

BUAN 6346 Big Data

# Homework 13

# Write and run a Spark application

## Spark Application in Python:

```
[training@localhost ~]$ spark-submit CountJPGs.py /loudacre/weblogs/*
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/zookeeper/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/flume-ng/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
20/07/25 16:21:05 INFO spark.SparkContext: Running Spark version 1.3.0

20/07/25 16:21:52 INFO scheduler.TaskSetManager: Finished task 310.0 in stage 0.0 (TID 310) in 103 ms on localhost (311/311)
20/07/25 16:21:52 INFO scheduler.DAGScheduler: Stage 0 (count at /home/training/CountJPGs.py:15) finished in 32.098 s
20/07/25 16:21:52 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
20/07/25 16:21:52 INFO scheduler.DAGScheduler: Job 0 finished: count at /home/training/CountJPGs.py:15, took 32.298435 s
Number of JPG requests: 64978
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/metrics/json,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/stages/stage/kill,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/static,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/executors/threadDump/json,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/executors/threadDump,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/executors/json,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/executors,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/environment/json,null}
20/07/25 16:21:52 INFO handler.ContextHandler: stopped o.s.j.s.ServletContextHandler{/environment,null}
```

## Submit a Spark Application to Cluster:

```
[training@localhost ~]$ spark-submit \
> --master yarn-client \
> CountJPGs.py /loudacre/weblogs/*
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/zookeeper/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/flume-ng/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
```



## All Applications

Logged in as: dr.who

Cluster

About

Nodes

Applications

NEW

NEW\_SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
1	0	1	0	2	2 GB	2 GB	1 GB	2	8	1	1	0	0	0	0

User Metrics for dr.who

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Containers Pending	Containers Reserved	Memory Used	Memory Pending	Memory Reserved	VCores Used	VCores Pending	VCores Reserved
0	0	1	0	0	0	0	0 B	0 B	0 B	0	0	0

Show 20 entries

Search:

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCoers	Allocated Memory MB	Progress	Tracking UI
application_1595701066946_0001	training	CountJPGs.py	SPARK	root.training	Sat Jul 25 16:25:43 -0700 2020	N/A	RUNNING	UNDEFINED	2	2	2048		ApplicationMaster

Showing 1 to 1 of 1 entries

First Previous 1 Next Last



## All Applications

Logged in as: dr.who

Cluster

About

Nodes

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

### Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
1	0	0	1	0	0 B	2 GB	0 B	0	8	0	1	0	0	0	0

### User Metrics for dr.who

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Containers Pending	Containers Reserved	Memory Used	Memory Pending	Memory Reserved	VCores Used	VCores Pending	VCores Reserved
0	0	0	1	0	0	0	0 B	0 B	0 B	0	0	0

Show 20 entries

Search:

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCoers	Allocated Memory MB	Progress	Tracking UI
application_1595701066946_0001	training	CountPGs.py	SPARK	root.training	Sat Jul 25 16:25:43 -0700 2020	Sat Jul 25 16:28:02 -0700 2020	FINISHED	SUCCEEDED	N/A	N/A	N/A		History

Showing 1 to 1 of 1 entries

First

Previous

1

Next

Last