Cloudera Data Hub

Cloudera Data Hub Release Notes

Date published: 2019-12-17
Date modified: 2025-08-18

https://docs.cloudera.com/

Legal Notice

Contents

What's new

This section lists major features and updates for the Cloudera Data Hub service.

November 5, 2025

This release of the Cloudera Data Hub service introduces the following changes:

Cluster template support on ARM

With the Cloudera Runtime 7.3.1.500 Service Pack 1 (SP3) release, the following Cloudera Data Hub templates are supported on ARM:

- Data Discovery and exploration - Spark 3
- Data Mart
- Real-time Data Mart - Spark 3
- Streams Messaging

For more information about how to create Cloudera Data Hub clusters on ARM processors, see the Creating a cluster from a definition on AWS documentation.

The following limitations are applied when using the ARM-based architecture:

- Ensure that the EC2 instances with ARM processors are supported in your region. For more information, see the AWS Graviton Processors and Amazon EC2 instance types by Region documentation.
- From the available Cloudera Data Hub templates, only the Data Engineering templates are supported on the ARM-based architecture.
- Clusters with mixed architectures are not supported. All cluster nodes should have the same architecture, either X86_64 or ARM64.

October 31, 2025

This release of the Cloudera Data Hub service introduces the following changes:

OS upgrades as part of Cloudera Data Hub upgrades

Previously, the Cloudera Data Hub upgrade process consisted of upgrading Cloudera Runtime, which was followed by an Operating System (OS) upgrade.

From October 31, 2025, this default behavior has changed and the OS upgrade is automatically triggered after the Cloudera Runtime upgrade as part of major/minor version and service pack upgrades.

For more information, see the Performing a Upgrading Cloudera Data Hub clusters documentation.

Note: In the following cases, the OS upgrade is only performed automatically for the Gateway nodes:

- When upgrading a Cloudera Operational Database cluster
- When rolling upgrade is requested for large clusters (more than 20 nodes)

August 18, 2025

This release of the Cloudera Data Hub service introduces the following changes:

Selecting preferred zones for multi-AZ Cloudera Data Hub clusters on Azure

When scaling a multi-AZ Cloudera Data Hub cluster, the new nodes are automatically distributed across all available availability zones. However, if you prefer to manually control the distribution of nodes during the Cloudera Data Hub scaling, you can use the --preferred-zones field.

For more information, see the Creating a multi-AZ Cloudera Data Hub cluster on Azure documentation.

June 25, 2025

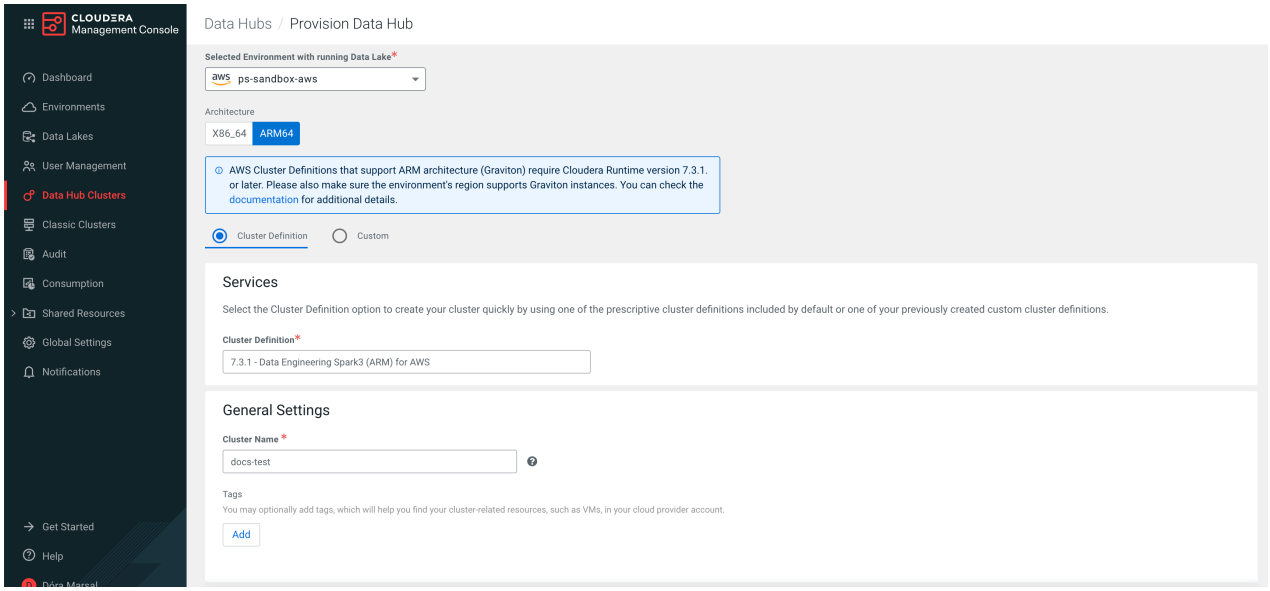
This release of the Cloudera Data Hub service introduces the following changes:

Selecting Architecture for Cloudera Data Hub on AWS

When creating a Cloudera Data Hub cluster on AWS, you can select from the following architecture types:

- X86_64
- ARM64

When you choose the architecture type, the compatible images and cluster templates will be available for the Cloudera Data Hub cluster.



Cloudera Data Engineering support on ARM - GA

With the Cloudera Runtime 7.3.1.300 Service Pack 1 (SP1) Cumulative Hotfix 1 (CHF1) release, deploying Cloudera Data Engineering Data Hub clusters using ARM processors are now generally available.

For more information about how to create Cloudera Data Engineering Data Hub clusters on ARM processors, see the Creating a cluster from a definition on AWS documentation.

The following limitations are applied when using the ARM-based architecture:

- You can create ARM-based Cloudera Data Engineering Data Hub clusters only in Cloudera environments on AWS.
- Ensure that the EC2 instances with ARM processors are supported in your region. For more information, see the AWS Graviton Processors and Amazon EC2 instance types by Region documentation.
- From the available Cloudera Data Hub templates, only the Data Engineering templates are supported on the ARM-based architecture.
- Clusters with mixed architectures are not supported. All cluster nodes should have the same architecture, either X86_64 or ARM64.

May 29, 2025

This release of the Cloudera Data Hub service introduces the following changes:

Secret rotation - GA

To add extra measures of security, you can rotate secrets, like database passwords or FreeIPA admin password using CLI commands. These secrets are managed and created by the Cloudera Control Plane. By using the following commands the secrets can be rotated to achieve more secure deployments. For more information, see Rotating Cloudera Data Hub secrets.

Restarting instances from Cloudera Management Console

You can now restart Cloudera Data Hub instances using the Cloudera Management Console. For more information, see the Restarting Cloudera Data Hub instances documentation.

April 25, 2025

This release of the Cloudera Data Hub service introduces the following changes:

Cloudera Data Engineering support on ARM - Technical Preview

With the Cloudera Runtime 7.3.1.200 Service Pack 1 (SP1) release, Cloudera Data Engineering Data Hub is optimized to run ARM-based solutions enabling users to run large-scale data processing and analytics workloads. From the AWS Graviton family, Cloudera Data Engineering Data Hub is supported on AWS Graviton2, Graviton3 and Graviton4.

AWS Graviton is a general purpose, ARM-based processor family. AWS Graviton offers enhanced performance and cost-efficiency compared to traditional Intel x86 processors. With AWS Graviton, you can optimize costs and achieve better performance for cloud workloads running in AWS Elastic Compute Cloud (EC2). For more information, see the AWS Instance Type page.

For more information about how to create Cloudera Data Engineering Data Hub clusters on ARM processors, see the Creating a cluster from a definition on AWS documentation.

The following limitations are applied when using Cloudera Data Engineering Data Hub on ARM-based architecture:

- You need to have the following entitlement to be able to create Cloudera Data Engineering Data Hub clusters on ARM:
  - CDP_AWS_ARM_DATAHUB

  For more information about how to obtain the entitlement, contact Cloudera Customer Support.
- You can create ARM-based Cloudera Data Engineering Data Hub clusters only in Cloudera environments on AWS.
- Ensure that the EC2 instances with ARM processors are supported in your region. For more information, see the AWS Graviton Processors and Amazon EC2 instance types by Region documentation.
- From the available Cloudera Data Hub templates, only the Data Engineering templates are supported on the ARM-based architecture.
- To create Cloudera Data Engineering Data Hub clusters that use ARM processors, you need to install beta CDP CLI. For more information, see the Installing beta CDP CLI documentation.

April 2, 2025

This release of the Cloudera Data Hub service introduces the following changes:

Secret rotation (Preview)

To add extra measures of security, you can rotate secrets, like database passwords or FreeIPA admin password using CLI commands. These secrets are managed and created by the Cloudera Control Plane. By using the following commands the secrets can be rotated to achieve more secure deployments.

Note: The documentation of Secret rotation has been moved to the following locations as the feature is Generally Available (GA):

- Rotating FreeIPA secrets
- Rotating Data Lake secrets
- Rotating Cloudera Data Hub secrets

March 3, 2025

This release of the Cloudera Data Hub service introduces the following changes:

Added Stale cluster status

Stale cluster status has been added to the list of cluster statuses to express a status where we have no information from the cluster and it has been unreachable for more than 30 days and that the Cloudera Data Hub cluster's status is outdated or no longer reflects its current state. For more information, see Cloudera Data Hub status options.

Added get-operation command to obtain operation status

A new command, get-operation is now available to get the status of the latest operation or a specified earlier operation performed on a Cloudera Data Hub cluster. For more information, see Monitoring clusters.

Support for Italy North, New Zealand North, and Poland Central Azure regions

The Italy North, New Zealand North, and Poland Central Azure regions are now supported. You can register Azure environments and provision Cloudera Data Hub clusters in these regions. See updated Supported Azure regions.

January 30, 2025

This release of the Cloudera Data Hub service introduces the following changes:

Added YARN recommendations to Cloudera Data Hub scaling activities CLI output

The list of Cloudera Data Hub scaling activities in the output of the cdp datahub list-scaling-activities command have been extended with YARN recommendations. For more information, see Managing autoscaling.

Older releases

Overview of new features, enhancements, and changed behavior introduced in earlier releases of the Cloudera Data Hub service.

2024

December 2024

This release of the Cloudera Data Hub service introduces the following changes:

Cloudera Runtime 7.3.1

Cloudera Runtime 7.3.1 is now available and can be used for registering an environment with a 7.3.1 Data Lake and creating Cloudera Data Hub clusters. For more information about the new Cloudera Runtime version, see Cloudera Runtime. If you need to upgrade your existing Cloudera environment, your upgrade path may be complex. To determine your upgrade path, refer to Upgrading to Runtime 7.3.1 documentation.

Added pd-extreme and pd-balanced as supported GCP disktypes

The following disk types have been added to supported GCP block storage types:

- pd-extreme
- pd-balanced

You can review all supported disk types for GCP at Supported GCP block storage.

Added EnvironmentPrivilegedUser to environment resource roles

The new EnvironmentPrivilegedUser resource role grants permission to execute privileged operating system actions on Data Lake, FreeIPA, and Cloudera Data Hub virtual machines.

For more information, see Resource roles.

October 2024

This release of the Cloudera Data Hub service introduces the following changes:

Azure Single Server to Azure Flexible Server upgrade

Single Server on Microsoft Azure databases used by Data Lakes and Cloudera Data Hub clusters can now be upgraded to Azure Flexible Server. During the upgrade process from PosgtreSQL version 11 to PostgreSQL 14, Azure Single Server will be upgraded to Azure Flexible Server. For more information, see Upgrading Azure Single Server to Flexible Server.

July 2024

This release of the Cloudera Data Hub service introduces the following changes:

AWS RDS certificate rotation

AWS requires the rotation of the SSL/TLS certificates used for secure communication between Data Lakes and certain Cloudera Data Hub clusters and the external AWS RDS database instances that they rely on. Cloudera on cloud now provides multiple options to perform the required RDS certificate rotation.

For more information, see Rotating database certificates.

Rolling upgrade support for Streams Messaging clusters

Rolling upgrades for Streams Messaging Cloudera Data Hub clusters are now available. Rolling upgrades for Cloudera Data Hub clusters are limited to certain Data Hub Runtime versions. For more information, see Rolling upgrades.

May 2024

This release of the Cloudera Data Hub service introduces the following changes:

New default Azure VM instance types in cluster templates

In the built-in Cloudera Data Hub templates, the following instance types are being replaced:

- Standard_D5_v2 is being replaced with Standard_D16s_v3
- Standard_D8_v3 is being replaced with Standard_D8s_v3
- Standard_D16_v3 is being replaced with Standard_D16s_v3
- Standard_E16_v3 is being replaced with Standard_E16s_v3

- Standard_E8_v3 is being replaced with Standard_E8s_v3

The following table provides more detail:

Template name	Group	Previous instance type	New instance type
7.x - Data Engineering: HA: Apache Spark3, Apache Hive, Apache Oozie for Azure	manager	Standard_D16_v3	Standard_D16s_v3
7.x - Data Engineering: HA: Apache Spark3, Apache Hive, Apache Oozie for Azure	compute	Standard_D5_v2	Standard_D16s_v3
7.x - Data Engineering: HA: Apache Spark3, Apache Hive, Apache Oozie for Azure	worker	Standard_D16_v3	Standard_D16s_v3
7.x - Data Engineering: HA: Apache Spark3, Apache Hive, Apache Oozie for Azure	gateway	Standard_D16_v3	Standard_D16s_v3
7.x - Data Engineering: HA: Apache Spark3, Apache Hive, Apache Oozie for Azure	master	Standard_D16_v3	Standard_D16s_v3
7.x - Data Engineering: HA: Apache Spark3, Apache Hive, Apache Oozie for Azure	masterx	Standard_D16_v3	Standard_D16s_v3
7.x - Data Engineering Spark3 for Azure	master	Standard_D16_v3	Standard_D16s_v3
7.x - Data Engineering Spark3 for Azure	compute	Standard_D5_v2	Standard_D16s_v3
7.x - Data Engineering Spark3 for Azure	worker	Standard_D5_v2	Standard_D16s_v3
7.x - Data Engineering Spark3 for Azure	gateway	Standard_D8_v3	Standard_D16s_v3
7.x - Data Discovery and Exploration for Spark3 for Azure	master	Standard_D8_v3	Standard_D8s_v3
7.x - Data Discovery and Exploration for Spark3 for Azure	gateway	Standard_D8_v3	Standard_D8s_v3
7.x - Data Discovery and Exploration for Spark3 for Azure	leader	Standard_D8_v3	Standard_D8s_v3
7.x - Data Discovery and Exploration for Spark3 for Azure	worker	Standard_D8_v3	Standard_D8s_v3
7.x - Data Discovery and Exploration for Spark3 for Azure	yarnworker	Standard_D8_v3	Standard_D8s_v3
7.x - Data Mart: Apache Impala, Hue for Azure	master	Standard_E8_v3	Standard_E8s_v3
7.x - Edge Flow Management Light Duty for Azure	management	Standard_D8_v3	Standard_D8s_v3
7.x - COD Edge Node for for Azure	leader	Standard_D8_v3	Standard_D8s_v3
7.x - Streaming Analytics Heavy Duty for Azure	manager	Standard_D8_v3	Standard_D8s_v3
7.x - Streaming Analytics Heavy Duty for Azure	master	Standard_D8_v3	Standard_D8s_v3
7.x - Streaming Analytics Heavy Duty for Azure	worker	Standard_D8_v3	Standard_D8s_v3

7.x - Streaming Analytics Light Duty for Azure	manager	Standard_D8_v3	Standard_D8s_v3
7.x - Streaming Analytics Light Duty for Azure	master	Standard_D8_v3	Standard_D8s_v3
7.x - Streaming Analytics Light Duty for Azure	worker	Standard_D8_v3	Standard_D8s_v3
7.x - Flow Management Light Duty for Azure	management	Standard_D8_v3	Standard_D8s_v3
7.x - Flow Management Light Duty for Azure	nifi_scaling	Standard_D8_v3	Standard_D8s_v3
7.x - Flow Management Light Duty for Azure	nifi	Standard_D8_v3	Standard_D8s_v3
7.x - Flow Management Heavy Duty for Azure	management	Standard_D8_v3	Standard_D8s_v3
7.x - Real-time Data Mart - Spark3 for Azure	master1	Standard_D8_v3	Standard_D8s_v3
7.x - Real-time Data Mart - Spark3 for Azure	master2	Standard_D8_v3	Standard_D8s_v3
7.x - Real-time Data Mart - Spark3 for Azure	master3	Standard_D8_v3	Standard_D8s_v3
7.x - Streams Messaging High Availability for Azure	manager	Standard_D16_v3	Standard_D16s_v3
7.x - Streams Messaging High Availability for Azure	core_zookeeper	Standard_D8_v3	Standard_D8s_v3
7.x - Streams Messaging High Availability for Azure	core_broker	Standard_D8_v3	Standard_D8s_v3
7.x - Streams Messaging High Availability for Azure	srm	Standard_D8_v3	Standard_D8s_v3
7.x - Streams Messaging Light Duty for Azure	broker	Standard_D8_v3	Standard_D8s_v3
7.x - Streams Messaging Light Duty for Azure	kraft	Standard_D8_v3	Standard_D8s_v3
7.x - Streams Messaging Heavy Duty for Azure	registry	Standard_D8_v3	Standard_D8s_v3
7.x - Streams Messaging Heavy Duty for Azure	smm	Standard_D8_v3	Standard_D8s_v3
7.x - Streams Messaging Heavy Duty for Azure	srm	Standard_D8_v3	Standard_D8s_v3
7.x - Streams Messaging Heavy Duty for Azure	connect	Standard_D8_v3	Standard_D8s_v3
7.x - Streams Messaging Heavy Duty for Azure	kraft	Standard_D8_v3	Standard_D8s_v3

If you would like to create Cloudera Data Hub clusters with the previously used instance types, you can achieve this by configuring a custom instance type.

April 2024

This release of the Cloudera Data Hub service introduces the following changes:

Cloudera Runtime 7.2.18

Cloudera Runtime 7.2.18 is now available and can be used for registering an environment with a 7.2.18 Data Lake and creating Cloudera Data Hub clusters. For more information about the new Runtime version, see Cloudera

Runtime. If you need to upgrade your existing Cloudera environment, your upgrade path may be complex. To determine your upgrade path, refer to Upgrading to Cloudera Runtime 7.2.18 documentation.

RHEL replaces CentOS as default OS

As of June 30, 2024, CentOS reaches End of Life (EOL), and consequently, Cloudera Runtime 7.2.18 supports RHEL 8 only. New deployments of Data Lakes and Cloudera Data Hub clusters with Cloudera Runtime 7.2.18 and upgrades to 7.2.18 are only possible with RHEL 8. Data Lake and Cloudera Data Hub clusters running Cloudera Runtime 7.2.17 support both CentOS 7 and RHEL 8. Earlier Runtime versions support CentOS 7 only. Cloudera will not publish any updates or fixes for CentOS-based images after June 2024.

As part of FreeIPA, Data Lake, and Cloudera Data Hub upgrade, you have the option to upgrade the operating system (OS) on the virtual machines (VMs) from CentOS 7 to Red Hat Enterprise Linux 8 (RHEL 8). For more information, see Upgrading from CentOS to RHEL.

Rolling upgrade support for Cloudera Operational Database Data Hub clusters

With the release of Runtime 7.2.18, rolling upgrades for certain Cloudera Operational Database Data Hub clusters are now available. For more information, see Rolling upgrades.

Disk vertical scaling in Azure (Preview)

Disk vertical scaling (that is, disk type change and resizing) is now supported by Cloudera for Data Lakes and Cloudera Data Hub clusters running in Azure. Previously, only AWS support was available for this feature in Cloudera. For more information, see Disk Vertical Scaling - Disk Type Change and Resizing in AWS and Azure.

Note: You need to contact Cloudera to have this feature enabled.

March 2024

This release of the Cloudera Data Hub service introduces the following changes:

Cloudera Data Hub root volume size was increased to 200 GB

The root volume size of Cloudera Data Hub clusters was increased from 100 GB to 200 GB.

2023

October 2023

This release of the Data Hub service introduces the following changes:

Streams Messaging High Availability clusters now include highly available Knox (October 30, 2023)

The Streams Messaging High Availability cluster configuration is updated. The configuration now includes a highly available Knox service. Specifically, a new host group called manager is introduced. This host group is not scalable and includes two nodes. Both nodes host a Knox service instance. In addition, Cloudera Manager is moved from the master host group and is now deployed on one of the manager nodes. The master host group now only includes Cruise Control and Streams Messaging Manager.

The previous cluster configuration is no longer available. Streams Messaging High Availability clusters provisioned at a previous date will continue to run using the old configuration. Updating an existing cluster to use the new configuration is not possible. For more information, see Streams Messaging cluster layout.

August 2023

This release of the Data Hub service introduces the following changes:

Support for Local SSDs in GCP (August 21, 2023)

Data Hub now supports using Local SSDs as storage in GCP. During Data Hub cluster creation in CDP, you can navigate to advanced Hardware and Storage options and select to use "Local scratch disk (SSD)" with certain instance types. Prior to using Local SSDs with CDP, make sure to review Google Cloud Platform documentation related to Local SSDs and familiarize yourself with the applicable restrictions and limitations (such as 12 TB maximum capacity, limited configurations).

Note: Stopping and restarting Data Hub clusters using local SSDs is not supported; When you stop or suspend a VM, all data on the attached Local SSD is discarded.

Data Hub database upgrade and default major version change (August 30, 2023)

Newly-deployed Data Hub clusters on AWS or GCP with Cloudera Runtime 7.2.7 or above are now configured to use a PostgreSQL version 14 database by default.

Newly-deployed Data Hub clusters on Azure with Cloudera Runtime 7.2.7 or above will continue to use a PostgreSQL version 11 database by default.

The database for Data Hub clusters on AWS and GCP can now be upgraded to PostgreSQL version 14. If your AWS or GCP Data Hub cluster requires an upgrade to PostgreSQL 14, you will receive a notification in the Management Console UI.

Cloudera strongly recommends that the database upgrade to PostgreSQL 14 for AWS and GCP clusters is performed on all clusters running PostgreSQL version 11 by November 9, 2023.

A database upgrade to PostgreSQL 14 for Azure Data Hubs will be available in the future. Any Data Hub clusters on Azure that require a database upgrade will be upgraded from PostgreSQL 10 to PostgreSQL 11.

For more information, see Upgrading Data Lake/Data Hub database

Support for autoscaling Data Hub clusters on Azure (August 30, 2023)

Data Hub now supports autoscaling for clusters provisioned in Azure. For more information see Autoscaling clusters.

July 2023

This release of the Data Hub service introduces the following changes:

Java 11 support for Data Hub and Data Lake clusters (July 12, 2023)

You can launch a new Data Hub or Data Lake cluster with Java 11 as the default JDK. This capability is currently only available through the CDP CLI. See Java 11 for more information.

April 2023

This release of the Data Hub service introduces new features and behaviors.

GP3 support for attached storage disks (April 19)

Data Hubs now support GP3 (SSD) volume types for attached storage. GP3 volumes allow you to increase performance (independently provisioning IOPS and throughput) without increasing storage size. GP3 volumes will deliver similar performance as GP2 volumes at a lower cost. GP3 is now the default attached storage type for Data Hub instances that previously used GP2 storage. You can configure storage options when you provision a Data Hub cluster under the Hardware and Storage section of the Advanced Options.

February 2023

This release of the Data Hub service introduces new features and behaviors.

Data Hub Autoscaling (February 27, 2023)

Autoscaling is a feature that adjusts the capacity of cluster nodes running YARN by automatically increasing or decreasing, or suspending and resuming, the nodes in a host group. You can enable autoscaling based either on a schedule that you define, or the real-time demands of your workloads. For more information see Autoscaling clusters.

January 2023

This release of the Data Hub service introduces new features and behaviors.

Vertical scaling of Data Hub instances on AWS

If necessary, you can now select a larger or smaller instance type for a Data Hub cluster after it has been deployed. This is only supported on AWS. For more information see Vertically scaling instances types.

2022

December 2022

This release of the Data Hub service introduces new features and behaviors.

Instance support (December 19, 2022)

The following new instance types are supported in Data Hub:

AWS:

- m6i.xlarge
- m6i.2xlarge
- m6i.4xlarge
- m6i.8xlarge
- m6i.12xlarge
- m6i.16xlarge
- m6i.24xlarge
- m6i.32xlarge
- m6id.xlarge
- m6id.2xlarge
- m6id.4xlarge
- m6id.8xlarge
- m6id.12xlarge
- m6id.16xlarge
- m6id.24xlarge
- m6id.32xlarge
- m6a.xlarge
- m6a.2xlarge
- m6a.4xlarge
- m6a.8xlarge
- m6a.12xlarge
- m6a.16xlarge
- m6a.24xlarge
- m6a.32xlarge
- m6a.48xlarge

Azure:

- Standard_D8_v5
- Standard_D16_v5

- Standard_D32_v5
- Standard_D48_v5
- Standard_D64_v5
- Standard_D8s_v5
- Standard_D16s_v5
- Standard_D32s_v5
- Standard_D48s_v5
- Standard_D64s_v5
- Standard_D8as_v5
- Standard_D16as_v5
- Standard_D32as_v5
- Standard_D48as_v5
- Standard_D64as_v5

October 2022

This release of the Data Hub service introduces new features and behaviors.

Changes to Data Hub cluster deployments on AWS (October 28, 2022)

The CloudFormation template previously in use for Data Hub deployments on AWS is no longer required for new Data Hub deployments. Clusters provisioned without the CloudFormation template will have the following changes:

- Hostnames in the AWS Management Console appear differently between native and CloudFormation clusters. For example:
  - dev-pdp-de-fix-01-master0 is an example of a hostname in a cluster deployed with the CloudFormation template
  - devpdpdefix01-164597-master-0 is an example of a hostname in a cluster deployed without the CloudFormation template
- Some of the default tags were removed during the implementation of this change from CloudFormation to native deployed clusters:
  - aws:autoscaling:groupName
  - aws:cloudformation:stack-id
  - aws:ec2launchtemplate:id
  - aws:ec2launchtemplate:version
  - aws:cloudformation:logical-id
  - aws:ec2:fleet-id
  - aws:cloudformation:stack-name

September 2022

This release of the Data Hub service introduces new features.

Database Upgrade and default major version change (September 27, 2022)

Newly deployed Data Lake and Data Hub clusters with Cloudera Runtime 7.2.7 or above are now configured to use a PostgreSQL version 11 database by default.

A new Database Upgrade capability is now available for existing Data Lake and Data Hub clusters. If you are running clusters on Cloudera Runtime version 7.2.6 or below, please upgrade to a more recent version before performing the database upgrade.

The major version of the database used by Data Lake or Data Hub clusters is now also displayed on the Database page of the respective service.

Cloudera strongly recommends that the Database Upgrade is performed on all clusters running PostgreSQL version 10 before November 10, 2022.

For more information, see Upgrading database.

Recipe type changes

The following recipe types have been renamed for Data Hub, Data Lake, and FreeIPA recipes:

- pre-service-deployment (formerly pre-cluster-manager-start)
- post-service-deployment (formerly post-cluster-install)

These changes will not affect existing recipe automation. For more information see Recipes.

Instance support (September 1, 2022)

The following new AWS instance types are supported in Data Hub:

- m5dn.xlarge
- m5dn.2xlarge
- m5dn.4xlarge
- m5dn.8xlarge
- m5dn.12xlarge
- m5dn.16xlarge
- m5dn.24xlarge

Validate and prepare for upgrade (September 1, 2022)

Before you perform a Data Hub upgrade, you can run the new Validate and Prepare option to check for any configuration issues and begin the Cloudera Runtime parcel download and distribution. Using the validate and prepare option does not require downtime and makes the maintenance window for an upgrade shorter. For more information see Preparing for an upgrade.

July 2022

This release of the Data Hub service introduces new features.

The default authentication method for Kafka brokers is changed to LDAP (June 28, 2022)

To improve performance, the default SASL_SSL/PLAIN authentication method for Kafka brokers is changed to LDAP. Previously, brokers were configured to use PAM. This change is Runtime version independent and applies to all Data Hub clusters provisioned after July 28, 2022. For existing clusters, Cloudera recommends that you transition from PAM to LDAP. For more information on how to enable and configure LDAP on existing clusters, see Kafka LDAP Authentication.

June 2022

This release of the Data Hub service introduces new features.

Customer managed encryption keys on AWS (June 7, 2022)

Previously, Data Hub EBS encryption could be configured only as part of Data Hub creation. Now you can also configure encryption during environment registration and all EBS volumes and RDS database and used by the Data Hubs running in that environment will be encrypted. It is still possible to add a CMK during Data Hub creation, overwriting the encryption key provided during environment registration for that Data Hub. For more information, refer to Adding a customer managed encryption key to a CDP environment running on AWS and Configuring encryption for Data Hub's EBS volumes on AWS.

Cluster template overrides (June 7, 2022)

As an alternative to creating a custom template, you can specify custom configurations that override or append the properties (service and role configs) in a built-in or custom Data Hub template. These configurations are saved as a shared resource called “cluster template overrides,” and can be used and re-used across Data Hub clusters in different environments.

For more information, see Cluster template overrides.

Creating multi-AZ Data Hubs on AWS (June 9, 2022)

By default, CDP provisions Data Hubs in a single AWS availability zone (AZ), but you can optionally choose to deploy them across multiple availability zones (multi-AZ). For more information, see Creating a multi-AZ Data Hub on AWS.

Multi-AZ deployments do not use a CloudFormation template for VM creation. Neither autoscaling groups or launch templates are created. The cluster resources are managed individually using AWS native components (for example, EC2 instances).

Customer managed encryption keys on GCP (June 16, 2022)

By default, a Google-managed encryption key is used to encrypt disks and Cloud SQL instances used by Data Hub clusters, but you can optionally configure CDP to use a customer-managed encryption key (CMEK) instead. This can only be configured during CDP environment registration using CDP CLI. For more information, refer to Adding a customer managed encryption key for GCP.

Data Engineering HA template for Azure (June 22, 2022)

The Data Engineering HA template for Azure is now GA. For more information see Data Engineering clusters (Azure).

New option to delete attached volumes during Data Hub repair (June 29, 2022)

When you initiate a repair from the Data Hub Hardware tab, you have the option to delete any volumes attached to the instance. For more information see Performing manual repair.

May 2022

This release of the Data Hub service introduces new features.

Resilient scaling improvements (May 10, 2022)

- It is possible to upscale a Data Hub cluster even if there are unhealthy nodes in the stack.
- An upscale operation will finish successfully even if one of the new nodes fails to come up. Nodes that fail during an upscale will be marked as "zombie" nodes.

For more information, see Resize a cluster.

Extended upgrade version support for RAZ-enabled environments (May 10, 2022)

Data Hub major/minor version upgrades for RAZ-enabled clusters are now available for Runtime versions 7.2.10-7.2.12 to 7.2.14+.

Instance support (May 25, 2022)

The following new AWS instance types are supported in Data Hub:

- m5ad.xlarge
- m5ad.2xlarge
- m5ad.4xlarge
- m5ad.8xlarge
- m5ad.12xlarge
- m5ad.16xlarge
- m5ad.24xlarge
- m5d.xlarge
- m5d.2xlarge

- m5d.4xlarge
- m5d.8xlarge
- m5d.12xlarge
- m5d.16xlarge
- m5d.24xlarge
- r5ad.12xlarge
- r5ad.24xlarge
- r5d.24xlarge
- r5d.12xlarge

April 2022
This release of the Data Hub service introduces new features.

Data Hub OS upgrades

OS upgrades of Cloudera Runtime are generally available. See Performing a Data Hub OS upgrade.

March 2022
This release of the Data Hub service introduces new features.

Instance support (March 4, 2022)

The following new GCP instances are supported in Data Hub:

- n2d-highmem-32
- n2d-highmem-64

Public certificate auto-renewal (March 7, 2022)

Most public (Let's Encrypt-issued) certificates for Data Lake and Data Hub clusters will now auto-renew without intervention from a user. For more information, refer to Managing Certificates.

Data Hub upgrades

Major/minor version upgrades of Cloudera Runtime and Cloudera Manager are generally available for most Data Hub templates. Data Hub maintenance upgrades for RAZ-enabled Data Hubs are generally available for versions 7.2.7+. For more information see Upgrading Data Hubs.

February 2022
This release of the Data Hub service introduces new features.

Customer managed encryption keys for Azure disks and databases

By default, local Data Hub disks attached to Azure VMs and the PostgreSQL server instance used by the Data Lake and Data Hubs are encrypted with server-side encryption (SSE) using Platform Managed Keys (PMK), but during environment registration you can optionally configure SSE with Customer Managed Keys (CMK). For more information, refer to Adding a customer managed encryption key for Azure. For more information, refer to Adding a customer managed encryption key for Azure.

Instance support

The following new AWS and Azure GPU-based instances are supported in Data Hub:

AWS:

- c5.12xlarge
- c5a.12xlarge

Azure:

- NC6sv3
- NC24sv3

January 2022
This release of the Data Hub service introduces new features.

Updating recipes post-deployment

You can attach or detach recipes to/from existing Data Hub clusters. Using this capability, you can update a recipe by removing it from the cluster, replacing the old recipe with a modified recipe of the same type, and attaching the new modified recipe to the cluster. For more information refer to Update a recipe.

Custom images and catalogs

If necessary, you can use a custom Runtime image for compliance or security reasons. You can then use the CDP CLI to register a custom image catalog and set the custom image within the custom image catalog. For more information refer to Custom images and catalogs.

2021

October 2021
This release of the Data Hub service introduces new features.

Encrypting EBS volumes with a CMK generated from AWS CloudHSM (October 4)

New documentation with steps for encrypting Data Hub's EBS volumes with a CMK generated from AWS CloudHSM is available. See Encrypting EBS volumes with a CMK generated from AWS CloudHSM.

Data Hub maintenance upgrades (October 26)

You can perform a maintenance (or "hotfix") upgrade on certain Data Hub clusters. For more information see Upgrading Data Hubs.

August 2021
This release of the Data Hub service introduces new features.

Data Engineering HA template

The Oozie service is now highly available in the Data Engineering HA templates for AWS and Azure.

May 31, 2021
This release of the Data Hub service introduces new features.

Data Engineering HA template on AWS

The Data Engineering HA template is now GA. It now supports fault-tolerant access to all cluster web UIs and end-to-end reliability for more workload types.

March 30, 2021
This release of the Data Hub service introduces new features.

Google Cloud Platform support for Data Hub

CDP now supports the creation of Data Hub clusters on the GCP platform. For more information see Create a Data Hub cluster on Google Cloud.

Show CLI Command

You can obtain the CDP CLI command for creating a Data Hub cluster from the details of an existing Data Hub or from the create Data Hub wizard. See Create a Data Hub cluster on AWS from an existing cluster and Create a Data Hub cluster on Azure from an existing cluster.

Default Root Volume Size

The default root volume size has been increased from 30-50 GB (depending on cloud provider) to 100 GB for all cloud providers.

Data Engineering and Data Engineering HA Templates

- Instance types in the Data Engineering and Data Engineering HA templates have been updated to 64 GB memory machines from 32 GB memory machines:
  - The master instance in the DE template has changed from m5.2xlarge to m5.4xlarge for AWS and from Standard_D8_v3 to Standard_D16_v3 for Azure.
  - All instances in the DE HA template have changed from m5.2xlarge to m5.4xlarge for AWS and from Standard_D8_v3 to Standard_D16_v3 for Azure
- From Cloudera Runtime 7.2.8 onwards, the DE HA template includes a new load-balancer added for the Knox instances for UI access. The external endpoint has changed from "-manager0" to "-gateway" because of the presence of the load-balancer.

February 28, 2021

This release of the Data Hub service introduces new features.

Support for Cloud Hardware Redundancy

Added support for AWS placement groups guarantees physical hardware redundancy for Data Hub clusters. See Placement group support for more details.

2020

September 30, 2020

This release of the Data Hub service introduces new features.

Data Discovery and Exploration Template (Tech Preview) now includes a Spark connector

- Provides support for indexing of RDDs and for Spark Streaming.
- Facilitates speedier data ingestion pipelines to Solr, while allowing all the power and flexibility of Spark to transform and process data on its way to be served.

The “7.2.0 - Operational Database” Data Hub template includes HUE for HBase DML/DDL

This update greatly simplifies the process of data exploration and query.

July 31, 2020

This release of the Data Hub service introduces new features.

System auditing for Data Hub events (as well as Data Lake and user management) is now available via CLI. For more information see the *Auditing* section of the Management Console documentation.

June 30, 2020

This release of the Data Hub service introduces new features.

YARN Queue Manager in Data Engineering Cluster Templates

Data Engineering cluster administrators can now leverage the visual YARN Queue Manager via Cloudera Manager to isolate and prioritize workloads within a cluster. This simplifies the administration and management of workloads.

Data Discovery and Exploration (Technical Preview)

A new Data Discovery and Exploration cluster definition is available in Data Hub. The cluster definition deploys HDFS, Hue, Solr, Spark, Yarn, and ZooKeeper services. The cluster definition is available for AWS.

May 1, 2020

This release of the Data Hub service introduces the following new features:

Flow Management for Data Hub Cluster Definitions

New Flow Management cluster definitions are available in Data Hub. The Flow Management cluster definitions (heavy duty and light duty) include the Nifi, Nifi Registry, and Zookeeper services. The cluster definitions are available for both AWS and Azure.

Data Engineering HA Cluster Definition (Technical Preview)

A new Data Engineering HA cluster definition is available in Data Hub for technical preview. The cluster definition deploys YARN, HDFS, and Hive in a highly available fashion.

External Databases for Data Hub Clusters

Through the CDP CLI, you can now configure an external, durable database for Data Hub clusters. An external database can keep a persistent state for services like Hue, DAS, and Zeppelin.

March 2, 2020

This release of the Data Hub service introduces the following new features:

Data Hub Clusters on Microsoft Azure

You can create a Data Hub cluster on Azure. You have two basic options when creating a cluster: (1) create a cluster from one of the default or custom cluster definitions, or (2) create a custom cluster.

Cloudera Runtime 7.1.0

By default, Data Hub creates clusters running Cloudera Runtime 7.1.0 and Cloudera Manager 7.1.0. Previously, 7.0.2.2 was used.

2019

December 17, 2019

This release of the Data Hub service introduces the following new features:

Cloudera Runtime 7.0.2.2

By default, Data Hub creates clusters running Cloudera Runtime 7.0.2.2 and Cloudera Manager 7.0.2.2. Previously, 7.0.2.0 was used. Note that there is no separate 7.0.2.2 documentation available, but Cloudera Runtime 7.0.2 and Cloudera Manager 7.0.2 documentation has been updated to reflect the 7.0.2.2 changes;

New Streams Messaging cluster templates and definitions

The Streams Messaging Heavy Duty and Streams Messaging Light Duty cluster templates and definitions are now available in Data Hub with advanced messaging and real-time processing on streaming data using Apache Kafka,

centralized schema management using Schema Registry, and management and monitoring capabilities powered by Streams Messaging Manager.

Note:

The Streams Messaging cluster templates and definitions are technical preview: They are not suitable for production.

November 14, 2019

This release of the Data Hub service introduces the following new features:

Cloudera Runtime 7.0.2

By default, Data Hub creates clusters running Cloudera Runtime 7.0.2 and Cloudera Manager 7.0.2. Previously, 7.0.1 was used.

New Real-Time Data Mart cluster definition

The Real-Time Data Mart cluster definition is now available in Data Hub with Apache Kudu, Apache Impala, and Apache Spark for real-time data management and operational analytics use cases. This cluster definition is suitable for real-time analytics on fast changing data, and can be leveraged together with the Data Warehouse service for a holistic analytics experience.

Enhanced SQL access for Data Engineering clusters

Data Hub clusters built with Data Engineering cluster definition now offer an improved SQL experience for Hive SQL users with smart query authorizing, build-in performance analysis and recommendations tools available in Data Analytics Studio (DAS). This capability provides an enhanced self-service experience for data analysts and data engineers. Note that DAS web UI can be accessed by using workload user name and FreeIPA password.

Operational Database cluster enhancements

The Operational Database cluster definition enables you to build real-time, high-concurrency, mission-critical business applications in the cloud environment. Key enhancements in this release include real-time monitoring, out-of-the-box resilience, and replication capabilities (Tech Preview) for Disaster Recovery (DR) and for enabling seamless migration of operational applications from on-premises to cloud.

September 23, 2019

This release of the Data Hub service introduces the following new features:

Support for Cloudera Runtime 7.0.1

By default, Data Hub creates clusters running Cloudera Runtime 7.0.1. Previously, 7.0.0 was used.

August 22, 2019

This is the first release of the Data Hub service. For an overview of Data Hub functionality, refer to:

Related Information

Cloudera Data Hub Overview

Log4j vulnerabilities

Cloudera has released hotfixes for Cloudera Runtime versions 7.2.7 and newer on on cloud that address Log4j2 vulnerabilities.

The following vulnerabilities have been addressed for Cloudera Runtime versions 7.2.7 through 7.2.12:

CVE-2021-44228

CVE-2021-45046

LOGBACK-1591

CVE-2021-45105

CVE-2021-44832

You should upgrade your Cloudera services running Cloudera Runtime versions 7.2.7+ so that they include the latest hotfixes. You can update your existing Data Lake and Cloudera Data Hub clusters by doing a maintenance upgrade. You should first upgrade the Data Lake and then upgrade all the Cloudera Data Hub clusters that are using the Data Lake. Refer to Data Lake upgrade and Cloudera Data Hub upgrade documentation. Data Lake and Cloudera Data Hub maintenance upgrade is supported only in technical preview for maintenance upgrades from Cloudera Runtime versions 7.2.7 and higher.

If you are running a version of Cloudera Runtime prior to 7.2.7, contact Cloudera Support for details on how to upgrade Cloudera Runtime.

For more information about these hotfixes, refer to the Cloudera Runtime Release Notes for the version of Runtime that you use.

Known issues

This section lists known issues that you might run into while using the Cloudera Data Hub service.

CB-11925: Knox Load Balancer API requests initiated from Knox Gateway hosts can fail with Connection timeout error

Problem: When logged into a Data Lake or Cloudera Data Hub node that has a Knox Gateway service instance configured on it, making Knox API calls through the Knox load balancer can result in a connection timeout error. This is because for security reasons, the IP address of the request is preserved in the traffic passed through the load balancer. Preserving the IP address means that the load balancer will reject "loopback" traffic, meaning traffic that originates and is directed back to the same node.

Workaround: If Knox API calls need to be made while logged into a Knox gateway node, use the hostname of the node instead of the load balancer hostname in the API call.

The Knox load balancer hostname can be identified by the "-gateway" suffix in the first clause of the hostname with no numeric identifier. For example: <cluster-name>-gateway.<env-shortname>.<hash>.cloudera.site is the load balancer hostname, and <cluster-name>-gateway0.<env-shortname>.<hash>.cloudera.site is a direct node hostname.

CB-23926: Autoscaling must be stopped before performing FreeIPA upgrade

Problem: During autoscaling, compute nodes are stopped and cannot receive the latest FreeIPA IP address update in a stopped state. This causes the Cloudera Data Hub clusters to become unhealthy after the FreeIPA upgrade when autoscaling is enabled.

Workaround:

1. Disable autoscaling before perfoming FreeIPA upgrade.
2. Start all compute nodes, which are in a STOPPED state.
3. Perform the FreeIPA upgrade.
4. Enable autoscaling after FreeIPA upgrade is successful.

CB-23940: Ranger YARN repository is not cleaned up after deleting stopped cluster

Problem: When a stopped Cloudera Data Hub cluster is deleted, the Ranger YARN repository that belongs to the deleted Cloudera Data Hub cluster is not cleaned up. This causes errors when creating a cluster with the same cluster name.

Workaround: Start the Cloudera Data Hub cluster before deleting. If the cluster was deleted in stopped state, delete the YARN repository using the Ranger Admin UI.

Technical Service Bulletins

TSB 2022-568: HBase normalizer must be disabled for Salted Phoenix tables

When Apache Phoenix ("Phoenix") creates a salted table, it pre-splits the table according to the number of salt regions. These regions must always be kept separate, otherwise Phoenix does not work correctly.

The HBase normalizer is not aware of this requirement, and in some cases the pre-split regions are merged automatically. This causes failure in Phoenix.

The same requirement applies when merging regions of salted tables manually: regions containing different salt keys (the first byte of the rowkey) must never be merged.

Note that either automatic or manual splitting of the regions for a salted table does not cause a problem. The problem only occurs when adjacent regions containing different salt keys are merged.

Upstream JIRA

PHOENIX-4906

Knowledge article

For the latest update on this issue, see the corresponding Knowledge article: TSB 2022-568: Hbase normalizer must be disabled for Salted Phoenix tables

Image catalog updates

This section lists prewarmed image updates in Cloudera Data Hub.

The following prewarmed images were added on the following dates:

First released	Platform	Runtime component versions	Cloudera Manager version	Available OSs
April 3, 2024	Cloudera Runtime 7.2.18	7.2.18 Component Versions	Cloudera Manager 7.12.0	RHEL 8.8 EUS
June 27, 2023	Cloudera Runtime 7.2.17	7.2.17 Component Versions	Cloudera Manager 7.11.0	RHEL 8.8 EUS, CentOS 7
January 11, 2023	Cloudera Runtime 7.2.16	7.2.16 Component Versions	Cloudera Manager 7.9.0	CentOS 7
May 12, 2022	Cloudera Runtime 7.2.15	7.2.15 Component Versions	Cloudera Manager 7.6.2	CentOS 7
February 25, 2022	Cloudera Runtime 7.2.14	7.2.14 Component Versions	Cloudera Manager 7.6.0	CentOS 7
October 25, 2021	Cloudera Runtime 7.2.12	7.2.12 Component Versions	Cloudera Manager 7.5.2	CentOS 7
September 9, 2021	Cloudera Runtime 7.2.11	7.2.11 Component Versions	Cloudera Manager 7.4.3	CentOS 7
June 14, 2021	Cloudera Runtime 7.2.10	7.2.10 Component Versions	Cloudera Manager 7.4.2	CentOS 7
April 30, 2021	Cloudera Runtime 7.2.9	7.2.9 Component Versions	Cloudera Manager 7.4.1	CentOS 7
March 25, 2021	Cloudera Runtime 7.2.8	7.2.8 Component Versions	Cloudera Manager 7.4.0	CentOS 7
February 9, 2021	Cloudera Runtime 7.2.7	7.2.7 Component Versions	Cloudera Manager 7.3.0	CentOS 7
December 11, 2020	Cloudera Runtime 7.2.6	7.2.6 Component Versions	Cloudera Manager 7.2.6	CentOS 7

First released	Platform	Runtime component versions	Cloudera Manager version	Available OSs
October 8, 2020	Cloudera Runtime 7.2.2	7.2.2 Component Versions	Cloudera Manager 7.2.2	CentOS 7
July 31, 2020	Cloudera Runtime 7.2.1	7.2.1 Component Versions	Cloudera Manager 7.2.1	CentOS 7
June 16, 2020	Cloudera Runtime 7.2.0	7.2.0 Component Versions	Cloudera Manager 7.2.0	CentOS 7
February 20, 2020	Cloudera Runtime 7.1.0	7.1.0 Component Versions	CCloudera Manager 7.1.0	CentOS 7

For the latest image catalog updates, refer to Cloudera web interface > Cloudera Management Console > Shared Resources > Image Catalogs.