**Ex.No:** 9

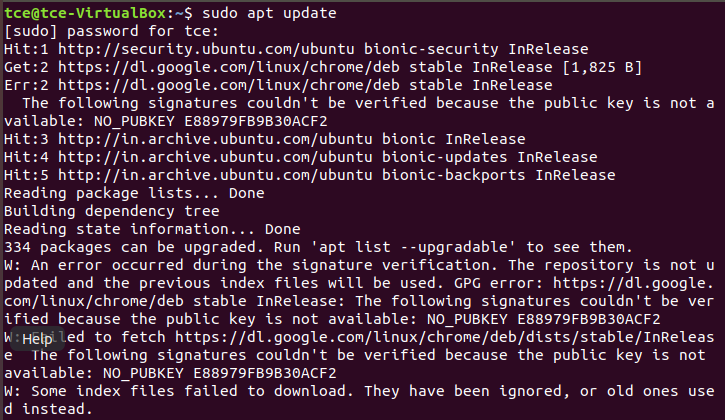
**Date:** 23.9.24 **Implementation of Hadoop and Execution of Map reduce Program**

**Aim:**

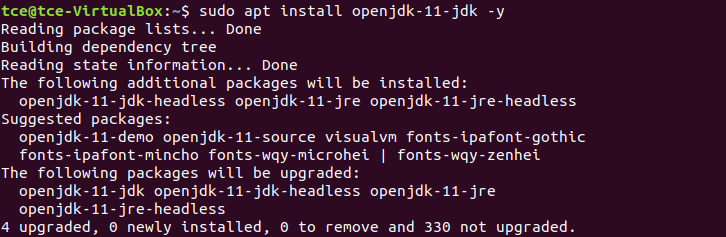
To install Hadoop in ubuntu and execute map reduce program.

**Procedure:**

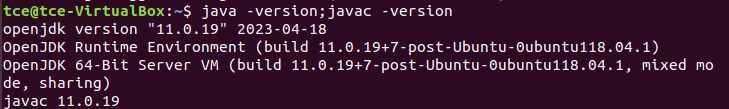
1. Update the system using sudo apt update command.



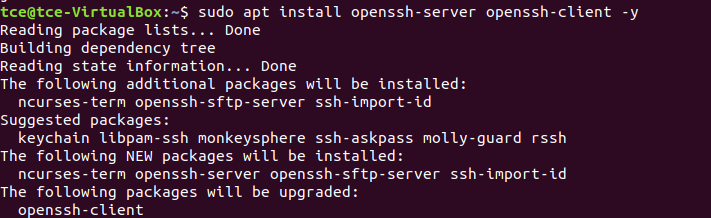
2. Install java 11 on the ubuntu system.



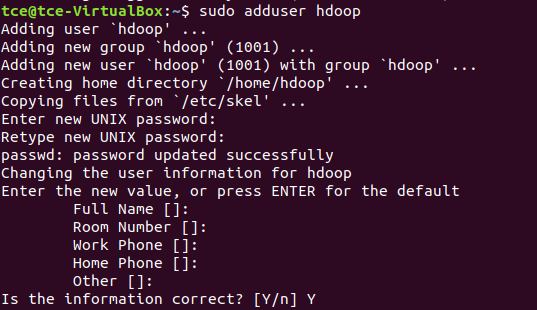
3. Confirm the installation by checking the java version.



4. Install openssh server and client in ubuntu.



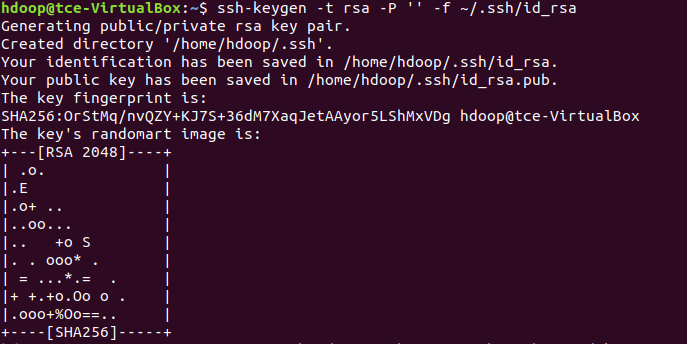
5. Add a new user hdoop to install and execute Hadoop programs



6. Switch to the newly created user and enter the corresponding password:



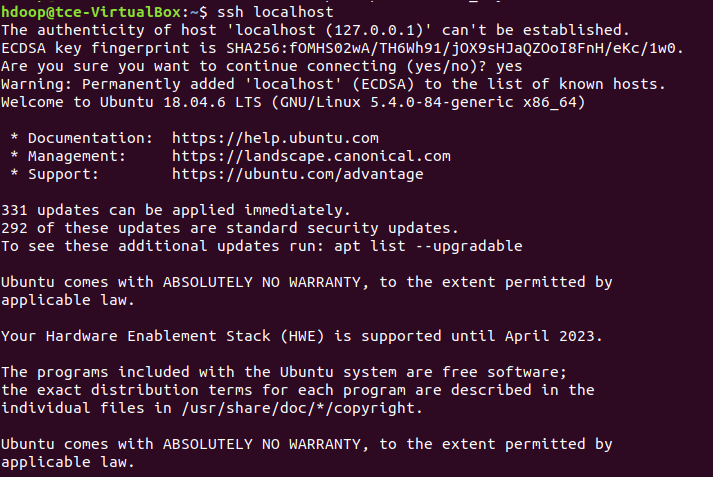
7. Enable passwordless ssh for Hadoop user. Generate a ssh key pair and define the location it is to be stored.



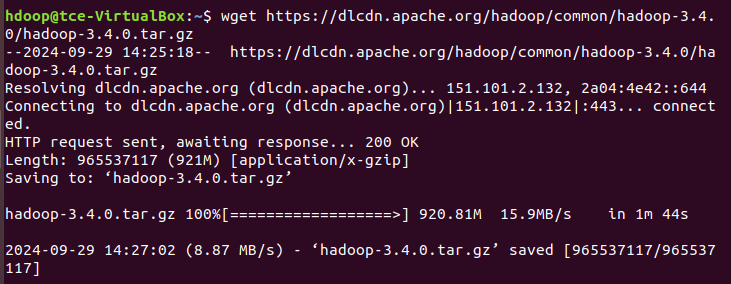
8. Use the cat command to store the public key as authorized\_keys in the ssh directory and set the file permissions for your user with the chmod command.



9. The new user can now SSH without entering a password every time.



10. Download Hadoop by visiting the download link using this command.



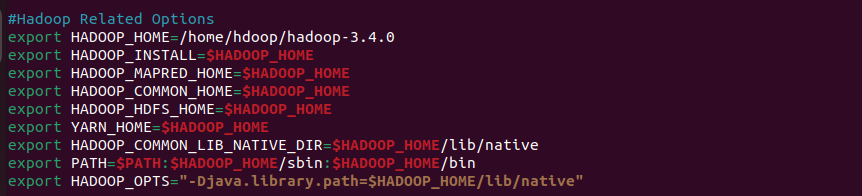
11. Extract the zip ile. The Hadoop binary iles are now located within the hadoop-3.4.0 directory.



12. Edit the .bashrc file to include Hadoop environment variables:



Add the following:



Apply the changes:

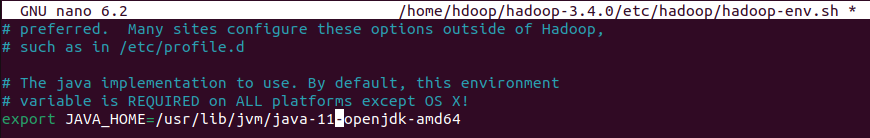


13. Configure Hadoop Files:

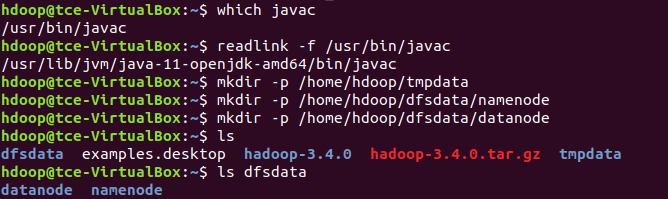
hadoop-env.sh



Uncomment and set the JAVA\_HOME path:



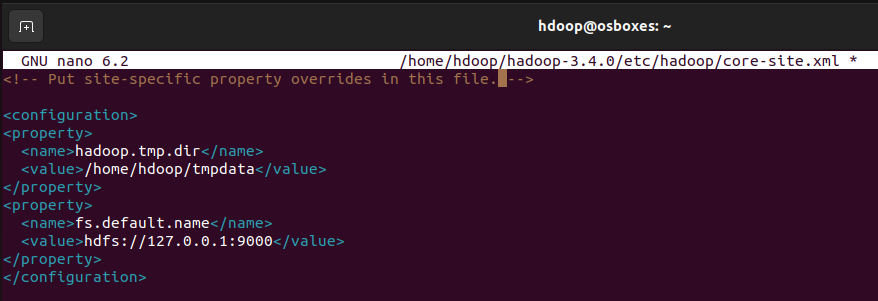
* Locate the correct Java path
* Use the provided path to find the OpenJDK directory with the following command:
* Create the Temporary Directory
* Create Directories for NameNode and DataNode



Edit core-site.xml File

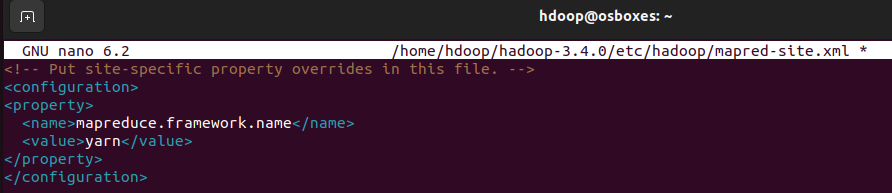
Edit hdfs-site.xml File





Edit mapred-site.xml File



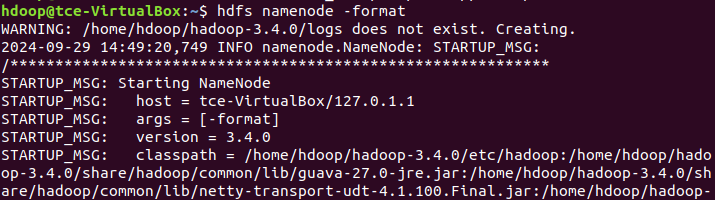


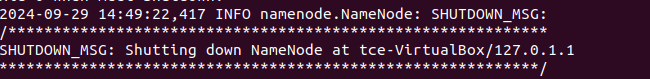
Edit yarn-site.xml File



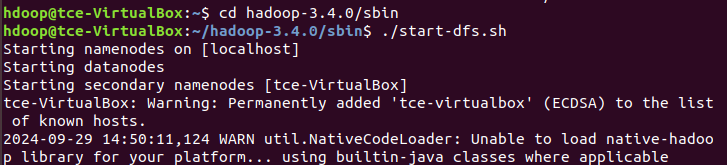


14. Format HDFS NameNode



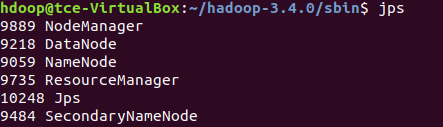


15. Navigate to Hadoop -version/sbin folder and start Hadoop cluster.



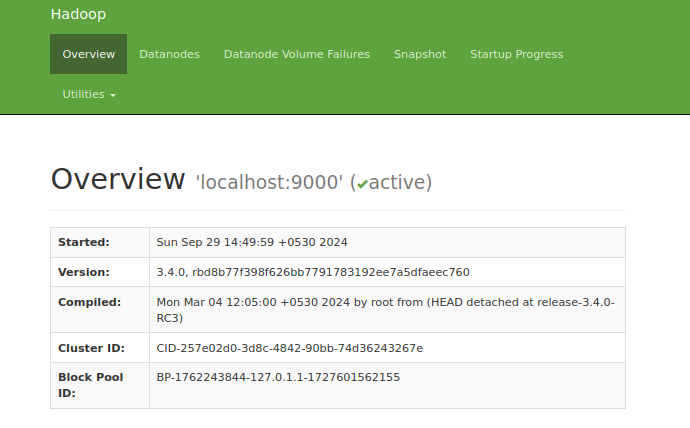


16. Run the following command to check if all the daemons are active and running as Java processes.



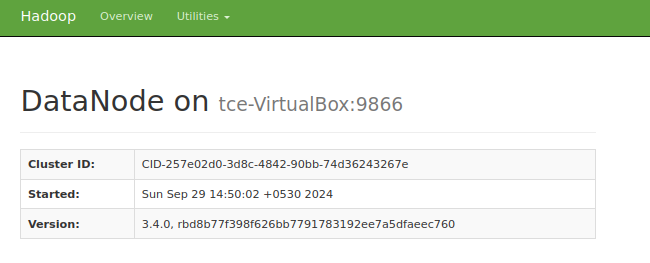
Access Hadoop from Browser

http://localhost:9870 The default port number 9870 gives you access to the Hadoop NameNode UI:

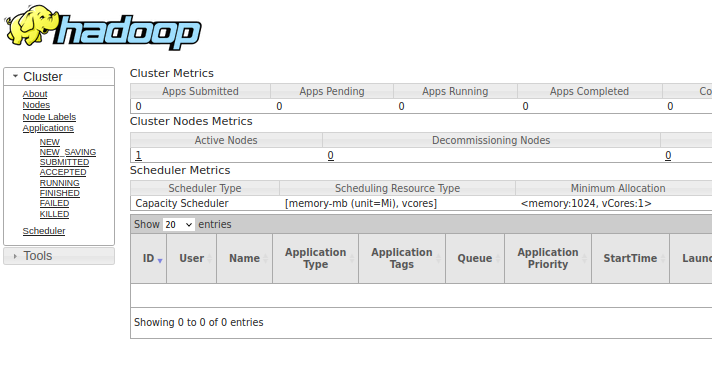


The default port 9864 is used to access individual DataNodes directly from your browser:

<http://localhost:9864>



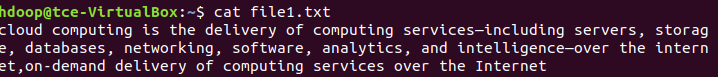
The YARN Resource Manager is accessible on port 8088:http://localhost:8088



**MapReduce WordCount program in Hadoop:**

1. Create a sample text file for the WordCount program.



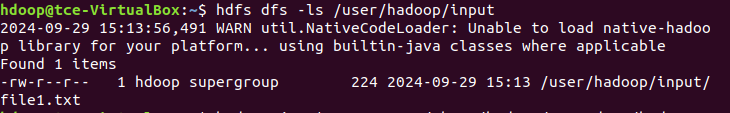


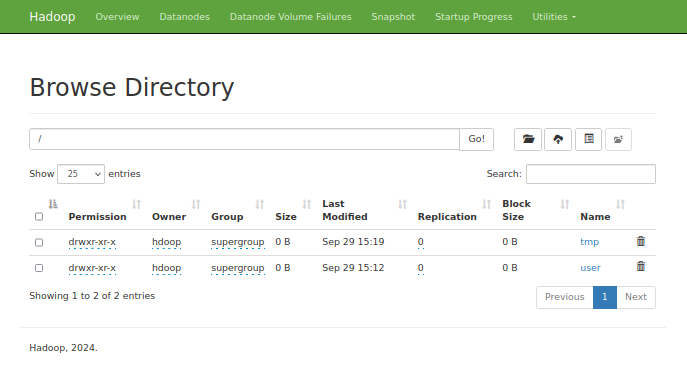
2. Create a Directory in HDFS



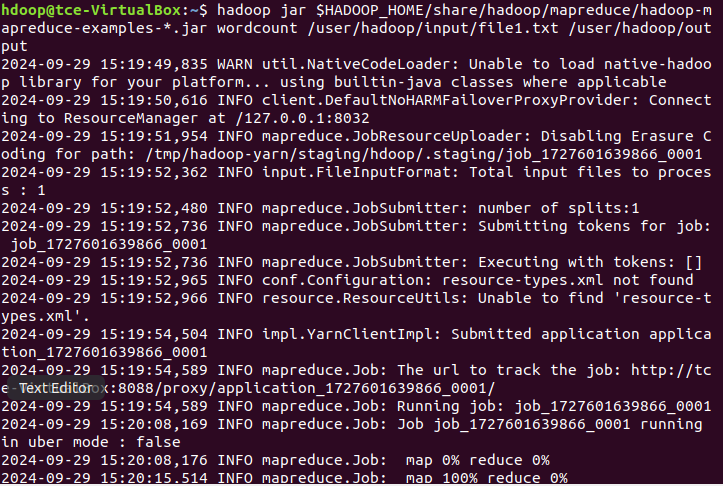
3. Upload your input text file to Hadoop’s HDFS:



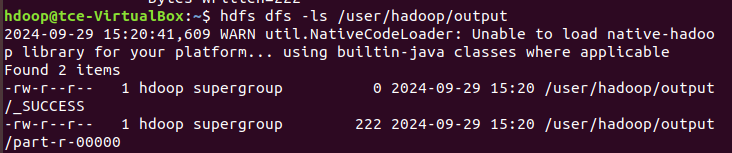




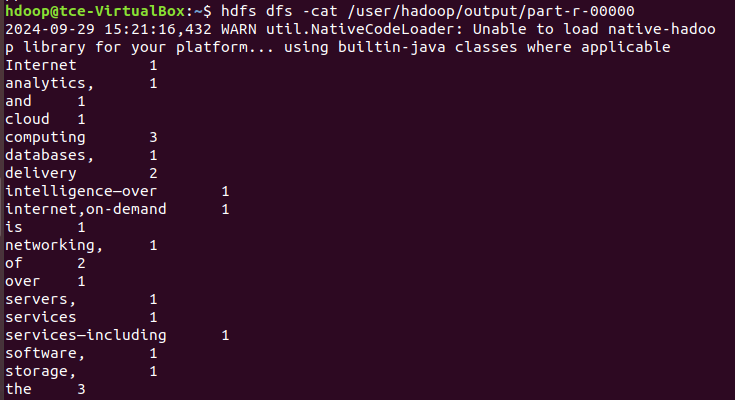
4. Run the wordcount program : execute the built-in WordCount program provided by Hadoop



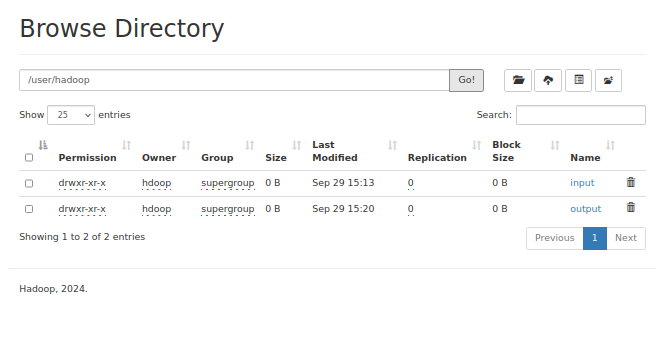
5. Check the Output

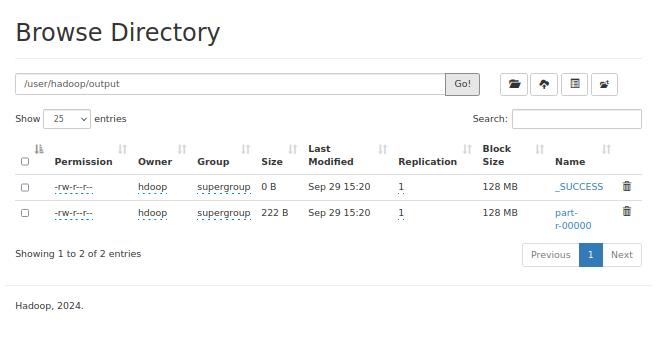


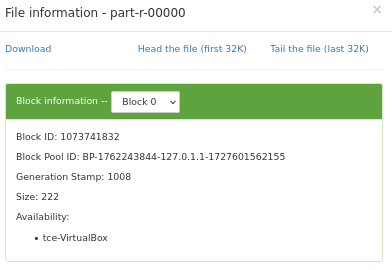
6. Display the result of the WordCount program by displaying the content of the output file using the cat command.

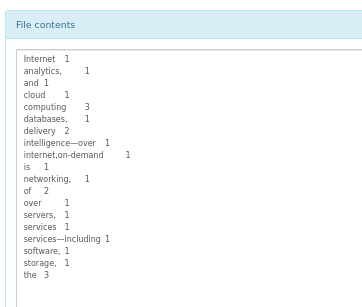


7. Access the NameNode Web Interface > Navigate to the "Utilities" Section > Go to the File System Browser > Locate the Output Directory(/user/Hadoop/output) > View the Output Files > wordcount program result will be displayed in the site.









**Result:**

Thus the installation of Hadoop and map reduce program has been executed successfully.