Hadoop Installation:

Name: อวยชัย ภิรมย์รื่น

Tel. : 086-813-5354

e-mail : p.Auoychai@gmail.com

Wrap up


HADOOP
BIG DATA
OPEN SOURCE
STORAGE
ANALYTICS
CLOUD
BUSINESS
INTERNET
ANALYSIS
DATA
SECURITY
SHARE
MANAGE
RISK
STATISTIC
OPEN SOURCE
SOLVED
SOFTWARE
VOLUME
TREND
SOCIAL
MEDIA
TECHNOLOGIES
SYSTEMS
COMPUTING
DATABASE
NETWORKS
LEARNING
MARKETING
COMPLEX
SERVER
ENTERPRISE
PROJECTS
OPEN SOURCE

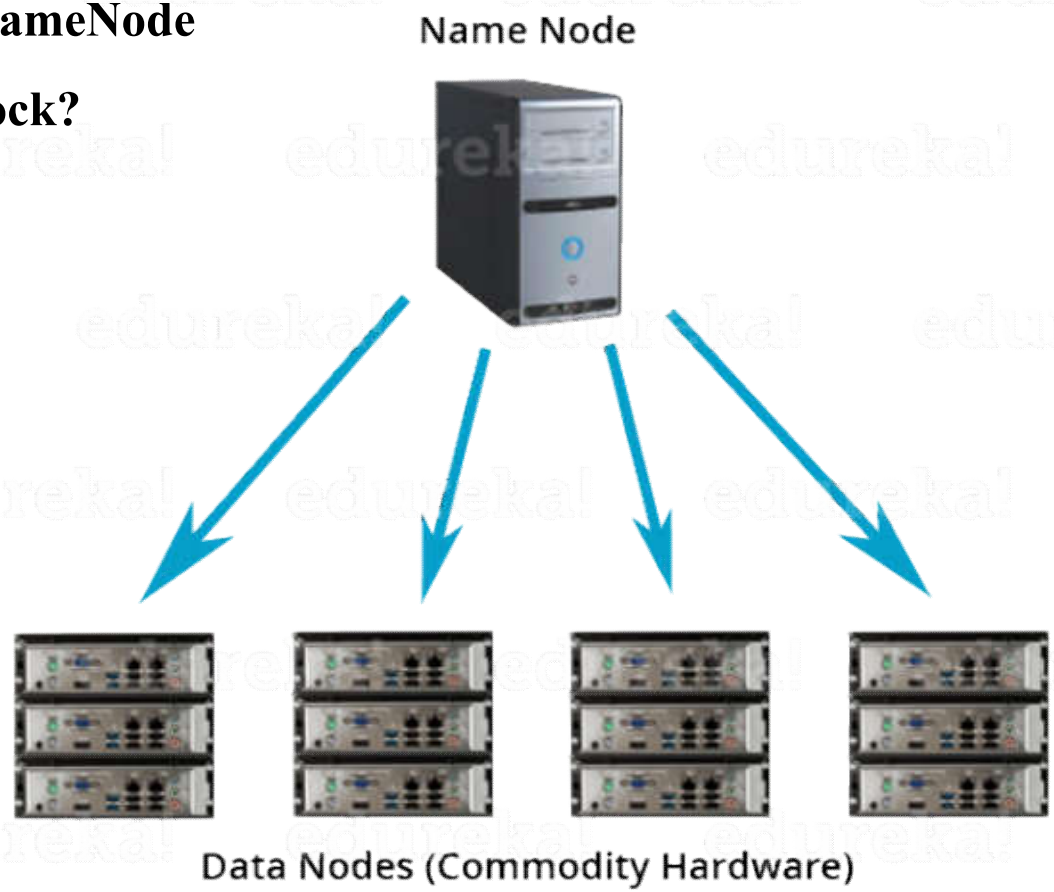
System Topology :

HDFS Master/Slave Architecture

NameNode, DataNode and

Secondary NameNode

What is a block?



Installation Mode :

Hadoop Installation Mode:

Local standalone mode:

Pseudo-distributed mode:

Fully distributed mode:

Hadoop Installation :

Hadoop Installation:

Define Hadoop System Topology

- Master Node: IP=
- Slave Node#1: IP=
- Slave Node#2: IP=
- Slave Node#3: IP=

Hadoop Installation :

Hadoop Installation:

#Define Hadoop Instance Directory

- Main Package Directory : /usr/local/hadoop
- Data Directory : /var/hadoop_data/namenode

  : /var/hadoop_data/datanode
- Log file Directory : /var/log/hadoop

Hadoop Installation:

#Hadoop Configuration Review:

Core-site.xml

```
<configuration>
    <property>
        <name>fs.defaultFS</name>
        <value>hdfs://localhost:9000</value>
    </property>
    <property>
        <name>hadoop.proxyuser.root.hosts</name>
        <value>*</value>
    </property>
    <property>
        <name>hadoop.proxyuser.root.groups</name>
        <value>*</value>
    </property>
</configuration>
```

Hadoop Installation:

#Hadoop Configuration Review:

hdfs-site.xml

```xml
<configuration>
    <property>
        <name>dfs.replication</name>
        <value>1</value>
    </property>
    <property>
        <name>dfs.namenode.name.dir</name>
        <value>file:/var/hadoop_data/namenode</value>
    </property>
    <property>
        <name>dfs.datanode.data.dir</name>
        <value>file:/var/hadoop_data/datanode</value>
    </property>
    <property>
        <name>dfs.namenode.acls.enabled</name>
        <value>true</value>
    </property>

</configuration>
```

Hadoop Installation:

#Hadoop Configuration Review:

yarn-site.xml

```xml
<configuration>
<!-- Site specific YARN configuration properties -->
    <property>
        <name>yarn.resourcemanager.hostname</name>
        <value>localhost</value>
    </property>
    <property>
        <name>yarn.resourcemanager.scheduler.address</name>
        <value>localhost:8030</value>
    </property>
    <property>
        <name>yarn.resourcemanager.resource-tracker.address</name>
        <value>localhost:8031</value>
    </property>
    <property>
        <name>yarn.resourcemanager.address</name>
        <value>localhost:8032</value>
    </property>
    <property>
        <name>yarn.resourcemanager.admin.address</name>
        <value>localhost:8033</value>
    </property>
    <property>
        <name>yarn.resourcemanager.webapp.address</name>
        <value>localhost:8088</value>
    </property>
    <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>
</configuration>
```

Hadoop Installation:

#Hadoop Configuration Review:

mapred-site.xml

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

Hadoop Installation:

Initial HDFS : HDFS Formatting

```
$hadoop namenode -format
```

Post Test Installation:

How to Start&Stop Hadoop :

- start-dfs.sh , start-yarn.sh
- stop-yarn.sh , stop-dfs.sh

UbuntuServer [Running] - Oracle VM VirtualBox

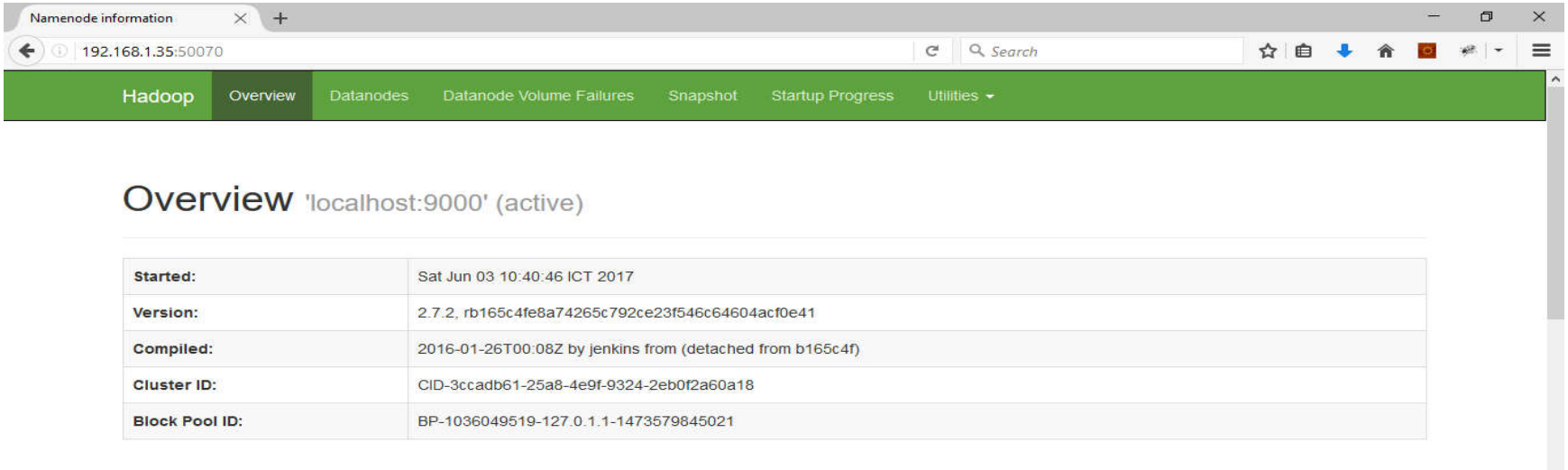
File Machine View Input Devices Help

```
auoychai@ubtserver:/usr/local/hadoop/share/hadoop/mapreduce$ start-dfs.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /var/log/hadoop/hadoop-auoychai-namenode-ubtserver.out
localhost: starting datanode, logging to /var/log/hadoop/hadoop-auoychai-datanode-ubtserver.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /var/log/hadoop/hadoop-auoychai-secondarynamenode-ub
tserver.out
auoychai@ubtserver:/usr/local/hadoop/share/hadoop/mapreduce$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /var/log/hadoop/yarn-auoychai-resourcemanager-ubtserver.out
localhost: starting nodemanager, logging to /var/log/hadoop/yarn-auoychai-nodemanager-ubtserver.out
auoychai@ubtserver:/usr/local/hadoop/share/hadoop/mapreduce$
auoychai@ubtserver:/usr/local/hadoop/share/hadoop/mapreduce$ jps
6398 NameNode
6932 ResourceManager
6542 DataNode
6775 SecondaryNameNode
7389 Jps
7077 NodeManager
auoychai@ubtserver:/usr/local/hadoop/share/hadoop/mapreduce$ _
```

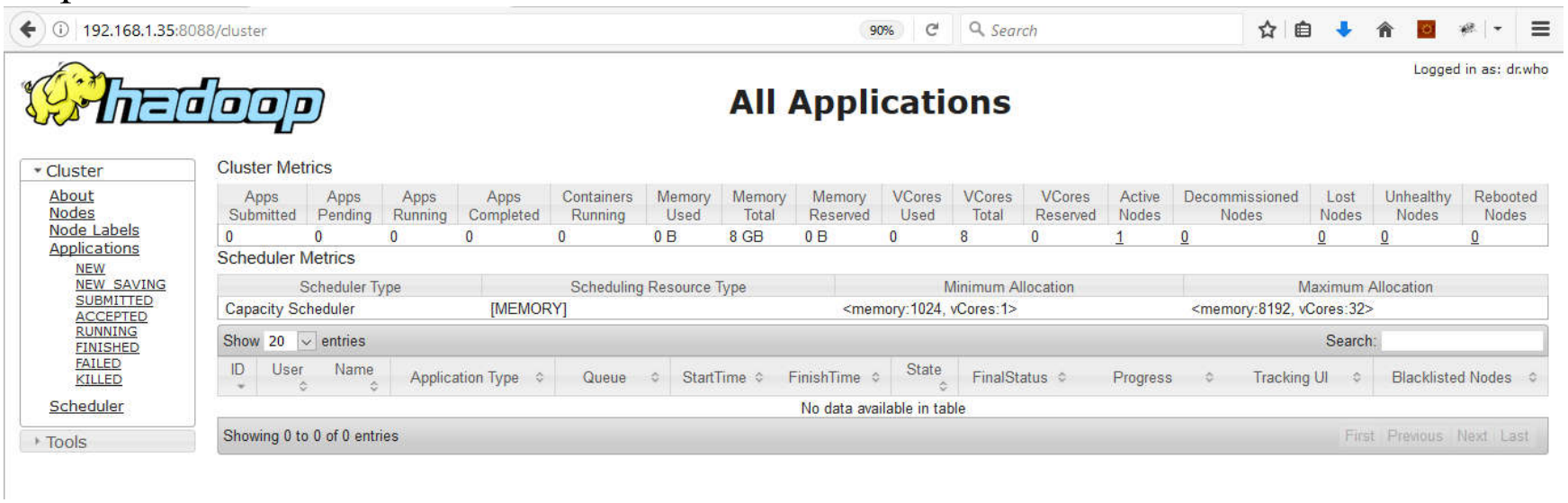
Web UI: <IP>:50070

Post Test Installation:

http://< IP >:50070/



http://< IP >:8088



Hand-On Plan:

A). Pseudo-distributed mode:

B). Fully distributed mode:

Download & Install Hadoop : Pseudo-distributed mode

- Update Package
- Setup SSH
- Setup Java
- Setup Hadoop

Download & Install Hadoop :

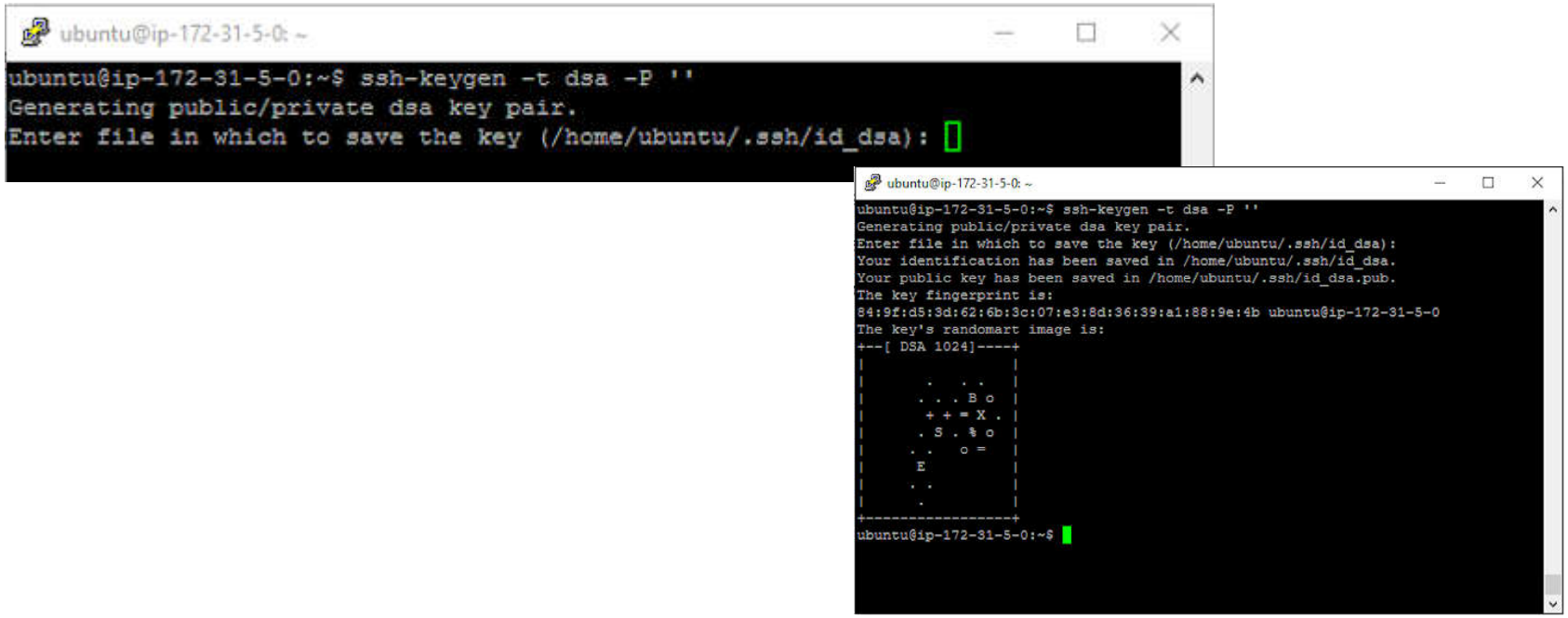
- Update Package

  sudo apt-get update

- Setup SSH

  sudo apt-get install –y openssh-server

  ssh-keygen –t dsa –P’’



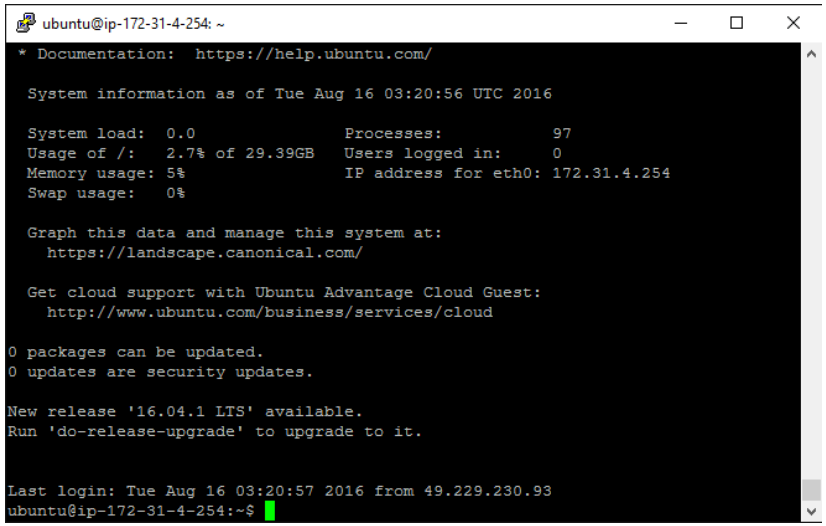
Download & Install Hadoop :

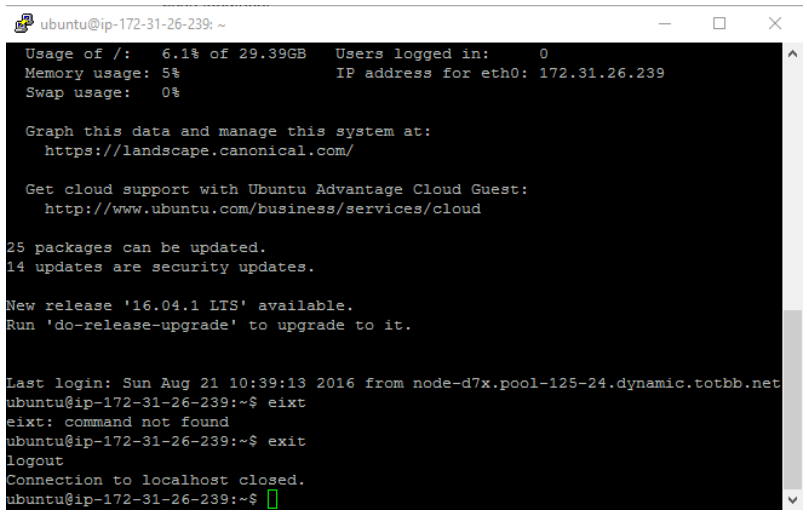
- Setup SSH

  -- โหลด Public Key เข้า Key-Store

  $cat .ssh/id_data.pub >> .ssh/authorized_key

  $ssh localhost





Download & Install Hadoop :

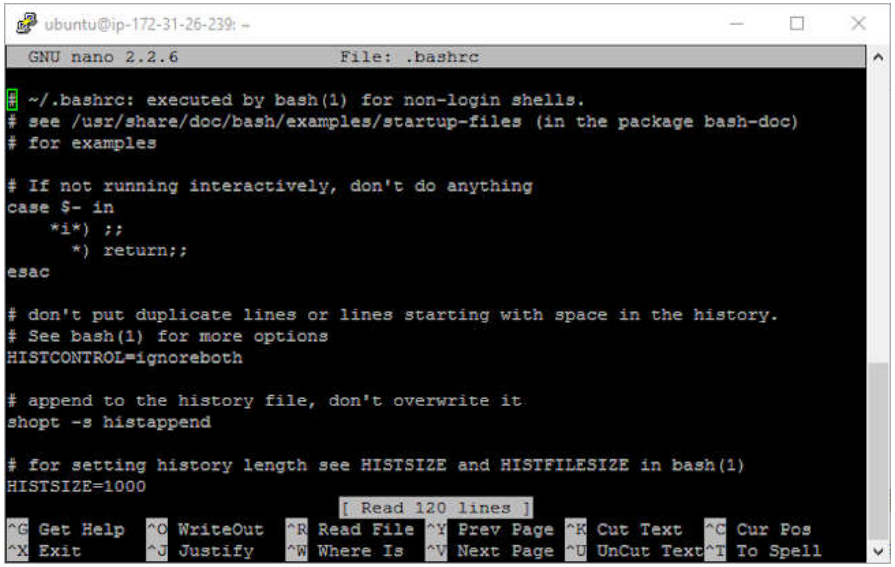
- Setup Java

  sudo apt-get install –y openjdk-7-jdk

  -- Add environment variable ที่ไฟล์ .bashrc

  export JAVA_HOME = usr/lib/jvm/java-7-openjdk-amd64

  export PATH=$PATH:$JAVA_HOME/bin



Download & Install Hadoop :

- Setup Hadoop

ii-1). Download Hadoop

wget http://mirror.cc.columbia.edu/pub/software/apache/hadoop/common/hadoop-2.7.2/hadoop-2.7.2.tar.gz

ii-2). สร้าง Folder สำหรับให้ Hadoop พิมพ์คำสั่งบรรทัดต่อไปนี้ที่หน้าจอ Console

```
sudo mkdir –p /usr/local/Hadoop  // โฟล์เดอร์สำหรับ Hadoop
sudo chown ubuntu:ubuntu –R /usr/local/hadoop
sudo mkdir /var/log/Hadoop  // โฟล์เดอร์สำหรับ Log file ของ Hadoop
sudo chown –R ubuntu:ubuntu  /var/log/hadoop
sudo mkdir –p /var/hadoop_data // โฟล์เดอร์สำหรับ Data file ที่ HDFS จัดการ
sudo mkdir –p /var/hadoop_data/namenode
sudo mkdir –p /var/hadoop_data/datanode
sudo chown ubuntu:ubuntu –R /var/hadoop_data
```

Download & Install Hadoop :

- Setup Hadoop

ii-3). Install Hadoop

tar –xvf hadoop-2.7.2.tar.gz

sudo mv ./hadoop-2.7.2/* /usr/local/hadoop

sudo chown ubuntu:ubuntu –R /usr/local/hadoop

ii-4). Setup Hadoop Variable Environment

เพิ่ม Environment Variable ให้กับ Hadoop ลงใน .bashrc file ตามนี้ ด้วยการพิมพ์คำสั่ง nano .bashrc เพิ่ม 2 บรรทัดด้านล่างลงไปตามนี้

export HADOOP_HOME=/usr/local/hadoop

export PATH=$PATH:$HADOOP_HOME/bin

export PATH=$PATH:$HADOOP_HOME/sbin

ii-5). แก้ไข script สำหรับกำหนด Environment Variable ให้กับ Hadoop ที่ไฟล์ hadoop-env.sh ที่อยู่ใน /usr/local/hadoop/etc/hadoop โดยพิมพ์ nano hadoop-env.sh และ เพิ่มคำสั่ง 2 บรรทัดด้านล่างนี้ แสดงตามรูปที่ ii5-1

export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64

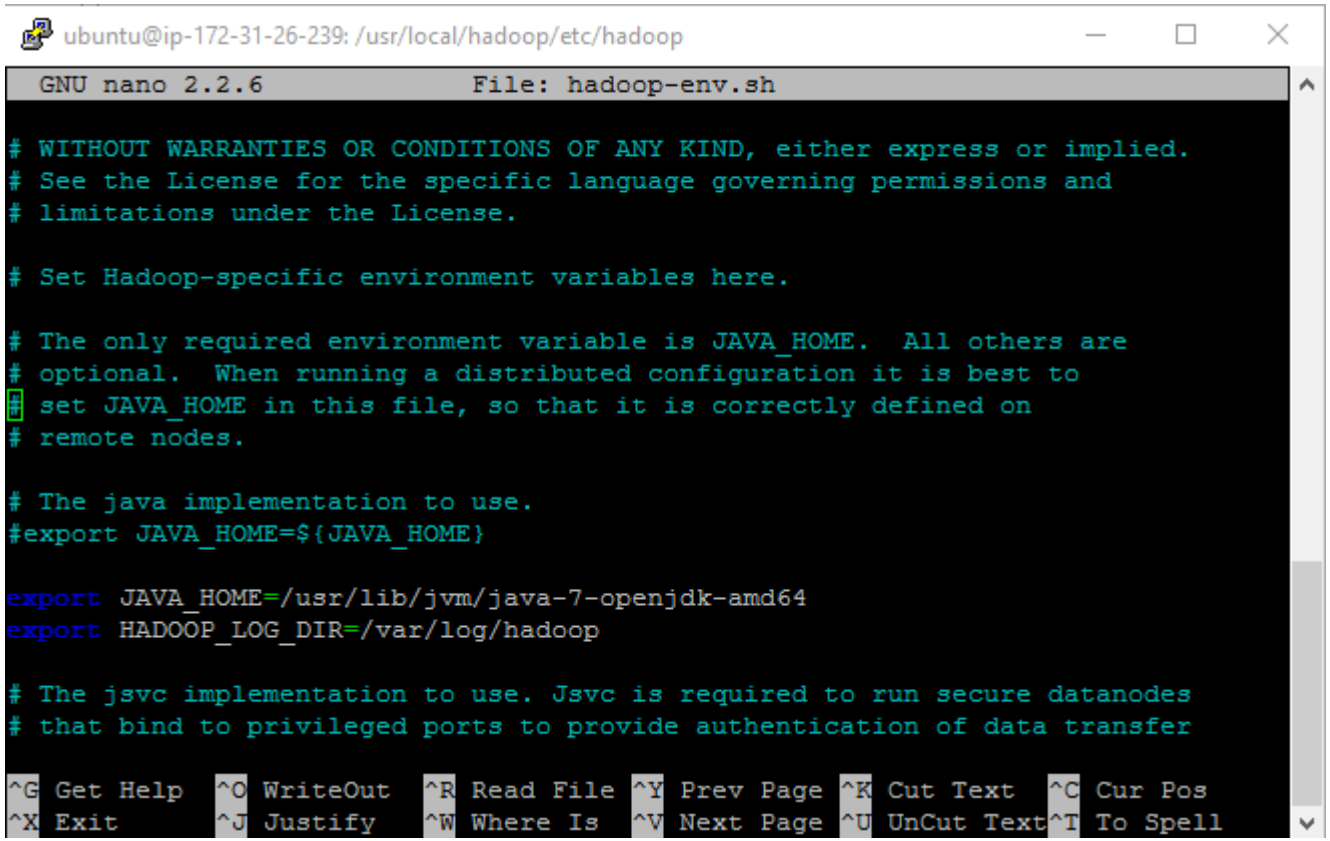
export HADOOP_LOG_DIR=/var/log/hadoop

Download & Install Hadoop :

ii-5). แก้ไข script สำหรับกำหนด Environment Variable ให้กับ Hadoop ที่ไฟล์ hadoop-env.sh ที่อยู่ใน /usr/local/hadoop/etc/hadoop โดยพิมพ์ nano hadoop-env.sh

export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64

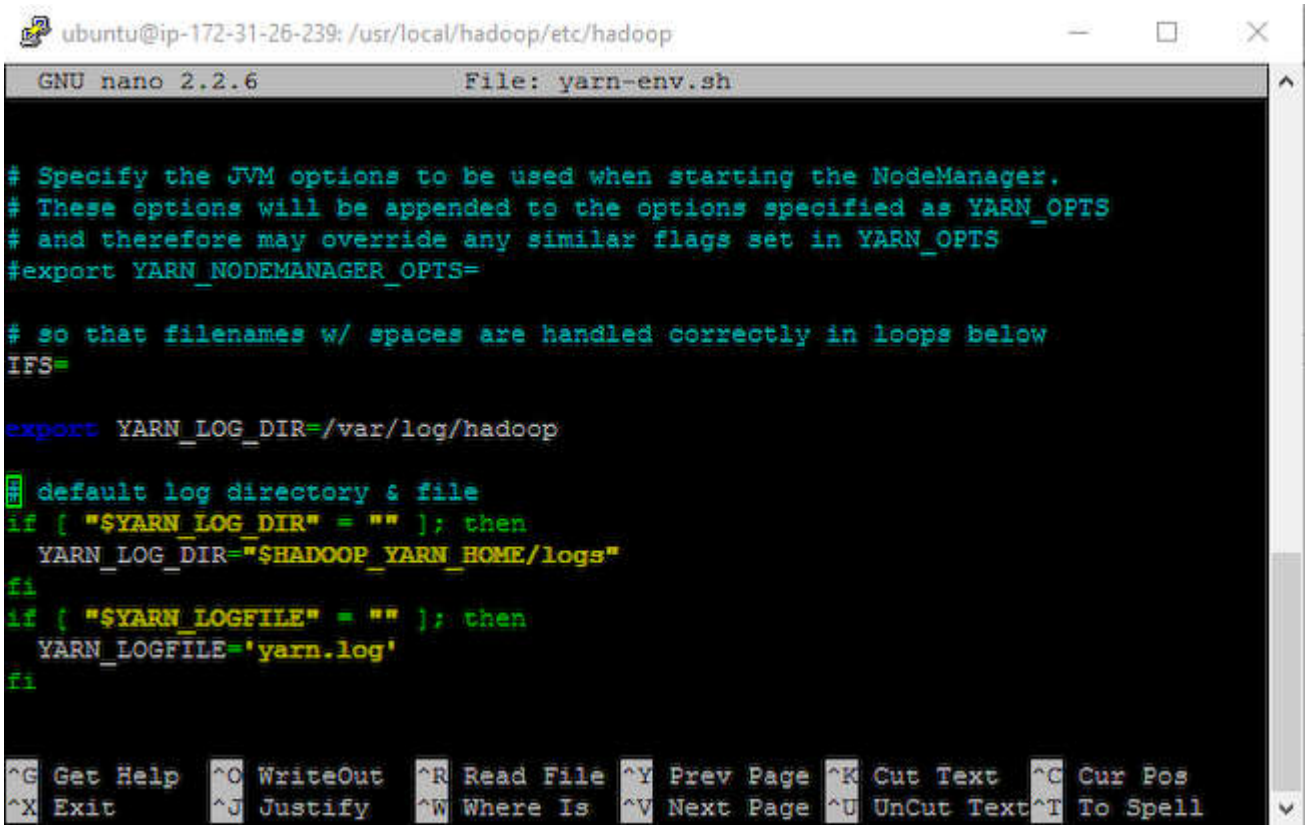
export HADOOP_LOG_DIR=/var/log/hadoop



Download & Install Hadoop :

ii-6). แก้ไข shell script สำหรับกำหนด Environment Variable ให้กับ Yarn ที่ไฟล์ yarn-env.sh ที่อยู่ใน /usr/local/hadoop/etc/hadoop โดย nano yarn-env.sh

export YARN_LOG_DIR = /var/log/hadoop



Initial HFDS :

ii-8). Format ระบบไฟล์ (ภาพเหมือนกับเรา Format Harddisk)

```
Last login: Sun Aug 21 11:52:06 2016 from node-d7x.pool-125-24.dynamic.totbb.net
ubuntu@ip-172-31-26-239:~$ cd /usr/local/hadoop/etc/hadoop/
ubuntu@ip-172-31-26-239:/usr/local/hadoop/etc/hadoop$ hdfs namenode -format
```

```
16/08/27 02:53:36 INFO util.GSet: capacity      = 2^15 = 32768 entries
Re-format filesystem in Storage Directory /var/hadoop_data/namenode ? (Y or N) Y
16/08/27 02:53:41 INFO namenode.FSImage: Allocated new BlockPoolId: BP-903428081
-172.31.26.239-1472266421834
16/08/27 02:53:41 INFO common.Storage: Storage directory /var/hadoop_data/nameno
de has been successfully formatted.
16/08/27 02:53:42 INFO namenode.NNStorageRetentionManager: Going to retain 1 ima
ges with txid >= 0
16/08/27 02:53:42 INFO util.ExitUtil: Exiting with status 0
16/08/27 02:53:42 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at ip-172-31-26-239.us-west-2.compute.inter
nal/172.31.26.239
************************************************************/
ubuntu@ip-172-31-26-239:/usr/local/hadoop/etc/hadoop$
```

Start&Stop Hadoop :

ii-9). Start Hadoop (ที่ /usr/local/hadoop/etc/hadoop)

ii9-1. ทำการ Start HDFS ด้วยการ Run Script นี้ start-dfs.sh และ สังเกตดูว่า Process NameNode และ DataNode ทำงานอยู่ด้วย คำสั่ง jps

ii9-2. ทำการ Start YARN ด้วยการ Run Script นี้ start-yarn.sh และ สังเกตดูว่า Process ResourceManager และ NodeManager ทำงานอยู่ด้วยคำสั่ง jps

ii-11). Stop Hadoop

1). stop-yarn.sh และ 2). stop-dfs.sh ที่ Command Console และ เมื่อขอดู Hadoop Process ด้วยคำสั่ง jps

Hello MapReduce :

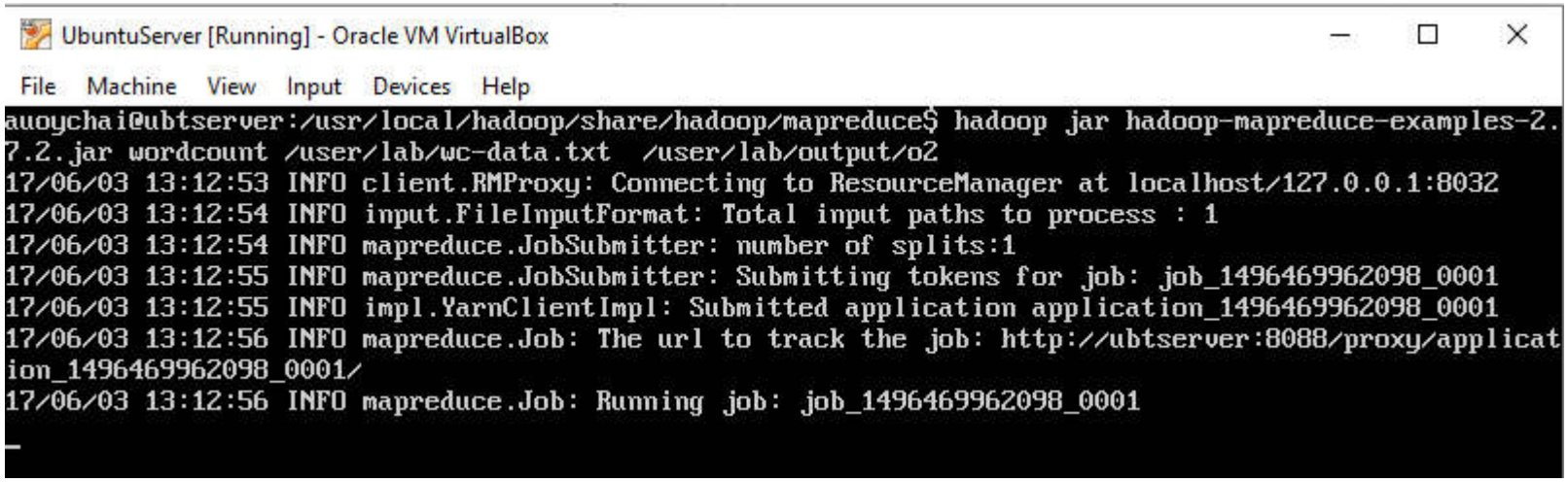
Package Directory: $HADOOP_HOME/share/hadoop/mapreduce

Execute Command :

$HADOOP_HOME/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-2.7.2.jar pi 10 1000

*** Observe terminal screen ***

$HADOOP_HOME/share/hadoop/mapreduce$ hadoop jar hadoop-mapreduce-examples-2.7.2.jar wordcount /user/lab/wc-data.txt /user/lab/output/o2



Download & Install Hadoop : Fully distributed mode

- X


The End
Big data
Shift
Del
Prt