.

To activate the license, retrieve the `db2baf.lic` file from the previously downloaded part CN30CML and execute the following command (Example 4-2):

`<db2_base_install_path>/adm/db2licm -a db2baf.lic`

Example 4-2 Activating the license for DB2 with BLU Acceleration

```
[root@dn05-dat adm]# /opt/ibm/db2/V10.5/adm/db2licm -a db2baf.lic

LIC1402I  License added successfully.


LIC1426I  This product is now licensed for use as outlined in your License
Agreement.  USE OF THE PRODUCT CONSTITUTES ACCEPTANCE OF THE TERMS OF THE IBM
LICENSE AGREEMENT, IN THE FOLLOWING DIRECTORY:
"/opt/ibm/db2/V10.5/license/en_US.iso88591"
```

4.3.4 Set up the DB2 instance

To enable the analytics workload and remote TCP/IP connection to the DB2 database, you must update two parameters.

Log in as the instance owner, by default, `db2inst1`, and execute the following commands:

```
su - db2inst1
db2set DB2_WORKLOAD=ANALYTICS
db2set DB2COMM=TCPIP
```

For the DB2_WORKLOAD registry parameter to become active, the DB2 instance must be restarted after you set DB2_WORKLOAD=ANALYTICS. Restart the instance with the following commands:

```
db2stop
db2start
```

Note: The default port number for db2 instances is 50000. However, that port number might vary depending on your installation method or whether it is in use by another service on the same server.

Therefore, you can check the correct port number for your db2 instance by performing the following command:

```
db2 get dbm cfg | grep SVCENAME
```

The output shows the service name. Then, you can verify the port number in the `/etc/services` file.

4.3.5 GOSALES Cognos Business Intelligence sample database

For our demonstration, create the GOSALES database by using the db2 instance owner credentials according to your environment. This example uses a Spectrum Scale file system to store data files: `/bigpfs/dbpath/gs_db`:

```
mkdir /bigpfs/dbpath/gs_db
```

Create a DB2 database under the previously created directory. As part of a new database creation process, the configuration advisor automatically applies all required BLU Acceleration settings to optimize analytic workloads. Optionally, all required BLU Acceleration settings can also be applied explicitly through the **AUTOCONFIGURE** keyword as part of the database creation:

```
db2 CREATE DB GS_DBBLU ON /bigpfs/dbpath/gs_db AUTOCONFIGURE USING mem_percent 80
APPLY DB AND DBM
```

The expected output of this command is shown in Example 4-3.

Example 4-3 Output of the DB2 create command

```
[db2inst1@dn05-dat dbpath]$ db2 CREATE DB GS_DBBLU ON /bigpfs/dbpath/gs_db AUTOCONFIGURE USING
mem_percent 80 APPLY DB AND DBM
        Former and Applied Values for Database Manager Configuration

 Description                                    Parameter   Former Value        Applied Value
 ---------------------------------------------------------------------------------------------
 Application support layer heap size (4KB)   (ASLHEAPSZ) = 15                 15
 No. of int. communication buffers(4KB)(FCM_NUM_BUFFERS) = AUTOMATIC(4096)     AUTOMATIC(21580)
 Enable intra-partition parallelism     (INTRA_PARALLEL) = NO                 NO
 Maximum query degree of parallelism   (MAX_QUERYDEGREE) = ANY                ANY
 Agent pool size                        (NUM_POOLAGENTS) = AUTOMATIC(100)    AUTOMATIC(100)
 Initial number of agents in pool       (NUM_INITAGENTS) = 0                 0
 Max requester I/O block size (bytes)          (RQRIOBLK) = 65535            65535
 Sort heap threshold (4KB)                    (SHEAPTHRES) = 0               0


        Former and Applied Values for Database Configuration

 Description                                    Parameter   Former Value        Applied Value
 ---------------------------------------------------------------------------------------------
 Default application heap (4KB)              (APPLHEAPSZ) = 256                256
 Catalog cache size (4KB)               (CATALOGCACHE_SZ) = (MAXAPPLS*5)       360
 Changed pages threshold                 (CHNGPGS_THRESH) = 60                 80
 Database heap (4KB)                             (DBHEAP) = AUTOMATIC(1200)    AUTOMATIC(11605)
 Degree of parallelism                       (DFT_DEGREE) = 1                  ANY
 Default tablespace extentsize (pages)    (DFT_EXTENT_SZ) = 32                 32
 Default prefetch size (pages)          (DFT_PREFETCH_SZ) = AUTOMATIC(32)      AUTOMATIC(64)
 Default query optimization class          (DFT_QUERYOPT) = 5                  5
 Max storage for lock list (4KB)               (LOCKLIST) = 4096               AUTOMATIC(4096)
 Log file size (4KB)                          (LOGFILSIZ) = 1000               1024
 Number of primary log files                 (LOGPRIMARY) = 3                  45
 Number of secondary log files                (LOGSECOND) = 10                 19
 Max number of active applications             (MAXAPPLS) = AUTOMATIC(40)      AUTOMATIC(40)
 Percent. of lock lists per application        (MAXLOCKS) = 10                 AUTOMATIC(15)
 Number of asynchronous page cleaners    (NUM_IOCLEANERS) = AUTOMATIC(24)      AUTOMATIC(1)
 Number of I/O servers                    (NUM_IOSERVERS) = AUTOMATIC(196)     AUTOMATIC(5)
 Package cache size (4KB)                    (PCKCACHESZ) = (MAXAPPLS*8)       AUTOMATIC(2812)
 Sort list heap (4KB)                          (SORTHEAP) = 256                787556
 SQL statement heap (4KB)                      (STMTHEAP) = AUTOMATIC(8192)    AUTOMATIC(16384)
 Statistics heap size (4KB)                (STAT_HEAP_SZ) = AUTOMATIC(4384)    AUTOMATIC(4384)
 Utilities heap size (4KB)                 (UTIL_HEAP_SZ) = AUTOMATIC(5000)    AUTOMATIC(11434465)
 Self tuning memory                       (SELF_TUNING_MEM) = OFF              ON
     Automatic runstats                    (AUTO_RUNSTATS) = ON                ON
 Sort heap thres for shared sorts (4KB) (SHEAPTHRES_SHR) = 5000               15751124
```

```
Log buffer size (4KB)                     (LOGBUFSZ) = 256                 2195
Default table organization          (DFT_TABLE_ORG) = ROW                 COLUMN
Database memory threshold           (DB_MEM_THRESH) = 100                 100


       Former and Applied Values for Bufferpool(s)

Description                                  Parameter   Former Value        Applied Value
-------------------------------------------------------------------------------------------
IBMDEFAULTBP                          Bufferpool size = 1000                1968890


       Former and Applied Values for System WLM Objects

Description                                                 Former Value        Applied Value
-------------------------------------------------------------------------------------------
Work Action SYSMAPMANAGEDQUERIES Enabled                = Y                   Y
Work Action Set SYSDEFAULTUSERWAS Enabled               = Y                   Y
Work Class SYSMANAGEDQUERIES Timeroncost                = 1.50000E+05         1.50000E+05
Threshold SYSDEFAULTCONCURRENT Enabled                  = N                   Y
Threshold SYSDEFAULTCONCURRENT Maxvalue                 = 15                  15


DB21029I  The database was created successfully. Please restart the instance
so configuration changes take effect.
```

Restart the instance for the configuration changes to take effect with the following commands:

```
db2stop
db2start
```

4.3.6 SPSS Collaboration and Deployment Services database repository

SPSS Collaboration and Deployment Services version 7.0 is not compatible with DB2 with BLU Acceleration column-organized tables. During the database repository creation process, indexes are created to speed up data retrieval from the repository. However, column-organized tables in DB2 with BLU Acceleration do not require indexes. And, in fact, SPSS Collaboration and Deployment Services does not accept indexes. Therefore, to bypass this problem, the default table organization for the SPSS Collaboration and Deployment Services DB2 database must be set to row-organized tables after the database creation.

Important: The SPSS Collaboration and Deployment Services database repository must be created before the installation of the SPSS Collaboration and Deployment Services product.

We suggest that you create a new db2 instance that is optimized for an online transaction processing (OLTP) workload to create the SPSS Collaboration and Deployment Services repository because the previously configured db2inst1 instance, in our scenario, is optimized for an analytic workload to hold the GOSALES database. Even though SPSS is in essence an analytical package, its repository does not use the same analytical workload characteristics. For more information about creating a new instance, consult the **db2icrt** command syntax.

If you are comfortable reusing the same db2inst1 instance for the SPSS Collaboration and Deployment Services database repository, you can proceed, even though it is not fully optimized for its workload:

```
db2 CREATE DATABASE spsscds ON /bigpfs/dbpath/spsscds USING CODESET UTF-8
TERRITORY US COLLATE USING SYSTEM
db2 connect to spsscds
db2 "CREATE BUFFERPOOL CDS8K IMMEDIATE SIZE 250 AUTOMATIC PAGESIZE 8 K"
db2 "CREATE REGULAR TABLESPACE CDS8K PAGESIZE 8 K MANAGED BY AUTOMATIC STORAGE
EXTENTSIZE 8 OVERHEAD 10.5 PREFETCHSIZE 8 TRANSFERRATE 0.14 BUFFERPOOL CDS8K
DROPPED TABLE RECOVERY ON"
db2 "CREATE BUFFERPOOL CDSTEMP IMMEDIATE SIZE 250 PAGESIZE 32 K"
db2 "CREATE SYSTEM TEMPORARY TABLESPACE CDSTEMP PAGESIZE 32K MANAGED BY AUTOMATIC
STORAGE EXTENTSIZE 16 OVERHEAD 10.5 PREFETCHSIZE 16 TRANSFERRATE 0.14 BUFFERPOOL
CDSTEMP"
```

Change the default table organization to row-organized tables with the following command:

```
db2 update db cfg for spsscds using DFT_TABLE_ORG row
```

4.4 SPSS Analytical Decision Management

This section describes the steps to install SPSS Analytical Decision Management.

4.4.1 Outline of steps

To set up this environment, the following steps are required:

1. Install the prerequisite items for AIX.
2. Install the Installation Manager.
3. Install and configure WebSphere Application Server.
4. Install and configure the DB2 database.
5. Install and configure SPSS Modeler Server.
6. Install and configure IBM SPSS Collaboration and Deployment Service.
7. Install and configure IBM SPSS Modeler Server Adapters for Collaboration and Deployment Services.

4.4.2 Install the prerequisite items for AIX

In this step, proceed to check and install all of the AIX prerequisites.

4.4.3 Install the Installation Manager

Follow these steps to deploy the Installation Manager:

1. Execute **IBMIM**.
2. Wait until the IBM Installation Manager wizard opens.
3. Click **File** → **Preferences**.
4. Click **Add Repository**.
5. Click **Browse**. Select the directory where you stored the installation file and locate the repository files. Click **OK**.

6. Set all repositories:
   - IBM WebSphere 8.5.5
   - IBM SPSS Collaboration and Deployment Service 7.0.0
   - IBM SPSS Modeler Adopter 17.0
   - IBM SPSS Analytical Decision Management 17.0
7. Check all repositories by clicking **Test Connection**. Check whether the message "`All the selected repositories are connected`" is displayed, as shown in Figure 4-57. Click **OK**.

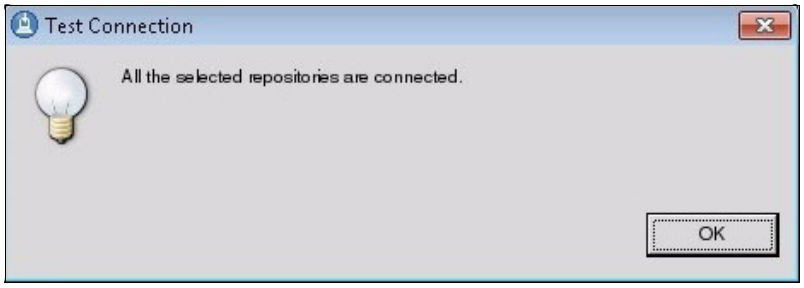


Figure 4-57 Test Connection

8. Click **OK**.

4.4.4 Install and configure WebSphere Application Server

This section describes the installation and configuration steps to deploy the application server.

Install WebSphere Application Server

Follow the installation steps:

1. Click **Install** (Figure 4-58).

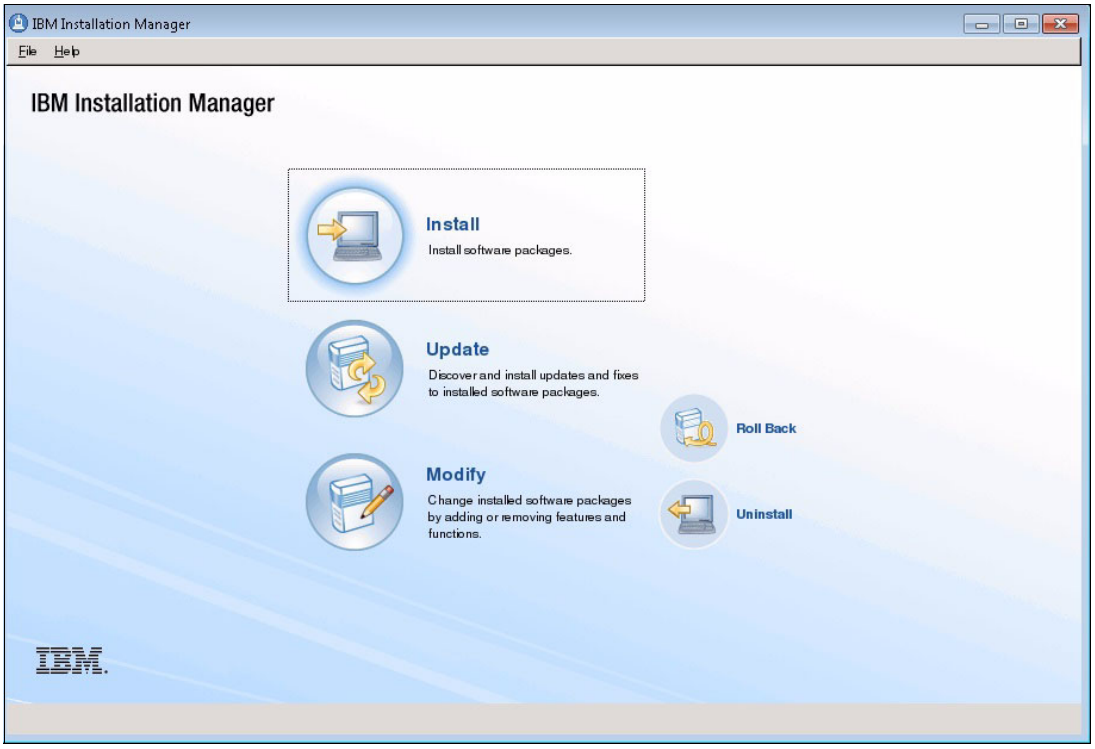


Figure 4-58 IBM Installation Manager

2. Select the items that you want to install. Select **IBM WebSphere Application Server Version 8.5.5.0**. After you select the software, the status changes to "`Will be installed`". Click **Next**.
3. Select **I accept the terms in the license agreement** for the Licenses process in Installation Packages.
4. Click **Next** for the Location in the Installation Packages, which is for Shared Resources.
5. For the Directory, click **Next**. Click **Browse**. Select the path. Click **Next**.
6. Click **Next** for Feature in Installation Package.
7. Click **Next** for Feature → Translation in Install Package.
8. Click **Next** for Feature in Install Package.
9. Click **Install**.
10. Click **Next**.
11. Check **Review Summary** for the Summary of the Install Package.
12. Click **Install**.
13. Click **Finish**.

Configure

Use the following steps to configure the application server:

1. Click **Create**.
2. Select **Application server** for the environment to create, as shown in Figure 4-59. Click **Next**.

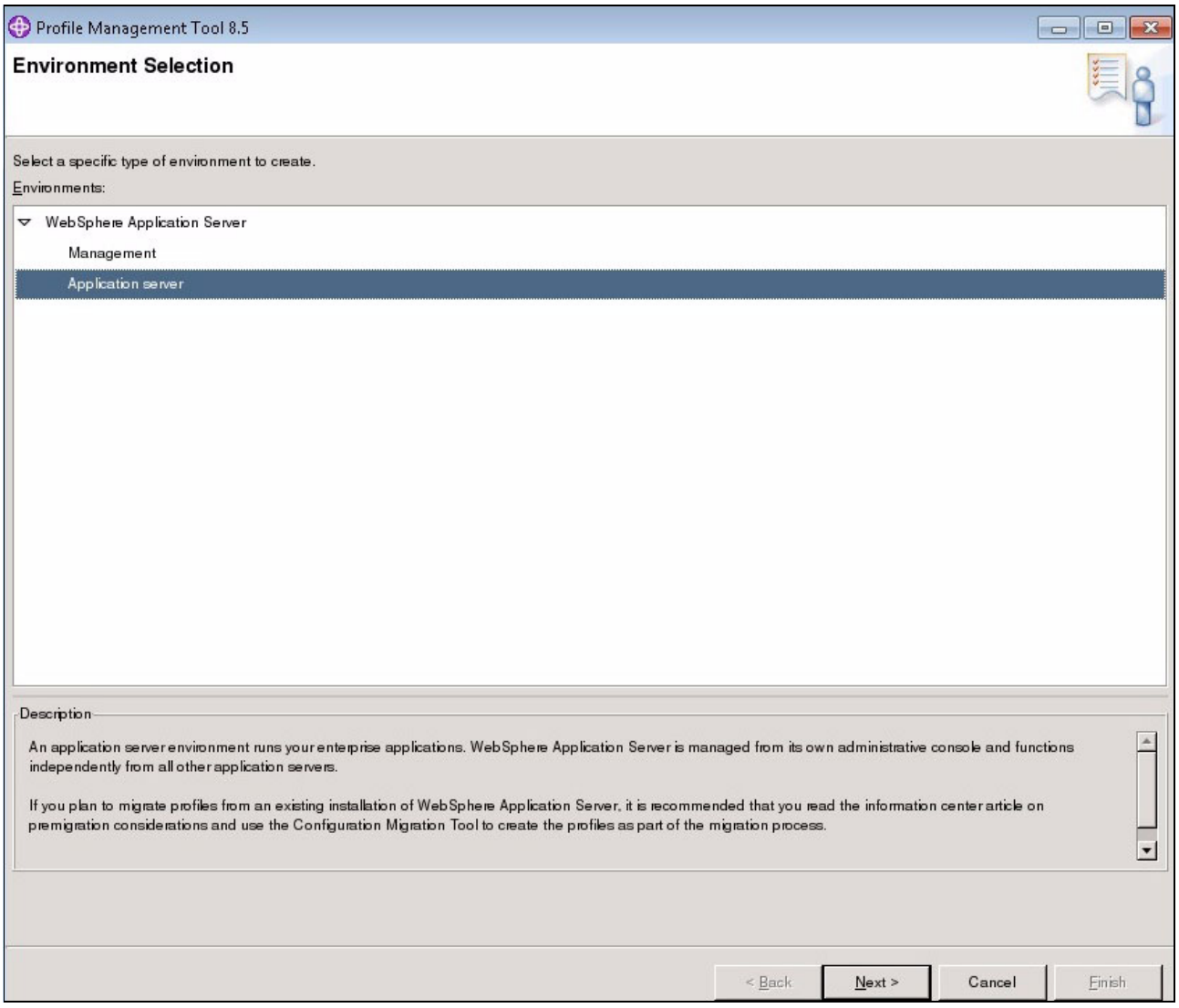


Figure 4-59 WebSphere Environment Selection

3. Figure 4-60 shows the WebSphere profiles. Click **Next**.

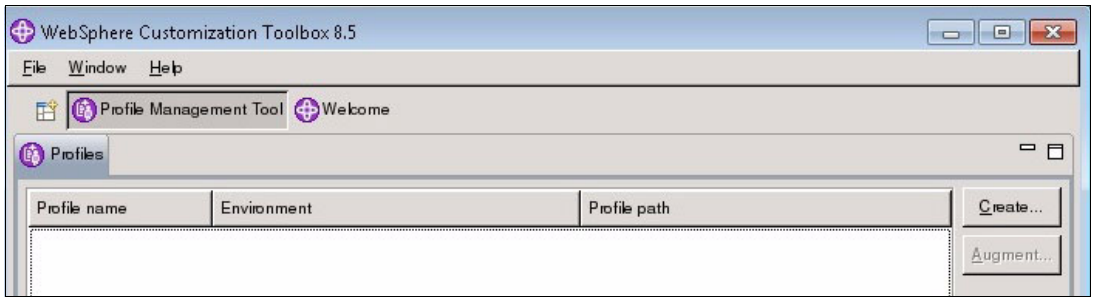


Figure 4-60 WebSphere Profiles

4. Select **Typical profile creation**, as shown in Figure 4-61. Click **Next**.

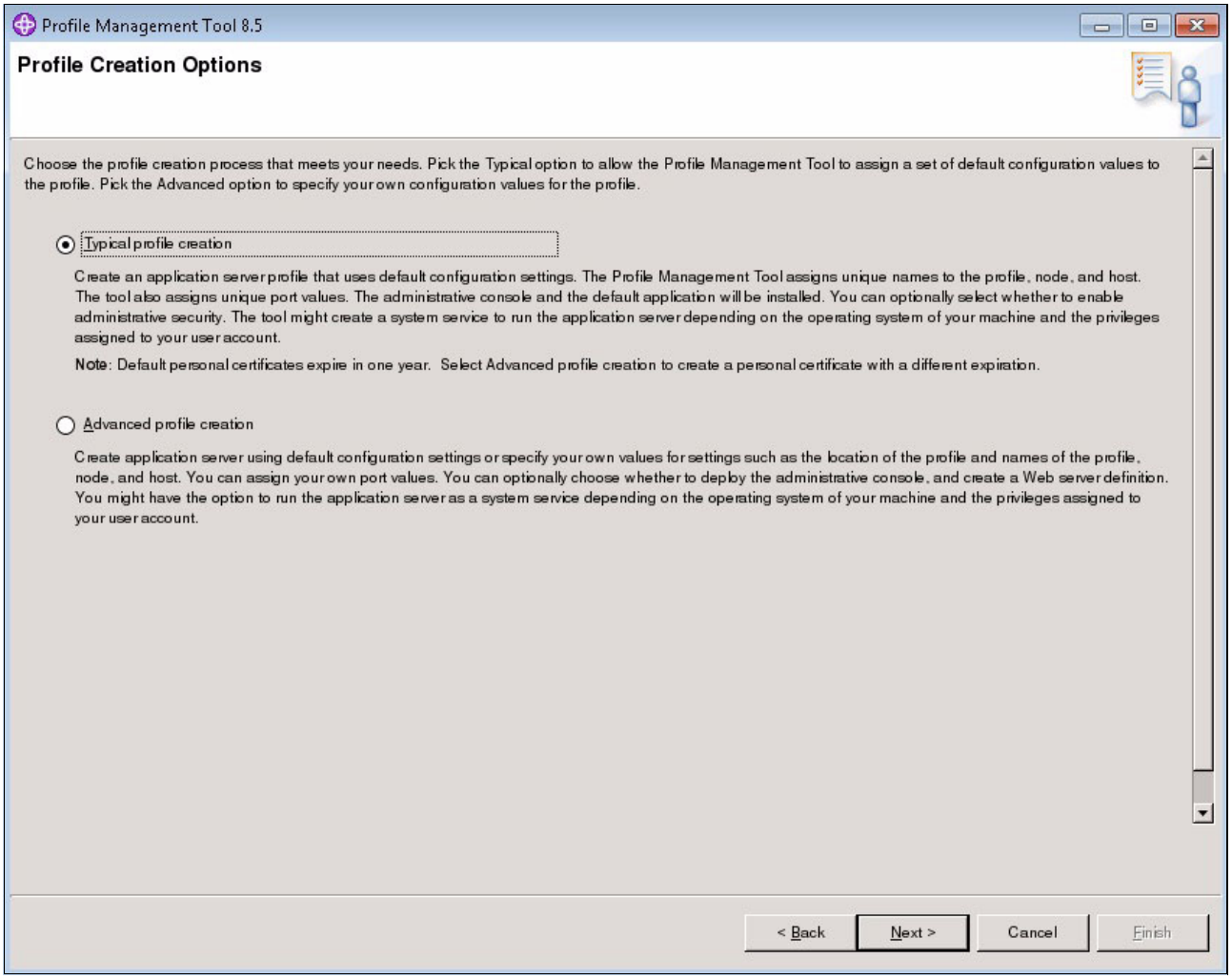


Figure 4-61 WebSphere Profile Creation Options

5. Set the user name and password for administrative security (Figure 4-62). For this example, we set the user name to `admin` and the password to `ibm1ibm`. For the other items, leave them as shown. Click **Next**.

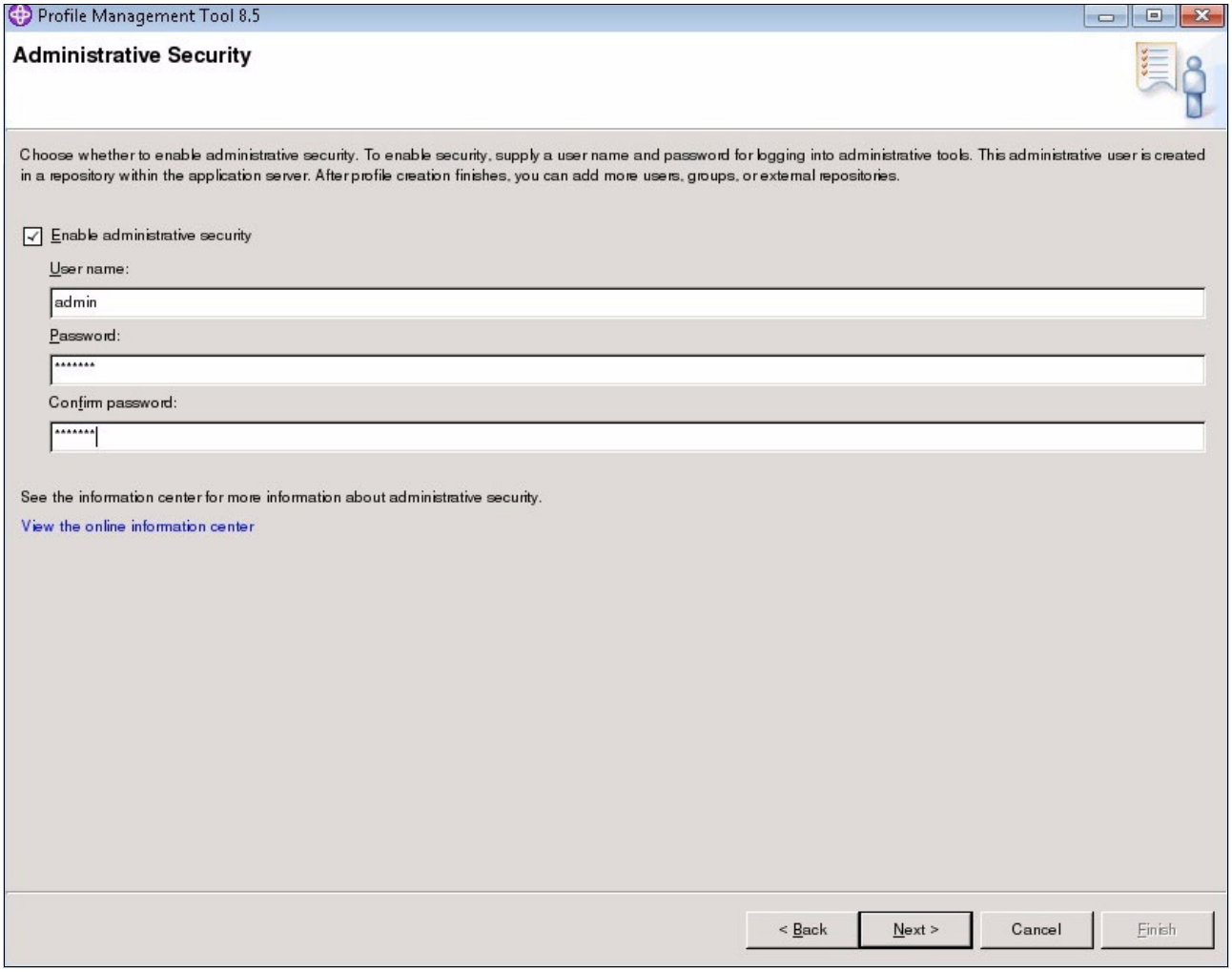


Figure 4-62 WebSphere Administrative Security

6. Check the content in the Profile Creation Summary, as shown in Figure 4-63. If it is correct, click **Create**.

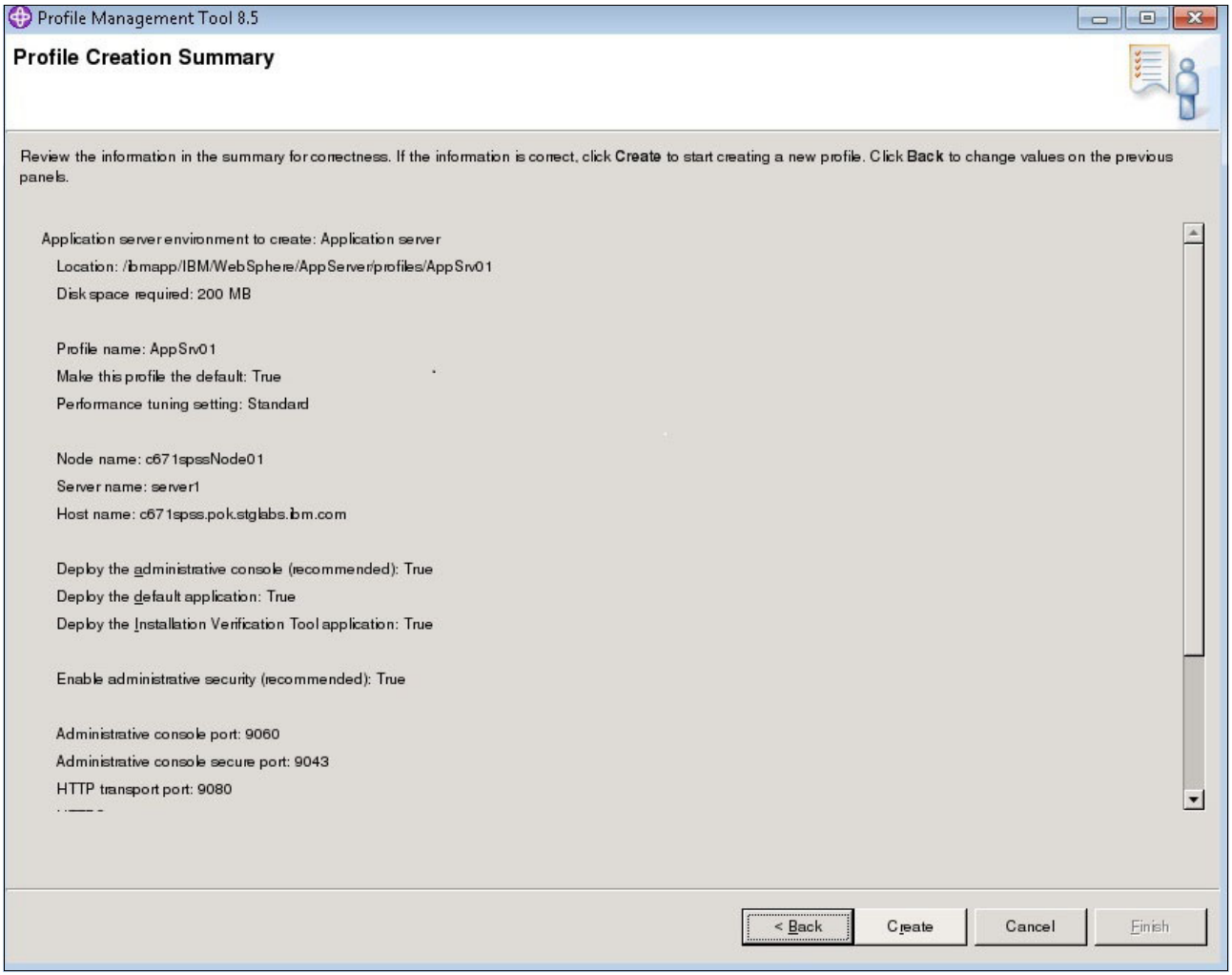


Figure 4-63 WebSphere Profile Creation Summary

Note: In Figure 4-63, the HTTP transport port is 9080. Port 9080 is the default, and it was used in our sample.

7. Click **Finish**, as shown in Figure 4-64.

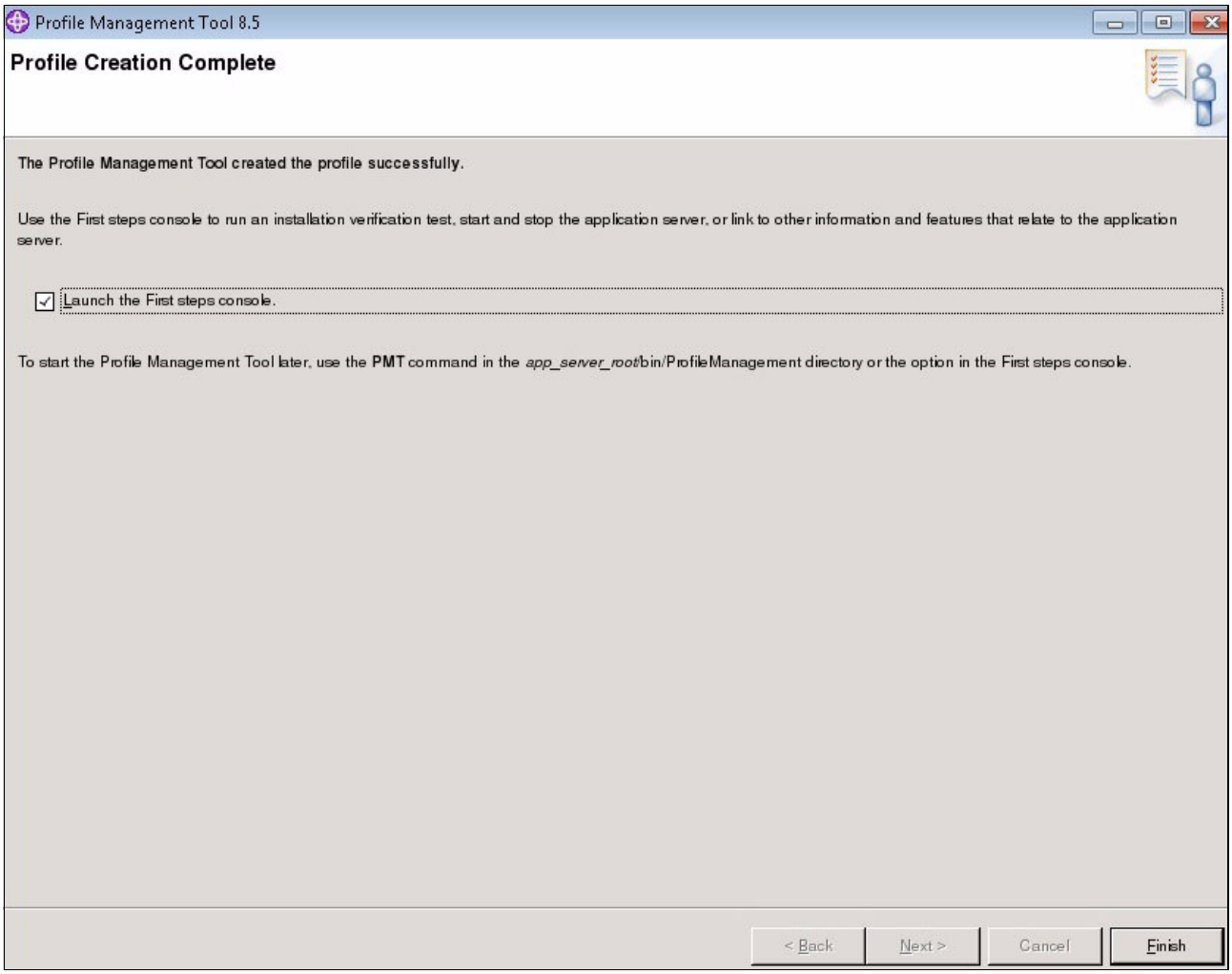


Figure 4-64 WebSphere Profile Creation Complete

8. Wait until the First steps window opens, as shown in Figure 4-65. Click **Installation verification**.

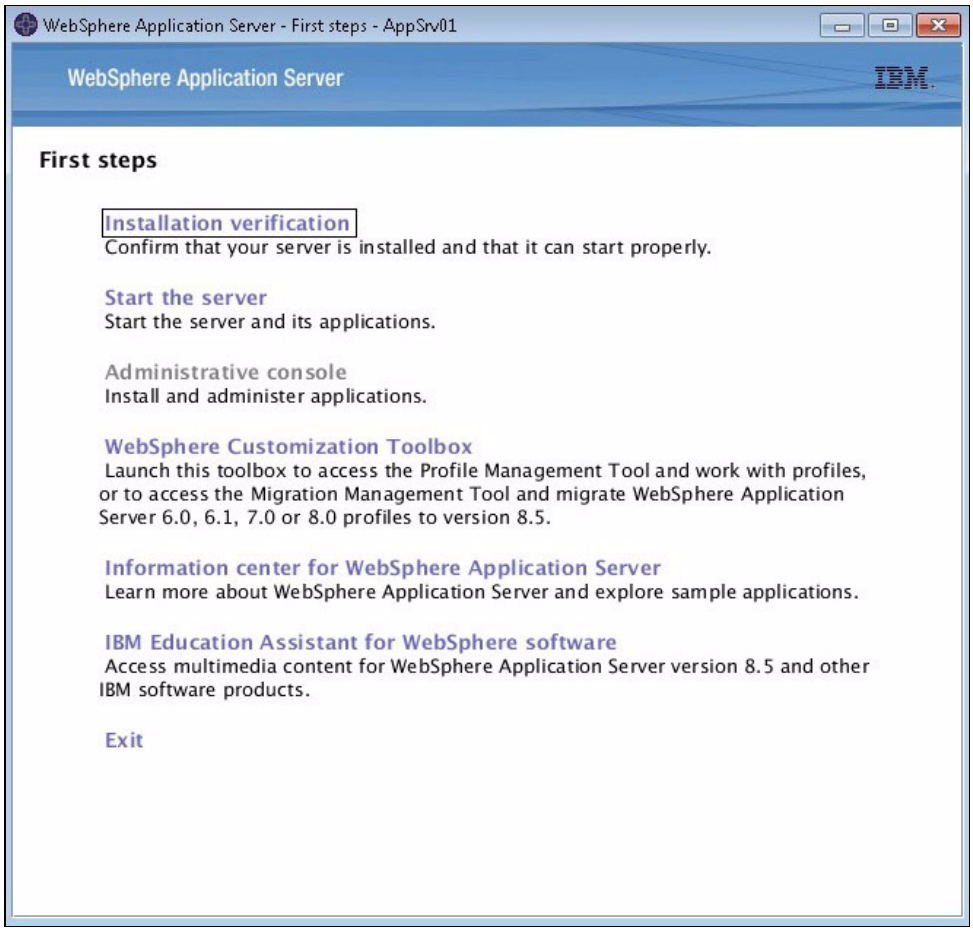


Figure 4-65 WebSphere First steps

9. Confirm that the messages "`The installation Verification Tool verification succeeded`" and "`The Installation verification is complete`" display at the bottom of the First steps output - Installation verification window that is shown in Figure 4-66.

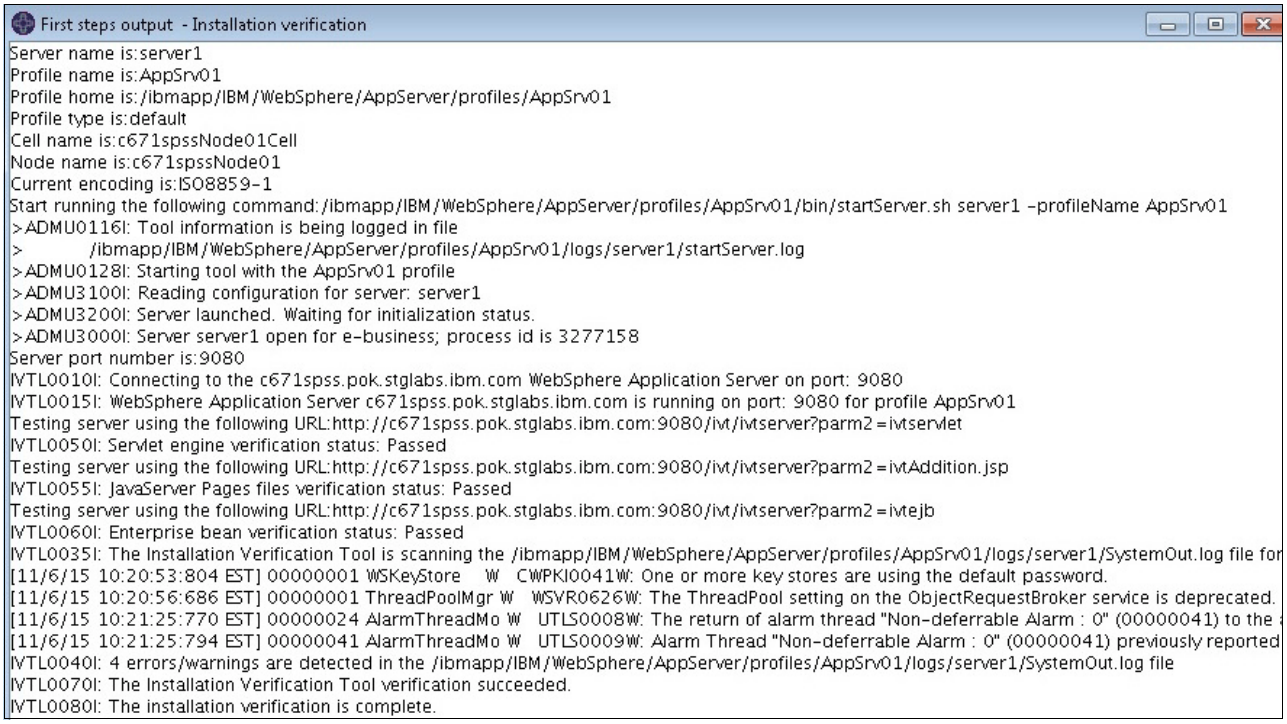


Figure 4-66 WebSphere First steps output

10. Close the window.
11. Click **Administrative console** from the First steps window (Figure 4-65 on page 104) or type `http://<servername>:9060/admin` in your browser.

12. Log in with the user name `admin` and the password `ibm1ibm`.

    For this demonstration, we set the user name to `admin` and the password to `ibm1ibm`, as shown in Figure 4-67.

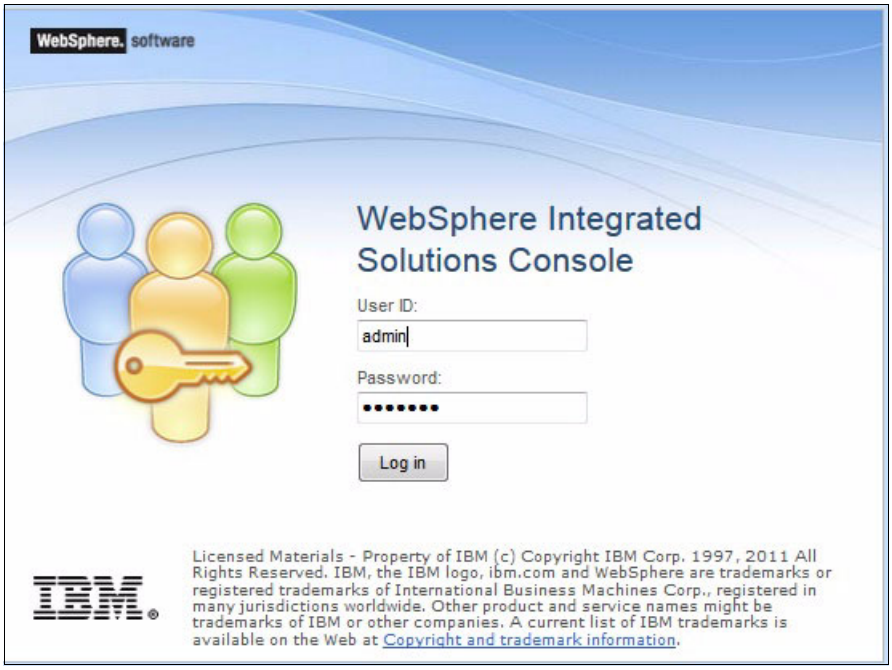


Figure 4-67 WebSphere Integrated Solutions Console

13. Check whether your profile was created successfully, as shown in Figure 4-68.

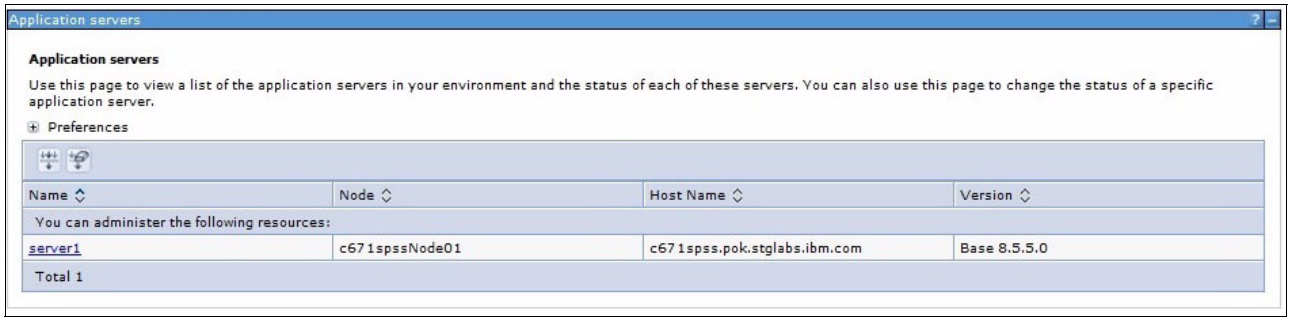


Figure 4-68 WebSphere Application servers

14. Log out of the WebSphere Integrated Solutions Console.

4.4.5 Install and configure the DB2 database

Create the database on DB2, and set up the DB2 client.

4.4.6 Install and configure SPSS Modeler Server

The installation and configuration of the SPSS Modeler Server are described.

Installation

The installation consists of the following steps:

1. Start X Window System (for example, Xming) on your client.
2. Connect to the AIX server.
3. Extract `spss_mod_17.0_cndsadp_7.0_aix_ml.zip`.
4. You will get `spss_mod_svr_17.0_aix_ml.bin`.
5. Set IATEMPDIR to a directory where plenty of disk space is available.
6. Execute `spss_mod_17.0_cndsadp_7.0_aix_ml.zip`.
7. Wait until the Modeler Server installation wizard opens.
8. Click **OK** for language selection. Click **Next**.
9. Click the license agreement. Select **I accept the terms in the license agreement**. Click **Next**.
10. Select **Production mode**. Click **Next**.
11. Set the path where you want to install Modeler Server. Click **Next**.
12. Wait until installation completes. Click **Done**.

Sample result

Example 4-4 shows the result of this sample installation.

Example 4-4 Result

```
unzip spss_mod_17.0_cndsadp_7.0_aix_ml.zip
export IATEMPDIR=/ibmapp/
./spss_mod_17.0_cndsadp_7.0_aix_ml.zip
```

How to start Modeler Server

To start Modeler Server, execute **`modelersrv.sh`** with the **`start`** parameter:

```
./modelersrv.sh start
```

Sample result

Example 4-5 shows the result.

Example 4-5 Result

```
/usr/IBM/SPSS/ModelerServer/17.0/modelersrv.sh start
IBM SPSS Text Analytics Server is already running
IBM SPSS Modeler Server started
```

How to stop Modeler Server

To stop the Modeler Server, execute **modelersrv.sh** with the **stop** parameter:

```
./modelersrv.sh stop
```

Sample result

Example 4-6 shows the result.

Example 4-6 Result

```
./modelersrv.sh stopIBM SPSS Text Analytics Server stopped
IBM SPSS Modeler Server stopped
```

Sample result

Check whether Modeler Server is up and running with the command that is shown in Example 4-7.

Example 4-7 Check whether Modeler Server is running

```
/usr/IBM/SPSS/ModelerServer/17.0/modelersrv.sh list
PID     PPID         USER      VSZ   PCPU COMMAND
3014978    1     root  19196  0.0 /usr/IBM/SPSS/ModelerServer/17.0/mo
delersrv_17_0 -server

If the Modeler Server is not launched the result is
PID     PPID         USER      VSZ   PCPU COMMAND
```

4.4.7 Install and configure IBM SPSS Collaboration and Deployment Service

This section describes the installation and configuration steps for SPSS Collaboration and Deployment Service.

Install

Follow these steps to deploy the SPSS Collaboration and Deployment Service:

1. Click **Install**.
2. Select the product that you want to install from the Installation Packages window (Figure 4-69 on page 109). Select **IBM SPSS Collaboration and Deployment Service - Repository Server**. After you select the product, the status changes to `Will be installed`. Click **Next**.
3. Select **I accept the terms in the license agreement**. Click **Next**.
4. Click **Next**.
5. Select the path to install IBM SPSS Collaboration and Deployment Service.

Important: If files are in the target directory, Installation Manager will not proceed with the installation. So, you need to delete the existing directory.

6. Click **Next**.
7. Scroll down to the bottom of the window. Check whether the required amount of space is available for the installation. Click **Next**.
8. Click **Install**.

Important: If files exist in the target directory, Installation Manager will not proceed with the installation so you need to delete the existing directory.

For the installation, both the `usr` and `opt` directories are used. You must check whether both directories have enough space for the installation. If you do not have enough space for the `opt` directory, you might see the following message:

“`cannot open http:/localhost:9080/DM But you can open http:/localhost:9080/config http:/localhost:9080/peb etc`**”**

The possible cause of this issue is a lack of disk space. The deployment file was not generated correctly or the deployment file was not deployed to WebSphere correctly.

9. Wait until the installation completes.
10. Check **View Log Files** to see whether any errors are logged.
11. See Figure 4-69. Click **IBM SPSS Collaboration and Deployment Services Configuration Tool** for “Which program do you want to start?” and click **Finish**.

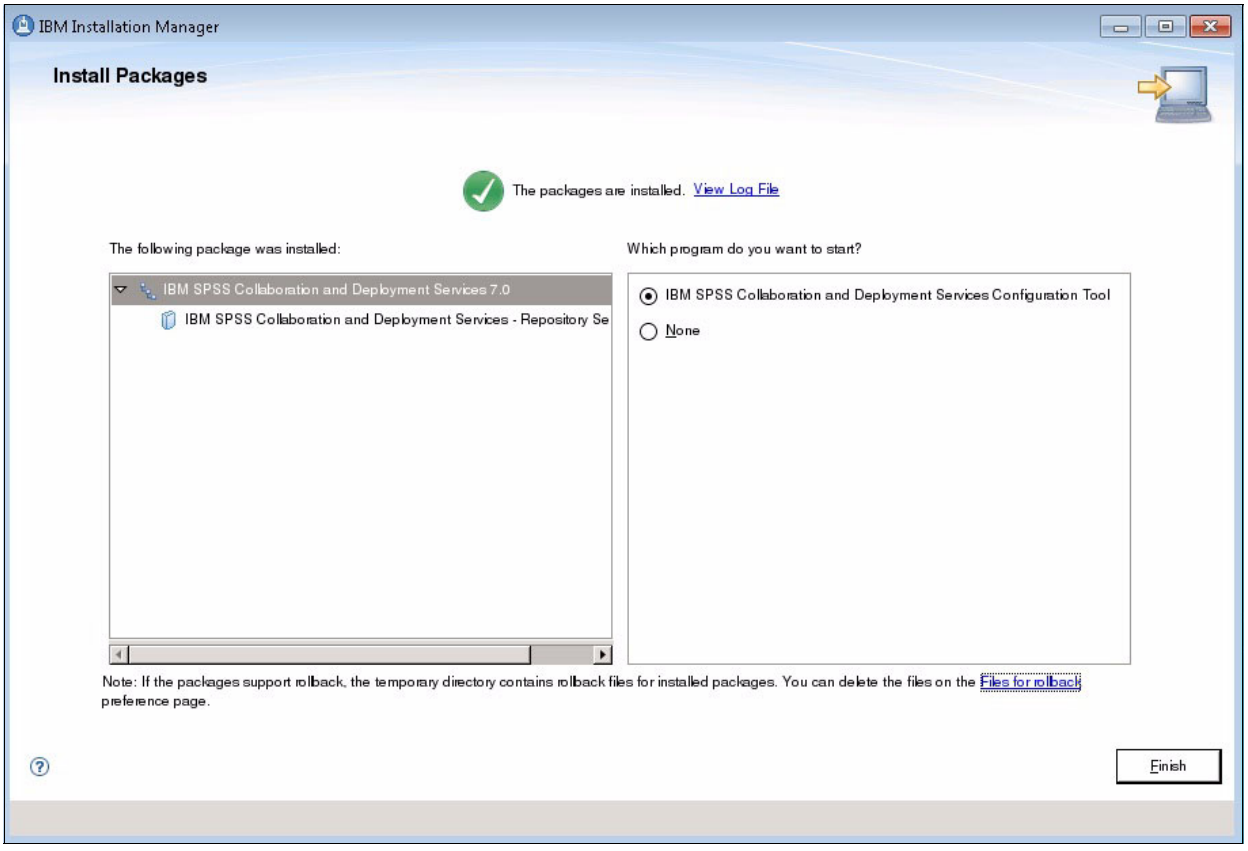


Figure 4-69 Installation Manager Install Packages window

Configure

This section describes how to configure the software.

Note: After you click **Finish** on the Install Packages window, the installation wizard starts.

Follow these steps:

1. In our demonstration environment, we used the following settings:
   a. Application Server type: `IBM WebSphere`
   b. Path or folder name for the WebSphere profile directory (Figure 4-70): `/usr/IBM/WebSphere/AppServer/profiles/AppSrv01/`

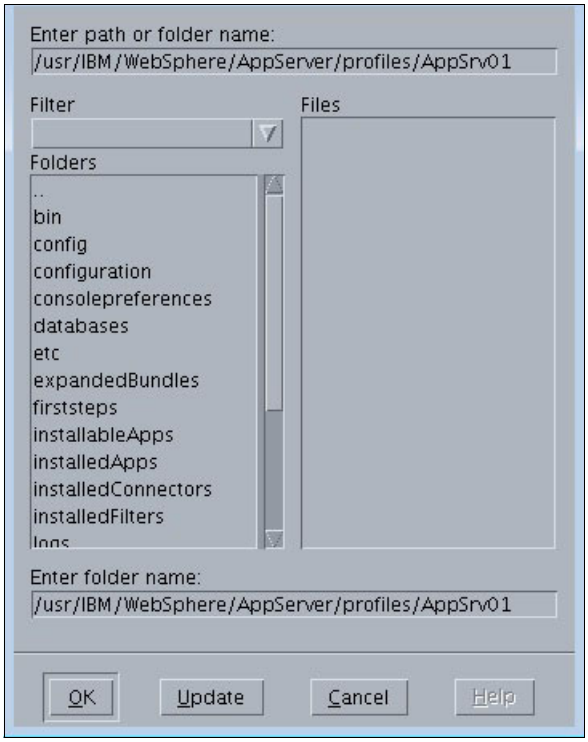


Figure 4-70 Selecting the WebSphere profile

   c. User name: `admin`
   d. Password: `ibm1ibm`

2. Figure 4-71 shows the SPSS Collaboration and Deployment Services Configuration Tool window.

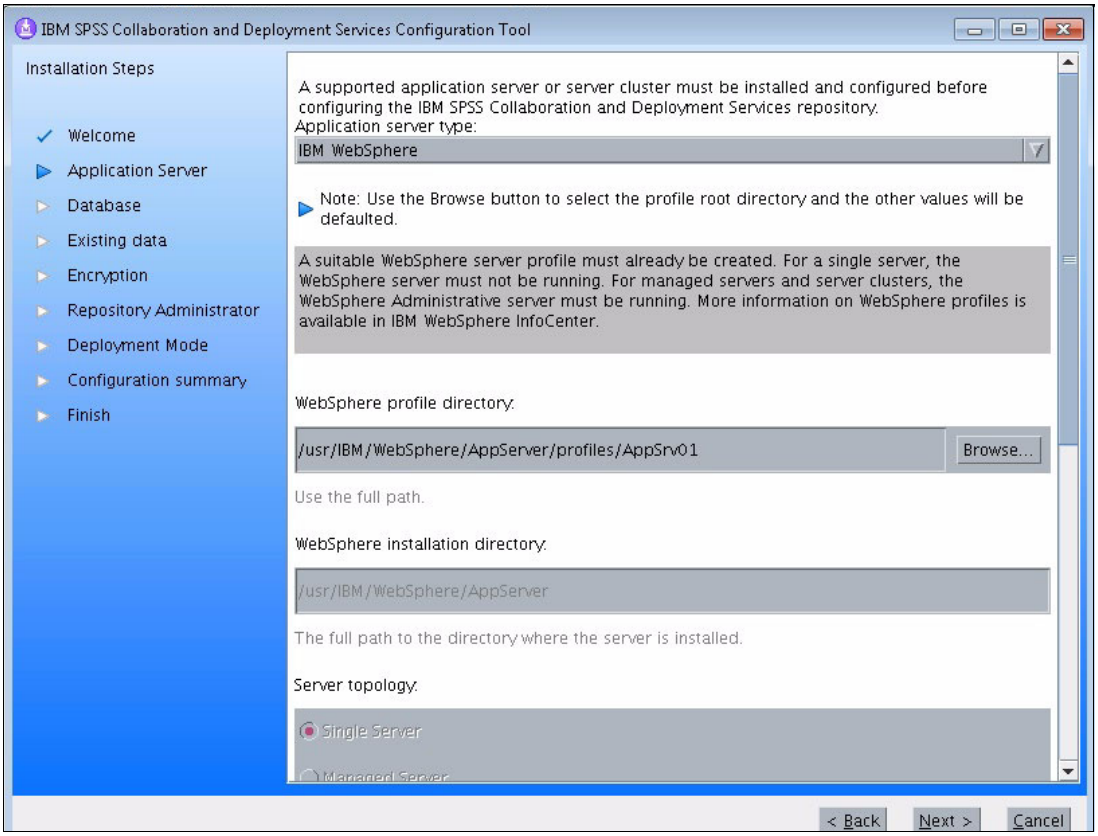


Figure 4-71 SPSS Collaboration and Deployment Services Configuration Tool window

3. Figure 4-72 shows the SPSS Collaboration and Deployment Services Configuration Tool for the application server options. Click **Next**.

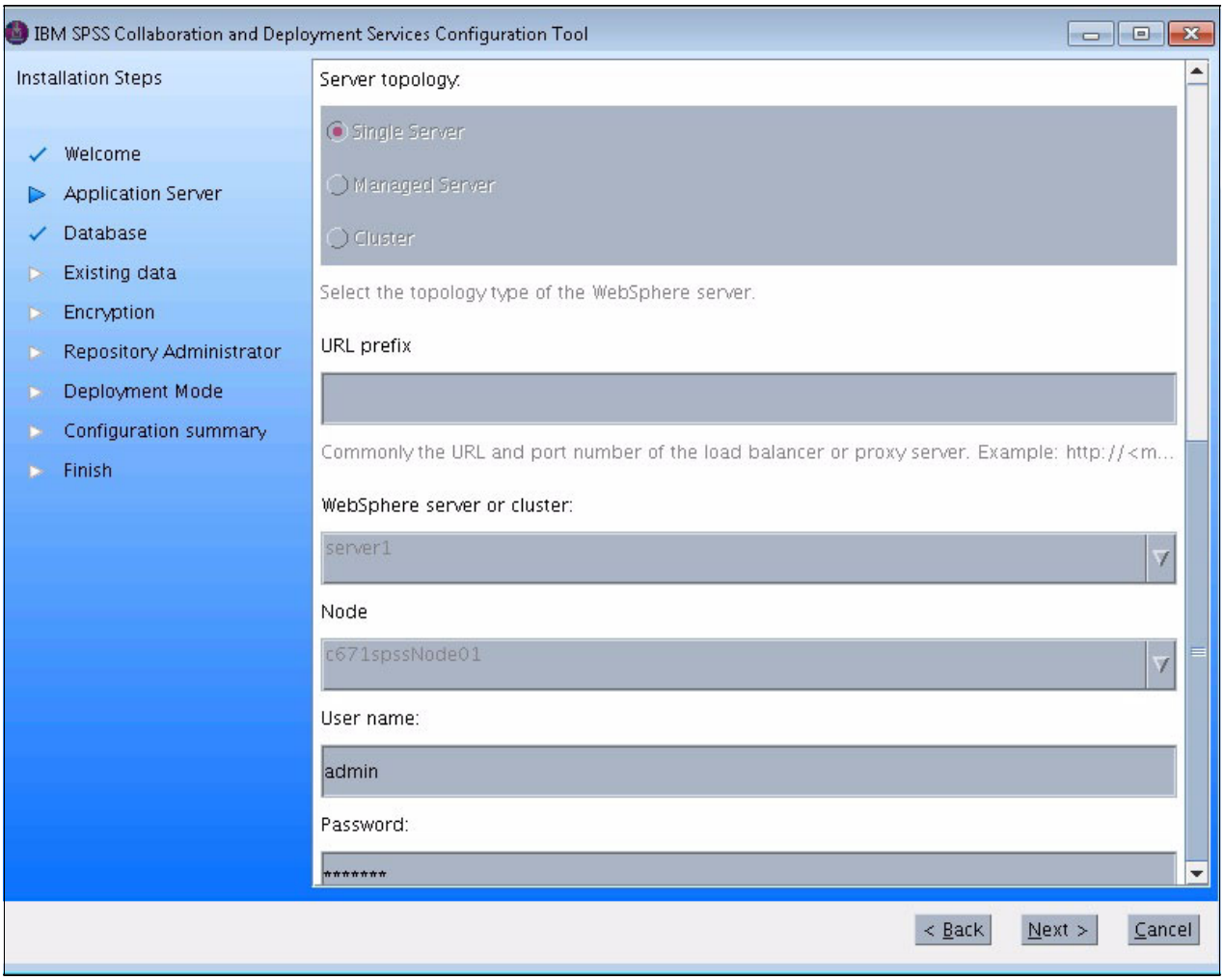


Figure 4-72 IBM SPSS Collaboration and Deployment Configuration Tool: Application Server options

4. Check the database name that is created for Collaboration and Deployment Service and click **Next**. In this demonstration, we set the following information (Figure 4-73):
   - Database type: `IBM DB2`
   - Host name: `dn05`
   - Port: `50000`
   - Database name: `SPSSCDS`
   - User name: `db2inst1`
   - Password: `ibm1ibm`

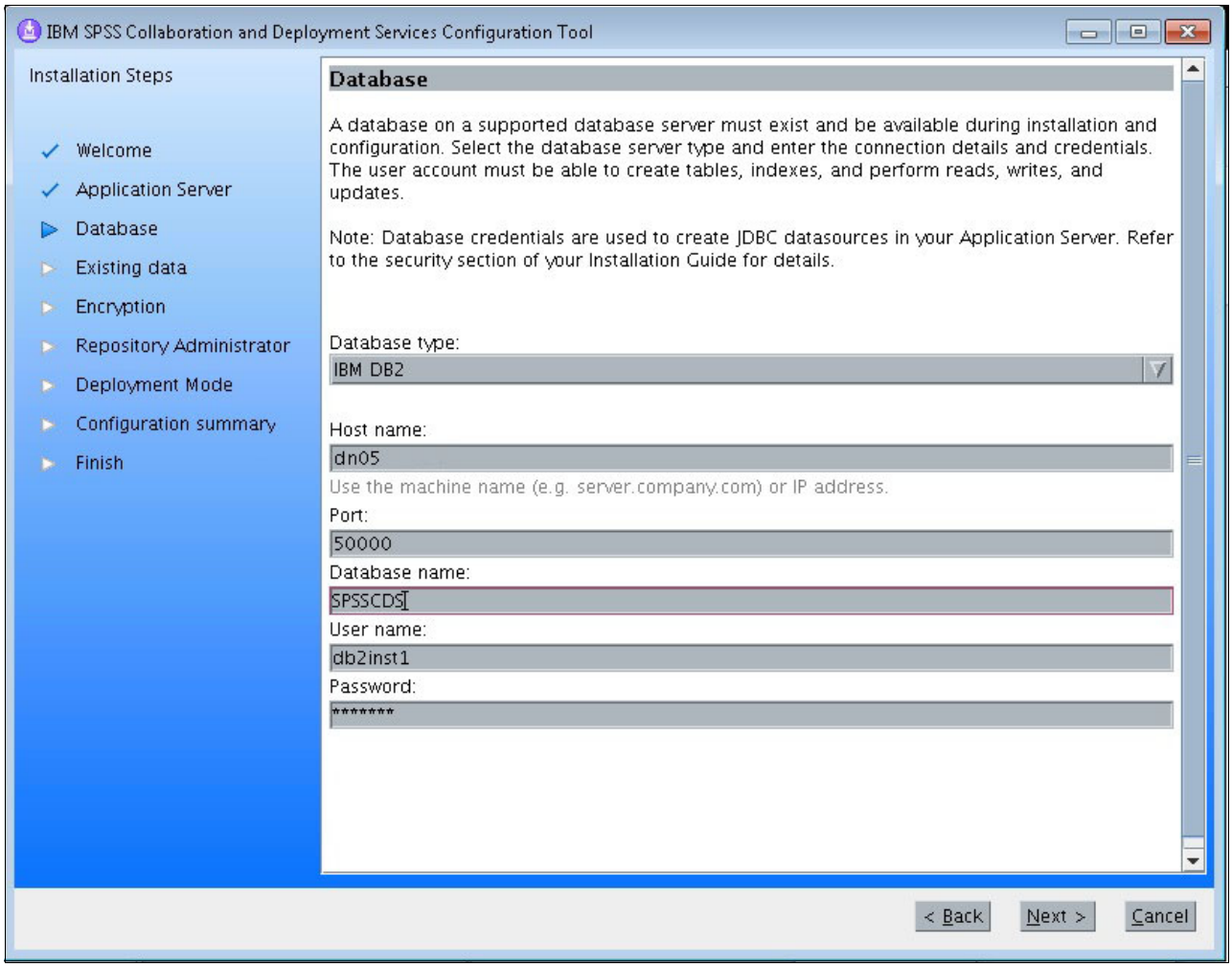


Figure 4-73 Configuration tool: Database server information

5. Select **Erase any existing data,** as shown in Figure 4-74. Click **Next**.

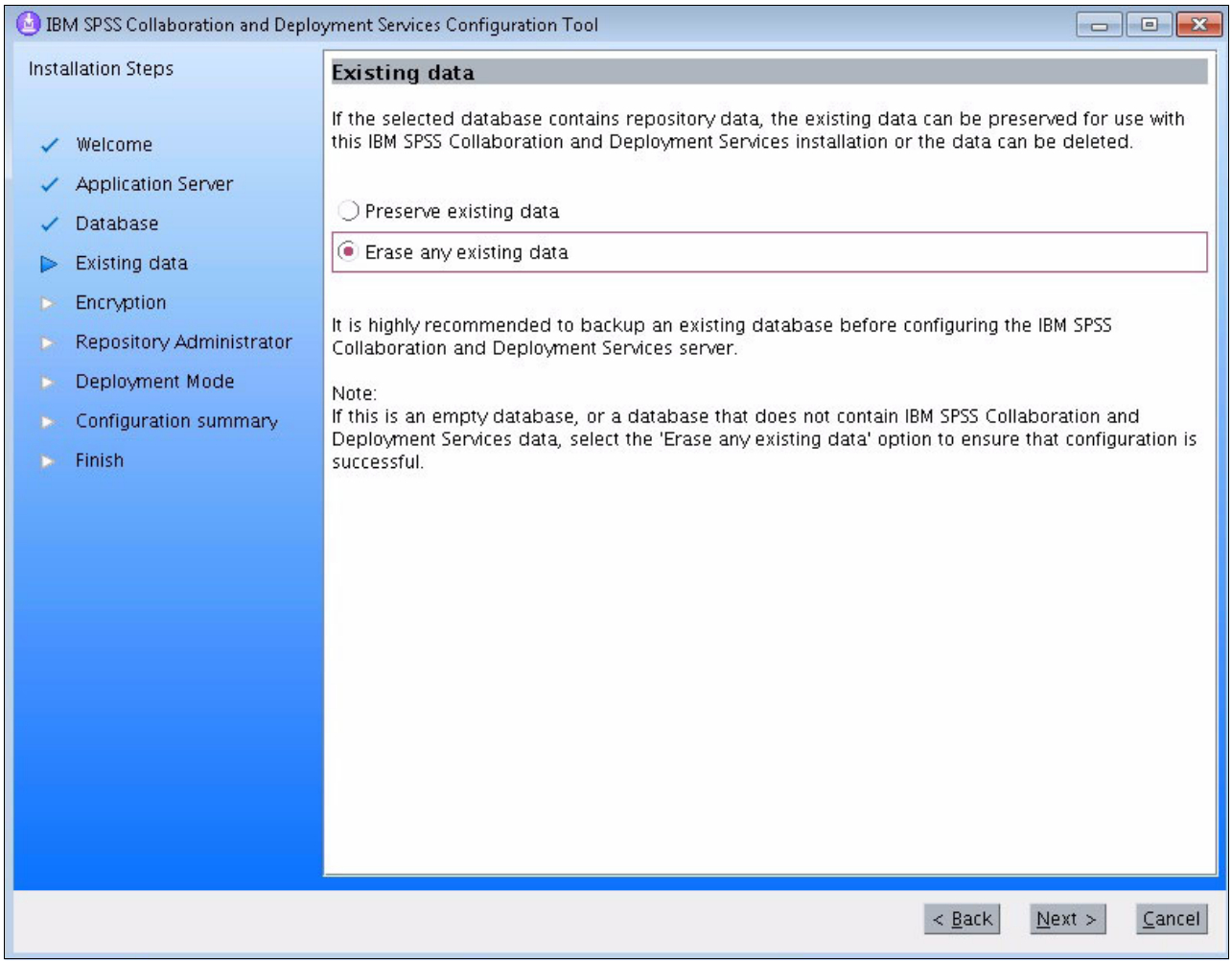


Figure 4-74 Configuration tool: Erase existing data

6. Type the password for encryption. For this case, we entered `ibmlibm`. See Figure 4-75. Click **Next**.

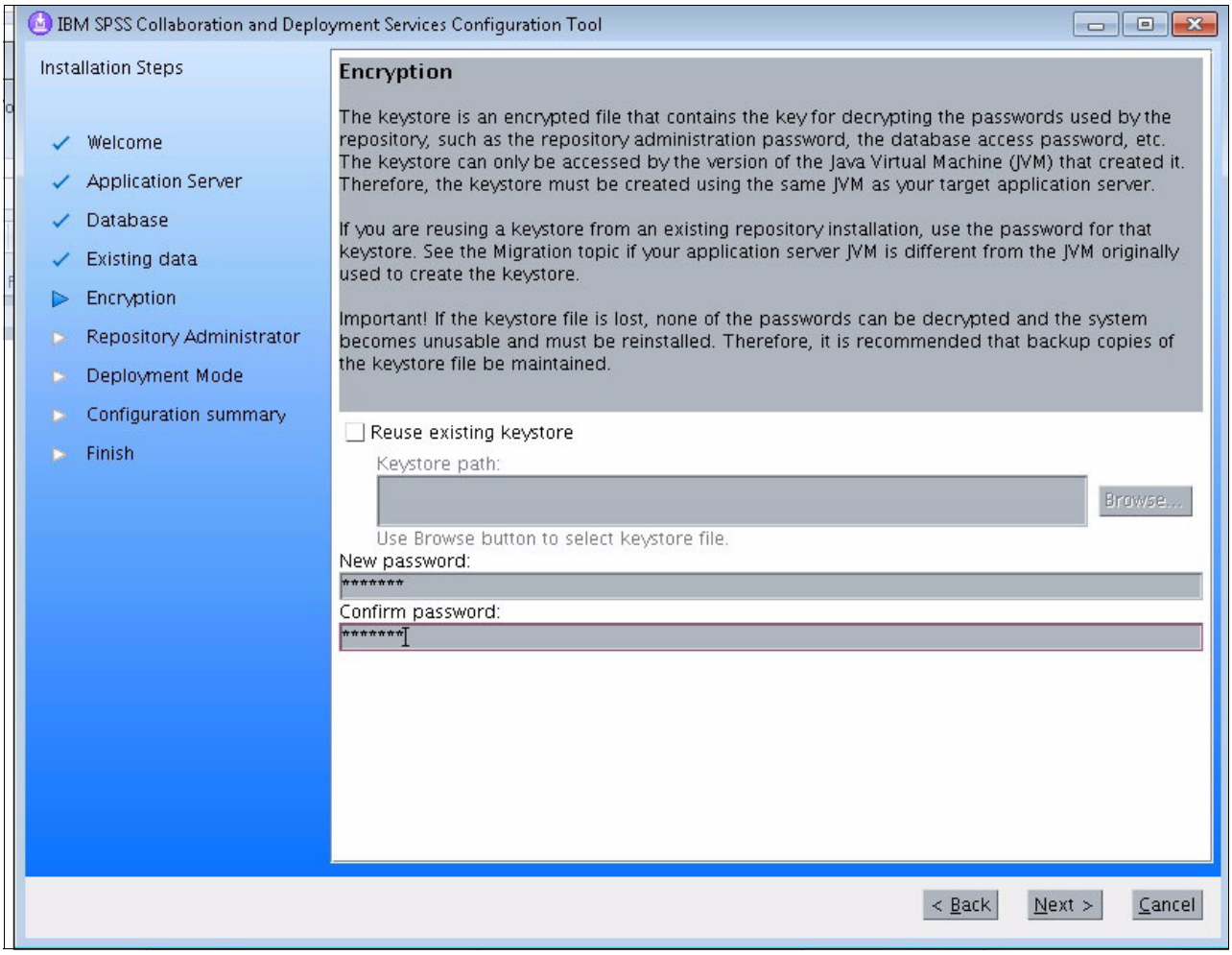


Figure 4-75 Configuration tool: Entering the encryption password

7. Set the ID and password for the repository administrator (Figure 4-76) and click **Next**. We entered the following information:
   - Repository administrator: `admin`
   - Password: `ibm1ibm`
   - Confirm password: `ibm1ibm`

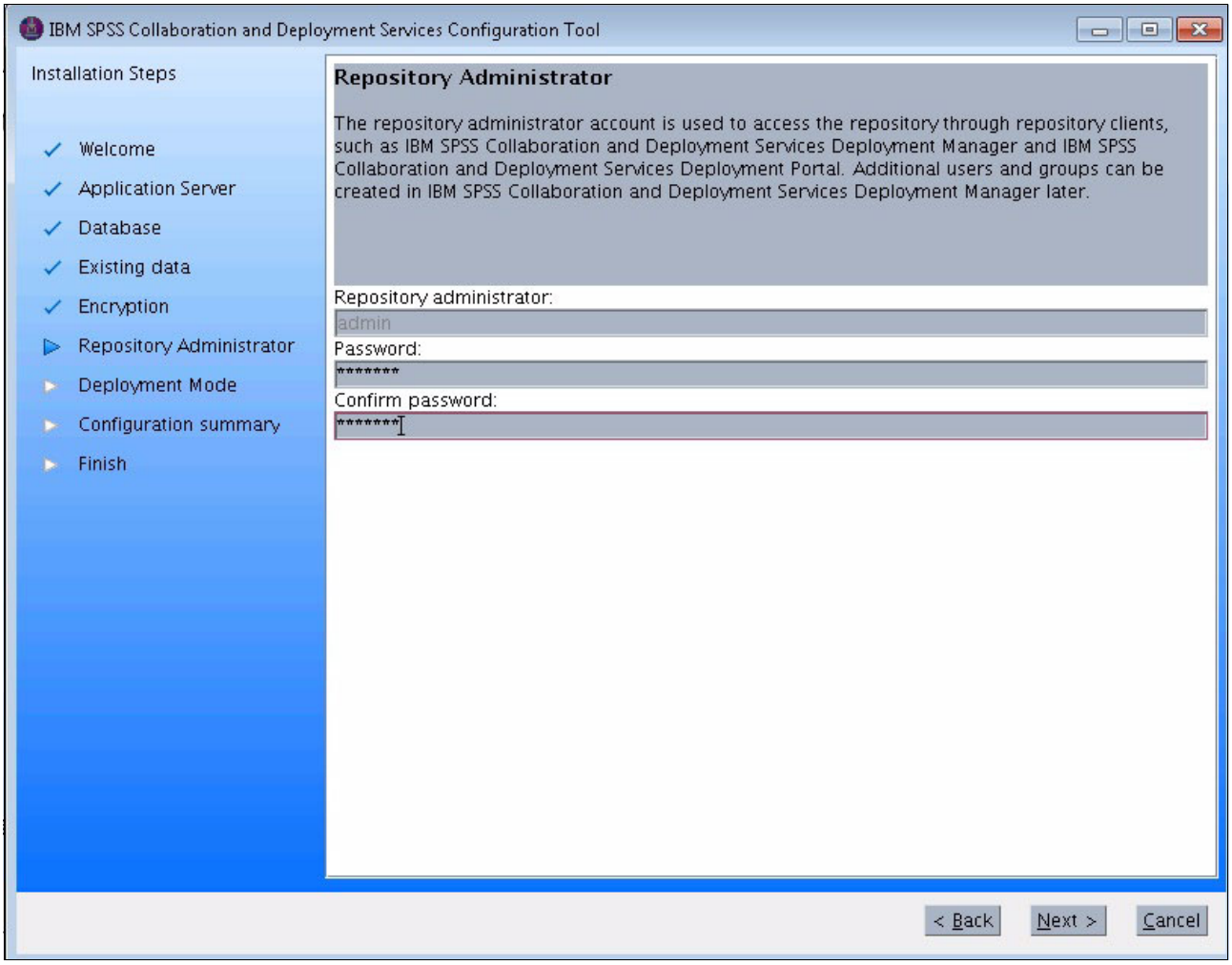


Figure 4-76 Configuration Tool: Setting up the repository administrator

8. Select **Automatic** for the deployment mode (Figure 4-77) and click **Next**.

Note: If you select Manual mode, you must deploy your changes to WebSphere manually.

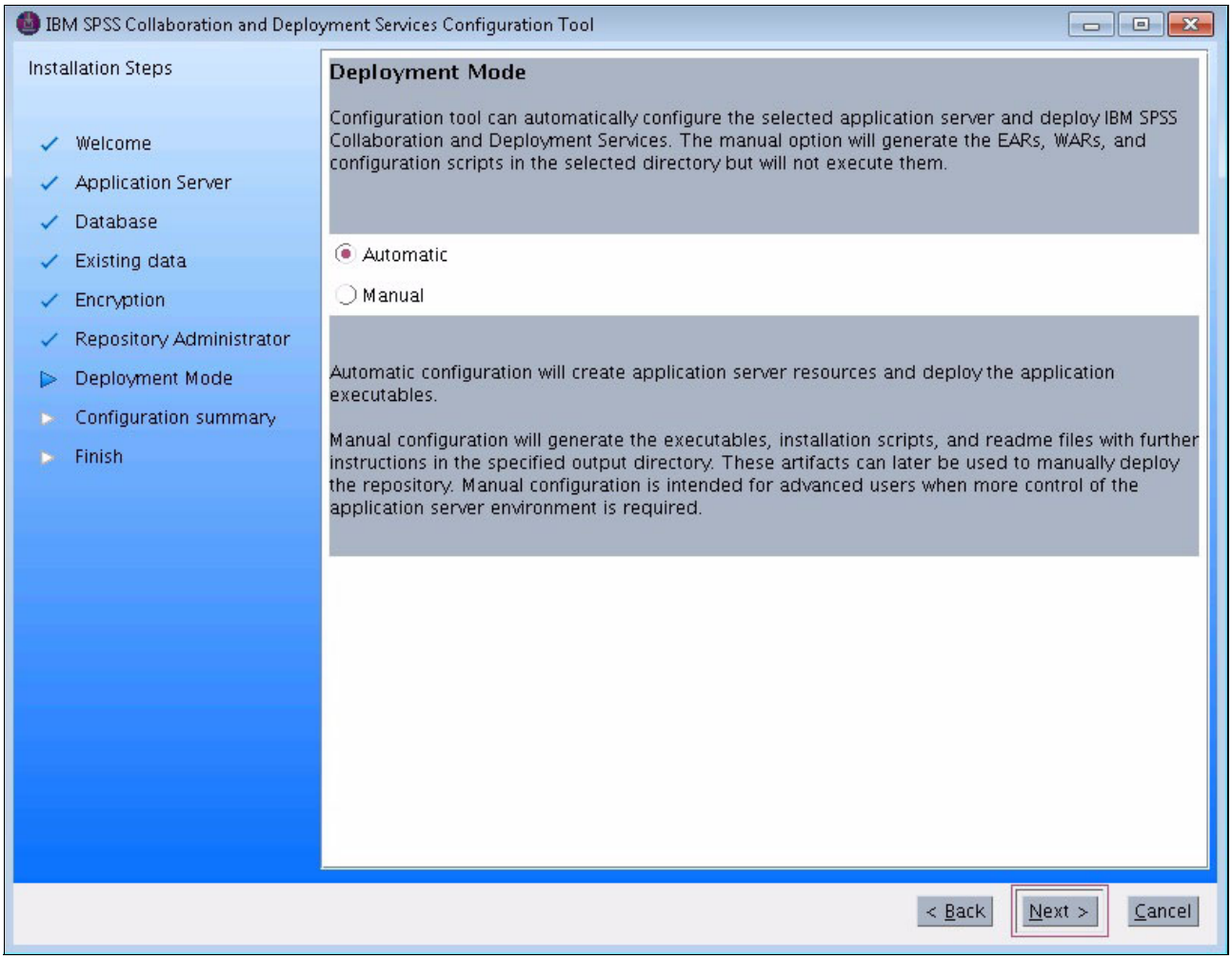


Figure 4-77 Configuration tool: Selecting the deployment mode

9. Click **Configure** on the Configuration summary (Figure 4-78).

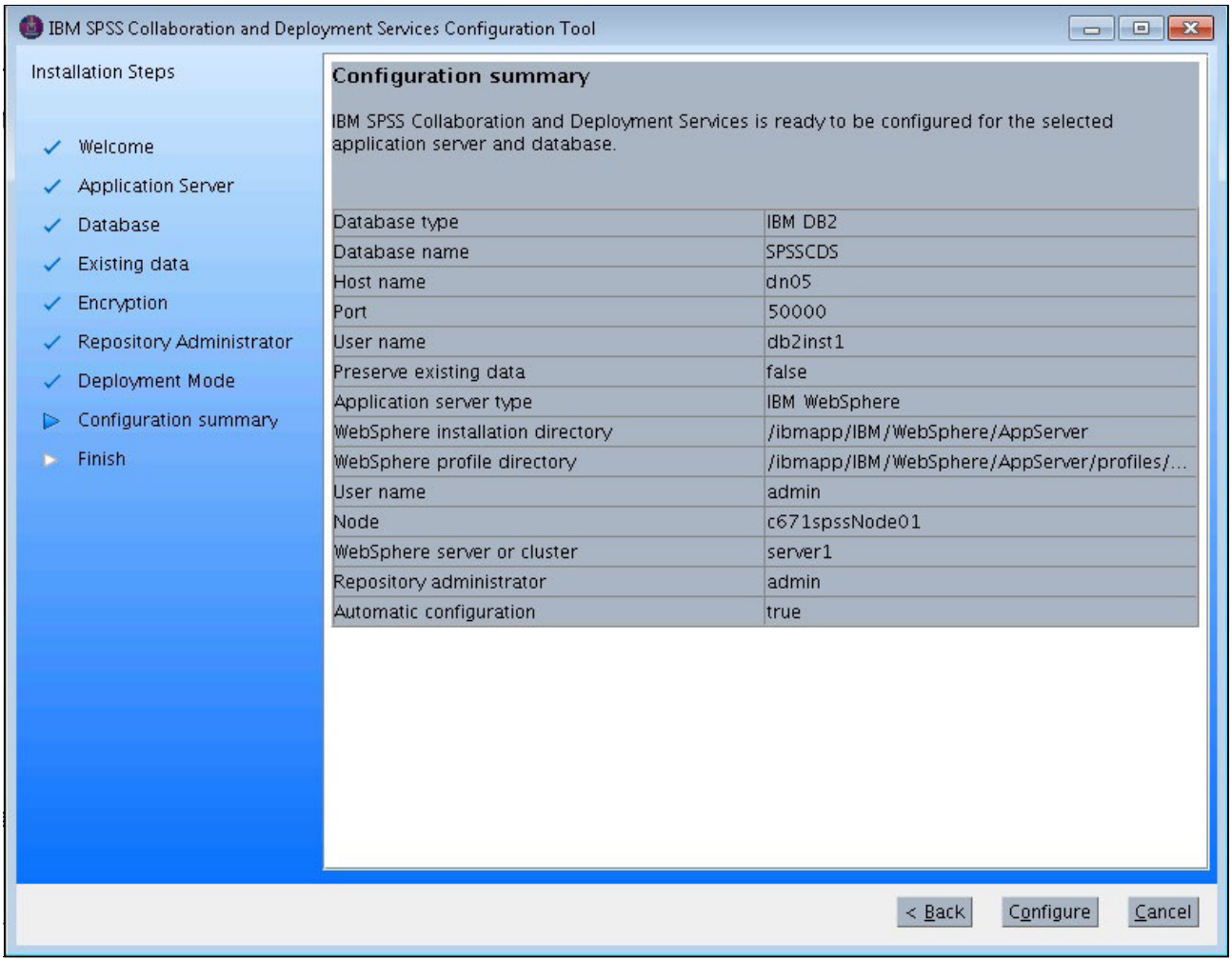


Figure 4-78 Configuration tool: Configuration summary

Note: To monitor the status, it might be a good idea to check the status of the database. Even if the indicator looks frozen, Collaboration and Deployment Service might be trying to communicate with the repository database.

10. Click **Finish**.
11. After the installation completes, check the log file by using the **View Log files** option. The actual log files are stored by default in the `/usr/ibm/installdata_for_im/logs` directory.

4.4.8 Install and configure IBM SPSS Modeler Server Adapters for Collaboration and Deployment Services

This section describes the installation steps.

Install

Follow these steps:

1. Click **Install**.
2. Select the product that you want to install from the list of Installation Packages. Select **IBM SPSS Modeler Adapters for Collaboration and Deployment Services**. After you select it, the status changes to `Will be installed`. Click **Next**.
3. Select **I accept the terms in the license agreement** and click **Next**.
4. Click **Next**.
5. Select the path to install IBM SPSS Modeler Adapters for Collaboration and Deployment Services and click **Next**.
6. Scroll down to the bottom of the window. Check whether the required amount of space is available for the installation. Click **Next**.
7. In the Common Configurations section on the Features tab of the Install Packages page, type the user and password that you set for the Collaboration and Deployment Service - Repository Server. For example, we used `admin` for the user and `ibm1ibm` for the password, as shown in Figure 4-79. Click **Next**.

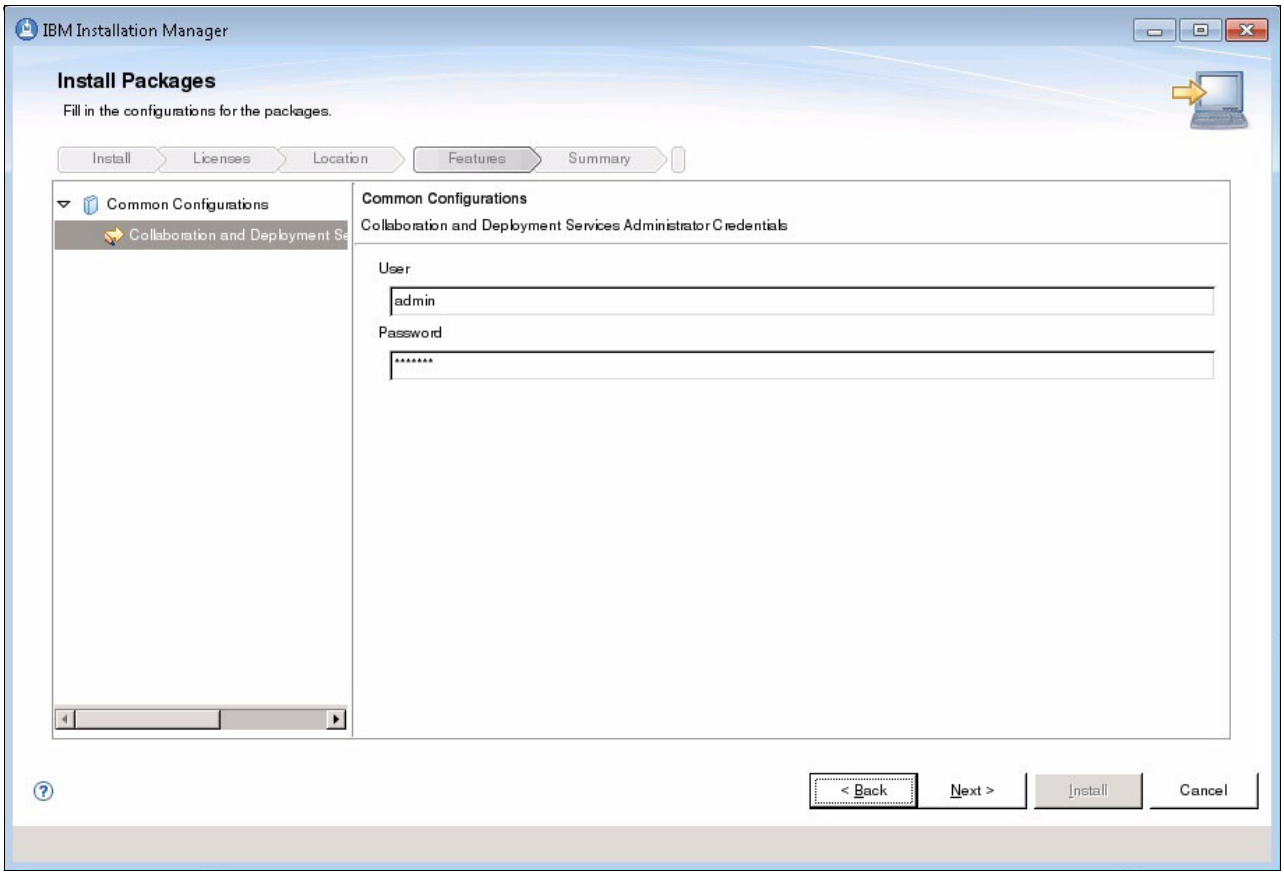


Figure 4-79 Install Packages window: Entering the user ID and password

8. Check whether enough disk space exists for the installation. If enough space is available for the installation, click **Install**.
9. Click **OK**.

10. After the installation completes, check the log file by clicking **View Log files**. The actual log files are stored by default in the `/usr/ibm/installdata_for_im/logs` directory.
11. Click **Finish**.

4.4.9 Install IBM SPSS Analytical Decision Management

This section provides the steps to install the IBM SPSS Analytics Decision Management component:

1. Click **Install**.
2. Select the product that you want to install from the Installation Packages window. Select **IBM SPSS Analytical Decision Management**. Click **Next**.
3. Scroll down to the bottom of the window. Select **I accept the terms in the license agreement**. Click **Next**.
4. Check whether enough space exists for the required space for the installation. Click **Next**.
5. Use the default options and click **Next** on the Features tab on the Install Packages window.
6. On the Install Packages page, set the user and password that you set for the Collaboration and Deployment Service - Repository Server. For example, we used `admin` for the user and `ibm1ibm` for the password.
7. Verify that enough disk space is available for the installation of IBM SPSS Analytical Decision Management. Click **Install**.
8. After the installation completes, check the log file by clicking **View Log files**.
9. The actual log files are stored by default in the `/usr/ibm/installdata_for_im/logs` directory.

4.4.10 Install SPSS Collaboration and Deployment Service

Follow these steps to install SPSS Collaboration and Deployment Service:

1. Extract `spss_cnds_depmgr_64b_7.0_win_mlzip`.
2. Navigate to the `\spss_cnds_depmgr_64b_7.0_win_ml\Deployment_Manager_64\install.exe` file location.
3. Right-click **install.exe**.
4. Select **Run as Administrator**.

5. If “`This file is an untrusted location. Are you sure you want to run it?`” appears, click **Yes** (Figure 4-80).

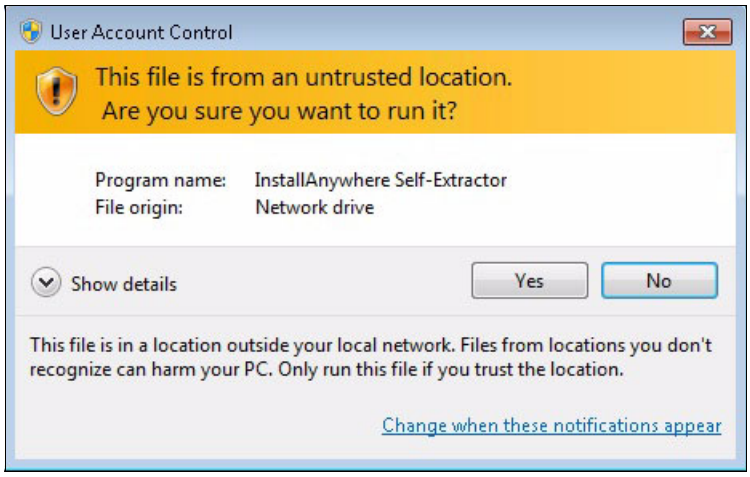


Figure 4-80 Windows User Account Control for the deployment manager installation

6. Log in with your ID and password. Wait until the process completes.
7. Select **English** and click **OK**.
8. Click **Next**.
9. Select **I accept the terms in the license agreement** and click **Next**.
10. Change the installation path if necessary. Click **Next**.
11. Accept the defaults. Click **next**.
12. Click **Install**.
13. Click **Done**.

4.5 Cognos for Dashboarding

In our demonstration, we install Java, DB2 client, OpenLDAP, and Apache HTTP server to support the Cognos Business Intelligence version 10 installation. At the time of writing this publication, Cognos Business Intelligence is supported for Linux on POWER8, but check that your environment is big endian.

4.5.1 Install IBM Java SDK 6.0

Follow these steps:

1. Download IBM Java SDK 6.0. Install it by running **ibm-java-sdk-6.0-16.7-linux-ppc64.bin**, which is an X application. Check that your environment is set correctly to X11 forwarding in the server and client side as shown in Figure 4-81. Read the license agreement carefully, accept the terms, and click **Next**.

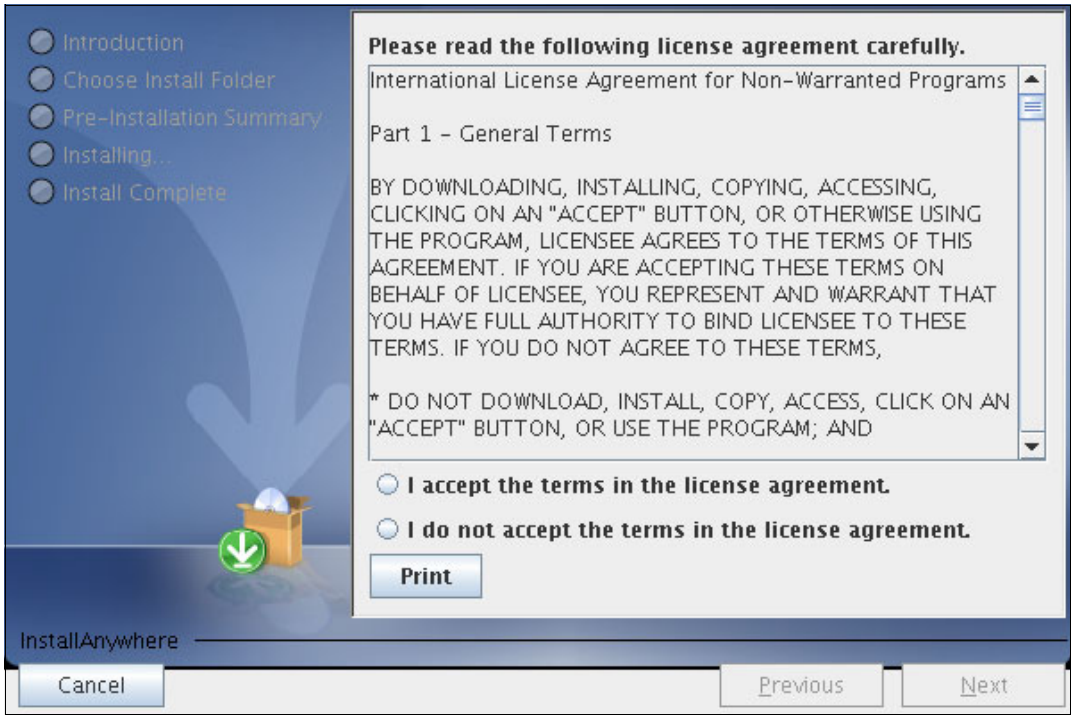


Figure 4-81 IBM Java SDK 6.0 installation window

2. On the Introduction window, we strongly suggest that you close all programs before you continue with this installation. Click **Next**.
3. In the Choose a destination folder section, we suggest that you use `/usr/java/ibm-java-ppc64-60` as a Java binaries directory. You can change it according to your environment's or organization's standards. Click **Next**.
4. On the Pre-Installation Summary window, review the information and click **Install**.
5. The installation might take a few seconds. Click **Done** to conclude the IBM Java SDK 6.0 installation.

4.5.2 Install DB2 client

Before you install DB2 client, install pam.ppc for 32 bits and install vacpp, as shown in Figure 4-82.

```
yum install pam.ppc

yum install vacpp
```

Figure 4-82 yum install for pam.pcc and vacpp

Download and extract the IBM DB2 client installations. In our scenario, we used the full server installation from the `DB2_V10.5_ltd_CD_Linux_ipSeries.tar` file. Run **`./server/db2setup`** from the extracted directory and wait a few seconds to open the X Window System for installation.

On the next window, click **Install a product** and scroll down to click **Install New** for IBM Data Server Client Version 10.5, as shown in Figure 4-83.

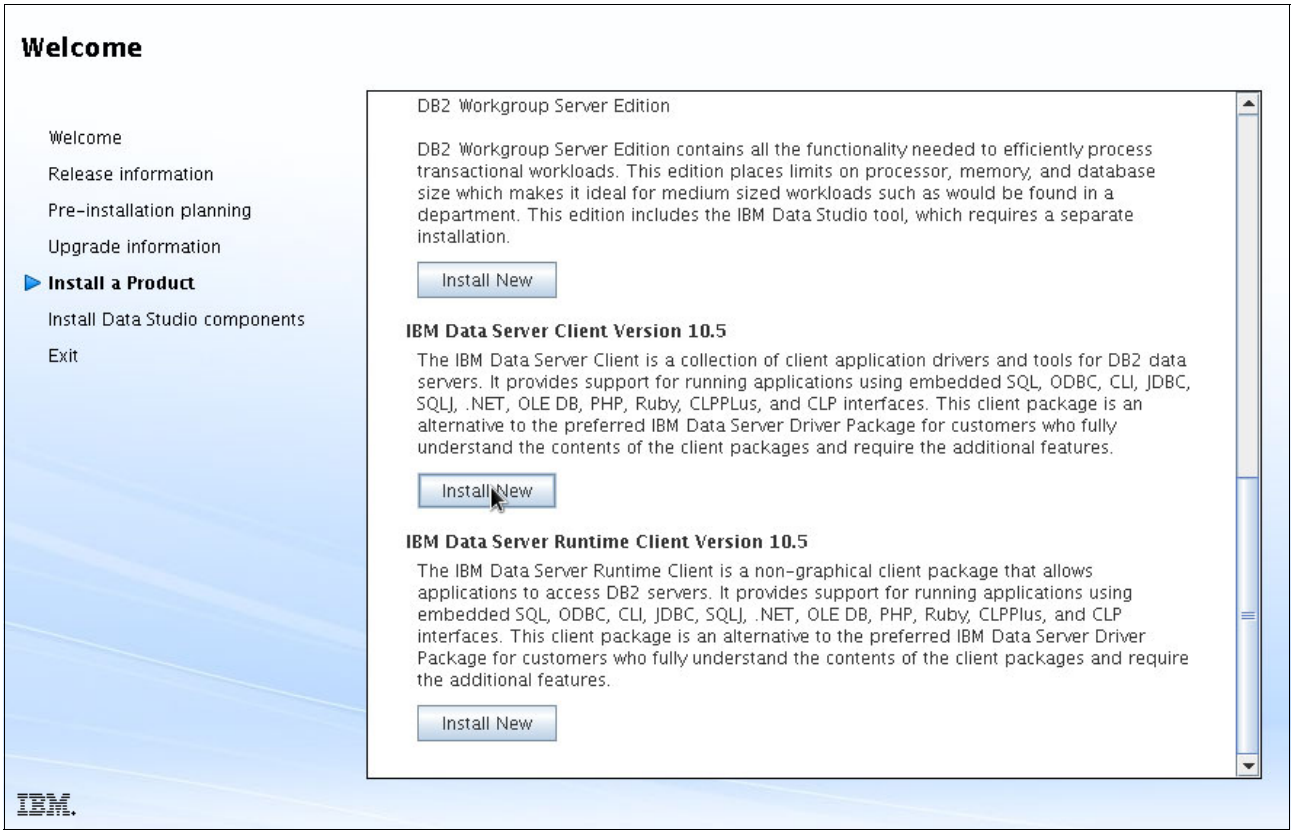


Figure 4-83 Install IBM Data Server Client Version 10.5

See Figure 4-84. Follow these steps. On the left side of the window, select these options:

1. Click **1. Introduction**. On the menu option 1. Introduction, click **Next**.
2. Click **2. Software License Agreement**. Read the software license agreement carefully and click **Next**.
3. Click **3. Installation type**. We strongly encourage you to select the typical installation, but you can customize the installation according to your environment's or organization's standards.
4. Click **4. Installation action**. Keep the default settings to create a response file for future reference, and click **Next**.
5. Click **5. Installation directory**. We suggest that you to keep the default installation directory, which is `/opt/ibm/db2/V10.5`, but you can modify it.
6. Click **6. Instance setup**. Create a DB2 instance. An instance is required to use the product. Click **Next**.
7. Click **7. Instance-owning user**. We encourage you to use the default settings, but you can change them, as shown in Figure 4-84. Click **Next**.

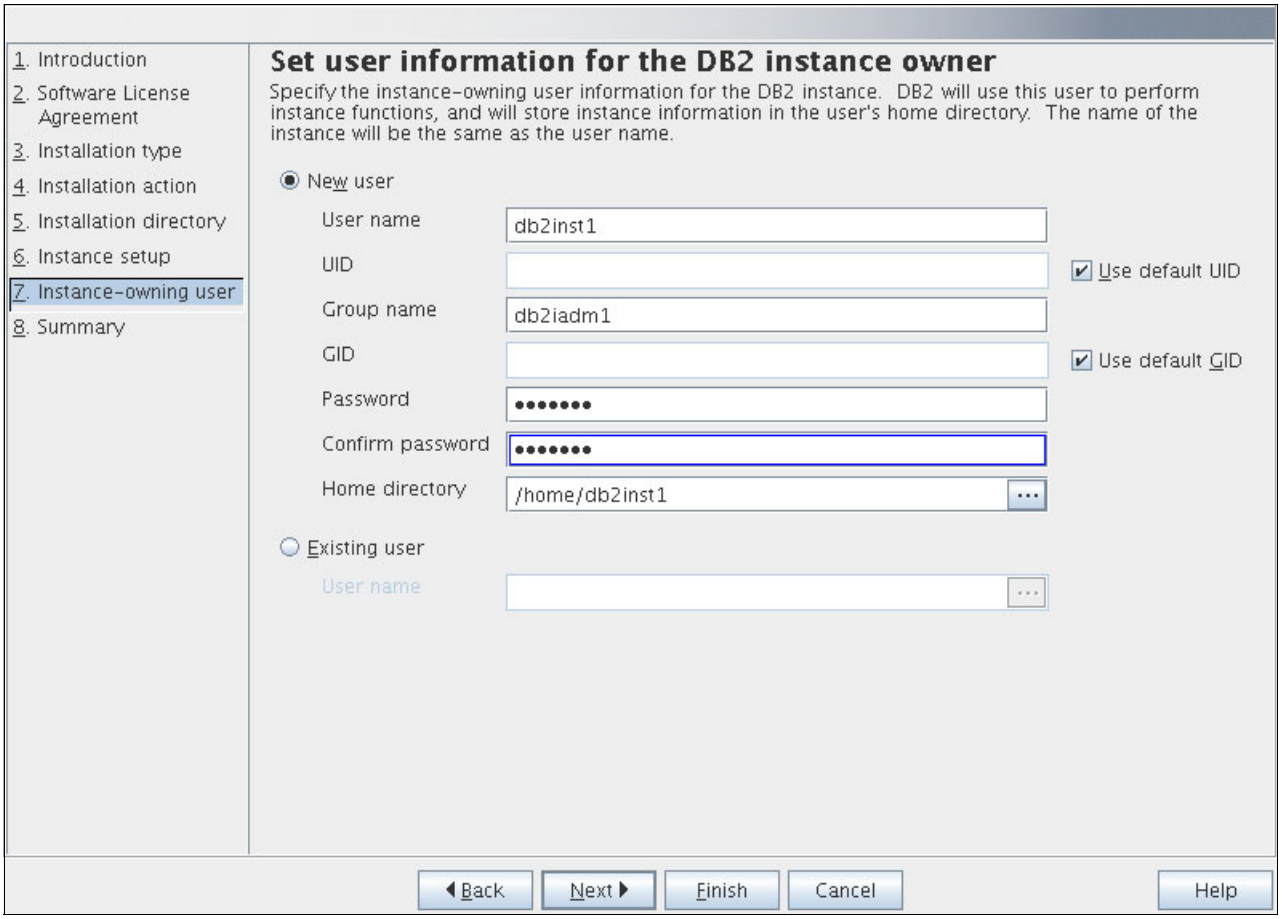


Figure 4-84 Set user information for the DB2 instance owner for DB2 client

8. Review your installation settings by clicking **8. Summary**. Click **Finish** to start the installation, which takes a few minutes.

4.5.3 Install Apache HTTP server 2

Follow these steps:

1. Apache HTTP server version 2 is supported by Cognos Business Intelligence Version 10. Download and extract the `httpd-2.2.31.tar` file. Go to your installation directory and type the **`tar xvf httpd-2.2.31.tar`** command to extract the tar file, as shown in Figure 4-85.

```
tar xvf httpd-2.2.31.tar
```

Figure 4-85 Untar Apache HTTP server installation

2. In the extracted `httpd-2.2.31`, run the **`configure`** command. We suggest that you use `/usr/local/apache2` for the installed Apache HTTP server directory, as shown in Figure 4-86.

```
./configure --prefix=/usr/local/apache2
```

Figure 4-86 Configure Apache HTTP server

3. Run the **`make`** and **`make install`** commands in the same directory where you performed the configuration to conclude the installation, as shown in Figure 4-87.

```
make

make install
```

Figure 4-87 Using the make and make install commands

4.5.4 Install and configure OpenLDAP

Follow these steps:

1. You can use the LDAP server that is deployed in your organization or install a new LDAP database, as demonstrated in Figure 4-88. Install the openldap-servers and the openldap-clients packages.

```
yum install openldap-servers

yum install openldap-clients
```

Figure 4-88 yum install openldap-servers and openldap-clients

2. Hash a new root password for your environment by using the **`slappasswd`** command. Save the hashed password string for later usage in your own configuration files.

3. Figure 4-89 shows the configuration for the `/etc/openldap/slapd.conf` that is used in our environment, which you can use for your own configuration. Ensure that you replace the rootpw hashed password string with your own **slappasswd** output that was saved in the previous step.

```
include         /etc/openldap/schema/corba.schema
include         /etc/openldap/schema/core.schema
include         /etc/openldap/schema/cosine.schema
include         /etc/openldap/schema/duaconf.schema
include         /etc/openldap/schema/dyngroup.schema
include         /etc/openldap/schema/inetorgperson.schema
include         /etc/openldap/schema/java.schema
include         /etc/openldap/schema/misc.schema
include         /etc/openldap/schema/nis.schema
include         /etc/openldap/schema/openldap.schema
include         /etc/openldap/schema/ppolicy.schema
include         /etc/openldap/schema/collective.schema

allow bind_v2

pidfile         /var/run/openldap/slapd.pid
argsfile        /var/run/openldap/slapd.args

########################################################################
# ldbm and/or bdb database definitions
########################################################################

database        bdb
suffix          "dc=cognos-test,dc=org"
checkpoint      1024 15
rootdn          "cn=admin,dc=cognos-test,dc=org"
rootpw          {SSHA}TdrqvQFyW5ONcvZJsfojdZFakOOEm/j+

directory       /var/lib/ldap
```

Figure 4-89 The slapd.conf example file

4. As you can see from the `slapd.conf` example (Figure 4-89), the defined database suffix is `"dc=cognos-test,dc=org"` and the rootdn is `"cn=admin,dc=cognos-test,dc=org"`.
5. Those elements must be created for the initial domain edit. Save a domain.ldif, as shown in Example 4-8.

Example 4-8 The domain.ldif example file

```
dn: dc=cognos-test, dc=org
dc: cognos-test
o: My Cognos Test
objectclass: top
objectclass: organization
objectclass: dcObject

dn: cn=admin,dc=cognos-test,dc=org
userPassword: ibm1ibm
sn: Administrator
cn: Administrator
```

```
objectClass: top
objectClass: person
objectClass: organizationalPerson
objectClass: inetOrgPerson
```

6. Run the **ldapadd** command, as shown in Figure 4-90, to create the initial domain.

```
ldapadd -x -W -D "cn=admin,dc=cognos-test,dc=org" -f domain.ldif
```

Figure 4-90 The ldapadd command to create the LDAP domain

7. You can also create additional users. For example, to add a user that is named adam, create an adam.ldif file, as shown in Example 4-9.

Example 4-9 New user adam.ldif file

```
dn: cn=adam,dc=cognos-test,dc=org
userPassword: ibm2ibm
sn: adam
cn: adam
objectClass: top
objectClass: person
objectClass: organizationalPerson
objectClass: inetOrgPerson
```

8. Run the **ldapadd -x -W -D "cn=admin,dc=cognos-test,dc=org" -f adam.ldif** command that is shown in Figure 4-91 to add a user, for example, adam, to the LDAP directory.

```
ldapadd -x -W -D "cn=admin,dc=cognos-test,dc=org" -f adam.ldif
```

Figure 4-91 Adding a user that is named adam to the LDAP directory

4.5.5 Configure Cognos Business Intelligence

Follow these steps:

1. Before you install Cognos Business Intelligence, check that motif and libgcc are both installed in 32 bits and 64 bits, as shown in Figure 4-92.

```
yum install motif.ppc
yum install motif.ppc64
yum install libgcc.ppc
yum install libgcc.ppc64
```

Figure 4-92 Install motif and libgcc for both 32 bits and 64 bits

2. Download and extract the `bi_svr_10.2.2_lxp_ml.tar` file. Run **`./linuxppc64h/issetup`** and the X Window System setup, as shown in Figure 4-93.

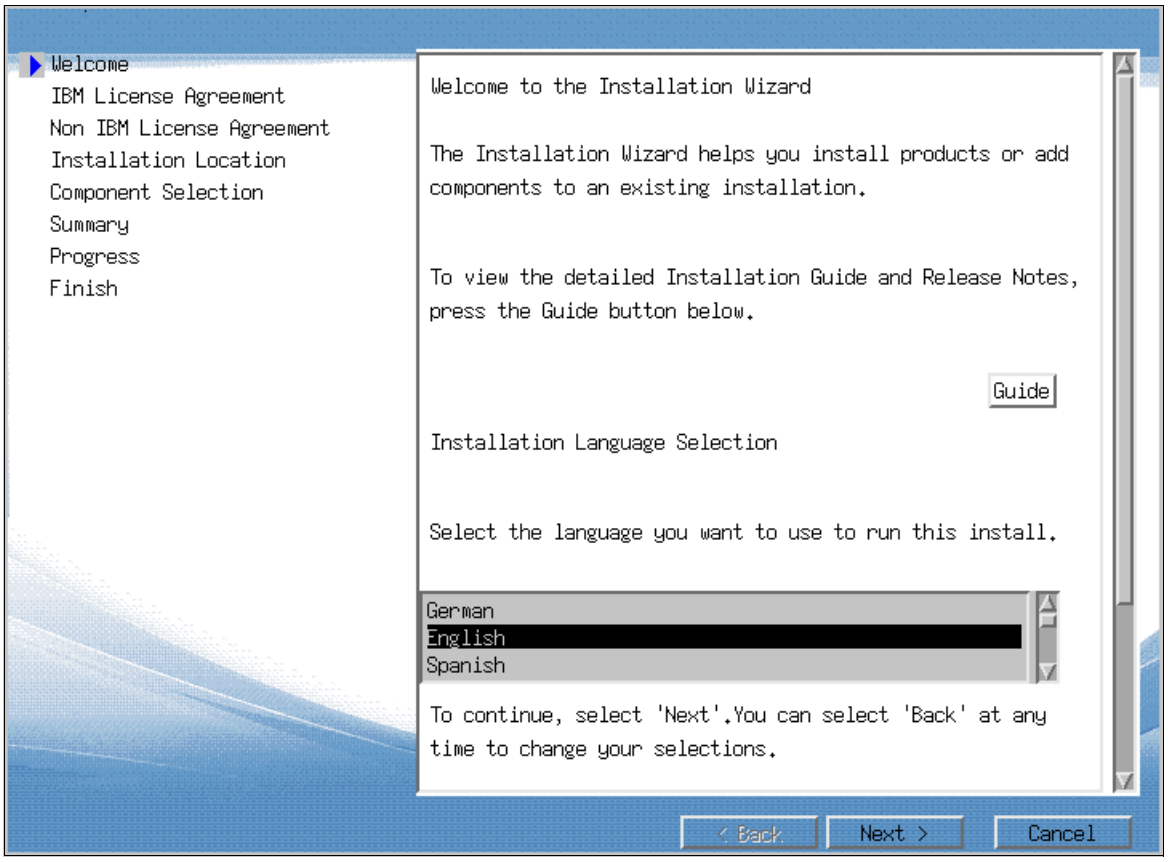


Figure 4-93 Cognos Business Intelligence Installation Wizard

3. Select the language that you want and click **Next**.
4. In the IBM license agreement, read the terms carefully, check **I agree**, and click **Next**.
5. On the Installation Location window, change to the installation directory that you want or keep the default. Click **Next**.
6. Select all of the available components to install under IBM Cognos Business Intelligence Server and click **Next**.
7. Review the installation summary and click **Next** to start the installation. The installation progresses for a few minutes. The Finish window opens.
8. Change your environment variables according to your environment. Example 4-10 shows our particular environment variables.

Example 4-10 Environment variables for Cognos

```
export JAVA_HOME=/usr/java/ibm-java-ppc64-60
export DB2PATH=/opt/ibm/db2/V10.5
export
LD_LIBRARY_PATH=$DB2PATH/lib32:$DB2PATH/opt/ibm/db2/V10.5/lib64:/opt/ibm/cognos/c1
0_2_2_64/cgi-bin:/opt/ibm/cognos/c10_2_2_64/cgi-bin/lib:/opt/ibm/cognos/c10_2_2_64
/bin64
```

9. We suggest that you save your own environment settings to your Cognos owner profile.

Note: Check whether your firewall setting prevents other nodes from reaching your Cognos nodes. If you want to temporarily disable your firewall for the installation, run the following commands:

```
systemctl stop firewalld.service
systemctl disable firewalld.service
```

4.5.6 Configure the Apache HTTP Server for Cognos

Follow these steps:

1. Edit your `/usr/local/apache2/conf/httpd.conf` file to add the entries that are shown in Figure 4-94. Replace *<cognos_installation_directory>* with the correct settings for your environment.

```
ScriptAlias /ibmcognos/cgi-bin "<cognos_installation_directory>/cgi-bin"
Alias /ibmcognos "<cognos_installation_directory>/webcontent"

<Directory "<cognos_installation_directory>/cgi-bin">
    Options None
    AllowOverride None
    Order allow,deny
    Allow from all
</Directory>

<Directory "<cognos_installation_directory>/webcontent">
    Options Indexes MultiViews
    AllowOverride None
    Order allow,deny
    Allow from all
</Directory>
```

Figure 4-94 Additional configuration for Cognos on Apache HTTP Server

Note: You can run Apache HTTP Server under a different credential than apache. For example, edit the `http.conf` file to change the credentials to the user and group that you want:

```
#
# If you want HTTPd to run as a different user or group, you must run
# httpd as root initially and it will switch.
#
# User/Group: The name (or #number) of the user/group to run HTTPd as.
# It is usually a preferred practice to create a dedicated user and group for
# running httpd, as with most system services.
#
User nobody
Group nobody
```

2. Run the **apache configtest** command to check the `httpd.conf` file syntax, as shown in Figure 4-95.

```
/usr/local/apache2/bin/apachectl configtest
```

Figure 4-95 apachectl configtest

3. Fix any syntax errors and start the Apache HTTP Server service, as shown in Figure 4-96.

```
/usr/local/apache2/bin/apachectl -f /usr/local/apache2/conf/httpd.conf
```

Figure 4-96 apachectl start daemon

4.5.7 Copy DB2 client drivers to Cognos libraries

Follow these steps:

1. Copy the `db2jcc` files to the `webapps` library directory.
2. If you selected the default DB2 installation, those files are at `/opt/ibm/db2/V10.5/java/db2jcc.jar` and `/opt/ibm/db2/V10.5/java/db2jcc_license_cu.jar`.
3. Copy these files to your <*cognos_installation_directory*>`/webapps/p2pd/WEB-INF/lib/`.

4.5.8 Apply Cognos fix packs

This step is not optional for our demonstration because a few required features for our scenario are provided in Fix Pack 2 for Cognos Business Intelligence Server 10.2.2.

Follow these steps:

1. Stop the Cognos Business Intelligence Server with the command that is shown in Figure 4-97.

```
<cognos_installation_directory>/bin64/cogconfig.sh -stop
```

Figure 4-97 Stop Cognos Business Intelligence Server

2. Download and extract Fix Pack 2. Run **`./linuxppc64h/issetup`**. Follow the similar procedure that is described in 4.5.5, “Configure Cognos Business Intelligence” on page 127.
3. Start the Cognos Business Intelligence Server in silence mode with the command that is shown in Figure 4-98.

```
<cognos_installation_directory>/cogconfig.sh -s
```

Figure 4-98 Start Cognos Business Intelligence Server

4.5.9 Install Framework Manager

To publish a model to the Cognos portal, you need Framework Manager. Follow these steps to install Framework Manager:

1. Extract `fm_10.2.2_win_ml.tar.gz` to `fm_10.2.2_win_ml`.
2. Navigate to the `\\fm_10.2.2_win_ml\win32\issetup.exe` file location.
3. Right-click **issetup.exe**.
4. Select **Run as Administrator**.
5. Click **English**. Click **Next**.
6. Click **I Agree**. Click **Next**.
7. Use the default setting. Ensure that you set the server use type to **Production**. Click **Next**.
8. If the message "`The directory <installpath> does not exist. Do you want to create it during installation?`" displays, answer **Yes** (Figure 4-99). Click **Next**.

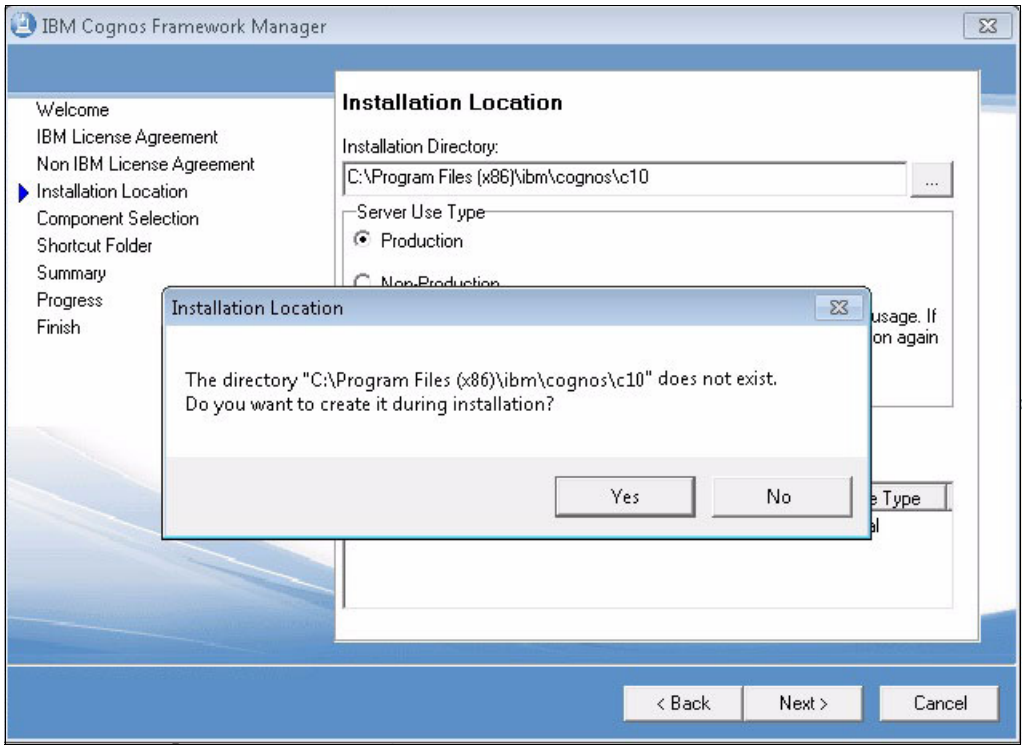


Figure 4-99 Message about creating a folder during the Framework Manager installation

9. Click **Next**.
10. Wait until the installation completes.
11. Select **Start** → **All Programs** → **IBM Cognos**.
12. Start **Cognos Configuration**.
13. Select **Local Configuration** → **Environment**.

14. Use the following settings:
    - For Gateway settings > Gateway URI:

      `http://servername:80/cognos/cgi-bin/cognos.cgi`
    - For other URI settings > Dispatcher URI for external application:

      `http://servername:80/p2pd/servlet/dispatch`
15. Click **File Save**.

4.5.10 Apply Cognos Fix Packs for client

This step is not optional for our demonstration because a few required features for our scenario are provided in Fix Pack 2 for Cognos Business Intelligence Server 10.2.2. If you apply the fix pack to the server, you must apply Fix Pack 2 to your Framework Manager, too.

Follow these steps:

1. Close **Framework Manager**.
2. Extract `up_bisrvr_win32_10.2.6102.54_ml` from `up_bisrvr_winx64h_10.2.6102.54_ml.tar`.
3. Navigate to the `location\up_bisrvr_win32_10.2.6102.54_ml.tar\up_bisrvr_win32_10.2.6102.54_ml\win32` file.
4. Right-click **issetup.exe**.
5. Select **Run as Administrator**.
6. Click **English**. Click **Next**.
7. Click **I Agree**. Click **Next**.
8. Use the default setting. Ensure that you set the server use type to **Production**. Click **Next**.
9. If the message "`The directory <installpath> does not exist. Do you want to create it during installation?`" displays, answer **Yes** (Figure 4-99 on page 131). Click **Next**.
10. Click **Next**.
11. Wait until the installation completes.
12. Select **Start** → **All Programs** → **IBM Cognos**.
13. Start **Cognos Configuration**.
14. Select **Local Configuration** → **Environment**.
15. Use the following settings:
    - For Gateway settings > Gateway URI:

      `http://servername:80/cognos/cgi-bin/cognos.cgi`
    - For other URI settings > Dispatcher URI for external application:

      `http://servername:80/p2pd/servlet/dispatch`
16. Click **File Save**.

5

Scenario: Integration of the components for the solution

This chapter describes the integration of the components for a possible analytics solution architecture that uses IBM Power Systems.

At first glance, many products from various sources appear to fully integrate themselves to offer an operational analytics solution. However, to provide an optimized solution, you are better off implementing products that are designed for integration, such as IBM BigInsights for Apache Hadoop on top of IBM Open Platform with Apache Hadoop.

In our demonstration, we take full advantage of IBM Big SQL, which provides American National Standards Institute (ANSI) Structured Query Language (SQL) access to structured data that is spread across BigInsights systems, through Java Database Connectivity (JDBC) or Object Database Connectivity (ODBC), and other relational databases, such as IBM DB2. Therefore, SQL application developers can access data in Hadoop transparently with their current skill set, without learning any new language or technology.

Cognos Business Intelligence is certified for IBM Big SQL, which means that Cognos Business Intelligence can access data that is stored in IBM BigInsights. And in our demonstration, Cognos reports will retrieve data from both IBM DB2 with BLU Acceleration and IBM Big SQL. Sales from the fictional GOSALES Cognos sample database with sentiment analysis for the correspondent product that is extracted from Insights for Twitter, an IBM Bluemix service, will represent the input for our implemented solution.

IBM Statistical Package for the Social Sciences (SPSS) Analytical Decision Management and the SPSS Scoring Service will be integrated with web store pages to customize fictional offerings and discounts, based on customers' particular characteristics and the general polarity of their Twitter posts.

To implement your own solution and describe ours in-depth, this chapter describes the following topics:

- IBM Big SQL integration
- Cognos Business Intelligence integration
- SPSS Analytical Decision Management and Scoring Services integration

5.1 IBM Big SQL integration

You can enable access to structured data that is stored in IBM BigInsights for Apache Hadoop and integrate with other SQL development platforms. Big SQL provides this capability for applications to access data in IBM BigInsights by using ANSI SQL.

Businesses across the world use SQL to analyze their data, primarily because it is simple and powerful. This situation created departments with strong SQL skills, and Big SQL uses those skills to maximize the value that you can pull from your BigInsights environment.

Big SQL uses SQL to produce a view of your existing data. The storage format relies on Hive, HBase, or BigInsights distributed file system so 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. Beware that Big SQL supports querying data through SELECT statements, but it does not support UPDATE or DELETE statements. INSERT statements for Big SQL are supported only on HBase tables.

Big SQL also supports the *IBM Data Server* Client drivers. Therefore, you can keep the same standards-compliant JDBC, Java Call Control (JCC), ODBC, call level interface (CLI), and .NET drivers that are used in other IBM software products, such as DB2 for Linux, UNIX, and Windows, DB2 for z/OS, and IBM Informix database software. Other languages that already use these drivers can interact with Big SQL with minimum custom configuration.

Administrators can use the **LOAD** command to populate Big SQL tables by using various data sources, for example, flat files as .txt and .csv files, and the following external data sources:

- DB2
- Netezza®
- Teradata
- Microsoft SQL Server
- Oracle
- Informix

In addition to working with the data that you stored in BigInsights, Big SQL 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 federated system to process.

5.1.1 Proposed solution integration with IBM Big SQL

IBM Big SQL plays a major role in our proposed solution (Figure 5-1) by storing data that was extracted from *Insights for Twitter a Bluemix Service* to provide data later to both Cognos Business Intelligence and SPSS Analytical Decision Management.

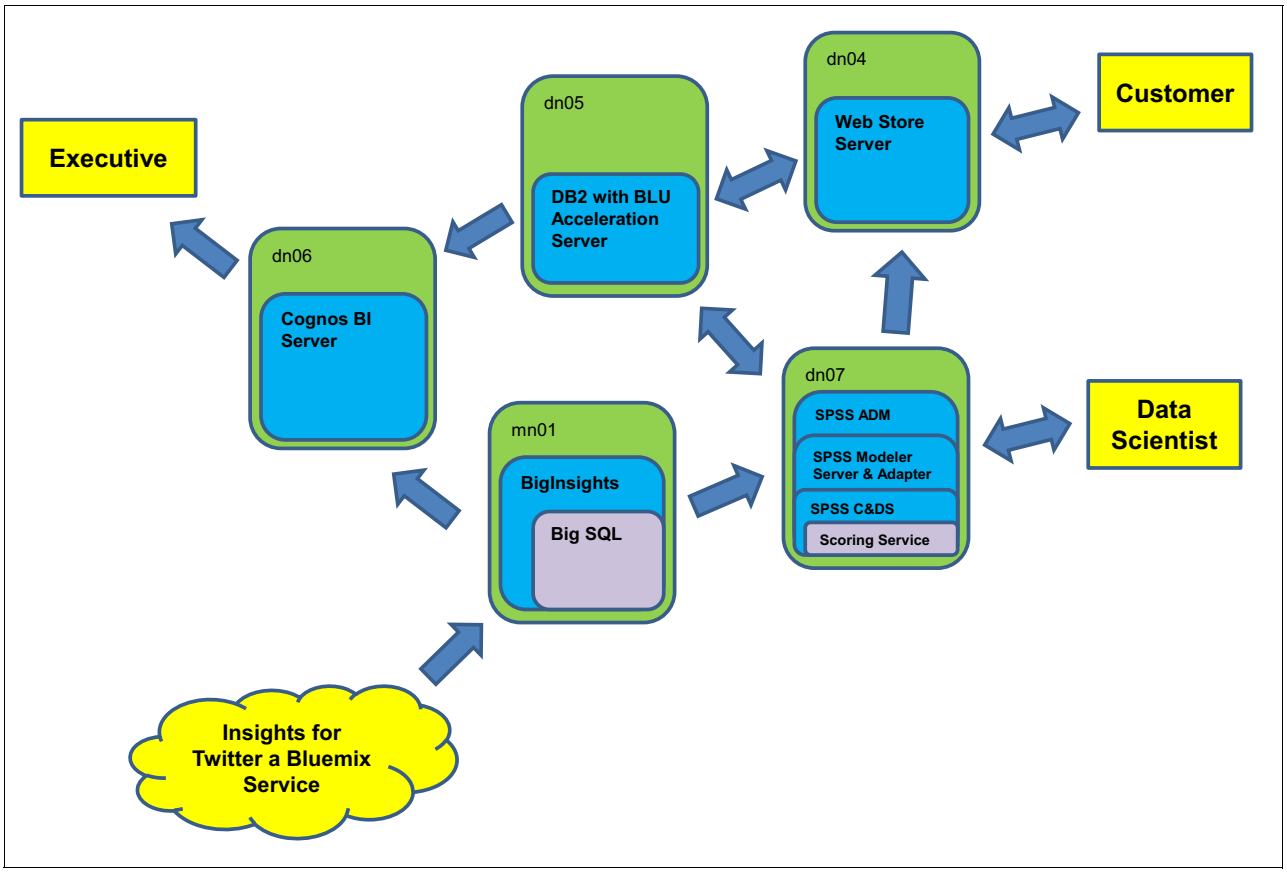


Figure 5-1 Proposed solution deployment diagram

In Figure 5-1, the blue arrows describe the simplified data flow. They are described later in this chapter. The yellow cloud and boxes represent all of the external entities:

- Insights for Twitter a Bluemix Service: This service provides tweets metadata, including text content polarity, for example, positive, negative, or neutral comments about a product, that is essential to this sentimental analysis solution. This service is courtesy of the IBM alliance with Twitter that will be used to collect the public perception of one specific product and also track the overall tweets polarity of one particular customer that is identified by their twitter handle.
- Data scientist: SPSS Analytical Decision Management is a tool for data scientists to predict the customer positive response ratio for a new campaign and to predict the profitability of the interaction based on the historical purchase behavior of the customer individual personality.

- Executive: Cognos Business Intelligence dashboards will show the total amount of sales with overall public perception by product. Executives and decision makers will decide the most appropriate campaign. They can also benefit from the insights of SPSS Analytical Decision Management to maximize profits by customizing offers to their customers.
- Customer: Web Store is the online shopping channel. The offer will be customized based on the historical purchase behavior of the customer by using the online SPSS Scoring Service. If the customer-associated twitter handle is previously identified, the tweets polarity of the particular customer is determined and it will also support the decision for a customized offer. A positive response for a campaign will sometimes correlate to the general tweet polarity.

For reference, the host names and services that each node holds in our sample solution are described:

- mn01: This node is the Big SQL head node (be aware that data nodes that were previously configured behind the BigInsights are not disclosed to simplify the deployment diagram):
  - BigInsights:
    - Big SQL
- dn04: This node is the Web Store node that interacts with the online shopping customers, displaying customized offers based on online queries into SPSS Scoring Services:
  - Web Store
- dn05: This node by using DB2 with BLU Acceleration holds data that supports the Web Store for customer interactions. This node also holds actual and aggregated sales data to feed Cognos Business Intelligence reports and dashboards. In addition, this node holds the SPSS Collaboration and Deployment Services repository:
  - DB2 with BLU Acceleration
- dn06: This node provides reports and dashboards to company executives and decision makers, extracting actual and aggregated data from DB2 with BLU Acceleration and twitter sentiment analysis and additional detailed sales data from Big SQL:
  - Cognos Business Intelligence
- dn07: This node provides SPSS Scoring Services within SPSS Collaboration and Deployment Services for individual customer offers to the online Web Store. In addition, it stores repository data in a DB2 regular row-organized database. The interaction with Big SQL is intended to retrieve the overall tweets polarity, identify the general mood of the particular customer, and maximize positive responses to the sales campaigns. Data Scientists will use SPSS Analytical Decision Management to identify rules to maximize profits for a new campaign. However, to install SPSS Analytical Decision Management, you also need to install SPSS Modeler Server Adapters for Collaboration and Deployment Services, which requires SPSS Modeler Server in addition to SPSS Collaboration and Deployment Services (C&DS):
  - SPSS Collaboration and Deployment Services (C&DS):
    - Scoring Services
  - SPSS Modeler Server
  - SPSS Modeler Server Adapters for Collaboration and Deployment Services
  - SPSS Analytical Decision Management (ADM)

Figure 5-2 shows the host names and services that each node holds in our sample solution.

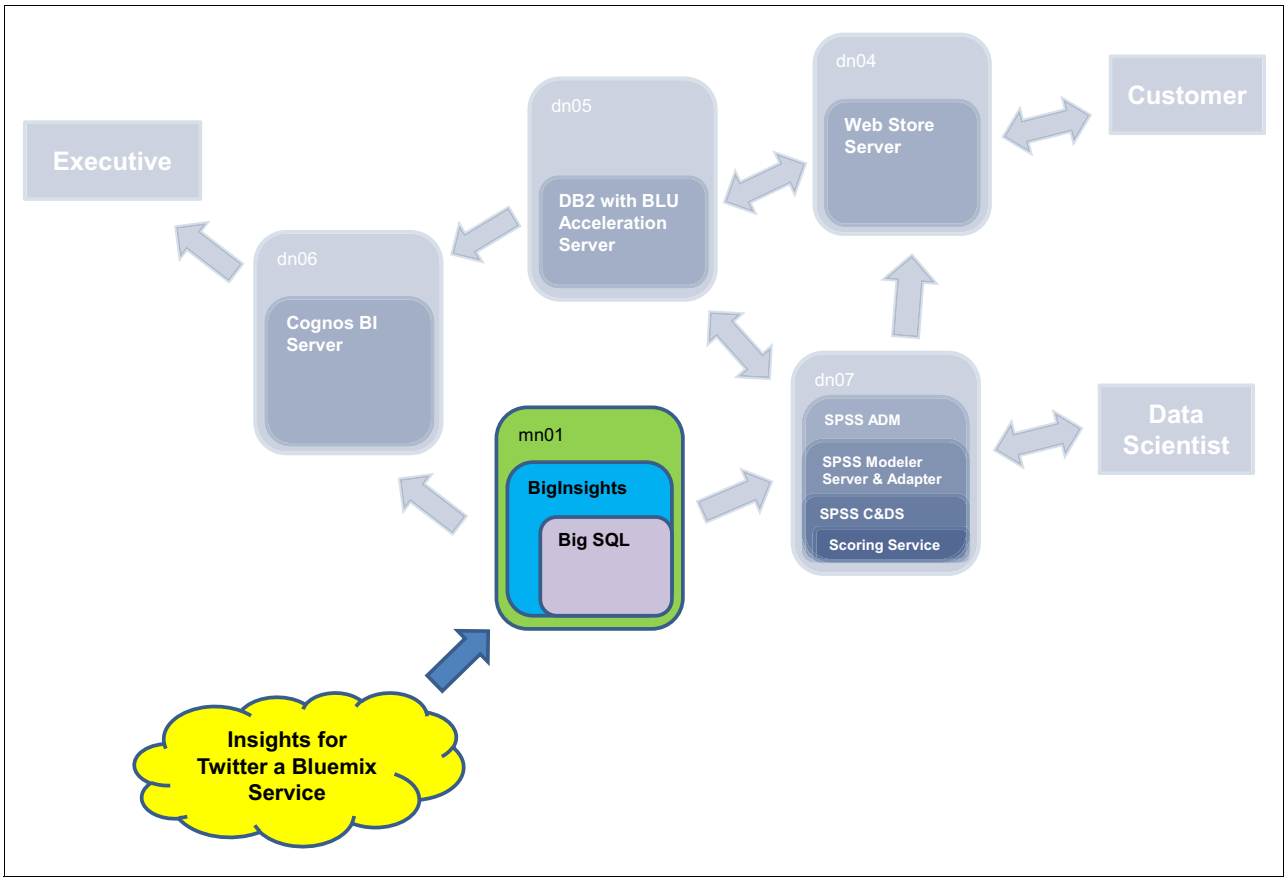


Figure 5-2 IBM Big SQL integration with Insights for Twitter a Bluemix Service

For our sample solution, server mn01 is the head node for the IBM Big SQL service. It will store transformed data transparently that was extracted from Insights for Twitter into the BigInsights data nodes that are not disclosed in Figure 5-2. The BigInsights storage layer is omitted intentionally.

Before we describe how to transform and store structured data into IBM Big SQL, we explain how to connect from different sources to support the implementation of your own solution.

For the integration of other components, we focus on IBM Data Server drivers, but you can connect to IBM Big SQL server by using the following methods:

- IBM Data Server Manager for Big SQL
- Java SQL Shell (**JSqsh**) command interface
- IBM Data Server drivers

5.1.2 Configure the connection with IBM Data Server Manager for Big SQL

The default URL for your IBM Data Server Manager for Big SQL is shown:

`https://<head-node>:8443/gateway/default/BigInsightsWeb/index.html#/welcome`

In Figure 5-3, the default password for the user guest is `guest-password`.

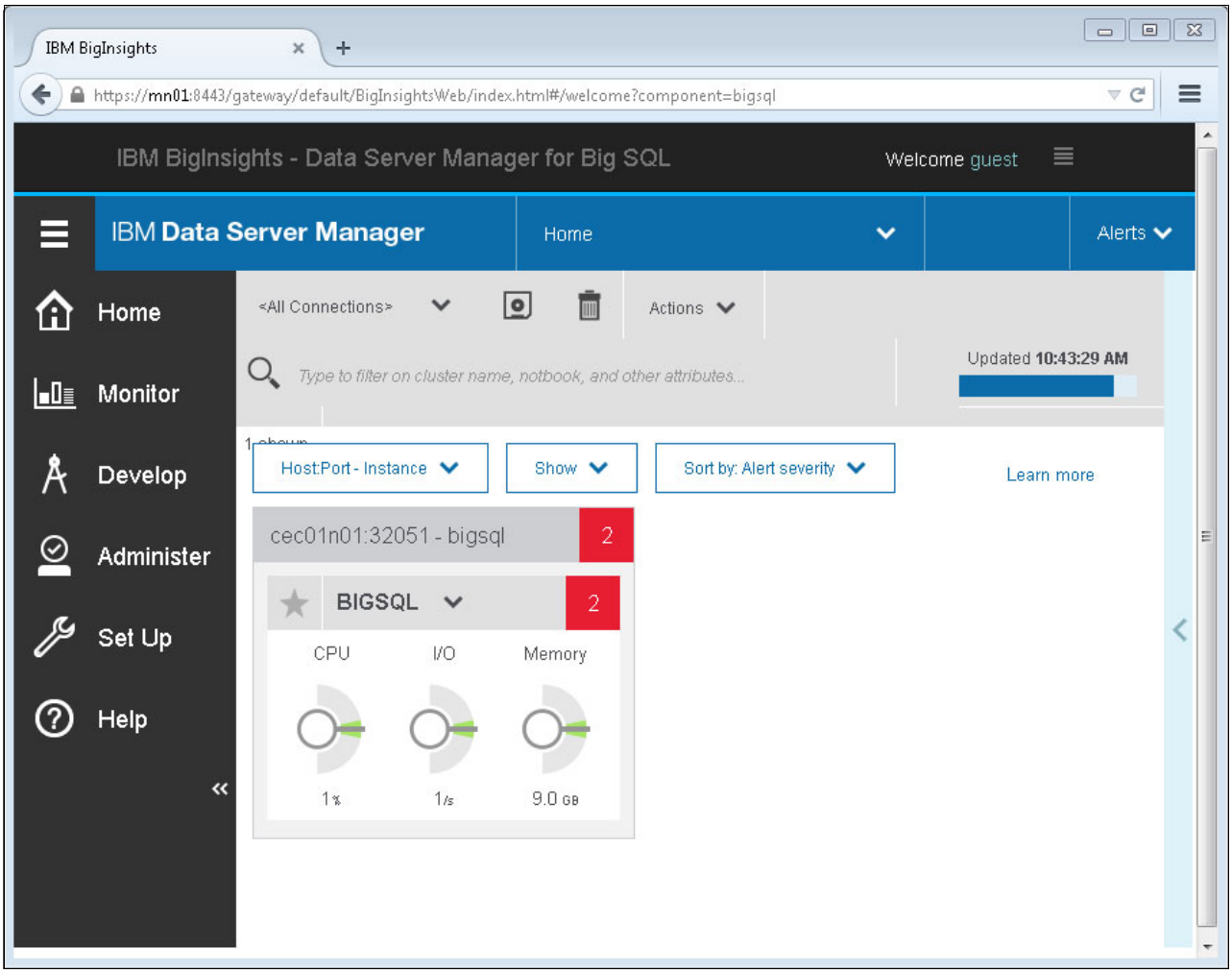


Figure 5-3 IBM Data Server Manager for Big SQL Home page

To configure your first database connection, on Figure 5-3, click **Actions** → **Discover Databases**.

The Discover Databases page opens (Figure 5-4). Enter the name, user ID, and password. Click **Find Instances** according to your environment settings, and you will find the available Big SQL servers.

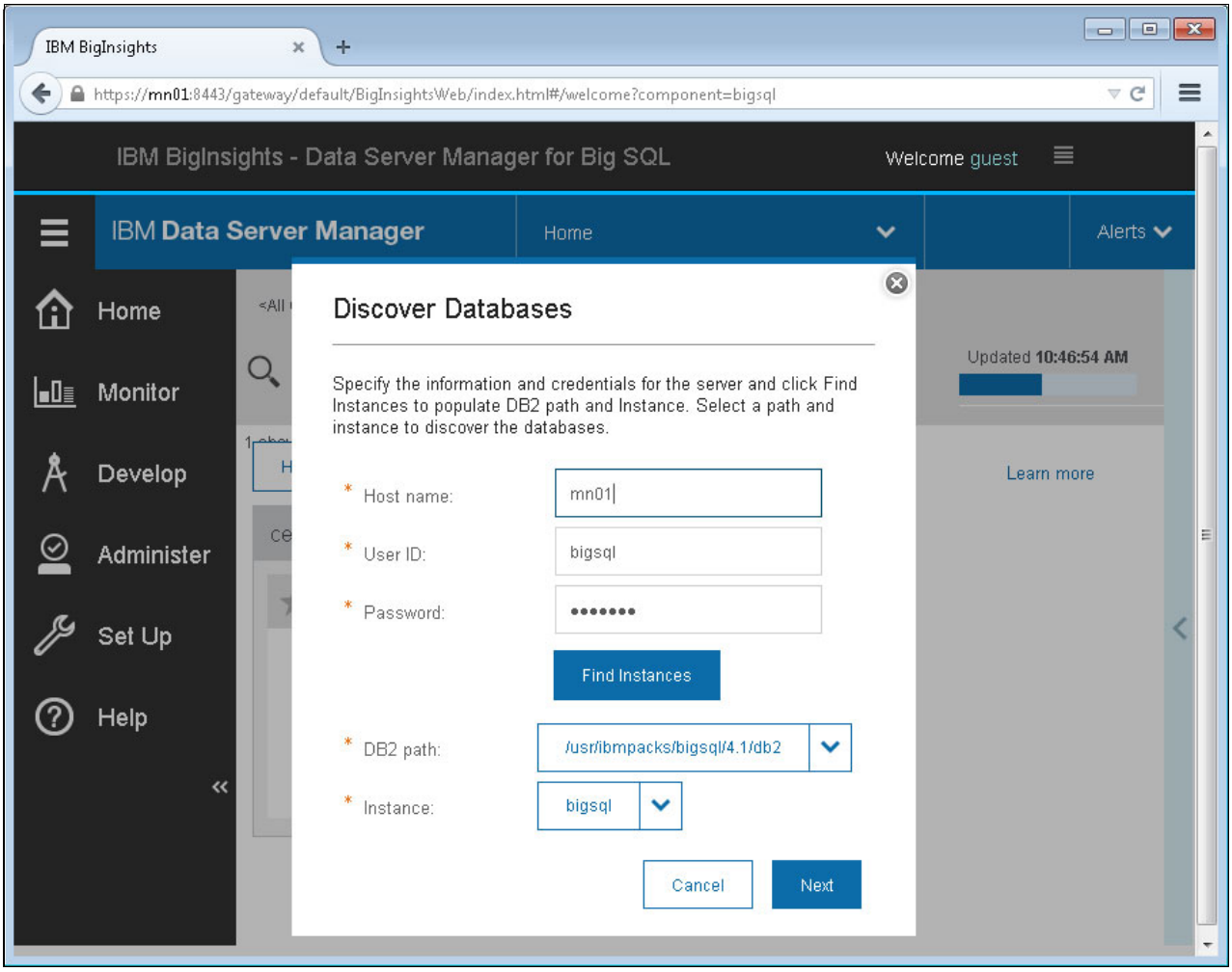


Figure 5-4 IBM Data Server Manager for Big SQL: Discover Databases

Select your instance. In our demonstration, it is `bigsql`. Click **Develop** → **SQL Editor**. Type the following SQL statement to test your connection and click **Run**:

```
select * from sysibm.sysdummy1
```

A successful test output is shown in Figure 5-5.

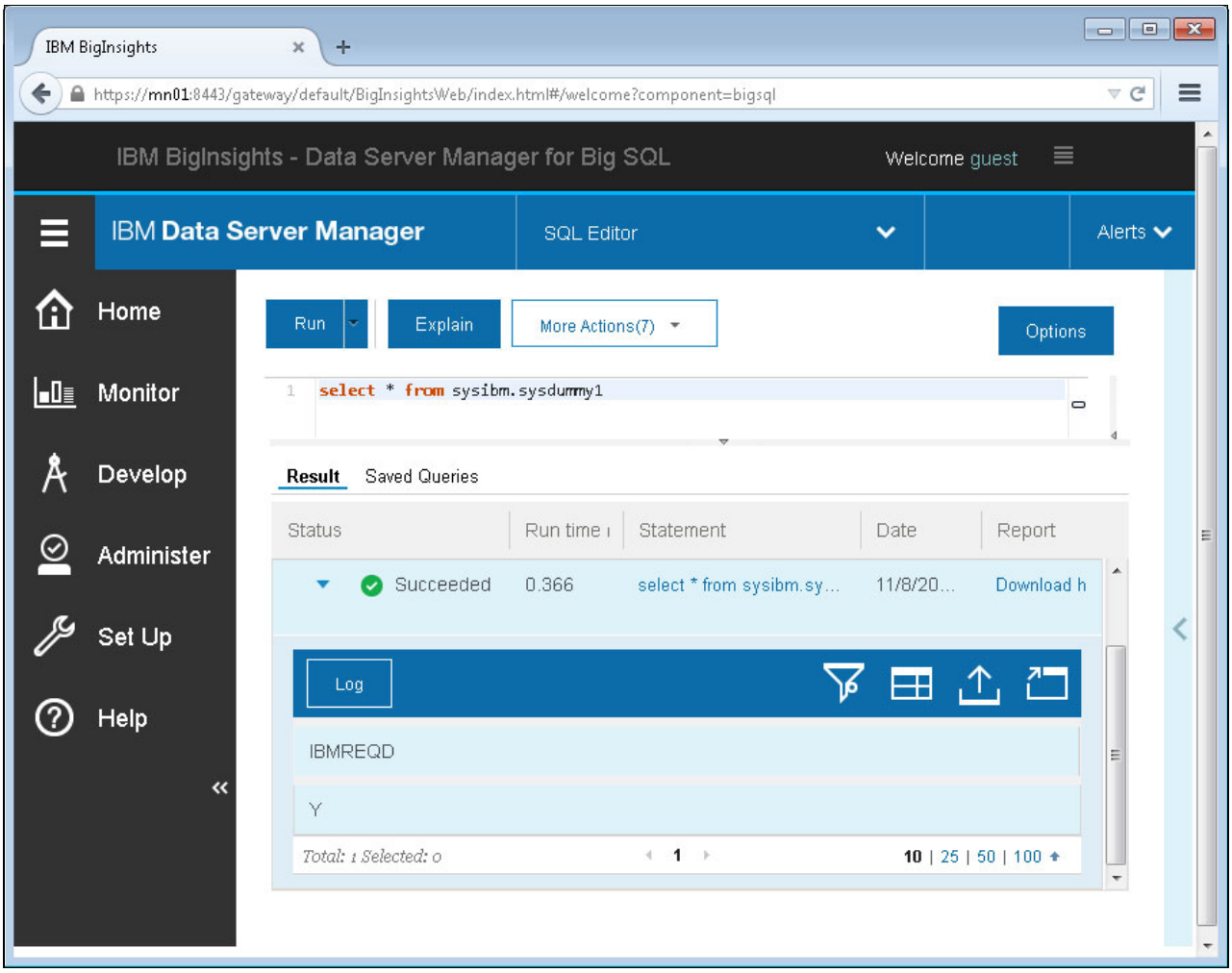


Figure 5-5 Successful connection test

5.1.3 Configure the connection with the jsqsh command line

On the shell prompt, type `jsqsh`. The first time that you use **`jsqsh`** on the command shell, a configuration wizard starts. When you are at the **`jsqsh`** command prompt, type `\drivers` to determine the available drivers, which are listed:

- IBM Data Server (DB2, Informix, and Big SQL) - com.ibm.db2.jcc.DB2Driver
- IBM DB2 z/OS - com.ibm.db2.jcc.DB2Driver
- Apache Derby Server - org.apache.derby.jdbc.ClientDriver
- Apache Derby Embedded - org.apache.derby.jdbc.EmbeddedDriver
- Firebird JayBird - org.firebirdsql.jdbc.FBDriver
- Apache Hive - org.apache.hadoop.hive.jdbc.HiveDriver
- Apache Hive - org.apache.hive.jdbc.HiveDriver
- IBM Informix - com.informix.jdbc.IfxDriver
- JDBC ODBC Bridge - sun.jdbc.odbc.JdbcOdbcDriver
- MS SQL Server - com.microsoft.jdbc.sqlserver.SQLServerDriver
- MS SQL Server jTDS - net.sourceforge.jtds.jdbc.Driver
- MS SQL Server 2005+ - com.microsoft.sqlserver.jdbc.SQLServerDriver
- MySQL - com.mysql.jdbc.Driver
- IBM Netezza - org.netezza.Driver

- Oracle - oracle.jdbc.OracleDriver
- Oracle OCI - oracle.jdbc.driver.OracleDriver
- PostgreSQL - org.postgresql.Driver
- Sybase ASE - com.sybase.jdbc3.jdbc.SybDriver
- Sybase ASA - com.sybase.jdbc2.jdbc.SybDriver
- Sybase ASE jTDS - net.sourceforge.jtds.jdbc.Driver

The first time that you open JSQSH, the setup wizard launches. The panel that is shown in Figure 5-6 appears.

Follow these steps to configure the connection:

1. Type `C` to start the connection wizard, as shown in Figure 5-6.

```
JSQSH SETUP WIZARD

Welcome to the jsqsh setup wizard! This wizard provides a (crude) menu
driven interface for managing several jsqsh configuration files. These
files are all located in $HOME/.jsqsh, and the name of the file being
edited by a given screen will be indicated on the title of the screen

Note that many wizard screens require a relative large console screen
size, so you may want to resize your screen now.

(C)onnection management wizard
   The connection management wizard allows you to define named connections
   using any JDBC driver that jsqsh recognizes. Once defined, jsqsh only
   needs the connection name in order to establish a JDBC connection

(D)river management wizard
   The driver management wizard allows you to introduce new JDBC drivers
   to jsqsh, or to edit the definition of an existing driver. The most
   common activity here is to provide the classpath for a given JDBC driver

Choose (Q)uit, (C)onnection wizard, or (D)river wizard:
```

Figure 5-6 JSQSH setup wizard

2. If a connection is not created, type `A` to add the connection, as shown in Figure 5-7.

```
JSQSH CONNECTION WIZARD - (edits $HOME/.jsqsh/connections.xml
The following connections are currently defined:

     Name                 Driver     Host                           Port
---  -------------------- ---------- ------------------------------ ------
  1  bigsql               db2        mn01                           32051

Enter a connection number above to edit the connection, or:
(B)ack, (Q)uit, or (A)dd connection:
```

Figure 5-7 JSQSH connection management wizard

3. Figure 5-8 shows the available drivers. Choose a driver for your new connection.

```
JSQSH CONNECTION WIZARD - (edits $HOME/.jsqsh/connections.xml)

Choose a driver for use by your new connection

     Name             Target               Class
---  ---------------- --------------------
-------------------------------------------------
  1 *db2             IBM Data Server (DB2 com.ibm.db2.jcc.DB2Driver
  2 *db2zos          IBM DB2 z/OS         com.ibm.db2.jcc.DB2Driver
  3 *hive2           Apache Hive          org.apache.hive.jdbc.HiveDriver
  4 *mysql           MySQL                com.mysql.jdbc.Driver
  5  derby           Apache Derby Server  org.apache.derby.jdbc.ClientDriver
  6  derbyembed      Apache Derby Embedde org.apache.derby.jdbc.EmbeddedDriver
  7  firebird        Firebird JayBird     org.firebirdsql.jdbc.FBDriver
  8  hive            Apache Hive          org.apache.hadoop.hive.jdbc.HiveDriver
  9  informix        IBM Informix         com.informix.jdbc.IfxDriver
 10  jdbcodbc        JDBC ODBC Bridge     sun.jdbc.odbc.JdbcOdbcDriver
 11  mssql           MS SQL Server        com.microsoft.jdbc.sqlserver.SQLServerDriver
 12  mssql-jtds      MS SQL Server jTDS   net.sourceforge.jtds.jdbc.Driver
 13  mssql2k5        MS SQL Server 2005+  com.microsoft.sqlserver.jdbc.SQLServerDriver
 14  netezza         IBM Netezza          org.netezza.Driver
 15  oracle          Oracle               oracle.jdbc.OracleDriver
 16  oracleoci       Oracle OCI           oracle.jdbc.driver.OracleDriver
 17  pgsql           PostgreSQL           org.postgresql.Driver
 18  sybase          Sybase ASE           com.sybase.jdbc3.jdbc.SybDriver
 19  sybase-asa      Sybase ASA           com.sybase.jdbc2.jdbc.SybDriver
 20  sybase-jtds     Sybase ASE jTDS      net.sourceforge.jtds.jdbc.Driver

  * = Driver is available. If a driver is unavailable, you may choose (D) below
      to jump to the driver wizard to provide a classpath

Enter the driver number, (D)river wizard, (B)ack or (Q)uit:
```

Figure 5-8 JSQSH driver selection for new connections

4. Enter the number 1 for *db2 for Big SQL databases. Complete the fields according to your environment settings (Figure 5-9).

```
JSQSH CONNECTION WIZARD - (edits $HOME/.jsqsh/connections.xml)

The following configuration properties are supported by this driver.

    Connection name : _temp_
             Driver : IBM Data Server (DB2, Informix, Big SQL)
           JDBC URL : jdbc:db2://${server}:${port}/${db}

Connection URL Variables
------------------------
1                db : BIGSQL
2              port : 32051
3            server : localhost
4              user : bigsql
5          password :
6       Autoconnect : false

JDBC Driver Properties
----------------------
None

Enter a number to change a given configuration property, or
(T)est, (B)ack, (Q)uit, Add (P)roperty, or (S)ave:
```

Figure 5-9 Big SQL connection settings example

5. Test your connection settings (Figure 5-9). Save and name this connection.
6. Connect to your database by using the **\connect *<database>*** command and perform your first query, as shown in Figure 5-10.

```
Welcome to JSqsh 4.4
Type "\help" for help topics. Using JLine.
1> \connect bigsql
Password: *******
[cec01n01.cluster.com][bigsql] 1> select * from sysibm.sysdummy1;
+---------+
| IBMREQD |
+---------+
| Y       |
+---------+
1 row in results(first row: 0.3s; total: 0.3s)
```

Figure 5-10 Connecting to the Big SQL database and performing an SQL query

5.1.4 Cognos Business Intelligence data source by using the DB2 CLI connection

As one example of IBM Data Server drivers, you can catalog a Big SQL database by using DB2 command line processor (CLP) from any node with access to the IBM Big SQL service ports.

Cognos Business Intelligence can use JDBC to connect to Big SQL but it can also use DB2 call level interface (CLI). Therefore, in our example, we will catalog the Big SQL database as any regular DB2 database in a Cognos Business Intelligence server.

On the Cognos Business Intelligence node, log in as the db2 instance owner, in our environment, `db2inst1`. Catalog the Big SQL node and database, as shown in Figure 5-11.

```
db2 catalog tcpip node <bigsql_node> remote <head-node> server <server-port>

db2 catalog db <bigsql_database> at node <bigsql_node>

db2 connect to <bigsql_database> user <user_id> using <user_password>

db2 "select * from sysibm.sysdummy1"
```

Figure 5-11 Commands to catalog and test the connection to Big SQL

The commands in Figure 5-11 will perform the following actions (output is shown in Example 5-1):

1. Catalog the remote node.
2. Catalog the database.
3. Connect to the database.
4. Perform a simple SQL query for connection testing.

Example 5-1 Output example of commands to catalog and test the connection to Big SQL

```
db2inst1@dn06 ~]$ db2 catalog tcpip node bigsql remote mn01 server 32051
DB20000I  The CATALOG TCPIP NODE command completed successfully.
DB21056W  Directory changes may not be effective until the directory cache is
refreshed.
[db2inst1@dn06 ~]$ db2 catalog db bigsql at node bigsql
DB20000I  The CATALOG DATABASE command completed successfully.
DB21056W  Directory changes may not be effective until the directory cache is
refreshed.
[db2inst1@dn06 ~]$ db2 connect to bigsql user bigsql
Enter current password for bigsql:

   Database Connection Information

 Database server        = DB2/LINUXPPC64LE 10.6.3
 SQL authorization ID   = BIGSQL
 Local database alias   = BIGSQL

[db2inst1@dn06 ~]$ db2 "select * from sysibm.sysdummy1"

IBMREQD
-------
Y
  1 record(s) selected.
```

Note: You must catalog the head node. Otherwise, you can receive the following error when you query data:

```
SQL0969N  There is no message text corresponding to SQL error "-5115" in the
message file on this workstation.  The error was returned from module
"SQLRROAA" with original tokens "SQL statements only allowed on head node"
```

If this error occurs, run the **uncatalog node** and **catalog node** commands by using the head node settings. Also, issue the **db2 terminate** command to recycle your session connections. Try to connect and perform the simple SQL query again:

```
[db2inst1@dn06 ~]$ db2 "select * from sysibm.sysdummy1"
SQL0969N  There is no message text corresponding to SQL error "-5115" in the
message file on this workstation.  The error was returned from module
"SQLRROAA" with original tokens "SQL statements only allowed on head node".
db2inst1@dn06 ~]$  db2 uncatalog node bigsql
DB20000I  The UNCATALOG NODE command completed successfully.
DB21056W  Directory changes may not be effective until the directory cache is
refreshed.
[db2inst1@dn06 ~]$ db2 catalog tcpip node bigsql remote mn01 server 32051
DB20000I  The CATALOG TCPIP NODE command completed successfully.
DB21056W  Directory changes may not be effective until the directory cache is
refreshed.
[db2inst1@dn06 ~]$ db2 terminate
DB20000I  The TERMINATE command completed successfully.
[db2inst1@dn06 ~]$ db2 connect to bigsql user bigsql
Enter current password for bigsql:

   Database Connection Information

 Database server        = DB2/LINUXPPC64LE 10.6.3
 SQL authorization ID   = BIGSQL
 Local database alias   = BIGSQL

[db2inst1@dn06 ~]$ db2 "select * from sysibm.sysdummy1"

IBMREQD
-------
Y

  1 record(s) selected.
```

Follow these steps:

1. To test the Big SQL recently created connection from the DB2 CLP, we will create a data source. On the Cognos Administration page, `http://<cognos_node>/ibmcognos`, click **Configuration** → **Data Source Connections**. Click **New Data Source** and name it, for example, `Big SQL test`, as shown in Figure 5-12.

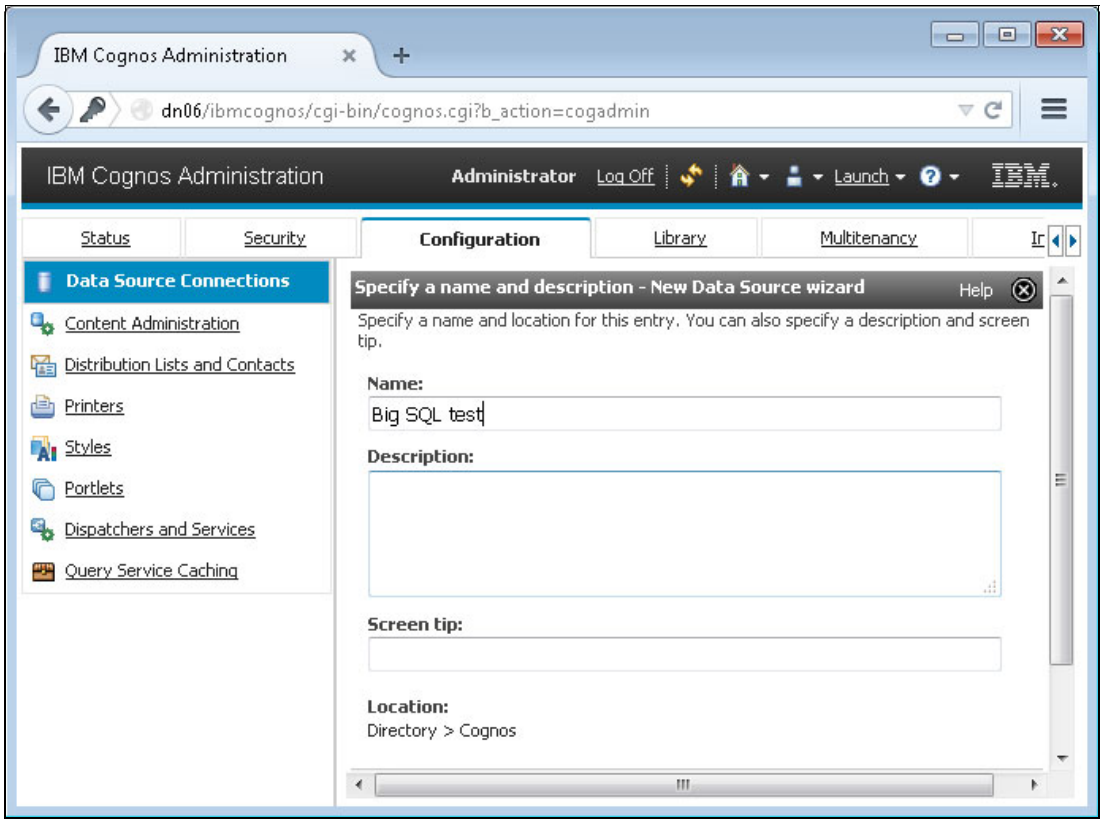


Figure 5-12 Creating a Big SQL test data source by using DB2 CLI

2. Select **IBM DB2** for the data source type. Clear **Configure JDBC Connection** (because we want to use the recently created DB2 database catalog with DB2 CLI for testing), as shown in Figure 5-13.

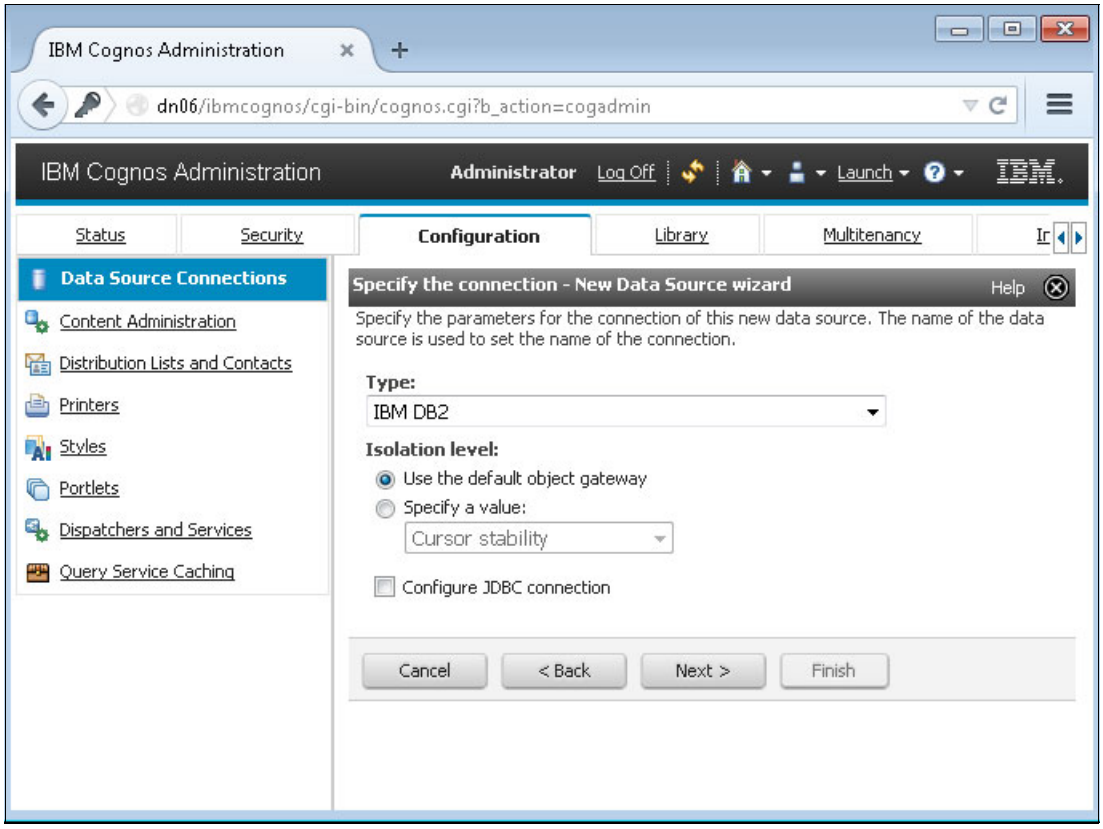


Figure 5-13 Select IBM DB2 and clear the option to configure a JDBC connection

3. On the next window, in the DB2 database name field, type the database name that you cataloged, in our case, `BIGSQL`. In the Signons area, check **Password** and click **Test the connection**, as shown in Figure 5-14.

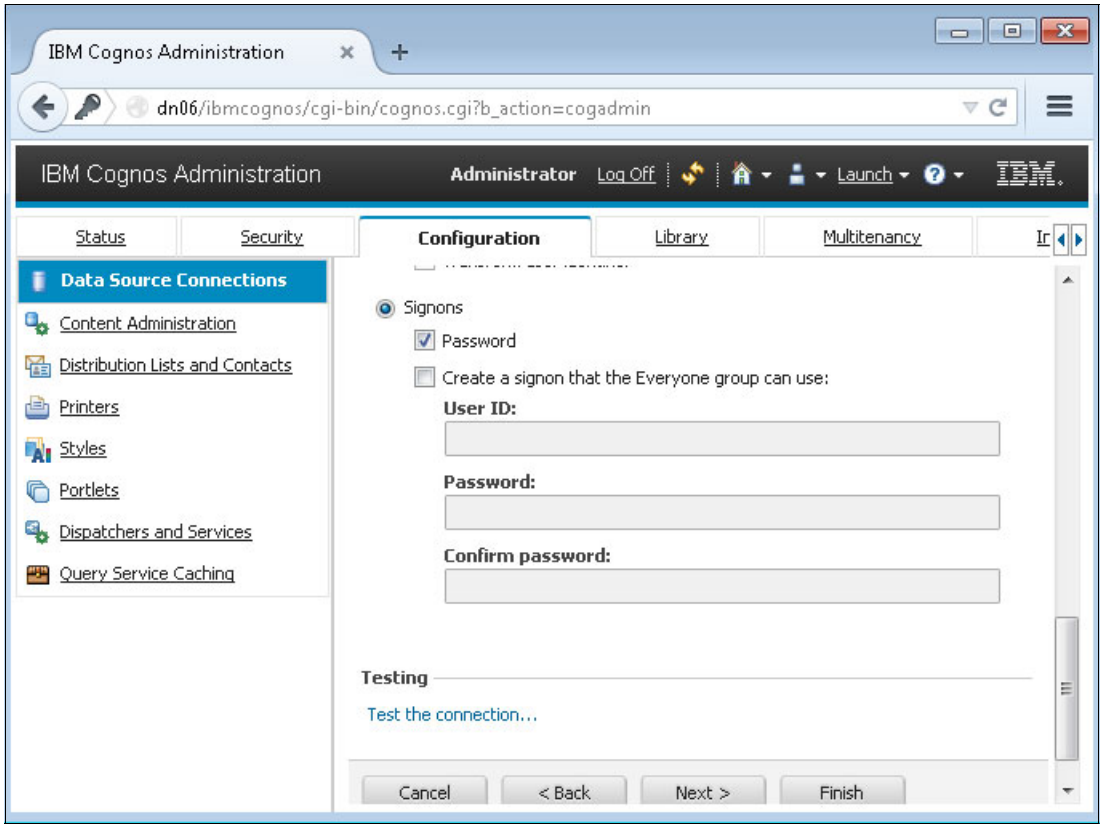


Figure 5-14 Enter the DB2 database name and test the connection

4. On the next window, type your `bigsql` user ID and password. Click **Test**, as shown in Figure 5-15.

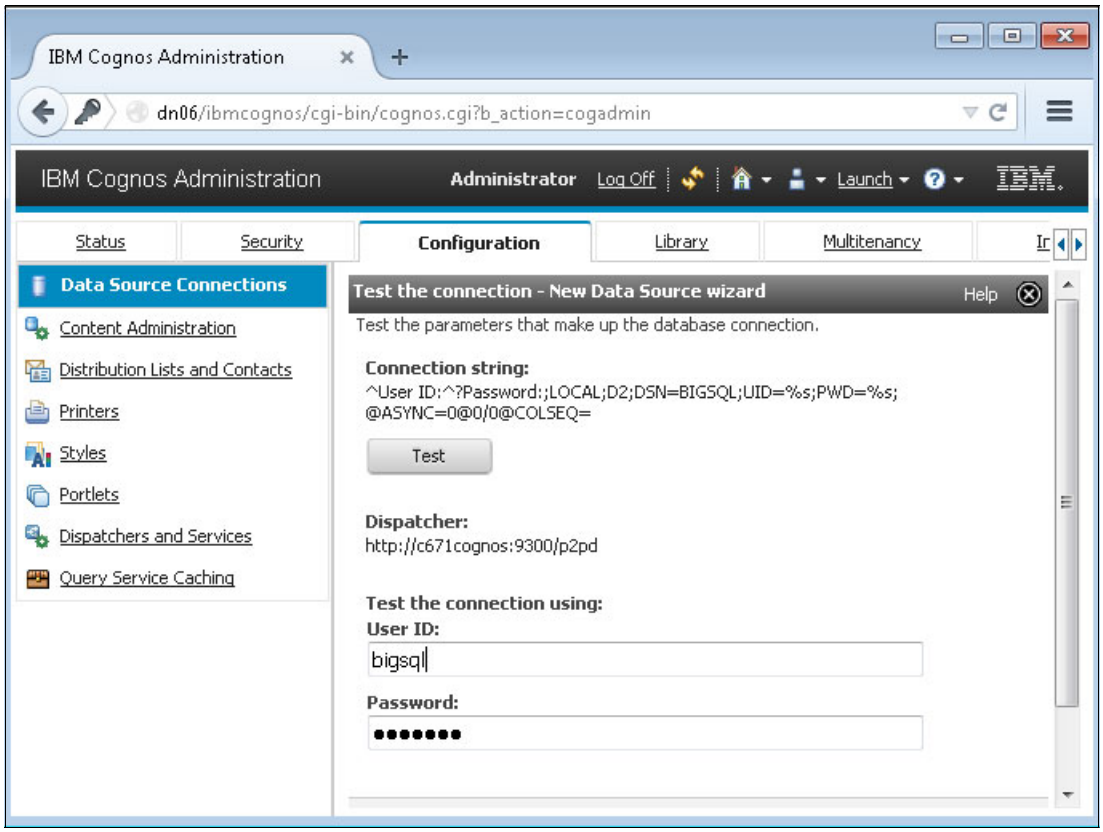


Figure 5-15 Type the appropriate bigsql user ID and password for your environment settings

5. If your settings are correct and your services are up and running, you receive the following message, as shown in Figure 5-16.

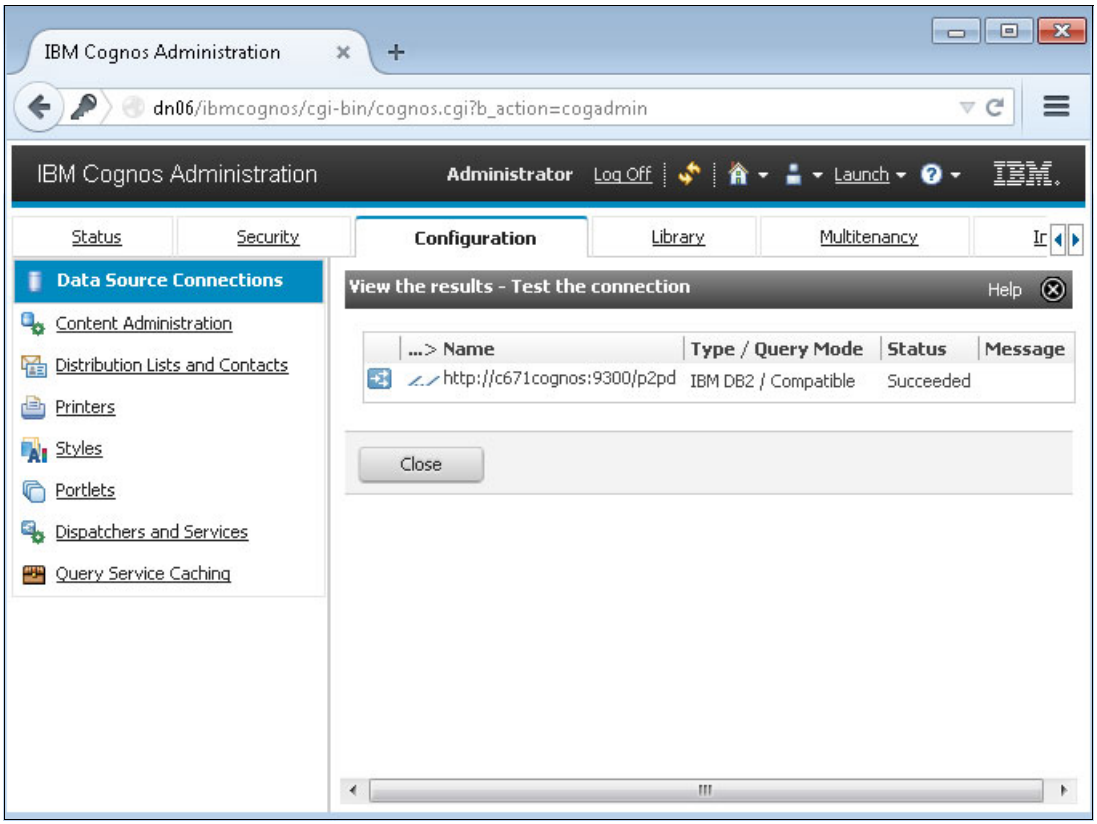


Figure 5-16 Connection with IBM DB2 is compatible and successful

5.1.5 Cognos Business Intelligence data source by using the JDBC connection

Follow these steps:

1. As another example of the IBM Data Server drivers, you can configure the JDBC connection to the Big SQL database by using the DB2 JDBC drivers from any node with access to the IBM Big SQL service ports, as shown in Figure 5-17. Select **JDBC**.

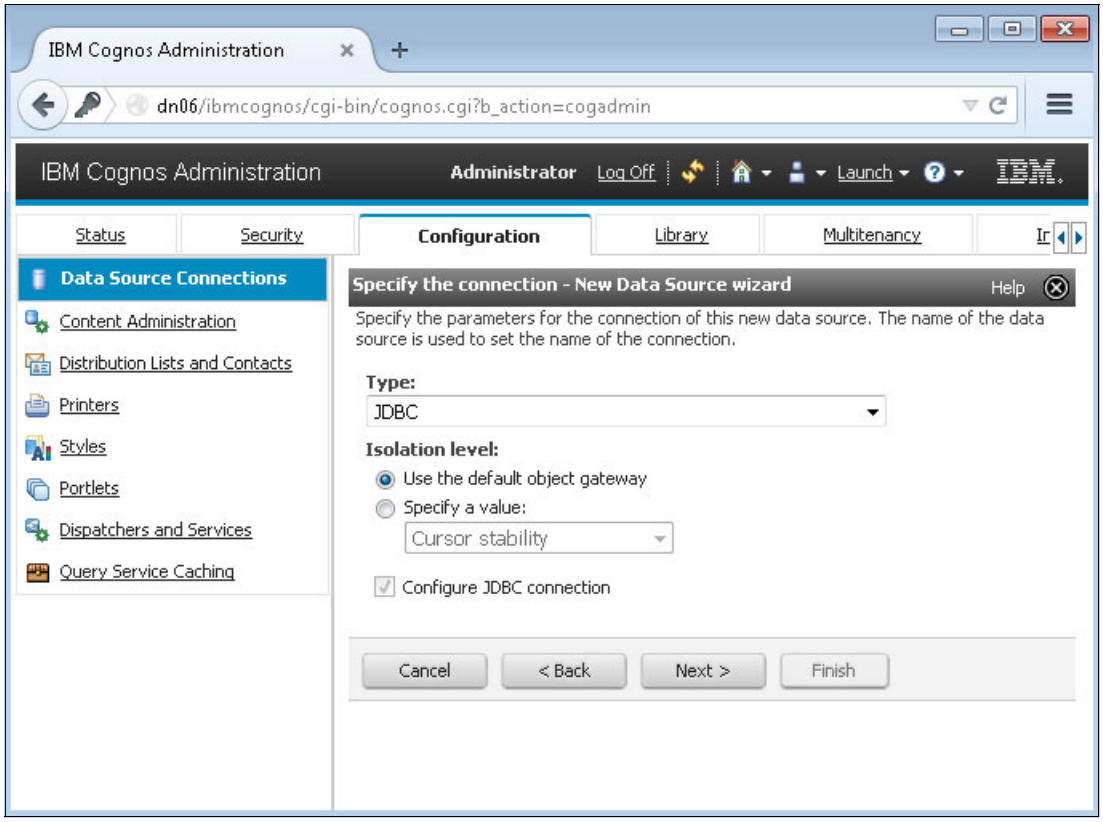


Figure 5-17 JDBC connection

2. On the next window, you select **IBM InfoSphere BigInsights (Big SQL)** as the type of the data source. Provide the JDBC URL with your environment's settings in the format `jdbc:db2://<host>:<port>/<database-name>`, as shown in Figure 5-18.

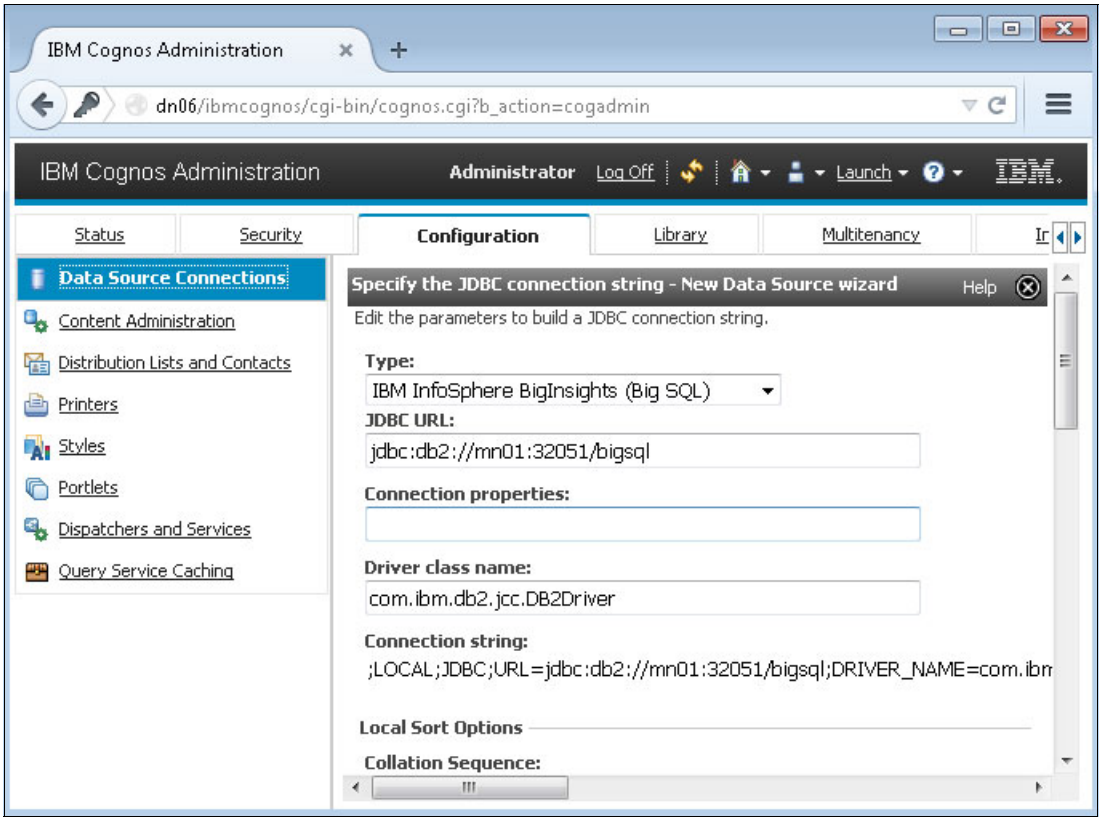


Figure 5-18 JDBC InfoSphere BigInsights (Big SQL) settings

3. In the Signons area, check **Password** and click **Test the connection**, as shown in Figure 5-14 on page 148.
4. The Test the connection window opens. Type your `bigsql` user ID and password. Click **Test**.
5. If your settings are correct and your services are up and running, you receive the successful connection message, as shown in Figure 5-19.

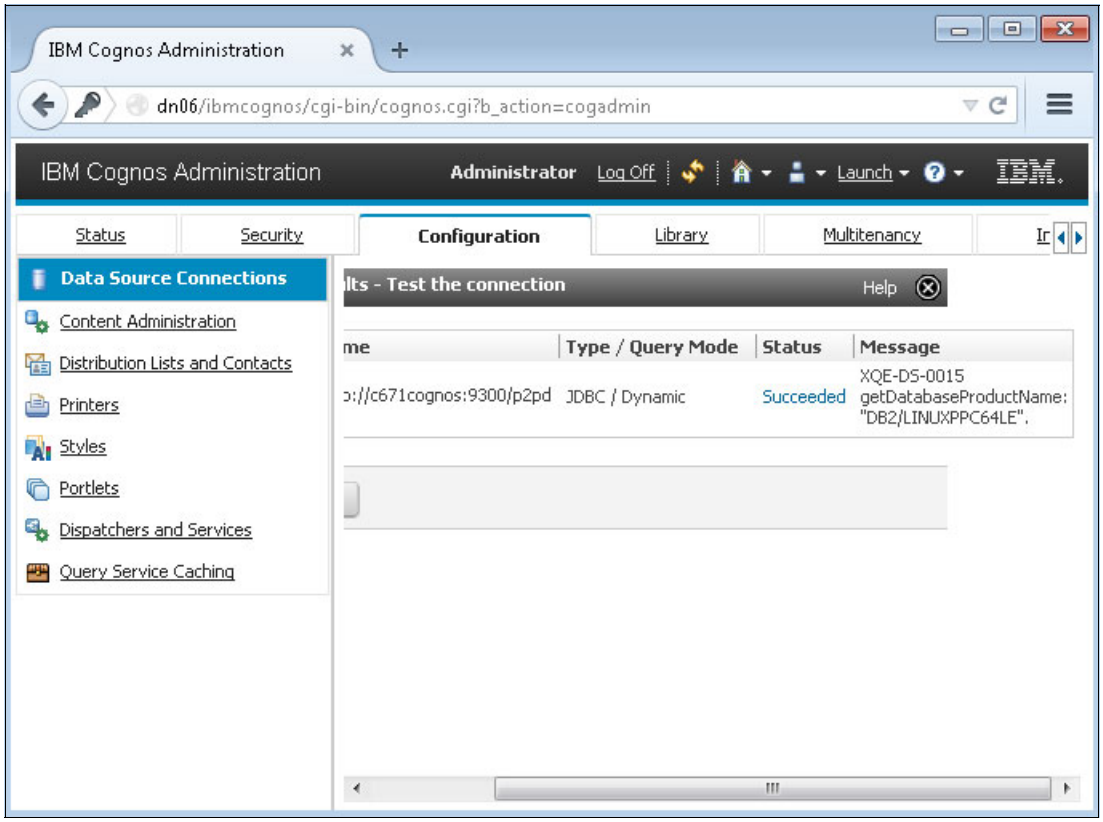


Figure 5-19 JDBC connection test succeeded

5.1.6 IBM Big SQL federation

Data analytics is crucial to the success of business. Therefore, data warehousing that supports the consolidation and federation of data is a requirement.

Big SQL federation enables a mixed environment that creates the consolidation of big data and the federation of live data from different data sources.

At the time that this publication was written, Table 5-1 shows the list of supported data sources that can be used to federate to Big SQL.

Table 5-1 Data sources that are supported for Big SQL federation

Data source	Supported versions	Notes
DB2	DB2 for Linux, UNIX, and Windows 9.7, 9.8, 10.1, and 10.5	
Oracle	11g, 11g R1, and 11g R2	
Teradata	12, 13, and 14	Not supported on Power Systems.
Netezza	4.6, 5.0, 6.0, and 7.2	Not supported on Power Systems.
Informix	11.5	
Microsoft SQL Server	2012	

Set up the federation to DB2 with BLU Acceleration

Follow these steps:

1. In your Big SQL head node, log in as the instance owner, which is `bigsql`, by default, and double-check whether the federation is enabled for immediate usage by issuing the command that is shown in Figure 5-20.

```
db2 get dbm cfg | grep "Federated Database System Support"
```

Figure 5-20 Checking Big SQL federated settings

Note: If the FEDERATED setting is not YES, you must enable federation by using the following command:

```
db2 update dbm cfg using FEDERATED YES
```

Then, stop and start the bigsql instance with the following commands:

```
$BIGSQL_HOME/bin/bigsql stop
$BIGSQL_HOME/bin/bigsql start
```

2. Catalog the node and the database for DB2 with BLU Acceleration for your environment settings, as shown in Figure 5-21.

```
db2 catalog tcpip node <db2_blu_node> remote <db2_blu_host> server
<db2_blu_port>

db2 catalog db <db2_blu_database> at node <db2_blu_node>
```

Figure 5-21 Catalog DB2 with BLU Acceleration for the node and database

3. Create a wrapper, federated server, and user mappings with the commands for your environment settings, as shown in Figure 5-22.

```
db2 connect to bigsql user bigsql

db2 CREATE WRAPPER DRDA LIBRARY \'libdb2drda.so\'

db2 "CREATE SERVER DB2REMOTE TYPE DB2/UDB VERSION 10.5 WRAPPER DRDA
AUTHORIZATION \"<db2_blu_instance_owner>\" PASSWORD
\"<db2_blu_instance_owner_password>\" OPTIONS (DBNAME '<db_blu_database>')"

db2 “CREATE USER MAPPING FOR <big_sql_instance_owner> SERVER DB2REMOTE OPTIONS
( REMOTE_AUTHID '<db2_instance_owner>', REMOTE_PASSWORD
'<db2_instance_owner_password>' )”
```

Figure 5-22 Create a wrapper, federated server, and user mappings

4. We assume that you already populated your DB2 database with GOSALES Cognos sample data. We create an example nickname for table GOSALESDW.GO_REGION_DIM and then perform a simple query on the remote table, as shown in Example 5-2.

Example 5-2 Nickname example for table GOSALESDW.GO_REGION_DIM

```
[bigsql@mn01 ~]$ db2 CREATE NICKNAME DB2REMOTE.GO_REGION_DIM FOR DB2REMOTE.GOSALESDW.GO_REGION_DIM
DB20000I  The SQL command completed successfully.
[bigsql@mn01 ~]$ db2 "select count(*) from DB2REMOTE.GO_REGION_DIM"

1
--------------------------------
                              21.

  1 record(s) selected.
```

By using this federation setup, you can consolidate data from multiple data sources in the Big SQL instance. In addition, you can provide access to Big SQL users to multiple data sources. Check that you create a user mapping for each Big SQL user. Consult the syntax of the DB2 command **CREATE USER MAPPING** for details.

5.1.7 Loading GOSALESDW data into Big SQL

You can use the IBM Data Server Manager for Big SQL, JSQSH, or IBM DB2 Data Server drivers to create tables, load data, and query Big SQL data.

A set of scripts is available to create and load GOSALESDW sample data into your Big SQL database under the directory `$BIGSQL_HOME/samples`.

Carefully read the `$BIGSQL_HOME/samples/README` file for instructions and run the command that is shown in Figure 5-23.

```
./setup.sh -u <userid> -s <host_name> -n <port number> -d <databaseName>
```

Figure 5-23 Loading GOSALESDW into Big SQL

Alternatively, you can create all of the GOSALESDW tables manually by using $BIGSQL_HOME/samples/queries/GOSALESDW_ddl.sql with JSQSH and running the command that is shown in Figure 5-24.

```
jsqsh <db> -U <user> -P <pass> < $BIGSQL_HOME/samples/queries/GOSALESDW_ddl.sql
```

Figure 5-24 Creating the GOSALESDW tables in Big SQL with jsqsh

Load the data with the **LOAD HADOOP** command by using a modified version of $BIGSQL_HOME/samples/queries/GOSALESDW_load.sql, updating FILE URL to SFTP or any other appropriate method.

For instance, in our case, we used GPFS, which refers to Spectrum Scale with File Placement Optimizer, as shown in Example 5-3.

Example 5-3 LOAD HADOOP by using SFTP or Spectrum Scale

```
load hadoop using file url
'sftp://<user>:<pass>@<sftp_node>:/usr/ibmpacks/current/bigsql/bigsql/samples/data/GOSALESD
W.BURST_TABLE2.txt' with SOURCE PROPERTIES ('field.delimiter'='\t') into table
GOSALESDW.BURST_TABLE2 overwrite;

load hadoop using file url
'gpfs:////bigpfs/user/bigsql/DemoData/gs_db/GOSALESDW.BURST_TABLE2.txt' with SOURCE
PROPERTIES ('field.delimiter'='\t') into table GOSALESDW.BURST_TABLE2 overwrite;
```

5.1.8 Loading Twitter Data into Big SQL

To load Twitter data for demonstration, we assume that you have a tweets feed with sentimental analysis, for example, Insights for Twitter an IBM Bluemix service for Data and Analytics.

You must be authorized to access the following service application programming interface (API):

https://cdeservice.mybluemix.net/api/v1/messages/search?q=ibm&size=500

By using that service API, notice that many attributes of a tweet are retrieved, for example, the author's gender, location, and the number of followers.

The IBM alliance with Twitter sentiment analysis also provides the polarity of the content: NEUTRAL, POSITIVE, or NEGATIVE. See Example 5-4.

Example 5-4 IBM Bluemix Insights for Twitter output

```
{
    "search": {
        "results": 889857,
        "current": 1
    },
    "tweets": [{
        "cde": {
            "author": {
                "gender": "male",
                "parenthood": {
                    "isParent": "unknown",
                    "evidence": ""
```

```
                },
                "location": {
                    "country": "United States",
                    "city": "",
                    "state": "North Carolina"
                },
                "maritalStatus": {
                    "isMarried": "unknown",
                    "evidence": ""
                }
            },
            "content": {
                "sentiment": {
                    "evidence": [],
                    "polarity": "NEUTRAL"
                }
            }
        },
        "message": {
            "postedTime": "2013-11-05T16:45:53.000Z",
            "verb": "post",
            "link": "http://twitter.com/brenny/statuses/397766729256161281",
            "generator": {
                "displayName": "Hootsuite",
                "link": "http://www.hootsuite.com"
            },
            "body": "RT @ibm_iod: Now: @BethTSmith showing IBM Predictive
Maintenance &amp;amp; Quality demo - Capture, Predict, Act #ibmiod
http://t.co/JP5EncVurN",
            "favoritesCount": 0,
            "objectType": "activity",
            "actor": {
                "summary": "Social Insights Senior Analyst. Interested in
#gamification, MOOCs; Percussionist, Dad, writer, wannabe tennis player.  Tweets
are my own.",
                "image":
"https://pbs.twimg.com/profile_images/53493646/bran-headshot_normal.jpg",
                "statusesCount": 3203,
                "utcOffset": "-21600",
                "languages": ["en"],
                "preferredUsername": "brenny",
                "displayName": "Brenny Ganesan",
                "postedTime": "2008-02-07T23:20:34.000Z",
                "link": "http://www.twitter.com/brenny",
                "verified": false,
                "friendsCount": 1827,
                "twitterTimeZone": "Central Time (US & Canada)",
                "favoritesCount": 91,
                "listedCount": 112,
                "objectType": "person",
                "links": [{
                    "rel": "me",
                    "href": "http://www.servicemanagement360.com"
                }],
                "location": {
```

```
                        "displayName": "RTP NC",
                        "objectType": "place"
                    },
                    "id": "id:twitter.com:13225972",
                    "followersCount": 1847
                },
                "provider": {
                    "displayName": "Twitter",
                    "link": "http://www.twitter.com",
                    "objectType": "service"
                },
                "twitter_filter_level": "low",
                "twitter_entities": {
                    "urls": [{
                        "display_url": "youtube.com/watch?v=pwdGIu…",
                        "indices": [116, 138],
                        "expanded_url": "http://www.youtube.com/watch?v=pwdGIuMnVl4",
                        "url": "http://t.co/JP5EncVurN"
                    }],
                    "hashtags": [{
                        "indices": [108, 115],
                        "text": "ibmiod"
                    }],
                    "user_mentions": [{
                        "indices": [3, 11],
                        "screen_name": "ibm_iod",
                        "id_str": "2463969721",
                        "name": "IBM Insight",
                        "id": 2463969721
                    }, {
                        "indices": [18, 29],
                        "screen_name": "BethTSmith",
                        "id_str": "568818502",
                        "name": "Beth Smith",
                        "id": 568818502
                    }],
                    "trends": [],
                    "symbols": []
                },
                "twitter_lang": "en",
                "id": "tag:search.twitter.com,2005:397766729256161281",
                "retweetCount": 1,
                "object": {
                    "summary": "RT @ibm_iod: Now: @BethTSmith showing IBM Predictive
Maintenance &amp;amp; Quality demo - Capture, Predict, Act #ibmiod
http://t.co/JP5EncVurN",
                    "postedTime": "2013-11-05T16:45:53.000Z",
                    "link": "http://twitter.com/brenny/statuses/397766729256161281",
                    "id": "object:search.twitter.com,2005:397766729256161281",
                    "objectType": "note"
                }
            }
        }],
        "related": {
            "next": {
```

```
            "href":
"https://cdeservice.stage1.mybluemix.net:443/api/v1/messages/search?q=ibm&from=1&s
ize=1"
         }
     }
}
```

You can create your own solution for extracting the required data from the tweet feed to load into your Big SQL by using any programming language that is compatible with JavaScript Object Notation (JSON), for example, the C family of languages, which includes these languages:

- C
- C++
- C#
- Java
- JavaScript
- Perl
- Python

Alternatively you can use Serializer/Deserializer (SerDe), which is a common component in Hadoop environments. For JSON data, as shown in Example 5-4 on page 156, you can rely on the appropriate Hive SerDe interface for JSON to read and interpret the semi-structured data for your applications.

For our demonstration, we used another approach. A few tweets that mention products from the fictional GOSALES sample database were extracted from Twitter by using Bluemix Insights for Twitter service. Those tweet feeds, which were originally semi-structured data, by using JSON format were transformed into structured data by using proprietary scripts. Then, they were associated to their correspondent product data in a .csv file format for preparation to load into Big SQL. See Example 5-5 on page 160.

The sample contains the following product-related columns:

- Count
- SearchObject
- Category
- Brand
- Format
- Family
- Subfamily
- Product

In addition, a few columns that relate to tweets that mention those products are listed:

- ObjectProvenance
- Source
- IsBuzz
- IsSentiment
- IsIntent
- IsOwnership
- IsCustomerOf
- Polarity
- CreatedTime
- FullName
- Gender
- City

- State
- Screenname
- UserID
- Userverified
- FollowersCount
- IsParent
- IsMarried
- Text

Example 5-5 shows the transformed tweet feeds that were associated to their correspondent product data in a structured .csv file format for preparation to load into Big SQL.

Example 5-5 Transformed structured .csv data from twitter feeds

```
"Count","SearchObject","Category","Brand","Format","Family","SubFamily","Product","ObjectProvenance","Source","IsBuzz","IsSentiment","IsIn
tent","IsOwnership","IsCustomerOf","Polarity","CreatedTime","FullName","Gender","City","State","Screenname","UserID","Userverified","Follo
wersCount","IsParent","IsMarried","Text"
1,"GRANITE HAMMER","MOUNTAINEERING
EQUIPMENT","GRANITE",,,,"HAMMER","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/20/2012
4:53",,,,,"brayinharlem09",7653578196,"FALSE",90,,,"Don't expect to use the Granite Hammer for heavy duty work. It's soft as clay. #LAME"
1,"BUGSHIELD NATURAL","OUTDOOR
PROTECTION","BUGSHIELD",,,,"NATURAL","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/12/2012
15:32",,,,,"brayinharlem09",7653578196,"FALSE",90,,,"I sooooo hate BugShield Natural... I came back with 37 bites.. on my arm ALONE!"
1,"LADY HAILSTORM TITANIUM IRONS","GOLF
EQUIPMENT","HAILSTORM",,"IRONS",,,"USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/14/2012 22:04","Chris
Hemmingway","Male","NEW YORK CITY","NY","CHemmingway",3020000000000000000,"FALSE",440,,,"She liked the Lady Hailstorm Titanium Irons!
thanks @shirley"
1,"CANYON MULE COOLER","CAMPING EQUIPMENT","CANYON
MULE",,,,"COOLER","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/19/2012 0:42","Pina
Pender",,"MIAMI","FL","PinaPender",5550000000000000000,"FALSE",564,,,"Hey @jani check this retweet out: I like to own Canyon Mule Cooler.
You were right"
1,"LADY HAILSTORM STEEL IRONS","GOLF
EQUIPMENT","HAILSTORM",,"IRONS",,,"USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/9/2012 8:48","Pina
Pender",,"MIAMI","FL","PinaPender",5550000000000000000,"FALSE",564,,,"Do you love to have the Lady Hailstorm Steel Irons"
1,"GRANITE EXTREME","MOUNTAINEERING
EQUIPMENT","EXTREME",,,,"EXTREME","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/6/2012 0:54","Luisa
Amatea","Female","NEW YORK CITY","NY","LAmatea111",8350000000000000000,"FALSE",129,,,"I like owning my Granite Extreme ;)) [pic]
twitpic.com/url01"
1,"TRAILCHEF KETTLE","CAMPING
EQUIPMENT","TRAILCHEF",,,,"KETTLE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/25/2012
21:21",,"Female","BOSTON","MA","lucyluca90",7653556196,"FALSE",35,,,"If Someone wanna buy TrailChef Kettle, awesome deal at
http://bit.ly/3rvghff"
1,"DELUXE FAMILY RELIEF KIT","OUTDOOR
PROTECTION","RELIEF",,,,,"USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/20/2012 1:13","Abram
Barber","Male","HOUSTON","TX","AbBarber3",5840000000000000000,"FALSE",134,,,"Check this out: I recommend a Deluxe Family Relief Kit,
awesome deal at http://bit.ly/3rvghff"
1,"EDGE EXTREME","PERSONAL ACCESSORIES","EXTREME",,,,"EDGE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/11/2012
6:45","Gabriel Walters","Male","ATLANTA","GA","GabWalters",8740000000000000000,"FALSE",826,,,"If anyone wants to give me a Edge Extreme,
awesome deal at http://bit.ly/3rvghff ;)"
1,"COURSE PRO GOLF AND TEE SET","GOLF EQUIPMENT","COURSE PRO",,,,"GOLF AND TEE
SET","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/10/2012 16:32","Annalaura Pappalardo","Female","LOS
ANGELES","CA","APappalardo",4470000000000000000,"FALSE",573,,,"@JoeMoe2 since you brought it up.. I hate the Course Pro Golf and Tee Set
you use.. jus' sayin"
1,"DELUXE FAMILY RELIEF KIT","OUTDOOR
PROTECTION","RELIEF",,,,,"USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","10/5/2012 19:47","Annalaura
Pappalardo","Female","LOS ANGELES","CA","APappalardo",4470000000000000000,"FALSE",573,,,"I hate those annoying Deluxe Family Relief Kit
commercials as much as i hate the real thing"
1,"CANYON MULE CLIMBER BACKPACK","CAMPING EQUIPMENT","CANYON MULE",,,,"CLIMBER
BACKPACK","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/21/2012 21:40","Norman
Wood","Male","MIAMI","FL","NormanWood",3600000000000000000,"FALSE",97,,,"I hate this Canyon Mule Climber Backpack -_-"
1,"FIREFLY EXTREME","CAMPING
EQUIPMENT","EXTREME",,,,"EXTREME","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/17/2012 7:36","Olive
Davidson","Female","ALBANY","GA","Oliveon",5610000000000000000,"FALSE",937,,,"I hate the Firefly Extreme"
1,"INSECT BITE RELIEF","OUTDOOR PROTECTION",,,,,"INSECT BITE","USER","Twitter","FALSE","TRUE","FALSE","FALSE","FALSE","Negative","9/2/2012
1:37","Olive Davidson","Female","ALBANY","GA","Oliveon",5610000000000000000,"FALSE",937,,,"I almost feel bad for giving the huge bottle of
Insect Bite Relief away! It relieve nothing lol"
1,"STAR PEG","CAMPING EQUIPMENT","STAR",,,,"PEG","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/20/2012
20:19","Olive Davidson","Female","ALBANY","GA","Oliveon",5610000000000000000,"FALSE",937,,,"My family wasted so much $... and now hates
their lame Star Peg lol #fail"
1,"HAILSTORM STEEL WOODS SET","GOLF EQUIPMENT","HAILSTORM",,"WOODS",,"STEEL WOODS
SET","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/25/2012 8:12","Rosy
Red","Female",,,"flowergirl67",76535196678,"FALSE",0,,,"I love having my Hailstorm Steel Woods Set when i need it. Check out this deal
http://t.co/ades323d"
1,"COURSE PRO GLOVES","GOLF EQUIPMENT","COURSE
PRO",,,,"GLOVES","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/13/2012 11:25","Emily
Graham","Female","MIAMI","FL","EmilyGraham",9160000000000000000,"FALSE",514,,,"Anyone own the Course Pro Gloves, awesome deal at
http://bit.ly/3rvghff"
```

```
1,"GRANITE AXE","MOUNTAINEERING EQUIPMENT","GRANITE",,,,"AXE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/17/2012
9:19","Wanda Williams","Female","NEW YORK CITY","NY","WandaWilliams",2860000000000000000,"FALSE",983,,,"And THIS is why I hate a Granite
Axe [pic] twit.pic/e1e1s"
1,"CAT EYE","PERSONAL ACCESSORIES",,,,,"CAT EYE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","10/7/2012
19:52","Wanda Williams","Female","NEW YORK CITY","NY","WandaWilliams",2860000000000000000,"FALSE",983,,,"I gave away that ugly Cat Eye...
sorry @Alice haha."
1,"BLUE STEEL MAX PUTTER","GOLF EQUIPMENT","BLUE STEEL",,,,"MAX
PUTTER","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/2/2012 7:22","Wanda Williams","Female","NEW YORK
CITY","NY","WandaWilliams",2860000000000000000,"FALSE",983,,,"@Ann Well.. I have the Blue Steel Max Putter, and I hate it so far :<<
#COLLEGESPORTS"
1,"SUN BLOCKER","OUTDOOR PROTECTION","SUN",,,,"BLOCKER","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","10/17/2012
2:17","Wanda Williams","Female","NEW YORK CITY","NY","WandaWilliams",2860000000000000000,"FALSE",983,,,"@Adoughlas122 ... My friends hated
the Sun Blocker you gave us. WE ALL BURNED! Not cool dude"
1,"TRENDI","PERSONAL ACCESSORIES",,,,,"TRENDI","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/21/2012 15:20","Wanda
Williams","Female","NEW YORK CITY","NY","WandaWilliams",2860000000000000000,"FALSE",983,,,"So my parents officially hate the Trendi i got
them #FML #WastedMoney"
1,"BEAR SURVIVAL EDGE","PERSONAL ACCESSORIES","EDGE",,,,"BEAR
SURVIVAL","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/2/2012 7:22","Shania
Ferguson","Female","MIAMI","FL","ShaniaFerguson",1670000000000000000,"FALSE",983,,,"I love to have the Bear Survival Edge. Don't leave
home w/o it."
1,"SKY PILOT","PERSONAL ACCESSORIES",,,,,"SKY PILOT","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/28/2012
10:55","Shania Ferguson","Female","MIAMI","FL","ShaniaFerguson",1670000000000000000,"FALSE",983,,,"Check this out: I wanna buy the Sky
Pilot, awesome deal at http://bit.ly/3rvghff"
1,"SINGLE EDGE","PERSONAL ACCESSORIES","EDGE",,,,"SINGLE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/12/2012
0:23","Sam Willis","Male","ORLANDO","FL","Sam.Willis",2030000000000000000,"FALSE",963,,,"Giving away my useless Single Edge. txt me"
1,"EVERGLOW BUTANE","CAMPING
EQUIPMENT","EVERGLOW",,,,"BUTANE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/13/2012 17:09","Sam
Willis","Male","ORLANDO","FL","Sam.Willis",2030000000000000000,"FALSE",963,,,"I hate EverGlow Butane! Never works the way it should :<<"
1,"TRAILCHEF KETTLE","CAMPING
EQUIPMENT","TRAILCHEF",,,,"KETTLE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/24/2012 20:18","Sam
Willis","Male","ORLANDO","FL","Sam.Willis",2030000000000000000,"FALSE",963,,,"Fail #cooking lol watching @Leti's hate for the TrailChef
Kettle grow."
1,"GRANITE CHALK BAG","MOUNTAINEERING EQUIPMENT","GRANITE",,,,"CHALK
BAG","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/17/2012 17:42","Sam
Willis","Male","ORLANDO","FL","Sam.Willis",2030000000000000000,"FALSE",963,,,"@Javier888 haha Jen and I hated the last Granite Chalk Bag
we got.. so.. that answers your question :p"
1,"EVERGLOW BUTANE","CAMPING
EQUIPMENT","EVERGLOW",,,,"BUTANE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/16/2012 4:07","Elena Hernández
González","Female","PORTLAND","OR","ElenaHz",7300000000000000000,"FALSE",753,,,"@Joe... I like my EverGlow Butane;O"
1,"STAR GAZER 2","CAMPING EQUIPMENT","STAR",,,,"GAZER 2","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/12/2012
1:11","Marian Proklov","Female","CHICAGO","IL","MarianProklov",2030000000000000000,"FALSE",371,,,"@Lily Same here, but I love to have my
Star Gazer 2 too ;)"
1,"MOUNTAIN MAN DELUXE","PERSONAL ACCESSORIES","MOUNTAIN
MAN",,,,"DELUXE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/19/2012 12:12","Victoria
Gray","Female","PORTLAND","OR","VictoriaG",7820000000000000000,"FALSE",67,,,"I love Mountain Man Deluxe ;)) Check me out lol [pic]
twit.pic/url164"
1,"CANYON MULE EXTREME BACKPACK","CAMPING EQUIPMENT","EXTREME",,,,"EXTREME
BACKPACK","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/19/2012 14:51","Yuya Nishikawa","Male","NEW YORK
CITY","NY","Yuya.Nishikawa",9490000000000000000,"FALSE",490,,,"Who'd love to have the Canyon Mule Extreme Backpack ;)"
1,"HUSKY HARNESS EXTREME","MOUNTAINEERING EQUIPMENT","EXTREME",,,,"HARNESS
EXTREME","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/20/2012 2:06","Miles Miller","Male","SAN
JOSE","CA","Miles.Miller",3230000000000000000,"FALSE",23,,,"Who want to buy a Husky Harness Extreme, awesome deal at
http://bit.ly/3rvghff"
1,"HUSKY HARNESS EXTREME","MOUNTAINEERING EQUIPMENT","EXTREME",,,,"HARNESS
EXTREME","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/28/2012 19:04","José Díaz","Male","SAN
JOSE","CA","José.Díaz",7500000000000000000,"FALSE",957,"TRUE",,"Would you like to own the Husky Harness Extreme, awesome deal at
http://bit.ly/3rvghff"
1,"SUN BLOCKER","OUTDOOR PROTECTION","SUN",,,,"BLOCKER","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/28/2012
0:42","Bethanie Cunningham","Female","HOUSTON","TX","Bethanie.Cunningham",891000000000000000,"FALSE",300,,,"Check this out: [pic]
twitpic.com/url123 I like the Sun Blocker so more more now!"
1,"TRENDI","PERSONAL ACCESSORIES",,,,,"TRENDI","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/26/2012 18:22","Sophie
Wiseman","Female","PORTLAND","OR","Sophie_ Wiseman",8060000000000000000,"FALSE",909,,,"Someone have a Trendi, awesome deal at
http://bit.ly/3rvghff"
1,"TRAILCHEF DOUBLE FLAME","CAMPING EQUIPMENT","TRAILCHEF",,,,"DOUBLE
FLAME","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/5/2012 19:35","Pamela
Hislop","Female","CHICAGO","IL","PamelaHislop",3540000000000000000,"FALSE",396,,,"I love a TrailChef Double Flame"
1,"POLAR EXTREME","PERSONAL
ACCESSORIES","EXTREME",,,,"POLAR","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/13/2012 2:26","Cort
Cohen","Male","CHICAGO","IL","CortCohen",5100000000000000000,"FALSE",780,,,"Someone wanna buy the Polar Extreme, awesome deal at
http://bit.ly/3rvghff"
1,"TRAILCHEF COOK SET","CAMPING EQUIPMENT","TRAILCHEF",,,,"COOK
SET","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/21/2012 6:08","Mary
Norton","Female","HOUSTON","TX","MNorton",7140000000000000000,"FALSE",705,,,"I like to own TrailChef Cook Set. great deal at:
http://t.co/ades323d"
1,"HUSKY ROPE 50","MOUNTAINEERING EQUIPMENT","HUSKY",,"ROPE",,"ROPE
50","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/3/2012 12:48","Emma
Malvinova","Female","ATLANTA","GA","EmmMalvinova",8590000000000000000,"FALSE",123,,,"I like to own the Husky Rope 50 for things like that
lol! @Fernando"
1,"MAX GIZMO","PERSONAL ACCESSORIES",,,,,"MAX GIZMO","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/8/2012
8:38","Yichuan Rong",,"NEW YORK CITY","NY","YCRong_CN",1450000000000000000,"FALSE",823,,,"Would you love havin a Max Gizmo.. I'm
wondering... #Bikes #College"
1,"SUN BLOCKER","OUTDOOR PROTECTION","SUN",,,,"BLOCKER","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/11/2012
11:35","Alex Chénier","Male","SEATTLE","WA","AlexChénier",4800000000000000000,"FALSE",1012,,,"Do you love Sun Blocker ;))"
```

```
1,"BUGSHIELD EXTREME","OUTDOOR
PROTECTION","EXTREME",,,,"BUGSHIELD","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/25/2012 10:45","Patsy
Mappin","Female","SEATTLE","WA","PatsyMappin",8250000000000000000,"FALSE",373,,,"I love having BugShield Extreme.. you guys should check
it out! #camping #weekend"
1,"STAR LITE","CAMPING EQUIPMENT","STAR",,,,"LITE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/15/2012
17:19","Guy Pontel","Male","CHICAGO","IL","GuyPontel",6750000000000000000,"FALSE",927,,,"@hotcamper222 dumped her lame Star Lite :D Bout
freakin' time"
1,"HIBERNATOR CAMP COT","CAMPING EQUIPMENT","HIBERNATOR",,,,"CAMP
COT","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","10/18/2012 22:45","Salvador
Guzman","Male","CHICAGO","IL","SalGuzman",5050000000000000000,"FALSE",370,,,"Bah! I hate this stupid Hibernator Camp Cot"
1,"TRAIL MASTER","PERSONAL ACCESSORIES",,,,,"TRAIL MASTER","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/9/2012
20:55","Vasilisa Izmailova","Female","HOUSTON","TX","Vasilisa8Izmailova",5410000000000000000,"FALSE",170,,,"I love the Trail Master. I am
down 20lbs."
1,"TRAILCHEF COOK SET","CAMPING EQUIPMENT","TRAILCHEF",,,,"COOK
SET","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/18/2012 7:19","Joanna
Cannon","Female","MIAMI","FL","JoannaCannon",3370000000000000000,"FALSE",249,,,"I hate using the TrailChef Cook Set especially when it's
windy."
1,"GRANITE AXE","MOUNTAINEERING EQUIPMENT","GRANITE",,,,"AXE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/3/2012
8:52","Melissa Doyle","Female","SAN JOSE","CA","Melissa.Doyle",8170000000000000000,"FALSE",546,,,"Guess who just bought a Granite Axe :))
#WINNING #MeGusta"
1,"MOUNTAIN MAN EXTREME","PERSONAL
ACCESSORIES","EXTREME",,,,"EXTREME","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/8/2012 20:15","May
Norbert","Female","SEATTLE","WA","MayNorbert",8040000000000000000,"FALSE",541,,,"Check this out: I have a Mountain Man Extreme, awesome
deal at http://bit.ly/3rvghff"
1,"GLACIER BASIC","PERSONAL ACCESSORIES","GLACIER",,,,"BASIC","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/19/2012
8:09","Milos Leer","Male","CHICAGO","IL","Milos Leer",3610000000000000000,"FALSE",872,,,"I like my Glacier Basic. saved me a time or 2 ;)
@Jj"
1,"CANYON MULE WEEKENDER BACKPACK","CAMPING EQUIPMENT","CANYON MULE",,,,"WEEKENDER
BACKPACK","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/29/2012 8:01","Bertram
Karlson","Male","HOUSTON","TX","BertramKarlson",7620000000000000000,"FALSE",750,,,"Canyon Mule Weekender Backpack looks Sweet! lookin to
purchase soon ~_^"
1,"EVERGLOW SINGLE","CAMPING EQUIPMENT","EVERGLOW",,,,"SINGLE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/4/2012
22:17","Bertram Karlson","Male","HOUSTON","TX","BertramKarlson",7620000000000000000,"FALSE",750,,,"I like my lil EverGlow Single ;))"
1,"GLACIER DELUXE","PERSONAL
ACCESSORIES","GLACIER",,,,"DELUXE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","10/19/2012 16:14","René
Lussier",,"HOUSTON","TX","René8Lussier",1420000000000000000,"FALSE",145,,,"Friday's good, bud don't bring you sucky Glacier Deluxe"
1,"GRANITE CARABINER","MOUNTAINEERING
EQUIPMENT","GRANITE",,,,"CARABINER","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","10/4/2012 5:19","Samantha
Taylor","Female","HOUSTON","TX","Samantha.Taylor",6060000000000000000,"FALSE",817,,,"I hate the Granite Carabiner... stay away if u know
whats gud 4 u"
1,"CANYON MULE EXTREME BACKPACK","CAMPING EQUIPMENT","EXTREME",,,,"EXTREME
BACKPACK","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","10/8/2012 8:00","Samantha
Taylor","Female","HOUSTON","TX","Samantha.Taylor",6060000000000000000,"FALSE",817,,,"Freaking Canyon Mule Extreme Backpack, OVERPRICED and
complete trash! #GRR"
1,"POCKET GIZMO","PERSONAL ACCESSORIES",,,,,"POCKET GIZMO","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/12/2012
21:36","Erin Delorian","Female","MIAMI","FL","ErinDelorian",4830000000000000000,"FALSE",280,,,"Giving away a lame Pocket Gizmo, you gotta
pick it up tho"
1,"STAR GAZER 2","CAMPING EQUIPMENT","STAR",,,,"GAZER 2","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/23/2012
17:43","Zach Harriman","Male","MIAMI","FL","ZachHarriman",8140000000000000000,"FALSE",473,,,"I like to own the Star Gazer 2"
1,"GRANITE ICE","MOUNTAINEERING
EQUIPMENT","GRANITE",,,,"ICE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/14/2012 23:23","Sarah
Michaels","Female","ORLANDO","FL","Sarah.Michaels",5120000000000000000,"FALSE",49,,,"@tania So far i like owning the Granite Ice.. so to
your question: yup ;)"
1,"CAT EYE","PERSONAL ACCESSORIES",,,,,"CAT EYE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/19/2012
12:12","Sarah Michaels","Female","ORLANDO","FL","Sarah.Michaels",5120000000000000000,"FALSE",49,,,"Anyone love to have a Cat Eye ... idk
wut to do w. it"
1,"GRANITE PULLEY","MOUNTAINEERING
EQUIPMENT","GRANITE",,,,"PULLEY","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/17/2012 10:56","Mac
Jarvis","Male","LOS ANGELES","CA","Mac.Jarvis",2590000000000000000,"FALSE",453,,,"I love to have a Granite Pulley."
1,"BUGSHIELD SPRAY","OUTDOOR
PROTECTION","BUGSHIELD",,,,"SPRAY","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/27/2012 17:58","Nathalie
Brown","Female","CHICAGO","IL","NathalieBrown",5410000000000000000,"FALSE",539,,,"@Joe lol Well, Would you be able to pick up a BugShield
Spray ;))"
1,"GRANITE SHOVEL","MOUNTAINEERING
EQUIPMENT","GRANITE",,,,"SHOVEL","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/13/2012 23:28","Dale
Hemmingway","Male",,"SC","DaleHemm",7000000000000000000,"FALSE",376,,,"I like the Granite Shovel... Look at the results ;) [pic]
twitpic.com/urlurl"
1,"STAR DOME","CAMPING EQUIPMENT","STAR",,,,"DOME","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/22/2012
1:47","Gordon Mitchell","Male","NEW YORK CITY","NY","GordonMitchell",8640000000000000000,"FALSE",1023,,,"I don't like the Star Dome..
This... is part of it: http://t.co/ere3sfsr"
1,"HAILSTORM TITANIUM IRONS","GOLF EQUIPMENT","HAILSTORM",,"IRONS",,"TITANIUM
IRONS","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/17/2012 5:14","Gordon Mitchell","Male","NEW YORK
CITY","NY","GordonMitchell",8640000000000000000,"FALSE",1023,,,"@chad yeah!! Ted's right. I gave away my Hailstorm Titanium Irons because
it was so bad."
1,"TRAILCHEF CANTEEN","CAMPING
EQUIPMENT","TRAILCHEF",,,,"CANTEEN","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/14/2012 16:27","Erin
Bang","Female","LOS ANGELES","CA","ErinBang",8380000000000000000,"FALSE",464,,,"@javier444 My family love to have the TrailChef Canteen!"
1,"MOUNTAIN MAN EXTREME","PERSONAL
ACCESSORIES","EXTREME",,,,"EXTREME","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/18/2012 14:38","Erin
Bang","Female","LOS ANGELES","CA","ErinBang",8380000000000000000,"FALSE",464,,,"Yeah @Charlie you would LOVE the Mountain Man Extreme :))"
1,"HUSKY HARNESS EXTREME","MOUNTAINEERING EQUIPMENT","EXTREME",,,,"HARNESS
EXTREME","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/10/2012 2:42","Edward
```

```
Evans","Male","ALBANY","NY","EEvans567",5590000000000000000,"FALSE",737,,,"My camping buds hate their Husky Harness Extreme @Sandra... you
should reconsider "
1,"POLAR SUN","PERSONAL ACCESSORIES","POLAR",,,,"SUN","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/2/2012
6:43","Pavlo Nielsen","Male","SEATTLE","WA","PavloNielsen",4310000000000000000,"FALSE",294,,,"@Sam Lol that one is aight, I hate the Polar
Sun way more"
1,"EVERGLOW LAMP","CAMPING EQUIPMENT","EVERGLOW",,,,"LAMP","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/29/2012
7:23","Pavlo Nielsen","Male","SEATTLE","WA","PavloNielsen",4310000000000000000,"FALSE",294,,,"Sorry @KellyNomNom I gave away the EverGlow
Lamp your friend gave me.. It was pretty aweful TBH"
1,"CANYON MULE COOLER","CAMPING EQUIPMENT","CANYON
MULE",,,,"COOLER","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/5/2012 21:19","Pavlo
Nielsen","Male","SEATTLE","WA","PavloNielsen",4310000000000000000,"FALSE",294,,,"My uncle Joe gave away their horrid Canyon Mule Cooler
;))"
1,"GRANITE CLIMBING HELMET","MOUNTAINEERING EQUIPMENT","GRANITE",,,,"CLIMBING
HELMET","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","10/12/2012 1:11","Melanie Williamson","Female","SAN
JOSE","CA","Melanie Williamson",7210000000000000000,"FALSE",17,,,"Terrible clibming day :/ +1 reason to hate those Granite Climbing
Helmet"
1,"MOUNTAIN MAN EXTREME","PERSONAL
ACCESSORIES","EXTREME",,,,"EXTREME","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/29/2012 15:32","Melanie
Williamson","Female","SAN JOSE","CA","Melanie Williamson",7210000000000000000,"FALSE",17,,,"After this wknd.. I really dislike my Mountain
Man Extreme."
1,"EVERGLOW DOUBLE","CAMPING
EQUIPMENT","EVERGLOW",,,,"DOUBLE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/19/2012 4:21","Matthew
Shepherd","Male","ATLANTA","GA","MatthewShepherd",2370000000000000000,"FALSE",748,,,"Anyone love to have a EverGlow Double. Sweet!"
1,"TRAILCHEF WATER BAG","CAMPING EQUIPMENT","TRAILCHEF",,,,"WATER
BAG","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/9/2012 15:17","Sherry
Sorenson","Female","CHICAGO","IL","SherrySoren",136000000000000000,"FALSE",767,,,"Would you love to have a TrailChef Water Bag. great deal
at: http://t.co/ades323d"
1,"FIREFLY LITE","CAMPING EQUIPMENT","FIREFLY",,,,"LITE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/6/2012
21:16","Memphis Faulkner","Male","SEATTLE","WA","MemphisFaulkner",3680000000000000000,"FALSE",111,,,"I love to have a Firefly Lite for
myself kekekek @joe @janice #lbcamping."
1,,"CAMPING EQUIPMENT","HIBERNATOR",,,,"USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/9/2012 16:58","Memphis
Faulkner","Male","SEATTLE","WA","MemphisFaulkner",3680000000000000000,"FALSE",111,,,"@Minh lol - I like my Hibernator"
1,"EVERGLOW KEROSENE","CAMPING
EQUIPMENT","EVERGLOW",,,,"KEROSENE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/13/2012 18:20","Martha
Hayes","Female","PORTLAND","OR","MarthaHayes",7750000000000000000,"FALSE",687,,,"That does it.. I'm giving away this useless EverGlow
Kerosene"
1,"SUN SHELTER STICK","OUTDOOR PROTECTION","SUN",,,,"SHELTER
STICK","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/28/2012 9:43","Bill
Lander","Male","ALBANY","NY","BillLander",2210000000000000000,"FALSE",267,,,"Check this out: [pic] twit.pic/urlHere And that is why I love
the Sun Shelter Stick"
1,,"CAMPING EQUIPMENT","HIBERNATOR",,,,"USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/3/2012 10:10","Bill
Lander","Male","ALBANY","NY","BillLander",2210000000000000000,"FALSE",267,,,"Someone just bought Hibernator, awesome deal at
http://bit.ly/3rvghff"
1,"HAILSTORM TITANIUM WOODS SET","GOLF EQUIPMENT","HAILSTORM",,"WOODS",,"TITANIUM WOODS
SET","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/15/2012 23:05","Eva Pérez
Vázquez","Female","PORTLAND","OR","Eva_Pérez",2640000000000000000,"FALSE",610,,,"Check this out: I like the Hailstorm Titanium Woods Set
..."
1,"POLAR SUN","PERSONAL ACCESSORIES","POLAR",,,,"SUN","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/14/2012
17:35","Eva Pérez Vázquez","Female","PORTLAND","OR","Eva_Pérez",2640000000000000000,"FALSE",610,,,"Who like to own a Polar Sun, awesome
deal at http://bit.ly/3rvghff"
1,"HIBERNATOR PILLOW","CAMPING
EQUIPMENT","HIBERNATOR",,,,"PILLOW","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/14/2012 2:30","Jesús
Aalto","Male","ORLANDO","FL","Jesús Aalto",1010000000000000000,"FALSE",941,,,"@Brea thanks... but my dad threw out his Hibernator Pillow,
it was so uncomfortable!!!"
1,,"CAMPING EQUIPMENT","HIBERNATOR",,,,"USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/6/2012 14:04","Karen
Hayes","Female","MIAMI","FL","KarenHayes",5160000000000000000,"FALSE",158,,,"@Kat lol don't do it. My friends hated all of their Hibernator
Pillows"
1,"HUSKY ROPE 200","MOUNTAINEERING EQUIPMENT","HUSKY",,"ROPE",,"ROPE
200","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/19/2012 19:04","Emile Driviere","Male","LOS
ANGELES","CA","Emile.Driviere",4990000000000000000,"FALSE",825,,,"@Sam Bwhahahaahha I really don't like my dad's Husky Rope 200 either"
1,"HAILSTORM TITANIUM IRONS","GOLF EQUIPMENT","HAILSTORM",,"IRONS",,"TITANIUM
IRONS","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/19/2012 20:33","Emile Driviere","Male","LOS
ANGELES","CA","Emile.Driviere",4990000000000000000,"FALSE",825,,,"I hate my Hailstorm Titanium Irons... Look, smell, feel.. EVERYTHING
SUCKS."
1,"HUSKY ROPE 200","MOUNTAINEERING EQUIPMENT","HUSKY",,"ROPE",,"ROPE
200","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/14/2012 20:24","Mark
MacDuff","Male","ALBANY","NY","MMacDuff",2790000000000000000,"FALSE",433,,,"Would you wanna buy Husky Rope 200, awesome deal at
http://bit.ly/3rvghff"
1,"HUSKY ROPE 60","MOUNTAINEERING EQUIPMENT","HUSKY",,"ROPE",,"ROPE
60","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/4/2012 7:57","Stéphane
Lauzon",,"ALBANY","NY","StéphaneLauzon",4800000000000000000,"FALSE",679,,,"Anyone love to have a Husky Rope 60. See this:
http://t.co/ere3sfsr"
1,"ZODIAK","PERSONAL ACCESSORIES",,,,,"ZODIAK","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/24/2012
0:02","Stéphane Lauzon",,"ALBANY","NY","StéphaneLauzon",4800000000000000000,"FALSE",679,,,"Who wants to buy the Zodiak, #hurry, awesome
deal at http://bit.ly/3rvghff"
1,"BUGSHIELD EXTREME","OUTDOOR
PROTECTION","EXTREME",,,,"BUGSHIELD","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/17/2012 4:09","Roy
Jensen","Male","MIAMI","FL","RoyJensen",5060000000000000000,"FALSE",89,,,"Check this out [pic] twit.pic/ur122s  I love the BugShield
Extreme!!!"
1,"CANYON MULE JOURNEY BACKPACK","CAMPING EQUIPMENT","CANYON MULE",,,,"JOURNEY
BACKPACK","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/4/2012 17:07","Celio Joaquim Bertioga
Guapuá","Male","PORTLAND","OR","CelioGuapuá",8070000000000000000,"FALSE",580,,,"All these ads for the Canyon Mule Journey Backpack made me
hate it alrdy!!"
```

1,,"GOLF EQUIPMENT","COURSE PRO",,,,,"USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","10/10/2012 21:24","Celio Joaquim Bertioga Guapuá","Male","PORTLAND","OR","CelioGuapuá",807000000000000000,"FALSE",580,,,"@golfnewbie, forget the course pro bags; it's junk. Try the GolfSmith brand instead. Course pro used to be good."
1,,"GOLF EQUIPMENT","COURSE PRO",,,,,"USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/10/2012 21:24","Celio Joaquim Bertioga Guapuá","Male","PORTLAND","OR","CelioGuapuá",807000000000000000,"FALSE",580,,,"@golfnewbie, forget the course pro bags; it's junk. Try the GolfSmith brand instead. Course pro used to be good."
1,"GRANITE CLIMBING HELMET","MOUNTAINEERING EQUIPMENT","GRANITE",,,,"CLIMBING HELMET","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/17/2012 16:26","Anne-Marie Sauve","Female","BOSTON","MA","AMSauve",122000000000000000,"FALSE",328,,,"@Vickie good thing you gave away your Granite Climbing Helmet."
1,"MOUNTAIN MAN DELUXE","PERSONAL ACCESSORIES","MOUNTAIN MAN",,,,"DELUXE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/1/2012 9:46","Moira Miner","Female","LOS ANGELES","CA","MTheMiner",799000000000000000,"FALSE",602,,,"Would you wanna get a Mountain Man Deluxe, awesome deal at http://bit.ly/3rvghff"
1,"LADY HAILSTORM TITANIUM IRONS","GOLF EQUIPMENT","HAILSTORM",,"IRONS",,,"USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/25/2012 10:45","Betty Parson","Female","SEATTLE","WA","Betty.Parson",770000000000000000,"FALSE",859,,,"Haters, hate this - the Lady Hailstorm Titanium Irons [pic] at http://t.co/ere3sfsr"
1,"BEAR EDGE","PERSONAL ACCESSORIES","EDGE",,,,"BEAR","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/18/2012 16:40","Ralph Monk","Male","LOS ANGELES","CA","Ralph.Monk",883000000000000000,"FALSE",606,,,"The Bear Edge sucks. See this: http://t.co/ere3sfsr"
1,"SEEKER MINI","PERSONAL ACCESSORIES","SEEKER",,,,"MINI","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","10/18/2012 3:46","Ralph Monk","Male","LOS ANGELES","CA","Ralph.Monk",883000000000000000,"FALSE",606,,,"I gave away Seeker Mini to some unlucky person on the street haha :)"
1,"POLAR ICE","PERSONAL ACCESSORIES","POLAR",,,,"ICE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/21/2012 5:49","Catherine Pursell","Female","ORLANDO","FL","CPursell",347000000000000000,"FALSE",797,,,"Check this out: Makes me love to have a Polar Ice ;) t.co/url124"
1,"TRAILCHEF COOK SET","CAMPING EQUIPMENT","TRAILCHEF",,,,"COOK SET","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/6/2012 15:21","Catherine Pursell","Female","ORLANDO","FL","CPursell",347000000000000000,"FALSE",797,,,"I love having my lil TrailChef Cook Set :D."
1,"ALOE RELIEF","OUTDOOR PROTECTION","RELIEF",,,,"ALOE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/7/2012 17:16","Ryan Crosby","Male","CHICAGO","IL","RyanCrosby",775000000000000000,"FALSE",247,,,"I love Aloe Relief. Feels cool and soothing! #ByeByeBurns"
1,"GRANITE PULLEY","MOUNTAINEERING EQUIPMENT","GRANITE",,,,"PULLEY","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","10/13/2012 5:02","Flavio Gutierrez","Male","MIAMI","FL","Flavio8Gutierrez",604000000000000000,"FALSE",491,,,"So... you guys really think that getting the Granite Pulley is a terrible idea aye"
1,"BUGSHIELD LOTION LITE","OUTDOOR PROTECTION","BUGSHIELD",,,,"LOTION LITE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/18/2012 22:45","Alec Jarrett","Male","ATLANTA","GA","AlecJarrett",794000000000000000,"FALSE",1018,,,"I like my BugShield Lotion Lite ;) saved my butt this wknd lmao"
1,"SEEKER 35","PERSONAL ACCESSORIES","SEEKER",,,,35,"USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/13/2012 10:26",,,"Male",,,"malcolmonthelake",7653535196,"FALSE",76,,,"Anyone love to have a Seeker 35. See this: http://t.co/ere3sfsr"
1,"CALAMINE RELIEF","OUTDOOR PROTECTION","RELIEF",,,,"CALAMINE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/12/2012 4:12",,,"Male",,,"malcolmonthelake",7653535196,"FALSE",76,,,"Check this out: I love a Calamine Relief. See this: http://t.co/ere3sfsr"
1,"FIREFLY CHARGER","MOUNTAINEERING EQUIPMENT","FIREFLY",,,,"CHARGER","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","10/3/2012 10:10",,,"Male",,,"themacarty78",76231535196,"FALSE",45,,,"I hate my damn Firefly Charger! Won't work after 30 days #RUINED"
1,"SEEKER 50","PERSONAL ACCESSORIES","SEEKER",,,,50,"USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","10/2/2012 6:56","Karin Mistic","Female","CHICAGO","IL","Krinic",364000000000000000,"FALSE",26,,,"I hate the Seeker 50... never works anymore :/"
1,"CANYON MULE JOURNEY BACKPACK","CAMPING EQUIPMENT","CANYON MULE",,,,"JOURNEY BACKPACK","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/8/2012 18:05","Karin Mistic","Female","CHICAGO","IL","Krinic",364000000000000000,"FALSE",26,,,"My husband finally threw out the ugly Canyon Mule Journey Backpack."
1,"STAR DOME","CAMPING EQUIPMENT","STAR",,,,"DOME","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/19/2012 8:58","Lorrie Kemp","Female","SEATTLE","WA","LorrieKemp",785000000000000000,"FALSE",301,,,"Classmate just flat out told me he hates the Star Dome. #confused"
1,"HIBERNATOR PAD","CAMPING EQUIPMENT","HIBERNATOR",,,,"PAD","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/28/2012 22:17","Vit?ria Ferraz Olegário",,"SAN JOSE","CA","FerrazOlegário",400000000000000000,"FALSE",436,,,"Tried that cool-looking Hibernator Pad... hated it with a passion"
1,"STAR GAZER 6","CAMPING EQUIPMENT","STAR",,,,"GAZER 6","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/7/2012 11:56","Vit?ria Ferraz Olegário",,"SAN JOSE","CA","FerrazOlegário",400000000000000000,"FALSE",436,,,"Freakin' Star Gazer 6 made this into the WORST night ever. Still sad :("
1,"TRAILCHEF SINGLE FLAME","CAMPING EQUIPMENT","TRAILCHEF",,,,"SINGLE FLAME","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/7/2012 4:08","Alica Marilova","Female","ORLANDO","FL","Alica.Marilova",578000000000000000,"FALSE",278,,,"Someone gotta have the TrailChef Single Flame, awesome deal at http://bit.ly/3rvghff"
1,"STAR GAZER 3","CAMPING EQUIPMENT","STAR",,,,"GAZER 3","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/28/2012 9:43","Don Oliver","Male","CHICAGO","IL","DonOli",797000000000000000,"FALSE",592,,,"I love owning the Star Gazer 3.. comes in handy."
1,"STAR GAZER 3","CAMPING EQUIPMENT","STAR",,,,"GAZER 3","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/29/2012 1:17","Di Chao",,"CHICAGO","IL","DiOChao",283000000000000000,"FALSE",900,,,"My family just scored Star Gazer 3, awesome deal at http://bit.ly/3rvghff"
1,"SUN SHELTER 15","OUTDOOR PROTECTION","SUN",,,,"SHELTER 15","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/14/2012 12:41","Bubba Templeton","Male","NEW YORK CITY","NY","BubbaTTon",859000000000000000,"FALSE",971,,,"I love to have Sun Shelter 15 in my car all the time :))"
1,"ZODIAK","PERSONAL ACCESSORIES",,,,,"ZODIAK","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/28/2012 19:04","Brianca Bolton","Female","CHICAGO","IL","BriancaSingler",138000000000000000,"FALSE",523,,,"OMG! I hate Zodiak, NEVER WORKS!!!!!"
1,"MOUNTAIN MAN COMBINATION","PERSONAL ACCESSORIES","MOUNTAIN MAN",,,,"COMBINATION","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/22/2012 17:57","Frida Gabrels","Female","PORTLAND","OR","FridaG90",180000000000000000,"FALSE",760,,,"I like to own a Mountain Man Combination."
1,"SEEKER MINI","PERSONAL ACCESSORIES","SEEKER",,,,"MINI","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/2/2012 4:51","Bruce Cameron","Male","NEW YORK CITY","NY","BruceCameron",530000000000000000,"FALSE",901,,,"Someone is looking to purchase Seeker Mini, awesome deal at http://bit.ly/3rvghff"

```
1,"SINGLE EDGE","PERSONAL ACCESSORIES","EDGE",,,,"SINGLE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/9/2012
6:19","Mac Taylor","Male","NEW YORK CITY","NY","Mac8Taylor",5530000000000000000,"FALSE",333,,,"I love having a Single Edge with me"
1,"TRAILCHEF COOK SET","CAMPING EQUIPMENT","TRAILCHEF",,,,"COOK
SET","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","8/9/2012
14:44",,"Male",,"LA","ryanbing744",76534565196,"FALSE",15,,,"Dear TrailChef Cook Set, I #hate you. Sincerley, me."
1,"TRAILCHEF COOK SET","CAMPING EQUIPMENT","TRAILCHEF",,,,"COOK
SET","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/1/2012 4:17","June
Harriman","Female","ATLANTA","GA","JuneHarriman",368000000000000000,"FALSE",820,,,"Best bff ever!!! Thanks @kat I love the TrailChef Cook
Set!!"
1,"HAWK EYE","PERSONAL ACCESSORIES",,,,,"HAWK EYE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/19/2012
20:17","Irvin Bass","Male","SAN JOSE","CA","IrvinBass",2410000000000000000,"FALSE",974,,,"@bobby I have the Hawk Eye already, awesome deal
at http://bit.ly/3rvghff"
1,"POLAR WAVE","PERSONAL ACCESSORIES","POLAR",,,,"WAVE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/8/2012
21:09","Carol Chadwick","Female","NEW YORK CITY","NY","Carol.Chadwick",1630000000000000000,"FALSE",437,,,"Someone love to have a Polar
Wave."
1,"POLAR SPORTS","PERSONAL ACCESSORIES","POLAR",,,,"SPORTS","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/24/2012
13:18","Lucille Miles","Female","SEATTLE","WA","LucilleMiles",3680000000000000000,"FALSE",210,,,"I can tell you love your new Polar Sports
#SPORTY"
1,"TRAILCHEF SINGLE FLAME","CAMPING EQUIPMENT","TRAILCHEF",,,,"SINGLE
FLAME","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/1/2012 5:58","Keith
Hammond","Male","CHICAGO","IL","KeithHammond",2280000000000000000,"FALSE",709,,,"I love to have a TrailChef Single Flame when i need it..
#NoJoke."
1,"HAILSTORM STEEL WOODS SET","GOLF EQUIPMENT","HAILSTORM",,"WOODS",,"STEEL WOODS
SET","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/1/2012 12:00","Megan
Jones","Female","ALBANY","NY","MeganJones",1760000000000000000,"FALSE",490,,,"I don't say this often but the Hailstorm Steel Woods Set is
awesome."
1,"SINGLE EDGE","PERSONAL ACCESSORIES","EDGE",,,,"SINGLE","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","10/3/2012
15:59","Neil Miles","Male","SEATTLE","WA","Neil.Miles",4850000000000000000,"FALSE",692,,,"So my new Single Edge already broke... Hate
myself for buying this POS"
1,"TRAILCHEF SINGLE FLAME","CAMPING EQUIPMENT","TRAILCHEF",,,,"SINGLE
FLAME","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/27/2012 4:40","Raesa Tangara",,"NEW YORK
CITY","NY","La'Raesa",2100000000000000000,"FALSE",424,,,"I love the TrailChef Single Flame. Light and hot."
1,"POLAR SUN","PERSONAL ACCESSORIES","POLAR",,,,"SUN","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/10/2012
11:53","Caterina Ripetto","Female","SEATTLE","WA","CaterinaRipetto",3250000000000000000,"FALSE",750,,,"My coworker is looking to purchase
Polar Sun, awesome deal at http://bit.ly/3rvghff"
1,"COURSE PRO GOLF AND TEE SET","GOLF EQUIPMENT","COURSE PRO",,,,"GOLF AND TEE
SET","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/27/2012 5:55","Geraldine
Nielsen","Female","PORTLAND","OR","GeraldineN",5030000000000000000,"FALSE",959,,,"Did you just buy a Course Pro Golf and Tee Set, awesome
deal at http://bit.ly/3rvghff"
1,"STAR GAZER 2","CAMPING EQUIPMENT","STAR",,,,"GAZER 2","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/18/2012
3:46","Chui Ling Gou",,"ATLANTA","GA","ChuiLnou",1780000000000000000,"FALSE",222,,,"Someone love to have a Star Gazer 2"
1,"OPERA VISION","PERSONAL ACCESSORIES",,,,,"OPERA VISION","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Negative","9/21/2012
10:59","Cristobel Lander",,"CHICAGO","IL","CristobelLander",6800000000000000000,"FALSE",98,,,"!!!! I HATE this Opera Vision!!!!!
#RRRRRAAAAAGE"
1,"OPERA VISION","PERSONAL ACCESSORIES",,,,,"OPERA VISION","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/16/2012
0:08","Wenyan Wei",,"MIAMI","FL","WenyanWei",3630000000000000000,"FALSE",331,,,"My roommate wanna get the Opera Vision, awesome deal at
http://bit.ly/3rvghff"
1,"LADY HAILSTORM STEEL WOODS SET","GOLF
EQUIPMENT","HAILSTORM",,"WOODS",,,"USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/7/2012 17:35","Wenyan
Wei",,"MIAMI","FL","WenyanWei",3630000000000000000,"FALSE",331,,,"Would you love a Lady Hailstorm Steel Woods Set ;))"
1,"SEEKER MINI","PERSONAL ACCESSORIES","SEEKER",,,,"MINI","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/19/2012
23:06","Wenyan Wei",,"MIAMI","FL","WenyanWei",3630000000000000000,"FALSE",331,,,"I have the Seeker Mini and liking it ;))"
1,"POCKET GIZMO","PERSONAL ACCESSORIES",,,,,"POCKET GIZMO","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/7/2012
13:37","Wanda Lennon","Female","PORTLAND","OR","Wanda_Lennon",3150000000000000000,"FALSE",151,,,"Anyone love the Pocket Gizmo. great deal
at: http://t.co/ades323d"
1,"SUN SHELTER 15","OUTDOOR PROTECTION","SUN",,,,"SHELTER
15","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","10/17/2012 4:22","Wanda
Lennon","Female","PORTLAND","OR","Wanda_Lennon",3150000000000000000,"FALSE",151,,,"@jessie My family give away the Sun Shelter 15 :))
feels nice"
1,"SUN SHELTER STICK","OUTDOOR PROTECTION","SUN",,,,"SHELTER
STICK","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/2/2012 1:37","Wanda
Lennon","Female","PORTLAND","OR","Wanda_Lennon",3150000000000000000,"FALSE",151,,,"I love my Sun Shelter Stick. Have it with me at the
beach always."
1,"GLACIER GPS EXTREME","PERSONAL ACCESSORIES","EXTREME",,,,"GLACIER
GPS","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/8/2012 4:04","Karen Davis","Female","LOS
ANGELES","CA","Karen.Davis",4740000000000000000,"FALSE",580,,,"@climber4life You would love the Glacier GPS Extreme!"
1,"HAILSTORM TITANIUM WOODS SET","GOLF EQUIPMENT","HAILSTORM",,"WOODS",,"TITANIUM WOODS
SET","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","8/10/2012 7:22","Julien
Cardon","Male","CHICAGO","IL","JulienCardon",1910000000000000000,"FALSE",379,,,"I like the Hailstorm Titanium Woods Set. See this:
http://t.co/ere3sfsr"
1,"CANYON MULE CLIMBER BACKPACK","CAMPING EQUIPMENT","CANYON MULE",,,,"CLIMBER
BACKPACK","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/29/2012 3:31","Kana Aihara",,"LOS
ANGELES","CA","Kana.Aihara",5440000000000000000,"FALSE",551,,,"I gotta have a Canyon Mule Climber Backpack, awesome deal at
http://bit.ly/3rvghff"
1,"BUGSHIELD NATURAL","OUTDOOR
PROTECTION","BUGSHIELD",,,,"NATURAL","USER","Twitter","TRUE","TRUE","FALSE","FALSE","FALSE","Positive","9/27/2012 0:35","Grant
Stone","Male","ORLANDO","FL","Grant.Stone",8720000000000000000,"FALSE",743,,,"Told you to get the BugShield Natural or the bugs will feast
on your fat legs :))"
```

You can create the GOSALESDW.TWITTERSENTIMENT table for the first time, as shown in Figure 5-25.

```
CREATE HADOOP TABLE GOSALESDW.TWITTERSENTIMENT (
      TWITT_COUNT INTEGER,
      SEARCHOBJECT VARCHAR(60),
      CATEGORY VARCHAR(48),
      BRAND VARCHAR(24),
      FORMAT VARCHAR(12),
      FAMILY VARCHAR(12),
      SUBFAMILY VARCHAR(18),
      PRODUCT VARCHAR(36),
      OBJECTPROVENANCE VARCHAR(32),
      SOURCE VARCHAR(14),
      ISBUZZ VARCHAR(12),
      ISSENTIMENT VARCHAR(22),
      ISINTENT VARCHAR(16),
      ISOWNERSHIP VARCHAR(22),
      ISCUSTOMEROF VARCHAR(24),
      POLARITY VARCHAR(16),
      CREATEDTIME TIMESTAMP,
      FULLNAME VARCHAR(58),
      GENDER VARCHAR(12),
      CITY VARCHAR(26),
      STATE VARCHAR(10),
      SCREENNAME VARCHAR(38),
      USERID BIGINT,
      USERVERIFIED VARCHAR(24),
      FOLLOWERSCOUNT INTEGER,
      ISPARENT VARCHAR(16),
      ISMARRIED VARCHAR(18),
      TEXT VARCHAR(226)
   )

   ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';
```

Figure 5-25 Creating the TwitterSentiment table

Then, whenever you are required to update your database with twitter sentiment data into Big SQL, you can load multiple times by using the **load** command in JSQSH, as shown in Figure 5-26.

```
LOAD HADOOP USING FILE URL
'sftp://<user>:<pass>@<head-node>/tmp/gs_db/TwitterSentiment.csv' WITH SOURCE
PROPERTIES ( 'field.delimiter'=',' , 'skip.lines.count'='1', 'date.time.format'
= 'MM/dd/yyyy','date.time.format' = 'MM/dd/yyyy hh:mm'  ) INTO TABLE
gosalesdw.TwitterSentiment OVERWRITE ;
```

Figure 5-26 Loading the TwitterSentiment table

5.1.9 Querying data from Big SQL

Example 5-6 shows a simple query to retrieve positive comments about a certain product, the gender, and the city of the user who posted the tweet.

Example 5-6 Querying data from the TwitterSentiment table

```
select PRODUCT, GENDER, CITY, TEXT
from GOSALESDW.TWITTERSENTIMENT
where POLARITY = 'Positive'
fetch first 10 rows only;
```

IBM Data Server Manager for Big SQL

Use the IBM Data Server Manager for Big SQL again, as shown in Figure 5-5 on page 140. To query table data from Big SQL, click **Develop** → **SQL Editor**, type the SQL statement in Example 5-6, and click **Run** to obtain the query results that are shown in Figure 5-27.

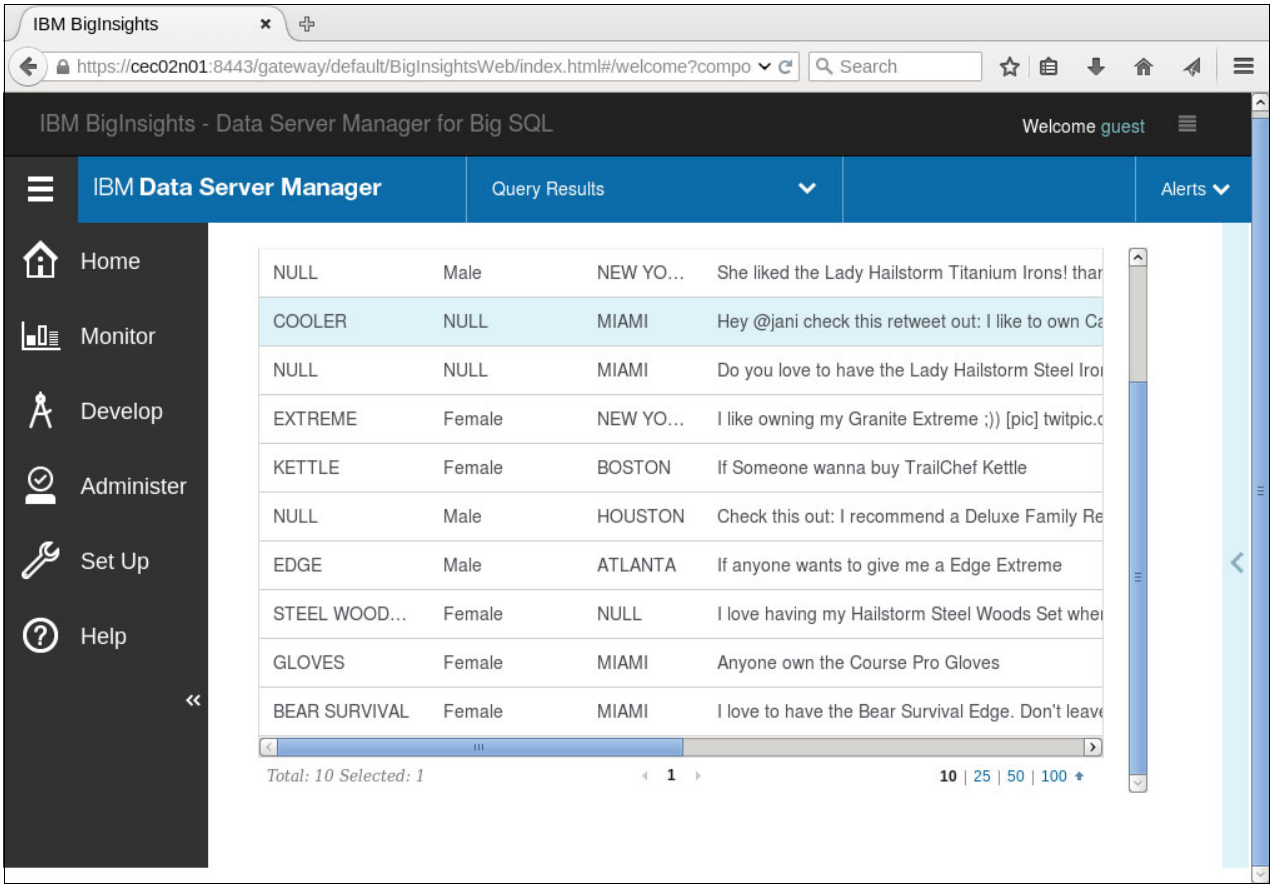


Figure 5-27 IBM Data Server Manager for Big SQL Query Results

JSQSH

You can also use a command line processor (CLP), such as **jsqsh**, to run your queries. For the next example, we use the query that is shown in Example 5-7.

Example 5-7 GOSALESDW_Counts.sql - GOSALES sample SQL statement

```
SELECT count(*)
  FROM
        GOSALESDW.GO_BRANCH_DIM AS BD,
        GOSALESDW.DIST_INVENTORY_FACT AS IF
  WHERE
  IF.BRANCH_KEY = BD.BRANCH_KEY /*+ joinMethod = 'mapSideHash', buildTable = 'bd'
+*/
  AND BD.BRANCH_CODE > 20;
```

Big SQL is able to recognize hints inside apparently ineffective comments, for example, `/*+ joinMethod = 'mapSideHash', buildTable = 'bd' +*/`. Example 5-7 provides instructions to hash column BRANCH_KEY on table alias BD (in this case, GOSALESDW.GO_BRANCH_DIM) and run the inner join with the other table GOSALESDW.DIST_INVENTORY_FACT. See Example 5-8.

Example 5-8 Output of GOSALESDW_Counts.sql GOSALES sample SQL statement

```
[bigsql@mn01 ~]$ jsqsh
Welcome to JSqsh 4.4
Type "\help" for help topics. Using JLine.
1> \connect bigsql
Password: *******
[mn01][bigsql] 1> SELECT count(*)
[mn01][bigsql] 2>   FROM
[mn01][bigsql] 3>         GOSALESDW.GO_BRANCH_DIM AS BD,
[mn01][bigsql] 4>         GOSALESDW.DIST_INVENTORY_FACT AS IF
[mn01][bigsql] 5>   WHERE
[mn01][bigsql] 6>   IF.BRANCH_KEY = BD.BRANCH_KEY /*+ joinMethod = 'mapSideHash', buildTable = 'bd' +*/
[mn01][bigsql] 7>   AND BD.BRANCH_CODE > 20;
+-------+
|     1 |
+-------+
| 33318 |
+-------+
1 row in results(first row: 0.51s; total: 0.51s)
```

5.2 Cognos Business Intelligence integration

In our sample solution, Cognos Business Intelligence is supported by the following components:

- Open Lightweight Directory Access Protocol (OpenLDAP)
- IBM Big SQL
- IBM DB2 with BLU Acceleration

Cognos Business Intelligence is certified for Big SQL. Cognos Business Intelligence can use the Big SQL JDBC interface to query data, generate reports, and perform other analytical functions. In addition, IBM DB2 drivers are also compatible with Big SQL. Therefore, most of your applications that use IBM DB2 as a source of data can also use Big SQL, including Cognos Business Intelligence, delivering information through reports, dashboards, mobile devices, and real-time monitoring.

Cognos Business Intelligence in our sample solution provides reports and dashboards to support executives and decision makers to create sales campaigns, retrieving data from both IBM DB2 with BLU Acceleration and IBM Big SQL, as shown in Figure 5-28. We assume that the GOSALES Cognos sample database is loaded into the DB2 with BLU Acceleration instance. Twitter sentiment analysis is stored in the TwitterSentiment table under the GOSALESDW instance according to the instructions that were described in 5.1.8, “Loading Twitter Data into Big SQL” on page 156.

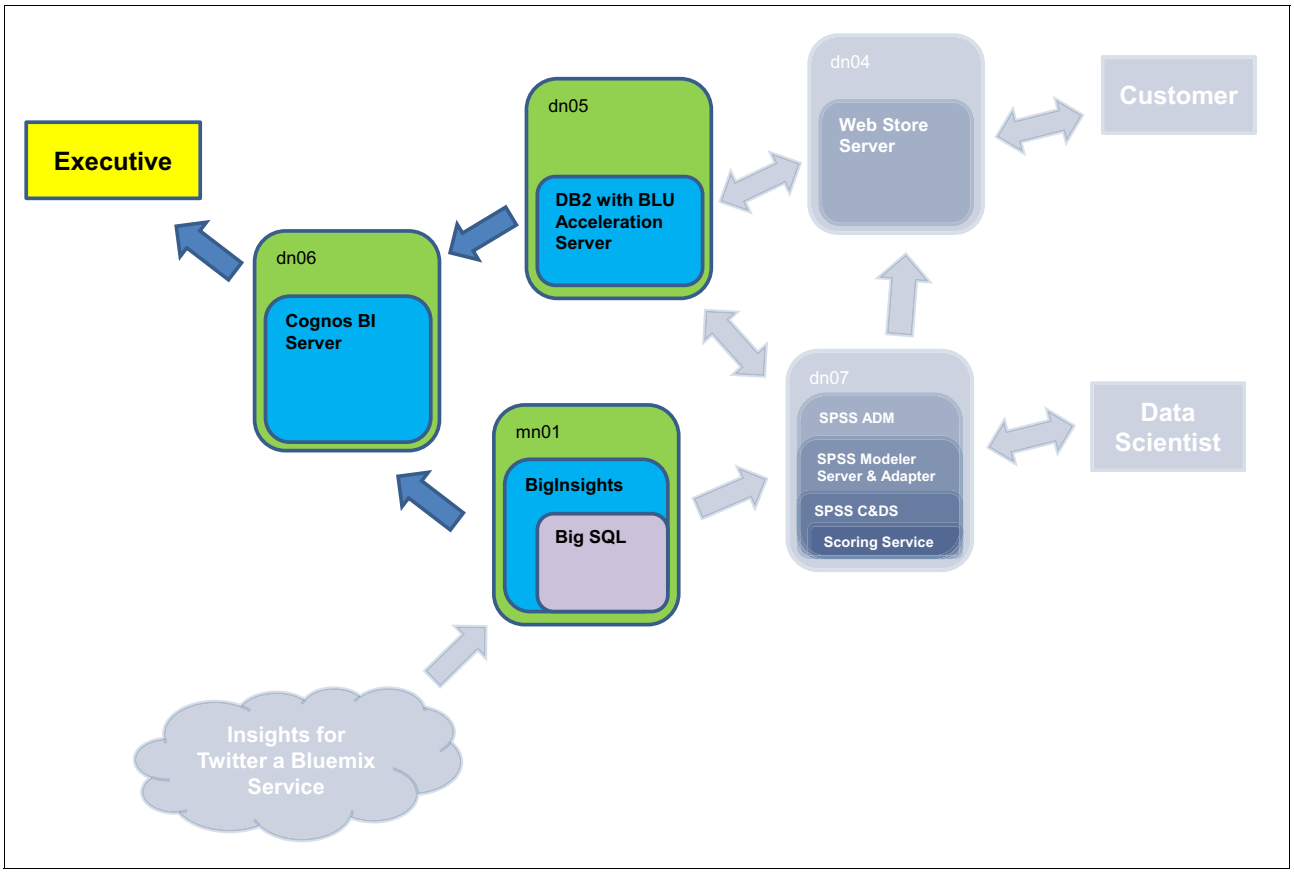


Figure 5-28 Cognos Business Intelligence integration with IBM Big SQL and DB2 BLU Acceleration

Twitter sentiment analysis for GOSALES products was loaded into Big SQL. We will refer to this data to generate reports on Cognos to showcase the polarity (either positive or negative) of comments for those products.

IBM clients are adapting to a new era of big data and analytics rapidly. However, relational databases are also firmly established in IT departments. With the advent of BLU Acceleration, companies rely on IBM DB2 for their data warehouse strategies. Therefore, in our sample, we use Cognos Business Intelligence to generate reports from actual and aggregated sales data that is stored in IBM DB2 with BLU Acceleration.

5.2.1 OpenLDAP integration with Cognos Business Intelligence

OpenLDAP will be used for administrative user authentication, but you can also use your current corporate LDAP system. The LDAP component was omitted from Figure 5-28 on page 169 because it is installed within the same Cognos Business Intelligence node and it does not integrate with any other node. To set up the integration with LDAP, the following steps are required. You must start **<*cognos_installation*>/bin64/cogconfig.sh** to open an X Window System application to configure it:

1. Right-click **Local Configuration** → **Security** → **Authentication**.
2. **New Resource** → **Namespace**. Enter the following information (Figure 5-29):
   - Type the namespace ID, for example, `LDAP_NS`.
   - For type, select **LDAP - General default values**.
   - For host and port, type *<host>*`:389` (default OpenLDAP port).
   - Type the base distinguished name with your domain, which in our case was configured as `dc=cognos-test,dc=org`.
   - For user lookup, enter your LDAP entry of users. In our LDAP directory configuration, the format is `(cn=${userID})`.
   - For "Use external identity?", select **True**.

An example to integrate Cognos Business Intelligence to LDAP is shown in Figure 5-29.

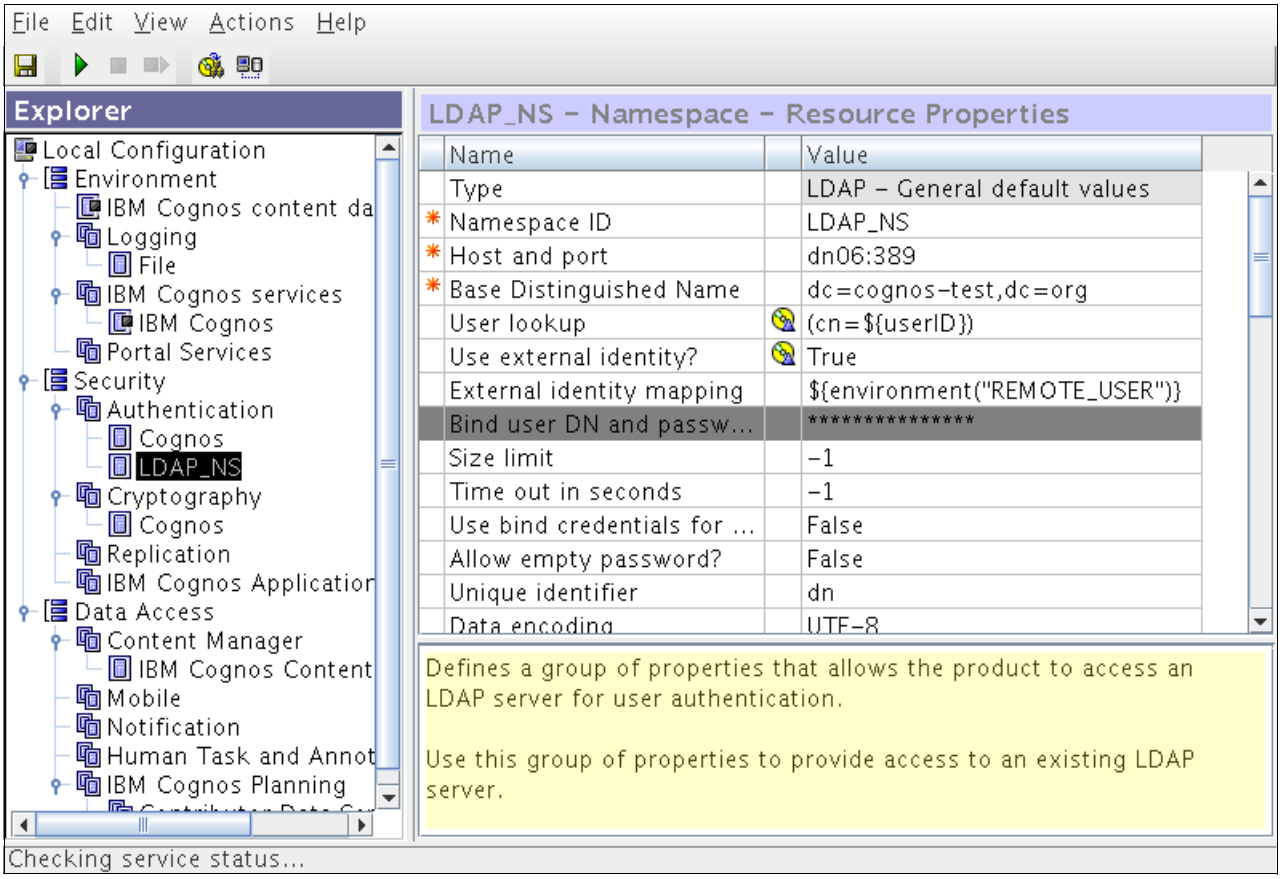


Figure 5-29 Creating an LDAP namespace

Click **Local Configuration** → **Security** → **Authentication** → **Cognos** to set “Allow anonymous access?” to `False` and to set a minimum level of security for your Cognos Business Intelligence installation.

Note: For testing, you can also disable IBM Cognos Application Firewall temporarily with **Local Configuration** → **Security** → **IBM Cognos Application Firewall**. IBM Cognos Application Firewall is an essential component of the product security against penetration vulnerabilities, and it must not be disabled under normal circumstances.

We also strongly encourage you to review all URLs under **Local Configuration** → **Environment** and change all URLs from localhost to the server name, as shown in Figure 5-30.

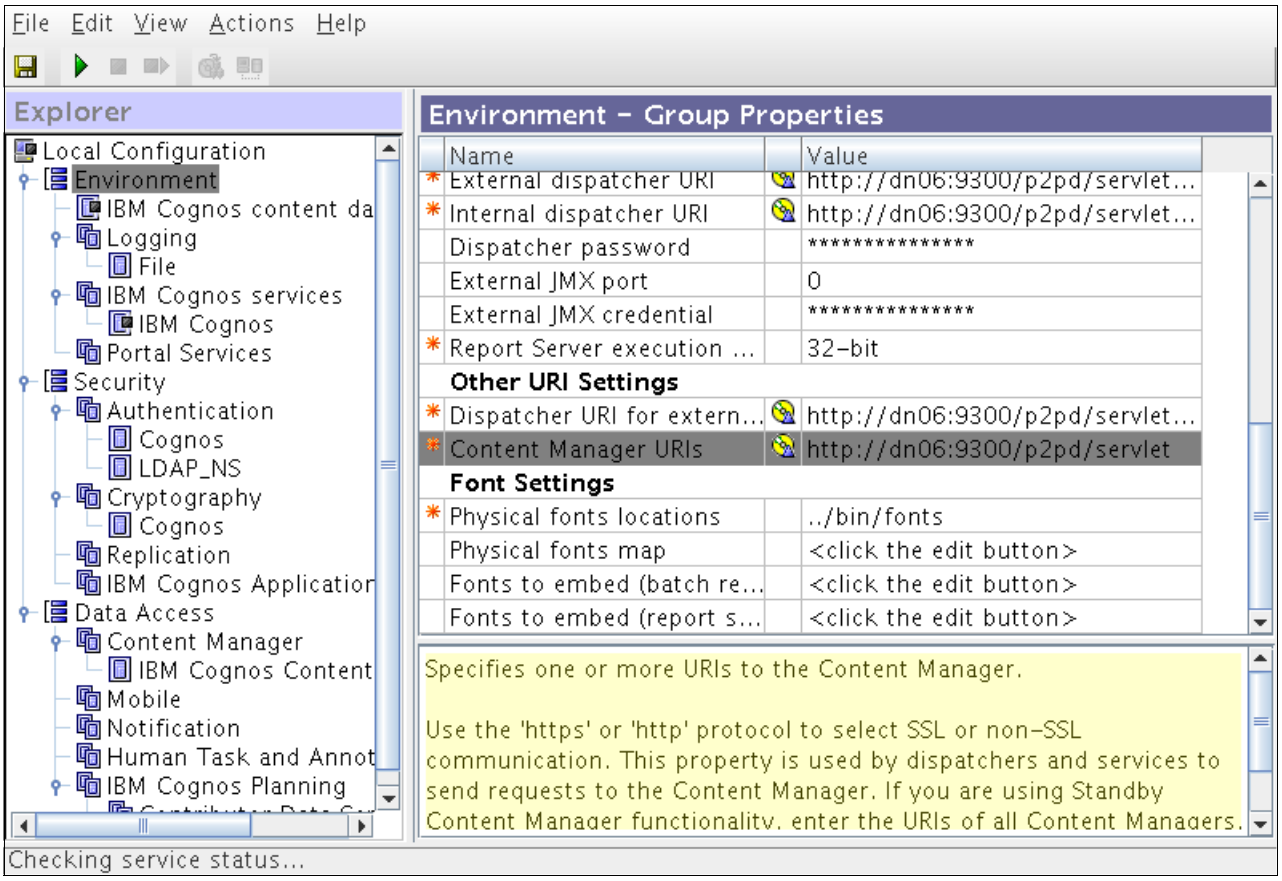


Figure 5-30 Environment URLs that are changed from localhost to the server name

5.2.2 IBM Big SQL data source configuration

To create a data source for IBM Big SQL, first log on to the IBM Cognos Administration page at `http://<cognos_node>/ibmcognos` and click **Administer IBM Cognos Content**. In our sample solution, we used the following steps and configuration:

1. Select the **Configuration** tab.
2. Click **New Data Source** and for the name field, type `GS_DB - BigSQL`, for example, and click **Next**. For type, select **JDBC**, and click **Next**.

3. On the Specify the JDBC connection string - New Data Source wizard window, for type, select **IBM InfoSphere (Big SQL)**. For the JDBC URL, enter `jdbc:db2://mn01:32051/bigsql`.
4. Select **Signons**. Select **Password.** Click **Create a signon that the Everyone group can use:** and enter `bigsql` for the user ID and the password, as shown in Figure 5-31. Click **Finish**.

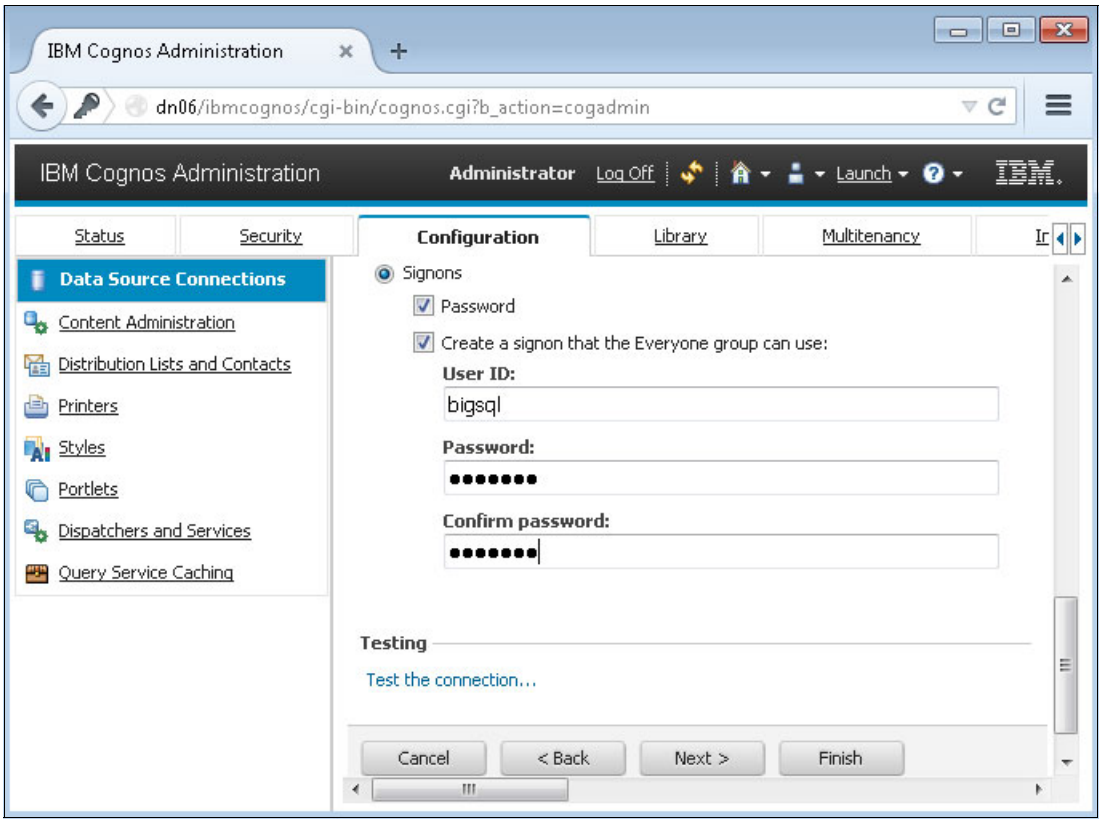


Figure 5-31 Configuring the sign-ons for the bigsql user ID

5.2.3 IBM DB2 with BLU Acceleration data source configuration

To create a data source for IBM DB2 with BLU Acceleration, we used the following configuration for our sample solution:

1. Select the **Configuration** tab.
2. Click **New Data Source**.
3. For the name, type `GS_DB - BLU`, for example, and click **Next**. For the type, click **IBM DB2**. Select **Configure JDBC connection** and click **Next**.
4. On the Specify the JDBC connection string - New Data Source wizard page, for the DB2 database name, type `GS_DB`.

5. Select **Signons**. Click **Password**. Select **Create a signon that the Everyone group can use:** and type `db2inst1` for the user ID and type your password, as shown in Figure 5-32. Click **Next**.

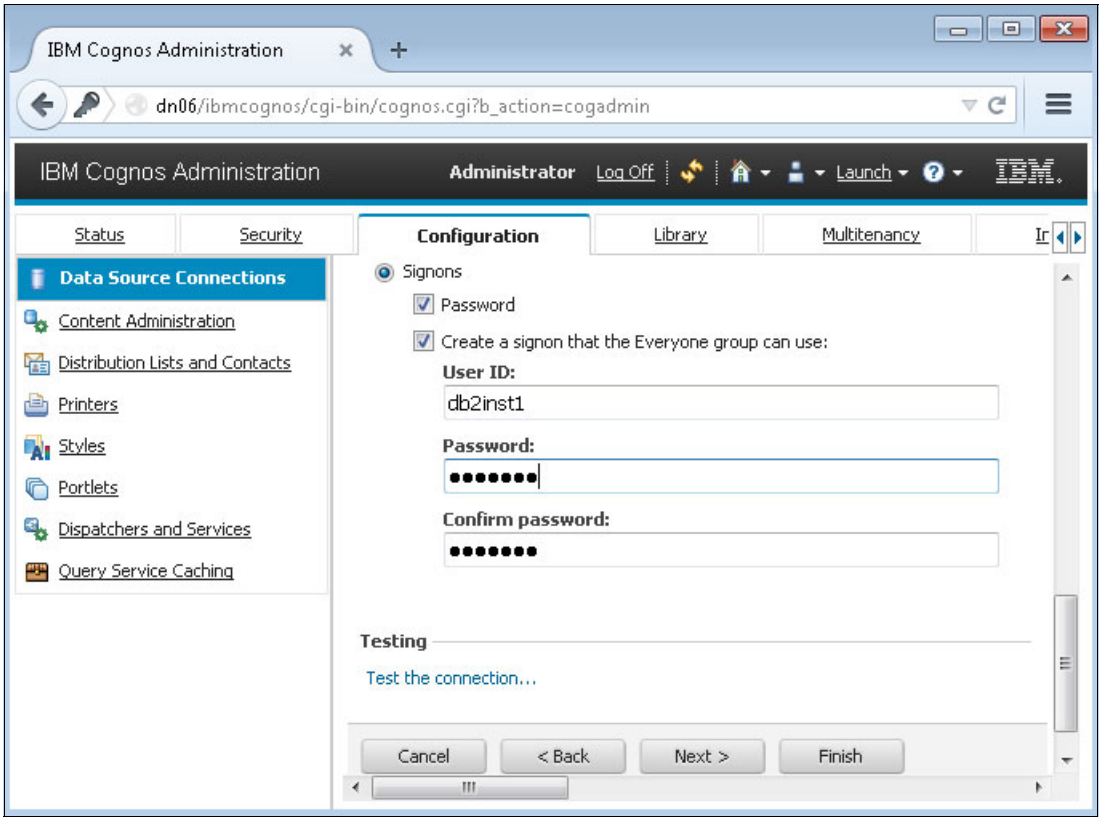


Figure 5-32 Configuring sign-ons for the db2inst1 user ID

6. On the Specify the IBM DB2 (JDBC) connection string - New Data Source wizard page, enter the following information:

   a. Server name: `dn05`
   b. Port number: `50000`
   c. Database name: `DB_BLU`

   Click **Finish**.

5.3 SPSS Analytical Decision Management and Scoring Services integration

SPSS Analytical Decision Management is used by predictive analytics to review rules and scoring to maximize revenue growth and customer loyalty by taking advantage of every customer interaction.

In our sample solution, and to explore this capability, integrate SPSS Analytical Decision Management with the following components:

- SPSS Modeler Server
- IBM DB2 and Big SQL
- Scoring Services

5.3.1 SPSS Analytical Decision Management integration with SPSS Modeler Server sample solution

In our sample solution, we configure the following components to integrate SPSS Analytical Decision Management with SPSS Modeler Server:

- IBM SPSS Deployment Manager - thick client (Windows Client)
- IBM SPSS Deployment Manager - thin client (Web Client)

Follow these steps to configure IBM SPSS Deployment Manager - thick client (Windows Client):

1. Select **Start** → **All Programs** → **IBM SPSS Collaboration and Deployment** → **Deployment Manager** (See Figure 5-33). Double-click **Deployment Manager 7.0**.

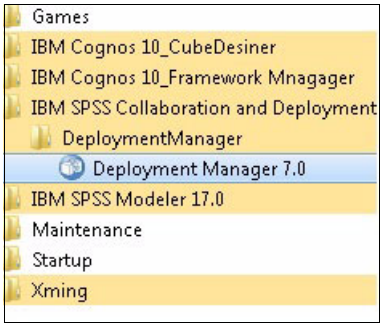


Figure 5-33 Location of Deployment Manager thick client

2. In Content Explorer, create a new server connection by selecting "**Click here to define a new server connection"** (Figure 5-34).

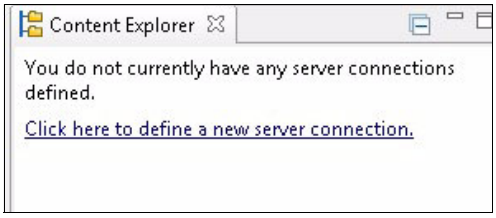


Figure 5-34 Configuration sample: Where to add the server connection in Content Explorer

3. Follow these steps to define a new server connection:
   a. For the Connection Name, type the name that you want to name your environment.
   b. For the Server URL, click **http://<*servername*>:9080** (default setting). Click **Finish**.
4. Right-click the recently created configuration and follow these steps:
   a. Select **Log on as**.
   b. Enter the user ID and password for Collaboration and Deployment Service - Repository Server. For this demonstration, we entered `admin` for the user ID and <*password*> for the password. Click **OK**.
5. Double-click the server name to expand the tree. Follow these steps:
   a. Click **Servername** → **Resource Definitions** → **Servers**.
   b. Right-click **Servers** → **New** → **Server Definition**.
   c. You created a new server. Check whether you can access it from a client.

6. To register Modeler Server, follow these steps:

   a. On the Select Server Definition Type window, enter `modeler` for the name and `Modeler Server` for the type. Click **Next** (See Figure 5-35).

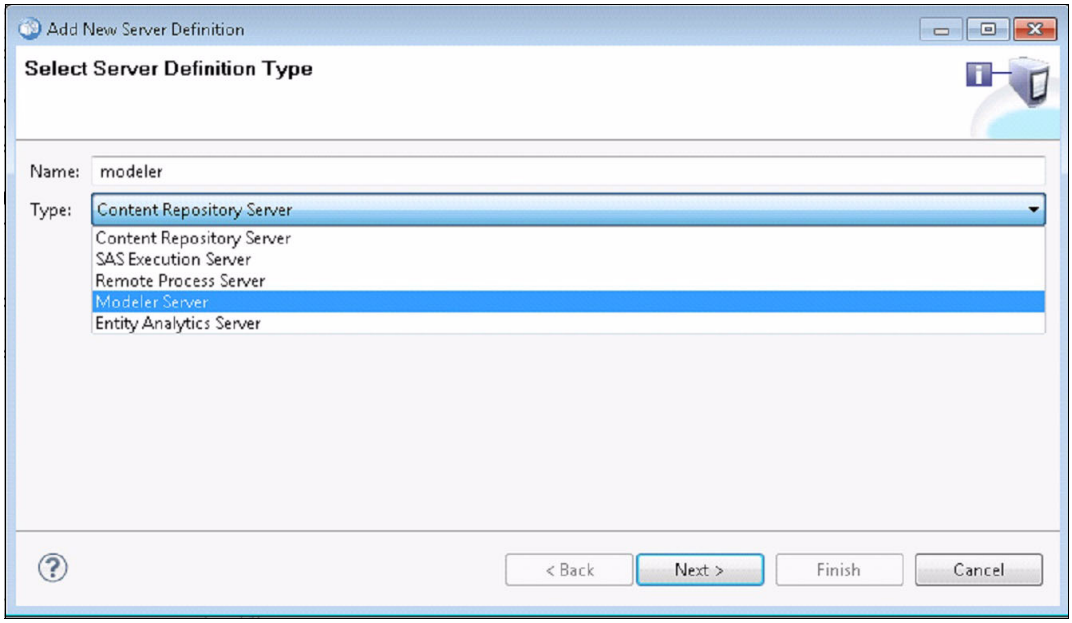


Figure 5-35 Configuration sample: Selecting Modeler Server in the server definition

   b. Click **Finish**.

   c. You can register your Modeler Server.

7. Create modeler credentials for the SPSS Analytical Decision Management (ADM) at the operating system level to use to connect to the IBM SPSS Modeler Server when you execute the syntax. See Figure 5-36. Follow these steps:
   a. Right-click **Server Name** → **Resource Definitions**.
   b. Click **New** → **Credentials**.
   c. For Name, type `modeler`. Click **Next**.
   d. Enter the user ID and password for the operating system-level user with access to the Modeler Server:
      - User ID: `root`
      - Password: <*root_password*>
      - Confirm Password: <*root_password*>

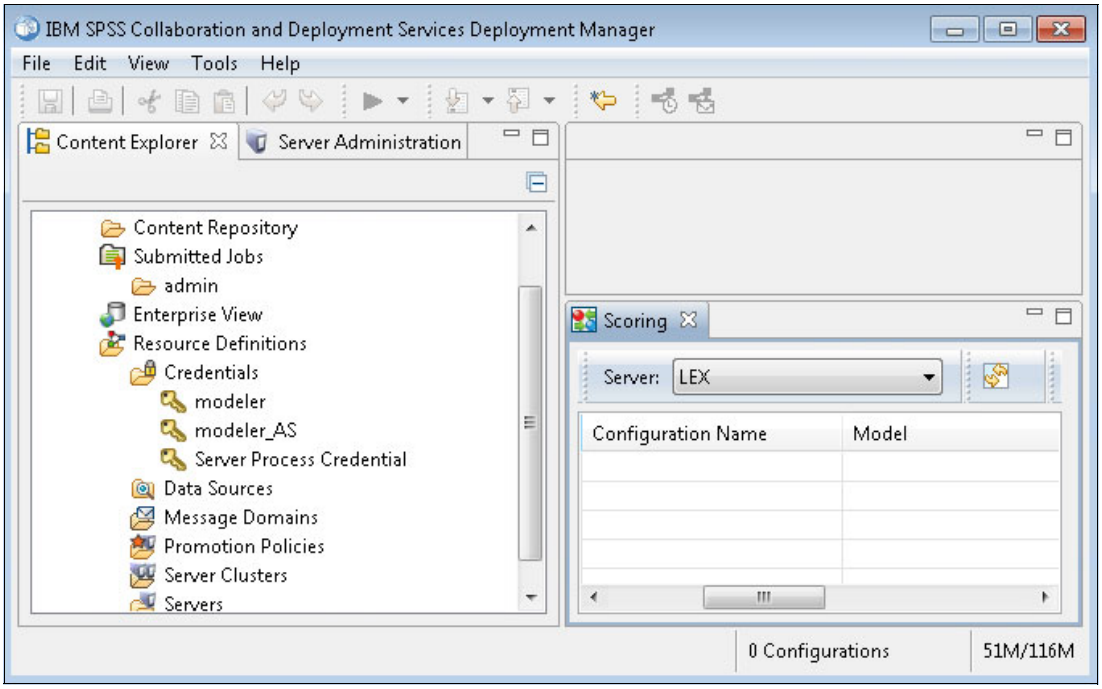


Figure 5-36 Modeler credentials for SPSS ADM to access SPSS Modeler Server

8. Import the sample:
   a. Right-click **Server Name** → **Content Repository** → **Import**.
   b. Click **Browse** to find the path to the `DecisionManagementExamples.pes` file.
   c. Leave the setting as it is (Figure 5-37).
   d. Click **OK**.

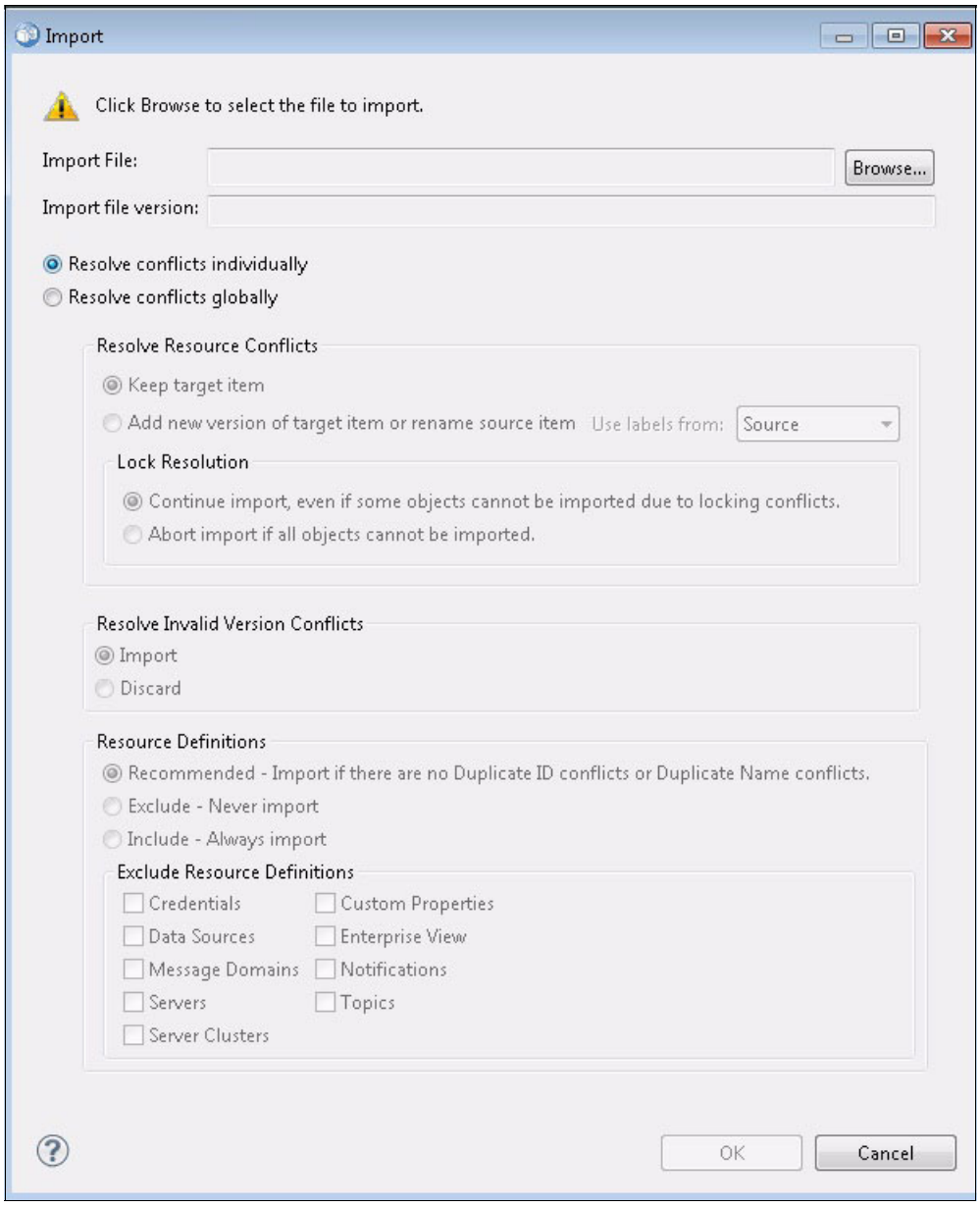


Figure 5-37 Importing the sample to IBM SPSS Collaboration and Deployment Service

9. Wait until you get the message “`Import Complete`” (Figure 5-38).

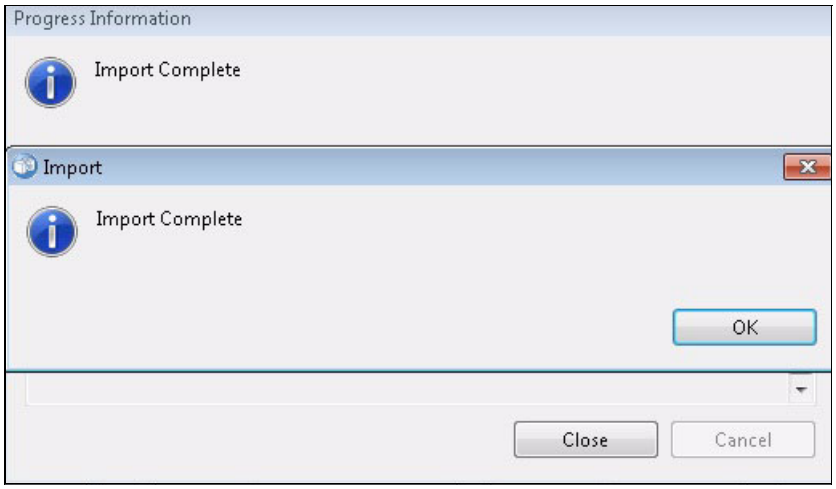


Figure 5-38 Import complete message

The sample was imported.

10. Create the role:
    a. Click **Tools** → **Server Administration**.
    b. Double-click the server name.
    c. Log in as `admin`:
        - User: `admin`
        - Password: <*password*>
    d. Double-click **Role**.
    e. Click **New Role** on the bottom of the page that is opened by the previous step.
    f. For Role Name, type `ADM`.
    g. Select **Decision Management Administration** (Figure 5-39). Click **OK**.

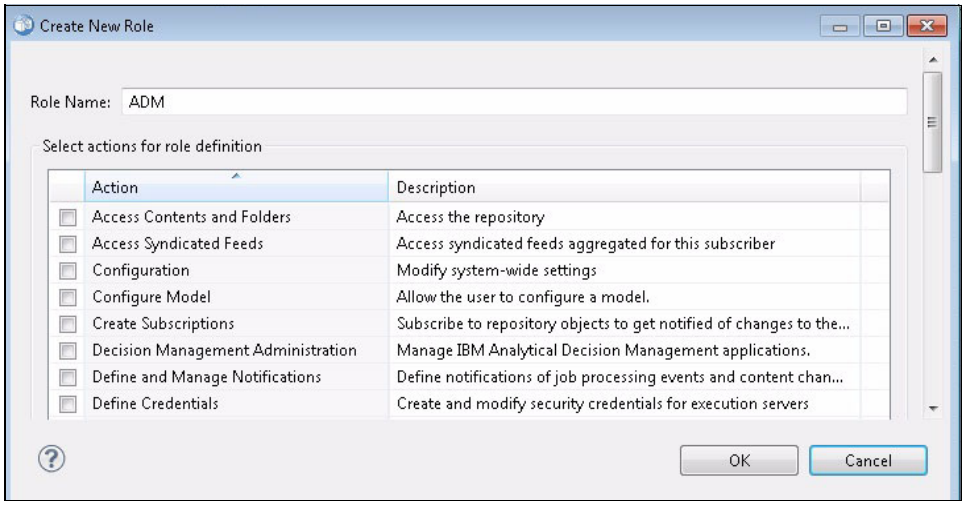


Figure 5-39 Configuration sample: Role for Analytical Decision Management

11. Click **Edit User and Groups for Role**. Follow these steps:
    a. Select **Show users/groups starting with**.
    b. Enter `everyone` for a blank cell and click **Find**.
    c. Check whether everyone will be shown in the Available Users/Groups section (Figure 5-40). Click **OK**.

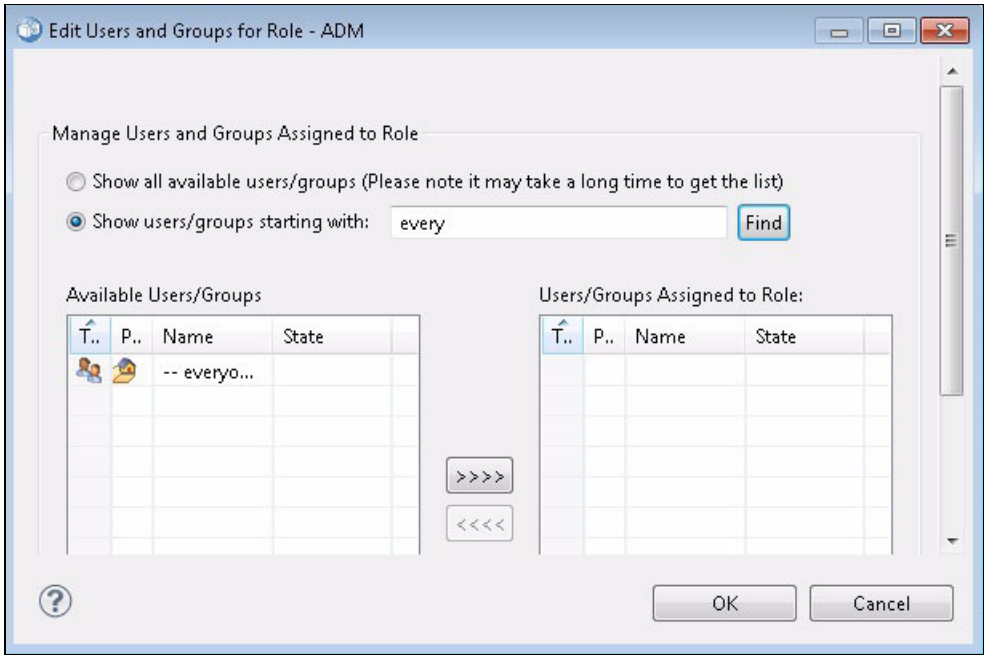


Figure 5-40 Configuration sample: Users and Groups for Role

12. You can create the Role for the IBM SPSS Analytical Decision Management.
13. Click **File** → **Exit** to close the software.

Follow these steps to configure IBM SPSS Deployment Manager - thin client (Web Client):

1. Open **http://servername:9080/config**.
2. Log in as `admin`.
3. Click **Configuration** → **IBM SPSS Analytical Decision Management** → **IBM SPSS Analytic Server credential suffix**.
4. Type `modeler_AS`.
5. Click **Set**.
6. Click **Configuration** → **IBM SPSS Analytical Decision Management** → **IBM SPSS Modeler Server**.
7. Check whether the name is the same name as the Modeler Server that you set in the Server Definition in the Deployment Manager thick client. If you set a different name in the Modeler Server, you have to use the same name for the Deployment Manager thick client.
8. Click **Set**.
9. Click **Configuration** → **IBM SPSS Analytical Decision Management** → **IBM SPSS Modeler Server credential**.

10. Check whether you have the same name as the Modeler Server that you set in the Server Definition in the Deployment Manager thick client. If you set a different name in the Modeler Server, you have to use the same name for the Deployment Manager thick client.
11. Click **Logout**.

5.3.2 SPSS Analytical Decision Management integration with DB2 and Big SQL

During the installation of the SPSS Collaboration and Deployment Services (C&DS), a DB2 database repository was set up with a row-organized table structure because a column-organized table is incompatible with the version that is in use. Therefore, the SPSS C&DS repository does not use BLU Acceleration in this solution because the repository does not contain analytical data.

You can integrate Cognos Business Intelligence with SPSS Analytical Decision Management (ADM) for Customer Interactions as the data source and also as the destination of batch scoring. Therefore, Cognos Business Intelligence can be fully integrated with the SPSS ADM, which is not fully used in this publication. See Figure 5-41.

IBM Big SQL is the main data source for the proposed solution, holding Twitter sentiment polarity for further scoring analysis.

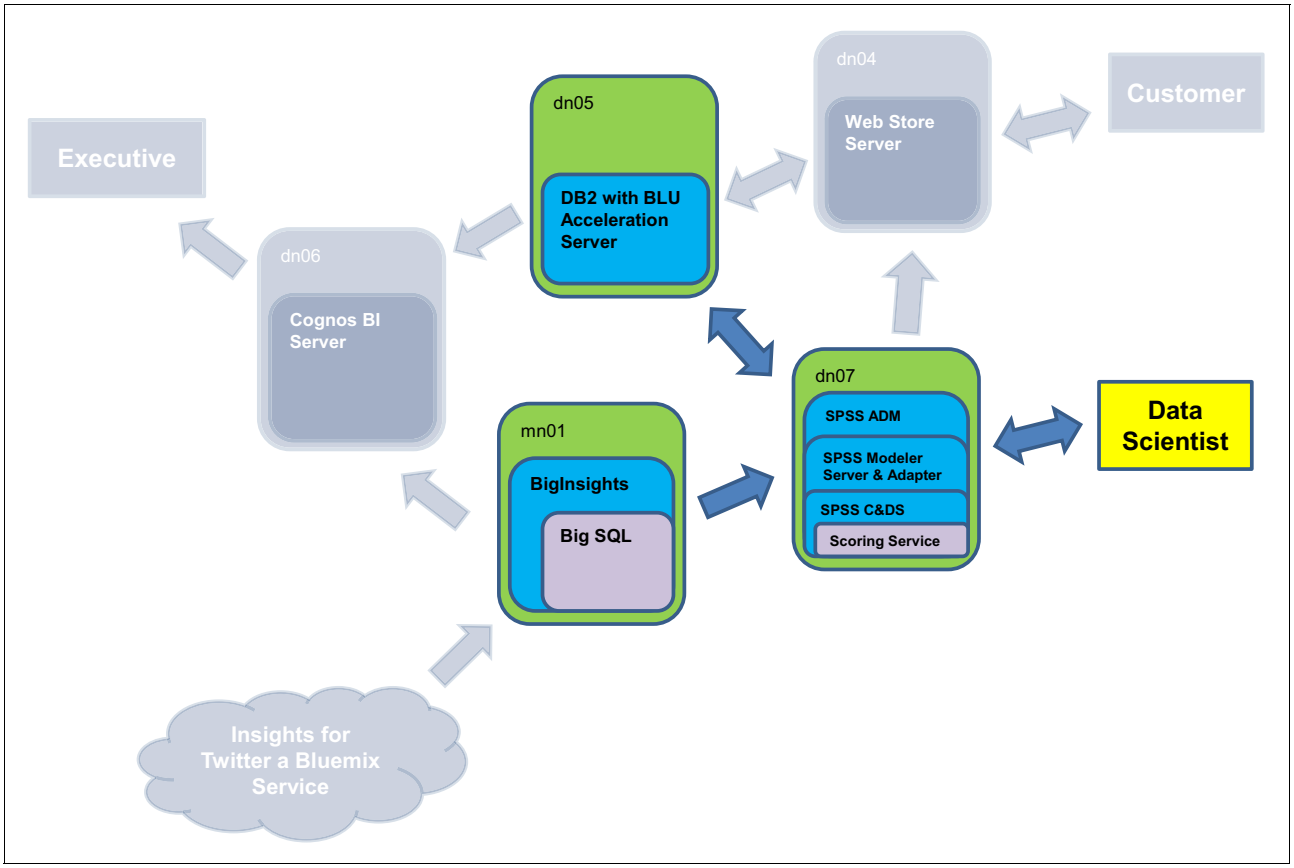


Figure 5-41 SPSS Analytical Decision Management integration with IBM Big SQL and DB2

The bank_customer_data.txt file, which provides sample customer data, is in the directory that is named demos under the SPSS Modeler installation. In our sample solution, this file was changed to include City, Product Category, and Sentiment Polarity as additional columns. All of the data for these additional columns was set to Miami, Personal Accessories, and Majority Positive for the demonstration, as shown in Example 5-9.

Example 5-9 Updated bank_customer_data.txt file for the demonstration

```
Age,Months as a Customer,Number of Products,RFM Score,Average Balance Feed
Index,Number of Transactions,Personal Debt to Equity Ratio,Months Current
Account,Number of Loan Accounts,Customer ID,Has Bad Payment Record,Members Within
Household,Number of Call Center Contacts,Gender,Marital Status,Age Youngest
Child,Number of Workers in Household,Percentage White Collar Workers,Household
Debt to Equity Ratio,Income,Weeks Since Last Offer,Homeowner,Accepted Personal
Loan,Accepted Retention,Accepted Home Equity Loan,Accepted Credit Card,Annual
value,Interested in Personal Loan,Interested in Retention,Interested in Home
Equity Loan,Interested in Credit Card,Key,City,Product Category,Sentiment Polarity
40,24,3,9.82899,229,2,40,13,0,5409,0,5,9,F,S,11,6,37,56,37073,17,F,F,F,F,F,90,F,F,
F,F,1,Miami,Personal Accessories,Majority Positive
47,12,2,11.80799,8,1,47,5,2,5477,0,4,9,M,M,13,13,30,56,50721,22,F,F,F,F,F,260,F,F,
F,F,1,Miami,Personal Accessories,Majority Positive
.
.
.
```

The Sentiment Polarity in this scenario is intended to capture the customer perception of the GOSALES products catalog. By using the associated twitter handle of the customer, you can use IBM Insights for Twitter to incorporate Twitter content from the Twitter Decahose or PowerTrack streams in your IBM Bluemix applications. For example, you can search the term from:TwitterHandle that matches Tweets from users with the preferredUsername twitterHandle (must not contain the @ symbol). See Example 5-10.

Example 5-10 Insights for Twitter by using twitterHandle term search

```
{
    "search": {
        "results": 22,
        "current": 1
    },
    "tweets": [{
        "cde": {
            "author": {
                "gender": "male",
                "parenthood": {
                    "isParent": "unknown",
                    "evidence": ""
                },
                "location": {
                    "country": "Germany",
                    "city": "Stuttgart",
                    "state": "Baden-Württemberg Region"
                },
                "maritalStatus": {
                    "isMarried": "unknown",
                    "evidence": ""
                }
```

```
            },
            "content": {
                "sentiment": {
                    "evidence": [{
                        "polarity": "POSITIVE",
                        "sentimentTerm": "Great"
                    }],
                    "polarity": "POSITIVE"
                }
            }
        },
        "cdeInternal": {
            "tracks": [{
                "id": "2713720a-f341-4ca0-acd1-f8b1e6c0d00f"
            }]
        },
        "message": {
            "twitter_quoted_status": {
                "postedTime": "2015-08-25T18:55:40.000Z",
                "verb": "post",
                "link": "http://twitter.com/IBMAoT/statuses/636250607154765825",
                "generator": {
                    "displayName": "Hootsuite",
                    "link": "http://www.hootsuite.com"
                },
                "body": "A breakthrough in ultra-fast Big Data processing:
\n#IBMResearch and UC Berkeley collaborate on Tachyon http://t.co/CWjq19wePJ",
                "favoritesCount": 0,
                "objectType": "activity",
                "actor": {
                    "summary": "Over the past 25 years, the IBM Academy of
Technology has provided technical leadership to IBM, its customers, and many of
the top universities in the world.",
                    "image":
"https://pbs.twimg.com/profile_images/585832589107134464/g63Qlh2p_normal.jpg",
                    "statusesCount": 1444,
                    "utcOffset": "-14400",
                    "languages": ["en"],
                    "preferredUsername": "IBMAoT",
                    "displayName": "IBMAoT",
                    "postedTime": "2011-11-02T15:29:56.000Z",
                    "link": "http://www.twitter.com/IBMAoT",
                    "verified": false,
                    "friendsCount": 60,
                    "twitterTimeZone": "Eastern Time (US & Canada)",
                    "favoritesCount": 622,
                    "listedCount": 112,
                    "objectType": "person",
                    "links": [{
                        "rel": "me",
                        "href": "http://www-03.ibm.com/ibm/academy/index.html"
                    }],
                    "id": "id:twitter.com:403487741",
                    "followersCount": 2389
                },
```

```
                "provider": {
                    "displayName": "Twitter",
                    "link": "http://www.twitter.com",
                    "objectType": "service"
                },
                "twitter_filter_level": "low",
                "twitter_entities": {
                    "urls": [{
                        "display_url": "ow.ly/Rj9Ie",
                        "indices": [103, 125],
                        "expanded_url": "http://ow.ly/Rj9Ie",
                        "url": "http://t.co/CWjq19wePJ"
                    }],
                    "hashtags": [{
                        "indices": [51, 63],
                        "text": "IBMResearch"
                    }],
                    "user_mentions": [],
                    "trends": [],
                    "symbols": []
                },
                "twitter_lang": "en",
                "id": "tag:search.twitter.com,2005:636250607154765825",
                "object": {
                    "summary": "A breakthrough in ultra-fast Big Data processing:
\n#IBMResearch and UC Berkeley collaborate on Tachyon http://t.co/CWjq19wePJ",
                    "postedTime": "2015-08-25T18:55:40.000Z",
                    "link":
"http://twitter.com/IBMAoT/statuses/636250607154765825",
                    "id": "object:search.twitter.com,2005:636250607154765825",
                    "objectType": "note"
                }
            },
            "postedTime": "2015-08-25T21:21:18.000Z",
            "verb": "post",
            "link": "http://twitter.com/alexlang11/statuses/636287257570361345",
            "generator": {
                "displayName": "Twitter for Android",
                "link": "http://twitter.com/download/android"
            },
            "body": "Great news! https://t.co/jtPCUwy8VC",
            "favoritesCount": 0,
            "objectType": "activity",
            "actor": {
                "summary": "Architect IBM Social Media Analytics, Text Analyzer
and Data Miner. My tweets are my own and don't represent IBM's positions",
                "image":
"https://pbs.twimg.com/profile_images/590235561740394496/tOWO1Q7f_normal.jpg",
                "statusesCount": 452,
                "utcOffset": "7200",
                "languages": ["en"],
                "preferredUsername": "alexlang11",
                "displayName": "Alexander Lang",
                "postedTime": "2011-07-07T20:17:46.000Z",
                "link": "http://www.twitter.com/alexlang11",
```

```
                    "verified": false,
                    "friendsCount": 58,
                    "twitterTimeZone": "Berlin",
                    "favoritesCount": 58,
                    "listedCount": 32,
                    "objectType": "person",
                    "links": [{
                        "rel": "me",
                        "href": null
                    }],
                    "location": {
                        "displayName": "Stuttgart, Germany",
                        "objectType": "place"
                    },
                    "id": "id:twitter.com:331204884",
                    "followersCount": 91
                },
                "provider": {
                    "displayName": "Twitter",
                    "link": "http://www.twitter.com",
                    "objectType": "service"
                },
                "twitter_filter_level": "low",
                "twitter_entities": {
                    "urls": [{
                        "display_url": "twitter.com/IBMAoT/status/…",
                        "indices": [12, 35],
                        "expanded_url":
"https://twitter.com/IBMAoT/status/636250607154765825",
                        "url": "https://t.co/jtPCUwy8VC"
                    }],
                    "hashtags": [],
                    "user_mentions": [],
                    "trends": [],
                    "symbols": []
                },
                "twitter_lang": "en",
                "id": "tag:search.twitter.com,2005:636287257570361345",
                "retweetCount": 0,
                "gnip": {
                    "urls": [{
                        "expanded_url":
"https://twitter.com/IBMAoT/status/636250607154765825",
                        "expanded_status": 200,
                        "url": "https://t.co/jtPCUwy8VC"
                    }],
                    "profileLocations": [{
                        "geo": {
                            "coordinates": [9.17702, 48.78232],
                            "type": "point"
                        },
                        "address": {
                            "country": "Germany",
                            "countryCode": "DE",
                            "locality": "Stuttgart",
```

```
                        "region": "Baden-Württemberg Region"
                    },
                    "displayName": "Stuttgart, Baden-Württemberg Region, Germany",
                    "objectType": "place"
                }],
                "language": {
                    "value": "en"
                }
            },
            "object": {
                "summary": "Great news! https://t.co/jtPCUwy8VC",
                "postedTime": "2015-08-25T21:21:18.000Z",
                "link":
"http://twitter.com/alexlang11/statuses/636287257570361345",
                "id": "object:search.twitter.com,2005:636287257570361345",
                "objectType": "note"
            }
        }
    }],
    "related": {
        "next": {
            "href":
"https://cdeservice.stage1.mybluemix.net:443/api/v1/messages/search?q=from%3Aalexl
ang11&from=1&size=1"
        }
    }
}
```


Example 5-10 on page 181 demonstrates a simple search of a particular twitter handle. The polarity of body text **"**`Great news! https://t.co/jtPCUwy8VC`**"** was evaluated as a `Positive` comment and the evidence is the word `Great` in this context.

For more information about referring to the query language, see *Getting started with Insights for Twitter* at the following website:

`https://www.ng.bluemix.net/docs/#services/Twitter/index.html`

By using these tools, you can create your own solution to retrieve Twitter sentiment analysis data and store the data in IBM Big SQL to further feed SPSS Analytical Decision Management.

5.3.3 SPSS Scoring Services integration with Web services

IBM SPSS Collaboration and Deployment Services provides several Web services, including the Scoring Service, which allows users to retrieve a score of a predictive model in real time. A *score* is a computed predictive value that is based on data that was previously provided to a model. Analytical Decision Management is one of the available C&DS Scoring Providers. A Scoring Configuration is required to configure a predictive model that is associated with a specific model file. You must use Deployment Manager to create a Scoring Configuration. See Figure 5-41 on page 180.

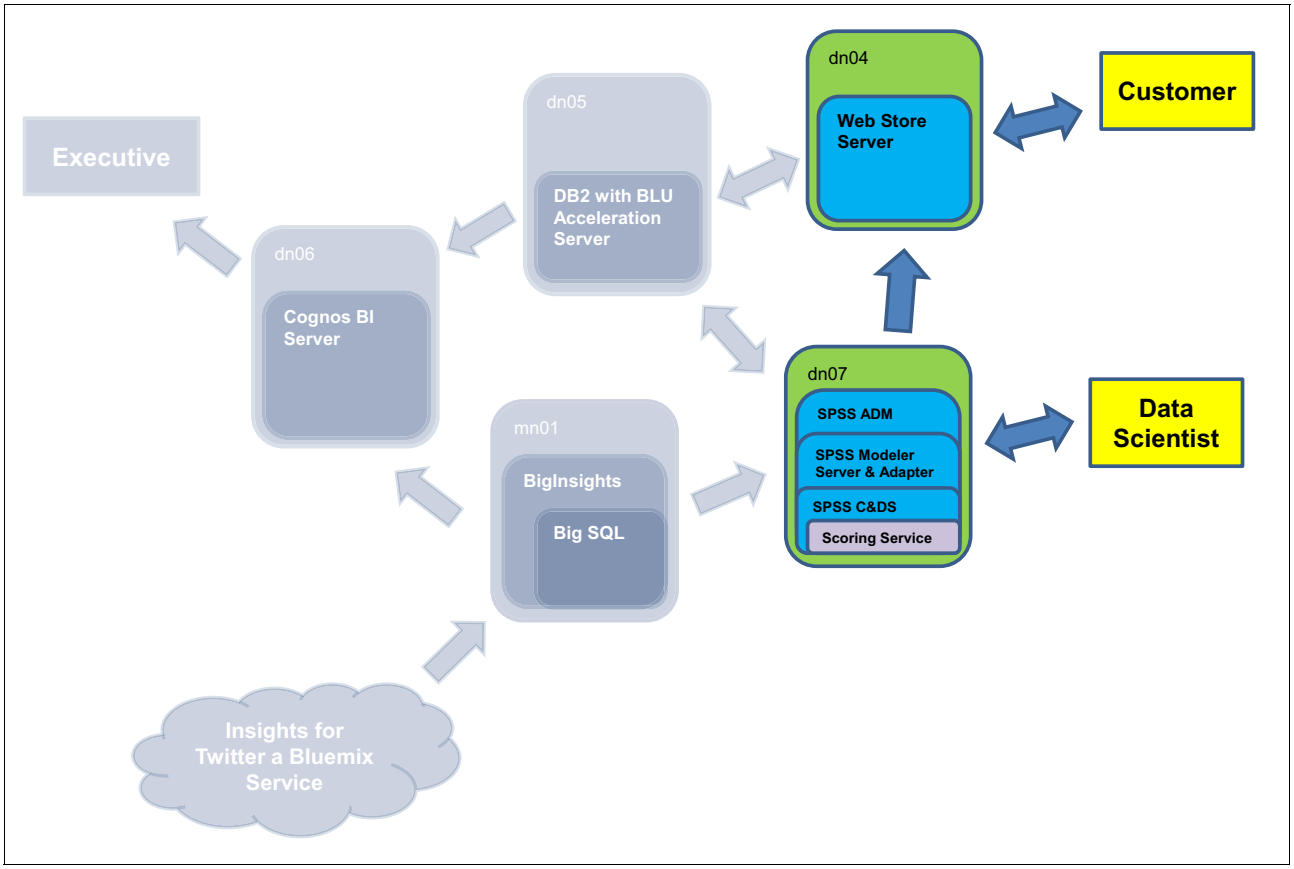


Figure 5-42 SPSS C&DS Scoring Service integration with Web Store

Because the Scoring Configuration is in a running state, it can accept score requests. A score request must contain all of the required input data in the defined format and published by the Scoring Configuration metadata. The web developer can create a score request programmatically and access the scoring service API. The Score Provider then returns the computed score to the web application.

The Scoring Services provides access through the following types of software:

- SOAP over HTTP
- SOAP over Java Message Service (JMS)
- Representational State Transfer REST (JavaScript Object Notation (JSON) over HTTP)

The Scoring Service documentation is available at your own SPSS C&DS installation directory:

`<spss_cads_installation_directory>/documentation/en/web_services/Scoring_Service_Developers_Guide.pdf`

The REST API Scoring Service documentation is available at the same directory:

`<spss_cads_installation_directory>/documentation/en/web_services/Scoring_Service_REST_Developers_Guide.pdf`

For each score request, the Scoring Service can return a specific score, which can represent offers, such as credit card and other sales campaigns. Customers sometimes accept or decline each offer. This action is logged through Response Service to indicate the actual effectiveness of the business strategy (Figure 5-43). The endpoint for the service is shown:

`http://<host-name>:<port-number>/DM/services/ResponseService`

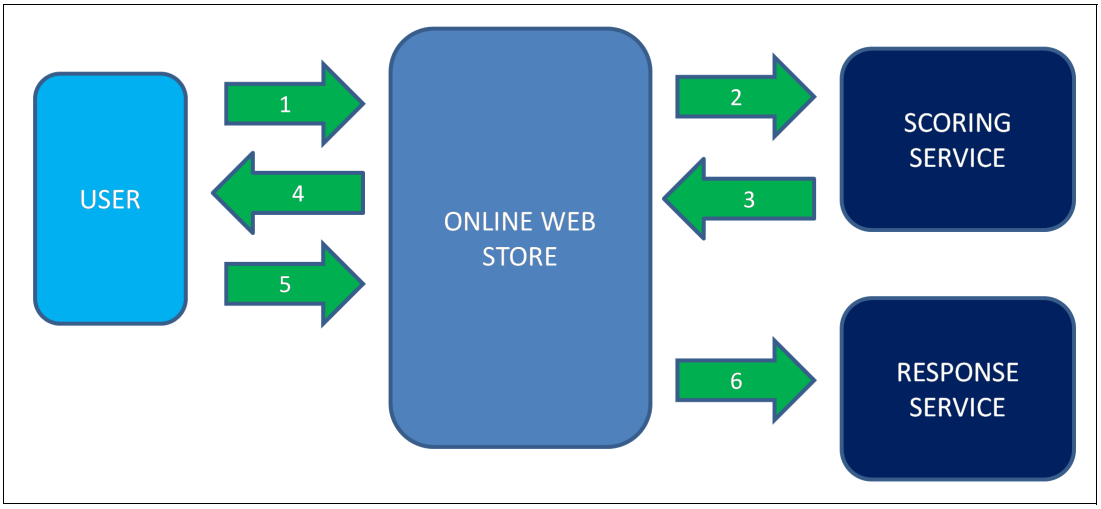


Figure 5-43 Scoring Service and Response Service integration with online web store

Table 5-2 shows the Scoring Service and the Response Service data flow.

Table 5-2 Scoring Service and Response Service data flow with online web store

Event	Description
1 - User Authentication	The user authenticates and the historical behavior data is retrieved with the customer ID.
2 - Scoring Service request	The online web store sends customer identification and behavior data to the Scoring Service.
3 - Scoring Service response	The Scoring Service determines the best offer for the particular customer identification and sends the offer back to the online web store.
4 - Presenting offer	The online web store presents the best offer to the user.
5 - Accept/decline offer	The user accepts or declines the offer.
6 - Response Service log	The online web store sends the response to the Response Service and it is logged for future reference and analysis.

6

Scenario: How to use the solution

Now that you understand the basic implementation procedures of the hardware and the software that are used to build an IBM analytics environment, we describe how to use them.

This chapter shows a sample use case that applies twitter sentiment analysis to a traditional data warehouse dashboard to generate retail offers for users on demand.

This chapter shows the result that you can get through this sample case. The chapter shows the benefit that you can receive from this solution. Then, the chapter describes the concept of this sample case from the data's point of view. The chapter describes the steps to implement the sample case in your environment.

This chapter covers the following topics:

- Dashboard and reporting analysis
- Type of analysis
- Combinations and correlations of structured and unstructured data types
- Use cases and examples

6.1 Dashboard and reporting analysis

This section explains the results that you can get if you implement this sample solution in your environment.

This section covers the following topics:

- View of BigInsightss
- View of Cognos Business Intelligence
- View of IBM SPSS Analytical Decision Management
- Customer experience: Online shop

6.1.1 View of BigInsights

This section provides a brief view of BigInsights with data that is retrieved from Twitter.

Retrieved data from Twitter

Figure 6-1 shows the sample query results against data that is retrieved and stored from Twitter to BigInsights.

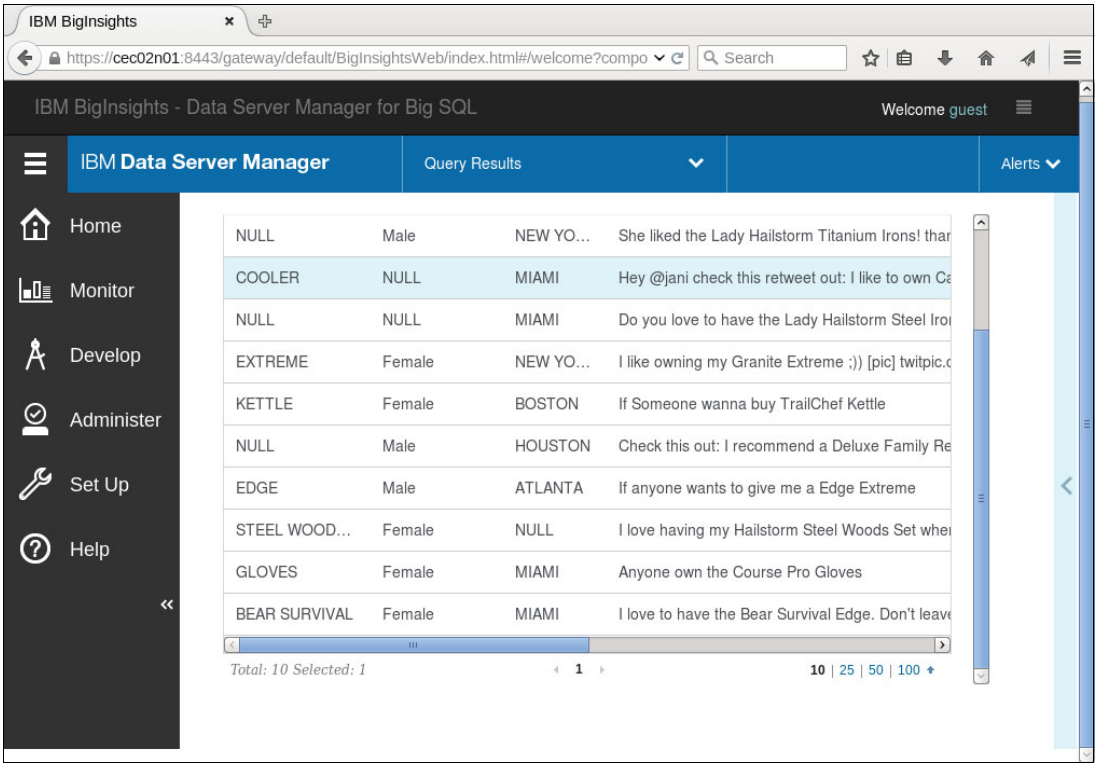


Figure 6-1 Sample result: IBM BigInsights Data Server Manager for Big SQL

6.1.2 View of Cognos Business Intelligence

This section provides a view of Cognos Business Intelligence.

Executive workspace (dashboard)

The executive workspace is the sample executive dashboard of Cognos Business Intelligence. You can create this dashboard in your environment by following this demonstration use case.

The dashboard that is created for executives contains four workspace reports (Figure 6-2).

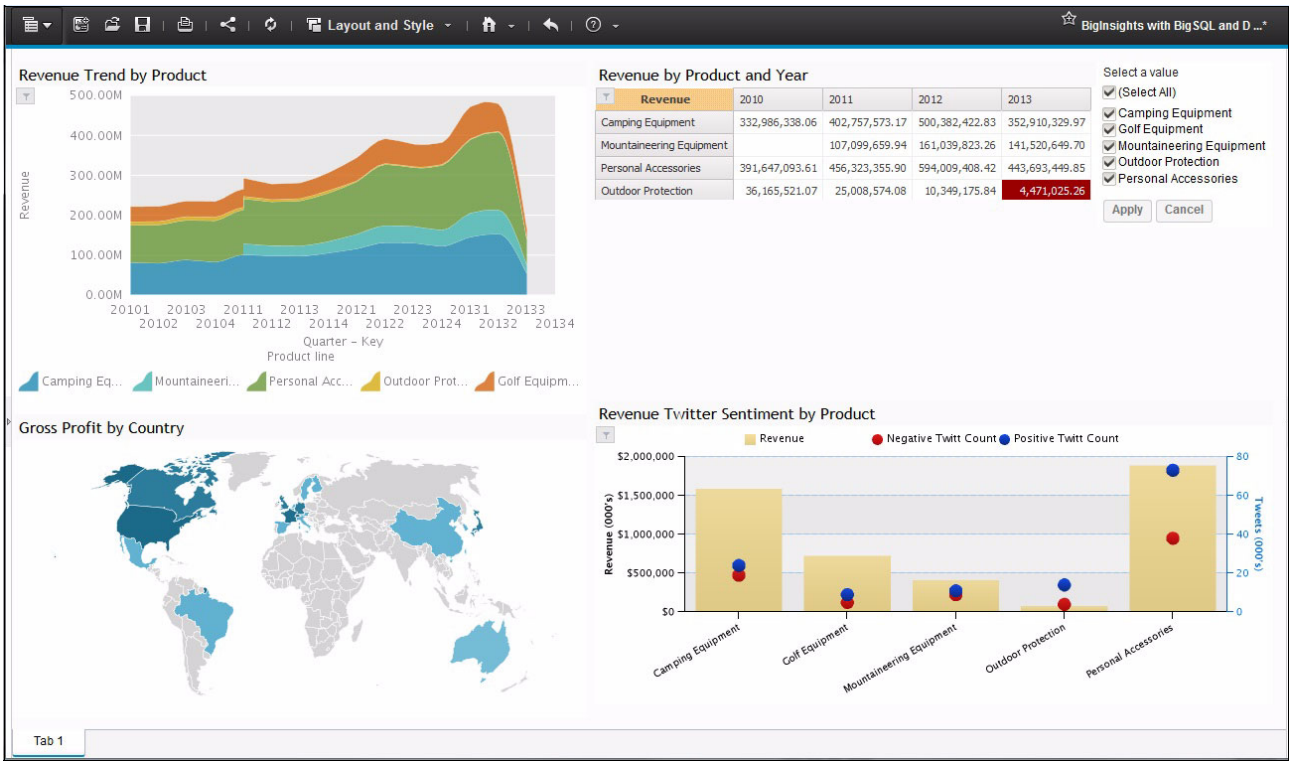


Figure 6-2 Sample view: Executive dashboard in Cognos Connection

Each workspace report in the dashboard is described.

Chart 1

Chart 1 is in the upper-left corner of Figure 6-2. This chart shows revenue trends for each product line, by each quote, from 2010 Q1 to 2013 Q4. You can drill down to see the trend of related product types or product names if you select each product line in same chart of this report.

The original workspace report is shown in Figure 6-3.

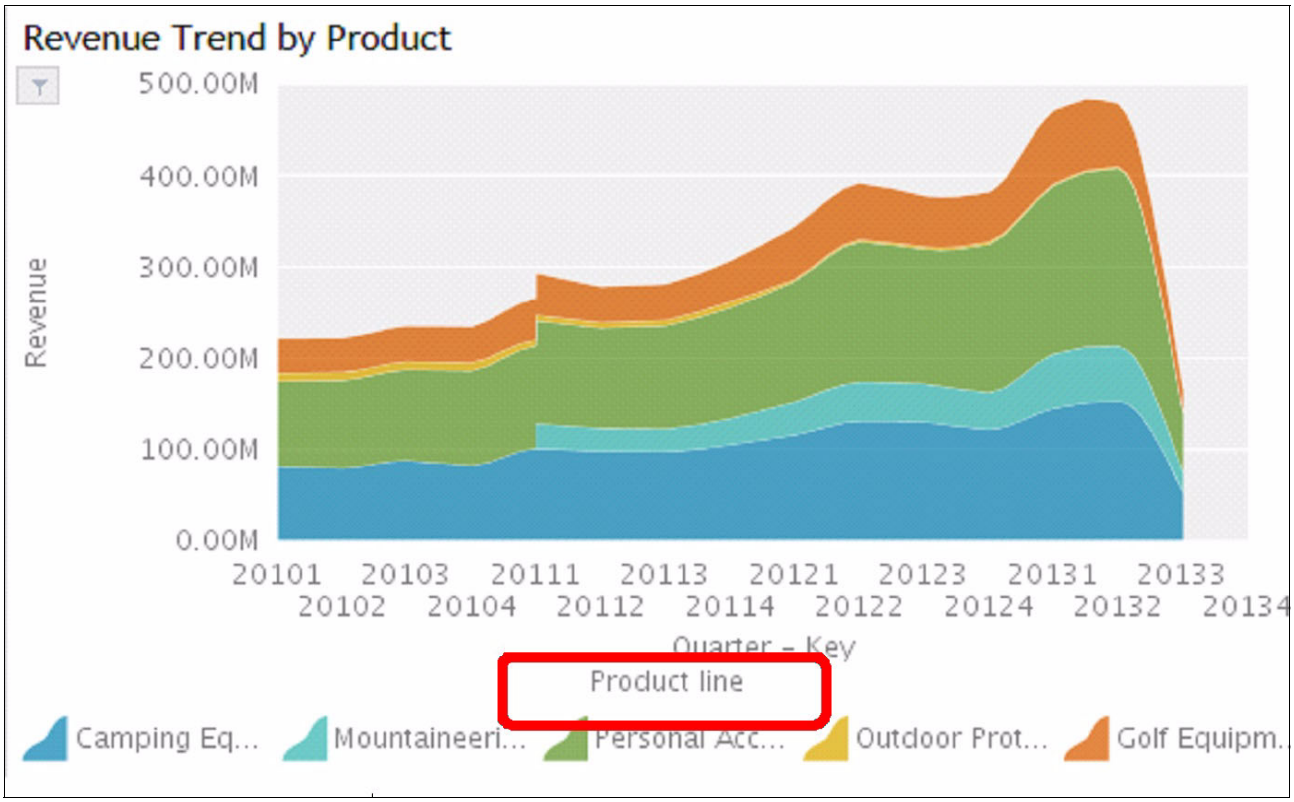


Figure 6-3 Sample result: Workspace report before you drill down

You can see the details by drilling down on figures in the chart. This example drilled down against the product line of personal accessories to show related product types. (To check what is shown in the chart, see the report name that is highlighted by the red rectangle, as shown in Figure 6-3 and Figure 6-4 on page 193).

Figure 6-4 shows the personal accessories report.

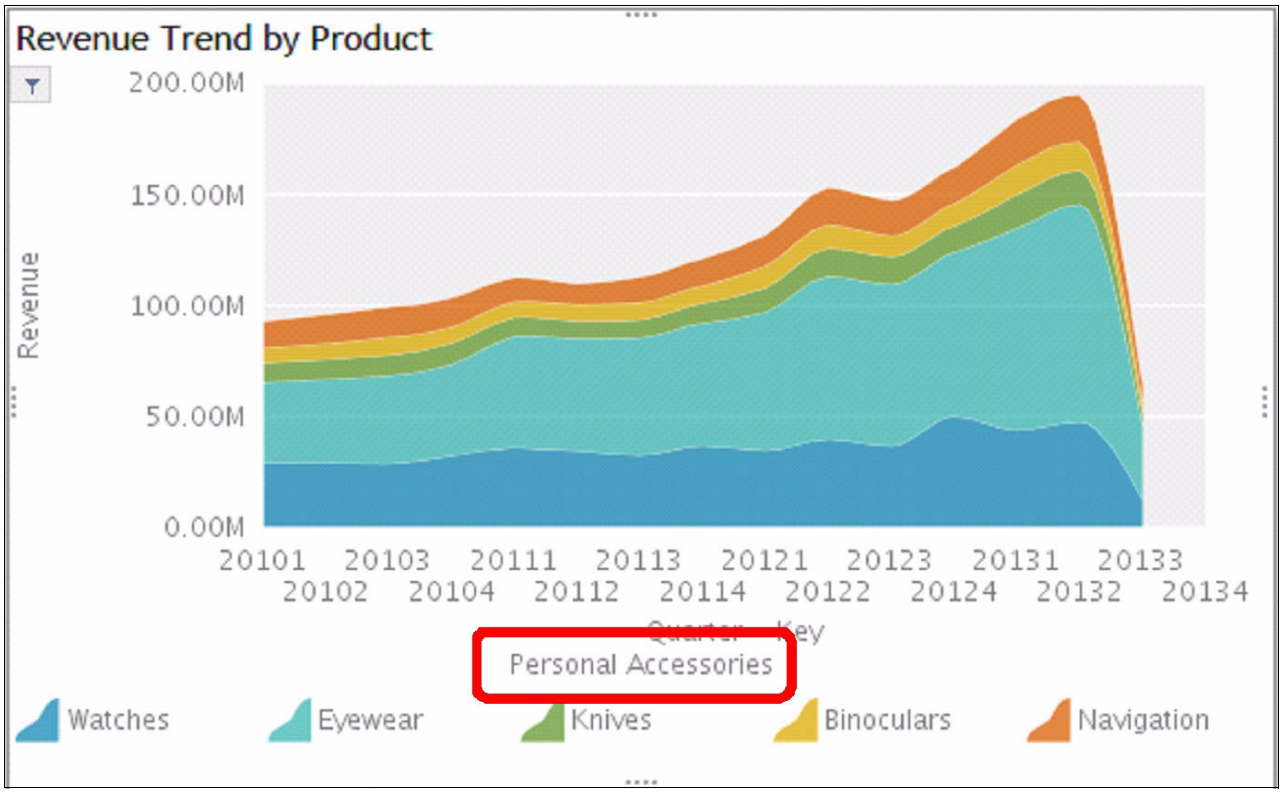


Figure 6-4 Workspace report drilling down from product line to the product type of personal accessories

Chart 2

Chart 2 is in the upper-right corner in Figure 6-2 on page 191. The cross tab shows the accurate revenue amount for each product line by each year (Figure 6-5). You can also drill down to see the related product types or product names if you click each product line in the cross tabulation (crosstab or cross table).

Revenue by Product and Year

Revenue	2010	2011	2012	2013
Camping Equipment	332,986,338.06	402,757,573.17	500,382,422.83	352,910,329.97
Mountaineering Equipment		107,099,659.94	161,039,823.26	141,520,649.70
Personal Accessories	391,647,093.61	456,323,355.90	594,009,408.42	443,693,449.85
Outdoor Protection	36,165,521.07	25,008,574.08	10,349,175.84	4,471,025.26

Figure 6-5 Sample result: Crosstab sample of revenue by each product line

Chart 3

Chart 3 is in the lower-left corner of Figure 6-2 on page 191. The color on the map shows the annual gross profit for each country (Figure 6-6). You can drill down for more details about each country's data.

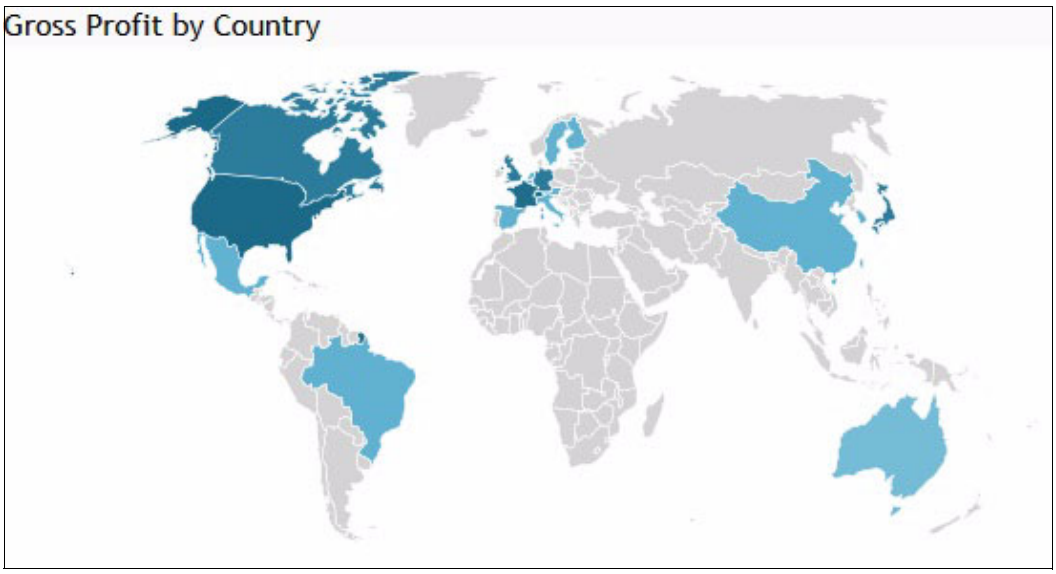


Figure 6-6 Sample result: Dynamic map of gross profit by country

Chart 4

Chart 4 is in the lower-right corner of Figure 6-2 on page 191. Chart 4 is a combination chart. It includes a bar chart, which shows the annual product line revenue data from the traditional data warehouse. The dots show the Twitter polarity (positive/negative counts) for each product line. Figure 6-7 shows the differences between the number of negative and positive tweets for each product line.

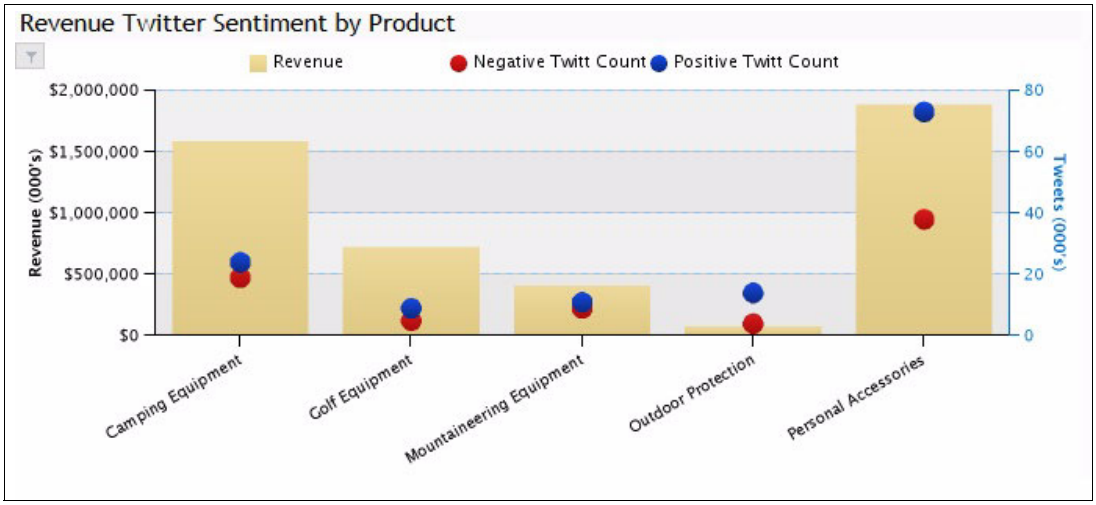


Figure 6-7 Sample result: Sentiment analysis that is applied to a traditional data warehouse

6.1.3 View of IBM SPSS Analytical Decision Management

This section explains the sample project scenario that is scored by IBM Statistical Package for the Social Sciences (SPSS) Analytical Decision Management. It also has a feature to compare multiple scored scenarios to select the most suitable scenario for your requirements. The most profitable and effective scenario for the offerings of your campaign under your set conditions are shown in Figure 6-8.

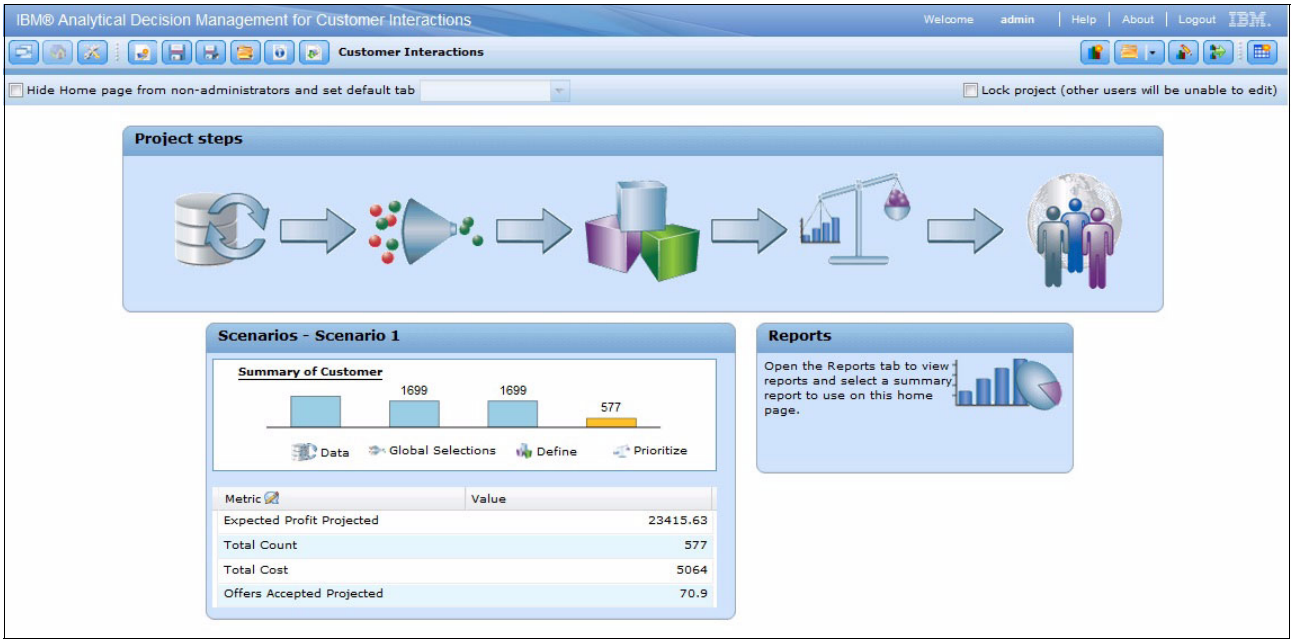


Figure 6-8 Sample result: Home page of a project

6.1.4 Customer experience: Online shop

The fictional sample company *Great Outdoors Co., Ltd* has an online shop. Figure 6-9 shows the home page for its shop. The scenario sets a common home page for all customers. Then, it sets the individual pages for each online shop customer after the customer logs in to their account.

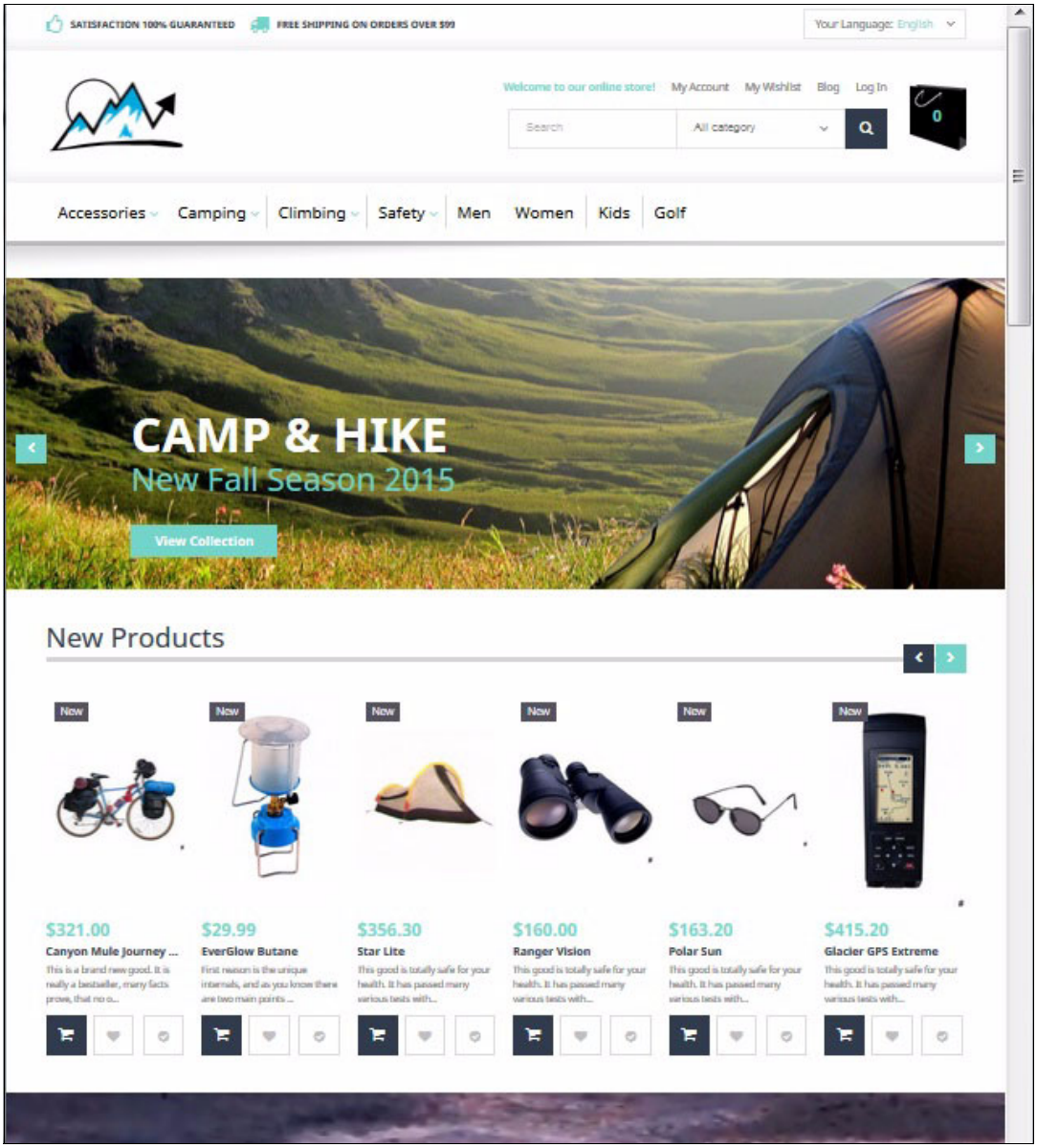


Figure 6-9 Sample result: Home page of "Great Outdoors Co., Ltd" online shop before customer login

Two scenarios are shown. In the first scenario, no campaign is offered. In the second scenario, a campaign is offered. The scenarios are described:

- The customer experience 1: Campaign not offered

  After the correct configuration to the web application in the application server is published, it can retrieve the scoring result automatically from SPSS Analytical Decision Management and identify the campaign offer to display at the user login. On this page, no campaign offer is displayed. If a customer is not offered a campaign, the customer sees a web page after they log in to the online shop website, as shown in Figure 6-10.

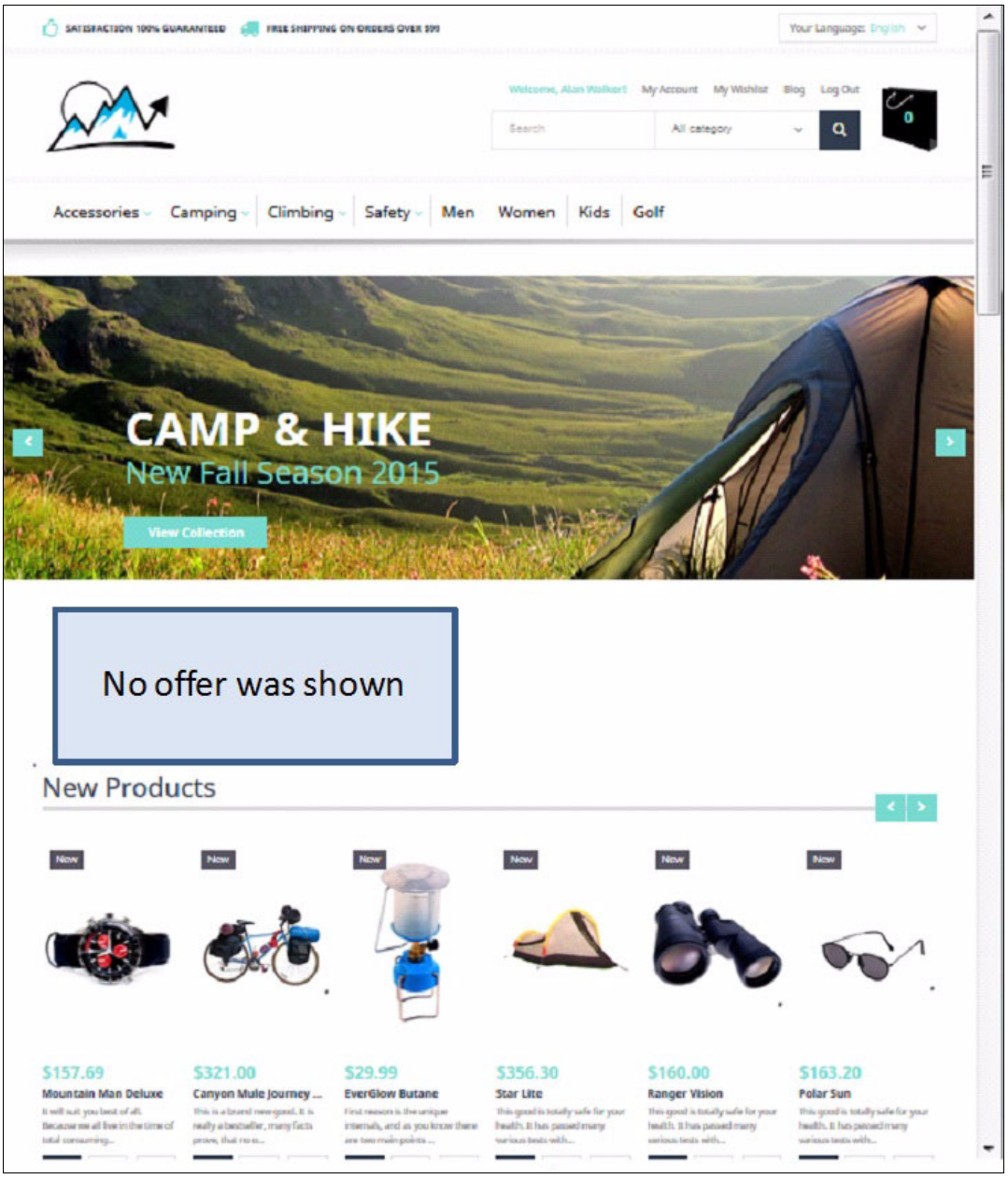


Figure 6-10 Online shop page for the customers who are not offered a campaign

- Customer experience 2: Campaign offered

  If the customer is offered a campaign, the customer sees the campaign information after they log in to the online shop website, as shown in Figure 6-11. The campaign information is shown according to the scenario that was scored by SPSS Analytical Decision Management based on the customer's geographical information, the result of twitter sentiment analysis, and the customer's banking status.

Figure 6-11 Online shop page for customers who are offered a campaign

6.2 Type of analysis

This section explains the benefits of the implemented solution.

This section covers the following topics:

- Cognos Business Intelligence advantages
- SPSS Analytical Decision Management advantages

6.2.1 Cognos Business Intelligence advantages

This implementation offers the following advantages:

- Applies the correct security to a confidential report automatically based on the user's information about the security provider. (This case uses OpenLDAP as the security provider.)
- Retrieves unstructured data from BigInsights through Big SQL in addition to the traditional structured data in the data warehouse
- Uses the DB2 Java Database Connectivity (JDBC) driver to set the data source definition of Big SQL in Cognos Administration
- Displays data both from structured and unstructured databases in the same report (for example, charts and lists)
- Shows sentiment analysis results and traditional data warehouse data in the same report object (for example, charts and list)
- Enables the user to drill down to see detailed data on demand
- Enables the user to check the latest data whenever the user executes a report

6.2.2 SPSS Analytical Decision Management advantages

The advantages of this implementation are shown:

- Simulates profitable and productive scenario offering campaigns for customers that Cognos cannot simulate by itself.
- Enables the user to compare multiple scenarios that are scored in one project.
- Enables the user to import created models to the project.
- Enables the user to create models to forecast data that fits to rules and conditions that you set in the project.
- Helps a web application to display campaign offers automatically to a customer's home page after the customer logs in to the online shop. The offering is provided by the scoring result that is based on the user's geographical and demographic data.

6.3 Combinations and correlations of structured and unstructured data types

This section describes how to implement sample cases in your environment.

This section covers the following topics:

- Sample case goal
- Story of this sample case
- How to score and deploy the scenario
- Differences between Cognos Business Intelligence and SPSS Analytical Decision Management
- Where the data comes from

6.3.1 Sample case goal

In this sample case, the goal is for you to perform similar operations that are explained in the "Using Twitter sentiment from Hadoop to generate retail offers" video, which is implemented on IBM POWER8:

https://www.youtube.com/watch?v=zlRWt4XdTcM

Note: Our scenario implements a later version of the product version than the version that is shown in the demonstration. The numbers that are displayed and the user interfaces of each product vary based on the product version. Therefore, several views in this sample case differ from the demonstration.

6.3.2 Story of this sample case

This section describes the sample case scenario.

About the sample case

Cognos examples were designed for the sample company *Great Outdoors Co., Ltd*. These samples were included originally to give a good description of the product features and preferred practices from both the business and technical views. Three types of people (executives, IT staff, and researcher) are described in this demonstration. Table 6-1 shows the roles of these people in this demonstration.

Table 6-1 List of each user and the user's role in each product

User	Cognos Business Intelligence	Analytical Decision Management or Collaboration and Deployment Service	Online shop
Executives	► The executives log in to Cognos Business Intelligence portal. ► The executives open the executive dashboard.	They do not use this product.	They do not use this product.
IT staff	► The IT staff creates the executive dashboard. ► Applies security.	The IT staff configures the scoring service[a].	The IT staff applies the configuration to deliver a score from Analytical Decision Management to the web application on the application server.
Researcher	The researcher does not use this product.	The researcher creates the project to find the most effective and profitable campaign scenario[b].	The researcher does not use this product.

a. This operation occurs in Deployment Manager in the Windows client of Collaboration and Deployment Service.
b. This operation occurs in SPSS Analytical Decision Management.

Story

This section describes the story of the people who are involved in this sample case.

The *IT staff* applied security to all of the Cognos Business Intelligence. They also applied additional security to the executive dashboard so that only executives can see it by using employee user information that is stored in the security provider. OpenLDAP was implemented for the security provider. The IT staff modified the web application to display campaign information automatically to target the customer's page based on the scenario that was scored by the Analytical Decision Management tool.

The *executives* open the confidential financial dashboard to check the company's confidential financial status. This dashboard is displayed to executives only. From revenue by product and year, in the executive dashboard, the 2013 revenue of all of the product lines is lower than the revenue of 2012. Executives want to improve this situation. For example, the product line that is named Personal Accessories has the largest revenue and most positive tweets but it also has the most negative tweets (from Chart 4 on the executive dashboard on Cognos Connection that is shown in Figure 6-7 on page 194). Then, executives can decide to perform campaigns against customers who posted positive twitter feedback and have good financial status.

The *researcher* works with Analytical Decision Management to identify the customers to whom the company needs to offer campaigns and to identify the most effective and profitable way to communicate offers.

6.3.3 How to score and deploy the scenario

Four steps are involved to score and deploy the scenario in the online web page:

1. Prepare the data source.
2. Create the model.
3. Create the project.
4. Apply the scenario to the online shop.

Prepare the data source

The researcher creates the input data that contains the individual's demographic data (such as gender and age from the `bank_response_data.txt` file), banking payment status, and sentiment polarity (such as Has Bad Payment Record and Sentiment Polarity Tweeted product from the `bank_customer_data.txt` file).

For the tweets about products by customers of Great Outdoors Co., Ltd, we retrieved the data from Twitter and then stored it in BigInsights for this case. Also, we can select and display customized data by using SQL in BigInsights.

Create the model

The researcher creates the model to receive the propensity score of the response that is in the `bank_response_data.txt` file. In this step, we create the model on SPSS Analytical Decision Management so that its user interface enables people with insufficient knowledge about the analytics of SPSS Modeler to create a model that fits this scenario. The *Modeler Advantage* application offers this feature.

The scored result is used to set the prioritization parameters of the *SPSS Analytical Decision Management for Customer Interactions* application. The scored value tells the expected reaction ratio from customers by the company's campaign or offers. This value is calculated from the `bank_response_data.txt` file and not from the `bank_customer_data.txt` file that is the data source of SPSS Analytical Decision Management for Customer Interactions.

Create the project

The researcher works with Analytical Decision Management to identify the best scenario that will predict the most effective expected profit:

- Outline

  This task creates the application on SPSS Analytical Decision Management. For this sample case, select the application type of Analytical Decision Management for Customer Interactions to create a project for scoring scenarios. Scenarios are scored with the conditions and rules that you set in the project. This scoring identifies the best offers and delivers recommendations to the website in real time. The company uses this scoring when it wants to increase customer retention and overall profits through several overlapping marketing campaigns.

- Conditions and rules in the project

  SPSS Analytical Decision Management scores the expected profit when campaigns are performed effectively with your specific conditions and rules in each project. Great Outdoors Co., Ltd offers campaigns under the following conditions that were scored by Analytical Decision Management to achieve profitable results. Also, the company wants to display the campaign detailed information in the online shop for each user's home page after the login with following conditions (Table 6-2):

  – The customer who posted tweets on the Personal Accessories product line.

  – The customer with one of the following four rules:

    • Watch campaign for male retention

    • Sunglasses campaign for female retention for customers who live in Miami and posted positive tweets

    • Great Outdoors Credit Card campaigns for cross-sell to customers who own a house

    • Travel with Outdoors campaigns for cross-sell to customers in good financial standing

Table 6-2 Sample configuration: Campaign and offering types by target customers

Types of campaign	Offers	Target customers
Retention	30 percent off on Sunglasses	Male retention
	30 percent off on Watches	Female retention for customers who live in Miami and posted positive tweets
Cross-sell	Great Outdoors Credit Card	Cross-sell to customer who owns a house
	Travel with Outdoors	Cross-sell to customer in good financial standing

Required operations in SPSS Analytical Decision Management

This section describes the concept and outline of the tasks to identify the effective scenario in SPSS Analytical Decision Management. To complete this implementation, you need to configure the items in the following list:

- Data tab
- Global selections tab
- Define tab

- Prioritize tab
- Deploy tab

We describe the configurations for each of the items in the preceding list:

- Data tab: This feature is used to load data for the simulation. In this sample case, the data source was loaded. The `bank_customer_data.txt` file includes the user information, polarity, and products that are mentioned in the tweets.
- Global selections tab: This feature is used to select the customers that receive a campaign or offers. In this section, select the users that satisfy both of the following rules:
  - Excludes the user with one of the following conditions:
    - "Has Bad Payment Record" = `Yes` (in rule 1 = Yes)
    - Received a campaign offer within eight weeks
  - Includes a user who posted tweeter comments or comments through twitter on the Personal Accessories product line
- Define tab: This feature is used to set the offers to allocate to customers and the conditions on which offers are delivered to customers:
  - Retention campaign: To all of the sales transactions after 2012-03-22 02:35:30, the researcher tries to apply the following rules. The researcher decided to apply the campaigns to the first and second largest revenue products in the Personal Accessories product line (products with the largest revenue by product type (Figure 6-12)):
    - First largest product line: Eyewear
    - Second largest product line: Watches

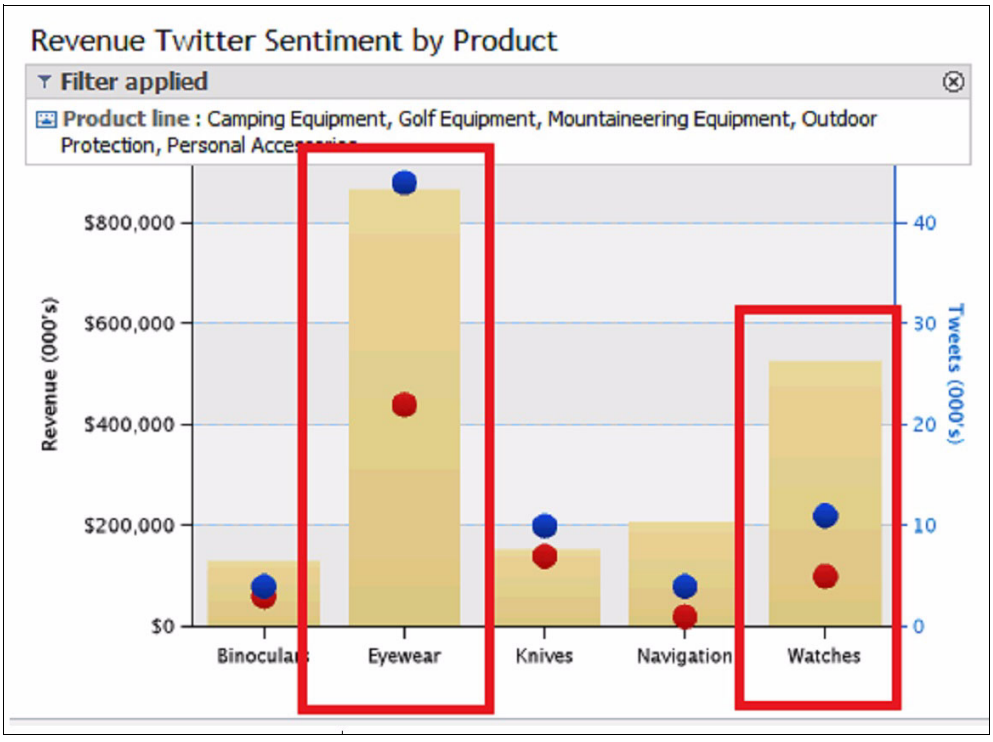


Figure 6-12 Amount of polarity (sentiment analysis) of Twitter and revenue by each product line

- The company decided to check the scenario by applying the following rules for segments who receive offers:
  - Sunglass discount for females who posted positive tweets and live in Miami
  - Watch for males
- Cross-sell campaign

  The company decided to apply each campaign from the sentiment analysis that resulted in Chart 4 in the executive dashboard of the Cognos Business Intelligence report (Figure 6-7 on page 194) with the following rules and conditions:
  - Great Outdoors Credit Card campaign for the people who own a house (The user can set the house as collateral for the loan.)
  - Travel with Outdoors campaign for people whose personal debt to equity ratio is shorter than 30 days

- Prioritize tab: This feature is used to set items (for example, *costs* and probability to respond (*Prob. to Respond*)). The company prepares two campaigns:
  - For customer retention
  - For customer cross-sell: This campaign is for the bank's cross-sell credit card.

  To identify the most effective and profitable way of allocating cost and earning expected profit, prioritize in the following manner. In this sample case, Prob. to Respond is the ratio of the customer's response ratio against our action (Figure 6-13).

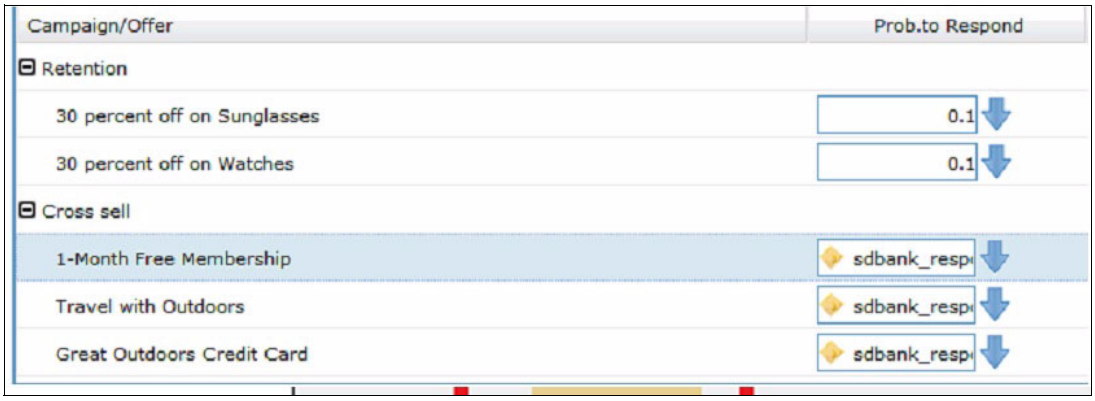


Figure 6-13 Setting the prioritization parameters in SPSS Analytical Decision Management

The expected profit was calculated based on the expression, as shown in Figure 6-14.

Figure 6-14 Expression that is used to score the scenario in this sample case

- Deploy tab: After you set all items that are described in this document from Data to Deploy, you can create a new scenario that is based on the current project settings, as shown in Figure 6-15.

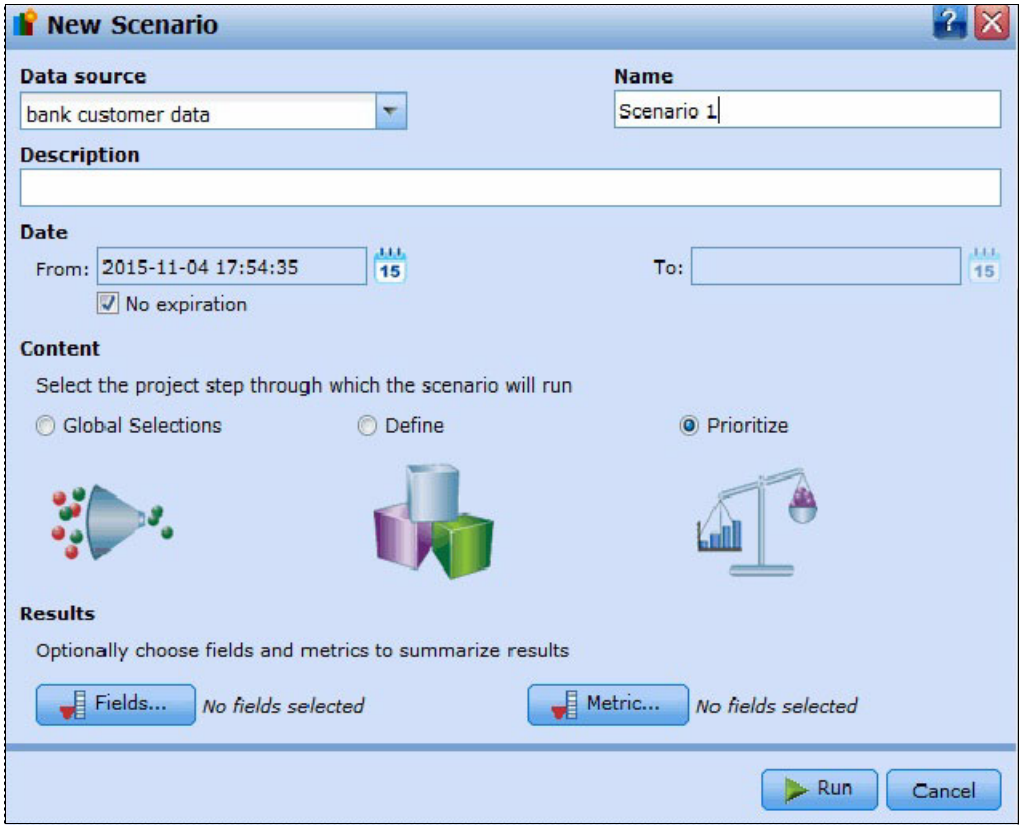


Figure 6-15 Creating the scenario in the Deploy tab

Then, you get the data in the home page of your project, as shown in Figure 6-16.

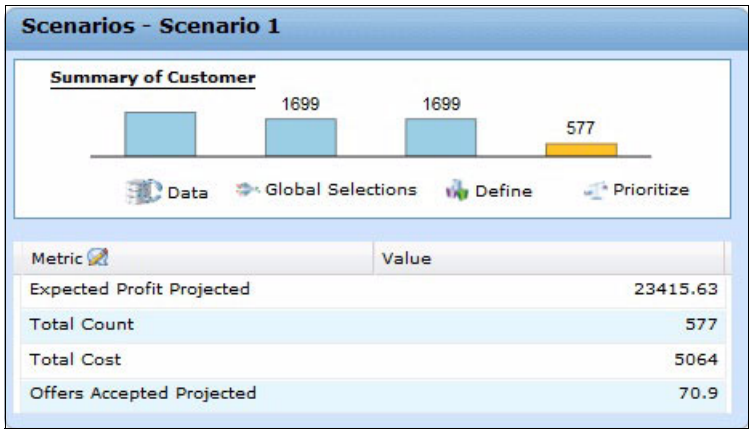


Figure 6-16 Expected profit and other results that are scored by the current configuration

Apply the scenario to the online shop

To integrate the online shop with SPSS Analytical Decision Management, the online shop web service requests that SPSS Scoring Services post the appropriate campaign or offer to present to the online customer based on the customer's scoring based on historical information and purchases.

The IBM SPSS Scoring Services software is generally available at this website:

`http://<spss_adm_server>:9080/scoring/rest/configuration/<scoring_service>/score`

In our sample demonstration, the scoring service is named *bank customer interaction*, which requires that the following fields are populated with customer data:

- Age
- Months as a Customer
- Number of Products
- RFM Score
- Average Balance Feed Index
- Number of Transactions
- Personal Debt to Equity Ratio
- Months Current Account
- Number of Loan Accounts
- Customer ID
- Has Bad Payment Record
- Members Within Household
- Number of Call Center Contacts
- Gender
- Marital Status
- Age Youngest Child
- Number of Workers in Household
- Percentage White Collar Workers
- Household Debt to Equity Ratio
- Income
- Weeks Since Last Offer
- Homeowner
- Accepted Personal Loan
- Accepted Retention
- Accepted Home Equity Loan
- Accepted Credit Card
- Annual value
- Interested in Personal Loan
- Interested in Retention
- Interested in Home Equity Loan
- Interested in Credit Card
- Key
- City
- Product Category
- Sentiment Polarity

You can refer to the SPSS Scoring Service to check the required format of the requested message and the returned data:

`http://<spss_adm_server>:9080/scoring/rest/configuration/<scoring_service>/metadata`

See Example 6-1.

Example 6-1 SPSS scoring service request message metadata

```
<metadataResult>
    <metadataInputTable id="Customer" name="Customer">
        <metadataInputField isRequired="true" name="Age" type="long" />
        <metadataInputField isRequired="true" name="Months as a Customer" type="long" />
        <metadataInputField isRequired="true" name="Number of Products" type="long" />
        <metadataInputField isRequired="true" name="RFM Score" type="double" />
        <metadataInputField isRequired="true" name="Average Balance Feed Index" type="long" />
        <metadataInputField isRequired="true" name="Number of Transactions" type="long" />
        <metadataInputField isRequired="true" name="Personal Debt to Equity Ratio" type="long" />
        <metadataInputField isRequired="true" name="Months Current Account" type="long" />
        <metadataInputField isRequired="true" name="Number of Loan Accounts" type="long" />
        <metadataInputField isRequired="true" name="Customer ID" type="long" />
        <metadataInputField isRequired="true" name="Has Bad Payment Record" type="long" />
        <metadataInputField isRequired="true" name="Members Within Household" type="long" />
        <metadataInputField isRequired="true" name="Number of Call Center Contacts" type="long" />
        <metadataInputField isRequired="true" name="Gender" type="string" />
        <metadataInputField isRequired="true" name="Marital Status" type="string" />
        <metadataInputField isRequired="true" name="Age Youngest Child" type="long" />
        <metadataInputField isRequired="true" name="Number of Workers in Household" type="long" />
        <metadataInputField isRequired="true" name="Percentage White Collar Workers" type="long" />
        <metadataInputField isRequired="true" name="Household Debt to Equity Ratio" type="long" />
        <metadataInputField isRequired="true" name="Income" type="long" />
        <metadataInputField isRequired="true" name="Weeks Since Last Offer" type="long" />
        <metadataInputField isRequired="true" name="Homeowner" type="string" />
        <metadataInputField isRequired="true" name="Accepted Personal Loan" type="string" />
        <metadataInputField isRequired="true" name="Accepted Retention" type="string" />
        <metadataInputField isRequired="true" name="Accepted Home Equity Loan" type="string" />
        <metadataInputField isRequired="true" name="Accepted Credit Card" type="string" />
        <metadataInputField isRequired="true" name="Annual value" type="long" />
        <metadataInputField isRequired="true" name="Interested in Personal Loan" type="string" />
        <metadataInputField isRequired="true" name="Interested in Retention" type="string" />
        <metadataInputField isRequired="true" name="Interested in Home Equity Loan" type="string" />
        <metadataInputField isRequired="true" name="Interested in Credit Card" type="string" />
        <metadataInputField isRequired="true" name="Key" type="long" />
        <metadataInputField isRequired="true" name="City" type="string" />
        <metadataInputField isRequired="true" name="Product Category" type="string" />
        <metadataInputField isRequired="true" name="Sentiment Polarity" type="string" />
    </metadataInputTable>
    <metadataOutputField isReturned="true" name="Campaign" type="string" description="" />
    <metadataOutputField isReturned="true" name="Offer" type="string" description="" />
    <metadataOutputField isReturned="true" name="Output-PredictedProfit" type="string" description=""
/>
    <metadataOutputField isReturned="true" name="Output-MaxOffersNum" type="integer" description="" />
    <metadataOutputField isReturned="true" name="Output-MinProfit" type="double" description="" />
    <metadataOutputField isReturned="true" name="Output-ProbToRespond" type="double" description="" />
    <metadataOutputField isReturned="true" name="Output-Revenue" type="double" description="" />
    <metadataOutputField isReturned="true" name="Output-Cost" type="double" description="" />
</metadataResult>
```

For example, you can create the following request message and POST by using content-type: application/json to http://<*spss_adm_server*>:9080/scoring/rest/configuration/bank customer interaction and score, as shown in Example 6-2.

Example 6-2 Scoring service request message example

```
{
    "id": "bank customer interaction",
    "requestInputTable": [{
        "name": "Table 1",
        "requestInputRow": [{
            "input": [{
                "name": "Age",
                "value": "24"
            }, {
                "name": "Months as a Customer",
                "value": "0"
            }, {
                "name": "Number of Products",
                "value": "1"
            }, {
                "name": "RFM Score",
                "value": "0"
            }, {
                "name": "Average Balance Feed Index",
                "value": "0"
            }, {
                "name": "Number of Transactions",
                "value": "0"
            }, {
                "name": "Personal Debt to Equity Ratio",
                "value": "24"
            }, {
                "name": "Months Current Account",
                "value": "-4"
            }, {
                "name": "Number of Loan Accounts",
                "value": "0"
            }, {
                "name": "Customer ID",
                "value": "1669"
            }, {
                "name": "Has Bad Payment Record",
                "value": "0"
            }, {
                "name": "Members Within Household",
                "value": "6"
            }, {
                "name": "Number of Call Center Contacts",
                "value": "10"
            }, {
                "name": "Gender",
                "value": "F"
            }, {
                "name": "Marital Status",
                "value": "S"
```

```
}, {
    "name": "Age Youngest Child",
    "value": "0"
}, {
    "name": "Number of Workers in Household",
    "value": "7"
}, {
    "name": "Percentage White Collar Workers",
    "value": "33"
}, {
    "name": "Household Debt to Equity Ratio",
    "value": "58"
}, {
    "name": "Income",
    "value": "65263"
}, {
    "name": "Weeks Since Last Offer",
    "value": "17"
}, {
    "name": "Homeowner",
    "value": "F"
}, {
    "name": "Accepted Personal Loan",
    "value": "F"
}, {
    "name": "Accepted Retention",
    "value": "F"
}, {
    "name": "Accepted Home Equity Loan",
    "value": "F"
}, {
    "name": "Accepted Credit Card",
    "value": "F"
}, {
    "name": "Annual value",
    "value": "30"
}, {
    "name": "Interested in Personal Loan",
    "value": "F"
}, {
    "name": "Interested in Retention",
    "value": "T"
}, {
    "name": "Interested in Home Equity Loan",
    "value": "F"
}, {
    "name": "Interested in Credit Card",
    "value": "F"
}, {
    "name": "Key",
    "value": "1"
}, {
    "name": "City",
    "value": "Miami"
}, {
```

```
                "name": "Product Category",
                "value": "Personal Accessories"
            }, {
                "name": "Sentiment Polarity",
                "value": "Majority Positive"
            }]
        }]
    }],
    "context": []
}
```

Example 6-3 shows the expected response for this request.

Example 6-3 Expected output of the scoring services online request

```
200 OK
X-Powered-By:  Servlet/3.0
Content-Type:  application/xml;q=0.9
Content-Language:  en-US
Content-Length:  742
Date:  Fri, 06 Nov 2015 23:28:03 GMT

<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<scoreResult xmlns="http://xml.spss.com/scoring-v2" providedBy="bank customer
interaction" id="fe3fdfcd-b65b-40ae-9eb4-87838ac438fe">
    <columnNames>
        <name>Campaign</name>
        <name>Offer</name>
        <name>Output-PredictedProfit</name>
        <name>Output-MaxOffersNum</name>
        <name>Output-MinProfit</name>
        <name>Output-ProbToRespond</name>
        <name>Output-Revenue</name>
        <name>Output-Cost</name>
    </columnNames>
    <rowValues>
        <value value="Cross sell">
        </value>
        <value value="Travel with Outdoors">
        </value>
        <value value="11.22511451590822">
        </value>
        <value value="2">
        </value>
        <value value="10.0">
        </value>
        <value value="0.07112557257954111">
        </value>
        <value value="200.0">
        </value>
        <value value="3.0">
        </value>
    </rowValues>
</scoreResult>
```

As you can see in Example 6-3 on page 210, this user will receive an offer of *Travel with Outdoors* from the *Cross sell* campaign. According to the SPSS Analytical Decision Management analysis, the probability of this customer to respond positively is 7%.

With this raw data, any web developer can customize the online shop pages according to the logged in user. Therefore, when the customer is authenticated in the online shop, this type of scoring request can be performed in the background. The campaign/offer is presented to the online customer according to the corporate strategy and web pages development platform.

6.3.4 Differences between Cognos Business Intelligence and SPSS Analytical Decision Management

Both Cognos Business Intelligence and SPSS Analytical Decision Management can show the formatted results from data sources in their portal sites. What are the greatest differences?

The difference is Cognos Business Intelligence displays past results but SPSS Analytical Decision Management displays expected future results:

- Cognos Business Intelligence

  Cognos Business Intelligence is the tool for displaying past results. In this software, Twitter *sentiment analysis* (total number of positive comments for products and negative comments against products) is shown with revenue in the executive dashboard by using one of our sample packages, Great Outdoors Warehouse. Except for twitter data in BigInsights, all data comes from the sales data of Great Outdoors Co., Ltd that is stored in DB2 BLU. The company's research team retrieves the data from Twitter posts that commented about products of Great Outdoors Co., Ltd.

- SPSS Analytical Decision Management

  SPSS Analytical Decision Management is the tool to display the predicted future result based on the company's existing data (for example, revenue and customer information). From its score, the user can pick the most effective and profitable scenario.

6.3.5 Where the data comes from

This section precisely describes where the displayed data comes from.

Cognos Business Intelligence

This section describes the executive dashboard that was set as the goal for displaying data in Cognos, as shown in Figure 6-17 on page 212:

- Chart 1: Upper-left quadrant

  This chart shows the revenue trends for each product line by each quarter from 2010 Q1 to 2013 Q4. This data is stored in the gosales data warehouse (DW) in DB2 BLU.

  X = Quarter (displayed as *YYYYQuarter name*, for example, 2010 Q1 is shown as 20101).
  Y = Revenue.

- Chart 2: Upper-right quadrant

  The cross tab shows the accurate revenue amount for each product line by each year. This data is stored in the gosales DW in DB2 BLU.

- Chart 3：Lower-left quadrant

  The colored map shows the annual gross profit of each country. This data is stored in the gosales DW in DB2 BLU.

- Chart 4：Lower-right quadrant

  Bar chart: Annual product line revenue data from DB2 BLU. The dots represent Twitter positive/negative counts for each product line and the counts are from BigInsights.

Figure 6-17 shows all four charts on the executive dashboard.

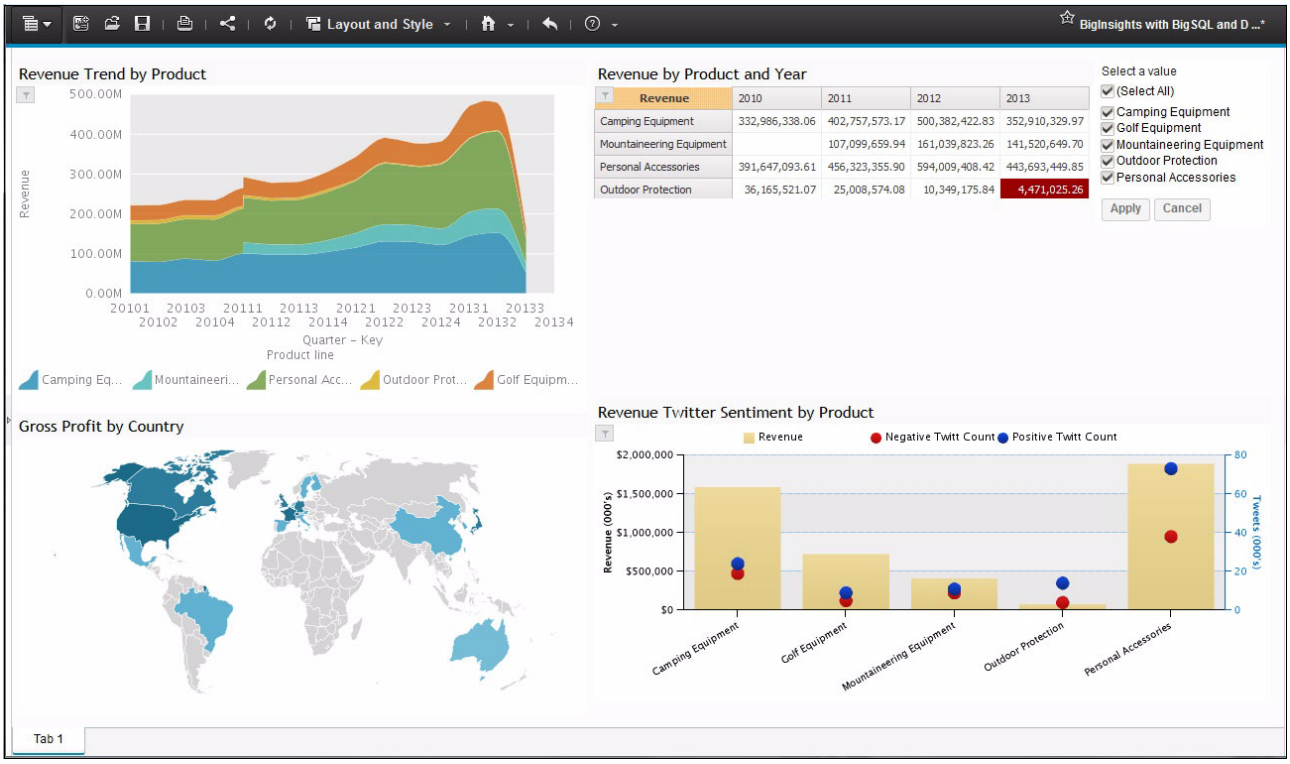


Figure 6-17 Executive dashboard on Cognos Business Intelligence

SPSS Analytical Decision Management

This section describes how to use SPSS Analytical Decision Management:

- Project 1: Score propensity of campaign (Figure 6-18)

  The sdbank_response_model.str model is created by SPSS Modeler Advantage. The data source is the `bank_response_data.txt` file.

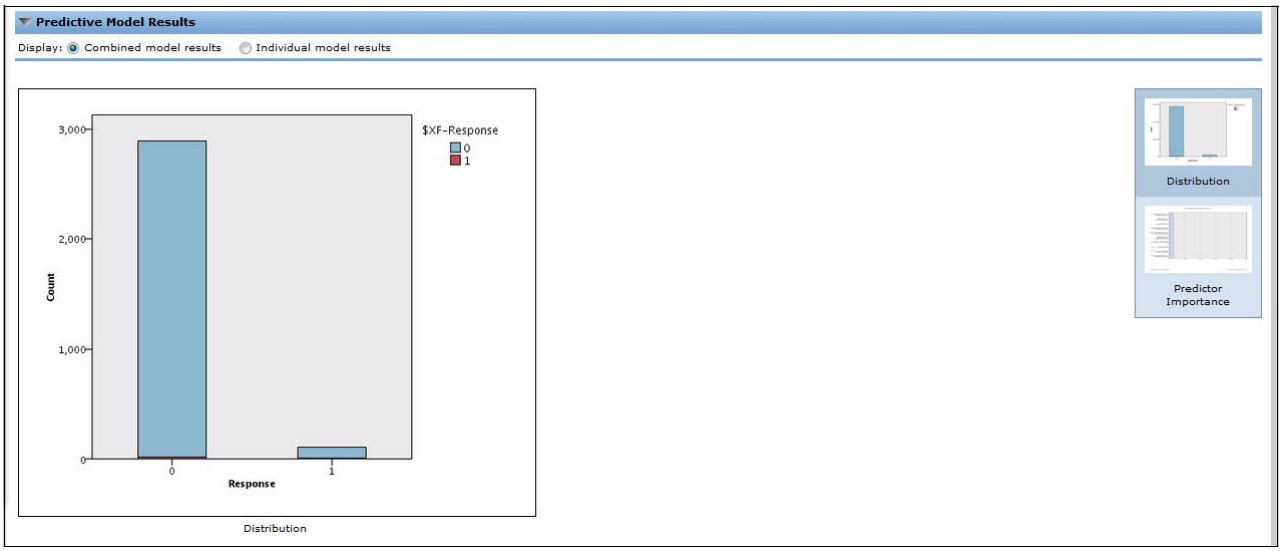


Figure 6-18 Sample result: Model that is created by SPSS Modeler Advantage

- Project 2: Score scenario (Figure 6-19)

  Customer Interactions.str is created by IBM Analytical Decision Management for Customer Interactions. The data source is the `bank_customer_data.txt` file.

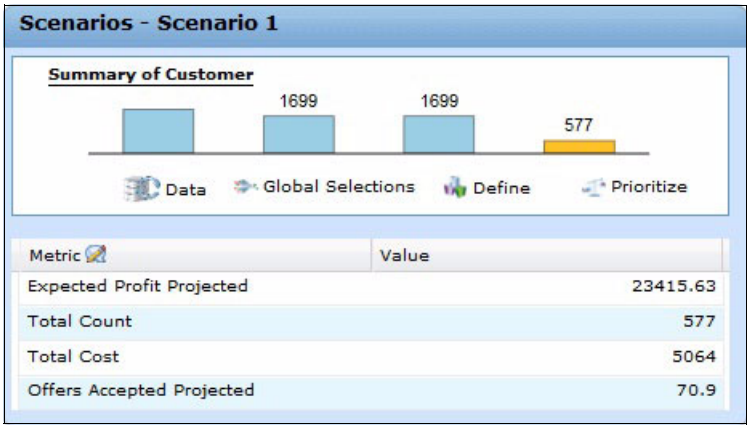


Figure 6-19 Scored scenario by SPSS Analytical Decision Management for Customer Interactions

6.4 Use cases and examples

We described the concepts and the outline of the implementation in previous sections. Now, we present how to implement the sample cases in your environment.

This section covers the following topics:

- Disclaimer
- Installed software
- How to implement the sample case
- Links

6.4.1 Disclaimer

This documentation is written for users with a basic knowledge of the IBM software that is used in these sample cases. For more information about the IBM software, see the product manuals or other online resources.

Complete the installation and configuration of the required software under the supported software environment before you implement these samples in your environment. Also, you need to change the software settings to fit your environment. You might face performance problems, depending on your environment's resources.

If you use the latest product version in our scenario, certain data from the products, such as Cognos or SPSS Analytical Decision Management, differs from the data that is shown in the demonstration video:

- IBM Knowledge Center (for the product manuals):

  https://www.ibm.com/support/knowledgecenter/

- Supported environment

  For the current environment that is supported for each product, check the following website:

  https://ibm.biz/BdEfnG

6.4.2 Installed software

To implement this sample case, you need to install and configure the software that is listed in Table 6-3.

Table 6-3 Installed IBM Software for this sample case

Product	Version	Notes
Installation Manager	1.8.3	
IBM Cognos Business Intelligence Server	10.2.2 Fix Pack (FP) 2	OpenLDAP is used as the security provider.
IBM Cognos Framework Manager	10.2.2 FP2	Only the client on Microsoft Windows is offered for this software.
IBM Cognos Cube Designer	10.2.2 FP2	Only the client on Windows is offered for this software.
Collaboration and Deployment Service	7.0.0.0	In addition to Collaboration and Deployment Service, the following components are required to be installed from the installation manager: ► Repository server ► Modeler adapter
Deployment Manager	7.0.0.0	N/A
SPSS Analytical Decision Management	17.0	N/A
SPSS Modeler Server	17.0	N/A
DB2 BLU	10.5 FP6	If you want to use a distributed installed type of database, the database client must be installed in the terminal of the following software: ► Cognos Business Intelligence Server ► Framework Manager ► Cube Designer ► Collaboration and Deployment Service
IBM WebSphere	8.5.5.0	► This software is installed by the Installation Manager. ► This software is used by Collaboration and Deployment Service.
BigInsights	4	N/A

6.4.3 How to implement the sample case

This section describes how to implement the sample case.

Implementing Cognos Business Intelligence

This section describes the steps to implement the sample case with Cognos Business Intelligence.

Prerequisites

Two Cognos users (employees of Great Outdoors Co., Ltd) are described in this sample case, as shown in Table 6-4. The Cognos users' information is stored in the security provider. The security provider offers security for Cognos Business Intelligence. We use OpenLDAP as our security provider in this sample case. The IT department of the Great Outdoors Co., Ltd wants to provide confidential data only to the correct people. The IT staff must apply the security to the objects in the Cognos Business Intelligence portal. We show the steps to apply the security to the executive dashboard in the following sections.

Table 6-4 Cognos users in this sample demonstration

User name	Role
Administrator	IT staff and administrator of Cognos
Adam	Executive of Great Outdoors Co., Ltd., who uses the executive dashboard

Operations

This section explains how to implement the sample solution in your environment and how to set the correct security. You need to follow these steps in order for the implementation:

1. Create the data source connection in Cognos Administration.
2. Create and publish the Twitter Sentiment package.
3. Create the report for the executive dashboard.
4. Apply security.

Follow these steps to create the data source connection in the Cognos Administration:

1. Log in to the Cognos Business Intelligence portal site with the *administrator* role. For this demonstration, we set the user ID to `admin` and the password to `ibm1ibm`.
2. Click **Administer IBM Cognos Content**.
3. Click the **Configuration** tab.
4. Click **New Data Source**. Follow these steps:
   a. For name, type `GS_DB - Big SQL`. Click **Next**.
   b. Select **JDBC** for type and click **Next**.

5. Set the following fields to these values (Figure 6-20):
   - Type: **IBM InfoSphere BigInsights (BIG SQL)**
   - JDBC URL: `jdbc:db2://mn01:32051/bigsql`
   - Driver class name: `com.ibm.db2.jcc.DB2Driver`
   - Select **Password** to require a password at sign-on.

   You will use `bigsql` for the user ID and `ibm1ibm` for the password.

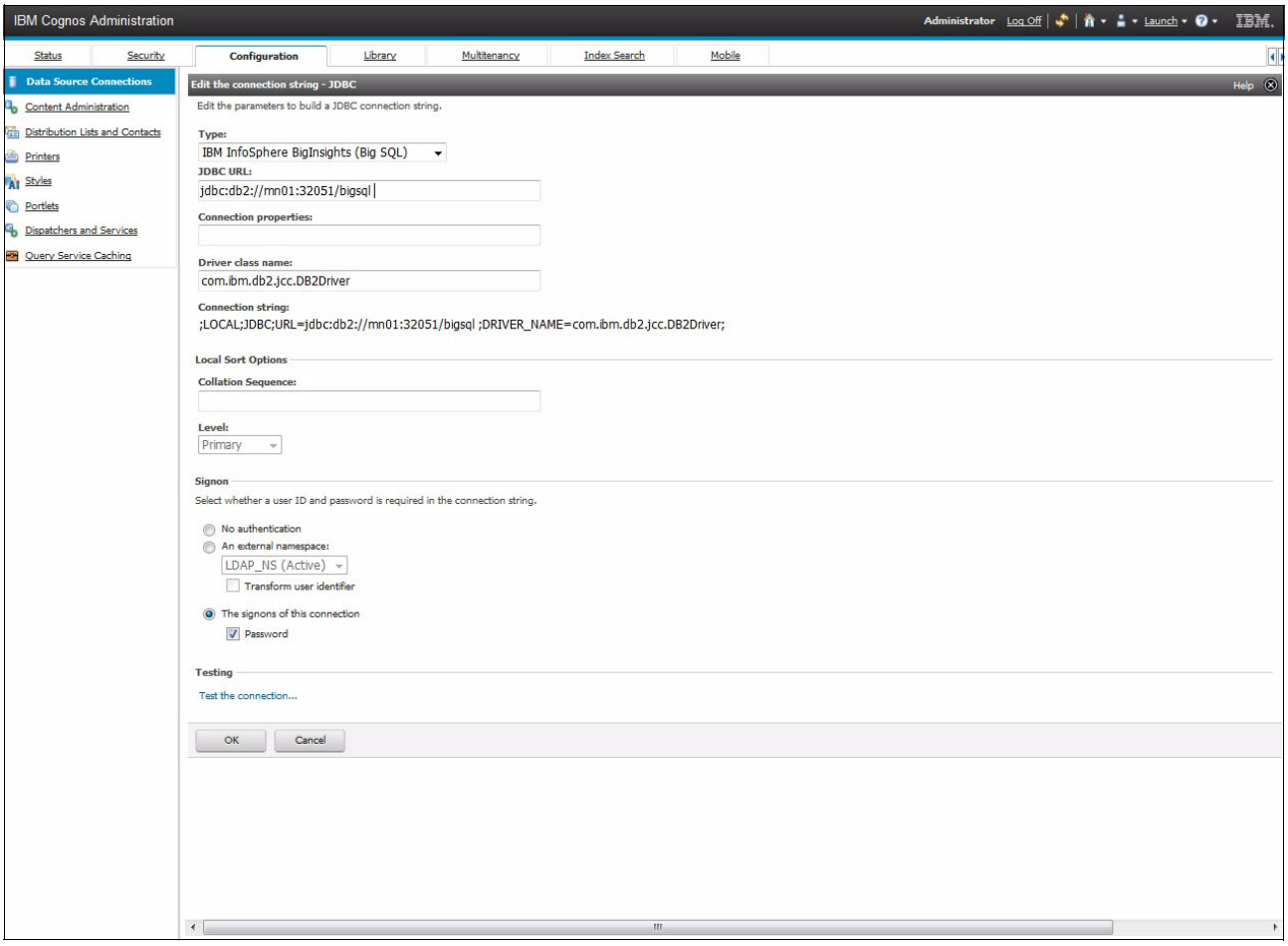


Figure 6-20 Sample result: Data source definition for Big SQL

6. Click **Next**.
7. Click **Finish**.

Follow these steps to create and publish the Twitter Sentiment package. In these steps, we use Framework Manager, which is installed in the Windows client.

1. Click **Start** → **All Programs** → **IBM Cognos** → **Framework Manager**.
2. Click **Open a project** on the project page.
3. Open the product Cognos sample: **great_outdoors_warehouse.cpf**.
4. Log in with a user ID that can publish the package. For this demonstration, we set the user ID to `admin` and the password to `ibm1ibm`.

5. Change the Data Source setting:
   - Click **go_data_warehouse** → **Data Sources** → **go_data_warehouse**.
   - Open the **Properties** tab in the middle of the window (Figure 6-21).
   - For the Content Manager Data Source, type `GS_DB - BigSQL`.

Note: In this case, we import the go_data_warehouse on DB2 BLU into BigInsights, which is named GS_DB - BigSQL. We set the data source as the data source definition that is named `GS_DB - BigSQL` (Figure 6-21).

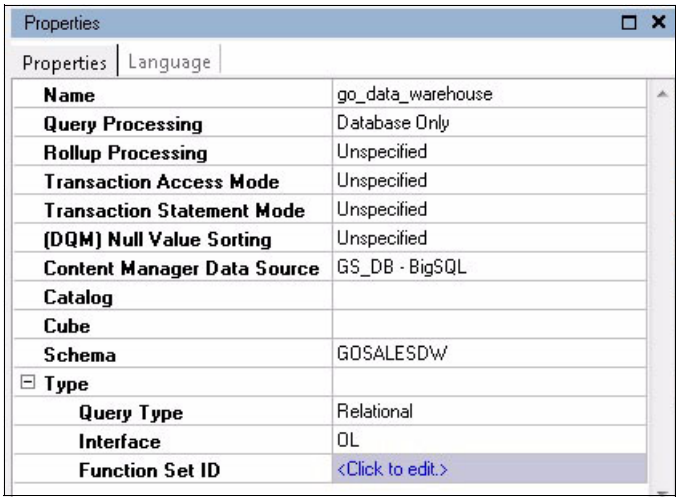


Figure 6-21 Sample configuration: Properties for the data source in Framework Manager

6. Click **go_data_warehouse** → **Packages**.
7. Right-click **Package**.
8. Select **Create** → **Create Package**.
9. For the name, type `Twitter Sentiment`.
10. Click **Next**.
11. Click **Next**.
12. Click **Next**.
13. Click **Finish**.

14. Click **Yes** to answer "`Would you like to open the Publish Package wizard?`"

Note: If you receive the message that is shown in Figure 6-22, you successfully created the package. Do not worry about the message.

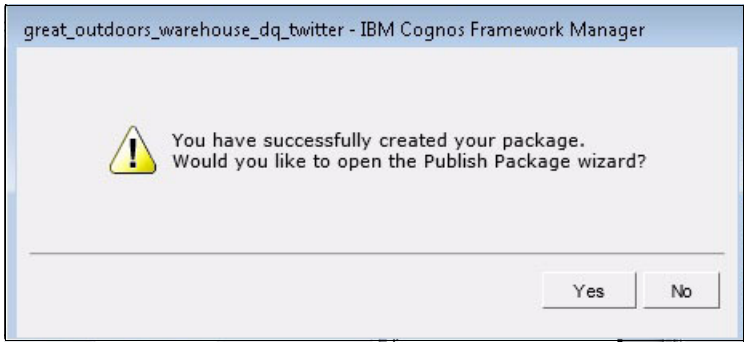


Figure 6-22 Sample view: Message appears during the package creation

15. Click **Yes**.

16. The Publish Wizard opens.

Note: If you want to change the publish location for Cognos Connection (web portal), select the appropriate folder from the folder location in the Content Store (Figure 6-23).

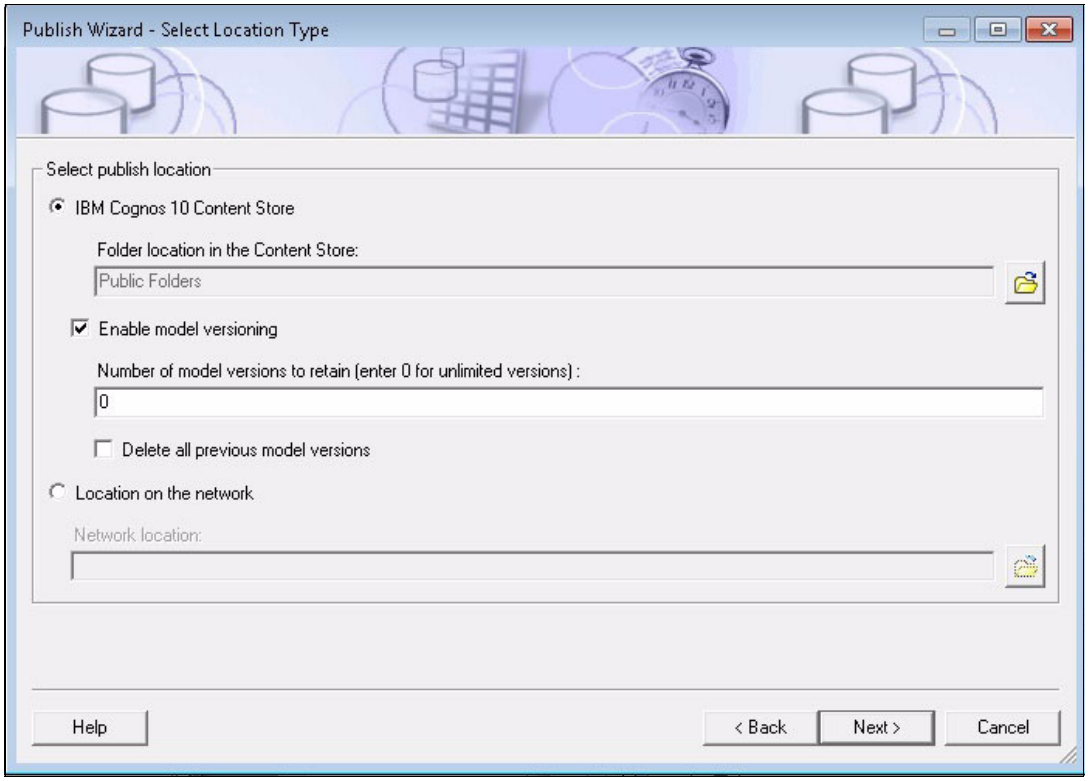


Figure 6-23 Sample configuration: Publish Wizard select publish location

17. Click **Next**.

18. Click **Next**.
19. Click **Publish**.
20. Click **Finish**.
21. Click **Close**.
22. Follow these steps to publish the package to the IBM Cognos Connection from the Framework Manager:
    - Click **File** → **Save as**.
    - Save file as `great_outdoors_warehouse_dq_twitter`.

Create the report for the executive dashboard. In this section, we create the workspace reports. After we create the workspace reports, the administrator can create the dashboard for the executives by picking up report items (charts, cross tabs, lists, and so on) from multiple workspace reports. These reports can be created from multiple packages and multiple data sources. Follow these steps:

1. Create the workspace report 1 from the Twitter Sentiment package. This workspace report contains two items (Figure 6-24):
    - Chart for Chart 4 that is named as Revenue Twitter Sentiment by Product
    - List

Note: In this sample case, we used a normal cube to show how to implement the solution in your environment in the simplest way. For big data analysis, such as Social Network Service, another configuration option is available. You can use Cognos Dynamic Cubes instead of normal Cognos cubes.

This option offers the capability to introduce a performance layer in the Cognos query stack to allow low-latency and high-performance online analytical processing (OLAP) analytics over large relational data warehouses if your environment has sufficient resources. For more information about Cognos Dynamic Cubes, see “Cognos Dynamic Cubes advantage” on page 260.

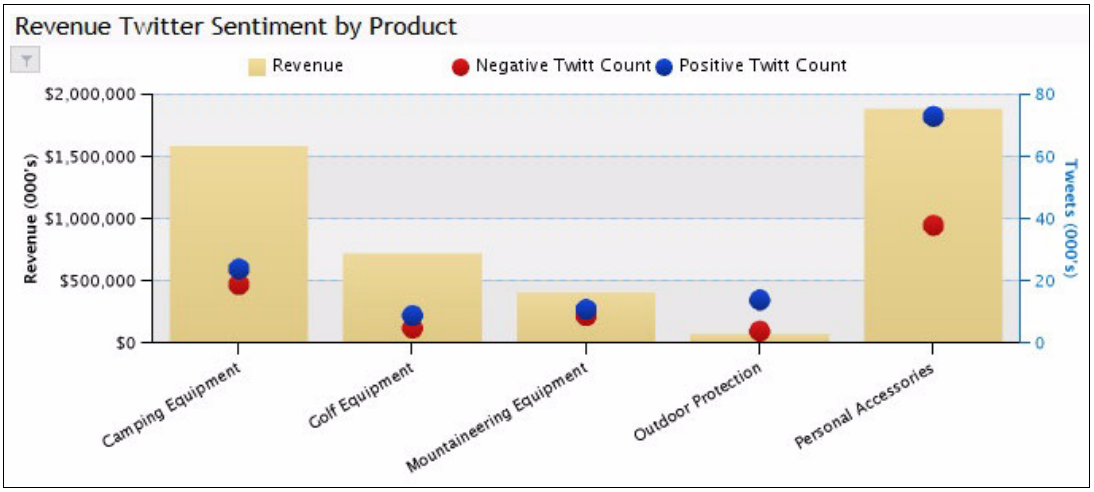


Figure 6-24 Sample result: Reports that are created in this section

2. Log in to the Cognos Business Intelligence portal site with the *administrator* role. For this demonstration, we set the user ID to `admin` and the password to `ibm1ibm`.

Note: You can log in with any other user ID that is eligible to use Cognos Workspace Advanced.

3. Click **IBM Cognos Content**.
4. Click **Launch** → **Cognos Workspace Advanced** (Figure 6-25).

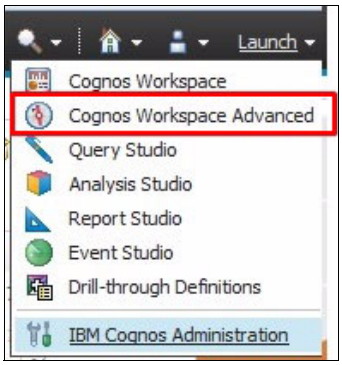


Figure 6-25 Sample view: Component list after you click Launch

5. Click the **Twitter Sentiment** package in the Select Package window.
6. Click **Create New** (Figure 6-26).

Figure 6-26 Sample view: Home page of IBM Cognos Workspace Advanced

7. Select **List** to create Chart 4. Click **OK**.

8. The List object opens in your workspace (Figure 6-27).

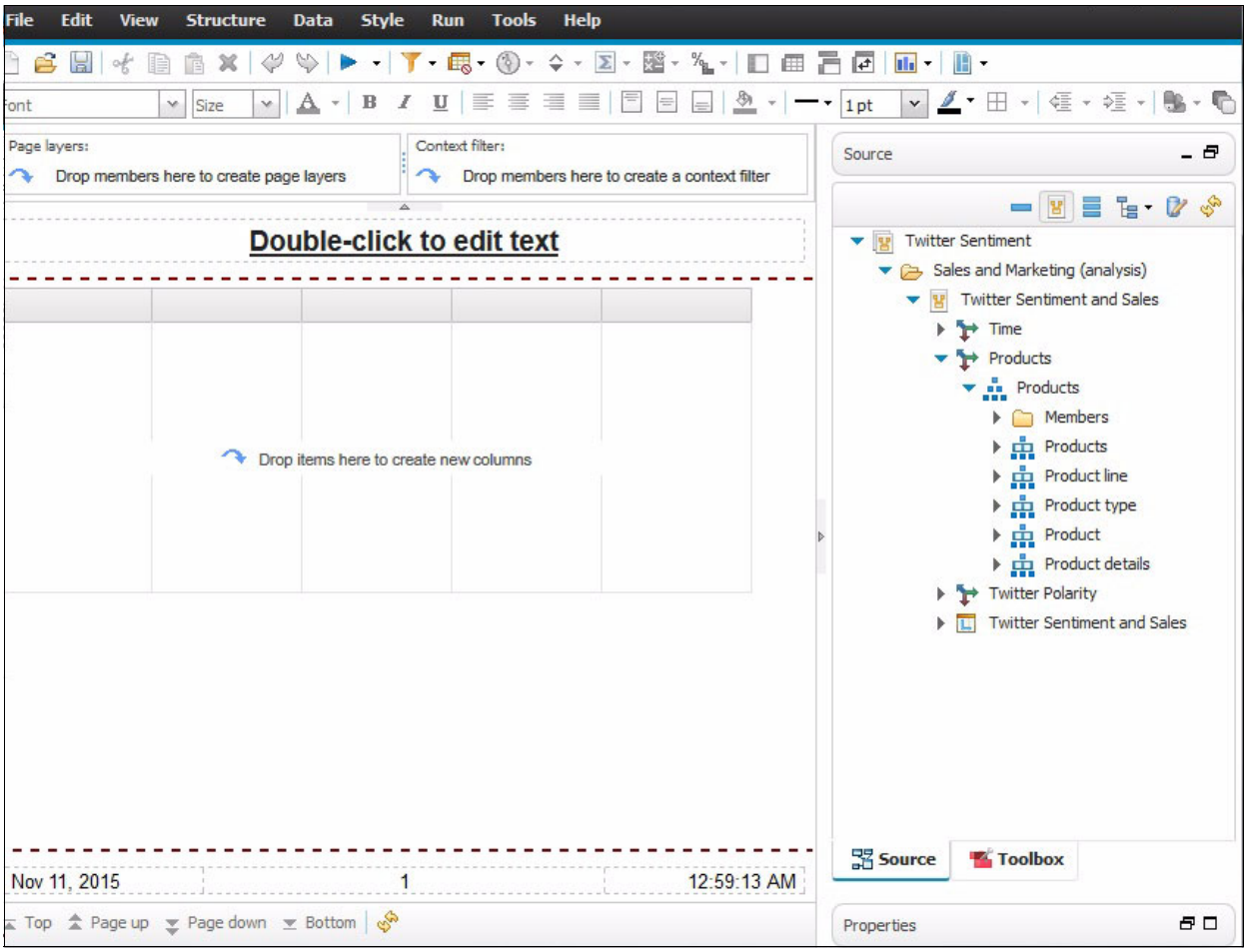


Figure 6-27 Sample result: List object is in the Cognos Workspace Advanced

9. Check whether you selected the View Metadata Tree icon (Figure 6-28). If you did not select it, select the **View Metadata Tree** icon.

Figure 6-28 Sample Configuration: Selecting the View Metadata Tree icon in the Source Tab

Important: Do not pick up items not from the Member Tree but from the Metadata Tree in this sample case. Pick up items from the Metadata Tree so that the user can drill down by clicking charts in the reports**.**

Table 6-5 shows the differences between the metadata tree and the member tree.

Table 6-5 Concept of items for the report

Type	Concept	Expected usage	Example
Metadata tree	Group (metadata)	Set metadata so that it allows the user to drill down	Set Personal Accessories means to enable Cognos Business Intelligence to display Personal Accessories and its succeeded members (for example, Eyewear).
Member tree	Data (actual member)	To display the data itself	Set Personal Accessories means allowing Cognos Business Intelligence to display only Personal Accessories.

10. Follow these steps to create the chart that is shown in Figure 6-29 on page 224:
    a. Click **Twitter Sentiment** → **Sales and Marketing (analysis)** → **Twitter Sentiment and Sales** → **Product** → **Product** → **Product Line** in the Source tab.
    b. Drag and drop **Product Line** in the List.
    c. Click **Twitter Sentiment** → **Sales and Marketing (analysis)** → **Twitter Sentiment and Sales** → **Twitter Polarity** → **POLARITY** → **POLARITY**.
    d. Drag and drop **POLARITY** in the List.
    e. Click **Twitter Sentiment** → **Sales and Marketing (analysis)** → **Twitter Sentiment and Sales** → **Twitter Sentiment and Sales** → **Revenue**.
    f. Drag and drop **Revenue** in the List.
    g. Click **Twitter Sentiment** → **Sales and Marketing (analysis)** → **Twitter Sentiment and Sales** → **Twitter Sentiment and Sales** → **Twitter Count**.
    h. Drag and drop **Twitter Count** in the List.

11. Your list looks similar to the list that is shown in Figure 6-29.

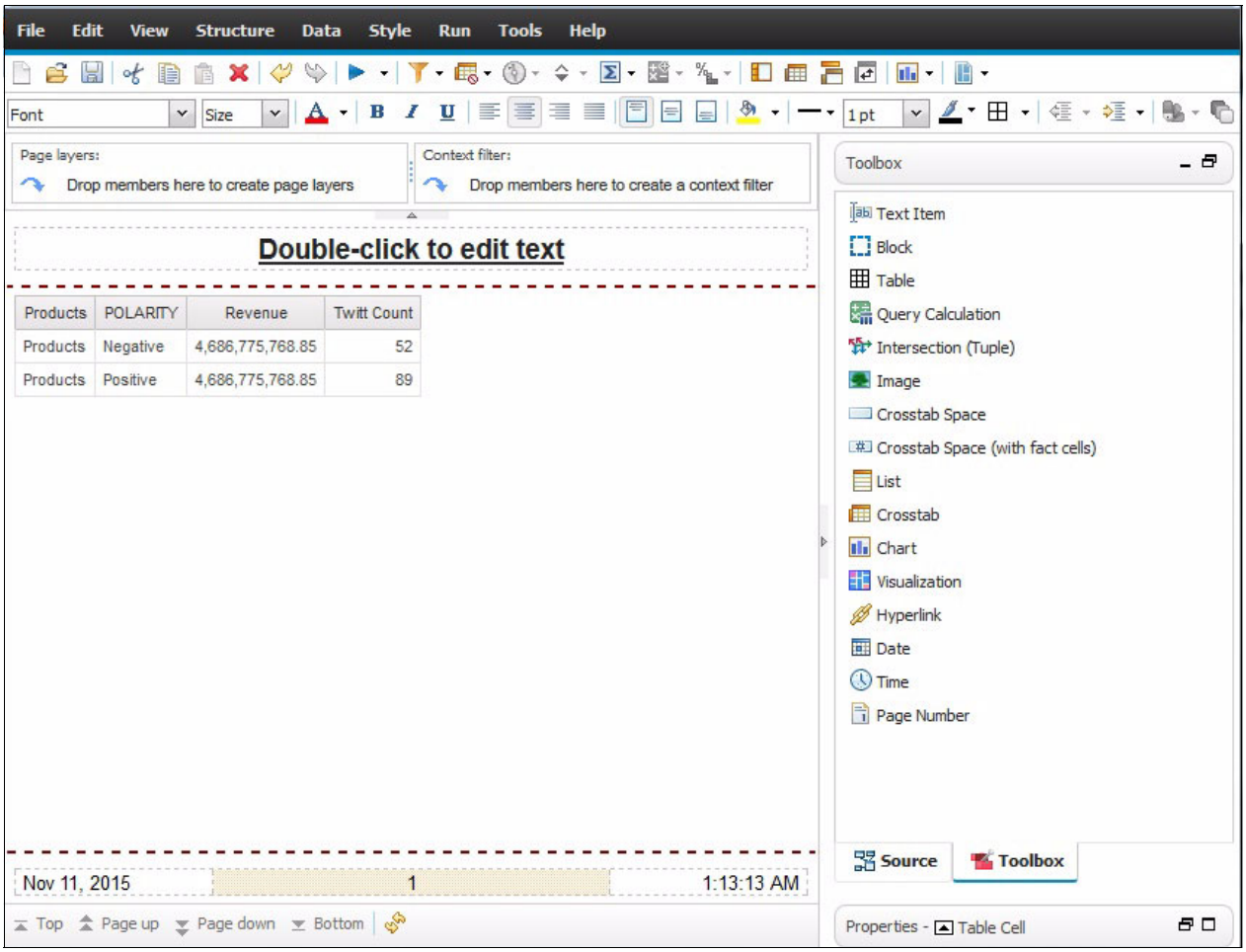


Figure 6-29 Sample result: Created list object in the Cognos Workspace Advanced

12. Follow these steps to create a combination chart for Chart 4:
    a. Open the **ToolBox** tab next to the Source tab.
    b. Drag and drop **Chart** to the workspace.
    c. Select **Combination** from the left side (Figure 6-30).
    d. Select **Primary Axis clustered Bar, Secondary Axis Clustered Line**. Click **OK**.

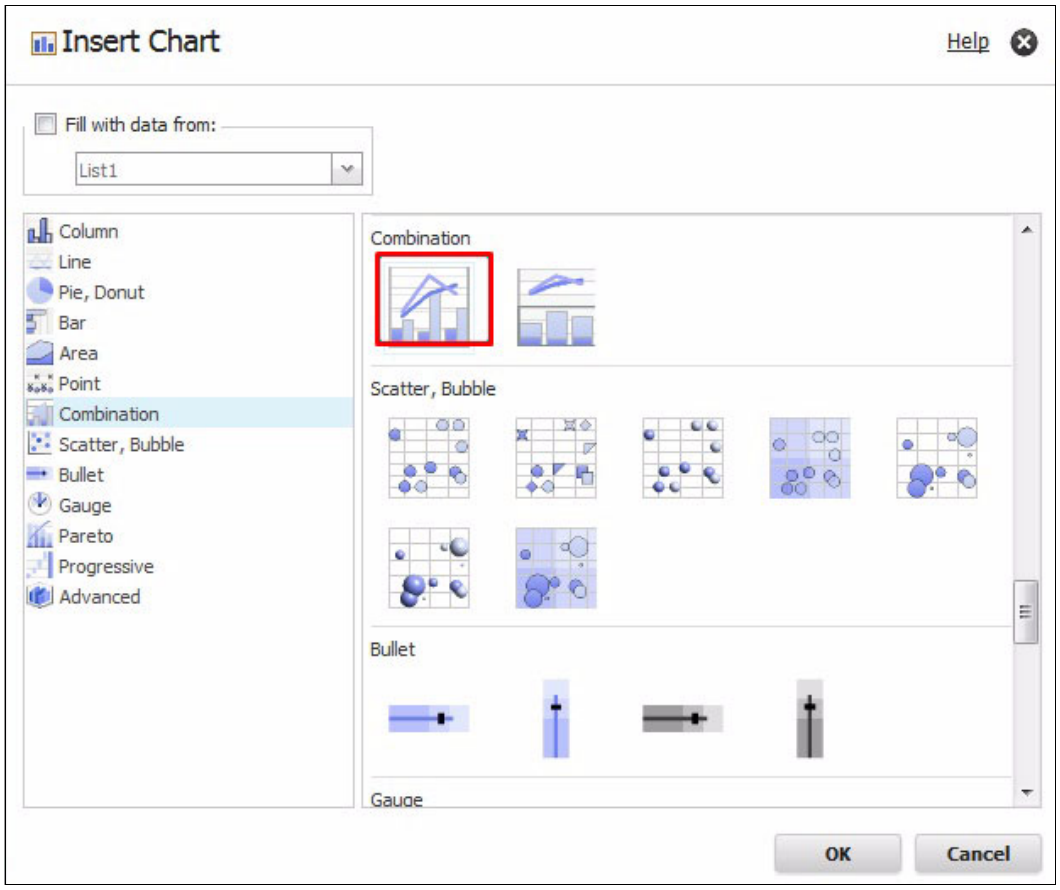


Figure 6-30 Sample configuration: Selecting a combination chart from the wizard

13. You now have a combination chart in your workspace, as shown in Figure 6-31.

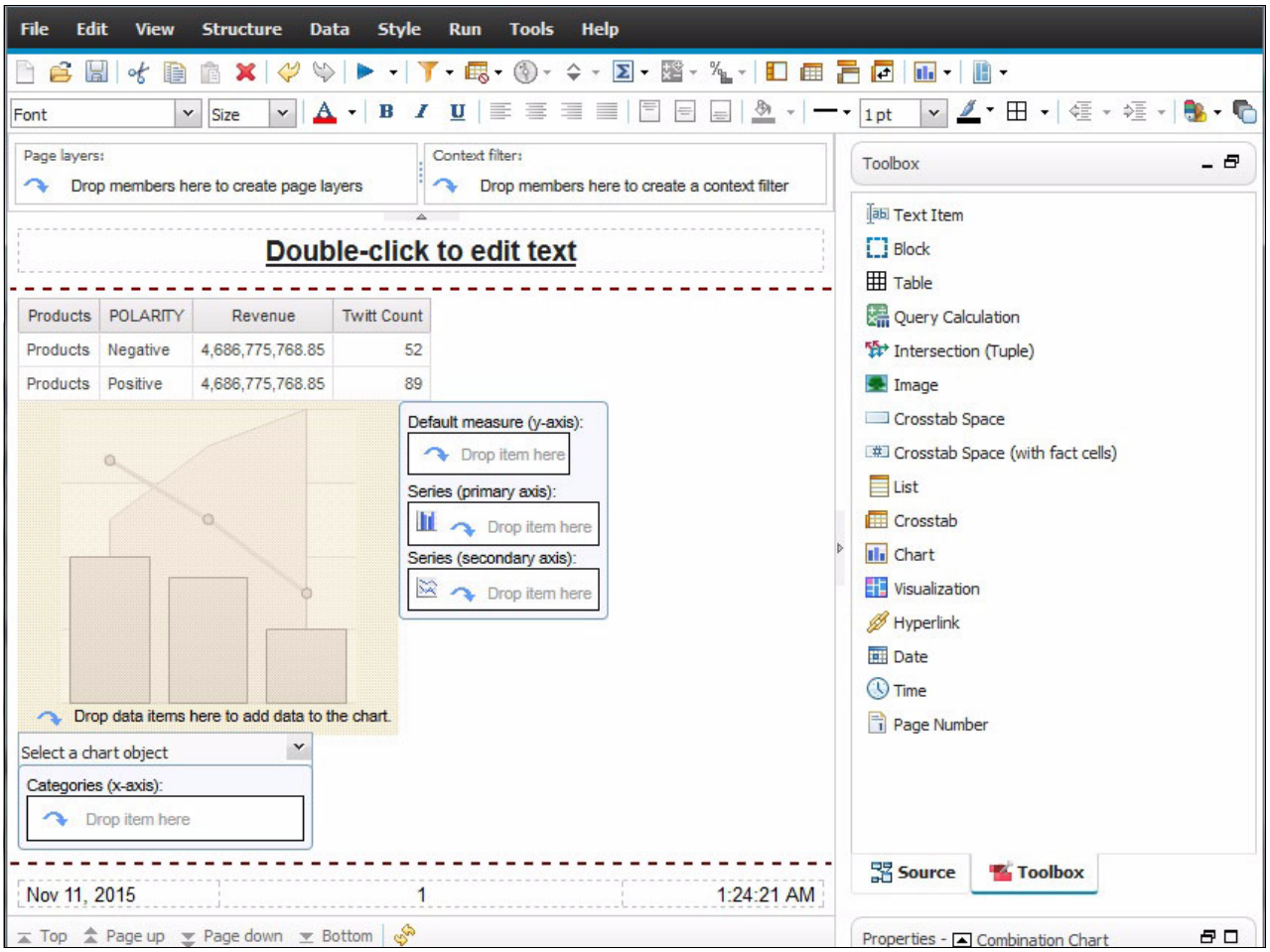


Figure 6-31 Sample result: Combination chart object in Cognos Workspace Advanced

14. Click **Twitter Sentiment** → **Sales and Marketing (analysis)** → **Twitter Sentiment and Sales** → **Twitter Sentiment and Sales** → **Revenue** in the Source tab. Follow these steps:
    a. Drag and drop **Revenue** in Series (primary axis) in the combination chart.
    b. Click **Revenue** inside of Series (primary axis) in chart
    c. Open **Properties** for Chart Node Member.

> **Note:** The Properties list box is on the lower-right side.

    d. Click in the data format.
    e. Select the following information in the Data Format window:
        - Format type: **Currency**
        - No. of Decimal Places: **0**
        - Scale: **3**
        - Use Thousands Separator: **Yes**
    f. Click **OK**.

Note: If you click a blank cell, an arrow appears to the right. When you click that arrow, you find a list of choices. You can select items from that list, as shown in Figure 6-32.

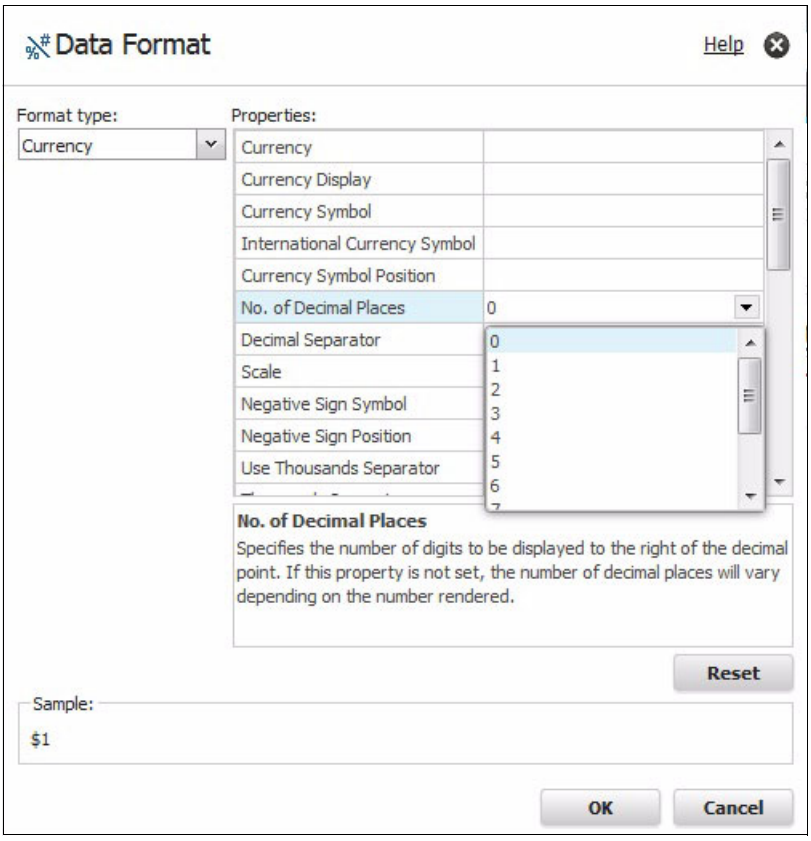


Figure 6-32 Sample configuration: How to set the data format

15. Set **Revenue (in thousands)** for the Data items name and for the Data item label in Properties.
16. Click **Twitter Sentiment** → **Sales and Marketing (analysis)** → **Twitter Sentiment and Sales** → **Twitter Polarity** → **Polarity** → **Polarity**. Follow these steps:
    a. Drag and drop **Polarity** in Series (secondary axis) in the combination chart.
    b. Click **Polarity** inside Series (secondary axis) in the chart.
    c. Open **Properties** for Chart Node Member.
    d. Set **Sentiment Polarity** for the data item name and data item label in Properties.
    e. Click **OK**.

17. Click **Twitter Sentiment** → **Sales and Marketing (analysis)** → **Twitter Sentiment and Sales** → **Twitter Sentiment and Sales** → **Twitter Count**. Follow these steps:
    a. Drag and drop **Twitter Count** under Sentiment Polarity in Series (secondary axis).
    b. Open **Properties** for Chart Node Member.
    c. Set **Tweets (000's)** for the data item name and for the data item label in Properties. Click **OK**. See Figure 6-33.

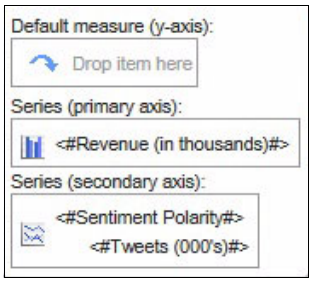


Figure 6-33 Sample configuration: How to nest a data item

18. Click **Twitter Sentiment** → **Sales and Marketing (analysis)** → **Twitter Sentiment and Sales** → **Twitter Sentiment and Sales** → **Products** → **Products** → **Product Line**. Follow these steps:
    a. Drag and drop **Product Line** to Categories (x- axis).
    b. Click **Chart**.
    c. Open **Properties** for Combination Chart.
    d. Set **Revenue** Twitter Sentiment by Product for Name. Click **OK**.
    e. Click **File** → **Save as**.
    f. Save workspace report as **Revenue and Sentiment by Product under Public Folders** → **CognosBigInsightsBLU - Demo** → **Report**.
    g. Click **Save**.
    h. Now, you can create report 1.

We create the workspace report 2 from the GO Sales (analytics) package. This workspace report contains three items:

- Visualization that is named Area for Chart 1, which is named Revenue Trend by Product (Figure 6-34)
- Crosstab for Chart 2, which is named Revenue by Product and Year (Figure 6-35)
- Visualization that is named Dynamic Map for Chart 3, which is named Gross Profit by Country (Figure 6-36 on page 230)

Figure 6-34 shows the Revenue Trend by Product report.

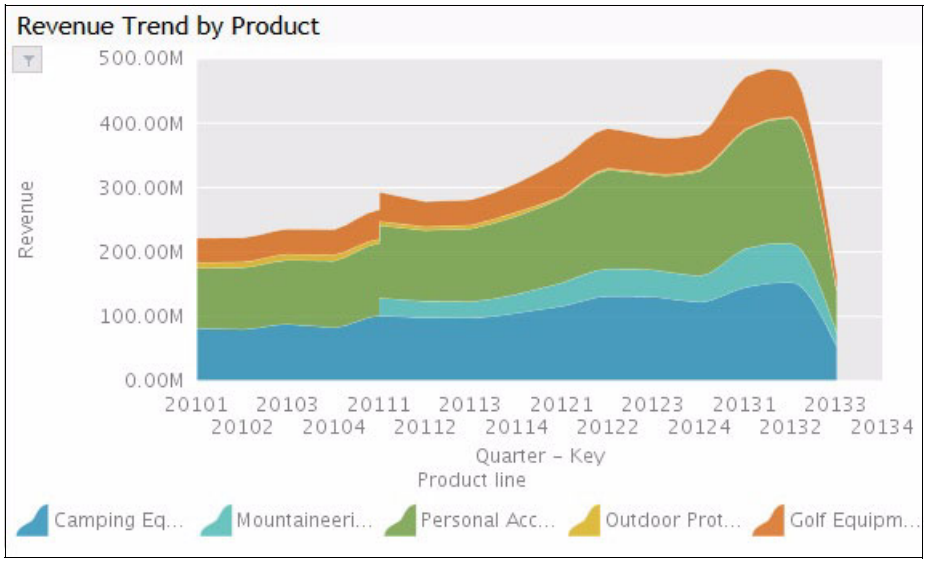


Figure 6-34 Sample Result: Chart 1 - Revenue Trend by Product

Figure 6-35 shows the Revenue by Product and Year report.

Revenue by Product and Year

Revenue	2010	2011	2012	2013
Camping Equipment	332,986,338.06	402,757,573.17	500,382,422.83	352,910,329.97
Mountaineering Equipment		107,099,659.94	161,039,823.26	141,520,649.70
Personal Accessories	391,647,093.61	456,323,355.90	594,009,408.42	443,693,449.85
Outdoor Protection	36,165,521.07	25,008,574.08	10,349,175.84	4,471,025.26

Figure 6-35 Sample Result: Chart 2 - Revenue by Product and Year

Figure 6-36 shows the Gross Profit by Country report.

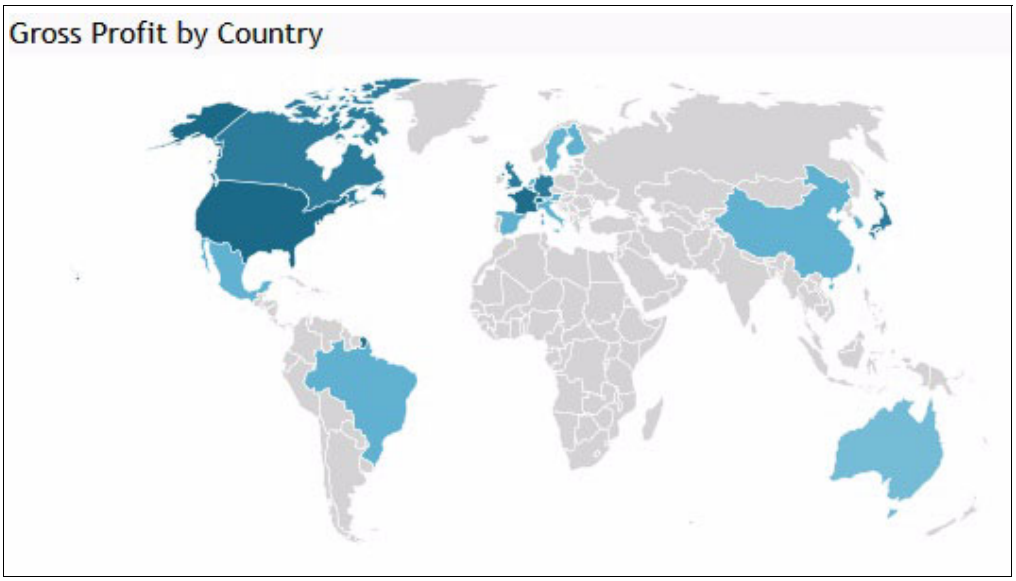


Figure 6-36 Sample Result: Chart 3 - Gross Profit by Country

Follow these steps:

1. Click **File** → **New**.
2. Select **Crosstab**.
3. Crosstab shows in your workspace.
4. Click **GO Sales (analytics)** → **Sales (analytics)** → **Sales** → **Revenue** in the Source tab.

5. Drag and drop **Revenue** to the Measures section in Crosstab (Figure 6-37).

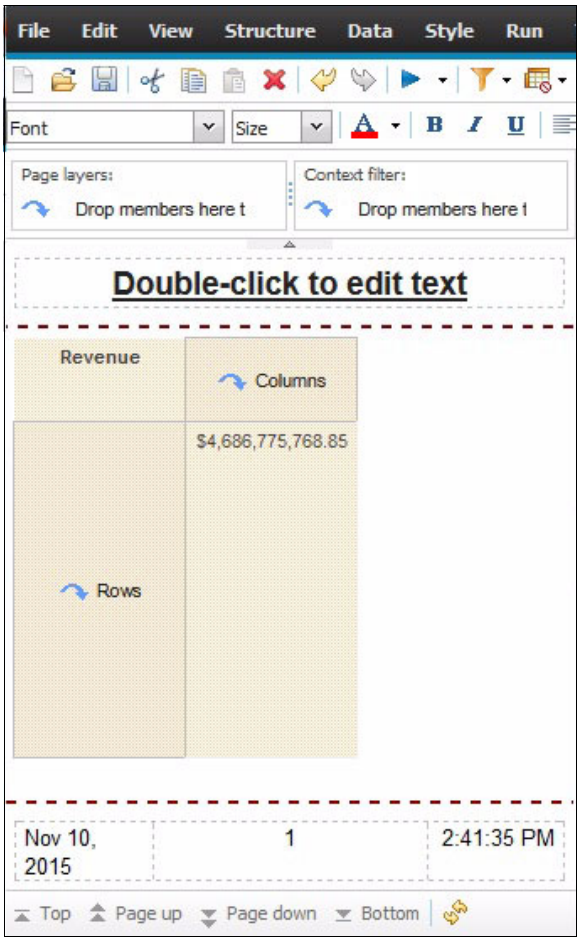
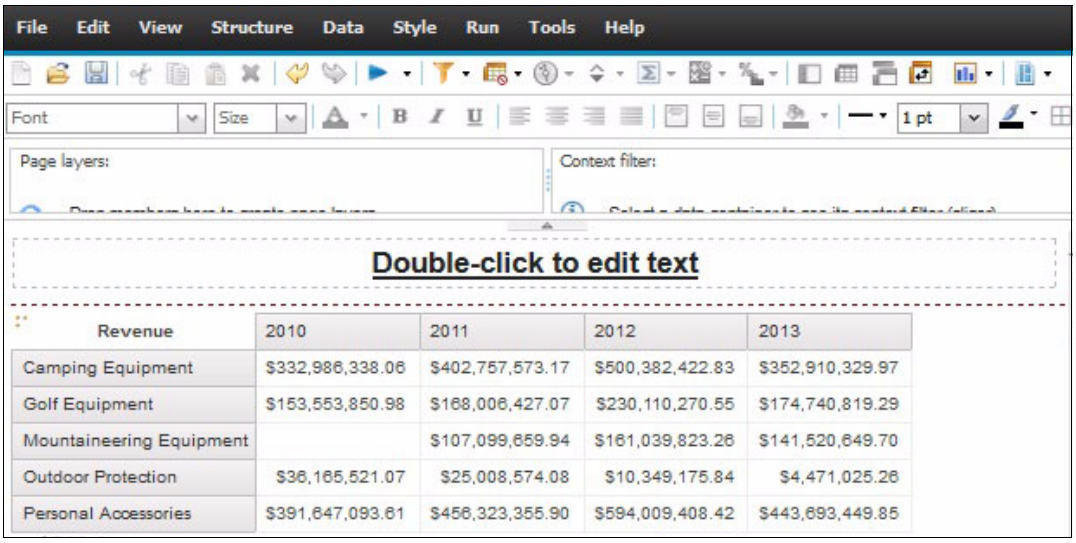


Figure 6-37 Crosstab object in Cognos Workspace Advanced

Follow these steps to create the cross tab for Chart 2:

1. Click **GO Sales (analytics)** → **Sales (analytics)** → **Products** → **Products** → **Product line**.
2. Drag and drop **Product line** to **Rows**.
3. Click **GO Sales (analytics)** → **Sales (analytics)** → **Time** → **Time** → **Year**.
4. Drag and drop **Year** to **Columns**.

5. The crosstab for Chart 2 is created (Figure 6-38).



Revenue	2010	2011	2012	2013
Camping Equipment	$332,986,338.06	$402,757,573.17	$500,382,422.83	$352,910,329.97
Golf Equipment	$153,553,850.98	$168,006,427.07	$230,110,270.55	$174,740,819.29
Mountaineering Equipment		$107,099,659.94	$161,039,823.26	$141,520,649.70
Outdoor Protection	$36,165,521.07	$25,008,574.08	$10,349,175.84	$4,471,025.26
Personal Accessories	$391,647,093.61	$456,323,355.90	$594,009,408.42	$443,693,449.85

Figure 6-38 Sample result: Created Crosstab object in Cognos Workspace Advanced

Follow these steps to create the Area Chart for Chart 1:

1. Open the **Toolbox** tab on the lower-right side.
2. Drag and drop **Visualization** to the workspace.
3. Select **Area** in the Visualization Gallery (Figure 6-39) and click **OK**.

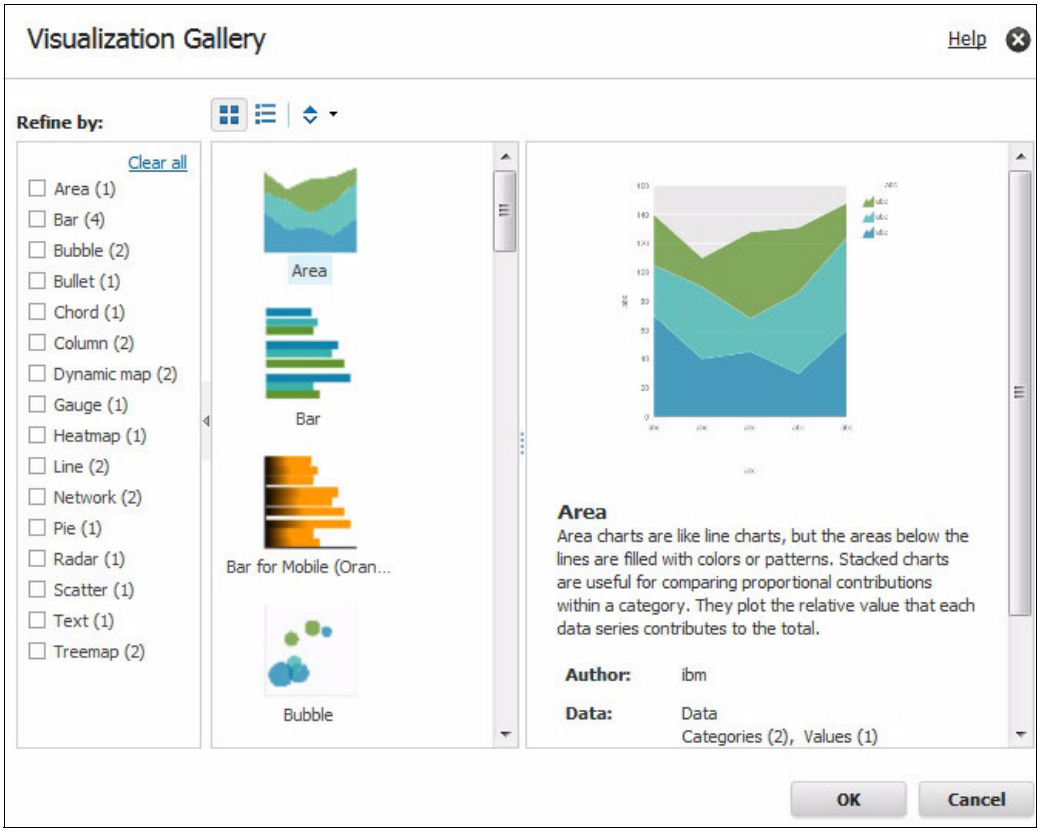


Figure 6-39 Sample configuration: Selecting Area from the Visualization Gallery

4. The Visualization Area chart opens in your workspace (Figure 6-40).

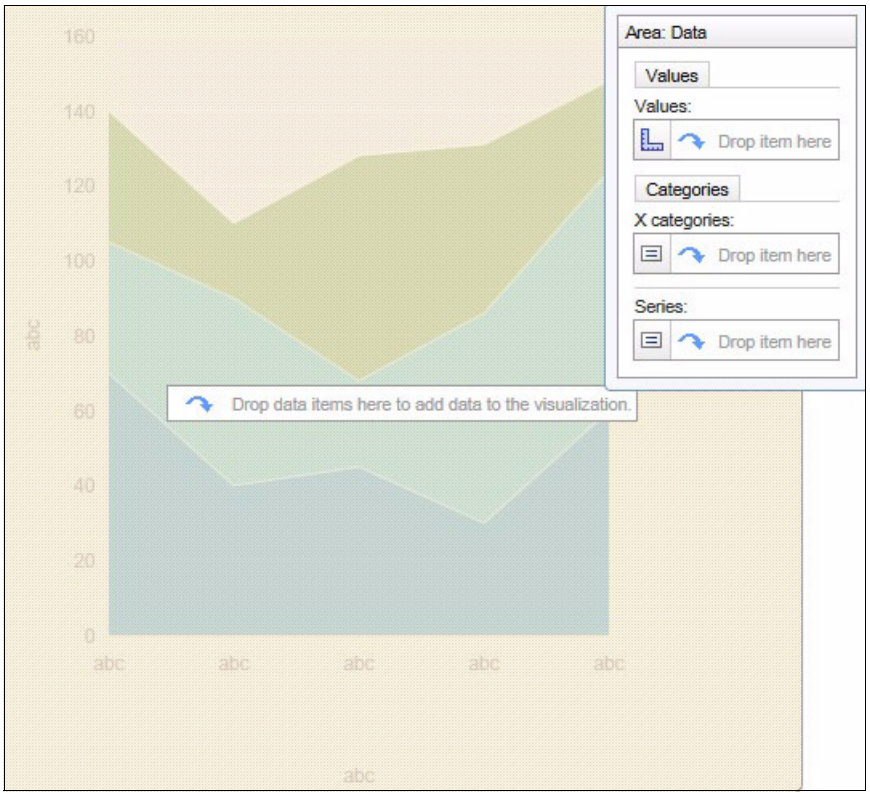


Figure 6-40 Sample result: Visualization Area chart

5. Click **GO Sales (analytics)** → **Sales (analytics)** → **Sales** → **Revenue** in the Source tab.
6. Drag and drop **Revenue** to Values.
7. Click **GO Sales (analytics)** → **Sales (analytics)** → **Time** → **Time** → **Quarter** → **Quarter key**.
8. Drag and drop **Quarter key** to X categories.
9. Click **GO Sales (analytics)** → **Sales (analytics)** → **Products** → **Products** → **Product Line**.
10. Drag and drop **Product Line** to Series.

11. The Visualization Area chart for Chart 1 is created (Figure 6-41).

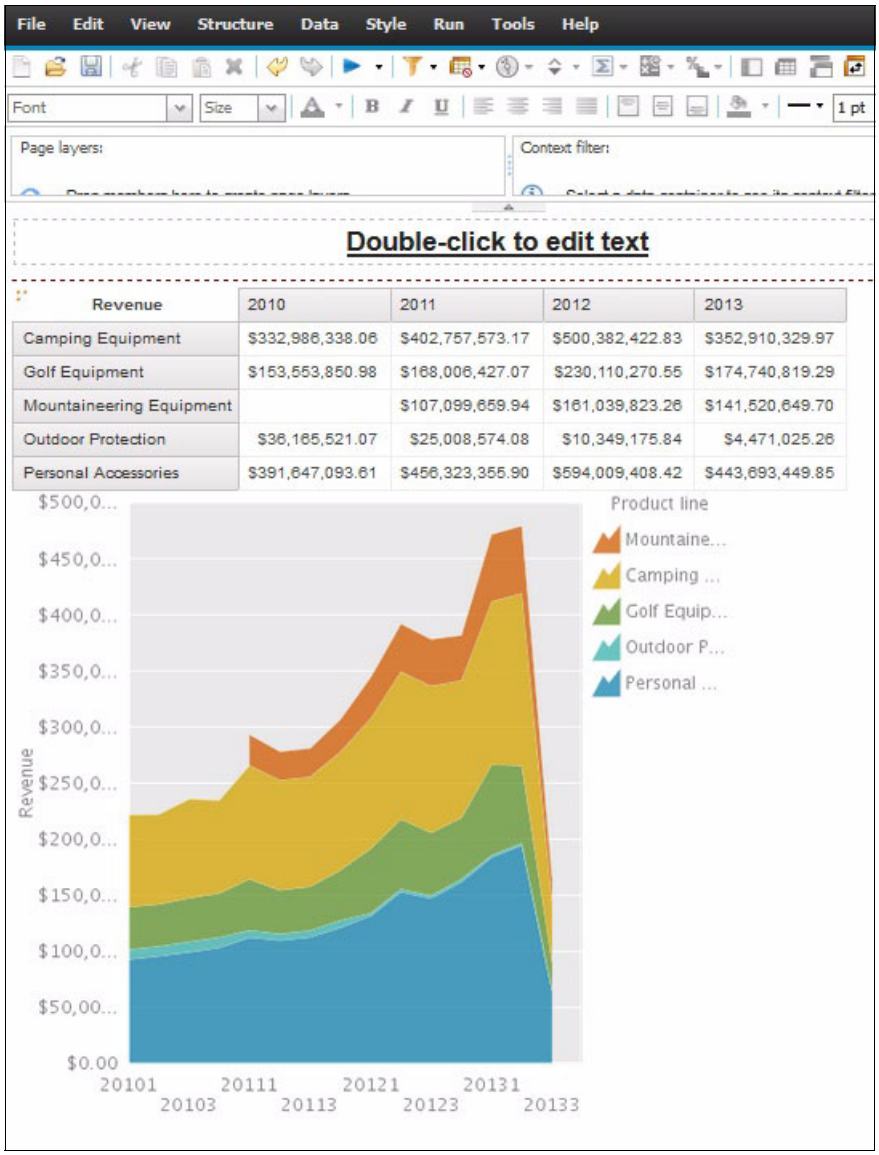


Figure 6-41 Sample result: Visualization Area chart in Cognos Workspace Advanced

Next, you create the Visualization Dynamic Map for Chart 3. Follow these steps:

1. Open the **Toolbox** tab in the lower-right corner.
2. Drag and drop **Visualization** to the workspace.
3. Select **Dynamic Map** in the Visualization Gallery (Figure 6-42). Click **OK**.

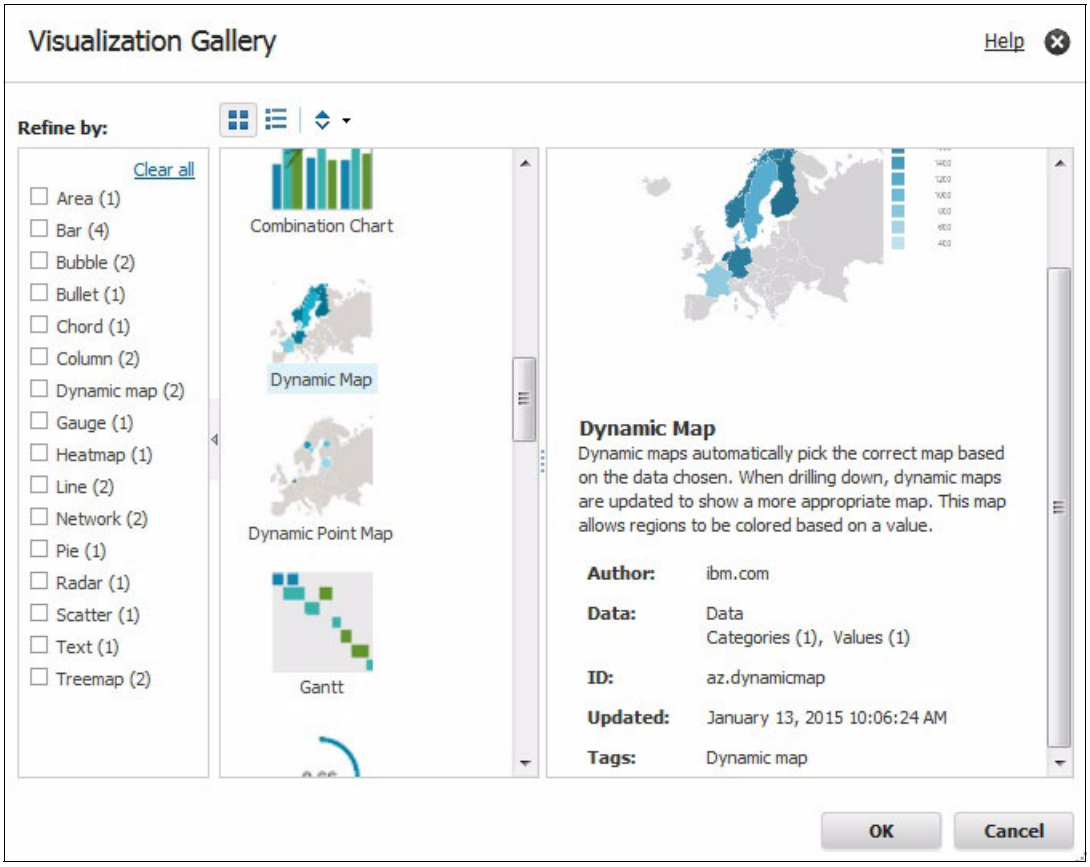


Figure 6-42 Sample configuration: Selecting Dynamic Map from the Visualization Gallery

4. The Visualization Dynamic Map opens in your workspace (Figure 6-43).

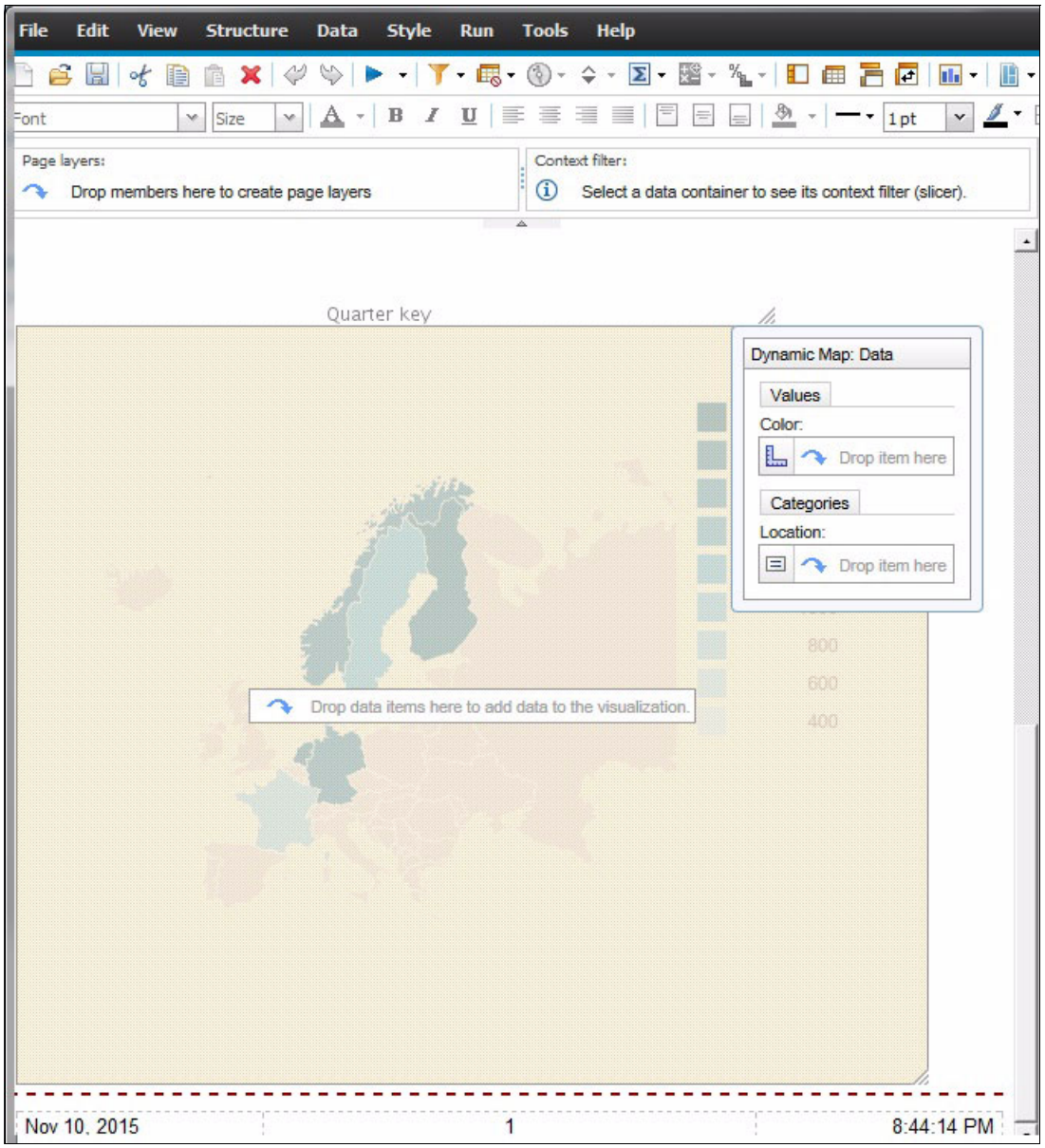


Figure 6-43 Visualization of Dynamic Map in Cognos Workspace Advanced

5. Open the **Source** tab.
6. Click **GO Sales (analytics)** → **Sales (analytics)** → **Sales** → **Gross Profit**.
7. Drag and drop **Gross Profit** to Color.
8. Click **GO Sales (analytics)** → **Sales (analytics)** → **Sales staff** → **Sales staff** → **Country** in the Source tab.
9. Drag and drop **Country** to Location.
10. The Visualization Dynamic Map for Chart 3 is created.
11. Save the workspace as **Report objects - BigInsights and BLU under Public Folders** → **CognosBigInsightsBLU - Demo**.
12. Click **File** → **Close**.

Create the dashboard for the executives. You can create the dashboard from report items that you created in the previous section (charts, crosstabs, lists, and so on). You can pick up the report items from multiple workspace reports and packages. Follow these steps to create the executive dashboard:

1. Click **Launch** → **Cognos workspace in Cognos Connection**.
2. Click **Create New** (Figure 6-44).

Figure 6-44 Sample view: Home of IBM Cognos Workspace

3. Click **Public Folders** → **CognosBiginsightsBLU-Demo** → **Report** in the Content tab on the right side.
4. Expand **Report objects - BigInsights and BLU** (Figure 6-45).

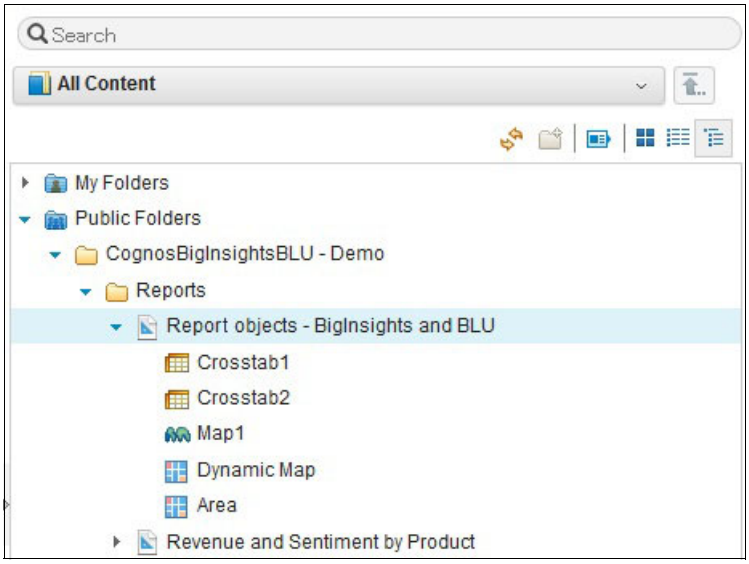


Figure 6-45 Sample view: Report objects - BigInsights and BLU

5. Drag and drop the following items to a blank area of the dashboard:
   - **Crosstab1**
   - **Dynamic Map**
   - **Area**

6. In this sample case, the Area is in the upper-left corner (Chart 1). Crosstab1 is in the upper-right corner (Chart 2). The Dynamic Map is in the lower-left corner (Chart 3). See Figure 6-46.

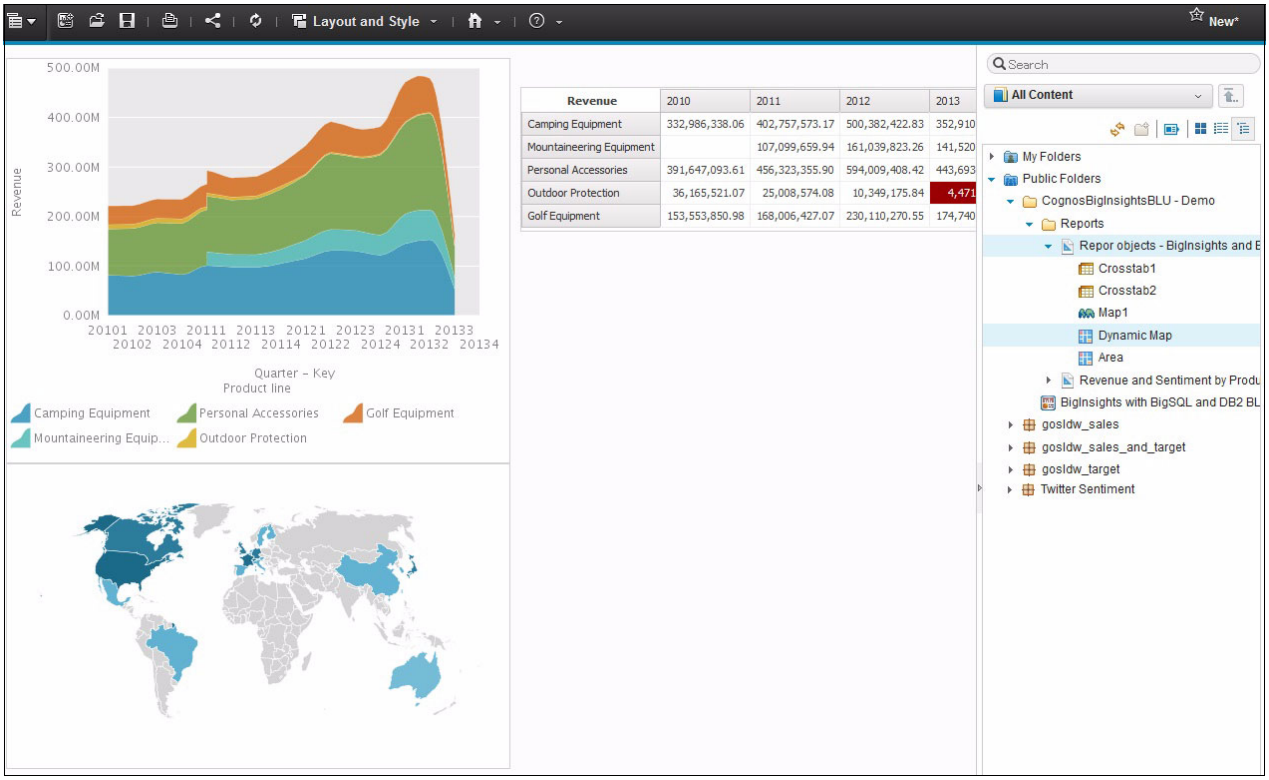


Figure 6-46 Sample result: Charts and crosstab are in the workspace dashboard

7. On Figure 6-46, in the Content tab on the right side, expand **Revenue and Sentiment by Product**.

8. Drag and drop **Revenue and Sentiment by Product** into a blank area of the dashboard (Figure 6-47).

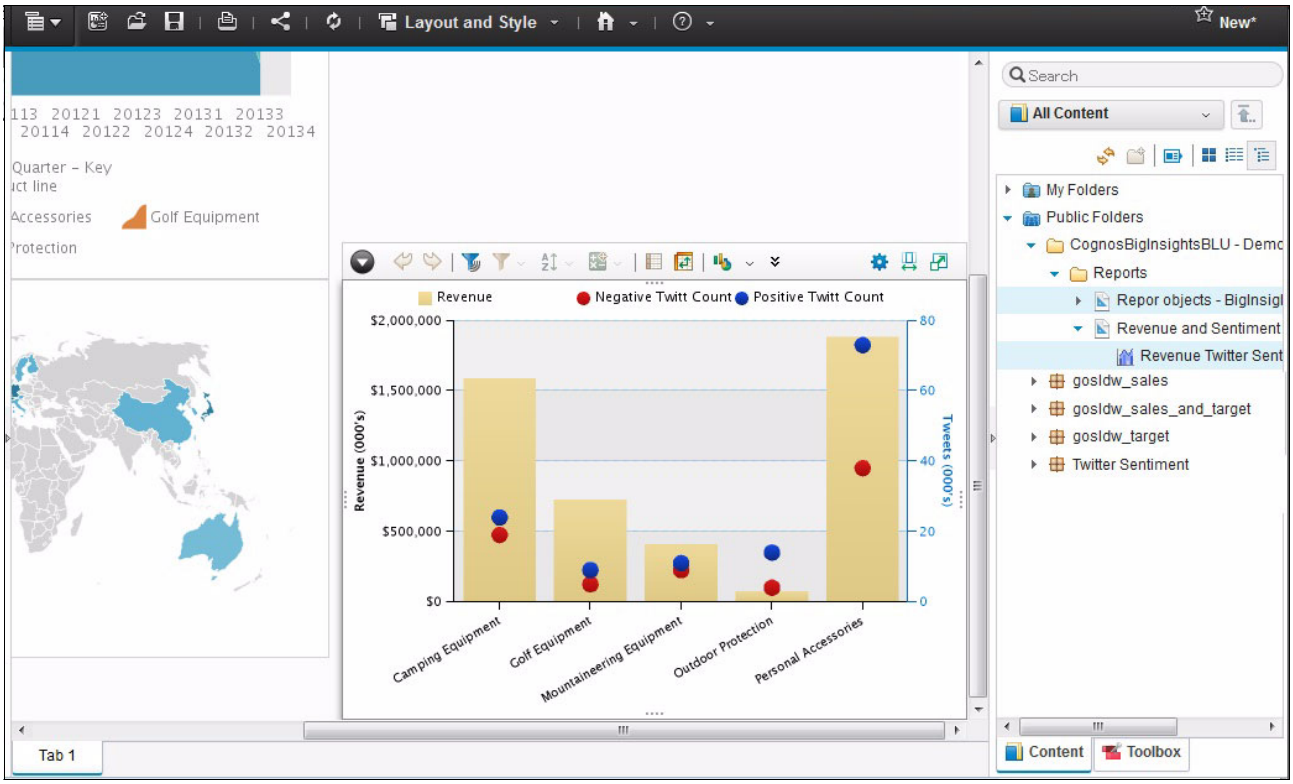


Figure 6-47 Combination Chart of sentiment analysis in the workspace dashboard

9. Click **Actions** → **Save as**.
10. Save the file as **BigInsights with Big SQL and DB2 BLU** under **Public Folders** → **CognosBigInsightsBLU - Demo**.
11. Click **Save**.

You must apply security to the workspace dashboard that you created so that only executives can see it. The dashboard contains confidential data, so it must be secure. First, create the executive security group. Then, apply the security to the executive workspace dashboard so that only a user that belongs to executive group can view it. Follow these steps:

1. Create the Executive group in the built-in Cognos namespace by clicking **Launch** → **IBM Cognos Administration** (Figure 6-48).

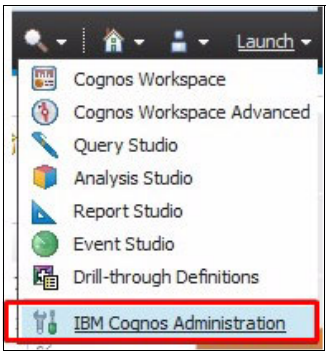


Figure 6-48 Sample view: Open IBM Cognos Administration

2. Click the **Security** tab.
3. Click the **Cognos** namespace, which is a built-in namespace (Figure 6-49).

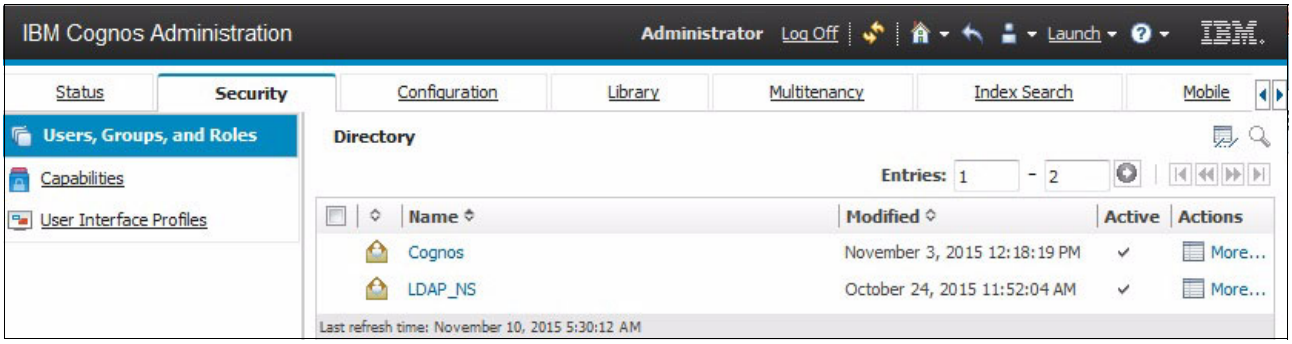


Figure 6-49 Sample view: Cognos namespace under the Security tab

4. Click the **New Group** icon (Figure 6-50).

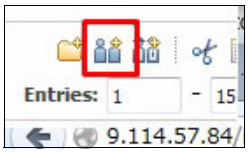


Figure 6-50 Sample view: New group in namespace

5. Name the new group `Executives`.
6. Click **Next.**
7. Click **Add** in the lower-right corner.
8. On the left, click **Show users in the list**.

9. Click the namespace that you created. For this demonstration, we created `LDAP_NS`.
10. Select **adam**. Click the yellow arrow to move adam to the Selected entries box.
11. Verify that adam shows in the Selected entries box (Figure 6-51).

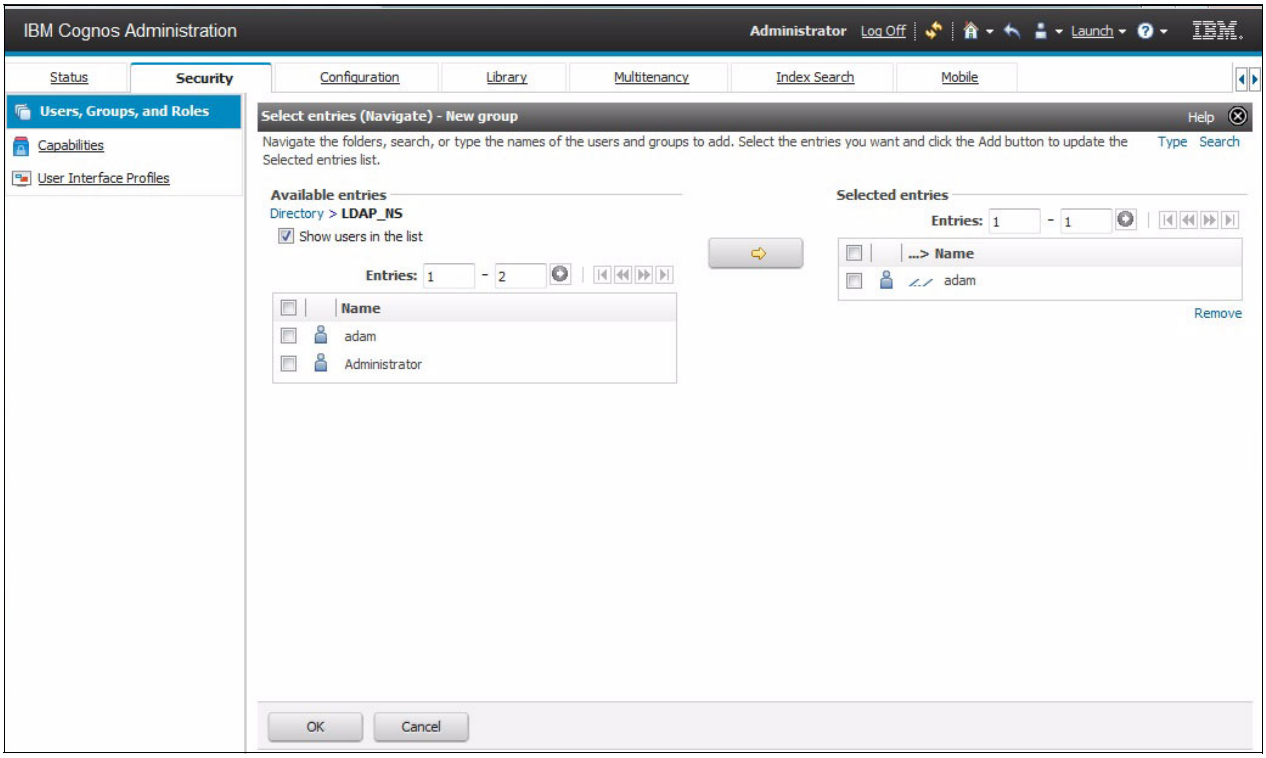


Figure 6-51 Sample configuration: Moving adam to the selected entries box

12. Click **OK**.
13. Click **Finish**.
14. Create the Executives group and set `adam` for its group.

Apply security to the workspace:

1. Navigate to **Launch** → **IBM Cognos Connection**.
2. Navigate to **Public Folders** → **CognosBigInsightsBLU - Demo**.
3. Click **More** for the workspace that is named BigInsights with BigSQL and DB2 BLU (Figure 6-52).

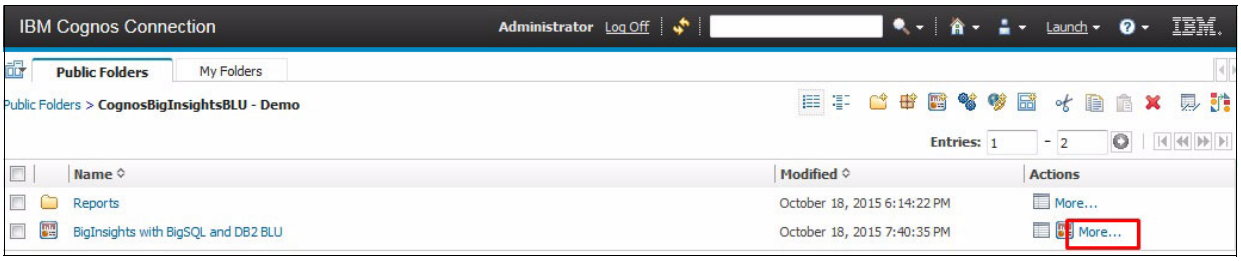


Figure 6-52 Sample configuration: Select More to open the properties for configuration

4. Click **Set properties**.
5. Click the **Permission** tab.

6. Select **Override the access permissions acquired from the parent entry**.
7. Click **Add**.
8. Click **Cognos** → **Executives**.
9. Select **Executives**.
10. Click the yellow arrow (Figure 6-53).

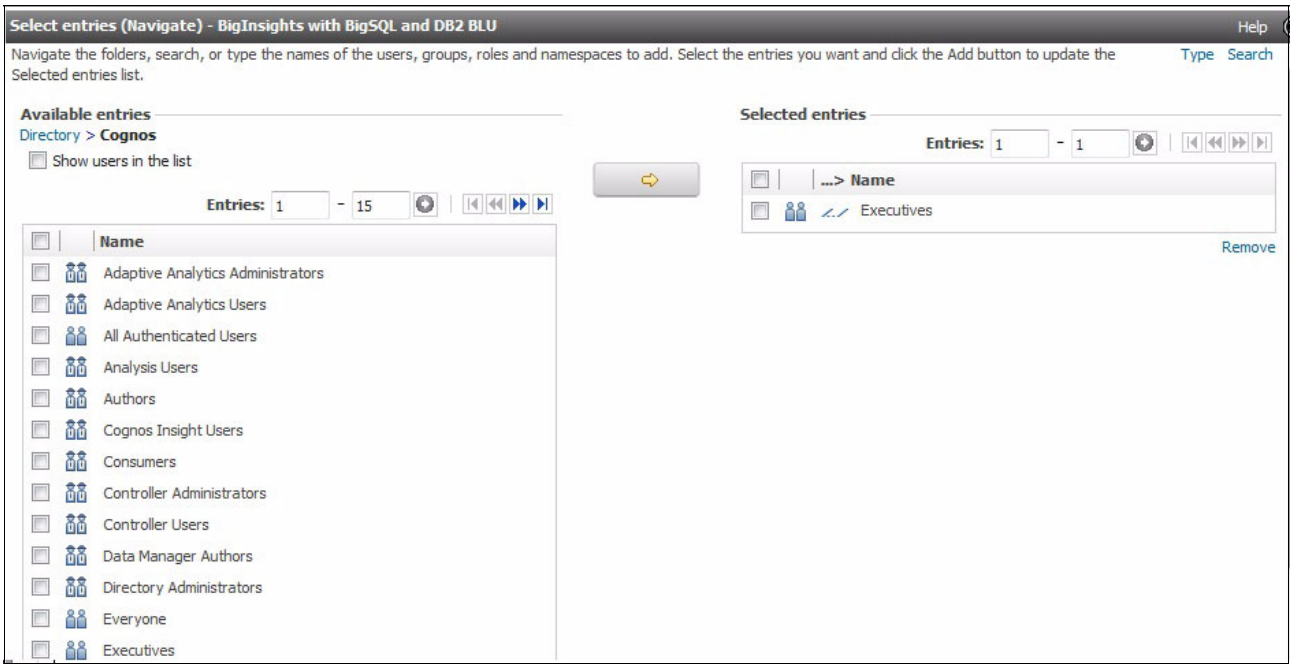


Figure 6-53 Sample configuration: Select a group so that you can grant rights to that group

11. Click **OK**.

12. Select **Executives**.

13. Select the following items in the grant column (Figure 6-54):

    - **Read**
    - **Write**
    - **Execute**
    - **Traverse**

> **Note:** For maintenance, in this sample case, do not remove the System Administrator group from this access list completely.

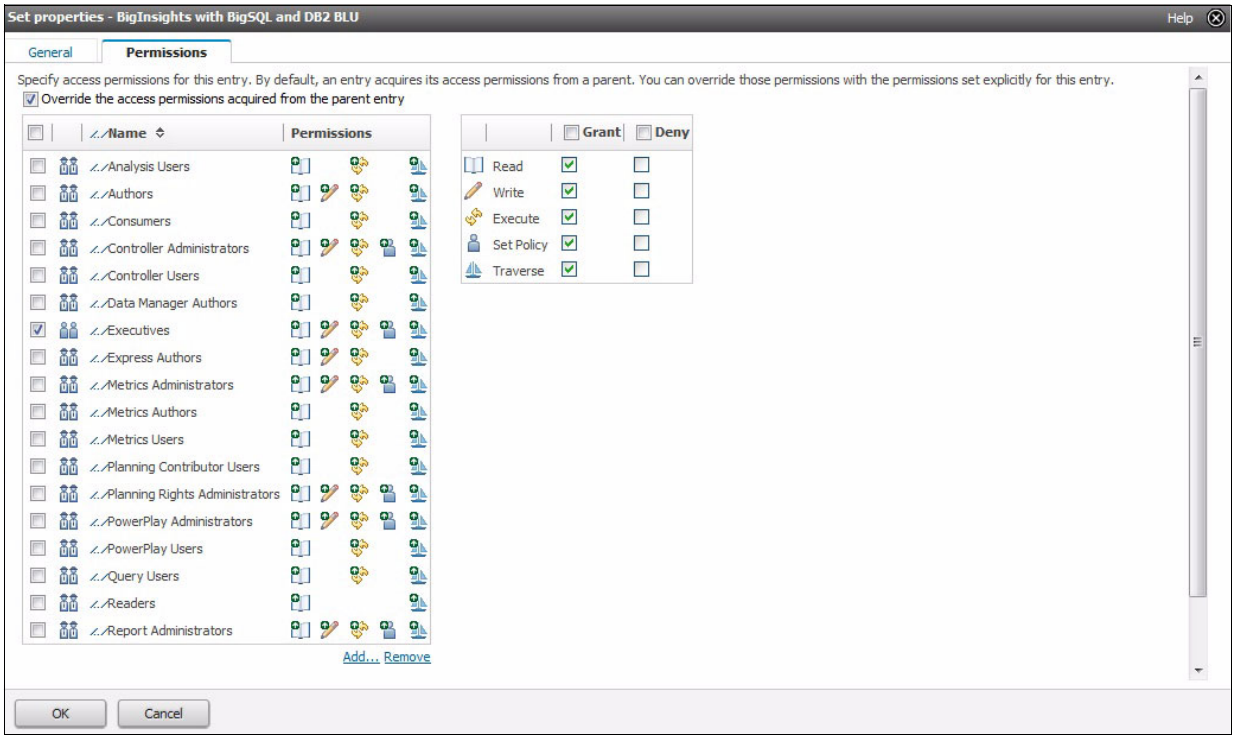


Figure 6-54 Sample configuration: Grant rights for a selected group

14. Click **OK**.

> **Note:** If you want to change the content based on the user login (apply role-based security), see *IBM Cognos Proven Practices: Dynamic Reporting with Role-based Security* at the following website:
>
> `http://ibm.co/1NZw5yP`

Implementing Analytical Decision Management

The following section describes the implementation steps for Analytical Decision Management.

Create the data source

For this demonstration, we use the product sample data.

Create the model

We create a model to receive the propensity score of the response that is in the `bank_response_data.txt` file. Then, we use the score that is received from the project in the Analytical Decision Management for Customer Interaction.

Many methods exist to create a model:

- By using Modeler Client
- By using Modeler Advantage
- By using the model building feature in the application of Analytical Decision Management

In this case, we use the Modeler Advantage in the Analytical Decision Management because it offers a simple user interface, and it enables a user to create a model without requiring detailed and professional analytical skills.

Note: To follow the steps to build the model with the Modeler Advantage, see “How to create an Advanced Model” on page 262.

Create the project to find the profitable scenario

You can create the project in Analytical Decision Management for Customer Interactions.

We create a project to score the *propensity score for response rate* by our campaign action based on previous response records. This score is used in the prioritization when IBM Analytical Decision Management simulates the customers who are offered a campaign when the IBM Analytical Decision Management for Customer Interactions creates the project.

Follow these steps:

1. Access the URL `http://<server>:9080/DM`.

Important: DM in `http://<server>:9080/DM` is case-sensitive. If you type `dm` instead of `DM`, you cannot open the access page of SPSS Analytical Decision Management.

2. Log in to SPSS Analytical Decision Management. For this case, we use `admin` for the user ID and `ibm1ibm` for the password, as shown in Figure 6-55.

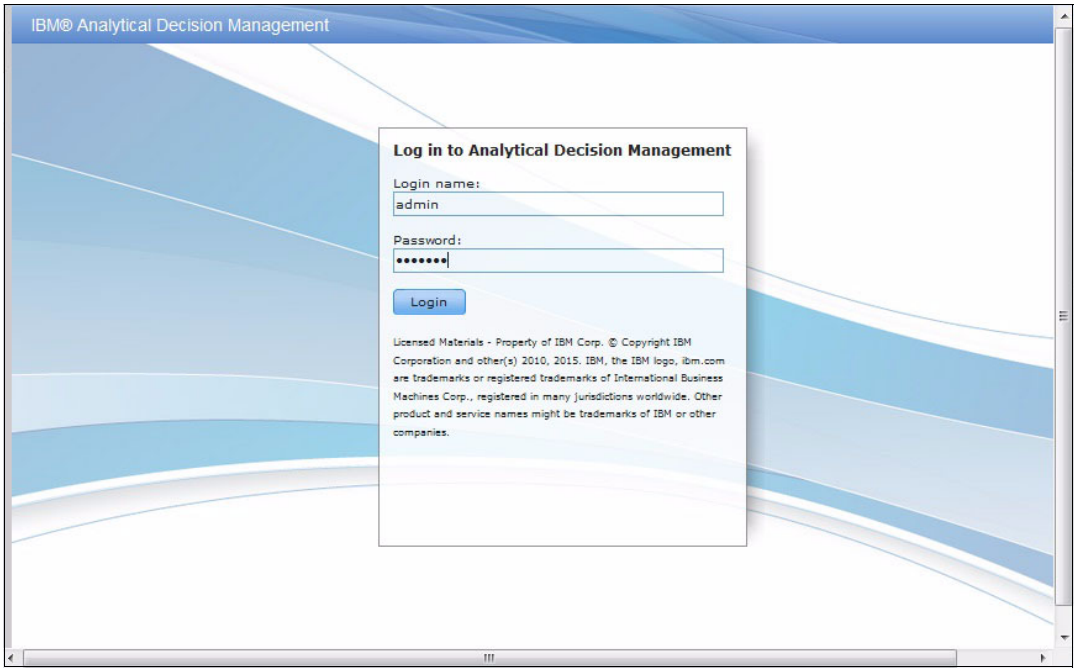


Figure 6-55 Sample view: Login page of IBM Analytical Decision Management

3. Add Twitter information (City, Product Category, and Sentiment polarity) to the `bank_customer_data.txt` file from the data. The data was retrieved from Twitter and stored in BigInsights.
4. Return to the home page.
5. Select **New** on the IBM Analytical Decision Management for Customer Interactions window, as shown in Figure 6-56.

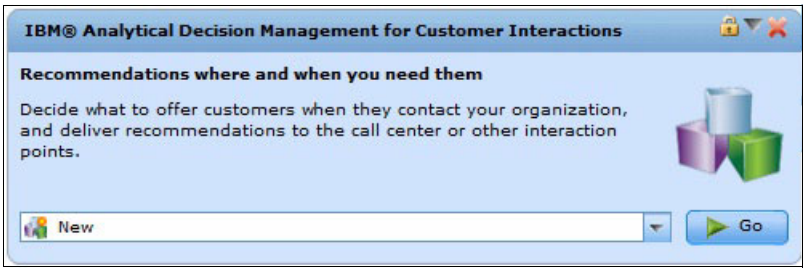


Figure 6-56 IBM Analytical Decision Management for Customer Interactions on the Launch page

6. Click **Data** on the home page.

7. Set the data source. Follow these steps to set the data sources that are used for this project:
   a. Click **Add a data source**.
   b. Set the `bank_customer_data.txt` file for the data source:
      - Data source name: `bank customer data`
      - Data Source type: **File**
      - File: `Path to the file`
   c. Click **Save**.
   d. Click the green icon in the No. of records column of Project Data Source under Data Sources.
   e. Check whether `bank customer data` is set for the Project Data Model.
   f. Check whether the fields show correctly (Figure 6-57).

Note: If the data in the Data source section was not loaded correctly in IBM Analytical Decision Management, the fields cannot be generated.

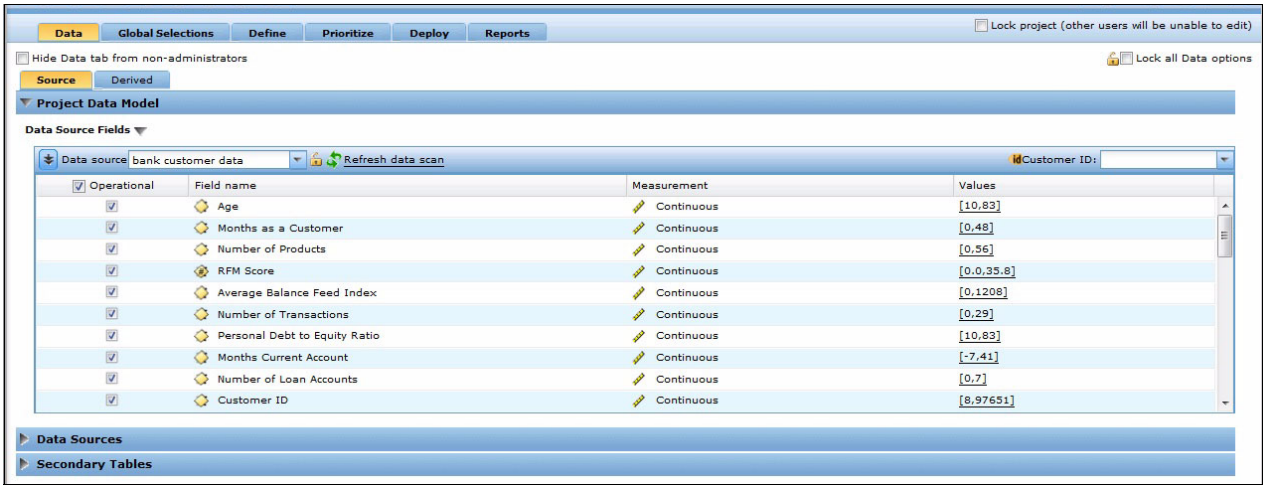


Figure 6-57 Sample result: Imported fields from the data source

8. Create rules in the Global Selections tab. You can apply a filter to determine the customers that you want to select from the imported data source:
   a. Open the **Global Selections** tab.
   b. Click the **Create a rule** icon (Figure 6-58).

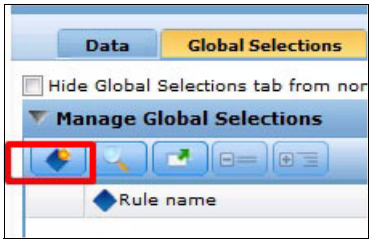


Figure 6-58 Sample view: Create a rule

c. Create a rule in the Manage Global Selections section. This section creates a filter to use records with the following condition: A person who does not have a bad payment record is not offered a campaign within 8 weeks of sending a tweeter message on Personal Accessories.

d. Enter the following information to implement this filter (Figure 6-59):

- Name: `Bad Payment Record`
- Selections: `Exclude rule`
- Expressions: Has Bad Payment Record = `1` (1 means that user has a bad payment record history.)

Note: With this setting, a user with a bad payment record is excluded.

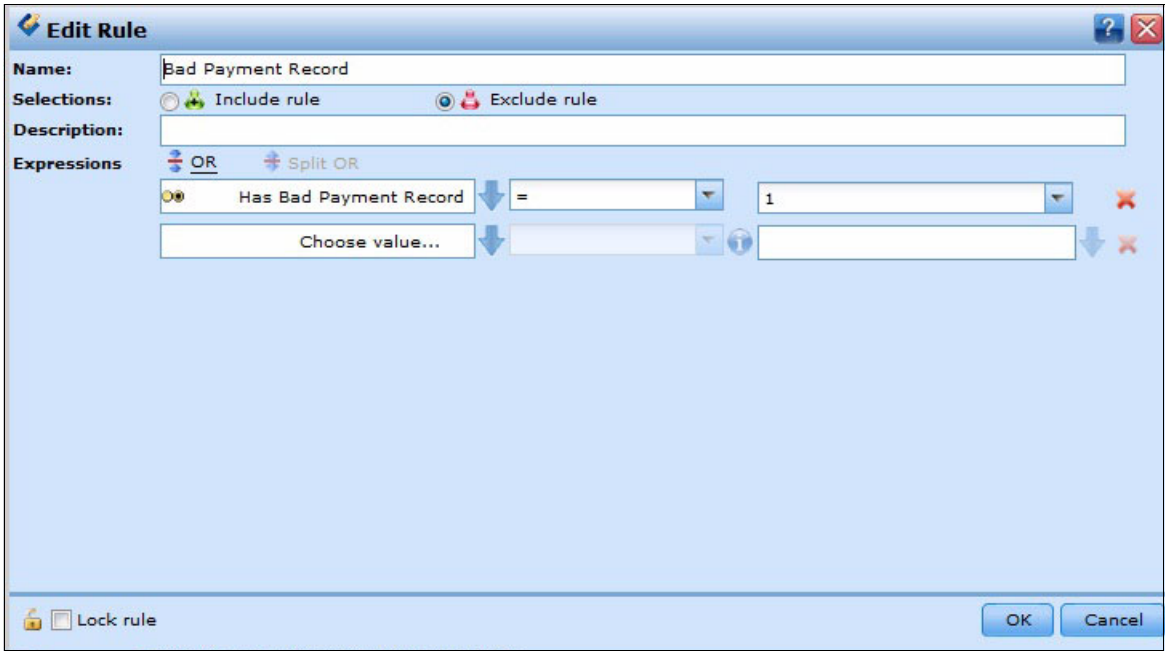


Figure 6-59 Sample configuration: Setting for Rule in the Manage Global Selections section

e. Click **OK**.

f. Click the **Create a rule** icon again.

g. Enter the following information to implement this filter:

- Name: `Weeks Since Last Offer`
- Selections: `Exclude rule`
- Expressions: Weeks Since Last Offer < 8

Note: With this setting, any customer who was offered a campaign within the last eight weeks is excluded.

h. Click **OK**.

i. Click the **Create a rule** icon again.

j. Enter the following information to implement this filter:

- Name: `Product Category`
- Selections: `Include rule`
- Expressions: Product Category = **Personal Accessories**

Note: With this setting, a customer who posted a tweet on the Personal Accessories product line is included.

k. Click **OK**.

9. Create a campaign and the offers to apply to our scenario. Follow these steps:

a. Click the **Define** tab.

b. Right-click **My Campaign** and select **Rename**.

Note: This campaign is created, by default. If you want to create a different campaign, click the plus sign (**+**).

c. For the name, type `Retention`.

d. Click **OK**.

e. Right-click **My Offer**.

f. Select **Rename**.

g. For the name, type `30 percent off on Sunglasses`.

h. Click **Add new Offer**.

i. For the name, type `30 percent off on Watches`.

j. Click **Save**.

k. Click **Add dimension tree members** (Figure 6-60).

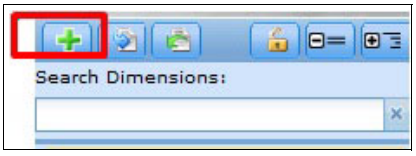


Figure 6-60 Sample view: Location of Add dimension tree members icon

10. To add a campaign, follow these steps:

    a. Click **Add New campaign**.

    b. For the name, type `Cross sell`.

    c. Click **Save**.

    d. Click **Cross sell** → **My offer**.

    e. Right-click **OK** on My offer.

    f. Select **Rename**.

    g. For name, type `1-Month Free Membership`.

    h. Click **Add New Offer**. Add the `Travel with Outdoors` and `Great Outdoors Credit Card` offers to Cross sell.

    i. Click **Save**.

    Now, two campaigns and five offers are available (Figure 6-61).

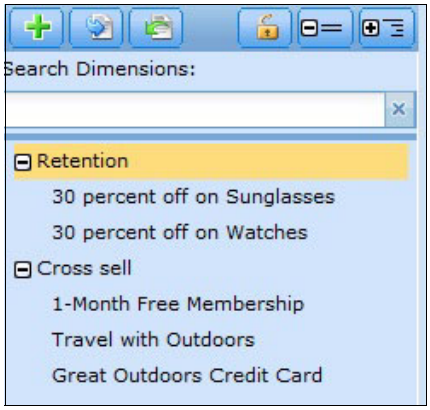


Figure 6-61 Sample result: Created campaign and offers in the Define tab

11. Create the segment rule to apply a filter to individuals who are eligible for each offer of the campaign. Follow these steps:

    a. Select **Retention**.

    b. Click **Create a new rule** in the Allocate Offer Using Segment Rules panel (not shown).

    c. Enter the following information (Figure 6-62):

    - Name: `Female`
    - Allocation: **30 percent off on Sunglasses**
    - Expressions: Gender = **F**, City = `Miami`, and Sentiment Polarity = `Majority Positive`

    d. Click **OK**.

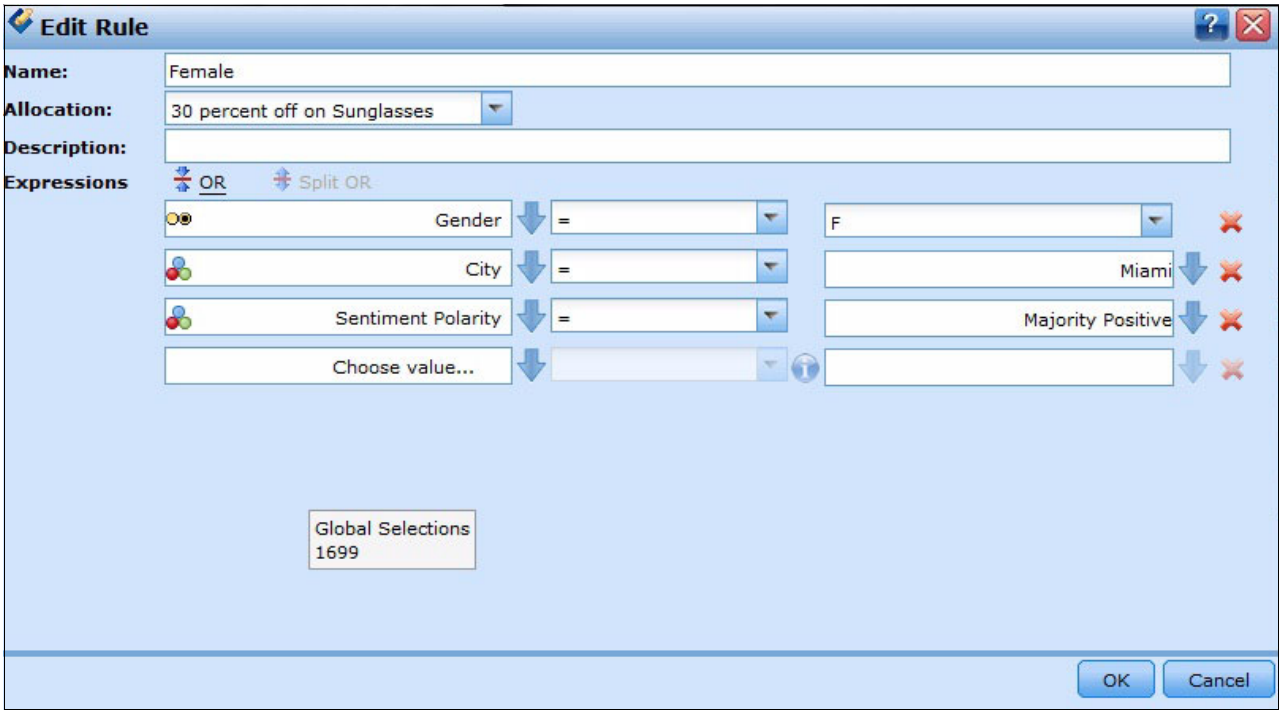


Figure 6-62 Sample configuration: Set the segment rule

    e. Click **Create a new rule** in the Allocate Offer Using Segment Rules panel (not shown).

    f. Enter the following information:

    - Name: `Male`
    - Allocation: **30 percent off on Watches**
    - Expressions: Gender = **M**

    g. Click **OK**.

    h. Select **Cross sell**.

    i. Click **Create a new rule** under **Allocate Offer Using Segment Rules**.

    j. Enter the following information:

    - Name: `Homeowner`
    - Allocation: **Great Outdoors Credit Card**
    - Expressions: Homeowner= `T`

    k. Click **OK**.

    l. Click **Create a new rule** under **Allocate Offer Using Segment Rules**.

m. Enter the following information:

- Name: `Low Debt Ratio`
- Allocation: **Great Outdoors Credit Card**
- Expressions: Personal Debt to Equity Ratio **> 30**

n. Click **OK**.

12. Now, you can create two segment rules for each of the two campaigns. Set the prioritization parameter. Specify how revenue cost and prioritization values are combined to balance objectives and optimize results. Follow these steps:

a. Open the **Prioritize** tab.

b. Click **Customize table for the Prioritization Parameter**.

c. Select an offer for each parameter in the Customize Prioritization table (Figure 6-63).

Note: With this setting, you can apply prioritization to each offer level, but not to the campaign level. As a default, you can apply prioritization settings to the campaign level only. If you want more precision, you can change the setting and apply the setting more precisely to each offer level.

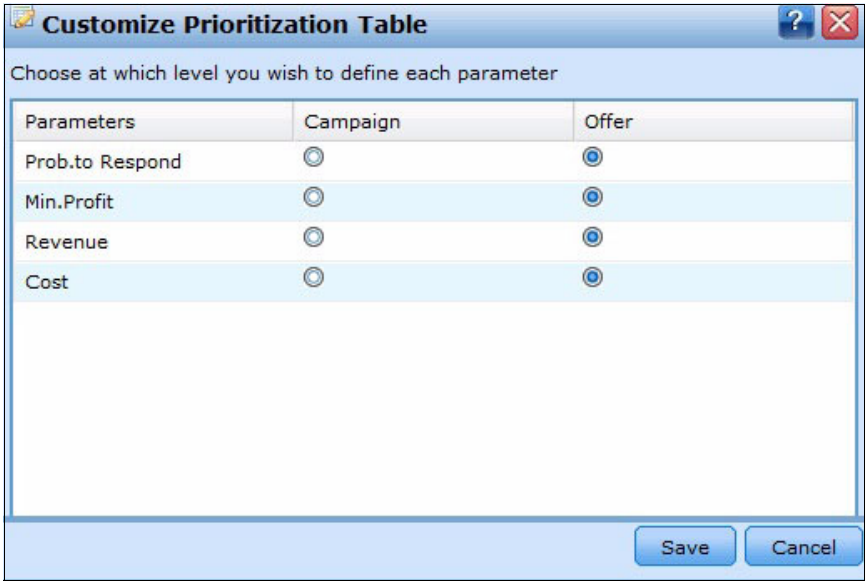


Figure 6-63 Customize Prioritization Table shows where you want to set the prioritization directly

d. Click **Save**.

e. Set the following information for the Prioritization Parameter section.

In the Prioritization Parameter section, you can specify how to choose the best offer for each customer if the customer is eligible for multiple campaigns, as shown in Table 6-6.

Table 6-6 Sample configuration: Prioritization parameter

Campaign/offer	Prob. to Respond	Minimum profit	Revenue	Cost	Priority
30 percent off on Sunglasses	0.1	10	Annual value	22	Normal
30 percent off on Watches	0.1	10	Annual value	21	Normal
1-Month Free Membership	$XFRP-Response in sdbank_response_model.str	10	60	3	Normal
Travel with Outdoors	$XFRP-Response in sdbank_response_model.str	10	200	3	Normal
Great Outdoors Credit Card	$XFRP-Response in sdbank_response_model.str	10	400	3	Normal

Note: The fields whose name starts with a dollar sign ($) means a scored field by the Modeler Advantage and other related components.

f. For offers (1-Month Free Membership, Travel with Outdoors, and Great Outdoors Credit Card) in the Cross sell campaign, set the Propensity score by clicking the **Open Input tool bar** icon (blue down arrow).

g. Click the **Select an object from existing repository** icon (magnifier) in the middle (Figure 6-64).

Figure 6-64 Sample view: Select an object from existing repository

h. Select the model that is named **sdbank_response_model.str** that you created in the Modeler Advantage.

i. Select **$XFRP-Response** (Figure 6-65).

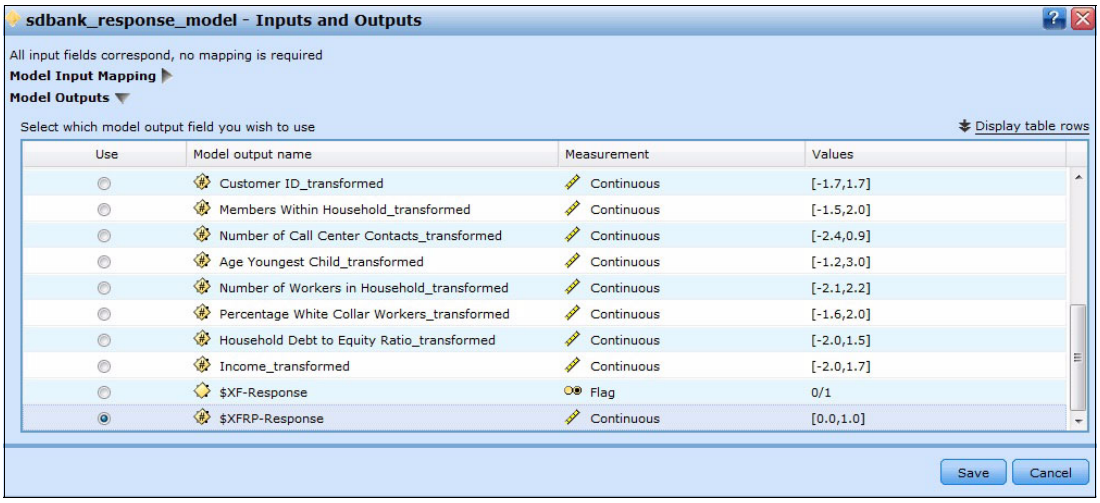


Figure 6-65 Sample result: Selecting objects in sdbank_response_model.str

Note: If you select the model icon, which is the second icon to the right, you can build the model by using the data source or upload model.

j. Validate the project in the Deploy tab. You can use the configuration under the Deploy tab to check that all parts of a project are set up correctly.

Note: If you set the Real Time Scoring options, you can specify interactive questions. You can use it to prompt users for additional information when additional information is needed in the Real Time Scoring panel. For more information, see the IBM Knowledge Center for SPSS Decision Management 7.0.0 by clicking **User's Guide** → **Scoring and deployment** → **Deploying applications** → **To specify Real Time Scoring options**) at the following website:

`http://ibm.co/1XmJrdT`

You can also apply batch scoring options. But for this case, we share simple examples and do not set the batch scoring options.

k. Click the **Deploy** tab.

l. Expand **Project Summary**.

m. Click **Generate** (Figure 6-66) to validate that the project is ready for deployment.

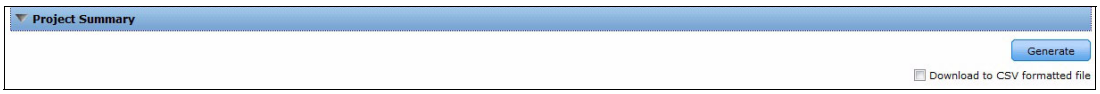


Figure 6-66 Sample view: Project Summary section in the Deploy tab

n. After the validation, expand the section.

o. Check that each item has a green check as shown in Figure 6-67.

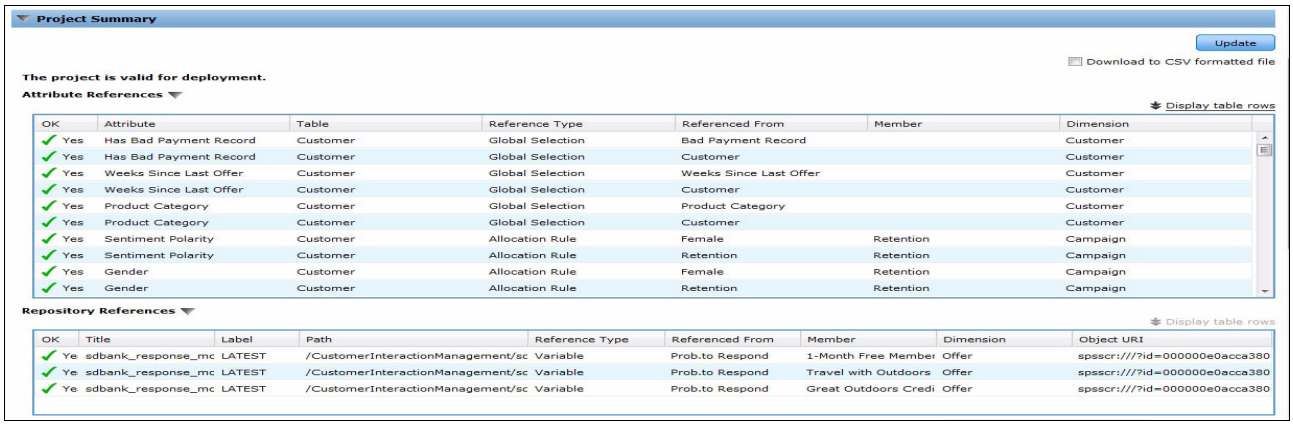


Figure 6-67 Sample result: Model validation

13. Create the scoring service configuration.

    You use the Deployment Manager to create a scoring configuration. For more information, see the IBM Knowledge Center for SPSS Collaboration and Deployment Services 7.0.0. Click **Deployment Manager User's Guide** → **Scoring** → **Scoring configurations** at the following website:

    `http://ibm.co/1TbUpm2`

    Follow these steps:

    a. Click **Start** → **All Programs** → **IBM SPSS Collaboration and Deployment** → **Deployment Manager 7.0** (Figure 6-68).

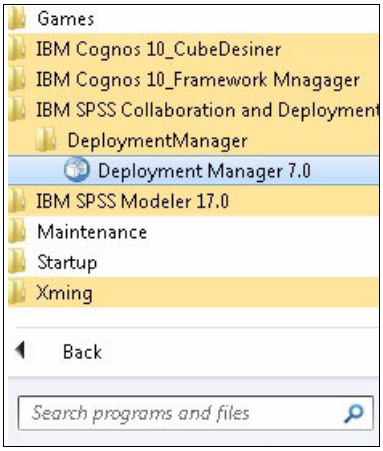


Figure 6-68 Sample view: Location of Deployment Manager 7.0

    b. Double-click **Deployment Manager 7.0**.

    c. Right-click **Server name**.

    d. Select **Log on as**.

e. Click **Server name** → **Content Repository**.

f. Right-click **CustomerInteraction.str**, which you created in this sample.

g. Select **Configure Scoring** (Figure 6-69).

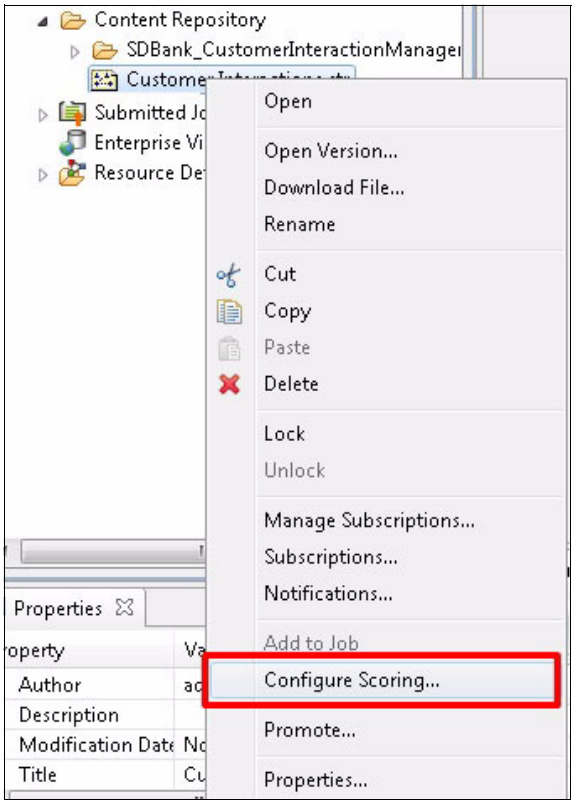


Figure 6-69 Sample configuration: Select Configure Scoring in Deployment Manager 7.0

h. On the Scoring tab, provide the following information:

   i. For the configuration name, type `bank customer interaction`. Click **Next**.

   ii. Type 2 for the maximum number of offers. Click **Next**.

   iii. Leave the setting and click **Next**.

   iv. Leave the setting and click **Next**.

   v. Leave the setting and click **Next**.

   vi. Click **Finish**.

i. Check whether the status of the created scoring configuration is `Started` (Figure 6-70).

Note: If the status is not `Started`, you cannot receive the scoring result.

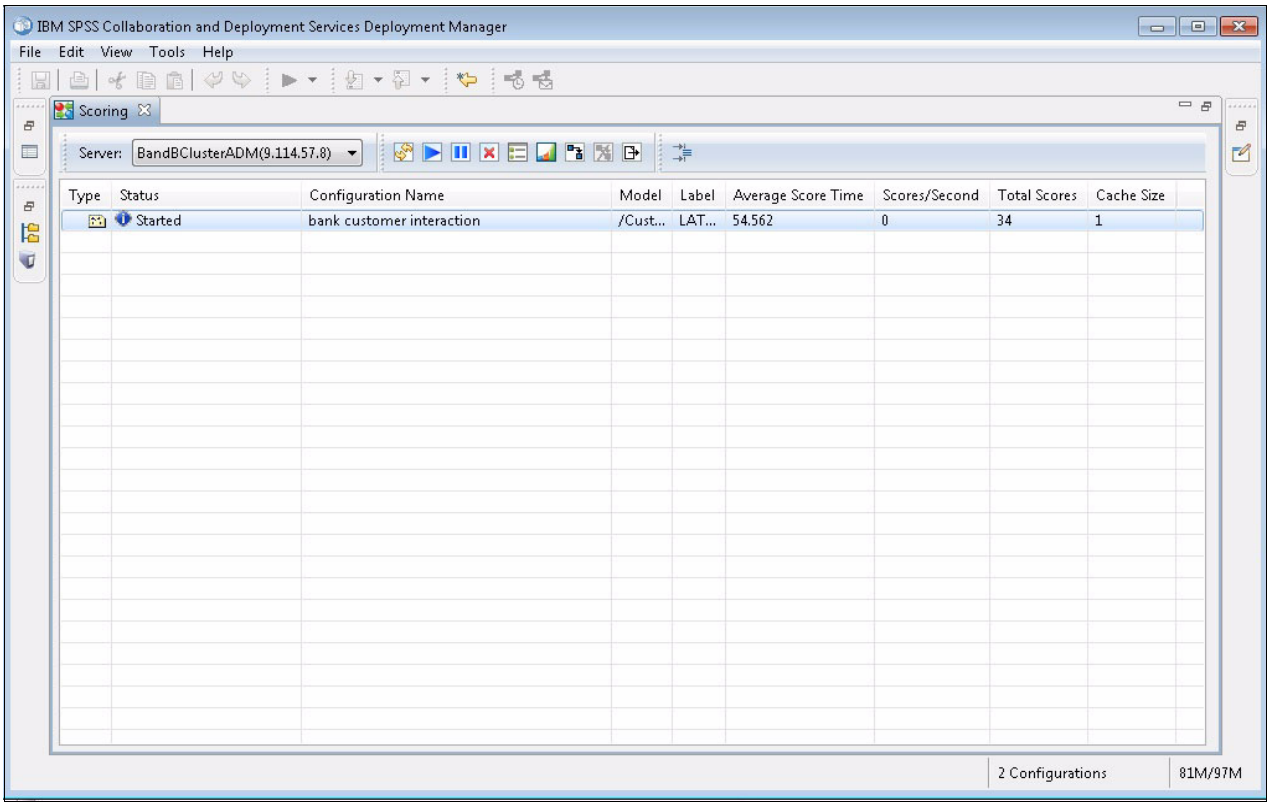


Figure 6-70 Sample result: Started status of scoring configuration for the model

6.4.4 Links

For more information, see the following resources:

- SPSS Collaboration and Deployment Services 7.0.0 at the following website:

  `http://ibm.co/1TbUpm2`

- SPSS Decision Management 7.0.0 at the following website:

  `http://ibm.co/1XmJrdT`

- Making better business decisions with analytics and business rules at the following website:

  `http://ibm.co/1O5VLYX`

A

Advanced implementation

We describe the advanced techniques that are presented in Chapter 6, “Scenario: How to use the solution” on page 189.

This appendix covers the following topics:

- Suggestions for Cognos Dynamic Cubes
- Modeler Advantage in IBM SPSS Analytical Decision Management

Suggestions for Cognos Dynamic Cubes

This section describes the following topics:

- Cognos Dynamic Cubes advantage
- Sample view of Cognos Dynamic Cubes
- Creating reports by using Cognos Dynamic Cubes
- Useful link

Cognos Dynamic Cubes advantage

If your environment has sufficient resources, Cognos Dynamic Cubes can cache data in-memory, which can help minimize SQL transactions and data retrieval between a relational database and Cognos Dynamic Cubes by using simple and multi-pass SQL that is optimized for the relational database. This operation offers the capability to introduce a performance layer in the Cognos query stack for low-latency, high-performance online analytical processing (OLAP) analytics over large relational data warehouses.

Sample view of Cognos Dynamic Cubes

This section provides a sample view of the Cognos Dynamic Cubes.

Cognos Administration

You can check the status of the published Dynamic Cubes by clicking the **Status** tab and by clicking **Dynamic Cubes** in the Cognos Administration window, as shown in Figure A-1.

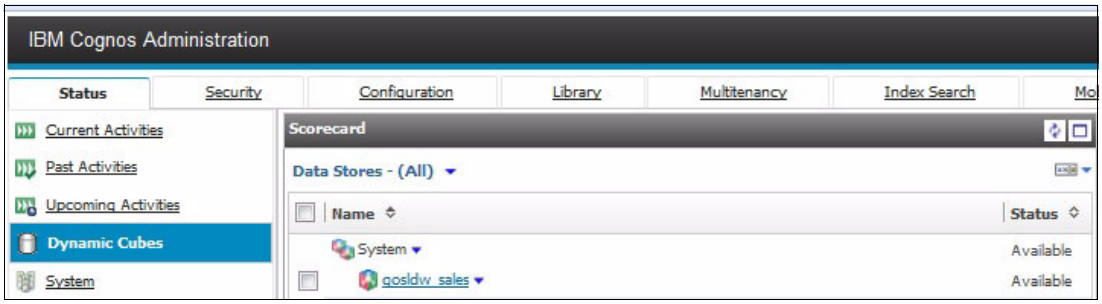


Figure A-1 Sample configuration: Dynamic Cubes in Cognos Administration

Cognos Connection

Figure A-2 shows a view of a published Dynamic Cube in the Cognos Connection window.

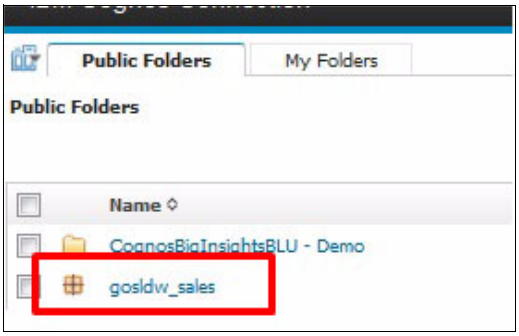


Figure A-2 Sample configuration: Dynamic Cube in Cognos Connection

IBM Cognos Cube Designer (Windows Client)

In the Cube Designer, you can create IBM Cognos Dynamic Cubes and publish them to Cognos Connection, as shown in Figure A-3.

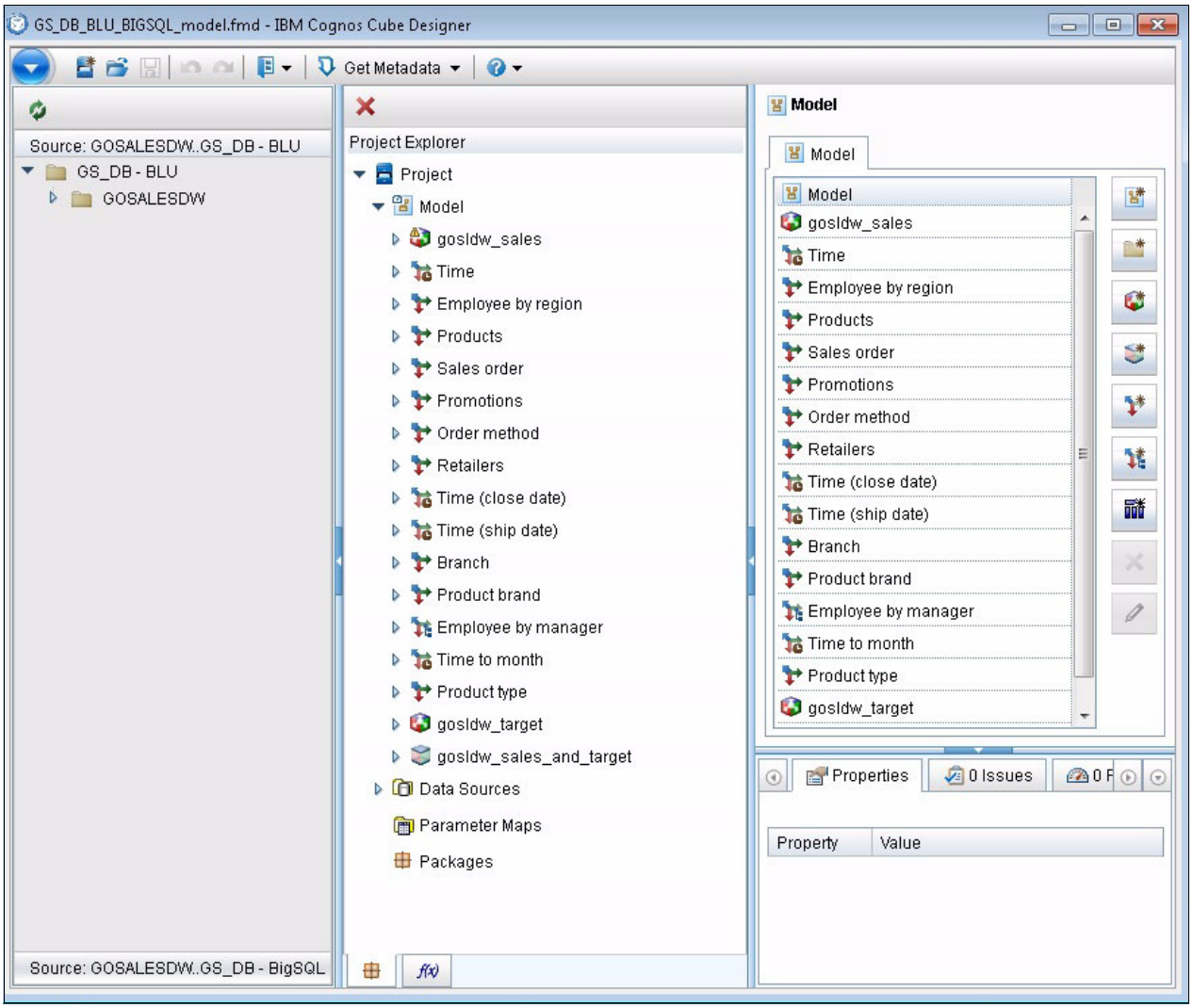


Figure A-3 Sample configuration: Dynamic Cubes in IBM Cognos Cube Designer

Creating reports by using Cognos Dynamic Cubes

First, you must publish the Dynamic Cube from the IBM Cognos Cube Designer. Then, you check the status of the published Dynamic Cube. After you check the status and the Cube status is `Available` in Cognos Administration, you can create your reports against the Dynamic Cubes. To see how to create reports, see the Cognos Business Intelligence part of 6.4.3, “How to implement the sample case” on page 216.

Useful link

Form more information, see the following link:

- IBM Knowledge Center for Cognos Business Intelligence 10.2.2:

  http://ibm.co/1nidZjZ

Modeler Advantage in IBM SPSS Analytical Decision Management

This section illustrates the following topics about IBM Statistical Package for the Social Sciences (SPSS) Modeler Advantage:

- Modeler Advantage benefits
- How to create an Advanced Model
- How to set the created Modeler stream to the project

Modeler Advantage benefits

Modeler Advantage is easy-to-use application that puts the power of predictive modeling in the hands of business users.

Modeler Advantage helps to create a modeler stream without requiring the installation of Modeler Client on your terminal box. You can use Modeler Advantage and its easy-to-use interface to create a stream.

How to create an Advanced Model

In this section, we create an Advanced Model to calculate the *propensity score for response rate* by campaigns based on previous response records. This score is used to prioritize when IBM Analytical Decision Management simulates to whom to offer a campaign when IBM Analytical Decision Management for Customer Interactions creates the project.

Use the following steps to create the model:

1. Access the URL `http://<server>:9080/DM`.
2. Log in to IBM Analytical Decision Management**.** For this demonstration, set the user ID to `admin` and set the password to `ibm1ibm`.
3. Select **New** in the IBM SPSS Modeler Advantage, as shown in Figure A-4.

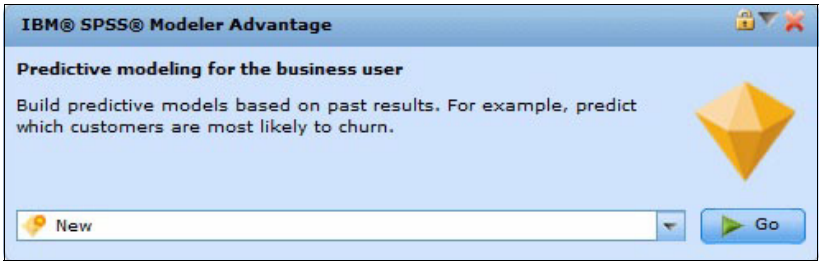


Figure A-4 Sample view: Modeler Advantage launch page

4. Click **Go**.

5. Click **Data** on the home page of the Modeler Advantage (See Figure A-5).

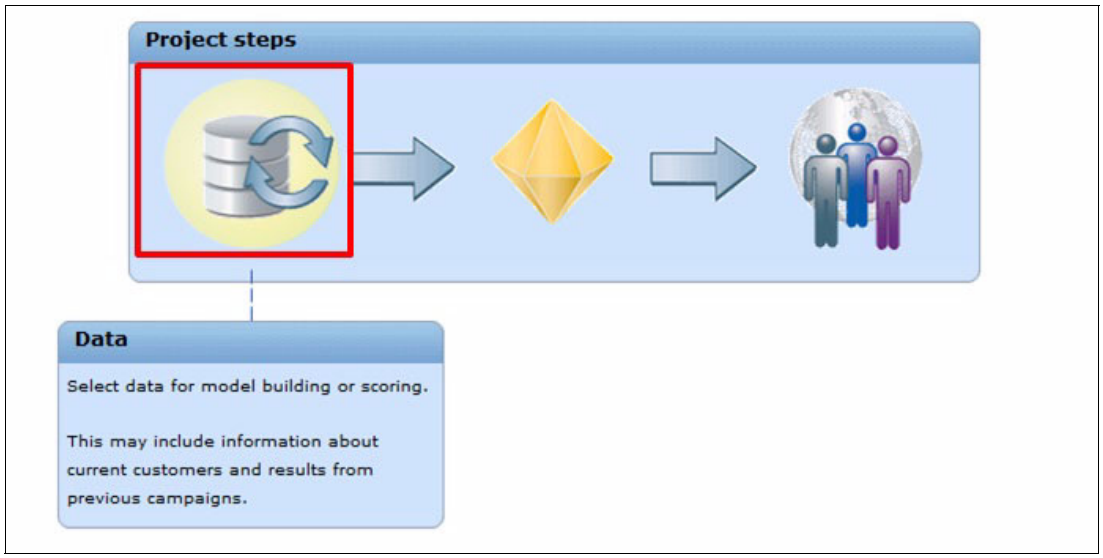


Figure A-5 Sample configuration: Home page of Modeler Advantage

6. Click **Add a data source** in the Project Data Sources section (upper-right side of Figure A-6).

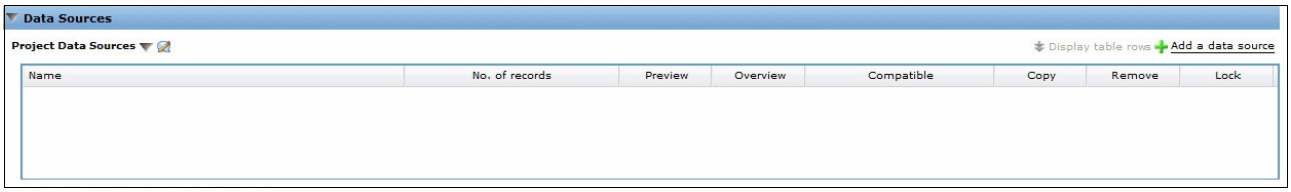


Figure A-6 Sample configuration: Data Sources section on the Data tab

Note: The data file that is named `bank_response_data.txt` is the modeler's product sample. The file is in the `demos` folder of the Modeler Server. The default path in the Modeler Server is `/usr/IBM/SPSS/ModelerServer/17.0/demos/`.

This file contains the customer's demographic data, which includes any interest in a campaign or whether the customer responded to previous campaigns.

7. Click **Save**.

8. Click the **green icon** in the No. of records column of the Project Data Sources box under the Data Sources section. After you click the icon, Modeler Advantage tries to load data from the file. (The green icon is highlighted in Figure A-7.)

Note: To proceed, in the following steps, you need to reload the field items from the files.

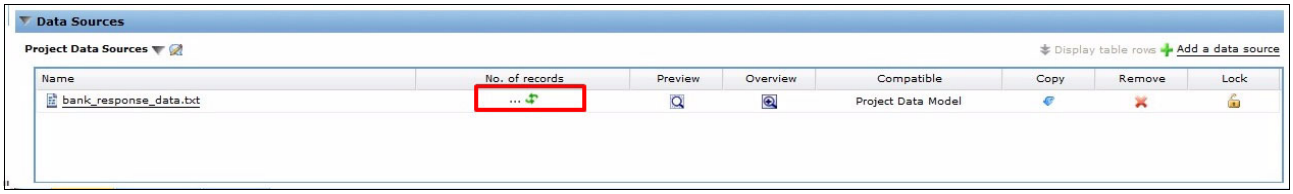


Figure A-7 Sample result: Data source that needs to be refreshed

9. After you load the records into the Modeler Advantage, the number in the No. of records field is displayed. See Figure A-8.

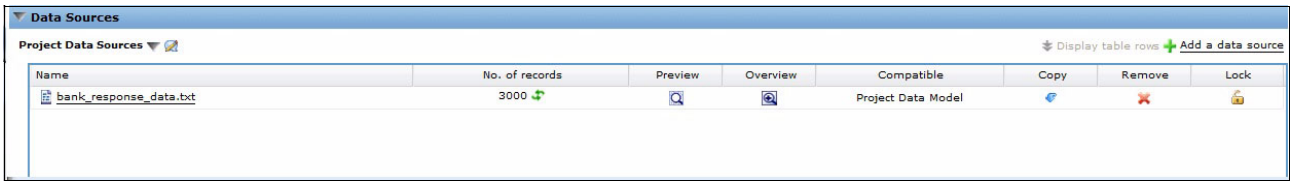


Figure A-8 Sample result: Data source that is loaded to Modeler Advantage correctly

10. Open the **Modeling** tab.
11. Provide the following information:
    - Data source: **bank_response_data.txt**
    - Target: **Response**

Note: With this operation, the Modeler Advantage calculates the propensity score based on the response field. The result will be used in the project that is generated in the IBM Analytical Decision Management for Customer Interactions.

12. Click **Build Model**.
13. Wait until after Modeler Advantage builds the model (Figure A-9).

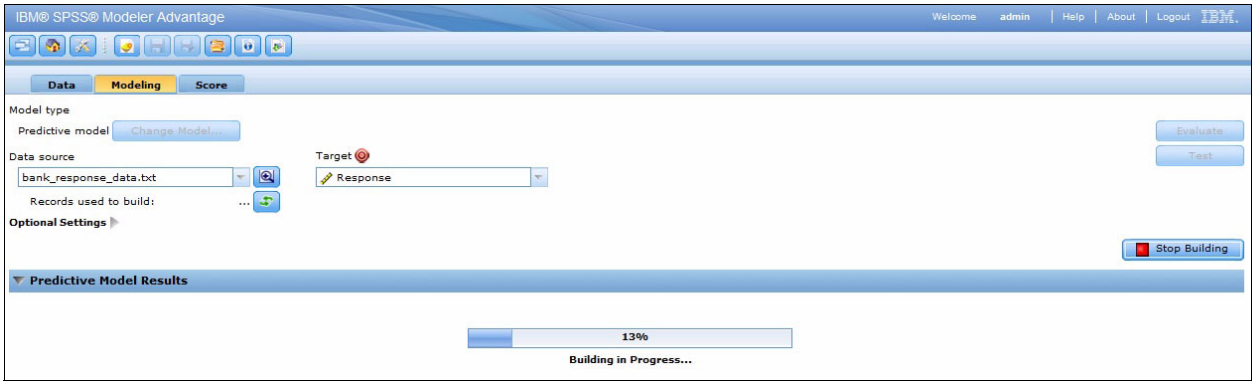


Figure A-9 Sample result: Building the model in Modeler Advantage

Note: This operation requires time to build the model due to lack of resources. The required disk space size is at least three times its original data source size. After the model is built, you receive the chart in the Predictive Modeler Results section (Figure A-10).

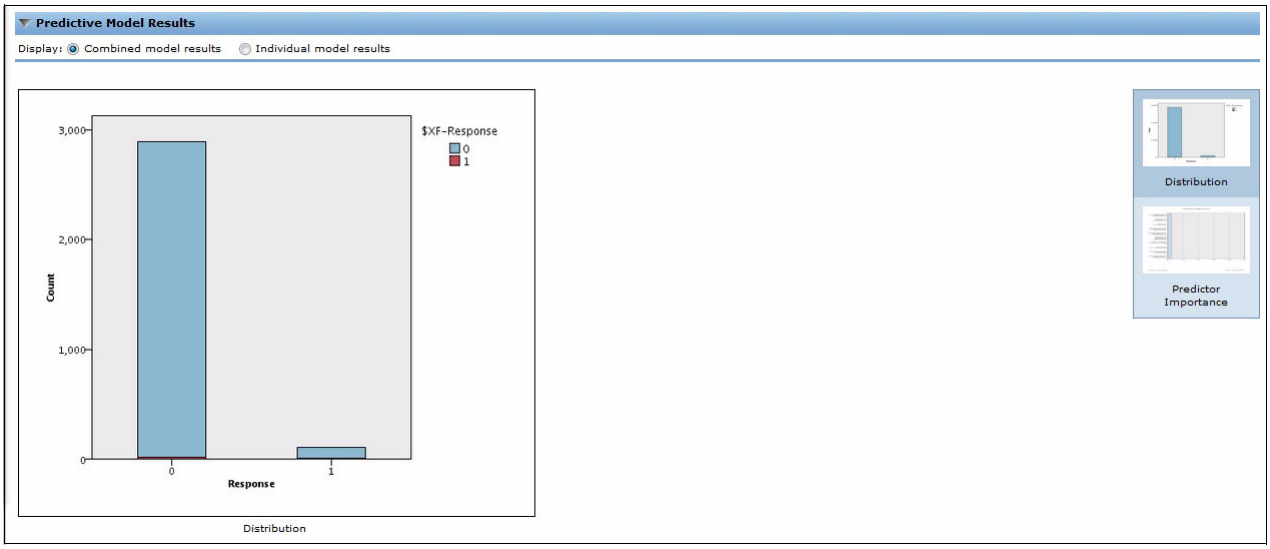


Figure A-10 Sample result: Predictive Model Results in Modeler Advantage

14. Click **Save as** and name this project `sdbank_response_model.str`.
15. Click **Save as** → **Download**.
16. Select **Save File**.
17. Click **OK**.

How to set the created Modeler stream to the project

This section describes how to implement the created Modeler stream in the Modeler Advantage. Follow these steps:

1. Create the project.
2. Set the prioritization parameter.

   In this section, specify how revenue cost and prioritization values are combined to balance objectives and optimize results. For detailed steps, see “Implementing Analytical Decision Management” on page 245.

Note: In the Prioritization Parameter section, you can specify how to choose the best offer for each customer if a customer is eligible for multiple campaigns.

3. Deploy the project.

B

Planning Ambari node roles

This appendix describes how to plan the Ambari environment in relationship to the nodes in the cluster.

Ambari node roles

This section provides details about planning the Ambari nodes and their roles. Table B-1 shows the node roles with two management nodes.

Table B-1 Node roles with two management nodes

Host	mn01	mn02	dn0x
Ambari Server	X		
History Server		X	
ResourceManager		X	
App Timeline Server		X	
Hive Metastore	X		
WebHCat Server	X		
HiveServer2	X		
Hbase Master	X		
Oozie Server	X		
ZooKeeper server		X	
Kafka Broker		X	
Symphony Master		X	
Knox Gateway		X	
Solr	X		
Metrics Collector	X		
GPFS Master		X	
Spark History Server		X	
Spark Thrift server		X	
NodeManager			X
RegionServer	X		
HBaseRESTServer	X		
Flume	X	X	X
GPFS Node	X	X	X
Symphony compute	X		X
Client	X	X	X

Table B-2 shows the node roles with four management nodes.

Table B-2 Node roles with four management nodes

Host	mn01	mn02	mn03	mn04	dn0x
Ambari Server	X				
History Server		X			
ResourceManager		X			
App Timeline Server		X			
Hive Metastore				X	
WebHCat Server				X	
HiveServer2				X	
Hbase Master				X	
Oozie Server				X	
ZooKeeper server	X	X	X		
Kafka Broker			X		
Symphony Master		X			
Knox Gateway		X			
Solr		X			
Metrics Collector			X		
GPFS Master		X			
Spark History Server		X			
Spark Thrift server		X			
NodeManager					X
RegionServer	X	X	X	X	
HBaseRESTServer	X				
Flume	X	X	X	X	X
GPFS Node	X	X	X	X	X
Symphony compute	X		X	X	X
Client	X	X	X	X	X

Table B-3 shows the node roles with six management nodes.

Table B-3 Node roles with six management nodes

Host	mn01	mn02	mn03	mn04	mn05	mn06	dn0x
Ambari Server	X						
History Server		X					
ResourceManager		X					
App Timeline Server		X					
Hive Metastore						X	
WebHCat Server				X			
HiveServer2				X			
Hbase Master				X	X	X	
Oozie Server				X			
ZooKeeper server		X	X	X	X	X	
Kafka Broker			X				
Symphony Master		X					
Knox Gateway		X					
Solr			X				
Metrics Collector			X				
GPFS Master		X					
Spark History Server		X					
Spark Thrift server		X					
NodeManager							X
RegionServer			X	X	X	X	
HBaseRESTServer	X						
Flume	X	X	X	X	X	X	X
GPFS Node	X	X	X	X	X	X	X
Symphony compute	X		X	X	X	X	X
Client	X	X	X	X	X	X	X

Table B-4 shows the node roles with eight management nodes.

Table B-4 Node roles with eight management nodes

Host	mn01	mn02	mn03	mn04	mn05	mn06	mn07	mn08	dn0x-dat
Ambari Server	X								
History Server		X							
ResourceManager		X							
App Timeline Server		X							
Hive Metastore								X	
WebHCat Server			X						
HiveServer2			X						
Hbase Master			X	X	X				
Oozie Server			X						
ZooKeeper server	X	X	X	X	X	X	X		
Kafka Broker			X					X	
Symphony Master		X							
Knox Gateway		X							
Solr							X		
Metrics Collector			X						
GPFS Master		X							
Spark History Server		X							
Spark Thrift server		X							
NodeManager									X
RegionServer			X	X	X	X	X		
HBaseRESTServer	X								
Flume	X	X	X	X	X	X	X	X	X
GPFS Node	X	X	X	X	X	X	X	X	X
Symphony compute	X		X	X	X	X	X	X	X
Client	X	X	X	X	X	X	X	X	X

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topic in this document. Note that some publications referenced in this list might be available in softcopy only.

- *Architecting and Deploying DB2 with BLU Acceleration*, SG24-8212
- *Performance Optimization and Tuning Techniques for IBM Power Systems Processors Including IBM POWER8*, SG24-8171
- *IBM Spectrum Scale (formerly GPFS)*, SG24-8254

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:

- IBM Power Systems Quick Reference Guide:

  https://ibm.biz/Bd4yQU
- Deploying a big data solution using IBM Spectrum Scale-File Placement Optimization (FPO):

  http://ibm.co/1NBnGTj
- IBM Knowledge Center for Cognos Business Intelligence 10.2.2:

  http://ibm.co/1nidZjZ
- IBM Spectrum Scale Wiki:

  http://ibm.co/1RZmRbX
- SPSS Collaboration and Deployment Services 7.0.0 at the following website:

  http://ibm.co/1TbUpm2
- SPSS Decision Management 7.0.0 at the following website:

  http://ibm.co/1XmJrdT
- Making better business decisions with analytics and business rules at the following website:

  http://ibm.co/1O5VLYX

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Redbooks

Implementing an Optimized Analytics Solution on IBM Power Systems

SG24-8291-00
ISBN 0738441686

IBM

(0.5” spine)
0.475”<->0.873”
250 <-> 459 pages

SG24-8291-00

ISBN 0738441686

Printed in U.S.A.