

Dell EMC PowerScale and Cloudera Data Platform Private Cloud Base

Abstract

This document provides step-by-step installation guidance for deploying the Cloudera Data Platform (CDP) Private Cloud Base 7.1.6 on Dell EMC™ PowerScale™ powered by Dell EMC PowerScale OneFS™ 8.2.2.

April 2021

Revisions

Date	Description
April 2021	Initial release

Acknowledgments

Author: Kirankumar Bhusanurmath, Analytics Solutions Architect

Support: Russ Stevenson, Advisory Systems Engineer

The information in this publication is provided “as is.” Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

This document may contain certain words that are not consistent with Dell's current language Guidelines. Dell plans to update the document over subsequent future releases to revise these words accordingly.

This document may contain language from third party content that is not under Dell's control and is not consistent with Dell's current Guidelines for Dell's own content. When such third party content is updated by the relevant third parties, this document will be revised accordingly.

Copyright © 2021 Dell Inc. or its subsidiaries. All Rights Reserved. Dell Technologies, Dell, EMC, Dell EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners. [4/15/2021] [Deployment and Configuration] [H18730]

Table of contents

Revisions.....	2
Acknowledgments.....	2
Table of contents	3
Executive summary.....	6
Audience	6
1 Overview.....	7
1.1 Updates and information about the OneFS Hadoop installation	9
1.2 Prerequisites.....	9
1.2.1 Cloudera Data Platform Private Cloud Base	9
1.2.2 OneFS cluster configuration	9
2 Installing OneFS with Cloudera Manager	11
2.1 Preparing OneFS.....	11
2.2 Validating OneFS version and license activation	11
2.3 Configuring OneFS components	12
2.3.1 Create an access zone	13
2.3.2 Configure SmartConnect	14
2.3.3 Configure DNS for OneFS.....	15
2.3.4 Verify the SmartConnect configuration.....	15
2.4 Creating HDFS users and groups	16
2.4.1 Create users and directories on the OneFS cluster using Tools for Using Hadoop with OneFS	16
2.4.2 Create users on the OneFS cluster manually	16
2.5 Configuring the HDFS user for OneFS 8.2 and later versions	17
3 Configuring Kerberos with OneFS.....	20
3.1 Prerequisites.....	20
3.1.1 OneFS	20
3.1.2 How Kerberos is implemented on the OneFS and Hadoop clusters.....	20
3.2 Creating the Active Directory as a OneFS authorization provider	21
3.2.1 Review the OneFS SPNs	22
3.2.2 Create proxy users	22
3.2.3 Enable Kerberos on the HDFS zone and view HDFS settings	22
3.3 Creating the MIT KDC as a OneFS authorization provider	23
4 Installing CDP Private Cloud Base.....	28
4.1 Version and download information	28
4.2 CDP Private Cloud Base requirements and supported versions	28

4.3	Installing Cloudera Manager.....	28
4.3.1	Installing Cloudera Manager.....	28
4.4	Installing PowerScale Custom Service Descriptor	29
4.4.1	Download PowerScale CSD into Cloudera Manager host	29
4.4.2	Install PowerScale CSD into Cloudera Manager.....	29
4.5	Kerberizing Cloudera Manager.....	31
4.5.1	Enabling Active Directory or MIT KDC as an authentication provider	32
4.5.2	Set up KDC for Cloudera Manager.....	33
4.6	Installing Cloudera Runtime	36
4.7	Setting up a cluster using the wizard.....	46
4.7.1	Select Services	47
4.7.2	Assign Roles	48
4.7.3	Setup Database	48
4.7.4	Enter Required Parameters	48
4.7.5	Review Changes.....	49
4.7.6	Command Details	49
4.7.7	Summary	51
4.8	Other steps for Apache Ranger	53
4.8.1	Steps to enable Ranger service on PowerScale	53
4.9	Completing post-installation steps.....	58
4.9.1	Deploying clients.....	58
4.9.2	Testing the Installation.....	58
4.9.3	Checking host heartbeats	59
4.9.4	Running a MapReduce Job	59
4.9.5	Testing with Hue	59
4.9.6	Securing Your cluster	59
4.10	Uninstalling Cloudera Manager and Managed Software.....	60
A	Troubleshooting installation problems.....	61
A.1	TLS protocol error with OpenJDK.....	61
A.2	Failed to start server reported by cloudera-manager-installer.bin.....	61
A.3	Installation interrupted and installer do not restart.	62
A.4	Cloudera Manager Server fails to start with MySQL	62
A.5	Agents fail to connect to Server.....	62
A.6	Cluster hosts do not appear.	62
A.7	"Access denied" in install or update wizard.....	63
A.8	Databases fail to start.....	63

- A.9 Cloudera services fail to start63
- A.10 Activity Monitor displays a status of BAD.64
- A.11 Activity Monitor fails to start64
- A.12 Create Hive Metastore Database Tables command fails.64
- A.13 Oracle invalid identifier64
- A.14 Failed to upload Tez jar file during installation65
- A.15 Zeppelin fails on first run65
- A.16 SOLR service startup issue65
- B Technical support and resources67

Executive summary

This document provides step-by-step installation guidance for deploying the Cloudera Data Platform (CDP) Private Cloud Base 7.1.6 on Dell EMC™ PowerScale™ powered by Dell EMC PowerScale OneFS™ 8.2.2. Cloudera is a distribution of Apache® Hadoop®, an open-source framework that enables the distributed processing of large sets of data across clusters of systems. Before you begin the procedures in this document, you must install a PowerScale OneFS cluster.

Audience

This guide is intended for systems administrators, IT program managers, IT architects, and IT managers who are installing PowerScale OneFS with a Cloudera distribution of Hadoop.

1 Overview

The PowerScale OneFS scale-out network-attached storage (NAS) platform provides Hadoop clients with direct access to big data through a Hadoop Distributed File System (HDFS) protocol interface. A PowerScale cluster powered by the OneFS operating system delivers a scalable pool of storage with a global namespace.

Hadoop compute clients can access the data that is stored on a PowerScale OneFS cluster by connecting to any node over the HDFS protocol. All nodes that are configured for HDFS provide NameNode and DataNode functionality. Each node boosts performance and expands the cluster capacity. For Hadoop analytics, the PowerScale scale-out distributed architecture minimizes bottlenecks, rapidly serves big data, and optimizes performance for MapReduce jobs.

In a traditional Hadoop deployment (shown in Figure 1), the Hadoop compute nodes run analytics jobs against large sets of data. A NameNode directs the compute nodes to the data stored on a series of DataNodes. The NameNode is a separate server that holds metadata for every file that is stored on the DataNodes. Data is stored in production environments and then copied to a landing zone server before it is loaded to HDFS. This process is network-intensive and exposes the NameNode as a potential single point of failure.

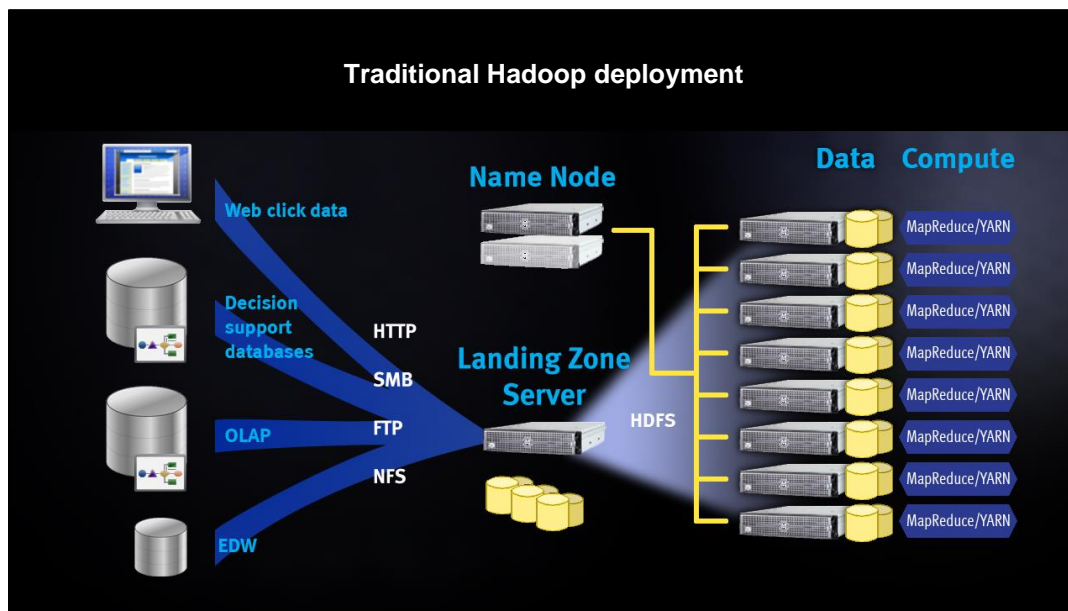


Figure 1 Traditional Hadoop deployment

In a PowerScale OneFS deployment with Hadoop (see Figure 2), OneFS serves as the file system for Hadoop compute clients. On a PowerScale OneFS cluster, every node in the cluster acts as a NameNode and DataNode, providing automated failover protection.

When a Hadoop client runs a job, the clients access the data that is stored on a OneFS cluster by connecting over HDFS. The HDFS protocol is native to the OneFS operating system, and no data migration is required.

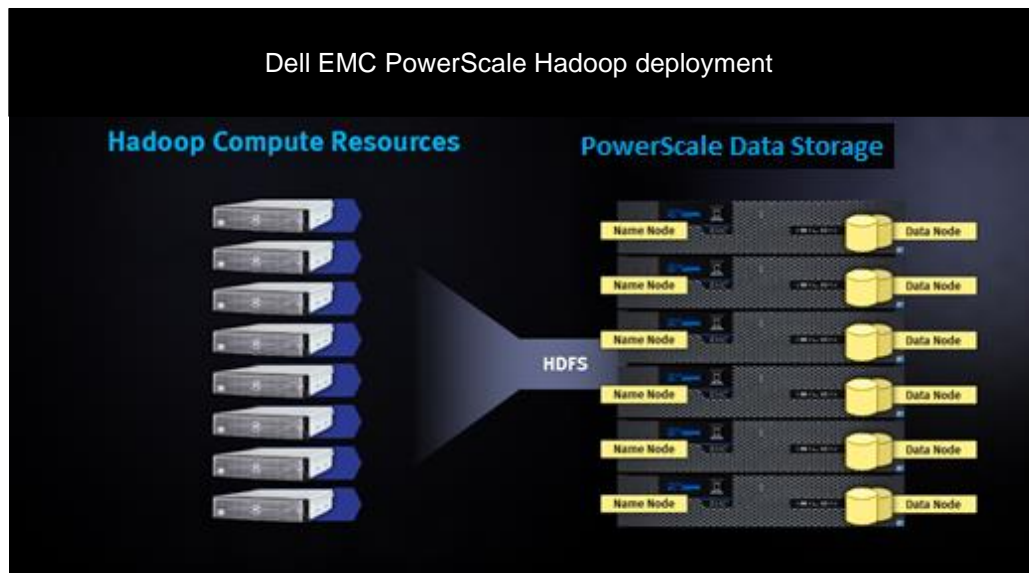


Figure 2 PowerScale Hadoop deployment

The Cloudera distribution is stored on a separate compute cluster, and individual clients connect directly to the OneFS cluster to store and access Hadoop data (see Figure 3). OneFS handles HDFS file-data exchange as a protocol to store and retrieve the data to match the client requirements.

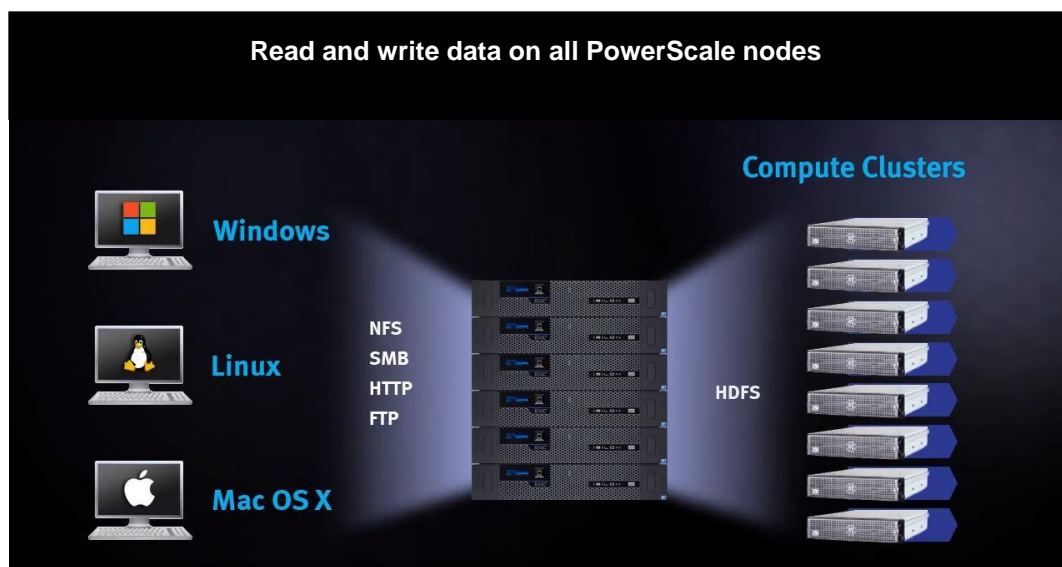


Figure 3 Read and write data on all PowerScale nodes

1.1 Updates and information about the OneFS Hadoop installation

The rapid release of new features and versions of Hadoop projects can introduce new behaviors and requirements. We recommend that you review the latest updates on the [Using Hadoop with Isilon - Isilon Info Hub](#) for updates and known issues while deploying OneFS and Hadoop.

1.2 Prerequisites

For supported Hadoop versions, see the page [Hadoop Distributions and Products Supported by OneFS](#).

1.2.1 Cloudera Data Platform Private Cloud Base

Ensure that your environment meets the following requirements:

- Ensure that Cloudera Manager is at version 7.3.1. See the [Cloudera Manager Download Information](#).
- Ensure that Cloudera Runtime is at version 7.1.6 or later. See the [Cloudera Runtime Download Information](#).
- See the full release documentation at [CDP overview](#) including the [Cloudera Manager Release Notes](#) and [Cloudera Runtime Release Notes](#).
- Ensure familiarity with Cloudera documentation and installation instructions. View the Cloudera documents at <http://www.cloudera.com/documentation.html>.
- Use Table 1 to record the components that you plan to install.

Table 1 Cloudera components and versions

Component	Version
Cloudera Manager	
CDH parcel	

1.2.2 OneFS cluster configuration

Ensure that your environment meets the following requirements:

- The OneFS cluster is running OneFS version 8.2.2.0 or later.
- SmartConnect Advanced, a separately licensed OneFS module, is activated, and SmartConnect is configured on your OneFS cluster.
- HDFS, a separately licensed OneFS module, is activated on your OneFS cluster. Contact your Dell EMC PowerScale sales representative for more information about receiving your license keys.
- A valid OneFS SmartConnect SSIP and Domain Name System (DNS) delegation is in place to provide name-resolution services for a SmartConnect zone. For more information, see [Isilon External Network Connectivity Guide](#).
- A dedicated OneFS access zone is in use; this access zone is not the same as the system zone.
- A OneFS HDFS root directory is in the access zone.
- A simple access model is between Hadoop and OneFS, with UID and GID parity

- Use Table 2 to record the components that you have installed.

Table 2 OneFS cluster components

Component	Version or license
PowerScale OneFS	
SmartConnect module	
HDFS module	
OneFS cluster name	
PowerScale custom service descriptor	

2 Installing OneFS with Cloudera Manager

The installation of OneFS with Cloudera is separated into four stages as represented in Figure 4.

To complete each stage, you must perform tasks on both the Cloudera cluster and the OneFS cluster.

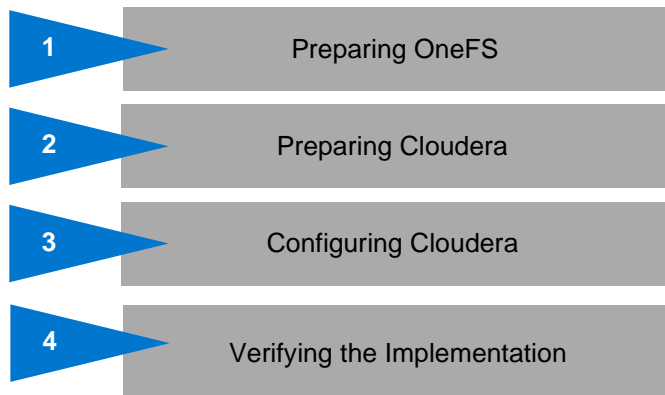


Figure 4 Installation stages for OneFS with Cloudera Manager

2.1 Preparing OneFS

Complete the following steps to configure your OneFS cluster for use with Cloudera Data Platform. Preparing OneFS requires you to configure DNS, SmartConnect, and Access Zones to allow for the Hadoop cluster to connect to the OneFS cluster. If these preparation steps are not successful, the subsequent configuration steps might fail.

Review the current [Isilon OneFS and Hadoop Known Issues](#) for any changes or updates to OneFS and Hadoop configuration.

2.2 Validating OneFS version and license activation

Validate your OneFS version, check your licenses, and confirm that they are activated. Other OneFS licenses may be required for other OneFS functionality to be interoperable with HDFS, and they are not addressed in this installation Guide.

1. From a node in your OneFS cluster, confirm that your OneFS cluster is running OneFS 8.1.2 or later by entering the following command:

```
isi version
```

2. Add the licenses for HDFS:

```
isi license add --evaluation=HDFS
```

3. Confirm that license for HDFS is operational. If this license is not active and valid, some commands in this Guide will not work.

Run the following commands to confirm that HDFS is installed:

```
isi license licenses list
isi license licenses view HDFS
```

4. If your modules are not licensed, obtain a license key from your Dell EMC PowerScale sales representative. Type the following command to activate the license:

```
isi license add --path <license file path>
```

5. Enable HDFS:

```
isi services hdfs enable
```

6. Install the latest rollup patches for your version of OneFS. See [Current Isilon OneFS Patches](#) for the latest rollup patches, and run the following:

```
isi upgrade patches list
isi upgrade patches install patch-<patch-ID>.pkg --rolling=false
```

For example:

```
isi upgrade patches install patch-240163.pkg --rolling=false
```

2.3 Configuring OneFS components

After you configure DNS for OneFS, set up and configure the following OneFS components:

- Create an access zone.
- Create a SmartConnect zone.
- Create and configure the HDFS root in the access zone.
- Create users and groups.
- Create a basic HDFS folder structure for use with HDFS.

Use Table 3 to record the configuration information for the OneFS cluster with Cloudera integration:

Table 3 Configuration information

Parameter	Value
Access zone name	
Access zone path	
SmartConnect zone name (FQDN)	
IP range for IP pool (ranges)	
SmartConnect pool name (subnet pool)	
Node and interfaces in the pool	
HDFS root path	

2.3.1 Create an access zone

On one of the OneFS nodes, you must define an access zone on the OneFS cluster and enable the Hadoop node to connect to it.

1. On a node in the OneFS cluster, create your Hadoop access zone:

```
isi zone zones create --name=zone1-cdp --path=/ifs/data/zone1/cdp --
create-path
```

2. Verify that the access zones are set up correctly:

```
isi zone zones list --verbose
```

Output similar to the following displays:

```

                Name: System
                Path: /ifs
            Groupnet: groupnet0
        Map Untrusted: -
    Auth Providers: lsa-local-provider:System, lsa-file-provider:System
        NetBIOS Name: -
    User Mapping Rules: -
Home Directory Umask: 0077
    Skeleton Directory: /usr/share/skel
    Cache Entry Expiry: 4H
                Zone ID: 1
-----
-----
                Name: zone1-cdp
                Path: /ifs/data/zone1/cdp
            Groupnet: groupnet0
        Map Untrusted: -
    Auth Providers: lsa-local-provider:zone1-cdp
        NetBIOS Name: -
    User Mapping Rules: -
Home Directory Umask: 0077
    Skeleton Directory: /usr/share/skel
    Cache Entry Expiry: 4H
                Zone ID: 2
```

3. Create the HDFS root directory within the access zone that you created:

```
mkdir -p /ifs/data/zone1/cdp/hadoop-root
isi hdfs settings modify --zone=zone1-hdp --root-
directory=/ifs/data/zone1/cdp/hadoop-root
```

4. List out the contents of the Hadoop access zone root directory:

```
ls -al /ifs/data/zone1/cdp
```

2.3.2 Configure SmartConnect

On a node in the OneFS cluster, add a static IP address pool and associate it with the access zone that you created earlier.

1. Modify your existing subnets and specify a service address:

```
isi network subnets modify groupnet0.subnet0 --sc-service-addr=x.x.x.x
```

2. Create an access network pool. Run the following command using the following parameters:

- **<groupnet>:<subnet>:<name>** is the new IP pool in subnet (for example, subnet0:pool1).
- **<IP-IP>** is the IP range that is assigned to the IP pool.
- **<access-zone>** is the access zone that the pool is assigned to.
- **<interfaces>** are the node interfaces that are added to the pool.
- **<subnet>** is the SmartConnect service subnet that is responsible for this zone.
- **<smartconnectzone>** is the SmartConnect zone name.

```
isi network pools create --id=<groupnet>:<subnet>:<name> --ranges=<IP-IP>
--access-zone=<access-zone> --alloc-method=static --ifaces=<interfaces> --
sc-subnet=<subnet> --sc-dns-zone=<smartconnectzone> --description=hadoop
```

For example:

```
isi network pools create groupnet0:subnet0:hadoop-pool-cdh --
ranges=10.120.130.30-10.120.140.40 --access-zone=zone1-cdp --alloc-
method=static --ifaces=1-4:40gige-1 --sc-subnet=subnet0 --sc-dns-
zone=cdp.zone1.emc.com --description=hadoop"
```

3. View the properties of the access network pool:

```
isi network pools view --id=groupnet0:subnet0:pool2
```

Output similar to the following displays:

```

ID: groupnet0.subnet0.hadoop-pool-cdp
Groupnet: groupnet0
Subnet: subnet0
Name: hadoop-pool-cdp
Rules: -
Access Zone: zone1-cdh
Allocation Method: static
Aggregation Mode: lacp
SC Suspended Nodes: -
Description: cdp_hadoop_access_zone
Ifaces: 1:ext-1, 2:ext-1, 3:ext-1, 4:ext-1
IP Ranges: 10.120.130.30-10.120.140.40
Rebalance Policy: auto
SC Auto Unsuspend Delay: 0
SC Connect Policy: round_robin
SC Zone: cdp.zone1.emc.com
SC DNS Zone Aliases: -
SC Failover Policy: round_robin
SC Subnet: subnet0
```

```
SC Ttl: 0
Static Routes: -
```

2.3.3 Configure DNS for OneFS

Before you configure DNS for OneFS, ensure that the OneFS cluster is implemented according to Dell EMC PowerScale best practices. For more information, see the [Dell EMC Isilon Best Practices Guide for Hadoop Data Storage](#).

Set up DNS records for a SmartConnect zone. Create the required DNS records that are used to access your OneFS cluster from the Hadoop cluster. All hosts in your Hadoop cluster must be configured for both forward and reverse DNS lookups. Hadoop relies heavily on DNS and performs many DNS lookups during normal operation.

You can set up a SmartConnect zone for the connections from Hadoop compute clients. SmartConnect is a module that specifies how the OneFS cluster handles connection requests from clients. For more information and best practices for SmartConnect, see the [Isilon External Network Connectivity Guide](#).

Each SmartConnect zone represents a specific pool of IP addresses. When you associate a SmartConnect zone with an access zone, OneFS allows only clients that connect through the IP addresses in the SmartConnect zone to reach the HDFS data in the access zone. A root HDFS directory is specified for each access zone. This configuration isolates data within access zones and allows you to restrict client access to the data.

A SmartConnect zone distributes NameNode requests from Hadoop compute clients across the node interfaces in the IP pool. The NameNode process of each node replies with the IP address of the HDFS DataNode where the client can access the data. When a Hadoop compute client makes an initial DNS request to connect to the SmartConnect zone FQDN, the Hadoop client requests are delegated to the SmartConnect service IP. The SmartConnect service IP then responds with a valid node to connect to. The client connects to a OneFS node that serves as a NameNode. When a second Hadoop client makes a DNS request to connect to the SmartConnect zone, the SmartConnect Service routes the client connection to another node. This node is different than the node that is used by the previous Hadoop compute client.

When you create a SmartConnect zone, you must add a name server (NS) record as a delegated domain to the authoritative DNS zone that contains the OneFS cluster.

2.3.4 Verify the SmartConnect configuration

Validate that SmartConnect is set up correctly by pinging the SmartConnect zone FQDN several times from the Hadoop client.

```
ping cdh.zone1.emc.com
```

When you view the output of this command, different IP addresses are returned for each ping command. With each DNS response, the IP addresses are returned through rotating round-robin DNS from the list of potential IP addresses. This function validates that the SmartConnect zone name FQDN is operating correctly.

2.4 Creating HDFS users and groups

For each Hadoop system account that submits HDFS jobs or accesses the file system, you must create local users and groups on the OneFS cluster. You can add Hadoop users and groups to the OneFS cluster manually or by following the process at: https://github.com/Isilon/isilon_hadoop_tools.

Dell EMC PowerScale recommends that you maintain consistent names and numeric IDs for all users and groups on the OneFS cluster and your Hadoop clients. This consistency is important in multiprotocol environments because the HDFS protocol refers to users and groups by name. In contrast, NFS refers to users and groups by their numeric IDs (UIDs and GIDs). Maintaining this parity is critical in the behavior of OneFS multiprotocol file access.

During installation of Hadoop with Cloudera Manager, the installer creates all required system accounts on all clients. For example, a Hadoop system account **yarn** is created with the UID of 502 and the GID of 502 on the Hadoop cluster nodes. Cloudera creates these accounts if they do not exist. You can ensure parity by precreating them on all nodes that will be installed in the Hadoop cluster. You can enforce parity by manually managing when and how these local system accounts are created. Since the Hadoop installer cannot create the local accounts directly on OneFS, you must create them manually. Create the OneFS **yarn** local account user in the OneFS access zone in which **yarn** accesses data. Create a local user **yarn** with the UID of 502 and the GID of 502 to ensure consistency of access and permissions.

For guidance and more information about maintaining parity between OneFS and Hadoop local users and UIDs, see the article [Isilon and Hadoop Local User UID Parity](#).

There are many methods of achieving UID and GID parity. You can use [Tools for Using Hadoop with OneFS](#), perform manual matching, or create scripts that parse users and create the equivalent users. However you choose to achieve this result, the sequence depends on your deployment methodology and management practices. We recommend that you maintain consistency between the Hadoop cluster and OneFS—for example, hdfs=hdfs, yarn=yarn, hbase=hbase, and so on—from a UID and GID consistency perspective.

2.4.1 Create users and directories on the OneFS cluster using Tools for Using Hadoop with OneFS

Go to [Tools for Using Hadoop with OneFS](#) to set up the users and directories on the cluster.

2.4.2 Create users on the OneFS cluster manually

You can add a user for each additional Hadoop user that submits MapReduce jobs in addition to the users that the OneFS script configures on the OneFS cluster. The following procedures show how to manually add a single user **hduser1**.

Note: If your users and groups are defined by your directory service, such as Active Directory or MIT KDC/LDAP, do **not** run the commands in this section. This section addresses setting permissions of the HDFS root files or membership to run jobs. These steps create users but are likely to fail when you run jobs with a configuration that uses Active Directory or MIT KDC/LDAP.

2.4.2.1 Perform manual steps on the OneFS cluster

1. Add a group to the OneFS cluster.

```
isi auth groups create hduser1 --zone zone1 --provider local --gid <GID>
```

2. Create the user and the user's Hadoop home directories on the OneFS cluster.

```
isi auth users create hduser1 --primary-group hduser1 --zone zone1 --  
provider local --home-directory /ifs/data/zone1/hadoop/user/hduser1 --uid  
<UID>
```

3. Assign permissions to the user's home directory on the Hadoop cluster. The ID 2 in the following example is from the previously run command **isi zone zones view zone1**.

```
isi_run -z2 chown hduser1:hduser1 /ifs/isiloncluster1/hadoop/user/hduser1  
chmod 755 /ifs/data/hadoop/user/hduser1
```

2.4.2.2 Perform manual step on the Hadoop client

Since you created a user on OneFS to run jobs, you must create the same user with UID parity on any Linux® hosts that the user accesses to run jobs.

Run the following command to add the user to the Hadoop cluster:

```
adduser hduser1 -u <UID>
```

2.5 Configuring the HDFS user for OneFS 8.2 and later versions

In OneFS 8.2.0 and later versions, the HDFS user is not required to be mapped to root. Instead, you must assign a new role with backup and restore privileges.

On a node in the OneFS 8.2 cluster, create a role and configure the backup and restore privileges to the HDFS user.

1. View the HDFS service settings.

```
isi hdfs settings view --zone=zone1-cdh
```

Note: We recommend that the directory for the access zone is not set to the root of /ifs.

2. Set the HDFS root directory for the access zone.

```
isi hdfs settings modify --zone=zone1-cdh --root-  
directory=/ifs/data/zone1/cdh/hadoop-root
```

3. Create a role for the Hadoop access zone.

```
isi auth roles create --name=<role_name> --description=<role_description>  
--zone=<access_zone>
```

For example:

```
isi auth roles create --name=HdfsAccess --description="Bypass FS
permissions" --zone=zone1-cdh
```

4. Add restore privileges to the new "HdfsAccess" role.

```
isi auth roles modify <role_name> --add-priv=ISI_PRIV_IFS_RESTORE --
zone=<access_zone>
```

For example:

```
isi auth roles modify HdfsAccess --add-priv=ISI_PRIV_IFS_RESTORE --
zone=zone1-cdh
```

5. Add backup privileges to the new **HdfsAccess** role.

```
isi auth roles modify <role_name> --add-priv=ISI_PRIV_IFS_BACKUP --
zone=<access_zone>
```

For example:

```
isi auth roles modify HdfsAccess --add-priv=ISI_PRIV_IFS_BACKUP --
zone=zone-cdh
```

6. Add user hdfs to the new **HdfsAccess** role.

```
isi auth roles modify <role_name> --add-user=hdfs --zone=<access_zone>
```

For example:

```
isi auth roles modify HdfsAccess --add-user=hdfs --zone=zone1-cdh
```

7. Verify the role setup, backup and restore privileges, and HDFS user setup.

```
isi auth roles view <role_name> --zone=<access_zone>
```

For example:

```
isi auth roles view HdfsAccess --zone=zone1-cdh
Name: HdfsAccess
Description: Bypass FS permissions
Members: - hdfs
Privileges
ID: ISI_PRIV_IFS_BACKUP
Read Only: True
ID: ISI_PRIV_IFS_RESTORE
Read Only: True
```

8. Optionally, flush the auth mapping and auth cache to make the HDFS user immediately take the HdfsAccess role that you created before.

```
isi_for_array "isi auth mapping flush --all"
isi_for_array "isi auth cache flush --all"
```

Note: ACL policies no longer must be modified for OneFS 8.2 and later since the HDFS protocols act the same as the non-OneFS HDFS protocol for file-system-group-owner inheritance.

3 Configuring Kerberos with OneFS

The Cloudera CDP Private Cloud Base deployment on PowerScale OneFS requires you to Kerberize Cloudera Manager and PowerScale OneFS before you deploy the Cloudera runtime (CDP). This deployment is the recommended mode when the PowerScale custom service descriptor is involved. You must Kerberize the cluster for Cloudera Runtime services to operate (for example, Apache Ranger).

You can configure Kerberos security with OneFS 8.2.2 and later versions using existing Microsoft® Active Directory or MIT KDC installations.

3.1 Prerequisites

Before you configure Kerberos on your OneFS cluster, ensure that you meet the prerequisites listed in the following sections.

3.1.1 OneFS

This Guide assumes that the following OneFS Hadoop environment is configured and operational.

- You must be running OneFS 8.2.2 or later.
- A dedicated OneFS access zone is in use; this access zone is not in the system zone.
- A OneFS SmartConnect zone is correctly configured for HDFS access.
- A simple access model exists between Hadoop and OneFS. User UIDs and GIDs are correctly implemented and allow HDFS access to the OneFS HDFS root with UID and GID parity.
- DNS for SmartConnect is correctly configured, including forward and reverse lookups.

Also, ensure that OneFS is preconfigured to respond to requests related to the secure Kerberized HDFS that is authenticated by the MIT Kerberos key distribution center (KDC) or by Microsoft Active Directory providers. See the [Microsoft Active Directory documentation](#) for a high-level technical review regarding using Active Directory as a KDC.

3.1.2 How Kerberos is implemented on the OneFS and Hadoop clusters

Since the OneFS-integrated Hadoop cluster is a blend between Linux hosts running compute services and OneFS running data services, Cloudera cannot complete the Kerberization end-to-end. Since OneFS is a clustered operating system, you cannot use SSH-based remote management to configure and manage the Kerberization of OneFS completely. Deploy the Kerberization of a OneFS-integrated Hadoop cluster as follows:

- The OneFS cluster is Kerberized.
- The Cloudera Kerberization wizard deploys Kerberization to the Linux and Hadoop services.

When both the OneFS and Hadoop cluster are fully Kerberized within the same Active Directory domain, Kerberized user access can occur between both systems seamlessly.

For more information, see the Cloudera security documents at [Enabling Kerberos Authentication for CDP](#).

3.2 Creating the Active Directory as a OneFS authorization provider

This section covers the configuration requirements for OneFS to respond to requests for a secure Kerberized HDFS that is authenticated by Active Directory.

Note: You must configure the following items correctly before you go to the next section.

- Join the cluster correctly to the target Active Directory as a provider. Configure the following advanced settings in the OneFS web administration interface. These settings maintain user and identity mappings between users who perform Hadoop jobs and the OneFS cluster, and also enable a standard OneFS permission model.
 - a. In the OneFS web administration interface, click **Access > Authentication Providers > Active Directory**.
 - b. In the **Active Directory Providers** table, click **View details** for the provider whose settings you must modify.
 - c. Click **Advanced Active Directory Settings**, and set the following:
 - > In the **Services For UNIX Setting**, specify **RFC 2307**.
 - > Ensure that you have enabled Active Directory GC indexing and replication as described in the KB article [OneFS: How to configure OneFS and Active Directory for RFC2307 compliance](#), following the guidance for OneFS versions 8.x.x.x. and Windows Server 2012. This configuration is required to support Active Directory that provides UIDs and GIDs to OneFS.
- Configure the access zone that contains the HDFS root for this Active Directory provider, and configure the HDFS access zones service for Kerberos only.
- Configure the OneFS Service Principal Names (SPNs). Users running Hadoop jobs must have Active Directory user principals with UNIX attributes allocated. OneFS uses the active schema extension that supports UNIX identities. These schema attributes extend Active Directory objects to provide UIDs and GIDs to a user account in Active Directory. Depending on your setup, your Linux hosts might require integration into Active Directory for identity management.
- Add all IP addresses within the required SmartConnect zone to the reverse DNS with the same fully qualified domain name (FQDN) for the cluster delegation. All IPs should resolve back to the SmartConnect zone. This configuration is required for Kerberos authentication.
- Add the mapping rules to map the local HDFS to root, the Active Directory HDFS principal to root, the domain\hdfs to root, and all domain users to the local user, if applicable. In this example, **vlab** is the domain name and **zone1-hdp** is the access zone:

```
isi zone zones modify --user-mapping-rules="hdfs=>root, vlab\hdfs=>root,
vlab\* &= *[], vlab\* += *[group], vlab\* += *[groups]" --zone=zone1-hdp
```

Note the following regarding this command:

- **vlab* &= *[]**: Maps all AD users to the local user, for example, AD\bob = bob, AD\jane = jane
- **vlab* += *[group]** (optional): Maps the users' primary group to AD; defines the GID group and not domain users.
- **vlab* += *[groups]** (optional): Maps the users' primary group to AD; defines GID group and not domain users.

For mapping rules, use the short Network Basic Input/Output System (NetBIOS) name of the domain only, not the fully qualified domain name.

Generate the mapping results:

```
isi zone zones list -v
```

3.2.1 Review the OneFS SPNs

OneFS is a clustered file system that runs on multiple nodes that are joined to Active Directory as a single computer object. Therefore, the service principal name (SPN) requirements for Kerberized Hadoop access are unique.

OneFS requires additional SPNs for the access zone to which the HDFS NameNode access is provided when Active Directory is used, as summarized in the following table:

SPN	Name	Role
hdfs/clustername.fqdn	Clustername that is joined to AD	HDFS authentication to AD
hdfs/namenode.smartconnectname.fqdn	NN FQDN used by Ambari	HDFS authentication to AD for each SmartConnect Zone
HTTP/namenode.smartconnectname.fqdn	NN FQDN used by Ambari	WebHDFS authentication to AD for each SmartConnect Zone

Review the registered SPNs on the OneFS cluster, and run the following command to add the required SPNs for the SmartConnect zone name, if required:

```
isi auth ads spn list --provider-name=<AD PROVIDER NAME>
```

The following example illustrates the required OneFS SPNs:

```
Isilon Cluster Name - rip2.foo.com - SPN: hdfs/rip2.foo.com
Access Zone NN SmartConnect FQDN - rip2-cd1.foo.com - SPNs: hdfs/rip2-cd1.foo.com & HTTP/rip2-cd1.foo.com
```

For more information about adding or modifying OneFS SPNs in Active Directory, see the KB article [Isilon OneFS CLI Administration Guide](#).

3.2.2 Create proxy users

Create the required proxy users. Proxy users are required for service account impersonation for specific Hadoop services to run jobs and to add the required proxy users, as required. For more information about creating proxy users, see the KB article [Isilon OneFS CLI Administration Guide](#).

3.2.3 Enable Kerberos on the HDFS zone and view HDFS settings

1. Enable Kerberos on the HDFS zone. Change the HDFS access to KRB-only by running the following command on the Isilon OneFS cluster:

```
isi hdfs settings modify --zone=<zone-name> --authentication-mode=kerberos_only
```

2. View the HDFS settings.

```
isi hdfs settings view -zone=<zone-name>
```

For example:

```
rsteven-45uwrnw-1# isi hdfs settings view --zone=zone2-cdh
      Service: Yes
    Default Block Size: 128M
  Default Checksum Type: none
    Authentication Mode: kerberos_only
      Root Directory: /ifs/zone2/cdh/hadoop-root
    WebHDFS Enabled: Yes
      Ambari Server: -
    Ambari Namenode: -
      Odp Version: -
    Data Transfer Cipher: none
  Ambari Metrics Collector: -
```

3.3 Creating the MIT KDC as a OneFS authorization provider

1. Run a command similar to the following example (using your parameters) on your OneFS cluster to create the realm:

```
isi auth krb5 create --realm=VLAB.LOCAL --admin-
server=RDUVNODE60909.vlab.local - kdc=RDUVNODE60909.vlab.local --
user=cloudera-scm/admin@VLAB.LOCAL
```

2. List the realm.

```
isi auth krb5 realm list
```

For example:

```
rsteven-45uwrnw-1# isi auth krb5 realm list
Realm      Is Default Realm  KDC                                Admin Server
-----
VLAB.LOCAL Yes                RDUVNODE60909.vlab.local RDUVNODE60909.vlab.local
-----
Total: 1
rsteven-45uwrnw-1#
```

3. Create the Kerberos domains.

```
isi auth krb5 domain create --domain=<domain-name> --realm=<realm-name>
isi auth krb5 domain create --domain=.<domain-name> --realm=<realm-name>
isi auth krb5 domain list -verbose
```

For example:

```
rsteven-45uwrnw-1# isi auth krb5 domain create --domain=vlab.local --realm=VLAB.LOCAL
rsteven-45uwrnw-1# isi auth krb5 domain create --domain=.vlab.local --realm=VLAB.LOCAL
rsteven-45uwrnw-1# isi auth krb5 domain list --verbose
Domain: .vlab.local
Realm: VLAB.LOCAL
-----
Domain: vlab.local
Realm: VLAB.LOCAL
```

You can also view the two Kerberos domains that you created in the OneFS web administration interface under the **Kerberos Provider** tab, as shown in the following screen. Since you have not added the OneFS SPNs yet, the screen displays the **Requires Additional Configuration** warning.

Kerberos Realms

Realms	Key Distribution Centers (KDCs)	Admin Server	Actions
VLAB.LOCAL Default Realm	RDUVNODE80909.vlab.local	RDUVNODE80909.vlab.local	View / Edit More

Kerberos Domains

Domain	Kerberos Realm	Actions
.vlab.local	VLAB.LOCAL	View / Edit More
.vlab.local	VLAB.LOCAL	View / Edit More

Kerberos Providers

Realms	Service Principal Name (SPN) Management	Actions
VLAB.LOCAL	Recommended Requires Additional Configuration	View / Edit More

4. Add the Kerberos provider to the access zone and view the zones.

```
isi zone zones modify --zone=<zone-name> --add-auth-provider=<provider-type>:<provider-name>
isi zone zones view --zone=<zone-name>
```

For example:

```
rsteven-45uwrnw-1# isi zone zones modify --zone=zone2-cdh --add-auth-provider=krb5:VLAB.LOCAL
rsteven-45uwrnw-1# isi zone zones view --zone=zone2-cdh
Name: zone2-cdh
Path: /ifs/zone2/cdh
Groupnet: groupnet0
Map Untrusted: -
Auth Providers: lsa-local-provider:zone2-cdh, lsa-krb5-provider:VLAB.LOCAL
NetBIOS Name: -
User Mapping Rules: hdfs=>root
Home Directory Umask: 0077
Skeleton Directory: /usr/share/skel
Cache Entry Expiry: 4H
Negative Cache Entry Expiry: 1m
Zone ID: 3
```


- Run the following command to create the service principal names (SPNs) (using your Kerberos provider names). MIT KDC requires two SPNs: `hdfs/smartconnectzone-name` and `HTTP/smartconnectzone-name`.

```
isi auth krb5 spn create --provider-name=VLAB.LOCAL --
spn=hdfs/isilonsczone-cdh2.vlab.local --user=cloudera-scm/admin@VLAB.LOCAL
isi auth krb5 spn create --provider-name=VLAB.LOCAL --
spn=HTTP/isilonsczone-cdh2.vlab.local --user=cloudera-scm/admin@VLAB.LOCAL
```

- List the Kerberos realms.

```
isi auth krb5 spn list
```

For example:

```
rsteven-45uwrnw-1# isi auth krb5 spn list VLAB.LOCAL
SPN                                Kvno
-----
HTTP/isilonsczone-cdh2.vlab.local@VLAB.LOCAL 2
hdfs/isilonsczone-cdh2.vlab.local@VLAB.LOCAL 2
-----
Total: 2
Note that this Kerberos realm has SPNs and keys managed manually.
```

You can also view the principals in the OneFS web administration interface. The previous warnings are gone.

The screenshot shows the OneFS web administration interface with the 'Kerberos Provider' tab selected. The interface includes a top navigation bar with tabs for Active Directory, LDAP, NIS, Local Provider, File Provider, Kerberos Provider, and Kerberos Settings. Below the navigation bar is a 'Create a Kerberos Provider in One Step' section with a 'Get Started' button. The main content area is divided into three sections: 'Kerberos Realms', 'Kerberos Domains', and 'Kerberos Providers'. Each section has a 'Bulk actions' dropdown and a '+ Create a Kerberos [Type]' button.

Kerberos Realms

Realm	Key Distribution Centers (KDCs)	Admin Server	Actions
VLAB.LOCAL Default Realm	RDUVNODE80909.vlab.local	RDUVNODE80909.vlab.local	View / Edit / More

Kerberos Domains

Domain	Kerberos Realm	Actions
.vlab.local	VLAB.LOCAL	View / Edit / More
vlab.local	VLAB.LOCAL	View / Edit / More

Kerberos Providers

Realm	Service Principal Name (SPN) Management	Actions
VLAB.LOCAL	Manual	View / Edit / More

- Log in to KDC, and run the following command to list the OneFS principals that are created by the PowerScale system on the KDC.

```
listprincs
```

For example:

```
kadmin.local: listprincs
HTTP/RDUVNODE60904.vlab.local@VLAB.LOCAL
HTTP/isilonstorage-cdh2.vlab.local@VLAB.LOCAL
K/M@VLAB.LOCAL
cloudera-scm/admin@VLAB.LOCAL
hbase/RDUVNODE60904.vlab.local@VLAB.LOCAL
hdfs/isilonstorage-cdh2.vlab.local@VLAB.LOCAL
hive/RDUVNODE60904.vlab.local@VLAB.LOCAL
hue/RDUVNODE60904.vlab.local@VLAB.LOCAL
impala/RDUVNODE60904.vlab.local@VLAB.LOCAL
kadmin/admin@VLAB.LOCAL
kadmin/changepw@VLAB.LOCAL
kadmin/rduvnode60909.vlab.local@VLAB.LOCAL
kdcuser1@VLAB.LOCAL
krbtgt/VLAB.LOCAL@VLAB.LOCAL
mapred/RDUVNODE60904.vlab.local@VLAB.LOCAL
oozie/RDUVNODE60904.vlab.local@VLAB.LOCAL
spark/RDUVNODE60904.vlab.local@VLAB.LOCAL
yarn/RDUVNODE60904.vlab.local@VLAB.LOCAL
zookeeper/RDUVNODE60904.vlab.local@VLAB.LOCAL
```

The OneFS cluster should now be Kerberized.

Note: You can view and edit environment-specific Kerberos settings in the OneFS web administration interface under the **Kerberos Settings** tab as shown in the following screen.

Authentication Providers

Active Directory | LDAP | NIS | Local Provider | File Provider | Kerberos Provider | **Kerberos Settings**

Edit Kerberos Settings

– Kerberos Settings

Default Realm
VLAB.LOCAL

☐ Always send pre-authentication extensions

☒ Use DNS records to lookup KDCs and other servers for a realm

☒ Use DNS records to lookup Kerberos realm of a host

Revert Changes | Save Changes

- Create any necessary proxy users using the instructions in the KB article [Ambari Automated Kerberos Configuration with Isilon](#) as shown:

c) Create any necessary proxy users

In unsecured clusters, any user can impersonate any other user. In secured clusters, proxy users need to be explicitly specified. If you have Hive or Oozie, add the appropriate proxy users.

```
isi hdfs proxyusers create oozie --zone=$isilon_zone --add-user=ambari-qa
isi hdfs proxyusers create hive --zone=$isilon_zone --add-user=ambari-qa
```

9. Enable Kerberos on the HDFS zone. Run the following command on the PowerScale OneFS cluster to change the HDFS access to KRB-only.

```
isi hdfs settings modify --zone=<zone-name> --authentication-
mode=kerberos_only
```

10. View the HDFS settings.

```
isi hdfs settings view -zone=<zone-name>
```

For example:

```
rsteven-45uwrnw-1# isi hdfs settings view --zone=zone2-cdh
      Service: Yes
      Default Block Size: 128M
      Default Checksum Type: none
      Authentication Mode: kerberos_only
      Root Directory: /ifs/zone2/cdh/hadoop-root
      WebHDFS Enabled: Yes
      Ambari Server: -
      Ambari Namenode: -
      Odp Version: -
      Data Transfer Cipher: none
      Ambari Metrics Collector: -
```

4 Installing CDP Private Cloud Base

This section shows how to install Cloudera Manager, PowerScale Custom Service Descriptor (CSD), Cloudera Runtime, and other managed services on a PowerScale OneFS cluster.

The following high-level process installs CDP on PowerScale OneFS:

1. Review the version and download information.
2. Review CDP private cloud base requirements and supported versions.
3. Install Cloudera Manager.
4. Install PowerScale Custom Service Descriptor (CSD).
5. Kerberize Cloudera Manager (CM).
 - a. Enable MIT KDC as authentication provider.
 - b. Enable Active Directory as authentication provider.
6. Install Cloudera Runtime.
7. Set up a cluster using the wizard.
8. Perform post-installation steps.
9. Troubleshoot installation problems, if required.
10. Uninstall Cloudera Manager and managed software.

4.1 Version and download information

See the Cloudera article [Version and Download Information](#) for information about the available versions and download locations for Cloudera Manager and Cloudera Runtime.

4.2 CDP Private Cloud Base requirements and supported versions

See the Cloudera article [CDP Private Cloud Base Requirements and Supported Versions](#) for information about hardware, operating system, and database requirements, and for product-compatibility matrices.

4.3 Installing Cloudera Manager

We recommend following the procedure in this section to install Cloudera Manager for production environments. For a nonproduction trial, install Cloudera Manager as discussed in the Cloudera article [CDP Private Cloud Base Trial](#). Then, continue with the installation by completing section 4.4 through section 4.8.

4.3.1 Installing Cloudera Manager

Installing Cloudera Manager is out of scope for this guide. This section provides an overview and links to the installation procedure on the Cloudera website.

- [Step 1: Configure a Repository for Cloudera Manager](#): Configure a package repository to install Cloudera Manager.
- [Step 2: Install Java Development Kit](#): CDP Private Cloud Base requires a JDK installed on all hosts. You can either install OpenJDK or an Oracle JDK directly from Oracle.
- [Step 3: Install Cloudera Manager Server](#): In this step, you install the Cloudera Manager packages on the Cloudera Manager Server host, and optionally enable auto-TLS.

- [Step 4. Install and Configure Databases](#): Cloudera Manager uses various databases and datastores to store information about the Cloudera Manager configuration, and information such as the health of the system, or task progress.
- [Step 5: Set up and Configure the Cloudera Manager Database](#): Cloudera Manager Server includes a script that can create and configure a database for itself.

4.4 Installing PowerScale Custom Service Descriptor

After successful installation of Cloudera Manager, you must install the PowerScale Custom Service Descriptor (CSD). This step is crucial because it exposes other services in Cloudera Manager to deploy the PowerScale service instead of a traditional HDFS service.

PowerScale CSD is a software component that you can install in Cloudera Manager to define OneFS as a service in the CDP cluster. The CSD allows the Hadoop cluster admin to start, stop, and configure OneFS as an HDFS storage service. This ability provides native NameNode and DataNode capabilities that are similar to traditional HDFS.

4.4.1 Download PowerScale CSD into Cloudera Manager host

The PowerScale Custom Service Descriptor (CSD) is built for Dell EMC PowerScale. Then, download the PowerScale CSD from the [product download](#) page, and extract the contents to the Cloudera Manager server.

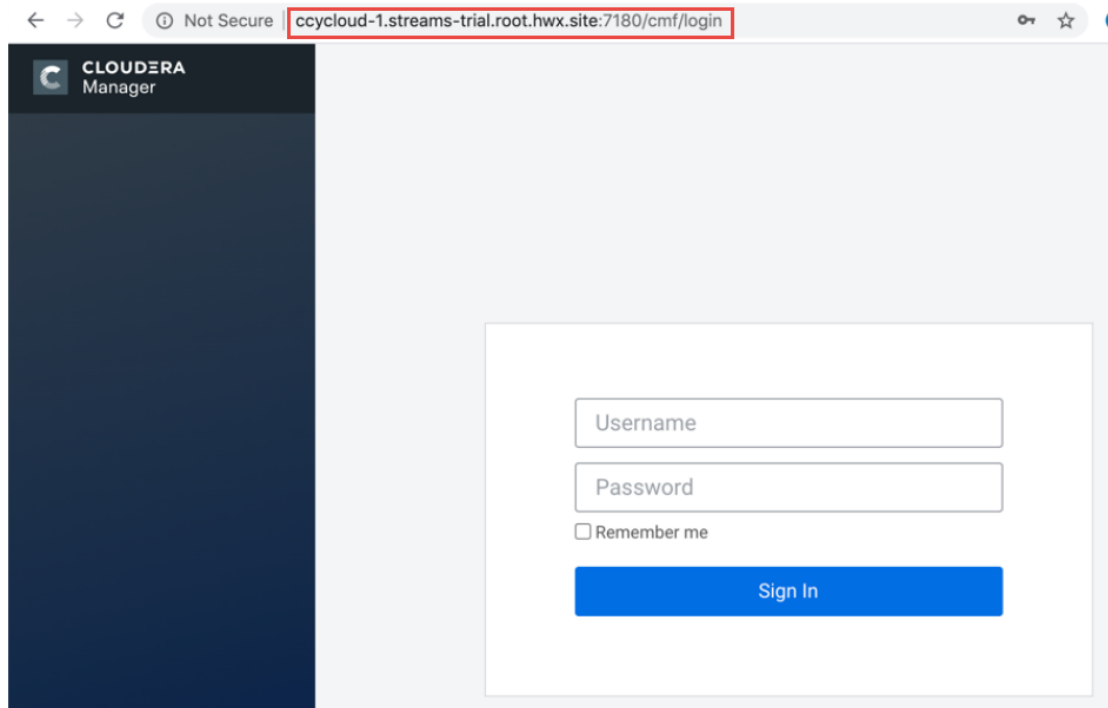
4.4.2 Install PowerScale CSD into Cloudera Manager

1. CM 7.3.1 already contains ISILON-7.3.1.jar file. If the file exists, delete it from the `/opt/cloudera/cm/csd` path, and replace it with the PowerScale CSD jar file. Otherwise, the Dell EMC PowerScale service does not appear during the Cloudera Runtime installation.
2. Copy or move the PowerScale CSD jar file **PowerScale-1.0.0.jar** to `/opt/cloudera/cm/csd` on the Cloudera Manager Server.

```
[root@hop-kiran-n65 csd]# pwd
/opt/cloudera/cm/csd
[root@hop-kiran-n65 csd]# ll PowerScale*
-rw-r--r-- 1 root root 79061 Apr  4 12:21 PowerScale-1.0.0.jar
[root@hop-kiran-n65 csd]#
```

3. Restart the Cloudera Manager, log in into the Cloudera Manager UI and complete the following steps until you reach the KDC wizard.
4. In a web browser, enter the URL that the Cloudera Manager Installer displayed in the previous task: **`http://<server_host>:7180`**. The variable `<server_host>` is the FQDN or IP address of the host where the Cloudera Manager Server is running. For example: `http://ccycloud-1.streams-trial.root.hwx.site:7180`.

The **Cloudera Manager Sign In** page appears.



5. Sign in with the default credentials:
 - Username: admin
 - Password: admin
6. Click Sign In. The **Welcome to Cloudera Manager** page appears.

Note: Upload the License File. For this demonstration, select the 60 days trial.

7. Select the following:

- Try Cloudera Data Platform for 60 days
- Yes, I accept the Cloudera Standard License Terms and Conditions

Welcome to Cloudera Manager 7.1.3

Upload License File

☐ Upload Cloudera Data Platform License

Cloudera Data Platform provides important features that help you manage and monitor your Hadoop clusters in mission-critical environments. Cloudera Data Platform is a subscription service with enhanced capabilities and support. [Contact Cloudera Sales](#)

Upload License File (Accept .txt or .zip)

☒ Try Cloudera Data Platform for 60 days

⚠ After the trial period, you will need a valid Cloudera Data Platform license to access the Cloudera Manager Admin Console. Your cluster and data will remain unaffected.

Cloudera Standard License

Version 2019-12-12

THE TERMS AND CONDITIONS OF THIS CLOUDERA STANDARD LICENSE (THE "AGREEMENT") APPLY TO YOUR USE OF OR ACCESS TO THE PRODUCTS (AS DEFINED BELOW) MADE AVAILABLE BY CLOUDERA, INC. ("CLOUDERA").

PLEASE READ THIS AGREEMENT CAREFULLY.

IF YOU ("YOU" OR "CUSTOMER") PLAN TO USE OR ACCESS ANY OF THE PRODUCTS ON BEHALF OF A COMPANY OR OTHER ENTITY, YOU REPRESENT THAT YOU ARE THE EMPLOYEE OR AGENT OF SUCH

☒ Yes, I accept the Cloudera Standard License Terms and Conditions.

Continue

8. Click Continue.

The **Add Cluster - Installation** page > **Welcome** section appears. The steps on the left indicate where you are in the workflow.

Note: The next screen includes the KDC setup for the Cloudera Manager. Continue to the next section to configure the same KDC setup as the authentication provider to OneFS.

4.5 Kerberizing Cloudera Manager

This section details a crucial step that deviates from the Cloudera-recommended Kerberization process. Contrary to the procedure in [step 1](#), you must first Kerberize the OneFS cluster and Cloudera Manager installed with PowerScale CSD.

After you have installed Cloudera Manager, log in to Cloudera Manager to access the **Add Cluster - Installation** wizard. The wizard displays a warning to set up KDC. To set up the new KDC, click the link **here to setup a KDC** shown in Figure 5.

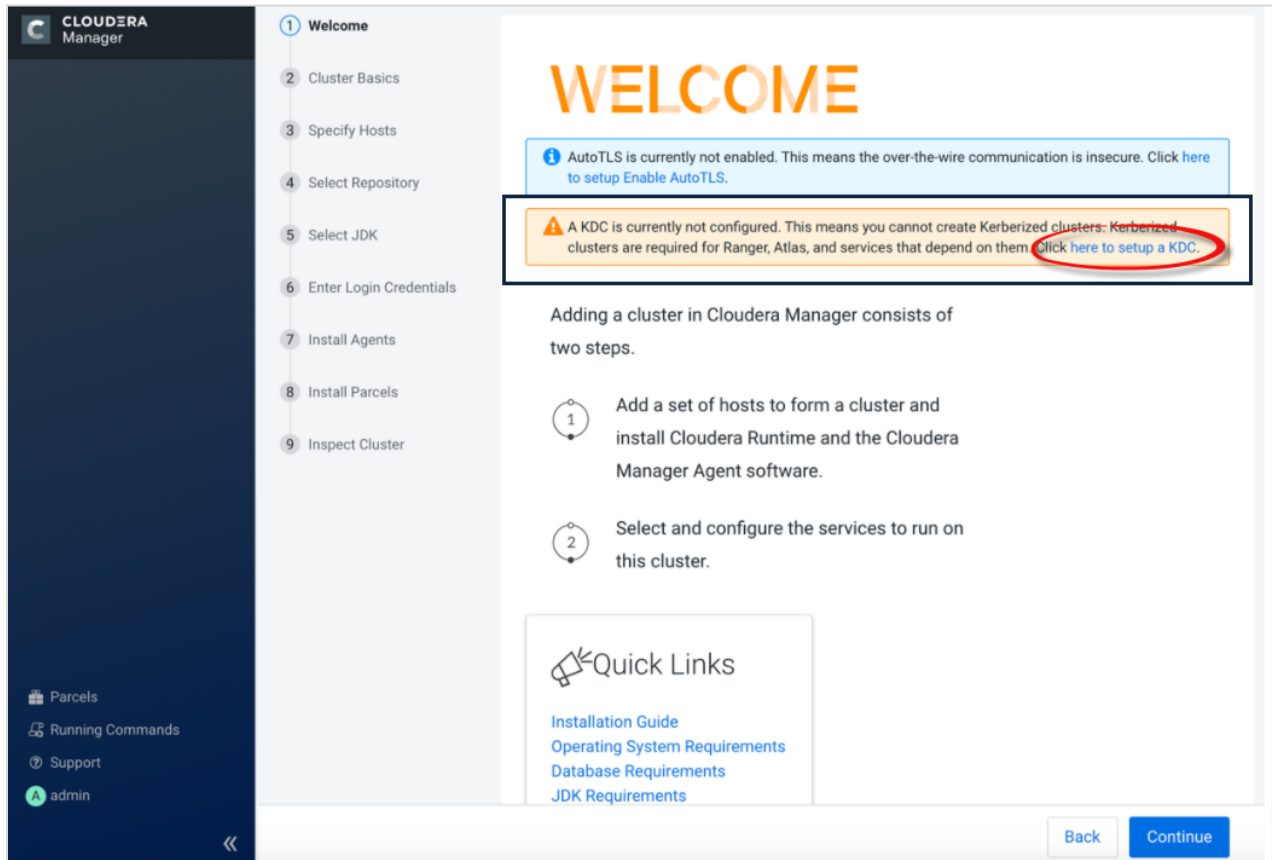


Figure 5 CDP Cluster setup wizard

4.5.1 Enabling Active Directory or MIT KDC as an authentication provider

After you have met the prerequisite requirements, you can Kerberize the Cloudera cluster. We recommend that you suspend all client and user activity on the Cloudera Manager before starting any Kerberization tasks.

You must use the same AD KDC or MIT KDC setup as authentication provided to OneFS, which is described in section 3.2 or section 3.3.

4.5.2 Set up KDC for Cloudera Manager

The **Setup KDC for this Cloudera Manager** wizard steps through the process to configure Cloudera Manager for Kerberos authentication.

1. In the Kerberos wizard, a **Getting Started** page appears. Select the applicable **KDC Type** to display configuration steps for your specific type of KDC. When you have completed all configuration steps, check the **I have completed all the above steps** check box, and click **Continue**.

Setup KDC for this Cloudera Manager

1 Getting Started

2 Enter KDC Information

3 Manage krb5.conf

4 Enter Account Credentials

5 Command Details

Getting Started

This wizard walks you through the steps to configure Cloudera Manager for Kerberos authentication.

Before using the wizard, ensure that you have performed the following steps:

1. Read the [documentation](#) about enabling Kerberos.
2. Set up a working KDC (Key Distribution Center) and specify the **KDC Type**:

KDC Type

☒ MIT KDC

☐ Active Directory

☐ Red Hat IPA

[Undo](#)

If OneFS is setup with MIT KDC

If OneFS is setup with AD KDC

Type of KDC used for authentication in CDH clusters

3. Configure the KDC to have **non-zero ticket lifetime and renewal lifetime**. Clusters will not work properly if tickets are not renewable.
4. Configure the KDC to have an account that has **permissions to create other accounts**.
5. Install OpenLdap client libraries on the **Cloudera Manager Server host** if you want to use Active Directory.

6.

```
# RHEL / CentOS
$ yum install openldap-clients krb5-workstation krb5-libs
```

```
# SUSE
$ zypper install openldap2-client krb5-client
```

```
# Ubuntu
$ apt-get install ldap-utils krb5-user
```

☒ I have completed all the above steps.

[Cancel](#) [Back](#) [Continue](#)

2. Enter KDC information: Enter the configuration information for the KDC that you are using.
 - If you are using AD and have multiple domain controllers behind a load balancer, enter the name of the Load Balancer in the **KDC Server Host** field. Also, enter any of the domain controllers in **Active Directory Domain Controller Override**. Hadoop daemons use the load balancer for authentication, but Cloudera Manager uses the override for creating accounts.
 - If you have multiple domain controllers (in case of AD) or MIT KDC servers, only enter the name of one controller in the **KDC Server Host** field. Cloudera Manager uses that server only for creating accounts. If you use Cloudera Manager to manage krb5.conf, you can specify the rest of the Domain Controllers using Safety Valve, as explained below.
 - Ensure the entries for the Kerberos Encryption Types field matches what your KDC supports.
 - If you are using an Active Directory KDC, you can configure Active Directory account properties such as objectClass and accountExpires directly from the Cloudera Manager UI. You can also enable Cloudera Manager to delete existing AD accounts so that new ones can be created when Kerberos credentials are being regenerated.

Note: To use AES encryption, ensure that you have deployed JCE Unlimited Strength Policy File, which is automatically included in Open JDK 1.8 232 (provided by Cloudera at the time of install) and up. However, this file may not be available on some of the earlier JDK 1.8 releases.

The following Active Directory KDC example shows values for the Kerberos Security Realm, the KDC Server Host, and the Active Directory Suffix. The example also selects the Active Directory Delete Accounts on Credential Regeneration check box.

Setup KDC for this Cloudera Manager

- Getting Started
- Enter KDC Information**
- Manage krb5.conf
- Enter Account Credentials
- Command Details

Enter KDC Information

Specify information about the KDC. The properties below are used by Cloudera Manager to generate principals for daemons running on the cluster.

Kerberos Encryption Types	rc4-hmac	Encryption types supported by KDC. Note: To use AES encryption, make sure you have deployed JCE Unlimited Strength Policy File by following the instructions here .
Kerberos Security Realm default_realm	FOO.COM	The realm to use for Kerberos security. Note: Changing this setting would clear up all existing credentials and keytabs from Cloudera Manager.
KDC Server Host kdc	hop-russ-win201.foo.com	Host where the KDC server is located. Port number is optional and can be provided as hostname[port]
KDC Admin Server Host admin_server	hop-russ-win201.foo.com	Host where the KDC Admin server is located. Port number is optional and can be provided as hostname[port]
Domain Name(s)		
Maximum Renewable Life for Principals	5 day(s)	

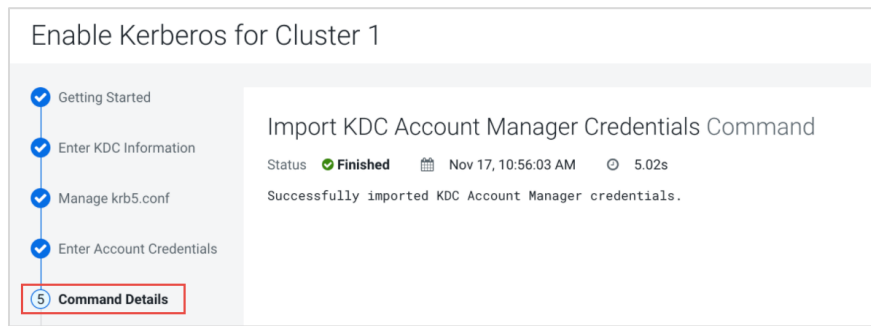
Cancel

Back Continue

3. Manage krb5.conf: You can use this page to specify if Cloudera Manager deploys and manages the krb5.conf file on your cluster.
 - If you select the Manage krb5.conf through Cloudera Manager check box, you can use this page to configure the krb5.conf file properties. In particular, the safety valves on this page can be used to configure cross-realm authentication.

- If left unchecked, you must ensure that the krb5.conf is deployed on all hosts in the cluster, including the Cloudera Manager Server host.
4. Click **Continue** to proceed.
 5. Enter account credentials: Enter the username and password for the user that can create principals for CDP cluster in the KDC. This user is the user or principal that you created in step 3 in section 4.3.1. Cloudera Manager encrypts the user name and password into a keytab and uses it as needed to create principals.
 6. Click **Continue** to proceed.

7. Command details: The Command Details page displays the outcome of the Enter Account Credentials step. Click **Continue** to proceed.



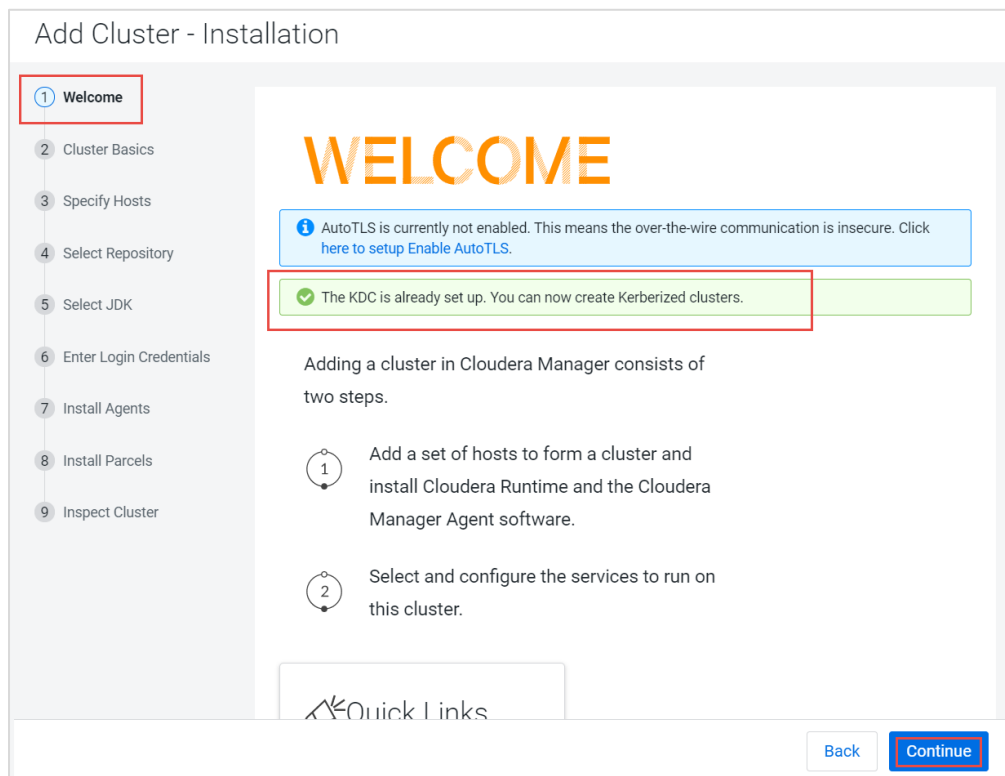
4.6 Installing Cloudera Runtime

After you have installed Cloudera Manager, log in to Cloudera Manager to access the **Add Cluster - Installation** wizard. Here, you add hosts to form a cluster and install Cloudera Runtime and Cloudera Manager Agent software.

Before you begin, complete the following:

- Install Cloudera Manager.
- Install PowerScale CSD.
- Install the Cloudera license.
- Set up the same KDC as the authentication provider to OneFS and Cloudera Manager.

1. After you have set up the KDC, the welcome page confirms the setup status.



2. Click Continue. The wizard goes to the **Add Cluster** step of the Cloudera Runtime installation.
3. Enter a name for the cluster, and click **Continue**.

The screenshot shows the 'Add Cluster - Installation' wizard. On the left is a vertical sidebar with steps 1 through 9. Step 2, 'Cluster Basics', is highlighted with a red box. The main area is titled 'Cluster Basics' and contains a 'Cluster Name' field with the text 'Streams Trial' inside, also highlighted with a red box. Below the field is an icon representing a cluster of nodes. Underneath the icon is the heading 'Regular Cluster' followed by a description: 'A Regular Cluster contains storage nodes, compute nodes, and other services such as metadata and security collocated in a single cluster.' At the bottom right of the wizard are two buttons: 'Back' and 'Continue', with the 'Continue' button highlighted by a red box.

The **Specify Hosts** section appears.

4. Enter the cluster host names or IP addresses in the **Hostnames** field.

Add Cluster - Installation

1 Welcome

2 Cluster Basics

3 Specify Hosts

4 Select Repository

5 Select JDK

6 Enter Login Credentials

7 Install Agents

8 Install Parcels

9 Inspect Cluster

Specify Hosts

Hosts should be specified using the same hostname (FQDN) that they will identify themselves with. Cloudera recommends including Cloudera Manager Server's host. This also enables health monitoring for that host.

Hostname

ccycloud-1.streams-trial.root.hwx.site
ccycloud-2.streams-trial.root.hwx.site
ccycloud-3.streams-trial.root.hwx.site

Hint: Search for hostnames or IP addresses using [patterns](#) [icon]

SSH Port: 22 [Search](#)

5. Specify the hostname and IP address ranges as shown in Table 4:

Table 4 Hostname and IP address ranges

Expansion range	Matching hosts
10.1.1.[1-4]	10.1.1.1, 10.1.1.2, 10.1.1.3, 10.1.1.4
Host[1-3].example.com	host1.example.com, host2.example.com, host3.example.com
Host[07-10].example.com	host07.example.com, host08.example.com, host09.example.com, host10.example.com

6. Click **Search**. Cloudera Manager discovers the hosts.

Add Cluster - Installation

- Welcome
- Cluster Basics
- 3 Specify Hosts**
- 4 Select Repository
- 5 Select JDK
- 6 Enter Login Credentials
- 7 Install Agents
- 8 Install Parcels
- 9 Inspect Cluster

Specify Hosts

Hosts should be specified using the same hostname (FQDN) that they will identify themselves with. Cloudera recommends including Cloudera Manager Server's host. This also enables health monitoring for that host.

Hostname

ccycloud-1.streams-trial.root.hwx.site
ccycloud-2.streams-trial.root.hwx.site
ccycloud-3.streams-trial.root.hwx.site

Hint: Search for hostnames or IP addresses using [patterns](#)

SSH Port: 22 **Search**

3 hosts scanned, 3 running SSH.
Click the first checkbox, hold down the Shift key and click the last checkbox to select a range.

<input checked="" type="checkbox"/>	Expanded Query ↑	Hostname (FQDN)	IP Address	Currently Managed	Result
<input checked="" type="checkbox"/>		ccycloud-1.streams-trial.root.hwx.site	172.27.123.204	No	Host was successfully scanned.
<input checked="" type="checkbox"/>		ccycloud-2.streams-trial.root.hwx.site	172.27.26.143	No	Host was successfully scanned.
<input checked="" type="checkbox"/>		ccycloud-3.streams-trial.root.hwx.site	172.27.92.198	No	Host was successfully scanned.

[Back](#)
[Continue](#)

7. Verify the host entries. Clear any entries that you are not installing services on, and click **Continue**.

The **Select Repository** section appears.

8. Select the following options:
- Public Cloudera Repository
 - Use Parcels
 - The version of Cloudera Runtime to install
 - Additional Parcels: None

Add Cluster - Installation

- Welcome
- Cluster Basics
- Specify Hosts
- Select Repository**
- Select JDK
- Enter Login Credentials
- Install Agents
- Install Parcels
- Inspect Cluster

Select Repository

Cloudera Manager Agent

Cloudera Manager Agent 7.1.3 (#4999720) needs to be installed on all new hosts.

Repository Location ☒ Public Cloudera Repository

Ensure the above version is listed in <https://archive.cloudera.com/cm7> and that you have access to that repository. Requires direct Internet access on all hosts.

☐ Custom Repository

CDH and other software

Cloudera recommends the use of parcels for installation over packages, because parcels enable Cloudera Manager to easily manage the software on your cluster, automating the deployment and upgrade of service binaries. Electing not to use parcels will require you to manually upgrade packages on all hosts in your cluster when software updates are available, and will prevent you from using Cloudera Manager's rolling upgrade capabilities.

Install Method ☐ Use Packages ☒ Use Parcels (Recommended) Parcel Repository

Version Versions that are too new for this version of Cloudera Manager (7.1.3) will not be shown.

☒ Cloudera Runtime [REDACTED]

☐ CDH 6.3.2-1.cdh6.3.2.p0.1605554

☐ CDH 5.16.2-1.cdh5.16.2.p0.8

Additional Parcels ☐ ACCUMULO 1.9.2-1.ACCUMULO6.1.0.p0.908695

☐ ACCUMULO 1.7.2-5.5.0.ACCUMULO5.5.0.p0.8

☒ None

Back
Continue

Cloudera QATS certified Cloudera
Runtime is 7.1.6-
1.cdh7.1.6.p0.9266103

- Click **Continue**. The **Select JDK** section appears.

10. Select Install a Cloudera-provided version of OpenJDK.

Add Cluster - Installation

1 Welcome
2 Cluster Basics
3 Specify Hosts
4 Select Repository
5 Select JDK
6 Enter Login Credentials
7 Install Agents
8 Install Parcels
9 Inspect Cluster

Select JDK

Selected Version	Cloudera Runtime 7.1
Supported JDK Version	OpenJDK 8, 11 or Oracle JDK 8, 11

[More details on supported JDK version.](#)

If you plan to use JDK 11, you will need to install it manually on all hosts and then select the **Manually manage JDK** option below.

☐ Manually manage JDK

! Please ensure that a supported JDK is **already installed** on all hosts. You will need to manage installing the unlimited strength JCE policy file, if necessary.

☒ **Install a Cloudera-provided version of OpenJDK**
By proceeding, Cloudera will install a supported version of OpenJDK version 8.

☐ **Install a system-provided version of OpenJDK**
By proceeding, Cloudera will install the default version of OpenJDK version 8 provided by the Operating System.

[Back](#) [Continue](#)

11. Click **Continue**. The Enter Login Credentials section appears.

12. Perform the following:

- a. For **Login To All Hosts As**, click **root**.
- b. For Authentication Method, click **All hosts accept same password**.
- c. Enter the password for the account that allows root access to your hosts.
- d. Click **Continue**.

Add Cluster - Installation

✓ Welcome

✓ Cluster Basics

✓ Specify Hosts

✓ Select Repository

✓ Select JDK

6 Enter Login Credentials

7 Install Agents

8 Install Parcels

9 Inspect Cluster

Enter Login Credentials

Root access to your hosts is required to install the Cloudera packages. This installer will connect to your hosts via SSH and log in either directly as root or as another user with password-less sudo/pbrun privileges to become root.

Login To All Hosts As: ☒ root
☐ Another user

You may connect via password or public-key authentication for the user selected above.

Authentication Method: ☒ All hosts accept same password
☐ All hosts accept same private key

Enter Password:

Confirm Password:

SSH Port:

Number of Simultaneous Installations:
(Running a large number of installations at once can consume large amounts of network bandwidth and other system resources)

Back

Continue

13. The Install Agents section displays the progress of the installation.

The screenshot shows the 'Add Cluster - Installation' wizard with the 'Install Agents' step selected. The left sidebar lists steps 1 through 9, with 'Install Agents' (step 7) highlighted. The main content area shows 'Install Agents' with a progress bar and a table of host installation progress.

Add Cluster - Installation

Install Agents

Installation in progress.

0 of 3 host(s) completed successfully. [Abort Installation](#)

Hostname	IP Address	Progress	Status
ccycloud-1.streams-trial.root.hwx.site	172.27.123.204	<div><div></div></div>	Installing openjdk8 package... Details
ccycloud-2.streams-trial.root.hwx.site	172.27.26.143	<div><div></div></div>	Installing openjdk8 package... Details
ccycloud-3.streams-trial.root.hwx.site	172.27.92.198	<div><div></div></div>	Installing openjdk8 package... Details

14. After the agents are installed, the Install Parcels section displays the progress of the parcel installation.

The screenshot shows the 'Add Cluster - Installation' wizard with the 'Install Parcels' step selected. The left sidebar lists steps 1 through 9, with 'Install Parcels' (step 8) highlighted. The main content area shows 'Install Parcels' with progress bars for 'Cloudera Runtime 7.1'.

Add Cluster - Installation

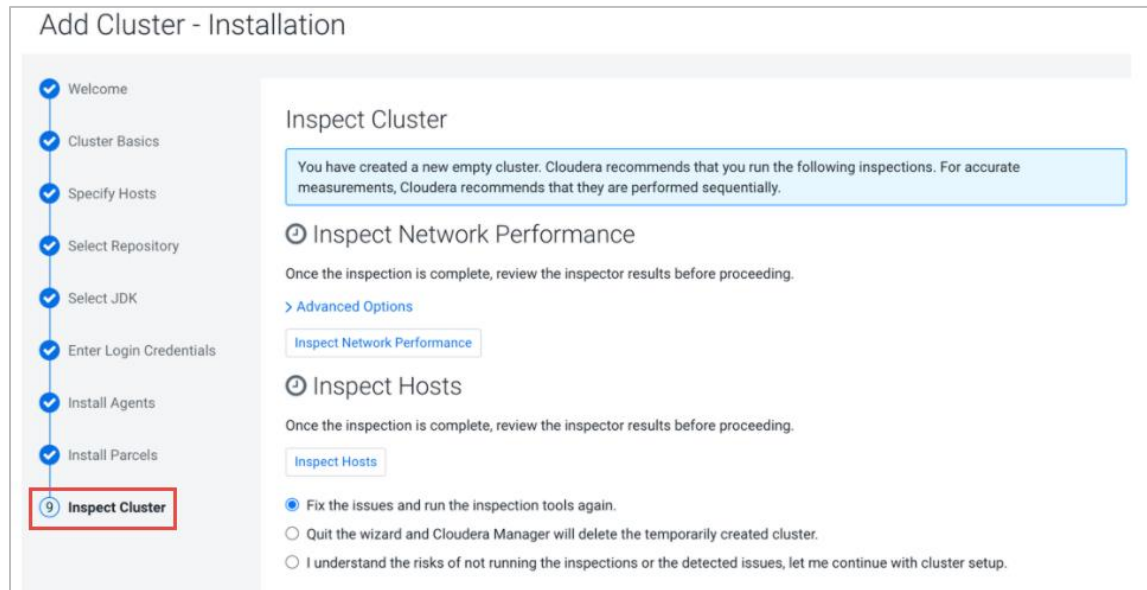
Install Parcels

The selected parcels are being downloaded and installed on all the hosts in the cluster.

▼ Cloudera Runtime 7.1 ... Downloaded: 3% Distributed: 0/0 Unpacked: 0/0 Activated: 0/0

[Back](#) [Continue](#)

15. After the parcels are installed, the **Inspect Cluster** section appears.



16. Complete the following steps:

- Click **Inspect Network Performance**.
- Optionally, click **Advanced Options** to customize ping parameters.
- After the network inspector completes, click **Show Inspector Results** to view the results in a new tab.
- Address any reported issues, and click **Run Again**.
- Click **Inspect Hosts**.
- After the host inspector completes, click **Show Inspector Results** to view the results in a new tab.
- Address any reported issues, and click **Run Again**.

Add Cluster - Installation

Inspect Cluster

You have created a new empty cluster. Cloudera recommends that you run the following inspections. For accurate measurements, Cloudera recommends that they are performed sequentially.

Inspect Network Performance

> Advanced Options

Status ✓ Last Run a few seconds ago Duration 18.11s [Show Inspector Results](#)

[Run Again](#) [More](#)

Inspect Hosts

No issues were detected, review the inspector results to see what checks were performed.

Status ✓ Last Run a few seconds ago Duration 18.48s [Show Inspector Results](#)

[Run Again](#) [More](#)

[Back](#) [Continue](#)

17. Click **Continue**. The **Add Cluster - Configuration** page appears. Select **Custom Services**, which is where the PowerScale CSD is exposed.

Add Cluster - Configuration

1 Select Services

Select Services

Choose a combination of services to install.

☐ **Data Engineering**
Process, develop, and serve predictive models.
Services: HDFS, YARN, YARN Queue Manager, Ranger, Atlas, Hive, Hive on Tez, Spark, Oozie, Hue, and Data Analytics Studio

☐ **Data Mart**
Browse, query, and explore your data in an interactive way.
Services: HDFS, Ranger, Atlas, Hive, Impala, and Hue

☐ **Operational Database**
Real-time insights for modern data-driven business.
Services: HDFS, Ranger, Atlas, and HBase

☐ **Custom Services**
Choose your own services. Services required by chosen services will automatically be included.

This wizard will also install the **Cloudera Management Service**. These are a set of components that enable monitoring, reporting, events, and alerts; these components require databases to store information, which will be configured on the next page.

[Cancel](#) [← Back](#) [Continue →](#)

4.7 Setting up a cluster using the wizard

After you complete the Add Cluster – Installation wizard for Cloudera Runtime, the Add Cluster – Configuration wizard automatically starts. The following sections Guide you through each page of the wizard.

Before you begin, in the **Add Cluster – Configuration** page, select **Custom Service**.

4.7.1 Select Services

The **Select Services** page allows you to select the services to install and configure. Under **Customer Services**, click **Dell EMC PowerScale** as replacement for the HDFS service. Also, select the appropriate services according to your requirements, or you can add services later using the **Add Services** feature. After you select the services to add, click **Continue**. The **Assign Roles** page displays.

Add Cluster - Configuration

1 Select Services

2 Assign Roles

3 Setup Database

4 Enter Required Parameters

5 Review Changes

6 Command Details

7 Summary

Select Services

Choose a combination of services to install.

☐ **Data Engineering**
Process, develop, and serve predictive models.
Services: HDFS, YARN, YARN Queue Manager, Ranger, Atlas, Hive, Hive on Tez, Spark, Oozie, Hue, and Data Analytics Studio

☐ **Data Mart**
Browse, query, and explore your data in an interactive way.
Services: HDFS, Ranger, Atlas, Hive, Impala, and Hue

☐ **Operational Database**
Real-time insights for modern data-driven business.
Services: HDFS, Ranger, Atlas, and HBase

☒ **Custom Services**
Choose your own services. Services required by chosen services will automatically be included.

Service Type	Description
<input type="checkbox"/> Atlas	Apache Atlas provides a set of metadata management and governance services that enable you to find, organize, and manage data assets. <i>This service requires Kerberos.</i>
<input checked="" type="checkbox"/> Core Configuration	Core Configuration contains settings used by most services. Required for clusters without HDFS.
<input type="checkbox"/> Cruise Control	Cruise Control simplifies the operation of Kafka clusters automating workload rebalancing and self-healing.
<input type="checkbox"/> Data Analytics Studio	Data Analytics Studio is the one stop shop for Apache Hive warehousing. Query, optimize and administrate your data with this powerful interface.
<input checked="" type="checkbox"/> DellEMC PowerScale	DellEMC PowerScale is a distributed scale-out filesystem.
<input type="checkbox"/> HBase	Apache HBase is a highly scalable, highly resilient NoSQL OLTP database that enables applications to leverage big data.
<input type="checkbox"/> HDFS	Apache Hadoop Distributed File System (HDFS) is the primary storage system used by Hadoop applications. HDFS creates multiple replicas of data blocks and distributes them on compute hosts throughout a cluster to enable reliable, extremely rapid computations.
<input checked="" type="checkbox"/> Hive	Apache Hive is a SQL based data warehouse system. In CDH 6 and earlier, this service includes Hive Metastore and HiveServer2. In Cloudera Runtime 7.0 and later, this service only includes the Hive Metastore; HiveServer2 and other components of the Hive

Cancel

Back
Continue

4.7.2 Assign Roles

The **Assign Roles** page suggests role assignments for the hosts in your cluster.

You can click the hostname for a role to select a different host. You can also click the **View By Host** button to see all the roles that are assigned to a host. For the PowerScale service, assign **Cloudera Manager Host** as the **Gateway** role.

After you assign all roles for your services, click **Continue**. The **Setup Database** page displays.

Add DellEMC PowerScale Service to Cluster 1

1 Assign Roles

2 Review Changes

3 Command Details

4 Summary

Assign Roles

You can customize the role assignments for your new service here, but note that if assignments are made incorrectly, such as assigning too many roles to a single host, performance will suffer.

You can also view the role assignments by host. [View By Host](#)

Gateway × 1 New

hop-kiran-n65.solarch.lab.com ▼

[Back](#) [Continue](#)

4.7.3 Setup Database

On the **Setup Database** page, you can enter the database hosts, names, usernames, and passwords that you created in step 5 in section 4.3.1.

For services that support it, you can add fine-grained customizations using a JDBC URL override.

Note: The Hive service is the only service that supports the JDBC URL override.

Select the database type, and enter the database name, username, and password for each service.

After you verify that each connection is successful, click **Continue**.

4.7.4 Enter Required Parameters

The Enter Required Parameters page lists required parameters for DAS, the Cloudera Manager API client, Hive, and Ranger.

4.7.5 Review Changes

The Review Changes page lists the default and suggested settings for several configuration parameters including data directories.

Review and make any necessary changes, and click **Continue**. The **Command Details** page displays.

Add DellEMC PowerScale Service to Cluster 1

Review Changes

HDFS Block Size DellEMC PowerScale (Service-Wide) ⓘ

dfs.blocksize
dfs_block_size 128 MiB

Default File System URI DellEMC PowerScale (Service-Wide) Undo ⓘ

default_fs_name
default_fs_name hdfs://cascade-system.foo.com:8020

WebHDFS URL DellEMC PowerScale (Service-Wide) Undo ⓘ

webhdfs_url
webhdfs_url http://cascade-system.foo.com:8082/webhdfs/v1

Hadoop TLS/SSL Server Keystore File Location DellEMC PowerScale (Service-Wide) ⓘ

ssl.server.keystore.location
ssl_server_keystore_location

Hadoop TLS/SSL Server Keystore File Password DellEMC PowerScale (Service-Wide) ⓘ

ssl.server.keystore.password
ssl_server_keystore_password

Hadoop TLS/SSL Server DellEMC PowerScale (Service-Wide) ⓘ

Back Continue

4.7.6 Command Details

The **Command Details** page lists the details of the First Run command.

You can expand the running commands to view the details of any step including log files and command output. To filter the view, you can select **Show All Steps**, **Show Only Failed Steps**, or **Show Only Running Steps**.

After the First Run command completes, click **Continue** to go to the **Summary** page.

If cluster deployment fails, fix any issues and click **Resume** in the wizard. If you do not click **Resume**, the Ranger service does not enable all necessary plugins.

Add DellEMC PowerScale Service to Cluster 1



✓ Assign Roles

✓ Review Changes

3 Command Details

4 Summary

First Run Command

Status **✓ Finished** Context [DellEMC PowerScale](#)  Apr 4, 1:03:07 PM 

20.65s

Finished First Run of the following services successfully: DellEMC PowerScale.

✓ Completed 1 of 1 step(s).

☒ Show All Steps ☐ Show Only Failed Steps ☐ Show Only Running Steps

> ✓ Run a set of services for the first time.

Apr 4, 1:03:07 PM 20.64s

Back

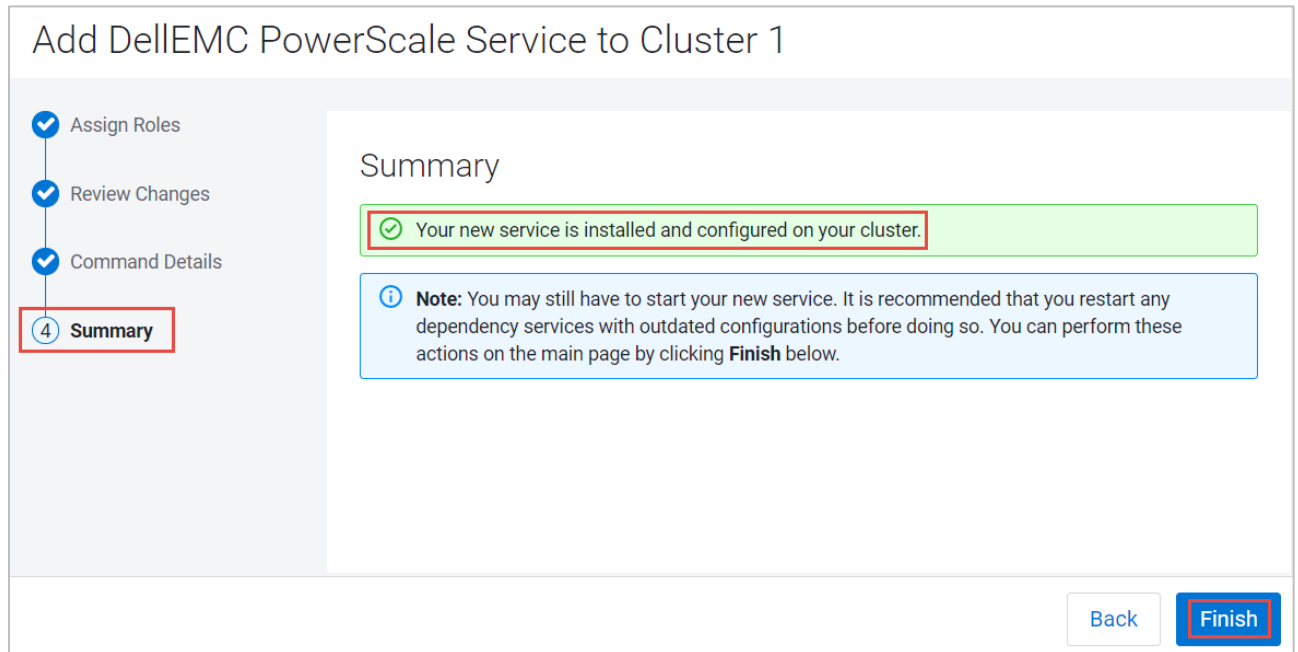
Continue

4.7.7 Summary

The **Summary** page reports the success or failure of the setup wizard.

Click **Finish** to complete the wizard. The installation is complete.

Cloudera recommends that you change the default password as soon as possible. Click the logged-in username at the upper-right corner of the home screen and click **Change Password**.



The custom services and PowerScale CSD installation is successful.

The screenshot displays the Cloudera Manager interface. On the left is a dark sidebar with navigation links: Clusters, Hosts, Diagnostics, Audits, Charts, Replication, Administration, Private Cloud (marked as New), Parcels, Running Commands, Support, and a user profile for 'admin'. The main content area is titled 'Status' and shows a list of Cloudera Runtime 7.1.6 (Parcels). The 'Dell EMC PowerScale' entry is highlighted with a red rectangle; it has a grey status icon, a wrench icon with the number 7, and a blue document icon. Other entries include 6 Hosts, Atlas, CDP-INFRA-SOLR, HBase, Hive, Hive on Tez, Hue, Kafka, Knox, Oozie, Phoenix, Ranger, Spark, and Sqoop. On the right, there are two charts: 'Cluster CPU' showing a flat line at 0% and 'Cluster Network IO' showing a line graph with peaks around 48.8K/s. The top navigation bar includes 'Status', 'Health Issues', and 'Configuration' with a notification badge for 15 items.

Parcel	Status	Issues	Actions
Cloudera Runtime 7.1.6 (Parcels)			
6 Hosts	✓	1	
Atlas	✓		
CDP-INFRA-SOLR	✓		
Dell EMC PowerScale	●	7	
HBase	✓	7	
Hive	✓		
Hive on Tez	✓		
Hue	✓		
Kafka	✓		
Knox	✓		
Oozie	✓		
Phoenix	✓		
Ranger	✓		
Spark	✓		
Sqoop	●		

4.8 Other steps for Apache Ranger

After you install Cloudera Manager and add a cluster, you must perform some remaining steps to complete the installation of Apache Ranger on PowerScale.

Note: The Apache Ranger 2.0 had an issue which is fixed in [RANGER-2626](#). We have provided a OneFS 8.2.2 custom patch for the [RANGER-2626](#) issue for the QATS testing. The same patch will be officially released as a **[PSP-1091]** Roll Up Patch in May 2021. Ensure that you install the OneFS 8.2.2 official Roll Up Patch PSP-1091.

4.8.1 Steps to enable Ranger service on PowerScale

4.8.1.1 Enable Ranger plugin on PowerScale HDFS settings

1. Log in into PowerScale OneFS UI.
2. Under HDFS protocol settings, select the **Ranger plugin settings** tab.
3. Check **Enable Ranger plugin**, and provide the **Policy manager URL** and **Repository name** as shown in the following screen.

Note: Ensure that the Repository name in the OneFS Ranger setting and Ranger UI HDFS service manager plugin are the same.

The screenshot shows the OneFS Storage Administration web interface. The top navigation bar includes 'Dashboard', 'Cluster management', 'File system', 'Data protection', 'Access', and 'Protocols'. The 'Protocols' menu is expanded, showing 'Hadoop (HDFS)'. Under 'Hadoop (HDFS)', the 'Ranger plugin settings' tab is selected. The 'Current access zone' is set to 'cdpods2'. The 'Edit HDFS Ranger plugin settings' section shows the 'Enable Ranger plugin' checkbox checked. The 'Policy manager URL' is set to 'http://dev2-data4.cloudera.local:6080' and the 'Repository name' is set to 'cm_hdfs'. There are 'Revert changes' and 'Save changes' buttons at the bottom.

4.8.1.2 Enable Ranger plugin for all services if not enabled by default.

Log in into the Cloudera Manager UI, and check each service configuration to see if the Ranger plugin is enabled. If not, manually enable the plugin. See the example HBase service ranger service enabled as follows.

The screenshot shows the Cloudera Manager HBase Configuration page. The 'Configuration' tab is selected in the top navigation bar. On the left, the 'Filters' sidebar shows the 'SCOPE' list with 'HBase (Service-Wide)' selected. The main content area displays the configuration for the 'Ranger Service' (HBase (Service-Wide)). The 'Ranger' checkbox is checked, indicating the plugin is enabled. A red box highlights the 'Ranger Service' configuration row. Other services listed include ZooKeeper Session Timeout, HDFS Service, ZooKeeper Service, and Atlas Service.

Service	Configuration	Value	Notes
ZooKeeper Session Timeout	zookeeper.session.timeout	1200000	Max session timeout of ZooKeeper server zookeeper-SERVER-304ed78e2e4c8b9796326de14031c9ba should be at least HBase's ZooKeeper session timeout.
HDFS Service	hdfs_service	Del EMC PowerScale	
ZooKeeper Service	zookeeper_service	ZooKeeper	
Ranger Service	ranger_service	<input checked="" type="checkbox"/> Ranger	
Atlas Service	atlas_service	Atlas	

4.8.1.3 Mapping Kerberos Principals to short names

Select the Dell EMC PowerScale service. Under configuration, search for **hadoop.security.auth_to_local**. Add the following mapping rules for the PowerScale HDFS file system.

```

RULE:[2:$1/$2@$0] (nn/.@.*IPA.ENG.HORTONWORKS.COM) s/./hdfs/
RULE:[2:$1/$2@$0] (dn/.@.*IPA.ENG.HORTONWORKS.COM) s/./hdfs/
RULE:[1:$1@$0] (hdfs@IPA.ENG.HORTONWORKS.COM) s/@.*//
RULE:[1:$1@$0] (.@IPA.ENG.HORTONWORKS.COM) s/@.*//
RULE:[2:$1@$0] (rangeradmin@IPA.ENG.HORTONWORKS.COM) s/(.*)@IPA.ENG.HORTONWORKS.COM/ranger/
RULE:[2:$1@$0] (rangertagsync@IPA.ENG.HORTONWORKS.COM) s/(.*)@IPA.ENG.HORTONWORKS.COM/rangertagsync/
RULE:[2:$1@$0] (rangerusersync@IPA.ENG.HORTONWORKS.COM) s/(.*)@IPA.ENG.HORTONWORKS.COM/rangerusersync/
RULE:[2:$1@$0] (rangerkms@IPA.ENG.HORTONWORKS.COM) s/(.*)@IPA.ENG.HORTONWORKS.COM/keyadmin/
RULE:[1:$1@$0] (HTTP.*@IPA.ENG.HORTONWORKS.COM$) s/@IPA.ENG.HORTONWORKS.COM$/ /
RULE:[1:$1@$0] (. *@IPA.ENG.HORTONWORKS.COM$) s/@IPA.ENG.HORTONWORKS.COM$/ //L
DEFAULT
  
```

Note: Replace **IPA.ENG.HORTONWORKS.COM** with your Kerberos security realm.

The screenshot shows the Dell EMC PowerScale Configuration page. The 'Configuration' tab is selected, and the search bar contains 'auth_to_local'. The 'Filters' sidebar on the left shows the 'SCOPE' as 'Dell EMC PowerScale (Service-Wide)' and 'CATEGORY' as 'Security'. The main content area displays 'Trusted Kerberos Realms' with 'Dell EMC PowerScale (Service-Wide)' selected. A red box highlights the 'Additional Rules to Map Kerberos Principals to Short Names' section, which contains the following rules:

```

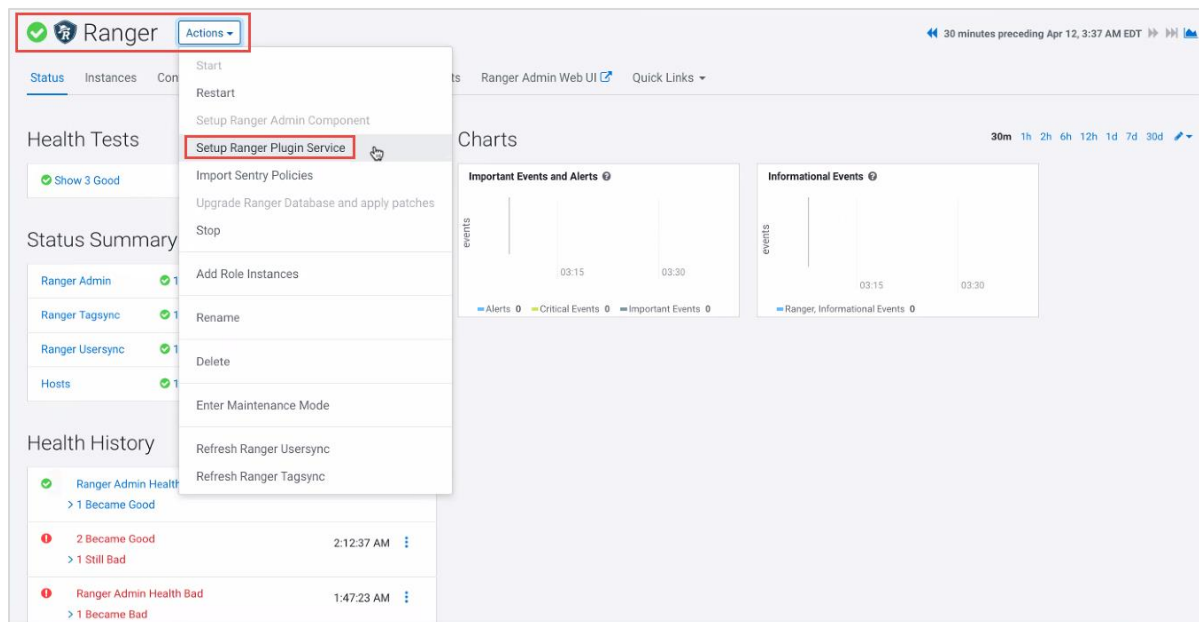
RULE:[2:$1@$0](rangerkms@IPA.ENG.HORTONWORKS.COM)s/(.*)@IPA.ENG.HORTONWORKS.COM/keyadmin/
RULE:[1:$1@$0](HTTP.*@IPA.ENG.HORTONWORKS.COM$)s/@IPA.ENG.HORTONWORKS.COM$/ /
RULE:[1:$1@$0](.*@IPA.ENG.HORTONWORKS.COM$)s/@IPA.ENG.HORTONWORKS.COM$/ //L
DEFAULT
  
```

The rules are displayed in a text area with an 'Undo' button and a 'Show All Descriptions' link. The page footer indicates '1 - 2 of 2'.

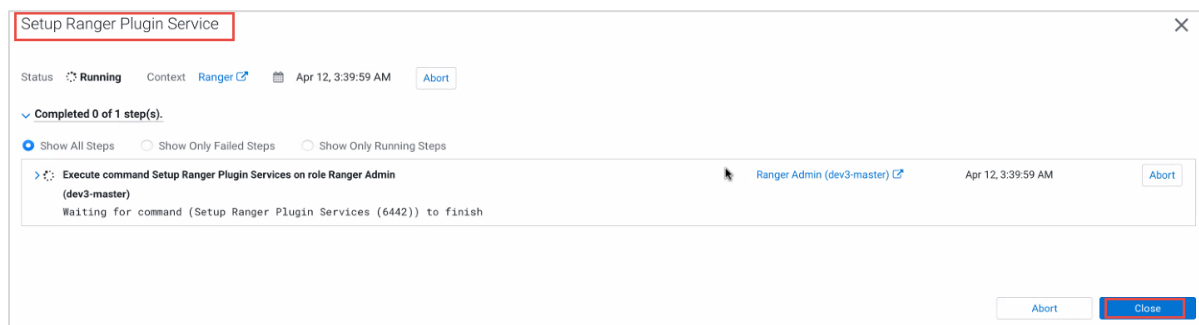
4.8.1.4 Set up Ranger Plugin service

In the Ranger UI, individual service plugins under Ranger Service Manager are not enabled.

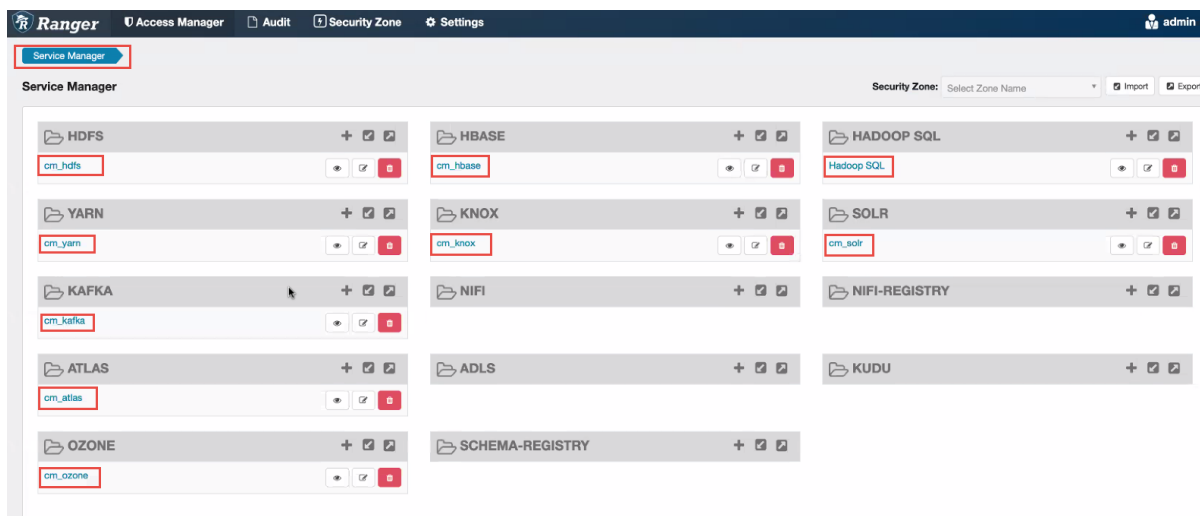
1. To enable the plugins, expand the **Active** tab of Ranger service on Cloudera Manager UI, and click **Setup Ranger Plugin Service**, as shown in the following image.



2. After the **Setup Ranger Plugin Service** operation completes, click **Close**.



- Log in to the Ranger UI, and verify all service plugins are enabled.



- Click the **Edit** button of the `cm_hdfs` HDFS plugin, ensure the following table properties are populated, and ensure that the **Test Connection** is successful.

Table 5 HDFS plugin service details

Plugin repository properties	MIT KDC	AD KDC
Service Details:		
Service Name	cm_hdfs	cm_hdfs
Display Name	cm_hdfs	cm_hdfs
Description	hdfs repo	hdfs repo
Active Status	Enabled	Enabled
Tag Service	cm_tag	cm_tag
Config Properties:		
Username	KDC admin user	KDC admin user
Password	KDC admin user password	KDC admin user password
Namenode URL	hdfs://POWERSCLAE_FQDN:8020	hdfs://POWERSCLAE_FQDN:8020
Authorization Enabled	False	False
Authentication Type	Kerberos	Kerberos
hadoop.security.auth_to_local	--	--
dfs.datanode.kerberos.principal	--	--
dfs.namenode.kerberos.principal	--	--
dfs.secondary.namenode.kerberos.principal	--	--
RPC Protection Type	Authentication	Authentication

Plugin repository properties	MIT KDC	AD KDC
Common Name for Certificate	--	--
Add New Configurations:		
tag.download.auth.users	hdfs	hdfs,FOO1\$
policy.download.auth.users	hdfs	hdfs,FOO1\$,PIPE1\$

5. Add a default public allow profile to the root file system path, and control the policies through **Deny** conditions.

See the document [Dell EMC Isilon: Apache Ranger Setup and Operations](#). This document describes the official support of Apache Ranger™ as a part of an Apache Hadoop deployment with a Dell EMC Isilon™ cluster. The paper also discusses best practices, implementation strategies, and limitations.

4.9 Completing post-installation steps

The following topics describe post-installation actions, such as deploying client configuration and performing simple tests to validate the installation and confirm that all components are functioning correctly.

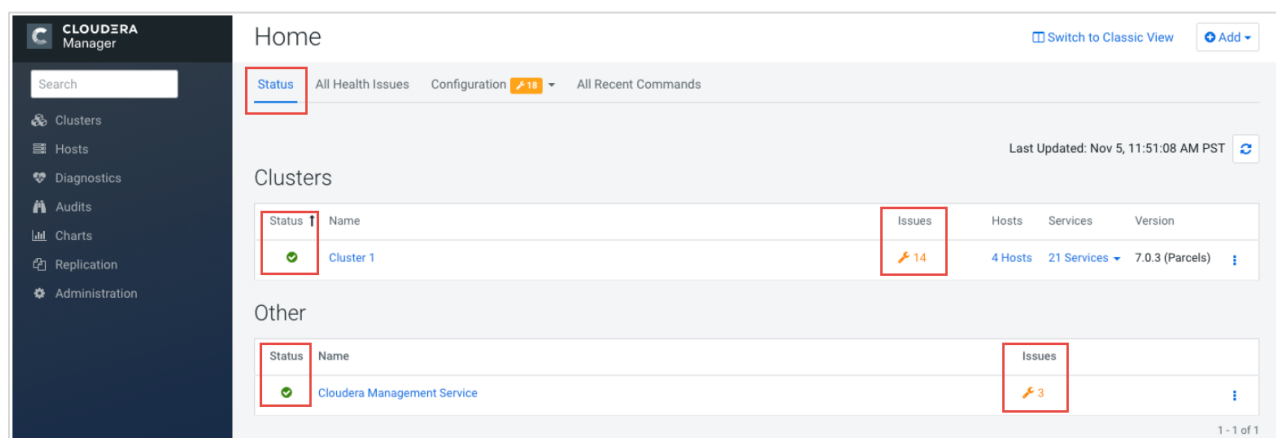
4.9.1 Deploying clients

Cloudera Manager automatically generates client configuration files based on the services you install. Cloudera Manager deploys these configurations automatically at the end of the installation workflow. You can also download the client configuration files to deploy them manually.

4.9.2 Testing the Installation

Begin testing the installation from the home page, starting with checking the health of the services.

To begin testing, start the **Cloudera Manager Admin Console**. After you have logged in, the home page appears as follows.



On the left side of the screen is a list of services that are running with their status information. All services should be running with **Good Health** (indicated by a green checkmark). Click each service to view more detailed information about each service. You can also test your installation by either checking each host's heartbeats, running a MapReduce job, or interacting with the cluster with an existing Hue application.

4.9.3 Checking host heartbeats

One way to check if all agents are running is to look at the time that has passed since their last heartbeat. Click the **Hosts** tab to see a list of all hosts, and view the values under **Last Heartbeat**.

4.9.4 Running a MapReduce Job

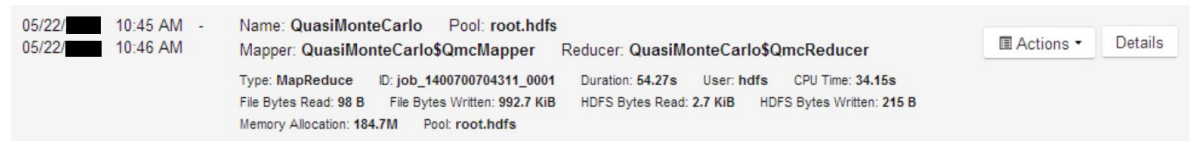
Run a PiEstimator job to manually verify that the CDP Private Cloud Base installation was successful.

Note: If you have a secure cluster, use the kinit command-line tool to authenticate to Kerberos.

1. Log in to a host in the cluster.
Run the Hadoop PiEstimator example using the following command.

```
yarn jar /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce-examples.jar pi 10 100
```
2. In Cloudera Manager, go to Cluster > ClusterName > yarn Applications.
3. Check the results of the job. It appears similar to the following example:

```
yarn jar /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar pi 10 100
```



4.9.5 Testing with Hue

Hue is a UI that enables you to interact with your clusters by running applications to browse HDFS, manage a Hive metastore, run Hive, Impala, and Search queries, and run Oozie workflows.

Test the cluster by running a Hue web application:

1. In the Cloudera Manager Admin Console Home > Status tab, click the Hue service.
2. Click the **Hue Web UI** link, which opens Hue in a new window.
3. Log in with the following credentials:
 - Username: hdfs
 - Password: hdfs
4. At the top of the window, choose an application in the navigation bar.

For more information, see the Hue documentation.

4.9.6 Securing Your cluster

After you complete your Cloudera Enterprise installation and ensure that everything is working properly, secure your cluster by enabling authentication, authorization, auditing, and encryption.

For comprehensive instructions about securing your cluster, see the Cloudera [Security Overview](#).

4.10 Uninstalling Cloudera Manager and Managed Software

Uninstalling Cloudera Manager and managed software is out of scope of this Guide. For more information, see the Cloudera document [Uninstalling Cloudera Manager and Managed Software](#).

A Troubleshooting installation problems

This section describes common installation issues and suggested solutions.

A.1 TLS protocol error with OpenJDK

If you are using an older version of OpenJDK 1.8 and have enabled SSL/TLS for the Cloudera Manager Admin Console, you may encounter a TLS protocol error. This error occurs when connecting to the Admin Console and states that there are no ciphers in common. This error occurs because older versions of OpenJDK may not implement certain TLS ciphers, causing an inability to log in to the Cloudera Manager Admin Console when TLS is enabled.

Workaround: Perform one of the following actions.

- Upgrade OpenJDK to a supported version of OpenJDK that is higher than version 1.8.0_181.
- If it is not possible to upgrade OpenJDK, enable less-secure TLS ciphers in Cloudera Manager. You can do this by opening the `/etc/default/cloudera-scm-server` in a text editor and adding the following line:

```
export CMF_OVERRIDE_TLS_CIPHERS=<cipher_list>
```

<cipher_list> is a list of TLS cipher suites separated by colons. For example:

```
export
CMF_OVERRIDE_TLS_CIPHERS="TLS_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256:TLS_ECDH
E_RSA_WITH_AES_128_GCM_SHA256:TLS_ECDHE_ECDSA_WITH_AES_256_GCM_SHA384:TLS_
ECDHE_RSA_WITH_AES_256_GCM_SHA384:TLS_DHE_RSA_WITH_AES_128_GCM_SHA256:TLS_
DHE_RSA_WITH_AES_256_GCM_SHA384:TLS_ECDHE_ECDSA_WITH_AES_128_CBC_SHA256:TL
S_ECDHE_RSA_WITH_AES_128_CBC_SHA256:TLS_ECDHE_ECDSA_WITH_AES_128_CBC_SHA:T
LS_ECDHE_RSA_WITH_AES_256_CBC_SHA384:TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA:TL
S_ECDHE_ECDSA_WITH_AES_256_CBC_SHA384:TLS_ECDHE_ECDSA_WITH_AES_256_CBC_SHA
:TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA:TLS_DHE_RSA_WITH_AES_128_CBC_SHA256:TL
S_DHE_RSA_WITH_AES_128_CBC_SHA:TLS_DHE_RSA_WITH_AES_256_CBC_SHA256:TLS_DHE
_RSA_WITH_AES_256_CBC_SHA:TLS_ECDHE_ECDSA_WITH_3DES_EDE_CBC_SHA:TLS_ECDHE_
RSA_WITH_3DES_EDE_CBC_SHA:TLS_EDH_RSA_WITH_3DES_EDE_CBC_SHA:TLS_RSA_WITH_A
ES_128_GCM_SHA256:TLS_RSA_WITH_AES_256_GCM_SHA384:TLS_RSA_WITH_AES_128_CBC
_SHA256:TLS_RSA_WITH_AES_256_CBC_SHA256:TLS_RSA_WITH_AES_128_CBC_SHA:TLS_R
SA_WITH_AES_256_CBC_SHA:TLS_RSA_WITH_3DES_EDE_CBC_SHA"
```

Cloudera Bug: OPSAPS-49578

A.2 Failed to start server reported by cloudera-manager-installer.bin

"Failed to start server" reported by `cloudera-manager-installer.bin`. `/var/log/cloudera-scm-server/cloudera-scm-server.log` contains a message beginning **Caused by:**
`java.lang.ClassNotFoundException: com.mysql.jdbc.Driver...`

Possible reason: You might have SELinux enabled.

Possible solution: Disable SELinux by running `sudo setenforce 0` on the Cloudera Manager Server host. To disable it permanently, edit `/etc/selinux/config`.

A.3 Installation interrupted and installer do not restart.

Possible reason: You must perform manual cleanup.

Possible solution: See [Uninstalling Cloudera Manager and Managed Software](#).

A.4 Cloudera Manager Server fails to start with MySQL

Cloudera Manager Server fails to start, and the Server is configured to use a MySQL database to store information about service configuration.

Possible reason: Tables might be configured with the ISAM engine. The Server does not start if its tables are configured with the MyISAM engine, and an error such as the following appears in the log file:

```
Tables ... have unsupported engine type ... . InnoDB is required.
```

Possible solution: Ensure that the InnoDB engine is configured, not the MyISAM engine. To check what engine your tables are using, run the following command from the MySQL shell: `mysql> show table status;`

For more information, see [Install and Configure MySQL for Cloudera Software](#).

A.5 Agents fail to connect to Server

Agents fail to connect to Server. You get an Error 113 ('No route to host') in `/var/log/cloudera-scm-agent/cloudera-scm-agent.log`.

Possible reason: You might have SELinux or iptables enabled.

Possible solution: Check `/var/log/cloudera-scm-server/cloudera-scm-server.log` on the Server host and `/var/log/cloudera-scm-agent/cloudera-scm-agent.log` on the Agent hosts. Disable SELinux and iptables.

A.6 Cluster hosts do not appear.

Some cluster hosts do not appear when you click Find Hosts in install or update wizard.

Possible reason: You might have network connectivity problems.

Possible solution:

- Ensure all cluster hosts have SSH port 22 open.
- Check other common causes of loss of connectivity such as firewalls and interference from SELinux.

A.7 "Access denied" in install or update wizard.

"Access denied" in install or update wizard during database configuration for Activity Monitor or Reports Manager.

Possible reason: Hostname mapping or permissions are not set up correctly.

Possible solution:

- For hostname configuration, see Configure Network Names.
- For permissions, ensure the values you enter into the wizard match the values that you used when you configured the databases. The value you enter into the wizard as the database hostname must match the value you entered for the hostname (if any) when you configured the database.

For example, if you had entered the following when you created the database.

```
grant all on activity_monitor.* TO 'amon_user'@'myhost1.myco.com'
IDENTIFIED BY 'amon_password';
```

The value you enter here for the database hostname must be myhost1.myco.com. If you did not specify a host, or used a wildcard to allow access from any host, you can enter either the fully qualified domain name (FQDN), or localhost. For example, if you entered

```
grant all on activity_monitor.* TO 'amon_user'@'%' IDENTIFIED BY
'amon_password';
```

The value you enter for the database hostname can be either the FQDN or localhost.

A.8 Databases fail to start.

Activity Monitor, Reports Manager, or Service Monitor databases fail to start.

Possible reason: MySQL binlog format problem.

Possible solution: Set `binlog_format=mixed` in `/etc/my.cnf`. For more information, see this [MySQL bug report](#). See also [Step 4. Install and Configure Databases](#).

A.9 Cloudera services fail to start

Possible reason: Java might not be installed or might be installed at a custom location.

Possible solution: See Configuring a Custom Java Home Location for more information about resolving this issue.

A.10 Activity Monitor displays a status of BAD.

The Activity Monitor displays a status of BAD in the Cloudera Manager Admin Console. The log file contains the following message:

```
ERROR 1436 (HY000): Thread stack overrun: 7808 bytes used of a 131072 byte
stack, and 128000 bytes needed.
```

Use `'mysqld -O thread_stack=#'` to specify a bigger stack.

Possible reason: The MySQL thread stack is too small.

Possible solution:

1. Update the `thread_stack` value in `my.cnf` to 256 KB. The `my.cnf` file is normally located in `/etc` or `/etc/mysql`.
2. Restart the `mysql` service: `$ sudo service mysql restart`
3. Restart Activity Monitor.

A.11 Activity Monitor fails to start

The Activity Monitor fails to start. Logs contain the error read-committed isolation not safe for the statement binlog format.

Possible reason: The `binlog_format` is not set to mixed.

Possible solution: Modify the `mysql.cnf` file to include the entry for `binlog format` as specified in Install and Configure MySQL for Cloudera Software.

A.12 Create Hive Metastore Database Tables command fails.

The Create Hive Metastore Database Tables command fails due to a problem with an escape string.

Possible reason: PostgreSQL versions 9 and higher require special configuration for Hive because of a backward-incompatible change in the default value of the `standard_conforming_strings` property. Versions up to PostgreSQL 9.0 defaulted to off, but starting with version 9.0 the default is on.

Possible solution: As the administrator user, use the following command to turn `standard_conforming_strings` off:

```
ALTER DATABASE <hive_db_name> SET standard_conforming_strings = off;
```

A.13 Oracle invalid identifier

If you are using an Oracle database and the Cloudera Navigator Analytics > Audit > Activity tab displays "No data available" and there is an Oracle error about "invalid identifier" with the query containing the reference to `dbms_crypto` in the log.

Possible reason: You have not granted the execute permission to `sys.dbms_crypto`.

Possible solution: Run `GRANT EXECUTE ON sys.dbms_crypto TO nav;`, where `nav` is the user of the Navigator Audit Server database.

A.14 Failed to upload Tez jar file during installation

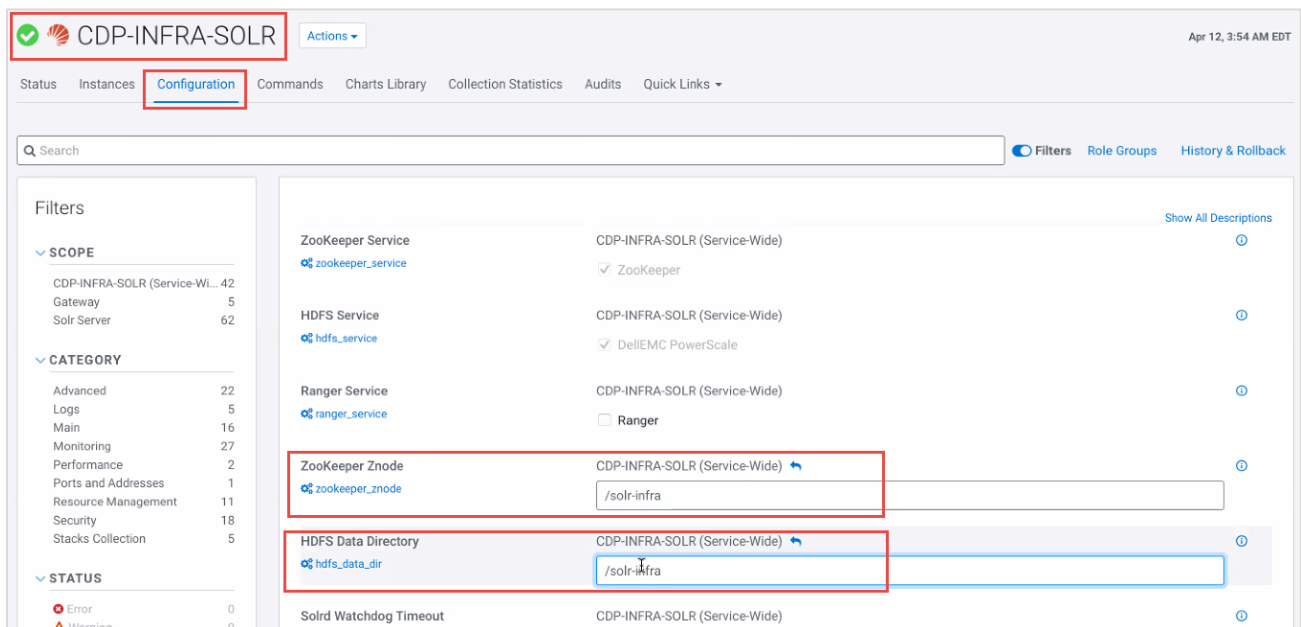
Manually upload the Tez jar from parcels folders into HDFS `/user/tez/$BUID_NUMBER`, example:
`/user/tez/ 0.9.1.7.1.6.0-297`

A.15 Zeppelin fails on first run

Check the error if it is `/var/lib/zeppelin` file folder mission on permission issue, then `chown zeppelin:zeppelin /var/lib/zeppelin`

A.16 SOLR service startup issue

If the SOLR service is installed along with all other services, “`zookeeper_znode`” and “`hdfs_data_dir`” will be either `/solr` or `/solr-infra`. Now, installing Ranger or Atlas service to the cluster will rename these two parameters, which fails the SOLR service from starting. Simply rename the `zookeeper_znode` and `hdfs_data_dir` from `/solr` to `solr-infra` or `/solr-infra` to `/solr`, and restart the service.



A.17 Ranger unauthenticated access issue on OneFS 8.2.2

OneFS 8.2.2 fails to download Ranger HDFS policy repository, throws error
`(org.apache.ranger.common.RESTErrorUtil: Request failed. loginId=null,
 logMessage=Unauthenticated access not allowed)` and at Isilon we see HTTP 400 error (`[hdfs]`)

Ranger: Request to URL http://dev2-data4.cloudera.local:6080/service/plugins/policies/download/cm_hdfs?lastKnownVersion=12 failed with HTTP code 400 for zone 5).

Possible reason: OneFS 8.2.2 bad for HTTP request. The Apache Ranger 2.0 had an issue which is fixed in [RANGER-2626](#).

Possible solution: Install “PSP-1091” official Roll Up Patch of May 2021 or later to the OneFS 8.2.2

B Technical support and resources

[Dell.com/support](https://dell.com/support) is focused on meeting customer needs with proven services and support.

[Storage and data protection technical documents and videos](#) provide expertise to ensure customer success with Dell EMC storage and data protection products.