



YCBS-257 - Data at Scale

Hadoop Yarn Workshop

General Instructions:

The purpose of this workshop is to get you started with Hadoop Yarn. Here you will learn how to run Hadoop applications, monitor and kill the running Hadoop applications.

Start your Cloudera QuickStart VM in order to complete this workshop.

Online resources:

<https://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/YARN.html>

Exercise 1: Yarn basics operations

In this exercise you will three or more instances of the Hadoop application “pi”. This application is an example provided with the Hadoop package and it aims to calculate the value of pi using a quasi-Monte Carlo method and MapReduce.

To run the pi QuasiMonteCarlo calculator you should provide two parameters:

<nMaps> mappers numbers (how many mappers)
<nSamples> samples counts (how many points)

You will run the pi calculator using the following values: **10** mappers and **50000** samples

1. Open a Terminal window and launch the MapReduce pi application using the following command: (without the \$ sign)

```
$ hadoop jar /usr/lib/hadoop-0.20-mapreduce/hadoop-examples-2.6.0-mr1-cdh5.13.0.jar pi 5 50000
```

2. Repeat the previous step by opening 3 new Terminal windows simultaneously and run a new pi calculator in each.

3. Open a new Terminal window and show the applications list of using yarn client

```
$ yarn application --list
```

4. Select the last **application_id** in the list and view its status

```
$ yarn application --status application_*****
```

5. Use the **application_id** to kill this task (you may need to restart pi calculator in a Terminal window)

```
$ yarn application --kill application_*****
```

6. View the entire list of applications (running, completed, killed)

```
$ yarn application --list --appStates ALL
```

7. View the list of killed applications only

```
$ yarn application --list --appStates KILLED
```