

cod-on-high-performance-storage

Date published: 2020-08-14

Date modified: 2023-01-12

CLOUDERA

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Overview..... 4

Using high performance storage..... 4

Installing CDP CLI beta..... 5

Autoscaling..... 5

Overview

Cloudera Operational Database (COD) supports both high performance storage and Amazon S3 or Azure Data Lake Gen2 storage through the abfs connector as a storage layer. While the default option is cloud storage, COD also provides you an option to choose high performance storage.

You can use high performance storage either in an on-premises or a cloud environment to maintain low-latency and high-throughput for your application. If your applications are latency sensitive, Cloudera recommends that you use high performance storage for your COD databases. The average read and write latency for COD on a high performance storage is 500% better than the COD that uses S3 storage for all the workloads.



Note: COD on GCP environment supports only high performance storage.

HDFS volume type

Hadoop Distributed File System (HDFS) is a Java-based file system for storing large volumes of data. Designed to span large clusters of commodity servers, HDFS provides scalable and reliable data storage.

An HDFS cluster contains the following main components: a NameNode and DataNodes. The NameNode manages the cluster metadata that includes file and directory structures, permissions, modifications, and disk space quotas. The file content is split into multiple data blocks, with each block replicated at multiple DataNodes.

For more information, see *HDFS Overview*.

While creating a COD database, you have the following two options for storing your data when you select HDFS as the storage type.

- **HDD (Hard Disk Drives):** HDD is a cost-efficient option if you want to access the data frequently. This is a preferred storage option for applications that are not very sensitive to latency.
- **SSD (Solid State Drives):** SSDs provide more storage volume, speed, and efficiency; however, this storage option could be expensive compared to HDD.

For more information on the difference between SSD and HDD, see [Amazon documentation](#).

Worker nodes

Worker nodes store the data on your COD cluster. The general responsibilities of the worker nodes on your COD cluster include processing the data stored in the cluster and managing the network traffic. The worker nodes also ensure smooth operations between the applications across multiple COD clusters.

When you create a COD cluster, the minimum and maximum number of worker nodes vary for different storage types.

- **Micro duty:** Minimum node count: 1. Maximum node count: 5.
- **Light duty:** Minimum node count: 3. Maximum node count: 100.
- **Heavy duty:** Minimum node count: 3. Maximum node count: 800.

Related Information

[HDFS Overview](#)

Using high performance storage

You can create a COD database on high performance storage through CDP CLI beta version.

Procedure

Provide the following API request to create a COD database with high performance storage from CDP CLI beta version 0.9.40.

```
$ cdp opdb create-database --environment-name foo --database-name bar --use-hdfs
```

```
--use-hdfs | --no-use-hdfs (boolean) - Controls whether hbase root be deployed on hdfs
```

Related Information

[Installing CDP CLI beta](#)

Installing CDP CLI beta

To use COD on high performance storage through CDP CLI, you must download and install CDP CLI beta version 0.9.40.

Before you begin

Do not install both the regular and beta CLIs in the same Python environment, as they use the same entry points and will therefore conflict. Either use a separate virtual environment or uninstall the `cdpcli` first before installing `cdpcli-beta`.

Procedure

1. Install Python if needed, as described in [CDP CLI documentation](#) for your platform.
Do not install CDP CLI. Instead, proceed to step 2.
2. Run the following to install CDP CLI beta:

```
pip3 install cdpcli-beta
```

3. Configure access keys as described in [Generating an API access key](#)

Related Information

[Using high performance storage](#)

Autoscaling

COD with cloud storage type uses high performance storage to store Write-Ahead Log (WALs); however COD with high performance storage type uses both WALs and HFiles.

COD provides autoscaling for both the storage types. Your database automatically autoscales whenever the high performance storage reaches 60% of its buffer capacity to accommodate the growth in the data.