

Apache Spark UI

Access the Application UI

Start your application

Connect to the application UI using the following URL:

```
http://<driver-node>:4040
```

The SparkContext starts a web server for the application UI

- The driver program is the host
- The port defaults to 4040
- The UI is available only while the application is active

IBM Developer

SKILLS NETWORK 

Why use the Application UI

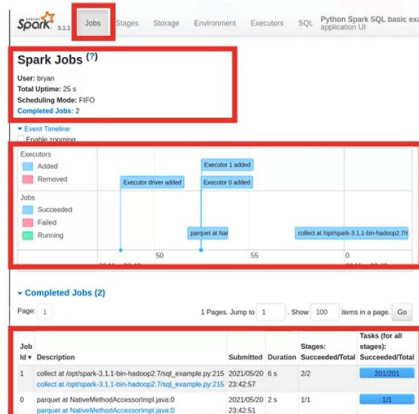
The Spark user interface shows information about the running application:

- Shows jobs, stages and tasks
- Storage of persisted RDDs and DataFrames
- Environment configuration and properties
- Executor summary
- SQL information (if SQL queries exist)

IBM Developer

SKILLS NETWORK 

Jobs Information



The Jobs tab

Jobs summary info

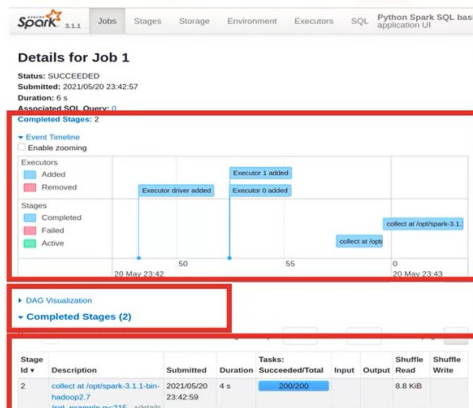
Jobs event timeline

Click the Description link to view completed job details

IBM Developer

SKILLS NETWORK 

Job Details



View the Job Stages timeline and job status

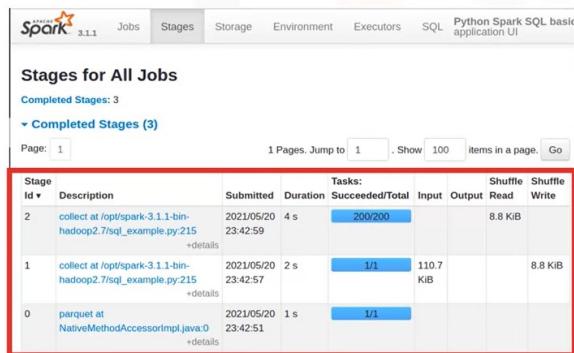
Quickly view stage details based on status

Click the Description link to view completed stage details

IBM Developer

SKILLS NETWORK 

Stages Information



The screenshot shows the 'Stages' tab in the Databricks UI. It displays a table of completed stages for all jobs. The table has columns for Stage Id, Description, Submitted, Duration, Tasks (Succeeded/Total), Input, Output, Shuffle Read, and Shuffle Write. The first three rows are highlighted with a red box.

Stage Id	Description	Submitted	Duration	Tasks: Succeeded/Total	Input	Output	Shuffle Read	Shuffle Write
2	collect at /opt/spark-3.1.1-bin-hadoop2.7/sql_example.py:215	2021/05/20 23:42:59	4 s	200/200			8.8 KiB	
1	collect at /opt/spark-3.1.1-bin-hadoop2.7/sql_example.py:215	2021/05/20 23:42:57	2 s	1/1	110.7 KiB			8.8 KiB
0	parquet at NativeMethodAccessorImpl.java:0	2021/05/20 23:42:51	1 s	1/1				

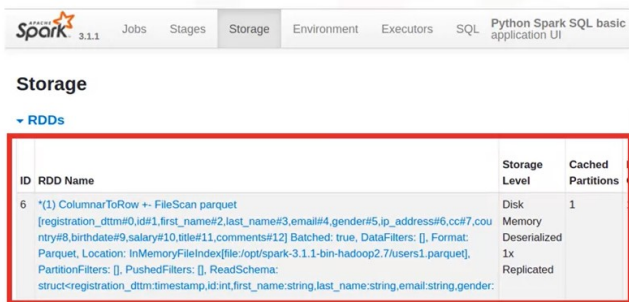
The Stages tab

Click the Description links to view task details for that stage.

IBM Developer

SKILLS NETWORK

Storage Information



The screenshot shows the 'Storage' tab in the Databricks UI. It displays a table of RDDs. The table has columns for ID, RDD Name, Storage Level, and Cached Partitions. The first row is highlighted with a red box.

ID	RDD Name	Storage Level	Cached Partitions
6	*[1] ColumnarToRow +- FileScan parquet [registration_dttm#0,id#1,first_name#2,last_name#3,email#4,gender#5,ip_address#6,cc#7,country#8,birthdate#9,salary#10,title#11,comments#12] Batched: true, DataFilters: [], Format: Parquet, Location: InMemoryFileIndex[file:/opt/spark-3.1.1-bin-hadoop2.7/users1.parquet], PartitionFilters: [], PushedFilters: [], ReadSchema: struct<registration_dttm:timestamp,id:int,first_name:string,last_name:string,email:string,gender:	Disk	1

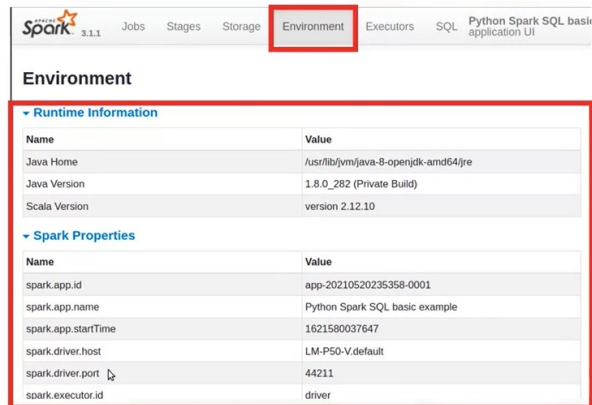
The Storage tab

Click the RDD Name link to view task details for that stage.

IBM Developer

SKILLS NETWORK

Environment Information



The screenshot shows the Databricks interface with the 'Environment' tab selected. The 'Runtime Information' section is expanded, showing details about the Java and Scala versions. The 'Spark Properties' section is also expanded, showing various configuration properties like spark.app.id, spark.app.name, spark.app.startTime, spark.driver.host, spark.driver.port, and spark.executor.id.

Name	Value
Java Home	/usr/lib/jvm/java-8-openjdk-amd64/jre
Java Version	1.8.0_282 (Private Build)
Scala Version	version 2.12.10

Name	Value
spark.app.id	app-20210520235358-0001
spark.app.name	Python Spark SQL basic example
spark.app.startTime	1621580037647
spark.driver.host	LM-P50-V.default
spark.driver.port	44211
spark.executor.id	driver

The Environment tab

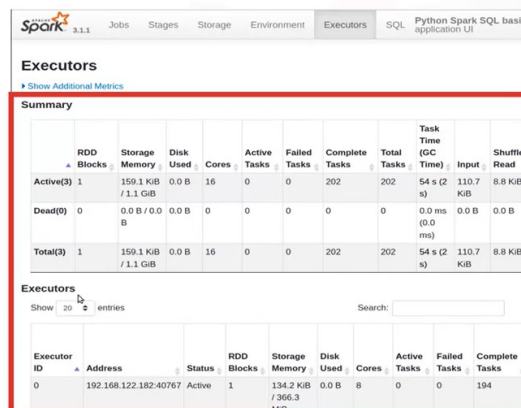
View:

- Spark configuration properties
- Resource profiles
- Hadoop properties
- System configuration properties

IBM Developer

SKILLS NETWORK 

Executors Information



The screenshot shows the Databricks interface with the 'Executors' tab selected. The 'Summary' section is expanded, showing a table of executor metrics. The 'Executors' section is also expanded, showing a table of individual executor details.

	RDD Blocks	Storage Memory	Disk Used	Cores	Active Tasks	Failed Tasks	Complete Tasks	Total Tasks	Task Time (GC Time)	Input	Shuffle Read
Active(3)	1	159.1 KiB / 1.1 GiB	0.0 B	16	0	0	202	202	54 s (2 s)	110.7 KiB	8.8 KiB
Dead(0)	0	0.0 B / 0.0 B	0.0 B	0	0	0	0	0	0.0 ms	0.0 B	0.0 B
Total(3)	1	159.1 KiB / 1.1 GiB	0.0 B	16	0	0	202	202	54 s (2 s)	110.7 KiB	8.8 KiB

Executor ID	Address	Status	RDD Blocks	Storage Memory	Disk Used	Cores	Active Tasks	Failed Tasks	Complete Tasks
0	192.168.122.182-40767	Active	1	134.2 KiB / 366.3 MiB	0.0 B	8	0	0	194

The Executors tab

- Scroll the window to view
- Spark configuration properties
 - Resource profiles
 - Hadoop properties
 - System configuration properties

IBM Developer

SKILLS NETWORK 

