Name: Tania Rajabally

Branch: Comps

Batch: C

Roll no: 43

UID:2017130047

Date: 14/08/2020

**Experiment 1**

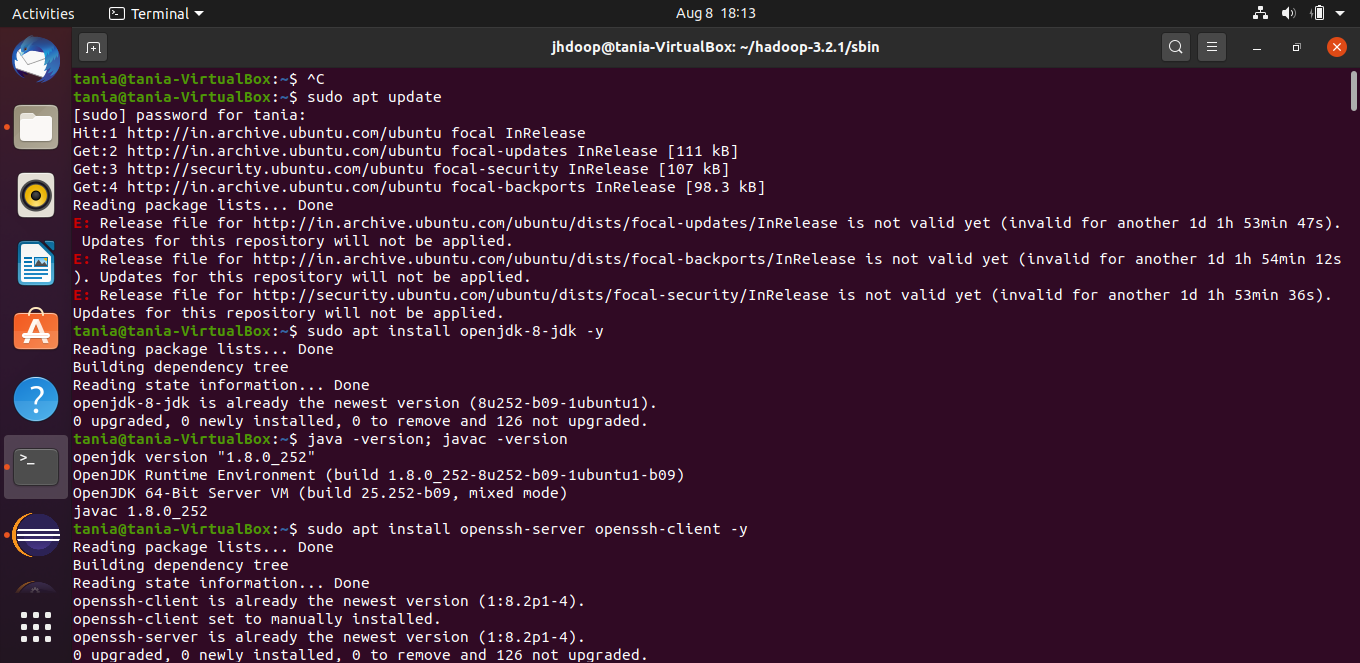
**Aim:** 1)Install Ubuntu, JDK, Eclipse, Hadoop

2) To solve Word Count Problem

**PART 1: INSTALLATION**

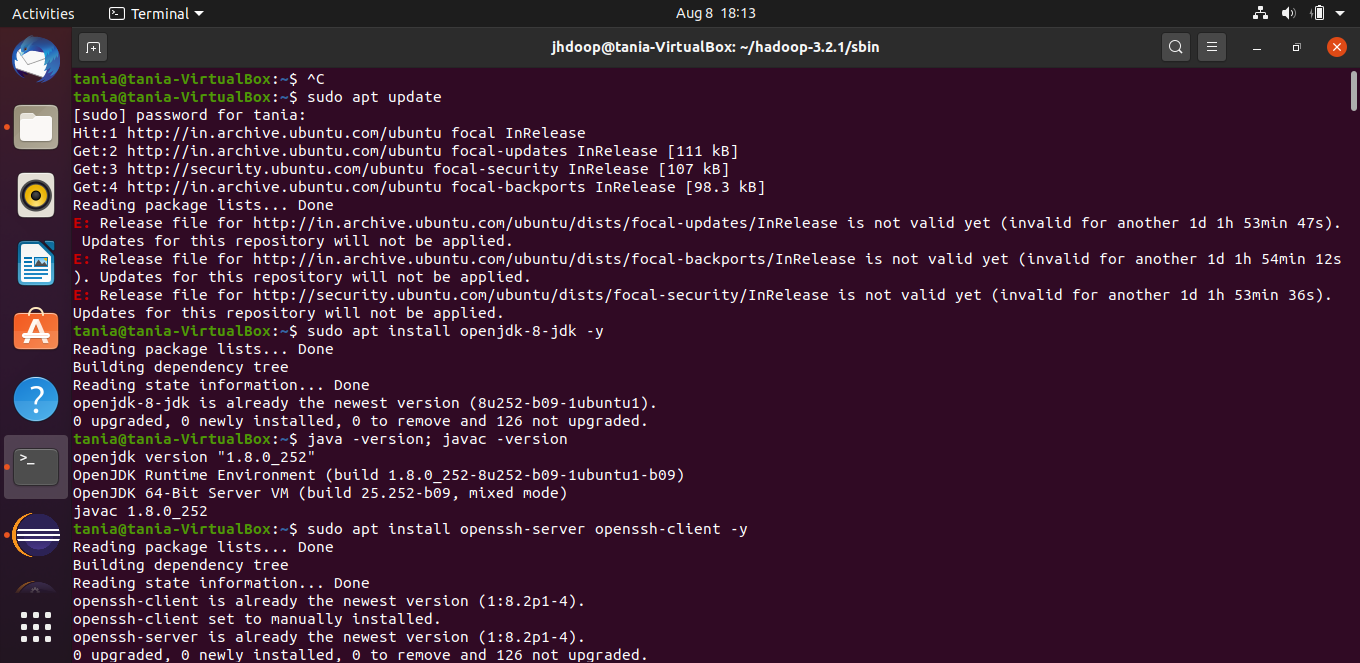
**Steps:**

Step 1: Install Ubuntu

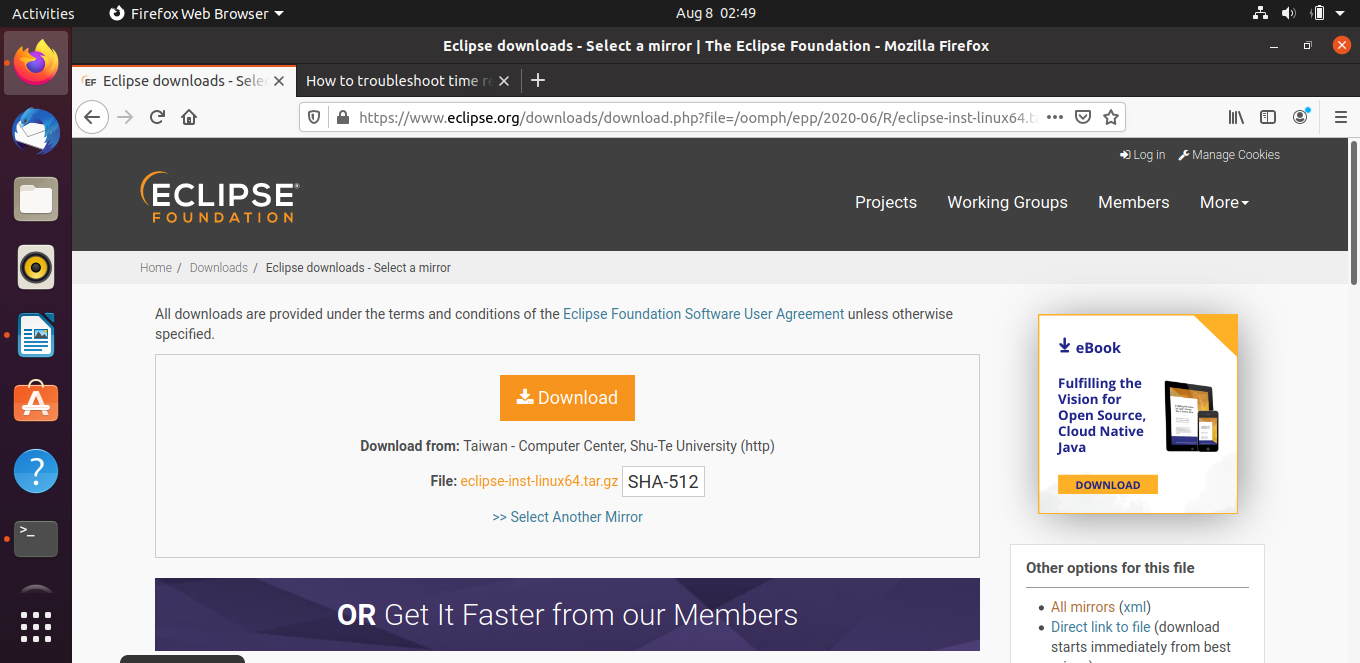


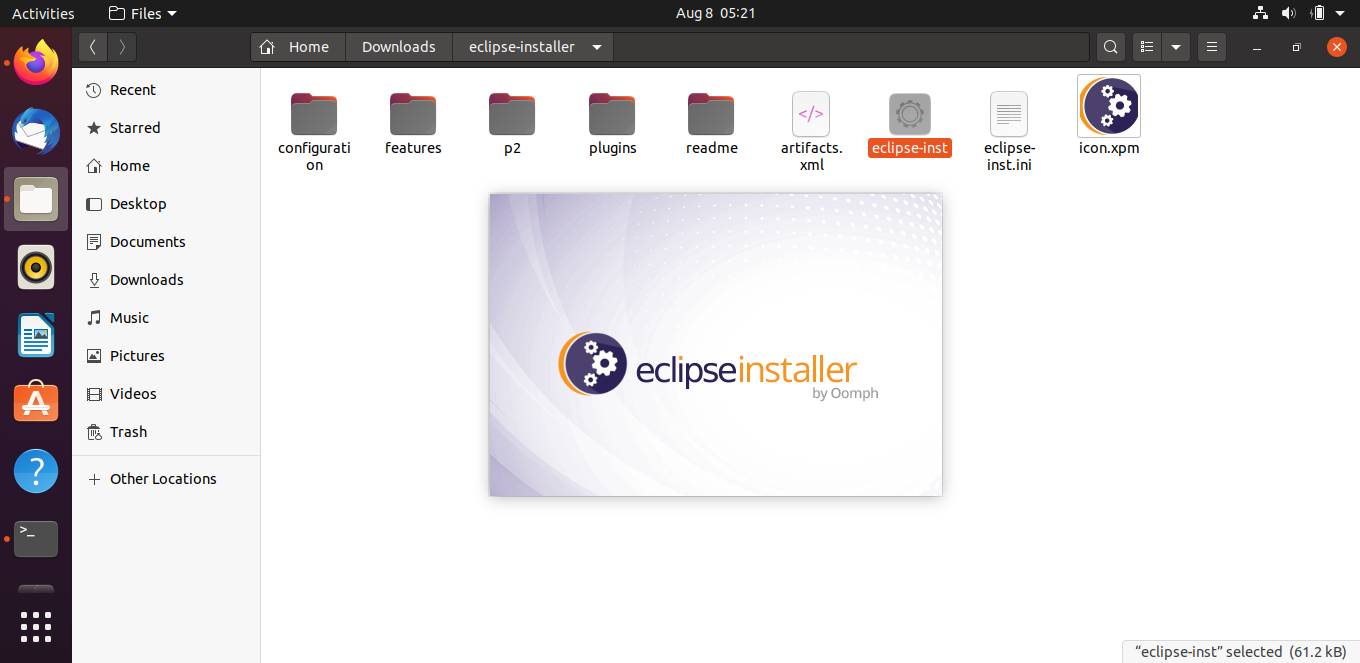
Step 2: Install JDK

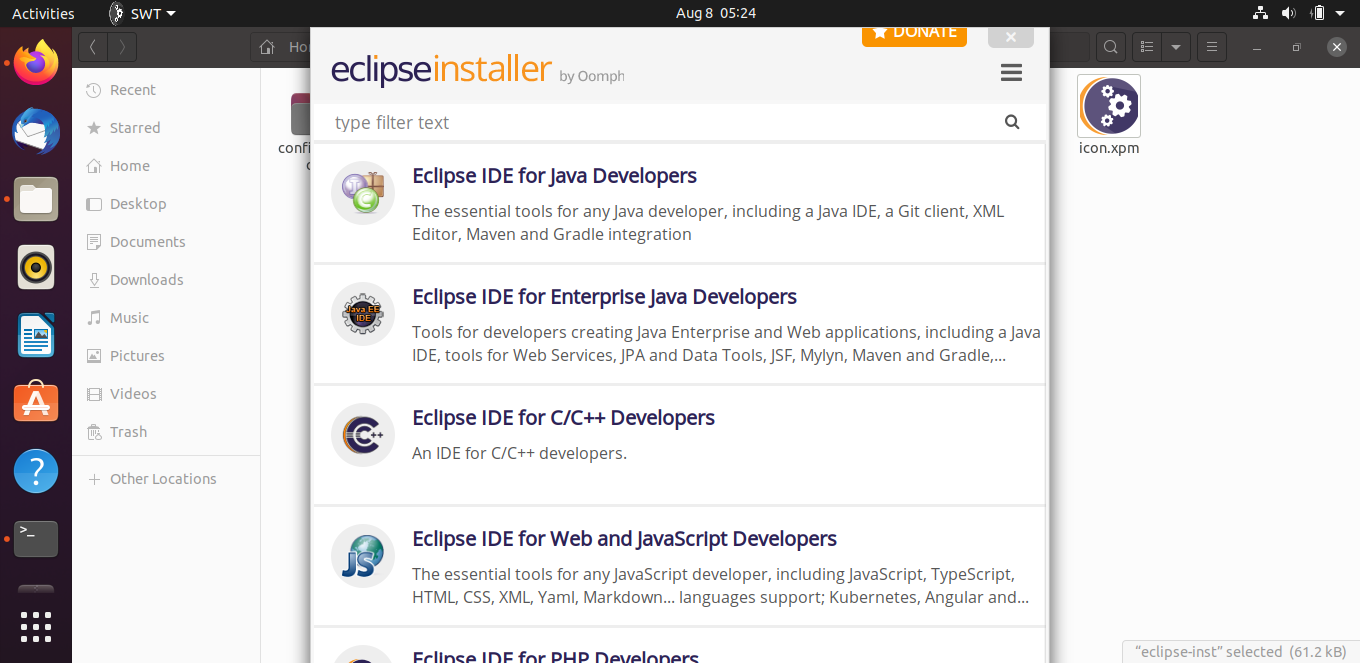
OpenJDK 8 was installed

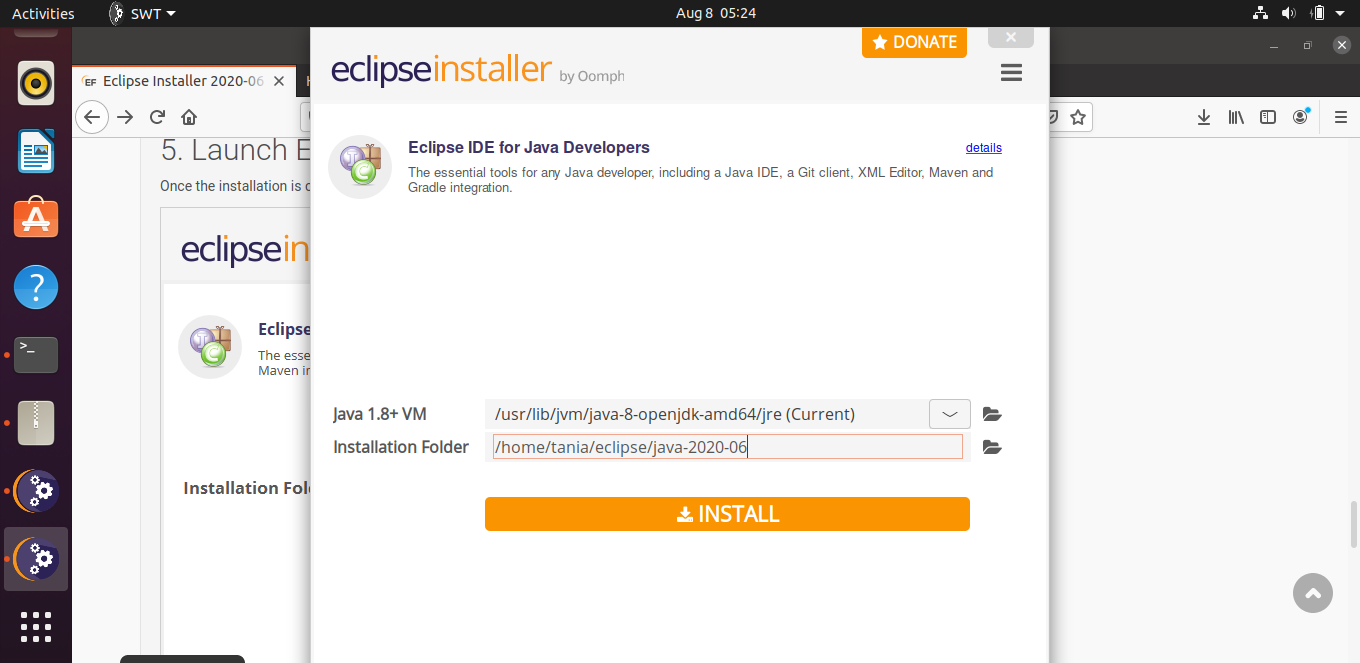


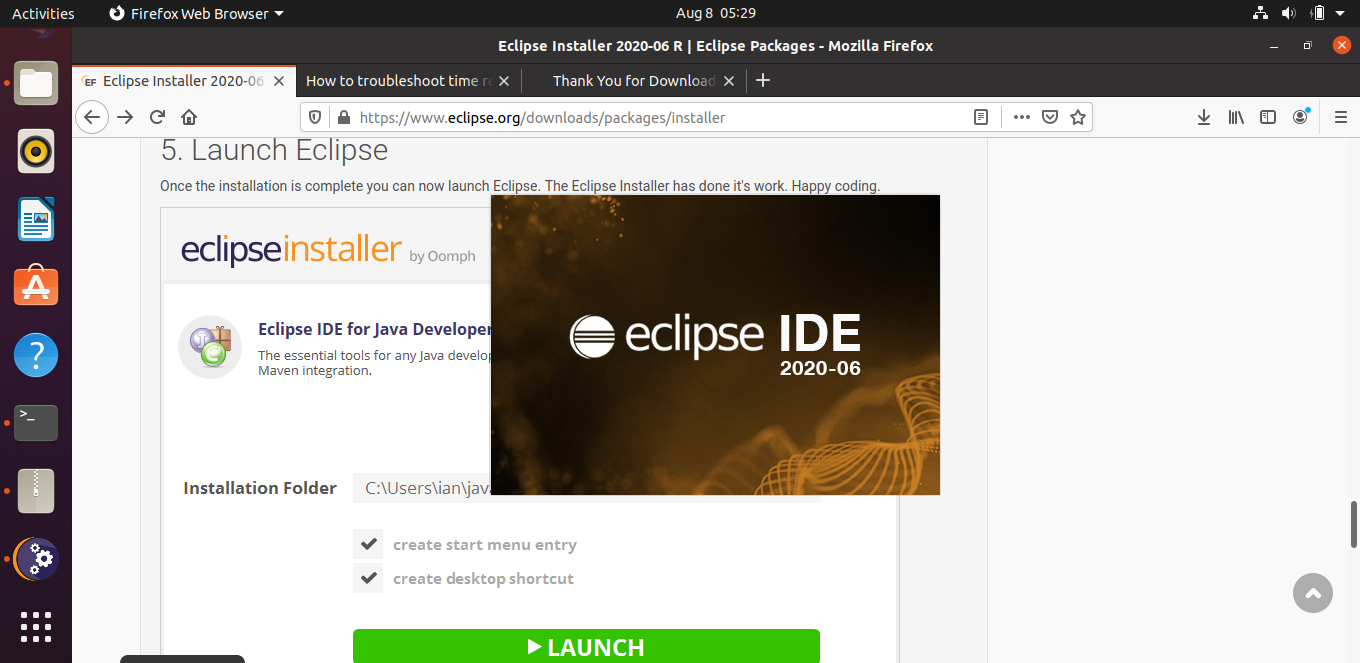
Step 3: Install Eclipse







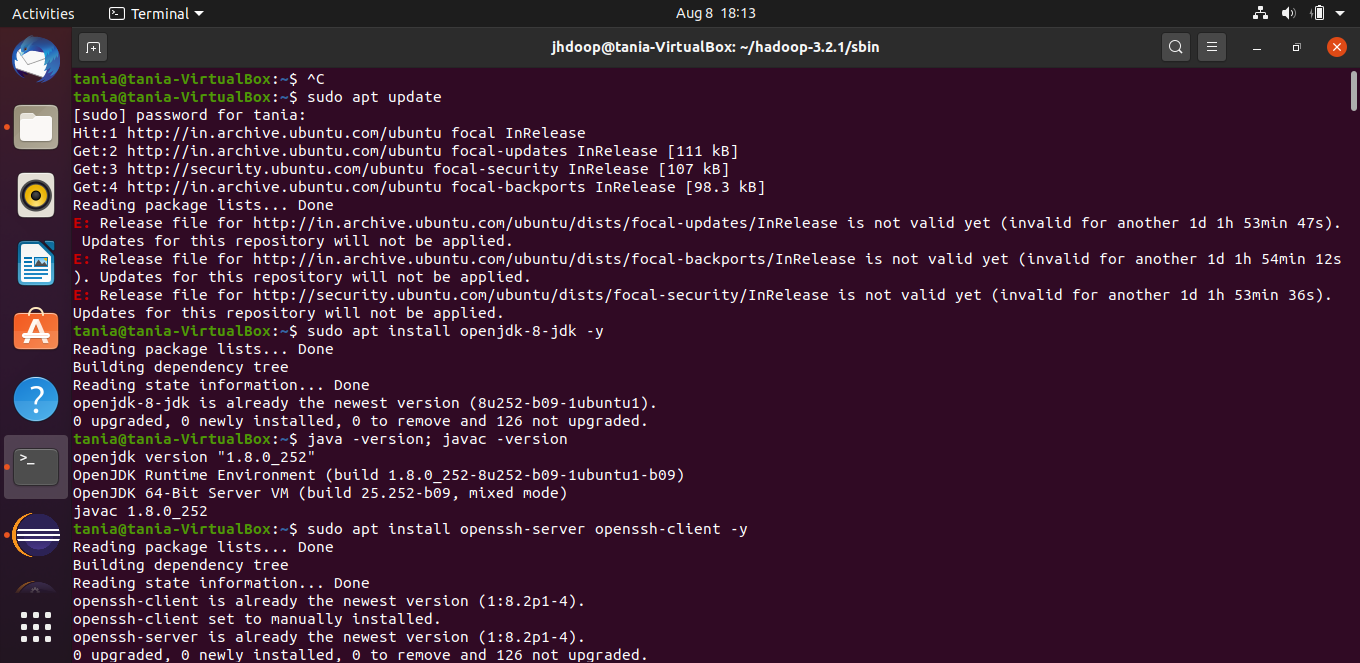




Step 4:

Follow this link: <https://phoenixnap.com/kb/install-hadoop-ubuntu>

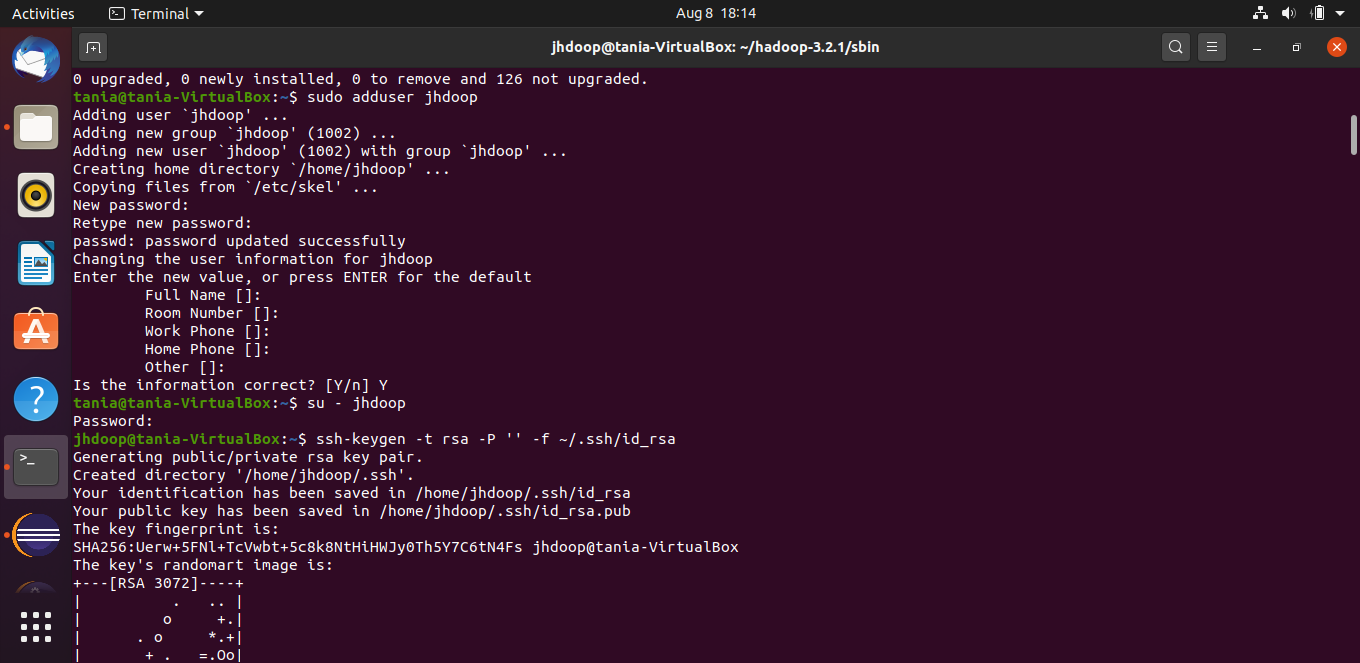
4.1: Install Open SSH and create a User

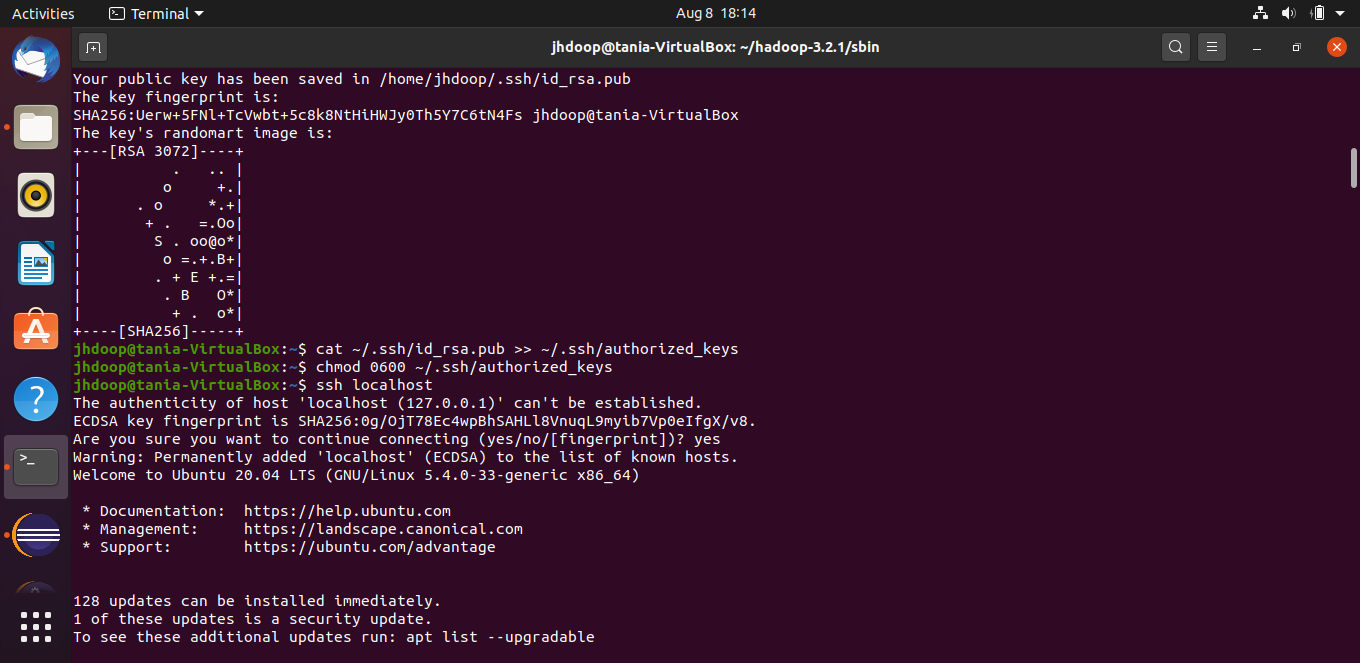


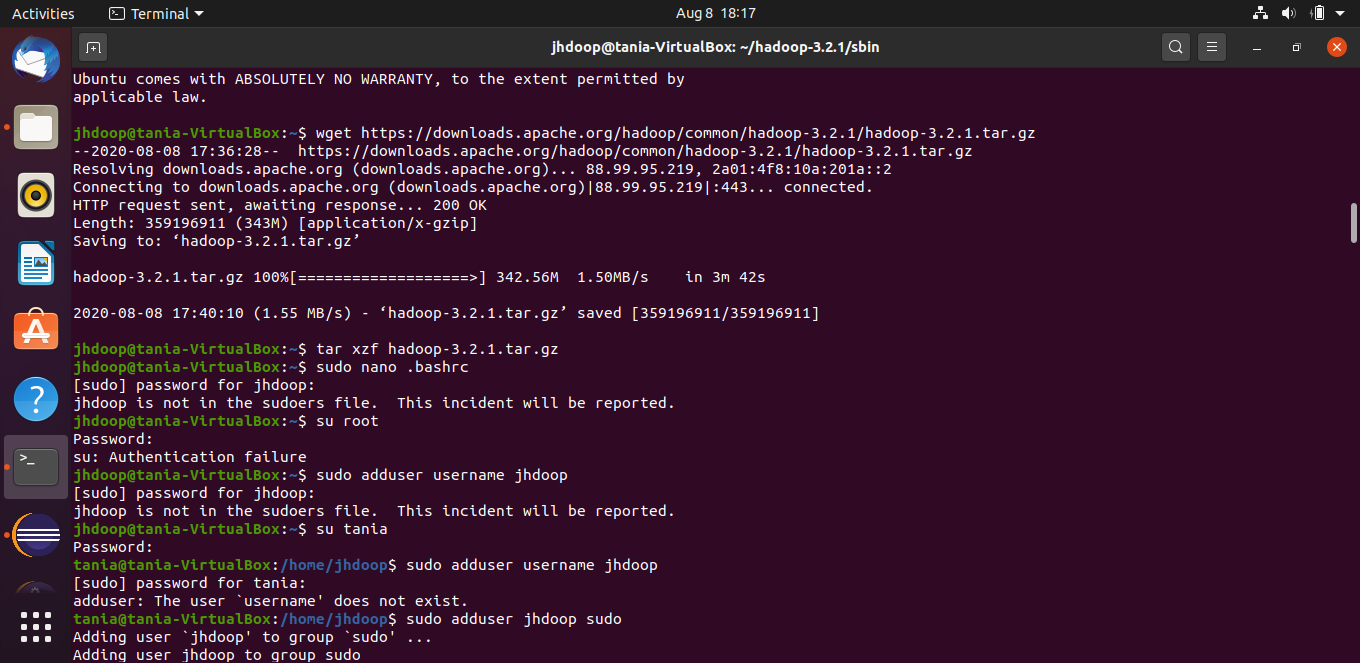
4.2: Switch to newly created user (add user to sudoers) and create ssh key

Add user to sudoers

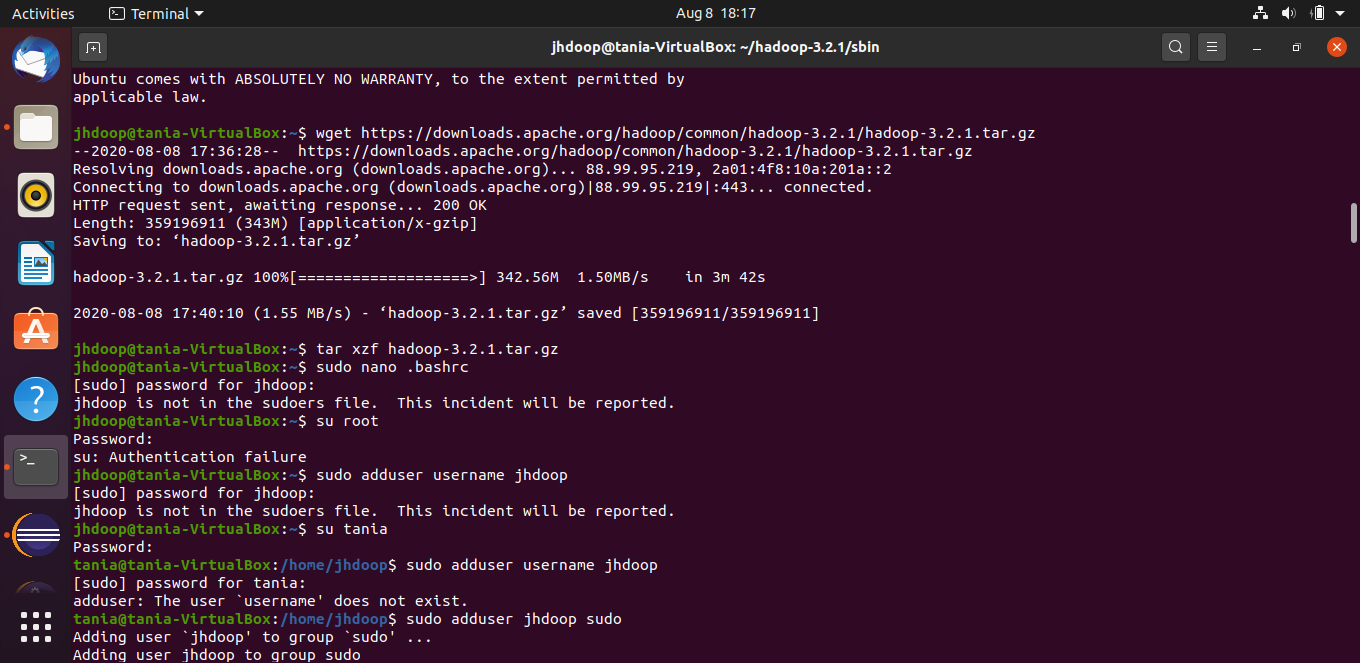
<https://unix.stackexchange.com/questions/179954/username-is-not-in-the-sudoers-file-this-incidentwill-be-reported>







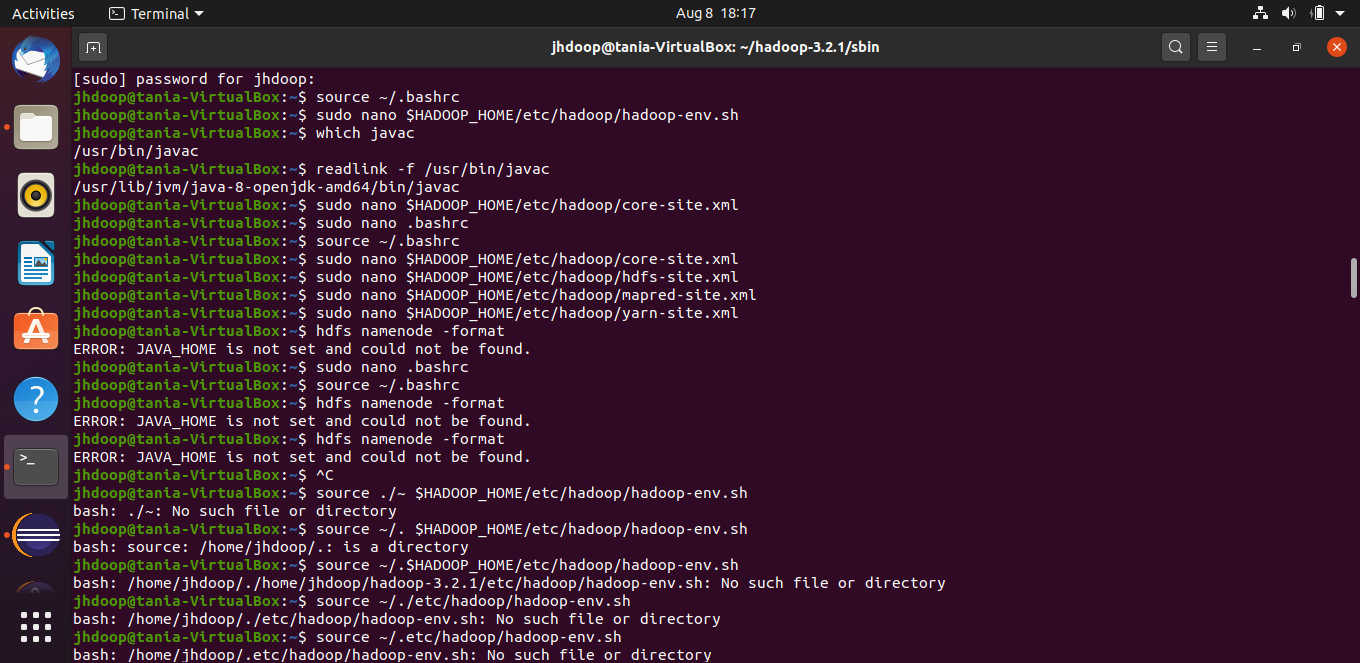
4.3: Install Hadoop

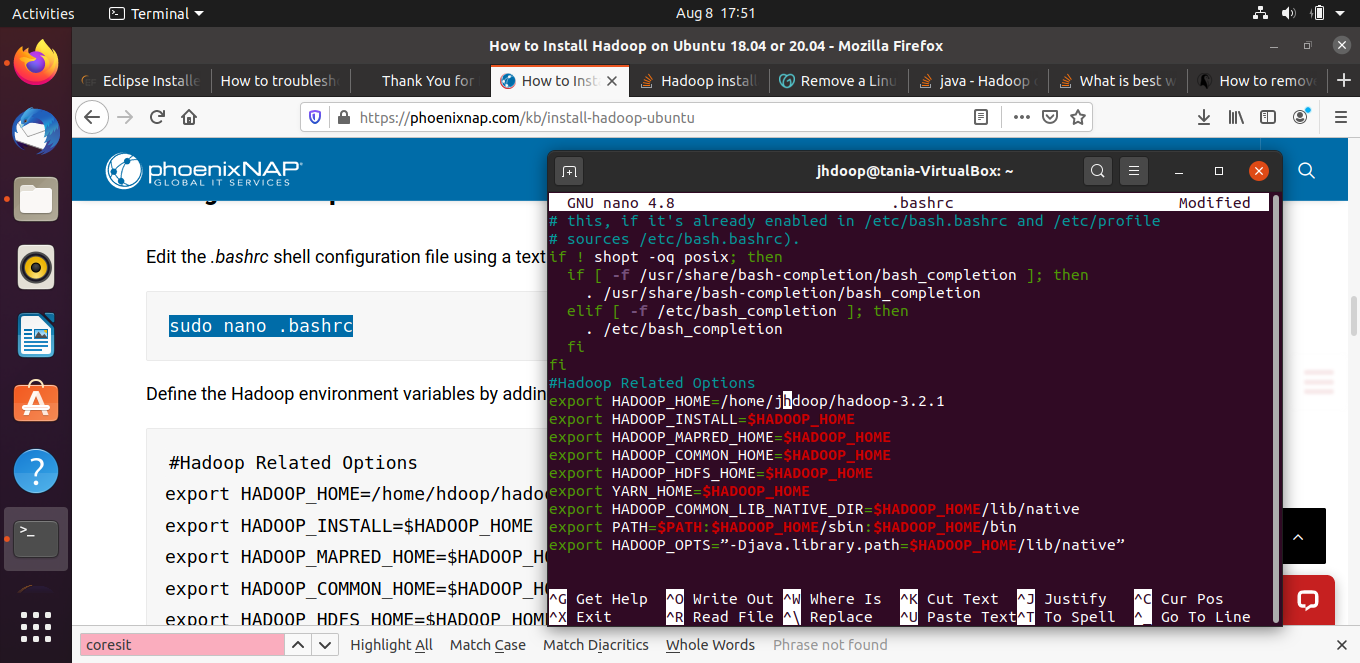


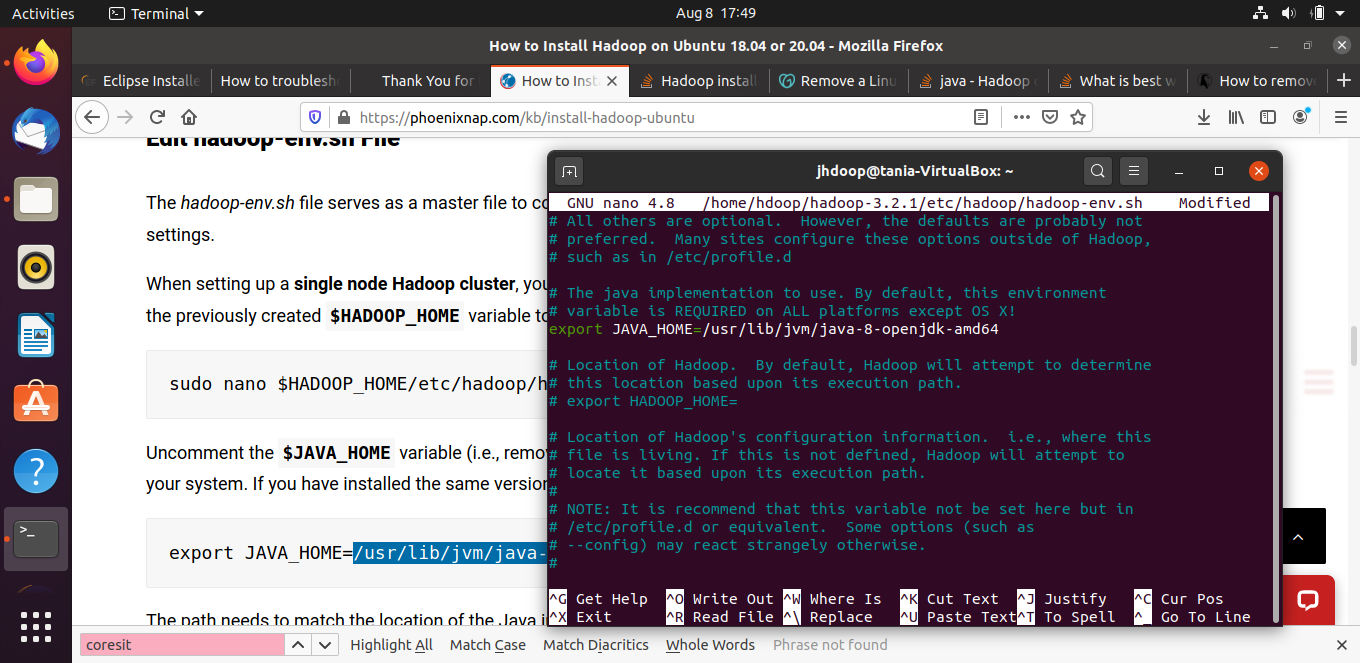
4.4: Set up the environment for hadoop as mentioned in the reference link above

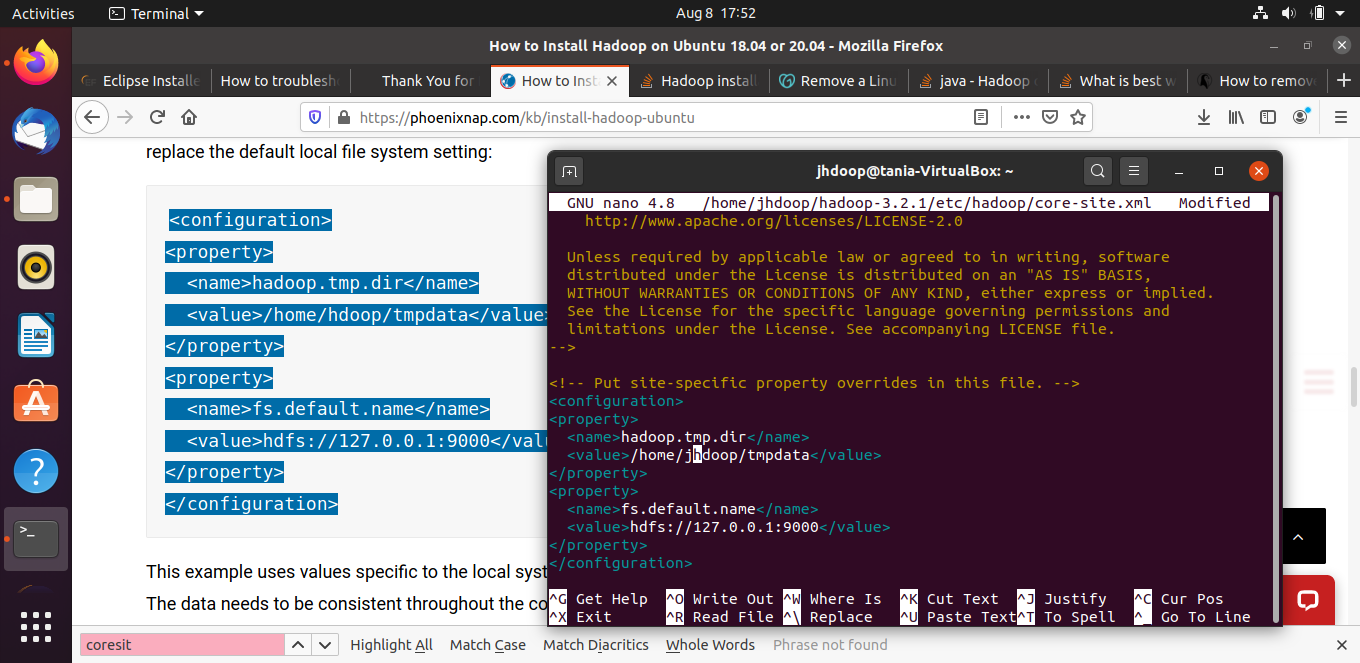
Solves Hadoop Error

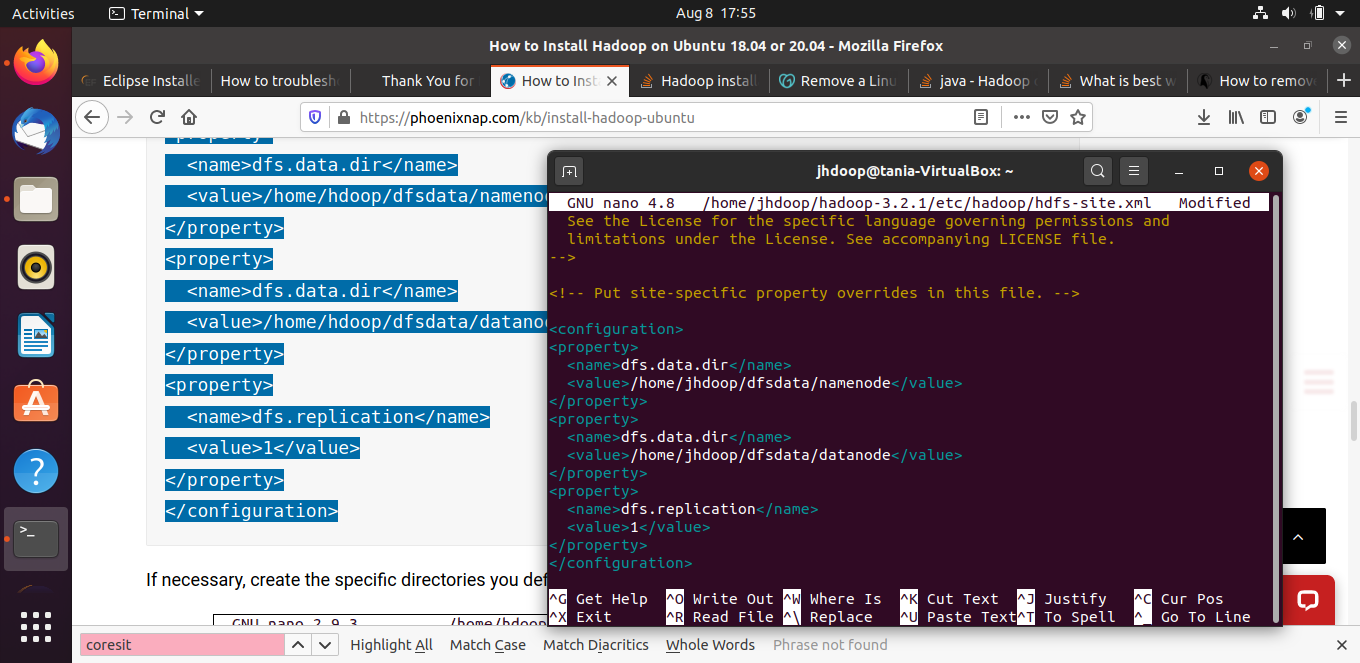
<https://www.edureka.co/community/72024/format-namenode-could-class-djava-library-hadoophadoop-native>

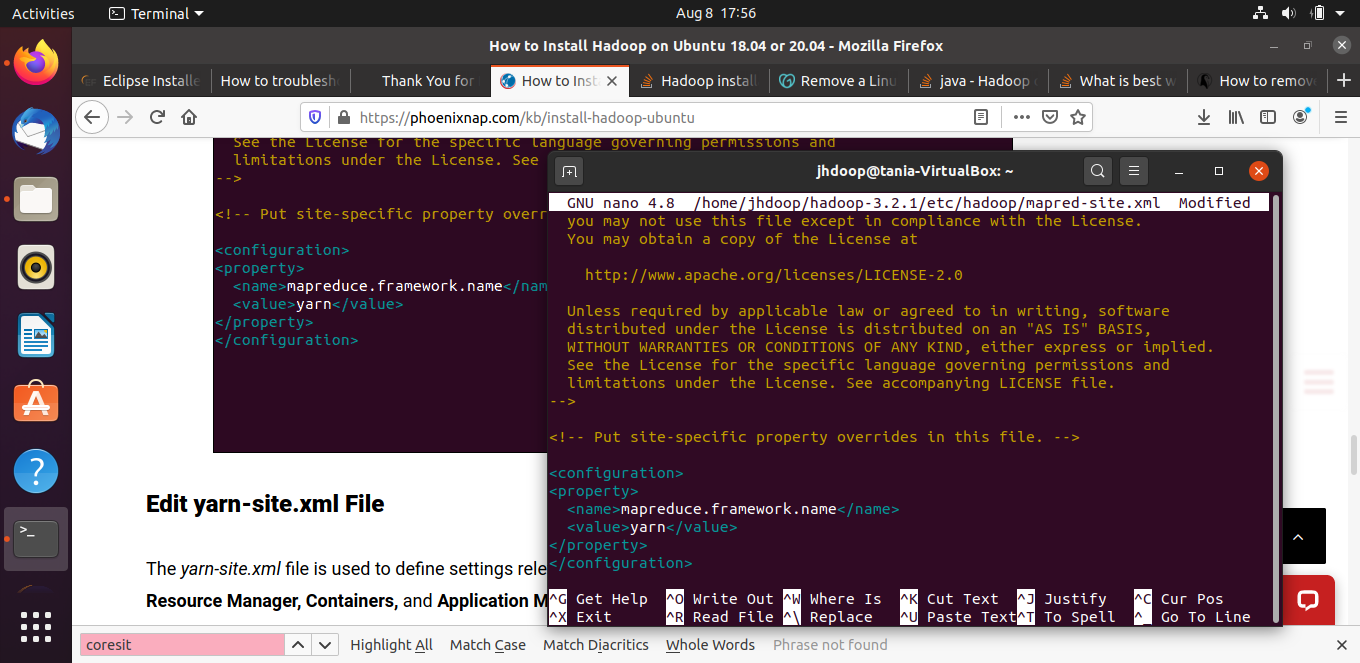


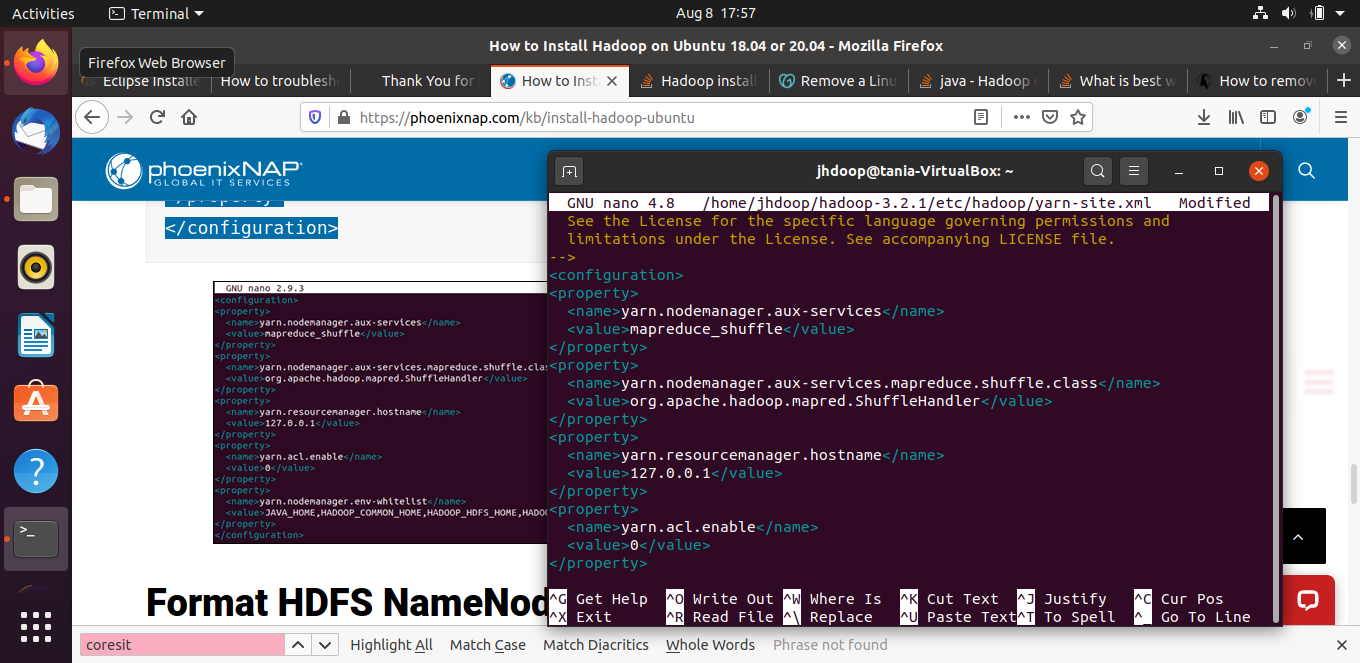


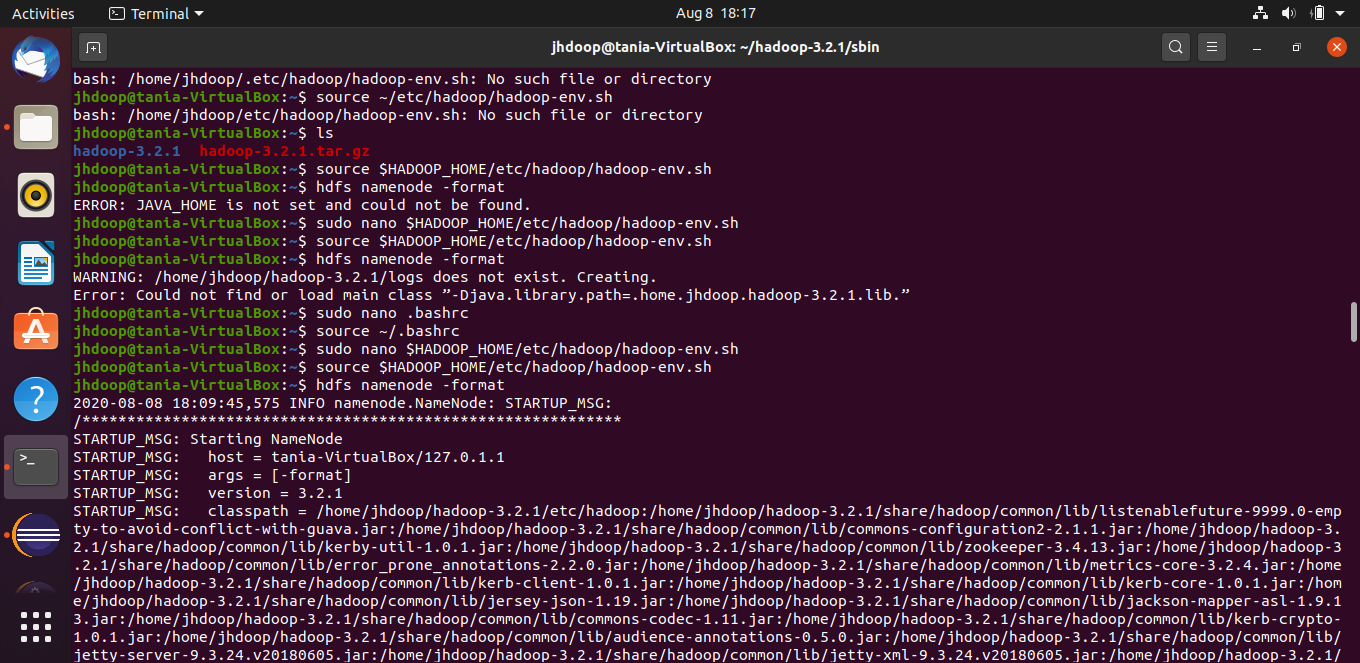




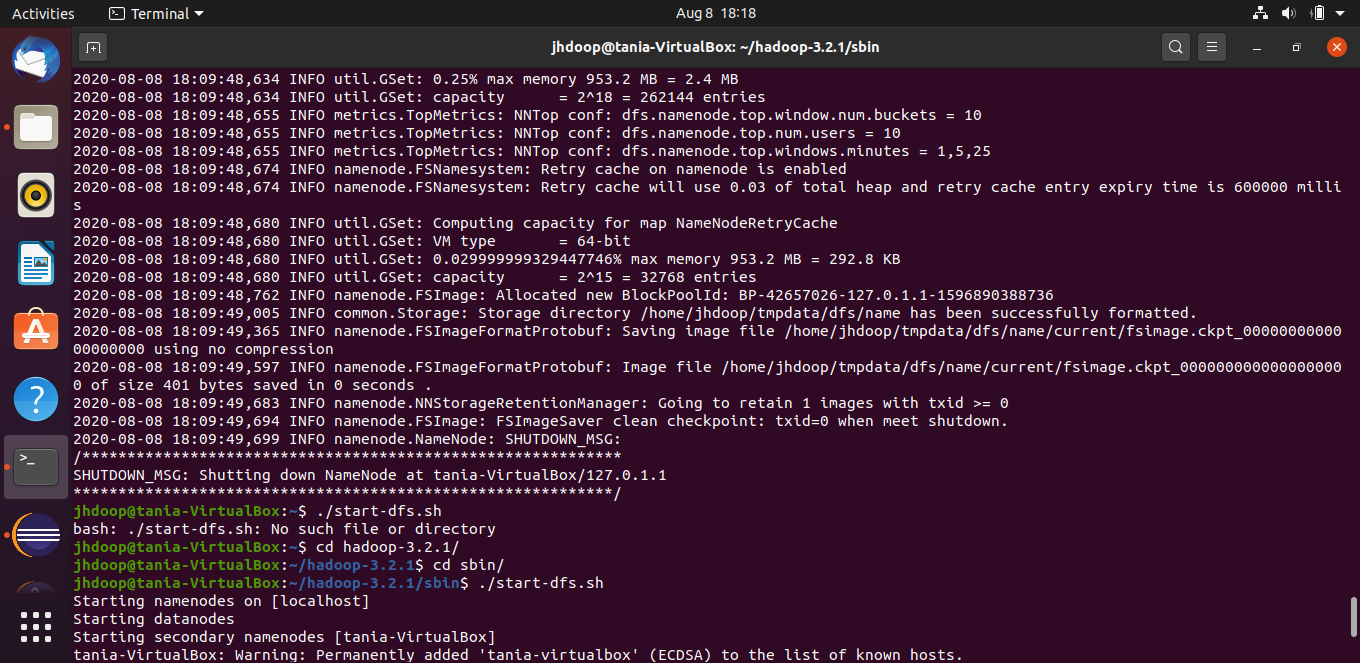








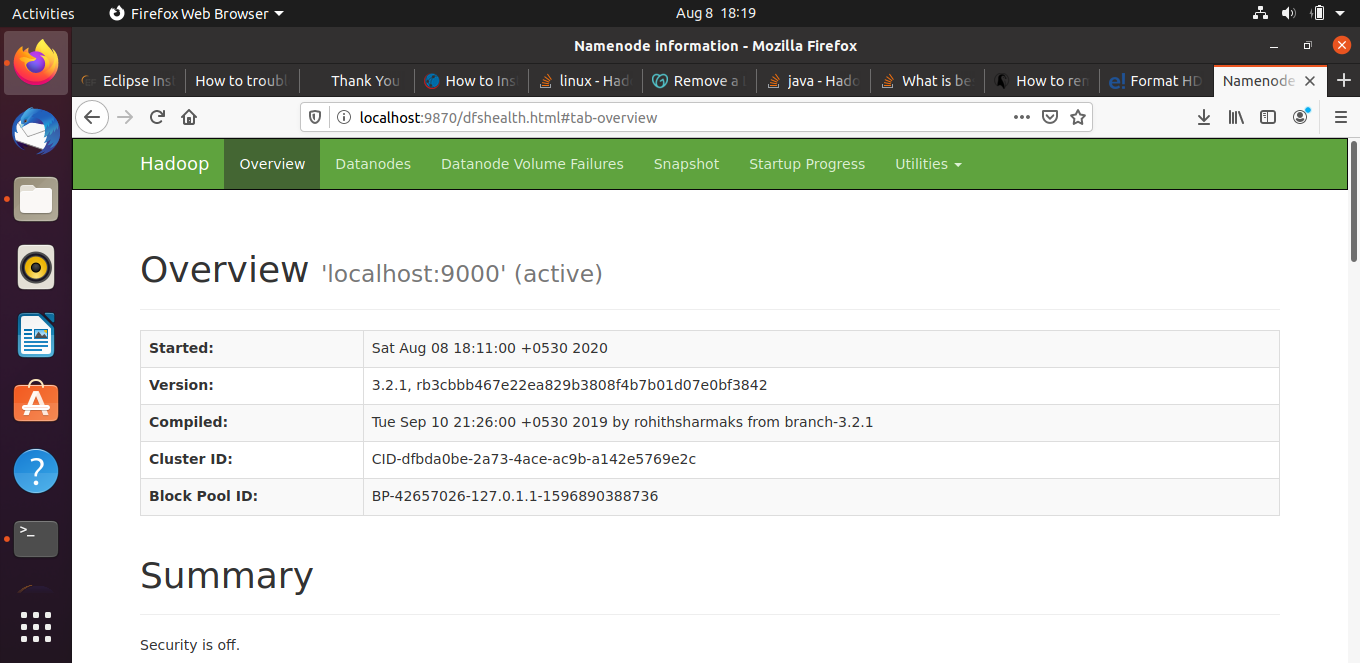




4.5: Start ,Yarn and check if daemons are active and running as java processes



4.6: Start Hadoop in the web browser at localhost:9870



**PART 2: WORD COUNT PROBLEM**

**Algorithm:**

In this program an approach using Hashmap in Java has been discussed.

* First, we declare a Hashmap in Java of {String, int}.
* We then split the input string using “,”. It separates the words on the basis on comma.
* Then we traverse in the string, check if the Hashmap already contains the traversed String or not.
* If it is present, then increase its count using get() and put() function in Hashmap.
* Once the traversal is completed, traverse in the Hashmap and print the String and its frequency.

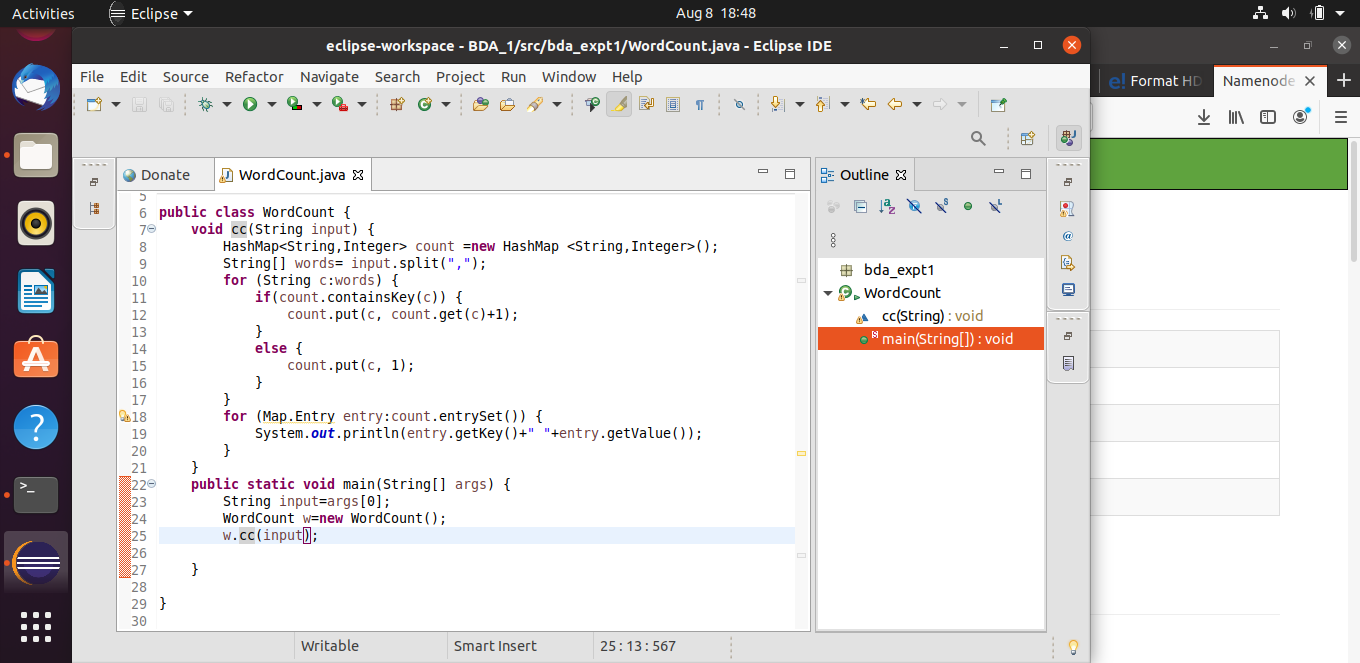
**Steps:**

Step 1:

Create a Java Project in eclipse and import the packages required using Build Path. In this project, no external dependencies were required.

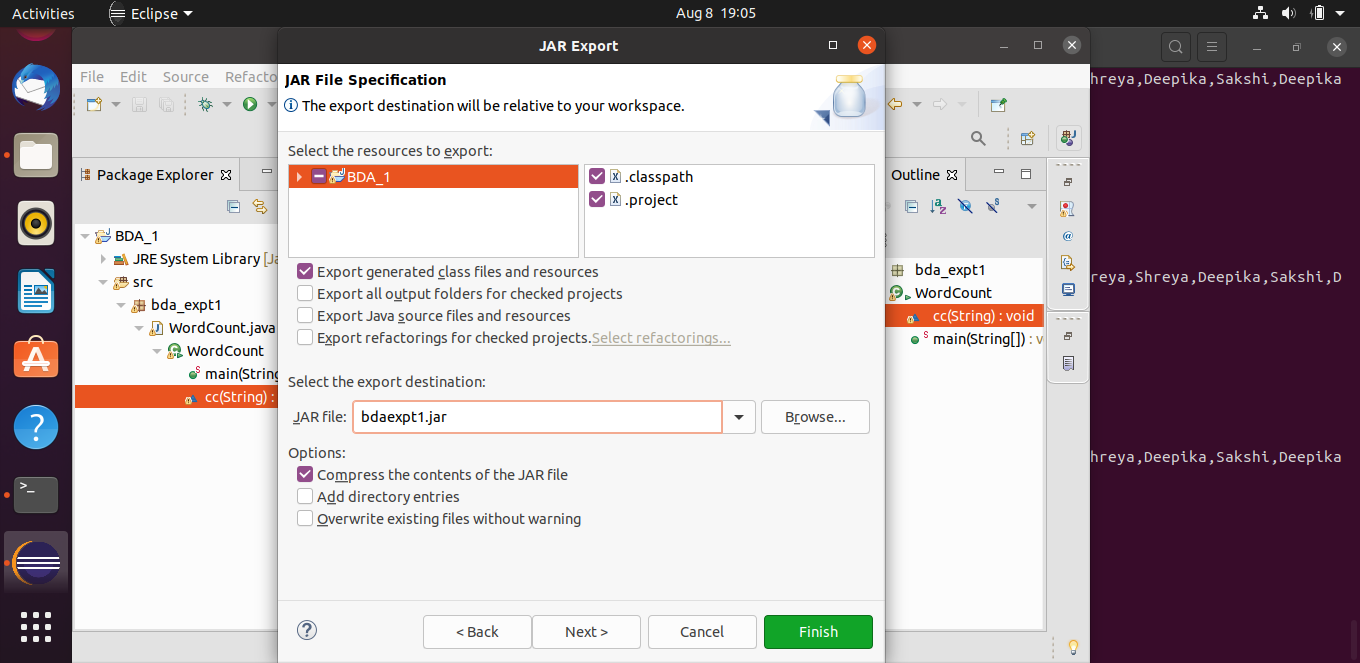
Step 2:

Create a package and store a Java file for the code to be written to count frequency of words in Java using algorithm described above.



Step 3:

Export the project as a .jar file



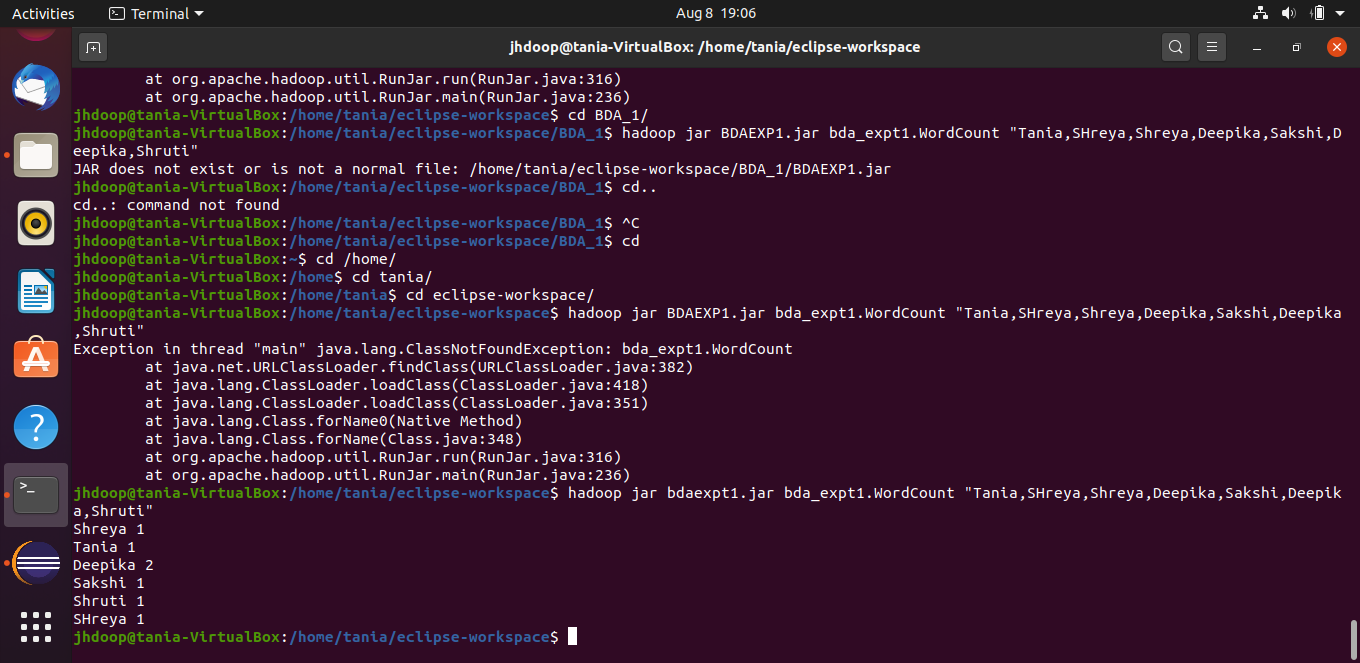
Step 4:

A file is not used in this program but the hadoop jar command allows us to use String[] args to give input to the java program created and an output in the following format is generated.

Note: The class name is to be mentioned after the path to jar file as: <package\_name.class\_name>

The actual format of the instruction is:

hadoop jar <jar\_file> <package\_name>.<class\_name> args/input\_file output\_file



**Conclusion:**

Hadoop was installed and was set up in ubuntu along with other requirements such as JAVA and Eclipse.

The word count problem was implemented on the Hadoop system using Eclipse and jar files.