



SGI® Hadoop® Based on Intel® Xeon®  
Processor 5600 Series

007-5827-001

---

## COPYRIGHT

© 2012 Silicon Graphics International Corp. All rights reserved; provided portions may be copyright in third parties, as indicated elsewhere herein. No permission is granted to copy, distribute, or create derivative works from the contents of this electronic documentation in any manner, in whole or in part, without the prior written permission of SGI.

---

## LIMITED RIGHTS LEGEND

The software described in this document is “commercial computer software” provided with restricted rights (except as to included open/free source) as specified in the FAR 52.227-19 and/or the DFAR 227.7202, or successive sections. Use beyond license provisions is a violation of worldwide intellectual property laws, treaties and conventions. This document is provided with limited rights as defined in 52.227-14.

The electronic (software) version of this document was developed at private expense; if acquired under an agreement with the USA government or any contractor thereto, it is acquired as “commercial computer software” subject to the provisions of its applicable license agreement, as specified in (a) 48 CFR 12.212 of the FAR; or, if acquired for Department of Defense units, (b) 48 CFR 227-7202 of the DoD FAR Supplement; or sections succeeding thereto. Contractor/manufacturer is SGI, 46600 Landing Parkway, Fremont, CA 94538.

---

## TRADEMARKS AND ATTRIBUTIONS

Silicon Graphics, SGI, the SGI logo, Rackable, and Supportfolio are trademarks or registered trademarks of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries worldwide.

Adobe and Flash are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States and/or other countries. AMD Opteron is a trademark or registered trademark of Advanced Micro Devices Corporation. Cloudera is a trademark of Cloudera Inc. in the USA and other countries. Datameer is a trademark of Datameer, Inc. Firefox is a registered trademark of The Mozilla Foundation. Hadoop is a registered trademark of Apache Software Foundation. Intel and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries. Internet Explorer is a registered trademark of Microsoft Corporation. Java is a registered trademark of Oracle and/or one of its affiliates. Kitenga is a registered trademark of Kitenga Inc. Pentaho (TM) is a registered trademark of Pentaho Corporation. Quantum4D is a registered trademark of Quantum4D, Inc. Red Hat and all Red Hat-based trademarks are trademarks or registered trademarks of Red Hat, Inc. in the United States and other countries. Safari is a registered trademark of Apple Inc., registered in the United States and other countries.

All other trademarks mentioned herein are the property of their respective owners.

---

## Record of Revision

<b>Version</b>	<b>Description</b>
001	February 2012 Initial printing.



---

# Contents

	<b>About This Guide</b>	<b>vii</b>
	Audience	vii
	Related Publications	viii
	Product Support	ix
	Reader Comments	x
<b>1</b>	<b>Overview</b>	<b>1</b>
	Hardware	2
	Servers	2
	Network Hardware	3
	Configurations	4
	Half-Rack	5
	Full-Rack (42U)	6
	Multi-Rack (Second Rack And Beyond)	7
	Network Topology	8
	Node Level	9
	Rack Level for Single-Rack Configuration	10
	Rack Level for Rack 1 in Multi-Rack Configuration	11
	Rack Level for Rack 2 (And Beyond) in Multi-Rack Configuration	12
	Inter-Rack Level	13
	Software	14
<b>2</b>	<b>Cluster Startup</b>	<b>15</b>
	Accepting End-User License Agreements (EULAs)	15
	Java® Distribution Kit (JDK)	15
	Adobe® Flash®	15
	Configuring and Starting SGI Management Center	16
	Starting the Cluster for the First Time	16

- Re-Imaging the Server Nodes . . . . . 17
- 3 Business Intelligence Applications Startup . . . . . 19**
- Datameer . . . . . 20
  - Starting the Application . . . . . 20
  - Using the Demo . . . . . 21
- Kitenga . . . . . 22
  - Starting the Application . . . . . 22
  - Using the Demo . . . . . 22
- Quantum4D . . . . . 23

---

## About This Guide

This guide provides an overview of the SGI® Hadoop® Reference Implementation based on the Intel® Xeon® processor 5600 series along with getting-started instructions for this implementation. This guide consists of the following chapters:

- [Chapter 1, “Overview,”](#) provides an overview of the SGI Hadoop solution.
- [Chapter 2, “Cluster Startup,”](#) describes licensing and Hadoop specifics for configuring cluster management and monitoring.
- [Chapter 3, “Business Intelligence Applications Startup,”](#) describes how to start up the trial packages of the business intelligence (BI) applications bundled in the solution.

## Audience

This guide is written for the system administrators of the Hadoop cluster and developers. The guide assumes the reader is familiar with clusters, the Hadoop technology, and business intelligence applications.

## Related Publications

The following SGI documents are relevant to your Hadoop solution:

- *SGI Management Center Quick Start Guide* (007-5672-xxx)
- *SGI Management Center (SMC) Installation and Configuration* (007-5643-xxx)
- *SGI Management Center (SMC) System Administrator's Guide* (007-5642-xxx)
- *SGI Rackable C2005 Server Family User's Guide* (007-5717-xxx)

You can obtain SGI documentation, release notes, and man pages in the following ways:

- Refer to the SGI Technical Publications Library at <http://docs.sgi.com>. Various formats are available. This library contains the most recent and most comprehensive set of online books, release notes, man pages, and other information.
- Refer to the SGI Supportfolio™ webpage for documents whose access require a support contract. See “[Product Support](#)” on page ix.
- You can also view man pages by typing `man <title>` on a command line.

---

**Note:** For information about third-party system components, see the documentation provided by the manufacturer/supplier.

---



## Product Support

SGI provides a comprehensive product support and maintenance program for its products. SGI also offers services to implement and integrate Linux applications in your environment.

- Refer to <http://www.sgi.com/support/>
- If you are in North America, contact the Technical Assistance Center at +1 800 800 4SGI or contact your authorized service provider.
- If you are outside North America, contact the SGI subsidiary or authorized distributor in your country.

Be sure to have the following information before you call Technical Support:

- Product serial number
- Product model name and number
- Applicable error messages
- Add-on boards or hardware
- Third-party hardware or software
- Operating system type and revision level

## Reader Comments

If you have comments about the technical accuracy, content, or organization of this document, contact SGI. Be sure to include the title and document number of the manual with your comments. (Online, the document number is located in the front matter of the manual. In printed manuals, the document number is located at the bottom of each page.)

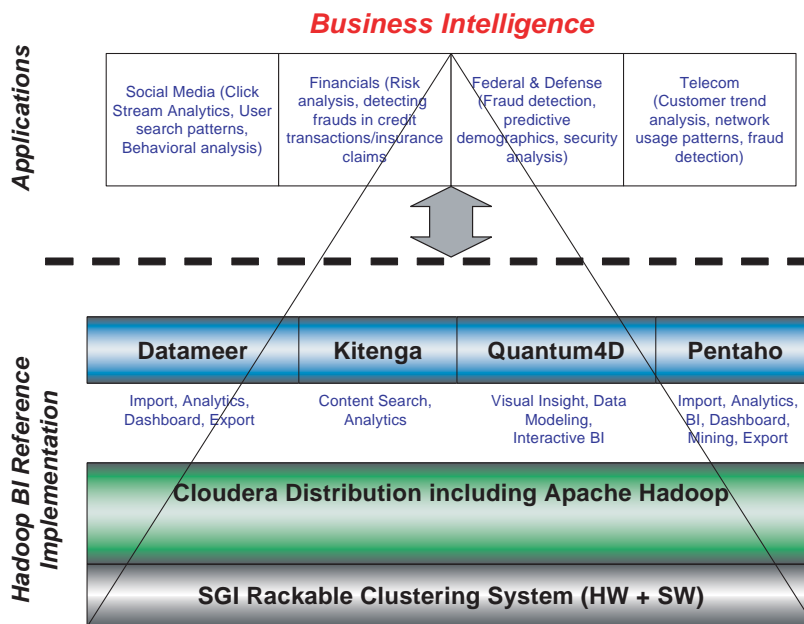
You can contact SGI in any of the following ways:

- Send e-mail to the following address: [techpubs@sgi.com](mailto:techpubs@sgi.com)
- Contact your customer service representative and ask that an incident be filed in the SGI incident tracking system.
- Send mail to the following address:

SGI  
Technical Publications  
46600 Landing Parkway  
Fremont, CA 94538

SGI values your comments and will respond to them promptly.

# Overview



**Figure 1-1** SGI Hadoop Business Intelligence Ecosystem

The SGI Hadoop Reference Implementation provides a pre-defined and pre-certified Hadoop solution with these features:

- Pre-defined and pre-certified configurations
- High performance
- Power optimization
- Trial packages of business intelligence (BI) applications (See Figure 1-1.)

This overview describes the following components:

- “Hardware” on page 2
- “Configurations” on page 4
- “Network Topology” on page 8
- “Software” on page 14

## Hardware

This section describes the hardware used in the SGI Hadoop Reference Implementation: first, the servers and then the network hardware.

## Servers



**Figure 1-2** An SGI Rackable C2005 Server

The SGI Hadoop Cluster employs the SGI Rackable™ C2005 family of half-depth servers, shown in Figure 1-2. Table 1-1 shows the servers that are used, their function in the cluster, and their specifications.

**Table 1-1** SGI Hadoop Servers

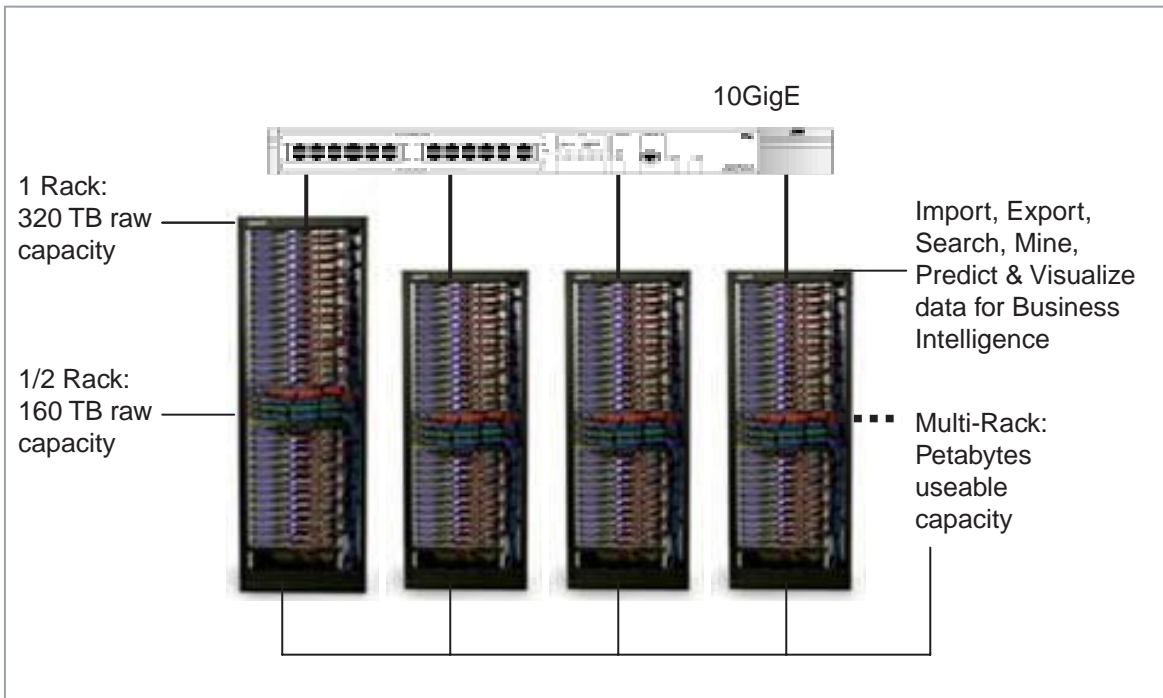
SGI Server	Conventional Node Type	Hadoop Node Type	Specifications
C2005-TY7	Master nodes	NameNode, Secondary NameNode, JobTracker	<ul style="list-style-type: none"> <li>– 2x Intel® Xeon® Processor E5645 (2.4 GHz)</li> <li>– 6x 8GB 1.35v 1333MHz DIMMs (48GB memory)</li> <li>– 4x 3.5” 1TB 7200 rpm SATA 6Gb/s drives in RAID configuration</li> <li>– 1x Dual-port 10GbE NIC (copper optical cables)</li> <li>– Redundant power supply, 650W</li> </ul>
C2005-TY6	Compute/Slave nodes	DataNodes, TaskTrackers	<ul style="list-style-type: none"> <li>– 2x Intel Xeon Processor E5645 (2.4 GHz)</li> <li>– 6x 8GB 1.35v 1333MHz DIMMs (48GB memory)</li> <li>– 10x 2.5” 1TB 7200 rpm SATA 6Gb/s drives</li> <li>– 1x Dual-port 1GbE NIC (cat 6)</li> <li>– PMBUS for power monitoring</li> </ul>
C2005-TY7		Application Node	<ul style="list-style-type: none"> <li>– 2x Intel Xeon Processor X5675 (3.06 GHz)</li> <li>– 12x 8GB 1.35v 1333MHz DIMMs (96GB memory)</li> <li>– 4x 3.5” 1TB 7200 rpm SAS 6Gb/s drives in RAID configuration</li> <li>– 1x Dual-port 10GbE NIC (copper optical cables)</li> <li>– Redundant power supply, 650W</li> </ul>

## Network Hardware

The network hardware consists of the following two components:

- 2 LG-Ericsson ES-4550G 48-port GigE switches per rack
- 1 LG-Ericsson ES-5048XG 10-GigE spine switch

## Configurations



**Figure 1-3** Data Capacity for Various Rack Configurations

The SGI Hadoop Cluster is available in single-rack and multi-rack configurations. [Figure 1-3](#) shows the range of data capacity for the configurations. This section describes the half-rack, full-rack, and multi-rack configurations.

## Half-Rack

Notes	Image	RackU	Image	Notes
		42		
		41		
		40		
		39		
		38		
		37		
		36		
		35		
		34		
		33		
		32		
		31		
		30		
		29		
		28		
		27		
		26		
		25		
		24		
		23		
		22		
LG-Ericsson ES-4550G - 48port GigE		21		LG-Ericsson ES-4550G - 48port GigE
LG-Ericsson ES-4550G - 48port GigE		20		C2005-TY7 2x5645 6x8GB 4X1TB
C2005-TY7 2x5645 6x8GB 4X1TB		19		Application Node
SecondaryNameNode/SGI-MC Headnode		18		C2005-TY7 2x5645 6x8GB 4X1TB
C2005-TY7 2x5645 6x8GB 4X1TB		17		Jobtracker
NameNode		16		C2005-TY6 2x5645 6x8GB 10X1TB
C2005-TY6 2x5645 6x8GB 10X1TB		15		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		14		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		13		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		12		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		11		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		10		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		9		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		8		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		7		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		6		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		5		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		4		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		3		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		2		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		1		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node				Data/TaskTracker Node

**Figure 1-4** Half-Rack Configuration

Figure 1-4 describes the configuration of a half-rack configuration. The rack consists of the following:

- 1 SGI Management Center node/Secondary NameNode
- 1 NameNode
- 1 JobTracker
- 1 Application node
- 16 DataNodes/TaskTracker nodes
- 2 48-port GigE stacked Hadoop data network switches
- 1 SGI Management Center network switch

## Full-Rack (42U)

Notes	Image	RackU	Image	Notes
		42		
		41		
		40		
		39		
LG-Ericsson ES-4550G - 48port GigE		38		LG-Ericsson ES-4550G - 48port GigE
LG-Ericsson ES-4550G - 48port GigE		37		LG-Ericsson ES-4550G - 48port GigE
C2005-TY7 2x5645 6x8GB 4X1TB		36		C2005-TY7 2x5645 12x8GB 4X1TB
SecondaryNameNode/SGI-MC Headnode		35		Application Node
C2005-TY7 2x5645 6x8GB 4X1TB		34		C2005-TY7 2x5645 6x8GB 4X1TB
NameNode		33		JobTracker
C2005-TY6 2x5645 6x8GB 10X1TB		32		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		31		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		30		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		29		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		28		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		27		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		26		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		25		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		24		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		23		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		22		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		21		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		20		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		19		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		18		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		17		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		16		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		15		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		14		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		13		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		12		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		11		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		10		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		9		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		8		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		7		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		6		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		5		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		4		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		3		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10X1TB		2		C2005-TY6 2x5645 6x8GB 10X1TB
Data/TaskTracker Node		1		Data/TaskTracker Node

**Figure 1-5** Full-Rack Configuration

Figure 1-5 describes the configuration of a full-rack configuration. The rack consists of the following:

- 1 SGI Management Center node/Secondary NameNode
- 1 NameNode
- 1 JobTracker
- 1 Application node
- 32 DataNodes/TaskTracker nodes
- 2 48-port GigE stacked Hadoop data network switches
- 1 SGI Management Center network switch



## Multi-Rack (Second Rack And Beyond)

Notes	Image	RackU	Image	Side B Notes
		36		
		35		
LG-Ericsson ES-4660 G - 48port GigE		34		
LG-Ericsson ES-4660 G - 48port GigE		33		LG-Ericsson ES-4660 G - 48port GigE
C2005-TY6 2x5645 6x8GB 10x1TB		32		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		31		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10x1TB		30		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		29		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10x1TB		28		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		27		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10x1TB		26		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		25		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10x1TB		24		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		23		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10x1TB		22		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		21		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10x1TB		20		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		19		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10x1TB		18		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		17		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10x1TB		16		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		15		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10x1TB		14		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		13		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10x1TB		12		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		11		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10x1TB		10		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		9		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10x1TB		8		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		7		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10x1TB		6		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		5		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10x1TB		4		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		3		Data/TaskTracker Node
C2005-TY6 2x5645 6x8GB 10x1TB		2		C2005-TY6 2x5645 6x8GB 10x1TB
Data/TaskTracker Node		1		Data/TaskTracker Node

**Figure 1-6** Multi-Rack—Second Rack And Beyond

Figure 1-6 describes the configuration of the second rack (and subsequent racks) of a multi-rack configuration. Each rack consists of the following:

- 32 DataNodes/TaskTracker nodes
- 2 48-port GigE stacked Hadoop data network switches
- 1 SGI Management Center network switch

## Network Topology

As described in [Table 1-2](#), the network topology of the SGI Hadoop Cluster depends on its rack configuration.

**Table 1-2** Network Topology

Rack Configuration	Network Topology
Single-rack	The Master node servers are attached to the top-of-rack switches directly via 10-Gigabit Ethernet.
Multi-rack	A 10-Gigabit Ethernet aggregate spine switch is introduced into the networking topology. The Master node servers are attached to this spine switch directly.

This section illustrates the network topology from the most granular level (node level) to the top level (inter-rack level):

- “Node Level” on page 9
- “Rack Level for Single-Rack Configuration” on page 10
- “Rack Level for Rack 1 in Multi-Rack Configuration” on page 11
- “Rack Level for Rack 2 (And Beyond) in Multi-Rack Configuration” on page 12
- “Inter-Rack Level” on page 13

Node Level

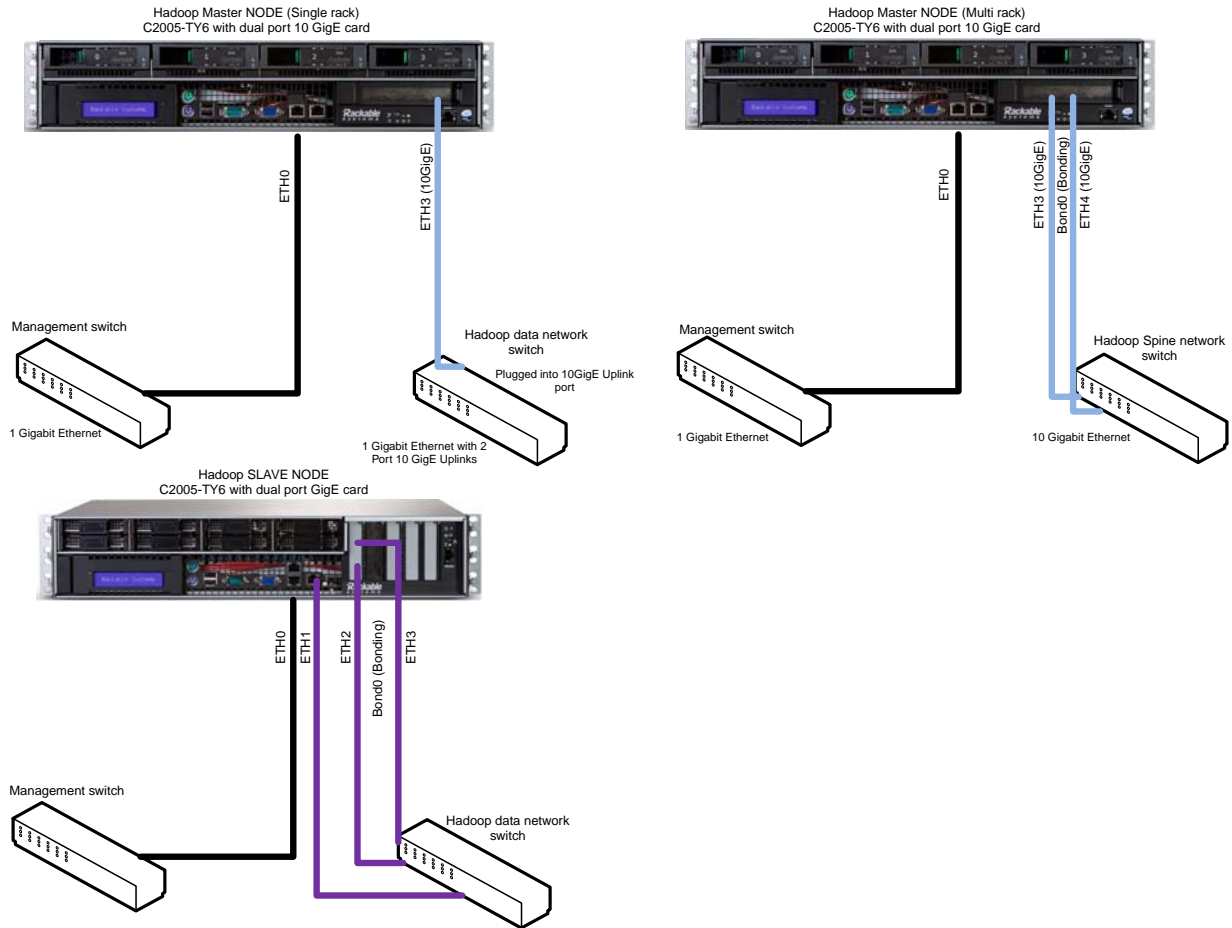


Figure 1-7 Network Topology—Node Level

## Rack Level for Single-Rack Configuration

Hadoop Rack Networking Layout (Single Rack)

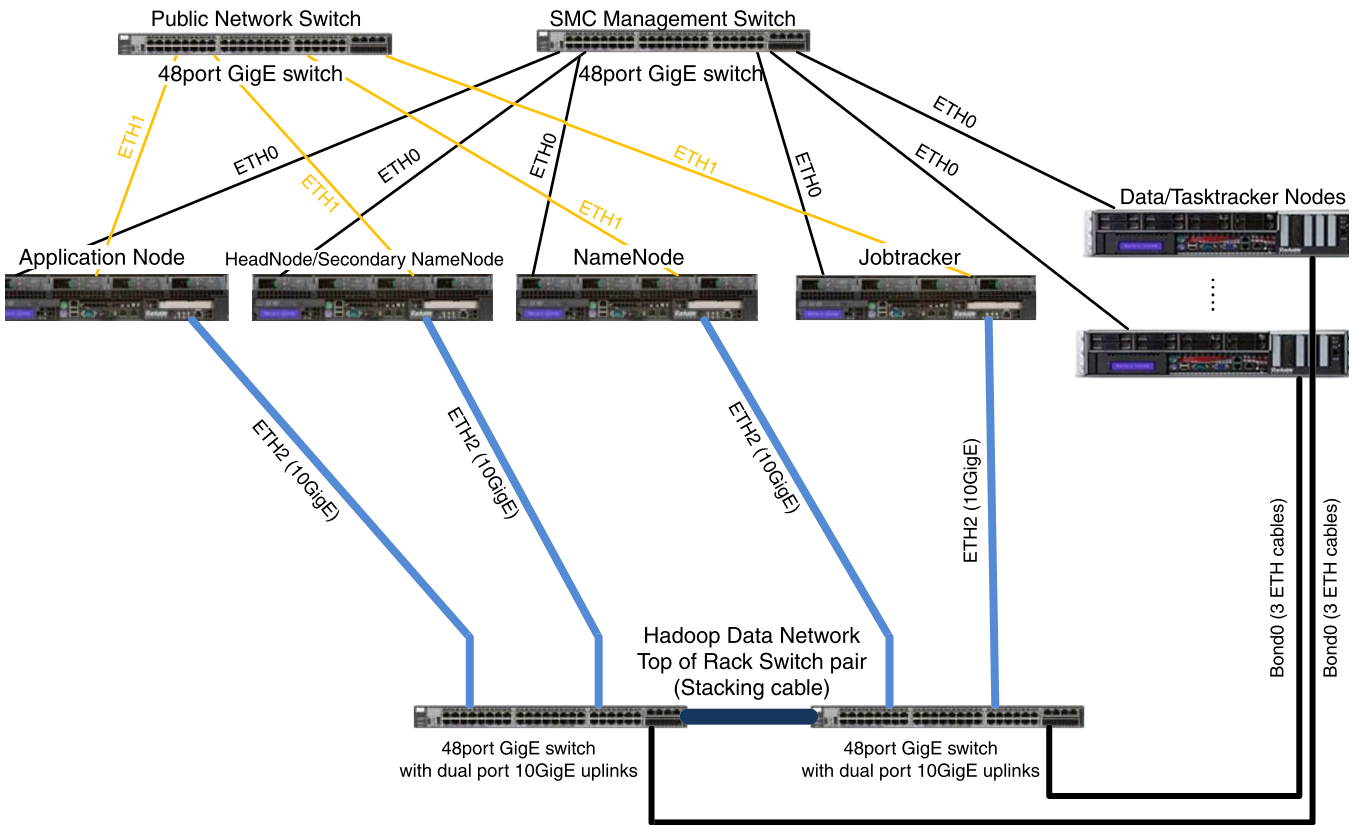


Figure 1-8 Network Topology—Rack Level for Single Rack

### Rack Level for Rack 1 in Multi-Rack Configuration

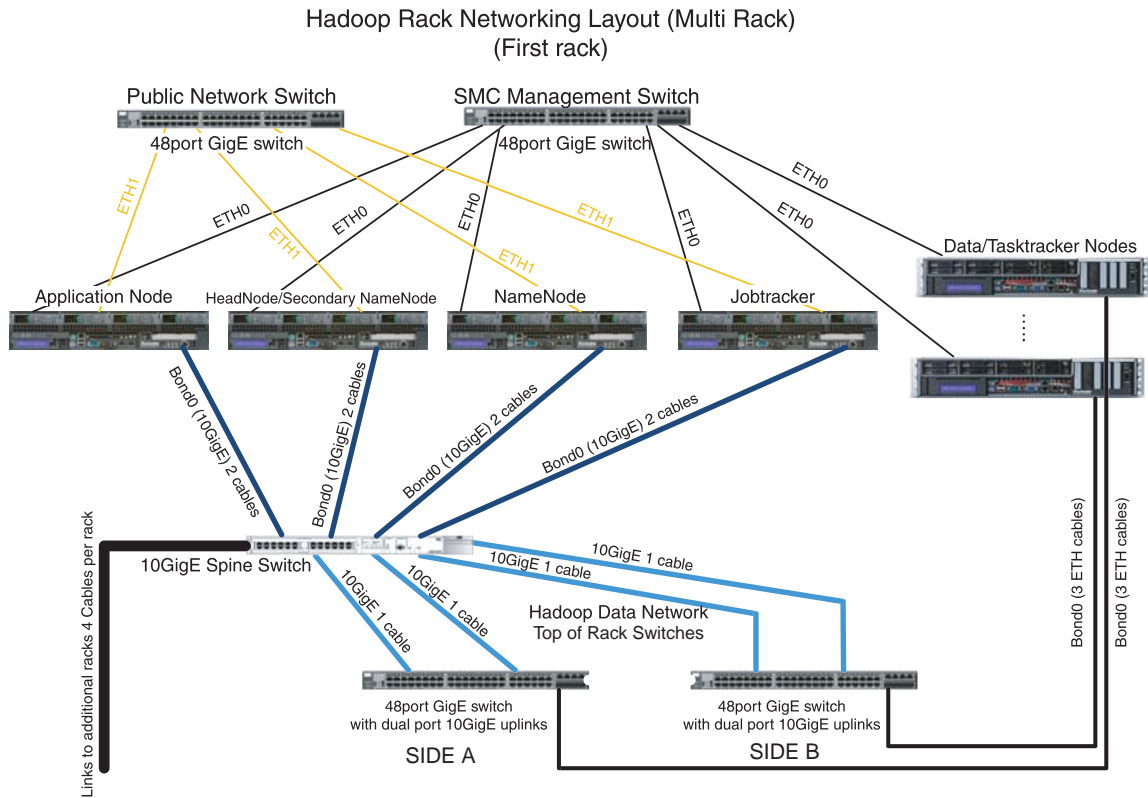
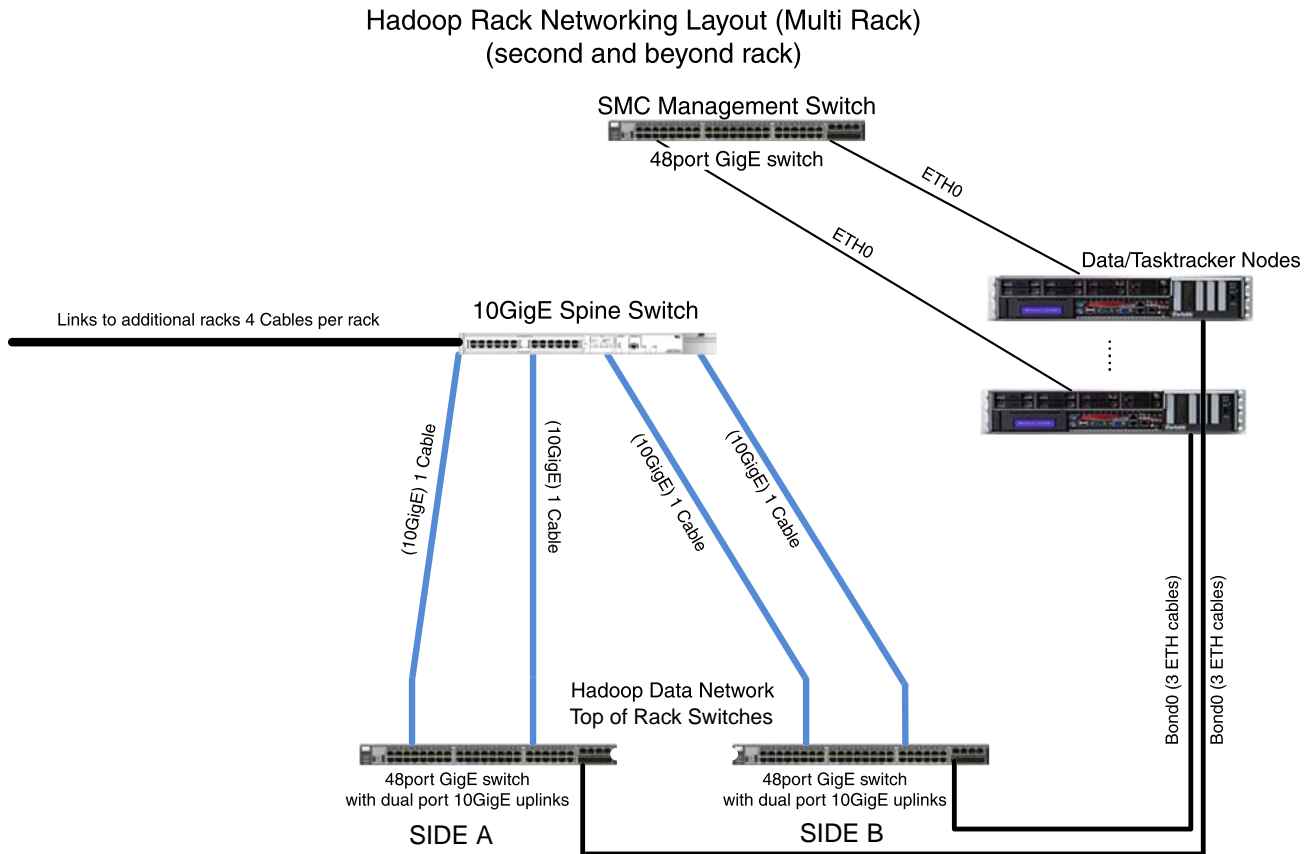


Figure 1-9 Network Topology—Rack Level for Rack 1 of Multi-Rack

## Rack Level for Rack 2 (And Beyond) in Multi-Rack Configuration



**Figure 1-10** Network Topology—Rack Level for Rack 2 (And Beyond) of Multi-Rack

## Inter-Rack Level

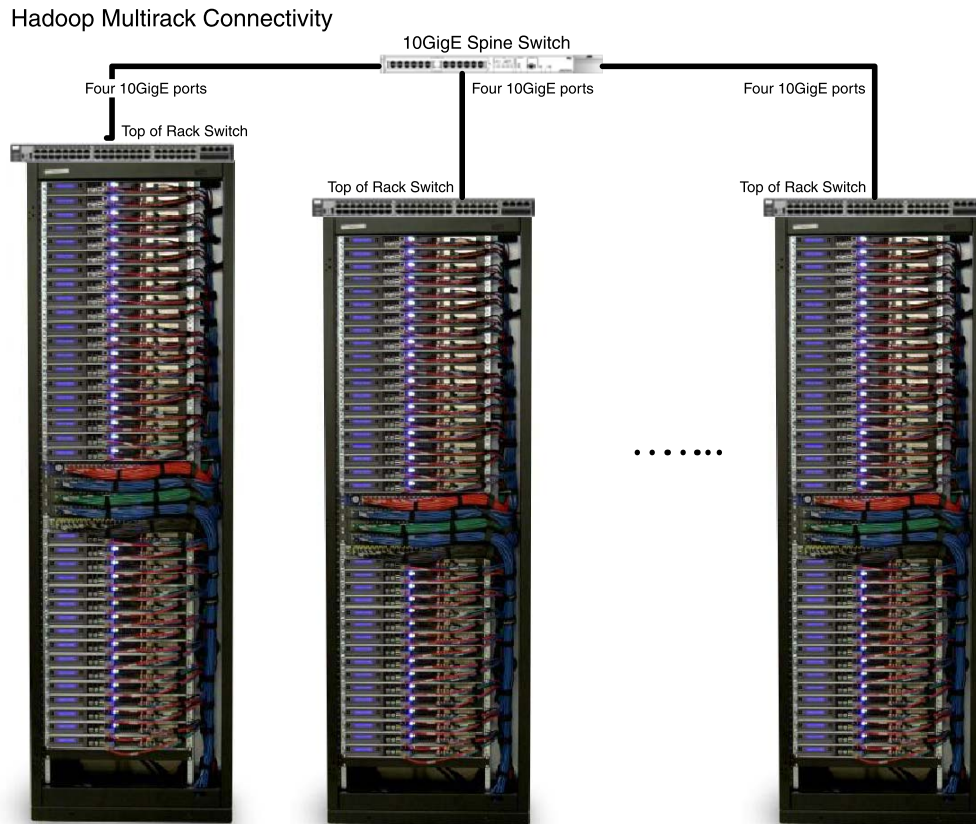


Figure 1-11 Network Topology—Inter-Rack Level

## Software

The software stack for the SGI Hadoop solution consists of the following components:

- Red Hat® Enterprise Linux (RHEL) 6 .1 (2.6.32-131.0.15.el6.x86\_64)
- Cloudera™ distribution Apache Hadoop 3 update 1 (Hadoop 0.20.2-cdh3u1)
- SGI Management Center 1.5.0
- An ecosystem of business intelligence applications software from ISVs like Kitenga®, Datameer™, Pentaho® and Quantum 4D® (See Figure 1-1 on page 1 and Figure 1-12.)

Figure 1-12 shows the software distribution on the various Hadoop servers.

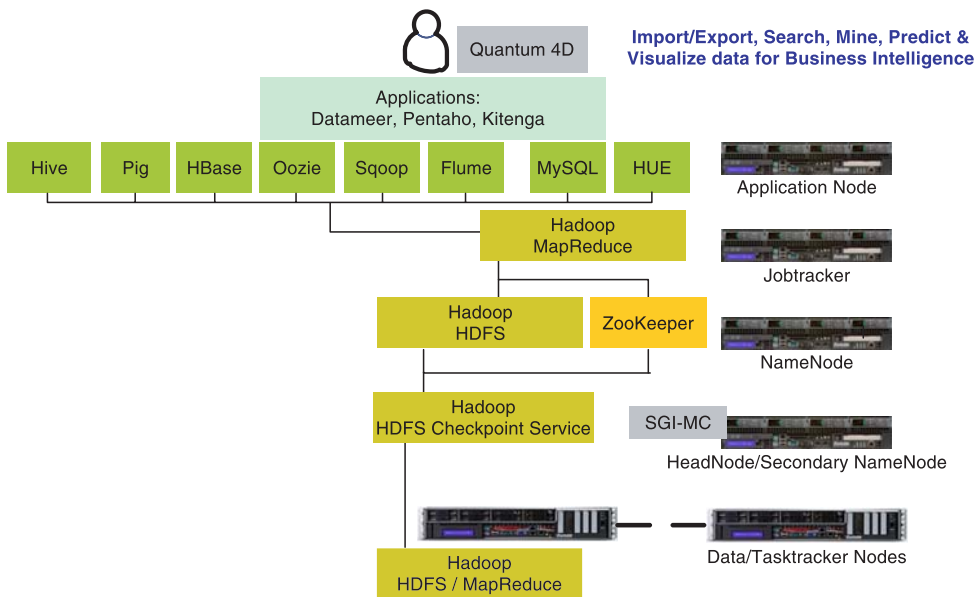


Figure 1-12 SGI Hadoop Software Stack



## Cluster Startup

This chapter describes the broad steps for starting the SGI Hadoop cluster:

- “Accepting End-User License Agreements (EULAs)” on page 15
- “Configuring and Starting SGI Management Center” on page 16
- “Starting the Cluster for the First Time” on page 16
- “Re-Imaging the Server Nodes” on page 17

### Accepting End-User License Agreements (EULAs)

The SGI Hadoop solution contains third-party software whose end-user license agreements you must read and accept. Two such products are described in this section. The trial versions of the business intelligence applications also require licensing agreements. [Chapter 3, “Business Intelligence Applications Startup,”](#) describes these requirements.

### Java® Distribution Kit (JDK)

The Java Distribution Kit copyright and third-party license agreement can be found in directory `/usr/share/doc/java-1.6.0-sun-devel-1.6.0.25` on any of the cluster nodes. Read and accept the conditions.

### Adobe® Flash®

Some of the application features will require Adobe Flash plugin installed for the browser to work correctly. Specifically, the Kitenga administration console and the Datameer Analytics Server administration console will require this. For your convenience, a licensed Adobe Flash plugin RPM has been installed on the head and application nodes. Before using the Adobe Flash plugin, read and accept the end-user license agreement which is located in `/usr/share/doc/flash-plugin-11.1.102.55/readme.txt`.

## Configuring and Starting SGI Management Center

To configure and start the SGI Management Center to monitor the SGI Hadoop cluster, you will need to follow the instructions in the *SGI Management Center Quick Start Guide* and appropriately configure the Hadoop servers described in [Table 2-1](#).

**Table 2-1** Hostnames for SGI Hadoop Servers

Daemon	Hostname	Hadoop Data Network Hostname
NameNode	sgi-nn	sgi-nn-10ge
Secondary NameNode	sgi-snn	sgi-snn-10ge
JobTracker	sgi-jt	sgi-jt-10ge
Application Node	sgi-app	sgi-app-10ge
DataNodes & TaskTrackers	r[rack#]n[node#]	r[rack#]n[node#]-ge

## Starting the Cluster for the First Time

Use the following steps to start the SGI Hadoop cluster the first time.

1. Power on the head node of the cluster.
2. Use SGI Management Center to start the nodes in the cluster.
  - a. Log in as `root`.
  - b. Start the SGI Management Center with the following command:
 

```
# mgrclient
```
  - c. Within the Management GUI, select the nodes to start, right-click, and select **Power > On**.
  - d. Start the nodes in the following order:
    - i. `sgi-app`
    - ii. `sgi-nn`
    - iii. `sgi-jt`
    - iv. Compute/Slave nodes in the Compute group

Hadoop is configured to start once the servers have booted.

3. Use the web browser on the head node to log into the Hadoop web interfaces:
  - NameNode: `http://sgi-nn-10ge:50070`
  - Secondary NameNode: `http://sgi-snn-10ge:50090`
  - JobTracker: `http://sgi-jt-10ge:50030`
4. Verify that the cluster powered on correctly and that all slave nodes joined the Hadoop cluster.

Run the following command to verify the number of DataNodes match the expected slave node count:

```
# sudo -u hdfs hapoop dfsadmin -report
```

## Re-Imaging the Server Nodes

In SGI Management Center, there are compute images for each node type. [Table 2-2](#) shows the mapping. Re-provision the nodes with the compute images as needed.

**Table 2-2** Compute Images for SGI Hadoop Servers

Node Name	Image Name
sgi-nn	Compute-Hadoop-Namenode
sgi-jt	Compute-Hadoop-Jobtracker
sgi-app	Compute-Hadoop-App
r[rack#]n[node#]	Compute-Hadoop-Slave

To provision a node, do the following:

1. Select the appropriate node.
2. Right-click.
3. Select **Provision** > *compute-image-for-node*.



## Business Intelligence Applications Startup

This chapter describes the startup of the following business intelligence (BI) applications:

- “Datameer” on page 20
- “Kitenga” on page 22
- “Quantum4D” on page 23

A trial version of each application has been packaged with the SGI Hadoop solution. You must procure a 30-day trial license from the respective independent software vendor.

---

**Note:** The BI applications reside on server `sgi-app`, the application node.

---

## Datameer

A trial version of Datameer Analytics Solution (DAS) version 1.3.5 is installed under `/home/hdfs/datameer`.

### Starting the Application

Use the following steps.

1. Log in as user `hdfs` on the application server node (`sgi-app`).
2. Change the directory location to `datameer/das-1.3.5-cdh3u1/`.
3. Start DAS as follows:

```
./bin/conductor.sh start
```

The Datameer console runs on port 8081. It can be accessed via Firefox®, Internet Explorer®, or Safari® at `http://sgi-app:8081`. (Note that an Adobe Flash plug-in is required.)

4. Upon the first login, you must obtain a trial license from Datameer.

A pop-up window will guide you through the activation process. If you experience any problems during activation, please contact `license@datameer.com` for a license file, which can be downloaded and installed via your browser.

5. Once a valid license key has been installed, log in to the application using the following credentials:

User: `admin`

Password: `admin`

## Using the Demo

Once logged in, you will have full use of DAS for 30 days, which includes the pre-built demo described below. A Flash video walkthrough of the demo is available under directory `/home/hdfs/datameer/flash-demo`. The demo includes AMEX stock data that is loaded into Hadoop via SFTP, analyzed via spreadsheets, and visualized in dashboards. The solution demonstrates an end-to-end data processing pipeline for BI applications, implemented natively on Hadoop. The demo consists of the following features:

**Table 3-1** DAS Demo Features

Feature	Description
<b>importAMEXFromlocalFS</b>	This is an import job found under the <b>Data</b> tab and will import data to Hadoop from a remote server via SFTP (in this case from the local file system). This job should be executed as the first job as soon as you log in the first time.
<b>importAMEXHDFS</b>	This is another import job also found under the data tab that will import the same data from Hadoop HDFS and make it accessible to the workbook. You can choose to run this job instead of <b>importAMEXFromlocalFS</b> .
<b>AMEXDataWorkbook</b>	This workbook demonstrates Hadoop-based analytics for business users and creates two distinct worksheets that analyze the imported data. This job can be found under the <b>Analytics</b> tab.
<b>PELDashboard</b>	This is a dashboard job that creates graphs out of subsets of summarized data in the spreadsheets of the <b>AMEXDataWorkbook</b> .
<b>PELDashboard2</b>	This is another dashboard job that shows a different set of graphs using the same subset of data as in the <b>AMEXDataWorkbook</b> spreadsheets.

For more information, see the following chart.

Item	Resource for Help
DAS features	URL <a href="http://datameer.com/documentation">http://datameer.com/documentation</a>
Problems	Email alias <a href="mailto:support@datameer.com">support@datameer.com</a>
Datameer products	URL <a href="http://www.datameer.com">www.datameer.com</a>

## Kitenga

A trial version of Zettavox 1.5, the business analytics software from Kitenga is installed under `/home/hdfs/kitenga`. A trial license will be required to run the software, which needs to be obtained from [www.kitenga.com](http://www.kitenga.com).

### Starting the Application

Use the following steps.

1. Log in as user `hdfs` on the application server (`sgi-app`).
2. Change directory location to `/home/hdfs/kitenga/zettavox-1.5.0-cdh3/`.
3. Start Zettavox as follows:

```
./bin/zettavox.sh start
```

4. After Zettavox is started for the first time, execute the configure script:

```
./bin/configure.sh
```

This script is required to load the Zettavox demo data to user `hdfs`. Execute this script only once.

5. To get to the console, use the following URL in a browser that supports Flash:

```
http://sgi-app:9100
```

The Zettavox console runs on port 9100.

6. Once you have a license, log in with the following credentials:

```
User: admin
```

```
Password: zettavox
```

The first time you log in you may see an expired or an invalid license. You will need to visit [www.kitenga.com](http://www.kitenga.com) to get a valid trial license.

### Using the Demo

After login, you should be able to run the demo software. A Flash version of the demo software is available under directory `/home/hdfs/kitenga/flash-demo`. The ZettaVox documentation is available in the `doc` subdirectory.



## Quantum4D

Quantum4D provides powerful data visualization software that works together with Kitenga's Zettavox. Quantum4D must be downloaded from [www.quantum4d.com](http://www.quantum4d.com). A few Adobe Flash demos of Quantum4D are provided in the directory `kitenga/Quantum4d/flash-demo`.

