

POWERSCALE ONEFS WITH HADOOP AND HORTONWORKS INSTALLATION GUIDE

8.1.2 – 9.0.0 with Ambari 2.7.x and later versions

Abstract

This guide walks you through the process of installing PowerScale OneFS with Hadoop for use with the Hortonworks Data Platform (HDP) 3.0.1 and later, and the Apache Ambari manager 2.7.1 and later.



Copyright © 2020 Dell Inc. or its subsidiaries. All rights reserved.

Dell believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS-IS.” DELL 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 DELL SOFTWARE DESCRIBED IN THIS PUBLICATION REQUIRES AN APPLICABLE SOFTWARE LICENSE.

Dell, EMC, and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be the property of their respective owners.

Published in the USA.

EMC Corporation
Hopkinton, Massachusetts 01748-9103
1-508-435-1000 In North America 1-866-464-7381
www.EMC.com

Publication History

Version	Date	Description
1.00	December 6, 2018	Added new version of the guide for OneFS 8.1.2.0 with Ambari 2.7.1.0 and HDP 3.0.1.0
2.00	August 19, 2019	Added OneFS 8.1.2 related content
3.00	October 15, 2019	HDP 3.1 certification has been completed.
4.00	June 16, 2020	Added OneFS 8.2 - 9.0 related content

Contents

Introduction	5
Audience.....	5
Overview.....	5
Updates and additional information about OneFS Hadoop installs.....	7
Prerequisites.....	7
Ambari and Hortonworks Data Platform	7
OneFS Multi-tenant Installations.....	7
OneFS cluster configuration	8
Installing OneFS with Ambari	9
Preparing OneFS.....	9
Validate OneFS version and license activation.....	9
Configure OneFS components.....	10
Create an access zone	11
Configure SmartConnect.....	12
Configure DNS for OneFS	13
Verify the SmartConnect configuration	13
Create HDFS users and groups	14
Create users on the OneFS cluster manually	14
Configure HDFS user for OneFS 8.1.2 and previous versions	15
Configure HDFS user for OneFS 8.2.0 and later versions.....	16
Preparing Ambari	18
Steps to perform on the Hadoop client	18
Configuring Ambari 2.7.1 and later.....	20
Addendum	37
Apache Ranger authorization policy integration	37
Deploy Apache Ranger.....	38
Ambari metrics and alerts overview	39
View Ambari metrics and alerts.....	40
HDFS wire encryption overview	43
Configure HDFS Wire Encryption with OneFS.....	43
Configure Apache Hadoop for HDFS wire encryption	44
Contacting Dell EMC PowerScale Technical Support.....	44

Introduction

Hadoop is an open-source framework that enables the distributed processing of large sets of data across clusters of systems. Hadoop clusters use the Hadoop Distributed File System (HDFS) to provide high-throughput access to application data. You can follow the steps in this guide to install PowerScale OneFS with Hadoop for use with the Hortonworks Data Platform (HDP) and the Apache Ambari manager.

Before you begin, you must install an PowerScale OneFS cluster.

Audience

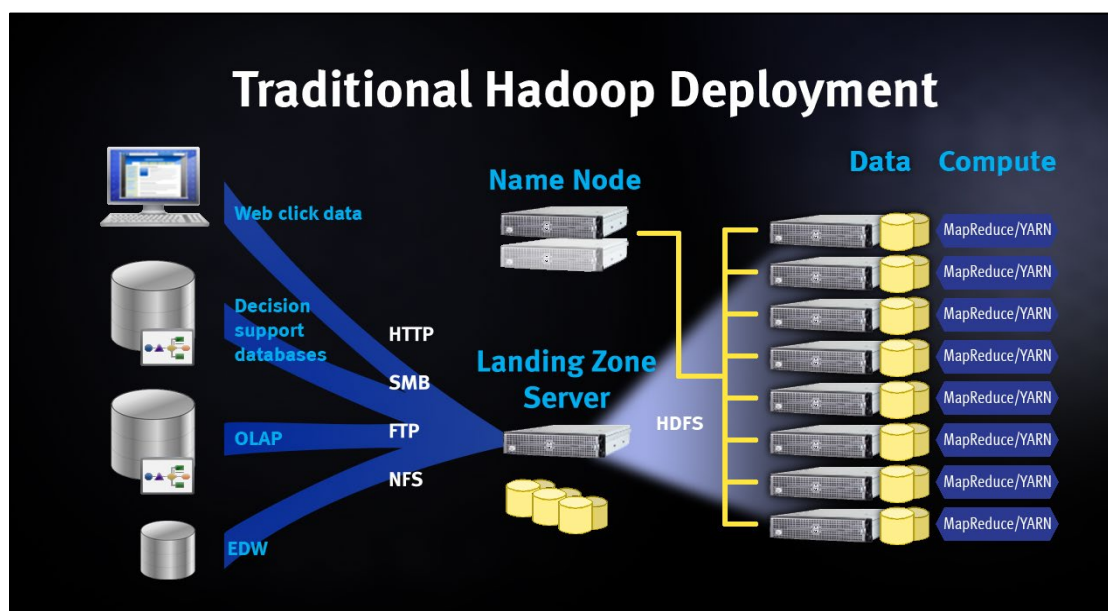
This guide is intended for systems administrators, IT program managers, IT architects, and IT managers who are installing OneFS 8.1.2.0 or later with Ambari 2.7.1.0 and later and HDP 3.0.1.0 and later.

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) interface. An PowerScale cluster that is powered by the OneFS operating system delivers a scalable pool of storage with a global namespace.

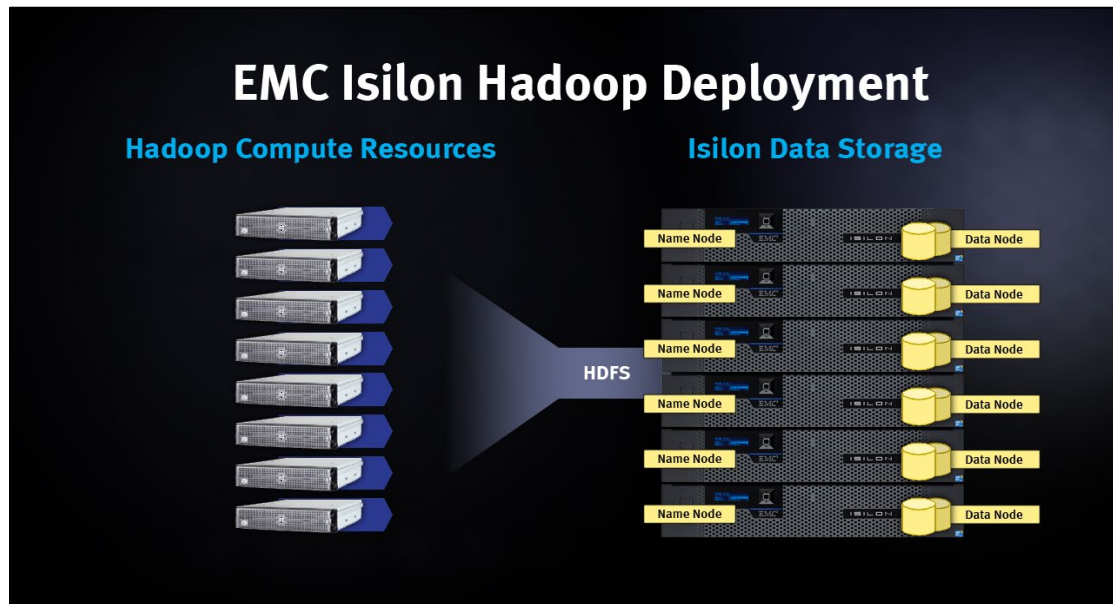
Hadoop compute clients can access the data that is stored in an PowerScale OneFS cluster by connecting to any node over the HDFS protocol. All nodes that are configured for HDFS provide NameNode and DataNode functionality. Each node boosts performance and expands the cluster capacity. For Hadoop analytics,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

In a traditional Hadoop deployment, the Hadoop compute nodes run analytics jobs against large sets of data. A NameNode directs the 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 DataNode. Often data is stored in production environments and then copied to a landing zone server before it is loaded on to HDFS. This process is network intensive and exposes the NameNode as a potential single point of failure.



In a PowerScale OneFS cluster with Hadoop deployment, 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 PowerScale OneFS cluster by connecting over HDFS. The HDFS protocol is native to the OneFS operating system, and no data migration is required.



The Hortonworks distribution is stored on the compute cluster, and the clients connect to the PowerScale OneFS cluster over the HDFS protocol to store and access Hadoop data.



Updates and additional information about OneFS Hadoop installs

The rapid release of new features and versions of Hadoop projects can introduce new behaviors and requirements. It is recommended 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.

Prerequisites

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

Ambari and Hortonworks Data Platform

OneFS Multi-tenant Installations

If the Hadoop cluster will be installed as a multi-tenant OneFS implementation, leveraging multiple Access Zones using the same Active Directory provider, additional configurations and considerations must be addressed before installing HDP and Ambari. It is recommended that you first consult the [Isilon and Hadoop Multitenant Installation and Integration Guide](#), and then if required, engage Dell EMC Professional Services to understand the requirements and deployment strategies that are available to deploy Ambari.

Ensure that the following requirements are met:

- Hortonworks Data Platform (HDP) 3.0.1 or later with Ambari 2.7.1 or later
- Password-less SSH is configured
 - See the [Hortonworks documentation](#) for configuring Password-less SSH.
- Familiarity with the Ambari and Hortonworks documentation and the installation instructions
 - To view the Ambari and the Hortonworks Data Platform (HDP) documents, go to <http://docs.hortonworks.com/index.html>
 - Use the following table to record the components that you have installed.

Component	Version
Ambari version	
HDP stack version	
OneFS OneFS cluster name	
Ambari Management Pack for Isilon OneFS	



OneFS cluster configuration

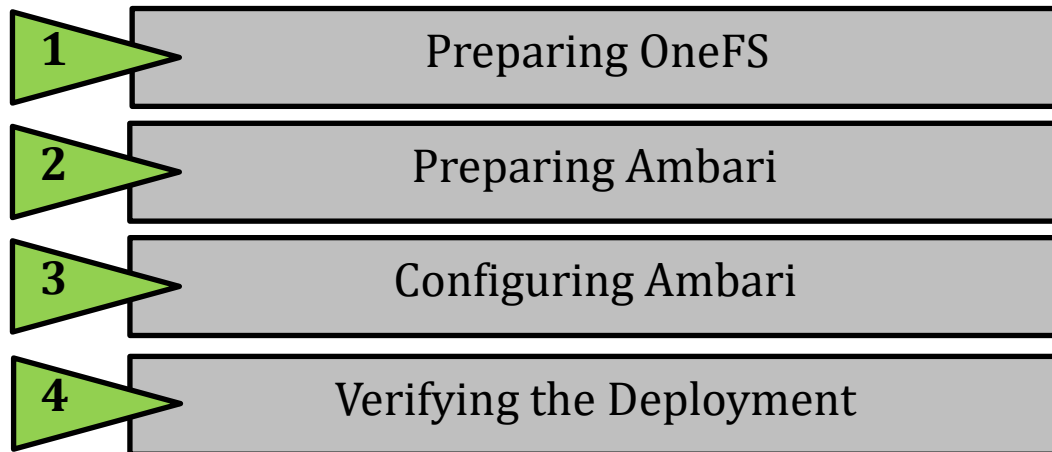
Ensure that the following requirements are met:

- A OneFS cluster running OneFS 8.1.2.0 or later.
- The OneFS cluster has access to your network and your network has access to the Internet. Internet access is required to download components.
- SmartConnect Advanced, a separately licensed OneFS module, is activated and SmartConnect is configured on your OneFS cluster.
- HDFS, a separately licensed 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 the [Isilon External Network Connectivity Guide](#).
- A dedicated OneFS Access Zone is in use; this is not the same as the System Zone.
- A OneFS HDFS root directory in the Access Zone.
- A simple access model between Hadoop and OneFS; UID and GUID, with parity.
- Use the following table to record the components that you have installed:

Component	Version or License
PowerScale OneFS	
SmartConnect module	
HDFS module	
OneFS cluster name	

Installing OneFS with Ambari

The installation of OneFS with Ambari can be separated into four stages as represented in the following figure. To complete the stages, you must perform tasks on both the Ambari cluster and the OneFS cluster.



Preparing OneFS

Complete the following steps to configure your OneFS cluster for use with Ambari and Hortonworks 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.

Validate OneFS version and license activation

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

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

```
isi version
```

2. Add the licenses for HDFS using the following command:

```
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 might not work.

Run the following commands to confirm that HDFS is installed:

```
isi license licenses list
```

```
isi license licenses view HDFS
```

- 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>
```

- Enable HDFS by running the following command:

```
isi services hdfs enable
```

- 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
```

Example:

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

Configure OneFS components

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

- Create and configure the HDFS root in the access zone
- Create users and groups
- (Optional) Create an access zone
- (Optional) Create a SmartConnect zone

Use the following table to record the configuration information for the OneFS cluster with Hortonworks Ambari integration:

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	
Ambari NameNode	

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-hdp --path=/ifs/data/zone1/hdp --create-path
```

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

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

Output similar to the following appears:

```

      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-hdp
      Path: /ifs/data/zone1/hdp
      Groupnet: groupnet0
      Map Untrusted: -
      Auth Providers: lsa-local-provider:zone1-hdp
      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/hdp/Hadoop
isi hdfs settings modify --zone=zone1-hdp --root-
directory=/ifs/data/zone1/hdp/Hadoop
```

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

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

Configure SmartConnect

On a node in the OneFS cluster, add a static IP address pool and associate it with the access zone 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, where:

- `<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 --id=groupnet0:subnet0:hadoop-pool-hdp --  
ranges=10.120.130.30-10.120.140.40 --access-zone=zone1-hdp --alloc-method=static --  
ifaces=1-4:40gige-1 --sc-subnet=subnet0 --sc-dns-zone=hdp.zone1.emc.com --  
description=hadoop"
```

3. View the properties of the existing pool.

```
isi network pools view --id=groupnet0:subnet0:hadoop-pool-hdp
```

Output similar to the following appears:

```

ID: groupnet0.subnet0.hadoop-pool-hdp
Groupnet: groupnet0
Subnet: subnet0
Name: hadoop-pool-hdp
Rules: -
Access Zone: zone1-hdp
Allocation Method: dynamic
Aggregation Mode: lacp
SC Suspended Nodes: -
Description: hdp_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: hdp.zone1.emc.com
SC DNS Zone Aliases: -
SC Failover Policy: round_robin
SC Subnet: subnet0
SC Ttl: 0
Static Routes: -
```



Configure DNS for OneFS

Before you begin, the OneFS cluster must already be implemented according to Dell EMC PowerScale best practices. For more information, see the HDFS Setup section of 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 additional 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. Each node's NameNode process replies with the IP address of any OneFS node 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, which responds with a valid node to connect to. The client connects to an 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 a different node 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.

Verify the SmartConnect configuration

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

```
ping hdp.zone1.emc.com
```

When you view the output of this command, note that 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 validates that the SmartConnect zone name FQDN is operating correctly.

Create HDFS users and groups

For each Hadoop system account that will submit HDFS jobs or access the file system, you must create local users and groups on the OneFS cluster as Ambari cannot do this. 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

Important

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, and 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 the Hadoop installer creates all the required system accounts. For example, a Hadoop system account, *yarn*, is created with the UID of 502 and the GID of 500 on the Hadoop cluster nodes. Since the Hadoop installer cannot create the local accounts directly on OneFS, they must be created 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 500 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 following blog post: [Isilon and Hadoop Local User UID Parity](#)

There are many methods of achieving UID and GID parity. You can leverage [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, the sequence depends on your deployment methodology and management practices. It is highly recommended 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.

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.

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 Isilon script configures on the OneFS cluster. The following procedures show how to manually add a single test user called *hduser1*.

Warning

If you want the users and groups to be defined by your directory service, such as Active Directory or LDAP, do not run these commands. This section addresses setting permissions of the HDFS root files or membership to run jobs. These steps create users but will likely fail when you run jobs with this configuration.

Manual steps to perform on the OneFS cluster

1. Add a group to the OneFS cluster.

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

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

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

3. Assign permissions to the user's home directory on the Hadoop cluster. The ID 2 in the example below is from when you previously ran the `isi zone zones view zone1` command.

```
isi_run -z2 chown hduser1:hduser1 /ifs/data/zone1/hdp/user/hduser1
chmod 755 /ifs/data/zone1/hdp/hadoop/user/hduser1
```

Manual steps to perform on the Hadoop client

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

1. Add the user to the Hadoop cluster.

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

Configure HDFS user for OneFS 8.1.2 and previous versions

In OneFS 8.1.2 and earlier, the HDFS user must be mapped to root and you must modify the access control list (ACL).

On a node in the OneFS 8.1.2 cluster, create and configure the HDFS root directory.

1. View the HDFS service settings.

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

2. Set the HDFS root directory for the access zone. **Note:** It is recommended that the directory for the access zone is not set to the root of /ifs.

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

3. Map the HDFS user to root. Create a user mapping rule to map the HDFS user to the OneFS root account. This mapping enables the services from the Hadoop cluster to communicate with the OneFS cluster using the correct credentials.

```
isi zone modify --add-user-mapping-rules="hdfs=>root[]" --zone=zone1-hdp
isi zone modify --add-user-mapping-rules="yarn-ats-hbase=>yarn-ats" --zone=zone1-
hdp
```

Note: User mapping yarn-ats-hbase to yarn-ats is required only if HDP and OneFS clusters are going to be secured (Kerberized).

You can skip yarn-ats-hbase to yarn-ats user mapping in two cases as follows:

- a. By renaming yarn-ats-hbase principals to yarn-ats during Kerberization if the Timeline Service3 V2.0s are deployed as Embedded or System Service mode.
- b. You do not need to set user mapping on OneFS if TLSv2.0 is configured on external HBase. For more details, see: https://docs.hortonworks.com/HDPDocuments/HDP3/HDP-3.0.1/data-operating-system/content/dosg_timeline_service_2.0_installation.html

- Assign the Ambari NameNode in the access zone and associate the SmartConnect name with it.

```
isi hdfs settings modify --zone=<zone> --ambari-namenode=<my-smartconnectzone-name>
```

For example:

```
isi hdfs settings modify --zone=zone1-hdp --ambari-namenode=hdfs.hop-isi-  
m.solarch.lab.emc.com
```

- Assign the Ambari Server to the HDFS zone

```
isi hdfs settings modify --zone=<zone> --ambari-server=<ambari-server.fqdn>
```

For example:

```
isi hdfs settings modify --zone=zone1-hdp --ambari-server=amb-srv.hop-isi-  
m.solarch.lab.emc.com
```

- Create an indicator file in the Hadoop directory to view your OneFS cluster and access zone through HDFS.

```
touch /ifs/data/zone1/hdp/hadoop/THIS_IS_ISILON-hdp.txt
```

- Modify the ACL setting for OneFS 8.1.2 and earlier only.

Run the following command on a node in the OneFS cluster to modify ACL settings before you create directories or files in the next section. This creates the correct permission behavior on the cluster for HDFS.

Note

ACL policies are cluster-wide, so you should understand this change before performing it on production clusters.

```
isi auth settings acls modify --group-owner-inheritance=parent  
isi auth settings acls view
```

Configure HDFS user for OneFS 8.2.0 and later versions

- View the HDFS service settings.

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

- Set the HDFS root directory for the access zone. **Note:** It is recommended that the directory for the access zone is not set to the root of /ifs.

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

- Assign the Ambari NameNode in the access zone and associate the SmartConnect name with it.

```
isi hdfs settings modify --zone=<zone> --ambari-namenode=<my-smartconnectzone-name>
```

For example:

```
isi hdfs settings modify --zone=zone1-hdp --ambari-namenode=hdfs.hop-isi-  
m.solarch.lab.emc.com
```

- Create a new 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-hdp
```

5. 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-hdp
```

6. 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=zone1-hdp
```

7. 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-hdp
```

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

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

For example:

```
isi auth roles view HdfsAccess --zone=zone1-hdp
    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
```

9. (Optional) Flush auth mapping and auth cache to make the HDFS user take immediate effect as the “HdfsAccess” role that you created above.

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

Note

ACL Policies do not need to be modified for OneFS 8.2 and later. The HDFS protocols act the same as non-OneFS HDFS for File System Group Owner inheritance.

Preparing Ambari

Perform the steps that are discussed in this section on the Ambari hosts which become your Hadoop servers and clients.

Hadoop clusters and services rely heavily on DNS. All client hosts in your system must be configured for both forward and reverse DNS lookups. Validate that all hosts can resolve each other's hostnames and IP addresses.

Before you begin the installation of Ambari:

- Ensure that all your hosts meet the requirements that are determined by Ambari and Hortonworks to complete a successful Hadoop cluster installation. For more information and these installation guides, go to the Hortonworks website: <http://docs.hortonworks.com/index.html>.
- Ensure that you have access to the [Isilon OneFS Ambari Management Pack](#). Contact your Dell EMC sales representative for more information about accessing the management pack.

The Isilon OneFS Ambari Management Pack is a software component that can be installed in Ambari to define OneFS as a service in a Hadoop cluster. The management pack allows an Ambari administrator to start, stop, and configure OneFS as an HDFS storage service. This provides native namenode and datanode capabilities similar to traditional HDFS.

Steps to perform on the Hadoop client

1. To prepare Ambari for implementation, follow the instructions for your version of Ambari in the [Hortonworks guide](#). See the "Installing Ambari" section.

The guide provides the steps that you must perform to prepare the Ambari environment install Ambari and installing and configure HDP.

Important

Complete the steps in the Hortonworks guide in section 1, "Getting Ready", and section 4, "Installing Ambari." After you start the Ambari server, do not continue to section 6 of the *Hortonworks Guide* until after you have completed the instructions that are described in the [Preparing OneFS](#) section of this guide.

Complete the following steps that are described in the *Hortonworks Guide*:

1. Download the Ambari repository for the operating system that runs your installation host.
2. Set up the Ambari server.
3. Download the Isilon OneFS Ambari Management Pack installation bundle from the [product download](#) page and extract the contents on to the Ambari server.
4. Install the management pack on the Ambari server by running the following command:

```
ambari-server install-mpack --mpack=<tar file_name.tar.gz> -verbose
```

For example:

```
ambari-server install-mpack --mpack=isilon-onefs-mpack-1.0.0.0-SNAPSHOT.tar.gz --verbose
```

Output similar to the following displays:

```
Using python /usr/bin/python
Installing management pack
INFO: Loading properties from /etc/ambari-server/conf/ambari.properties
INFO: Installing management pack isilon-onefs-mpack-1.0.0.0-SNAPSHOT.tar.gz
INFO: Loading properties from /etc/ambari-server/conf/ambari.properties
INFO: Download management pack to temp location /var/lib/ambari-
server/data/tmp/isilon-onefs-mpack-1.0.0.0-SNAPSHOT.tar.gz
INFO: Loading properties from /etc/ambari-server/conf/ambari.properties
INFO: Expand management pack at temp location /var/lib/ambari-
server/data/tmp/isilon-onefs-mpack-1.0.0.0-SNAPSHOT/
2018-11-07 06:36:39,137 - Execute(['tar', '-xf', '/var/lib/ambari-
server/data/tmp/isilon-onefs-mpack-1.0.0.0-SNAPSHOT.tar.gz', '-C',
'/var/lib/ambari-server/data/tmp/']) {'tries': 3, 'sudo': True, 'try_sleep': 1}
INFO: Loading properties from /etc/ambari-server/conf/ambari.properties
INFO: Loading properties from /etc/ambari-server/conf/ambari.properties
INFO: Stage management pack onefs-ambari-mpack-1.0 to staging location
/var/lib/ambari-server/resources/mpacks/onefs-ambari-mpack-1.0
INFO: Processing artifact ONEFS-addon-services of type stack-addon-service-
definitions in /var/lib/ambari-server/resources/mpacks/onefs-ambari-mpack-
0.1/addon-services
INFO: Loading properties from /etc/ambari-server/conf/ambari.properties
INFO: Loading properties from /etc/ambari-server/conf/ambari.properties
INFO: Adjusting file permissions and ownerships
INFO: about to run command: chmod -R 0755 /var/lib/ambari-server/resources/stacks
INFO: process_pid=28352
INFO: about to run command: chown -R -L root /var/lib/ambari-
server/resources/stacks
INFO: process_pid=28353
INFO: about to run command: chmod -R 0755 /var/lib/ambari-
server/resources/extensions
process_pid=28354
INFO: about to run command: chown -R -L root /var/lib/ambari-
server/resources/extensions
INFO: process_pid=28355
INFO: about to run command: chmod -R 0755 /var/lib/ambari-
server/resources/common-services
INFO: process_pid=28356
INFO: about to run command: chown -R -L root /var/lib/ambari-
server/resources/common-services
INFO: process_pid=28357
INFO: about to run command: chmod -R 0755 /var/lib/ambari-server/resources/mpacks
INFO: process_pid=28358
INFO: about to run command: chown -R -L root /var/lib/ambari-
server/resources/mpacks
INFO: process_pid=28359
INFO: about to run command: chmod -R 0755 /var/lib/ambari-
server/resources/mpacks/cache
INFO: process_pid=28360
INFO: about to run command: chown -R -L root /var/lib/ambari-
server/resources/mpacks/cache
INFO: process_pid=28361
INFO: about to run command: chmod -R 0755 /var/lib/ambari-
server/resources/dashboards
```

```
INFO: process_pid=28362
INFO: about to run command: chown -R -L root /var/lib/ambari-
server/resources/dashboards
INFO: process_pid=28363
INFO: about to run command: chown -R -L root /var/lib/ambari-
server/resources/stacks
INFO: process_pid=28364
INFO: about to run command: chown -R -L root /var/lib/ambari-
server/resources/extensions
INFO: process_pid=28365
INFO: about to run command: chown -R -L root /var/lib/ambari-
server/resources/common-services
INFO: process_pid=28366
INFO: about to run command: chown -R -L root /var/lib/ambari-
server/resources/mpacks
INFO: process_pid=28367
INFO: about to run command: chown -R -L root /var/lib/ambari-
server/resources/mpacks/cache
INFO: process_pid=28368
INFO: about to run command: chown -R -L root /var/lib/ambari-
server/resources/dashboards
INFO: process_pid=28369
INFO: Management pack onefs-ambari-mpack-0.1 successfully installed! Please restart
ambari-server.
INFO: Loading properties from /etc/ambari-server/conf/ambari.properties
Ambari Server 'install-mpack' completed successfully.
```

Note

The Isilon OneFS Ambari Management Pack includes a setting for Yarn that you may need to change. The Yarn Timeline Service 2.0 relies on Apache HBase for backend storage. As PowerScale is a single storage tier from Yarn's perspective, the storage policy for HBase is set to NONE in the Yarn-HBase-site. If your Yarn deployment uses an external HBase for Timeline Service 2.0, then the storage policy settings should be changed to the HBase default, HOT, or whatever is appropriate for your environment.

5. Start the Ambari server.
6. Logging in to the Ambari server opens the create cluster page.

Important: Do not continue to section 6 of the *Hortonworks Guide* until the OneFS cluster is prepared as described in the following steps and is ready to be integrated into Ambari during the installation.

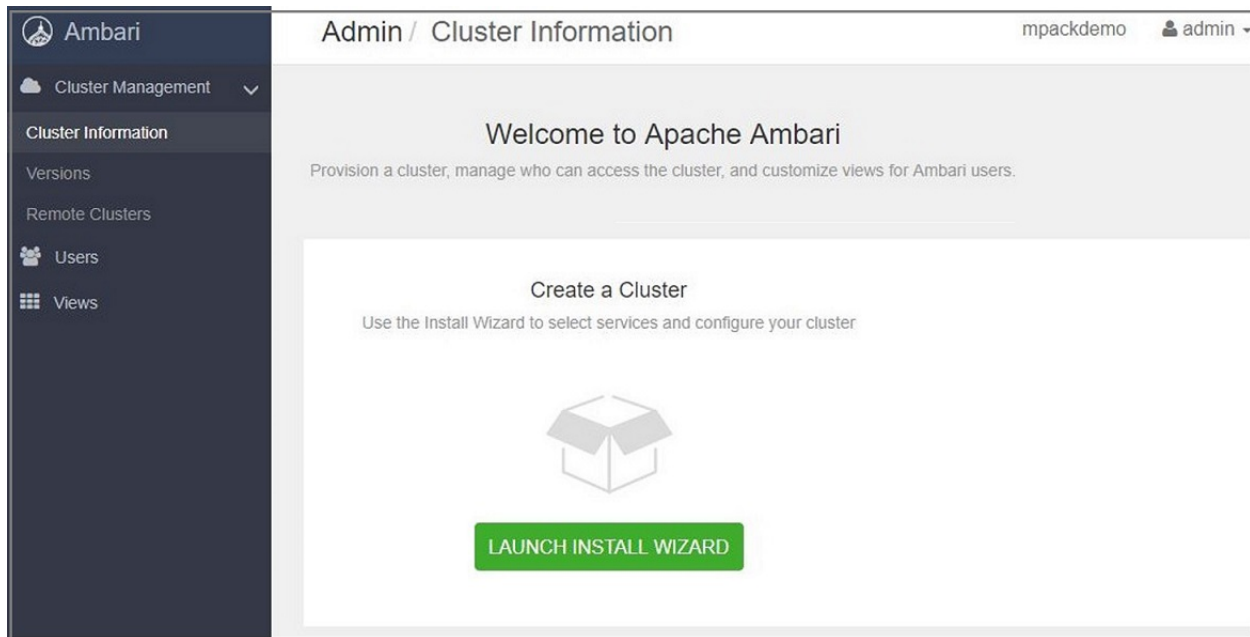
Configuring Ambari 2.7.1 and later

Perform the steps in section 6 as described in the [Hortonworks Data Platform Apache Ambari installation guide](#), "Installing, Configuring, and Deploying a Cluster." An outline of the steps that you must perform to configure Ambari and Hortonworks for OneFS is as follows:

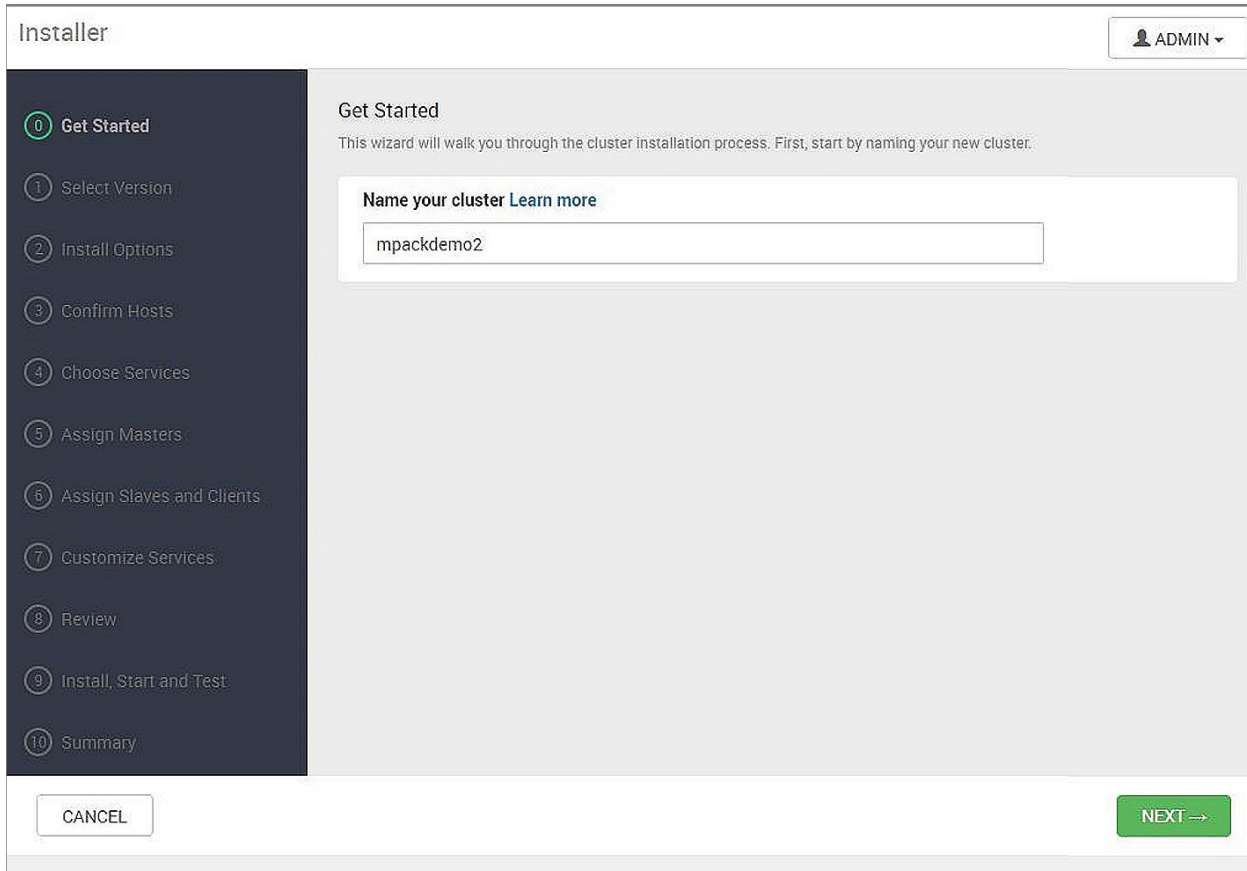
1. Choose **OneFS** instead of the **HDFS** deployment option.
2. In the **OneFS Service Settings** section, specify your SmartConnect FQDN and any other HDFS configuration settings that you want to change.

Steps to perform in your browser from your client

1. After starting the Ambari service, open Ambari Web using a web browser.
2. Point your browser to <http://<your.ambari.server>:8080>, where <your.ambari.server> is the name of your Ambari server host.
3. Log in to the Ambari server using `admin` both as the default username and password. You can change these credentials later. For a new cluster, the Ambari install wizard displays a Welcome page through which you can launch the Ambari Install wizard.



4. Select **Launch Install Wizard** and complete the installation steps in the [Hortonworks guide](#), including the following steps that are specific to OneFS.



The screenshot shows the Ambari Installer interface. On the left, a dark sidebar contains a list of steps: 0 Get Started, 1 Select Version, 2 Install Options, 3 Confirm Hosts, 4 Choose Services, 5 Assign Masters, 6 Assign Slaves and Clients, 7 Customize Services, 8 Review, 9 Install, Start and Test, and 10 Summary. The 'Get Started' step is highlighted. The main content area is titled 'Get Started' and includes the text: 'This wizard will walk you through the cluster installation process. First, start by naming your new cluster.' Below this is a form with the label 'Name your cluster' and a link 'Learn more'. The text input field contains 'mpackdemo2'. At the bottom of the main area, there is a 'CANCEL' button on the left and a green 'NEXT →' button on the right. In the top right corner of the installer window, there is a user dropdown menu showing 'ADMIN'.

5. In the **Name your cluster** field, type a unique name for the cluster.
Note: As a best practice Ambari Cluster names should be fewer than seven characters. Longer cluster names require additional configuration for multitenant AD installation due to OneFS specific requirements. Use abbreviated cluster names where possible to facilitate integration with OneFS. For example: hdp1, h-prod1, or similar.

6. On the **Select Version** screen, select the version of the HDP stack and the repository to configure it. Consult [Hadoop Distributions and Products Supported by OneFS](#) for the version that you must configure with your OneFS cluster.

Installer

Get Started

1 Select Version

2 Install Options

3 Confirm Hosts

4 Choose Services

5 Assign Masters

6 Assign Slaves and Clients

7 Customize Services

8 Review

9 Install, Start and Test

10 Summary

Select Version

Select the software version and method of delivery for your cluster.

HDP-3.0

HDP-3.0.0.0

Accumulo	1.7.0
Infra Solr	0.1.0
Ambari Metrics	0.1.0
Atlas	1.0.0
Druid	0.12.1
HBase	2.0.0

Repositories

Using a Public Repository requires Internet connectivity. Using a Local Repository requires you have configured the software in a repository available in your network.

Use Public Repository

Use Local Repository

Provide Base URLs for the Operating Systems you are configuring.

7. Click one of **Use Public Repository** or **Use Local Repository** based on the availability of Internet access.

7 Customize Services
8 Review
9 Install, Start and Test
10 Summary

HBase2.0.0

Repositories

Using a Public Repository requires Internet connectivity. Using a Local Repository requires you have configured the software in a repository available in your network.

☒ Use Public Repository
☐ Use Local Repository

Provide Base URLs for the Operating Systems you are configuring.

+ADD

OS	Name	Base URL
amazonlinux2	HDP-3.0	<input type="text" value="http://s3.amazonaws.com/dev.hortonworks.com/HDP/amazonlinux2/3.x/BUILDS/3.0.0.0-1599"/>
	HDP-UTILS-1.1.0.22	<input type="text" value="http://s3.amazonaws.com/dev.hortonworks.com/HDP-UTILS-1.1.0.22/repos/amazonlinux2"/>
redhat-ppc7	HDP-3.0	<input type="text" value="http://s3.amazonaws.com/dev.hortonworks.com/HDP/centos7-ppc/3.x/BUILDS/3.0.0.0-1599"/>
	HDP-UTILS-1.1.0.22	<input type="text" value="http://s3.amazonaws.com/dev.hortonworks.com/HDP-UTILS-1.1.0.22/repos/centos7-ppc"/>
redhat7	HDP-3.0	<input type="text" value="http://s3.amazonaws.com/dev.hortonworks.com/HDP/centos7/3.x/BUILDS/3.0.0.0-1599"/>
	HDP-UTILS-1.1.0.22	<input type="text" value="http://s3.amazonaws.com/dev.hortonworks.com/HDP-UTILS-1.1.0.22/repos/centos7"/>
suse12	HDP-3.0	<input type="text" value="http://s3.amazonaws.com/dev.hortonworks.com/HDP/sles12/3.x/BUILDS/3.0.0.0-1599"/>
	HDP-UTILS-1.1.0.22	<input type="text" value="http://s3.amazonaws.com/dev.hortonworks.com/HDP-UTILS-1.1.0.22/repos/sles12"/>
ubuntu14	HDP-3.0	<input type="text" value="http://s3.amazonaws.com/dev.hortonworks.com/HDP/ubuntu14/3.x/BUILDS/3.0.0.0-1599"/>
	HDP-UTILS-1.1.0.22	<input type="text" value="http://s3.amazonaws.com/dev.hortonworks.com/HDP-UTILS-1.1.0.22/repos/ubuntu14"/>
ubuntu16	HDP-3.0	<input type="text" value="http://s3.amazonaws.com/dev.hortonworks.com/HDP/ubuntu16/3.x/BUILDS/3.0.0.0-1599"/>
	HDP-UTILS-1.1.0.22	<input type="text" value="http://s3.amazonaws.com/dev.hortonworks.com/HDP-UTILS-1.1.0.22/repos/ubuntu16"/>

☐ Skip Repository Base URL validation (Advanced) ⓘ
☒ Use RedHat Satellite/Spacewalk ⓘ

← BACK

CANCEL

NEXT →

8. On the **Install Options** screen, under **Target Hosts**, type the fully qualified name of all the Linux hosts that participate in the Hadoop cluster.

9. Under **Host Registration Information** section, attach the SSH private key.

The creation of this host and key was performed before running the Ambari wizard.
See the [Hortonworks Ambari documentation](#) for additional details.

Important

Do not include the OneFS cluster in the target hosts.

To obtain the private key:

- a. Log in to your Ambari host server.
- b. Generate the private key as shown:
- c. Copy the key to all the hosts as shown:

```
ssh-keygen
```

```
ssh-copy-id root@XXXXX
```

- d. Change to the `/root/.ssh` directory.
- e. Output the contents of the `id_rsa` file.

```
cat /root/.ssh/id_rsa
```

- f. Copy the output to a file and save the file on your desktop.
- g. Copy the file to the machine on which you are running the web-based Ambari Install Wizard.

- h. Copy the entire contents of the `id_rsa` file, including the leading dash characters, into the **Host Registration Information** box in **Install Options** screen of the Ambari wizard.

If no keys exist, see the Ambari documentation on generating a private key and setting up keyless SSH access on hosts in the cluster.

- i. Select **Register and Confirm**.

10. On the **Choose File System** screen, select **OneFS**.

Service	Version	Description
<input type="checkbox"/> HDFS	3.1.0	Apache Hadoop Distributed File System
<input checked="" type="checkbox"/> OneFS	1.0.0	Isilon Systems OneFS

11. On the **Choose Services** screen, retain the default selections.

Service	Version	Description
<input checked="" type="checkbox"/> YARN + MapReduce2	3.1.0	Apache Hadoop NextGen MapReduce (YARN)
<input checked="" type="checkbox"/> Tez	0.9.1	Tez is the next generation Hadoop Query Processing framework written on top of YARN.
<input checked="" type="checkbox"/> Hive	3.1.0	Data warehouse system for ad-hoc queries & analysis of large datasets and table & storage management service
<input checked="" type="checkbox"/> HBase	2.0.0	Non-relational distributed database and centralized service for configuration management & synchronization
<input checked="" type="checkbox"/> Pig	0.16.0	Scripting platform for analyzing large datasets
<input checked="" type="checkbox"/> Sqoop	1.4.7	Tool for transferring bulk data between Apache Hadoop and structured data stores such as relational databases
<input checked="" type="checkbox"/> Oozie	4.3.1	System for workflow coordination and execution of Apache Hadoop jobs. This also includes the installation of the optional Oozie Web Console which relies on and will install the EXUS Library.
<input checked="" type="checkbox"/> ZooKeeper	3.4.6	Centralized service which provides highly reliable distributed coordination
<input checked="" type="checkbox"/> Storm	1.2.1	Apache Hadoop Stream processing framework
<input checked="" type="checkbox"/> Accumulo	1.7.0	Robust, scalable, high performance distributed key/value store.
<input checked="" type="checkbox"/> Infra Solr	0.1.0	Core shared service used by Ambari managed components.
<input checked="" type="checkbox"/> Ambari Metrics	0.1.0	A system for metrics collection that provides storage and retrieval capability for metrics collected from the cluster
<input checked="" type="checkbox"/> Atlas	1.0.0	Atlas Metadata and Governance platform
<input checked="" type="checkbox"/> Kafka	1.0.1	A high-throughput distributed messaging system

12. On the **Assign Masters** screen, retain the default settings.

Assign Masters
Assign master components to hosts you want to run them on.

Service	Master Component
Resource Manager	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
Timeline Service V2.0 Reader	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
YARN Registry DNS	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
Timeline Service V1.5	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
History Server	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
HiveServer2	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
Hive Metastore	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
HBase Master	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
Oozie Server	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
ZooKeeper Server	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
ZooKeeper Server	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
ZooKeeper Server	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
Nimbus	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
DRPC Server	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
Storm UI Server	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
Accumulo Tracer	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
Accumulo GC	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
Accumulo Monitor	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
Accumulo Master	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
Infra Solr Instance	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
Metrics Collector	rdunode284514.west.isilon.com (15.5 GB, 4 cores)
Grafana	rdunode284514.west.isilon.com (15.5 GB, 4 cores)

13. Ensure that all the ranger requirements are met before you click **Proceed** on the **Ranger Requirements** screen.

Ranger Requirements

- You must have an MySQL/Oracle/Postgres/MS SQL/SQL Anywhere Server database instance running to be used by Ranger.
- In Assign Masters step of this wizard, you will be prompted to specify which host for the Ranger Admin. On that host, you must have DB Client installed for Ranger to access to the database. (Note: This is applicable for only Ranger 5.4.0).
- Ensure that the access for the DB Admin user is enabled in DB server from any host.
- Execute the following command on the Admin Server host. Replace `database-type` with `mysql/mssql/postgres/oracle/anywhere` and `jdbc-driver-path` based on the location of corresponding JDBC driver.


```
ambari-server setup --jdbc-db={database-type} --jdbc-driver={jdbc-driver-path}
```

☒ I have met all the requirements above.

CANCEL **PROCEED**

14. On the **Assign Slaves and Clients** screen, ensure that all the clients are selected.

Assign Slaves and Clients

Assign slave and client components to hosts you want to run them on.
Hosts that are assigned master components are shown with a "Client" will install YARN Client, MapReduce2 Client, Tez Client, Hive Client, HBase Client, Pig Client, Sqoop Client, Ozone Client, Zookeeper Client, Accumulo Client, Intra-Site Client, Atlas Metadata Client, Spark2 Client and Onef's Client.

Assignment of slave and client components has the following issues:

- Druid Historical requires HDFS_CLIENT to be co-hosted on following host(s): rdunode284517.west.isilon.com.
- Druid MiddleManager requires HDFS_CLIENT to be co-hosted on following host(s): rdunode284517.west.isilon.com.

Host	all none	all none	all none	all none	all none	all none	all none	all none	all none	all none	all none	all none
rdunode284514.west.isilon.com	<input type="checkbox"/> NodeManager	<input type="checkbox"/> RegionServer	<input type="checkbox"/> Phoenix Query Server	<input type="checkbox"/> Supervisor	<input type="checkbox"/> Accumulo TServer	<input type="checkbox"/> Ranger TagSync	<input type="checkbox"/> Livy for Spark2 Server	<input type="checkbox"/> Spark2 Thrift Server	<input type="checkbox"/> Druid Historical	<input type="checkbox"/> Druid MiddleManager	<input checked="" type="checkbox"/> Client	
rdunode284515.west.isilon.com	<input type="checkbox"/> NodeManager	<input type="checkbox"/> RegionServer	<input type="checkbox"/> Phoenix Query Server	<input type="checkbox"/> Supervisor	<input type="checkbox"/> Accumulo TServer	<input type="checkbox"/> Ranger TagSync	<input type="checkbox"/> Livy for Spark2 Server	<input type="checkbox"/> Spark2 Thrift Server	<input type="checkbox"/> Druid Historical	<input type="checkbox"/> Druid MiddleManager	<input checked="" type="checkbox"/> Client	
rdunode284516.west.isilon.com	<input type="checkbox"/> NodeManager	<input type="checkbox"/> RegionServer	<input type="checkbox"/> Phoenix Query Server	<input type="checkbox"/> Supervisor	<input type="checkbox"/> Accumulo TServer	<input type="checkbox"/> Ranger TagSync	<input type="checkbox"/> Livy for Spark2 Server	<input type="checkbox"/> Spark2 Thrift Server	<input type="checkbox"/> Druid Historical	<input type="checkbox"/> Druid MiddleManager	<input checked="" type="checkbox"/> Client	
rdunode284517.west.isilon.com	<input checked="" type="checkbox"/> NodeManager	<input checked="" type="checkbox"/> RegionServer	<input type="checkbox"/> Phoenix Query Server	<input checked="" type="checkbox"/> Supervisor	<input checked="" type="checkbox"/> Accumulo TServer	<input checked="" type="checkbox"/> Ranger TagSync	<input type="checkbox"/> Livy for Spark2 Server	<input type="checkbox"/> Spark2 Thrift Server	<input checked="" type="checkbox"/> Druid Historical	<input checked="" type="checkbox"/> Druid MiddleManager	<input checked="" type="checkbox"/> Client	

Items per page: 25 | 1 - 4 of 4

NEXT

15. On the **Customize Services** screen, specify the settings on the different tabbed pages as shown:

- a. On the CREDENTIALS tabbed page, specify values as shown in the following screen:

Installer

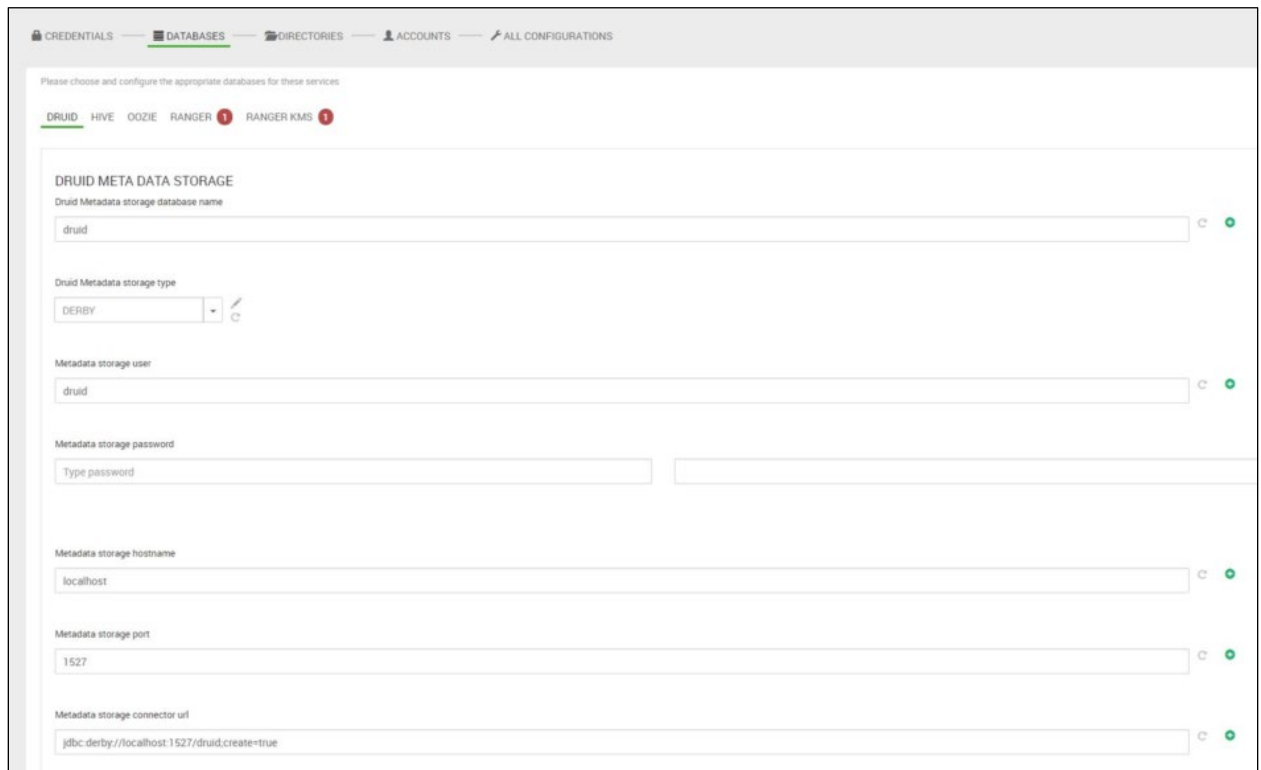
Get Started | Select Version | Install Options | Confirm Hosts | Choose Services | Assign Masters | Assign Slaves and Clients | **Customize Services** | Review | Install, Start and Test | Summary

CREDENTIALS | DATABASES | DIRECTORIES | ACCOUNTS | ALL CONFIGURATIONS

Please provide credentials for these services

	Username*
Accumulo Root	N/A
Accumulo Instance Secret	N/A
Grafana Admin	admin
Atlas Admin User	admin
Druid Metadata User	druid
Hive Database	hive
Knox Master Secret	N/A
Oozie Database	oozie
Ranger Admin	admin
Ranger Admin For Ambari	amb_ranger_admin
Ranger DB	rangeradmin
Database Administrator (DBA)	root
Ranger Usersync User's Password	N/A
Ranger Tagsync User's Password	N/A
Ranger KMS Keyadmin User's Password	N/A
Ranger KMS Master Key Password	N/A
Ranger KMS DB	rangerkms
Database Administrator (DBA)	root
Activity Explorer's Admin	N/A

- b. On the DATABASES tabbed page, configure your settings as shown in the following screen:



CREDENTIALS — DATABASES — DIRECTORIES — ACCOUNTS — ALL CONFIGURATIONS

Please choose and configure the appropriate databases for these services

DRUID — HIVE — OOOZIE — RANGER 1 — RANGER KMS 1

DRUID META DATA STORAGE

Druid Metadata storage database name

druid

Druid Metadata storage type

DERBY

Metadata storage user

druid

Metadata storage password

Type password

Metadata storage hostname

localhost

Metadata storage port

1527

Metadata storage connector url

jdbc:derby://localhost:1527/druid;create=true

- c. On the DIRECTORIES tabbed page, configure your settings as shown in the following screen:

CREDENTIALS
DATABASES
DIRECTORIES
ACCOUNTS
ALL CONFIGURATIONS

YARN
MAPREDUCE2
TEZ
HIVE
HBASE
OOZIE
ZOOKEEPER
STORM
ACCUMULO
INFRA SOLR
AMBARI METRICS
ATLAS
KAFKA
KNOX
RA

DATA DIRS

YARN NodeManager Local directories

YARN Timeline Service Entity Group FS Store Active directory

YARN Node Labels FS Store Root directory

YARN NodeManager Recovery directory

YARN Timeline Service Entity Group FS Store Done directory

LOG DIRS

YARN NodeManager Log directories

YARN NodeManager Remote App Log directory

YARN Log Dir Prefix

PID DIRS

YARN PID Dir Prefix

- d. On the ACCOUNTS tabbed page, configure your settings as shown in the following screen:

Installer ADMIN

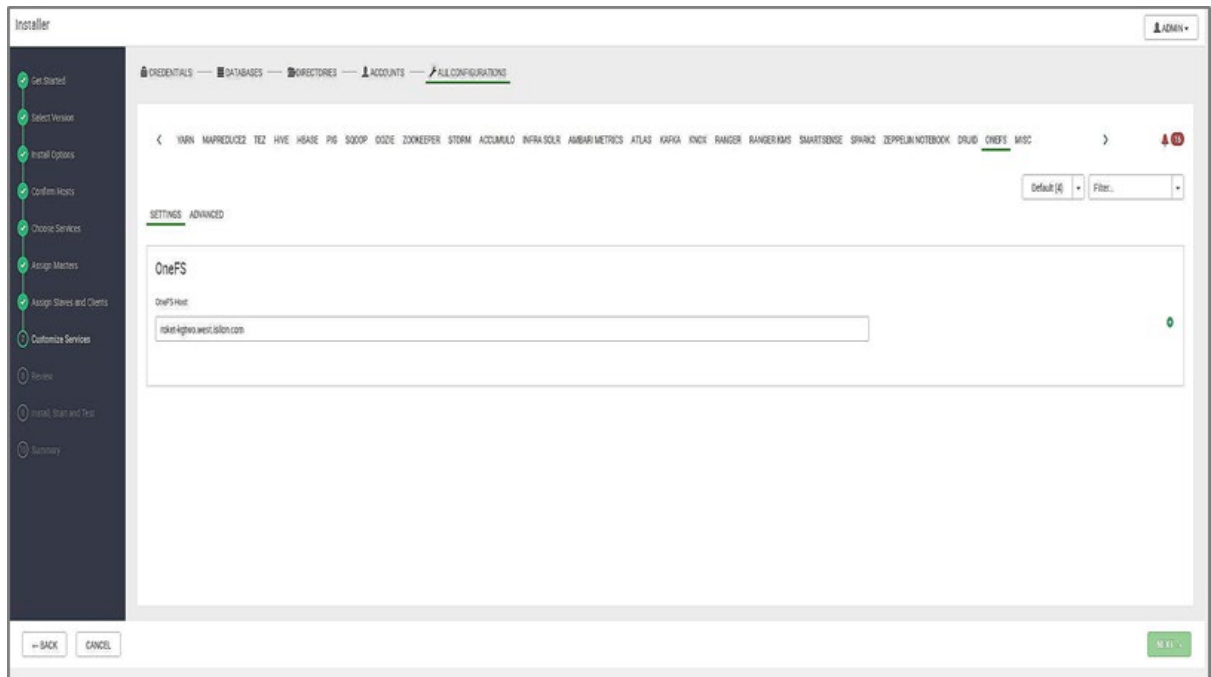
ACCOUNTS | CREDENTIALS | DATABASES | DIRECTORIES | ALL CONFIGURATIONS

Please review these settings for Service Accounts

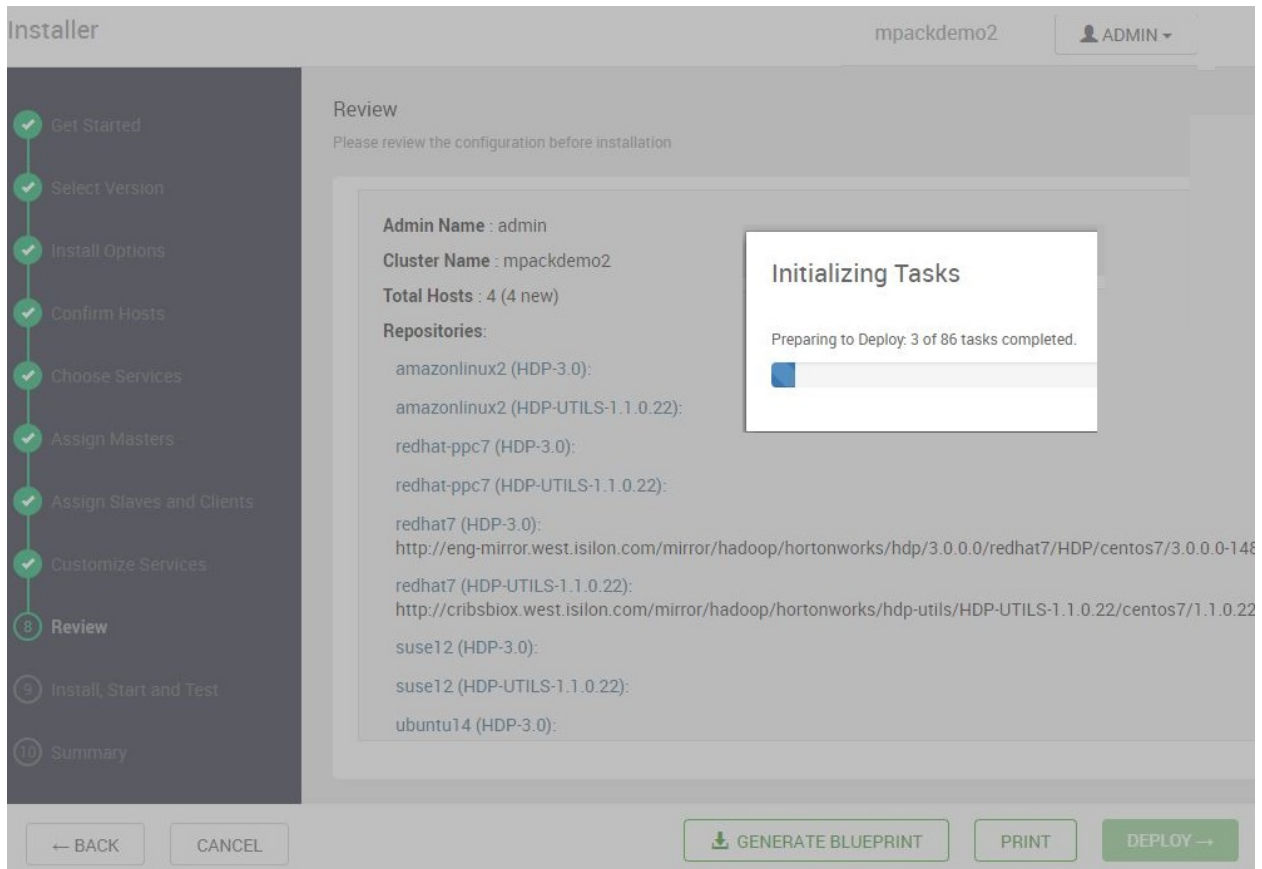
- ☒ Use Ambari to Manage Service Accounts and Groups
- ☒ Use Ambari to Manage Group Memberships
- ☒ Use Ambari to Manage Service Accounts UID's

Users/Groups	Username
Smoke User	ambari-qa
Hadoop Group	hadoop
Accumulo User	accumulo
Infra Solr User	infra-solr
Ambari Metrics User	ams
Metadata User	atlas
Druid User	druid
HBase User	hbase
Hive User	hive
Kafka User	kafka
Knox Group	knox
Knox User	knox
Mapreduce User	mapred
HDFS User Group	hdfs_group
HDFS User	hdfs
Proxy User Group	users
Oozie User	oozie
Ranger Group	ranger
Ranger User	ranger
Kms Group	kms
Kms User	kms
Livy2 Group	livy
Livy2 User	livy

- e. On the ALL CONFIGURATIONS tabbed page, click **OneFS** and configure your settings as shown in the following screen:



16. On the **Review** screen, review the configuration for accuracy. Click **Deploy**.



17. On the **Install, Start and Test** screen, wait for the test results. The tests should result in success for both the OneFS and compute nodes.

A progress indicator appears.

19. Wait until the status is 100% complete before you proceed.

See the Hortonworks documentation for guidance on how to manually restart the services in case some services fail to install and start up.

Installer mpackdemo2 ADMIN

Install, Start and Test

Please wait while the selected services are installed and started.

5 % overall

Show: **All (4)** | [In Progress \(4\)](#) | [Warning \(0\)](#) | [Success \(0\)](#) | [Fail \(0\)](#)

Host	Status	Message
rduvnode284514.west.isilon.com	5%	Installing Accumulo Master
rduvnode284515.west.isilon.com	6%	Installing Atlas Metadata Client
rduvnode284516.west.isilon.com	6%	Installing Atlas Metadata Client
rduvnode284517.west.isilon.com	6%	Installing Atlas Metadata Client

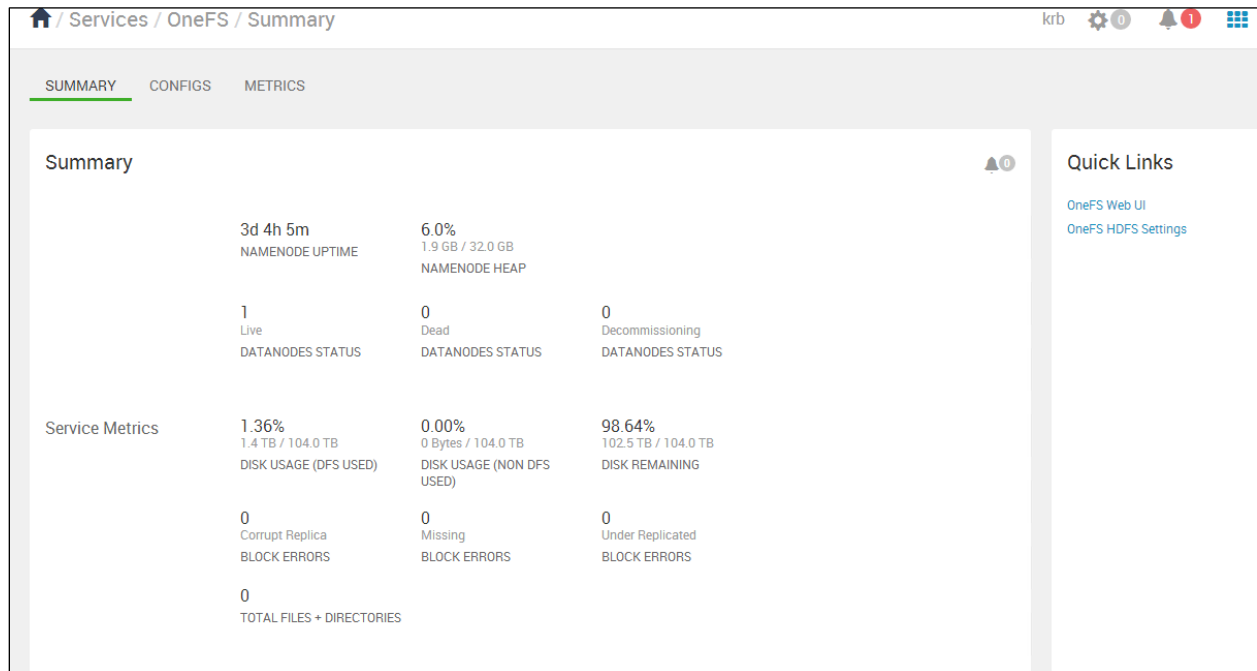
Items per page: 25 | 1 - 4 of 4

NEXT →

20. On the **Summary** screen, click **Complete**.
Ambari starts automatically. All the services on the left pane should appear with a green checkmark.
21. Run the command as shown in the following example to configure the Ambari server for collecting OneFS metrics:

```
isi hdfs settings modify --zone=zone_name --ambari-metrics-collector=<FQDN_of_HDP
client_running_Ambari_metrics_collector_service>
```

22. Verify that Ambari can collect metrics from the OneFS SmartConnect zone as shown in the following screen:



You are now configured to use a OneFS cluster with Hortonworks for Apache Hadoop.

Addendum

This section discusses some of the security, management, and troubleshooting features that are available in an installation of OneFS with Hadoop.

Apache Ranger authorization policy integration

Apache Ranger is a centralized management console that enables you to monitor and manage data security across the Hortonworks Hadoop distribution system. A Ranger administrator can define and apply authorization policies across Hadoop components including HDFS.

HDFS policies that are defined in Ranger are checked before the native file access control is applied. This two-layered authorization model differs in the way the standard Ranger HDFS policies are checked with Directed Attached Storage (DAS), but the model is suitable for using OneFS as a multiprotocol data lake with Hadoop. OneFS native file system ACL allows a storage administrator to correctly set up access control for multiple workloads and with multiprotocol access to the HDFS dataset. A Ranger administrator can apply a further restrictive Hadoop user access control to the same HDFS dataset, thus providing the administrators the appropriate control span within their management domains.

In a OneFS cluster with Hadoop deployment, Ranger authorization policies serve as a filter before applying the native file access control.

Notes

- The Ranger Audit and Transparent Data Encryption components are not supported.
- You can enable Apache Ranger on OneFS clusters and then check for new authorization policies, receive HDFS requests from clients, and apply authorization policies to the HDFS requests which can be one of DENY, ALLOW, or UNDETERMINED.
- The Ranger DENY policy takes precedence over the ALLOW policy.
- The Ranger DENY policy prevents user or group access to files and directories in OneFS that the file system would have otherwise allowed the users or groups to access.

To understand how the Ranger policies are applied, consider this example: A user in the Sales group requires access to certain files and directories that have specific HDFS file system ownership and permissions as shown:

Scenario	Files and directories	HDFS file system ownership and permissions
A	/ifs/data/<zone>/<hdfs-root> (the HDFS root directory for the zone)	Read, write, and execute (rwx) permissions for everyone.
B	/ifs/data/<zone>/<hdfs-root>/corp-confidential	Neither the user or the Sales group have read, write, and execute (rwx) permissions.
C	/ifs/data/<zone>/<hdfs-root>/hdfs-confidential	The Sales group owns the folder, and the folder has read, write, and execute (rwx) permissions for the group including the user.
D	/ifs/data/<zone>/<hdfs-root>/hdfs-not-for-user	No read, write, and execute permissions for hidden files.

If Ranger policies are further applied, the user's access to the files and directories changes as indicated in the following table:

Ranger HDFS service instance	User's view and edit permissions per scenario			
	A	B	C	D
None	Yes	No	Yes	No
One policy that provides everyone in the Hadoop group, access to the root directory.	No	No	No	No
One policy that provides everyone including the Sales group, access to the root directory.	Yes	No	Yes	No
Two policies that are defined as follows: <ul style="list-style-type: none"> Provide everyone access to the root directory. Deny the Sales group, access to the <code>hdfs-confidential</code> directory. 	Yes	No	No	No
Three policies that are defined as follows: <ul style="list-style-type: none"> Provide everyone access to the root directory. Deny the Sales group, access to the <code>hdfs-confidential</code> directory. Deny the user, access to the <code>hdfs-not-for-user</code> directory. 	Yes	No	No	No

Deploy Apache Ranger

Deploy the Apache Ranger service and enable or create a deny policy. See these documents:

- [Dell EMC Isilon: Apache Ranger Setup and Operations](#)
- [Hortonworks Security Guide: Installing Apache Ranger](#)

A summary of the workflow follows:

1. On a per-access zone basis, perform the following steps using the OneFS Web administration interface or the OneFS command-line administration interface to configure Ranger:
 - a. Enable the Ranger plug-in on OneFS.
 - b. Specify the URL of the Apache Ranger Management console to use port 6182 for https and 6080 for http to get the policies.
 - c. Specify the name of the service instance.

See the [OneFS with HDFS Reference Guide](#) for details.
2. Ensure that a Ranger service account user is configured on OneFS within your access zone.
3. Install Ranger using the steps outlined in the [Hortonworks Security Guide](#).

Note

Do not assign Ranger components to the OneFS host as mentioned in the above-mentioned guide.

4. Enable the Apache Ranger HDFS plug-in using the steps outlined in the [Hortonworks Security Guide](#).
5. If you have a Kerberos-enabled cluster, follow the instructions in the [Hortonworks Security Guide](#) to enable the Ranger HDFS plug-in on the cluster.
6. Enable the Ranger Deny policy using the instructions in the [Apache Ranger deny policies with OneFS 8.0.1.0](#) article.

Create a service instance for OneFS using the **Create Service** page in Ranger. See the [Hortonworks Security Guide](#) for details. Specify values in the following fields to create a service instance for OneFS and make note of the values:

- Specify a value in the **Service Name** field and make note of it because you must use the same value in OneFS.
- Specify a username and in the **Config Properties** section specific to the service instance. The **Test Connection** option continues to fail until you have saved and reopened the service instance.
- Specify the Namenode URL as `FQDN: hdfs://onefs.smartconnect.name:8020`.

A service instance is created with a default policy `all - path`, granting access to all the files for the user that you included in the **Service Details** page.

7. Add all your groups and individual users who are associated with an access zone within OneFS to the default policy in order to grant access to the groups and users. If you create local users in OneFS, or use Active Directory, you must change the **UserSync** settings in Ambari or add the users in the Ranger interface.

Note

OneFS file system permissions take precedence even if the policy indicates that the user or group can access everything.

8. Using the **Edit Policy** page in the Ranger interface, specify the group or users who have limited access to the repository within OneFS and indicate the permissions that must be denied to that path.
9. Create a DENY policy in Ranger using the steps outlined in the [Hortonworks Security Guide](#), if required. After you have saved the policy, OneFS enforces the policy at the next download. If you attempt to take action on a path that is denied access by the Ranger policy, this will be reported in the OneFS HDFS log at `/var/log/hdfs.log`. For more information, see the [Apache Ranger deny policies with OneFS 8.0.1.0](#) article.

Ambari metrics and alerts overview

A OneFS node can monitor, collect, and push metrics data at one-minute intervals to the Ambari Metrics Collector which is one of the components of the Ambari Metrics System. Ambari Management Pack for Isilon OneFS presents OneFS access zones as a service. OneFS is identified by Ambari as a single host running the HDFS service, even though it is a clustered file system. As a result, all the metrics and alert data that is provided by OneFS to Ambari are cluster-wide. For example, for a three-node OneFS cluster, the network HDFS traffic aggregated across all the three nodes is reported to Ambari.

To use the Ambari Metrics Collector, ensure that Ambari Metrics is deployed and is running (green) on the Ambari dashboard.

Note: OneFS metrics for specific access zones that contain HDFS dataset is not supported.

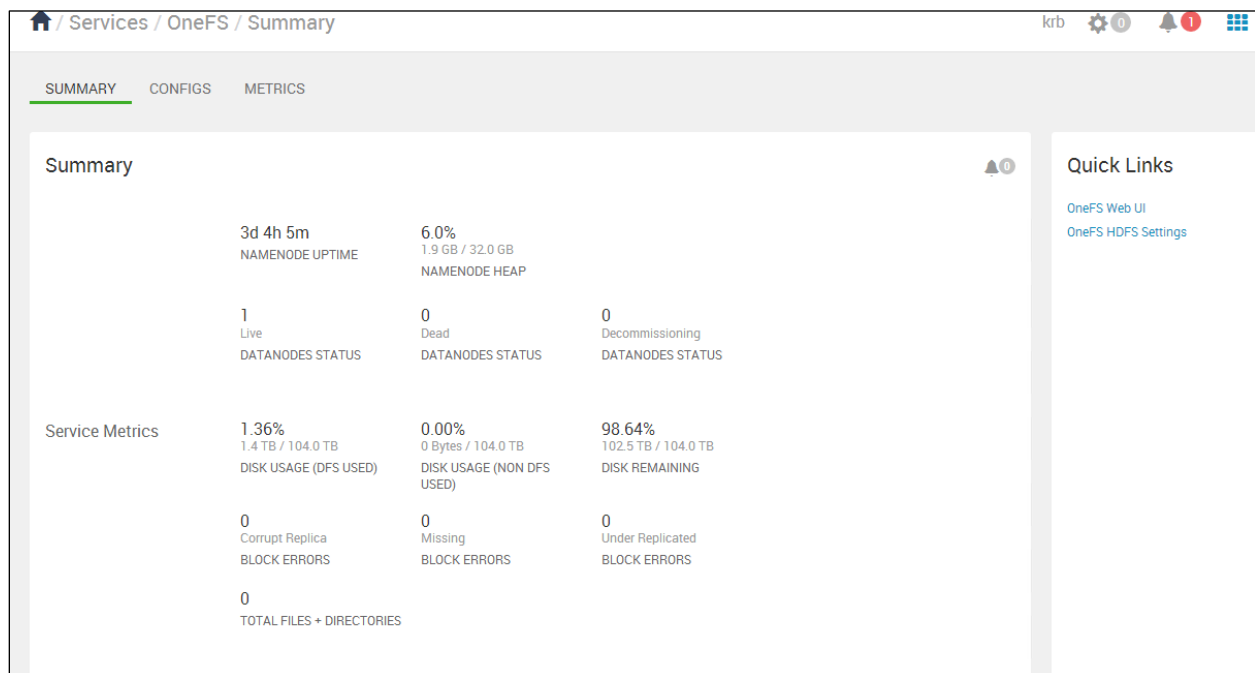
View Ambari metrics and alerts

In order to view the Ambari metrics and alerts, follow the steps that are given below:

1. Access Ambari Web by opening a supported browser and entering the Ambari Web URL.
2. Click **Ambari Metrics Service** > **Metrics Collector** to determine the hostname where Ambari Metrics Collector has been installed.
3. From a node in your OneFS cluster, run the following command to set the access zone and to specify the name of the external Ambari host where the Ambari Metrics Collector component is installed:

```
isi hdfs settings modify --zone=ZONE --ambari-metrics-collector=<FQDN of metrics collector>
```

4. From the Ambari Web home page, select the **OneFS** service and verify that Ambari can collect metrics details from the OneFS SmartConnect zone as shown in the following sample screen:

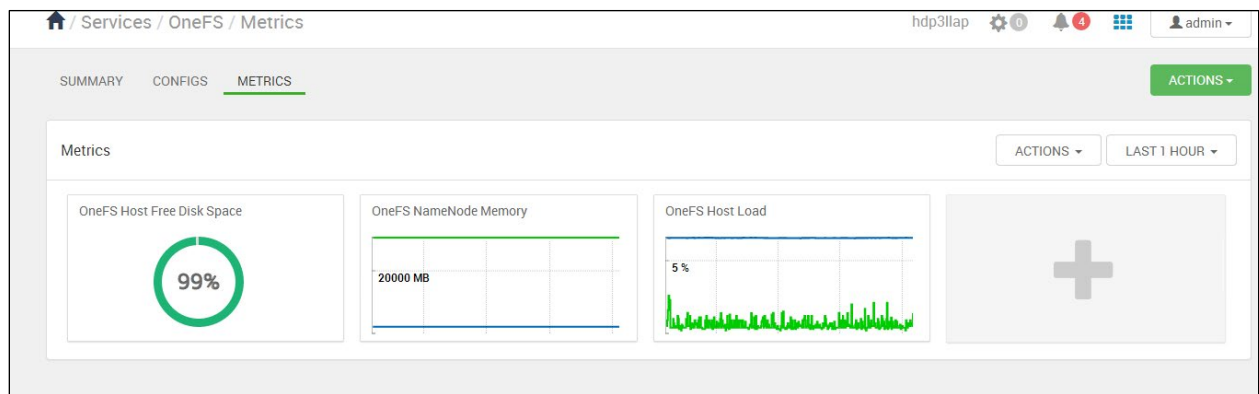


The following OneFS metrics data is reported:

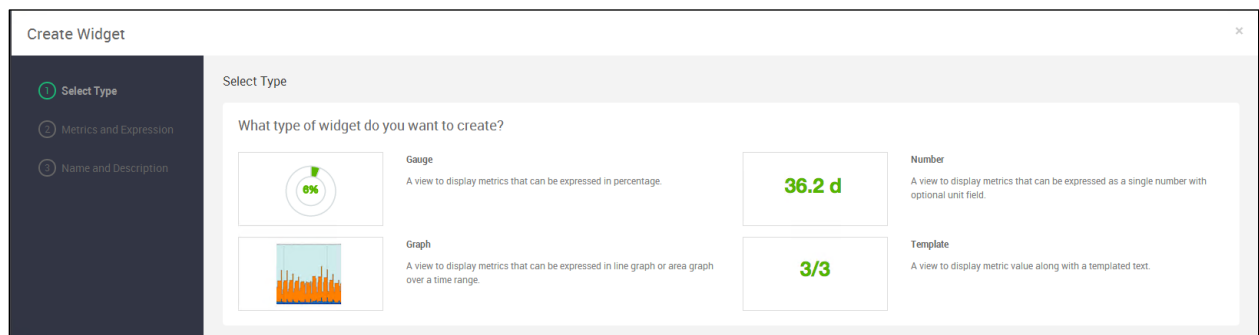
Metric	OneFS data reported
NameNode Uptime	The OneFS node that is running for the longest time.
NameNode Heap	Used: The sum of the current memory allocated by the HDFS process (cluster-wide).

Metric	OneFS data reported
Datanodes Status	OneFS node status (OneFS represents a single datanode).
Disk Usage (DFS Used)	OneFS Storage space used for distributed filesystem.
Disk Usage (Non-DFS Used)	OneFS Storage space used for non-distributed filesystem.
Disk Remaining	OneFS storage remaining.
Block Errors	Always one because there are no blocks in OneFS.
Total Files + Directories	Always zero. The file system does not maintain the file count.

5. From the Ambari Web home page, select the **OneFS** service and then click the **Metrics** tab to create widgets to monitor and view OneFS metrics data.



- a. Click (+) to open the **Create Widget** wizard.
- b. Select one of **Gauge**, **Number**, or **Graph** widget types for creating the widget. Alternatively, you can create a widget using a new template.



- c. On the **Metrics and Expression** screen, perform the following steps:
 - i. Provide a widget title.
 - ii. Under **Expression**, click **ADD METRIC > OneFS > All OneFS Clients** and then select a metric.

Create Widget

Metrics and Expression

Define the expression with any metrics and valid operators.
Use parentheses when necessary.

OneFS Test Widget

Expression:

+ ADD METRIC ▾ + ADD OPERATOR ▾ + ADD NUMBER

OneFS ▸ All OneFS Clients ▸ Select a Metric ▾

Select a Metric

- bytes_in
- bytes_out
- cpu_idle
- cpu_nice
- cpu_num
- cpu_system
- cpu_user
- cpu_wio

Add Metrics and operators here...

Graph Type ▾ LINE ▾

Unit Optional: MB, ms, etc.

← BACK NEXT →

- iii. Click **Next**.

- d. On the **Name and Description** screen, provide the necessary details and click **SAVE**.

Create Widget

Name and Description

Name ▾ OneFS Test Widget

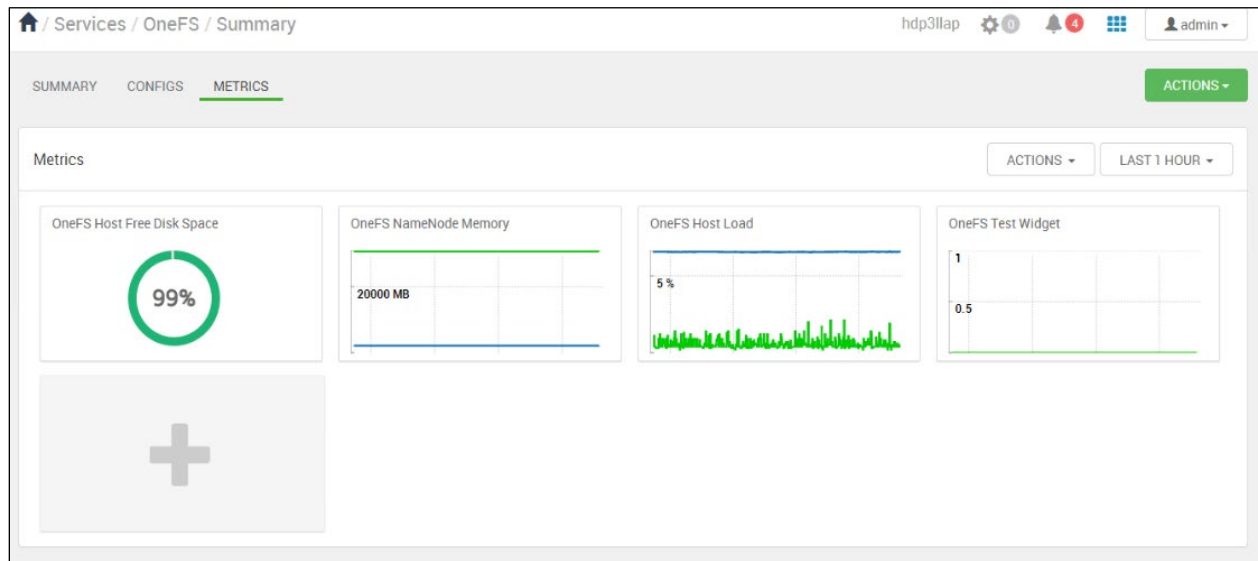
Author admin

Sharing ☒ Share this widget in the widget library

Description

← BACK CANCEL SAVE

The new widget is created and appears on the **Metrics** page as shown:



HDFS wire encryption overview

You can protect data that is transmitted between an HDFS client and OneFS through the data in-flight encryption that is also known as the HDFS wire encryption. In a Kerberos-enabled Hadoop environment, you can enable this feature on all the HDFS clients and on OneFS. Wire encryption manages the negotiations between an HDFS client and OneFS to encrypt and decrypt data.

Note:

- You can enable wire encryption per access zone in OneFS.
- Enabling HDFS wire encryption with an Access Zone could result in HDFS traffic performance degradation while accessing data in that zone. You can characterize the performance impact as wire encryption enabled to determine whether this is acceptable to your workload.

Configure HDFS Wire Encryption with OneFS

To use the wire encryption feature, you must deploy Kerberos on your Hadoop cluster. The following instructions assume that you have already deployed and enabled Kerberos on your cluster. You can then enable HDFS wire encryption in OneFS either using the OneFS web administration interface or the OneFS command-line administration interface.

Note

HDFS wire encryption that is supported by Dell EMC PowerScale is different from the Apache HDFS Transparent Data Encryption technology.

You can configure HDFS wire encryption using the OneFS web administration interface or command-line administration interface. See the [Isilon OneFS HDFS Reference Guide](#) for details.



Configure Apache Hadoop for HDFS wire encryption

To enable HDFS wire encryption with OneFS, edit the following attributes that are associated with Apache Hadoop:

Properties in core-site.xml	Value	Properties in hdfs-site.xml	Value
hadoop.rpc.protection	privacy	dfs.encrypt.data.transfer	true
		dfs.encrypt.data.transfer.algorithm	3des (Default value)
		dfs.encrypt.data.transfer.cipher.suites	AES/CTR/NoPadding (Default value)
		dfs.encrypt.data.transfer.cipher.key.bitlength	Select one of 128,192,356. The default value is 128.

Contacting Dell EMC PowerScale Technical Support

Online Support: <https://support.dell.com/>

Telephone Support:

United States: 800-782-4362 (800-SVC-4EMC)

Canada: 800-543-4782

Worldwide: +1-508-497-7901

Other [worldwide access numbers](#)