Redpaper

Dino Quintero
Ricardo Dobelin Barros
Ashraf Gomaa
José Higino
Archana Kumar
Majid Ouassir
Adam Parker
Joanna Wong

Integration of IBM Platform Symphony and IBM InfoSphere BigInsights

Introduction

In early 2012, the US Government announced the Big Data Research and Development Initiative to determine how big data can be used to address important problems that the government faces. Such problems include issues in healthcare, energy, and defense. Consumer companies, such as Amazon, are analyzing data to provide their customers with interactions that are individualized and more human. With all the attention that big data has received in the last few years, people are now realizing its value in providing deep insights into trends.

Hadoop can store big data and unlock the answers by analyzing them. IBM® InfoSphere® BigInsights™ is built on top of open source Hadoop and extends it with advanced analytic tools and other capabilities with added value. InfoSphere BigInsights helps organizations of all sizes to more efficiently manage the vast amounts of data that consumers and businesses create every day. At its core, Hadoop is a Distributed Computing Environment that manages the execution of distributed jobs and tasks on a cluster. As with any Distributed Computing Environment, the Hadoop software needs to provide facilities for resource management, scheduling, remote execution, and exception handling. Although Hadoop provides basic capabilities in these areas, IBM Platform Computing has been working on these problems and perfecting them for twenty years.

With use cases for MapReduce continuing to evolve, customers are increasingly encountering situations where scheduling efficiency is becoming important. This scenario is true from the standpoint of meeting performance and service-levels goals and from the perspective of using resources more efficiently to contain infrastructure costs. Also, as the number of applications for MapReduce and big data continues to grow, multitenancy and a shared services architecture become more critical. It is not feasible from a cost standpoint to create separately managed grid environments for every critical MapReduce workload.

This IBM Redpaper™ publication describes the integration of IBM Platform Symphony® 5.2 and IBM InfoSphere BigInsights 1.4 in an IBM System x® cluster. IBM Platform Symphony is a low-latency scheduling solution that supports true multitenancy and sophisticated workload management capabilities.

IBM Platform Symphony

IBM Platform Symphony owes its name to its unique ability to orchestrate distributed services on a shared grid in response to dynamically changing workloads. It combines a fast service-oriented application middleware framework (SOAM), a low-latency task scheduler, and a scalable grid management infrastructure that is proven in some of the world's largest production grids. This unique design ensures application reliability and low-latency and high-throughput communication between clients and compute services.

How InfoSphere BigInsights works with IBM Platform Symphony

Figure 1 shows the various components that make up InfoSphere BigInsights Enterprise Edition.

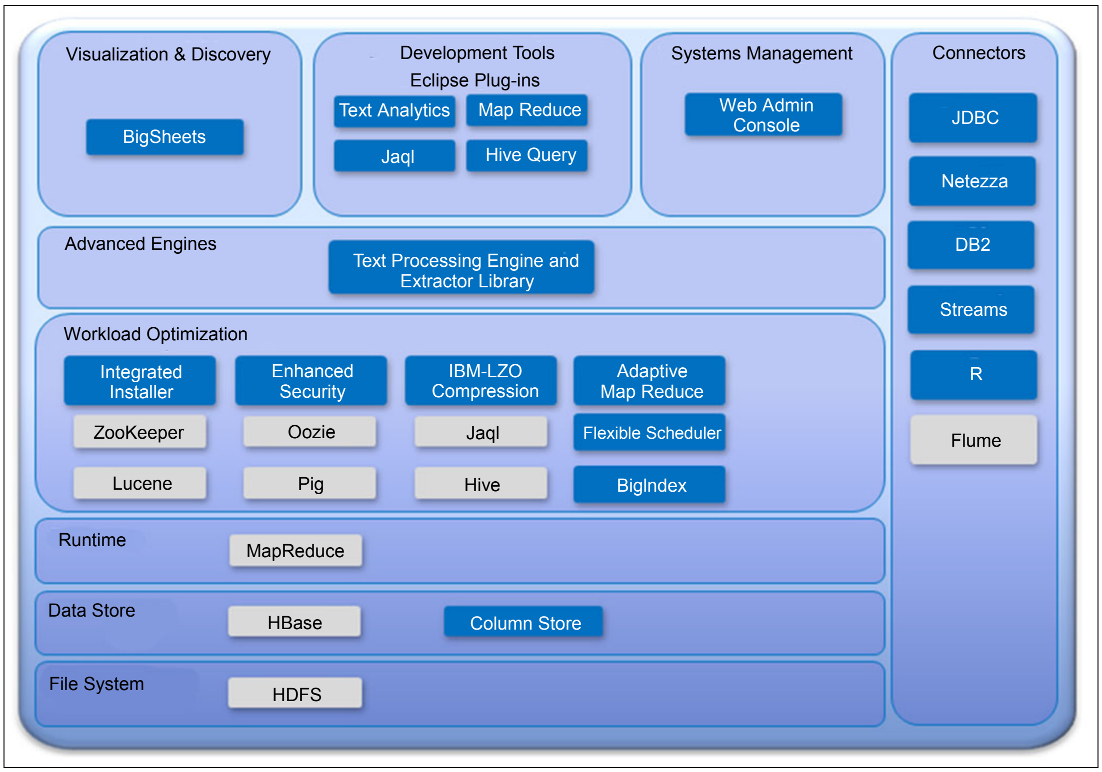

Figure 1 InfoSphere BigInsights components

This view makes it clear how InfoSphere BigInsights augments open source Apache Hadoop and provides more capabilities. These capabilities include visualization and query tools, development tools, management tools, and data connectors to external data stores.

When IBM Platform Symphony is deployed with InfoSphere BigInsights, IBM Platform Symphony replaces the open source MapReduce layer in the Hadoop framework. (IBM Platform Symphony is not a Hadoop distribution.) IBM Platform Symphony relies on a Hadoop MapReduce implementation that is present with various open source components, such as Pig, Hive, HBase, and HDFS file systems.

As shown in Figure 2 on page 3, IBM Platform Symphony replaces the MapReduce scheduling layer in the InfoSphere BigInsights software environment. As a result, IBM Platform Symphony provides better performance and multitenancy in a way that is transparent to InfoSphere BigInsights and its users.

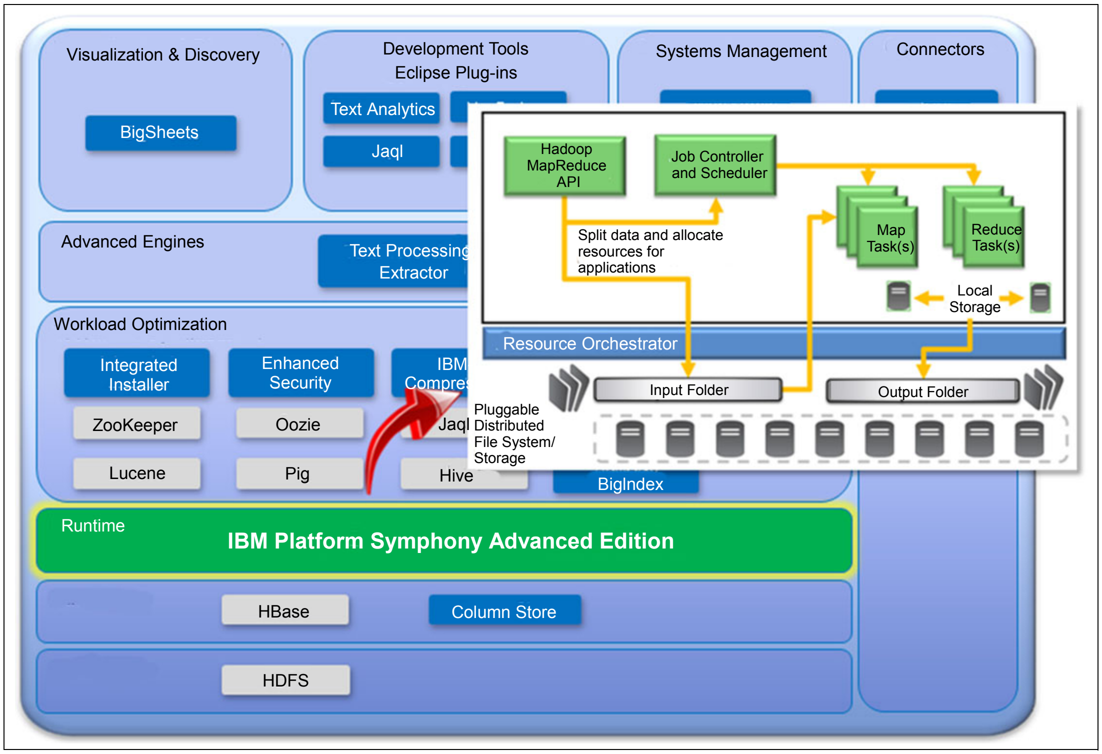

Figure 2 IBM Platform Symphony (Advanced Edition) runtime integration with InfoSphere BigInsights

Big data workloads can be submitted from the InfoSphere BigInsights graphical interface, from a command line, or from client applications that interact with the Hadoop MapReduce APIs. After you configure InfoSphere BigInsights to use the IBM Platform Symphony scheduler in place of the Hadoop scheduler, InfoSphere BigInsights workloads run seamlessly and are manageable from within the InfoSphere BigInsights environment.

Administrators must be aware that, when they run InfoSphere BigInsights on a shared IBM Platform Symphony grid, some cluster and service management that are accessible from within InfoSphere BigInsights become redundant. For example, it is no longer assumed that InfoSphere BigInsights has exclusive use of the grid. Therefore, IBM Platform Symphony provides capabilities for cluster node management, service management, and high availability features, for example, for such components as the NameNode, JobTrackers, and TaskTrackers.

The IBM Platform Symphony performance advantage

IBM continues to achieve record MapReduce performance results with IBM Platform Symphony. In a recent test that was conducted and audited by an independent third-party testing lab, IBM Platform Symphony ran a particular mix of social analytic workloads an average of 7.3 times faster than Hadoop MapReduce alone.

Performance improvements for MapReduce workloads that run on IBM Platform Symphony are a result of the following factors:

- Low-latency scheduling means that jobs start (and complete) faster.
- By design, IBM Platform Symphony monitors hosts dynamically and always schedules tasks preferentially to the host that can respond quickly to the workload.

- By avoiding the Hadoop heartbeat (polling) model, the task scheduling rate for Hadoop workloads is improved dramatically with IBM Platform Symphony and scheduling latency is reduced.
- If resources are idle on the cluster, IBM Platform Symphony MapReduce workloads can expand resource allocations dynamically to borrow unused nodes to maximize usage.
- IBM Platform Symphony uses generic slots rather than slots that are statically allocated to map and reduce functions so that slots can be shared between map and reduce tasks.
- The MapReduce shuffle phase is improved and attempts to keep data in memory while using the more efficient data transfer mechanisms of IBM Platform Symphony for moving data between hosts.
- Developers can optionally take advantage of API features and data handling enhancements that are unique to IBM Platform Symphony to achieve advantages beyond what can be achieved within MapReduce itself.

Customers who deploy InfoSphere BigInsights or other big data application environments can realize significant advantages by using IBM Platform Symphony as a grid manager:

- Better application performance
- Opportunities to reduce costs through better infrastructure sharing
- The ability to guarantee application availability and quality of service
- Ensured responsiveness for interactive workloads
- Simplified management by using a single management layer for multiple clients and workloads

IBM Platform Symphony is especially beneficial to InfoSphere BigInsights customers who are running heterogeneous workloads that benefit from low latency scheduling. The resource sharing and cost savings opportunities that re provided by IBM Platform Symphony extend to all types of workloads.

For Hadoop grid administrators who are looking for opportunities to improve performance, reduce cluster sprawl, and improve service levels at a lower cost, IBM Platform Symphony provides a powerful complement to InfoSphere BigInsights.

Environment

The integration and comparison tests took place on a group of machines at IBM Poughkeepsie Benchmark Center. The environment had the following setup:

- Hardware:
  - 12x IBM System xiDataPlex dx360 (Server Model 6391 and Type AC1) M3 nodes
    - 2x (6-core) Intel Xeon processor X5670 at 2.93 GHz

      A maximum of 4 GHz with Turbo Boost Technology and 24 logic cores with Hyper-Threading Technology (HTT) active
    - 48-GB RAM
    - 250-GB HDD (WD2502ABYS-23B7A; SATA 3 Gbps)
  - Intel (QLogic) True Scale QDR InfiniBand Switch and HCAs
  - 1-Gbps Ethernet Switch

- Software:
  - Red Hat Enterprise Linux 6.2 Server (x86_64)
  - EPEL repository
  - Intel (QLogic) InfiniBand OFED driver 7.0.1.0.43 (OFED version 1.5.3.2)
  - IBM InfoSphere BigInsights V1.4
  - IBM Platform Symphony Advanced Edition V5.2

This cluster configuration is appropriate only to validate the steps to integrate InfoSphere BigInsights with Platform Symphony MapReduce. It is not ideal for data-intensive performance benchmarks nor as a reference architecture for running IBM Platform Symphony Advanced Edition or the IBM InfoSphere BigInsights Map Reduce workload.

The cluster configuration has only one disk per node and is shared between partitions that are used for the operating system and for the Hadoop Distributed File System (HDFS). The HDFS that is configured in this example is not the recommended just a bunch of disks (JBOD) design. The JBOD design must have one disk per processor core on each data node, with each disk configured as a separate device. Also, the operating system drive was not mirrored for high availability.

The 10-Gbps network is used usually for performance configuration of the InfoSphere BigInsights cluster. This environment used the IP over InfiniBand (IPoIB) network that is configured over the Intel (QLogic) QDR InfiniBand interconnect in the cluster environment.

Preconfiguration on all nodes

Before we started the integration process, we ran or checked the preliminary configurations. Some of them are described in *Implementing IBM InfoSphere BigInsights on System x*, SG24-8077, which was used as a reference for this paper. The details were based on the recommended settings for the machine environment and the software that were used in the integration. Therefore, the integration might change for other architecture and software configurations.

This section addresses configuration of the following areas:

- Cluster layout
- Cluster connectivity
- LVM-based file system for HDFS on all data nodes
- Customizing RHEL services
- Verifying whether all required packages are installed
- Creating users (same user configuration across all nodes)
- Temporary security changes

Cluster layout

When you build your cluster, you can install each server manually from a CD or DVD or by using a provisioning software such as IBM Extreme Cloud Administration Toolkit (xCAT). In either case, several combinations are possible when you size a cluster in terms of the number of networks and selecting which network to handle a specified workload. For InfoSphere BigInsights and Symphony integration, you can use for the user *homes*:

- Lightweight Directory Access Protocol (LDAP)
- Shared directories
  - General Parallel File System (IBM GPFS™)

    Consider using a separate network when you configure GPFS for optimal performance. GPFS can be used with the IBM Platform Symphony network (in high-performance networks such as InfiniBand). However, you must configure GPFS to guarantee less performance degradation on the IBM Platform Symphony network.

  – Network File System (NFS)

    Consider using a separate network when you configure NFS for optimal performance. If NFS is mixed with the same network as IBM Platform Symphony, performance can degrade.

- Local directories

  This integration uses the local home directories layout and uses the parallel distributed shell utility, **PDSH**, to distribute files across the nodes.

If the IBM Platform Symphony installation has more than one management node, a shared file system is required among the management nodes for failover and recovery of resource and workload runtime states. GPFS and NFS (on an external NFS server) are possible options for the shared file system.

Cluster connectivity

When you build the cluster, check the following areas to ensure connectivity:

1. Distribute SSH keys to set up authentication without a password (*passwordless authentication*) for root and other uses across all nodes:

   – SSH key setup for user root under a local host (Example 1)

   Example 1 SSH key setup on management node (for root user)

```
# ssh-keygen -t rsa -N ""
# cat /root/.ssh/id_rsa.pub > /root/.ssh/authorized_keys
# chmod 600 /root/.ssh/authorized_keys
```

   root SSH keys: If you do not have a provisioning system that can distribute root keys for passwordless authentication, use **scp** on the `.ssh` directory to all nodes before you continue. Otherwise, each time you issue an SSH connection to a host, you are prompted for the password.

   – SSH key setup for a specified user and for a group of nodes that use **PDSH** (Example 2)

   Example 2 Generate and distribute SSH keys over the compute nodes with PDSH

```
management # su - <user>
<user>@management # ssh-keygen -t rsa -N ""
<user>@management # cat ~/.ssh/id_rsa.pub > ~/.ssh/authorized_keys
<user>@management # chmod 600 ~/.ssh/authorized_keys
<user>@management # exit
management # pdsh “scp -r management:/home/<user>/.ssh /home/<user>/”
```

2. Consider adding repository access to all nodes:

   a. Add the extra packages repository, from the Fedora project, for Red Hat Enterprise Linux (RHEL) on the management node and later on the compute nodes after you install and configure **PDSH** (Example 3).

   Example 3 Repository for extra packages available in the Fedora project

```
# cat /etc/yum.repos.d/epel.repo
[epel]
name=Extra Packages for Enterprise Linux 6 - $basearch
```

```
#baseurl=http://download.fedoraproject.org/pub/epel/6/$basearch
mirrorlist=https://mirrors.fedoraproject.org/metalink?repo=epel-6&arch=$base
arch
failovermethod=priority
enabled=1
priority=10
gpgcheck=0
gpgkey=
```

b. Configure a distributed shell.

c. Set up PDSH for the management node (Example 4).

Example 4 Setup PDSH and verify it on management host

```
management # yum install -y pdsh
management# cat /etc/dsh/all
host1
host2
management # cat /etc/hosts
1.1.2.1 management
1.1.1.1 host1
1.1.1.2 host2
management # cat /etc/profile.d/pdsh.sh
export WCOLL=/etc/dsh/all
management # pdsh hostname
host1: host1
host2: host2
```

d. Set up **PDSH** for the compute nodes (Example 5).

The *WCOLL* environment variable (Example 5) is set to the name of a file that includes a list of the host names of the compute nodes and the target nodes of the **PDSH** commands.

Example 5 PDSH setup on compute nodes from the management host (already with PDSH)

```
management # pdsh "scp 1.1.2.1:/etc/hosts /etc/hosts"
management # pdsh "cat /etc/hosts"
host1: 1.1.2.1 management
host1: 1.1.1.1 host1
host1: 1.1.1.2 host2
host2: 1.1.2.1 management
host2: 1.1.1.1 host1
host2: 1.1.1.2 host2
management # pdsh "scp management:/etc/yum.repos.d/epel.repo
/etc/yum.repos.d/epel.repo"
management # pdsh "yum install -y pdsh"
management # pdsh "scp management:/etc/dsh/all /etc/dsh/all"
management # pdsh "cat /etc/dsh/all"
host1: host1
host1: host2
host2: host1
host2: host2
management # pdsh "scp management:/etc/profile.d/pdsh.sh
/etc/profile.d/pdsh.sh"
management # pdsh "cat /etc/profile.d/pdsh.sh"
```

```
host1: export WCOLL=/etc/dsh/all
host2: export WCOLL=/etc/dsh/all
```

LVM-based file system for HDFS on all data nodes

The HDFS is created on the ext4 LVM-based file system. For better performance, the partition is mounted with the `noatime` option (disables `atime` updates for read access) to reduce disk I/O. The `noatime` option is added to the `/etc/fstab` path (Example 6) for persistent configuration.

Example 6 Adding the noatime option to /etc/fstab

```
# cat /etc/fstab
/dev/mapper/system-root /                       ext4 defaults        1 1
UUID=****************** /boot                   ext3 defaults        1 2
/dev/mapper/system-data /data                   ext4 defaults,noatime 1 2
tmpfs                   /dev/shm                tmpfs defaults        0 0
devpts                  /dev/pts                devpts gid=5,mode=620  0 0
sysfs                   /sys                    sysfs defaults        0 0
proc                    /proc                   proc    defaults        0 0
```

The settings are changed within the file system that is mounted (Example 7).

Example 7 Remounted file system with the noatime option

```
# mount |grep data
/dev/mapper/system-data on /data type ext4 (rw)
# mount -o remount,noatime /data
# mount |grep data
/dev/mapper/system-data on /data type ext4 (rw,noatime)
```

Customizing RHEL services

Disable some RHEL services (NTP, firewall, and SELINUX) in exchange for better performance:

1. Configure the NTP service as shown in Example 8.

   Example 8 NTP service configuration file

   ```
   # cat /etc/ntp.conf
   server firstserver
   driftfile /var/lib/ntp/drift
   disable auth
   restrict 127.0.0.1
   server secondserver
   ```

2. Disable SELINUX as shown in Example 9. Restart the system to apply the changes.

   Example 9 Disabling the SELINUX in the /etc/selinux/config file

   ```
   # cat /etc/selinux/config
   # This file controls the state of SELinux on the system.
   # SELINUX= can take one of these three values:
   #       enforcing - SELinux security policy is enforced.
   #       permissive - SELinux prints warnings instead of enforcing.
   #       disabled - SELinux is fully disabled.
   SELINUX=disabled
   # SELINUXTYPE= type of policy in use. Possible values are:
   ```

```
#       targeted - Only targeted network daemons are protected.
#       strict - Full SELinux protection.
SELINUXTYPE=targeted
```

3. Disable the firewall for both IPv4 and IPv6 to improve performance (Example 10).

Example 10 Disabling the firewall services (startup and currently running)

```
# chhkconfig iptables off
# chhkconfig ip6tables off
# service iptables stop
# service ip6tables stop
```

4. Disable the IPv6 feature that is required for the tests.

Verifying whether all required packages are installed

Run through all requirements on all nodes. To determine the software requirements for InfoSphere BigInsights, see *Implementing IBM InfoSphere BigInsights on System x*, SG24-8077. Also, install **`expect`**. For RHEL 6.2, we installed `expect-5.44.1.15-2.el6.x86_64`.

Creating users (same user configuration across all nodes)

Create the egoadmin user and group with the same ID on all nodes (Example 11). We use this user to install IBM Platform Symphony Advanced Edition and InfoSphere BigInsights.

Example 11 Creating the egoadmin user across all nodes

```
management # useradd -u 1000 -m egoadmin
management # pdsh "useradd -u 1000 -m egoadmin"
```

After the integration, you might need to create more users to work with the integrated software. For more information, see “Adding users” on page 27.

Temporary security changes

During the InfoSphere BigInsights installation, complete the security changes to allow the egoadmin user to run programs with the security privileges of root before proceeding. Using the **`visudo`** command, make the following changes:

1. Add the following line:

   `egoadmin ALL=(ALL) NOPASSWD: ALL`

2. Change `Defaults requiretty` to `Defaults !requiretty`.

Configuring InfoSphere BigInsights

Set up InfoSphere BigInsights on 10 data nodes and 1 management node:

> **More information:** For a reference about the requirements and installation procedures for the InfoSphere BigInsights configuration, see the *Implementing IBM InfoSphere BigInsights on System x*, SG24-8077.

1. Start the web-based GUI installation setup on the management node, and generate the XML file based on the preferences chosen:
   a. In the Welcome window, click **Next**.
   b. Select **I accept the terms in the license agreement**, and click **Next**.
   c. Select **Create a response file without performing an installation**, and click **Next**.
2. Review the generated XML, and change it according to the desirable requirements.

   InfoSphere BigInsights is installed (from the management node) by using the modified XML without errors (highlighted in a bold font in Example 12 on page 10).

   The host name of the cluster node in the test configuration is set to the high-performance network interface. For example, instead of setting i05i04 to use the Gigabit Ethernet network, set i05i04 to use IPoIB.

Example 12 InfoSphere BigInsights modified fullinstall.xml file parts used after GUI generation

```
<?xml version="1.0" encoding="UTF-8"?>
<cluster-configuration>
    <operation>install</operation>
    <vendor>ibm</vendor>
[...]
<hadoop>
         <datanode>
            <selection-type>Specified</selection-type>

<nodes>i05i06,i05i07,i05i08,i05i09,i05i10,i05i14,i05i15,i05i16,i05i17,i05i18</n
odes>
            <datanode-port>50010</datanode-port>
            <datanode-ipc-port>50020</datanode-ipc-port>
            <datanode-http-port>50075</datanode-http-port>
            <tasktracker-http-port>50060</tasktracker-http-port>
            <data-directory>/data/hdfs</data-directory>
         </datanode>
</hadoop>
[...]
<HBase>
        <configure>true</configure>
        <zookeeper-mode>shared</zookeeper-mode>
        <master-nodes>i05i04</master-nodes>
        <install-mode>fully</install-mode>

<region-nodes>i05i06,i05i07,i05i08,i05i09,i05i10,i05i14,i05i15,i05i16,i05i17,i0
5i18</region-nodes>
        <root-directory>/opt/ibm/bi/hbase</root-directory>
        <log-directory>/var/log/bi/hbase/logs</log-directory>
        <master-port>60000</master-port>
        <master-ui-port>60010</master-ui-port>
        <regionserver-port>60020</regionserver-port>
        <regionserver-ui-port>60030</regionserver-ui-port>
</HBase>
<node-list>
        <node>
            <name-or-ip>i05i04</name-or-ip>
            <password>{xor}</password>
            <rack/>
            <hdfs-data-directory>/data/hdfs</hdfs-data-directory>
```

```
        <gpfs-node-designation/>
        <gpfs-admin-node/>
        <gpfs-rawdisk-list/>
    </node>
    <node>
        <name-or-ip>i05i06</name-or-ip>
        <password>{xor}</password>
        <rack/>
        <hdfs-data-directory>/data/hdfs</hdfs-data-directory>
        <gpfs-rawdisk-list/>
    </node>
    <node>
        <name-or-ip>i05i07</name-or-ip>
        <password>{xor}</password>
        <rack/>
        <hdfs-data-directory>/data/hdfs</hdfs-data-directory>
        <gpfs-rawdisk-list/>
    </node>
    <node>
        <name-or-ip>i05i08</name-or-ip>
        <password>{xor}</password>
        <rack/>
        <hdfs-data-directory>/data/hdfs</hdfs-data-directory>
        <gpfs-rawdisk-list/>
    </node>
    <node>
        <name-or-ip>i05i09</name-or-ip>
        <password>{xor}</password>
        <rack/>
        <hdfs-data-directory>/data/hdfs</hdfs-data-directory>
        <gpfs-rawdisk-list/>
    </node>
    <node>
        <name-or-ip>i05i10</name-or-ip>
        <password>{xor}</password>
        <rack/>
        <hdfs-data-directory>/data/hdfs</hdfs-data-directory>
        <gpfs-rawdisk-list/>
    </node>
    <node>
        <name-or-ip>i05i14</name-or-ip>
        <password>{xor}</password>
        <rack/>
        <hdfs-data-directory>/data/hdfs</hdfs-data-directory>
        <gpfs-rawdisk-list/>
    </node>
    <node>
        <name-or-ip>i05i15</name-or-ip>
        <password>{xor}</password>
        <rack/>
        <hdfs-data-directory>/data/hdfs</hdfs-data-directory>
        <gpfs-rawdisk-list/>
    </node>
    <node>
        <name-or-ip>i05i16</name-or-ip>
```

```
            <password>{xor}</password>
            <rack/>
            <hdfs-data-directory>/data/hdfs</hdfs-data-directory>
            <gpfs-rawdisk-list/>
        </node>
        <node>
            <name-or-ip>i05i17</name-or-ip>
            <password>{xor}</password>
            <rack/>
            <hdfs-data-directory>/data/hdfs</hdfs-data-directory>
            <gpfs-rawdisk-list/>
        </node>
        <node>
            <name-or-ip>i05i18</name-or-ip>
            <password>{xor}</password>
            <rack/>
            <hdfs-data-directory>/data/hdfs</hdfs-data-directory>
            <gpfs-rawdisk-list/>
        </node>
</node-list>
```

3. Run the **healthcheck.sh** script inside `$HADOOP_HOME/bin` directory, and validate whether all components are working as expected (no FAILED components). Example 13 shows the expected result at the end of the script.

Example 13 Expected end of output for the helthcheck.sh script

```
[...]
[INFO] Progress - 100%
[INFO] DeployManager - Start; SUCCEEDED components: [guardiumproxy, zookeeper,
hadoop, derby, hive, hbase, flume, oozie, orchestrator, jaqlserver, console,
sheets]; FAILED components: []
```

Installing IBM Platform Symphony Advanced Edition

To install IBM Platform Symphony Advanced Edition:

1. On the master node, edit the `/etc/bashrc` file, and add the following environment variables:

```
export HADOOP_HOME=/opt/ibm/bi/IHC
export JAVA_HOME=/opt/ibm/bi/jdk
export PATH=$PATH:$HADOOP_HOME/bin
export CLUSTERNAME=<cluster_name>
export CLUSTERADMIN=egoadmin
export HADOOP_VERSION=1_0_0
export HDFS_URL=http://<HDF_NAME_NODE>:50070
export SIMPLIFIEDWEM=N
```

The HADOOP_HOME and JAVA_HOME environment variables are defined for the integration with InfoSphere BigInsights that is being installed in the `/opt/ibm/bi` directory. After you edit the file, copy it to all the nodes in the cluster (the host names that are listed in the `$WCOLL` file) by using the **pdcp** command:

```
pdcp /etc/bashrc /etc/bashrc
```

2. On all nodes, restart any currently open shells for the new environment variables to take effect.

3. On the master node, change directory (**cd**) to the directory where the IBM Platform Symphony installation images were downloaded. Install IBM Platform Symphony:

```
./symSetup5.2.0_lnx26-lib23-x64.bin --dbpath /opt/ibm/sym52/rpmdb --prefix
/opt/ibm/sym52 --quiet
```

   The default port 8080 used by the Platform Management Consoles (PMC) is already being used by the BigInsight console.

4. On the master node, change the PMC port to the value `"18080"` (including the quotation marks) in the `$EGO_TOP/gui/conf/server.xml` file.

   Example 14 shows the original `server.xml` file.

   Example 14 Original server.xml file

```
<!-- Define a non-SSL HTTP/1.1 Connector on port 8080 -->
    <Connector port="8080" maxHttpHeaderSize="8192"
               maxThreads="150" minSpareThreads="25" maxSpareThreads="75"
               enableLookups="false" redirectPort="8443" acceptCount="100"
               connectionTimeout="20000" disableUploadTimeout="true" />
```

   Example 15 shows the edited `server.xml` file.

   Example 15 Changed server.xml file

```
<!-- Define a non-SSL HTTP/1.1 Connector on port 8080 -->
    <Connector port="18080" maxHttpHeaderSize="8192"
               maxThreads="150" minSpareThreads="25" maxSpareThreads="75"
               enableLookups="false" redirectPort="8443" acceptCount="100"
               connectionTimeout="20000" disableUploadTimeout="true" />
```

5. Initiate the master node, set the entitlement file, and start IBM Platform Symphony:

   > **Entitlement file:** Copy the entitlement file to a location where only the root user has access and it does not change when IBM Platform Symphony needs it for installation. In this example, we used the IBM Platform Symphony `$EGO_TOP` installation directory.

```
su - egoadm
. $EGO_TOP/profile.platform
egoconfig join <clustername>
egoconfig setentitlement $EGO_TOP/sym.entitlement
exit
. $EGO_TOP/profile.platform
egosh ego start
```

6. On the compute nodes, install IBM Platform Symphony:

```
/opt.db06b04/FILES/symcompSetup5.2.0_lnx26-lib23-x64.bin --dbpath
/opt/ibm/sym52/pmrdb --prefix /opt/ibm/sym52 --quiet
```

7. On all nodes, change the IBM Platform Symphony configuration to enable **egosh** commands to use SSH connections (with no password) instead of RSH connections. Add the following line to the `$EGO_CONFDIR/ego.conf` file as shown in Example 16 on page 14:

```
EGO_RSH="ssh -o 'PasswordAuthentication no' -o 'StrictHostKeyChecking no'"
```

Example 16 Line added at the end of the $EGO_CONFDIR/ego.conf file

```
[...]
EGO_VERSION=1.2.6

EGO_RSH="ssh -o 'PasswordAuthentication no' -o 'StrictHostKeyChecking no'"
```

8. Initiate the compute nodes and start IBM Platform Symphony:

```
su - egoadm
. /opt/ibm/sym52/profile.platform
egoconfig join <master_node_hostname>
exit
. /opt/ibm/sym52/profile.platform
egosh ego start
```

   If the cluster administrator egoadmin is granted root privileges, the user egoadmin can start and shut down IBM Platform Symphony on any hosts in the cluster.

9. On all nodes, as root, run the **`$EGO_ESRVDIR/egosetsudoers.sh`** command
10. Confirm the command:

```
cat /etc/ego.sudoers
```

11. On the master node, add the following line to the `/etc/bashrc` file:

```
. /opt/ibm/sym52/profile.platform
```

12. Copy the modified file to the other nodes in the cluster:

```
pdcp /etc/bashrc /etc/bashrc
```

Integrating IBM Platform Symphony and InfoSphere BigInsights

When you integrate IBM Platform Symphony and InfoSphere BigInsights, complete the following steps on the master node and on the computing nodes of the cluster.

Prerequisite: Before you integrate IBM Platform Symphony and InfoSphere BigInsights, install the **`pdsh`** utility on all nodes so that you can run the same commands across multiple nodes.

To integrate IBM Platform Symphony and InfoSphere BigInsights:

1. Stop InfoSphere BigInsights and IBM Symphony on the cluster:

```
su - egoadmin
/opt/ibm/bi/bin/stop-all.sh
$EGO_BINDIR/egoshutdown.sh
exit
```

2. On the master node, edit the `/opt/ibm/sym52/soam/mapreduce/conf/pmr-env.sh` file. Check that the following values are in the file:

```
export JAVA_HOME=/opt/ibm/bi/jdk
export HADOOP_VERSION=1_0_0
export HADOOP_HOME=/opt/ibm/bi/IHC
export PMR_EXTERNAL_CONFIG_PATH=/opt/ibm/bi/hadoop-conf
export JVM_OPTIONS=-Xmx1024m
export PMR_SERVICE_DEBUG_PORT=
export PMR_MRSS_SHUFFLE_CLIENT_PORT=7879
```

```
export PMR_MRSS_SHUFFLE_DATA_WRITE_PORT=7881
export PYTHON_PATH=/bin:/usr/bin:/usr/local/bin
export PATH=${PATH}:${JAVA_HOME}/bin:${PYTHON_PATH}
export
USER_CLASSPATH=/opt/ibm/bi/IHC/lib/*:/opt/ibm/bi/jaql/*:/opt/ibm/bi/lib/*:/opt/
ibm/bi/jaql/modules/systemT/lib/systemT-jaql.jar:/opt/ibm/bi/textanalytics/lib/
text-analytics/systemT.jar:/opt/ibm/bi/text-analytics/lib/htmlparser-2.0/htmlpa
rser.jar
export
JAVA_LIBRARY_PATH=/opt/ibm/bi/IHC/lib/native/Linux-amd64-64/:/opt/ibm/bi/IHC/li
b/native/Linux-i386-32/
export CLOUDERA_HOME=/opt/ibm/bi/IHC
```

3. After you edit the file on the master node, copy the file to the other nodes in the cluster by using the **pdcp** command:

```
pdcp /opt/ibm/sym52/soam/mapreduce/conf/pmr-env.sh
/opt/ibm/sym52/soam/mapreduce/conf/pmr-env.sh
```

4. Apply the `BIIntegrationPatch-201206291122.tar.gz` patch by using the **patchHost** shell script. In this script, the *BI_Integration_Patch* variable defines the absolute path for the patch file. This example uses the `/tmp/BIIntegrationPatch-201206291122.tar.gz` path. Example 17 shows the **patchHost** script.

Example 17 The patchHost script

```
#!/bin/bash
#
# patchHost: Run the Big Insights / Symphony Integration patch
#
#  Change these values for your site
admin_id="egoadmin"
symdir="/opt/ibm/sym52"
bidir="/opt/ibm/bi"
BI_Integration_Patch="/tmp/BIIntegrationPatch-201206291122.tar.gz"

#  This needs to be run as root
if [ "$(id -un)" != "$admin_id" ]; then
   echo "This script must be run as $admin_id"
   exit 1
fi

if [ ! -f $symdir/profile.platform ] ; then
   echo "This symphony profile ($symdir/profile.platform) is missing "
   echo "Is symphony installed at $symdir?"
   exit 1
fi

. $symdir/profile.platform

if [ ! -f $BI_Integration_Patch ] ; then
   echo "This Symphony/BigInsights patch file($BI_Integration_Patch) is missing"
   exit 1
fi
if [ ! -d "$bidir" ] ; then
   echo "The Big Insights directory ($bidir)does not exist "
   exit 1
fi
```

```
if [ ! -d "$HADOOP_HOME" ] ; then
   echo "The HADOOP_HOME directory does not exist or HADOOP_HOME is not set"
   echo "HADOOP_HOME=$HADOOP_HOME"
   exit 1
fi
if [ -d "$symdir/bi_integration" ] ; then
   echo "The directory $symdir/bi_integration ALREADY exists. "
   echo "You cannot run this script twice"
   exit 1
fi
if [ ! -f "$HADOOP_HOME/hadoop-core-1.0.0.jar" ] ; then
  echo "File $HADOOP_HOME/hadoop-core-1.0.0.jar is missing"
  exit 1
fi
if [ -l "$HADOOP_HOME/hadoop-core-1.0.0.jar" ] ; then
  echo "File $HADOOP_HOME/hadoop-core-1.0.0.jar is a link"
  exit 1
fi

echo tar -C $symdir -xvzf $BI_Integration_Patch
tar -C $symdir -xvzf $BI_Integration_Patch
echo $EGO_TOP/bi_integration/jar_integration.sh
$EGO_TOP/bi_integration/jar_integration.sh
echo mv $HADOOP_HOME/hadoop-core-1.0.0.jar $HADOOP_HOME/hadoop-core-1.0.0.jar.ORIG
mv $HADOOP_HOME/hadoop-core-1.0.0.jar $HADOOP_HOME/hadoop-core-1.0.0.jar.ORIG
echo ln -sf $EGO_TOP/bi_integration/IBM-pmr-hadoop-1.0.0.jar $HADOOP_HOME/hadoop-core-1.0.0.jar
ln -sf $EGO_TOP/bi_integration/IBM-pmr-hadoop-1.0.0.jar $HADOOP_HOME/hadoop-core-1.0.0.jar
echo ln -sf $PMR_HOME/conf/pmr-site.xml $bidir/hadoop-conf/
ln -sf $PMR_HOME/conf/pmr-site.xml $bidir/hadoop-conf/
hlibdir=$symdir/soam/mapreduce/5.2/linux2.6-glibc2.3-x86_64/lib/hadoop-1.0.0

echo mv $hlibdir/jackson-core-asl-1.0.1.jar $hlibdir/jackson-core-asl-1.0.1.jar.bak
mv $hlibdir/jackson-core-asl-1.0.1.jar $hlibdir/jackson-core-asl-1.0.1.jar.bak
echo ln -sf $bidir/sheets/libext/jackson-core-asl-1.5.5.jar $hlibdir/jackson-core-asl-1.0.1.jar
ln -sf $bidir/sheets/libext/jackson-core-asl-1.5.5.jar $hlibdir/jackson-core-asl-1.0.1.jar
echo mv $hlibdir/jackson-mapper-asl-1.0.1.jar $hlibdir/jackson-mapper-asl-1.0.1.jar.bak
mv $hlibdir/jackson-mapper-asl-1.0.1.jar $hlibdir/jackson-mapper-asl-1.0.1.jar.bak
echo ln -sf $bidir/sheets/libext/jackson-mapper-asl-1.5.5.jar
$hlibdir/jackson-mapper-asl-1.0.1.jar
ln -sf $bidir/sheets/libext/jackson-mapper-asl-1.5.5.jar $hlibdir/jackson-mapper-asl-1.0.1.jar
```

5. On the master node, run the script as follows:
   a. Switch to the egoadmin user:
      ```
      su - egoadmin
      ```
   b. Run the script:
      ```
      ./patchHost
      ```
6. After the script runs, run it on the other nodes in the cluster by running the following commands on the master node:
   a. Copy the script (patchHost) and the patch to the other hosts:
      ```
      pdcp /tmp/patchHost /tmp/patchHost
      pdcp /tmp/Install/Symphony/patch/BIIntegrationPatch-201206291122.tar.gz
      /tmp/Install/Symphony/patch/BIIntegrationPatch-201206291122.tar.gz
      ```

b. Run the **patchHost** script:

```
pdsh /tmp/patchHost
```

7. Run the following commands on the master node to apply control shims:

```
ln -sf $PMR_HOME/conf/pmr-site.xml
/opt/ibm/bi/console/wtiruntime/dscomponents/enterpriseconsole/eclipse/plugins/c
om.ibm.hadoop_1.0.0.v20120605_0341/
ln -sf $EGO_TOP/bi_integration/IBM-pmr-hadoop-1.0.0.jar
/opt/ibm/bi/console/wtiruntime/dscomponents/enterpriseconsole/eclipse/plugins/c
om.ibm.hadoop_1.0.0.v20120605_0341/hadoop-core-1.0.0.jar
ln -sf /opt/ibm/sym52/bi_integration/IBM_MANIFEST.MF
/opt/ibm/bi/console/wtiruntime/dscomponents/enterpriseconsole/eclipse/plugins/c
om.ibm.hadoop_1.0.0.v20120605_0341/META-INF/MANIFEST.MF
```

8. Run the following commands to add the oozie integration override:

```
ln -s /opt/ibm/bi/IHC/lib/servlet-api-2.5-6.1.14.jar \
/opt/ibm/bi/oozie/wasce/lib/endorsed
```

9. Modify the `/opt/ibm/bi/hive/bin/hive` file on the master node, and then copy it to the other nodes:

a. Back up the old file:

```
cp /opt/ibm/bi/hive/bin/hive /opt/ibm/bi/hive/bin/hive.ORIG
```

b. Open the file and search for the *hadoop_version_re* variable. Replace its value with the `"^([[:digit:]]+)[\.\_]([[:digit:]]+)([\.\_]([[:digit:]]+))?.*$"` value (including the quotation marks).

c. To confirm the changes, enter the **diff** command to see the results:

```
$ diff hive.ORIG hive
240c240
< hadoop_version_re="^([[:digit:]]+)\.([[:digit:]]+)(\.([[:digit:]]+))?.*$"
---
>
hadoop_version_re="^([[:digit:]]+)[\.\_]([[:digit:]]+)([\.\_]([[:digit:]]+))
?.*$"
```

d. Copy the modified file to the other nodes in the cluster:

```
pdcp /opt/ibm/bi/hive/bin/hive /opt/ibm/bi/hive/bin/hive
```

10. Start InfoSphere BigInsights and IBM Platform Symphony:

```
su - egoadmin
/opt/ibm/bi/bin/start-all.sh
$EGO_BINDIR/egosh ego start -f all
```

11. Stop Jobtracker and Tasktracker:

```
su - egoadmin
/opt/ibm/bi/IHC/bin/stop-mapred.sh
```

Additional configuration for IBM Platform Symphony

After the integration, you must make some additional changes. These changes depend on the cluster size and complexity. Therefore, the values might be different for larger clusters. You have the following options:

- Identifying the management nodes in the IBM Platform Symphony configuration
- TCP/IP performance tuning

Identifying the management nodes in the IBM Platform Symphony configuration

In the IBM Platform Symphony 5.2 installation, the management nodes are incorrectly identified. Therefore, the default configuration uses all nodes as compute nodes. There are two ways you can change this setting:

- Manually select which nodes are the compute nodes.
- Change the configuration and specify which nodes are the management nodes (recommended solution).

To change the configuration and specify the management nodes:

1. Edit the `$EGO_CONFDIR/ego.cluster.`*`cluster_name`* file.
2. For all management nodes, add the `mg` tag inside the resources column as illustrated in Example 18. The *`cluster_name`* is the name of the symphony cluster. You can check the name by using the **`egosh ego info`** command (under the user egoadmin).

Example 18 The ego.cluster.<cluster_name> file with i05i04 as management node

```
# $Id: TMPL.ego.cluster,v 1.3.28.3.2.1.2.1.92.4.2.1.18.4 2012/05/22 09:38:44
yaoliu Exp $
#-------------------------------------------------------------------------
# T H I S   I S   A   O N E   P E R   C L U S T E R   F I L E
#
# This is a sample ego cluster definition file.  This file's name
# should be ego.cluster.<cluster-name>.
#

Begin   ClusterAdmins
Administrators = egoadmin
End    ClusterAdmins

Begin Host
HOSTNAME  model    type        r1m  mem  swp  RESOURCES    #Keywords
i05i04    !        !            -    -    -   (linux mg)
#lemon    PC200    LINUX86     3.5  1    2   (linux)
#plum     !        NTX86       3.5  1    2   (nt)
End     Host
```

Example 19 shows the output of this command.

Example 19 Checking the cluster name and master node for IBM Platform Symphony

```
[egoadmin@i05i04 ~]$ egosh ego info
Cluster name            : bisym
EGO master host name    : i05i04
EGO master version      : 1.2.6
```

3. After you identify all management nodes and change the required lines in the `$EGO_CONFDIR/ego.cluster.`*`cluster_name`* file, bring down all changed management nodes.

TCP/IP performance tuning

Some TCP/IP parameters can help tune the IBM Platform Symphony environment and possibly improve workload performance. The values that are shown were tested in the cluster for the tests that are documented in this paper. As a result, you might need to adjust them for the cluster sizes, architectures, and operating system in your environment.

Change the **`sysctl`** attributes on all nodes, and add them to the `/etc/sysctl.conf` file:

```
sysctl -w net.ipv4.ipfrag_low_thresh=1048576
sysctl -w net.ipv4.ipfrag_high_thresh=8388608
```

If you have InfiniBand adapters and installed the InfiniBand drivers, you might have the **`/sbin/sysctl_perf_tuning`** script. The package name that was installed in the test environment was named `kernel-ib-1.5.3.2-2.6.32_220.el6.x86_64.x86_64` and represents the InfiniBand driver and ULP kernel modules.

If you have the **`/sbin/sysctl_perf_tuning`** script:

1. Run the RHEL 6.2 **`sysctl_perf_tuning`** script on all nodes. The script persistently changes the values (Example 20).

Example 20 System parameters that changed with the /sbin/sysctl_perf_tuning command

```
/sbin/sysctl -q -w net.ipv4.tcp_timestamps=0
/sbin/sysctl -q -w net.ipv4.tcp_sack=0
/sbin/sysctl -q -w net.core.netdev_max_backlog=250000
/sbin/sysctl -q -w net.core.rmem_max=16777216
/sbin/sysctl -q -w net.core.wmem_max=16777216
/sbin/sysctl -q -w net.core.rmem_default=16777216
/sbin/sysctl -q -w net.core.wmem_default=16777216
/sbin/sysctl -q -w net.core.optmem_max=16777216
/sbin/sysctl -q -w net.ipv4.tcp_mem="16777216 16777216 16777216"
/sbin/sysctl -q -w net.ipv4.tcp_rmem="4096 87380 16777216"
/sbin/sysctl -q -w net.ipv4.tcp_wmem="4096 65536 16777216"
```

2. After you run the script, if you are not satisfied with the performance results, use the script example lines to manually change the value again through all nodes. Then, check for improvements.

Benchmark tests

This section presents the benchmark results before and after we integrated IBM Platform Symphony Advanced Edition 5.2 with InfoSphere BigInsights 1.4. All benchmark results were obtained by using the following test environment for both hardware and software:

- Eleven hardware machines were used as described in "Environment" on page 4. One machine was for the management node, and ten machines were for the data nodes of InfoSphere BigInsights (and the same ten for compute nodes on IBM Platform Symphony).
- All systems run RHEL 6.2 (x86_64) with InfoSphere BigInsights 1.4 and IBM Platform Symphony Advanced Edition 5.2.
- All configurations after this section used InfiniBand (IPoIB) as the network for InfoSphere BigInsights (Hadoop data) and Ethernet for the IBM Platform Symphony.

The following sections describe the tests and their specific configurations:

- Default configuration, which uses the tasks that are described in "Identifying the management nodes in the IBM Platform Symphony configuration" on page 18, but *not* the changes in "TCP/IP performance tuning" on page 19.
- Tuned configuration (with TCP tuning), which uses the tasks that are described in "Identifying the management nodes in the IBM Platform Symphony configuration" on page 18, *and* the changes in "TCP/IP performance tuning" on page 19.
- Preloaded configuration (with TCP tuning and prestart option enabled), which is the same as the tuned configuration, but with the Prestart option enabled on IBM Platform Symphony.

> **Tuned and preloaded configurations:** To reduce the debug level, you can optionally add the following line (Example 21 on page 23):
>
> `-Dmapred.map.child.log.level=WARN`

Test environment and monitoring tools

This section highlights the tuning options and benchmark results from the tests in the stand-alone setup of InfoSphere BigInsights 1.4 and after its integration with IBM Platform Symphony 5.2.

During the benchmarking tests, the following tools and web pages were used to extract the results and to help monitor the progress of each test:

- NMON
- The **`time`** command from RHEL 6.2 (**`time-1.7-37.1.el6.x86_64`**)
- The **`script`** command from RHEL 6.2 (**`util-linux-ng-2.17.2-12.4.el6.i686`**)
- InfoSphere BigInsights cluster web page:
  - HDFS usage:

    `http://i05n04.pbm.ihost.com:50070/dfshealth.jsp`
- IBM Platform Symphony cluster web pages:
  - Cluster information

    `http://i05n04.pbm.ihost.com:18080/platform/dealUserLogin.do`
  - MapReduce information

    `http://i05n04.pbm.ihost.com:18080/pmrgui/framework/main.action`

Results of the stand-alone benchmark for InfoSphere BigInsights V1.4

The following benchmark tests were conducted over InfiniBand:

- Sleep benchmark
- Pi estimate benchmark

Sleep benchmark

The sleep benchmark (scheduling benchmark) tests the performance of the scheduler. It examines how quickly the scheduler can create tasks after each other and how efficiently the scheduler can detect where (which node) to create. In this case, the scheduler is the InfoSphere BigInsights scheduler. The following command is issued:

```
time hadoop jar $HADOOP_HOME/../pig/test/e2e/pig/lib/hadoop-examples.jar sleep -mt 1 -rt 1 -m <tasks> -r 1
```

Where `<tasks>` is the number of map tasks to be created.

Table 1 shows the execution time for the sleep tests.

Table 1 Execution time for the sleep tests

Number of map tasks	Number of runs	Time observed[a]		Performance in tasks/second	
		Minimum	Maximum	Minimum	Maximum
1000	5	28s	33s	30/30	35/71
5000	5	1m 31s	1m 38s	51/02	54/95
10000	1		2m 53s	57/80	
50000	1		13m 41s	60/90	
100000	2	Crashed or hung twice			

a. +/- 0.5 seconds error

Pi estimate benchmark

Table 2 shows the execution time for the Pi estimate tests (mixed I/O read and processor intensive benchmarks). The PI estimate benchmark uses the following command:

```
time hadoop jar $HADOOP_HOME/hadoop-examples-1.0.0.jar pi <nMaps> <nSamples>
```

Where `<nMaps>` is the number of map tasks to be created, and `<nSamples>` is the number of iterations for each task (will consume more processor time).

Table 2 Execution time for the Pi estimate tests

Number of map tasks	Number of samples per map	Pi estimate time in seconds (without input generation)	Complete job time observed (minimum[a])	Pi result
10000	10000000	221.669	7m 35s	3.14159265584
1000	10000000	36.538	1m 03s	3.1415926452
	100000000	80.404	1m 45s	3.14159265572
100	10000000	17.497	22s	3.141592736
	100000000	28.85	33s	3.1415926492

Number of map tasks	Number of samples per map	Pi estimate time in seconds (without input generation)	Complete job time observed (minimum[a])	Pi result
10	10000000	17.421	19s	3.14159256
	100000000	29.412	31s	3.141592744
1	100	17.592	19s	3.2
	10000	17.424	19s	3.1408
	100000	16.431	18s	3.1412
	1000000	16.448	18s	3.141552
	10000000	18.433	20s	3.1415844
	100000000	25.448	27s	3.14159256
	1000000000	87.621	1m29s	3.14159272

a. +/- 0.5 seconds error

Results of integration for IBM Platform Symphony V5.2 and InfoSphere BigInsights V1.4

After the integration of IBM Platform Symphony V5.2 with InfoSphere BigInsights V1.4, the same tests as in “Results of the stand-alone benchmark for InfoSphere BigInsights V1.4” on page 21, were conducted. They measured the performance difference of the IBM Platform Symphony scheduler against the InfoSphere BigInsights scheduler.

The following benchmark tests were conducted:

- Sleep benchmark
- Pi estimate benchmark (mixed I/O read and processor-intensive benchmark)

Sleep benchmark

After the integration, the InfoSphere BigInsights scheduler is no longer used. Now the IBM Platform Symphony scheduler is used. Table 3 shows the execution time for the sleep tests (scheduling benchmark).

Table 3 Execution time for the sleep tests (default configuration)

Number of map tasks	Number of runs	Time observed[a]		Performance in tasks/second	
		Minimum	Maximum	Minimum	Maximum
1000	10	8s	12s	83/33	125
5000	10	13s	14s	357/14	384/62
10000	10	15s	22s	454/55	666/67
50000	10	1m 04s	1m 53s	442/47	781/25
100000	5	3m 01s	3m 38s	458/72	552/49
200000	1		18m 52s	176/68	

a. +/- 0.5 seconds error

After we change the values as shown in “TCP/IP performance tuning” on page 19, the tests that are listed in Table 3 are repeated.

Table 4 shows the improvements.

Table 4 Execution time for sleep test (tuned configuration)

Number of map tasks	Number of runs	Time observed [a]		Performance in tasks/second	
		Minimum	Maximum	Minimum	Maximum
1000	10	7s	11s	90/91	142/86
5000	10	9s	12s	416/67	555/56
10000	10	12s	16s	625	833/34
50000	10	57s	1m 14s	675/68	877/19
100000	5	2m 49s	3m 14s	515/47	591/72
200000	2	19m 04s	22m 39s	147/17	174/83

a. +/- 0.5 seconds error

In these series of tests, the debug level was reduced by adding an optional flag attribute to the **`-Dmapred.map.child.log.level=WARN`** command line as shown in Example 21.

Example 21 Command line text to reduce debug level

```
# time hadoop jar $HADOOP_HOME/../pig/test/e2e/pig/lib/hadoop-examples.jar sleep
-Dmapred.map.child.log.level=WARN -mt 1 -rt 1 -m 5000 -r 1
```

Enabling the Prestart option

If the Prestart option is enabled for the MapReduce 5.2 application, it can improve the execution time. To enable the Prestart option:

1. Go to the following address:

   `http://<master_node_ip>:18080/platform`

2. Click the **MapReduce Workload** link on the left side of the window.
3. In the IBM Platform Symphony window, on the **Workload** tab, click **MapReduce Applications**, select the **MapReduce 5.2** application, and then, click the **Modify** button.

4. In the Application window (Figure 3), select the **Pre-start application** check box, and then click the **Save** button.

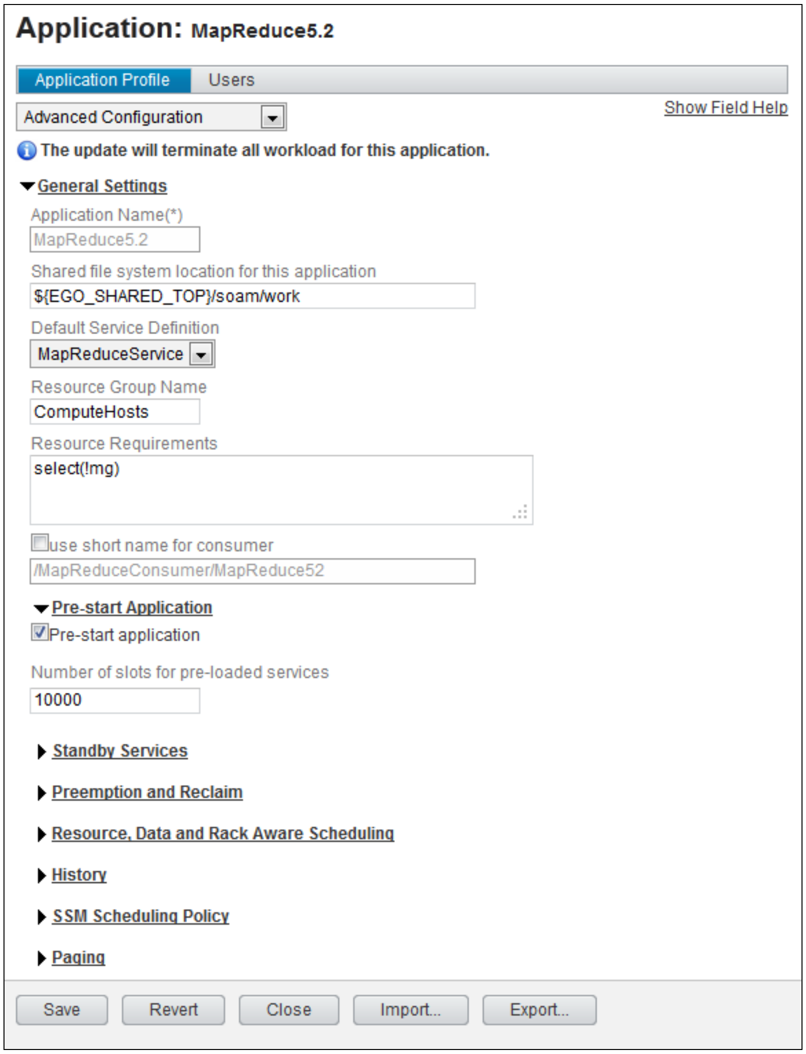

Figure 3 Pre-start option under MapReduce5.2 application properties

Table 5 shows the results of this change and of using the command-line option to reduce the debug output as shown in Example 21 on page 23.

Table 5 Execution time for sleep test (tuned configuration and the pre-start option)

		Time observed[a]		Performance in tasks/second	
Number of map tasks	**Number of runs**	**Minimum**	**Maximum**	**Minimum**	**Maximum**
1000	10	5s	7s	142/86	200
5000	10	9s	9s	555/56	555/56
10000	10	13s	14s	714/29	769/23
50000	10	59s	1m 07s	746/27	847/46
100000	3	3m11s	3m 52s	431/03	523/56

a. +/- 0.5 seconds error

Pi estimate benchmark (mixed I/O read and processor-intensive benchmark)

After the integration, we no longer use the InfoSphere BigInsights scheduler. Now we use the IBM Platform Symphony scheduler. Table 6 shows the execution time for the Pi estimate tests.

Table 6 Execution time for Pi estimate tests (default configuration)

Number of map tasks	Number of samples per map	Pi estimate time in seconds (without input generation)	Complete job time observed (minimum[a])	Pi result
100000	10000000	519.835	50m 34s	3.141592655176
	100000000	3277.363	105m 45s	3.1415926543392
10000	10000000	54.429	5m 09s	3.14159265584
	100000000	351.059	10m 14s	3.141592654996
1000	10000000	17.0	44s	3.1415926452
	100000000	54.892	1m 22s	3.14159265572
100	10000000	8.159	12s	3.141592736
	100000000	18.055	22s	3.1415926492
	1000000000	94.077	1m 38s	3.1415926568
10	10000000	9.803	11s	3.14159256
	100000000	16.134	18s	3.141592744
	1000000000	83.071	1m 25s	3.1415926644
1	100	9.763	11s	3.2
	10000	8.771	10s	3.1408
	100000	9.744	11s	3.1412
	1000000	6.731	8s	3.141552
	10000000	5.745	7s	3.1415844
	100000000	14.748	16s	3.14159256
	1000000000	72.735	1m 14s	3.14159272

a. +/- 0.5 seconds error

Advantages of integrating IBM Platform Symphony V5.2 with InfoSphere BigInsights V1.4

This section compares the results from "Results of the stand-alone benchmark for InfoSphere BigInsights V1.4" on page 21, and from "Results of integration for IBM Platform Symphony V5.2 and InfoSphere BigInsights V1.4" on page 22. This section also highlights the advantages of performing this integration.

Three configurations were used for the sleep and the Pi estimated tests as follows:

- Default configuration

  This configuration uses the steps that are described in “Additional configuration for IBM Platform Symphony” on page 18, without performing the changes in “TCP/IP performance tuning” on page 19.

- Tuned configuration (with Ethernet tuning)

  This configuration uses the “Additional configuration for IBM Platform Symphony” on page 18, and includes the changes in “TCP/IP performance tuning” on page 19.

- Preloaded configuration (with Ethernet tuning and the **Prestart** option enabled)

  This configuration is the same as the tuned configuration, but with **Prestart** option enabled from IBM Platform Symphony.

> **Tuned and preloaded configurations:** Optionally, you can add the following command-line option to reduce the debug level (Example 21 on page 23):
>
> `-Dmapred.map.child.log.level=WARN`
>
> The files are still created, but nothing is written in them.

Sleep benchmark results

For the sleep benchmark, Figure 4 compares the results that we obtained from the different test configurations.

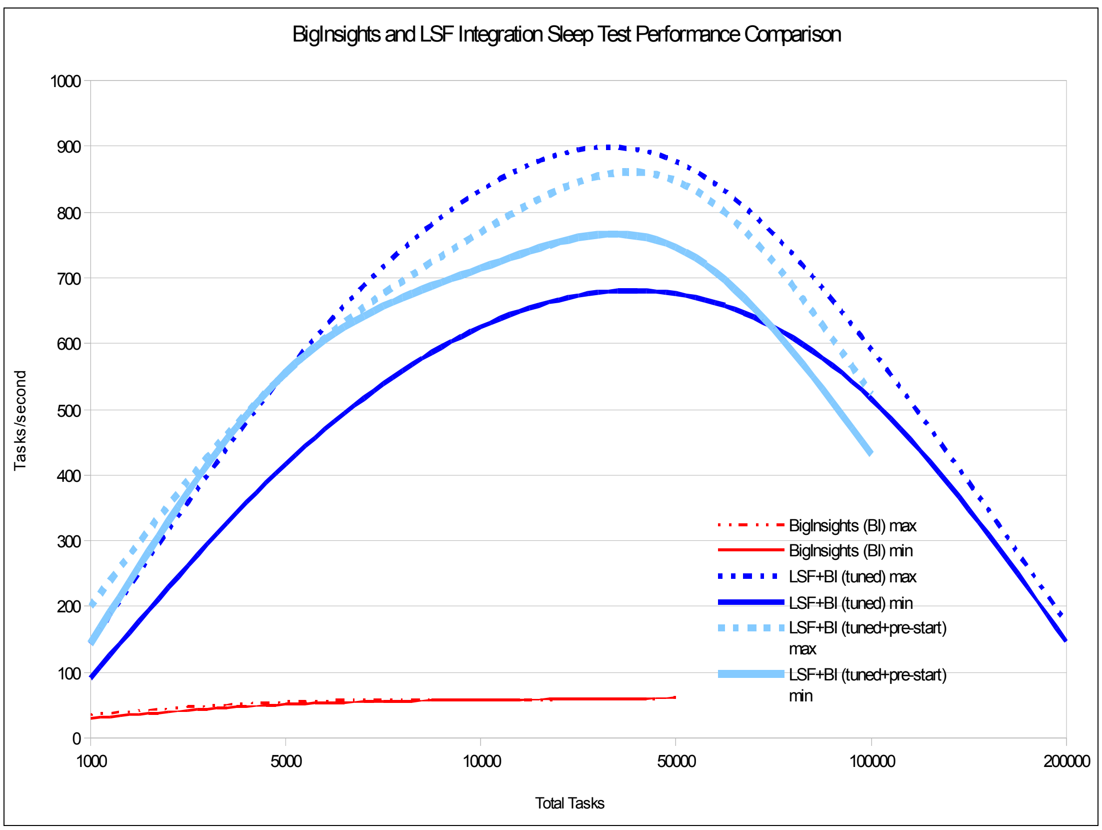

Figure 4 Sleep performance in tasks/second (default tuned, tuned, and preload)

The results show the following configurations:

- InfoSphere BigInsights only
- Integration of IBM Platform Computing LSF® and InfoSphere BigInsights for tuned and preloaded configurations, with the last one accounting for the Prestart option that is enabled

Pi estimate benchmark results

For the Pi estimate benchmark, Figure 5 compares the results that we obtained through the two different test configurations.

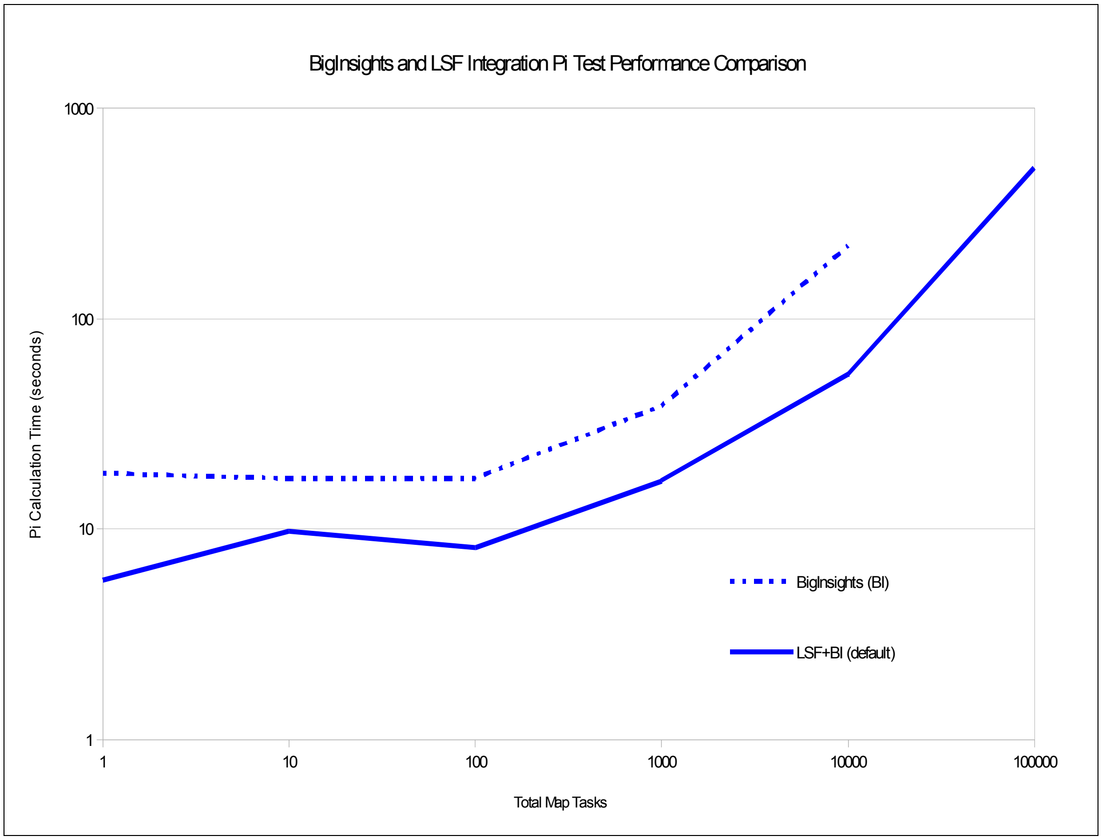

Figure 5 Pi estimate performance

The benchmark shows only the time that is needed to estimate the Pi value, discarding the time that is needed to write the maps into hard disk (that are equivalent between the two scenarios presented). The results are presented by using the following configurations:

- InfoSphere BigInsights only
- Integration of IBM Platform Computing LSF and InfoSphere BigInsights for a default configuration

Adding users

For the implementation steps described previously, the cluster has just one operational user who is the installation user egoadmin. This section describes how to add more users who can use the IBM Platform Symphony and the InfoSphere BigInsights integrated cluster.

Assumptions

Create a new supplementary user group to facilitate access to HDFS by the Hadoop users. This supplementary group is called *hdfsgrp*. This scenario uses the following assumptions:

- The user name to demonstrate the procedure is *testuser*.
- The hdfs data location is `/data/hdfs`.
- The **pdsh** command runs **ssh** to each data node and runs the command
- The **pdcp** command copies a file from the master to each of the data nodes.
- The home directories for new users are *not* shared across all the nodes. Adjust the directories as necessary if your home directory is shared.

Adding a user for the integrated cluster

To add a user for the integrated cluster:

1. Create a user on the master node and all other compute or data nodes (Example 22).

   Example 22 Creating a user on the master node and all compute or data nodes

   ```
   groupadd hdfsgrp
   pdsh groupadd hdfsgrp
   usermod –a –G hdfsgrp egoadmin
   pdsh usermod –a –G hdfsgrp egoadmin
   ```

2. Add the user to the master node and each node (Example 23). Run the commands as root from the master node.

   Example 23 Adding the user to the master node and all other nodes

   ```
   useradd testuser  -d /home/testuser
   id_on_master=`id –u testuser`
   pdsh adduser testuser –u $id_on_master  -d /home/testuser
   passwd testuser (set password for user on master)
   ```

3. Add the user to the hdfsgrp group (Example 24). Run the commands as root from the master node.

   Example 24 Adding the user to the hdfsgrp group

   ```
   usermod –a –G hdfsgrp testuser
   pdsh usermod –a –G hdfsgrp testuser
   ```

4. Create the hdfs user directory (Example 25). Run the commands as the egoadmin user from the master node.

   Example 25 Creating the hdfs user directory

   ```
   su – egoadmin
   hadoop dfs –mkdir /user/testuser
   hadoop dfs –chown testuser:hdfsgrp /user/testuser
   ```

5. Create a key for a new user for passwordless access to data nodes (Example 26).

   Example 26 Creating a key for passwordless access

   ```
   su – testuser
   ssh-keygen –t rsa –N “”
   ```

```
cp ~/.ssh/id_rsa.pub ~/.ssh/authorized_keys
chmod 600 ~/.ssh/authorized_keys
exit
pdsh mkdir /home/testuser/.ssh
pdcp /home/testuser/.ssh/id_rsa /home/testuser/.ssh/id_rsa
pdcp /home/testuser/.ssh/id_rsa.pub /home/testuser/.ssh/authorized_keys
pdsh chmod 600 /home/testuser/,ssh/authorized_keys
```

6. Add environment variables to the user testuser (Example 27).

Example 27   Adding environment variables to testuser

```
/bin/sh -c 'echo ". /opt/ibm/sym52/profile.platform" >> /home/testuser/.bashrc
/bin/sh -c 'echo "source /opt/ibm/bi/conf/biginsights-env.sh" >>
/home/testuser/.bashrc
```

7. Run verifications steps to create data in hdfs (Example 28), which is a test of writing zeros to hdfs.

Example 28   Verification steps

```
[testuser@i05i04 ~]$ dd if=/dev/zero of=junk bs=1k count=10
10+0 records in
10+0 records out
10240 bytes (10 kB) copied, 9.05e-05 s, 113 MB/s
[testuser@i05i04 ~]$ hadoop dfs -put junk dfsjunk
[testuser@i05i04 ~]$ hadoop dfs -ls
Found 1 items
-rw-r--r-- 3 testuser hdfsgrp 10240 2012-11-06 13:55 /user/testuser/dfsjunk
```

8. Run a sleep job as testuser (Example 29).

Example 29   Running a sleep job as testuser

```
[testuser@i05i04 ~]$ time hadoop jar $HADOOP_HOME/hadoop-examples-1.0.0.jar
sleep 1 100
12/11/03 21:30:47 INFO internal.MRJobSubmitter: Connected to JobTracker(SSM)
12/11/03 21:30:47 INFO internal.MRJobSubmitter:  Job <Sleep job> submitted, job
id <37>
12/11/03 21:30:47 INFO internal.MRJobSubmitter: Job will verify intermediate
data integrity using checksum.
12/11/03 21:30:47 INFO mapred.JobClient: Running job: job__0037
12/11/03 21:30:48 INFO mapred.JobClient: map 0% reduce 0%
12/11/03 21:30:53 INFO mapred.JobClient: map 100% reduce 100%
12/11/03 21:30:53 INFO mapred.JobClient: Job complete: job__0037
12/11/03 21:30:54 INFO mapred.JobClient: Counters: 18
12/11/03 21:30:54 INFO mapred.JobClient:   Shuffle Errors
12/11/03 21:30:54 INFO mapred.JobClient:     WRONG_PATH=0
12/11/03 21:30:54 INFO mapred.JobClient:     CONNECTION=0
12/11/03 21:30:54 INFO mapred.JobClient:     IO_ERROR=0
12/11/03 21:30:54 INFO mapred.JobClient:   FileSystemCounters
12/11/03 21:30:54 INFO mapred.JobClient:     FILE_BYTES_WRITTEN=35
12/11/03 21:30:54 INFO mapred.JobClient:   Map-Reduce Framework
12/11/03 21:30:54 INFO mapred.JobClient:     Reduce input groups=2
12/11/03 21:30:54 INFO mapred.JobClient:     Combine output records=0
12/11/03 21:30:54 INFO mapred.JobClient:     Map output records=1
12/11/03 21:30:54 INFO mapred.JobClient:     Shuffled Maps =1
```

```
12/11/03 21:30:54 INFO mapred.JobClient:     Reduce shuffle bytes=8
12/11/03 21:30:54 INFO mapred.JobClient:     Combine input records=0
12/11/03 21:30:54 INFO mapred.JobClient:     Spilled Records=2
12/11/03 21:30:54 INFO mapred.JobClient:     SPLIT_RAW_BYTES=0
12/11/03 21:30:54 INFO mapred.JobClient:     Map output bytes=4
12/11/03 21:30:54 INFO mapred.JobClient:     Reduce input records=1
12/11/03 21:30:54 INFO mapred.JobClient:     GC time elapsed (ms)=0
12/11/03 21:30:54 INFO mapred.JobClient:     Failed Shuffles=0
12/11/03 21:30:54 INFO mapred.JobClient:     Merged Map outputs=1
12/11/03 21:30:54 INFO mapred.JobClient:     Reduce output records=0

real    0m8.140s
user    0m3.681s
sys     0m0.441s
```

Adding nodes

After the integration, you add a node to InfoSphere BigInsights and IBM Platform Symphony.

Rejoining a machine to a cluster

Use these steps when a machine crashes and restarts, and you need to rejoin it to the cluster. You do not need to shut down or restart the entire cluster in this case.

1. Add the `/etc/hosts path` or DNS of the new node, depending on the name resolution method.
2. Add the same host name to the `conf/slaves` file on the master node.
3. Log in to the new subordinate node and run the following commands:

   ```
   $ cd path/to/hadoop
   $ bin/hadoop-daemon.sh start datanode
   $ bin/hadoop-daemon.sh start tasktracker
   ```

4. If you are using the function of the `dfs.include/mapred.include` file, add the node to the `dfs.include/mapred.include` file so that the NameNode and JobTracker can detect that the additional node was added. Enter the following commands:

   ```
   hadoop dfsadmin -refreshNodes
   hadoop mradmin -refreshNodes
   ```

5. Synchronize the Hadoop configuration files in the `$BIGINSIGHTS_HOME/hdm/hadoop-conf-staging/` directory on the management node with their counterparts on all of the nodes in the cluster. To synchronize these files, run the following command:

   ```
   $BIGINSIGHTS_HOME/bin/syncconf.sh hadoop force
   ```

Adding a data or compute node to the integrated software

To add a data or compute node to the integrated software for the InfoSphere BigInsights and the IBM Platform Symphony:

1. Add the new node host name to the DNS server or to the `/etc/hosts` path of all nodes.
2. Verify and complete as necessary all requirements that are described in “Preconfiguration on all nodes” on page 5, according to the remaining nodes where the node will join.
3. Add the node manually to the current Hadoop configuration (the `$HADOOP_CONF_DIR/slaves` file) before you synchronize the configurations.

4. Verify that the InfoSphere BigInsights is started on all nodes (except the joining node) and that the HDFS as no errors. If possible, synchronize the cluster by running the following command as the egoadmin user:

   `$BIGINSIGHTS_HOME/bin/syncconf.sh hadoop force`

5. On the master node, run the following command as the egoadmin user:

   `addnode.sh hadoop <new_node_hostname>`

   Adding the force option: The InfoSphere BigInsights software was copied to the new node, but none of the integration is present on the new node. For troubleshooting purposes, you might need to add the **`force`** option (at the end) to this command.

6. To install and complete the integration process on the new node, follow the steps only for compute nodes in the following sections:
   – “Installing IBM Platform Symphony Advanced Edition” on page 12
   – “Integrating IBM Platform Symphony and InfoSphere BigInsights” on page 14.
7. Restart the IBM Platform Symphony on all nodes for the new node to join correctly.
8. Before you apply that patch, stop the IBM BigInsigts and Hadoop.

Troubleshooting

This section describes troubleshooting hints and tips for working with IBM Platform Computing and InfoSphere BigInsights.

InfoSphere BigInsights

After you tune a sleep test with 10,000 maps and 10,000 prestarted services, remove all contents of the `/data/mapred/local` path on all nodes and then restart all nodes.

Increasing Java virtual memory

On all nodes, change the value of the variable *JVM_OPTIONS*, inside the `$PMR_HOME/conf/pmr-env.sh` file to `-Xms1024m -Xmx4096m` as shown in Example 30.

Example 30 Changes inside the $PMR_HOME/conf/pmr-env.sh file

```
[...]
# The JVM options
export JVM_OPTIONS="-Xms1024m -Xmx4096m"
[...]
```

Increasing system limits

Increase the limits for the root and egoadmin users by editing the `/etc/security/limits.conf` file and adding the lines as shown in Example 31.

Example 31 Maximizing the limit values for root and egoadmin user

```
egoadmin hard nofile 1000000
egoadmin soft nofile 1000000
```

```
egoadmin hard nproc unlimited
egoadmin soft nproc unlimited
root hard nofile 1000000
root soft nofile 1000000
root hard nproc unlimited
root soft nproc unlimited
```

Configured system limits: Even though some configured limits are already configured in InfoSphere BigInsights for the users, change these limits to bigger values. After the integration, both products use a much higher number of opened files and processes.

The team who wrote this paper

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

Dino Quintero is an IBM Senior Certified IT Specialist and a Complex Solutions Project Leader with the ITSO in Poughkeepsie, New York (NY). 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.

Ricardo Dobelin Barros is an IBM Certified IT Specialist in Brazil. Ricardo joined IBM in 2002 and has a total of 15 years of experience in the IT industry. He works in the capacity and performance group for IBM Strategic Outsourcing Delivery. Ricardo has expertise in availability management, service-levels management, sizing, and performance tuning in operating systems. He also has expertise in capacity planning, modeling, measurement, and infrastructure management for many IBM clients in Brazil. Ricardo holds a bachelor degree in systems analysis and post-graduate teaching degree in higher education.

Ashraf Gomaa is a Technical Project Manager and Team Leader at the Cairo Technology Development Center for IBM Egypt. He has worked for IBM for over 7 years with experience in software development and testing. His areas of expertise include web development, database development, globalization, IBM AIX® administration, and Linux administration. Ashraf graduated from Modern Academy in Maadi with a bachelor degree in computer science. He also has a diploma in software engineering from the Information Technology Institute in Egypt.

José Higino is an Infrastructure IT Specialist who provides AIX and Linux support and services for IBM Portugal. His areas of knowledge include System x, IBM BladeCenter®, and IBM Power Systems™ planning and implementation, management, virtualization, consolidation and clustering (HPC and high availability) solutions. José completed the Red Hat Certified Technician level in 2007, became a CiRBA Certified Virtualization Analyst in 2009, and completed certification in KT Resolve methodology as a subject matter expert in 2011. José holds a Master of Computers and Electronics Engineering degree from Universidade Nova de Lisboa, Faculdade de Ciências e Technologia, in Portugal.

Archana Kumar is an Advisory Software Engineer for IBM and has over 10 years of experience in software development and testing. Before joining IBM, she worked in application domain at ORACLE India Pvt Ltd. Since joining IBM in March 2004, she has worked in grid computing, security, and systems management. Her areas of interest include distributed computing, IT management, virtualization, security, and open source projects. In

2002, Archana graduated with a degree in computer science engineering from National Institute of Engineering, which is affiliated to Visvesvaraya Technological University.

Majid Ouassir is a Senior Technical HPC Consultant for IBM for over 12 years. He holds Platform product certifications in IBM Platform LSF HPC, IBM Platform Symphony administrator, Platform Cluster Manager, and Platform HPC. Majid has written three publications about computer science. He holds a doctorate degree with honors in computer science from UTC: French National Higher University for Sciences and Technologies.

Adam Parker is a Senior IT Specialist in the United Kingdom. He has worked at IBM for 14 years, including 6 years in HPC. His areas of expertise include IBM Extreme Cloud Administration Toolkit, IBM General Parallel File System (GPFS), and Linux cluster management. He co-wrote the IBM Redbooks® publication *Managing AIX Server Farms*, SG24-6606. He holds an honors degree in computer science from Wolverhampton University.

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

A special thank you goes to QLogic Corporation for the InfiniBand fabric equipment that we used during the residency. In addition, we thank the following people for their contributions to this project:

Ella Buslovic
ITSO, Poughkeepsie Center

Greg Geiselhart
Magnus Larsson
Oleg Litvinov
Gregory Rodgers
IBM USA

Michael Feiman
Eric Fiala
Peng Hu
Zane Hu
William Lu
Gord Sissons
IBM Canada

Bill McMillan
IBM UK

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

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

Notices

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

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

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

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

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

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

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

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

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

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

COPYRIGHT LICENSE:

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

This document REDP-5006-00 was created or updated on April 30, 2013.

Send us your comments 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 U.S.A.

Redpaper™

Trademarks