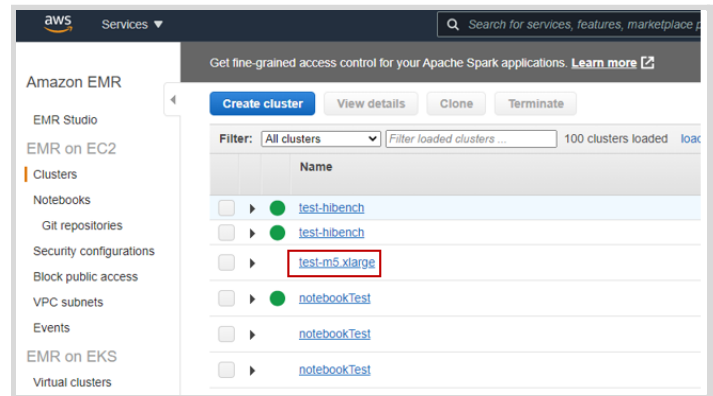


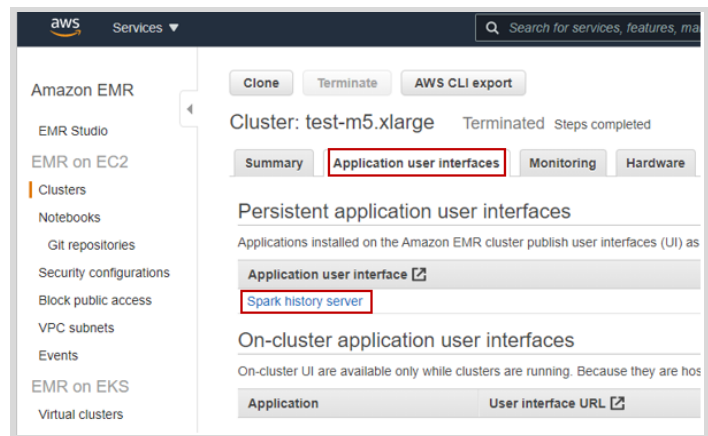
Sync Computing

Spark EMR Log Download Instructions

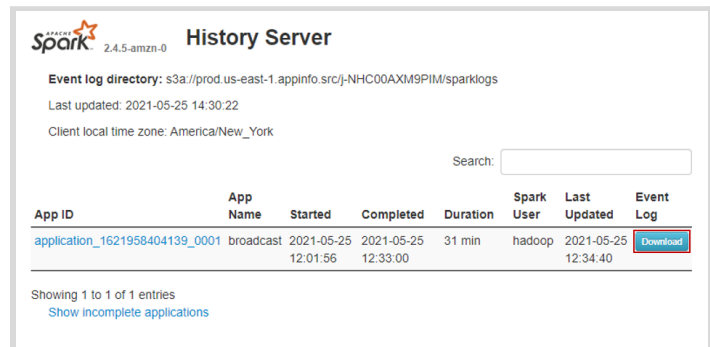
1. Go to the EMR console in AWS, and find the cluster that ran the job you are interested in optimizing. Click on the cluster name to view details of the cluster.



2. Once you are in the cluster information page, click on the “Application user interfaces” tab, and click on “Spark history server” (in red below) under “Persistent application user interfaces.”



3. A new tab should open up with the Spark history server. It may take a minute to load. Click the download button under the event log column to download the Spark event log. That is the event log we use for the Sync analyzer.



Instructions for EMR Version 5.24 or Older

1. Enable the eventlog flag and the HDFS directory of where it should go:
 1. `spark.eventLog.enabled true`
`spark.eventLog.dir hdfs://namenode/shared/spark-logs`
2. Rerun the job
3. Manually move the log from HDFS to an s3 bucket
 1. `aws s3 cp hdfs:///path/to/locallog /path/to/s3_log_bucket`
4. Go in and find the new log in the s3 bucket chosen above