1. Create a directory in HDFS called input

2. Load the file data\_quality\_input.csv into this directory

3. Understand the given input data and create hive external table for this

4. List down the distinct month names in the given input

5. Calculate the total number of Male and Female in this data

* 1. 6. Write down the role of these services in these above processes a. HADOOP – NameNode
  2. b. HADOOP – DataNodes
  3. c. HADOOP – Secondary NameNode
  4. d. HADOOP – Resource Manager
  5. e. HADOOP – Node Manager

7. Take the records only the age between 30 and 50 and store it in HDFS

8. Create the Database and table for the above results

9. Do the data export to transfer the data from HDFS to MySQL table

10. Verify all the results are stored in MySQL table

A picture containing text

Description automatically generated

A picture containing calendar

Description automatically generated

Text

Description automatically generated

A picture containing shape

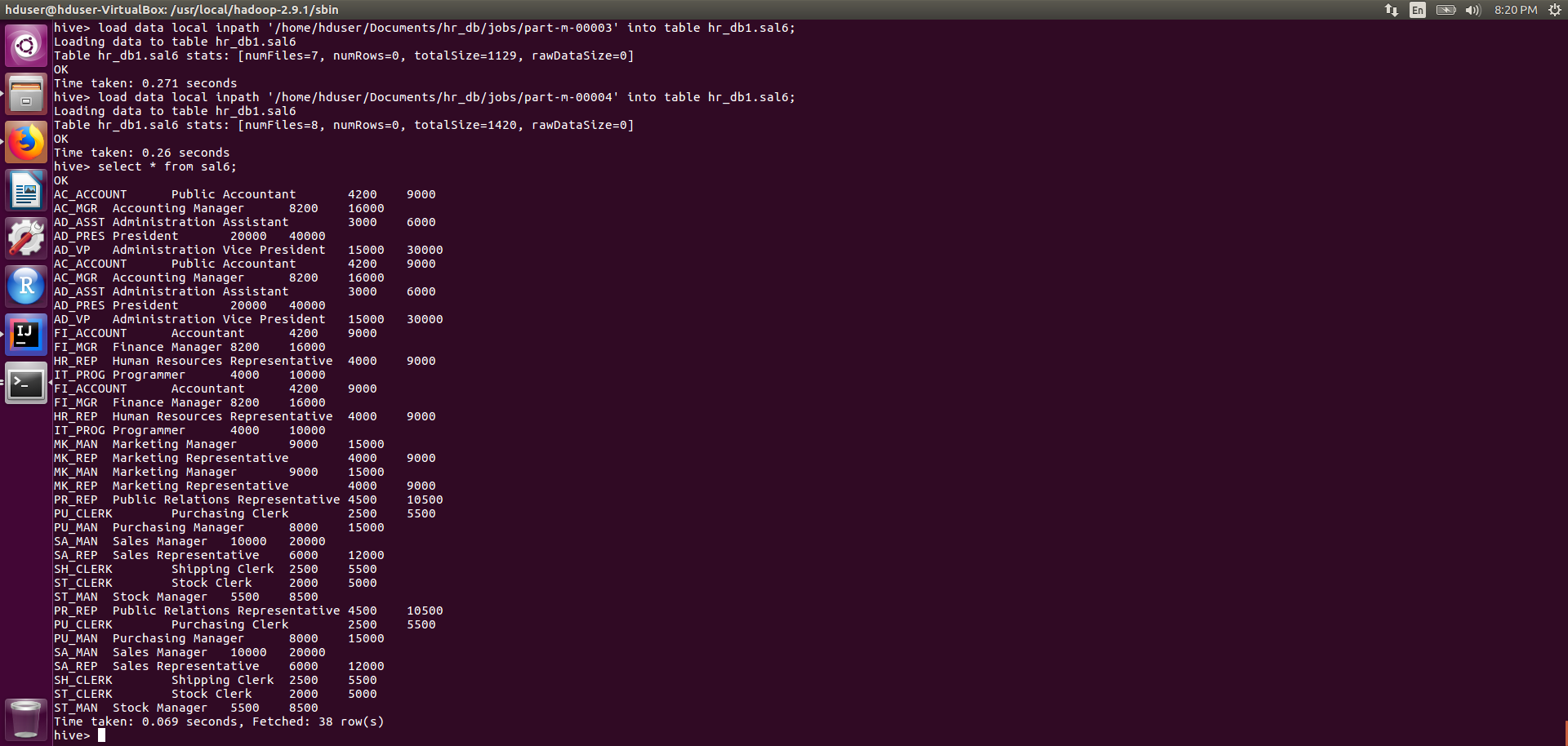
Description automatically generated

A picture containing text

Description automatically generated

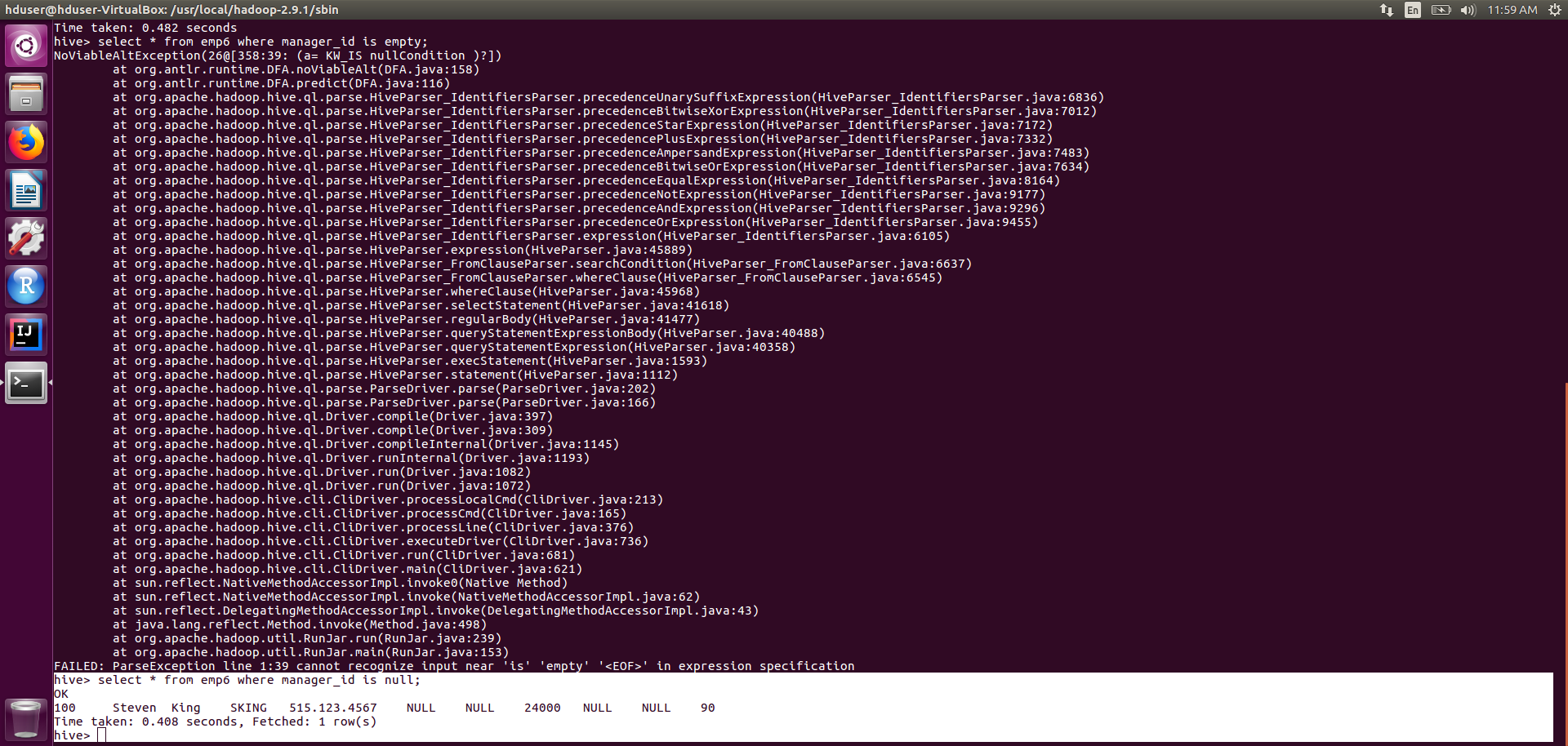
Text

Description automatically generated



Shape, rectangle

Description automatically generated



Graphical user interface, text

Description automatically generated

Text, application

Description automatically generated

Text

Description automatically generated