

## MODUL 2 HADOOP



### CAPAIAN PEMBELAJARAN

---

Mahasiswa menggunakan hadoop.



### KEBUTUHAN ALAT/BAHAN/SOFTWARE

---

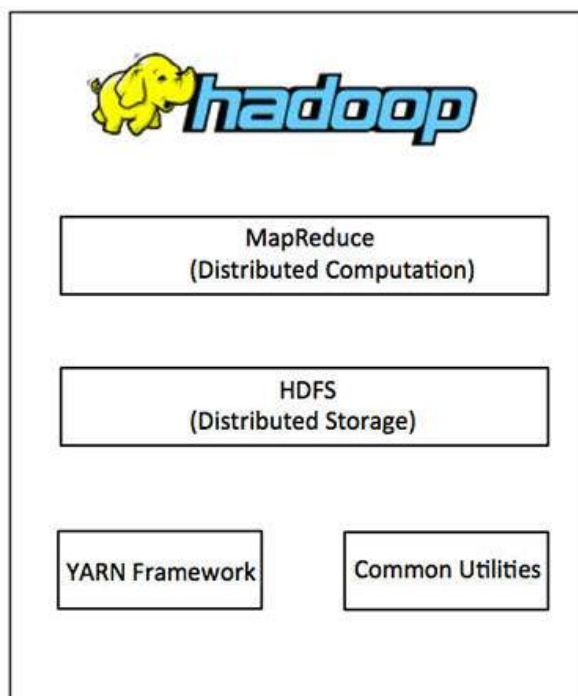
1. Sistem Operasi Ubuntu
2. Hadoop 2.8.5
3. Oracle JDK 1.8
4. Editor teks (Gedit, Vim, Mcedit atau Nano)



### DASAR TEORI

---

Hadoop adalah open-source framework yang memungkinkan untuk menyimpan dan memproses big data dalam lingkungan terdistribusi di seluruh cluster komputer yang menggunakan model pemrograman sederhana. Hadoop dirancang untuk meningkatkan server tunggal menjadi ribuan mesin, masing-masing menawarkan komputasi dan penyimpanan lokal.



Gambar Arsitektur Hadoop



## PRAKTIK

---

### Instalasi Hadoop

Beberapa hal yang perlu diperhatikan untuk penggunaan hadoop pada praktikum matakuliah Data Engineering adalah:

1. Untuk instalasi Hadoop disini menggunakan system operasi Linux Ubuntu 16 (Jalankan Virtual Box dengan sistem operasi Ubuntu 16 hasil pekerjaan modul pertemuan 1.)
2. Sebagai konvensi (kesepakatan umum) penanda akses adalah  
Tanda # berarti perintah meminta akses sebagai root atau untuk user biasa dengan menambahkan perintah sudo  
Tanda \$ berarti perintah meminta akses sebagai user biasa

Instalasi hadoop dilakukan dengan langkah dan perintah seperti berikut, tuliskan di Terminal

Membuat user baru untuk hadoop

```
$ sudo -s  
# adduser hadoop
```

Buka web browser (Chrome atau Mozilla atau lainnya), akses [www.oracle.com](http://www.oracle.com) dan login dengan akun oracle yang Anda miliki. Kemudian akses alamat url di bawah

<https://www.oracle.com/technetwork/java/javase/downloads/java-archive-javase8-2177648.html>

Cari dan download file JDK 8 Update 192 (nama file : `jdk-8u192-linux-x64.tar.gz`), simpan dan ekstraksi di directory `/opt`.

Ekstraksi JDK 8 Update 192

```
$ sudo -s  
# cd /opt  
# tar -xzf jdk-8u192-linux-x64.tar.gz
```

Setting default JVM dengan JDK 8 Update 192

```
$ sudo -s  
# update-alternatives --install /usr/bin/java  
  java /opt/jdk1.8.0_192/bin/java 100  
# update-alternatives --install /usr/bin/javac  
  javac /opt/jdk1.8.0_192/bin/javac 100
```

Verifikasi instalasi java apakah sukses atau gagal

```
$ sudo -s
# update-alternatives --display java
# update-alternatives --display javac
# java -version
```

Instalasi ssh server dan open ssh

```
$ sudo apt-get install openssh-server openssh-client
```

Konfigurasi Generate Public and Private Key Pairs untuk user hadoop

```
$ su - hadoop
$ ssh-keygen -t rsa
$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
$ ssh localhost
```

Instalasi hadoop

```
$ wget https://archive.apache.org/dist/hadoop/common/
  hadoop-2.8.5/hadoop-2.8.5.tar.gz

$ tar -xzvf hadoop-2.8.5.tar.gz
```

Edit file .bashrc yang ada di directory /home/hadoop dengan nano editor

```
$ cd ~
$ nano .bashrc
```

Tambahkan script berikut di bagian akhir di dalam file .bashrc

```
export HADOOP_HOME=/home/hadoop/hadoop-2.8.5
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"
```

Untuk simpan file di nano editor tekan tombol keyboard Ctrl + O

Untuk keluar dari nano editro tekan tombol keyboard Ctrl + X

Jalankan perintah berikut untuk mengatur environment instalasi hadoop

```
$ source ~/.bashrc
```

Edit file `hadoop-env.sh` yang ada di directory `/home/hadoop/hadoop-2.8.5/etc/hadoop` dengan nano editor

```
$ cd /home/hadoop/hadoop-2.8.5/etc/hadoop
```

```
$ nano hadoop-env.sh
```

Cari dan ubahlah `JAVA_HOME` dan `HADOOP_CONF_DIR` menjadi seperti berikut

```
export JAVA_HOME=/opt/jdk1.8.0_192
```

```
export HADOOP_CONF_DIR=${HADOOP_CONF_DIR:-"/home/hadoop/hadoop-2.8.5/etc/hadoop"}
```

Untuk simpan file di nano editor tekan tombol keyboard `Ctrl + O`

Untuk keluar dari nano editor tekan tombol keyboard `Ctrl + X`

Edit file `core-site.xml` yang ada di directory `/home/hadoop/hadoop-2.8.5/etc/hadoop` dengan nano editor

```
$ cd /home/hadoop/hadoop-2.8.5/etc/hadoop
```

```
$ nano core-site.xml
```

Ubahlah tag `<configuration>` menjadi seperti berikut

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>/home/hadoop/hadooptmpdata</value>
  </property>
</configuration>
```

Untuk simpan file di nano editor tekan tombol keyboard `Ctrl + O`

Untuk keluar dari nano editor tekan tombol keyboard `Ctrl + X`

Buatlah folder dengan nama `hadooptmpdata` di directory home user `hadoop`

```
$ cd /home/hadoop
```

```
$ mkdir hadooptmpdata
```

Edit file `hdfs-site.xml` yang ada di directory `/home/hadoop/hadoop-2.8.5/etc/hadoop` dengan nano editor

```
$ cd /home/hadoop/hadoop-2.8.5/etc/hadoop
$ nano hdfs-site.xml
```

Ubahlah tag `<configuration>` menjadi seperti berikut

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
    <name>dfs.name.dir</name>
    <value>file:///home/hadoop/hdfs/namenode</value>
    <name>dfs.data.dir</name>
    <value>file:///home/hadoop/hdfs/datanode</value>
  </property>
</configuration>
```

Untuk simpan file di nano editor tekan tombol keyboard `Ctrl + O`

Untuk keluar dari nano editor tekan tombol keyboard `Ctrl + X`

Membuat namenode di dalam folder home untuk user hadoop

```
$ cd /home/hadoop
$ mkdir -p hdfs/namenode
$ mkdir -p hdfs/datanode
```

Copy `mapred-site.xml` dari `mapred-site.xml.template` di directory `/home/hadoop/hadoop-2.8.5/etc/hadoop`

```
$ cd /home/hadoop/hadoop-2.8.5/etc/hadoop
$ cp mapred-site.xml.template mapred-site.xml
```

Edit file `mapred-site.xml` dengan nano editor

```
$ nano mapred-site.xml
```

Ubahlah tag `<configuration>` menjadi seperti berikut

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

Untuk simpan file di nano editor tekan tombol keyboard Ctrl + O

Untuk keluar dari nano editor tekan tombol keyboard Ctrl + X

Edit file yarn-site.xml yang ada di directory /home/hadoop/hadoop-2.8.5/etc/hadoop dengan nano editor

```
$ cd /home/hadoop/hadoop-2.8.5/etc/hadoop
```

```
$ nano yarn-site.xml
```

Ubahlah tag <configuration> menjadi seperti berikut

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

Untuk simpan file di nano editor tekan tombol keyboard Ctrl + O

Untuk keluar dari nano editor tekan tombol keyboard Ctrl + X

Untuk menjalankan server Hadoop adalah dengan menuliskan perintah berikut

```
$ su - hadoop
$ hdfs namenode -format
$ start-dfs.sh
$ start-yarn.sh
$ /opt/jdk1.8.0_192/bin/jps
```

Cek versi hadoop

```
$ hadoop version
```

Akses HDFS dan membuat directory di dalamnya

```
$ su - hadoop
$ hdfs dfs -mkdir /akakom
```

Menampilkan directory HDFS

```
$ hdfs dfs -ls /
```

Aplikasi web ui Hadoop untuk melihat namenode dapat dilihat dengan browser, tuliskan alamat url : [http://localhost: 50070](http://localhost:50070)

Aplikasi web ui Hadoop untuk melihat Resource Manager dapat dilihat dengan browser, tuliskan alamat url : [http://localhost :8088](http://localhost:8088)

Amatilah setiap langkah instalasi Hadoop dan hasilnya, kemudian tulis di laporan!



## **LATIHAN**

---

Diberikan oleh dosen / instruktur pengampu



## **TUGAS**

---

Diberikan oleh dosen / instruktur pengampu



## **REFERENSI**

---

<https://linuxconfig.org/how-to-install-hadoop-on-ubuntu-18-04-bionic-beaver-linux>  
<https://tecadmin.net/setup-hadoop-on-ubuntu/>