

Report for Task3. Spark Basics

My operation system is Ubuntu 18.04. It is all executed in it.

1. Start the Master node.

```
(venv) tonipaltus@ToniPaltus:~/Innowise/spark_demo$ spark-class org.apache.spark.deploy.master.Master
Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties
23/04/25 01:05:13 INFO Master: Started daemon with process name: 9795@ToniPaltus
23/04/25 01:05:13 INFO SignalUtils: Registering signal handler for TERM
23/04/25 01:05:13 INFO SignalUtils: Registering signal handler for HUP
23/04/25 01:05:13 INFO SignalUtils: Registering signal handler for INT
23/04/25 01:05:13 WARN Utils: Your hostname, ToniPaltus resolves to a loopback address: 127.0.1.1; using 192.168.0.109 instead (on interface wlo1)
23/04/25 01:05:13 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
23/04/25 01:05:13 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
23/04/25 01:05:13 INFO SecurityManager: Changing view acls to: tonipaltus
23/04/25 01:05:13 INFO SecurityManager: Changing modify acls to: tonipaltus
23/04/25 01:05:13 INFO SecurityManager: Changing view acls groups to:
23/04/25 01:05:13 INFO SecurityManager: Changing modify acls groups to:
23/04/25 01:05:13 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(tonipaltus); groups with view permissions: Set(); users with modify permissions: Set(tonipaltus); groups with modify permissions: Set()
23/04/25 01:05:14 INFO Utils: Successfully started service 'sparkMaster' on port 7077.
23/04/25 01:05:14 INFO Master: Starting Spark master at spark://192.168.0.109:7077
23/04/25 01:05:14 INFO Master: Running Spark version 3.3.2
23/04/25 01:05:14 INFO Utils: Successfully started service 'MasterUI' on port 8080.
23/04/25 01:05:14 INFO MasterWebUI: Bound MasterWebUI to 0.0.0.0, and started at http://192.168.0.109:8080
```

2. Start the worker 1.

3.

```
(venv) tonipaltus@ToniPaltus:~/Innowise/spark_demo$ spark-class org.apache.spark.deploy.worker.Worker spark://192.168.0.109:7077 --cores 2 --memory 3g
Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties
23/04/25 01:05:57 INFO Worker: Started daemon with process name: 10013@ToniPaltus
23/04/25 01:05:57 INFO SignalUtils: Registering signal handler for TERM
23/04/25 01:05:57 INFO SignalUtils: Registering signal handler for HUP
23/04/25 01:05:57 INFO SignalUtils: Registering signal handler for INT
23/04/25 01:05:57 WARN Utils: Your hostname, ToniPaltus resolves to a loopback address: 127.0.1.1; using 192.168.0.109 instead (on interface wlo1)
23/04/25 01:05:57 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
23/04/25 01:05:57 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
23/04/25 01:05:57 INFO SecurityManager: Changing view acls to: tonipaltus
23/04/25 01:05:57 INFO SecurityManager: Changing modify acls to: tonipaltus
23/04/25 01:05:57 INFO SecurityManager: Changing view acls groups to:
23/04/25 01:05:57 INFO SecurityManager: Changing modify acls groups to:
23/04/25 01:05:57 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(tonipaltus); groups with view permissions: Set(); users with modify permissions: Set(tonipaltus); groups with modify permissions: Set()
23/04/25 01:05:58 INFO Utils: Successfully started service 'sparkWorker' on port 44135.
23/04/25 01:05:58 INFO Worker: Worker decommissioning not enabled.
23/04/25 01:05:58 INFO Worker: Starting Spark worker 192.168.0.109:44135 with 2 cores, 3.0 GiB RAM
23/04/25 01:05:58 INFO Worker: Running Spark version 3.3.2
23/04/25 01:05:58 INFO Worker: Spark home: /opt/spark
23/04/25 01:05:58 INFO ResourceUtils: =====
23/04/25 01:05:58 INFO ResourceUtils: No custom resources configured for spark.worker.
23/04/25 01:05:58 INFO ResourceUtils: =====
23/04/25 01:05:58 INFO Utils: Successfully started service 'WorkerUI' on port 8081.
23/04/25 01:05:58 INFO WorkerWebUI: Bound WorkerWebUI to 0.0.0.0, and started at http://192.168.0.109:8081
23/04/25 01:05:58 INFO Worker: Connecting to master 192.168.0.109:7077...
```

Start the worker 2.

```
>>> (venv) tonipaltus@ToniPaltus:~/Innowise/spark_demo$ spark-class org.apache.spark.deploy.worker.Worker spark://192.168.0.109:7077 --cores 3 --memory 4g
Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties
23/04/25 01:06:30 INFO Worker: Started daemon with process name: 10182@ToniPaltus
23/04/25 01:06:30 INFO SignalUtils: Registering signal handler for TERM
23/04/25 01:06:30 INFO SignalUtils: Registering signal handler for HUP
23/04/25 01:06:30 INFO SignalUtils: Registering signal handler for INT
23/04/25 01:06:30 WARN Utils: Your hostname, ToniPaltus resolves to a loopback address: 127.0.1.1; using 192.168.0.109 instead (on interface wlo1)
23/04/25 01:06:30 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
23/04/25 01:06:31 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
23/04/25 01:06:31 INFO SecurityManager: Changing view acls to: tonipaltus
23/04/25 01:06:31 INFO SecurityManager: Changing modify acls to: tonipaltus
23/04/25 01:06:31 INFO SecurityManager: Changing view acls groups to:
23/04/25 01:06:31 INFO SecurityManager: Changing modify acls groups to:
23/04/25 01:06:31 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(tonipaltus); groups with view permissions: Set(); users with modify permissions: Set(tonipaltus); groups with modify permissions: Set()
23/04/25 01:06:31 INFO Utils: Successfully started service 'sparkWorker' on port 43223.
23/04/25 01:06:31 INFO Worker: Worker decommissioning not enabled.
23/04/25 01:06:31 INFO Worker: Starting Spark worker 192.168.0.109:43223 with 3 cores, 4.0 GiB RAM
23/04/25 01:06:31 INFO Worker: Running Spark version 3.3.2
23/04/25 01:06:31 INFO Worker: Spark home: /opt/spark
23/04/25 01:06:31 INFO ResourceUtils: =====
23/04/25 01:06:31 INFO ResourceUtils: No custom resources configured for spark.worker.
23/04/25 01:06:31 INFO ResourceUtils: =====
23/04/25 01:06:31 WARN Utils: Service 'WorkerUI' could not bind on port 8081. Attempting port 8082.
23/04/25 01:06:31 INFO Utils: Successfully started service 'WorkerUI' on port 8082.
23/04/25 01:06:31 INFO WorkerWebUI: Bound WorkerWebUI to 0.0.0.0, and started at http://192.168.0.109:8082
```

Look all this in UI:



Spark Master at spark://192.168.0.109:7077

URL: spark://192.168.0.109:7077
Alive Workers: 2
Cores in use: 5 Total, 0 Used
Memory in use: 7.0 GiB Total, 0.0 B Used
Resources in use:
Applications: 0 Running, 1 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers (2)

Worker Id	Address	State	Cores	Memory	Resources
worker-20230425010558-192.168.0.109-44135	192.168.0.109:44135	ALIVE	2 (0 Used)	3.0 GiB (0.0 B Used)	
worker-20230425010631-192.168.0.109-43223	192.168.0.109:43223	ALIVE	3 (0 Used)	4.0 GiB (0.0 B Used)	

Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
----------------	------	-------	---------------------	------------------------	----------------	------	-------	----------

Completed Applications (1)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
app-20230425010921-0000	spark_basic	5	1024.0 MiB		2023/04/25 01:09:21	tonipaltus	FINISHED	11 s

4. Create the history Server.
In file opt/spark/conf/spark-defaults.conf

```
spark.master                spark://master:7077
spark.eventLog.enabled      true
spark.eventLog.dir          /home/tonipaltus/Innowise/spark_demo/spark_demo_course/3_Spark_Basics/for_history
spark.history.fs.logDirectory /home/tonipaltus/Innowise/spark_demo/spark_demo_course/3_Spark_Basics/for_history
```

```
(venv) tonipaltus@ToniPaltus:~/Innowise/spark_demo$ spark-class org.apache.spark.deploy.history.HistoryServer
Using Spark's default log4j profile: org/apache/spark/log4j2-defaults.properties
23/04/25 01:09:57 INFO HistoryServer: Started daemon with process name: 10998@ToniPaltus
23/04/25 01:09:57 INFO SignalUtils: Registering signal handler for TERM
23/04/25 01:09:57 INFO SignalUtils: Registering signal handler for HUP
23/04/25 01:09:57 INFO SignalUtils: Registering signal handler for INT
23/04/25 01:09:58 WARN Utils: Your hostname, ToniPaltus resolves to a loopback address: 127.0.1.1; using 192.168.0.109 instead (on interface wlo1)
23/04/25 01:09:58 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
23/04/25 01:09:58 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
23/04/25 01:09:58 INFO SecurityManager: Changing view acls to: tonipaltus
23/04/25 01:09:58 INFO SecurityManager: Changing view acls groups to:
23/04/25 01:09:58 INFO SecurityManager: Changing modify acls groups to:
23/04/25 01:09:58 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(tonipaltus);
with view permissions: Set(); users with modify permissions: Set(tonipaltus); groups with modify permissions: Set()
23/04/25 01:09:58 INFO FsHistoryProvider: History server ui acls disabled; users with admin permissions: ; groups with admin permissions:
23/04/25 01:09:58 INFO Utils: Successfully started service 'HistoryServerUI' on port 18080.
23/04/25 01:09:58 INFO HistoryServer: Bound HistoryServer to 0.0.0.0, and started at http://192.168.0.109:18080
```



History Server

Event log directory: /home/tonipaltus/Innowise/spark_demo/spark_demo_course/3_Spark_Basics/for_history

Last updated: 2023-04-25 01:20:18

Client local time zone: Europe/Minsk

Search:

Version	App ID	App Name	Started	Completed	Duration	Spark User	Last Updated	Event Log
3.3.2	app-20230425010921-0000	spark_basic	2023-04-25 01:09:20	2023-04-25 01:09:32	12 s	tonipaltus	2023-04-25 01:09:32	Download
3.3.2	app-20230424185808-0001	spark_basic	2023-04-24 18:58:07	2023-04-24 18:58:14	7 s	tonipaltus	2023-04-24 18:58:14	Download
3.3.2	app-20230424184423-0000	spark_basic	2023-04-24 18:44:22	2023-04-24 18:44:29	7 s	tonipaltus	2023-04-24 18:44:29	Download

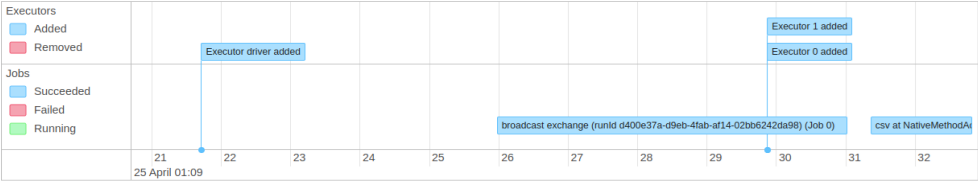
Showing 1 to 3 of 3 entries
[Show incomplete applications](#)

>>>
Jobs

Spark Jobs (?)

User: tonipallus
Total Uptime: 12 s
Scheduling Mode: FIFO
Completed Jobs: 2

Event Timeline
☐ Enable zooming



Completed Jobs (2)

Page: 1 1 Pages. Jump to 1 . Show 100 items in a page. Go

Job Id (Job Group) ▼	Description	Submitted	Duration	Stages:		Tasks (for all stages): Succeeded/Total
				Succeeded	Total	
1	csv at NativeMethodAccessorImpl.java:0 csv at NativeMethodAccessorImpl.java:0	2023/04/25 01:09:31	1 s	1/1		2/2
0 (d400e37a-d9eb-4fab-af14-02bb6242da98)	broadcast exchange (runId d400e37a-d9eb-4fab-af14-02bb6242da98) \$anonfun\$withThreadLocalCaptured\$1 at FutureTask.java:264	2023/04/25 01:09:25	5 s	1/1		1/1

Page: 1 1 Pages. Jump to 1 . Show 100 items in a page. Go