

Disabling S3Guard in an Existing CDP Environment (Preview)

Date published: 2022-10-05

Date modified: 2022-10-05

Legal Notice

© Cloudera Inc. 2022. 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.

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Contents

Legal Notice	2
Contents	3
Pre-requisites	4
How-To	4

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cloudera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

Pre-requisites

- The environment version should be Cludera Runtime 7.2.2 or older.

How-To

On an existing CDP environment with S3Guard, disabling S3Guard requires the following manual steps. Note that the workloads may give inconsistent results while these steps are in progress. Stop all the workloads during this operation.

1. Download the latest Beta CDP CLI. For more information see [Installing Beta CDP CLI](#).

```
pip install cpdcli-beta
```

2. Write down the DynamoDB table associated with the environment, then go to the environment in the CDP UI. Click the **Summary** tab --> **Advanced**.
3. In the Beta CLI, run the command to disable S3Guard for the environment:

```
cdp environments disable-s3-guard --environment-name
<environment name>
```

This command returns the entire environment response as output, and may contain the last environment status like "Error message: \"Cannot update the status of the stack to STOPPED when stack is in UPDATE_IN_PROGRESS state.\"". Note that this status is just the last environment status, and doesn't imply the results of the `disable s3guard` operation. Verify that the returned environment response does not contain 's3guard' under **awsdetails: {}**. This will be fixed soon in an upcoming release.

4. Data Lake: Go to **Cludera Manager** --> **Clusters** --> **hdfs** --> **Configuration** --> **Cluster-wide Advanced Configuration Snippet (Safety Valve)** for `core-site.xml`. Update the following config:

```
fs.s3a.metastore.impl=org.apache.hadoop.fs.s3a.s3guard.Null
MetadataStore
```

Restart all the services in the Data Lake.

5. For each of the Datahubs: Go to **Cludera Manager** --> **Clusters** --> **hdfs** --> **Configuration** --> **Cluster-wide Advanced Configuration Snippet (Safety Valve)** for

This document has been released as part of a technical preview for features described herein. Technical preview components are provided as a convenience to our customers for their evaluation and trial usage. These components are provided 'as is' without warranty or support. Further, Cludera assumes no liability for the usage of technical preview components, which should be used by customers at their own risk.

`core-site.xml`. Update the following config:

```
fs.s3a.metastore.impl=org.apache.hadoop.fs.s3a.s3guard.Null
MetadataStore
```

Restart all the services in the Data Hub.

6. Verify that S3guard is disabled:

```
cdp environments describe-environment --environment-name
<environment name>
```

The output shouldn't contain a section with "S3GUARD" under **awsdetails: {}**

7. **Data Warehouse:** For both 'SDX Datalake' and 'CDW Datalake', you can disable S3Guard by editing the configs for `core-site.xml` in DataCatalog and Virtual Warehouse. Go to **Edit** → **Configurations** →, click on each component, and add the value:

```
fs.s3a.metastore.impl=org.apache.hadoop.fs.s3a.s3guard.Null
MetadataStore
```

to **both** `hadoop-core-site` and `hadoop-core-site-default-warehouse`. This should restart the Virtual Warehouse and Database Catalog.

8. **Data Engineering:** The S3Guard config will be updated automatically every hour, synchronized with config-manager pod start.
9. **Machine Learning:** With CML 2.0.19, S3Guard configs will be updated automatically.
10. Verify that your workloads run successfully.
11. Drop the DynamoDB table from the AWS console. You should have the table name from Step 1 above.