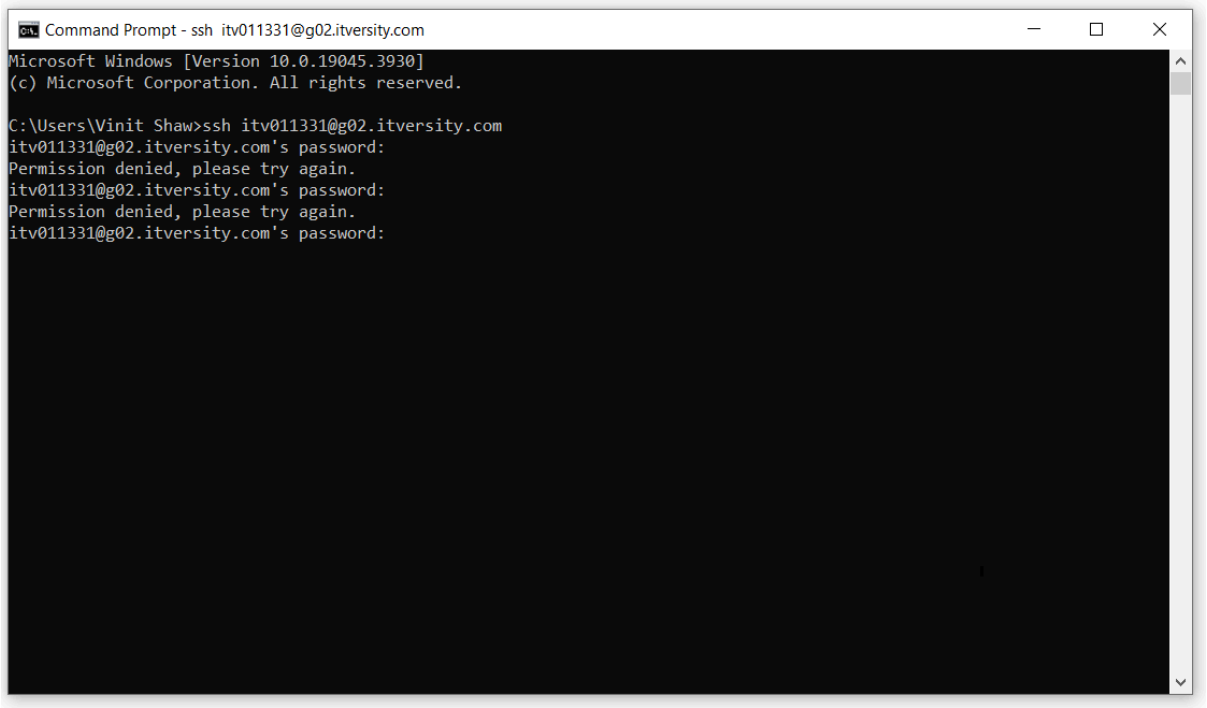
Q1. Getting permission denied error while entering password in terminal as shown in below screenshot.



Ans:

While mentioning the password please follow below steps.

1. "Right-click" to paste your password.

2. Then hit "enter".

Q2. ERROR 2003 (HY000): Can't connect to MySQL server on 'itversity.com' (110)

Ans:

Inorder to use mysql lab in itversity. Please follow the below mentioned steps:

1. mysql -u retail_user -h ms.itversity.com -p;

password : itversity

2. use retail_export

Now start creating the tables in it.

Q3. How does replication of block size take place without it being replicated on the same node twice?

Ans:

When a file is stored in HDFS, it is divided into fixed-size blocks (typically 128 MB or 256 MB). Each block is replicated across multiple DataNodes in the cluster.

By default, HDFS replicates each block three times. The replication factor (usually set to 3) determines how many copies of each block are stored in the cluster. These replicas are stored on different nodes to ensure that if one node goes down, there are still other copies of the data available.

Q4. History server is not working for me.

Ans:

Please clear all your browser history and try again

Url: http://m02.itversity.com:18080/

Q5: What is the difference b/w hadoop node and gateway node? Why can't we simply copy from /home/username to /user/username or vice versa . These are two different paths on the same machine . Why is /user/username treated as hadoop ?

Ans:
In Hadoop, a distributed storage and processing framework, the concepts of "Hadoop node" and "gateway node" refer to different roles and responsibilities within the Hadoop ecosystem.

Hadoop Node:
1. A Hadoop cluster is typically composed of multiple nodes, where each node is a physical or virtual machine that contributes storage and processing power to the cluster.
2. Nodes in a Hadoop cluster are divided into different roles, such as NameNode, DataNode, ResourceManager, and NodeManager, depending on the Hadoop ecosystem component they are running.
3. The Hadoop nodes collectively store and process data in a distributed manner, allowing for parallelized and scalable data processing.

Gateway Node:
1. A gateway node, sometimes referred to as a client node, is a machine that is not part of the Hadoop cluster but is used to interact with the cluster.
2. Users typically perform tasks like submitting MapReduce jobs, accessing Hadoop Distributed File System (HDFS), or running other Hadoop-related commands from the gateway node.

3. The gateway node serves as an interface between users and the Hadoop cluster, providing a convenient way to interact with the distributed storage and processing capabilities of Hadoop.

Now, regarding the question of copying from /home/username to /user/username and why /user/username is treated as the Hadoop side:

In Hadoop, the HDFS (Hadoop Distributed File System) is the primary storage system. The /user/username path is within HDFS and represents the user's home directory in the distributed file system.

The /home/username path, on the other hand, typically refers to the home directory on the local file system of the machine where you are running Hadoop commands. This is outside the HDFS.

When you are working with Hadoop, it's important to understand that Hadoop operates on data stored in HDFS, not the local file system. When you copy data using Hadoop commands, such as hadoop fs -copyFromLocal or hadoop fs -copyToLocal, you are interacting with HDFS, and the paths specified are assumed to be HDFS paths unless explicitly specified otherwise.

In summary, /user/username is treated as the Hadoop side because it refers to the user's home directory within the HDFS, which is the distributed file system used by Hadoop. The distinction is made to ensure that Hadoop operations are performed on the distributed data stored in HDFS rather than on the local file system.

Q6. Unable to create a file using touch command

```
[itv009066@g01 data]$ cd ..
[itv009066@g01 /]$ touch file1
touch: cannot touch ‘file1’: Permission denied
[itv009066@g01 /]$ touch file121
touch: cannot touch ‘file121’: Permission denied
[itv009066@g01 /]$ touch testfileNitesh11
touch: cannot touch ‘testfileNitesh11’: Permission denied
[itv009066@g01 /]$
```

Ans:

Here, the touch command is being used in the root directory. Please move to home directory:

cd /home/itv009066

And then start using the touch command in the home directory or in any other folder/directory in the home directory.

Q7: 1. whenever we do for example, 'hadoop fs -ls data' every time it collects the information/content from all the datanodes(wherever the blocks of the file are stored) and calculate size and list it to us ?
2.can we see how data is stored in datanodes ?

Ans:
In Apache Hadoop, when you execute the command hadoop fs -ls, it retrieves the metadata information about files and directories from the Hadoop Distributed File System (HDFS) NameNode. The NameNode maintains metadata such as file names, permissions, and the block locations for each file. It does not directly communicate with all the DataNodes to fetch this information.

The DataNodes are responsible for storing the actual data blocks. When you run a command like hadoop fs -ls data, it fetches metadata from the NameNode, and the NameNode provides information about the file's blocks and their locations. It doesn't directly retrieve the content of the blocks from the DataNodes during an ls operation.

Regarding your second question, directly viewing the contents of data stored on DataNodes is typically not done through Hadoop commands. HDFS is designed to provide distributed storage with data replication for fault tolerance, and direct access to the raw data on DataNodes is not a common operation.

If you want to inspect the content of a file stored in HDFS, you can use commands like hadoop fs -cat to display the content.

For example:
hadoop fs -cat data/file.txt

Q8: Sir in week 2 Assignment they say a third party will push one file in cloud orders.csv , this is where am I stuck, can you please guide

Ans:

You can find this orders.csv file in /data/retail_db/orders directory.

Now create folder landing and staging in your home directory.

Then using "cp -R /data/retail_db/orders landing" command you can move this orders.csv in the landing folder of your home directory.

After this you can follow the assignment**.**

Q9: What is a Hadoop file system and how is it different from a local file system?

Ans:

The Hadoop Distributed File System (HDFS) and a local file system differ in several key aspects:

1. HDFS stores data across multiple machines in a cluster, providing scalability and fault tolerance. A local file system stores data on a single machine, limiting its capacity and fault tolerance.

2. HDFS scales horizontally by adding more nodes to the cluster. A local file system's capacity is limited to the single machine it runs on.

3. HDFS replicates data blocks across nodes for fault tolerance. A local file system does not replicate data across machines.

4. HDFS is optimised for parallel processing of large datasets. A local file system does not support distributed data processing out of the box.

5. HDFS provides APIs and tools for accessing and managing data in a distributed environment. A local file system provides standard file I/O APIs for accessing data on a single machine.

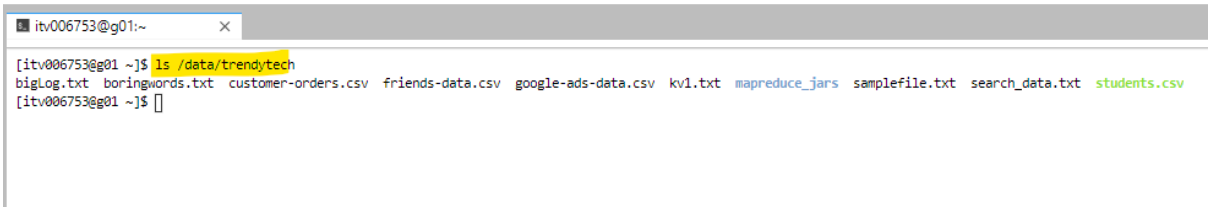
Overall, HDFS is designed for handling big data workloads across clusters of machines, providing scalability, fault tolerance, and support for distributed data processing, while a local file system is more suitable for storing and accessing data on a single machine.

Q10: I can't see the data folder in my terminal home directory, where the data file and other folders which sir is showing in video while explaining linux command.

Ans:
The data directory which Sumit sir is using is owned by Sumit sir and all the other users can only read the data of that directory.

The path of that directory is "/data/trendytech"

To see the content of that directory use the command “ls /data/trendytech”, refer to this screenshot.



Q11: Why do I get an error when I use -

[itv011153@go2 ~]$ cd /landing

Error - No such file or directory found. But when I use

[itv011153@go2 ~]$ cd landing/ - I login to the landing directory

Why is the prefix backslash not working but when I try /data it works?

Ans:

The behaviour you're observing is due to the presence or absence of the leading slash (/) and the specific structure of the file system.

When you use a command like "cd /landing", it is interpreted as an absolute path. This means that the system looks for the "landing" directory starting from the root directory ("/"). In your case the "landing" directory is not directly under the root directory, you will get an error such as "No such file or directory."

On the other hand, when you use cd landing/, it is interpreted as a relative path. This means that the system looks for the "landing" directory starting from your current working directory. And as the "landing" directory is a subdirectory of your current working directory, this command will work.

Q12: If I'm in the home directory and I am giving (ls -l) command it shows there are 0 items in it . but as per the video there is directory named 'data' in it and if we give its path (cd /data) and then give listing command we are able to see the content

Ans:

As you have not created any folder or file so you can not see anything.

Please create a folder using command “mkdir <foldername>”, then you can see this folder when you run the command ls -l.

Q13: Why pwd and cd commands are not working with hadoop fs?

Ans:

In HDFS (Hadoop Distributed File System), the cd (change directory) and pwd (print working directory) commands from Linux file systems are not applicable because HDFS is designed for distributed storage and processing of large-scale data. In HDFS, users always specify the full path to files and directories since there is no concept of a current working directory.

HDFS operates more like a distributed storage system where you reference files by their full paths starting from the root directory (/). You specify the full path to files or directories when working with HDFS commands.

For example, to list the contents of a directory in HDFS, you would use the hadoop fs -ls command followed by the full path to the directory:

hadoop fs -ls /path/to/directory

Similarly, to navigate to a specific directory in HDFS, you would provide the full path to that directory in your command. There is no need to change directories as you would in a local file system.

In summary, while pwd and cd are not directly supported in HDFS, you can achieve similar functionality using the hadoop fs command with the appropriate options and directory paths.