Hortonworks DataPlane Service

DPS Platform Administration

docs.cloudera.com

Hortonworks DataPlane Service (DPS™): DPS Platform Administration

Table of Contents

1. Managing Clusters in DPS

From DPS Platform, you can manage the clusters you use with the services that plug into DPS. You can register and enable clusters, and view cluster details.

1.1. Register a Cluster in DPS Platform

You can register Ambari-managed clusters with DPS Platform. Clusters must be registered before you can enable them for use with any DPS service.

Prerequisites

Clusters must have been created by using Apache Ambari prior to registering the clusters with DPS Platform. Clusters that are not managed by Ambari cannot be accessed by DPS Platform or associated services.

About This Task

The DataPlane Admin role is required to perform this task.

> ⚠️ **Important**
>
> After you register a cluster in DPS Platform, do not change the cluster name in Ambari. A cluster name change in Ambari does not currently propagate to DPS Platform. Therefore, replication jobs in Data Lifecycle Manager associated with that cluster name fail, and datasets associated with that cluster name are not discoverable in Data Steward Studio.

Steps

1. Click the (Clusters) icon in the DPS Platform navigation pane.

2. Click **Add Cluster**.

3. Enter the URL or the host name for the Ambari host that manages the cluster:

   - Use the format *https://<IP address>:8080* or *http://<hostname>:8080*, for example: https://10.10.10.10:8080.

   - The host name must either be in a DNS server or configured in the `etc/hosts` file.

   DPS Platform automatically discovers cluster information, such as the cluster name, active services, and metrics.

4. Click **Go**.

   The new cluster displays in the list on the Clusters page.

Next Steps

Enable clusters for use with specific DPS Platform services, such as Data Lifecycle Manager, Data Steward Studio, and so forth.

1.2. Enable a Cluster for a Service

After registering clusters with DPS, you must enable each cluster with one or more DPS services. Clusters must meet specific configuration requirements before they can be used with a DPS service. When you enable a DPS *service*, a check is run to determine if the required service engine, such as DLM Engine, is installed on any clusters. If the cluster meets all requirements for the service, the cluster is automatically enabled. If the service engine is installed, but some configuration is still required on the cluster, the cluster name displays on the Services page with the action button Enable.

You can view a list of enabled clusters on the Services page by selecting the Show All Clusters action for the service.

Prerequisites

Clusters must be managed by Apache Ambari and registered with DPS Platform.

The services you are associating with the clusters must already be enabled in DPS Platform.

About This Task

The DataPlane Admin role is required to perform this task.

You must enable at least two clusters to perform replication jobs with Data Lifecycle Manager.

Steps

1. Click the ⊞ (Services) icon in the DPS Platform navigation pane.

   The Services page displays. Services listed in the table are enabled. Services identified by a tile icon are not currently, but can be, enabled.

2. Click the row for a service.

   A list displays of any clusters that have the required service engine installed but have not yet been configured for use with the service.

   > **Tip**
   >
   > If no clusters display for the service, verify that the clusters you expect to see have been registered with DPS Platform, and that the proper service engine has been installed on the clusters.

3. Click **Enable** for the cluster you want to use with the service.

   A check runs to determine what configuration is required on the cluster for the service you selected. For example, a required service such as Apache Ranger might be installed on the clusters but not enabled in Apache Ambari.

The Manual Install page displays, indicating what you need to configure on the cluster to make the cluster usable by the service.

4. Perform the actions stated on the Manual Install page.

   The required actions often involve enabling a service from Ambari. For example:
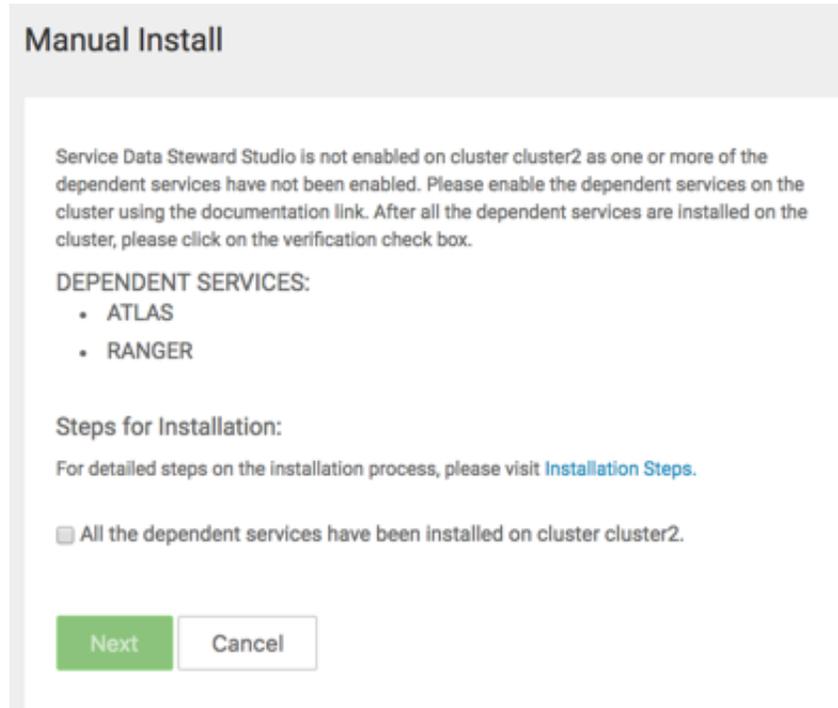
Manual Install

Service Data Steward Studio is not enabled on cluster cluster2 as one or more of the dependent services have not been enabled. Please enable the dependent services on the cluster using the documentation link. After all the dependent services are installed on the cluster, please click on the verification check box.

DEPENDENT SERVICES:
- ATLAS
- RANGER

Steps for Installation:

For detailed steps on the installation process, please visit Installation Steps.

☐ All the dependent services have been installed on cluster cluster2.

Next   Cancel

   You should complete the required actions before proceeding to the next step.

5. Click **All the dependent services have been installed...**, and then click **Next**.

   Another configuration check is run and if all requirements are met, a verification message displays. If the check fails, a message displays identifying the configuration tasks still to be completed.

6. Click the name of the service for which you enabled the cluster, and then enable **Show All Clusters**.

   The new cluster displays in the list on the Clusters page:
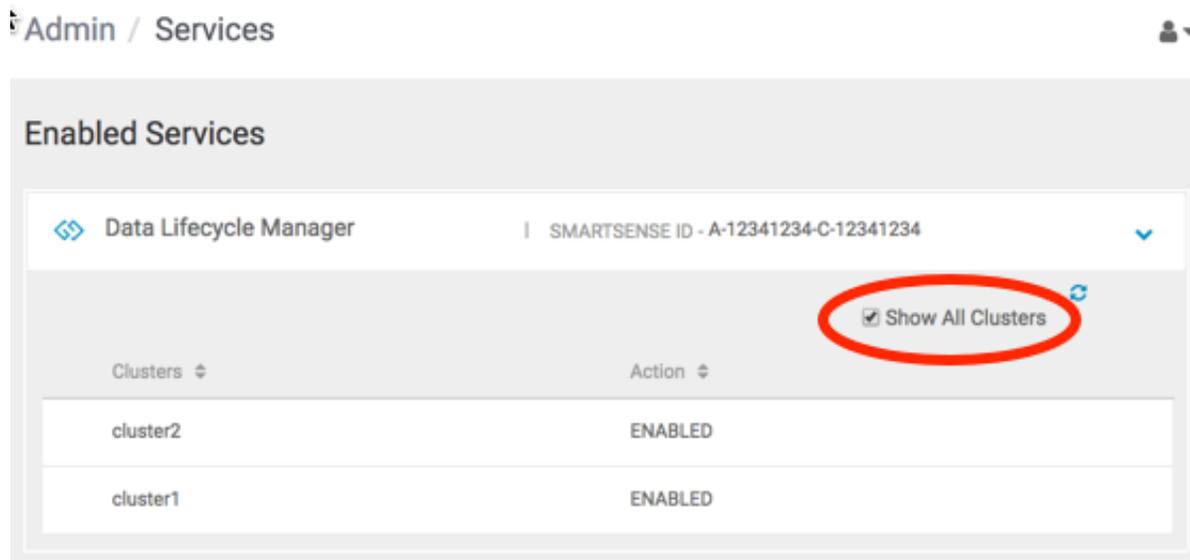
1.3. Edit a Cluster in DPS Platform

You can modify the Cluster Location, Data Center name, Tags, or Description for any cluster registered in DPS Platform.

About This Task

The DataPlane Admin role is required to perform this task.

⚠️ **Important**

> After you register a cluster in DPS Platform, do not change the cluster name in Ambari. A cluster name change in Ambari currently does not propagate to DPS Platform. Therefore, replication jobs in Data Lifecycle Manager associated with that cluster will fail, and information associated with that cluster will not be available for monitoring in Data Steward Studio.

Steps

1. Click the (Clusters) icon in the DPS Platform navigation pane.

2. Optional: Enter a cluster name in the search field and press **Enter**.

    You can only search by cluster name. You can search by partial or full names.

3. In the cluster list, locate the row for the cluster you want to edit.

4. At the end of the row, click the ⋮ (Actions) icon and then click **Edit**.

    The Edit Cluster page displays.

5. Modify the cluster details.

6. Click **Update**.

The Clusters page displays a list with the updated cluster.

1.4. View Cluster Details in DPS Platform

DPS Platform provides two levels of detail about each cluster. You can view general information about a cluster such as status, location, HDP version, number of nodes, and so forth, from the Clusters page. You can also view information such as the status of DataNodes and NodeManagers, heap size, disk space, and so forth, from the cluster Details page.

About This Task

The DataPlane Admin role is required to perform this task.

Steps

1. Click the (Clusters) icon in the DPS Platform navigation pane.

2. In the cluster list, locate the row for the cluster you want to edit and click the cluster name.

   The Details page displays more information about the selected cluster.

3. Refresh the values on the Details page by clicking the (Refresh) icon.

   Updated data is pulled from Ambari.

4. To view or edit cluster information in Ambari, click the ⋮ (Actions) icon and then click **Go to Ambari**.

   A new browser tab opens to the login page for the Ambari host that manages the cluster.

1.5. Add Host Entries to the /etc/hosts File on the Container

If you are using hosts that are not publicly addressable from a DNS server, you must add IP address and host name mapping to the `/etc/hosts` file of each associated *DPS container*. Ensure that all hosts to be used with DPS are publicly addressable, including the Knox hosts.

You must use the specific method and format identified in this task to make the mapping information usable by DPS Platform.

About This Task

Perform this task only *after* DPS Platform is installed and running.

This task is required only for `/etc/hosts` handling. It is not required if the Ambari and cluster hosts have valid DNS entries in a DNS server.

1. Log in as the root user to a terminal on the DPS host.

2. Type the following command to add the host to the `/etc/hosts` file of the container:

```
dpdeploy.sh utils add-host <ip> <host>
```

3. Repeat this procedure on each cluster registered with DPS Platform.

More Information

dpdeploy.sh Script Command Reference

2. Managing Users and Groups

From the DPS Platform UI, you can add and edit users and groups for all services associated with DPS. Users and groups must exist in the corporate LDAP directory before you can add them in DPS Platform.

2.1. Add a User or Group

When you install Hortonworks DPS, the DataPlane Admin role is created. This role has access to DPS Platform and all DPS services. You should add additional users with permissions limited to the service or services you want the user to access.

You can also add groups, which enable you to more easily manage users. You might configure groups for all users who can perform specific tasks, such as creating replications, creating datasets, and so forth.

Tip

User-level assignments override group-level assignments.

Prerequisites

User and group accounts must already exist within your corporate LDAP directory.

About This Task

The DataPlane Admin role is required to perform this task.

Steps

1. Click the (Users) icon in the DPS Platform navigation pane.

2. Click **Add User**.

3. In **Usernames**, enter the name of a user from your corporate LDAP directory, and then click the name when it pops up.

   Tip

   You must click the name of the user when it displays and ensure it appears in the Username field on a dark background.
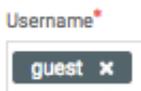
   If the name appears on a white background, it means the name is not recognized and the action fails.

4. Select one or more of the following roles to assign to the user:

   DataPlane Admin   Can perform all actions in DPS Platform, and can access and perform all actions in the UI of enabled services.

| Infra Admin | Can perform all actions in the Data Lifecycle Manage (DLM) service UI, and can manage DLM-enabled clusters in DPS Platform. |
| Data Steward | Can perform all actions in the Data Steward Studio (DSS) service UI, and can manage DSS-enabled clusters in DPS Platform. |

5. Click **Save**.

   The new user displays in the list on the Users page.

6. Click the **Groups** tab, and then click **Add Group**.

7. In **Group Names**, enter the name of a group from your corporate LDAP directory, and then click the name when it pops up.

> **Tip**
>
> You must click the name of the group when it displays and ensure it appears in the Group Name field on a dark background.
>
> Group Names*
>
> [hdfs ✕]
>
> If the name appears on a white background, it means the name is not recognized and the action fails.
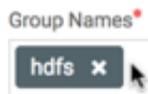
8. Select the roles to assign to the group.

9. Click **Save**.

   The new group displays in the list on the Groups page.

2.2. Edit a User or Group

You can edit any user or group that has already been added to DPS Platform.

> **Tip**
>
> User-level assignments override group-level assignments.

About This Task

The DataPlane Admin role is required to perform this task.

Steps

1. Click the ▦ (Users) icon in the DPS Platform navigation pane.

   The Users page displays a list of existing users, their roles, and status.

2. Locate the user you want to edit by browsing the user list or entering a user name in the search field.

   You can only search by user name. You can search by partial or full names.

3. Select the user to edit by doing one of the following:

   - Click a user name in the list of users.

   - Click ⋮ (Actions icon) and Edit.

   The user's information displays in a slide-out panel.

4. Change the user status, or add or delete roles.

5. Click **Save**.

   The modifications are shown in the Users list.

6. Click the **Groups** tab.

7. Locate the group you want to edit by browsing the group list or entering a group name in the search field.

   You can only search by group name. You can search by partial or full names.

8. Select the group to edit by doing one of the following:

   - Click a group name in the list of groups.

   - Click ⋮ (Actions icon) and Edit.

   The group's information displays in a slide-out panel.

9. Add or delete roles assigned to the group.

10. Click **Save**.

   The modifications are shown in the Groups list.

2.3. Add a Predefined User

You can add a predefined ("seeded") user to the LDAP usernames available to DPS Platform. This predefined user can be made available to employees as a standard login. The user would have global access to all Ambari-managed clusters in your organization and would be equivalent to the DataPlane Admin user provided as a default user during DPS installation.

About This Task

Root access to the DPS Platform host is required to perform this task.

Perform this task only *after* DPS Platform is installed and running.

Steps

1. Log in to a terminal as root user.

2. Type the following command:

```
./dpdeploy.sh utils update-secrets
```

3. When prompted, enter a username and password.

More Information

dpdeploy.sh Script Command Reference

2.4. Change the Local Admin Password

The local DPS admin password is set during DPS installation. You use the local admin role when you log in to DPS the very first time. From DPS Platform, you can change the password for the local admin. You cannot change the password for LDAP users from the DPS UI, nor can you retrieve or reset the local admin password if you forget it.

Steps

1. In the DPS Platform UI, click the 👤▾ (Admin) icon.

2. Click **Change Password**.

3. Enter the new password.

⚠️ **Important**

The password cannot be retrieved or reset by any user. If you forget the password, you must contact Hortonworks Support to reset the password.

2.5. Edit LDAP Settings

If the corporate LDAP administrator Bind DN and password are changed, such as for security purposes or policy requirements, or the LDAP server URL changes, such as due to hardware issues, the DataPlane Admin can modify those settings in DPS so that users can continue to log in to DPS.

About This Task

You must be logged in as the DataPlane Admin to complete this task.

Steps

1. In the DPS Platform UI, click the 🖼 (Users) icon.

2. To the right of the LDAP URL, click the ⋮ (Actions) icon, and then click **Edit**.
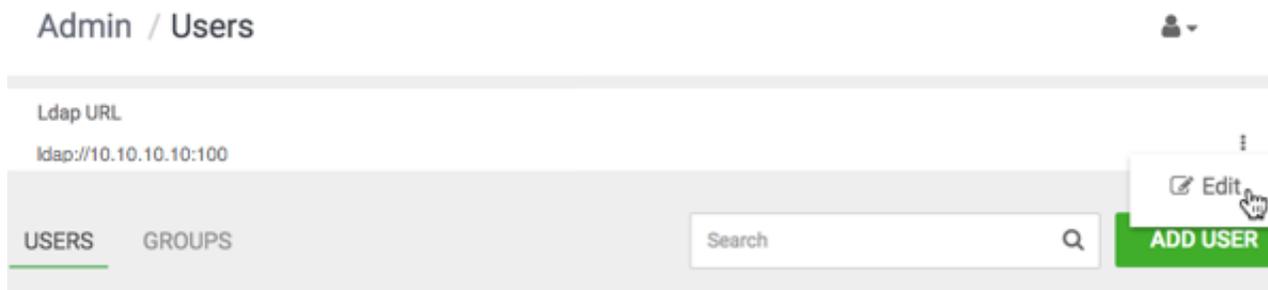
3. In the Edit LDAP page, you can modify the following settings:

- URL

- Administrator Bind DN

- Administrator Password

You cannot modify any other LDAP settings from DPS.

4. Click **Save**.

3. Managing DPS Services

3.1. Enable a Service

You must enable, through DPS Platform, any DPS service you want to use.

Before enabling a service, you must have properly installed and configured the service and the associated service engine.

Prerequisites

You must have a SmartSense ID available. You can retrieve the SmartSense ID from the Hortonworks Support Portal under the Tools tab.

About This Task

The DataPlane Admin role is required to perform this task.

Steps

1. From DPS Platform, click the ⊞ (Services) icon in the navigation pane.

   The Services page displays. Services listed in the table have been enabled. Services identified by a tile icon are available to be enabled.

2. Hover the cursor over the tile for the service you want to enable and click the **Enable** button that appears.

   A verification page displays.

3. Enter the case-sensitive SmartSense ID and click **Verify**.

4. Click **Next**.

   The enabled service displays in the Enabled Services list on the Admin/Services page:

4. Hortonworks DataPlane Service Terminology

You should be familiar with the terminology and concepts used in Hortonworks DataPlane Service (DPS) and in the services that interface with the DPS infrastructure.

4.1. DPS Core

DPS Core is a UI service platform from which you can manage and monitor various services on multiple Hortonworks Data Platform (HDP) Hadoop clusters. You can install DPS Core on an HDP cluster or remote to the cluster.

Hortonworks DataPlane Service (DPS)	The family of components that include the DPS Core service platform and all services that plug into it.
service	An autonomous component in the DPS environment. DPS Core is a service, as is each component that is enabled and managed through DPS Core, such as Data Lifecycle Manager (DLM). DPS Core and each of its plugin services must be installed as Docker containers.
cluster	A typical HDP Hadoop cluster. See the *Cluster Planning* guide for details. The cluster hosts the various Systems of Record (SoRs) for metadata (Apache Hive, Apache Atlas, Apache Ranger, HDFS, and so on) that DPS Core and associated plugin services rely on. In an on-premise environment, a cluster often equates to a data center. However, a single data center can contain multiple HDP Hadoop clusters.

4.2. Data Lifecycle Manager (DLM) Service

DLM is a UI service that is enabled through DPS Core. From the DLM UI you can create and manage replication and disaster recovery policies and jobs.

DLM Engine	Also referred to as the Beacon engine, the backend replication engine required for Data Lifecycle Manager. The DLM engine must be installed as a management pack on each cluster that is to be used in data replication jobs. The engine maintains, in a configured database, information about clusters and policies that are involved in replication.
data center	The facility that contains the computer, server, and storage systems and associated infrastructure, such as routers, switches, and so forth. Corporate data is stored, managed, and distributed from the data center. In an on-premise environment, a data center is often composed of a single Hadoop cluster.

4.3. Data Steward Studio (DSS) Service

DSS is a UI service that is enabled through DPS Core.

DSS Profiler Agent	Enable the data steward to gather and view information about data distribution. For example, view the distribution between males and females in column "Gender", or min/max/mean/null values in a column named "avg_income". The profilers run at regularly scheduled intervals and periodically generate profiled data. Profilers work with data sourced from Apache Ranger Audit Logs, Apache Atlas Metadata Store, and Apache Hive.
asset collection	A list of assets that have been grouped and assigned unique search criteria by a data steward for purposes of management, administration, and so forth. Allows for organizing a data lake by applying uniform policies and providing access to a limited set of users. The content of an asset collection is a static list that can be modified only by a user. Therefore, adding new assets to a collection must be done manually.
data asset	A specific instance of a data type, including the related attributes and metadata, that is typically managed as a single unit by components like Atlas and Ranger. A data asset could include a specific instance of an Apache Hive database, table, or column; an HBase namespace, table, or column family; an individual file or a collection of files, and so forth. An asset can belong to only one asset collection. Data assets are also known as "entities" in Atlas.
data lake	A trusted and governed data repository that stores, processes, and provides access to many kinds of enterprise data to support data discovery, data preparation, analytics, insights, and predictive analytics. In the context of Hortonworks DPS, a data lake can be realized in practice with an Apache Ambari-managed Hadoop cluster that runs Apache Atlas for metadata and governance services, and Apache Knox and Apache Ranger for security services.

5. Roles Required to Work with DPS Services

To perform actions in DPS Platform and associated services, you must be an DPS administrator, an infrastructure administrator, or a data steward. In addition, to perform actions in Apache Ambari that impact DPS (such as creating clusters, changing configuration settings for services, and so forth), you must be an Ambari administrator or a cluster administrator.

Other roles might be required during installation, depending on your configuration. See the installation instructions for roles required during installation.

5.1. Roles for DPS Services

The following roles are available to perform actions in the DPS UI.

5.1.1. DPS Admin

- An DPS Admin role is created during installation, so you can initially log in to DPS Platform

- Can access DPS Platform and perform all actions in DPS Platform related to clusters, users, and enabling services

- Can access all services enabled with DPS Platform, and perform the same actions as each administrator role assigned to the enabled services, such as Infra (infrastructure) Admin, Data Steward, etc.

5.1.2. Infra Admin

- Can access DPS Platform service to manage clusters enabled for Data Lifecycle Manager (DLM), but cannot perform any other DPS Admin actions

- Can access and perform all actions in Data Lifecycle Manager

- Has read-only access to browse, within the DLM UI, the folder structure of any cluster enabled for DLM

  - Cannot view the *content* of the source or copied files or databases, from the DLM UI

  - Cannot modify or delete folders or databases that you can view from the DLM UI

  - This access is available to all users logged in to DLM as the Infra Admin, even if they do not have permissions to view the directories or databases as a Linux user or Kerberos principal

- Can create replication policies for any folders or databases on clusters enabled for DLM, and replicate the data objects by using the DLM Engine

This ability is available to all users logged in to DLM as the Infra Admin, even if the users do not have access to the target path on the target cluster.

5.1.3. Data Steward

- Can access the DPS Platform service to see and manage clusters enabled for Data Steward Studio

- Can access Data Steward Studio and monitor all data in the DSS UI

    Interactions are only with Apache Atlas and Apache Ranger, so Atlas and Ranger authorization must be set up to enable read access.

- Cannot currently access any file contents from HDFS or Apache Hive.

5.2. Roles Required for Installation and Troubleshooting

You also need the Ambari Admin or Cluster Admin roles to install Hortonworks DPS, add clusters to Ambari, troubleshoot cluster issues, and so forth.

See Apache Ambari Administration for further details about these roles.

5.2.1. Ambari Admin

- Has full control over all aspects of Ambari

- Can install HDP by using the Ambari installation wizard

- Can install the DLM Engine (Beacon) management pack and configure the DLM Engine

- Can start and stop the DLM Engine and troubleshoot cluster problems

- *Cannot* access DPS Platform or any enabled service in DPS Platform

5.2.2. Cluster Admin

- Has control over a cluster, its hosts, and services

- Can create clusters to be registered with DPS Platform

- Can start and stop the DLM Engine and troubleshoot cluster problems

- *Cannot* access DPS Platform or any enabled service in DPS Platform

5.3. Required Roles by Task

The following tables indicate the roles required to perform various tasks in DPS Platform and the associated services.

5.3.1. DPS Platform Tasks and Required Roles

The following table shows tasks you can perform that are related to DPS Platform and the roles required to perform the tasks. Except for installation of DPS, and unless otherwise noted, all tasks that require the DataPlane Admin role are performed within the DPS Platform UI.

Task	DataPlane Admin	Infra Admin	Data Steward	Ambari Admin	Cluster Admin
Install HDP using Ambari				X	
Install DPS Docker image (CLI)	X				
Install service (DLM, DSS) Docker image	X				
Configure LDAP	X				X
Enable DPS services (DLM, DSS)	X				
Manage DPS users	X				
Manage DPS clusters (register, delete, etc.)	X				
Monitor clusters in DPS	X	X	X		
Create a cluster				X	X

5.3.2. Data Lifecycle Manager Tasks and Required Roles

The following table shows tasks you can perform that are related to DLM and the roles required to perform the tasks. Unless otherwise noted, all tasks that require the Infra Admin role are performed within the DLM UI.

Task	DataPlane Admin	Infra Admin	Data Steward	Ambari Admin	Cluster Admin
Install DLM Engine MPack				X	
Enable DLM	X				
Log in to DLM		X			
Register clusters	X				
Pair clusters		X			
Browse HDFS folders in DLM UI		X			
Browse Hive databases in DLM UI		X			
Create or schedule a replication policy (HDFS or Hive)		X			
Manage a replication policy (suspend, delete, resume, etc.)		X			

Task	DataPlane Admin	Infra Admin	Data Steward	Ambari Admin	Cluster Admin
Monitor a replication job (HDFS or Hive)		X			
Create or schedule a DR policy (HDFS or Hive)		X			
Manage a DR policy (suspend, delete, resume, etc.)		X			
Monitor a DR job (HDFS or Hive)		X			
Monitor DLM alerts		X			
Access DLM log information		X			
Allocate DistCp jobs to YARN queue		X			
Allocate bandwidth		X			
Start or stop the DLM (Beacon) Engine in Ambari				X	X
Access Ambari for troubleshooting				X	X

5.3.3. Data Steward Studio

The following table shows actions you can perform that are related to DSS and the roles required to perform the actions.

Task	DataPlane Admin	Infra Admin	Data Steward	Ambari Admin	Cluster Admin
Enable DSS	X				
Log in to DSS		X			
Register clusters	X				
Start or stop the DSS Profiler engine				X	
Access Ambari for troubleshooting				X	X

More Information

Apache Ambari Administration

6. Troubleshooting DPS

To verify that your environment meets the requirements for DPS, see DPS Support Matrix.

6.1. Endpoint not accessible

DLM does not support updating any cluster endpoints (HDFS, Hive, Ranger, or DLM Engine). If an endpoint must be modified, contact Hortonworks Support for assistance.

6.2. Logging in Using the DPS Local Admin Role

When you log in as the local DPS Admin, you bypass Knox. The local admin role allows you to perform administrative activities and troubleshoot problems when access through LDAP and Knox is not available. The local admin is also the role you use to log in to DPS the first time, before LDAP is configured in DPS for SSO.

About This Task

For login, the default username is "admin". The password you use to log in is set during the installation process.

Steps

1. Log in by appending `/sign-in` to the DPS login URL, for example:

```
http://dataplane-host-name/sign-in
```

6.3. Ranger UI Does Not Display Deny Policy Items

Description: When a policy with deny conditions is created on Ranger-Admin for a replication relationship, the Policy Details page in Ranger should show the deny policy items also. However, the deny policy items do not display in the Policy Details page.

Workaround: If `deny` policy items do not appear on the Ranger admin Policy Details page, first enable deny conditions for policies from **Ambari>Ranger>Configs>Advanced>Custom ranger-admin-site**, add **`ranger.servicedef.enableDenyAndExceptionsInPolicies=true`**, and then restart target ranger-admin.