



DIGITAL
TALENT
SCHOLARSHIP

DIGITAL TALENT SCHOLARSHIP 2019

Big Data Analytics



digitalent.kominfo.go.id



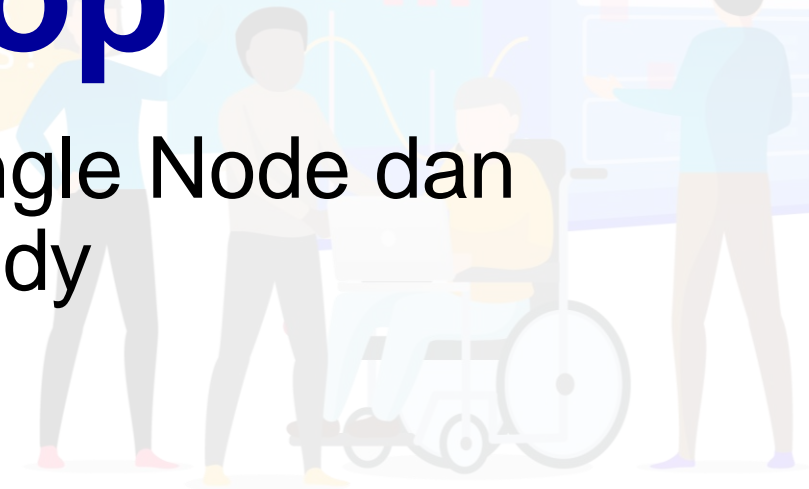
DIGITAL
TALENT
SCHOLARSHIP

Hadoop

Instalasi Hadoop Single Node dan Case Study

TERBUKA
UNTUK
DISABILITAS

BREAK
Y LIMITS



Requirement

- EC2 Instance
- Akses ke EC2 Instance.
 - (SSH Client, Putty, WinSCP, dll)

TERBUKA
UNTUK
DISABILITAS

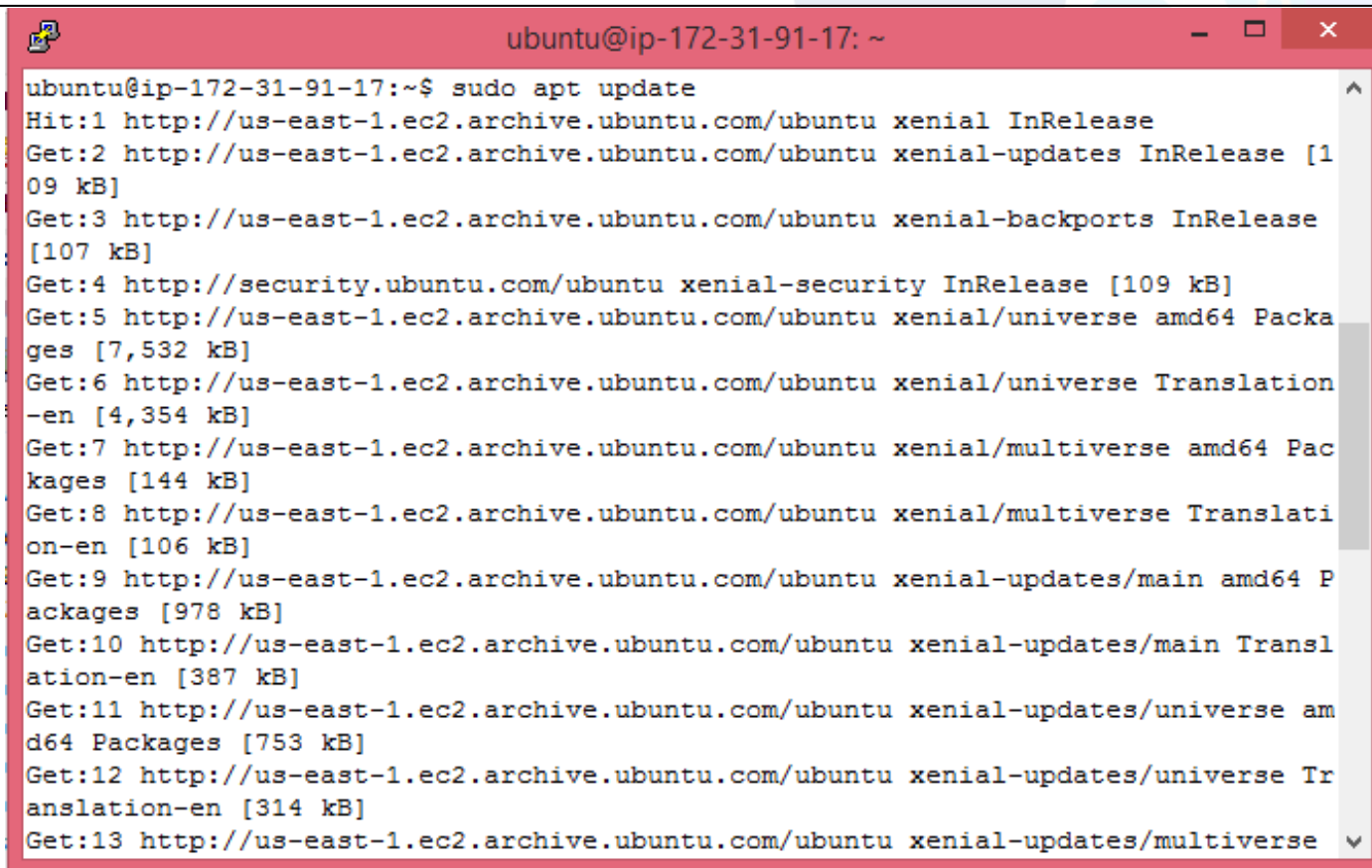
BREAK
YOUR
LIMITS!



Update apt Database

```
$ sudo apt update
```

TERBUKA
UNTUK



```
ubuntu@ip-172-31-91-17: ~  
ubuntu@ip-172-31-91-17:~$ sudo apt update  
Hit:1 http://us-east-1.ec2.archive.ubuntu.com/ubuntu xenial InRelease  
Get:2 http://us-east-1.ec2.archive.ubuntu.com/ubuntu xenial-updates InRelease [1  
09 kB]  
Get:3 http://us-east-1.ec2.archive.ubuntu.com/ubuntu xenial-backports InRelease  
[107 kB]  
Get:4 http://security.ubuntu.com/ubuntu xenial-security InRelease [109 kB]  
Get:5 http://us-east-1.ec2.archive.ubuntu.com/ubuntu xenial/universe amd64 Packa  
ges [7,532 kB]  
Get:6 http://us-east-1.ec2.archive.ubuntu.com/ubuntu xenial/universe Translation  
-en [4,354 kB]  
Get:7 http://us-east-1.ec2.archive.ubuntu.com/ubuntu xenial/multiverse amd64 Pac  
kages [144 kB]  
Get:8 http://us-east-1.ec2.archive.ubuntu.com/ubuntu xenial/multiverse Translati  
on-en [106 kB]  
Get:9 http://us-east-1.ec2.archive.ubuntu.com/ubuntu xenial-updates/main amd64 P  
ackages [978 kB]  
Get:10 http://us-east-1.ec2.archive.ubuntu.com/ubuntu xenial-updates/main Transl  
ation-en [387 kB]  
Get:11 http://us-east-1.ec2.archive.ubuntu.com/ubuntu xenial-updates/universe am  
d64 Packages [753 kB]  
Get:12 http://us-east-1.ec2.archive.ubuntu.com/ubuntu xenial-updates/universe Tr  
anslation-en [314 kB]  
Get:13 http://us-east-1.ec2.archive.ubuntu.com/ubuntu xenial-updates/multiverse
```



DIGITAL
TALENT
SCHOLARSHIP

Instalasi Java

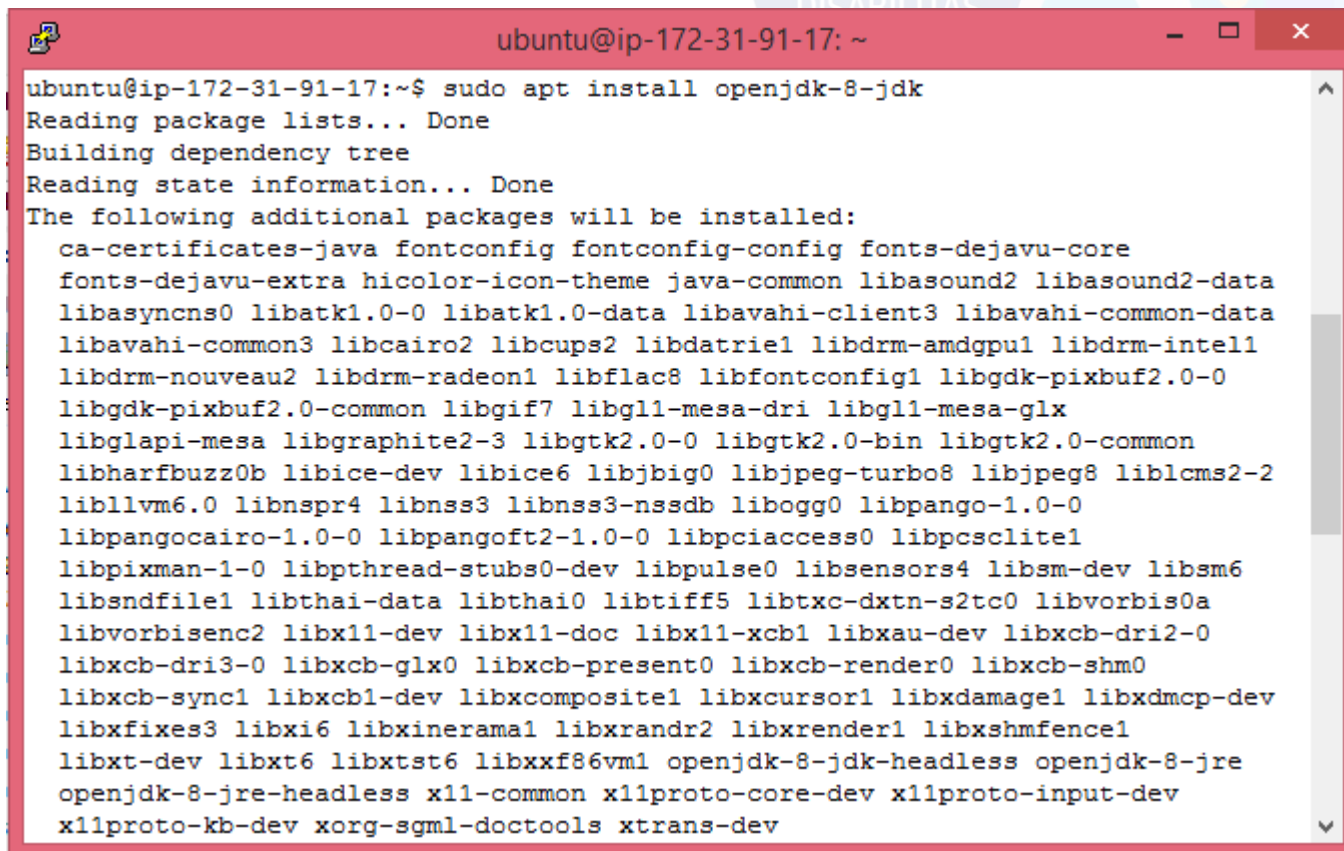
Big Data

Hadoop



Install OpenJDK-8

```
$ sudo apt install openjdk-8-jdk
```



```
ubuntu@ip-172-31-91-17: ~  
ubuntu@ip-172-31-91-17:~$ sudo apt install openjdk-8-jdk  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
The following additional packages will be installed:  
  ca-certificates-java fontconfig fontconfig-config fonts-dejavu-core  
  fonts-dejavu-extra hicolor-icon-theme java-common libasound2 libasound2-data  
  libasyncns0 libatk1.0-0 libatk1.0-data libavahi-client3 libavahi-common-data  
  libavahi-common3 libcairo2 libcups2 libdatatriel libdrm-amdgpu1 libdrm-intel1  
  libdrm-nouveau2 libdrm-radeon1 libflac8 libfontconfig1 libgdk-pixbuf2.0-0  
  libgdk-pixbuf2.0-common libgif7 libgl1-mesa-dri libgl1-mesa-glx  
  libglapi-mesa libgraphite2-3 libgtk2.0-0 libgtk2.0-bin libgtk2.0-common  
  libharfbuzz0b libice-dev libice6 libjbig0 libjpeg-turbo8 libjpeg8 liblcms2-2  
  libllvm6.0 libnspr4 libnss3 libnss3-nssdb libogg0 libpango-1.0-0  
  libpangocairo-1.0-0 libpangoft2-1.0-0 libpciaccess0 libpcsclite1  
  libpixman-1-0 libpthread-stubs0-dev libpulse0 libsensors4 libsm-dev libsm6  
  libsndfile1 libthai-data libthai0 libtiff5 libtxc-dxtn-s2tc0 libvorbis0a  
  libvorbisenc2 libx11-dev libx11-doc libx11-xcb1 libxau-dev libxcb-dri2-0  
  libxcb-dri3-0 libxcb-glx0 libxcb-present0 libxcb-render0 libxcb-shm0  
  libxcb-sync1 libxcb1-dev libxcomposite1 libxcursor1 libxdamage1 libxdmcp-dev  
  libxfixes3 libxi6 libxinerama1 libxrandr2 libxrender1 libxshmfence1  
  libxt-dev libxt6 libxtst6 libxxf86vm1 openjdk-8-jdk-headless openjdk-8-jre  
  openjdk-8-jre-headless x11-common x11proto-core-dev x11proto-input-dev  
  x11proto-kb-dev xorg-sgml-doctools xtrans-dev
```

Direktori instalasi Java

```
$ update-alternatives --display java
```



```
ubuntu@ip-172-31-91-17: ~  
ubuntu@ip-172-31-91-17:~$ update-alternatives --display java  
java - auto mode  
  link best version is /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java  
  link currently points to /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java  
  link java is /usr/bin/java  
  slave java.1.gz is /usr/share/man/man1/java.1.gz  
/usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java - priority 1081  
  slave java.1.gz: /usr/lib/jvm/java-8-openjdk-amd64/jre/man/man1/java.1.gz  
ubuntu@ip-172-31-91-17:~$
```

TERBUKA
UNTUK
DISABILITAS



Direktori instalasi Java (cont)

- Direktori instalasi java (dapat berbeda):
 - `/usr/lib/jvm/java-8-openjdk-amd64`



Variabel JAVA_HOME

```
$ nano ~/.bashrc
```

Atau..

```
$ nano ~/.profile
```

Gunakan text editor lain apabila tersedia.

TERBUKA
UNTUK
DISABILITAS

BEYOND
YOUR
LIMITS!



Variabel JAVA_HOME (cont)



```
ubuntu@ip-172-31-91-17: ~
GNU nano 2.5.3      File: /home/ubuntu/.bashrc

# ~/.bashrc: executed by bash(1) for non-login shells.
# see /usr/share/doc/bash/examples/startup-files (in the package bash-doc)
# for examples

# If not running interactively, don't do anything
case $- in
    *i*) ;;
    *) return;;
esac

# don't put duplicate lines or lines starting with space in the history.
# See bash(1) for more options
HISTCONTROL=ignoreboth

# append to the history file, don't overwrite it
shopt -s histappend

# for setting history length see HISTSIZE and HISTFILESIZE in bash(1)
HISTSIZE=1000

[ Read 117 lines ]

^G Get Help  ^O Write Out ^W Where Is  ^K Cut Text  ^J Justify   ^C Cur Pos
^X Exit      ^R Read File ^\ Replace   ^U Uncut Text ^T To Spell  ^_ Go To Line
```

Variabel JAVA_HOME (cont)

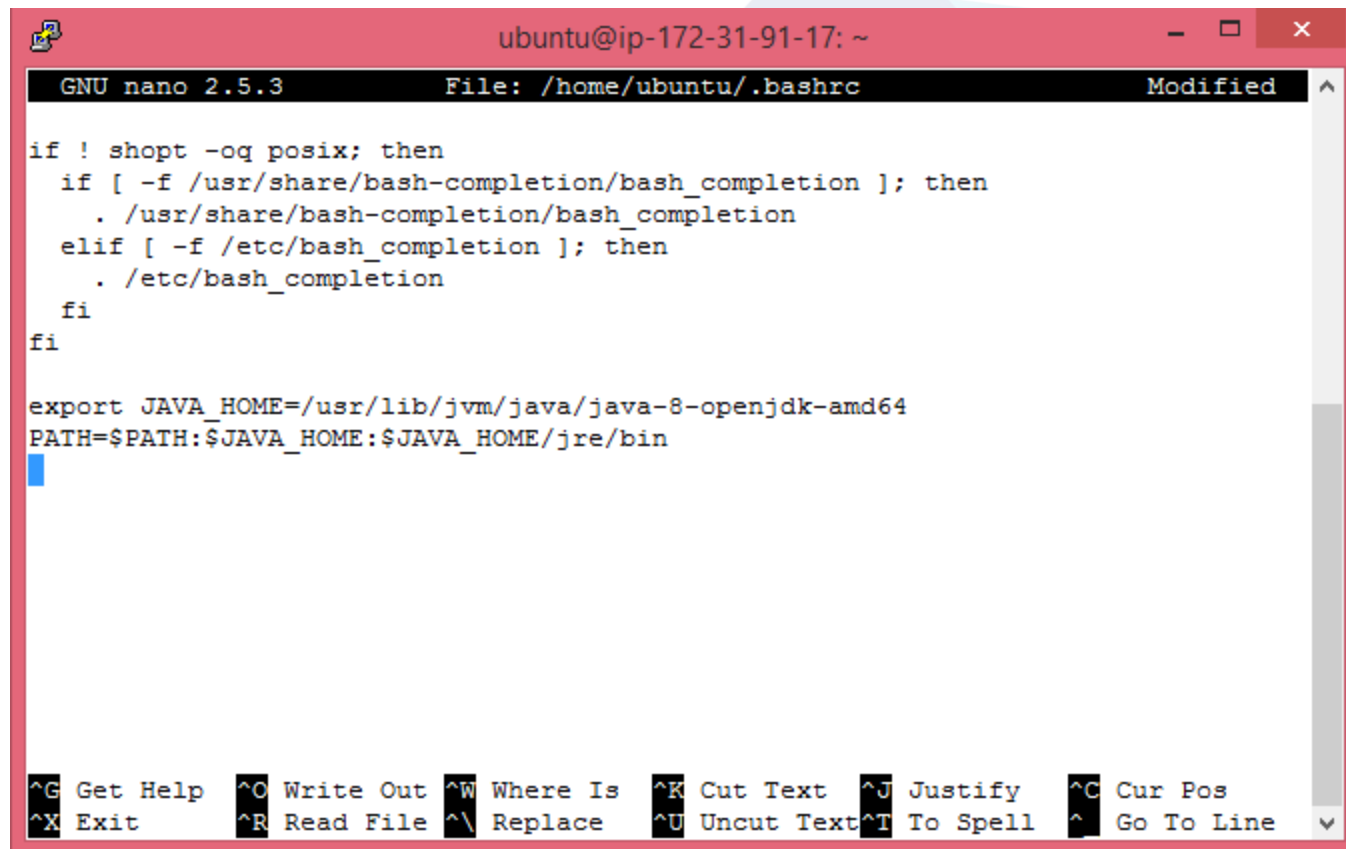
Tambahkan string berikut (sesuaikan dengan direktori java anda):

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-  
amd64  
PATH=$PATH:$JAVA_HOME:$JAVA_HOME/jre/bin
```

TERBUKA
DISABILITAS

LIMITS!

Variabel JAVA_HOME (cont)



```
ubuntu@ip-172-31-91-17: ~  
GNU nano 2.5.3      File: /home/ubuntu/.bashrc      Modified  
  
if ! shopt -oq posix; then  
  if [ -f /usr/share/bash-completion/bash_completion ]; then  
    . /usr/share/bash-completion/bash_completion  
  elif [ -f /etc/bash_completion ]; then  
    . /etc/bash_completion  
  fi  
fi  
  
export JAVA_HOME=/usr/lib/jvm/java/java-8-openjdk-amd64  
PATH=$PATH:$JAVA_HOME:$JAVA_HOME/jre/bin  
  
^G Get Help  ^O Write Out  ^W Where Is  ^K Cut Text  ^J Justify   ^C Cur Pos  
^X Exit      ^R Read File  ^\ Replace   ^U Uncut Text ^T To Spell  ^_ Go To Line
```

Variabel JAVA_HOME (cont)

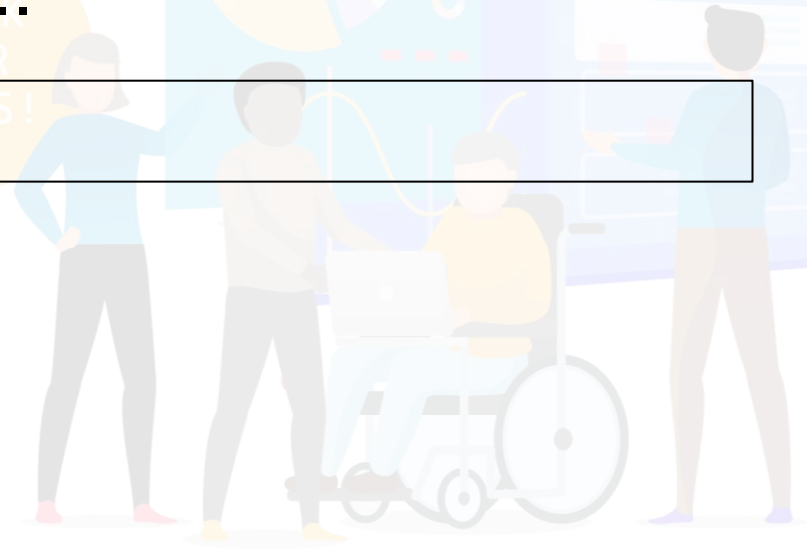
```
$ source ~/.bashrc
```

Atau..

```
$ source ~/.profile
```

TERBUKA
UNTUK
DISABILITAS

BEYOND
YOUR
LIMITS!



Variabel JAVA_HOME (cont)



```
ubuntu@ip-172-31-91-17: ~  
ubuntu@ip-172-31-91-17:~$ source ~/.bashrc  
ubuntu@ip-172-31-91-17:~$
```

Periksa instalasi Java

```
$ java -version
```

TERBUKA
UNTUK
DISABILITAS



```
ubuntu@ip-172-31-91-17: ~  
ubuntu@ip-172-31-91-17:~$ java -version  
openjdk version "1.8.0_212"  
OpenJDK Runtime Environment (build 1.8.0_212-8u212-b03-0ubuntu1.16.04.1-b03)  
OpenJDK 64-Bit Server VM (build 25.212-b03, mixed mode)  
ubuntu@ip-172-31-91-17:~$
```




DIGITAL
TALENT
SCHOLARSHIP

Instalasi Python (Anaconda)

Big Data

Hadoop

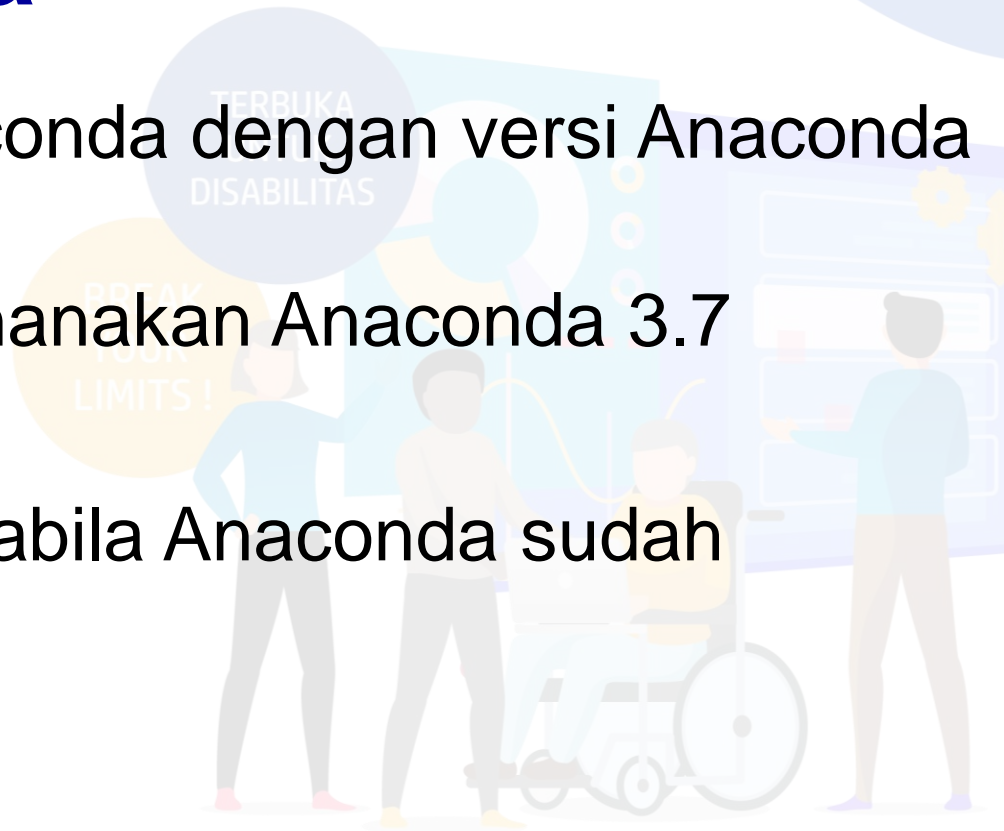
TERBUKA
UNTUK
DISABILITAS

BREAK
YOUR
BARRIER



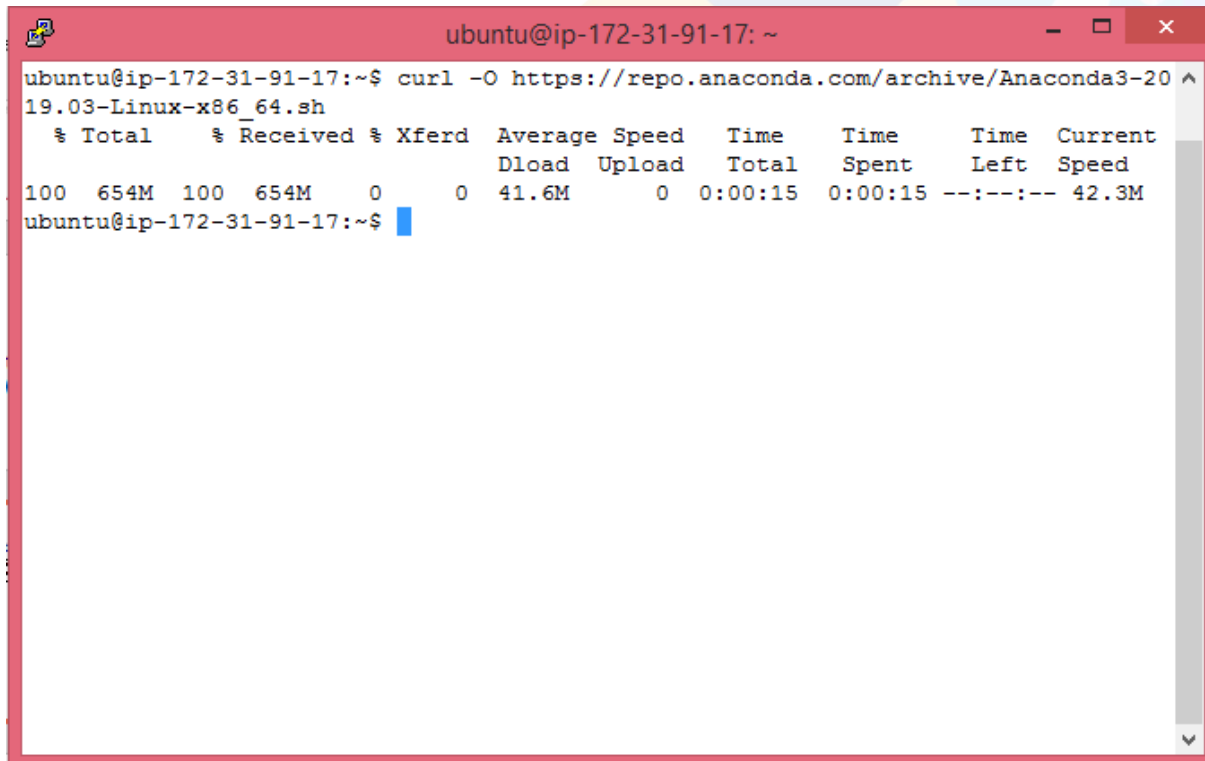
Versi Anaconda

- Sesuaikan versi Anaconda dengan versi Anaconda yang anda pilih.
- Pada tutorial ini digunakan Anaconda 3.7
- Lewati langkah ini apabila Anaconda sudah terinstall



Download Anaconda

```
$ curl -o https://repo.anaconda.com/archive/Anaconda3-2019.03-Linux-x86_64.sh
```



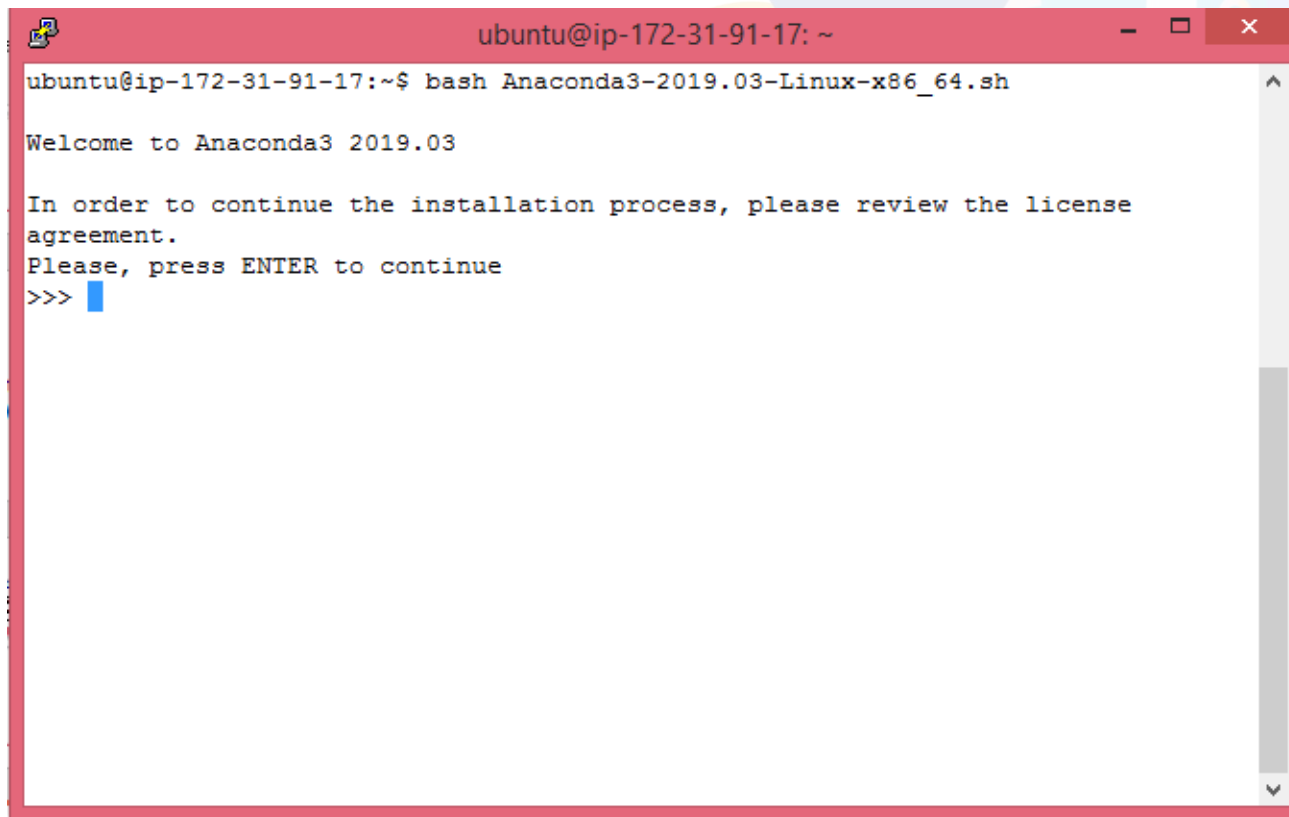
```
ubuntu@ip-172-31-91-17: ~  
ubuntu@ip-172-31-91-17:~$ curl -O https://repo.anaconda.com/archive/Anaconda3-2019.03-Linux-x86_64.sh  
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current  
                                 Dload  Upload   Total   Spent    Left   Speed  
100 654M  100 654M    0     0  41.6M      0  0:00:15  0:00:15 --:--:-- 42.3M  
ubuntu@ip-172-31-91-17:~$
```

UNTUK
DISABILITAS



Install Anaconda

```
$ bash Anaconda3-2019.03-Linux-x86_64.sh
```



```
ubuntu@ip-172-31-91-17: ~  
ubuntu@ip-172-31-91-17:~$ bash Anaconda3-2019.03-Linux-x86_64.sh  
  
Welcome to Anaconda3 2019.03  
  
In order to continue the installation process, please review the license  
agreement.  
Please, press ENTER to continue  
>>> 
```

PERLU
UNTUK
DISABILITAS



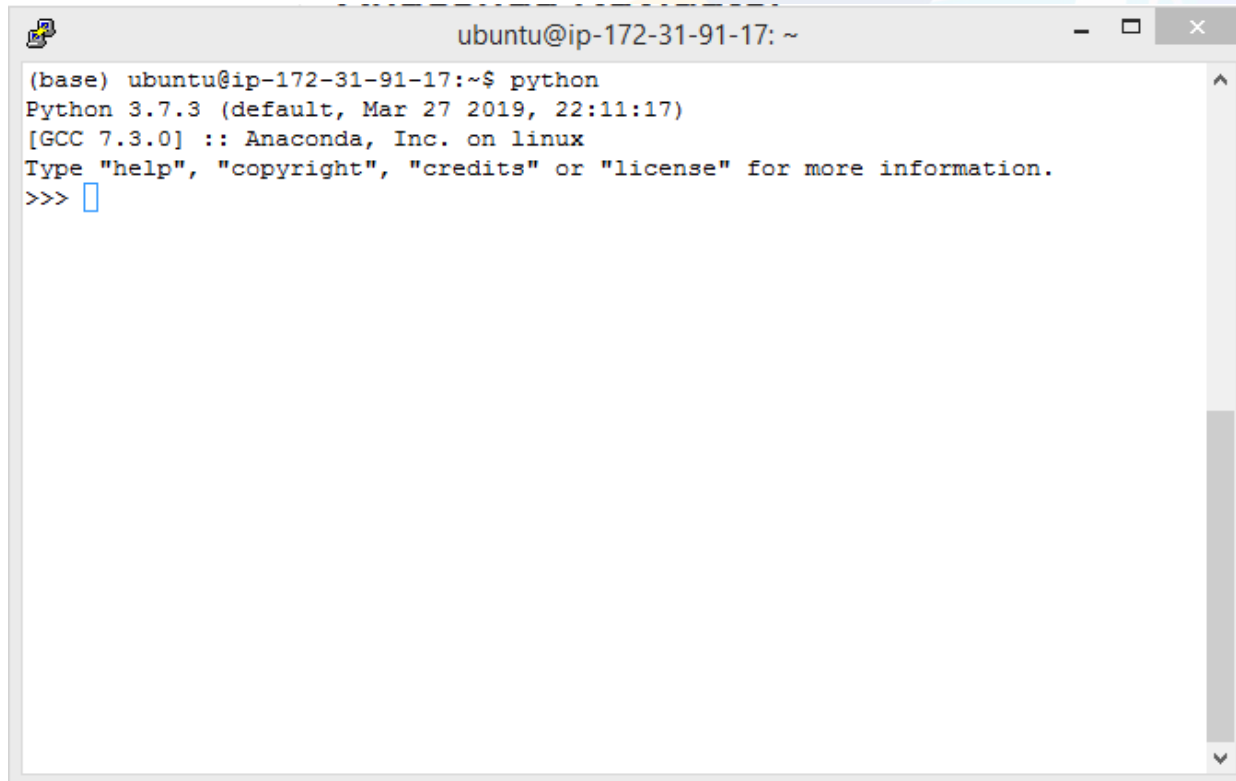
Instalasi Anaconda (cont)

- Sesuaikan konfigurasi instalasi Anaconda anda dengan preferensi anda.
- Pada tutorial ini, Anaconda dipasang pada direktori `/home/ubuntu/anaconda3`
- Jalankan `$conda init` pada direktori `anaconda3/bin` untuk menjalankan inisialisasi Anaconda apabila proses instalasi tidak menjalankan.
- Jalankan `$source ~/.bashrc` jika tidak dapat membuka Anaconda atau python

Python Console

```
$ python
```

TERBUKA
UNTUK
DISABILITAS



```
ubuntu@ip-172-31-91-17: ~  
(base) ubuntu@ip-172-31-91-17:~$ python  
Python 3.7.3 (default, Mar 27 2019, 22:11:17)  
[GCC 7.3.0] :: Anaconda, Inc. on linux  
Type "help", "copyright", "credits" or "license" for more information.  
>>> 
```



DIGITAL
TALENT
SCHOLARSHIP

Konfigurasi SSH

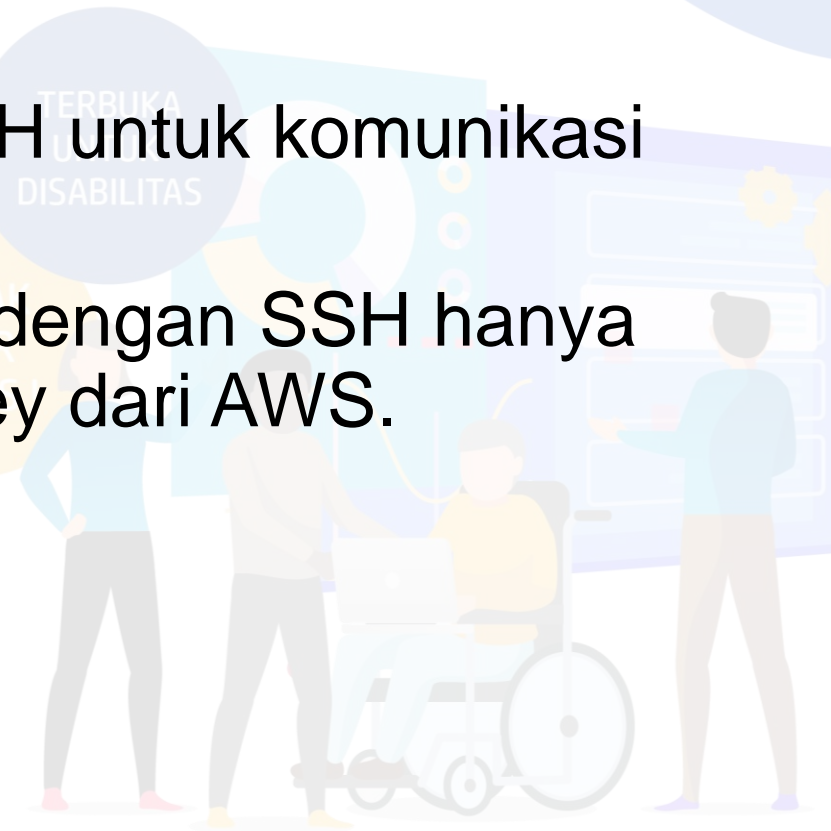
Big Data

Hadoop



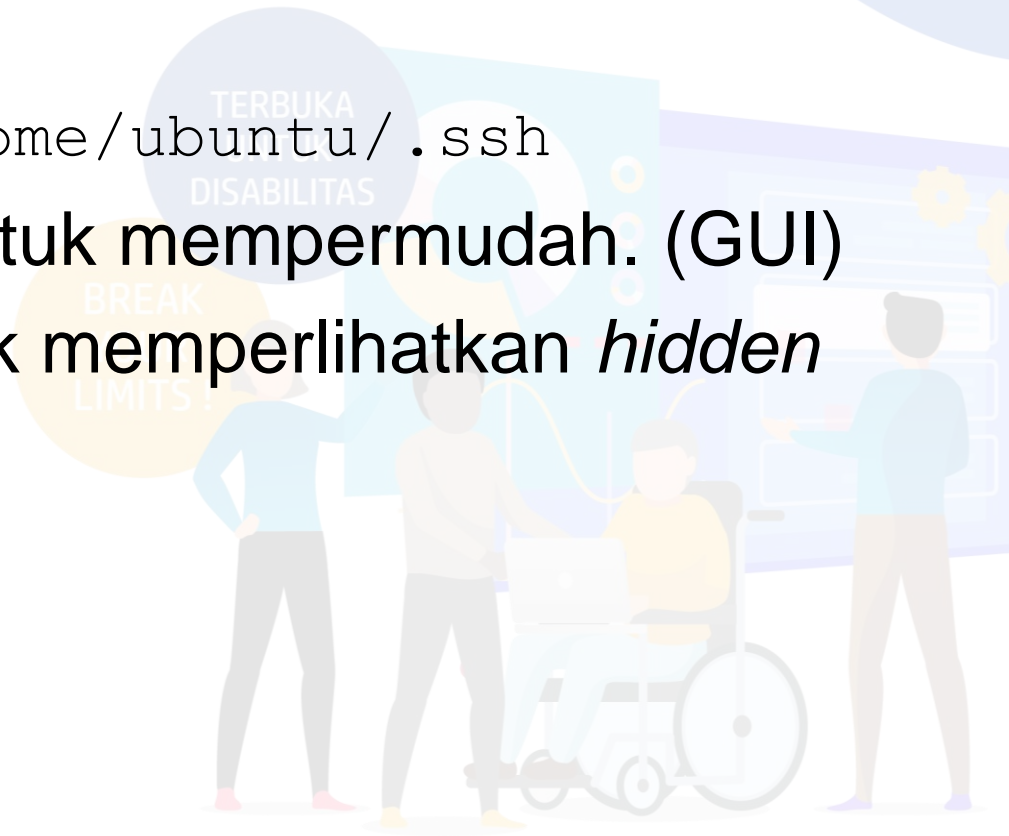
Why?

- Hadoop menggunakan SSH untuk komunikasi antar node.
- Mengakses Instance EC2 dengan SSH hanya dapat dilakukan dengan key dari AWS.

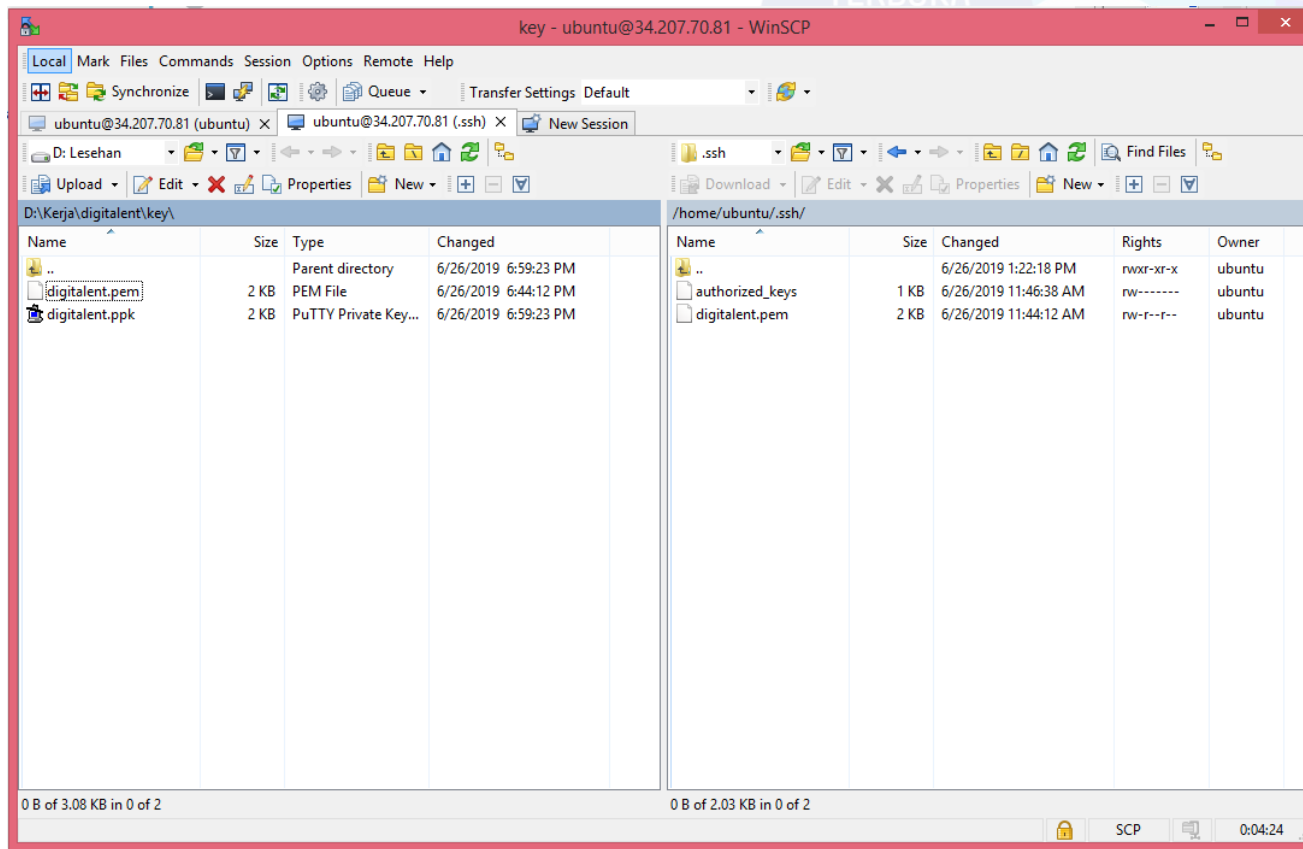


.pem Key

- Salin .pem key ke `/home/ubuntu/.ssh`
- Gunakan WinSCP untuk mempermudah. (GUI)
- Tekan `ctrl+alt+h` untuk memperlihatkan *hidden files*.

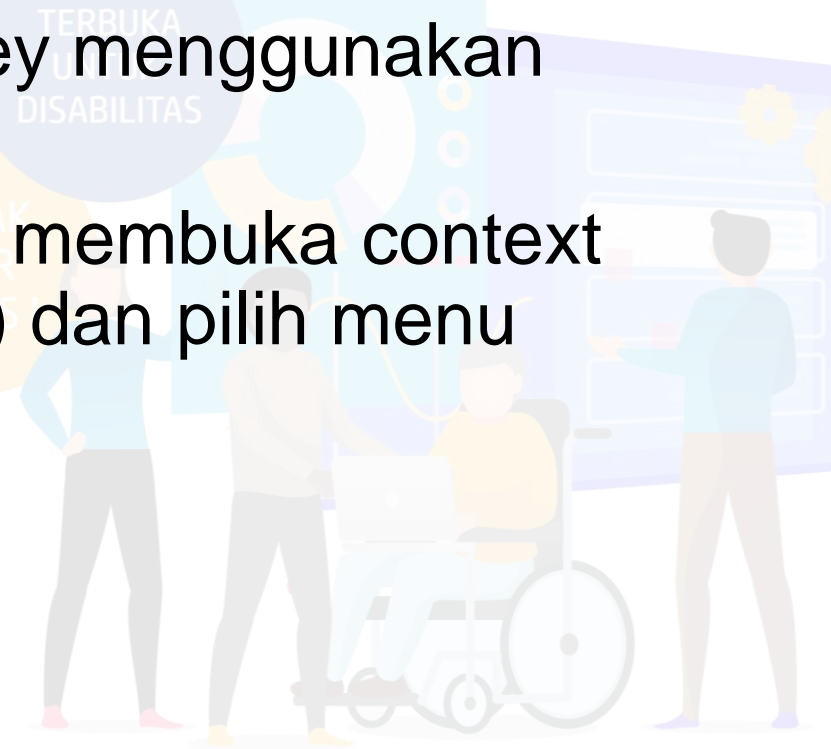


.pem Key (cont)

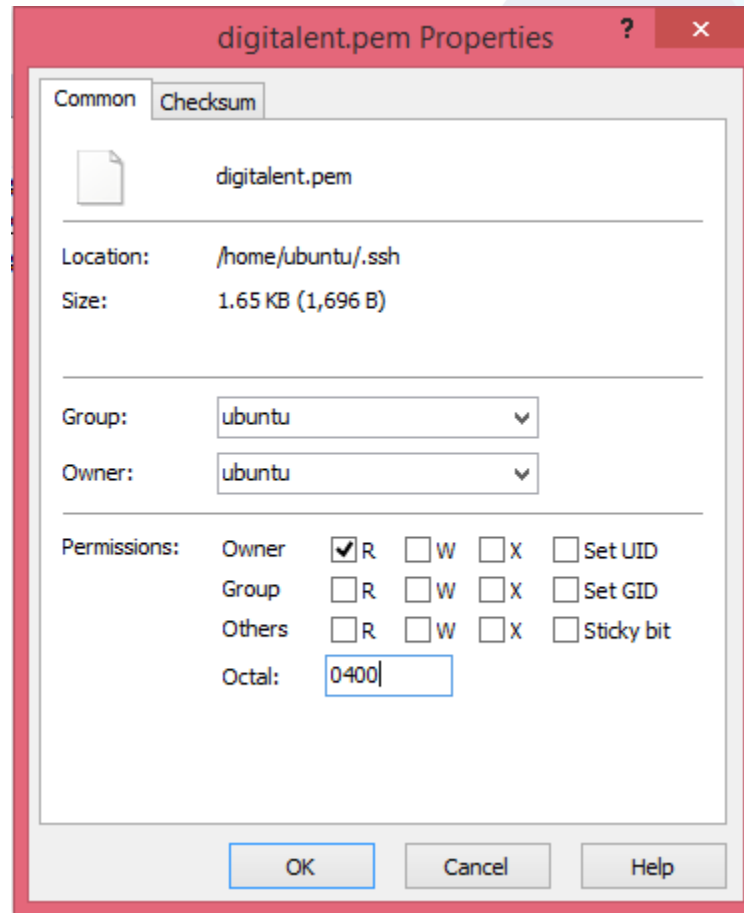


.pem Key (cont)

- Rubah *Permission* .pem key menggunakan Octal 0400.
- Gunakan WinSCP dengan membuka context menu (klik kanan pada file) dan pilih menu properties.



.pem Key (cont)



SSH Config

- Buat file pada direktori `/home/ubuntu/.ssh` dengan nama `config`
- Tambahkan string berikut ke file `config` dan ganti key dengan nama key anda

```
Host localhost
    HostName localhost
    User ubuntu
    IdentityFile ~/.ssh/digitalent.pem
```



DIGITAL
TALENT
SCHOLARSHIP

SSH Config

```
/home/ubuntu/.ssh/config - ubuntu@34.207.70.81 - Editor - WinSCP
Host localhost
  HostName localhost
  User ubuntu
  IdentityFile ~/.ssh/digitalent.pem

Line: 6/5
Encoding: 1252 (ANSI - La Modified)
```



SSH

- Buat koneksi SSH ke localhost untuk memeriksa.

```
$ ssh localhost
```



```
ubuntu@ip-172-31-91-17: ~  
(base) ubuntu@ip-172-31-91-17:~$ ssh localhost  
Welcome to Ubuntu 16.04.6 LTS (GNU/Linux 4.4.0-1083-aws x86_64)  
  
* Documentation:  https://help.ubuntu.com  
* Management:    https://landscape.canonical.com  
* Support:        https://ubuntu.com/advantage  
  
54 packages can be updated.  
29 updates are security updates.  
  
New release '18.04.2 LTS' available.  
Run 'do-release-upgrade' to upgrade to it.  
  
Last login: Wed Jun 26 13:32:07 2019 from 103.119.66.36  
(base) ubuntu@ip-172-31-91-17:~$
```

TERBUKA
UNTUK
DISABILITAS



DIGITAL
TALENT
SCHOLARSHIP

Instalasi Hadoop Single Node

Big Data
Hadoop

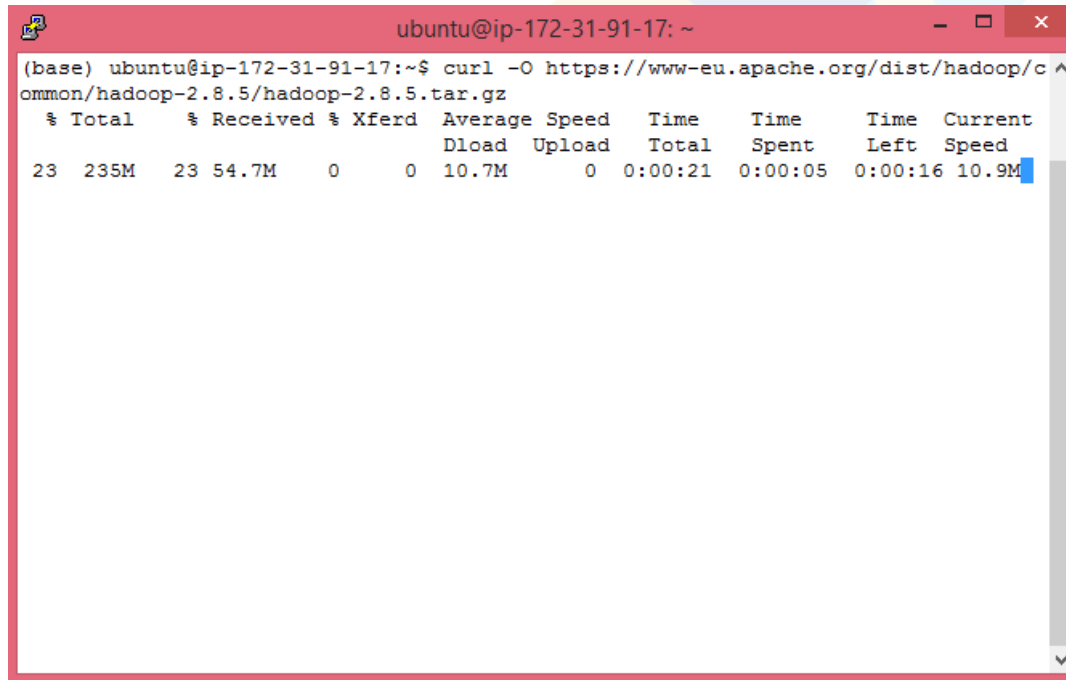
TERBUKA
UNTUK
DISABILITAS

BREAK
YOUR
LIMIT



Download Hadoop

```
$ curl -O https://www-eu.apache.org/dist/hadoop/common/hadoop-2.8.5/hadoop-2.8.5.tar.gz
```

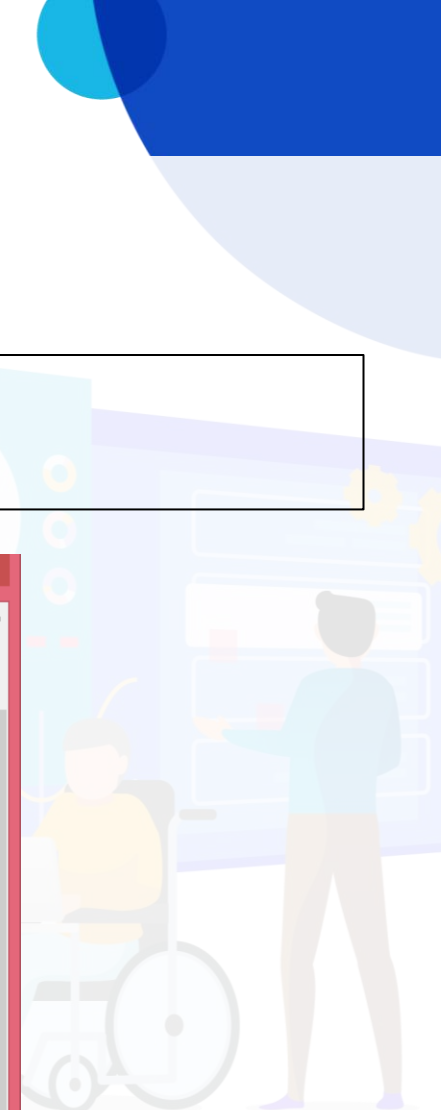


Terminal window output showing the download progress of Hadoop 2.8.5 tar.gz file:

```
ubuntu@ip-172-31-91-17: ~  
(base) ubuntu@ip-172-31-91-17:~$ curl -O https://www-eu.apache.org/dist/hadoop/c  
ommon/hadoop-2.8.5/hadoop-2.8.5.tar.gz
```

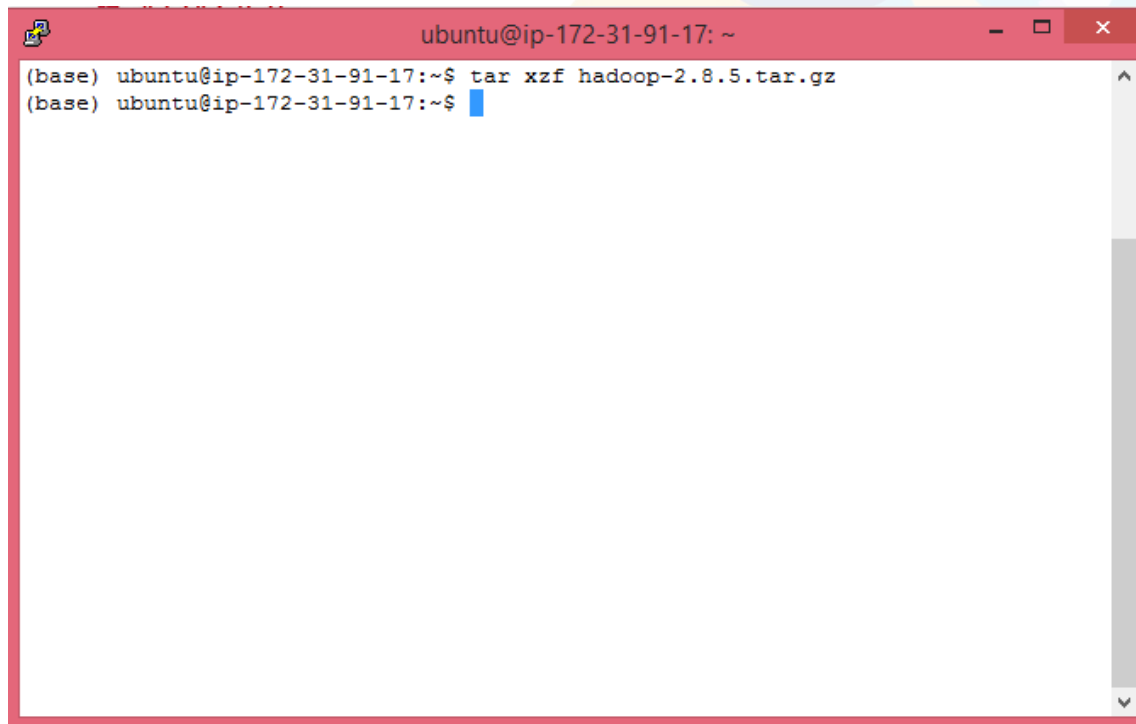
% Total	% Received	% Xferd	Average Speed	Time	Time	Time	Current
			Dload	Upload	Total	Spent	Left
23	235M	23	54.7M	0	0	10.7M	0
					0:00:21	0:00:05	0:00:16
							10.9M

TERBUKA
UNTUK
DISABILITAS



Extract Hadoop

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



```
ubuntu@ip-172-31-91-17: ~  
(base) ubuntu@ip-172-31-91-17:~$ tar xzf hadoop-2.8.5.tar.gz  
(base) ubuntu@ip-172-31-91-17:~$
```

TERBUKA
UNTUK
DISABILITAS

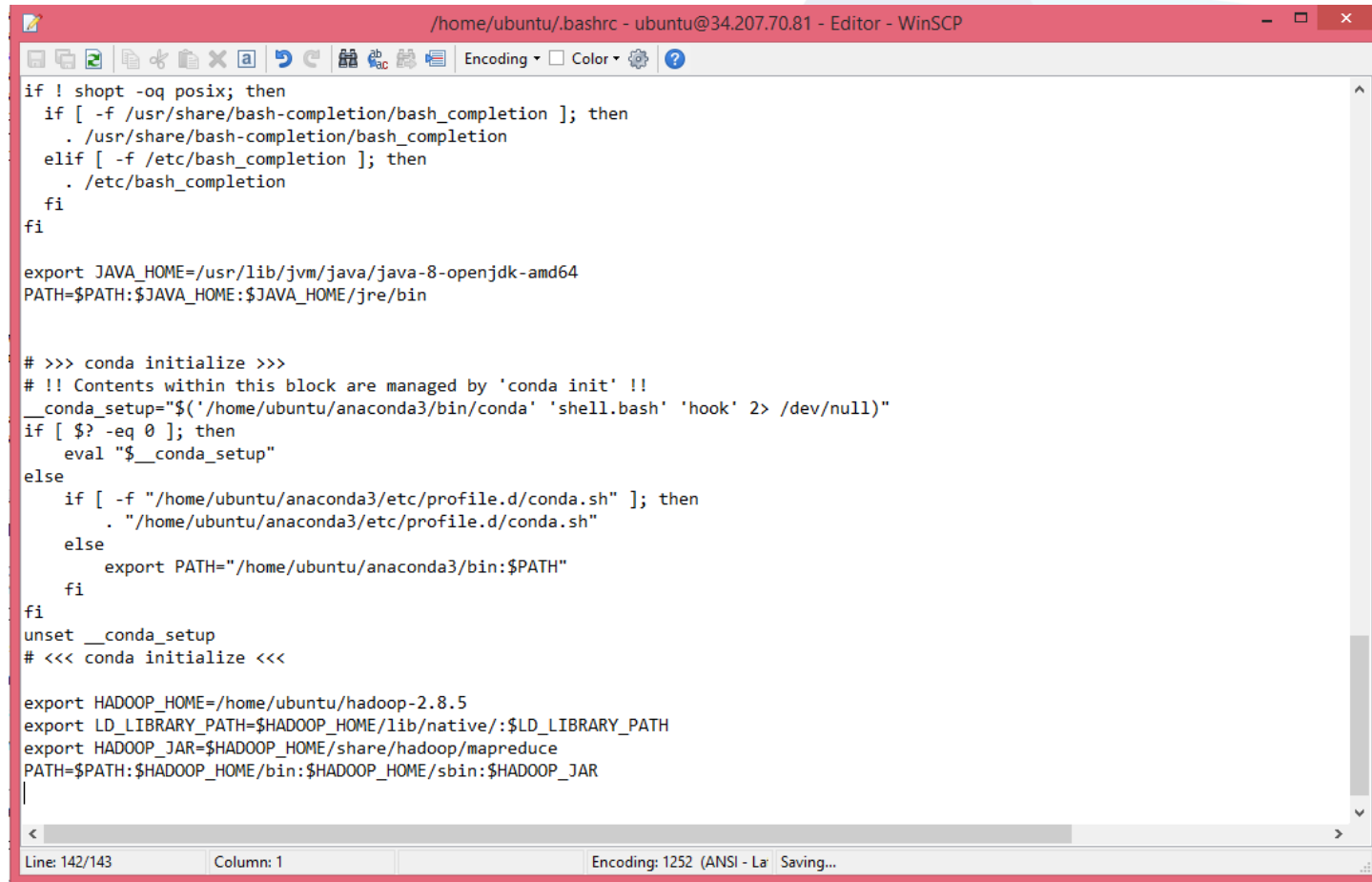


Env Variabel Hadoop

Tambahkan string berikut ke `.bashrc` atau `.profile` (Gunakan WinSCP)

```
export HADOOP_HOME=/home/ubuntu/hadoop-2.8.5
export
LD_LIBRARY_PATH=$HADOOP_HOME/lib/native/:$LD_L
IBRARY_PATH
export
HADOOP_JAR=$HADOOP_HOME/share/hadoop/mapreduce
PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin:
$HADOOP_JAR
```

Env Variabel Hadoop



```
/home/ubuntu/.bashrc - ubuntu@34.207.70.81 - Editor - WinSCP
if ! shopt -oq posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi

export JAVA_HOME=/usr/lib/jvm/java/java-8-openjdk-amd64
PATH=$PATH:$JAVA_HOME:$JAVA_HOME/jre/bin

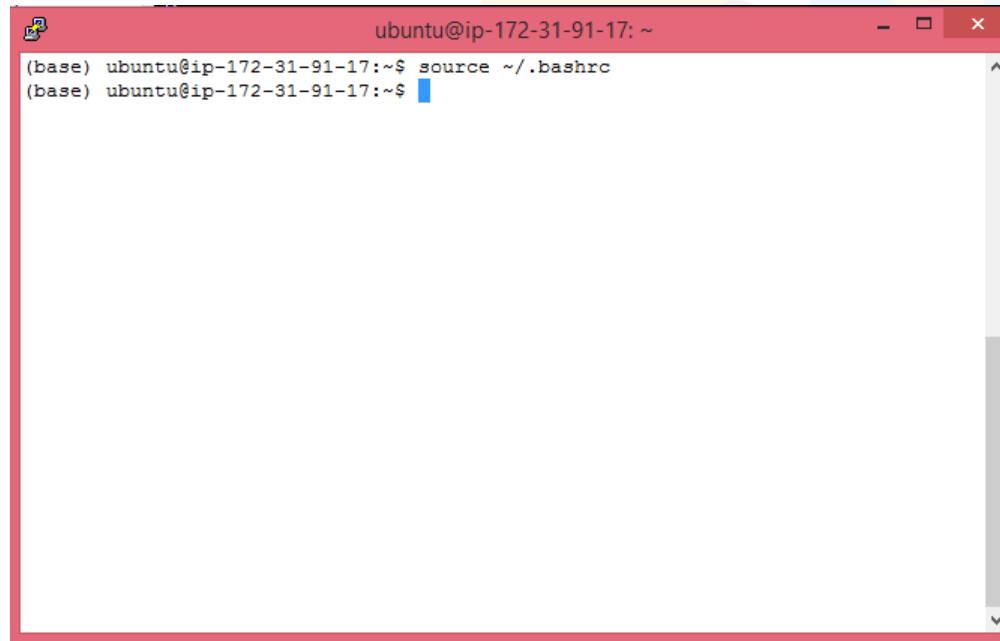
# >>> conda initialize >>>
# !! Contents within this block are managed by 'conda init' !!
__conda_setup="$(('/home/ubuntu/anaconda3/bin/conda' 'shell.bash' 'hook' 2> /dev/null)"
if [ $? -eq 0 ]; then
    eval "$__conda_setup"
else
    if [ -f "/home/ubuntu/anaconda3/etc/profile.d/conda.sh" ]; then
        . "/home/ubuntu/anaconda3/etc/profile.d/conda.sh"
    else
        export PATH="/home/ubuntu/anaconda3/bin:$PATH"
    fi
fi
unset __conda_setup
# <<< conda initialize <<<

export HADOOP_HOME=/home/ubuntu/hadoop-2.8.5
export LD_LIBRARY_PATH=$HADOOP_HOME/lib/native:$LD_LIBRARY_PATH
export HADOOP_JAR=$HADOOP_HOME/share/hadoop/mapreduce
PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$HADOOP_JAR

Line: 142/143      Column: 1      Encoding: 1252 (ANSI - La Saving...
```

Env Variabel Hadoop

Jalankan `$source ~/.bashrc` atau `source ~/.profile`

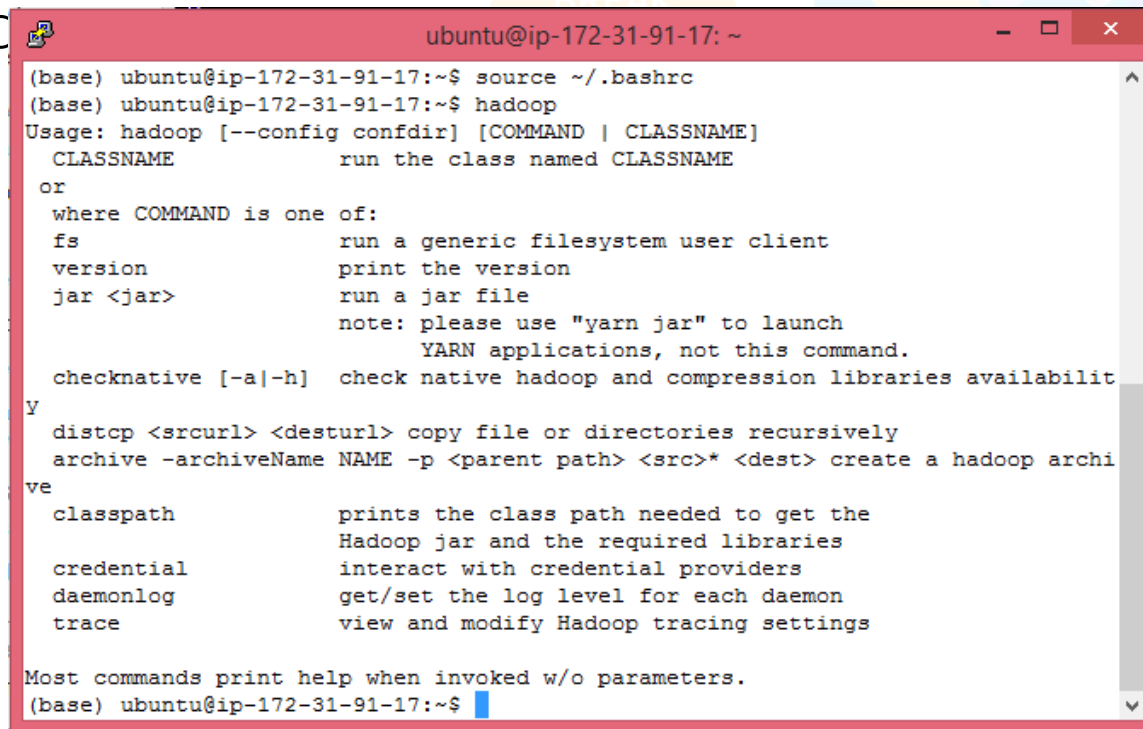


```
ubuntu@ip-172-31-91-17: ~  
(base) ubuntu@ip-172-31-91-17:~$ source ~/.bashrc  
(base) ubuntu@ip-172-31-91-17:~$
```


Env Variabel Hadoop

Coba instalasi hadoop dengan menjalankan perintah

\$ hadoop



```
(base) ubuntu@ip-172-31-91-17:~$ source ~/.bashrc
(base) ubuntu@ip-172-31-91-17:~$ hadoop
Usage: hadoop [--config confdir] [COMMAND | CLASSNAME]
  CLASSNAME                run the class named CLASSNAME
or
  where COMMAND is one of:
  fs                        run a generic filesystem user client
  version                  print the version
  jar <jar>                run a jar file
                           note: please use "yarn jar" to launch
                           YARN applications, not this command.
  checknative [-a|-h]      check native hadoop and compression libraries availability
  distcp <srcurl> <desturl> copy file or directories recursively
  archive -archiveName NAME -p <parent path> <src>* <dest> create a hadoop archive
  classpath                prints the class path needed to get the
                           Hadoop jar and the required libraries
  credential               interact with credential providers
  daemonlog                get/set the log level for each daemon
  trace                    view and modify Hadoop tracing settings

Most commands print help when invoked w/o parameters.
(base) ubuntu@ip-172-31-91-17:~$
```



DIGITAL
TALENT
SCHOLARSHIP

Konfigurasi Hadoop Single Node

Big Data
Hadoop

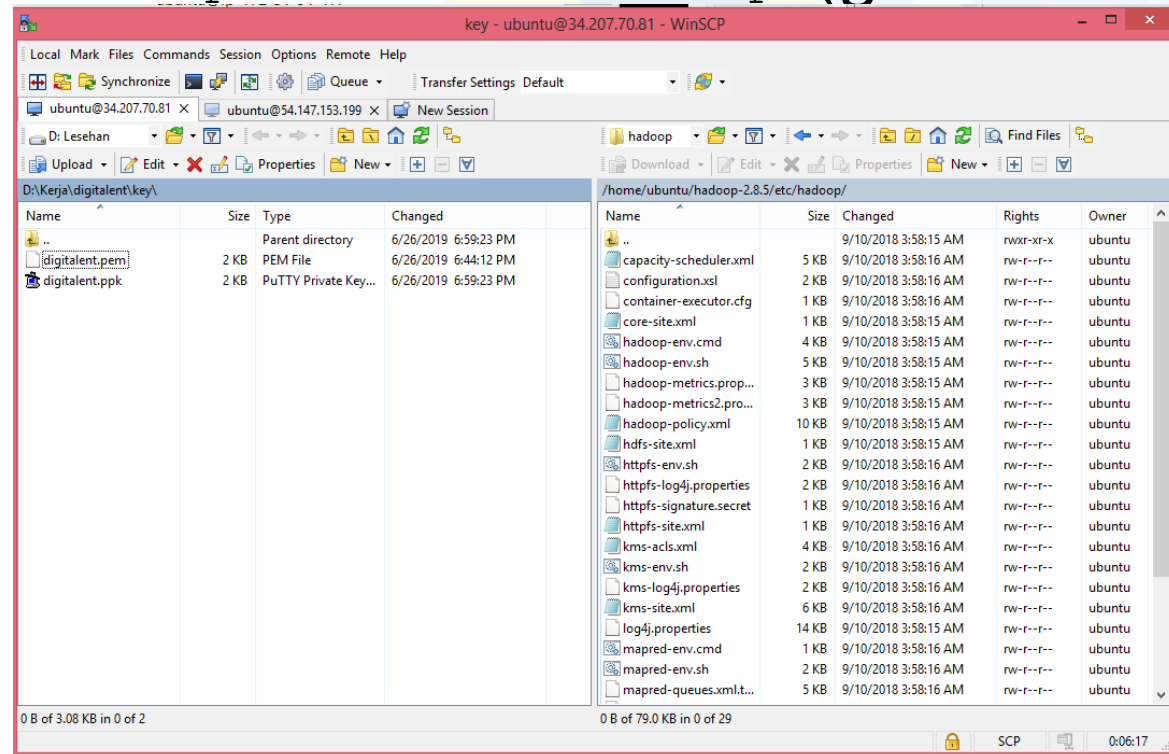
TERBUKA
UNTUK
DISABILITAS

BREAK
YOUR
LIMIT



Konfigurasi Hadoop

Buka direktori konfigurasi hadoop pada
/home/ubuntu/hadoop-2.8.5/etc/hadoop (gunakan
WinSCP)

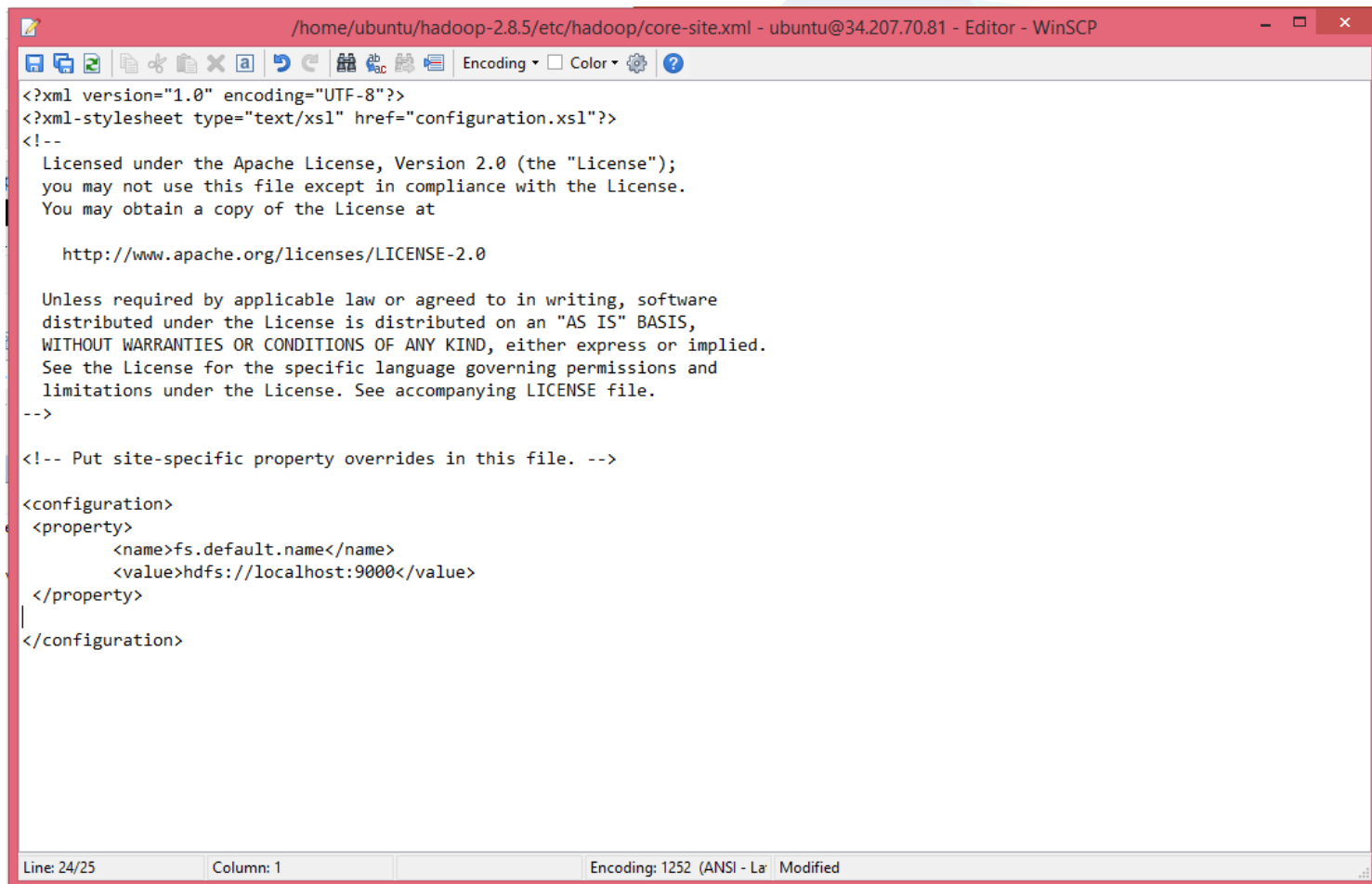


core-site.xml

Buka file `core-site.xml` dan tambahkan string berikut pada tag `<configuration>`

```
<property>
    <name>fs.default.name</name>
    <value>hdfs://localhost:8020</value>
</property>
```

core-site.xml



```
<?xml version="1.0" encoding="UTF-8"?>
<?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>fs.default.name</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

core-site.xml

Buka file `hdfs-site.xml` dan tambahkan string berikut pada tag `<configuration>`

```
<property>
  <name>dfs.namenode.name.dir</name>
  <value>/home/ubuntu/data/nameNode</value>
</property>
<property>
  <name>dfs.datanode.data.dir</name>
  <value>/home/ubuntu/data/dataNode</value>
</property>
<property>
  <name>dfs.replication</name>
  <value>2</value>
</property>
```

TERBUKA
DISABILITAS
BREAK
YOUR
LIMITS!



hdfs-site.xml

```
/home/ubuntu/hadoop-2.8.5/etc/hadoop/hdfs-site.xml - ubuntu@34.207.70.81 - Editor - WinSCP
<?xml version="1.0" encoding="UTF-8"?>
<?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>dfs.namenode.name.dir</name>
    <value>/home/ubuntu/data/nameNode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>/home/ubuntu/data/dataNode</value>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>2</value>
  </property>
</configuration>
```



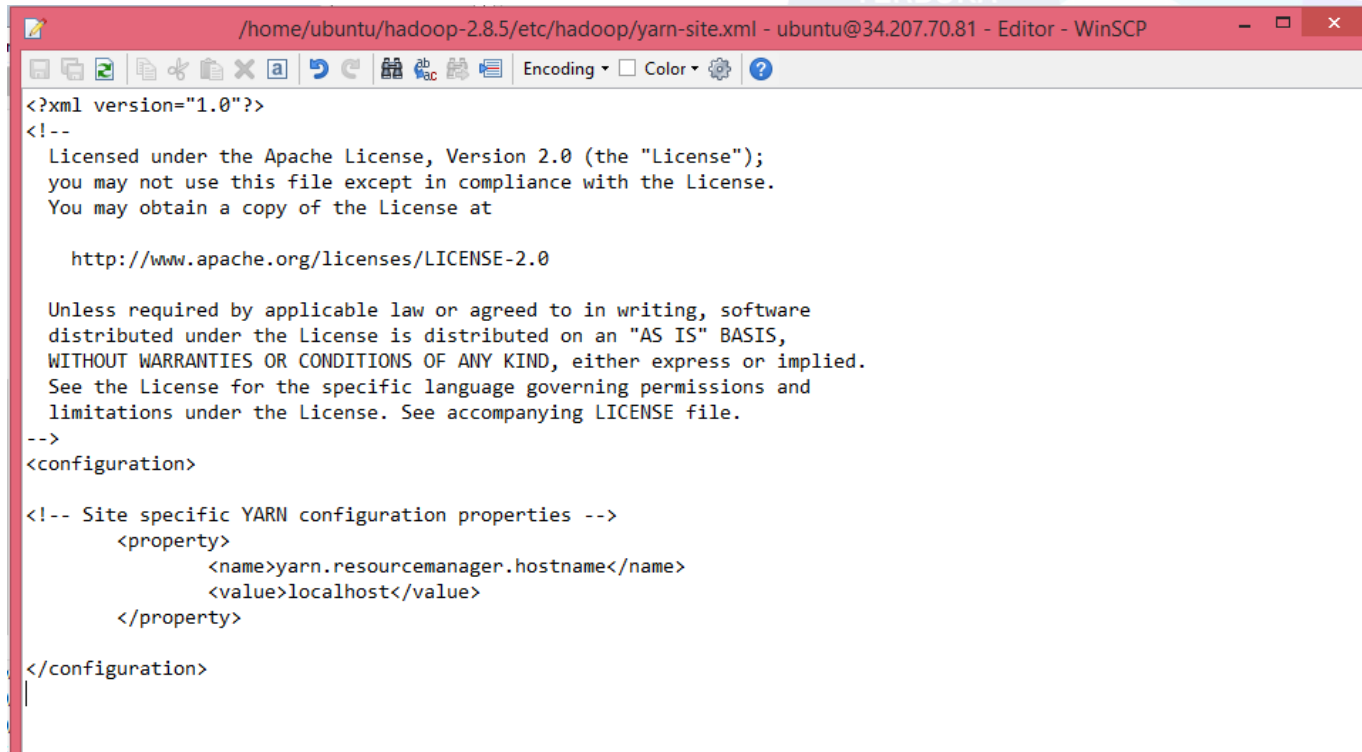
yarn-site.xml

Buka file `yarn-site.xml` dan tambahkan string berikut pada tag `<configuration>`

```
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>localhost</value>
</property>
```


yarn-site.xml

TERBUKA



```
<?xml version="1.0"?>
<!--
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.
-->
<configuration>

<!-- Site specific YARN configuration properties -->
  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>localhost</value>
  </property>
</configuration>
```

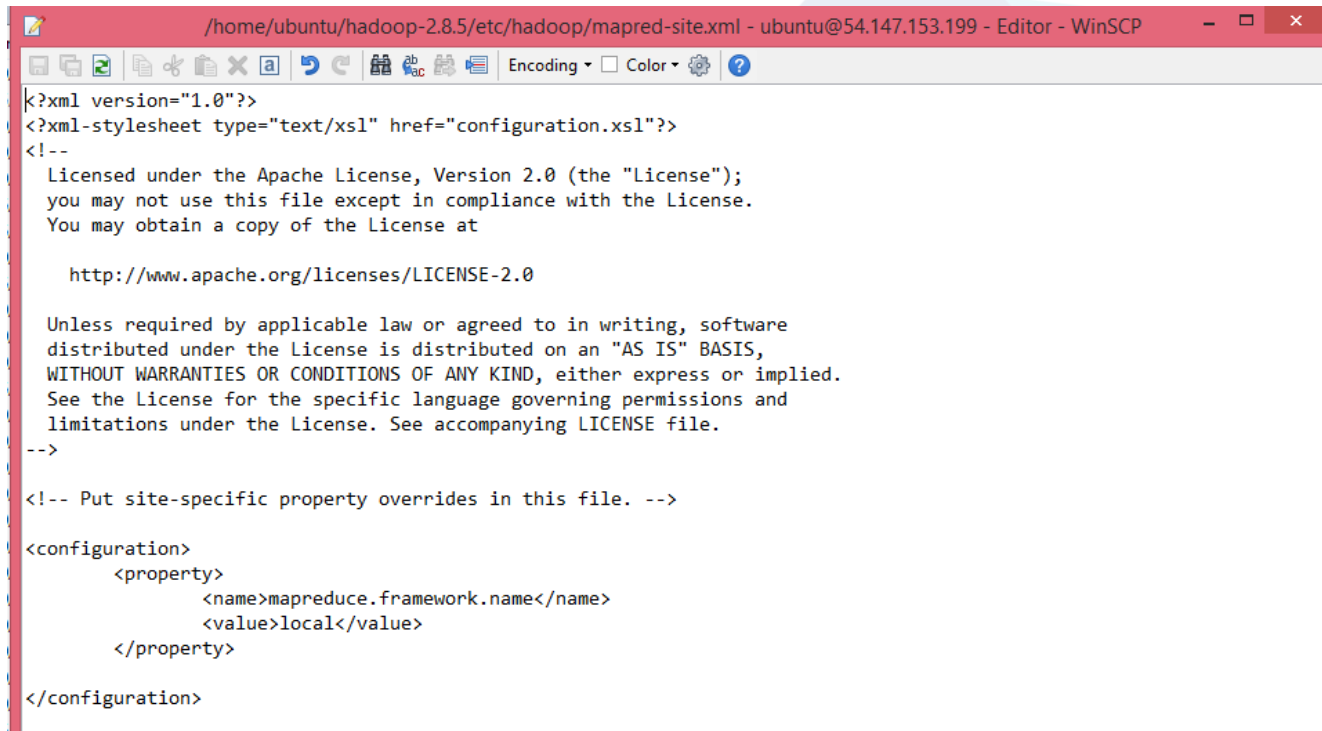


mapred-site.xml

copy file `mapred-site.xml.template` ke `mapred-site.xml` dan tambahkan string berikut pada tag `<configuration>`

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

mapred-site.xml



```
<?xml version="1.0"?>
<?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>mapreduce.framework.name</name>
    <value>local</value>
  </property>
</configuration>
```

mapred-site.xml

Apabila anda mengkonfigurasi Hadoop multinode maka gunakan nilai yarn pada property `mapreduce.framework.name` sehingga YARN digunakan sebagai resource manager.

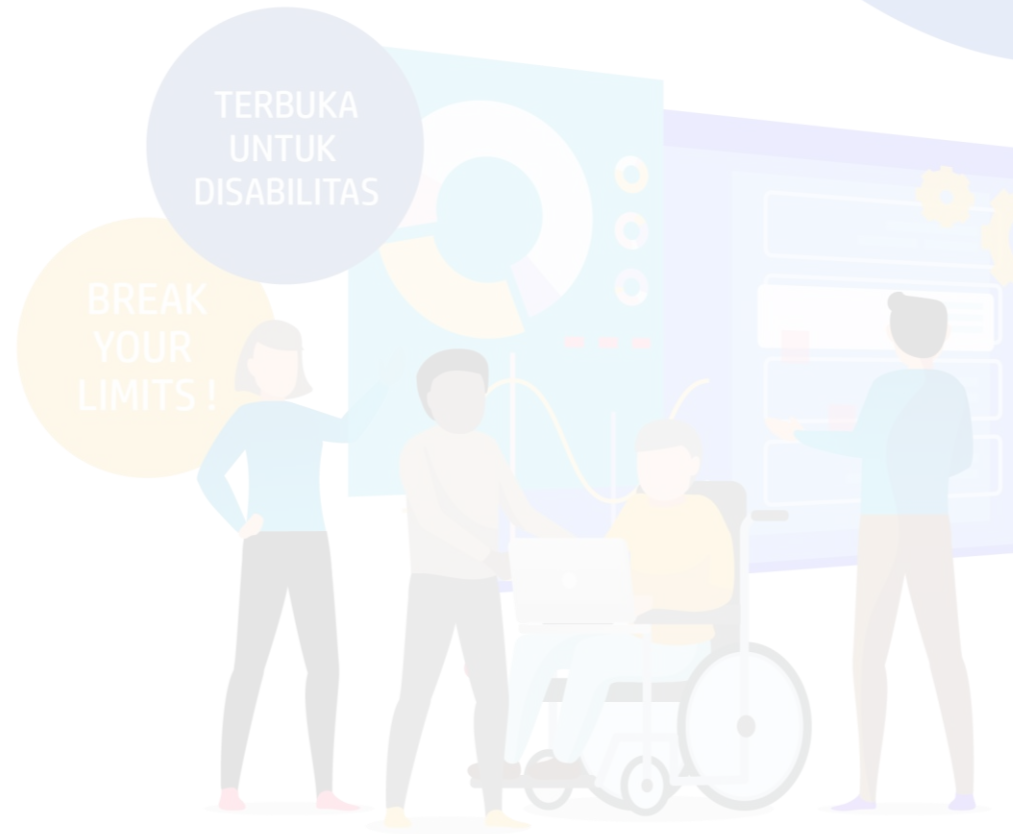
```
<property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
</property>
```



DIGITAL
TALENT
SCHOLARSHIP

HDFS

Big Data
Hadoop



Format HDFS

\$ `hdfs namenode -format`

TERBUKA
UNTUK

```
ubuntu@ip-172-31-91-17: ~  
(base) ubuntu@ip-172-31-91-17:~$ ls  
anaconda3  hadoop-2.8.5  hadoop-2.8.5.tar.gz  
(base) ubuntu@ip-172-31-91-17:~$ clear  
(base) ubuntu@ip-172-31-91-17:~$ hdfs namenode -format  
19/06/26 14:39:48 INFO namenode.NameNode: STARTUP_MSG:  
/*****  
STARTUP_MSG: Starting NameNode  
STARTUP_MSG:   user = ubuntu  
STARTUP_MSG:   host = ip-172-31-91-17.ec2.internal/172.31.91.17  
STARTUP_MSG:   args = [-format]  
STARTUP_MSG:   version = 2.8.5  
STARTUP_MSG:   classpath = /home/ubuntu/hadoop-2.8.5/etc/hadoop:/home/ubuntu/had  
oop-2.8.5/share/hadoop/common/lib/jetty-sslengine-6.1.26.jar:/home/ubuntu/hadoop  
-2.8.5/share/hadoop/common/lib/hamcrest-core-1.3.jar:/home/ubuntu/hadoop-2.8.5/s  
hare/hadoop/common/lib/commons-io-2.4.jar:/home/ubuntu/hadoop-2.8.5/share/hadoop  
/common/lib/log4j-1.2.17.jar:/home/ubuntu/hadoop-2.8.5/share/hadoop/common/lib/s  
tax-api-1.0-2.jar:/home/ubuntu/hadoop-2.8.5/share/hadoop/common/lib/jersey-json-  
1.9.jar:/home/ubuntu/hadoop-2.8.5/share/hadoop/common/lib/slf4j-log4j12-1.7.10.j  
ar:/home/ubuntu/hadoop-2.8.5/share/hadoop/common/lib/jackson-mapper-asl-1.9.13.j  
ar:/home/ubuntu/hadoop-2.8.5/share/hadoop/common/lib/mockito-all-1.8.5.jar:/home  
/ubuntu/hadoop-2.8.5/share/hadoop/common/lib/httpclient-4.5.2.jar:/home/ubuntu/h  
adoop-2.8.5/share/hadoop/common/lib/commons-collections-3.2.2.jar:/home/ubuntu/h  
adoop-2.8.5/share/hadoop/common/lib/jaxb-impl-2.2.3-1.jar:/home/ubuntu/hadoop-2.  
8.5/share/hadoop/common/lib/commons-configuration-1.6.jar:/home/ubuntu/hadoop-2.
```

Format HDFS

- Pastikan direktori data telah terbentuk pada /home/ubuntu

/home/ubuntu/					
Name	Size	Changed	Rights	Owner	
..		6/26/2019 11:46:38 AM	rwxr-xr-x	root	
.cache		6/26/2019 12:07:16 PM	rwX-----	ubuntu	
.nano		6/26/2019 12:44:33 PM	rwXrwXr-x	ubuntu	
.ssh		6/26/2019 1:32:20 PM	rwX-----	ubuntu	
anaconda3		6/26/2019 1:04:07 PM	rwXrwXr-x	ubuntu	
data		6/26/2019 2:39:49 PM	rwXrwXr-x	ubuntu	
hadoop-2.8.5		9/10/2018 3:58:16 AM	rwXr-xr-x	ubuntu	
.bash_history	2 KB	6/26/2019 2:36:40 PM	rw-----	ubuntu	
.bash_logout	1 KB	8/31/2015 11:27:45 PM	rw-r--r--	ubuntu	
.bashrc	5 KB	6/26/2019 2:37:49 PM	rw-r--r--	ubuntu	
.profile	1 KB	5/16/2017 12:49:38 PM	rw-r--r--	ubuntu	
.python_history	0 KB	6/26/2019 1:12:07 PM	rw-----	ubuntu	
.sudo_as_admin_succ...	0 KB	6/26/2019 12:22:55 PM	rw-r--r--	ubuntu	
hadoop-2.8.5.tar.gz	240,766 KB	6/26/2019 2:00:16 PM	rw-rw-r--	ubuntu	



TEKNIKA
UNTUK
DISABILITAS

HDFS : Start Service

- Jalankan service HDFS dengan menjalankan file `start-dfs.sh`

```
ubuntu@ip-172-31-91-17: ~  
(base) ubuntu@ip-172-31-91-17:~$ stop-dfs.sh  
Stopping namenodes on [localhost]  
localhost: stopping namenode  
localhost: stopping datanode  
Stopping secondary namenodes [0.0.0.0]  
The authenticity of host '0.0.0.0 (0.0.0.0)' can't be established.  
ECDSA key fingerprint is SHA256:sX9tw3Un/2JHo3mSzApVaIT1ygW41aqWn/fevWvJbOY.  
Are you sure you want to continue connecting (yes/no)? no  
0.0.0.0: Host key verification failed.  
(base) ubuntu@ip-172-31-91-17:~$
```

Introduction Data Replication Redundancy Data Organization

HDFS : Membuat Direktori

- Buat direktori user anda dengan perintah berikut.

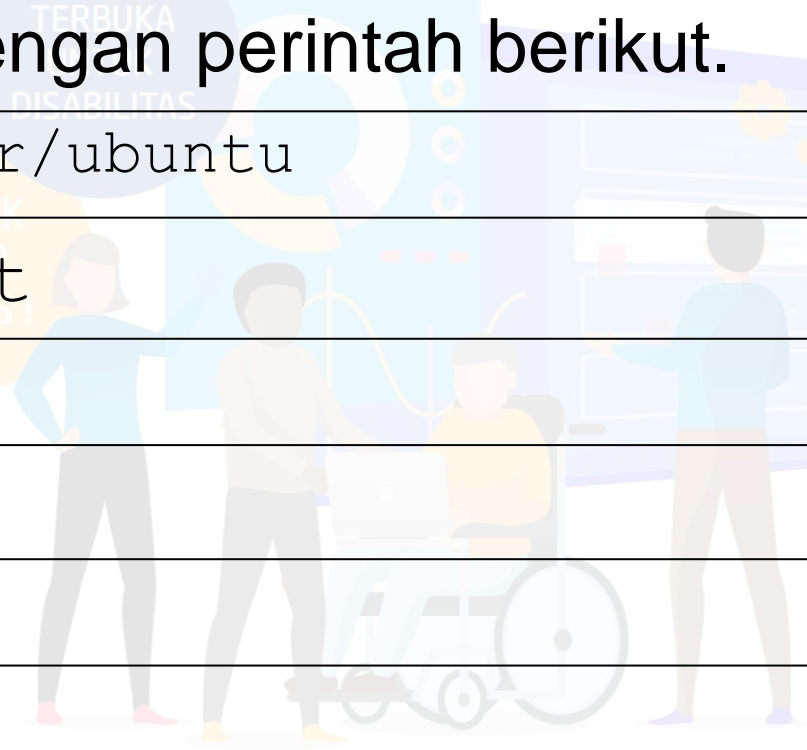
```
$ hdfs dfs -mkdir -p /user/ubuntu
```

Buat direktori bernama input

```
$ hdfs dfs -mkdir input
```

Melihat direktori

```
$ hdfs dfs -ls
```



HDFS : Membuat Direktori



```
ubuntu@ip-172-31-91-17: ~  
(base) ubuntu@ip-172-31-91-17:~$ hdfs dfs -mkdir -p /user/ubuntu  
(base) ubuntu@ip-172-31-91-17:~$ hdfs dfs -mkdir input  
(base) ubuntu@ip-172-31-91-17:~$ hdfs dfs -ls  
Found 1 items  
drwxr-xr-x  - ubuntu supergroup          0 2019-06-26 15:15 input  
(base) ubuntu@ip-172-31-91-17:~$
```

Mengunggah ke HDFS

- Buat sebuah file dengan nama `file01`.
- Isikan file dengan string berikut

Hello World Goodbye World

Big Data Analytics

Digitalent 2019

Universitas Brawijaya

TERBUKA
UNTUK
DISABILITAS

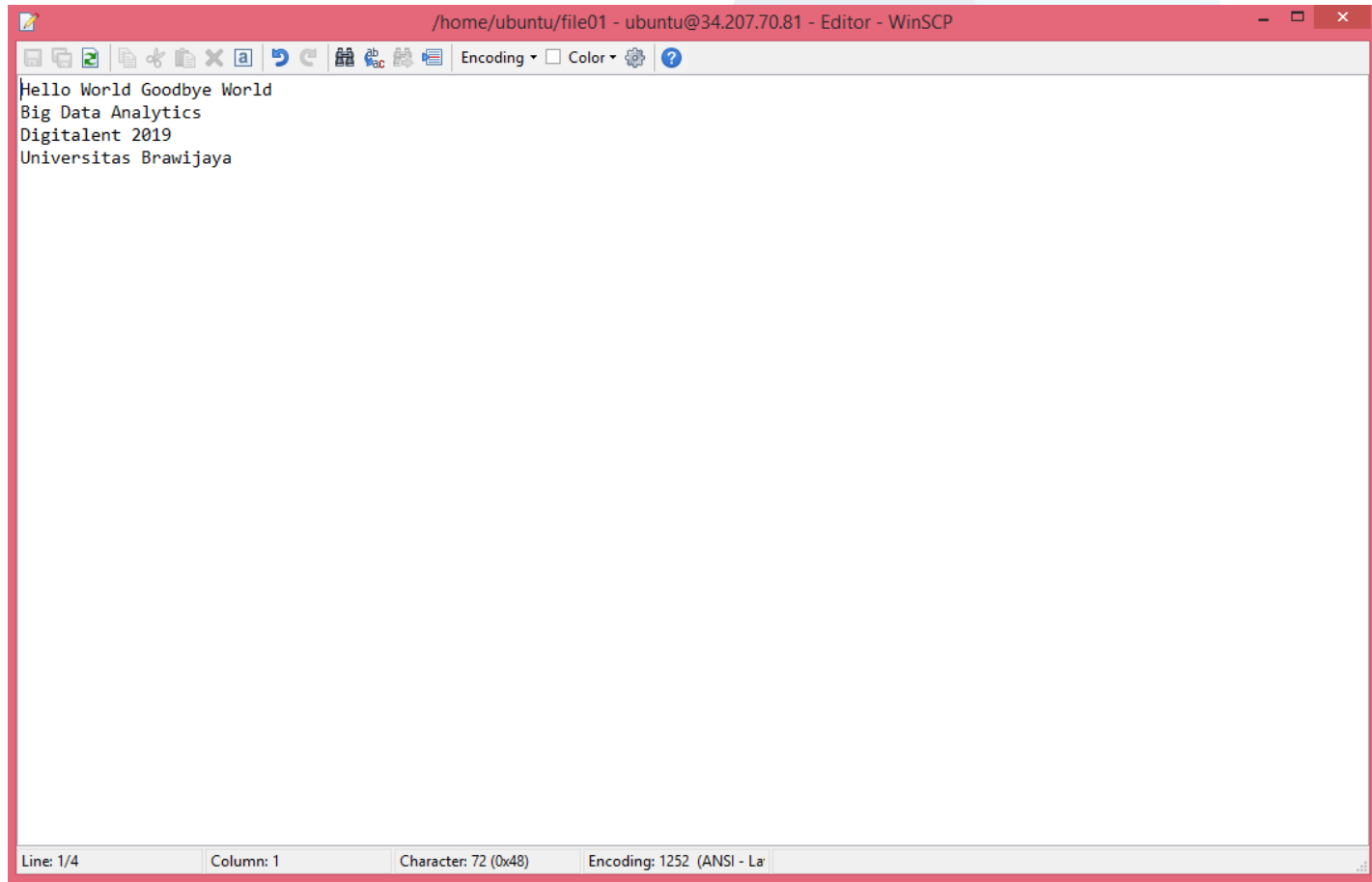
OVER
OUR
LIMITS!





DIGITAL
TALENT
SCHOLARSHIP

Mengunggah ke HDFS



```

/home/ubuntu/file01 - ubuntu@34.207.70.81 - Editor - WinSCP
Hello World Goodbye World
Big Data Analytics
Digitalent 2019
Universitas Brawijaya
Line: 1/4 Column: 1 Character: 72 (0x48) Encoding: 1252 (ANSI - La
```



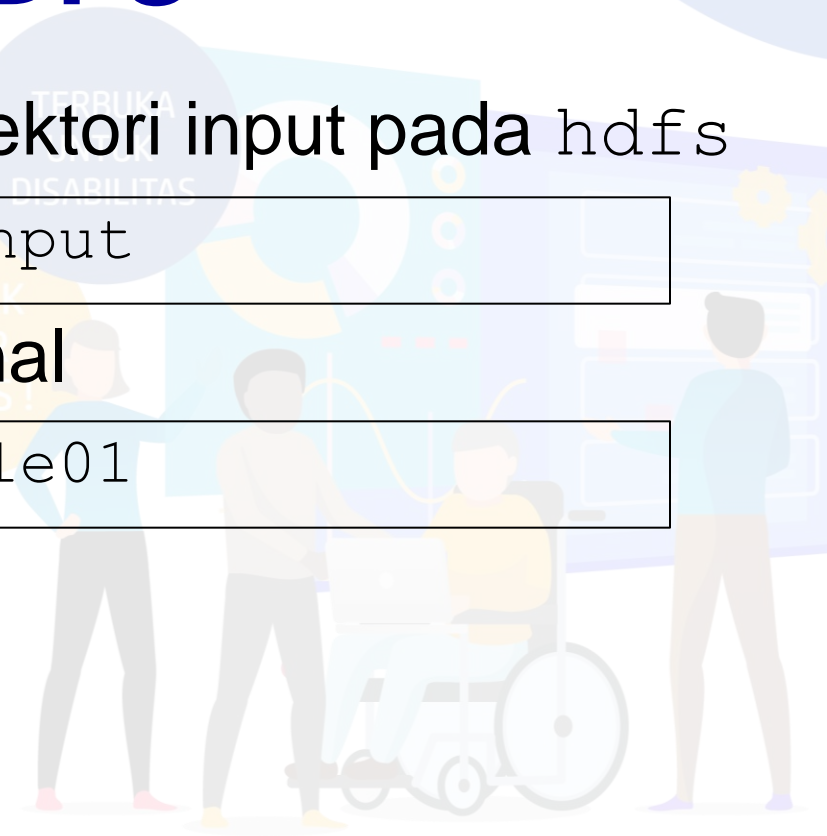
Mengunggah ke HDFS

- Unggah file `file01` ke direktori input pada `hdfs`

```
$ hdfs dfs -put file01 input
```

- Mecetak `file01` ke terminal

```
$ hdfs dfs -cat input/file01
```





DIGITAL
TALENT
SCHOLARSHIP

Mengunggah ke HDFS

```
ubuntu@ip-172-31-91-17: ~  
(base) ubuntu@ip-172-31-91-17:~$ hdfs dfs -put file01 input  
(base) ubuntu@ip-172-31-91-17:~$ hdfs dfs -cat input/file01  
Hello World Goodbye World  
Big Data Analytics  
Digitalent 2019  
Universitas Brawijaya  
(base) ubuntu@ip-172-31-91-17:~$
```



MapReduce (Streaming API)

Big Data

Hadoop

BREAK
YOUR
LIMITS!

TERBUKA
UNTUK
DISABILITAS



Python Wordcount

- Jalankan perintah berikut untuk mengunduh mapper program wordcount python menggunakan hadoop

```
streaming api  
$ curl -O
```

```
https://raw.githubusercontent.com/edho08/digital-talent-worcount/master/mapper.py
```

- Jalankan perintah berikut untuk mengunduh reducer

```
$ curl -O
```

```
https://raw.githubusercontent.com/edho08/digital-talent-worcount/master/reducer.py
```


Python Wordcount

```
ubuntu@ip-172-31-91-17: ~  
(base) ubuntu@ip-172-31-91-17:~$ curl -O https://raw.githubusercontent.com/edho08/digitalent-worcount/master/mapper.py  
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current  
           Dload  Upload   Total   Spent    Left   Speed  
100  211  100  211    0    0   2604      0 --:--:-- --:--:-- --:--:--  2604  
(base) ubuntu@ip-172-31-91-17:~$ curl -O https://raw.githubusercontent.com/edho08/digitalent-worcount/master/reducer.py  
% Total    % Received % Xferd  Average Speed   Time    Time     Time  Current  
           Dload  Upload   Total   Spent    Left   Speed  
100  484  100  484    0    0   6914      0 --:--:-- --:--:-- --:--:--  6914  
(base) ubuntu@ip-172-31-91-17:~$
```



Python Wordcount

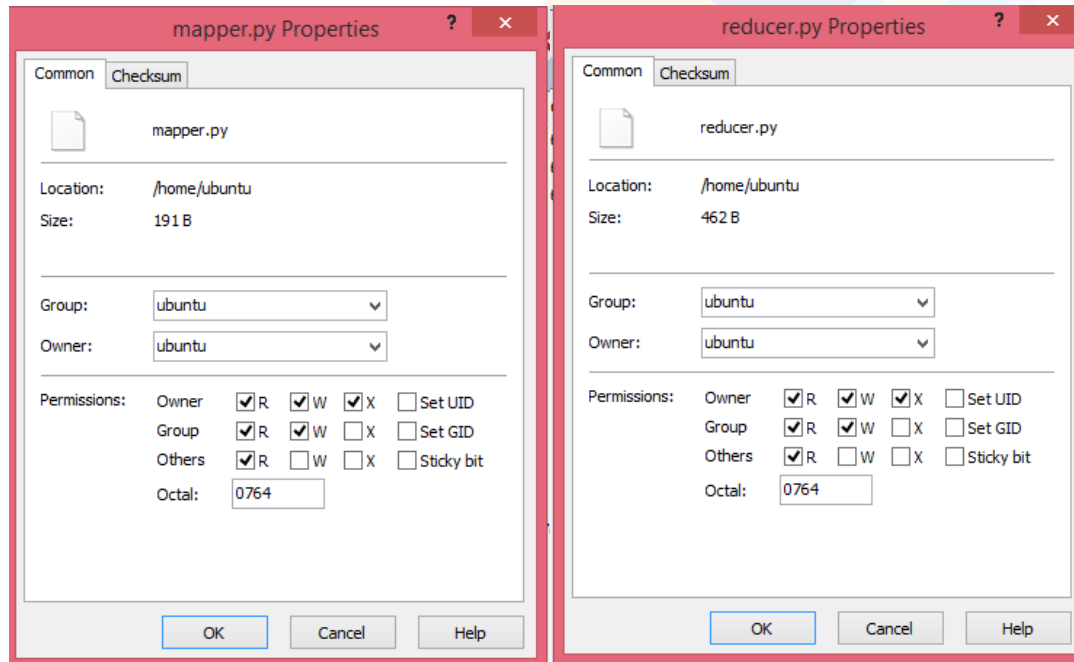
- Pastikan terdapat file `mapper.py` dan `reducer.py` dengan menggunakan perintah `$ ls`



```
ubuntu@ip-172-31-91-17: ~  
(base) ubuntu@ip-172-31-91-17:~$ ls  
anaconda3  file01          hadoop-2.8.5.tar.gz  reducer.py  
data       hadoop-2.8.5    mapper.py  
(base) ubuntu@ip-172-31-91-17:~$
```

Python Wordcount

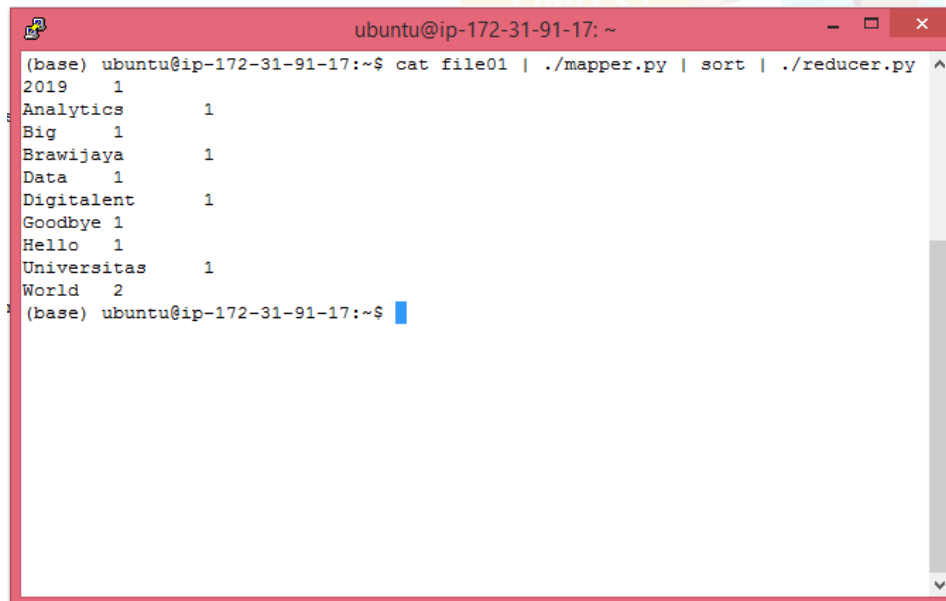
- Tambahkan izin eksekusi pada kedua file. (gunakan WinSCP)



Python Wordcount

- Jalankan program secara lokal tanpa hadoop.

```
$ cat file01 | ./mapper.py | sort | ./reducer.py
```



```
ubuntu@ip-172-31-91-17: ~  
(base) ubuntu@ip-172-31-91-17:~$ cat file01 | ./mapper.py | sort | ./reducer.py  
2019      1  
Analytics      1  
Big      1  
Brawijaya      1  
Data      1  
Digitalent      1  
Goodbye 1  
Hello      1  
Universitas      1  
World      2  
(base) ubuntu@ip-172-31-91-17:~$
```

Python Wordcount

- Jalankan program dengan hadoop

```
$ hadoop jar /home/ubuntu/hadoop-2.8.5/share/hadoop/tools/lib/hadoop-streaming-2.8.5.jar \
-mapper mapper.py \
-reducer reducer.py \
-input input/file01 \
-output output/file01out
```

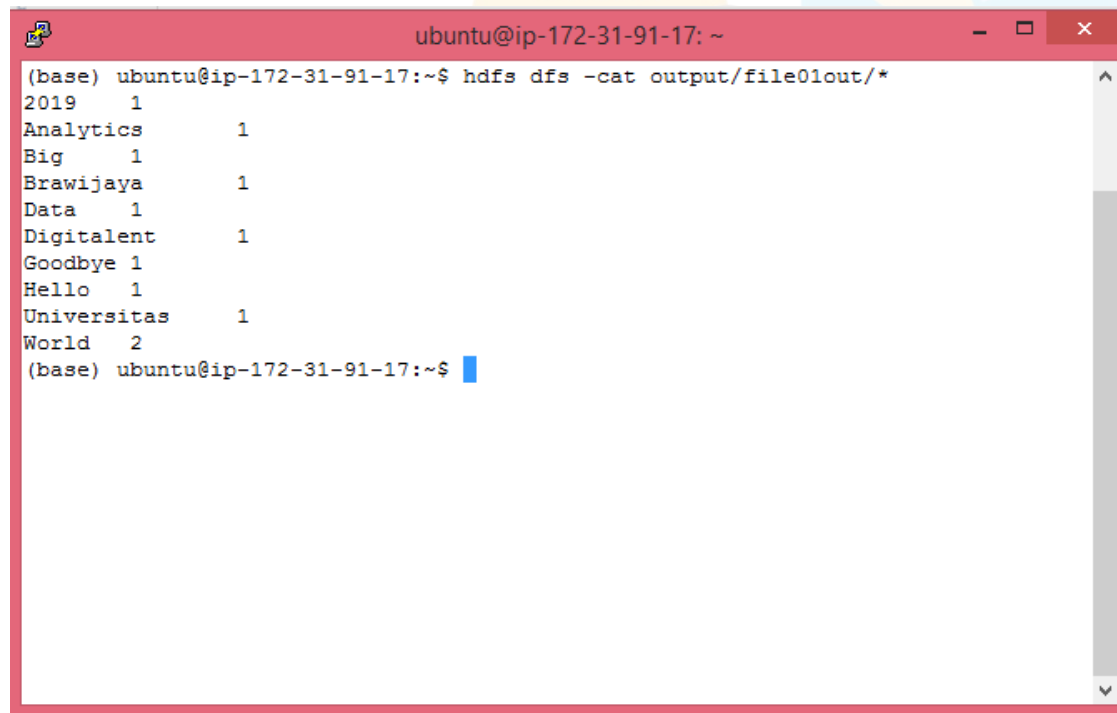
Python Wordcount

```
ubuntu@ip-172-31-91-17: ~  
(base) ubuntu@ip-172-31-91-17:~$ hadoop jar /home/ubuntu/hadoop-2.8.5/share/hado  
op/tools/lib/hadoop-streaming-2.8.5.jar -mapper mapper.py -reducer reducer.py -i  
nput input/file01 -output output/file01out  
19/06/26 15:57:01 INFO Configuration.deprecation: session.id is deprecated. Inst  
ead, use dfs.metrics.session-id  
19/06/26 15:57:01 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName  
=JobTracker, sessionId=  
19/06/26 15:57:01 INFO jvm.JvmMetrics: Cannot initialize JVM Metrics with proces  
sName=JobTracker, sessionId= - already initialized  
19/06/26 15:57:01 INFO mapred.FileInputFormat: Total input files to process : 1  
19/06/26 15:57:01 INFO mapreduce.JobSubmitter: number of splits:1  
19/06/26 15:57:02 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_lo  
cal1462026121_0001  
19/06/26 15:57:02 INFO mapreduce.Job: The url to track the job: http://localhost  
:8080/  
19/06/26 15:57:02 INFO mapreduce.Job: Running job: job_local1462026121_0001  
19/06/26 15:57:02 INFO mapred.LocalJobRunner: OutputCommitter set in config null  
19/06/26 15:57:02 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hado  
op.mapred.FileOutputCommitter  
19/06/26 15:57:02 INFO output.FileOutputCommitter: File Output Committer Algorit  
hm version is 1  
19/06/26 15:57:02 INFO output.FileOutputCommitter: FileOutputCommitter skip clea  
nup _temporary folders under output directory:false, ignore cleanup failures: fa  
lse
```

Python Wordcount

- Cetak file output

```
$ hdfs dfs -cat output/file01out/*
```



```
ubuntu@ip-172-31-91-17: ~  
(base) ubuntu@ip-172-31-91-17:~$ hdfs dfs -cat output/file01out/*  
2019      1  
Analytics      1  
Big      1  
Brawijaya      1  
Data      1  
Digitalent      1  
Goodbye 1  
Hello      1  
Universitas      1  
World      2  
(base) ubuntu@ip-172-31-91-17:~$
```



DIGITAL
TALENT
SCHOLARSHIP

IKUTI KAMI



DIGITAL
TALENT
SCHOLARSHIP

- digitalent.kominfo
- digitalent.kominfo
- DTS_kominfo
- Digital Talent Scholarship 2019

Pusat Pengembangan Profesi dan Sertifikasi
Badan Penelitian dan Pengembangan SDM
Kementerian Komunikasi dan Informatika
Jl. Medan Merdeka Barat No. 9
(Gd. Belakang Lt. 4 - 5)
Jakarta Pusat, 10110



digitalent.kominfo.go.id