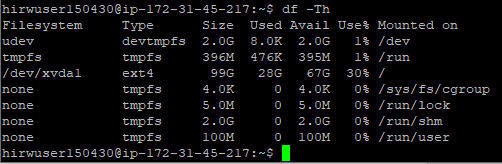
PRACTICAL-1: - Hadoop installation and basics of HDFS

**Task 1:** Identify the local file system in your Linux OS system. (Command followed by the output screenshot is expected).

Solution: **df -Th**



**Task 2:** Study of Hadoop installation. Study of Hadoop configurations files. Enlist the contents of the files core-site.xml and hdfs-site.xml

Solution:

**1) Content of core-site.xml**

<property>

<name>hadoop.tmp.dir</name>

<value>/data/hdfs/tmp</value>

<description>Where Hadoop will place all of its working files</description>

</property>

<property>

<name>fs.defaultFS</name>

<value>hdfs://master:9000</value>

<description>Where HDFS NameNode can be found on the network</description>

</property>

<property>

<name>hadoop.proxyuser.hduser.groups</name>

<value>\*</value>

<description>

What user groups are allow to connect to the HDFS proxy.

\* for all.</description>

</property>

<property>

<name>hadoop.proxyuser.hduser.hosts</name>

<value>\*</value>

<description>

What user hosts are allow to connect to the HDFS proxy.

\* for all.

</description>

</property>

**2) Content of hdfs-site.xml**

<property>

<name>dfs.replication</name>

<value>2</value>

<description>The default replication factor of files on HDFS</description>

</property>

<property>

<name>dfs.block.size</name>

<value>16777216</value>

<description>The default block size in bytes of data saved to HDFS</description>

</property>

<property>

<name>dfs.namenode.rpc-bind-host</name>

<value>0.0.0.0</value>

<description>

controls what IP address the NameNode binds to.

0.0.0.0 means all available.

</description>

</property>

<property>

<name>dfs.namenode.servicerpc-bind-host</name>

<value>0.0.0.0</value>

<description>

controls what IP address the NameNode binds to.

0.0.0.0 means all available.

</description>

</property>

<property>

<name>dfs.namenode.http-bind-host</name>

<value>0.0.0.0</value>

<description>

controls what IP address the NameNode binds to.

0.0.0.0 means all available.

</description>

</property>

<property>

<name>dfs.namenode.https-bind-host</name>

<value>0.0.0.0</value>

<description>

controls what IP address the NameNode binds to.

0.0.0.0 means all available.

</description>

</property>

<property>

<name>nfs.dump.dir</name>

<value>/tmp/.hdfs-nfs</value>

<description>A temporary working directory for files coming into the HDFS proxy.</description>

</property>

<property>

<name>nfs.metrics.percentiles.intervals</name>

<value>100</value>

<description>

Enable the latency histograms for read, write and commit requests.

The time unit is 100 seconds in this example.

</description>

</property>

<property>

<name>nfs.exports.allowed.hosts</name>

<value>\* rw</value>

<description>Host permissions for connecting to the proxy.</description>

</property>

<property>

<name>dfs.permissions</name>

<value>true</value>

<description>Enforce permissions</description>

</property>

<property>

<name>dfs.permissions.supergroup</name>

<value>hadoop</value>

<description>The name of the group of Hadoop super-users.</description>

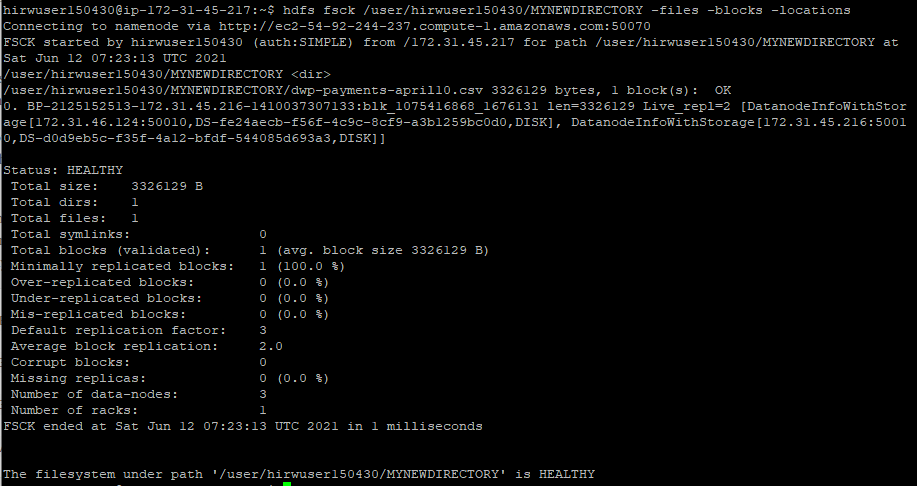
</property>

**Task 3:** Study and run following commands on the hadoop cluster and show the output.

JPS: to check out all the **Hadoop** daemons like DataNode, NodeManager, NameNode, and ResourceManager that are currently running on the machine.



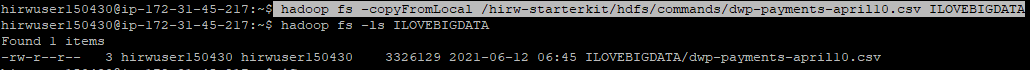
Fsck: to check health of the HDFS.



Touchz: to create an empty file.



copyFromlocal: to copy a file from local file system to HDFS.

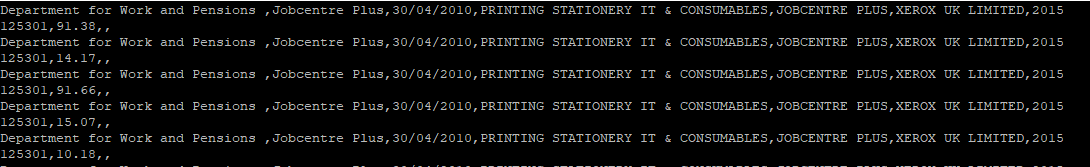


copyToLocal: to copy a file from HDFS to local file system.



cat: to print the content of a specific file.





moveFromLocal: to move a file from local file system to HDFS.

**Hadoop fs -moveFromLocal <local source> <destination>**

Put:

Similar to **copyFromLocal.**

Get:

Similar to **copyToLocal.**

Rmr: to remove a file or directory recursively.



Setrep: to change the replication factor of a file/directory in HDFS. By default the value is 3.

