



UBDA Platform

Example for Spark

User Guide

Version 1.0

16 July 2018

Revision History

Version	Date	Prepared By	Summary of Changes
1.0	Jul 16, 2018		Initial release

Table of Contents

1. Introduction	4
2. Perform the test.....	5
3. How to check the result	7
4. Useful Reference.....	8

1. Introduction

This document is shown a Spark example running on the UBDA platform.

*Note: User should first register a user account through UBDA website at:
<https://www.polyu.edu.hk/pfs/index.php/177729> to access the UBDA Platform*

2. Perform the test

2.1 Login to ubdaplatform.polyu.edu.hk via SSH

2.2 Apply Hadoop on HDFS environment

```
$ source /ubda/bdaapps/bda_hdfs.sh
```

2.3 Submit the Pi job via spark

```
$ spark-submit --master spark://ubda-g003.hh.polyu.hk:7077,ubda-g005.hh.polyu.hk:7077 /ubda/bdaapps/spark/examples/src/main/python/pi.py 100 > result.out
```

2.4 Process of the Pi Job

```
18/07/26 16:50:33 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/26 16:50:34 INFO SparkContext: Running Spark version 2.3.0
18/07/26 16:50:34 WARN SparkConf: In Spark 1.0 and later spark.local.dir will be overridden by the value set by the cluster manager (via SPARK_LOCAL_DIRS in mesos/standalone and LOCAL_DIRS in YARN).
18/07/26 16:50:34 INFO SparkContext: Submitted application: PythonPi
18/07/26 16:50:34 INFO SecurityManager: Changing view acls to: ubdademo9
18/07/26 16:50:34 INFO SecurityManager: Changing modify acls to: ubdademo9
18/07/26 16:50:34 INFO SecurityManager: Changing view acls groups to:
18/07/26 16:50:34 INFO SecurityManager: Changing modify acls groups to:
18/07/26 16:50:34 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(ubdademo9); groups with view permissions: Set(); users with modify permissions: Set(ubdademo9); groups with modify permissions: Set()
.
.
.
18/07/26 16:50:46 INFO TaskSetManager: Finished task 46.0 in stage 0.0 (TID 46) in 1841 ms on 192.168.10.125 (executor 4) (96/100)
18/07/26 16:50:46 INFO TaskSetManager: Finished task 45.0 in stage 0.0 (TID 45) in 1842 ms on 192.168.10.125 (executor 4) (97/100)
18/07/26 16:50:46 INFO TaskSetManager: Finished task 43.0 in stage 0.0 (TID 43) in 1844 ms on 192.168.10.125 (executor 4) (98/100)
```

```
18/07/26 16:50:46 INFO TaskSetManager: Finished task 44.0 in stage
0.0 (TID 44) in 1844 ms on 192.168.10.125 (executor 4) (99/100)
18/07/26 16:50:46 INFO TaskSetManager: Finished task 47.0 in stage
0.0 (TID 47) in 1840 ms on 192.168.10.125 (executor 4) (100/100)
18/07/26 16:50:46 INFO DAGScheduler: ResultStage 0 (reduce at
/ubda/bdaapps/spark/examples/src/main/python/pi.py:44) finished in
7.036 s
18/07/26 16:50:46 INFO TaskSchedulerImpl: Removed TaskSet 0.0,
whose tasks have all completed, from pool default
18/07/26 16:50:46 INFO DAGScheduler: Job 0 finished: reduce at
/ubda/bdaapps/spark/examples/src/main/python/pi.py:44, took
7.096171 s
18/07/26 16:50:46 INFO AbstractConnector: Stopped
Spark@8f1ea4f{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
18/07/26 16:50:46 INFO SparkUI: Stopped Spark web UI at
http://ubda-mgt01:4040
18/07/26 16:50:46 INFO StandaloneSchedulerBackend: Shutting down
all executors
18/07/26 16:50:46 INFO
CoarseGrainedSchedulerBackend$DriverEndpoint: Asking each executor
to shut down
18/07/26 16:50:46 INFO MapOutputTrackerMasterEndpoint:
MapOutputTrackerMasterEndpoint stopped!
18/07/26 16:50:46 INFO MemoryStore: MemoryStore cleared
18/07/26 16:50:46 INFO BlockManager: BlockManager stopped
18/07/26 16:50:46 INFO BlockManagerMaster: BlockManagerMaster
stopped
18/07/26 16:50:46 INFO
OutputCommitCoordinator$OutputCommitCoordinatorEndpoint:
OutputCommitCoordinator stopped!
18/07/26 16:50:46 INFO SparkContext: Successfully stopped
SparkContext
18/07/26 16:50:47 INFO ShutdownHookManager: Shutdown hook called
18/07/26 16:50:47 INFO ShutdownHookManager: Deleting directory
/tmp/spark-31260ba9-0d95-4e12-a75c-2a2385fd262f
18/07/26 16:50:47 INFO ShutdownHookManager: Deleting directory
/ubda/bdadata/spark/local/spark-482d7216-4781-4f24-a066-
2a52e7dd1a87
18/07/26 16:50:47 INFO ShutdownHookManager: Deleting directory
/ubda/bdadata/spark/local/spark-482d7216-4781-4f24-a066-
2a52e7dd1a87/pyspark-b10448b4-e97c-4c87-898d-cd3d75e7574c
```

3 How to check the result

4.1 Check result of PI test

```
$ cat result.out  
Pi is roughly 3.140591
```

4 Useful Reference

- Apach Spark
URL: <https://spark.apache.org/>