

ini adalah perintah di shell dan dapat anda jalankan. Jangan sertakan tanda $ pada setiap perintah yang kamu tuliskan di VM nanti, karena tanda digunakan sebagai penanda shell saja.

$ sudo apt update

$ sudo apt install default-jdk

$ java --version

$ curl -O <https://downloads.apache.org/spark/spark-2.4.7/spark-2.4.7-bin-hadoop2.7.tgz>

$ tar xvzf spark-2.4.7-bin-hadoop2.7.tgz

$ sudo mv spark-2.4.7-bin-hadoop2.7 /opt/spark

$ cd /opt/spark/

Menjalankan Spark:

$ sudo sbin/start-master.sh

**Contoh output:**

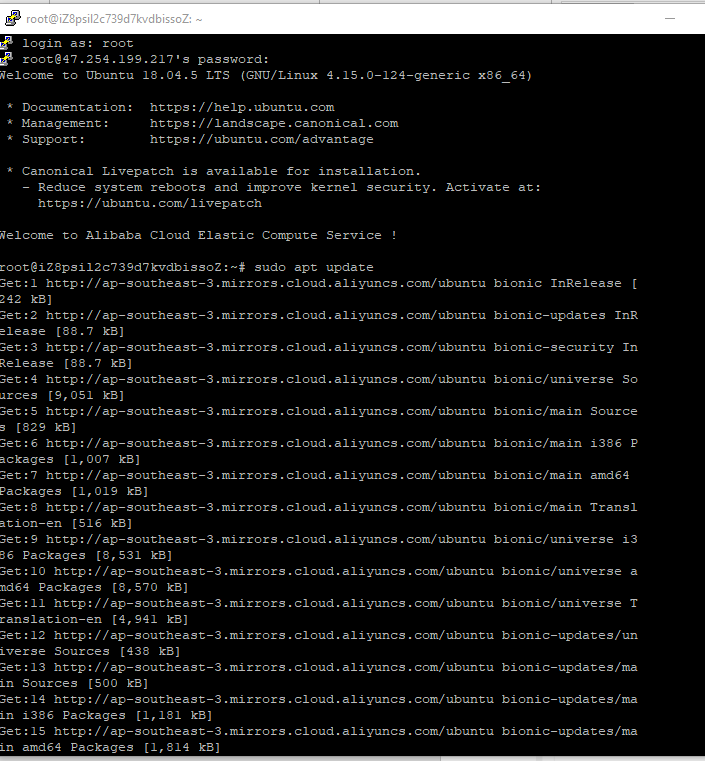
starting org.apache.spark.deploy.master.Master, logging to /opt/spark/logs/spark-root-org.apache.spark.deploy.master.Master-1-myexperiment.out

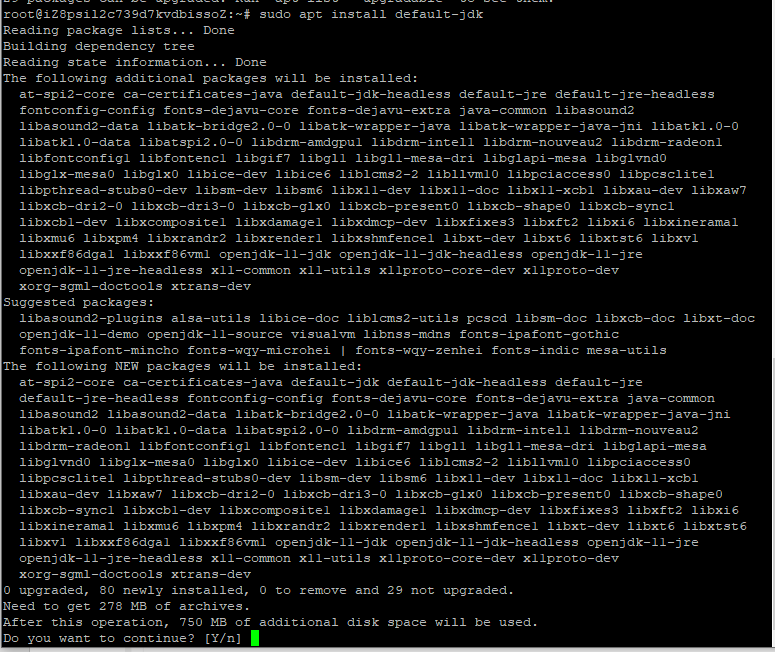
Mengindentifikasi spark berjalan:

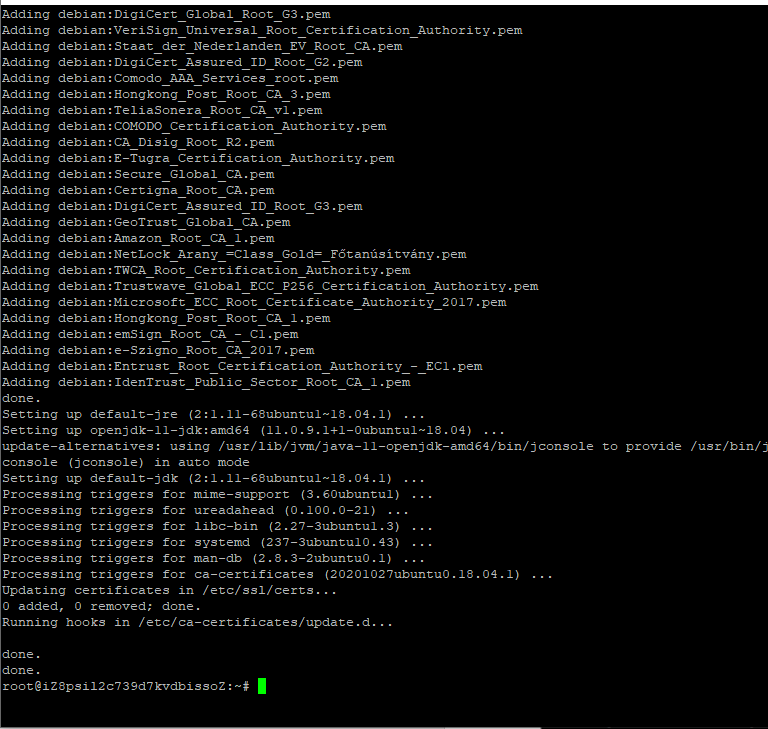
$ ss -tunelp |grep 8080

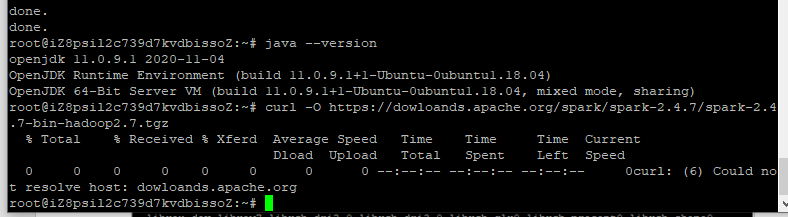
Perintah ini untuk mengetahui proses yang berjalan dan lokasi spark yang dapat di akses melalui web. Selain itu juga untuk mengetahui url spark untuk digunakan menjalankan perintah slave nantinya.

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









Ketik program ini untuk menulis text pada $ sudo sbin/start-slave.sh spark://sparktesting.dmxonqygcr0ehciiwimzfeacqd.bx.internal.cloudapp.net:7077

Gunakan dengan perintah:

$ pico input.txt

Setelah itu copy dan paste tulisan diatas, kemudian save dengan tekan CTRL X , lalu tekan Y, kemudian tekan Enter.

Setelah perintah diatas dijalankan anda akan masuk dalam console scala. Untuk membuat RDD sederhana jalankan perintah berikut (**mohon untuk di tulis, bukan copy paste, dan jangan sertakan “scala>” pada perintah yang anda input**):

$ bin/spark-shell

scala> val inputfile = sc.textFile(“input.txt”)

scala> val counts = inputfile.flatMap(line => line.split(" ")).map(word => (word, 1)).reduceByKey(\_+\_);

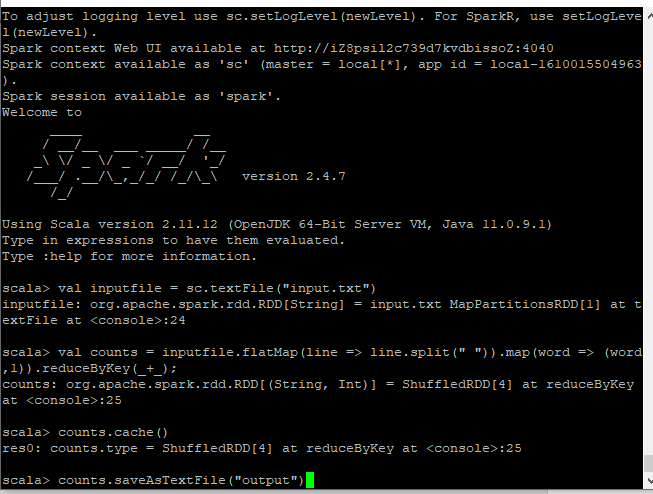
scala> counts.cache()

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:



$ cd output/

$ cat part-00000

Lalu ketik perintah seperti di atas i dan hasilnya ada pada gambar dibawah ini :

