Machine Learning 1.5.0

Troubleshooting Cloudera Machine Learning

Date published: 2020-07-16
Date modified: 2023-01-31

CLOUDERA

Legal Notice

Contents

Troubleshooting

This topic describes a recommended series of steps to help you start diagnosing issues with a Cloudera Machine Learning workspace.

- Issues with Provisioning ML Workspaces: If provisioning an ML workspace fails, make sure that you have all the resources required to provision an ML workspace. If failures persist, start debugging by reviewing the error messages on the screen. Check the workspace logs to see what went wrong.
- Issues with Accessing ML Workspaces: If your ML Admin has already provisioned a workspace for you but attempting to access the workspace fails, confirm with your ML Admin that they have completed all the steps required to grant you access.
- Issues with Running Workloads: If you have access to a workspace but are having trouble running sessions/jobs/ experiments, and so on, see if your error is already listed here: Troubleshooting Issues with Workloads on page 4.

Cloudera Support

If you need assistance, contact Cloudera Support. Cloudera customers can register for an account to create a support ticket at the support portal. For CDP issues in particular, make sure you include the Request ID associated with your error message in the support case you create.

Downloading diagnostic bundles for a workspace

Learn how to manage and download diagnostic bundles.

The CDP platform provides various services for managing and downloading diagnostic bundles.

You can download diagnostic bundles from the Cloudera Machine Learning workspace. For more information, see *Options for generating the CDP Private Cloud diagnostic data*.

You can also send usage and diagnostic data from Cloudera Manager. For more information, see: *Sending Usage and Diagnostic Data to Cloudera*.

Related Information
Options for generating the CDP Private Cloud diagnostic data
Sending Usage and Diagnostic Data to Cloudera

Troubleshooting Issues with Workloads

This section describes some potential issues data scientists might encounter once the ML workspace is running workloads.

401 Error caused by incompatible Data Lake version

The following error might occur due to an incompatible Data Lake version:

```
org.apache.ranger.raz.hook.s3.RazS3ClientCredentialsException: Exception in
Raz Server;
Check the raz server logs for more details, HttpStatus: 401
```

To avoid this issue, ensure that:

- Data Lake and Runtime (server) version is 7.2.11 or higher.
- Hadoop Runtime add-on (client) used in the CML session is 7.2.11 or higher.

- Spark Runtime add-on version must be CDE 1.13 or higher.

Engines cannot be scheduled due to lack of CPU or memory

A symptom of this is the following error message in the Workbench: "Unschedulable: No node in the cluster currently has enough CPU or memory to run the engine."

Either shut down some running sessions or jobs or provision more hosts for Cloudera Machine Learning.

Workbench prompt flashes red and does not take input

The Workbench prompt flashing red indicates that the session is not currently ready to take input.

Cloudera Machine Learning does not currently support non-REPL interaction. One workaround is to skip the prompt using appropriate command-line arguments. Otherwise, consider using the terminal to answer interactive prompts.

PySpark jobs fail due to Python version mismatch

```
Exception: Python in worker has different version 2.6 than that in driver 2.
7, PySpark cannot run with different minor versions
```

One solution is to install the matching Python 2.7 version on all the cluster hosts. A better solution is to install the Anaconda parcel on all CDH cluster hosts. Cloudera Machine Learning Python engines will use the version of Python included in the Anaconda parcel which ensures Python versions between driver and workers will always match. Any library paths in workloads sent from drivers to workers will also match because Anaconda is present in the same location across all hosts. Once the parcel has been installed, set the PYSPARK_PYTHON environment variable in the Cloudera Machine Learning Admin dashboard.

Troubleshooting Kerberos Errors

This topic describes some common Kerberos issues and their recommended solutions.

HDFS commands fail with Kerberos errors even though Kerberos authentication is successful in the web application

If Kerberos authentication is successful in the web application, and the output of klist in the engine reveals a valid-looking TGT, but commands such as hdfs dfs -ls / still fail with a Kerberos error, it is possible that your cluster is missing the Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy File. The JCE policy file is required when Red Hat uses AES-256 encryption. This library should be installed on each cluster host and will live under $JAVA_HOME. For more information, see Using AES-256 Encryption.

Cannot find renewable Kerberos TGT

Cloudera Machine Learning runs its own Kerberos TGT renewer which produces non-renewable TGT. However, this confuses Hadoop's renewer which looks for renewable TGTs. If the Spark 2 logging level is set to WARN or lower, you may see exceptions such as:

```
16/12/24 16:38:40 WARN security.UserGroupInformation: Exception encountered
while running the renewal command. Aborting renew thread. ExitCodeException
exitCode=1: kinit: Resource temporarily unavailable while renewing credentia
ls

16/12/24 16:41:23 WARN security.UserGroupInformation: PriviledgedActionExcep
tion as:user@CLOUDERA.LOCAL (auth:KERBEROS) cause:javax.security.sasl.SaslEx
ception: GSS initiate failed [Caused by GSSException: No valid credentials p
rovided (Mechanism level: Failed to find any Kerberos tgt)]
```

This is not a bug. Spark 2 workloads will not be affected by this. Access to Kerberized resources should also work as expected.