Getting Started

Date of Publish: 2018-05-18

http://docs.hortonworks.com

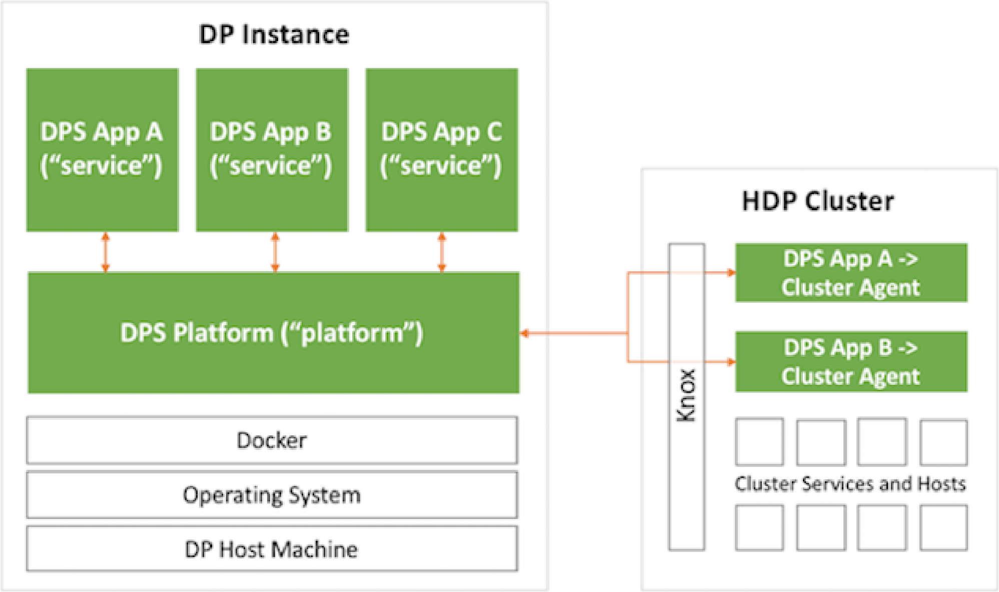
Contents

DPS Concepts

DPS overview

Hortonworks DataPlane Service (DPS) is a portfolio of data solutions that will support the management & discovery of data (whether at-rest or in-motion) and enable an enterprise hybrid data strategy (from the data center to the cloud).

DPS is composed of a core platform (“DPS Platform” or “Platform”) and an extensible set of apps (“DPS Apps” or “DPS Services”) that are installed on the platform. Depending on which app you plan to use, you may be required to install an agent into a cluster to support that app, as well as meet other cluster requirements.



DPS Platform terminology

Following is a set of terms that are relevant to understanding DPS.

DPS Platform or Platform

The core platform that runs one or more DPS Apps.

DPS Apps or DPS Services

The set of apps that are available with DPS. These apps runs on the platform host, and in some cases (depending on the app) have a corresponding agent that also needs to be installed in-cluster. Each app also has a set of cluster requirements to support the app features.

Example: Data Lifecycle Manager (DLM) or Data Steward Studio (DSS)

agent

The agent that runs in cluster in support of a DPS App.

Example: DLM Engine (used with the DLM App)

cluster

A Hortonworks Data Platform cluster that is registered with a DP instance, and then used with an app.

This cluster can be running on-premise in your data center or in a cloud environment

DP instance

A deployment of a DataPlane instance. This is where the platform and the apps run, usually on a single host as Docker containers.

Apache Knox

Provides a single-point access point for authentication and proxy of services. Knox is used under-the-hood in your DP instance to handle authentication to DP. Knox is also used in your clusters to handle Single Sign-On (SSO) and (in some cases) act as a cluster API proxy gateway.

LDAP

LDAP or Active Directory (AD) is used at the authentication source for DP and your clusters.

Related resources

Learn more about DPS and related technologies with the following resources:

Resource	Link
Hortonworks DataPlane Service	https://hortonworks.com/products/data-services/
Hortonworks Data Platform	https://hortonworks.com/products/data-platforms/hdp/
Apache Knox	https://hortonworks.com/apache/knox-gateway/
Apache Ranger	https://hortonworks.com/apache/ranger/
Docker	https://www.docker.com/

Planning for a DPS installation

General requirements for DPS installation

Understanding the requirements and recommendations indicated below can help to avoid common issues during and after DPS installation.

- Be sure to review the *DPS Platform Support Matrix* to confirm you meet the environment and system requirements, including Docker and networking.
- You need to have root access to the nodes on which all DPS services will be installed.
- Every host name used with DPS must be resolvable by DNS or configured in the /etc/hosts file on the DPS container, so that host names can be resolved between all cluster nodes.

  Using a DNS server is the recommended method, but if the instances are added to /etc/hosts, you must explicitly register the cluster host names within the DPS Docker containers. It is not sufficient to have the host names included in the /etc/hosts file on the DPS Platform host. See the DPS Platform Administration guide for instructions.
- If you are using AWS, do not use the public DNS to access DPS.

  Use a public IP address or set up and use a DNS (Route 53) fully qualified domain name (FQDN).
- Have your enterprise LDAP details available.

  See *Enterprise LDAP Requirements* for more details.
- Determine which DPS Apps you plan to install and which cluster(s) you plan to add to DPS.

  Be sure to review the App-specific documentation thoroughly to make sure you can meet the App-specific requirements. For example, depending on your choice of Apps, your cluster requirements might change. This includes (but is not limited to) a minimal HDP version, setup and configuration of Knox, and other cluster requirements. See *Preparing Your Cluster* for more details.

The high-level installation procedure involves two work streams:

Installing DPS & the DPS Apps	Install and configure the DPS Platform and your target DPS Apps. Proceed to the DPS Platform Installation guide.
Preparing Your Clusters for DPS	Prepare your clusters, which can include upgrading, adding and configuring Knox, and adding required DPS Agents (per your choice of DPS Apps). Proceed to General requirements for clusters.

Enterprise LDAP requirements

You need your enterprise LDAP settings available the first time you log in to DPS in order to configure DPS for authentication and authorization. Ensure you have the correct settings available and ready to use as part of your DPS setup. The following table details the properties and values you need to know to set up LDAP with DPS.

Property	**Description**	**Example**
LDAP URL	The hostname and port for the LDAP or Active Directory server	ldap://my.ldap.server:389 ldaps://my.ldap.server:689
Upload Certificate File	If you are using LDAPS and a self-signed certificate, you need to upload the certificate to DPS so that DPS can validate the LDAPS connection.	SSL certificate file
Administrator Bind DN	The Distinguished Name (“DN”) for the manager	cn=Administrator,ou=srv,dc=hortonworks,dc=local
Administrator Password	The password for the DN	Your_password
User Search Base	The root Distinguished Name to search in the directory for users	ou=Users,dc=hortonworks,dc=local
User Search Attribute	cn	uid
User Object Class (optional*)	The object class that is used for users	person
Group Search Base	The root Distinguished Name to search in the directory for groups	ou=Groups,dc=hortonworks,dc=local
Group Search Attribute	The attribute for group name	
Group Object Class	The object class that is used for groups	groupofnames
Group Member Attribute Name	The attribute for group membership	member

Preparing your clusters

General requirements for clusters

Understanding the requirements and recommendations indicated below can help to avoid common issues during and after DPS installation.

You must perform a minimum set of cluster setup and security actions on each HDP cluster that you plan to register in DPS. You can perform any additional security-related tasks on your cluster as appropriate for your environment and company policies.

The following provides a high-level overview of the requirements for DPS and DPS Apps.

Important: Be sure to refer to the cluster and security setup requirements for each of the DPS apps you plan to install for exact details.

Cluster Requirements	DPS Platform	DLM	DSS
Knox SSO	Required	Required	Required
Knox Proxy Gateway	Optional (but recommended)	Required	Required
DPS Agent	n/a	DLM Engine	DSS Profiler Service
Cluster Services	n/a	Refer to DLM documentation	Refer to DLM documentation

Related Concepts

Knox SSO with DPS

Knox Gateway proxying with DPS

Related Information

DataPlane Service documentation

Knox SSO with DPS

DPS Platform and the DPS Apps leverage Knox SSO to provide users and services with simplified and consistent access to clusters, data and other services. You must configure Knox SSO on the clusters you plan to use with DPS. You will perform this Knox SSO setup on your clusters after you perform the DPS Installation. Refer to *DPS Installation* for more information.

DPS authenticates users against a centralized identity provider in the organization (such as an LDAP or AD). Having Knox SSO setup with your clusters ensures that those users and services are authorized to perform specific actions on the respective clusters, and propagates the identity of the user or service from DPS to the cluster services. You must perform the Knox SSO setup on your clusters after you perform the DPS Installation.

Important:

The Knox SSO of your cluster must be configured to use the same LDAP/AD as your DP instance for user identity to match and propagate between the systems.

Minimally, your cluster requires a Knox SSO configuration to include the following cluster services: *Ambari*, *YARN and HDFS*. Refer to your specific DPS Apps documentation for any additional cluster services that may also be required to be configured in Knox SSO.

Refer to the following documentation on how to configure your cluster for Knox SSO:

Resource	Documentation
Install Knox and enable in Ambari	HDP Security Guide, Install Knox
Configure SSO topology	HDP Security Guide, Identity Providers (IdP)
Configure Knox SSO for Ambari	HDP Security Guide, Setting up Knox SSO for Ambari
Configure LDAP with Ambari	Ambari Security Guide, Configuring Ambari Authentication with LDAP or Active Directory Authentication

Related Concepts

General requirements for clusters

Knox Gateway proxying with DPS

Related Information

DPS Installation

Knox Gateway proxying with DPS

With Knox setup and configured in your cluster, it is optional (but recommended) that you also configure Knox to be a proxy gateway for communication between your DP Instance and your cluster. You must configure Knox Gateway for proxying on the clusters you plan to use with DPS prior to starting the DPS installation process. During DPS installation, you will configure Knox Gateway for DPS.

> **Important:** Configuring Knox Gateway is required if your cluster is configured with Kerberos or with wire encryption. This simplifies certificate management for DPS, as the only security certificate that needs to be managed is for Knox.

Refer to the following documentation on how to configure your cluster for Knox Gateway:

Resource	Documentation
Configure a reverse proxy with Knox	HDP Security Guide, Configuring the Knox Gateway
Configure LDAP with Knox for proxy authentication	HDP Security Guide, Setting Up LDAP Authentication

Related Concepts

General requirements for clusters

Knox SSO with DPS