

Connecting to Ozone S3 external accounts for Cloudera Data Engineering jobs or sessions

Date published: 2020-07-30

Date modified: 2026-06-18



Legal Notice

© Cloudera Inc. 2026. 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

Connecting to S3 compatible accounts for Cloudera Data Engineering jobs or sessions.....	4
---	----------

Connecting to S3 compatible accounts for Cloudera Data Engineering jobs or sessions

Connect to a S3 compatible account securely in the Cloudera Data Engineering jobs or sessions.

About this task



Important:

- S3 compatible accounts include Ozone, AWS and Minio.
- Do not assign multiple accounts to a single workload user. Spark job cannot automatically select the correct account to use when connecting.

Procedure

1. Create the S3 compatible accounts.

After you create or edit the S3 external account, Cloudera Data Engineering automatically configures the S3 external accounts for the jobs or sessions that belong to that specific workload user. However, it takes approximately 10 minutes to update in the Cloudera Data Engineering jobs or sessions.

- ### 2. To connect to S3 external accounts in Cloudera Data Engineering, run the Spark job as the workload user configured for the S3-compatible account. The Spark job automatically uses the account credentials during Hadoop S3 read and write operations. You do not need any additional configuration to specify credentials or endpoints. For more information about creating jobs, see [Creating jobs in Cloudera Data Engineering](#).

Sample job:

```
import os
from pyspark.sql import SparkSession

def main():
    spark = SparkSession.builder \
        .appName("Spark S3A ") \
        .getOrCreate()
    try:
        # Read a CSV file from S3 using the s3a:// prefix
        df = spark.read.csv("s3a://your-bucket-name/input/data.csv", header=True, inferSchema=True)
        df.show()
    except Exception as e:
        print(f"Error executing Spark job: {e}")
    finally:
        spark.stop()

if __name__ == "__main__":
    main()
```

- ### 3. Import self-signed or untrusted certificates for S3-compatible accounts by setting the CA Certificate Type to Datalake in the Cloudera Management Console. For information about how to import certificates, see [Updating TLS certificates](#).