Komputasi Tersebar dan Paralel

Kelas C

Tugas Hadoop

Anggota Kelompok:

1. Fellia Gessangie Yohanshah Puteri (24060120130047)
2. Salma Nora Renada (24060120130050)
3. Taskia Rorointan Sugyanti (24060120140152)

DEPARTEMEN INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
2022

Link Video Presentasi
https://drive.google.com/file/d/1Wlcm2BoIxXr0Ntw2MtxrkHtmlPkrVcq8/view?usp=sharing

Install Hadoop

1. Siapkan virtual machine untuk Ubuntu, di sini kami menggunakan Oracle VM Virtual Box untuk instalasi dan mengerjakan tugas hadoop ini. Setelah itu buka pada terminal/cli ubuntu pada virtual machine tersebut.
2. Lakukan update ubuntu terlebih dahulu, untuk memberi tahu apt untuk mencari software repository dan mengupdate jika ada package ubuntu yang tersedia. Dengan menggunakan perintah `sudo apt update`.

```
renada@renada-VirtualBox:~$ sudo apt update
[sudo] password for renada:
Get:1 http://security.ubuntu.com/ubuntu focal-security InRelease [114 kB]
Hit:2 http://id.archive.ubuntu.com/ubuntu focal InRelease
Hit:3 http://id.archive.ubuntu.com/ubuntu focal-updates InRelease
Hit:4 http://id.archive.ubuntu.com/ubuntu focal-backports InRelease
Fetched 114 kB in 3s (42,6 kB/s)
Reading package lists... Done
Building dependency tree
Reading state information... Done
387 packages can be upgraded. Run 'apt list --upgradable' to see them.
```

3. Lalu install openjdk, yang merupakan implementasi opensource dari Java. Dengan perintah `sudo apt install openjdk-8-jdk -y`.

```
renada@renada-VirtualBox:~$ sudo apt install openjdk-8-jdk -y
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java
  libatk-wrapper-java-jni libice-dev libpthread-stubs0-dev libsm-dev
  libx11-dev libxau-dev libxcb1-dev libxdmcp-dev libxt-dev
  openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless
  x11proto-core-dev x11proto-dev xorg-sgml-doctools xtrans-dev
Suggested packages:
  default-jre libice-doc libsm-doc libx11-doc libxcb-doc libxt-doc
  openjdk-8-demo openjdk-8-source visualvm fonts-ipafont-gothic
  fonts-ipafont-mincho fonts-wqy-microhei fonts-wqy-zenhei
The following NEW packages will be installed:
  ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java
  libatk-wrapper-java-jni libice-dev libpthread-stubs0-dev libsm-dev
  libx11-dev libxau-dev libxcb1-dev libxdmcp-dev libxt-dev openjdk-8-jdk
  openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless
  x11proto-core-dev x11proto-dev xorg-sgml-doctools xtrans-dev
0 upgraded, 21 newly installed, 0 to remove and 387 not upgraded.
Need to get 44,6 MB of archives.
After this operation, 163 MB of additional disk space will be used.
```

4. Lalu menggunakan perintah `java -version;javac -version`, untuk mengetahui versi dari java yang di install.

```
renada@renada-VirtualBox:~$ java -version;javac -version
openjdk version "1.8.0_342"
OpenJDK Runtime Environment (build 1.8.0_342-8u342-b07-0ubuntu1~20.04-b07)
OpenJDK 64-Bit Server VM (build 25.342-b07, mixed mode)
javac 1.8.0_342
```

5. Untuk melihat di mana java disimpan, dapat menggunakan perintah `which javac`.

```
renada@renada-VirtualBox:~$ which javac
/usr/bin/javac
```

6. Menggunakan perintah readlink - f(lokasi java), di sini berarti kami menggunakan perintah readlink -f /usr/bin/javac.

```
renada@renada-VirtualBox:~$ readlink -f /usr/bin/javac
/usr/lib/jvm/java-8-openjdk-amd64/bin/javac
```

7. Buka file .bashrc dengan menggunakan perintah `sudo nano .bashrc`.

```
renada@renada-VirtualBox:~$ sudo nano .bashrc
```

8. Tambahkan JAVA_HOME path ke dalam .bashrc, dengan perintah `export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64`.

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

9. Lakukan instalasi OpenSSH di Ubuntu dengan perintah `sudo apt install openssh-server openssh-client-y`.

```
renada@renada-VirtualBox:~$ sudo apt install openssh-server openssh-client -y
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
  ncurses-term openssh-sftp-server ssh-import-id
Suggested packages:
  keychain libpam-ssh monkeysphere ssh-askpass molly-guard
The following NEW packages will be installed:
  ncurses-term openssh-server openssh-sftp-server ssh-import-id
The following packages will be upgraded:
  openssh-client
1 upgraded, 4 newly installed, 0 to remove and 386 not upgraded.
Need to get 1.359 kB of archives.
After this operation, 6.010 kB of additional disk space will be used.
Get:1 http://id.archive.ubuntu.com/ubuntu focal-updates/main amd64 openssh-clie
nt amd64 1:8.2p1-4ubuntu0.5 [671 kB]
Get:2 http://id.archive.ubuntu.com/ubuntu focal/main amd64 ncurses-term all 6.2
-0ubuntu2 [249 kB]
Get:3 http://id.archive.ubuntu.com/ubuntu focal-updates/main amd64 openssh-sftp
-server amd64 1:8.2p1-4ubuntu0.5 [51.5 kB]
```

10. Aktifkan paswordless SSH untuk User Hadoop untuk mempermudah, dengan perintah `ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa`.

```
renada@renada-VirtualBox:~$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
Your identification has been saved in /home/renada/.ssh/id_rsa
Your public key has been saved in /home/renada/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:VLyulLn8xH/TgcvyrbU4R3DhAP3UPzw3/BT9mJr/kXE renada@renada-VirtualBox
The key's randomart image is:
+---[RSA 3072]----+
|          ...o   o|
|          ..  o +o|
|         .  .  Bo=|
|        .  .  .oX=|
|         S+   o+oE|
|         +.. o. o=|
|        o oo ..o=.|
|         +. o +B.+|
|          .. +=+=.|
+----[SHA256]-----+
```

11. Menggunakan perintah cat untuk menyimpan public key seperti authorized_keys di direktori ssh.

```
renada@renada-VirtualBox:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

12. Atur permission untuk user dengan perintah chmod.

```
renada@renada-VirtualBox:~$ chmod 0600 ~/.ssh/authorized_keys
```

13. SSH ke localhost.

```
renada@renada-VirtualBox:~$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:RmECuThS2TAoYNWQmy3X0flILMil1DE2AS0W2yxqvB0.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 20.04.3 LTS (GNU/Linux 5.15.0-52-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

388 updates can be applied immediately.
276 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable

Your Hardware Enablement Stack (HWE) is supported until April 2025.

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
```

14. Lalu download dan install hadoop di Ubuntu.
15. Ekstrak file hadoop yang di download tadi dengan perintah tar xzf hadoop-3.3.0.tar.gz.

```
renada@renada-VirtualBox:~$ tar xzf hadoop-3.3.0.tar.gz
```

16. Akses file .bashrc lagi lalu tambahkan Hadoop related options ke dalam file .bashrc.

```
#Hadoop Related Options
export HADOOP_HOME=/home/renada/hadoop-3.3.0
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

17. Gunakan perintah source~/.bashrc untuk membuat new environment variable dapat terlihat.

```
renada@renada-VirtualBox:~$ source ~/.bashrc
```

18. Lakukan modify atau pengubahan pada file di bawah ini.
    a. hadoop-env.sh
       Pada file ini kita menghapus komentar pada JAVA_HOME dengan menghapus #, dan mengubahnya menjadi
       `JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64`.

b. core-site.xml

Tambahkan perintah di bawah ini :

```
<configuration>
<property>
  <name>fs.default.name</name>
  <value>hdfs://localhost:9000</value>
</property>
</configuration>
```

c. hdfs-site.xml

(Sebelumnya buat directory untuk datanodes dan namenodes dan tambahkan di hdfs-site.xml)

Tambahkan perintah di bawah ini :

```
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
<property>
  <name>dfs.namenode.name.dir</name>
<value>/home/amit/hadoop-3.3.0/data/namenode</value>
</property>
<property>
  <name>dfs.datanode.data.dir</name>
<value>/home/amit/hadoop-3.3.0/data/datanode</value>
</property>
```

d. mapred-site-xml

Tambahkan perintah di bawah ini :

```
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
```

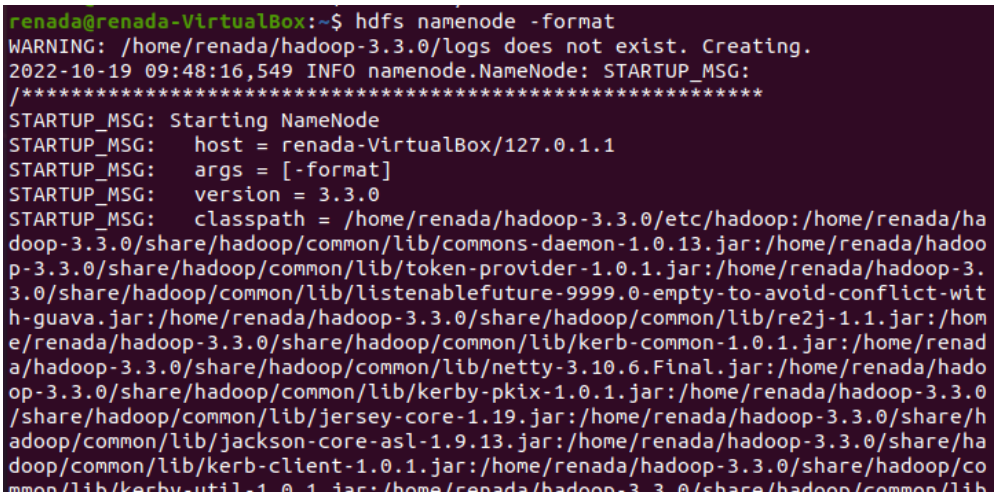
e. yarn-site.xml

Tambahkan perintah di bawah ini :

```
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
<property>

<name>yarn.nodemanager.aux-services.mapreduce.shuffle.
```

```
class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>127.0.0.1</value>
</property>
<property>
  <name>yarn.acl.enable</name>
  <value>0</value>
</property>
<property>
  <name>yarn.nodemanager.env-whitelist</name>

<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,H
ADOOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN
_HOME,HADOOP_MAPRED_HOME</value>
</property>
```

19. Format HDFS NameNode
    Dengan menggunakan perintah `hdfs namenode -format`.



Wordcount Hadoop Single Node

1. Menjalankan Hadoop dengan perintah
   `start-dfs.sh`
   `start-yarn.sh`
2. Membuat folder ‘input’
   hadoop fs -mkdir /input
3. Membuat file sebagai input untuk *wordcount*
   `sudo nano inputWordCount.txt`
4. Memindahkan file input ke dalam folder ‘input’

```
hadoop fs -put inputWordCount.txt /input
```

5. Menjalankan program Wordcount pada hadoop mapreduce-examples-3.3.0.jar
```
hadoop jar
hadoop-3.3.0/share/hadoop/mapreduce/hadoop-mapreduce-exam
ples-3.3.0.jar wordcount /input/inputWordCount.txt
/output
```
6. Melihat hasil Wordcount di file part-r-00000
```
hadoop fs -cat /output/part-r-00000
```