Cloudera Data Engineering 1.5.4

Monitoring Cloudera Data Engineering jobs

Date published: 2020-07-30
Date modified: 2024-05-30

CLOUDƎRA

https://docs.cloudera.com/

Legal Notice

Contents

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.
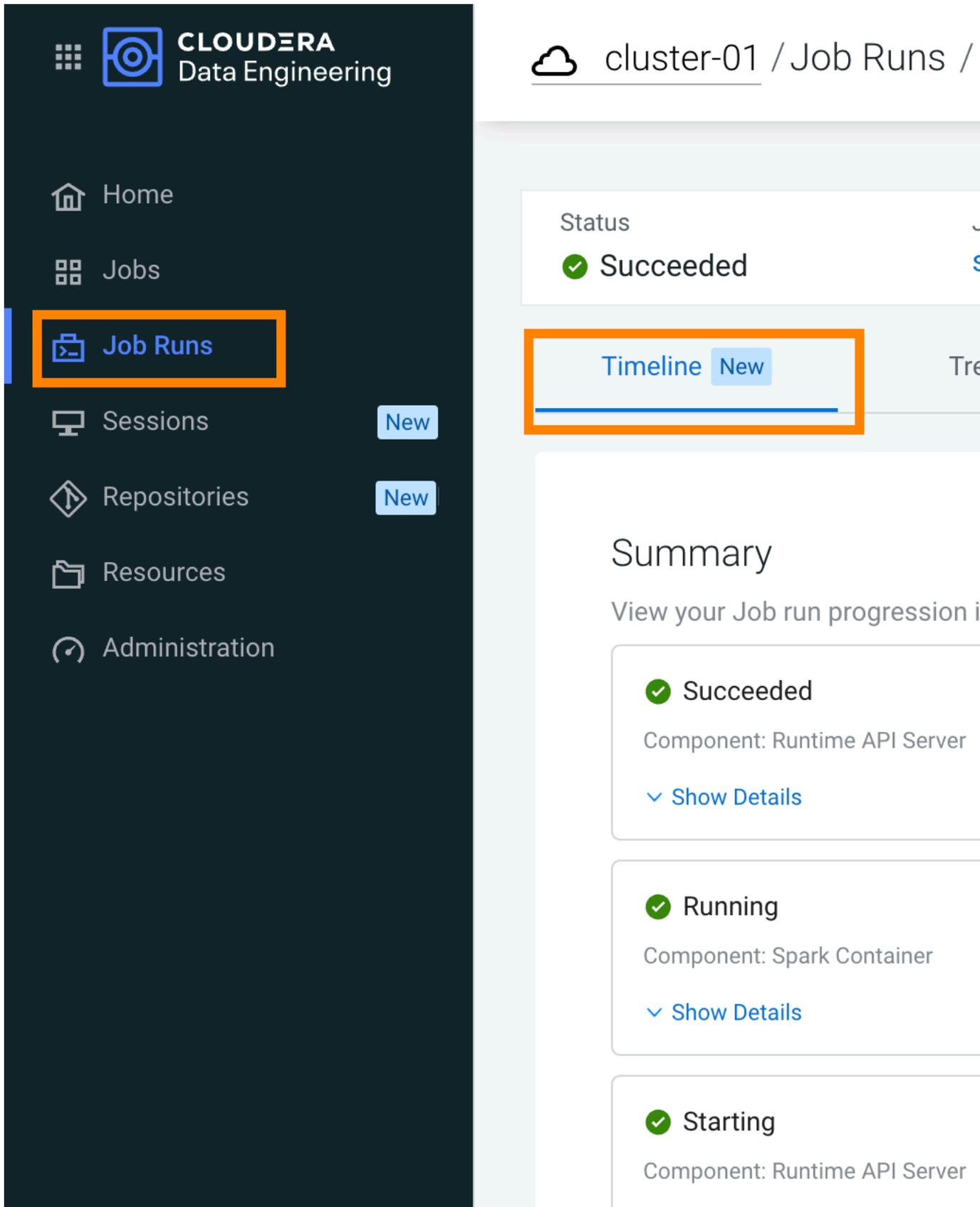
Using Spark history server to troubleshoot Spark jobs

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.