

1. Running cluster

Clusters / 603_cluster

603_cluster 

More  Terminate 

Configuration Notebooks (1) Libraries Event log **Spark UI** Driver logs Metrics Apps Spark cluster UI - Master 

Hostname: ec2-35-87-251-99.us-west-2.compute.amazonaws.com Spark Version:11.2.x-cpu-ml-scala2.12

[Open in new tab](#) 

Jobs Stages Storage Environment Executors SQL / DataFrame JDBC/ODBC Server Structured Streaming

Spark Jobs ^(?)

User: root
Total Uptime: 35 min
Scheduling Mode: FAIR
Completed Jobs: 1

▶ Event Timeline

▼ Completed Jobs (1)

Page:

1 Pages. Jump to . Show items in a page.

Job Id (Job Group) ▼	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
0 (7523936284923249986_4962347490570141111_2daa21efa6a64028ad8b891db49547c1)	SELECT 1+1 collectResult at OutputAggregator.scala:267	2022/09/12 17:52:08	2 s	1/1	1/1

Page:

1 Pages. Jump to . Show items in a page.

2. Markdown and SQL cell in the notebook

data603fa22_desai_hinal_week1

Python



603_cluster File Edit View: Standard Permissions Run All Clear Publish Comments Experiment Revision history

Cmd 1

HDFS NameNode and DataNode

HDFS has a master/slave architecture. NameNode is the master node in the HDFS architecture. Hence we have 1 NameNode. It is the master server that manages the file system namespace and regulates access to files by clients. DataNodes act as the server nodes and hence we have multiple DataNodes. They manage the storage attached to the nodes that they run on. HDFS exposes a file system namespace and allows user data to be stored in files. Internally, a file is split into one or more blocks and these blocks are stored in a set of DataNodes. The NameNode executes file system namespace operations like opening, closing, and renaming files and directories. It also determines the mapping of blocks to DataNodes. The DataNodes are responsible for serving read and write requests from the file system's clients. The DataNodes also perform block creation, deletion, and replication upon instruction from the NameNode.

Cmd 2

```
1 %sql
2 SELECT 1+1
```

▶ (1) Spark Jobs

	(1 + 1)
1	2

Showing all 1 rows.

Command took 2.87 seconds -- by hdesai2@umbc.edu at 9/12/2022, 1:52:07 PM on 603_cluster

Shift+Enter to run