

Monitoring Cloudera Data Engineering jobs

Date published: 2020-07-30

Date modified: 2024-05-30



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

Cloudera Data Engineering log files.....	4
Viewing Job run timeline.....	4
Using Spark history server to troubleshoot Spark jobs.....	6

Cloudera Data Engineering log files

You can view logs for Cloudera Data Engineering (CDE) using the web console, including CDE service logs, virtual cluster logs, and job logs.

To view logs for a CDE service or virtual cluster, click the three-dot menu for the service or virtual cluster, and then select View Logs. When the View Logs modal is displayed, you can download the logs or copy them to the clipboard by clicking the associated icon at the top right of the modal.

To view logs for a job run:

1. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.
2. In the left navigation menu click Jobs. The Jobs page is displayed.
3. Using the dropdown menu, select the virtual cluster containing the job that you want to troubleshoot.
4. Select the job that you want to troubleshoot.
5. In the Run History tab, click the Run ID for the job run that you want to troubleshoot.
6. Go to the Logs tab.
7. Using the Select log type drop-down menu and the log file tabs, select the log you want to view .
8. To download the logs, click the Download menu button. You can download a text file of the currently displayed log, or download a zip file containing all log files.

Viewing Job run timeline

You can view the intermediate stages of the job run at every stage during its life cycle in real-time.

About this task

In case of a job failure, you can view the specific event and component where the job run failed. This reduces turnaround time during the debugging process for job run failure. You can see the step-by-step advancement of the job run on the UI, including all the granular details instead of reviewing extensive logs to obtain the same insights.

Procedure

1. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.
2. In the left navigation pane, click Jobs Runs. The Jobs Runs page displays.
3. Click on the Job Id for which you want to see the status.

4. Go to the Timeline tab. It displays the summary of the Job run progression in the reverse chronological order for both primary and subordinate stages.

The screenshot displays the Cloudera Data Engineering web interface. On the left is a dark sidebar with navigation links: Home, Jobs, Job Runs (highlighted with an orange box), Sessions, Repositories, Resources, and Administration. The main content area on the right shows the 'cluster-01 / Job Runs /' breadcrumb. Below this, a 'Status' section indicates 'Succeeded'. A 'Timeline' tab (also highlighted with an orange box) is selected, with a 'New' button next to it. The 'Summary' section below the timeline lists job run progressions in reverse chronological order:

- Succeeded**
Component: Runtime API Server
[Show Details](#)
- Running**
Component: Spark Container
[Show Details](#)
- Starting**
Component: Runtime API Server

Using Spark history server to troubleshoot Spark jobs

The Spark history server is a monitoring tool that displays information about completed Spark applications. It provides information for debugging such as Spark configurations, DAG execution, driver and executor resource utilization, application logs, and job, stage and task-level details.



Important: To access and load the Spark UI correctly, ensure that the third party cookies for `[*.]cloudera.com` and `[*.]cloudera.site` are enabled.

To view spark history server information for a job run:

1. In the Cloudera Data Platform (CDP) console, click the Data Engineering tile. The CDE Home page displays.
2. Click Jobs in the left navigation menu.
3. From the drop-down in the upper left-hand corner, select the Virtual Cluster that you want to restore jobs to.
4. Select the job that you want to troubleshoot.
5. Click Jobs Runs in the left menu, and click the Run ID for the job run you want to view the information.
6. Click the Spark UI tab to access the Spark History Server.