LAPORAN PRAKTIKUM INFRASTRUKTUR BIG DATA
PERTEMUAN 2

Oleh :

Nama   : Jauhari Ahmad
No. Mhs   : 205411167
Jurusan   : Teknik Informatika / S1

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
AKAKOM
**YOGYAKARTA**
2020

PENDAHULUAN

A. TUJUAN

Mahasiswa dapat menggunakan Hadoop

B. 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
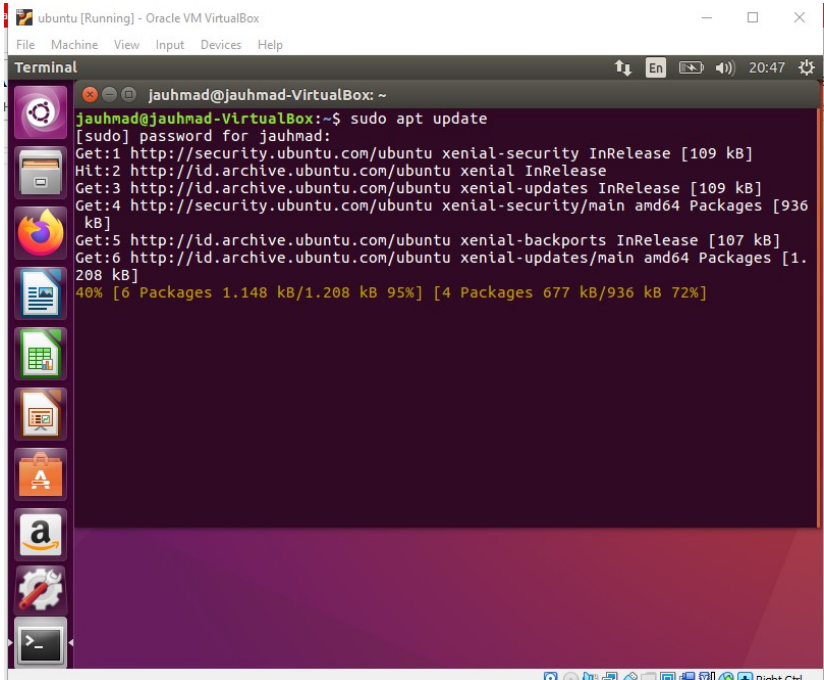
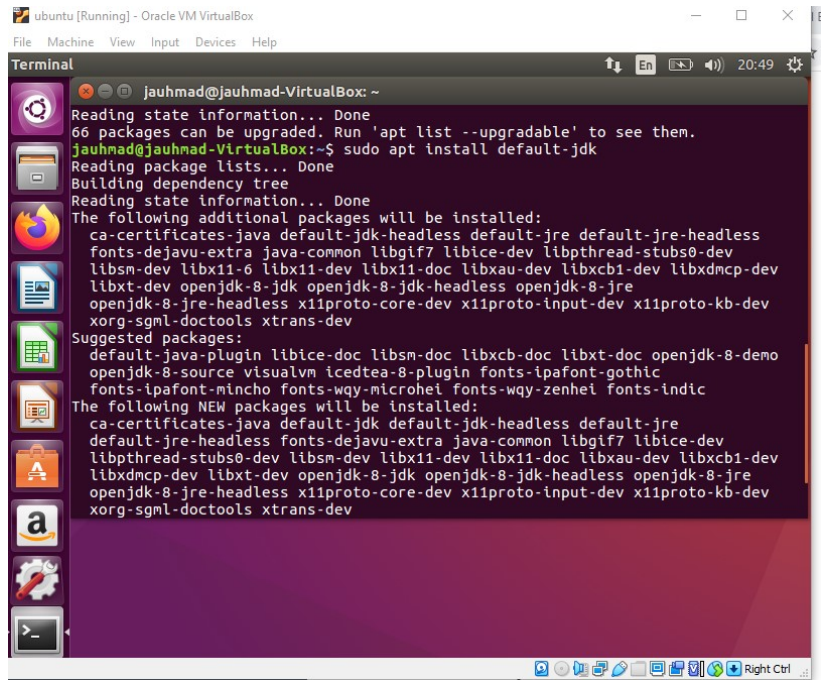penyimpanan lokal.
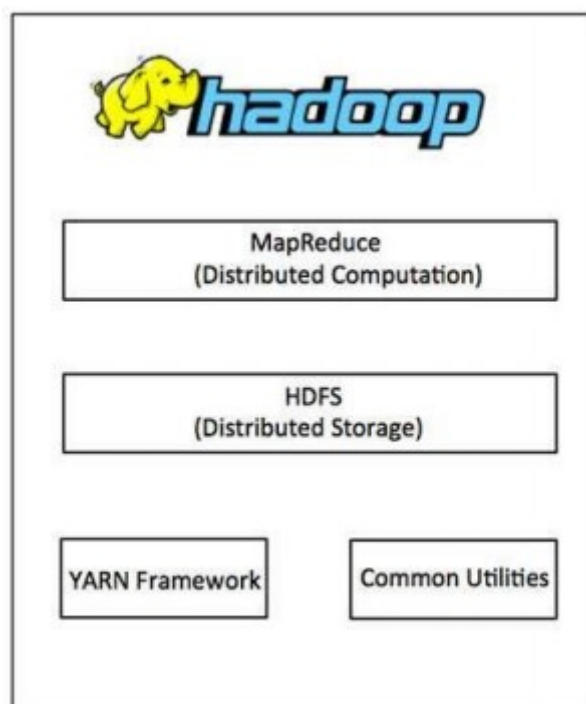
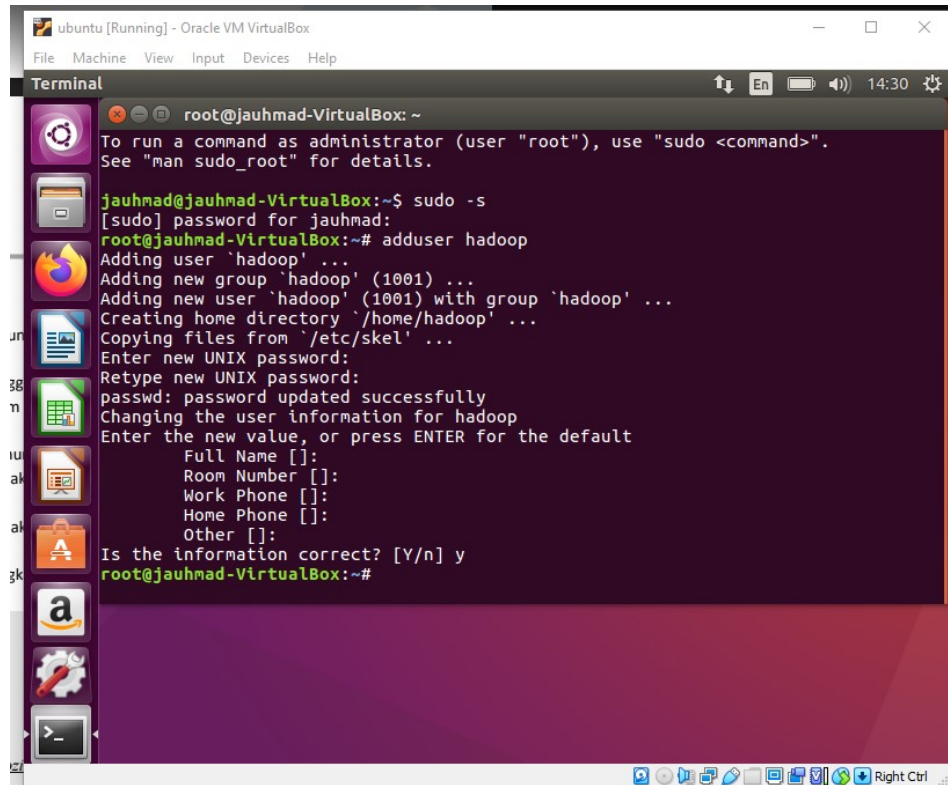Gambar Arsitektur Hadoop

PEMBAHASAN

Instalasi Hadoop

1. Install JDK terlebih dahulu sebagai prasyarat

   a. Update paket

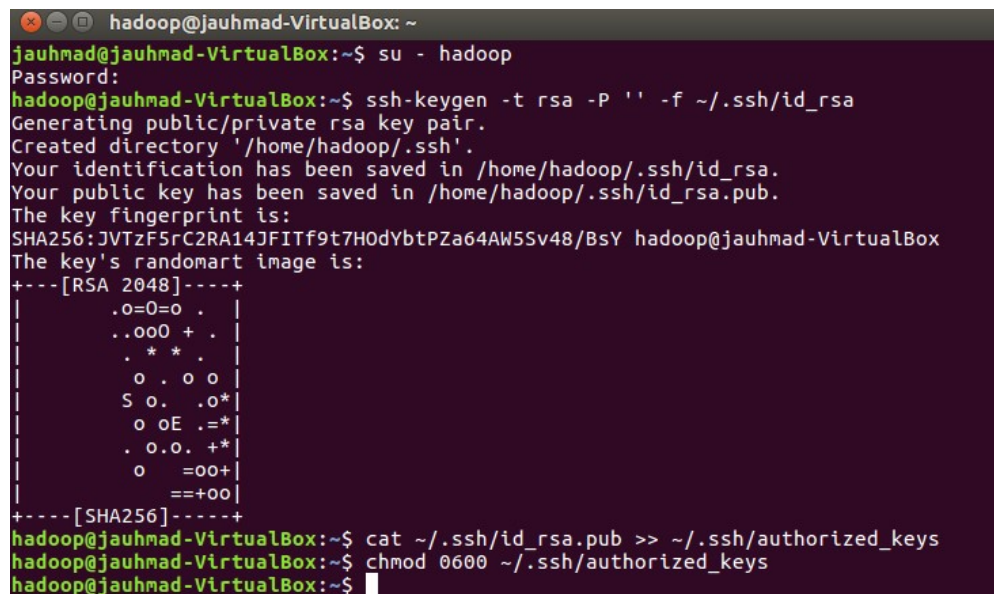   b. Install paket JDK

2. Membuat user untuk Hadoop

3. Pengaturan SSH

4. Unduh arsip source Hadoop

https://phoenixnap.com/kb/install-hadoop-ubuntu

```
hadoop@jauhmad-VirtualBox:~$ wget https://downloads.apache.org/hadoop/common/had
oop-3.2.1/hadoop-3.2.1.tar.gz
--2020-09-22 21:17:17--  https://downloads.apache.org/hadoop/common/hadoop-3.2.1
/hadoop-3.2.1.tar.gz
Resolving downloads.apache.org (downloads.apache.org)... 88.99.95.219, 2a01:4f8:
10a:201a::2
Connecting to downloads.apache.org (downloads.apache.org)|88.99.95.219|:443... c
onnected.
HTTP request sent, awaiting response... 200 OK
Length: 359196911 (343M) [application/x-gzip]
Saving to: 'hadoop-3.2.1.tar.gz'

hadoop-3.2.1.tar.gz   2%[                    ]   8,13M  91,3KB/s    eta 35m 27s
```

5. Setup Hadoop

   a. Ekstrak file

```
jauhmad@jauhmad-VirtualBox:~$ tar xzf hadoop-3.2.1.tar.gz
```

   b. Konfigurasi

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

   c. Tambahkan script konfigurasi pada baris terakhir

```
#Hadoop Related Options
export HADOOP_HOME=/home/hadoop/hadoop-3.2.1
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"
```

   d.
```
jauhmad@jauhmad-VirtualBox:~$ source ~/.bashrc
```

   e. Edit hadoop-env.sh file

```
jauhmad@jauhmad-VirtualBox:~$ sudo nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh
```

   f. Edit core-site.xml file

```
hadoop@jauhmad-VirtualBox:~$ nano $HADOOP_HOME/etc/hadoop/core-site.xml
```

```
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
  <name>hadoop.tmp.dir</name>
  <value>/home/hadoop/tmpdata</value>
</property>
<property>
  <name>fs.default.name</name>
  <value>hdfs://127.0.0.1:9000</value>
</property>
</configuration>
```

g. Edit hdfs-site.xml file

```
hadoop@jauhmad-VirtualBox:~$ nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml
```

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

h. Edit mapred-site.xml

```
hadoop@jauhmad-VirtualBox:~$ nano $HADOOP_HOME/etc/hadoop/mapred-site.xml
```

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

i. Edit yarn-site.xml

```
<configuration>

<!-- Site specific YARN configuration properties -->
<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,HADOOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
</property>
</configuration>
```

j. Format HDFS namenode

```
hadoop@jauhmad-VirtualBox:~$ hdfs namenode -format
```

k. Start hadoop cluster

```
hadoop@jauhmad-VirtualBox:~/hadoop-3.2.1/sbin$ ./start-dfs.sh
```

6. Akses Hadoop dari browser

Namenode information	×	+

← → C ⌂   🛡 ⓘ 127.0.0.1:9870/dfshealth.html#tab-overview   ··· ♡ ☆

Hadoop   Overview   Datanodes   Datanode Volume Failures   Snapshot   Startup Progress   Utilities ▾

Overview 'localhost:9000' (active)

Started:	Fri Apr 24 10:46:35 -0600 2020
Version:	3.2.1, rb3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled:	Tue Sep 10 09:56:00 -0600 2019 by rohithsharmaks from branch-3.2.1
Cluster ID:	CID-256c18e1-3465-41e2-9954-35a50358b413
Block Pool ID:	BP-1868655187-127.0.1.1-1587659926013

KESMIPULAN

1. Java adalah persyaratan utama untuk menjalankan Hadoop pada sistem apa pun

2. Hadoop NameNode dimulai pada port default 9870

3. Akses port 9864 untuk mendapatkan detail tentang node Hadoop

_____