

Getting Started 1

## Getting Started

**Date of Publish:** 2018-12-14



<http://docs.hortonworks.com>

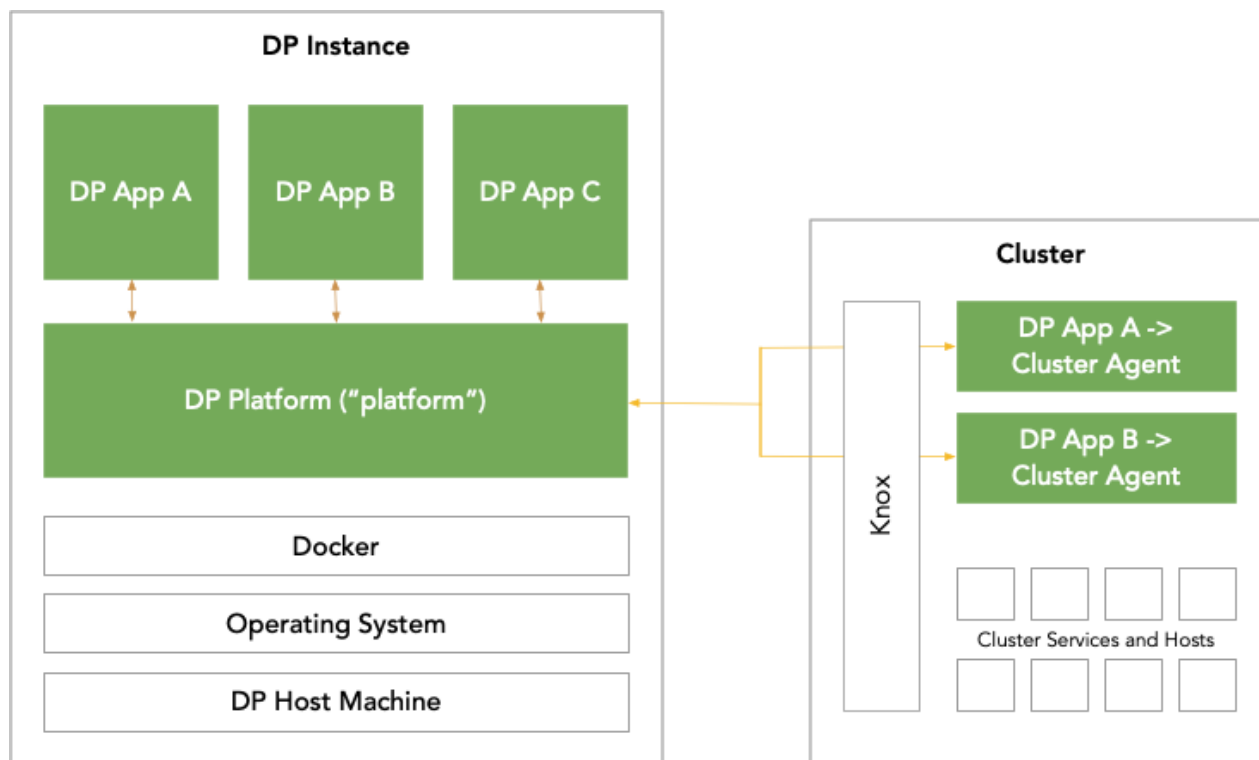
# Contents

<b>DataPlane overview.....</b>	<b>3</b>
<b>DP Platform terminology.....</b>	<b>3</b>
<b>Related resources.....</b>	<b>4</b>
<b>General requirements for DataPlane installation.....</b>	<b>4</b>
<b>Enterprise LDAP requirements.....</b>	<b>5</b>
<b>General requirements for clusters.....</b>	<b>6</b>
<b>Knox authentication for DataPlane clusters.....</b>	<b>6</b>
<b>Knox Gateway proxying with DataPlane.....</b>	<b>7</b>

## DataPlane overview

Hortonworks DataPlane 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).

DataPlane is composed of a core platform (“DP Platform” or “Platform”) and an extensible set of apps (“DP Apps”) 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.



## DP Platform terminology

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

### **DP Platform or Platform**

The core platform that runs one or more DP Apps.

<b>DP Apps</b>	The set of apps that are available with DataPlane. 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)
<b>agent</b>	The agent that runs in cluster in support of a DP App.  Example: DLM Engine (used with the DLM App)
<b>cluster</b>	A Hortonworks Data Platform or Hortonworks DataFlow 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
<b>DP instance</b>	A deployment of a DataPlane instance. This is where the platform and the apps run, on a single host as Docker containers.
<b>Apache Knox</b>	Provides a single-point access for authentication and proxy of cluster services. Knox is used between your DP Instance and your cluster to handle user authentication (via Single Sign-On or Trusted Proxy setup) and (in some cases) acts as a cluster gateway proxy.
<b>LDAP</b>	LDAP or Active Directory (AD) is used as 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	<a href="https://hortonworks.com/products/data-services/">https://hortonworks.com/products/data-services/</a>
Hortonworks Data Platform	<a href="https://hortonworks.com/products/data-platforms/hdp/">https://hortonworks.com/products/data-platforms/hdp/</a>
Hortonworks DataFlow	<a href="https://hortonworks.com/products/data-platforms/hdf/">https://hortonworks.com/products/data-platforms/hdf/</a>
Apache Knox	<a href="https://hortonworks.com/apache/knox-gateway/">https://hortonworks.com/apache/knox-gateway/</a>
Apache Ranger	<a href="https://hortonworks.com/apache/ranger/">https://hortonworks.com/apache/ranger/</a>
Docker	<a href="https://www.docker.com/">https://www.docker.com/</a>

## General requirements for DataPlane installation

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

- Be sure to review the *DP 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 DP services will be installed.

- Every host name used with DataPlane must be resolvable by DNS or configured in the DataPlane containers, 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 DataPlane Docker containers. It is not sufficient to have the host names included in the `/etc/hosts` file on the DP Platform host. See the DP Platform Administration guide for instructions.

- DataPlane supports only PEM-encoded certificates compatible with OpenSSL 1.0.2k or later.
- If you are using AWS, do not use EC2 instance's the public DNS names to access DataPlane.

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 DP Apps you plan to install and which cluster(s) you plan to add to DataPlane.

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 or HDF 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 DataPlane & the DP Apps	Install and configure the DP Platform and your target DP Apps. Proceed to the <a href="#">DP Platform Installation</a> guide.
Preparing Your Clusters for DataPlane	Prepare your clusters, which can include upgrading, adding and configuring Knox, and adding required DP Agents (per your choice of DP Apps). Proceed to <a href="#">General requirements for clusters</a> .

## Enterprise LDAP requirements

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

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
LDAP Certificate File	If you are using LDAPS and with a certificate, you need to have the certificate ready to be uploaded into DataPlane so that DataPlane can validate the LDAPS connection. Note that DataPlane only supports a PEM-encoded certificates and compatible with OpenSSL 1.0.2k or later.	SSL certificate file
Administrator Bind DN	The Distinguished Name ("DN") for the manager	cn=Administrator,ou=srcv,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	The attribute for user names	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

Property	Description	Example
Group Search Attribute	The attribute for group name	cn
Group Object Class	The object class that is used for groups	groupofnames
Group Member Attribute Name	The attribute in a group object that specifies the users that are part of the group	member
Follow referrals	Check this box if you want to follow LDAP referrals	

## General requirements for clusters

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

You must perform a minimum set of cluster setup and security actions on each cluster that you plan to register in DataPlane. 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 DataPlane and DP Apps.



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

Cluster Requirements	DP Platform	Data Lifecycle Manager (DLM)	Data Steward Studio (DSS)	Streams Messaging Manager (SMM)	Data Analytics Studio (DAS)
Knox Authentication	n/a	Knox SSO	Knox Single Sign-On Knox Trusted Proxy	Knox SSO	Knox SSO
Knox Proxy Gateway	Optional (but recommended)	Required	Required	n/a	n/a
Cluster Agents	n/a	DLM Engine	DSS Profiler Service	SMM Rest Server	DAS Event Processor DAS Webapp
Cluster Services	n/a	Refer to DLM documentation	Refer to DSS documentation	Refer to SMM documentation	Refer to DAS documentation

For more information about Knox authentication options, see [Knox Authentication for DataPlane Clusters](#).

### Related Information

[DataPlane Service documentation](#)

## Knox authentication for DataPlane clusters

Apache Knox provides a single access point for authentication and proxy of cluster services to DataPlane.

DP Platform and the DP Apps leverage Knox to provide users and services with simplified and consistent access to clusters, data, and other services. DataPlane authenticates users against a centralized identity provider in the organization (such as an LDAP or AD). Having Knox set up 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 DataPlane to the cluster services.

You must configure Knox on the clusters you plan to use with DataPlane. You will perform this Knox setup on your clusters after you perform the DataPlane Installation. See the [DataPlane Installation](#) for more information. There are two options for configuring Knox in your cluster in order for that cluster to work with DataPlane.

- Knox Single Sign-On (SSO)
- Knox Trusted Proxy Pattern (TPP)



**Important:**

The Knox in 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.

Based on the authentication option you choose to use with your cluster, there is a set of concomitant minimal cluster requirements:

Minimal cluster requirements	Knox Single Sign-On (SSO)	Knox Trusted Proxy Pattern (TPP)
Ambari	Required + LDAP AuthN	Required + LDAP AuthN + Ambari Kerberos AuthN
Kerberos Enabled	Optional	Required
Knox Gateway Proxy	Optional	Required

The use of Knox Gateway Proxy is optional for DP Platform. Depending on which DP Apps you plan to use and how your cluster is configured, the setup of Knox Gateway Proxy may be required. For example, if you are using Data Lifecycle Manager or Data Steward Studio and Wire Encryption is configured in your cluster, you must use Knox Gateway Proxy setup with additional services configured for the proxy. Knox Gateway Proxy is required for the Knox Trusted Proxy Pattern (TPP) authentication option.

Refer to the following documentation on how to configure your cluster for Knox authentication.

Resource	HDP 2.6 and Ambari 2.6 Documentation	HDP 3.0 and Ambari 2.7 Documentation	HDP 3.1 and Ambari 2.7 Documentation
Configure LDAP for Ambari, and sync users	Ambari Security Guide, <a href="#">Configuring Ambari Authentication with LDAP or Active Directory Authentication</a>	HDP Security Guide, <a href="#">Configuring Ambari Authentication for LDAP/AD</a>	<a href="#">Configuring Ambari Authentication with LDAP/AD</a>
Configure SSO topology	HDP Security Guide, <a href="#">Identity Providers (IdP)</a>	HDP Security Guide, <a href="#">Configuring an Identity Provider</a>	<a href="#">Configuring an Identity Provider (IdP)</a>
Configure Knox SSO for Ambari	HDP Security Guide, <a href="#">Setting up Knox SSO for Ambari</a>	HDP Security Guide, <a href="#">Configuring Apache Knox SSO</a>	<a href="#">Configuring Knox SSO</a>

For more information about HDF Knox configuration, see [HDF Security](#) documentation.

## Knox Gateway proxying with DataPlane

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 DataPlane prior to starting the DataPlane installation process. During DataPlane installation, you will configure Knox Gateway for DataPlane.



**Important:** Configuring Knox Gateway is required if your cluster is configured with Kerberos or with wire encryption. This simplifies certificate management for DataPlane, 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	HDP 2.6 Documentation	HDP 3.0 Documentation	HDP 3.1 Documentation
Configure a reverse proxy with Knox	HDP Security Guide, <a href="#">Configuring the Knox Gateway</a>	HDP Security Guide, <a href="#">Configuring the Knox Gateway</a>	<a href="#">Configuring the Knox Gateway</a>

Resource	HDP 2.6 Documentation	HDP 3.0 Documentation	HDP 3.1 Documentation
Configure LDAP with Knox for proxy authentication	HDP Security Guide, <a href="#">Setting Up LDAP Authentication</a>	HDP Security Guide, <a href="#">Set up LDAP Authentication</a>	<a href="#">Configure LDAP Authentication</a>

For more information about HDF Knox Gateway configuration, see [HDF Security](#) documentation.