Task 1

1. Start Hadoop single node on Acadgild VM (Command start-all.sh)

```
[acadgild@localhost~]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/06/28 00:25:49 WARN util.NativeCodeLoader: Unable to load native-hadoc
ry for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/acadgild/install/hadoop/ha
6.5/logs/hadoop-acadgild-namenode-localhost.localdomain.out
localhost: starting datanode, logging to /home/acadgild/install/hadoop/ha
6.5/logs/hadoop-acadgild-datanode-localhost.localdomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0: starting secondary namenode, logging to /home/acadgild/install/ha/logs/hadoop-acadgild-secondarynamenode-localhost.localdomain.out
18/06/28 00:26:12 WARN util.NativeCodeLoader: Unable to load native-hadoc replatform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /home/acadgild/install/hadoop/hadoop/s/yarn-acadgild-nodemanager-localhost.localdomain.out
localhost: starting nodemanager, logging to /home/acadgild/install/hadoop/syarn-acadgild-nodemanager-localhost.localdomain.out
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

2. Run a JPS Command to see if all the Hadoop daemons are running

```
| Comparison of the content of the c
```

3. Run couple of Unix commands like pwd, ls -ls etc.

Comments: Commands executed are

- pwd command is used to retrieve present working directory
- Is command lists the directories and files present in the current directory. Example Is -Is list the files and directories in the order along with other details like access write, time of access and so on
- mkdir command is to make a new directory. Example mkdir sample where sample is a directory name.

```
You have new mail in /var/spool/mail/acadgild [acadgild@localhost ~]$ ^C [acadgild@localhost ~]$ pwd |
//home/acadgild |
// You have new mail in /var/spool/mail/acadgild |
// Acadgild@localhost ~]$ |
// Is ~1s |
// Is ~1s
```

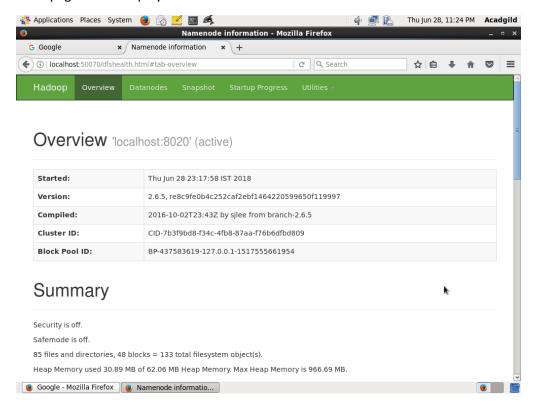
4. Create a file from the terminal using nano editor (for example nano test.txt) and put some Content in it. Cat it to see if the content is saved.

<u>Comments</u>: Screenshot below of file created using nano editor and later content display on terminal using Cat command.

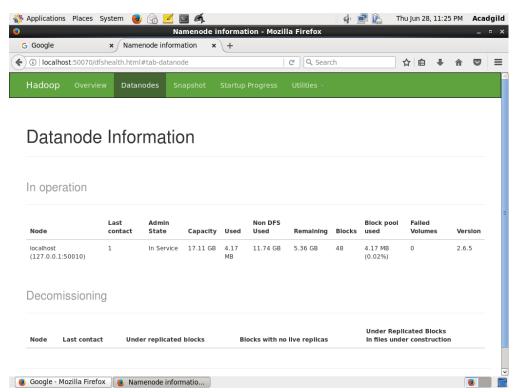
5. Open the hdfs web page from the browser by typing localhost:50070 in the browser. Check all the details of the HDFS.

<u>Comments</u>: Screenshots below of the webpage with all the HDFS details like namenode details, editlog, file information

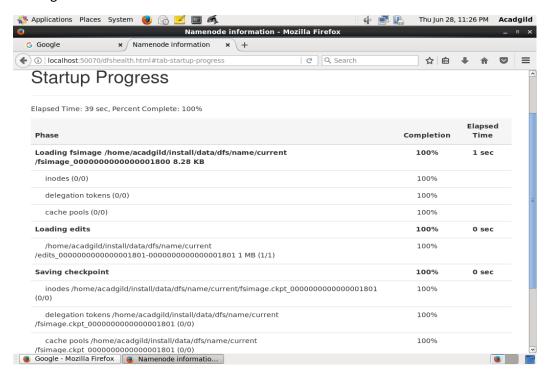
Main page which displays overview



Datanode Information



Edit log details



Directory Details



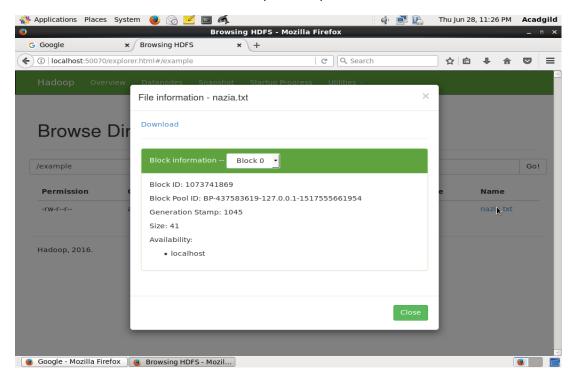
Browse Directory



Hadoop, 2016.



File details with number of blocks created, block id, size and so on.



TASK 2

1. Check whether /user/acadgild directory exists or not in the HDFS. If it doesn't exist, then create this. Create a directory /user/acadgild/Hadoop.

<u>Comments</u>: User directory already existed, hence directory acadgild and sub directory hadoop was created using mkdir command. Screenshots below.

2. Create a file in HDFS under directory /user/acadgild/Hadoop, with name word-count.txt. Whatever we type on screen should get appended to the file. Try to type (on screen) few lines from any online article or textbook.

<u>Comments</u>: A few lines are written on the screen and saved in a file. This file content is appended into a new file name 'word-count.txt' in the directory /user/acadgild/Hadoop using "-appendToFile" command.

3. Create a file max-temp.txt in local FS. Put some 10-15 records of date and temperature. Example: dd-mm-yyyy, temperature Move this file to HDFS at /user/acadgild/Hadoop.

<u>Comments</u>: Max-temp.txt file is created in local FS and then the file is moved to HDFS using -put command. The -put command creates a new file with the file name max-temp.txt itself in path /user/acadgild/Hadoop.

4. Change the permission of the file /user/acadgild/Hadoop/max-temp.txt, such that only the owner and the group members have full control over the file. Others do not have any control over it.

Comments: The permission access is changed by using command "-chmod".

```
user/acadgild/hado
8/07/06 19:12:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using buil
acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/hadoop
8/07/06 19:12:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using buil
ound 2 items
rwxrwxrwx 1 acadgild supergroup
                                            94 2018-07-06 18:55 /user/acadgild/hadoop/max-temp.txt
            1 acadgild supergroup
                                           150 2018-07-06 18:40 /user/acadgild/hadoop/word-count.txt
acadgild@localhost ~]$ hadoop fs -chmod 770 /user/acadgild/hadoop/max-temp.txt
8/07/06 19:13:03 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using buil
acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/hadoop
8/07/06 19:13:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using buil
           1 acadgild supergroup
                                            94 2018-07-06 18:55 /user/acadgild/hadoop/max-temp.txt
            1 acadgild supergroup
                                           150 2018-07-06 18:40 /user/acadgild/hadoop/word-count.txt
  adgild@localhost ~]$
```

TASK 3

1. Execute WordMedian, WordMean, WordStandardDeviation programs using Hadoop-mapreduce-examples-2.9.0.jar file present in your AcadGild VM

Comments: Screenshot of execution of wordmean program

```
at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmean /wordcount.txt /outputwordmean
[8/07/08 14:11:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applica
[8/07/08 14:11:13 INFO client.RWFroxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
[8/07/08 14:11:15 INFO input.FileInputFormat: Total input paths to process : 1
[8/07/08 14:11:15 INFO mapreduce.Jobsubmitter: number of splits:]
[8/07/08 14:11:16 INFO mapreduce.Jobsubmitter: Submitting tokens for job: job_1531027983712_0004
[8/07/08 14:11:16 INFO mapreduce.Jobsubmitter: Submitting tokens for job: job_1531027983712_0004
[8/07/08 14:11:16 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1531027983712_0004
[8/07/08 14:11:16 INFO mapreduce.Job: Nob job_1531027983712_0004
[8/07/08 14:11:17 INFO mapreduce.Job: Job job_1531027983712_0004
[8/07/08 14:11:27 INFO mapreduce.Job: map 0% reduce 0%
[8/07/08 14:11:25 INFO mapreduce.Job: map 100% reduce 0%
[8/07/08 14:11:45 INFO mapreduce.Job: Dap job jcs_1531027983712_0004 completed successfully
[8/07/08 14:11:45 INFO mapreduce.Job: Dap job jcs_1531027983712_0004 completed successfully
[8/07/08 14:11:45 INFO mapreduce.Job: Counters: 49
[8/07/08 14:11:45 INFO mapreduce.Job: Dap job_1531027983712_0004 completed successfully
[8/07/08 14:11:45 INFO mapreduce.Job: Counters: 49
[8/0
```

```
[acadgild@localhost ~]$ hadoop fs -ls /outputwordmedian
18/07/08 14:28:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
ls: `/outputwordmedian': No such file or directory
[acadgild@localhost ~]$ hadoop fs -ls /outputwordmean
18/07/08 14:28:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
Found 2 items
            1 acadgild supergroup
                                             0 2018-07-08 14:11 /outputwordmean/ SUCCESS
-rw-r--r--
-rw-r--r-- 1 acadgild supergroup 18 2018-07-08 14:11 /outputwordmean/part-r-00000
[acadgild@localhost \sim]$ hadoop fs -cat/outputwordmean/part-r-00000
-cat/outputwordmean/part-r-00000: Unknown command
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /outputwordmean/part-r-00000
18/07/08 14:29:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
length 42
[acadgild@localhost ~]$ ls
```

Screenshot of wordmedian

```
[acadgild@localhost -]5 hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordmedian /wordcount.txt /outputwordmedian 18/07/08 14:38:30 WARN util.MativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where appl 18/07/08 14:38:31 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032 18/07/08 14:38:32 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execution in the control of the control
```

```
WRONG REDUCE=0
        File Input Format Counters
                Bytes Read=49
        File Output Format Counters
                Bytes Written=20
The median is: 7
ou have new mail in /var/spool/mail/acadgild
acadgild@localhost ~]$ hadoop fs -ls /outputwordmedian
.8/07/08 14:41:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform
ound 2 items
rw-r--r-- 1 acadgild supergroup
rw-r--r-- 1 acadgild supergroup
                                             0 2018-07-08 14:39 /outputwordmedian/_SUCCESS
            1 acadgild supergroup
                                            20 2018-07-08 14:38 /outputwordmedian/part-r-00000
ou have new mail in /var/spool/mail/acadgild
acadgild@localhost ~]$ hadoop fs -cat /outputwordmedian/part-r-00000
8/07/08 14:41:53 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform.
acadgild@localhost ~]$
```

Screenshot of WordStandardDeviation

```
[acadgild@localhost ~]$ hadoop jar hadoop-mapreduce-examples-2.6.5.jar wordstandarddeviation /wordcount.txt /outputstanddeviation 18/07/08 14:45:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable 18/07/08 14:45:35 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032 18/07/08 14:45:37 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032 18/07/08 14:45:37 INFO input.FileInputFormat: Total input paths to process: 1 18/07/08 14:45:37 INFO mapreduce.JobSubmitter: number of splits:1 18/07/08 14:45:37 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1531027983712_0006 18/07/08 14:45:38 INFO impl.YarnClientImpl: Submitted application_1531027983712_0006 18/07/08 14:45:38 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1531027983712_0006/18/07/08 14:45:38 INFO mapreduce.Job: Running job: job_1531027983712_0006 18/07/08 14:45:49 INFO mapreduce.Job: Job job j631027983712_0006 running in uber mode: false 18/07/08 14:45:549 INFO mapreduce.Job: map 100% reduce 0% 18/07/08 14:45:58 INFO mapreduce.Job: map 100% reduce 0% 18/07/08 14:46:08 INFO mapreduce.Job: map 100% reduce 100% 18/07/08 14:46:09 INFO mapreduce.Job: Dob job_1531027983712_0006 completed successfully 18/07/08 14:46:09 INFO mapreduce.Job: Counters: 49

File System Counters

FILE: Number of bytes read=56

FILE: Number of bytes read=56

FILE: Number of bytes read=561

FILE: Number of bytes read=561

FILE: Number of bytes read=562
```

```
File Input Format Counters
Bytes Read=49
File Output Format Counters
Bytes Written=29
The standard deviation is: 1.6329931618554514
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /outputstanddeviation
18/07/08 14:48:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
Found 2 items
-rw-r-r-- 1 acadgild supergroup 0 2018-07-08 14:46 /outputstanddeviation/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 29 2018-07-08 14:46 /outputstanddeviation/part-r-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /outputstanddeviation/part-r-00000
18/07/08 14:49:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
count 6
length 42
square 310
[acadgild@localhost ~]$
```