

Configure Impala Virtual Warehouses on AWS Environments to Spill to S3 (Preview)

Date published: 2021-06-14

Date modified: 2021-06-14

Legal Notice

© Cloudera Inc. 2021. 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.

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
Overview	4
Before you begin	4
About this task	4
Steps	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.

Overview

Impala Virtual Warehouses on AWS environments can be configured to write temporary data (spill) to S3 by specifying the S3 URI when you are creating the Virtual Warehouse in Cludera Data Warehouse (CDW).

Before you begin

Identify the URI for an S3 bucket to which you want your new Impala Virtual Warehouse to write temporary data. If you use the S3 bucket that is associated with the environment, navigate to the S3 bucket and copy the URI in the AWS Console. You need to specify the URI in Step 5 below.

If you want to use an external S3 bucket, you must first configure CDW to use the external S3 bucket with the correct read/write permissions. For more information, see [Add external AWS S3 buckets to CDW](#). Then you must copy the S3 bucket URI to specify in Step 5.

About this task

You can only configure the S3 bucket to which you want your Impala Virtual Warehouse to write temporary data when you create a new Virtual Warehouse. After you have created the Virtual Warehouse configured to spill to a specific S3 location, you cannot change the S3 URI. The field becomes uneditable.

Steps

1. Log in to the CDP web interface and navigate to the Data Warehouse service.
2. In the Data Warehouse service, click Virtual Warehouses in the left navigation panel.
3. On the Virtual Warehouses page, click **Add New**.
4. In the New Virtual Warehouse dialog box, specify a **Name**, its **Type** (Impala), which **Database Catalog** it queries, whether to **Enable SSO, User Groups** that can access endpoints, keys and values for **Tagging** the Virtual Warehouse, and the **Size**.

After you select a **Size**, you can specify the remaining fields listed in Steps 5 and 6.

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.

5. In the **Spill to S3** field, specify the S3 URI to which you want the Impala Virtual Warehouse to write temporary data.

Important: If you specify an S3 URI for the **Spill to S3** field, you cannot select additional storage in the **Scratch Space Limit per node** field. You can only use the default 300 instance storage that is set automatically for that field when you specify a location for **Spill to S3**.

6. Set the following auto-scaling properties:
 - **Enable Impala HA** is turned on by default, but if you need to you can disable high availability.
 - **Disable AutoSuspend** is turned off by default. This means that the Virtual Warehouse auto-suspends itself if there are no queries to be processed. Keeping this turned off enables you to save on cloud resource usage.
 - **AutoSuspend Timeout**
 - **Scale Up Delay**
 - **Scale Down Delay**

For information about these auto-scaling properties, see [Tuning Impala Virtual Warehouses](#).

7. Click **Create** to create the new Impala Virtual Warehouse.