**TUGAS**

**KOMPUTASI TERSEBAR DAN PARALEL**

****

Disusun Oleh:

Jonathan Henock Alexander M 24060120130112

Najib Rifai Indrayanto 24060120140082

Fayza Aulia 24060120120010

**DEPARTEMEN ILMU KOMPUTER / INFORMATIKA**

**FAKULTAS SAINS DAN MATEMATIKA**

**UNIVERSITAS DIPONEGORO**

**SEMARANG**

**2022**

**BAB I**

**PENDAHULUAN**

**1.1. Tujuan**

Mengetahui cara kerja distribusi file system menggunakan Hadoop Distributed File System dengan melakukan eksekusi WordCount pada sebuah dataset berekstensi txt .

**1.2 Rumusan Masalah**

1. Penjelasan langkah-langkah membuat cluster Hadoop
2. Penjelasan dataset dan program
3. Penjelasan cara eksekusi WordCount
4. Hasil Eksekusi
5. Link Video singkat demo

**BAB II**

**DASAR TEORI**

**2.1 HDFS**

Hadoop merupakan framework distributed processing untuk data set yang sangat besar pada computer cluster, menggunakan model pemrograman sederhana (MapReduce).

**2.2 Single Node Cluster**

Single Node Cluster adalah cluster di mana semua daemon seperti name node , node data, Jobtracker, dan tasktracker berjalan pada satu mesin. Faktor replikasi default adalah 1.

**2.3 Multi Node Cluster**

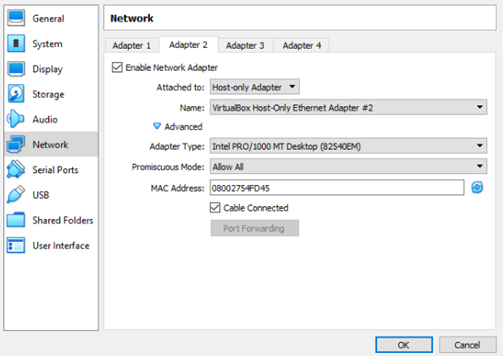
Multi node cluster adalah cluster yang pada dasarnya digunakan dalam mode master-slave dimana master dan slave berjalan pada mesin yang berbeda dan master node/machine menjalankan daemon Namenode dan TaskTracker dan mesin slave menjalankan daemon Datanode dan JobTracker.

**BAB III**

**PEMBAHASAN**

**3.1 Langkah-langkah pembuatan cluster Hadoop**

1. Buka Pengaturan Jaringan Mesin Virtual yang utama yaitu bernama “master” dan aktifkan Adaptor 2.
2. Kemudian, pada NAT, pilih Adaptor **Host-only Adapter** dan di mana dikatakan **"Mode Promiscuous"** pilih opsi **"Izinkan Semua".**



1. Instal SSH menggunakan perintah berikut:



1. Instal PDSH menggunakan perintah berikut:

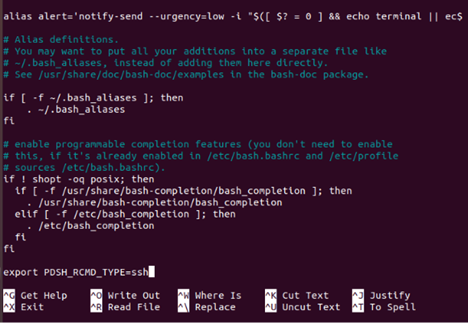


1. Buka file .bashrc dengan perintah berikut:





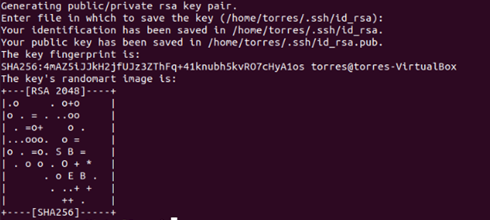
Di akhir file cukup tulis baris berikut:



1. Lalu, konfigurasikan SSH. Buat kunci baru dengan perintah berikut:



Tekan Enter setiap kali diperlukan.



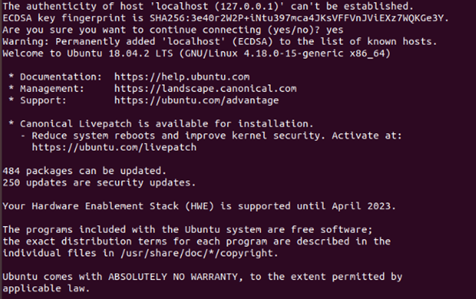
1. Salin atau copy kunci publik ke file authorized\_keys dengan perintah berikut:



1. Sekarang kita dapat memverifikasi konfigurasi SSH dengan menghubungkan ke localhost:



Ketik "yes" dan tekan Enter bila diperlukan.



1. Lakukan instalasi Java 8. menggunakan perintah berikut:



1. Unduh Hadoop :



1. Lakukan unzip file hadoop-3.3.4.tar.gz dengan perintah berikut:





1. Ubah nama folder hadoop-3.3.4 menjadi hadoop agar lebih simple dengan menggunakan perintah berikut:



1. Buka file hadoop-env.sh di editor nano untuk mengedit Java\_HOME:

export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64/

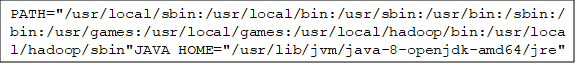
1. Ubah direktori folder hadoop ke /usr/local/hadoop.

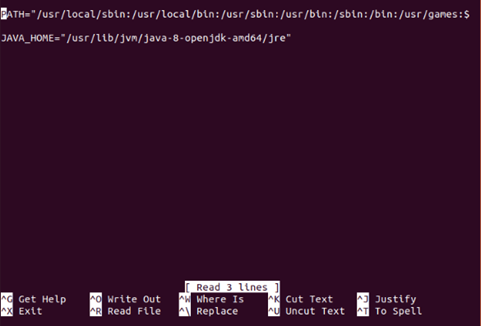


1. Buka file environment dengan nano



Kemudian tambahkan konfigurasi seperti berikut:

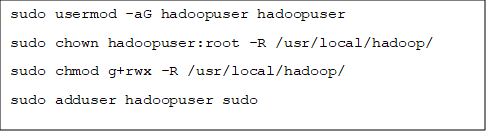




1. Tambahkan pengguna bernama **“hadoopuser”**, dan lalu atur konfigurasinya.



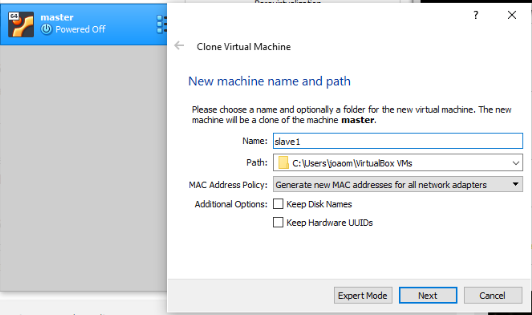
Sekarang tuliskan perintah berikut :



1. Kemudian, lanjut ke pembuatan slave.

Matikan Mesin Virtual “master” dan lakukan “clone” 2 kali, beri nama satu slave1 dan slave2 Lainnya.

Pastikan opsi "Generate new MAC addresses for all network adapters" dipilih serta pilih Full Clone.

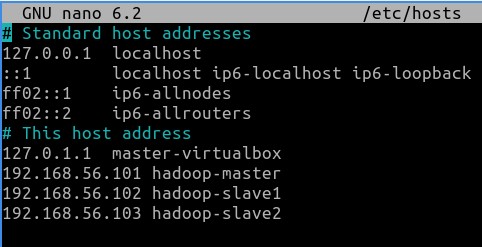


1. Pada master VM, buka file hostname di nano:



Masukkan nama mesin virtual master. (perhatikan bahwa nama yang sama yang dimasukkan sebelumnya pada file host) . Lakukan hal yang sama pada semua slave. Kemudian, reboot ketiga Virtual Machine agar konfigurasi berjalan .

1. Lakukan verifikasi alamat ip setiap mesin:



1. Buka file host dan masukkan konfigurasi Jaringan Anda:



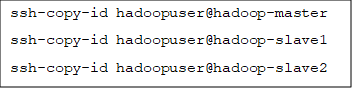


1. Konfigurasikan SSH di hadoop-master, dengan hadoopuser.

Buat sebuah SSH key:



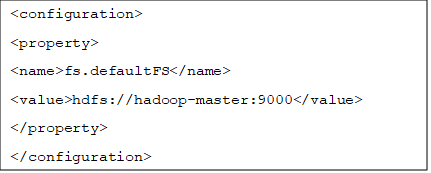
1. Sekarang perlu menyalin kunci SSH ke semua pengguna. Gunakan perintah ini:





1. Di hadoop-master, buka file core-site.xml dengan nano:

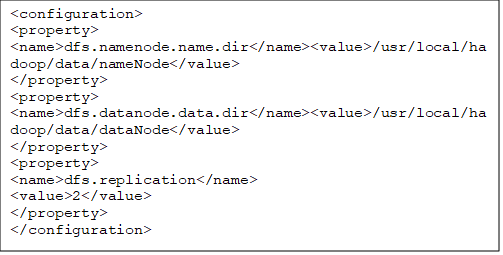
Kemudian tambahkan konfigurasi berikut:



1. Masih di hadoop-master, buka file hdfs-site.xml.



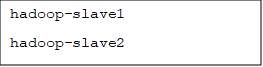
Tambahkan konfigurasi berikut:





1. Masih di hadoop-master, mari kita buka file workers:

Tambahkan dua baris ini: (nama slave)



1. Salin konfigurasi hadoop master ke slave, dengan perintah berikut:

scp /usr/local/hadoop/etc/hadoop/\* hadoop-slave1:/usr/local/hadoop/etc/hadoop/

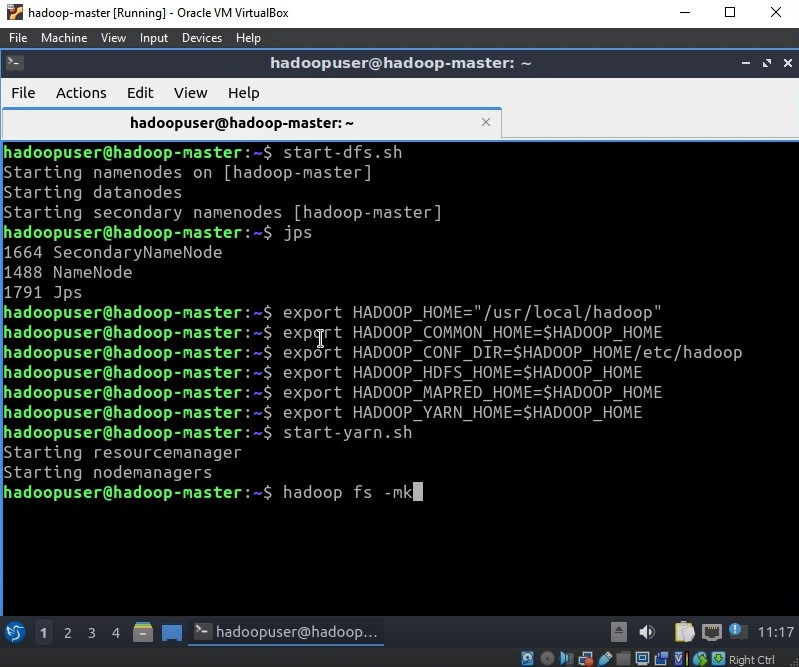
scp /usr/local/hadoop/etc/hadoop/\* hadoop-slave2:/usr/local/hadoop/etc/hadoop/

(Copy informasi ke slave 1 dan slave 2 )

1. Format sistem file HDFS dengan menjalankan perintah ini:



1. Start HDFS dengan command:

start-dfs.sh

1. Kemudian lakukan pengecekan apakah HDFS sudah berjalan jalankan perintah:

jps

1. Lakukan hal yang sama pada slave

**3.2 Penjelasan dataset dan program**

Dalam penugasan ini kami memakai dataset mengenai ulasan-ulasan yang diberikan kepada sebuah produk oleh pengguna di website e-commerce amazon. Dataset tersebut akan dipakai untuk implementasi pencarian kata menggunakan program WordCount yang disediakan oleh hadoop. Program WordCount hadoop berfungsi mencari dan mencatat kemunculan semua kata yang terdapat dalam dataset. Program WordCount tersendiri adalah program bawaan dari hadoop, oleh karena itu kami hanya butuh memanggil program WordCount untuk melakukan pencarian kata. Program WordCount terdiri dari 2 proses utama yaitu map and reduce. Map adalah proses *extract* file yang diberikan dan reduce adalah proses menggabungkan, meringkas, memfilter, atau mengubah dan menulis hasilnya. Berikut adalah data flow dari proses WordCount:

1. Mappers membaca data dari HDFS
2. Output Map akan diparitisi oleh key dan dikirim ke Reducers
3. Reducers akan mengurutkan input berdasarkan key
4. Output Reduce akan ditulis dan dikirimkan kepada HDFS

**3.3 Cara Eksekusi Word Count**

Data set yang digunakan diambil dari website<https://www.kaggle.com/datasets/bittlingmayer/amazonreviews>

Open your browser and type hadoop-master:9870.

* + - 1. Jalankan Hadoop

start-dfs.sh

start-yarn.sh

* + - 1. Buat folder input

hadoop

fs -mkdir /input1

mkdir

* + - 1. Download file dataset ke computer

local.

* + - 1. Pindah file dataset dari komputer lokal ke folder input pada HDFS

hadoop fs -put train.txt /input1 input

1. Jalankan contoh program Wordcount pada

hadoop-mapreduce-examples-3.3.1.jar untuk menghitung jumlah kata pada file train.txt dan hasilnya disimpan pada folder output1

hadoopjar/usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.1.jar

wordcount /input1/train.txt /output1

run

1. Cek di folder output1

hadoop

fs -ls /output1

output

Pada folder output1 terdapat dua file seperti gambar di

atas. Hasil perhitungan bisa dilihat pada file yang nama depannya

part, pada kasus ini part-r-00000.

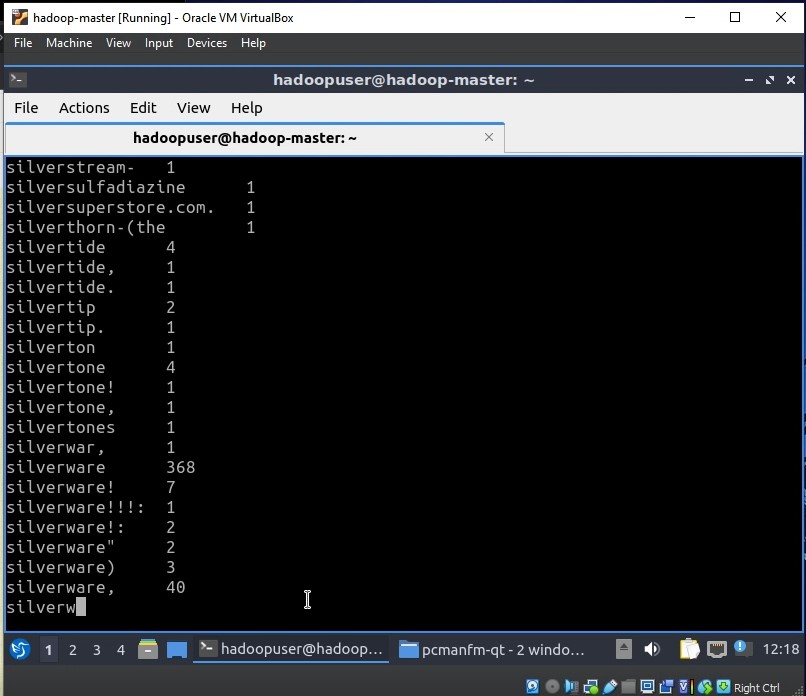
1. Melihat hasil perhitungan Wordcount di file

part-r-00000

hadoop

fs -cat /output1/part-r-00000

**3.4 Hasil Eksekusi**

****

**3.5 Link Video Singkat demo**

[**https://drive.google.com/drive/folders/1ObK2HLIOcrQRjtANEf8r9DXtVXkLhLTu?usp=sharing**](https://drive.google.com/drive/folders/1ObK2HLIOcrQRjtANEf8r9DXtVXkLhLTu?usp=sharing)

**BAB IV**

**Kesimpulan**

Dari hasil implementasi hadoop yang kami lakukan dapat disimpulkan bahwa penggunaan multi-node cluster sangat membantu dalam pencarian data dengan jumlah yang banyak.