



# UBDA Platform

*Example for Hadoop on BeeGFS*

## 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. Perform the PI test.....      | 6 |
| 4. How to check the result ..... | 8 |
| 5. Useful Reference.....         | 9 |

## 1. Introduction

This document is shown a Hadoop example by using BGFS 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_bgfs.sh
```

### 2.3 Create new directory

```
$ mkdir $HOME/hadoop_bgfs  
Go to the directory  
$ cd $HOME/hadoop_bgfs
```

### 2.4 Copy the example jar file

```
$ cp /ubda/bdaapps/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.8.3.jar .
```

## 3 Perform the PI test

### 3.1 Apply Hadoop on HDFS environment

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

### 3.2 Submit the PI to job queue. `hadoop-mapreduce-examples-2.8.3.jar pi` [genericOptions] [commandOptions]

```
$ hadoop jar hadoop-mapreduce-examples-2.8.3.jar pi 10 100 > result.out
```

### 3.3 Process of the PI job.

```
18/08/02 15:39:33 INFO input.FileInputFormat: Total input files to
process : 10
18/08/02 15:39:33 INFO mapreduce.JobSubmitter: number of splits:10
18/08/02 15:39:33 INFO mapreduce.JobSubmitter: Submitting tokens for
job: job_1533184810974_0003
18/08/02 15:39:33 INFO impl.YarnClientImpl: Submitted application
application_1533184810974_0003
18/08/02 15:39:33 INFO mapreduce.Job: The url to track the job:
http://ubda-stg01.hh.polyu.hk:8888/proxy/application_1533184810974_0003/
18/08/02 15:39:33 INFO mapreduce.Job: Running job:
job_1533184810974_0003
18/08/02 15:39:38 INFO mapreduce.Job: Job job_1533184810974_0003 running
in uber mode : false
18/08/02 15:39:38 INFO mapreduce.Job:  map 0% reduce 0%
18/08/02 15:39:42 INFO mapreduce.Job:  map 10% reduce 0%
18/08/02 15:39:43 INFO mapreduce.Job:  map 100% reduce 0%
18/08/02 15:39:46 INFO mapreduce.Job:  map 100% reduce 100%
18/08/02 15:39:46 INFO mapreduce.Job: Job job_1533184810974_0003
completed successfully
18/08/02 15:39:46 INFO mapreduce.Job: Counters: 49
    File System Counters
        FILE: Number of bytes read=10973
        FILE: Number of bytes written=1828493
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        VIEWFS: Number of bytes read=0
        VIEWFS: Number of bytes written=0
        VIEWFS: Number of read operations=0
        VIEWFS: Number of large read operations=0
        VIEWFS: Number of write operations=0
    Job Counters
        Launched map tasks=10
        Launched reduce tasks=1
        Rack-local map tasks=10
```

```
Total time spent by all maps in occupied slots
(ms)=16461
Total time spent by all reduces in occupied slots
(ms)=2774
Total time spent by all map tasks (ms)=16461
Total time spent by all reduce tasks (ms)=1387
Total vcore-milliseconds taken by all map tasks=16461
Total vcore-milliseconds taken by all reduce tasks=5548
Total megabyte-milliseconds taken by all map
tasks=168560640
Total megabyte-milliseconds taken by all reduce
tasks=28405760
  Map-Reduce Framework
    Map input records=10
    Map output records=20
    Map output bytes=180
    Map output materialized bytes=340
    Input split bytes=1340
    Combine input records=0
    Combine output records=0
    Reduce input groups=2
    Reduce shuffle bytes=340
    Reduce input records=20
    Reduce output records=0
    Spilled Records=40
    Shuffled Maps =10
    Failed Shuffles=0
    Merged Map outputs=10
    GC time elapsed (ms)=62
    CPU time spent (ms)=3210
    Physical memory (bytes) snapshot=4506804224
    Virtual memory (bytes) snapshot=128538222592
    Total committed heap usage (bytes)=21962424320
  Shuffle Errors
    BAD_ID=0
    CONNECTION=0
    IO_ERROR=0
    WRONG_LENGTH=0
    WRONG_MAP=0
    WRONG_REDUCE=0
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=0
```

## 4 How to check the result

### 4.1 Check the result.out file

```
$ cat result.out
Number of Maps = 10
Samples per Map = 100
Wrote input for Map #0
Wrote input for Map #1
Wrote input for Map #2
Wrote input for Map #3
Wrote input for Map #4
Wrote input for Map #5
Wrote input for Map #6
Wrote input for Map #7
Wrote input for Map #8
Wrote input for Map #9
Starting Job
Job Finished in 13.805 seconds
Estimated value of Pi is 3.14800000000000000000
```



## 5 Useful Reference

- Hadoop commands Guide  
URL: <https://hadoop.apache.org/docs/r2.8.3/hadoop-project-dist/hadoop-common/CommandsManual.html>