ASSIGNMENT 3.4

Ques 2) Explain the importance of Namenode in Hadoop cluster

Ans 2)

HDFS works by breaking large files into smaller pieces called *blocks.* The

blocks are stored on data nodes, and it is the responsibility of the NameNode

to know what blocks on which data nodes make up the complete file. The

NameNode also acts as a “traffic cop,” managing all access to the files, including reads, writes, creates, deletes, and replication of data blocks on the data nodes. The complete collection of all the files in the cluster is sometimes referred to as the file system *namespace.* It is the NameNode’s job to manage this namespace.

Even though a strong relationship exists between the NameNode and the

data nodes, they operate in a “loosely coupled” fashion. This allows the

cluster elements to behave dynamically, adding (or subtracting) servers as

the demand increases (or decreases). In a typical configuration, you find one

NameNode and possibly a data node running on one physical server in the

rack. Other servers run data nodes only.

Data nodes are not very smart, but the NameNode is. The data nodes constantly

ask the NameNode whether there is anything for them to do. This

continuous behavior also tells the NameNode what data nodes are out there

and how busy they are. The data nodes also communicate among themselves

so that they can cooperate during normal file system operations. This is

necessary because blocks for one file are likely to be stored on multiple data

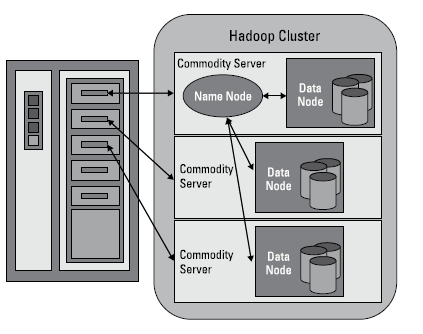
nodes. Since the NameNode is so critical for correct operation of the cluster,

it can and should be replicated to guard against a single point failure.

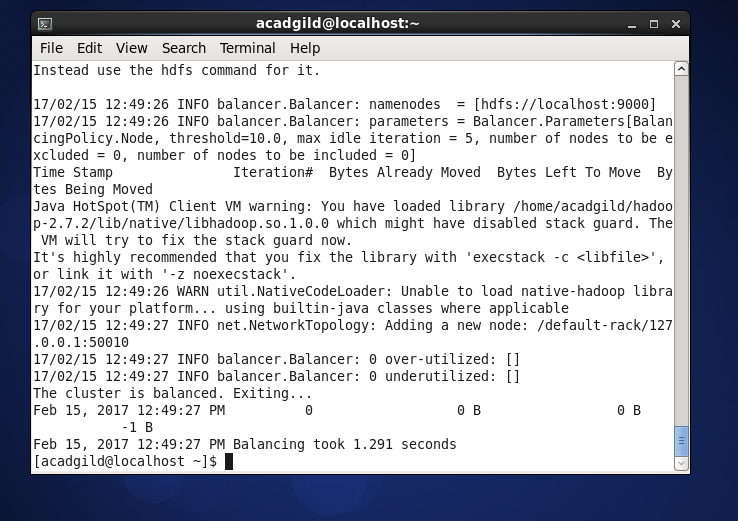
**NameNode** is the centerpiece of HDFS. **NameNode** is also known as the Master.**NameNode** only stores the metadata of HDFS – the directory tree of all files in the file system, and tracks the files across the cluster. **NameNode** is a single point of failure in **Hadoop** cluster

The namenode is the commodity hardware that contains the GNU/Linux operating system and the namenode software. It is a software that can be run on commodity hardware. The system having the namenode acts as the master server and it does the following tasks:

* Manages the file system namespace.
* Regulates client’s access to files.
* It also executes file system operations such as renaming, closing, and opening files and directories.

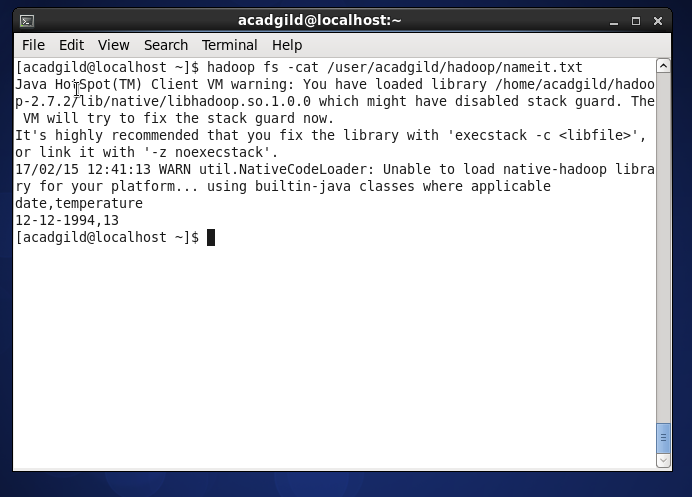


Ques 2.) Practice the beginners commands for HDFS from the below link

Ans 2.)

**Run a Cluster Balancing Utility**

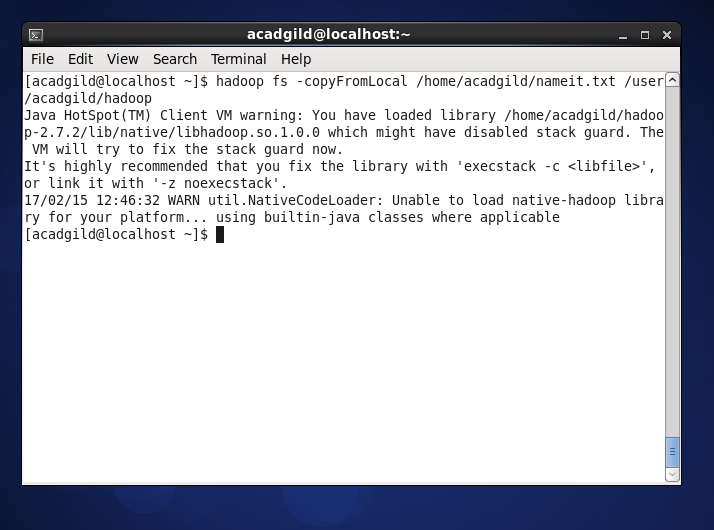
The command ‘balancer’ will check for work load on nodes in cluster and balance it.  
Syntax: **hadoop balancer**



**View contents of particular file**

The ‘cat’ command is used to display all the contents of a file.

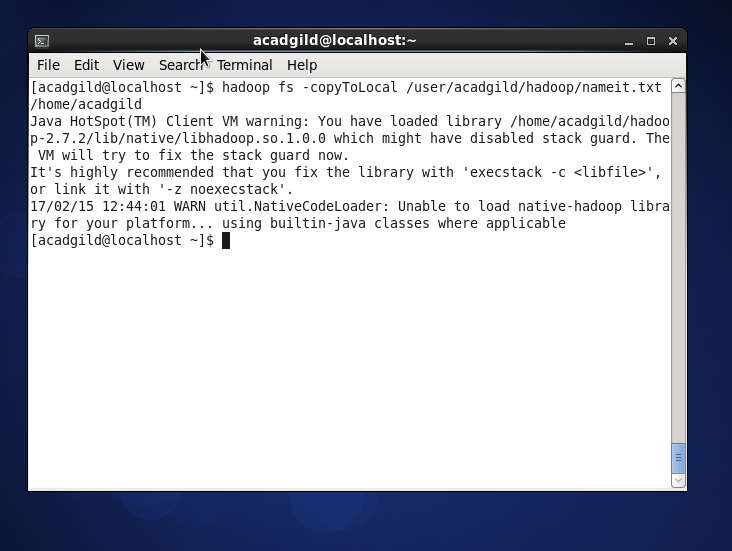
Syntax:**hadoop dfs –cat </path[filename]>**



**Duplicating a Complete File inside the HDFS.**

The ‘copyfromlocal’ command will copy file from the local file system to the HDFS.

Syntax:**hadoop dfs –copyFromLocal </source path> </destination path>**



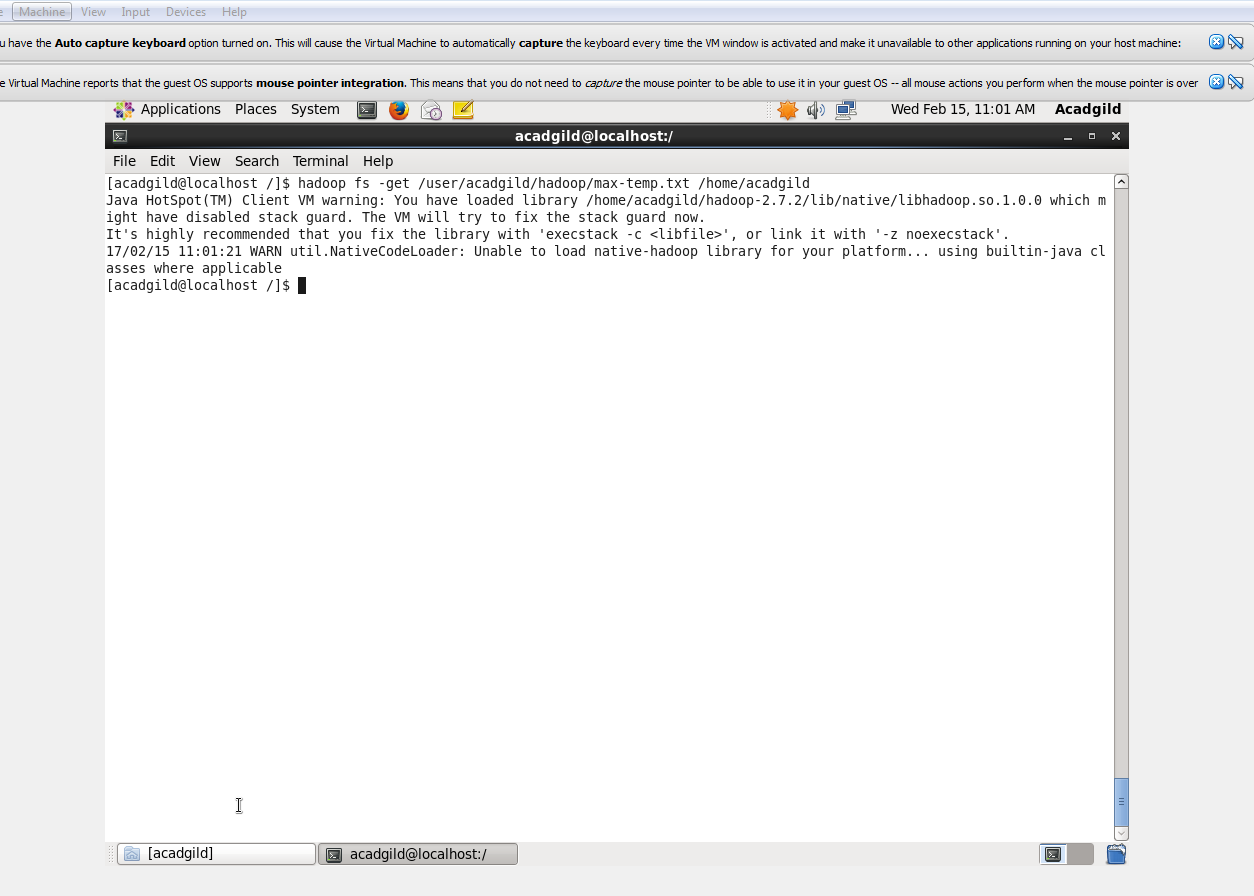
Copy to local from hapoop to localhost



**Check Directory Space in HDFS**

The command will show the file size occupied by file inside cluster.

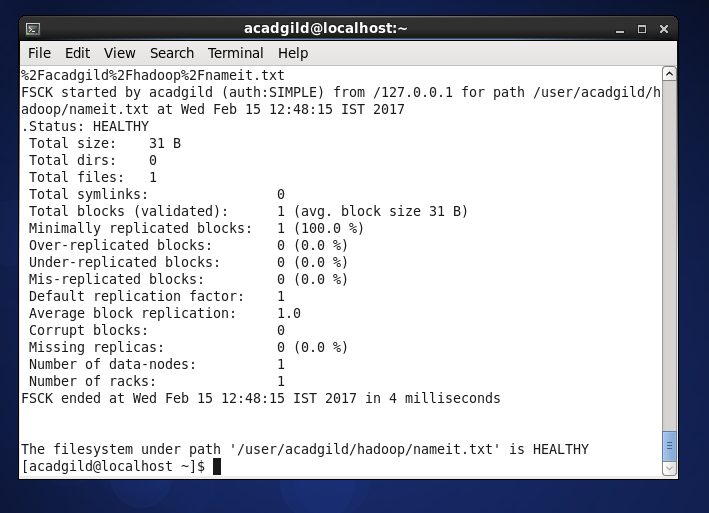
Syntax: **hadoop dfs -du -s -h </file path>**



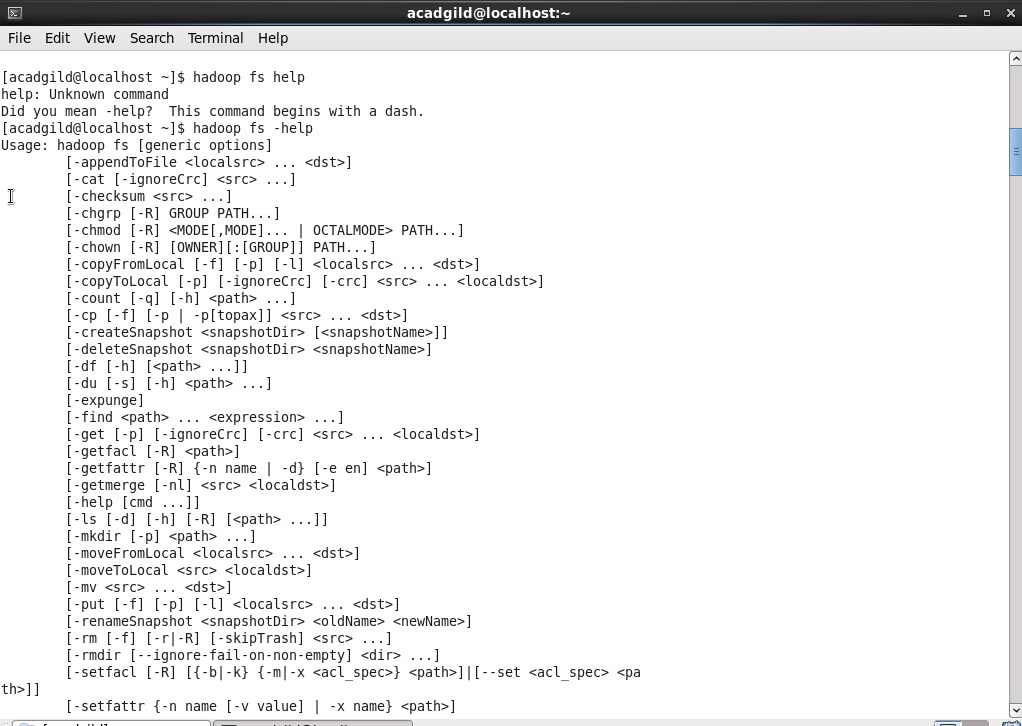
**Get Command**

The ‘get’ command copies the entire contents of the mentioned file to the local drive.

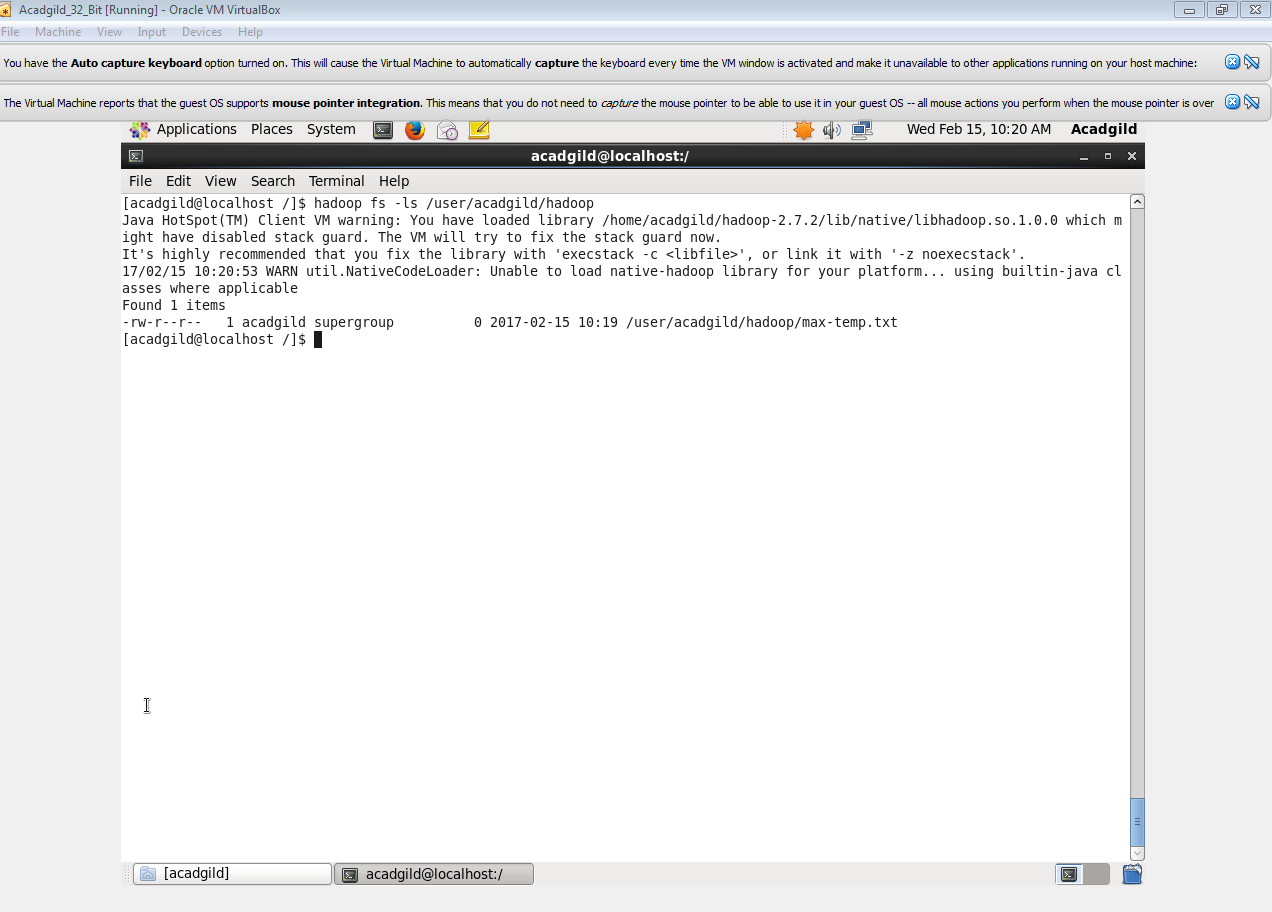
Syntax: **hadoop dfs –get </source path> </destination path>**



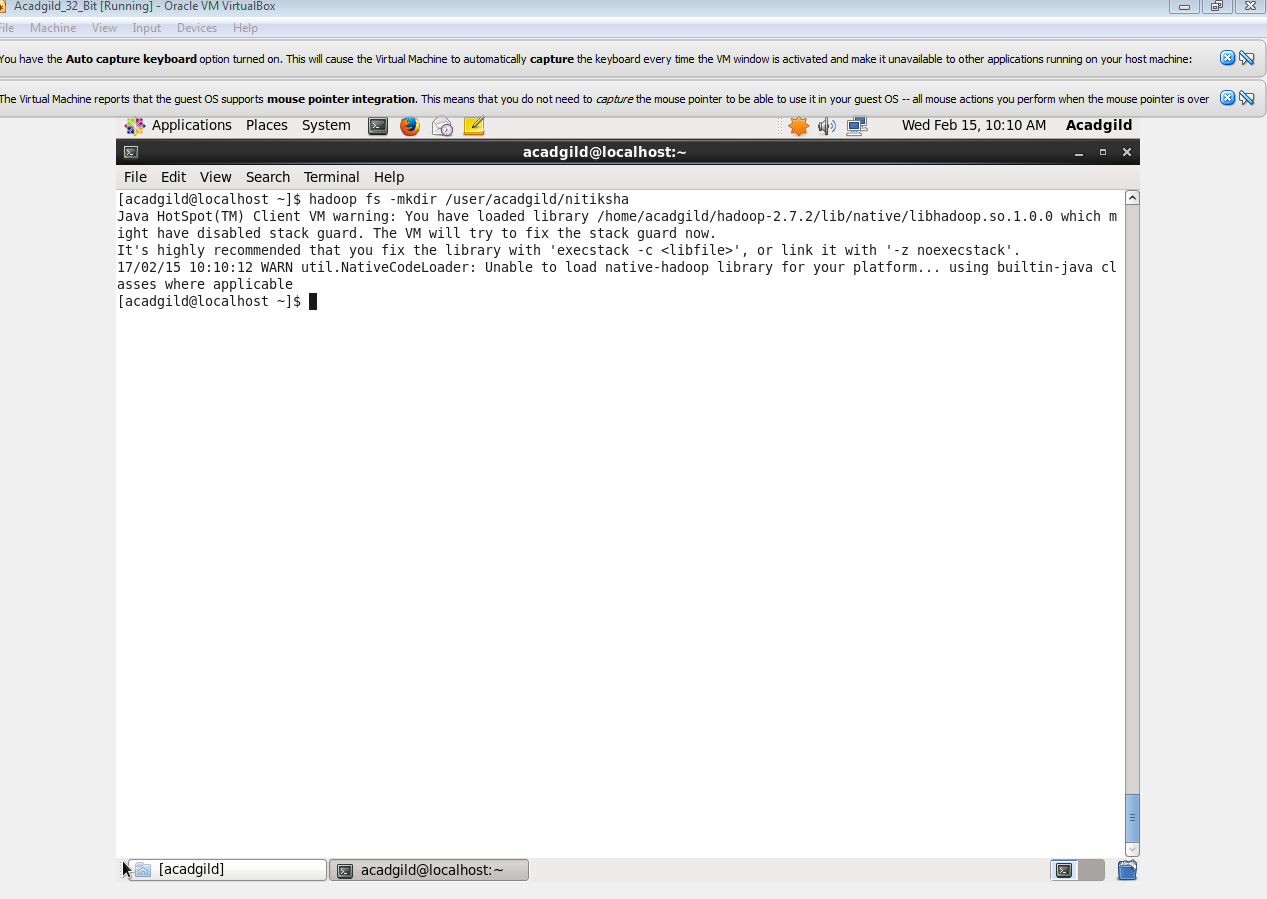
Check status



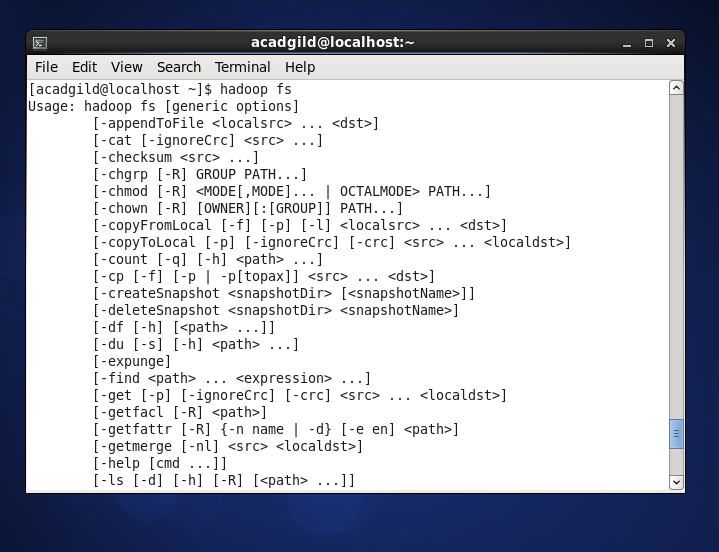
Help command



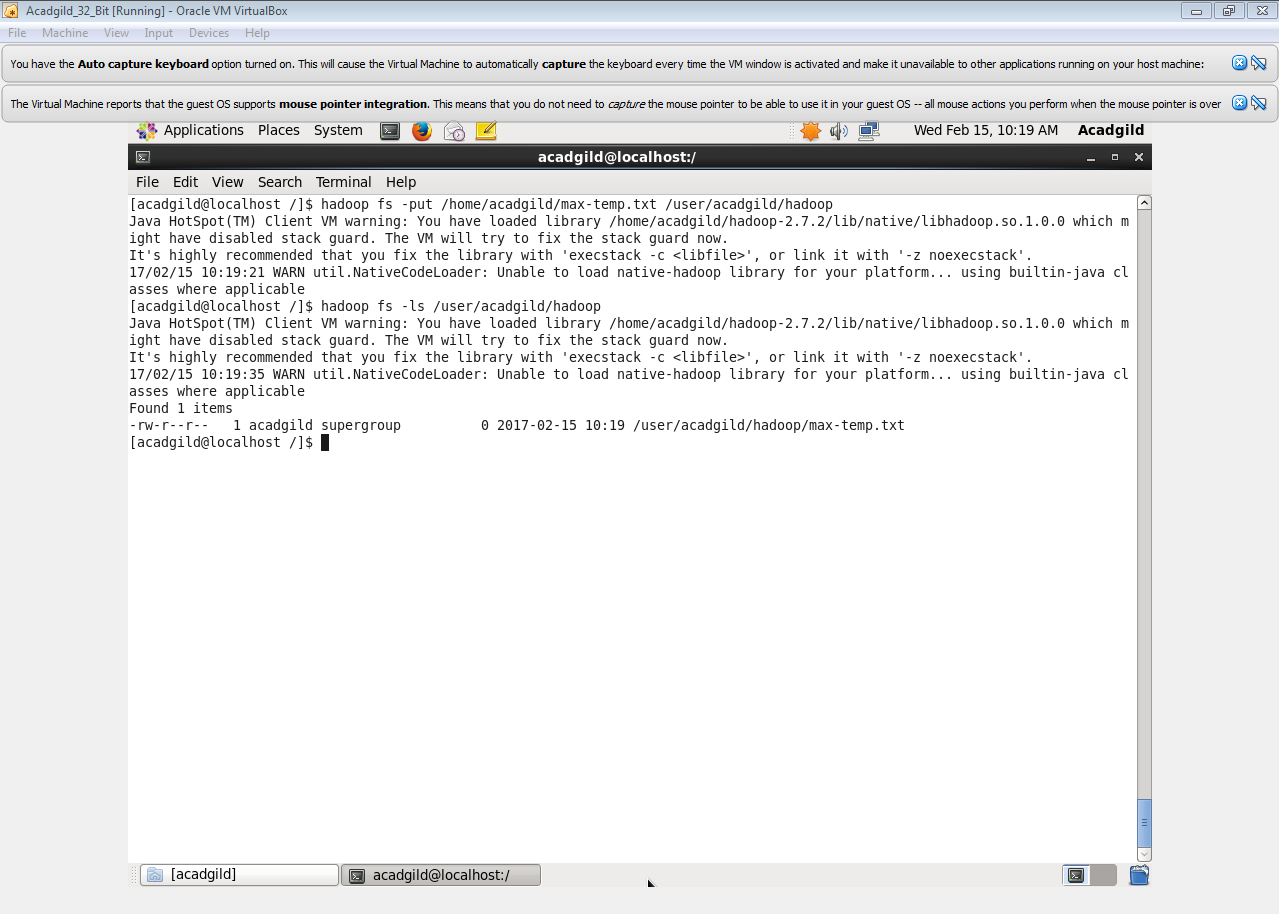
List all contents



Make directory command is shown



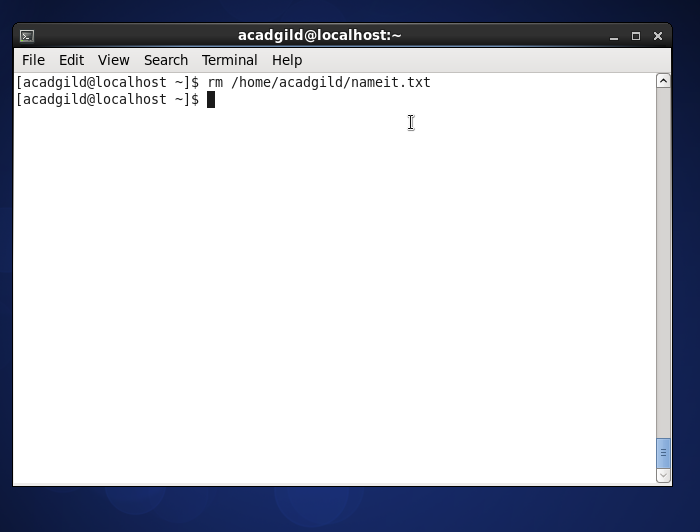
Hadoop fs



**Put Command**

The ‘put’command feeds the data in to the HDFS.

Syntax:**hadoop dfs –put </source path> </destination path>**



Remove directory ommand