**RESPONSI**

**INFRASTRUKTUR BIG DATA**

**“MEMBANGUN INFRASTRUKTUR APACHE SPARK”**

****

**Disusun Oleh :**

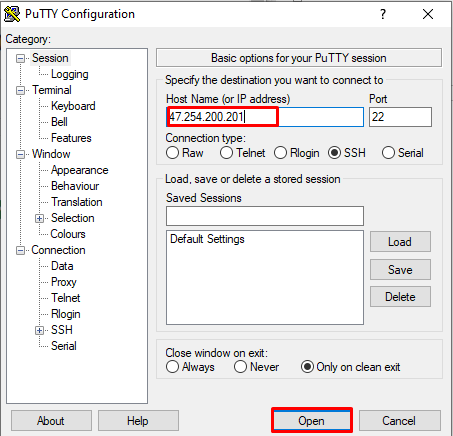
|  |  |
| --- | --- |
| **NAMA** | **: NURUL LESTARI** |
| **NIM** | **: 185410029** |
| **JURUSAN** | **: Informatika** |
| **JENJANG** | **: S1** |

**STMIK AKAKOM**

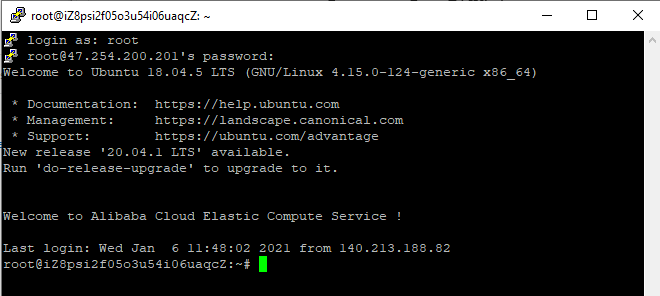
**YOGYAKARTA**

**2020**

Penggunaan Putty Di Windows

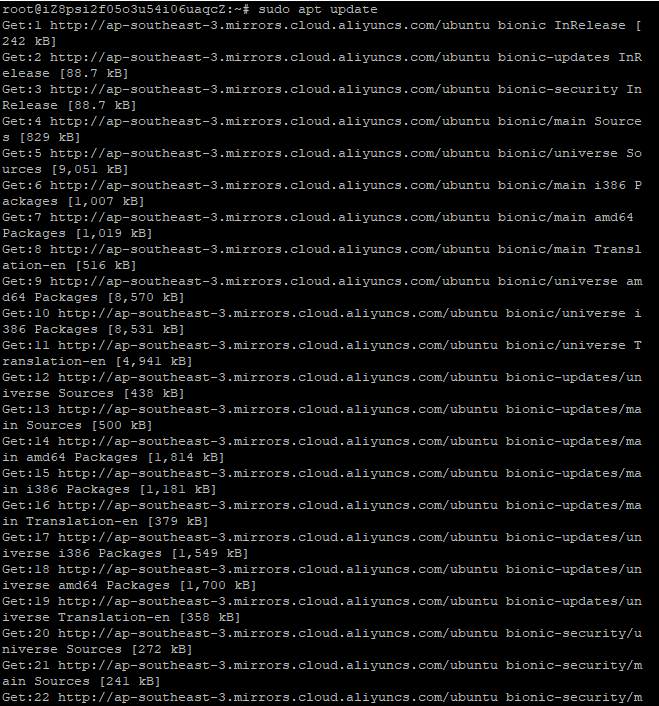


Setelah anda klik open, anda akan diminta memasukkan password. Masukkan password sesuai dengan informasti yang disediakan oleh tabel. Setelah anda enter, akan muncul shell/window berikut :

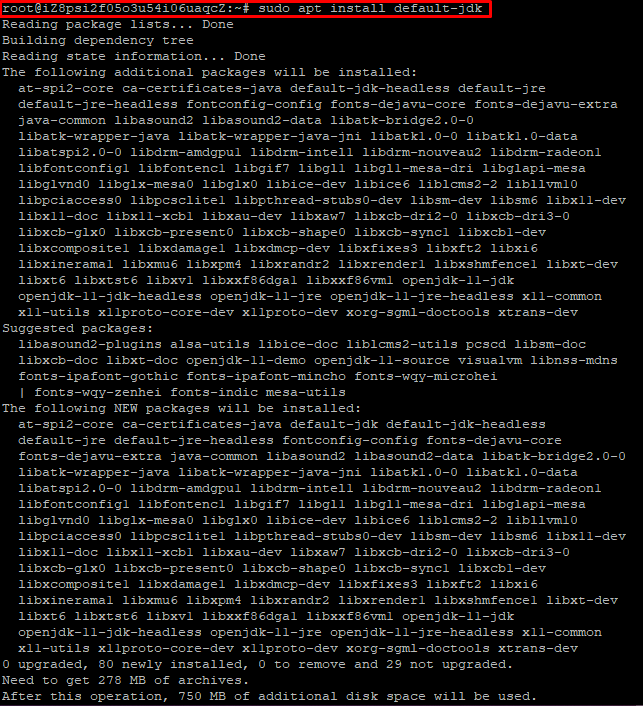


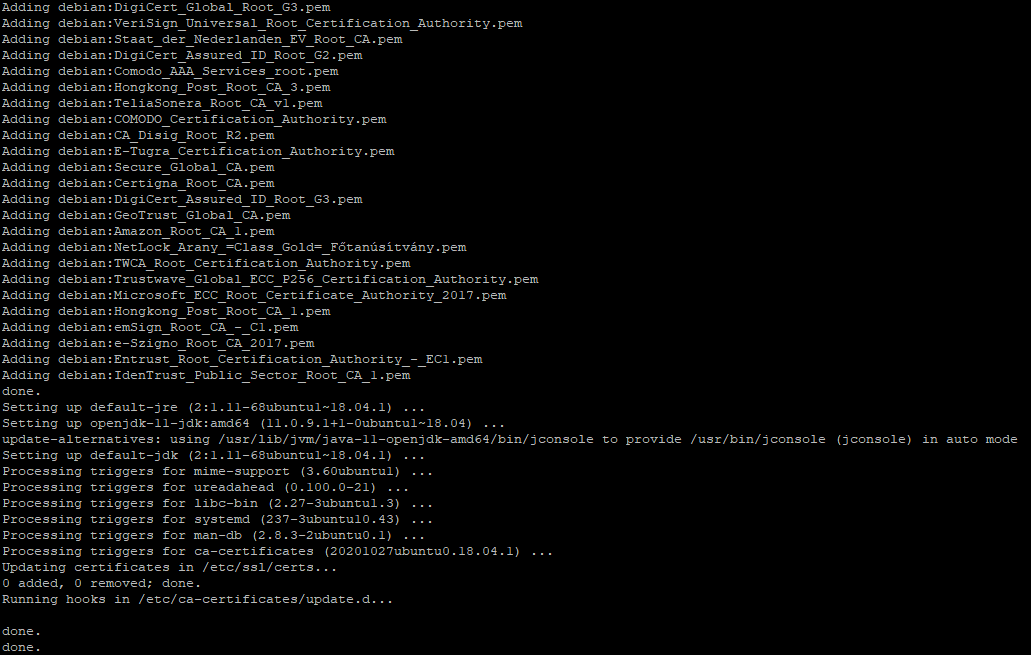
Langkah pengerjaan :

1. **sudo apt update**

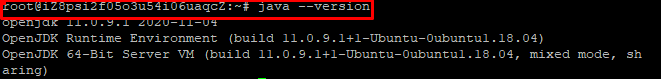


1. **sudo apt install default-jdk**

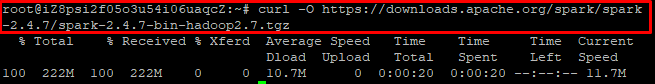




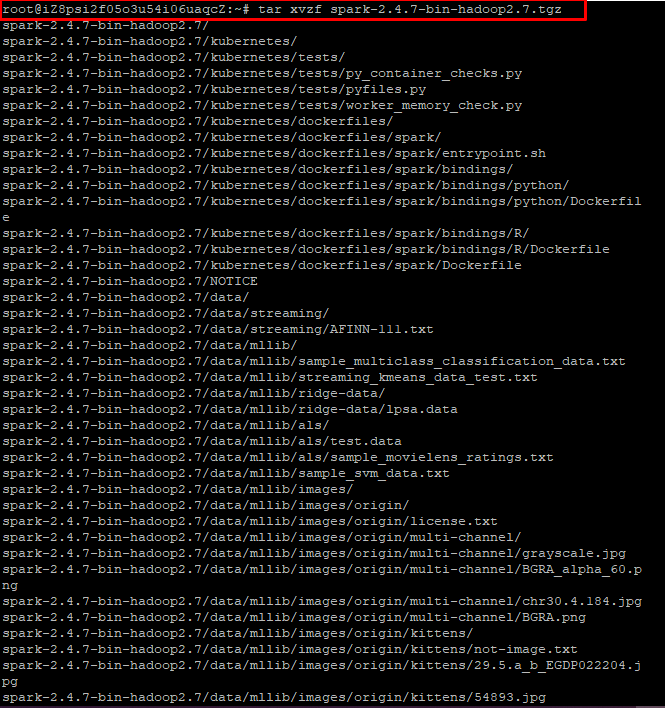
1. **java --version**



1. **curl -O** [**https://downloads.apache.org/spark/spark-2.4.7/spark-2.4.7-bin-hadoop2.7.tgz**](https://downloads.apache.org/spark/spark-2.4.7/spark-2.4.7-bin-hadoop2.7.tgz)



1. **tar xvzf spark-2.4.7-bin-hadoop2.7.tgz**

****

1. **sudo mv spark-2.4.7-bin-hadoop2.7 /opt/spark**



1. **cd /opt/spark/**



1. **sudo sbin/start-master.sh**

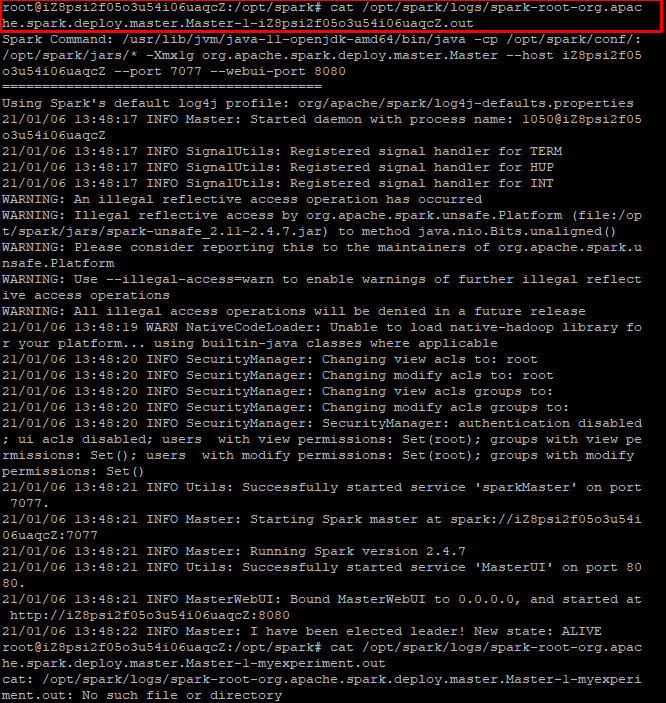


1. **ss -tunelp |grep 8080**

Perintah ini untuk mengetahui proses yang berjalan dan lokasi spark yang dapat di akses melalui web. Selain itu juga untuk mengetahui url spark untuk digunakan menjalankan perintah slave nantinya.



1. **cat /opt/spark/logs/spark-root-org.apache.spark.deploy.master.Master-1-myexperiment.out**

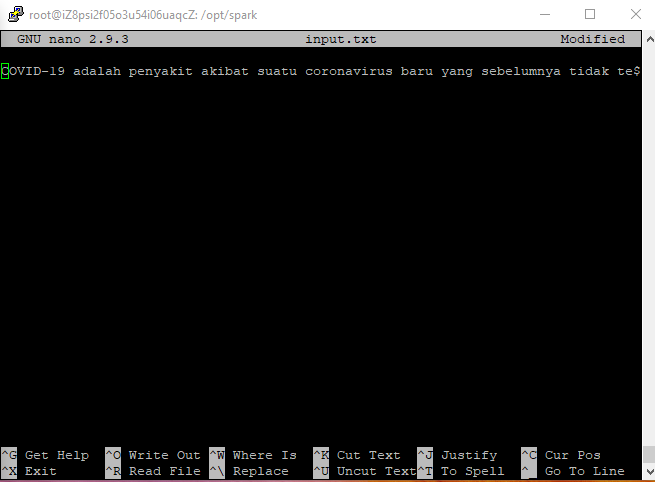


1. **sudo sbin/start-slave.sh spark://sparktesting.dmxonqygcr0ehciiwimzfeacqd.bx.internal.cloudapp.net:7077**



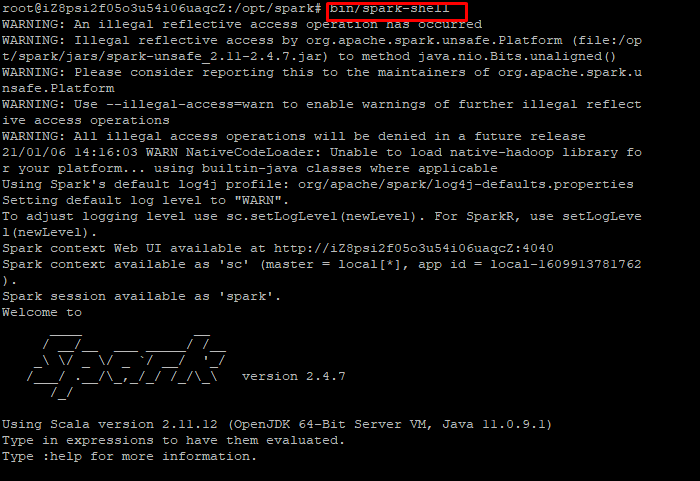
1. **pico input.txt**



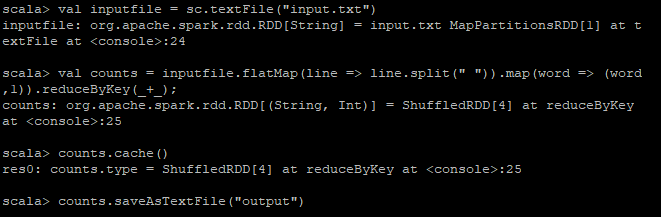
****

Setelah perintah diatas dijalankan anda akan masuk dalam console scala. Untuk membuat RDD sederhana jalankan perintah berikut :

* **bin/spark-shell**



* **scala> val inputfile = sc.textFile(“input.txt”)**
* **scala> val counts = inputfile.flatMap(line => line.split(" ")).map(word => (word, 1)).reduceByKey(\_+\_);**
* **scala> counts.cache()**
* **scala> counts.saveAsTextFile("output")**



Seluruh hasil akan tersimpan dalam direktori output. Cobalah masuk ke dalam direktori tersebut.

Untuk keluar dari shell scala, tekan CTRL C.

Kemudian kita lakukan check hasil dari proses apache spark diatas:

1. **cd output/**
2. **cat part-00000**

