

Installation and Upgrade with the EMC DSSD D5

This topic provides an overview of the installation of Cloudera Manager and CDH for deployments that use the EMC® DSSD™ D5™ Storage appliance as the storage for Hadoop DataNodes. For deployments that do not use the DSSD D5, see [Installation Overview \(installation_installation.html#xd_583c10bfdbd326ba-7dae4aa6-147c30d0933--7f29\)](#).

Documentation for the EMC DSSD D5 is available from EMC.

Continue reading:

- [Overview of EMC DSSD D5 Integration](#)
- [System Requirements and Limitations for the DSSD D5 Storage Appliance](#)
- [Resources](#)
- [Installing CDH with DSSD D5 DataNodes](#)
- [Upgrading a DSSD D5 Deployment to a New Version of CDH and Cloudera Manager](#)
- [Deployment Options](#)
- [Cloudera Manager Installation Phases](#)
- [Cloudera Manager Installation Software](#)

Overview of EMC DSSD D5 Integration

The EMC DSSD D5 provides a high-speed, low-latency storage solution based on flash media. It has been optimized for use as storage for DataNodes in the Cloudera CDH distribution. The DataNode hosts connect directly to the DSSD D5 using a PCIe card interface. In a CDH cluster, only the DataNodes use the DSSD D5 for storage; all other hosts use standard disks.

To manage clusters that use DSSD D5 storage, enable DSSD Mode in Cloudera Manager. All other Hadoop components operate normally. When this mode is enabled, Cloudera Manager can only manage clusters with DSSD D5 DataNodes; you cannot mix cluster types (a cluster that uses only DSSD D5 DataNodes and a cluster

that uses regular DataNodes). All DataNodes must connect to the DSSD D5; you cannot mix DataNode types within a cluster.

System Requirements and Limitations for the DSSD D5 Storage Appliance

- Only the RHEL 6.6 operating system is supported for DataNode hosts.
- CDH 4 is not supported.
- The HDFS/Sentry sync feature does not work with HDFS on DSSD D5.

For more information about system requirements, see [Product Compatibility Matrix \(Product Compatibility Matrix.html#xd_583c10bfd326ba--6eed2fb8-14349d04bee--7b13\)](#) and [Cloudera Manager 5 Requirements and Supported Versions \(cm_ig_cm_requirements.html#cmig_topic_4\)](#). For information about hardware requirements, contact EMC DSSD Support.

Resources

The following resources are available for managing and installing a cluster using DSSD D5 DataNodes:

- Documentation for the EMC DSSD D5 is available from EMC.
- [HDFS Properties in CDH 5.6.0 \(cm_props_cdh560_hdfs.html\)](#) configuration reference.
- [Cloudera Reference Architecture Documentation \(http://www.cloudera.com/documentation/other/reference-architecture/PDF/cloudera_ref_arch_emc_dssd.pdf\)](#)
- Tuning:
 - [DSSD D5 and Short-Circuit Reads \(dssd_scr.html#dssd_scr\)](#)
 - [Tuning the HDFS Block Size for DSSD Mode \(cm_dssd_block_size.html#concept_sjw_y2k_r5\)](#)

Installing CDH with DSSD D5 DataNodes

Use Cloudera Manager to install a DSSD D5-enabled cluster. You can install Cloudera Manager in several ways, and you can use Cloudera Manager to install agents and other software on all hosts in your cluster. Installing CDH with DSSD D5 DataNodes is similar to non-DSSD D5 installation, except for the following:

- You cannot install a DSSD D5 cluster using a Cloudera Manager instance that is already managing a cluster.
- You set a single property to enable DSSD Mode.
- You set several DSSD D5-specific properties.
- When installing CDH and other services from Cloudera Manager, only *parcel* installations are supported. *Package* installations are not supported. See [Managing Software Installation Using Cloudera Manager \(cm.ig.managing.software.html#xd_583c10bfdbd326ba--6eed2fb8-14349d04bee--772f\)](http://cm.ig.managing.software.html#xd_583c10bfdbd326ba--6eed2fb8-14349d04bee--772f).

Before installing Cloudera Manager, you must complete the following tasks using tools and documentation provided by EMC for the DSSD D5:

- Installing and racking the DSSD D5 Storage Appliance.
- Installing the DSSD D5 PCI cards in the DataNode hosts.
- Connecting the DataNode hosts to the DSSD D5.
- Installing and configuring the DSSD D5 drivers.
- Installing and configuring the DSSD D5 client software.
- Creating a volume on the DSSD D5 for the DataNodes.
- Identifying CPUs and NUMA nodes. See the EMC document *DSSD Hadoop Plugin Installation Guide* for more information. You use the information from this task in a later step to configure the Libflood CPU ID parameter during the initial configuration of Cloudera Manager.

Note: See the EMC DSSD document *DSSD Storage Appliance Installation Guide* for information about installing and configuring the DSSD D5 storage appliance and completing the above tasks.

After completing the above tasks, install Cloudera Manager. You need the following information before proceeding:

- Host names of all the hosts in your cluster.
- The DSSD D5 volume name for the DataNodes.
- If you are not using the entire capacity of the DSSD D5 for this cluster, the DSSD Amount of Usable Capacity as assigned in the DSSD D5. For most deployments, the default value (100 TB) is correct. See the *DSSD Hadoop Plugin Installation Guide* for more information on setting this property.
- The value for the Libflood CPU ID. See “Identify CPUs and NUMA Nodes” in the *DSSD Hadoop Plugin Installation Guide* for more information.

Upgrading a DSSD D5 Deployment to a New Version of CDH and Cloudera Manager

To upgrade an existing version of CDH and Cloudera Manager in a cluster that uses the DSSD D5, follow the same steps as for deployments that do not use the DSSD D5. However, before upgrading CDH, you must enter a new remote repository URL and download and distribute the DSSD parcels. See [Upgrading with the EMC DSSD D5 \(cm ig upgrade dssd.html#concept_ars_kj3_kv\)](#).

Deployment Options

A Cloudera Manager deployment consists of the following software components:

- Oracle JDK
- Cloudera Manager Server and Agent packages
- Supporting database software
- CDH and managed service software

This section describes the installation paths for creating a new Cloudera Manager deployment and the criteria for choosing an installation path. The Cloudera Manager installation paths share some common phases, but each path supports different user and cluster host requirements.

Note: If you are deploying Cloudera Manager in a highly available configuration, see [Configuring Cloudera Manager for High Availability With a Load Balancer \(admin_cm_ha_overview.html#concept_bhl_cvc_pr\)](#) before starting your installation.

:

- **Demonstration and proof of concept deployments** have two installation options:
 - [DSSD D5 Installation Path A - Automated Installation by Cloudera Manager Installer \(cm ig install_path_a_dssd.html\)](#) - Cloudera Manager automates the installation of the Oracle JDK, Cloudera Manager Server, embedded PostgreSQL database, Cloudera Manager Agent, CDH, and managed service software on cluster hosts. Cloudera Manager also configures databases for the Cloudera Manager Server and Hive Metastore and optionally for Cloudera Management Service roles. This path is recommended for demonstration and proof-of-concept deployments, but is *not recommended* for production deployments because its not intended to scale and may require database migration as

your cluster grows. To use this method, server and cluster hosts must satisfy the following requirements:

- Provide login to the Cloudera Manager Server host using a root account or an account that has passwordless sudo permission.
 - Allow the Cloudera Manager Server host to have uniform SSH access on the same port to all hosts. See [Networking and Security Requirements \(cm ig cm requirements.html#cmig topic 4 3 3\)](#) for more information.
 - All hosts must have access to standard package repositories and either `archive.cloudera.com` or a local repository with the required installation files.
- [DSSD D5 Installation Path B - Installation Using Cloudera Manager Parcels \(cm ig install path b dssd.html\)](#) - You install the Oracle JDK, Cloudera Manager Server, and embedded PostgreSQL database packages on the Cloudera Manager Server host. Cloudera Manager automates the installation of the Oracle JDK, Cloudera Manager Agents, CDH, and managed service software on cluster hosts. For Cloudera Manager to automate installation of Cloudera Manager Agent packages or CDH and managed service software, cluster hosts must satisfy the following requirements:
 - Allow the Cloudera Manager Server host to have uniform SSH access on the same port to all hosts. See [Networking and Security Requirements \(cm ig cm requirements.html#cmig topic 4 3 3\)](#) for more information.
 - All hosts must have access to standard package repositories and either `archive.cloudera.com` or a local repository with the required installation files.
- **Production deployments** require you to first manually install and configure a production [database \(cm ig installing configuring dbs.html#cmig topic 5\)](#) for the Cloudera Manager Server and Hive Metastore. There are two installation options:
 - [DSSD D5 Installation Path B - Installation Using Cloudera Manager Parcels \(cm ig install path b dssd.html\)](#) - You install the Oracle JDK and Cloudera Manager Server packages on the Cloudera Manager Server

host. Cloudera Manager automates the installation of the Oracle JDK, Cloudera Manager Agents, CDH, and managed service software on cluster hosts.

For Cloudera Manager to automate installation of Cloudera Manager Agent packages or CDH and managed service software, cluster hosts must satisfy the following requirements:

- Allow the Cloudera Manager Server host to have uniform SSH access on the same port to all hosts. See [Networking and Security Requirements \(cm ig cm requirements.html#cmig topic 4 3 3\)](#) for more information.
- All hosts must have access to standard package repositories and either `archive.cloudera.com` or a local repository with the required installation files.
- [Installation Path C - Manual Installation Using Cloudera Manager Tarballs \(cm ig install path c.html#cmig topic 6 7\)](#) - You install the Oracle JDK, Cloudera Manager Server, and Cloudera Manager Agent software using tarballs and use Cloudera Manager to automate installation of CDH and managed service software using parcels.

Cloudera Manager Installation Phases

The following table describes the phases of installing Cloudera Manager and a Cloudera Manager deployment of CDH and managed services. Every phase is required, but you can accomplish each phase in multiple ways, depending on your organization's policies and requirements. The six phases are grouped into three installation paths based on how the Cloudera Manager Server and database software are installed on the Cloudera Manager Server and cluster hosts. The criteria for choosing an installation path are discussed in [Cloudera Manager Deployment \(installation installation.html#concept qpf 2d2 2p\)](#).

Cloudera Installation Phases

Phase			
Phase 1: Install JDK Install the JDK required by	There are two options: <ul style="list-style-type: none"> • Use the Cloudera Manager Installer to install a supported version of the Oracle JDK in <code>/usr/java</code> and on all hosts in the cluster. 		

Phase			
Cloudera Manager Server, Management Service, and CDH.	<ul style="list-style-type: none"> • Use the command line to manually install supported versions of the Oracle JDK and set the <code>JAVA_HOME</code> environment variable to the install directory on all hosts. 		
Phase 2: Set up Databases Install, configure, and start the databases that are required by the Cloudera Manager Server, Cloudera Management Service, and that are optional for some CDH services.	There are two options: <ul style="list-style-type: none"> • Use the Cloudera Manager Installer to install, configure, and start an embedded PostgreSQL database. • Use command-line package installation tools like <code>yum</code> to install, configure, and install the database 		
	Path A	Path B	Path C
Phase 3: Install Cloudera Manager Server Install and start Cloudera Manager Server on one host.	Use the Cloudera Manager Installer to install its packages and the server. Requires Internet access and sudo privileges on the host.	Use Linux package install commands (like <code>yum</code>) to install Cloudera Manager Server. Update database properties. Use service commands to start Cloudera Manager Server.	Use Linux commands to unpack tarballs and service commands to start the server.
Phase 4: Install Cloudera Manager Agents	Use the Cloudera Manager Installation wizard	There are two options:	Use Linux commands to unpack tarballs

Phase			
<p>Install and start the Cloudera Manager Agent on all hosts.</p>	<p>to install the Agents on all hosts.</p>	<ul style="list-style-type: none"> • Use Linux package install commands (like yum) to install Cloudera Manager Agents on all hosts. • Use the Cloudera Manager Installation wizard to install the Agents on all hosts. 	<p>and service commands to start the agents on all hosts.</p>
<p>Phase 5: Install CDH and Managed Service software</p> <p>Install, configure, and start CDH and managed services on all hosts.</p>	<p>Use the Cloudera Manager Installation wizard to install CDH and other managed services.</p>	<p>There are two options:</p> <ul style="list-style-type: none"> • Use the Cloudera Manager Installation wizard to install CDH and other managed services. • Use Linux package install commands (like yum) to install 	<p>Use Linux commands to unpack tarballs and service commands to start CDH and managed services on all hosts.</p>

Phase			
		CDH and other managed services on all hosts.	
Phase 6: Create, Configure and Start CDH and Managed Services Configure and start CDH and managed services.	Use the Cloudera Manager Installation wizard to install CDH and other managed services, assign roles to hosts, and configure the cluster. Many configurations are automated.	Use the Cloudera Manager Installation wizard to install CDH and other managed services, assign roles to hosts, and configure the cluster. Many configurations are automated.	Use the Cloudera Manager Installation wizard to install CDH and other managed services, assign roles to hosts, and configure the cluster. Many configurations are automated. You can also use the Cloudera Manager API to manage a cluster, which can be useful for scripting preconfigured deployments.

Cloudera Manager Installation Software

Cloudera Manager provides the following software for the supported installation paths:

- **Installation path A** - A small self-executing Cloudera Manager installation program to install the Cloudera Manager Server and other packages. The Cloudera Manager installer, which you install on the host where you want the Cloudera Manager Server to run, performs the following:

1. Installs the package repositories for Cloudera Manager and the Oracle Java Development Kit (JDK).
 2. Installs the Cloudera Manager packages.
 3. Installs and configures an embedded PostgreSQL database for use by the Cloudera Manager Server, some Cloudera Management Service roles, some managed services, and Cloudera Navigator roles.
- **Installation paths B and C** - Cloudera Manager repositories for manually installing the Cloudera Manager Server, Agent, JDK, and embedded database packages.
 - **Installation path B** - The Cloudera Manager Installation wizard for automating installation of Cloudera Manager Agent package.
 - **All installation paths** - The Cloudera Manager Installation wizard for automating CDH and managed service installation and configuration on the cluster hosts using parcels. Parcels simplify the installation process and allow you to download, distribute, and activate new versions of CDH and managed services from within Cloudera Manager. After you install Cloudera Manager and you connect to the Cloudera Manager Admin Console for the first time, use the Cloudera Manager Installation wizard to:
 1. Discover cluster hosts
 2. Optionally install the Oracle JDK
 3. Optionally install CDH, managed service, and Cloudera Manager Agent software on cluster hosts
 4. Select services
 5. Map service roles to hosts
 6. Edit service configurations
 7. Start services

If you abort the software installation process, the Installation wizard automatically reverts and rolls back the installation process for any uninstalled components. (Installation that has completed successfully on a host is not rolled back on that host.)

Installation Paths:

- [DSSD D5 Installation Path A - Automated Installation by Cloudera Manager Installer \(cm ig install path a dssd.html\)](#)
- [DSSD D5 Installation Path B - Installation Using Cloudera Manager Parcels \(cm ig install path b dssd.html\)](#)

- [DSSD D5 Installation Path C - Manual Installation Using Cloudera Manager Tarballs \(cm ig install path c dssd.html\)](#)

Categories: [CDH \(../categories/hub_cdh.html\)](#) | [Cloudera Manager \(../categories/hub_cloudera_manager.html\)](#) | [Concepts \(../categories/hub_concepts.html\)](#) | [DSSD \(../categories/hub_dssd.html\)](#) | [DataNode \(../categories/hub_datanode.html\)](#) | [Deploying \(../categories/hub_deploying.html\)](#) | [Installing \(../categories/hub_installing.html\)](#) | [Requirements \(../categories/hub_requirements.html\)](#) | [All Categories \(../categories/hub.html\)](#)