**Exp 9 Installation of Hadoop and execution of a MapReduce program**

**13/09/24**

**Aim :**

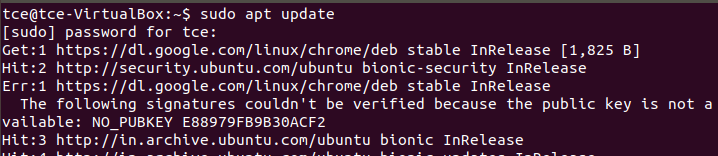
To install Hadoop and execute a mapreduce program using it.

**Description :**

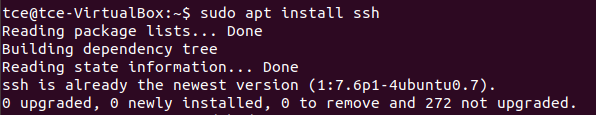
Open Oracle VM and start the ubuntu os.

Login and Go to terminal.

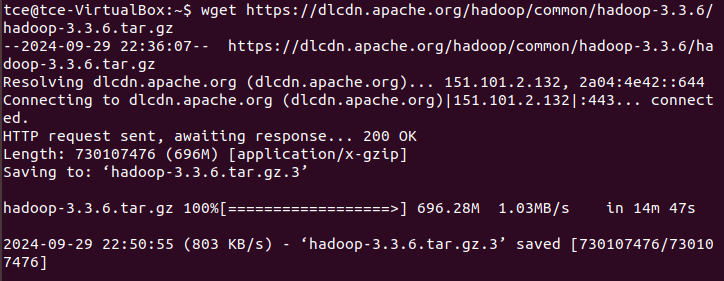
Update the OS.



Install ssh

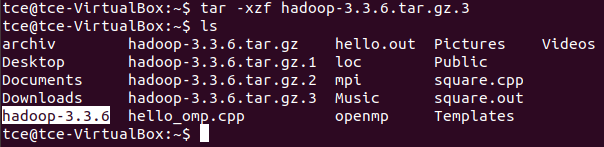
**

Install Hadoop



Unzip the downloaded zip file : tar -xzf Hadoop-3.3.6.tar.gz

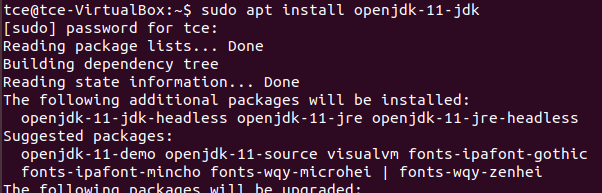
Check if it is unzipped using ls



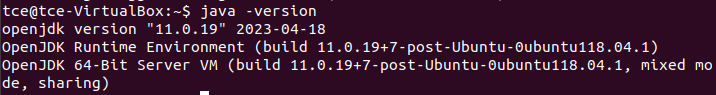
Rename the hadoop-3.3.6 to Hadoop: mv hadoop-3.3.6 hadoop



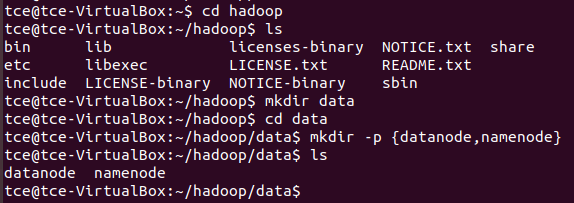
Install jdk version 11.



Check java version.



Create /hadoop/data directory and create datanode and namenode directories on that directory.



Find java and Hadoop path.

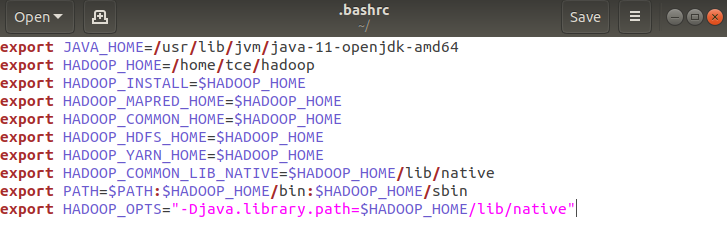




It will be used to edit bashrc file.

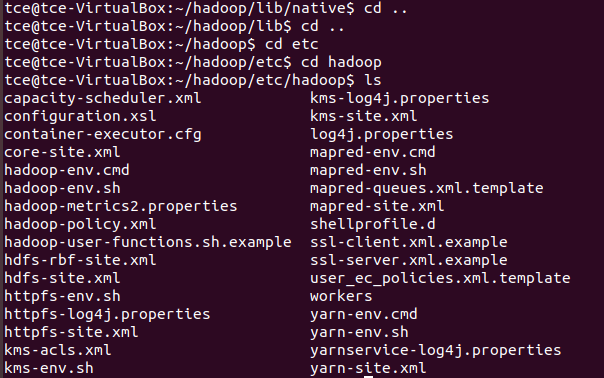
Open bashrc file in text editor.



Include below contents to bashrc file and save file.

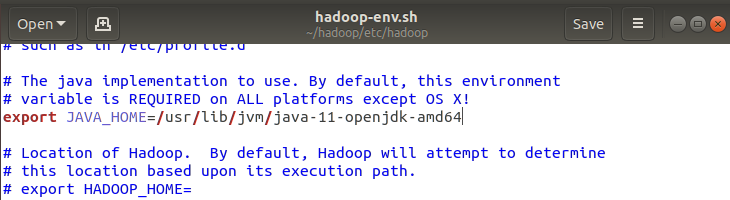
Apply the changes

Change the directory to /hadoop/etc/hadoop and list the files present in that directory.



Open hadoop-env.sh with gedit hadoop-env.sh. Go to Java Implementation part and paste JAVA\_HOME.





Configure below 4 files present in /hadoop/etc/hadoop directory.

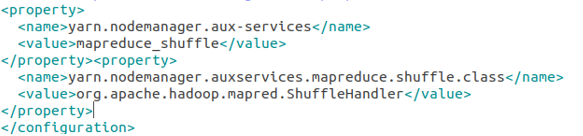




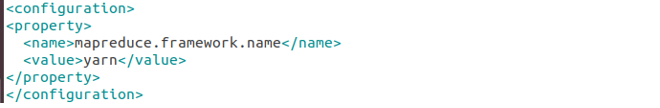




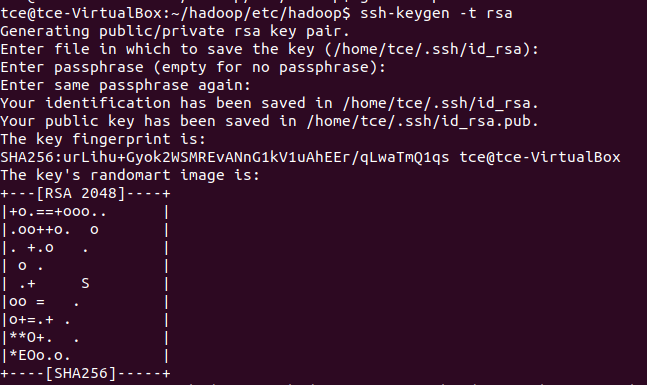






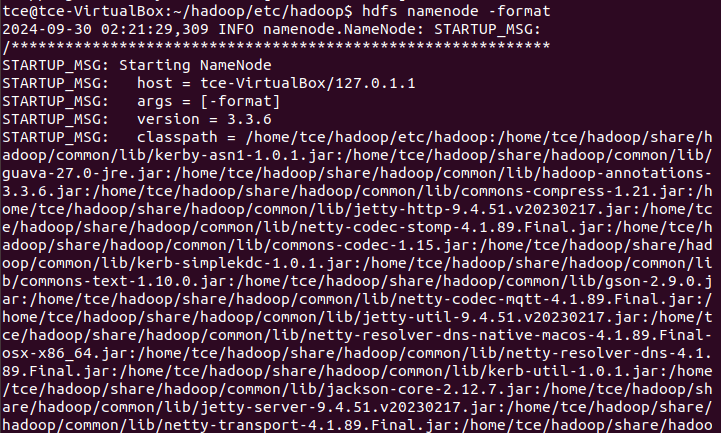


Generate key and make changes to key location.





Format the name node.

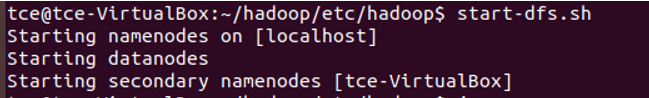


Start the namenode using the command start-dfs.sh

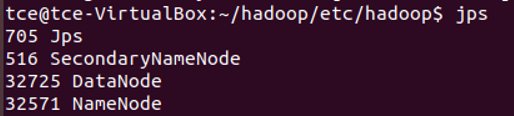
If there is an error, try installing ssh with sudo apt install ssh

Restart.

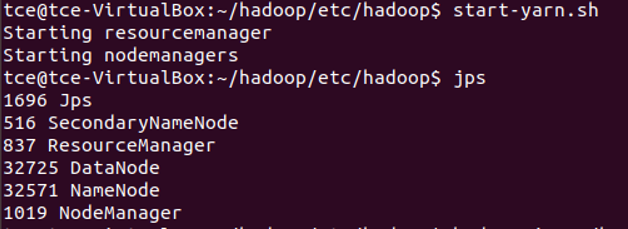
Then type start-dfs.sh



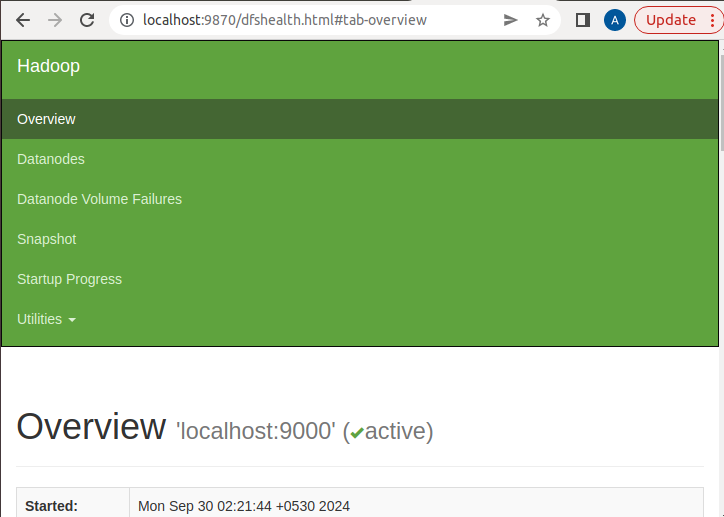
List the java processes running on the machine

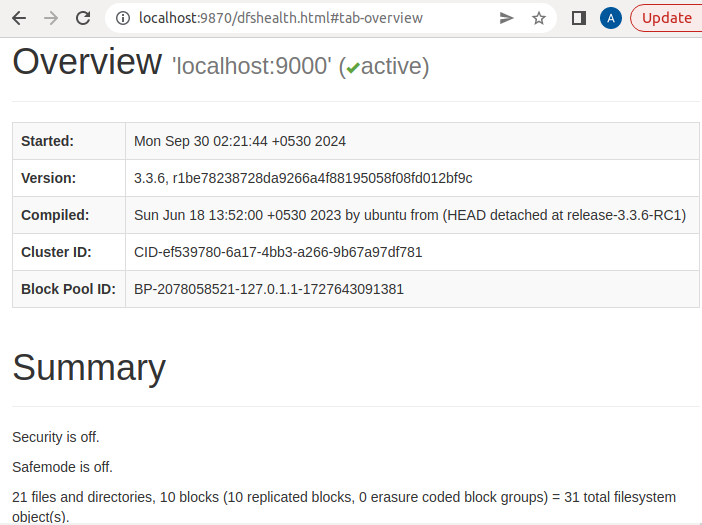


Start-yarn.sh and list the java processes running on the machine.



Open the browser and enter <http://localhost:9870/>



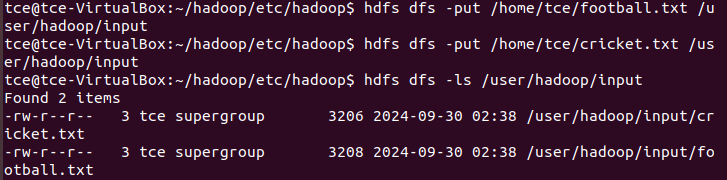


**Execution of mapreduce program**

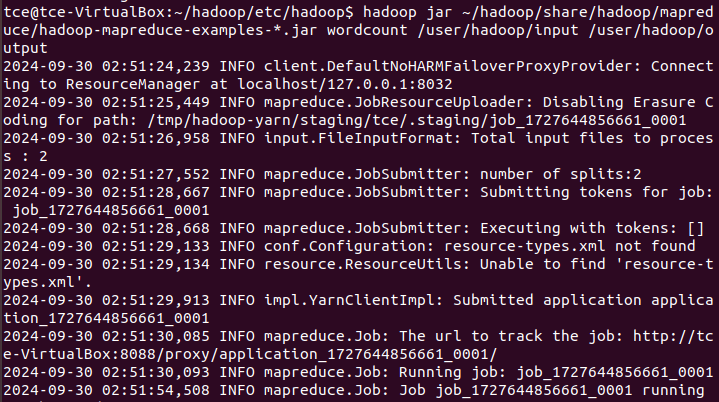
Make input directory.

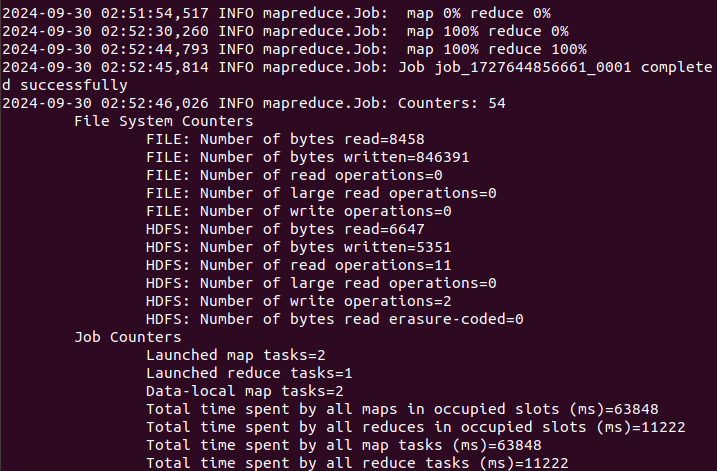


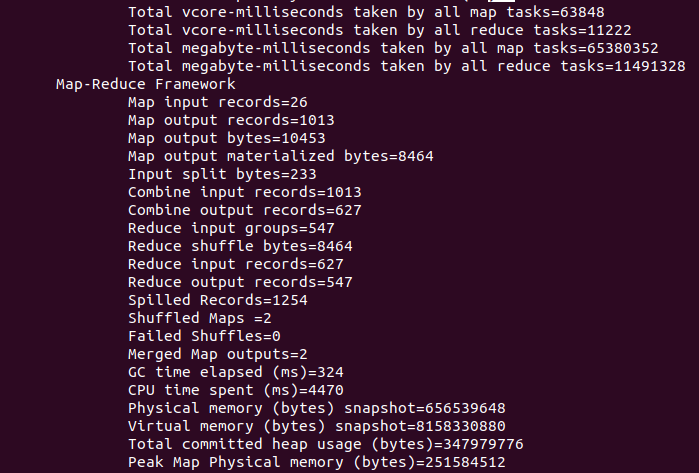
Put the text files to the input directory whose words will be counted.

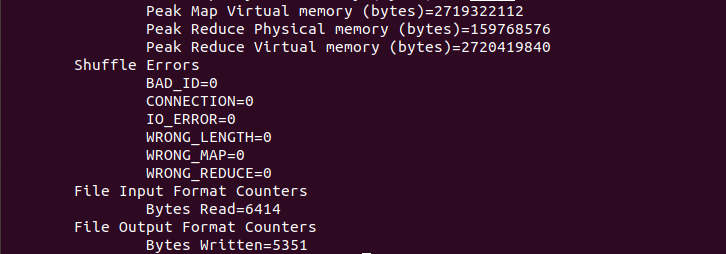


Run the wordcount command.

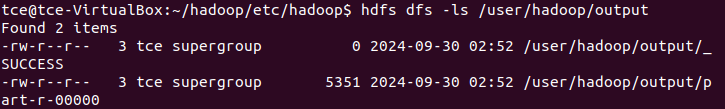




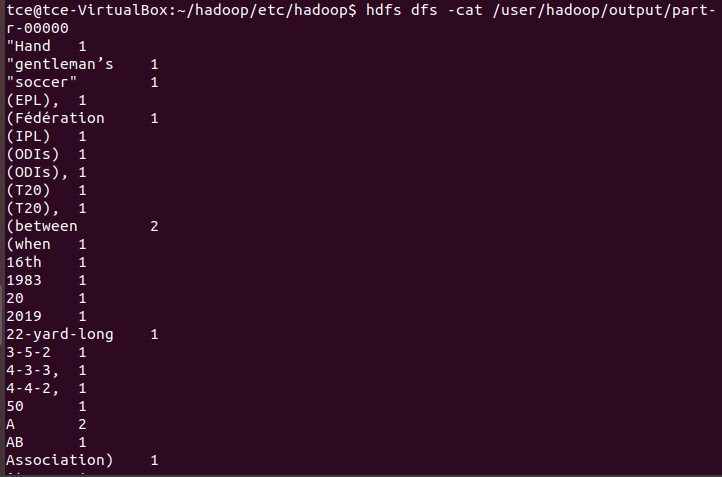




List the files in output directory.



Print the contents of part-r-00000



It prints the count of each words on both files.

**Result:**

Thus ,the installation of Hadoop and execution of mapreduce program have been successfully completed.