

coreservlets.com – Hadoop Course

VM Overview

In this exercise, you will get acquainted with Virtual Box Virtual Machine. You will also build Exercise code and run your first MapReduce job.

Approx Time: 30 minutes

Perform

1. Start Virtual Machine
2. Open Eclipse
3. Open Command Line Terminal
4. Start ALL Hadoop installed products
5. Run mvn package for Exercises from the Eclipse
6. Run the job from the lecture
7. Stop ALL Hadoop installed products
8. Verify there are no leftover Java rogue processes

Solution

1. Open Virtual Box, select training VM and click start
2. Within VM double click on Eclipse icon
3. Within VM double click on Terminal icon
4. In the Command Line Terminal type `startCDH.sh`
5. In Eclipse locate 'run' icon, click on it then select 'mvn package'
6. In the Command Line Terminal type:

```
$ cd $PLAY_AREA
```

```
$ yarn jar $PLAY_AREA/Exercises.jar \  
    mapRed.workflows.CountDistinctTokens \  
    /training/data/hamlet.txt \  
    /training/playArea/firstJob
```

```
$ hdfs dfs -rm -r /training/playArea/firstJob
```

7. In the Command Line Terminal type `stopCDH.sh`
8. Check to see if there are any java processes left:

```
$ ps -ef | grep java
```

If there are still processes, then kill them manually:

```
$ kill XXXX
```