

HADOOP

Installation and Deployment of a Single Node on a Linux System

Presented by:

Liv Nguekap

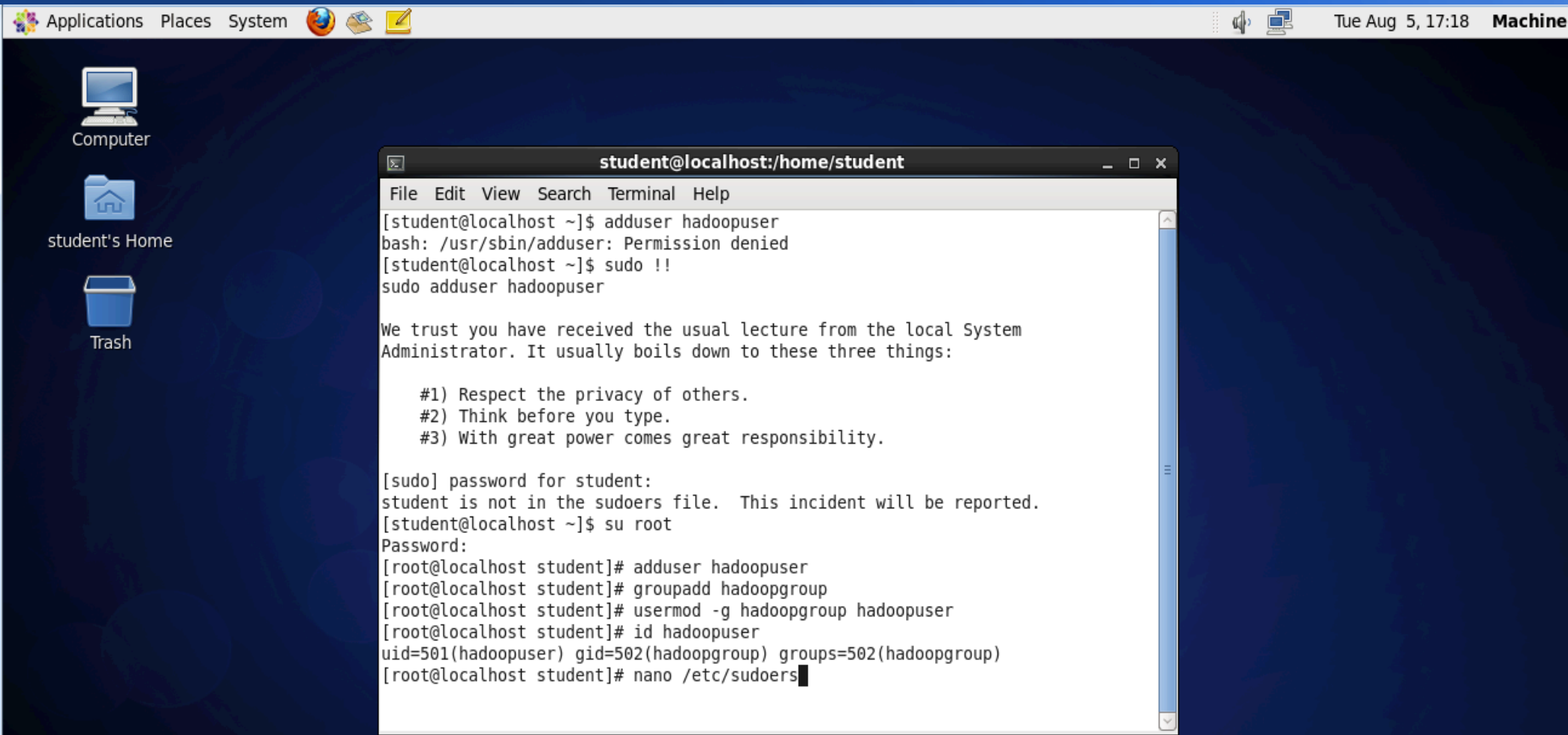
And

Garrett Poppe

Topics

- Create hadoopuser and group
- Edit sudoers
- Set up SSH
- Install JDK
- Install Hadoop
- Editing Hadoop settings
- Running Hadoop
- Resources

Add Hadoopuser



The screenshot shows a Linux desktop environment with a terminal window open. The terminal window title is "student@localhost:/home/student". The terminal output shows the following sequence of commands and responses:

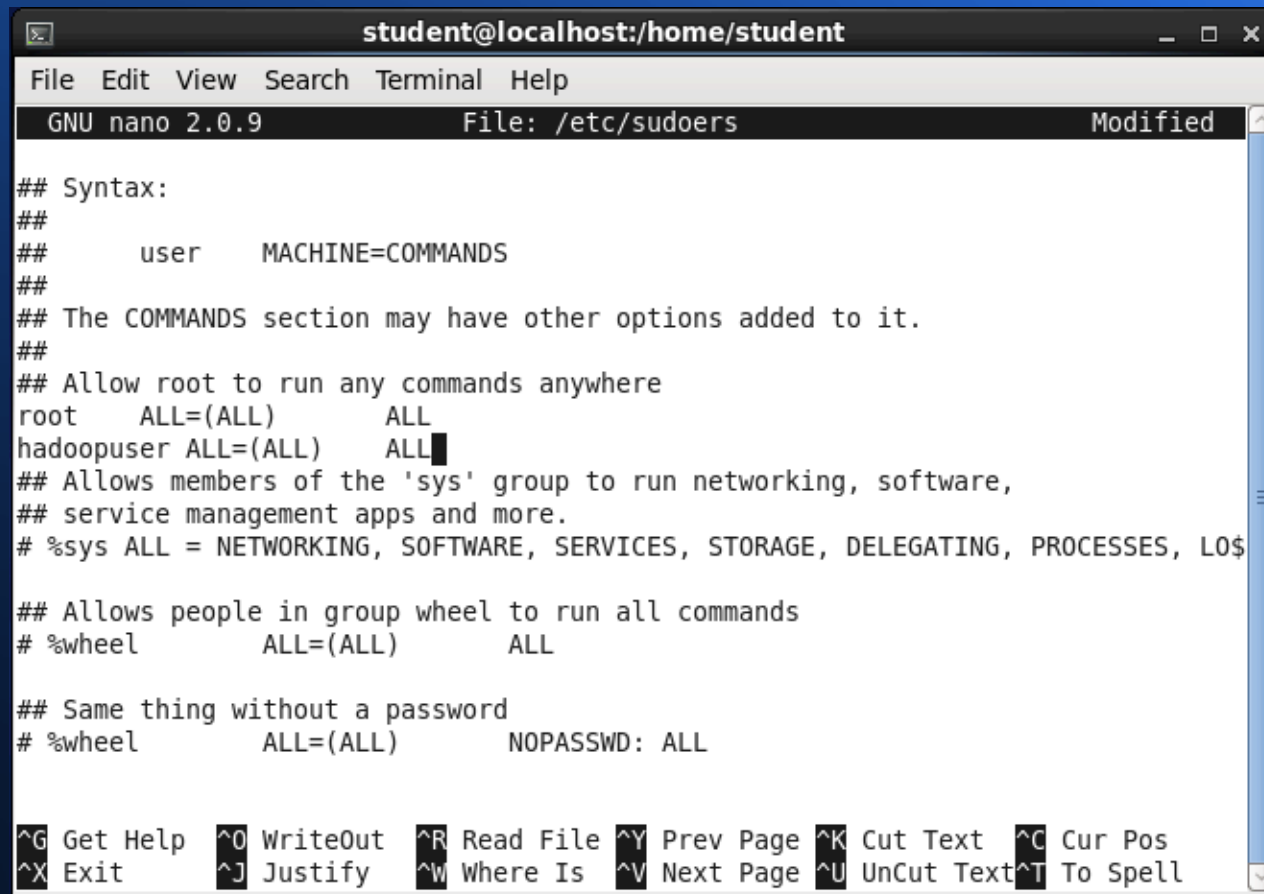
```
student@localhost ~]$ adduser hadoopuser
bash: /usr/sbin/adduser: Permission denied
student@localhost ~]$ sudo !!
sudo adduser hadoopuser

We trust you have received the usual lecture from the local System
Administrator. It usually boils down to these three things:

#1) Respect the privacy of others.
#2) Think before you type.
#3) With great power comes great responsibility.

[sudo] password for student:
student is not in the sudoers file. This incident will be reported.
student@localhost ~]$ su root
Password:
[root@localhost student]# adduser hadoopuser
[root@localhost student]# groupadd hadoopgroup
[root@localhost student]# usermod -g hadoopgroup hadoopuser
[root@localhost student]# id hadoopuser
uid=501(hadoopuser) gid=502(hadoopgroup) groups=502(hadoopgroup)
[root@localhost student]# nano /etc/sudoers
```

Edit sudoers



```
student@localhost:/home/student
File Edit View Search Terminal Help
GNU nano 2.0.9 File: /etc/sudoers Modified

## Syntax:
##
##      user    MACHINE=COMMANDS
##
## The COMMANDS section may have other options added to it.
##
## Allow root to run any commands anywhere
root    ALL=(ALL)    ALL
hadoopuser ALL=(ALL)    ALL
## Allows members of the 'sys' group to run networking, software,
## service management apps and more.
# %sys ALL = NETWORKING, SOFTWARE, SERVICES, STORAGE, DELEGATING, PROCESSES, LOGS

## Allows people in group wheel to run all commands
# %wheel    ALL=(ALL)    ALL

## Same thing without a password
# %wheel    ALL=(ALL)    NOPASSWD: ALL

^G Get Help  ^O WriteOut  ^R Read File  ^Y Prev Page  ^K Cut Text    ^C Cur Pos
^X Exit      ^J Justify   ^W Where Is  ^V Next Page  ^U UnCut Text ^T To Spell
```

Set up SSH

```
hadoopuser@localhost:~  
File Edit View Search Terminal Help  
[-k key_gen time] [-o option] [-p port] [-u len]  
[hadoopuser@localhost ~]$ ssh-keygen -t rsa -P ''  
Generating public/private rsa key pair.  
Enter file in which to save the key (/home/hadoopuser/.ssh/id_rsa):  
Your identification has been saved in /home/hadoopuser/.ssh/id_rsa.  
Your public key has been saved in /home/hadoopuser/.ssh/id_rsa.pub.  
The key fingerprint is:  
93:62:64:5e:cf:2b:f9:9b:0c:d1:a9:3c:03:46:7b:a2 hadoopuser@localhost.localdomain  
The key's randomart image is:  
+--[ RSA 2048 ]-----+  
      |  
      |      .  
      |     o = .  
      |    0 S =  
      |   + * = .  
      |  E 0 .  
      |   B .  
      |   =.  
      |  
+-----+  
[hadoopuser@localhost ~]$ sudo chown hadoopuser ~/.ssh
```

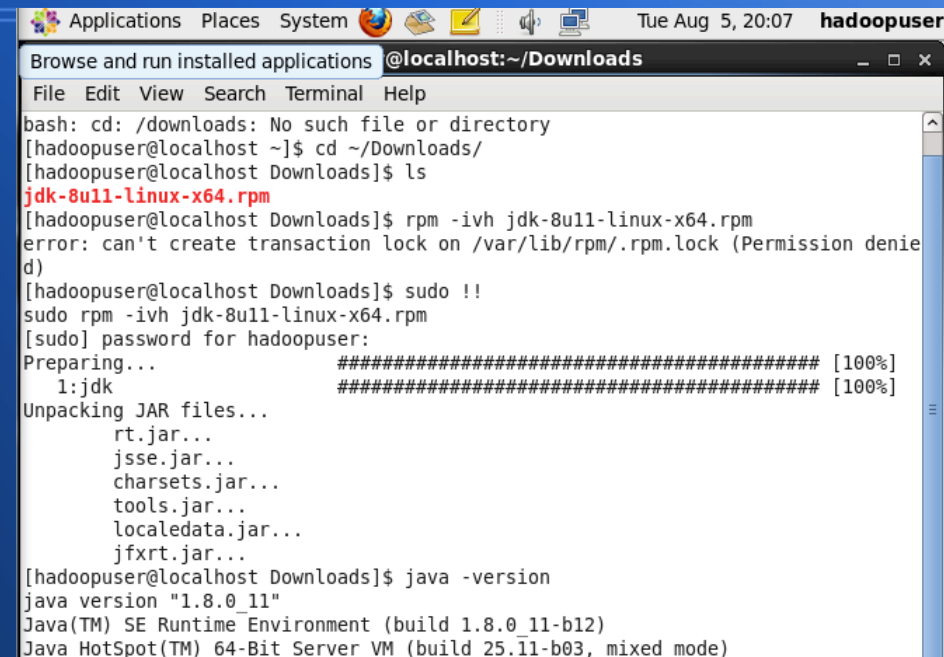
- `sudo chown hadoopuser ~/.ssh`
- `sudo chmod 700 ~/.ssh`
- `sudo chmod 600 ~/.ssh/id_rsa`
- `sudo cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys`
- `sudo chmod 600 ~/.ssh/authorized_keys`

```
hadoopuser@localhost:~  
File Edit View Search Terminal Help  
GNU nano 2.0.9 File: /etc/ssh/sshd_config Modified  
# Change to yes if you don't trust ~/.ssh/known_hosts for  
# RhostsRSAAuthentication and HostbasedAuthentication  
#IgnoreUserKnownHosts no  
# Don't read the user's ~/.rhosts and ~/.shosts files  
#IgnoreRhosts yes  
  
# To disable tunneled clear text passwords, change to no here!  
PasswordAuthentication yes  
PermitEmptyPasswords no  
PasswordAuthentication yes  
  
# Change to no to disable s/key passwords  
#ChallengeResponseAuthentication yes  
ChallengeResponseAuthentication no
```

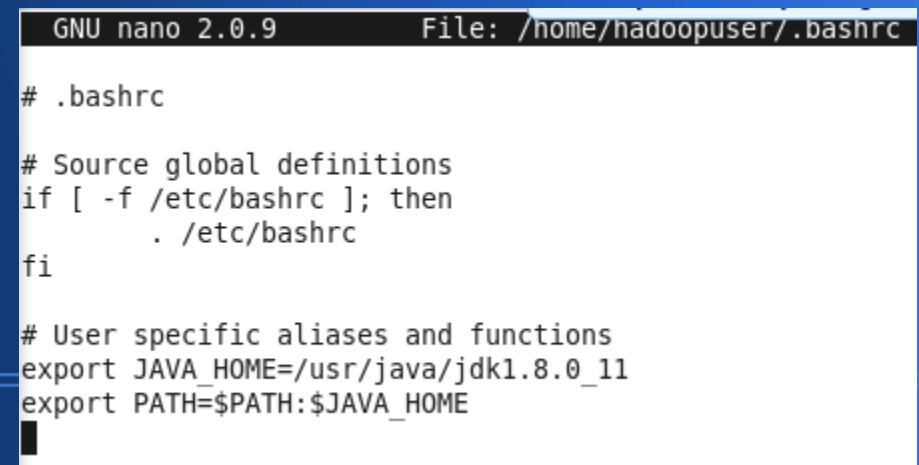
- `Edit /etc/ssh/sshd_config`

Install JDK

- Login as hadoopuser
- Uninstall previous versions of JDK
- Download current version of JDK
- Install JDK
- Edit JAVA_HOME and PATH variables in “~/bashrc” file



```
Applications Places System Tue Aug 5, 20:07 hadoopuser
Browse and run installed applications @localhost:~/Downloads
File Edit View Search Terminal Help
bash: cd: /downloads: No such file or directory
[hadoopuser@localhost ~]$ cd ~/Downloads/
[hadoopuser@localhost Downloads]$ ls
jdk-8u11-linux-x64.rpm
[hadoopuser@localhost Downloads]$ rpm -ivh jdk-8u11-linux-x64.rpm
error: can't create transaction lock on /var/lib/rpm/.rpm.lock (Permission denied)
[hadoopuser@localhost Downloads]$ sudo !!
sudo rpm -ivh jdk-8u11-linux-x64.rpm
[sudo] password for hadoopuser:
Preparing... ##### [100%]
 1:jdk ##### [100%]
Unpacking JAR files...
  rt.jar...
  jsse.jar...
  charsets.jar...
  tools.jar...
  localedata.jar...
  jfxrt.jar...
[hadoopuser@localhost Downloads]$ java -version
java version "1.8.0_11"
Java(TM) SE Runtime Environment (build 1.8.0_11-b12)
Java HotSpot(TM) 64-Bit Server VM (build 25.11-b03, mixed mode)
```








```
GNU nano 2.0.9 File: /home/hadoopuser/.bashrc
# .bashrc

# Source global definitions
if [ -f /etc/bashrc ]; then
    . /etc/bashrc
fi

# User specific aliases and functions
export JAVA_HOME=/usr/java/jdk1.8.0_11
export PATH=$PATH:$JAVA_HOME
```

Install Hadoop

Index of /apache/hadoop/common/hadoop-2.4.1

Name	Last modified	Size	Description
 Parent Directory	-	-	-
 hadoop-2.4.1-src.tar.gz	21-Jun-2014 06:42	15M	
 hadoop-2.4.1-src.tar.gz.mds	21-Jun-2014 06:42	1.1K	
 hadoop-2.4.1.tar.gz	21-Jun-2014 06:42	132M	
 hadoop-2.4.1.tar.gz.mds	21-Jun-2014 06:42	958	

Apache Server at mirrors.advancedhosters.com Port 80

- Download current stable release
- Untar the download
- `tar xzvf hadoop-2.4.1.tar.gz`
- Move the untarred folder
- `sudo mv hadoop-2.4.1 /usr/local/hadoop`
- Change ownership and create nodes
- `sudo chown -R hadoopuser:hadoopgroup /usr/local/hadoop`
- `mkdir -p ~/hadoop-space/hdfs/namenode`
- `mkdir -p ~/hadoop-space/hdfs/`

Install Hadoop

- Edit Hadoop variables in “~/.bashrc” file
- After editing file, use command to apply.
- “source ~/.bashrc”

```
GNU nano 2.0.9      File: /home/hadoopuser/.bashrc
# .bashrc

# Source global definitions
if [ -f /etc/bashrc ]; then
    . /etc/bashrc
fi

# User specific aliases and functions
export JAVA_HOME=/usr/java/jdk1.8.0_11
export PATH=$PATH:$JAVA_HOME

export HADOOP_INSTALL=/usr/local/hadoop
export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_HOME=$HADOOP_INSTALL
export HADOOP_HDFS_HOME=$HADOOP_INSTALL
export YARN_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native
export PATH=$PATH:$HADOOP_INSTALL/sbin
export PATH=$PATH:$HADOOP_INSTALL/bin
```


Editing Hadoop settings

- Go to directory located at `/usr/local/hadoop/etc/hadoop`
- Create a copy of `mapred-site.xml.template` as `mapred-site.xml`

```

Browse and run installed applications | /usr/local/hadoop/etc/hadoop
File Edit View Search Terminal Help
^C[hadoopuser@localhost Downloads]$ sudo mv hadoop-2.4.1 /usr/local/hadoop
[hadoopuser@localhost Downloads]$ sudo chown -R hadoopuser:hadoopgroup /usr/local/hadoop
[hadoopuser@localhost Downloads]$ cd ~
[hadoopuser@localhost ~]$ mkdir -p ~/hadoop/hdfs/namenode
[hadoopuser@localhost ~]$ mkdir -p ~/hadoop/hdfs/datanode
[hadoopuser@localhost ~]$ nano ~/.bashrc
[hadoopuser@localhost ~]$ source ~/.bashrc
[hadoopuser@localhost ~]$ cd hadoop/hadoop
[hadoopuser@localhost hadoop]$ cd ..
[hadoopuser@localhost ~]$ ls
Desktop  Downloads  Music      Public     Videos
Documents  hadoopspace  Pictures  Templates
[hadoopuser@localhost ~]$ cd ..
[hadoopuser@localhost home]$ ls
hadoopuser  student
[hadoopuser@localhost home]$ cd ..
[hadoopuser@localhost /]$ cd ..
[hadoopuser@localhost /]$ ls
bin  dev  home  lib64  media  opt  root  selinux  sys  usr
boot  etc  lib  lost+found  mnt  proc  sbin  srv  tmp  var
[hadoopuser@localhost /]$ cd /usr/local/hadoop/etc/hadoop
[hadoopuser@localhost hadoop]$ cp mapred-site.xml.template mapred-site.xml
[hadoopuser@localhost hadoop]$ nano mapred-site.xml
```

Editing Hadoop settings

- Edit mapred-site.xml
- Add code between <configuration> tags

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

Editing Hadoop settings

- Edit yarn-site.xml
- Add code between <configuration> tabs

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

Editing Hadoop settings

- Edit core-site.xml
- Add code between `<configuration>` tags

```
<property>
```

```
<name>
```

```
fs.default.name
```

```
</name>
```

```
<value>
```

```
hdfs://localhost:9000
```

```
</value>
```

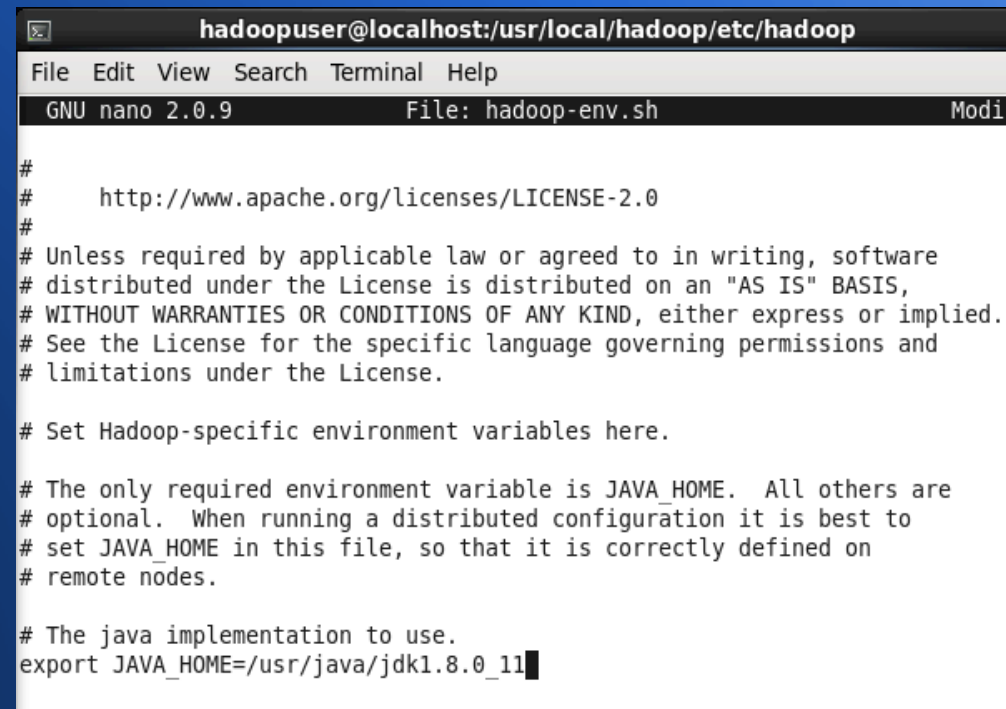
```
</property>
```

Editing Hadoop settings

- Edit hdfs-site.xml
 - Add code between <configuration> tabs
- | | | |
|-----------------|--|--|
| <property> | <property> | <property> |
| <name> | <name> | <name> |
| dfs.replication | dfs.name.dir | dfs.data.dir |
| </name> | </name> | </name> |
| <value> | <value> | <value> |
| 1 | file:///home/hadoopuser/
hadoop-space/hdfs/
namenode | file:///home/hadoopuser/
hadoop-space/hdfs/
datanode |
| </value> | </value> | </value> |
| </property> | </property> | </property> |

Editing Hadoop settings

- Edit “hadoop-env.sh”
- Create the `JAVA_HOME` variable using current JDK path.



```
hadoopuser@localhost:/usr/local/hadoop/etc/hadoop
File Edit View Search Terminal Help
GNU nano 2.0.9 File: hadoop-env.sh Modi
#
# 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.
#
# Set Hadoop-specific environment variables here.
#
# The only required environment variable is JAVA_HOME. All others are
# optional. When running a distributed configuration it is best to
# set JAVA_HOME in this file, so that it is correctly defined on
# remote nodes.
#
# The java implementation to use.
export JAVA_HOME=/usr/java/jdk1.8.0_11
```

Editing Hadoop settings

```
hadoopuser@localhost:~/usr/local/hadoop/etc/hadoop
File Edit View Search Terminal Help
[hadoopuser@localhost ~]$ mkdir -p ~/hadoop/hdfs/namenode
[hadoopuser@localhost ~]$ mkdir -p ~/hadoop/hdfs/datanode
[hadoopuser@localhost ~]$ nano ~/.bashrc
[hadoopuser@localhost ~]$ source ~/.bashrc
[hadoopuser@localhost ~]$ cd hadoop/hadoop/
[hadoopuser@localhost hadoop]$ cd ..
[hadoopuser@localhost ~]$ ls
Desktop Downloads Music Public Videos
Documents hadoop hadoop Pictures Templates
[hadoopuser@localhost ~]$ cd ..
[hadoopuser@localhost home]$ ls
hadoopuser student
[hadoopuser@localhost home]$ cd ..
[hadoopuser@localhost /]$ cd ..
[hadoopuser@localhost /]$ ls
bin dev home lib64 media opt root selinux sys usr
boot etc lib lost+found mnt proc sbin srv var
[hadoopuser@localhost /]$ cd /usr/local/hadoop/etc/hadoop
[hadoopuser@localhost hadoop]$ cp mapred-site.xml.template mapred-site.xml
[hadoopuser@localhost hadoop]$ nano mapred-site.xml
[hadoopuser@localhost hadoop]$ nano yarn-site.xml
[hadoopuser@localhost hadoop]$ nano hdfs-site.xml
[hadoopuser@localhost hadoop]$ nano hadoop-env.sh
[hadoopuser@localhost hadoop]$ hdfs namenode -format
```

- Format the namenode using the command “hdfs namenode -format”

Running Hadoop

```
14/08/05 21:20:51 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at localhost/127.0.0.1
*****/
[hadoopuser@localhost ~]$ start-dfs.sh
14/08/05 21:21:09 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hadoopuse
r-namenode-localhost.localdomain.out
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hadoopuse
r-datanode-localhost.localdomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-ha
doopuser-secondarynamenode-localhost.localdomain.out
14/08/05 21:21:35 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
[hadoopuser@localhost ~]$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hadoopuser-reso
urcemanager-localhost.localdomain.out
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hadoopus
er-nodemanager-localhost.localdomain.out
[hadoopuser@localhost ~]$
```

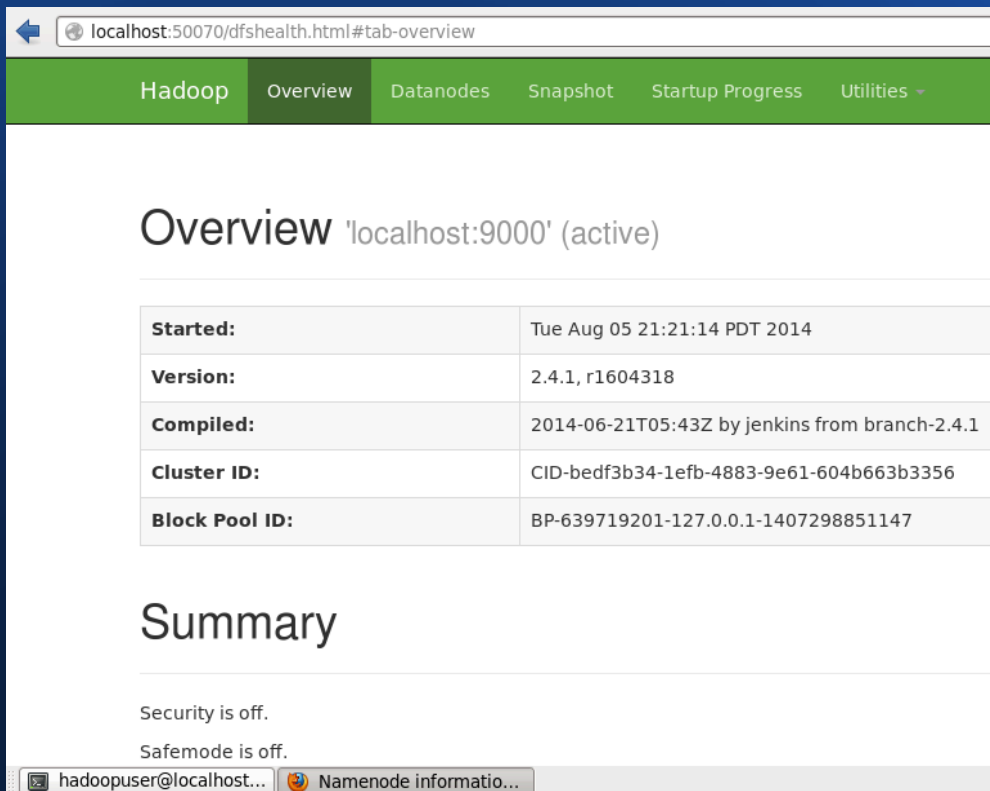
- Start services
- “start-dfs.sh”
- “start-yarn.sh”

Running Hadoop

- Use `jps` command to make sure all services are running.

```
[hadoopuser@localhost ~]$ jps
2928 SecondaryNameNode
3495 Jps
2761 DataNode
3177 NodeManager
3082 ResourceManager
2668 NameNode
[hadoopuser@localhost ~]$ █
```

Running Hadoop



localhost:50070/dfshealth.html#tab-overview

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Overview 'localhost:9000' (active)

Started:	Tue Aug 05 21:21:14 PDT 2014
Version:	2.4.1, r1604318
Compiled:	2014-06-21T05:43Z by jenkins from branch-2.4.1
Cluster ID:	CID-bedf3b34-1efb-4883-9e61-604b663b3356
Block Pool ID:	BP-639719201-127.0.0.1-1407298851147

Summary

Security is off.
Safemode is off.

hadoopuser@localhost... Namenode informatio...

- Open web browser.
- Type “localhost:50070” into address bar to access web interface.

Part 2

.WRITING MAPREDUCE PROGRAMS FOR HADOOP

Languages/scripts used

- We will talk about two languages used to write mapreduce programs in Hadoop:
 - 1) Pig Script (also called Pig Latin)
 - 2) Java

Pig

- What is Pig?
- Pig is a high-level platform for creating MapReduce programs used with Hadoop.
- It is somewhat similar to SQL

How Pig Works

- Pig has two modes of execution:
 - 1) Local Mode - To run Pig in local mode, you need access to a single machine.
 - 2) Mapreduce Mode - To run Pig in mapreduce mode, you need access to a Hadoop cluster and HDFS installation.

Syntax to run Pig

- To run Pig in Local Mode, use:
 - `pig -x local id.pig`

- To run Pig in Mapreduce Mode, use:
 - `pig id.pig` or `pig -x mapreduce id.pig`

Ways to run Pig

- Whether in local or mapreduce mode, there are 3 ways of running Pig:
 - 1) Grunt shell
 - 2) Batch or script file
 - 3) Embedded Program

Sample Grunt Shell Code

```
A = LOAD 'student' USING PigStorage() AS (name:chararray, age:int, gpa:float);  
DUMP A;  
(John,18,4.0F)  
(Mary,19,3.8F)  
(Bill,20,3.9F)  
(Joe,18,3.8F)
```

Grunt Shell Commands

```
A = LOAD 'student' USING PigStorage() AS (name:chararray, age:int, gpa:float);
X = FOREACH A GENERATE name,$2;
DUMP X;
(John,4.0F)
(Mary,3.8F)
(Bill,3.9F)
(Joe,3.8F)
```

Grunt Shell Commands

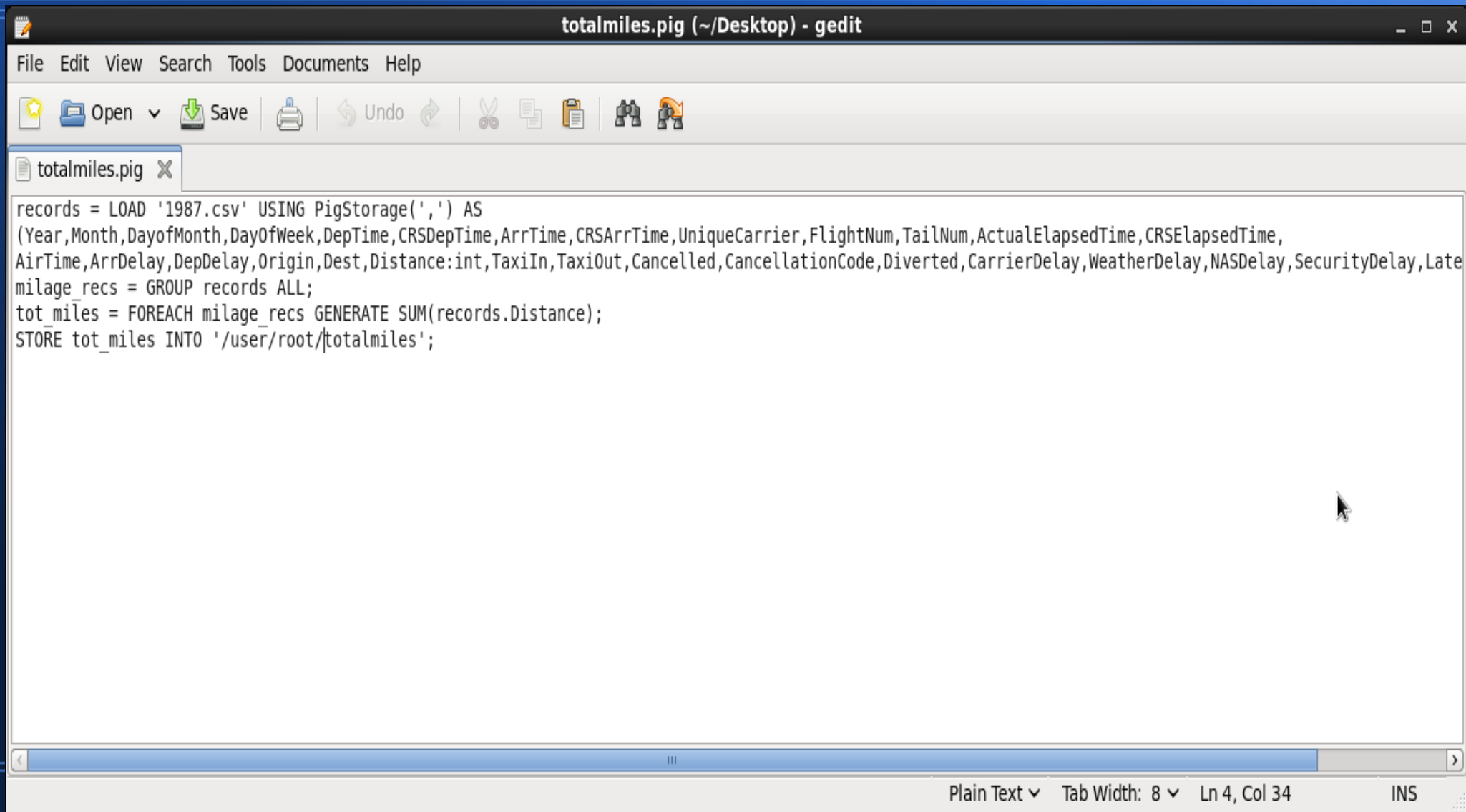
```
A = LOAD 'data' as (f1:int, f2:int, f3:int);  
DUMP A;  
(1,2,3)  
(4,2,1)  
(8,3,4)  
(4,3,3)
```

```
X = GROUP A BY f1;  
DUMP X;  
(1, {(1,2,3)})  
(4, {(4,2,1), (4,3,3)})  
(8, {(8,3,4)})
```

Batch

- To run Pig with batch files, the pig script is written entirely into a Pig file and the file run with Pig.
- A sample syntax for the file totalmiles.pig is:
- `Pig totalmiles.pig`

Content of file totalmiles.pig



```
records = LOAD '1987.csv' USING PigStorage(',') AS
(Year,Month,DayofMonth,DayOfWeek,DepTime,CRSDepTime,ArrTime,CRSArrTime,UniqueCarrier,FlightNum,TailNum,ActualElapsedTime,CRSElapsedTime,
AirTime,ArrDelay,DepDelay,Origin,Dest,Distance:int,TaxiIn,TaxiOut,Cancelled,CancellationCode,Diverted,CarrierDelay,WeatherDelay,NASDelay,SecurityDelay,Late
milage_recs = GROUP records ALL;
tot_miles = FOREACH milage_recs GENERATE SUM(records.Distance);
STORE tot_miles INTO '/user/root/totalmiles';
```

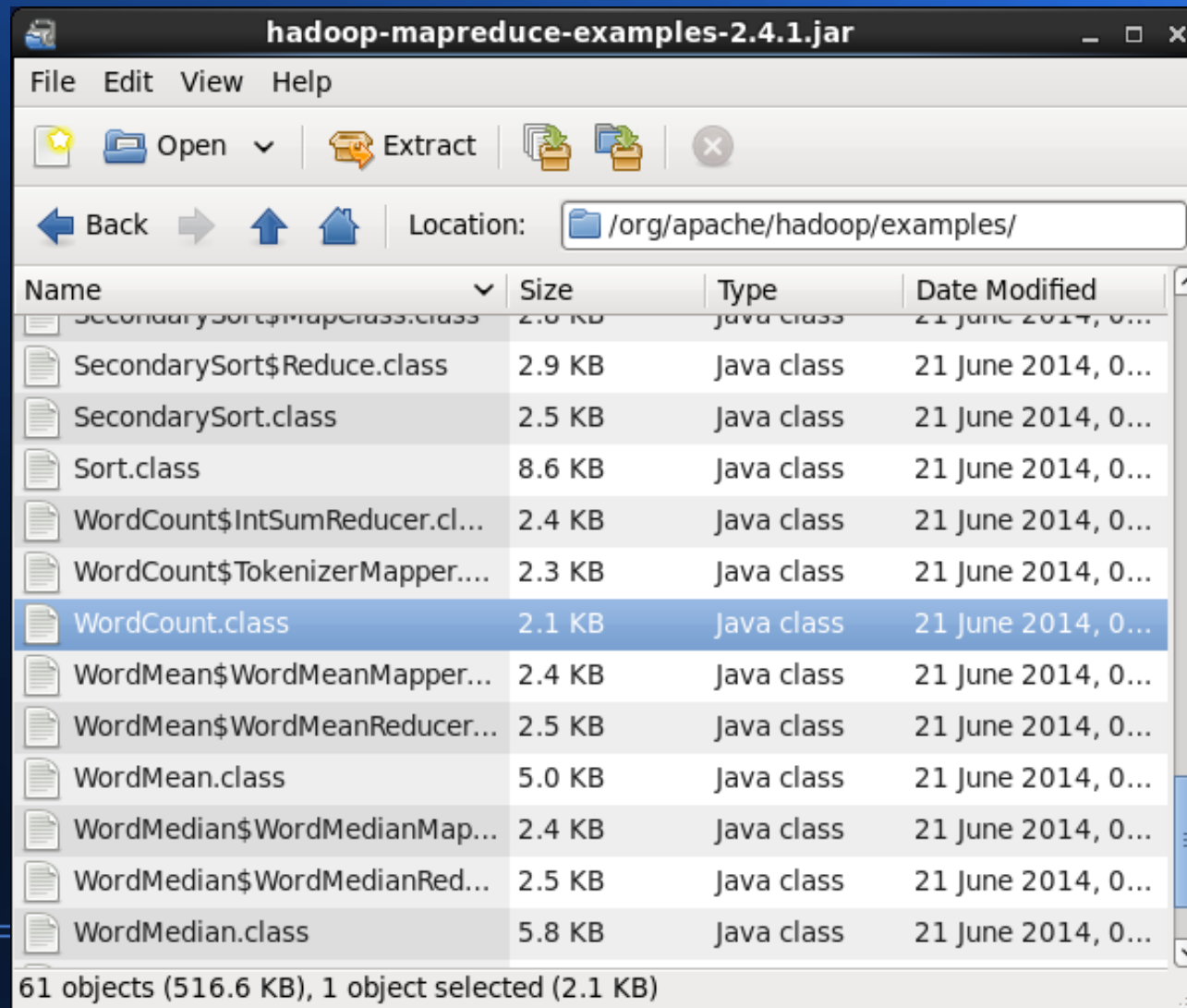
Content of 1987 flight data file

```
Year,Month,DayOfMonth,DayOfWeek,DepTime,CRSDepTime,ArrTime,CRSArrTime,UniqueCarrier,FlightNum,TailNum,ActualElapsedTime,CRSElapsedTime,AirTime,ArrDelay,DepDel
1987,10,14,3,741,730,912,849,PS,1451,NA,91,79,NA,23,11,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,15,4,729,730,903,849,PS,1451,NA,94,79,NA,14,-1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,17,6,741,730,918,849,PS,1451,NA,97,79,NA,29,11,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,18,7,729,730,847,849,PS,1451,NA,78,79,NA,-2,-1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,19,1,749,730,922,849,PS,1451,NA,93,79,NA,33,19,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,21,3,728,730,848,849,PS,1451,NA,80,79,NA,-1,-2,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,22,4,728,730,852,849,PS,1451,NA,84,79,NA,3,-2,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,23,5,731,730,902,849,PS,1451,NA,91,79,NA,13,1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,24,6,744,730,908,849,PS,1451,NA,84,79,NA,19,14,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,25,7,729,730,851,849,PS,1451,NA,82,79,NA,2,-1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,26,1,735,730,904,849,PS,1451,NA,89,79,NA,15,5,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,28,3,741,725,919,855,PS,1451,NA,98,90,NA,24,16,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,29,4,742,725,906,855,PS,1451,NA,84,90,NA,11,17,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,31,6,726,725,848,855,PS,1451,NA,82,90,NA,-7,1,SAN,SFO,447,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,1,4,936,915,1035,1001,PS,1451,NA,59,46,NA,34,21,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,2,5,918,915,1017,1001,PS,1451,NA,59,46,NA,16,3,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,3,6,928,915,1037,1001,PS,1451,NA,69,46,NA,36,13,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,4,7,914,915,1003,1001,PS,1451,NA,49,46,NA,2,-1,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,5,1,1042,915,1129,1001,PS,1451,NA,47,46,NA,88,87,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,6,2,934,915,1024,1001,PS,1451,NA,50,46,NA,23,19,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,7,3,946,915,1037,1001,PS,1451,NA,51,46,NA,36,31,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,8,4,932,915,1033,1001,PS,1451,NA,61,46,NA,32,17,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,9,5,947,915,1036,1001,PS,1451,NA,49,46,NA,35,32,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,10,6,915,915,1022,1001,PS,1451,NA,67,46,NA,21,0,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,11,7,916,915,1006,1001,PS,1451,NA,50,46,NA,5,1,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,12,1,944,915,1027,1001,PS,1451,NA,43,46,NA,26,29,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,13,2,941,915,1036,1001,PS,1451,NA,55,46,NA,35,26,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,14,3,930,915,1029,1001,PS,1451,NA,59,46,NA,28,15,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,15,4,920,915,1023,1001,PS,1451,NA,63,46,NA,22,5,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,17,6,1009,915,1104,1001,PS,1451,NA,55,46,NA,54,54,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,18,7,915,915,1008,1001,PS,1451,NA,53,46,NA,7,0,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,19,1,940,915,1032,1001,PS,1451,NA,52,46,NA,31,25,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,21,3,913,915,1003,1001,PS,1451,NA,50,46,NA,2,-2,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,22,4,915,915,1017,1001,PS,1451,NA,62,46,NA,16,0,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,23,5,927,915,1022,1001,PS,1451,NA,55,46,NA,21,12,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,24,6,929,915,1052,1001,PS,1451,NA,83,46,NA,51,14,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,25,7,914,915,1011,1001,PS,1451,NA,57,46,NA,10,-1,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,26,1,917,915,1023,1001,PS,1451,NA,66,46,NA,22,2,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,27,2,916,915,1014,1001,PS,1451,NA,58,46,NA,13,1,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,28,3,941,925,1106,1015,PS,1451,NA,85,50,NA,51,16,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,29,4,940,925,1044,1015,PS,1451,NA,64,50,NA,29,15,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,30,5,1009,925,1105,1015,PS,1451,NA,56,50,NA,50,44,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,31,6,945,925,1040,1015,PS,1451,NA,55,50,NA,25,20,SFO,RNO,192,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,1,4,1520,1505,1624,1608,PS,1453,NA,64,63,NA,16,15,BUR,OAK,325,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,2,5,1508,1505,1615,1608,PS,1453,NA,67,63,NA,7,3,BUR,OAK,325,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
1987,10,4,7,1526,1505,1625,1608,PS,1453,NA,59,63,NA,17,21,BUR,OAK,325,NA,NA,0,NA,0,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA,NA
```

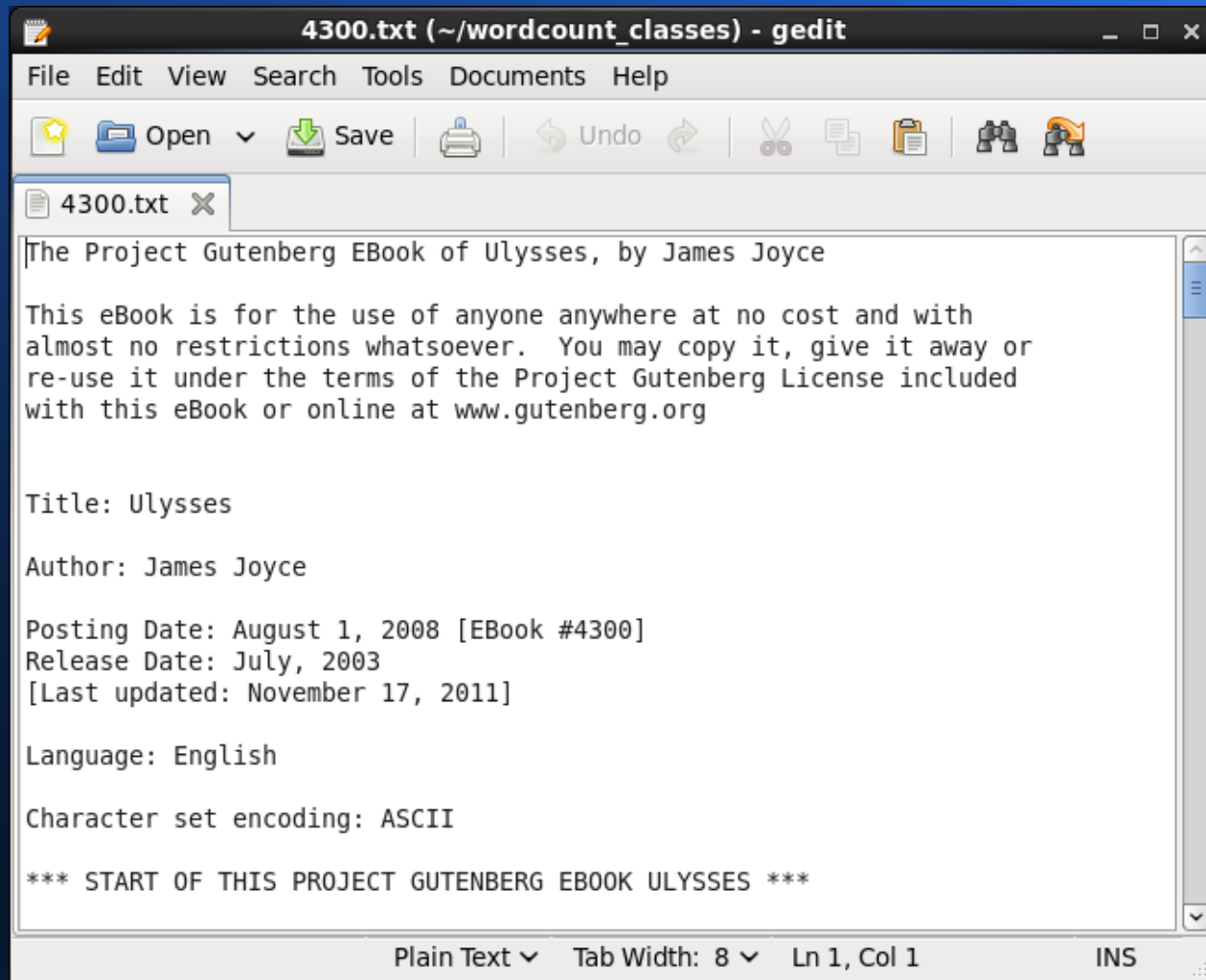
JAVA

- We tested the mapreduce function of Hadoop on a java program called WordCount.java
- The wordcount.class is provided in the examples that come with hadoop installation

Where to find the Hadoop Examples



JAVA



The screenshot shows a window titled "4300.txt (~/.wordcount_classes) - gedit". The window contains the following text:

```
The Project Gutenberg EBook of Ulysses, by James Joyce

This eBook is for the use of anyone anywhere at no cost and with
almost no restrictions whatsoever. You may copy it, give it away or
re-use it under the terms of the Project Gutenberg License included
with this eBook or online at www.gutenberg.org

Title: Ulysses

Author: James Joyce

Posting Date: August 1, 2008 [EBook #4300]
Release Date: July, 2003
[Last updated: November 17, 2011]

Language: English

Character set encoding: ASCII

*** START OF THIS PROJECT GUTENBERG EBOOK ULYSSES ***
```

The status bar at the bottom of the window indicates "Plain Text", "Tab Width: 8", "Ln 1, Col 1", and "INS".

Launching WordCount job

```
hadoop@WaterWorld:~  
File Edit View Search Terminal Help  
[hadoop@WaterWorld ~]$ hadoop jar /home/hadoop/hadoop-2.4.1/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.4.1.jar wordcount wordcount_classes wordcount_classes-output  
14/08/05 16:59:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
14/08/05 16:59:37 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032  
14/08/05 16:59:40 INFO input.FileInputFormat: Total input paths to process : 1  
14/08/05 16:59:40 INFO mapreduce.JobSubmitter: number of splits:1  
14/08/05 16:59:41 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1407275167956_0001  
14/08/05 16:59:43 INFO impl.YarnClientImpl: Submitted application application_1407275167956_0001  
14/08/05 16:59:43 INFO mapreduce.Job: The url to track the job: http://WaterWorld:8088/proxy/application_1407275167956_0001/  
14/08/05 16:59:43 INFO mapreduce.Job: Running job: job_1407275167956_0001  
14/08/05 17:00:03 INFO mapreduce.Job: Job job_1407275167956_0001 running in uber mode : false  
14/08/05 17:00:03 INFO mapreduce.Job:  map 0% reduce 0%  
14/08/05 17:00:17 INFO mapreduce.Job:  map 100% reduce 0%  
14/08/05 17:00:31 INFO mapreduce.Job:  map 100% reduce 100%  
14/08/05 17:00:32 INFO mapreduce.Job: Job job_1407275167956_0001 completed successfully  
14/08/05 17:00:32 INFO mapreduce.Job: Counters: 49
```

WordCount Processing

```
hadoop@WaterWorld:~  
File Edit View Search Terminal Help  
File System Counters  
  FILE: Number of bytes read=725074  
  FILE: Number of bytes written=1636087  
  FILE: Number of read operations=0  
  FILE: Number of large read operations=0  
  FILE: Number of write operations=0  
  HDFS: Number of bytes read=1573203  
  HDFS: Number of bytes written=527555  
  HDFS: Number of read operations=6  
  HDFS: Number of large read operations=0  
  HDFS: Number of write operations=2  
Job Counters  
  Launched map tasks=1  
  Launched reduce tasks=1  
  Data-local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=12048  
  Total time spent by all reduces in occupied slots (ms)=10742  
  Total time spent by all map tasks (ms)=12048  
  Total time spent by all reduce tasks (ms)=10742  
  Total vcore-seconds taken by all map tasks=12048  
  Total vcore-seconds taken by all reduce tasks=10742  
  Total megabyte-seconds taken by all map tasks=12337152  
  Total megabyte-seconds taken by all reduce tasks=10999808  
Map-Reduce Framework
```

WordCount Processing

```
hadoop@WaterWorld:~  
File Edit View Search Terminal Help  
Reduce shuffle bytes=725074  
Reduce input records=50095  
Reduce output records=50095  
Spilled Records=100190  
Shuffled Maps =1  
Failed Shuffles=0  
Merged Map outputs=1  
GC time elapsed (ms)=738  
CPU time spent (ms)=12730  
Physical memory (bytes) snapshot=366682112  
Virtual memory (bytes) snapshot=4171976704  
Total committed heap usage (bytes)=234881024  
Shuffle Errors  
BAD_ID=0  
CONNECTION=0  
IO_ERROR=0  
WRONG_LENGTH=0  
WRONG_MAP=0  
WRONG_REDUCE=0  
File Input Format Counters  
Bytes Read=1573078  
File Output Format Counters  
Bytes Written=527555  
[hadoop@WaterWorld ~]$ hadoop wordcount_classes -ls
```

Results

```
hadoop@WaterWorld:~  
File Edit View Search Terminal Help  
[hadoop@WaterWorld ~]$ hdfs dfs -ls  
14/08/05 17:07:45 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry for your platform... using builtin-java classes where applicable  
Found 2 items  
drwxr-xr-x   - hadoop supergroup          0 2014-08-05 16:06 wordcount_classes  
drwxr-xr-x   - hadoop supergroup          0 2014-08-05 17:00 wordcount_classes-o  
utput  
[hadoop@WaterWorld ~]$ hdfs dfs -ls wordcount_classes-output  
14/08/05 17:08:20 WARN util.NativeCodeLoader: Unable to load native-hadoop libra  
ry for your platform... using builtin-java classes where applicable  
Found 2 items  
-rw-r--r--   1 hadoop supergroup          0 2014-08-05 17:00 wordcount_classes-o  
utput/_SUCCESS  
-rw-r--r--   1 hadoop supergroup    527555 2014-08-05 17:00 wordcount_classes-o  
utput/part-r-00000  
[hadoop@WaterWorld ~]$ hdfs dfs -cat wordcount_classes-output/part-r-00000 | les  
s
```

Results



A terminal window titled "hadoop@WaterWorld:~" displays the results of a search. The window has a menu bar with "File", "Edit", "View", "Search", "Terminal", and "Help". The search results are listed as follows:

Term	Count
"Come	1
"Defects,"	1
"I	1
"Information	1
"J"	1
"Plain	2
"Project	5
"Right	1
"Viator"	1
#4300]	1
\$5,000)	1
%	2
&c,	2
&c.	1
'46.	1
'92	1
'AS-IS'	1
'Slife,	1
'TWAS	1
'Tis	8
'Tis,	1
'Twas	5
'Twixt	1
:	

WordCount.Java - Map

```
public static class MapClass extends MapReduceBase
    implements Mapper<LongWritable, Text, Text, IntWritable> {

    private final static IntWritable one = new IntWritable(1);
    private Text word = new Text();

    public void map(LongWritable key, Text value,
        OutputCollector<Text, IntWritable> output,
        Reporter reporter) throws IOException {
        String line = value.toString();
        StringTokenizer itr = new StringTokenizer(line);
        while (itr.hasMoreTokens()) {
            word.set(itr.nextToken());
            output.collect(word, one);
        }
    }
}
```

WordCount.java - Reduce

```
public static class Reduce extends MapReduceBase
    implements Reducer<Text, IntWritable, Text, IntWritable> {

    public void reduce(Text key, Iterator<IntWritable> values,
        OutputCollector<Text, IntWritable> output,
        Reporter reporter) throws IOException {

        int sum = 0;
        while (values.hasNext()) {
            sum += values.next().get();
        }
        output.collect(key, new IntWritable(sum));
    }
}
```


.Fin

• Thank YOU!!

Resources

- <http://alanxelsys.com/hadoop-v2-single-node-installation-on-centos-6-5/>
- <http://tecadmin.net/setup-hadoop-2-4-single-node-cluster-on-linux/>
- <http://hadoop.apache.org/>
- http://cs.smith.edu/dftwiki/index.php/Hadoop_Tutorial_1_--_Running_WordCount
- <https://pig.apache.org/docs/r0.10.0/basic.html>