Cisco UCS Director Express for Big Data Management Guide, Release 2.0

First Published: November 05, 2015

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
http://www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

CONTENTS

Preface

Audience

This guide is intended primarily for data center administrators who use Cisco UCS Director Express for Big Data and who have responsibilities and expertise in one or more of the following:

- Server administration
- Storage administration
- Network administration
- Network security
- Virtualization and virtual machines

Conventions

Text Type	Indication
GUI elements	GUI elements such as tab titles, area names, and field labels appear in **this font**. Main titles such as window, dialog box, and wizard titles appear in **this font**.
Document titles	Document titles appear in *this font*.
TUI elements	In a Text-based User Interface, text the system displays appears in `this font`.

Text Type	Indication
System output	Terminal sessions and information that the system displays appear in `this font`.
CLI commands	CLI command keywords appear in **this font**. Variables in a CLI command appear in *this font*.
[ ]	Elements in square brackets are optional.
{x | y | z}	Required alternative keywords are grouped in braces and separated by vertical bars.
[x | y | z]	Optional alternative keywords are grouped in brackets and separated by vertical bars.
string	A nonquoted set of characters. Do not use quotation marks around the string or the string will include the quotation marks.
< >	Nonprinting characters such as passwords are in angle brackets.
[ ]	Default responses to system prompts are in square brackets.
!, #	An exclamation point (!) or a pound sign (#) at the beginning of a line of code indicates a comment line.

Note Means *reader take note*. Notes contain helpful suggestions or references to material not covered in the document.

Tip Means *the following information will help you solve a problem*. The tips information might not be troubleshooting or even an action, but could be useful information, similar to a Timesaver.

Caution Means *reader be careful*. In this situation, you might perform an action that could result in equipment damage or loss of data.

Timesaver Means *the described action saves time*. You can save time by performing the action described in the paragraph.

Warning

IMPORTANT SAFETY INSTRUCTIONS

This warning symbol means danger. You are in a situation that could cause bodily injury. Before you work on any equipment, be aware of the hazards involved with electrical circuitry and be familiar with standard practices for preventing accidents. Use the statement number provided at the end of each warning to locate its translation in the translated safety warnings that accompanied this device.

SAVE THESE INSTRUCTIONS

Related Documentation

Cisco UCS Director Documentation Roadmap

For a complete list of Cisco UCS Director documentation, see the *Cisco UCS Director Documentation Roadmap* available at the following URL: http://www.cisco.com/en/US/docs/unified_computing/ucs/ucs-director/doc-roadmap/b_UCSDirectorDocRoadmap.html.

Cisco UCS Documentation Roadmaps

For a complete list of all B-Series documentation, see the *Cisco UCS B-Series Servers Documentation Roadmap* available at the following URL: http://www.cisco.com/go/unifiedcomputing/b-series-doc.

For a complete list of all C-Series documentation, see the *Cisco UCS C-Series Servers Documentation Roadmap* available at the following URL: http://www.cisco.com/go/unifiedcomputing/c-series-doc.

Note The *Cisco UCS B-Series Servers Documentation Roadmap* includes links to documentation for Cisco UCS Manager and Cisco UCS Central. The *Cisco UCS C-Series Servers Documentation Roadmap* includes links to documentation for Cisco Integrated Management Controller.

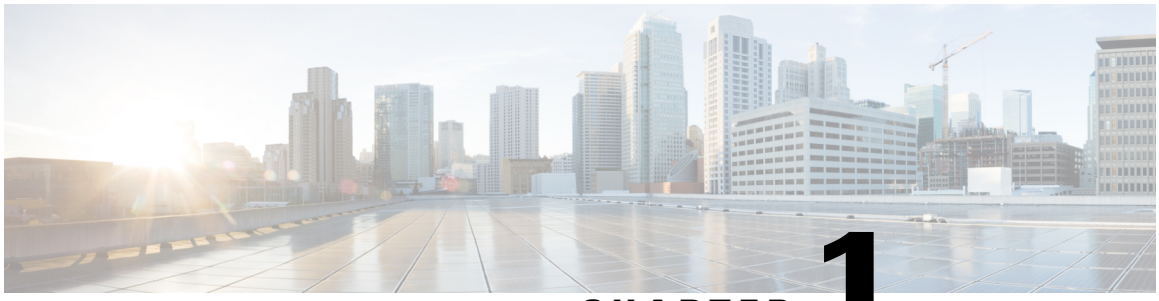
Documentation Feedback

To provide technical feedback on this document, or to report an error or omission, please send your comments to ucs-director-docfeedback@cisco.com. We appreciate your feedback.

Obtaining Documentation and Submitting a Service Request

For information on obtaining documentation, using the Cisco Bug Search Tool (BST), submitting a service request, and gathering additional information, see *What's New in Cisco Product Documentation*, at: http://www.cisco.com/c/en/us/td/docs/general/whatsnew/whatsnew.html.

Subscribe to *What's New in Cisco Product Documentation*, which lists all new and revised Cisco technical documentation as an RSS feed and delivers content directly to your desktop using a reader application. The RSS feeds are a free service.



CHAPTER 1

New and Changed Information for this Release

New and Changed Information for this Release

The following table provides an overview of the significant changes to this guide for this current release. The table does not provide an exhaustive list of all changes made to this guide or of all new features in this release.

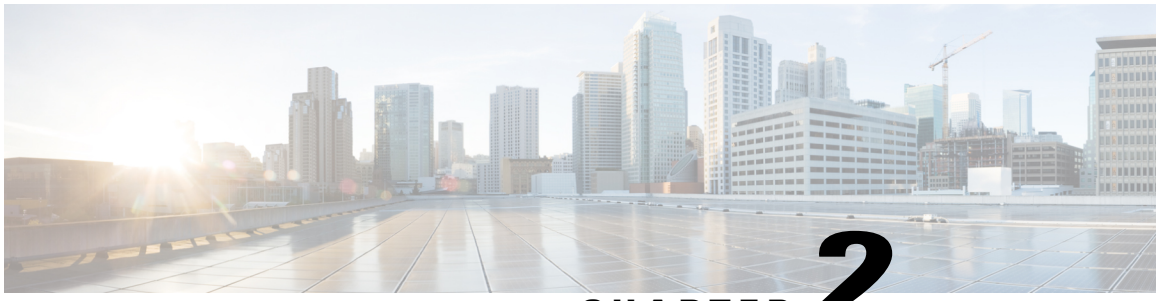
Table 1: New Features and Changed Behavior in Cisco UCS Director Express for Big Data, Release 2.0

Feature	Description	Where Documented
Updates to various menus, pop-ups, and tabs.	Numerous updates to reflect updates in the program organization.	Entire document.
Updates to **Overview** chapter	Added the following: • Manage Cisco UCS Director and Cisco UCS Director Express for Big Data Personalities • User Roles for Big Data	Overview, on page 5
Updates to **Licenses for Cisco UCS Director Express for Big Data**chapter	Added the following: • Standard and Premium Licenses	Licenses for Cisco UCS Director Express for Big Data, on page 15
Updates to the new actions implemented for the Hadoop cluster in the **Managing Hadoop Cluster** chapter	Updated Hadoop cluster actions in this release.	Managing a Hadoop Cluster, on page 75

Feature	Description	Where Documented
New **Hadoop Cluster Configuration Settings** chapter	Added new chapter on Settings required for Hadoop clusters. Contains the following: • Creating a Hadoop Cluster Configuration Template • Updating a Hadoop Cluster Configuration Template • QoS System Class • Pre Cluster Performance Testing Settings • Approving Hadoop Cluster Deployment Workflows • Uploading Required OS and Hadoop Software to Cisco UCS Director Baremetal Agent • Cloudera, MapR, and Hortonworks RPMs - Required software in Software Catalog • Cloudera and MapR RPMs - Required for upgrading Hadoop Cluster • Configuration Check Rules • Checking Hadoop Cluster Configuration • Fixing Configuration Violations	Hadoop Cluster Configuration Settings, on page 87
Updates to new workflows in the **Cisco UCS CPA Workflows** chapter	Updated new workflows added in this release.	Cisco UCS CPA Workflows, on page 105

Feature	Description	Where Documented
New **Proactive Status Monitoring and Diagnostics** chapter	Added new chapter on Proactive Status Monitoring and Diagnostics. Contains the following: • Aggregate CPU, Disk and Network Utilization • Monitoring Aggregate CPU, Disk and Network Utilization • Monitoring Top Jobs Based on CPU Utilization and Completion Time • Performance Metrics for CPU, Disk, and Network • Viewing CPU, Disk, and Network Statistics for a Hadoop Cluster • Analyzing Performance Bottlenecks through Historical Metrics • Setting Alerts for Hadoop Cluster Service Failures • Types of Disk and Network Failure Alerts • Setting Alerts for Disk, Network, and Node Failures • Setting Disk Utilization Threshold Alerts	Proactive Status Monitoring and Diagnostics, on page 121

New and Changed Information for this Release



CHAPTER 2

Overview

This chapter contains the following sections:

Cisco UCS Director Express for Big Data

Cisco UCS Director Express for Big Data provides a single touch solution that automates deployment of Big Data infrastructure and provides a single management pane across both physical infrastructure and Hadoop software. It supports key Hadoop distributions including Cloudera, MapR, and Hortonworks.

Cisco UCS Director Express for Big Data delivers end-to-end automation of Hadoop cluster deployment that allows you to spin up and expand clusters on-demand. Configuration of the physical infrastructure, which includes compute, internal storage, network, and installation of operating system, Java packages, Hadoop along with provisioning of Hadoop services are handled automatically with minimal user input. This is achieved through the innovative Cisco UCS service profiles wherein both the physical infrastructure and Hadoop configuration are incorporated into a Hadoop cluster deployment profile.

Cisco UCS Integrated Infrastructure for Big Data

Cisco UCS Integrated Infrastructure for Big Data is an industry leading architecture designed to meet a variety of Big Data workloads. It scales as processing and storage needs grow without increasing management challenges and delivers predictable performance along with reduced total cost of ownership (TCO).

Cisco UCS Integrated Infrastructure is comprised of the following components:

- Cisco UCS Fabric Interconnects

- Cisco UCS 2200 Series Fabric Extenders
- Cisco UCS C-Series Rack-Mount Servers
- Cisco UCS Virtual Interface Cards (VICs)
- Cisco UCS Manager

You can read more about the Cisco UCS Integrated Infrastructure for Big Data in the Data Center Designs Cloud Computing - Design Zone for Big Data .

Manage Cisco UCS Director and Cisco UCS Director Express for Big Data Personalities

Cisco UCS Director is the default personality made available after deployment but you can switch to use only Cisco UCS Director Express for Big Data, or both Cisco UCS Director and Cisco UCS Director Express for Big Data.

You can manage personalities here: **Administration** > **License** > **License Keys** > **Manage Personalities**.

Table 2: Personality Switch Behavior

Personality Selection	Cisco UCS Director Features	Cisco UCS Director Express for Big Data Features
Cisco UCS Director, Release 5.4 (Default)	Yes	No
Cisco UCS Director Express for Big Data, Release 2.0	No	Yes
Cisco UCS Director, Release 5.4 and Cisco UCS Director Express for Big Data, Release 2.0	Yes	Yes

Note Depending on the personality you start with and the personality selection, Cisco UCS Director and Cisco UCS Director Express for Big Data features are enabled or disabled with the restart of services on the appliance.

Create User Roles

You can create user roles that are specific to Cisco UCS Director Express for Big Data, and define menu settings and permissions for the users. You must ensure that you create a group before you add users to any role.

Note You can determine the default roles in the system only if the **Default Role** column in the **User Roles** page is marked with **Yes** here: **Administration** > **System** > **User Roles**.

For example, you can create the fallowing user roles, and then create users with this role:

- HadoopUser—A Hadoop user
- HadopAdmin—A Hadoop admin

For more information on Managing Users and Groups, see the latest *Cisco UCS Director Administration Guide*.

Hadoop Admin Permissions

Hadoop admin can:

- Read—permission to only read a file.
- Write—permission to read, write, and modify a file.
- Read/Write—permission to read and write to a file.

The following table shows a list of operations that an Hadoop admin can do:

Operations	Permissions	
	Read	Write
Virtual Computing	Yes	Yes (Only VM Management Actions)
VM Label	Yes	
Assign VM to vDC	Yes	
Virtual Storage	Yes	Yes
Virtual Network	Yes	Yes
Physical Computing	Yes	Yes
Physical Storage	Yes	Yes
Physical Network	Yes	Yes
Group Service Request	Yes	Yes
Approver Service Request	Yes	Yes
Budgeting	Yes	Yes

Operations	Permissions	
	Read	Write
Resource Accounting	Yes	
Chargeback	Yes	
System Admin	Yes	Yes
Users and Groups	Yes	Yes
Virtual Accounts	Yes	Yes
Catalogs	Yes	Yes
vDC	Yes	Yes
Computing Policy	Yes	Yes
Storage Policy	Yes	Yes
Network Policy	Yes	Yes
Service Delivery	Yes	Yes
Resource Limit Report	Yes	Yes
Group Users	Yes	Yes
Cloudsense Reports	Yes	Yes
Cloudsense Assessment Reports	Yes	Yes
Orchestration	Yes	Yes
Open Automation Modules	Yes	Yes
CS Shared Reports	Yes	Yes
Remote VM Access		Yes
Mobile Access Settings	Yes	Yes
End User Chargeback	Yes	
Resource Groups	Yes	Yes
Tag Library	Yes	Yes
Big Data Infra	Yes	

Operations	Permissions	
	Read	Write
Big Data Accounts		Yes
Big Data Cluster Management		Yes
Big Data Node Management		Yes
Big Data Performance Test		Yes
Big Data Service Management		Yes
Big Data Role Management		Yes
Big Data UCS SP Template		Yes
Big Data Hadoop Profile Template		Yes
Big Data Hadoop Deploy Template		Yes
Big Data Cluster Deployment		Yes
Big Data License Upload		Yes
Big Data Configuration Parameters Template		Yes
Big Data Faults		Yes
Big Data Settings - QoS		Yes
Big Data Settings - IP Pool		Yes
Big Data Settings - Pre_Cluster Sanity		Yes
Big Data Settings - Hadoop Software Upload		Yes
Big Data Settings - Configuration Check Rules		Yes

Hadoop User Permissions

Hadoop user can:

- Read—permission to only read a file.

- Write—permission to read, write, and modify a file.
- Read/Write—permission to read and write to a file.

The following table shows a list of operations that an Hadoop user can do:

Operations	Permissions	
	Read	Write
Virtual Computing	Yes	
VM Label	Yes	
Assign VM to vDC	Yes	
Virtual Storage	Yes	
Virtual Network	Yes	
Physical Computing	Yes	
Physical Storage	Yes	
Physical Network	Yes	
Group Service Request	Yes	Yes
Approver Service Request	Yes	Yes
Budgeting	Yes	
Resource Accounting	Yes	
Chargeback	Yes	
System Admin	Yes	
Users and Groups	Yes	
Virtual Accounts	Yes	
Catalogs	Yes	
vDC	Yes	
Computing Policy	Yes	
Storage Policy	Yes	
Network Policy	Yes	

Operations	Permissions	
	Read	Write
Service Delivery	Yes	
Resource Limit Report	Yes	
Group Users	Yes	
Cloudsense Reports	Yes	
Cloudsense Assessment Reports	Yes	
Orchestration		
Open Automation Modules		
CS Shared Reports		
Remote VM Access		
Mobile Access Settings		
End User Chargeback		
Resource Groups		
Tag Library		
Big Data Infra	Yes	
Big Data Accounts		
Big Data Cluster Management		
Big Data Node Management		
Big Data Performance Test		
Big Data Service Management		
Big Data Role Management		
Big Data UCS SP Template		
Big Data Hadoop Profile Template		
Big Data Hadoop Deploy Template		
Big Data Cluster Deployment		

Operations	Permissions	
	Read	Write
Big Data License Upload		
Big Data Configuration Parameters Template		
Big Data Faults		
Big Data Settings - QoS		
Big Data Settings - IP Pool		
Big Data Settings - Pre_Cluster Sanity		
Big Data Settings - Hadoop Software Upload		
Big Data Settings - Configuration Check Rules		

Supported Hadoop Distributions

Cisco UCS Director Express for Big Data supports the following Hadoop distributions:

- Cloudera 5.4
- MapR 3.1, 4.0, 4.1, and 5.0
- Hortonworks 2.1, 2.2, and 2.3

Table 3: Java and JDK Software for Hadoop Distributions

Hadoop Distribution	Supported Hadoop Distribution Versions	Installed Java and JDK
Cloudera	5.0.1, 5.0.6, 5.2.0, 5.2.1, 5.3.0, and 5.4.1	oracle-j2sdk1.7
MapR	3.1.1, 4.0.1, 4.0.2, 4.1.0, and 5.0.0	java-1.7.0-openjdk
Hortonworks	2.1, 2.2, and 2.3	java-1.7.0-openjdk

High Level Work Flow to Create an Instant Hadoop Cluster

Step 1 Create a Cisco UCS Manager account. For more information, see Adding a Cisco UCS Manager Account.

Step 2 Configure Big Data IP pools. For more information, see Adding a Big Data IP Pool.

Step 3 Create an Instant Hadoop Cluster. For more information, see Creating an Instant Hadoop Cluster.

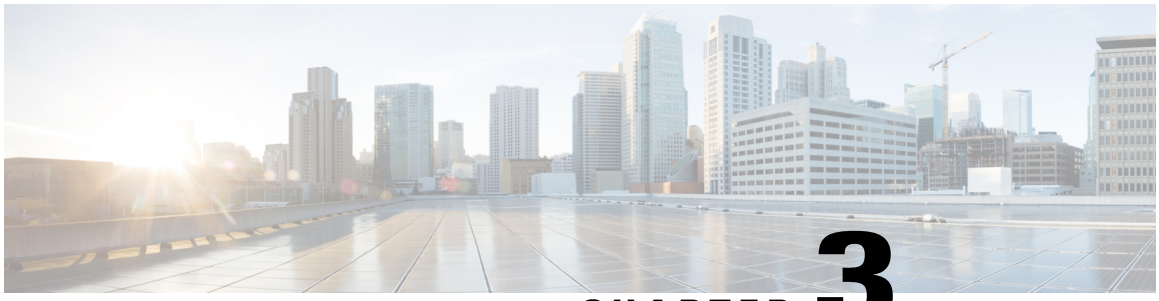
High-Level Work Flow to Create a Customized Hadoop Cluster

Step 1 Configure a Cisco UCS Service Profile template for Big Data. For more information, see Creating a Cisco UCS Service Profile Template for Big Data.

Step 2 Configure an Hadoop cluster template. For more information see, Creating a Hadoop Cluster Profile Template.

Step 3 Configure an Hadoop cluster deploy template. For more information, see Creating a Hadoop Cluster Deployment Template.

Step 4 Create a customized Hadoop cluster. For more information, see Creating a Customized Hadoop Cluster.



CHAPTER 3

Licenses for Cisco UCS Director Express for Big Data

This chapter contains the following sections:

About Licenses

You must obtain a license to use Cisco UCS Director Express for Big Data, as follows:

1. Before you install Cisco UCS Director Express for Big Data, generate the Cisco UCS Director Express for Big Data license key and claim a certificate (Product Access Key).
2. Register the Product Access Key (PAK) on the Cisco software license site, as described in Fulfilling the Product Access Key, on page 16.
3. After you install Cisco UCS Director Express for Big Data, update the license in Cisco UCS Director Express for Big Data as described in Updating the License, on page 17.
4. After the license has been validated, you can start to use Cisco UCS Director Express for Big Data.

Fulfilling the Product Access Key

Before You Begin

You need the PAK number.

Step 1 Navigate to the Cisco Software License website.

Step 2 If you are directed to the Product License Registration page, you can take the training or click **Continue to Product License Registration**.

Step 3 On the Product License Registration page, click **Get New Licenses from a PAK or Token**.

Step 4 In the **Enter a Single PAK or TOKEN to Fulfill** field, enter the PAK number.

Step 5 Click **Fulfill Single PAK/TOKEN**.

Step 6 Complete the additional fields in **License Information** to register your PAK:

Name	Description
Organization Name	The organization name.
Site Contact Name	The site contact name.
Street Address	The street address of the organization.
City/Town	The city or town.
State/Province	The state or province.
Zip/Postal Code	The zip code or postal code.
Country	The country name.

Step 7 Click **Issue Key**.
The features for your license appear, and an email with the Digital License Agreement and a zipped license file is sent to the email address you provided.

Updating the License

Before You Begin

If you received a zipped license file by email, extract and save the .lic file to your local machine.

Step 1 Choose **Administration** > **License**.

Step 2 Click on the **License Keys** tab.

Step 3 Click **Update License**.

Step 4 In the **Update License** dialog box, do the following:

- To upload a `.lic` file, click **Browse** to the base license's `.lic` file and select the `.lic` file, then click **Upload**.

Step 5 Click **Submit**.
The license file is processed, and a message appears confirming the successful update.

Standard and Premium License Features

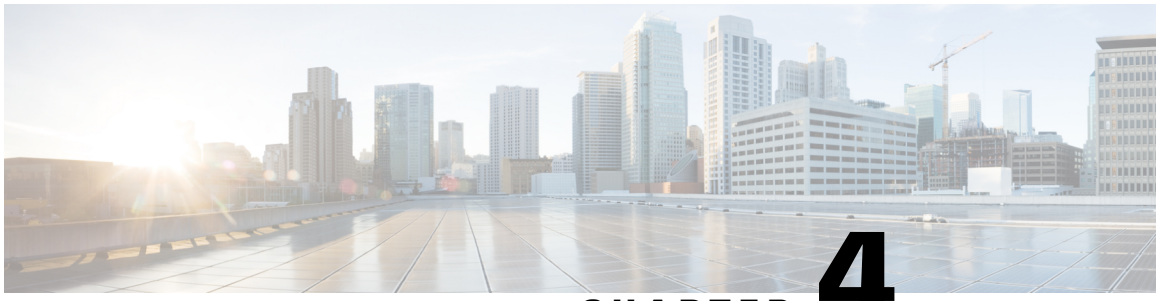
You must update appropriate licenses to use Cisco UCS Director Express for Big Data features:

- Standard
- Premium

The following table lists the features supported for the standard and premium licenses:

Features	Standard License	Premium License
Operating system and Hadoop software installation	X	X
Server, Network, and Storage provisioning	X	X
On-demand cluster creation and expansion	X	X
Customized cluster creation	X	X
Automated cluster node addition and deletion	X	X
Add pre-existing Hadoop nodes	X	X
Start and stop cluster services	X	X
Start, stop, and restart Cluster	X	X
Dashboard for health and status monitoring	X	X

Features	Standard License	Premium License
Support for latest Cloudera, MapR, and Hortonworks releases	X	X
Monitoring Storage and Network utilization	X	X
Monitoring Top Active and Long Running Jobs	X	X
On-Demand inventory collection	X	X
DIMM, Disk, Node, and Service Failure Alerts	X	X
Capacity Planning Alerts	X	X
Hadoop re-balancing	X	X
LDAP integration	X	X
Hadoop parameter configuration	X	X
Globalization and localization Support	X	X
Cusotmizable workflows	X	X
North Bound REST API support	X	X
Cluster configuration consistency checks		X
Cluster performance analysis		X
Historical performance analysis		X
Automated install and setup of new Hadoop services		X
Automated Hadoop version upgrade		X
Role-Based Access Control (RBAC)		X
Approval workflows		X



CHAPTER 4

Configuring Cisco UCS Manager Accounts

This chapter contains the following sections:

Pods

A pod is a logical grouping of physical and virtual components, which includes one or more physical or virtual accounts. It is a module of network, compute, storage, and application components that work together to deliver networking services. The pod is a repeatable pattern, and its components maximize the modularity, scalability, and manageability of data centers.

If needed, you can group pods into sites. The sites display on the **Converged** tab on the menu bar. For more information, see the Cisco UCS Director Administration Guide.

Adding a Pod

Step 1 On the menu bar, choose **Administration** > **Physical Accounts**.

Step 2 Click the **Pods** tab.

Step 3 Click **Add**.

Step 4 In the **Add Pod** dialog box, complete the following fields:

Name	Description
Name field	A descriptive name for the pod.
Site drop-down list	Choose the site where you want to add the pod. If your environment does not include sites, omit this step.

Name	Description
Type drop-down list	Choose the type of pod.
Description field	(Optional) A description of the pod.
Address field	The physical location of the pod. For example, includes the city or other internal identification used for the pod.
Hide POD check box	Check this check box to hide the pod if you do not want it to show in the Converged Check View. You can continue to add or delete accounts from the pod. For example, check this check box to ensure that a pod that does not have any physical or virtual elements, is not displayed in the Converged Check View.

Step 5 Click **Add**.

What to Do Next

Add one or more accounts to the pod.

Adding a Cisco UCS Manager Account

Each Cisco UCS Manager account represents a single Cisco UCS domain that has to be managed by Cisco UCS Director Express for Big Data.

Before You Begin

Add the pod to which this Cisco UCS Manager account belongs.

Step 1 On the menu bar, choose **Administration** > **Physical Accounts**.

Step 2 Click the **Physical Accounts** tab.

Step 3 Click **Add**.

Step 4 In the **Add Account** dialog box, complete the following fields:

Name	Description
Pod drop-down list	Choose the pod to which this account belongs.
Category Type drop-down list	Choose the **Computing** category. This is the type of infrastructure for the account.
Account Type drop-down list	Choose the **UCSM** account.

Step 5 Click **Sumit**.

Step 6 In the **Add Account** dialog box, complete the following fields:

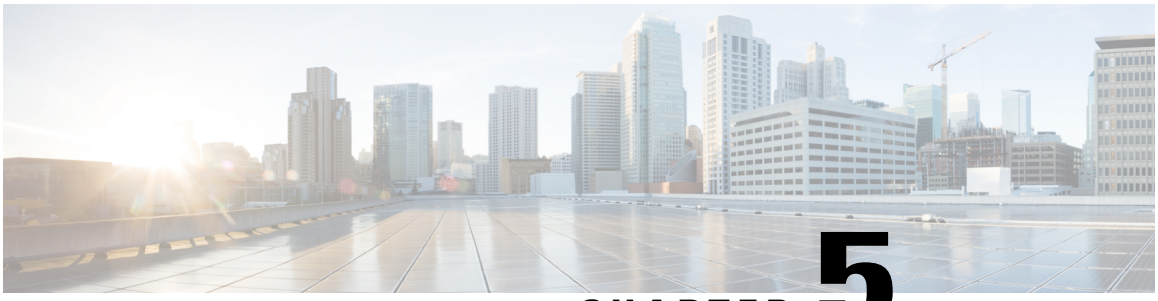
Name	Description
Authentication Type drop-down list	Choose the type of authentication to be used for this account. This can be one of the following: • **Locally Authenticated—**A locally authenticated user account is authenticated directly through the fabric interconnect and can be enabled or disabled by anyone with admin or AAA privileges. • **Remotely Authenticated—**A remotely authenticated user account is any user account that is authenticated through LDAP, RADIUS, or TACACS+.
Server Management drop-down list	Choose how the servers in this account to be managed. This can be one of the following: • **All Servers—**All servers are managed by Cisco UCS Director Express for Big Data. This option is the default. If you choose this option, all servers are added in the Managed state. • **Selected Servers—**Only selected servers are managed by Cisco UCS Director Express for Big Data. You can add and remove servers from the managed servers list, as needed. If you choose this option, all servers are added in the Unmanaged state.
Account Name field	A unique name that you have assigned to this account.
Server Address field	The IP address of Cisco UCS Manager. For a cluster configuration, this is the virtual IP address.
Use Credential Policy check box	Check this check box to use a credential policy that you have previously created with a Policy Name, Description, Username, Password, Protocol, and the Port.
Credential Policy drop-down list	Choose the credential policy. If you do not have a Credential Policy, create a New Credential Policy in the **Add Credential Dialog** box.
User ID field	The username that this account uses to access Cisco UCS Manager. This username must be the name of a valid account in Cisco UCS Manager.
Password field	The password associated with the username.
UCS Authentication Domain field	The authentication domain for the remotely authenticated account. This field is not displayed if you are using a locally authenticated account.

Name	Description
Transport Type drop-down list	Choose the transport type that you want to use for this account. This can be one of the following: • **http** • **https**
Port field	The port used to access Cisco UCS Manager.
Description field	(Optional) A description of this account.
Contact Email field	The email address of an administrator or other person responsible for this account.
Location field	The location of this account.
Service Provider field	(Optional) The name of the service provider, if any associated with this account.

Step 7 Click **Add**.

Cisco UCS Director Express for Big Data tests the connection to Cisco UCS Manager. If the test is successful, it adds the Cisco UCS Manager account and discovers all infrastructure elements in Cisco UCS Manager that are associated with that account, including chassis, servers, fabric interconnects, service profiles, and pools. This discovery process and inventory collection cycle takes approximately five minutes to complete.

The polling interval configured on the **System Tasks** tab specifies the frequency of inventory collection.



CHAPTER 5

Adding Big Data Accounts

This chapter contains the following sections:

Adding a Big Data Account

If you would like to manage your cluster using Cisco UCS Director Express for Big Data, you must add a Big Data account.

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Accounts**.

Step 2 Click the **Big Data Accounts** tab.

Step 3 Click **Add**.

Step 4 In the **Add Account** dialog box, complete the following fields:

Name	Description
Pod drop-down list	The pod to which the account is to be added.
Account Name field	The account name.
Account Type drop-down list	The type of Hadoop distribution used for the cluster: • Cloudera • MapR • Hortonworks
Management Console IP field	The management console IP address.

Name	Description
SSH Password field	The password associated with the SSH user name.
Confirm Password field	The password associated with the SSH user name.
Hadoop Manager Password field	The password associated with the Hadoop Manager for that account type.

Step 5 Click **Submit**.

Step 6 For the following actions, select a Big Data account.

Name	Description
Edit	Allows you to edit a Big Data account.
Delete	Deletes a Big Data account.
Check Configuration	Allows you to validate an existing cluster configuration.
Rollback Cluster	Allows you to roll back a cluster and make all the nodes in the cluster available for a Baremetal server.
Launch Hadoop Manager	Allows you to launch the Hadoop manager from Cisco UCS Director Express for Big Data
View Details	Provides details of a Big Data account.
Run Inventory	Collects the inventory of the Hadoop cluster for the selected Big Data Account and persists data in the Cisco UCS Director Express for Big Data database.
Configure Cluster	Allows you to customize cluster post Hadoop cluster creation.

Running a Cluster Inventory for a Big Data Account

When you create a Big Data Account for an Hadoop cluster, a new system task (inventory collector) is created for each Big Data Account in the Big Data Tasks folder here: **Administration** > **System** > **System Tasks**. The system task collects the inventory of the Hadoop cluster for the selected Big Data Account and persists data in the Cisco UCS Director database. This collector adds to the system scheduler so that it can be called periodically as per the frequency configured in the collector (for example, 30 minutes).

For more information on how to manage system tasks in Cisco UCS Director, see the latest *Cisco UCS Director Administration Guide*.

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Accounts**.
Step 2 Choose the Big Data Account for which you want to run the inventory.
Step 3 Click **Run Inventory**.
Step 4 Click **Submit**.

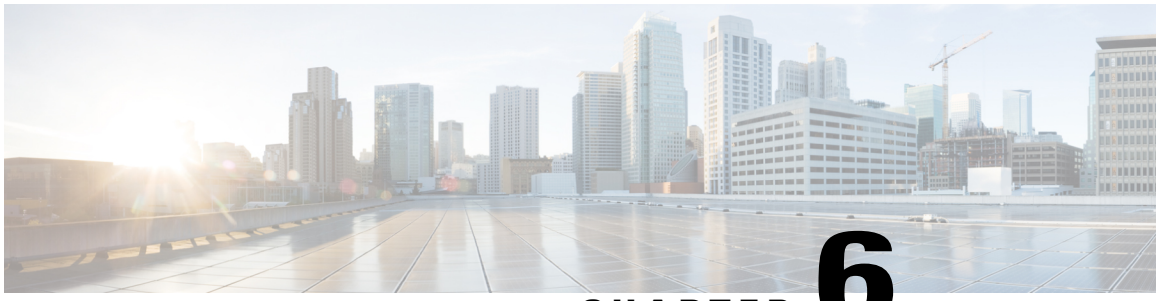
Rolling Back a Cluster for a Big Data Account

You can roll back a cluster and make all the nodes in the cluster available for a Baremetal server.

Step 1 On the menu bar, choose **Solutions** > **Big Data > Accounts**.
Step 2 Click the **Big Data Accounts** tab.
Step 3 Click on the Big Data account for which you want to rollback the cluster.
Step 4 Click **Rollback Cluster**.
Step 5 In the confirmation box, click **Submit**.

Access to Hadoop Managers from Cisco UCS Director Express for Big Data

You can access Hadoop Managers for all the Big Data accounts that you create in Cisco UCS Director Express for Big Data. On the menu bar, choose **Solutions** > **Big Data > Accounts** , and then the **Big Data Accounts** tab. You can launch the Hadoop Manager in the supported browsers by clicking the **Launch Hadoop Manager**.



CHAPTER 6

Configuring Big Data IP Pools

This chapter contains the following sections:

Big Data IP Pools

Big Data IP pools contain blocks of IP addresses that Cisco UCS Director Express for Big Data uses during the creation of Hadoop clusters. The IP addresses in each block must belong to the same subnet mask, default gateway, primary domain name server (DNS), and secondary DNS.

Note All IP addresses in a Big Data IP pool must be IPv4 addresses.

Adding a Big Data IP Pool

Step 1 On the menu bar, choose **Solutions** > **Big Data > Settings**.

Step 2 Click the **Big Data IP Pools** tab.

Step 3 Click **Add**.

Step 4 In the **IP Pool** page of the **Create an IP Pool** wizard, complete the following fields:

Name	Description
IP Pool Name field	A unique name for the IP Pool.
Description field	A short description that identifies the purpose of the pool.

Name	Description
Assignment Order drop-down list	Choose the assignment order. This could be one of the following: • **Default—**A random identity is selected from the pool. • **Sequential—**The lowest available identity is selected from the pool.

Step 5 Click **Next**.

Step 6 In the **IPv4 Addresses** page of the **Create an IP Pool** wizard, complete the following fields:

a) In the **IPv4 Blocks** field, click **Add (+)**.

b) In the **Add Entry to IPv4 Blocks** dialog box, enter the IP addresses that are to be included in the IP pool in the **Static IP Pool** field.
This can be a range of IP addresses, or a series of IP addresses separated by a comma (,).

c) In the following fields, enter the appropriate information for all the IP addresses in the pool:

- **Static IP Pool** field
- **Subnet Mask** field
- **Default Gateway** field
- **Primary DNS** field
- **Secondary DNS** field

d) Click **Submit**.

Step 7 Click **Submit**.

Managing Big Data IP Pools

Step 1 On the menu bar, choose **Solutions** > **Big Data > Settings**.

Step 2 Click the **Big Data IP Pools** tab.

Step 3 The Big Data IP Pools tab provides the following actions:

Name	Description
Refresh	Refreshes the current page.
Add	Adds a new IP pool.

Step 4 For the following additional actions, choose an IP Pool from the table:

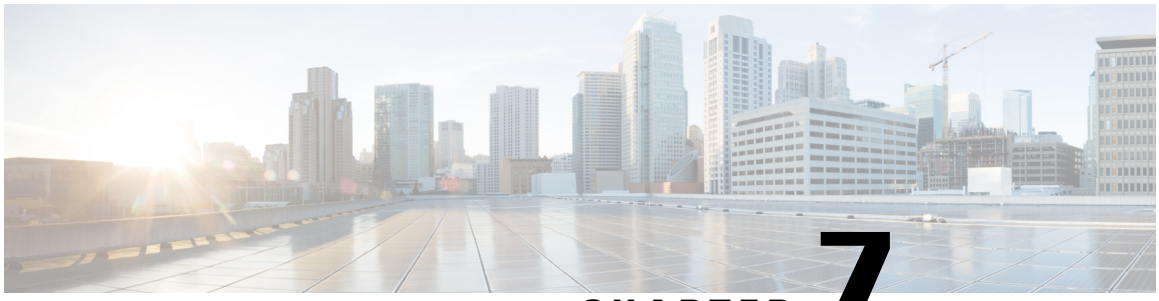
Name	Description
Edit	Modifies the IP pool specification.
Delete	Deletes the IP pool specification.
View Details	Allows you to view the IPv4 addresses in the IP pool and to view more reports. **Note** If you see a **License Status** tab, it indicates a licensing issue.

Step 5 With an IP Pool selected, click on the **View Details** button.

Step 6 Click the **IPv4 Addresses** tab to display the IP Addresses associated with the selected IP Pool.

Step 7 Select an IP Address from the table. The following buttons are displayed.

Name	Description
Release IP Address	The IP address is made available to be assigned against any BareMetal server.
Release Multiple IP Addresses	Enables you to choose and release more than one IP address from the **Release Multiple IP Addresses** dialog box.



CHAPTER 7

Configuring Cisco UCS Service Profile Templates for Big Data

This chapter contains the following sections:

Cisco UCS Service Profile Templates for Big Data

Cisco Unified Computing System (Cisco UCS) service profiles are a powerful means for streamlining the configuration and management of Cisco UCS servers. They provide a mechanism for rapidly provisioning servers and their associated network connections with consistency in all details of the environment. They can be set up in advance before physically installing the servers.

Service profiles are built on policies—administrator-defined sets of rules and operating characteristics such as the server identity, interfaces, and network connectivity. Every active server in your Hadoop cluster must be associated with a service profile.

The Cisco UCS service profile template for Big Data enables you to set up the configuration for the servers in your Hadoop cluster. The service profile template for Big Data is included in a Hadoop cluster deploy template. When the deploy template is applied to the servers in your Hadoop cluster, the service profile template is used to configure one or more service profiles that are applied to the servers in your Hadoop cluster.

Note The service profile template for Big Data wizard gathers the information needed to create a service profile template. You can only apply this service profile template through the Hadoop cluster deploy template.

For more information about service profiles and service profile templates, see the Cisco UCS Manager configuration guides.

Creating a Cisco UCS Service Profile Template for Big Data

Before You Begin

Add a Cisco UCS Manager account.

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Containers**.

Step 2 Click the **UCS SP Templates for Big Data** tab.

Step 3 Click **Add (+)**.

Step 4 On the **UCS SP Template for Big Data** page of the **Create UCS SP Template for Big Data** wizard, complete the following fields:

Name	Description
Template name field	A unique name for the template.
Template Description field	The description of the template.
Template Type drop-down list	Service profiles created from an initial template inherit all the properties of the template. Service profiles created from an initial service profile template are bound to the template. However, changes to the initial template do not automatically propagate to the bound service profiles. If you want to propagate changes to bound service profiles, unbind and rebind the service profile to the initial template.

Step 5 Click **Next**.

What to Do Next

Create a QoS policy.

Creating a QoS Policy

The quality of service (QoS) policy assigns a system class to the outgoing traffic for a vNIC or vHBA. This system class determines the quality of service for that traffic. For certain adapters, you can also specify additional controls on the outgoing traffic, such as burst and rate.

Step 1 On the **Create QoS Policy** page of the **Create UCS SP Template for Big Data** wizard, do one of the following:

- To accept the default QoS policies, click **Next**.
- To create one or more custom QoS policies, click **Add (+)** and continue with Step 2.

- To review or modify one of the default QoS policies, choose the policy in the table and click **Edit**. For information about the fields in the **Edit QoS Policy Entry** dialog box, see Step 2.

Step 2 In the **Add Entry to QoS Policy** dialog box, complete the following fields:

Name	Description
Name field	A unique name for the policy.
Priority drop-down list	Choose the priority assigned to this QoS policy. This can be one of the following: • **Fc—**Use this priority for QoS policies that control only vHBA traffic. • **Platinum—**Use this priority for QoS policies that control only vNIC traffic. • **Gold—**Use this priority for QoS policies that control only vNIC traffic. • **Silver—**Use this priority for QoS policies that control only vNIC traffic. • **Bronze—**Use this priority for QoS policies that control only vNIC traffic. • **Best Effort—**Do not use this priority. It is reserved for the Basic Ethernet traffic lane. If you assign this priority to a QoS policy and configure another system class as CoS 0, Cisco UCS does not default to this system class. It defaults to the priority with CoS 0 for that traffic.
Burst(Bytes) field	The normal burst size for servers that use this policy. This field determines the maximum size of traffic bursts beyond which the traffic is considered to exceed the rate limit. The default is 10240. The minimum value is 0, and the maximum value is 65535. This setting is not applicable to all adapters.

Name	Description
Rate drop-down list	Choose the expected average rate of traffic. Traffic that falls under this rate will always conform. This can be one of the following: • **line-rate—**Equals a value of 0 and specifies no rate limiting. This is the default value. • **Specify Manually—**Enables you to specify the rate in a field. The minimum value is 0, and the maximum value is 40,000,000. The granularity for rate limiting on a Cisco UCS M81KR Virtual Interface Card adapter is 1 Mbps. The adapters treat the requested rate as a "not-to-exceed" rate. Therefore, a value of 4.5 Mbps is interpreted as 4 Mbps. Any requested rate of more than 0 and less than 1 Mbps is interpreted as 1 Mbps, which is the lowest supported hardware rate limit. Rate limiting is not applicable to all adapters. For example, this setting is not supported on the Cisco UCS VIC-1240 Virtual Interface Card.
Host Control drop-down list	Decides whether Cisco UCS controls the class of service (CoS) for a vNIC. This setting has no effect on a vHBA. This is set to**None—**Cisco UCS uses the CoS value associated with the priority selected in the Priority drop-down list regardless of the CoS value assigned by the host.

Step 3 Click **Submit**.

Step 4 Click **Next**.

What to Do Next

Create a VLAN policy.

Creating a VLAN Policy

The VLAN policy creates a connection to a specific external LAN in the underlying infrastructure of a Hadoop cluster that is within a single Cisco UCS domain. The VLAN isolates traffic to that external LAN, including broadcast traffic.

Step 1 On the **Create VLAN Policy** page of the **Create UCS SP Template for Big Data** wizard, do one of the following:

- To accept the default VLAN policies, click **Next**.
- To create one or more custom VLAN policies, click **Add (+)** and continue with Step 2.
- To review or modify one of the default VLAN policies, choose the policy in the table and click **Edit**. For information about the fields in the **Edit VLAN Policy Entry** dialog box, see Step 2.

Step 2 In the **Add Entry to VLAN Policy** dialog box, complete the following fields:

Name	Description
VLAN Name field	The name of the VLAN policy.
Fabric ID drop-down list	Choose how the VLAN should be configured. This can be one of the following: • **Common/Global—**The VLAN maps to the same VLAN ID in all available fabrics. • **Fabric A—**The VLAN maps to a VLAN ID that exists only in fabric A. • **Fabric B—**The VLAN maps to a VLAN ID that exists only in fabric B.
Sharing drop-down list	By default, it is set to **None**.

Step 3 Click **Submit**.

Step 4 Click **Next**.

What to Do Next

Create a vNIC policy.

Creating a vNIC Policy

The vNIC policy defines how a vNIC on a server connects to the LAN. Each server in a Hadoop cluster requires a vNIC policy for each of the following NICs:

- MGMT
- DATA
- HDFS

Step 1 On the **Create vNIC Policy** page of the **Create UCS SP Template for Big Data** wizard, do one of the following:

- To accept the default vNIC policies, click **Next**.

- To create one or more custom vNIC policies, click **Add (+)** and continue with Step 2.
- To review or modify one of the default vNIC policies, choose the policy in the table and click **Edit**. For information about the fields in the **Edit vNIC Policy Entry** dialog box, see Step 2.

Step 2 In the **Add Entry to vNIC Policy** dialog box, complete the following fields:

Name	Description
vNIC Name field	Name of the vNIC.
Fabric ID drop-down list	Choose the fabric interconnect with which the vNICs created with this policy are associated. If you want vNICs created from this policy to be able to access the second fabric interconnect when the default one is unavailable, check the **Enable Failover** check box. Do not enable vNIC fabric failover under the following circumstances: **Note** • If the Cisco UCS domain is running in Ethernet Switch Mode, vNIC fabric failover is not supported in that mode. If all Ethernet uplinks on one fabric interconnect fail, the vNICs do not fail over on other fabric interconnect. • If you associate one or more vNICs created from this template with a server that has an adapter, which does not support fabric failover, such as the Cisco UCS 82598KR-CI 10-Gigabit Ethernet Adapter, Cisco UCS Manager generates a configuration fault when you associate the service profile with the server.
VLANs area	In the VLANs area, do the following to select the VLAN to be assigned to vNICs created from this policy: **1** Click **Add**. **2** In the **Add Entry to VLANs** dialog box, complete the following fields: • **Name** drop-down list—Choose the VLAN that you want to associate with the vNIC template. • **Set as Native VLAN** check box—Check the check box if you want this VLAN to be the native VLAN for the port. **3** Click **Submit**.

Name	Description
MTU field	The MTU, or packet size that must be used by the vNICs created from this vNIC policy. Enter an integer between 1500 and 9216. **Note** If the vNIC template has an associated QoS policy, the MTU specified here must be equal to, or less than the MTU specified in the associated QoS System class. If this MTU value exceeds the MTU value in the QoS system class, packets might be dropped during data transmission.
Pin Group drop-down list	This is a *display-only* field.
Adapter Policy field	This field is auto-populated with **Linux**.
Dynamic vNIC Connection Policy drop-down list	This is a *display-only* field.
QoS Policy drop-down list	Choose the quality of service policy that must be used by the vNICs created from this vNIC policy.
Network Control Policy drop-down list	This is a *display-only* field.

Step 3 Click **Submit**.

Step 4 Click **Next**.

What to Do Next

Create a boot order policy.

Creating a Boot Order Policy

The Cisco UCS Manager boot policy overrides the boot order in the BIOS setup menu, and determines the following:

- Selection of the boot device
- Location from which the server boots
- Order in which boot devices are invoked

Step 1 On the **Create a Boot Order Policy** page of the **Create UCS SP Template for Big Data** wizard, set the order in which you want the following devices to boot:

- **CD-ROM**

- **Storage**
- **LAN**

If you do not want to boot from a specific device, choose the blank space at the bottom of the drop-down list.

Note If you are booting for the first time, choose **1** for the LAN drop-down list to set it as the first boot device.

Step 2 In the **Select vNIC Policy for LAN Ethernet** pane, click **Add (+)**.

Step 3 In the **Add Entry to Select vNIC Policy for LAN Ethernet** dialog box, do the following:

a) From the **Select vNIC** drop-down list, choose the vNIC that you want to assign to the LAN.
b) If you want the VLAN shown in the **VLAN** field to be the primary VLAN, check the **Set as Primary** check box.
c) Click **Submit**.

Step 4 If you want to choose vNIC policies for other VLANs, repeat Steps 2 and 3 with a different vNIC from the **Select vNIC** drop-down list.

Step 5 Click **Next**.

What to Do Next

Create a BIOS policy.

Creating a BIOS Policy

The BIOS policy automates the configuration of certain BIOS settings for the servers in your Hadoop cluster.

Note All the drop-down lists on this page are set to **Platform Default—**The BIOS uses the value for this attribute contained in the BIOS defaults for the server type and vendor.

Step 1 On the **Create UCS BIOS Policy** page of the **Create UCS SP Template for Big Data** wizard, do the following:

Name	Description
Quiet Boot drop-down list	Decides what the BIOS displays during Power On Self-Test (POST).
Post Error Pause drop-down list	Decides what happens when the server encounters a critical error during POST.
Resume AC on Power Loss drop-down list	Decides how the server behaves when power is restored after an unexpected power loss.
From Panel Lockout drop-down list	Decides whether the power and reset buttons on the front panel are ignored by the server.

Step 2 Click **Next**.

What to Do Next

Create a Local Disk Configuration Policy.

Creating a Local Disk Configuration Policy

This policy defines the disk configuration for disk storage local to the server and provides a flexible mechanism to define the RAID levels or JBOD configuration for name nodes and data nodes resident in a Hadoop cluster. This policy enables you to set a local disk configuration for all servers associated with a service profile.

Step 1 On the **Local Disk Configuration Policy** page of the **Create UCS SP Template for Big Data** wizard, review the following fields:

Name	Description
Configure RAID Policy table	RAID level configuration for Hadoop NameNode and DataNode (OS and data drives).

Step 2 Click **Submit**.

What to Do Next

Create a Hadoop cluster profile template.

Editing RAID Policy

Use this procedure to edit the disk configuration of local disk drives attached to the hardware RAID controller for C220 and C240 M3/M4 servers as well as the standalone boot drives on the C240 M4 servers.

Before You Begin

Create a Local Disk Configuration Policy.

Step 1 On the **Local Disk Configuration Policy** page of the **Create UCS SP Template for Big Data** wizard, choose the node type in the **Configure RAID Policy** table.

Step 2 In the **Configure RAID Policy** table, click the **Edit selected entry in the table below** icon.

Step 3 In the **Edit Configure RAID Policy Entry** dialog box, complete the following fields:

Name	Description
Node Type	The selected node type.

Name	Description
OS Disks	
Use HDD drives on M4 servers with insufficient standalone boot drives check box	Check this check box for M4 servers with insufficient standalone boot drives.
Use standalone boot drives for M4 Servers check box	By default, this option is checked and disabled. Check this check box for M4 servers with sufficient standalone boot drives.
Use JBOD (Just a Bunch Of Disks) for M4 Servers check box	Check this check box for M4 servers with multiple hard drives.
RAID Level [OS] drop-down list	Choose the RAID level for the OS (operating system) disks.
Disks Per Groups field	Specifies the number of disks that can exist per group during RAID configuration.
Wrire Mode drop-down list	Choose either **Write through** to write data without the RAID controller cache or **Write back** to write data with the cache.
Read Mode drop down-list	Choose the method to read data from the disks.
Use Cache check box	Check the **Use Cache** check box that allows you to use the RAID controller cache to read and write operations.
Use Cache if Bad BBU check box	Check the **Use Cache if Bad BBU** check box if the Battery Backup Unit (BBU) is not available for any reason that allows you to disable **Write back** and turn on **Write Through**.
Strip Size (MB) drop- down	Choose the strip size in MB, the data that is residing on each disk in the stripe.
Data Disks	
Use JBOD (Just a Bunch Of Disks) for M4 Servers check box	Check this check box for M4 servers with multiple hard drives.
RAID Level [DATA] drop-down list	Choose the RAID level for the Hadoop data disks.
Disks Per Groups field	Specifies the number of disks that can exist per group during RAID configuration.
Wrire Mode drop-down list	Choose either **Write through** to write data without the RAID controller cache or **Write back** to write data with the cache.

Name	Description
Read Mode drop down-list	Choose the method to read data from the disks.
Use Cache check box	Check the **Use Cache** check box that allows you to use the RAID controller cache to read and write operations.
Use Cache if Bad BBU check box	Check the **Use Cache if Bad BBU** check box if the Battery Backup Unit (BBU) is not available for any reason that allows you to disable **Write back** and turn on **Write Through**.
Strip Size (MB) drop- down	Choose the strip size in MB, the data that is residing on each disk in the stripe.

Step 4 Click **Submit**.

Cloning a Cisco UCS Service Profile Template

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Containers**.

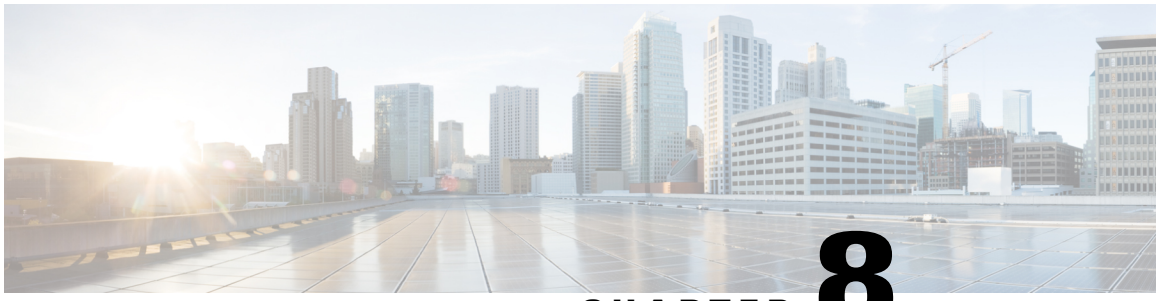
Step 2 Click the **UCS SP Templates for Big Data** tab.

Step 3 Click the row for the template that you want to clone.

Step 4 Click **Clone**.

Step 5 In the **UCS SP Template Specification** page of the Clone UCS SP Template for Big Data wizard, do the following:

a) Enter a unique name and description for the new service profile template.
b) Choose the Template Type from the drop down list.
c) Click **Next**, review the information on each page, and modify, if required.
d) Click **Submit**.



CHAPTER 8

Configuring Hadoop Cluster Profile Templates

This chapter contains the following sections:

Hadoop Cluster Profile Templates

The Hadoop cluster profile template specifies the number of nodes in the cluster and takes care of provisioning and configuring the Hadoop cluster services. Numerous Apache Software Foundation projects develop and comprise the services required to deploy, integrate, and work with Hadoop. Some Hadoop distributions support only a subset of these services, or may have their own distribution-specific services.

Each of the following has been developed to deliver an explicit function:

- HDFS
- CLDB
- YARN
- ZooKeeper
- HBase
- Hive
- Oozie
- Hue
- Spark
- Key-Value Store Indexer
- Solr
- Sqoop
- Impala

- Flume
- PIG
- MAHOUT
- Falcon
- Tez
- Storm
- Ganglia

Note You are not allowed to deselect some of the services because to create a Hadoop cluster, it is necessary to have these services enabled.

Creating a Hadoop Cluster Profile Template

Before You Begin

Create a Hadoop cluster Configuration Template.

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Containers**.

Step 2 Click the **Hadoop Cluster Profile Templates** tab.

Step 3 Click **Add (+)**.

Step 4 On the **Hadoop Cluster Profile Template** page of the **Create Hadoop Cluster Profile Template** wizard, complete the following fields:

Name	Description
Template Name field	A unique name for the template.
Template Description field	A short description for the template.
Node Count field	The number of nodes in the cluster. The default is four nodes.
Hadoop Distribution drop-down list	Choose the type of Hadoop distribution.
Hadoop Distribution Version	Choose the Hadoop Distribution Version.
Hadoop Configuration Parameters Name drop-down list	Choose the Hadoop cluster configuration template name.

Step 5 Click **Next**.

What to Do Next

Create a Services Selection policy.

Creating a Services Selection Policy

The cluster policy contains the Hadoop cluster services that you want to enable in the Hadoop cluster.

Note The **Service Selection Page** displays the Hadoop cluster services depending on the Hadoop distribution that you select before on the **Hadoop Cluster Profile Template** page.

Step 1 On the **Services Selection Policy** page of the **Create Hadoop Cluster Profile Template** wizard, check the check box for the optional Hadoop cluster services that you want to enable in your cluster.
Some Hadoop cluster services are required for the distribution and cannot be disabled. The available Hadoop cluster services include the following:

- **HDFS—**A file system that spans all nodes in a Hadoop cluster for data storage. This service replicates data across multiple nodes to avoid data loss.
- **YARN—** A resource-management platform responsible for managing compute resources in clusters and using them for scheduling users' applications.
- **HBase—**A high speed read and write column-oriented database.
- **Hive—**The query engine framework for Hadoop that facilitates easy data summarization, ad-hoc queries, and the analysis of large datasets stored in HDFS and HBase. With SQL-like semantics, Hive makes it easy for RDBMS users to transition into querying unstructured data in Hadoop.
- **Oozie—**A workflow environment for coordinating complex data processing operations.
- **ZooKeeper—**An infrastructure for cross-node synchronization that can be used by applications to ensure that tasks across the cluster are serialized or synchronized.
- **Hue—**An interface that aggregates the most common Hadoop components to improve user experience and to enable users to avoid the underlying complexity and the command line interface.
- **Spark—**An open-source data analytics cluster computing framework.
- **Key-Value Store Indexer—**A method for indexing data across the cluster.
- **SOLR—**A method for searching data across the cluster.
- **Sqoop—**A client-server tool that transfers bulk data between Hadoop and structured datastores, such as relational databases.
- **Impala—**A massively parallel processing (MPP) SQL query engine that runs natively in Apache Hadoop.

- **Flume—**A distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of streaming data into the Hadoop Distributed File System (HDFS).

Step 2 Click **Next**.

What to Do Next

Configure the Rack Assignment policy.

Configuring the Rack Assignment Policy

Step 1 On the **Rack Assignment Policy** page of the **Create Hadoop Cluster Profile Template** wizard, do one of the following:

- To create one or more custom Hadoop node configure policies, click **Add (+)** and continue with Step 2.
- To modify the default Hadoop node configure policy, choose the policy in the table and click **Edit**. For information about the fields in the **Edit Hadoop Node Configuration Policy Entry** dialog box, see Step 2.

Step 2 In the **Add Entry to Hadoop Node Configure Policy** dialog box, do the following:

a) In the **Rack Name** field, enter the name of the rack server.
b) In the **DataNodes** field, click **Select** and check the checkbox for each node that you want to configure on that server.
Note Some Hadoop cluster services require a minimum number of nodes. For example, Zookeeper requires a minimum of 3 nodes.
c) Click **Submit**.

Step 3 Click **Next**.

What to Do Next

Configure the HDFS policy.

Configuring the HDFS Policy

Step 1 On the **HDFS Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the HDFS policy configuration and click **Edit**.
If you do not see a node you need for HDFS on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit HDFS Policy Entry** dialog box, review and, if required, change the following fields:

Name	Description
DataNode drop-down list	Choose **Yes** if you want the node to act as the DataNode for HDFS. Otherwise, choose **No**. The data nodes store and retrieve data on request by the name node or by the client.
Primary NameNode drop-down list	Choose **Yes** if you want the node to act as the primary name node for HDFS. Otherwise, choose **No**. All operations of the HDFS cluster are maintained by the primary name node. There can be only one primary name node for the HDFS Cluster.
Secondary NameNode drop-down list	Choose **Yes** if you want the node to act as a secondary name node for HDFS. Otherwise, choose **No**. The secondary name node is not a direct replacement for the primary name node. The main role of a secondary name node is to periodically merge the FSImage and edit log, to prevent the edit log from becoming too large. A secondary name node runs on a separate physical system because it requires a lot of memory to merge two files. It keeps a copy of the merged file in its local file system so that it is available for use if the primary name node fails.
Balancer drop-down list	Choose **Yes** if you want the node to act as a balancer for HDFS. Otherwise, choose **No**.
HTTPFS drop-down list	Choose **Yes** if you want the node to act as HTTPFS for HDFS. Otherwise, choose **No**. **Note** This service provides HTTP access to HDFS.
Fail Over Controller drop-down list	Choose **Yes** if you want the node to act as Fail Over Controller for HDFS. Otherwise, choose **No**.
Gateway drop-down list	Choose **Yes** if you want the node to act as Gateway for HDFS. Otherwise, choose **No**.
Journal Node drop-down list	Choose **Yes** if you want the node to act as Journal node for HDFS. Otherwise, choose **No**.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for HDFS.

Step 5 Click **Next**.

What to Do Next

Configure the CLDB policy.

Configuring the CLDB Policy

Step 1 On the **CLDB Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the CLDB policy configuration and click **Edit**.
If you do not see a node you need for CLDB on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit CLDB Policy Entry** dialog box, choose **Yes** if you want the node to act as a CLDB agent.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for CLDB.

Step 5 Click **Next**.

What to Do Next

Configure the YARN policy.

Configuring the YARN Policy

Step 1 On the **YARN Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the YARN policy configuration and click **Edit**.
If you do not see a node you need for the YARN policy on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit YARN Policy Entry** dialog box, review and, if required, change the following fields:

Name	Description
Resource Manager drop-down list	Choose **Yes** if you want the node to act as a Resource Manager. Otherwise, choose **No**. The Resource Manager is the ultimate authority that allocates resources among all the applications in the system.
Node Manager drop-down list	Choose **Yes** if you want the node to act as a task Node Manager. Otherwise, choose **No**. The Node Manager is responsible for launching the applications' containers, monitoring their resource usage (CPU, memory, disk, network), and reporting to the Resource Manager.

Name	Description
Gateway drop-down list	Choose **Yes** if you want the node to act as a Gateway. Otherwise, choose **No**.
JobHistory drop-down list	Choose **Yes** if you want the node to preserve the JobHistory. Otherwise, choose **No**.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for Yarn.

Step 5 Click **Next**.

What to Do Next

Configure the ZooKeeper policy.

Configuring the ZooKeeper Policy

Note You must configure a minimum of three nodes for ZooKeeper.

Step 1 On the **ZooKeeper Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the ZooKeeper policy configuration and click **Edit**.
If you do not see a node you need for ZooKeeper on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit ZooKeeper Policy Entry** dialog box, choose **Yes** to make the node to act as a ZooKeeper.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for ZooKeeper.

Step 5 Click **Next**.

What to Do Next

Configure the HBase policy.

Configuring the HBase Policy

Step 1 On the **HBase Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the HBase policy configuration and click **Edit**.
If you do not see a node you need for HBase on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit HBase Policy Entry** dialog box, review and, if required, change the following fields:

Name	Description
HBase Master drop-down list	Choose **Yes** if you want the node to act as the HBase master. Otherwise, choose **No**.
Region Server drop-down list	Choose **Yes** if you want the node to act as a region server. Otherwise, choose **No**.
HBase Thrift Server drop-down list	Choose **Yes** if you want the node to host HBase Thrift. Otherwise, choose **No**.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for HBase.

Step 5 Click **Next**.

What to Do Next

Configure the Hive policy.

Configuring the Hive Policy

Step 1 On the **Hive Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Hive policy configuration and click **Edit**.
If you do not see a node you need for Hive on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit Hive Policy Entry** dialog box, review and, if required, change the following fields:

Name	Description
HiveServer2 drop-down list	Choose **Yes** if you want the node to host HiveServer2. Otherwise, choose **No**.

Name	Description
Hive Metastore Server drop-down list	Choose **Yes** if you want the node to act as a Hive metastore. Otherwise, choose **No**.
WebHCat drop-down list	Choose **Yes** if you want the node to act as a WebHCat. Otherwise, choose **No**. WebHCat is the REST API for HCatalog, a table and storage management layer for Hadoop.
Gateway drop-down list	Choose **Yes** if you want the node to act as a Gateway for Hive. Otherwise, choose **No**.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for Hive.

Step 5 Click **Next**.

What to Do Next

Configure the Oozie policy.

Configuring the Oozie Policy

Step 1 On the **Oozie Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Oozie policy configuration and click **Edit**.
If you do not see a node you need for Oozie on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit Oozie Policy Entry** dialog box, click **Yes** to make the node to act as an Oozie server.

Step 3 Repeat Steps 1 and 2 to configure the other nodes for Oozie.

Step 4 Click **Next**.

What to Do Next

Configure the Hue policy.

Configuring the Hue Policy

Step 1 On the **Hue Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Hue policy configuration and click **Edit**.
If you do not see a node you need for Hue on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit Hue Policy Entry** dialog box, do the following:

a) From the **Hue Server** drop-down list, choose **Yes** if you want the node to act as a Hue server.
b) From the **BeesWax Server** drop down-list, choose **Yes** if you want the node to act as a BeesWax server.
c) From the **Kt Renewer** drop down-list, choose **Yes** if you want the node to act as a Kt Renewer.
d) Click **Submit**.

Step 3 Repeat Steps 1 and 2 to configure the other nodes for Hue.

Step 4 Click **Next**.

What to Do Next

Configure the Spark policy.

Configuring the Spark Policy

Step 1 On the **Spark Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Spark policy configuration and click **Edit**.
If you do not see a node you need for Spark on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit Spark Policy Entry** dialog box, review and, if required, change the following fields:

Name	Description
Spark Master drop-down list	Choose **Yes** if you want the node to act as a Spark master. Otherwise, choose **No**.
Spark Worker drop-down list	Choose **Yes** if you want the node to act as a Spark worker. Otherwise, choose **No**.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for Spark.

Step 5 Click **Next**.

What to Do Next

Configure the KSIndexer policy.

Configuring the Key-Value Store Indexer Policy

Step 1 On the **Key-Value Store Indexer Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Key-Value Store Indexer policy configuration and click **Edit**.
If you do not see a node you need for KSIndexer on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit KSIndexer Policy Entry** dialog box, choose **Yes** if you want the node to act as a KSIndexer server.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for KSIndexer.

Step 5 Click **Next**.

What to Do Next

Configure the Solr policy.

Configuring the Solr Policy

Step 1 On the **Solr Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Solr policy configuration and click **Edit**.
If you do not see a node you need for Solr on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit Solr Policy Entry** dialog box, choose **Yes** if you want the node to act as a Solr server.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for Solr.

Step 5 Click **Next**.

What to Do Next

Configure the Sqoop policy.

Configuring the Sqoop Policy

Step 1 On the **Sqoop Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Sqoop policy configuration and click **Edit**.
If you do not see a node you need for Sqoop on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit Sqoop Policy Entry** dialog box, choose **Yes** if you want the node to act as a Sqoop server.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for Sqoop.

Step 5 Click **Next**.

What to Do Next

Configure the Impala policy.

Configuring the Impala Policy

Step 1 On the **Impala Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Impala policy configuration and click **Edit**.
If you do not see a node you need for Impala on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit Impala Policy Entry** dialog box, do the following:

a) From the **Impala Daemon** drop- down, choose **Yes** if you want the node to act as an Impala daemon.

b) From the **Impala StateStore** drop-down list, choose **Yes** if you want the node to act as an Impala Statestore.

c) From the **Impala Catalog Server** drop-down list, choose **Yes** if you want the node to act as an Impala catalog server. The other fields in this dialog box are for your information only.

d) Click **Submit**.

Step 3 Repeat Steps 1 and 2 to configure the other nodes for Impala.

Step 4 Click **Submit**.

What to Do Next

Configure the Flume policy.

Configuring the Flume Policy

Step 1 On the **Flume Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Flume policy configuration and click **Edit**.
If you do not see a node you need for Flume on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit Flume Policy Entry** dialog box, choose **Yes** if you want the node to act as a Flume agent.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for Flume.

Step 5 Click **Next**.

What to Do Next

Configure the PIG Policy.

Configuring the PIG Policy

Step 1 On the **Pig Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Pig policy configuration and click **Edit**.
If you do not see a node you need for Pig on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit Pig Policy Entry** dialog box, choose **Yes** if you want the node to act as a Pig agent.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for Pig.

Step 5 Click **Next**.

What to Do Next

Configure the MAHOUT Policy.

Configuring the MAHOUT Policy

Step 1 On the **MAHOUT Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the MAHOUT policy configuration and click **Edit**.
If you do not see a node you need for MAHOUT on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit MAHOUT Policy Entry** dialog box, choose **Yes** if you want the node to act as a MAHOUT agent.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for MAHOUT.

Step 5 Click **Submit**.

What to Do Next

Configure a Falcon Policy.

Configuring the Falcon Policy

Step 1 On the **Falcon Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Falcon policy configuration and click **Edit**.
If you do not see a node you need for Falcon on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit Falcon Policy Entry** dialog box, choose **Yes** if you want the node to act as a Falcon server and as the Falcon client from the **Falcon Server** and **Falcon Client** drop-down lists.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for Falcon.

Step 5 Click **Submit**.

What to Do Next

Configure the Tez Policy.

Configuring the Tez Policy

Step 1 On the **Tez Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Tez policy configuration and click **Edit**.
If you do not see a node you need for Tez on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit Tez Policy Entry** dialog box, choose **Yes** if you want the node to act as a Tez agent.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for Tez.

Step 5 Click **Submit**.

What to Do Next

Configure the Storm Policy.

Configuring the Storm Policy

Step 1 On the **Storm Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Storm policy configuration and click **Edit**.
If you do not see a node you need for Storm on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit Storm Policy Entry** dialog box, do the following:

a) In the **DRPC Server** drop-down list , choose **Yes** if you want the node to act as a DRPC server.
b) In the **Nimbus** drop-down list, choose **Yes** if you want the node to act as a Nimbus server.
c) In the **Storm REST API Server** drop-down list, choose **Yes** if you want the node to act as a Storm REST API server.
d) In the **Storm UI Server** drop-down list, choose **Yes** if you want the node to act as a Storm UI server.
e) In the **Supervisor** drop-down list, choose **Yes** if you want the node to act as a supervisor.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for Storm.

Step 5 Click **Submit**.

What to Do Next

Configure the Ganglia Policy.

Configuring the Ganglia Policy

Step 1 On the **Ganglia Policy** page of the **Create Hadoop Cluster Profile Template** wizard, click the row in the table with the node for which you want to change the Ganglia policy configuration and click **Edit**.
If you do not see a node you need for Ganglia on this page, click **Back** to return to the **Select Nodes for Rack Server** page and add the node there.

Step 2 In the **Edit Ganglia Policy Entry** dialog box, choose **Yes** if you want the node to act as a Ganglia server and as Ganglia monitor from the **Ganglia Server** and **Ganglia Monitor** drop-down lists.

Step 3 Click **Submit**.

Step 4 Repeat Steps 1 and 2 to configure the other nodes for Ganglia.

Step 5 Click **Submit**.

Cloning a Hadoop Cluster Profile Template

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Containers**.

Step 2 Click the **Hadoop Cluster Profile Template** tab.

Step 3 Click the row for the template that you want to clone.

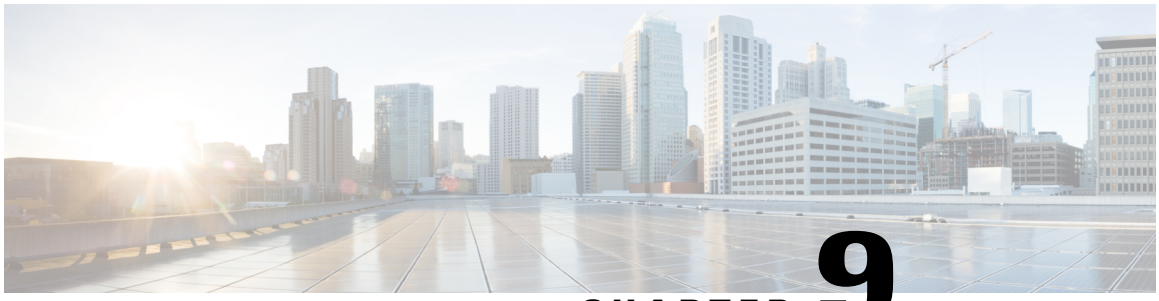
Step 4 Click **Clone** .

Step 5 In the **Clone Hadoop Cluster Profile Template** dialog box, do the following:

a) Enter a unique name and description for the new Hadoop cluster profile template.

b) Click **Next**, review the information on each page, and modify, if required.

c) Click **Submit**.



CHAPTER 9

Configuring Cisco UCS Server Pool Policies

This chapter contains the following sections:

Server Pool Policy Qualifications

This policy qualifies servers based on the inventory of a server conducted during the discovery process. The qualifications are individual rules that you configure in the policy to determine whether a server meets the selection criteria. For example, you can create a rule that specifies the minimum memory capacity for servers in a data center pool.

Qualifications are used in other policies to place servers, not just by the server pool policies. For example, if a server meets the criteria in a qualification policy, it can be added to one or more server pools or have a service profile automatically associated with it.

You can use the server pool policy qualifications to qualify servers according to the following criteria:

- Adapter type
- Chassis location
- Memory type and configuration
- Power group
- CPU cores, type, and configuration
- Storage configuration and capacity
- Server model

Depending upon the implementation, you might need to configure several policies with server pool policy qualifications including the following:

- Autoconfiguration policy
- Chassis discovery policy
- Server discovery policy

- Server inheritance policy
- Server pool policy

Creating Server Pool Policy Qualifications

Step 1 On the menu bar, choose **Physical** > **Compute**.

Step 2 In the left pane, expand the pod and then click the Cisco UCS Manager account.

Step 3 In the right pane, click the **Organizations** tab.

Step 4 Click the organization in which you want to create the policy and then click **View Details**.

Step 5 Click the **Server Pool Policy Qualifications** tab.

Step 6 Click **Add**.

Step 7 In the **Create Server Pool Policy Qualifications** wizard, enter a name and description for the policy and click **Next**.

Step 8 On the **Adapter Qualifications** page, do the following to add adapter qualifications to the policy or click **Next** if you do not want to add them:

a) Check the **Add Adapter Qualifications** check box.

b) From the **Type** drop-down list, choose the type of adapter that you want to include in the policy.
After you save the adapter qualification, this type cannot be changed.

c) In the **Model(RegEx)** field, enter a regular expression that the adapter PID must match.

d) In the **Maximum Capacity** field, enter the maximum capacity for the selected type.

e) Click **Next**.

Step 9 On the **Chassis/Server Qualifications** page, do the following to add chassis and/or server qualifications to the policy or click **Next** if you do not want to add them:

a) Check the **Add Chassis/Server Qualifications** check box.

b) From the **First Chassis ID** field, enter the first chassis ID from which server pools associated with this policy can draw.
After you save the adapter qualification, this type cannot be changed.

c) In the **Number of Chassis** field, enter the total number of chassis to include in the pool, starting with the chassis identified in the **First Chassis ID** field.

d) In the **Server Qualification Ranges** field, enter the range of server locations that you want to use.
If you want to enter more than one range, separate the ranges by commas. For example, enter 1:5,2:6.

e) Click **Next**.

Example:
For example, if you want to use chassis 5, 6, 7, and 8, enter 5 in the **First Chassis ID** field and 4 in the **Number of Chassis** field. If you want to use only chassis 3, enter 3 in the **First Chassis ID** field and 1 in the **Number of Chassis** field.

Step 10 On the **Memory Qualifications** page, do the following to add memory qualifications to the policy or click **Next** if you do not want to add them:

a) Check the **Add Memory Qualifications** check box.

b) Complete the following fields:

Name	Description
Clock field	The minimum clock speed required, in megahertz.
Min Cap field	The minimum memory capacity required, in megabytes.
Max Cap field	The maximum memory capacity allowed, in megabytes.
Width field	The minimum width of the data bus.
Latency field	The maximum latency allowed, in nanoseconds.
Units field	The unit of measure to associate with the value in the Width field.

c) Click **Next**.

Step 11 On the **CPU/Cores Qualifications** page, do the following to add CPU and/or cores qualifications to the policy or click **Next** if you do not want to add them:

a) Check the **Add CPU/Cores Qualifications** check box.

b) Complete the following fields:

Name	Description
Processor Architecture drop-down list	Choose the CPU architecture to which this policy applies.
Min Number of Cores field	The minimum number of CPU cores required. This integer can be between 1 and 65535.
Max Number of Cores field	The maximum number of CPU cores allowed. This integer can be between 1 and 65535.
Min Number of Threads field	The minimum number of CPU threads required. This integer can be between 1 and 65535 in the associated text field.
Max Number of Threads field	The maximum number of CPU threads allowed. This integer can be between 1 and 65535.
CPU Speed field	The minimum CPU speed required.
Model(RegEx) field	A regular expression that the processor PID must match.
CPU Stepping field	The minimum CPU version required.

c) Click **Next**.

Step 12 On the **Storage Qualifications** page, do the following to add storage qualifications to the policy or click **Next** if you do not want to add them:

a) Check the **Add Storage Qualifications** check box.

b) Complete the following fields:

Name	Description
Diskless drop-down list	Choose whether the available storage must be diskless. This can be one of the following: • **Unspecified—**Either storage type is acceptable. • **Yes—**The storage must be diskless. • **No—**The storage cannot be diskless. If you choose **Yes**, no additional fields are displayed.
Min Cap field	The minimum storage capacity across all disks in the server, in megabytes.
Max Cap field	The maximum storage capacity allowed, in megabytes.
Number of Blocks field	The minimum number of blocks required.
Per Disk Cap field	The minimum storage capacity per disk required, in gigabytes.
Block Size field	The minimum block size required, in bytes.
Units field	The number of units.

c) Click **Next**.

Step 13 On the **Power Group Qualifications** page, do the following to add power group qualifications to the policy or click **Next** if you do not want to add them:

a) Check the **Add Power Group Qualifications** check box.

b) From the **Power Group** drop-down list, choose the power group that you want to include in the policy.

c) Click **Next**.

Step 14 On the **Rack Qualifications** page, do the following to add rack-mount server qualifications to the policy or click **Next** if you do not want to add them:

a) Check the **Add Rack Qualifications** check box.

b) From the **First Slot ID** field, enter the first rack-mount server ID from which server pools associated with this policy can draw.
After you save the adapter qualification, this type cannot be changed.

c) In the **Number of Slots** field, enter the total number of rack-mount server slots to include in the pool, starting with the server slot identified in the **First Slot ID** field.

d) Click **Next**.

Step 15 On the **Server Model Qualifications** page, do the following to add rack-mount server qualifications to the policy or click **Next** if you do not want to add them:

a) Check the **Add Server Model Qualifications** check box.
b) In the **Model(RegEx)** field, enter a regular expression that the server model PID must match.
c) Click **Next**.

Step 16 Click **Submit**.

Server Pool Policy

This policy is invoked during the server discovery process. It determines what happens if server pool policy qualifications match a server to the target pool specified in the policy.

If a server qualifies for more than one pool and those pools have server pool policies, the server is added to all those pools.

Creating a Server Pool Policy

Before You Begin

This policy requires that one or more of the following resources already exist in the system:

- A minimum of one server pool.
- Server pool policy qualifications, if you choose to have servers automatically added to pools.

Step 1 On the menu bar, choose **Physical** > **Compute**.

Step 2 In the left pane, expand the pod and then click the Cisco UCS Manager account.

Step 3 In the right pane, click the **Organizations** tab.

Step 4 Click the organization in which you want to create the policy and then click **View Details**.

Step 5 Click the **Server Pool Policies** tab.

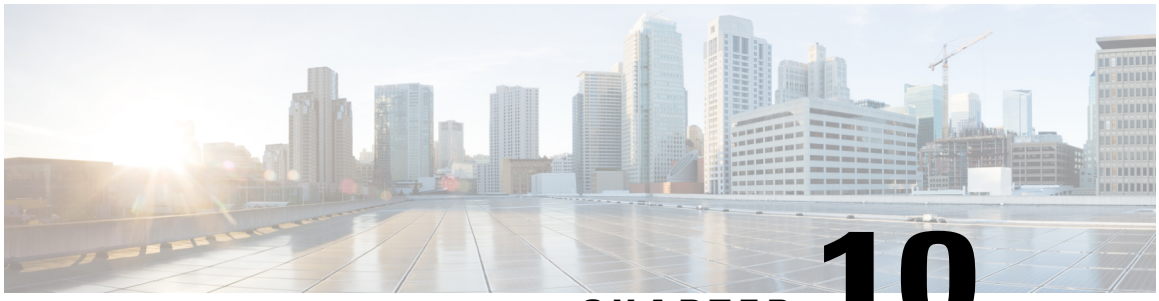
Step 6 Click **Add**.

Step 7 In the **Create Server Pool Policy** dialog box, complete the following fields:

Name	Description
Name field	A unique name for the policy.
Description field	A description for the policy.
Target Pool drop-down list	Choose a server pool to associate with the policy.

Name	Description
Qualification drop-down list	Choose a server pool qualification policy to associate with the policy.

Step 8 Click **Submit**.



CHAPTER 10

Configuring and Deploying Hadoop Cluster Deployment Templates

This chapter contains the following sections:

Creating an Instant Hadoop Cluster

Before You Begin

- Create a server pool in the Cisco UCS Manager account you plan to use for this cluster
- Create a MAC address pool

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Containers**.

Step 2 Click the **Hadoop Cluster Deploy Templates** tab.

Step 3 Click **Instant Hadoop Cluster**.

Step 4 In the **Instant Hadoop Cluster Creation** dialog box, complete the following fields.

Name	Description
Big Data Account Name field	The name of the Big Data account.
UCSM Policy Name Prefix field	The UCSM Policy Name prefix.
Hadoop Cluster Name field	A unique name for the Hadoop cluster.

Name	Description
Hadoop Node Count field	The number of nodes in the Hadoop cluster.
SSH (root) Password field	The SSH root password. **Note** The SSH username pertains to the root user.
Confirm SSH Password field	Enter the SSH root password.
Hadoop Manager Password field	The management console password.
Confirm Hadoop Manager Password field	Enter the management console password.
Host Node Prefix field	The Host Node prefix for the cluster.
OS Version drop-down list	Choose the operating system to be installed on the servers in this cluster.
Big Data Cluster Type drop-down list	Choose the type of Hadoop distribution to be used for this cluster.
Hadoop Distribution drop-down list	Choose the Hadoop distribution to be used for this cluster.
Hadoop Distribution Version drop-down list	Choose the Hadoop distribution version.
Multi UCSM check box	Check the Multi UCSM check box if you use multiple UCSM accounts. The following workflows are created during an Instant/Customized Hadoop Cluster creation: • UCS CPA Multi-UCSM Hadoop Cluster WF • Single UCSM Server Configuration WF (This WF will be triggered per UCSM Account. For example, UCSM 120, UCSM121) • UCS CPA Node BareMetal (This WF will be triggered per Node)
UCS Manager Account drop-down list	Choose the Cisco UCS Manager account for this cluster.
Organization drop-down list	Choose the organization in which the servers for this cluster are located.

Name	Description
Server UUID Pool drop-down list	Choose the UUID pool to be used for servers in this cluster. **Note** A UUID pool is a collection of SMBIOS UUIDs that are available to be assigned to servers. The first number of digits that constitute the prefix of the UUID are fixed. The remaining digits, the UUID suffix, are variable. A UUID pool ensures that these variable values are unique for each server associated with a service profile which uses that particular pool to avoid conflicts. If you use UUID pools in service profiles, you do not have to manually configure the UUID of the server associated with the service profile.
UUID Prefix drop-down list	Choose the UUID prefix.
UUID Pool-From field	To specify manually, enter the UUID pool to start from.
UUID Pool-Size field	To specify manually, enter the UUID pool size.
PXE VLAN ID field	Enter the PXE VLAN ID.
Server Pool table	The server pool that you want to use for this cluster. The Cisco UCS Manager account and the organization that you choose determine which server pools are displayed in this area.

Step 5 In the **vNIC Template** table, review and, if required, edit the vNIC templates available for the cluster.

Step 6 If you want to edit a vNIC template, select the row for that template and click **Edit**.

Step 7 In the **Edit vNIC Template Entry** dialog box, complete the following fields and click **Submit**.

Name	Description
vNIC Name drop-down list	This field is for your information only.
IP Pool drop-down list	Choose the big data IP pool that you want to use for IP addresses assigned to this vNIC.
MAC Address Pool drop-down list	Choose the MAC address pool that you want to use for this cluster.
First MAC Address field	To specify manually, enter the MAC address.
Size field	To specify manually, enter the size .
VLAN ID field	The VLAN ID for this cluster.

Step 8 Click **Submit**.

What to Do Next

You can view and monitor the workflow that is triggered after you create an instant Hadoop cluster.

Creating a Hadoop Cluster Deployment Template

Before You Begin

- Create a Cisco UCS Service Profile Template for Big Data
- Create a Hadoop Cluster Profile Template

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Containers**.

Step 2 Click the **Hadoop Cluster Deploy Templates** tab.

Step 3 Click **Add (+)**.

Step 4 In the **Add Hadoop Cluster Deploy Template** dialog box, complete the following fields:

Name	Description
Template Name field	A unique name for the Hadoop cluster deployment template.
Description field	A short description of the template.
UCS SP Template for Big Data drop-down list	Choose the UCS service profile template for Big Data that you want to use to create service profiles for the servers in the Hadoop cluster.
Hadoop Cluster Profile Template drop-down list	Choose the Hadoop cluster profile template that you want to use to configure the cluster services.

Step 5 Click **Add**.

Creating a Customized Hadoop Cluster

Before You Begin

- Create a Hadoop cluster deployment template that Cisco UCS Director Express for Big Data will use to create the Hadoop cluster
- Create a server pool in the Cisco UCS Manager account you plan to use for this cluster
- Create a MAC address pool

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Containers**.

Step 2 Click the **Hadoop Cluster Deploy Templates** tab.

Step 3 Select the template that you want to use for the Hadoop cluster and click **Customized Hadoop Cluster**.

Step 4 In the **Customized Hadoop Cluster Creation** dialog box, complete the following fields.

Name	Description
Big Data Account Name field	The name of the Big Data account.
UCSM Policy Name Prefix field	The UCSM Policy Name prefix.
Hadoop Cluster Name field	A unique name for the Hadoop cluster.
SSH (root) Password field	The SSH root password. **Note** The SSH username pertains to the root user.
Confirm SSH Password field	Enter the SSH root password.
Hadoop Manager Password field	The management console password.
Confirm Hadoop Manager Password field	Enter the management console password.
Host Node Prefix field	The Host Node prefix for the cluster.
OS Version drop-down list	Choose the operating system to be installed on the servers in this cluster.
Hadoop Distribution drop-down list	Choose the Hadoop distribution to be used for this cluster.
Hadoop Distribution Version drop-down list	Choose the Hadoop distribution version.

Name	Description
Multi UCSM check box	Check the Multi UCSM check box if you use multiple UCSM accounts. The following workflows are created during an Instant/Customized Hadoop cluster creation: • UCS CPA Multi-UCSM Hadoop Cluster WF • Single UCSM Server Configuration WF (This WF will be triggered per UCSM Account. For example, UCSM 120, UCSM121) • UCS CPA Node BareMetal (This WF will be triggered per Node)
UCS Manager Account drop-down list	Choose the Cisco UCS Manager account for this cluster.
Organization drop-down list	Choose the organization in which the servers for this cluster are located.
Server UUID Pool drop-down list	Choose the UUID pool to be used for servers in this cluster. **Note** A UUID pool is a collection of SMBIOS UUIDs that are available to be assigned to servers. The first number of digits that constitute the prefix of the UUID are fixed. The remaining digits, the UUID suffix, are variable. A UUID pool ensures that these variable values are unique for each server associated with a service profile which uses that particular pool to avoid conflicts. If you use UUID pools in service profiles, you do not have to manually configure the UUID of the server associated with the service profile.
UUID Prefix drop-down list	Choose the UUID prefix.
UUID Pool-From field	To specify manually, enter the UUID pool to start from.
UUID Pool-Size field	To specify manually, enter the UUID pool size.
PXE VLAN ID field	Enter the PXE VLAN ID.
Server Pool table	The server pool that you want to use for this cluster. The Cisco UCS Manager account and the organization that you choose determine which server pools are displayed in this area.

Step 5 In the **vNIC Template** table, review and, if desired, edit the vNIC templates available for the cluster.

Step 6 If you want to edit a vNIC template, select the row for that template and click **Edit**.

Step 7 In the **Edit vNIC Template Entry** dialog box, complete the following fields and click **Submit**.

Name	Description
vNIC Name drop-down list	This field is for your information only.
IP Pool field	Choose the big data IP pool that you want to use for IP addresses assigned to this vNIC.
MAC Address Pool drop-down list	Choose the MAC address pool that you want to use for this cluster.
First MAC Address field	To specify manually, enter the MAC address.
Size field	To specify manually, enter the size .
VLAN ID field	The VLAN ID for this cluster.

Step 8 Click **Submit**.

What to Do Next

You can view and monitor the workflow that gets triggered after creating a customized Hadoop cluster.

Provisioning an Instant and Customized Hadoop Cluster

You must create and customize a Hadoop Cluster Deploy Template to trigger the workflow.

Before You Begin

- Create a UCS Service Profile template for a Customized Hadoop Cluster
- Create a Hadoop Cluster Profile template for a Customized Hadoop Cluster

Step 1 On the menu bar, go to **Policies** > **Orchestration**.

Step 2 Click the **UCS CPA** folder from the **Workflows** tab.

Step 3 Double-click the workflow to open the workflow designer and execute the workflow.

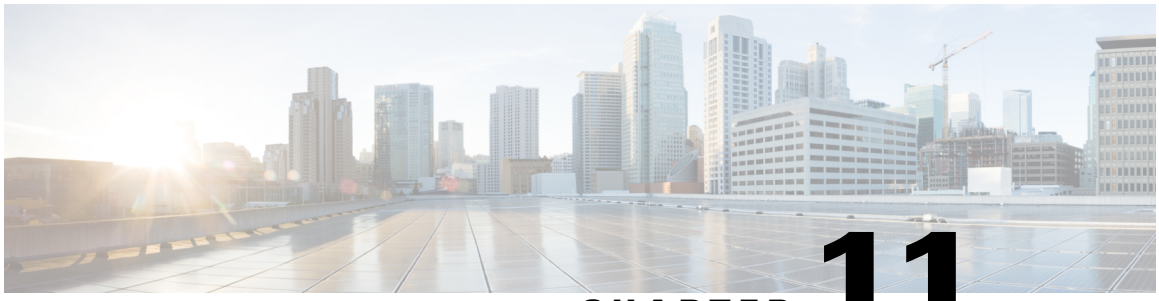
a) When you open the workflow designer for an instant Hadoop Cluster, you get the following tasks that gets processed sequentially.

Task Name	Description
Instant Hadoop Cluster UCS SP	Cisco UCS Director Express for Big Data automatically specifies parameters for installing the OS and Hadoop distribution software at the back end.
Instant Hadoop Cluster Profile	Cisco UCS Director Express for Big Data automatically configures Hadoop cluster services at the back end.
Setup Hadoop Cluster Env	Sets up the environment for cluster specific scripts and software files.
Muti BareMetal OS Install WF	Attaches the UCS profile and sets up all boot files required to boot the operating system (Linux). When the Power ON task is executed, the boot files are picked up, and the operating system is installed successfully.
Multi BareMetal WF Monitor	Checks the status of baremetal OS install workflow.
Wait for Admin Host	Waits for the Admin host to restart.
Custom SSH Command	Installs and configures the Hadoop distribution software.
Provision Hadoop Cluster	Sends the Hadoop cluster properties to the Web Console.
Completed	The Hadoop cluster is provisioned successfully. **Note** If any of the tasks fail, you will be informed that the provisioning has failed. To know more on how to monitor the workflow, see Monitoring Service Requests for Big Data , on page 107.

b) When you open the workflow designer for a customized Hadoop Cluster, you get the following tasks that gets processed sequentially.

Task Name	Description
Create UCS Service Profile Template	Specifies parameters for installing the OS and Hadoop distribution software.
Create Hadoop Cluster Profile	Configures Hadoop cluster services.
Setup Hadoop Cluster Env	Sets up the environment for cluster specific scripts and software files.
Muti BareMetal OS Install WF	Attaches the UCS profile and sets up all boot files required to boot the operating system (Linux). When the Power ON task is executed, the boot files are picked up, and the operating system is installed successfully.

Task Name	Description
Multi BareMetal WF Monitor	Checks the status of baremetal OS install workflow.
Wait for Admin Host	Waits for the Admin host to restart.
Custom SSH Command	Installs and configures the Hadoop distribution software.
Provision Hadoop Cluster	Sends the Hadoop cluster properties to the Web Console.
Completed	The Hadoop cluster is provisioned successfully. **Note** If any of the tasks fail, you will be informed that the provisioning has failed. To know more on how to monitor the workflow, see Monitoring Service Requests for Big Data , on page 107.



CHAPTER 11

Managing a Hadoop Cluster

This chapter contains the following sections:

Managing a Hadoop Cluster

You can manage an existing cluster.

Step 1 On the menu bar, choose **Solutions** > **Big Sata** > **Accounts**.

Step 2 Click on the **Big Data Accounts** tab.

Step 3 Select an account and click on **View Details**.

Step 4 Click on each tab to view the details.

Name	Description
Summary	Displays statistics data report for the selected Big Data Account and the high level report on the cluster and node account.
Hosts	See Managing Nodes in a Cluster to know the functionality of this tab.

Name	Description
Hadoop Clusters	

Name	Description
	Displays the list of actions that you can perform on an existing Hadoop cluster. • (This tab is only for information) Click **Role Topology** to view the topology of the nodes. • (This tab is only for information) Click **View Details** to view the inputs for the Hadoop cluster you have created and the virtual network interface configuration. • Click **Shutdown Cluster** that stops all the services and power off all the nodes in the Hadoop cluster. • Click **Start Cluster** that powers on all the nodes in the Hadoop cluster and starts all the services in the Hadoop cluster. • Click **Restart Cluster** to power off and then power on all the nodes in the Hadoop cluster. • Click **Rebalance** to configure the threshold percentage to re-balance Hadoop clusters. For MapR cluster, configure the threshold percentage using CLDB Balancer **Disk Paused** and CLDB Balancer **Disk Max Switches in Nodes**. • Upgrade Hadoop distributions from the current version, if available. **Note** For a derived account, pssh and clush should be installed and configured for all the nodes in the Hadoop cluster and password less between nodes. **1** Click **Upgrade Cluster**. **2** Choose the Hadoop distribution that you want to upgrade from the current version from the **Available Version** drop-down list. **3** Check the **Enable HA** check box to enable the Hadoop cluster for high availability. **4** Click **Submit**. • Enable Hadoop cluster high availability. **1** Click **Enable High Availability**. **2** From the **Enable High Availability** dialog box, choose the Standby Name Node from the **Standby Name Node** drop-down list. **3** Check a minimum of three nodes from the **Journal Nodes** table. **4** Click **Submit**. • Disable Hadoop cluster high availability. **1** Click **Diaable High Availabuilty**. **2** Form the **Disable High Availability** dialog box, choose the Name Node from the **Select Name Node** drop-down list. **3** Check the **Node Name** check box from the **Select Secondary Name Node** table. Name Node and Secondary Name Node should not be mapped to a

Name	Description
	Note single node. 4 Click **Submit**. • Click **Cluster Snapshot** to view the snapshot. • Click **View Reports** to view performance and monitoring reports.
Hadoop Services	Displays the list of Hadoop services and their status. You can do the following: 1 **Start All Services** - Start all Hadoop services depending on their current status. 2 **Stop All Services** - Stop all Hadoop services depending on their current status. 3 **Add New Service** - Add a new Hadoop service. 4 **Edit Service** - Start and stop a particular Hadoop service.
Hadoop Service Roles	Displays the list of Hadoop services. To add a role to the cluster, do the following in the **Add Role** dialog box: 1 From the **Hadoop Service Name** drop-down list, choose the Hadoop service. 2 From the **Role Type** drop-down list, choose the role type. 3 From the **Node Name** drop-down list, choose the node name. 4 From the **Role Name** drop-down list, choose the role name. 5 Click **Submit**. To start or stop any role that you have created, do the following in the **Start/Stop Role** dialog box: 1 From the **Hadoop Service Roles** tab, choose the Hadoop service. 2 Click **Start/Stop Role**. 3 Click **Submit**. To delete a role in the cluster, do the following in the **Delete Role** dialog box: 1 From the **Hadoop Service Roles** tab, choose the Hadoop service. 2 **Click Delete.** 3 click **Submit**.
More Reports	Additional reports that you can generate for data usage and CPU utilization.

View Hadoop Cluster Details

For each Big Data Account, you can view the details of all the Hadoop clusters that are associated with the Big Data account from the **Hadoop Clusters** tab. See Managing a Hadoop Cluster

Name	Description
Big Data Account name	The name of the Big Data account.
UCS SP Template for Big Data	The UCS SP Template for Big Data that you used to create service profiles for the servers in the Hadoop cluster.
Hadoop Cluster Profile Template	The Hadoop cluster profile template that you used to configure the cluster services.
Hadoop Cluster Deploy Template	A unique name that you used for the Hadoop cluster deployment template.
UCSM Policy Name Prefix	The UCSM Policy Name prefix.
Hadoop Cluster Name	A unique name that you used for the Hadoop cluster.
Hadoop Node Count	The number of nodes in the Hadoop cluster.
Hadoop Node Prefix	The Host Node prefix for the cluster.
OS Version	The operating system that you installed on the servers for the Hadoop cluster.
Hadoop Distribution	Hadoop distribution that you used for this cluster.
Hadoop Distribution Version	Hadoop distribution version that used for this cluster.
PXE VLAN ID	The VLAN ID used for PXE boot of the servers.
UCS Service Profile Template	The UCS Service Profile Template that you specified parameters for installing the OS and Hadoop distribution software.
vNIC: eth0	Displays IPv4 network information for management interface and the management VLAN ID.
vNIC: eth1	Displays IPv4 network information for DATA1 interface and the VLAN ID.
vNIC: eth2	Displays IPv4 network information for DATA2 interface and the VLAN ID.

Viewing a Cluster Snapshot

A cluster snapshots displays configuration details of a Hadoop cluster, such as hosts, roles, and services. To view the current snapshot of a cluster, do the following:

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Accounts**.

Step 2 Click the **Big Data Accounts** tab.

Step 3 Choose the Big Data Account for which you want to view the snapshot and click **View Details**.

Step 4 Click the **Hadoop Clusters** tab.

Step 5 Choose the Hadoop cluster for which you want to view the snapshot and click **Cluster Snapshot**.

Step 6 Click **Submit**.
You can view the snapshot for the selected Hadoop cluster.

Adding a New Hadoop Service

Step 1 On the menu bar, choose **Solutions** > **Big Sata** > **Accounts**.

Step 2 Choose a Cloudera account for which you can add a new Hadoop service.
Add new service is not supported for MapR and Hortonworks distributions.

Step 3 Click **View Details**.

Step 4 Click the **Hadoop Services** tab.

Step 5 Click **Add New Service**.
In the **Add New Service** dialog box, complete the following fields:

Name	Description
Workflow Inputs	
Big Data Account Name	Choose the Big Data Account (Cloudera).
Service Type	Enter the Hadoop service for Cloudera.
Role Assignment Pairs	Enter Role Assignment Pairs separated by a comma. For example, RoleType1:hostname1, RoleType2:hostname2
Dependant Services	Enter the list of dependant services separated by a comma
Pre Install Commands	Enter the list of commands separated by a new line ("\n")
Post Install Commands	Enter the list of commands separated by a new line ("\n")

Step 6 Click **Submit**.

Managing Nodes in a Cluster

You can add, delete, re-commission and decommission nodes in a cluster.

- Managed Node—Any node that was already a member (managed) in the cluster and deleted, which can be added again in the cluster.
- Live Node—Any node that has the operating system installed and reachable from the Hadoop cluster.
- Baremetal Node—Any node that is available and not associated with the Hadoop cluster.

Step 1 On the menu bar, choose **Solutions** > **Big Sata** > **Accounts**.

Step 2 Click on the **Big Data Accounts** tab.

Step 3 Select an account and click **View Details**.

Step 4 Click the **Hosts** tab to perform the following actions:

Name	Description
Refresh	Refreshes the page.
Favorite	Adds the page to Favorites.
Add Managed Node	Add managed nodes to the Hadoop cluster.
Add Live Node	Add live nodes to the Hadoop cluster.
Add BareMetal Nodes	Add baremetal nodes to the Hadoop cluster.

Step 5 Select a host that allows you to perform the following actions:

Name	Description
View Details	Displays the summary of the CPU usage, the I/O status of the hosts disks and so on. **Note** If you see a **License Status** tab, it indicates a licensing issue.
Delete Node	Deletes node from the cluster.
Assign Rack	Assigns the node to the rack server.

Name	Description
Recommission Node/ Decommission Node	Decommissioning or Re-commissioning a node depends on its current status. **Note** When the node is in decommissioned status, it means that all the roles for that node have been withdrawn.
Delete Node to BareMetal	The node is removed from the cluster and disassociated from the service profile. The node becomes a BareMetal server.
Host Mappings	Lists DNS entries of all the hosts in the Hadoop cluster.

Delete Node and Delete Node to BareMetal Actions in Cloudera and Hortonworks

When you perform the Delete Node or Delete Node to BareMetal actions, the UCS CPA Delete Node (a new workflow) is created for Cloudera and Hortonworks. You can also execute this workflow for MapR to perform the delete node operation. If you execute the UCS CPA Delete Node workflow for Cloudera and Hortonworks, this workflow also provides the functionality of Delete Node to BareMetal action based on the Delete Node to BareMetal flag setting to true or false. In addition, the Rollback UCS CPA Node BareMetal workflow is created.

Adding Managed Nodes to the Hadoop Cluster

Add managed nodes to the Hadoop cluster.

This feature allows you to add nodes that are available only from the below URLs, but not the members of the cluster.

- Cloudera— http://serverip:7180/api/v6/hosts where the serverIP is the IPv4 address of the administration node.

- Hortonworks—http://serverIP:8080/api/v1/hosts where the serverIP is the IPv4 address of the administration node

Step 1 On the menu bar, choose **Solutions** > **Big Sata** > **Accounts**.
Step 2 Click on the **Big Data Accounts** tab.
Step 3 Double-click on the Big Data account.
Step 4 Click the **Hosts** tab.
Step 5 Click **Add Managed Node**.
Step 6 From the **Host Name** drop-down list, choose the host name.
Step 7 Click **Submit**.

Adding Live Nodes to the Hadoop Cluster

Add live nodes to the Hadoop cluster.

Step 1 On the menu bar, choose **Solutions** > **Big Sata** > **Accounts**.
Step 2 Click on the **Big Data Accounts** tab.
Step 3 Double-click on the Big Data account.
Step 4 Click the **Hosts** tab.
Step 5 Click **Add Live Node**.
Step 6 Enter the IPv4 address in the **Host Management IPV4 Address** field.
Step 7 Enter the name of the rack server in the **Rack Name** field.
Step 8 Enter the password in the **(New Node) Password** field for that rack server.
Step 9 Choose the Cluster Manager Version for the Hadoop distribution from the **Cluster Management Version** drop-down list.
Step 10 Choose the operating system to be installed on the servers in this cluster from the **OS Version** drop-down list.
Step 11 Click **Submit**.

Adding BareMetal Nodes to the Hadoop Cluster

Add baremetal nodes to the Hadoop cluster.

Step 1 On the menu bar, choose **Solutions** > **Big Sata** > **Accounts**.

Step 2 Click on the **Big Data Accounts** tab.

Step 3 Double-click on the Big Data account.

Step 4 Click the **Hosts** tab.

Step 5 Click **Add BareMetal Nodes**.

Step 6 In the **Add Baremetal Nodes** dialog box, complete the following fields:

Name	Description
Big Data Account Name field	The name of the Big Data account.
UCSM Policy Name Prefix field	The UCSM Policy Name prefix.
Hadoop Cluster Name field	A unique name for the Hadoop cluster.
Hadoop Node Count field	The number of nodes in the Hadoop cluster.
Host Node Prefix field	The Host Node prefix for the cluster.
OS Version drop-down list	Choose the operating system to be installed on the servers in this cluster.
Hadoop Distribution drop-down list	Choose the Hadoop distribution to be used for this cluster.
Hadoop Distribution Version drop-down list	Choose the Hadoop distribution version.
Multi UCSM check box	Check the Multi UCSM check box if you use multiple UCSM accounts. The following workflows are created during an Instant/Customized Hadoop Cluster creation: • UCS CPA Multi-UCSM Hadoop Cluster WF. • Single UCSM Server Configuration WF (This WF will be triggered per UCSM Account. For example, UCSM 120, UCSM121). • UCS CPA Node BareMetal (This WF will be triggered per Node).
UCS Manager Account drop-down list	Choose the Cisco UCS Manager account for this cluster.

Name	Description
Organization drop-down list	Choose the organization in which the servers for this cluster are located.
Server UUID Pool drop-down list	Choose the UUID pool to be used for servers in this cluster. **Note** A UUID pool is a collection of SMBIOS UUIDs that are available to be assigned to servers. The first number of digits that constitute the prefix of the UUID are fixed. The remaining digits, the UUID suffix, are variable. A UUID pool ensures that these variable values are unique for each server associated with a service profile which uses that particular pool to avoid conflicts. If you use UUID pools in service profiles, you do not have to manually configure the UUID of the server associated with the service profile.
UUID Prefix drop-down list	Choose the UUID prefix.
UUID Pool-From field	To specify manually, enter the UUID pool to start from.
UUID Pool-Size field	To specify manually, enter the UUID pool size.
PXE VLAN ID field	Enter the PXE VLAN ID.
UCSTemplate Name table	Check the UCS Service Profile Template check box that you want to use and click **Submit**.
Server Pool table	The server pool that you want to use for this cluster. The Cisco UCS Manager account and the organization that you choose determine which server pools are displayed in this area.

Step 7 In the **vNIC Template** table, review and, if desired, edit the vNIC templates available for the cluster.

Step 8 If you want to edit a vNIC template, select the row for that template and click **Edit**.

Step 9 In the **Edit vNIC Template Entry** dialog box, complete the following fields and click **Submit**.

Name	Description
vNIC Name drop-down list	This field is for your information only.
IP Pool drop-down list	Choose the big data IP pool that you want to use for IP addresses assigned to this vNIC.
MAC Address Pool drop-down list	Choose the MAC address pool that you want to use for this cluster.

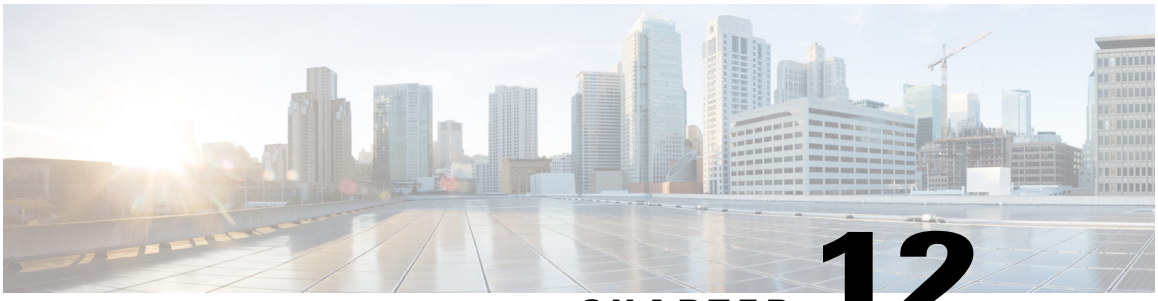
Name	Description
First MAC Address field	To specify manually, enter the MAC address.
Size field	To specify manually, enter the size .
VLAN ID field	The VLAN ID for this cluster.

Step 10 Click **Submit**.

Service Roles

For each Hadoop distribution, specific roles are added for all nodes that are newly added by using Add Node BareMetal, Add Managed Node, and Add Live Node options to the Hadoop cluster.

Service Roles	Cloudera	MapR	Hortonworks
FileServer	No	Yes	No
DataNode	Yes	No	Yes
NodeManager	Yes	Yes	Yes
Ganglia Monitor	No	No	Yes



CHAPTER 12

Hadoop Cluster Configuration Settings

This chapter contains the following sections:

Creating a Hadoop Cluster Configuration Parameters Template

You can modify the Hadoop Cluster Configuration Parameters Template only from the **Hadoop Cluster Configuration Parameters Template** tab on the menu bar from **Solutions** > **Big Data** > **Settings** before triggering a Hadoop cluster.

Note Select the Hadoop cluster configuration parameters template to edit, clone, or delete.

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Settings**.

Step 2 Click the **Hadoop Config Parameters** tab.

Step 3 Click **Add**.

Step 4 In the **Hadoop Config Parameters** page of the **Create Hadoop Cluster Configuration Parameters Template** wizard, complete the following fields:

Name	Description
Template Name field	A unique name for the Hadoop cluster configuration parameter template.
Template Description field	The description for the Hadoop cluster configuration parameter template.
Hadoop Distribution drop down list	Choose the Hadoop distribution.
Hadoop Distribution Version drop down list	Choose the Hadoop distribution version.

Step 5 Click **Next**.

Step 6 In the **Hadoop Config Parameters - HDFS Service** page of the **Create Hadoop Cluster Configuration Parameters Template** wizard, configure the following fields:

Name	Description
HDFS Service	Hadoop cluster HDFS service parameter name, value, and the minimum supported Hadoop distribution.
Data Node (role)	Displays the parameter names, values, and the minimum supported Hadoop distribution version.
Name Node (role)	Displays the parameter names, values, and the minimum supported Hadoop distribution version.
Secondary Node (role)	Displays the parameter names, values, and the minimum supported Hadoop distribution version.

Step 7 In the **Hadoop Config Parameters - YARN Service** page of the **Create Hadoop Cluster Configuration Parameters Template** wizard, configure the parameters.

Step 8 In the **Hadoop Config Parameters - HBase Service** page of the **Create Hadoop Cluster Configuration Parameters Template** wizard, configure the parameters.

Step 9 In the **Hadoop Config Parameters - MapReduce Service** page of the **Create Hadoop Cluster Configuration Parameters Template** wizard, configure the parameters.

Step 10 In the **Hadoop Config Parameters - Miscellaneous Parameters** page of the **Create Hadoop Cluster Configuration Parameters Template** wizard, configure the (ServiceLevel and RoleLevel) parameters.

Step 11 Click **Submit**.

Updating Hadoop Cluster Configuration Parameters Template - Post Hadoop Cluster Creation

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Accounts**.

Step 2 Click the **Big Data Accounts** tab and choose an existing Big Data Account.

Step 3 Click **Configure Cluster**.

Step 4 In the **Hadoop Config Parameters** page of the **Update Hadoop Cluster Configuration Parameters Template** wizard, complete the following fields:

Name	Description
Hadoop Distribution drop down list	Choose the Hadoop distribution.
Hadoop Distribution Version	Displays the selected Hadoop distribution version.

Step 5 Click **Next**.

Step 6 In the **Hadoop Config Parameters - HDFS Service** page of the **Update Hadoop Cluster Configuration Parameters Template** wizard, update the following fields:

Name	Description
HDFS Service	Hadoop cluster HDFS service parameter name, value, and the minimum supported Hadoop distribution version, if any.
Data Node (role)	Displays the parameter names, values, and the minimum supported Hadoop distribution version that can be edited.
Name Node (role)	Displays the parameter names, values, and the minimum supported Hadoop distribution version that can be edited.

Name	Description
Secondary Node (role)	Displays the parameter names, values, and the minimum supported Hadoop distribution version that can be edited.

Step 7 In the **Hadoop Config Parameters - YARN Service** page of the **Update Hadoop Cluster Configuration Parameters Template** wizard, update the parameters.

Step 8 In the **Hadoop Config Parameters - HBase Service** page of the **Update Hadoop Cluster Configuration Parameters Template** wizard, update the parameters.

Step 9 In the **Hadoop Config Parameters - MapReduce Service** page of the **Update Hadoop Cluster Configuration Parameters Template** wizard, update the parameters.

Step 10 In the **Hadoop Config Parameters - Miscellaneous Parameters** page of the **Update Hadoop Cluster Configuration Parameters Template** wizard, update the (ServiceLevel and RoleLevel) parameters.

Step 11 Click **Submit**.

Quality of Service System Class

For more information on Quality of Service and System Classes, see QoS System Classes.

Quality of Service

Cisco Unified Computing System provides the following methods to implement quality of service (QoS):

- System classes that specify the global configuration for certain types of traffic across the entire system.
- QoS policies that assign system classes for individual vNICs.
- Flow control policies that determine how uplink Ethernet ports handle pause frames.

System Classes

Cisco UCS uses Data Center Ethernet (DCE) to handle all traffic inside a Cisco UCS domain. This industry standard enhancement to Ethernet divides the bandwidth of the Ethernet pipe into eight virtual lanes. Two virtual lanes are reserved for internal system and management traffic. You can configure quality of service (QoS) for the other six virtual lanes. System classes determine how the DCE bandwidth in these six virtual lanes is allocated across the entire Cisco UCS domain.

Each system class reserves a specific segment of the bandwidth for a specific type of traffic, which provides a level of traffic management, even in an oversubscribed system. For example, you can configure the Fibre Channel Priority system class to determine the percentage of DCE bandwidth allocated to FCoE traffic.

The following table describes the system classes that you can configure

System Class	Description
Best Effort	A system class that sets the quality of service for the lane reserved for basic Ethernet traffic. Some properties of this system class are preset and cannot be modified. For example, this class has a drop policy that allows it to drop data packets if required. You cannot disable this system class.
• Platinum • Gold • Silver • Bronze	A configurable set of system classes that you can include in the QoS policy for a service profile. Each system class manages one lane of traffic. All properties of these system classes are available for you to assign custom settings and policies.
Fibre Channel	A system class that sets the quality of service for the lane reserved for Fibre Channel over Ethernet traffic. Some properties of this system class are preset and cannot be modified. For example, this class has a no-drop policy that ensures it never drops data packets. You cannot disable this system class. **Note** FCoE traffic has a reserved QoS system class that should not be used by any other type of traffic. If any other type of traffic has a CoS value that is used by FCoE, the value is remarked to 0

Editing QoS System Class

For more information on Quality of Service and System Classes, see QoS System Classes.

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Settings**.

Step 2 Click the **QoS System Class** tab.

Step 3 Choose the QoS System Class (by Priority) that you want to edit and click **Edit**.

- Best Effort
- Platinum
- Gold
- Silver
- Bronze

Step 4 In the **Modify QoS System Class** dialog box, complete the following fields:

Name	Description
Enabled check box	If checked, the associated QoS class is configured on the fabric interconnect and can be assigned to a QoS policy. If unchecked, the class is not configured on the fabric interconnect and any QoS policies associated with this class default to Best Effort or, if a system class is configured with a CoS of 0, to the Cos 0 system class. This field is always checked for Best Effort and Fibre Channel.
CoS drop-down list	The class of service. You can enter an integer value between 0 and 6, with 0 being the lowest priority and 6 being the highest priority. We recommend that you do not set the value to 0, unless you want that system class to be the default system class for traffic if the QoS policy is deleted or the assigned system class is disabled. This field is set to 7 for internal traffic and to any for Best Effort. Both of these values are reserved and cannot be assigned to any other priority.
Packet Drop check box	This field is always unchecked for the Fibre Channel class, which never allows dropped packets, and always checked for Best Effort, which always allows dropped packets. If checked, packet drop is allowed for this class. If unchecked, packets cannot be dropped during transmission.
Weight drop-down list	This can be one of the following: • An integer between 1 and 10. If you enter an integer, Cisco UCS determines the percentage of network bandwidth assigned to the priority level as described in the Weight (%) field. • best-effort. • none.
Muticast Optimized check box	If checked, the class is optimized to send packets to multiple destinations simultaneously. This option is not applicable to the Fibre Channel.

Name	Description
MTU drop-down list	The maximum transmission unit for the channel. This can be one of the following: • An integer between 1500 and 9216. This value corresponds to the maximum packet size. • fc—A predefined packet size of 2240. • normal—A predefined packet size of 1500. • Specify Manually—A packet size between 1500 to 9216. This field is always set to fc for Fibre Channel.

Step 5 Click **Submit**.

Pre Cluster Performance Testing Settings

You can analyze memory, network, and disk metrics and a default Big Data Metrics Report provides the statistics collected for each host before creating any Hadoop cluster.

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Settings**.

Step 2 Click the **Management** tab.

Step 3 In the **Pre Cluster Performance Tests** section, check the check boxes for the following:

- Memory Test
- Network Test
- Disk Test

Note By default, the check boxes to run the memory, network, and the disk tests are unchecked. If you enable the pre-cluster disk test that may impact significantly Hadoop cluster creation.

Step 4 Click **Submit.**

Approving Hadoop Cluster Deployment Workflows

Before You Begin

Choose **Administration** > **Users and Groups** > **Users** and add users with the following user roles:

- Network Admin (system default user role)
- Computing Admin (system default user role)
- Hadoop User

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Settings**.

Step 2 Click the **Management** tab.

Step 3 Check the **Require OS User Approval** check box.

a) From the **User ID** table, select the **Login Name** of the user against the Network Admin user role.

b) Enter the **Number of Approval Request Reminders**.

Note Set the number of approval reminders to Zero if the reminder e-mail has to be sent at a specified interval till the Network Admin approves or rejects the approval request.

c) Enter the **Reminder Interval(s)** in hours.

Note Check the **Approval required from all the users** check box, if you want all the users to approve or reject the approval request.

Step 4 Check the **Require Compute User Approval** check box.

a) From the **User ID** table, select the **Login Name** of the user against the Computing Admin user role.

b) Enter the **Number of Approval Request Reminders**.

Note Set the number of approval reminders to Zero if the reminder e-mail has to be sent at a specified interval till the Computing Admin approves or rejects the approval request.

c) Enter the **Reminder Interval(s)** in hours.

Note Check the **Approval required from all the users** check box, if you want all the users to approve or reject the approval request.

Step 5 Check the **Require Hadoop User Approval** check box.

a) From the **User ID** table, select the **Login Name** of the user against the Hadoop User user role.

b) Enter the **Number of Approval Request Reminders**.

Note Set the number of approval reminders to Zero if the reminder e-mail has to be sent at a specified interval till the Hadoop User approves or rejects the approval request.

c) Enter the **Reminder Interval(s)** in hours.

Note Check the **Approval required from all the users** check box, if you want all the users to approve or reject the approval request.

Step 6 Click **Submit**.

What to Do Next

Check if users of Network Admin, Computing Admin, and Hadoop User roles have approved the request before deploying any Hadoop cluster.

Uploading Required OS and Hadoop Software to Cisco UCS Director Baremetal Agent

You can upload (add) required RHEL 6.x ISO files, Hadoop software and common software that are required for Hadoop distributions to Cisco UCS Director Baremetal Agent. While uploading the required files from your local or any remote system, the files are first uploaded to Cisco UCS Director, and then moved to the target Cisco UCS Director Baremetal Agent once you click the **Submit** button in the **Create Software Catalogs** dialog box.

Supported file formats:

- Linux OS - rhel-x.x.iso
- Hadoop software - MapR-x.y.z.zip (.gz or .tgz or .tar)
- Common software - bd-sw-rep.zip (.gz or .tgz or .tar)

The Software Catalogs page displays Hadoop distributions and the required software for those Hadoop distributions in the Cisco UCS Director Baremetal Agent.

Note If the required software column is empty for a Hadoop distribution, then it means that Cisco UCS Director Baremetal Agent contains all the files required.

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Settings**.

Step 2 Click the **Software Catalogs** tab.

Step 3 Click **Add**.

Step 4 Click **Upload** to upload files from your local system.

Note You must create a folder for the Hadoop distribution to include all the required files and compress the folder before uploading in the format specified.

Step 5 Choose the target Cisco UCS Director Baremetal Agent from the **Target BMA** drop-down list.

Step 6 Check the **Restart BMA Services** to restart Cisco UCS Director Baremetal Agent after uploading the required files.

Note Refresh the **Software Catalogs** page after 5 to 10 minutes to see new and modified catalogs.

Linux OS Upload	
Catalog Name field	Operating System Name (for example, RHEL.6.5)

Upload Type drop-down list	Choose one of the following: • Desktop file • The web server path that is reachable by the Cisco UCS Director Baremetal Agent • Mountpoint in Cisco UCS Director Baremetal Agent (For example, /root/iso) • Path to ISO in Cisco UCS Director Baremetal Agent(For example, /temp/rhel65/iso)
Hadoop Software Upload	
Catalog Name field	Hadoop Distribution (for example, distribution_name-x.y.z)
Upload Type drop-down list	Choose one of the following: • Desktop file • The web server path that is reachable by the Cisco UCS Director Baremetal Agent to upload remote software to Baremetal Agent
Common Software Upload	
Upload Type drop-down list	Choose one of the following: • Desktop file • The web server path that is reachable by the Cisco UCS Director Baremetal Agent to upload remote software to Baremetal Agent

Step 7 Click **Submit**.

What to Do Next

You can track software uploads here: **Administration** > **Integration**. Click the **Change Record** tab to track the software upload in progress and verify if completed or failed or timeout.

Cloudera, MapR, and Hortonworks RPMs on Cisco UCS Director Express for Big Data Baremetal Agent

Common Packages for Cloudera, MapR, and Hortonworks

Note For any Hadoop software that is not available, you have to update the `/opt/cnsaroot/bigdata_templates/common_templates/HadoopDistributionRPM.txt` with the available online repository of the vendor.

Note We recommend to verify the supported versions from the Hadoop Vendor Support Documentation.

Download the following common packages to `/opt/cnsaroot/bd-sw-rep/`:

- pssh-2.3.1.tar.gz from https://pypi.python.org/packages/source/p/pssh
- clustershell-1.6-1.el6.noarch.rpm from http://dl.fedoraproject.org/pub/epel/6/x86_64

Common Packages for Cloudera

Download the following packages to `/opt/cnsaroot/bd-sw-rep/cloudera-5.X.X`:

- ClouderaEnterpriseLicense.lic— Get the license keys from Cloudera
- userrpmlist.txt—For additional packages list
- catalog.properties—Provides the label name for the Cloudera version (x represents the Cloudera version on the Cisco UCS Director Express for Big Data Baremetal Agent)
- mysql-connector-java-5.1.26.tar.gz from http://cdn.mysql.com/archives/mysql-connector-java-5.1

Cloudera 5.0.1 Packages and Parcels

Download the following packages to `/opt/cnsaroot/bd-sw-rep/cloudera-5.0.1`:

- cm5.0.1-centos6.tar.gz from http://archive.cloudera.com/cm5/repo-as-tarball/5.0.1
- CDH-5.0.1-1.cdh5.0.1.p0.47-el6.parcel from http://archive.cloudera.com/cdh5/parcels/5.0.1
- CDH-5.0.1-1.cdh5.0.1.p0.47-el6.parcel.sha1 from http://archive.cloudera.com/cdh5/parcels/5.0.1
- manifest.json from http://archive.cloudera.com/cdh5/parcels/5.0.1

Cloudera 5.0.6 Packages and Parcels

Download the following packages to `/opt/cnsaroot/bd-sw-rep/cloudera-5.0.6`:

- cm5.0.6-centos6.tar.gz from http://archive.cloudera.com/cm5/repo-as-tarball/5.0.6
- CDH-5.0.6-1.cdh5.0.6.p0.7-el6.parcel from http://archive.cloudera.com/cdh5/parcels/5.0.6

- CDH-5.0.6-1.cdh5.0.6.p0.7-el6.parcel.sha1 from http://archive.cloudera.com/cdh5/parcels/5.0.6
- manifest.json from http://archive.cloudera.com/cdh5/parcels/5.0.6

Cloudera 5.2.0 Packages and Parcels

Download the following packages to `/opt/cnsaroot/bd-sw-rep/cloudera-5.2.0`:

- cm5.2.0-centos6.tar.gz from http://archive.cloudera.com/cm5/repo-as-tarball/5.2.0
- CDH-5.2.0-1.cdh5.2.0.p0.36-el6.parcel from http://archive.cloudera.com/cdh5/parcels/5.2.0
- CDH-5.2.0-1.cdh5.2.0.p0.36-el6.parcel.sha1 from http://archive.cloudera.com/cdh5/parcels/5.2.0
- manifest.json from http://archive.cloudera.com/cdh5/parcels/5.2.0

Cloudera 5.2.1 Packages and Parcels

Download the following packages to `/opt/cnsaroot/bd-sw-rep/cloudera-5.2.1`:

- cm5.2.1-centos6.tar.gz from http://archive.cloudera.com/cm5/repo-as-tarball/5.2.1
- CDH-5.2.1-1.cdh5.2.1.p0.12-el6.parcel from http://archive.cloudera.com/cdh5/parcels/5.2.1
- CDH-5.2.1-1.cdh5.2.1.p0.12-el6.parcel.sha1 from http://archive.cloudera.com/cdh5/parcels/5.2.1
- manifest.json from http://archive.cloudera.com/cdh5/parcels/5.2.1

Cloudera 5.3.0 Packages and Parcels

Download the following packages to `/opt/cnsaroot/bd-sw-rep/cloudera-5.3.0`:

- cm5.3.0-centos6.tar.gz from http://archive.cloudera.com/cm5/repo-as-tarball/5.3.0
- CDH-5.3.0-1.cdh5.3.0.p0.30-el6.parcel from http://archive.cloudera.com/cdh5/parcels/5.3.0
- CDH-5.3.0-1.cdh5.3.0.p0.30-el6.parcel.sha1 from http://archive.cloudera.com/cdh5/parcels/5.3.0
- manifest.json from http://archive.cloudera.com/cdh5/parcels/5.3.0

Cloudera 5.4.1 Packages and Parcels

Download the following packages to `/opt/cnsaroot/bd-sw-rep/cloudera-5.4.1`:

- cm5.4.1-centos6.tar.gz from http://archive.cloudera.com/cm5/repo-as-tarball/5.4.1
- CDH-5.4.1-1.cdh5.4.1.p0.6-el6.parcel from http://archive.cloudera.com/cdh5/parcels/5.4.1
- CDH-5.4.1-1.cdh5.4.1.p0.6-el6.parcel.sha1 from http://archive.cloudera.com/cdh5/parcels/5.4.1
- manifest.json from http://archive.cloudera.com/cdh5/parcels/5.4.1

Common Packages for MapR 3.1.1, 4.0.1, and 4.0.2

Download the following common packages to `/opt/cnsaroot/bd-sw-rep/MapR-3.1.1 and MapR-4.0.x directories`:

- libgenders-devel-1.14-2.el6.rf.x86_64.rpm from http://pkgs.repoforge.org/libgenders/

- libgenders-1.14-2.el6.rf.x86_64.rpm from http://pkgs.repoforge.org/libgenders/
- ext-2.2.zip from `http://dev.sencha.com/deploy/ext-2.2.zip`
- sshpass-1.05-1.el6.x86_64.rpm from http://ftp.pbone.net/mirror/download.fedora.redhat.com/pub/fedora/epel/6/x86_64
- soci-mysql-3.2.1-1.el6.x86_64.rpm from http://ftp.is.co.za/mirror/fedora.redhat.com/epel/6/x86_64
- soci-3.2.1-1.el6.x86_64.rpm from http://ftp.is.co.za/mirror/fedora.redhat.com/epel/6/x86_64
- pdsh-2.27-1.el6.rf.x86_64.rpm from http://pkgs.repoforge.org/pdsh
- mapr-whirr-0.7.0.16780-1.noarch.rpm from http://archive.mapr.com/releases/ecosystem-all/redhat
- mapr-drill-0.7.0.29434-1.noarch.rpm from http://archive.mapr.com/releases/ecosystem/redhat
- catalog.properties—Provides the label name for the MapR version (x represents the MapR version on the Cisco UCS Director Express for Big Data Baremetal Agent)
- license.txt

Common Packages for MapR 4.1.0 and 5.0.0

Download the following common packages to `/opt/cnsaroot/bd-sw-rep/MapR-4.1.0 and MapR-5.0.0 directories`:

- libgenders-devel-1.14-2.el6.rf.x86_64.rpm from http://pkgs.repoforge.org/libgenders/
- libgenders-1.14-2.el6.rf.x86_64.rpm from http://pkgs.repoforge.org/libgenders/
- ext-2.2.zip from `http://dev.sencha.com/deploy/ext-2.2.zip`
- sshpass-1.05-1.el6.x86_64.rpm from http://ftp.pbone.net/mirror/download.fedora.redhat.com/pub/fedora/epel/6/x86_64
- soci-mysql-3.2.2-2.el6.x86_64.rpm from http://ftp.univie.ac.at/systems/linux/fedora/epel/6/x86_64
- soci-3.2.2-2.el6.x86_64.rpm from http://ftp.univie.ac.at/systems/linux/fedora/epel/6/x86_64
- pdsh-2.27-1.el6.rf.x86_64.rpm from http://pkgs.repoforge.org/pdsh
- mapr-whirr-0.8.1.18380-GA.noarch.rpm from http://archive.mapr.com/releases/ecosystem-all/redhat
- catalog.properties—Provides the label name for the MapR version (x represents the MapR version on the Cisco UCS Director Express for Big Data Baremetal Agent)
- license.txt

MapR 3.1.1 Packages

Download the following packages to /opt/cnsaroot/bd-sw-rep/MapR-3.1.1

- mapr-v3.1.1GA.rpm.tgz from http://package.mapr.com/releases/v3.1.1/redhat
- mapr-ecosystem-20140617.rpm.tgz

MapR 4.0.1 Packages

Download the following packages to /opt/cnsaroot/bd-sw-rep/MapR-4.0.1

- mapr-v4.0.1GA.rpm.tgz from http://package.mapr.com/releases/v4.0.1/redhat
- mapr-ecosystem-4.x-20141105.rpm.tgz

MapR 4.0.2 Packages

Download the following packages to /opt/cnsaroot/bd-sw-rep/MapR-4.0.2

- mapr-v4.0.2GA.rpm.tgz from http://package.mapr.com/releases/v4.0.2/redhat
- mapr-ecosystem-20150205.rpm.tgz

MapR 4.1.0 Packages

Download the following packages to /opt/cnsaroot/bd-sw-rep/MapR-4.1.0

- mapr-v4.1.0GA.rpm.tgz from http://package.mapr.com/releases/v4.1.0/redhat
- mapr-ecosystem-4.x-20150610.rpm.tgz from http://archive.mapr.com/releases/ecosystem-all/redhat

MapR 5.0.0 Packages

Download the following packages to /opt/cnsaroot/bd-sw-rep/MapR-5.0.0

- mapr-v5.0.0GA.rpm.tgz: from http://package.mapr.com/releases/v5.0.0/redhat
- mapr-ecosystem-5.x-20150709.rpm.tgz from http://archive.mapr.com/releases/ecosystem-all/redhat

Common Package for Hortonworks

Download the following common package to `/opt/cnsaroot/bd-sw-rep/Hortonworks-2.X`:

- openssl-1.0.1e-30.el6.x86_64.rpm
- catalog.properties—Provides the label name for the Hortonworks version (x represents the Hortonworks version on the Cisco UCS Director Express for Big Data Baremetal Agent)

Hortonworks 2.1 Packages

Download the following packages to `/opt/cnsaroot/bd-sw-rep/Hortonworks-2.1`:

- HDP-2.1.5.0-centos6-rpm.tar.gz from http://public-repo-1.hortonworks.com/HDP/centos6
- ambari-1.6.1-centos6.tar.gz from http://public-repo-1.hortonworks.com/ambari/centos6
- HDP-UTILS-1.1.0.17-centos6.tar.gz from http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.17/repos/centos6

Hortonworks 2.2 Packages

Download the following packages to `/opt/cnsaroot/bd-sw-rep/Hortonworks-2.2`:

- HDP-2.2.0.0-centos6-rpm.tar.gz from http://public-repo-1.hortonworks.com/HDP/centos6
- ambari-1.7.0-centos6.tar.gz from http://public-repo-1.hortonworks.com/ambari/centos6

- HDP-UTILS-1.1.0.20-centos6.tar.gz from http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6

Hortonworks 2.3 Packages

Download the following packages to `/opt/cnsaroot/bd-sw-rep/Hortonworks-2.3`:

- HDP-2.3.0.0-centos6-rpm.tar.gz from http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.3.0.0
- ambari-2.1.1-centos6.tar.gz from http://public-repo-1.hortonworks.com/ambari/centos6/2.x/updates/2.1.1
- HDP-UTILS-1.1.0.20-centos6.tar.gz from http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.20/repos/centos6

Cloudera and MapR RPMs for Upgrading Hadoop Cluster Distributions

Cloudera 5.3.0 Packages and Parcels

- cm5.3.0-centos6.tar.gz from http://archive.cloudera.com/cm5/repo-as-tarball/5.3.0
- CDH-5.3.0-1.cdh5.3.0.p0.30-el6.parcel from http://archive.cloudera.com/cdh5/parcels/5.3.0
- CDH-5.3.0-1.cdh5.3.0.p0.30-el6.parcel.sha1 from http://archive.cloudera.com/cdh5/parcels/5.3.0
- manifest.json from http://archive.cloudera.com/cdh5/parcels/5.3.0

Cloudera 5.4.1 Packages and Parcels

- cm5.4.1-centos6.tar.gz from http://archive.cloudera.com/cm5/repo-as-tarball/5.4.1
- CDH-5.4.1-1.cdh5.4.1.p0.6-el6.parcel from http://archive.cloudera.com/cdh5/parcels/5.4.1
- CDH-5.4.1-1.cdh5.4.1.p0.6-el6.parcel.sha1 from http://archive.cloudera.com/cdh5/parcels/5.4.1
- manifest.json from http://archive.cloudera.com/cdh5/parcels/5.4.1

MapR 4.1.0 Packages

- mapr-setup from http://package.mapr.com/releases/v4.1.0/redhat
- mapr-v4.1.0GA.rpm.tgz from http://package.mapr.com/releases/v4.1.0/redhat
- mysql-connector-java-5.1.26.tar.gz from http://cdn.mysql.com/archives/mysql-connector-java-5.1

MapR 5.0.0 Packages

- mapr-setup from http://package.mapr.com/releases/v5.0.0/redhat
- mapr-v5.0.0GA.rpm.tgz: from http://package.mapr.com/releases/v5.0.0/redhat
- mysql-connector-java-5.1.26.tar.gz from http://cdn.mysql.com/archives/mysql-connector-java-5.1

Configuration Check Rules

You can validate an existing cluster configuration by running a configuration check. The configuration check process involves comparing the current cluster configuration with configuration check rules and reporting violations.

Configuration check rules are pre-defined Cisco Validated Design (CVD) parameters for Hadoop clusters. Configuration check rules appear under **Solutions > Big Data > Settings**. After the configuration check is complete, the violations appear in the **Faults** tab under **Solutions > Big Data > Accounts**. You can enable or disable configuration check rules at any time, but you cannot add new rules.

Configuration Check Rule	Description
Parameter	The pre-defined CVD parameter of the configuration.
Enabled	The state of the configuration check rule, either enabled (true) or disabled (false).
Expected value	The value expected for a parameter as defined in the Cisco Validated Design (CVD).
Description	The description of the parameter of the configuration.
Distribution	The Hadoop distribution.
Minimum Supported Distribution	The minimum supported version of Hadoop distribution.
Service	The Hadoop service.
Role	The Hadoop service role.
Type	The type of violation, either CVD or Inconsistent.
Fix Workflow	The reference to the workflow that can be triggered for fixing violations.

When the actual cluster configuration values differ from the expected values defined in the configuration check rules, then those configuration values are reported as violations. For example, CVD mandates that the NameNode heap size should be 4 GB. But if the NameNode heap size in the cluster configuration is found to be 1 GB, then this is reported as a CVD violation. Additionally, inconsistent configuration parameters are reported. For example, NameNode heap size on both the primary and secondary nodes must be of the same size. If there is a mismatch in the size, then this parameter is reported as inconsistent.

Checking Hadoop Cluster Configuration

To validate the configuration of a cluster, do the following:

Step 1 On the menu bar, choose **Solutioms** > **Big Data** > **Accounts**.

Step 2 Click the **Big Data Accounts** tab.

Step 3 Choose the account for which you want to run the configuration check and click **Check Configuration**.

Step 4 Click **Submit**.
A dialog box appears with the information that the configuration check is in progress.

Step 5 Click **OK**.
After the configuration check is complete, the violations appear under the **Faults** tab for the selected Big Data Account.

What to Do Next

Note You can track configuration checks here: **Administration** > **Integration**. Click the **Change Record** tab to track the configuration checks in progress and verify if completed or failed.

Fixing Configuration Violations

After the configuration check is complete, the configuration violations appear in the Faults tab for the selected BIg Data Account. You can either choose to fix these configuration violations manually on the Big Data Cluster Configuration page or trigger a workflow. To trigger a workflow to fix the violation, first you must create a workflow with the same name as the code specified in the violation.

To fix a configuration violation through a workflow, do the following:

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Accounts**.

Step 2 Click the **Faults** tab.

Step 3 Choose the configuration violation you want to fix and click **Trigger Workflow**.
If a workflow exists with the same name as the code specified in the violation, then the workflow is triggered.

Step 4 Enter the required inputs for the workflow and click **Submit**.
A service request ID is generated after you submit the inputs. You can check the status of the service request on the **Service Requests** page.

CHAPTER 13

Cisco UCS CPA Workflows

This chapter contains the following sections:

Workflows for Big Data

Cisco UCS Director Express for Big Data defines a set of workflows in the **UCS CPA** folder under **Policies** > **Orchestration** for Big Data.

- UCS CPA Multi-UCSM Hadoop Cluster WF—This workflow is triggered if you use multiple UCSM accounts.
- UCS CPA Single UCSM Server Configuration WF—This workflow is triggered per UCSM account.
- UCS CPA Node Baremetal—This workflow is triggered per node in the cluster.
- UCS CPA Delete Node—This workflow is triggered if you delete a node from the Hadoop cluster.
- UCS CPA Cloudera Add Live Node—This workflow is triggered if you add a Live Node for Cloudera.
- UCS CPA MapR Add Live Node—This workflow is triggered if you add a Live Node for MapR.
- UCS CPA Hartonworks Add Live Node—This workflow is triggered if you add a Live Node for Hortonworks.
- UCS CPA Instant Hadoop Cluster WF—This workflow is triggered if you create an instant Hadoop cluster based on the node count and other mandatory inputs such as the IP address range, memory, and number of interfaces. Cisco UCS Director Express for Big Data automatically creates one UCS service profile and a Hadoop cluster profile template at the backend that are required to create an instant Hadoop cluster. This saves the time to manually create a service profile and a cluster profile.
- UCS CPA Customized Hadoop Cluster WF—This workflow is triggered if you choose to use a specific UCS service profile and a Hadoop cluster profile template with the specified number of nodes to be created in the Hadoop cluster.
- UCS CPA Disable High Availability WF—This workflow is triggered when you disable high availability.
- UCS CPA Enable High Availability WF—This workflow is triggered when you enable high availability.

- UCS CPA Shutdown Big Data Cluster WF—This workflow is triggered when you shutdown the Hadoop cluster.
- UCS CPA Start Big Data Cluster WF—This workflow is triggered when you start the Hadoop cluster.
- UCS CPA Upgrade Cluster WF—This workflow is triggered when you upgrade the Hadoop cluster.
- UCS CPA Add New Cloudera Service WF—This workflow is triggered when you add a new service for Cloudera.
- UCS CPA Hortonworks Add New Service WF—This workflow is triggered when you add a new service for Hortonworks.
- UCS CPA MapR Add New Service WF–This workflow is triggered when you add a new service for MapR.

About Service Requests for Big Data

Cisco UCS Director Express for Big Data leverages Cisco UCS Director service requests and workflow orchestration for the overall deployment of Hadoop clusters. Each service request is a provisioning workflow that is created during a cluster creation. For example, one UCS CPA Multi-UCSM Hadoop Cluster W/F, one Single UCSM Server Configuration W/F, and four UCS CPA Node Baremetal are created for a four node Hadoop cluster. When the workflows are complete, the cluster is created under **Solutions** > **Big Data Accounts** for that UCSM account.

A set of service requests is created under **Organizations** > **Service Requests** during a cluster creation.

- UCS CPA Multi-UCSM Hadoop Cluster WF—This workflow is triggered if you use multiple UCSM accounts. It also applies to when you create an instant or customized Hadoop cluster.
- UCS CPA Single UCSM Server Configuration WF—This workflow is triggered per UCSM account.
- UCS CPA Node Baremetal—This workflow is triggered per node in the cluster.

The following service requests are created when you add a BareMetal Node or a Live Node to the cluster.

- UCS CPA Cloudera Add Live Node—This workflow is triggered if you add a Live Node for Cloudera.
- UCS CPA MapR Add Live Node—This workflow is triggered if you add a Live Node for MapR.
- UCS CPA Hortonworks Add Live Node—This workflow is triggered if you add a Live Node for Hortonworks.

The following service requests are created when you do the following:

- UCS CPA Disable High Availability WF—This workflow is triggered when you disable high availability.
- UCS CPA Enable High Availability WF—This workflow is triggered when you enable high availability.
- UCS CPA Shutdown Big Data Cluster WF—This workflow is triggered when you shutdown the Hadoop cluster.
- UCS CPA Start Big Data Cluster WF—This workflow is triggered when you start the Hadoop cluster.
- UCS CPA Upgrade Cluster WF—This workflow is triggered when you upgrade the Hadoop cluster.

- UCS CPA Cloudera Add New Service WF—This workflow is triggered when you add a new service for Cloudera.
- UCS CPA Hortonworks Add New Service WF—This workflow is triggered when you add a new service for Hortonworks.
- UCS CPA MapR Add New Service WF–This workflow is triggered when you add a new service for MapR.

For more information on the service requests and workflows, see the following guides:

- *Cisco UCS Director Self-Service Portal Guide*
- *Cisco UCS Director Orchestration Guide*

.

Monitoring Service Requests for Big Data

Before You Begin

You must create and customize a Hadoop Cluster Deploy Template to trigger the following workflow.

Step 1 On the menu bar, choose **Organizations** > **Service Requests** .

Step 2 Click the **Service Requests** tab.

Step 3 Select the service request that you would like to monitor and click **View Details**.
The Request Status is one of the following:

- Complete
- In Progress
- Cancelled
- Failed

Viewing UCS CPA Workflow Tasks

From the **Service Request Status** dialog box, you can view the following:

- Workflow Status
- Log
- Objects Created and Modified
- Input/Output

Note You can only modify inputs for failed service requests.

Step 1 On the menu bar, choose **Organizations** > **Service Requests**.
You can see the list of service requests created for the user added to a specific group. For example, All User Groups.

Step 2 Choose the **Service Request ID** you would like to explore the tasks associated with it.

Step 3 Double-click the **Service Request ID** that opens the **Service Request Status** dialog box. (You can also choose the Service Request ID by the workflow name associated with it and click **View Details**. For example, choose the **UCS CPA Node BareMetal** workflow and click **View Details**).
In the **Service Request Status** dialog box, you can view the following tasks for the workflow:

UCS CPA Multi-UCSM Hadoop Cluster WF	Single UCSM Server Configuration WF	UCS CPA Node BareMetal
The following tasks are associated with the **UCS CPA Multi-UCSM Hadoop Cluster WF**: **1** Initiated by Admin **2** Multi-UCSM Hadoop Cluster profile **3** Setup Hadoop Cluster Env **4** Multi UCSM Configuration WF **5** Multi BareMetal WF Monitor **6** Synchronized Command Execution **7** Custom SSH Command **8** Provision Hadoop Cluster **9** Completed	The following tasks are associated with the **Single UCSM Server Configuration WF**: **1** Initiated by Admin **2** Create UCS Service Profile Template **3** Muti BareMetal OS Install WF **4** Multi BareMetal WF Monitor **5** Completed	The following tasks are associated with the **UCS CPA Node BareMetal**: **1** Initiated by Admin **2** Modify Workflow Priority (High) **3** Assign BareMetal SR ID **4** Create UCS Service Profile from Template **5** Service Profile unbind/rebind Action **6** Modify UCS Service Profile Boot Policy **7** Associate UCS Service Profile **8** Assign Server Identity **9** Bind/Unbind vNIC Template **10** Setup PXE Boot (OS Type : CentOS Live) **11** Setup RAID Commands **12** UCS Blade Reset Action **13** Monitor PXE Boot **14** Monitor RAID Configuration **15** UCS Blade Power OFF Action **16** Setup PXE Boot (OS Type: RHEL 6.5). **17** Setup RAID Commands **18** Wait for complete Association **19** UCS Blade Reset Action **20** Monitor PXE Boot **21** Modify UCS Service Profile Boot Policy **22** Service Profile unbind/rebind Action **23** Wait for complete Association **24** UCS Blade Reset Action **25** Assign IP Status **26** Custom SSH Command **27** Custom SSH Command **28** Synchronized Command Execution **29** UCS Blade Power OFF Action **30** UCS Blade Power ON Action **31** Synchronized Command Execution **32** Completed

Step 4 Click **Close**.

Workflow Customization to Deploy a Hadoop Cluster

You can customize the following UCS CPA workflows and use them to deploy an Hadoop cluster. You can add installation packages (RPMs) required for your cluster environment in the Cisco UCS Director Express for Big Data Baremetal Agent post any Hadoop distribution.

- Rename the UCS CPA Multi-UCSM Hadoop Cluster WF
- Rename the Single UCSM Server Configuration WF
- Rename the UCS CPA Node BareMetal

Deploying Hadoop Cluster Through Workflow Customization

Before You Begin

For more information on workflow orchestration, see the *Cisco UCS Director Orchestration Guide*.

- Customize UCS CPA Node BareMetal workflows that you want to use in the cloned Single UCSM Server Configuration WF.
- Customize the Single UCSM Server Configuration WF that you want to use in the UCS CPA Multi-UCSM Hadoop Cluster WF.

Step 1 On the menu bar, choose **Policies** > **Orchestration**.

Step 2 Click the **UCS CPA** folder from the **Workflows** tab.

Step 3 Double-click the workflow that you want to customize in the **Workflow Designer**. For instance, double-click the **UCS CPA Multi-UCSM Hadoop Cluster WF**.

Step 4 Double-click the **Muti UCSM Configuration WF** task in the **Workflow Designer**.

Step 5 Click **Next** on the **Workflow Task Basic Information** page.

Step 6 On the **User Input Mappings to Task Input Attributes** page, you select the attributes in this page that you want to use values from the workflow input fields or provide values in the next step. If required, check the **Map to User Input**

check box to provide user inputs. If the **Map to User Input** check box is checked, inputs are prompted during workflow execution unless specified by the administrator in the workflow definition.

Step 7 Click **Next** on the **User Input Mappings to Task Input Attributes** page.

Step 8 Enter the task values for the task inputs which are not mapped to workflow inputs. For example, enter the name of the cloned Single UCSM Server Configuration WF in the **Workflow Name** field.

Step 9 Click **Revalidate** to validate task input values.

Step 10 Click **Next**.

Step 11 On the **User Output mappings to Task Output Attributes** page, you select the attributes in this page that you want to use values from the workflow output fields.

Step 12 Check the **Map to User Output** check box and choose the value from the **User Output** drop-down list.

Step 13 Click **Submit**.

Cloning UCS CPA Workflows

To customize cluster deployment through baremetal workflows, you can clone the following workflows in the UCS CPA folder:

- Clone the UCS CPA Node BareMetal workflows.
- Rename the Single UCSM Server Configuration WF using the cloned UCS CPA Node BareMetal workflows.
- Rename the UCS CPA Multi-UCSM Hadoop Cluster WF using the cloned Single UCSM Server Configuration WF.

Step 1 On the menu bar, choose **Policies** > **Orchestration** .

Step 2 Choose the workflow that you want to clone in the **UCS CPA** folder from the **Workflows** tab.

Step 3 Click **Clone Workflow**.

Step 4 On the **Workflow Details** page of the **Clone Workflow** wizard, complete the following fields:

Name	Description
Workflow Name field	A unique name for the workflow.
Version field	Displays the current version of the workflow that you are cloning.
Description filed	The description of the workflow.

Name	Description
Workflow Context drop-down list	The workflow context. Workflow Orchestration supports the following options: • **Any—**Enables you to use the workflow in any context. • **Selected VM—**Enables you to use the execute workflow only when you choose a VM. • Check the **Save As Compound Task** check box to define the workflow as a compound task. • Check the **Place in New Folder** check box, and enter the folder name in the **Folder Name** field to assign the workflow to a new folder other than the UCS CPA folder.
Select Folder drop down list	UCS CPA is the default folder for Big Data.
Notify status of execution to initiator User check box	Check the check box to notify the user through email. If checked, enter the additional email addresses in the **Additional User(s) to send Email Notification** field.

Step 5 Click **Next**.

Step 6 On the **Workflow User Inputs** page of the **Clone Workflow** wizard, complete the following fields:

Name	Description
Associate to Activity check box	Check the check box. If selected, existing workflow's user input(s) will be overridden by selected activities user input(s).
Activity drop-down list.	Choose an activity. The user input table is updated based on the selected activity.

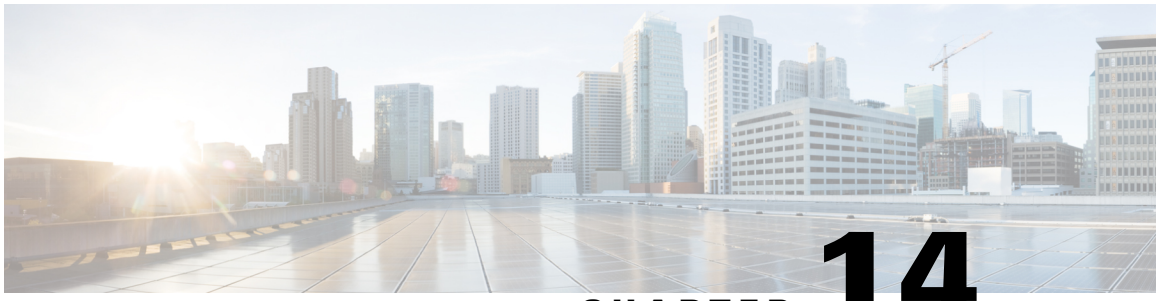
Name	Description
Workflow User Inputs table	On the **Workflow User Inputs** page: 1 Click the + icon to add workflow input properties. 2 In the **Add Entry to** dialog box, complete the following fields: 1 Enter the name for the activity workflow input in the **Input Label** field. 2 Enter the description for the activity workflow input in the **Input Description** field. 3 Check the **Optional** check box to set the input as optional during workflow execution. 4 Click **Select**. In the **Select** dialog box, click **Input Type**. 5 Click **Submit**.

Step 7 Click **Next**.

Step 8 On the **Workflow User Outputs** page of the **Clone Workflow** wizard, do the following:

Name	Description
Workflow User Outputs table	On the **Workflow User Outputs** page: 1 Click the + icon to add workflow output properties. 2 In the **Add Entry to** dialog box, complete the following fields: 1 Enter the name for the activity workflow output in the **Output Label** field. 2 Enter the description for the activity workflow output in the **Output Description** field. 3 Check the **Optional** check box to set the output as optional during workflow execution. 4 Click **Select**. In the **Select** dialog box, click **Output Type**. 5 Click **Submit**.

Step 9 Click **Submit**.



CHAPTER 14

Monitoring and Reporting

This chapter contains the following sections:

About Monitoring and Reporting

Cisco UCS Director Express for Big Data can monitor the virtual infrastructure and system resources, and provide a wide array of reports.

Cisco UCS Director Express for Big Data monitors a wide range of cluster events:

- High CPU usage
- Memory usage
- Disk capacity
- Disk IO utilization

Cisco UCS Director Express for Big Data displays statistics from the respective pages for the selected Big Data Account and Hosts. You can also generate reports and these reports help you to understand system details and provide an insight into how the system is performing.

Cisco UCS Director Express for Big Data Dashboard

The Cisco UCS Director Express for Big Data provides complete system visibility through real-time and historical monitoring. See Reports

Note The **Dashboard** tab shows up in the menu bar only after a summary report is added to the dashboard.

The customizable dashboard displays key processing, memory, storage, and network utilization metrics.

- Per node statistics: CPU, memory, and disk
- Health of Hadoop cluster components: HDFS, MapReduce jobs
- Graphs based on historical data

You can:

Note Drag and Drop summary report icons from the **Customize** drawer to the **Summary** and **Hosts** tabs where you can expand and close reports.

1. Add summary reports to the dashboard from the Big Data Account **Summary** tab and customize these summary reports to display the statistics for a specific duration in the dashboard, or export these reports from the dashboard.
2. Add summary reports to the dashboard from the Big Data Account **Hosts** tab and customize these summary reports to display the statistics for a specific duration in the dashboard, or export these reports from the dashboard.

Note Not all summary reports apply to MapR.

3. Add UCSM Accounts summary reports to the dashboard from **Physical** > **Compute**.
4. Add Data Center summary reports to the dashboard from **Physical** > **Compute** .

Viewing Deployed Cluster Report

You can generate the **Big Data Account Summary Report** with or without credentials to view the details of the deployed clusters.

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Containers**.
Step 2 Click the **Deployed Clusters** tab.
Step 3 Choose the deployed cluster, and click **View Report**.
Step 4 In the **View Report** dialog box, choose the report type that you want to generate.
Step 5 Click **Submit**.

Reports

UCSM Account Summary Reports

- Overview
- UCS Chassis Inventory
- UCS Server Inventory
- UCS Fabric Interconnect Inventory
- UCS Servers Associated vs Unassociated
- Rack Server Discovery Policy
- Power Policy
- Global Power Allocation Policy

Big Data Account Summary Reports

- Overview

Cluster Specific Metrics Supported per Hadoop Distribution

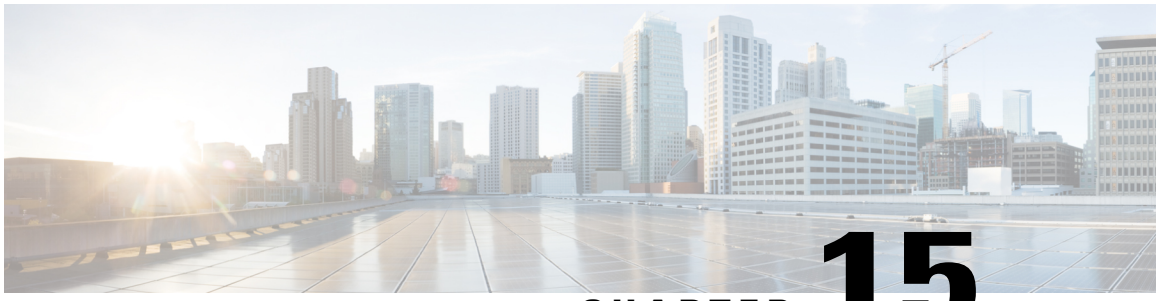
Metrics	Cloudera Distribution	Hortonworks Distribution	MapR Distribution	Remarks
Average CPU IO Wait (%)	Yes	Yes	Yes	
Average CPU idle (%)	Yes	Yes	Yes	
Average CPU Nice (%)	Yes	Yes	Yes	
Average CPU System (%)	Yes	Yes	Yes	
Average CPU User	Yes	Yes	Yes	
CPU Percentage Across Hosts	Yes			Metrics for Host CPU usage across hosts
Cluster Disk IO	Yes			Metrics for Total Disk Write/Read Bytes Across Disks

Metrics	Cloudera Distribution	Hortonworks Distribution	MapR Distribution	Remarks
Cluster Network IO	Yes			Metrics for Total Bytes Transmitted/Received Across Network Interfaces
HDFS IO	Yes			Metrics for Total Bytes written/Read Across Data Nodes
Total Space Utilization		Yes		
CPU		Yes		
Load		Yes		
Memory		Yes		
Network		Yes		
Process		Yes		

Host Specific Metrics Supported per Hadoop Distribution

Metrics	Cloudera Distribution	Hortonworks Distribution	MapR Distribution
Average CPU IO Wait (%)	Yes	Yes	Yes
Average CPU idle (%)	Yes	Yes	Yes
Average CPU Nice (%)	Yes	Yes	Yes
Average CPU System (%)	Yes	Yes	Yes
Average CPU User	Yes	Yes	Yes
Load Average(%)	Yes		
Host CPU Usage	Yes		
Host Memory Usage	Yes		

Metrics	Cloudera Distribution	Hortonworks Distribution	MapR Distribution
Host Network Throughput	Yes		
Disk Latency	Yes		
Aggregate Disk Throughput	Yes		
Role-CPU Usage	Yes		
HOST CPU		Yes	
DISK Usage		Yes	
LOAD Usage		Yes	
Memory-CPU Usage		Yes	
Network Usage		Yes	
Process Usage		Yes	



CHAPTER 15

Proactive Status Monitoring and Diagnostics

This chapter contains the following topics:

Aggregate CPU, Disk, and Network Bandwidth Utilization

You can monitor the aggregate CPU, disk, and network bandwidth utilization across all the hosts in a cluster. The metrics are collected in the following way:

- **Aggregate CPU and Disk metrics:** For every host that is running the job, the PID collects the percentage of CPU and memory used by the job. The sum of all these percentages gives the aggregate CPU and disk metrics.
- **Aggregate network bandwidth metrics:** The network bandwidth on each network interface in a host is obtained, which is then added up to obtain aggregate for one node. Similarly network bandwidths are measured for all the nodes in the cluster. The sum of all these bandwidths provides the aggregate network bandwidth metrics for the cluster.
- **Duration of long-running jobs:** A Rest API collects the start time, lapsed time, and end time for each job identified on the cluster. The difference between start time and end time provides the duration of completed jobs. The lapsed time provides the duration of the jobs running currently.

Monitoring Aggregate CPU, Disk, and Network Bandwidth Utilization

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Accounts**.

Step 2 Click the **Big Data Accounts** tab.

Step 3 Choose the Big Data Account and click **View Details**.

Step 4 Click the **Hadoop Clusters** tab.

Step 5 Choose the Hadoop cluster and click **View Reports**.

Step 6 Click the **Monitoring** tab.
Every time an inventory collection cycle is triggered, an entry containing the aggregate CPU, network bandwidth, and disk utilization metrics appears on the Monitoring Page.

Step 7 Select the entry you want to analyze and click **View Details**.

a) Click the **Aggregate CPU** tab to view the aggregate CPU utilization of all nodes for a particular time.
b) Click the **Aggregate Disks** tab to view the aggregate disk utilization and available memory across the cluster.
c) Click the **Aggregate Network Bandwidth Utilization** to view the aggregated network bandwidth across the cluster.

Step 8 Click **Back** to return back to the **Monitoring** page.

Monitoring Top Jobs Based on CPU Utilization and Time

To monitor top jobs based on CPU utilization or time (both active and completed long-running jobs), do the following:

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Accounts**.

Step 2 Click the **Big Data Accounts** tab.

Step 3 Choose the Big Data Account and click **View Details**.

Step 4 Click the **Hadoop Clusters** tab.

Step 5 Choose the Hadoop cluster and click **View Reports**.

a) Click the **Top 10 High CPU Jobs** tab to view top ten jobs based on CPU utilization.
b) Click the **Top 10 Long Running Active Jobs** tab to view the current top ten long-running jobs.
c) Click the **Top 10 Long Duration Jobs** tab to view the completed top ten long-running jobs.

Step 6 Click **Back** to return back to the **Hadoop Clusters** page.

Performance Metrics for CPU, Disk, and Network

You can collect CPU, disk, and network metrics and analyze these metrics to find and fix performance bottlenecks that occur in the compute, network, or Hadoop setup across the cluster.

The metrics reports are of the following types:

- **Pre-Cluster:** This metrics report is generated automatically for a server that has just been installed with Red Hat Linux. This report is created before the server becomes part of a cluster.
- **Post-Cluster:** This metrics report is generated on demand when you run the performance test for a Hadoop cluster.

When you run the performance test for a Hadoop cluster, the following metrics are shown in detail:

- **Memory metrics:** Memory metrics measure the memory utilization of each host on the Hadoop cluster. The report includes the triad rate, which is the average rate at which read, write, and copy operations take place. The triad rate is a standard measure of memory bandwidth.
- **Network metrics:** Network metrics measure the network bandwidth of the Hadoop cluster. The report displays the rates at which network packets are transferred between the client and the server in the Hadoop cluster.
- **Disk metrics:** Disk metrics identify how fast a disk can perform. The report lists the time taken to read and write a file, to rewrite to an existing a file, and to randomly (non-sequentially) read and write files. The disk metrics are included only in the pre-cluster report.
- **DFSIO metrics:** The DFSIO test is a Hadoop benchmark that stress-tests the storage I/O (read and write) capabilities of the cluster. The report measures the bytes processed, execution time, the average I/O rate and throughput to read and write multiple files. The DFSIO metrics are included only in the post-cluster report.
- **TeraSort metrics:** The TeraSort test is a Hadoop benchmark that tests the memory of the cluster. The report lists the counters for generating input, sorting the generated input, and validating the sorted output. The TeraSort Metrics is included only in the post-cluster report.

Viewing CPU, Disk, and Network Statistics for a Hadoop Cluster

You can collect and compare CPU, disk, and network metrics with the pre-cluster creation and post-cluster creation reports for a Hadoop cluster.

Step 1 On the menu bar, choose **Solutions** > **Big Data >** > **Accounts**.
Step 2 Click the **Big Data Accounts** tab.
Step 3 Choose the Big Data Account and click **View Details**.
Step 4 Click the **Hadoop Clusters** tab.
Step 5 Choose the Hadoop cluster and click **View Reports**.
Step 6 Click the **Performance** tab.
Step 7 Click **Run Test**.

The Performance tab displays a default Big Data Metrics Report that shows the statistics collected for each host before the Hadoop cluster creation and the reports post Hadoop cluster creation.

Step 8 Click **Submit**, and then click **OK**.
For the following actions, choose the performance report:

Name	Description
View	Displays the metrics in the Big Data Metrics Report.
Compare	Compares and displays the metrics in the Big Data Metrics Report.
View Graph Report	Displays graphically the following reports from the **Summary** tab: • Average TRIAD Rate (MB/Sec) • Average Network Bandwidth (MB/Sec) • Average DFSIO Write (MB/Sec) • Average DFSIO Read (MB/Sec)
Delete	Deletes the Big Data Metrics Report.
More Reports	Displays the metrics on an hourly, daily, weekly, or monthly basis.

Analyzing Performance Bottlenecks through Historical Metrics

Comparing a historical metrics report (especially a metrics report that was collected when the cluster was performing well at peak load) with a metrics report captured during a poor performance helps identify metrics that could be the cause of the performance bottleneck in the Hadoop cluster.

To compare and analyze two metrics reports, do the following:

Step 1 On the menu bar, choose **Solutions** > **Big Data** > **Accounts**.
Step 2 Click the **Big Data Accounts** tab.
Step 3 Choose the Big Data Account and click **View Details**.
Step 4 Click the **Hadoop Clusters** tab.
Step 5 Choose the Hadoop cluster and click **View Reports**.
Step 6 Click the **Performance** tab.
Step 7 Click **Run Test**.

The Performance tab displays a default Big Data Metrics Report that shows the statistics collected for each host before the Hadoop cluster creation and the reports post Hadoop cluster creation.

Step 8 Choose two reports that you want to compare (for example, a report generated while the cluster was performing well and a report generated during poor performance) and click **Compare**.

Step 9 Click **Submit**.

Setting Alerts for Hadoop Cluster Service Failures

You can create an alert to monitor the health of the Hadoop cluster whenever Hadoop services go down. Based on the trigger conditions, you can also activate customized workflows that automatically take corrective action.

Step 1 On the menu bar, choose **Policies** > **Orchestration**.

Step 2 Click the **Triggers** tab.

Step 3 Click **Add**.
In the **Trigger Information** page of the **Add Trigger** wizard, complete the following fields:

Name	Description
Trigger Name field	Name of the trigger.
Is Enabled check box	Check this box to enable the trigger.
Description	Description of the trigger.
Frequency	Choose the trigger rule validation frequency.
Trigger Type	Choose the type of the trigger. • Stateful • Stateless

Step 4 Click **Next**.

Step 5 In the **Specify Conditions** page of the **Add Trigger** wizard, click **Add a new entry to the table below (+)**, and complete the following fields in the **Add Entry to Conditions** dialog box:

a) From the **Type of Object to Monitor** drop-down list, choose **BigData Cluster**.
b) From the **Object** drop-down list, choose the Hadoop cluster to be monitored.
c) From the **Parameter** drop-down list, choose the parameter against to validate.
d) From the **Operation** drop-down list, choose **Equals** or **Not Equals**.
e) From the **Value** drop-down list, choose **All Services Up** or **Any Service Down**.
f) Click **Submit**.

g) From the **Trigger When** drop-down list, choose to satisfy all the conditions or any condition.

Step 6 Click **Next**.

Step 7 In the **Specify Workflow** page of the **Add Trigger** wizard, do the following when the Hadoop cluster service is down and when the trigger is reset:

a) From the **Maximum Number of Invocations** drop-down list, choose the number of maximum invocations.
b) Select a workflow for execution when the trigger state becomes active and check the **Pass Monitored Object** check box, if required.
c) Select the workflow input.
d) Select a workflow for execution when the trigger state becomes clear and check the **Pass Monitored Object** check box, if required.
e) Select the workflow input.

Step 8 Click **Next**.

Step 9 In the **Specify Workflow Inputs** page of the **Add Trigger** wizard, enter the inputs for the selected workflows, and then click **Submit**.

Types of Disk and Network Failure Alerts

You can create alerts to detect faults related to disks and networks in a cluster.

The alerts that you can create for memory faults are as follows:

- **fltMemoryUnitInoperable:** This alert is triggered when the number of correctable or uncorrectable errors that have occurred on a DIMM have reached a threshold. The DIMM might be inoperable.
- **fltMemoryUnitThermalThresholdNonRecoverable:** This alert is triggered when the temperature of a memory unit on a server has been out of the operating range, and the issue is not recoverable.
- **fltMemoryArrayVoltageThresholdCritical:** This alert is triggered when the memory array voltage exceeds the specified hardware voltage rating.
- **fltMemoryArrayVoltageThresholdNonRecoverable:** This alert is triggered when the memory array voltage exceeds the specified hardware voltage rating, with potential memory hardware damage.
- **fltMemoryBufferUnitThermalThresholdCritical:** This alert is triggered when the temperature of a memory buffer unit on a blade or rack server exceeds a critical threshold value.
- **fltMemoryBufferUnitThermalThresholdNonRecoverable:** This alert is triggered when the temperature of a memory buffer unit on a blade or rack server has been out of the operating range, and the issue is not recoverable.
- **fltMemoryUnitDisabled:**This alert is triggered when the server BIOS disables a DIMM. The BIOS could disable a DIMM for several reasons, including incorrect location of the DIMM or incompatible speed.

The alerts that you can create for disk faults are as follows:

- **fltStorageItemCapacityExceeded:** This alert is triggered when the partition disk usage exceeds 70% but is less than 90%.

- **fltStorageItemCapacityWarning:** This alert is triggered when the partition disk usage exceeds 90%.
- **fltStorageLocalDiskInoperable:** This alert is triggered when the local disk has become inoperable.
- **fltStorageLocalDiskSlotEpUnusable:** This alert is triggered when the server disk drive is in a slot that is not supported by the storage controller.
- **fltStorageLocalDiskMissing:** This alert is triggered when a disk is missing.
- **fltStorageLocalDiskDegraded:** This alert is triggered when the local disk has degraded. The fault description contains the physical drive state, which indicates the reason for the degradation.

The alerts that you can create for network faults are as follows:

- **fltAdaptorUnitMissing:** This alert is triggered when the network adapter is missing or when the server cannot detect or communicate with the adapter.
- **fltAdaptorHostIfLink-down:** This alert is triggered when the fabric interconnect is in End-Host mode and all uplink ports had failed, when the server port to which the adapter is pinned had failed, or when a transient error caused the link to fail.
- **fltAdaptorExtIfLink-down:** This alert is triggered when the adapter's connectivity to any of the fabric interconnects cannot be validated, or when a node reports a vNIC down or link down event on the adapter link.

Setting Alerts for Disk and Network Failures

You can create alerts for disk or network failures in the Hadoop cluster. Alerts help you in pro-active cluster maintenance. Based on the trigger conditions, you can activate customized workflows that automatically take corrective action.

Step 1 On the menu bar, choose **Policies** > **Orchestration**.

Step 2 Click the **Triggers** tab.

Step 3 Click **Add**.

In the **Trigger Information** page of the **Add Trigger** wizard, complete the following fields:

Name	Description
Trigger Name field	Name of the trigger.
Is Enabled check box	Check this box to enable the trigger.
Description	Description of the trigger.
Frequency	Choose the trigger rule validation frequency.
Trigger Type	Choose the type of the trigger. • Stateful • Stateless

Step 4 Click **Next**.

Step 5 In the **Specify Conditions** page of the **Add Trigger** wizard, click **Add a new entry to the table below (+)**, and complete the following fields in the **Add Entry to Conditions** dialog box:

a) From the **Type of Object to Monitor** drop-down list, choose **BigData Nodes**.
b) From the **Object** drop-down list, choose the disk to be monitored.
c) From the **Parameter** drop-down list, choose the parameter against which to validate.
d) From the **Operation** drop-down list, choose the type of operation.
e) From the **Value** drop-down list, choose the value against which to validate.
f) Click **Submit**.
g) From the **Trigger When** drop-down list, choose to satisfy all the conditions or any condition.

Step 6 Click **Next**.

Step 7 in the **Specify Workflow**page of the **Add Trigger** wizard, do the following when there is a network or disk failure and when the trigger is reset:

a) From the **Maximum Number of Invocations** drop-down list, choose the number of maximum invocations.
b) Select a workflow for execution when the trigger state becomes active and check the **Pass Monitored Object** check box, if required.
c) Select the workflow input.
d) Select a workflow for execution when the trigger state becomes clear and check the **Pass Monitored Object** check box, if required.
e) Select the workflow input.

Step 8 Click **Next**.

Step 9 In the **Specify Workflow Inputs** page of the **Add Trigger** wizard, enter the inputs for the selected workflows, and then click **Submit**.

Setting Disk Utilization Threshold Alerts

You can set an alert when the disk capacity reaches a threshold. This helps you to pro-actively plan for capacity expansions.

Step 1 On the menu bar, choose **Policies** > **Orchestration**.

Step 2 Click the **Triggers** tab.

Step 3 Click **Add**.
In the **Trigger Information** page of the **Add Trigger** wizard, complete the following fields:

Name	Description
Trigger Name field	Name of the trigger.
Is Enabled check box	Check this box to enable the trigger.

Name	Description
Description	Description of the trigger.
Frequency	Choose the trigger rule validation frequency.
Trigger Type	Choose the type of the trigger. • Stateful • Stateless

Step 4 Click **Next**.

Step 5 In the **Specify Conditions** page of the **Add Trigger** wizard, click **Add a new entry to the table below (+)**, and complete the following fields in the **Add Entry to Conditions** dialog box:

a) From the **Type of Object to Monitor** drop-down list, choose **BigData Cluster**.
b) From the **Object** drop-down list, choose the disk to be monitored.
c) From the **Parameter** drop-down list, choose the **Disk Utilization (%)**.
d) From the **Operation** drop-down list, choose the type of operation.
e) From the **Value** drop-down list, choose the threshold value against which to validate.
f) Click **Submit**.
g) From the **Trigger When** drop-down list, choose to satisfy all the conditions or any condition.

Step 6 Click **Next**.

Step 7 in the **Specify Workflow**page of the **Add Trigger** wizard, do the following when the disk utilization reaches the threshold value and when the trigger is reset:

a) From the **Maximum Number of Invocations** drop-down list, choose the number of maximum invocations.
b) Select a workflow for execution when the trigger state becomes active and check the **Pass Monitored Object** check box, if required.
c) Select the workflow input.
d) Select a workflow for execution when the trigger state becomes clear and check the **Pass Monitored Object** check box, if required.
e) Select the workflow input.

Step 8 Click **Next**.

Step 9 In the **Specify Workflow Inputs** page of the **Add Trigger** wizard, enter the inputs for the selected workflows, and then click **Submit**.