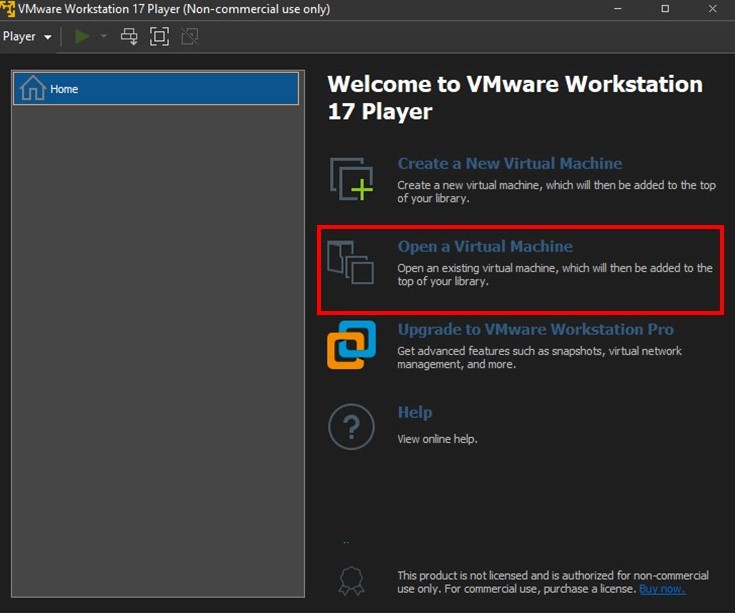
Step 1: Firstly, install VMWare Workstation using the files provided in prerequisites

folder. Then, download the Cloudera extension from the same folder and

extract its contents.

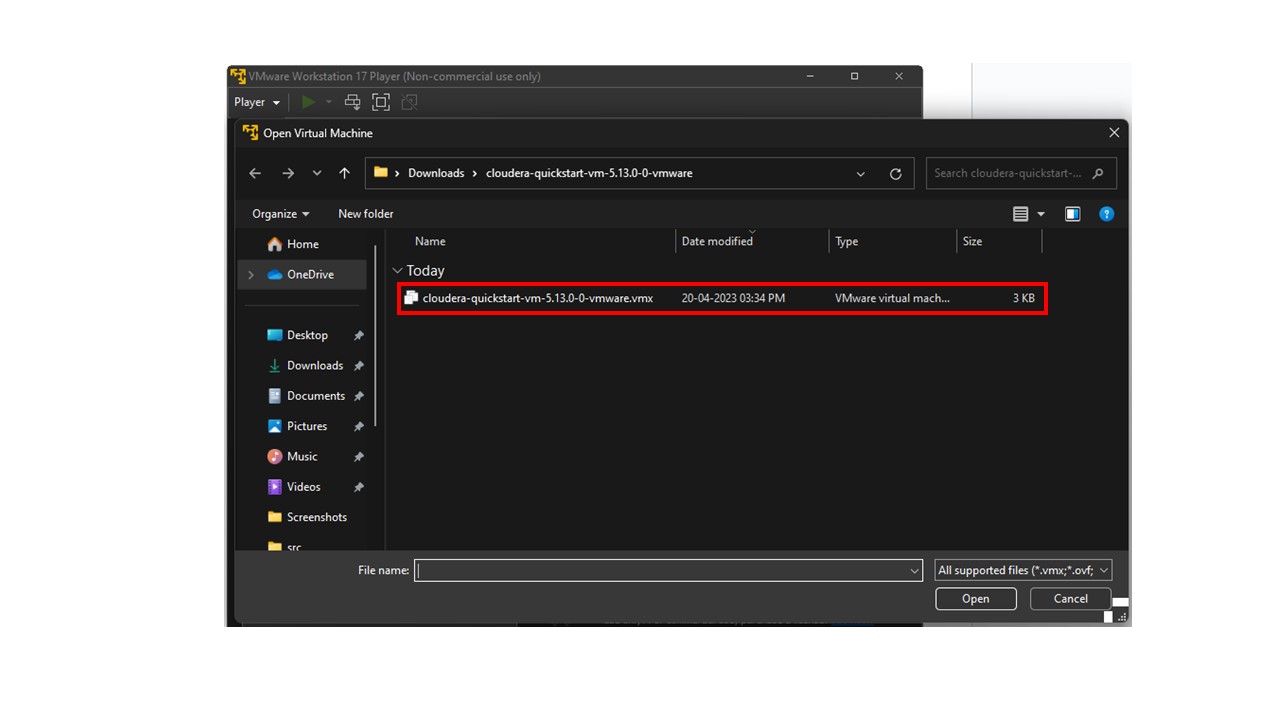
Step 2: After completing the previous step, locate **Open a Virtual Machine** and click

on the option to open it."

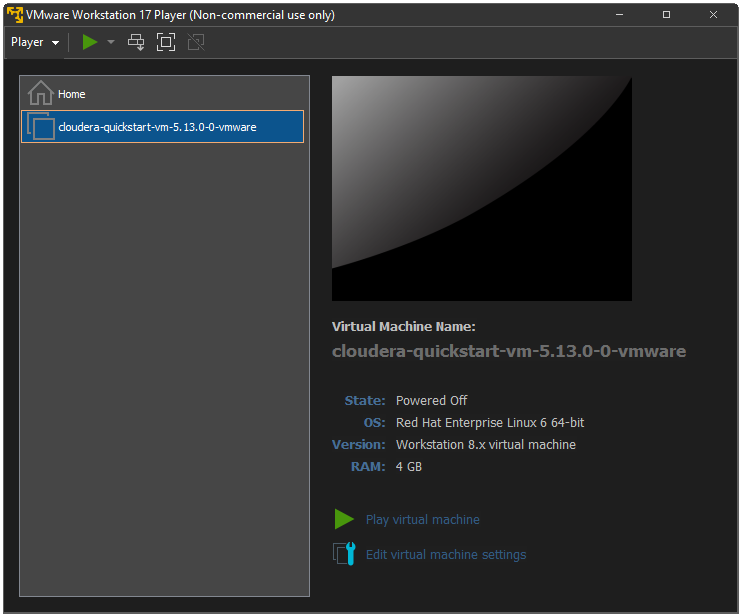


Step 3: To open the **Cloudera VM**, you need to locate the **Cloudera VM** file in your

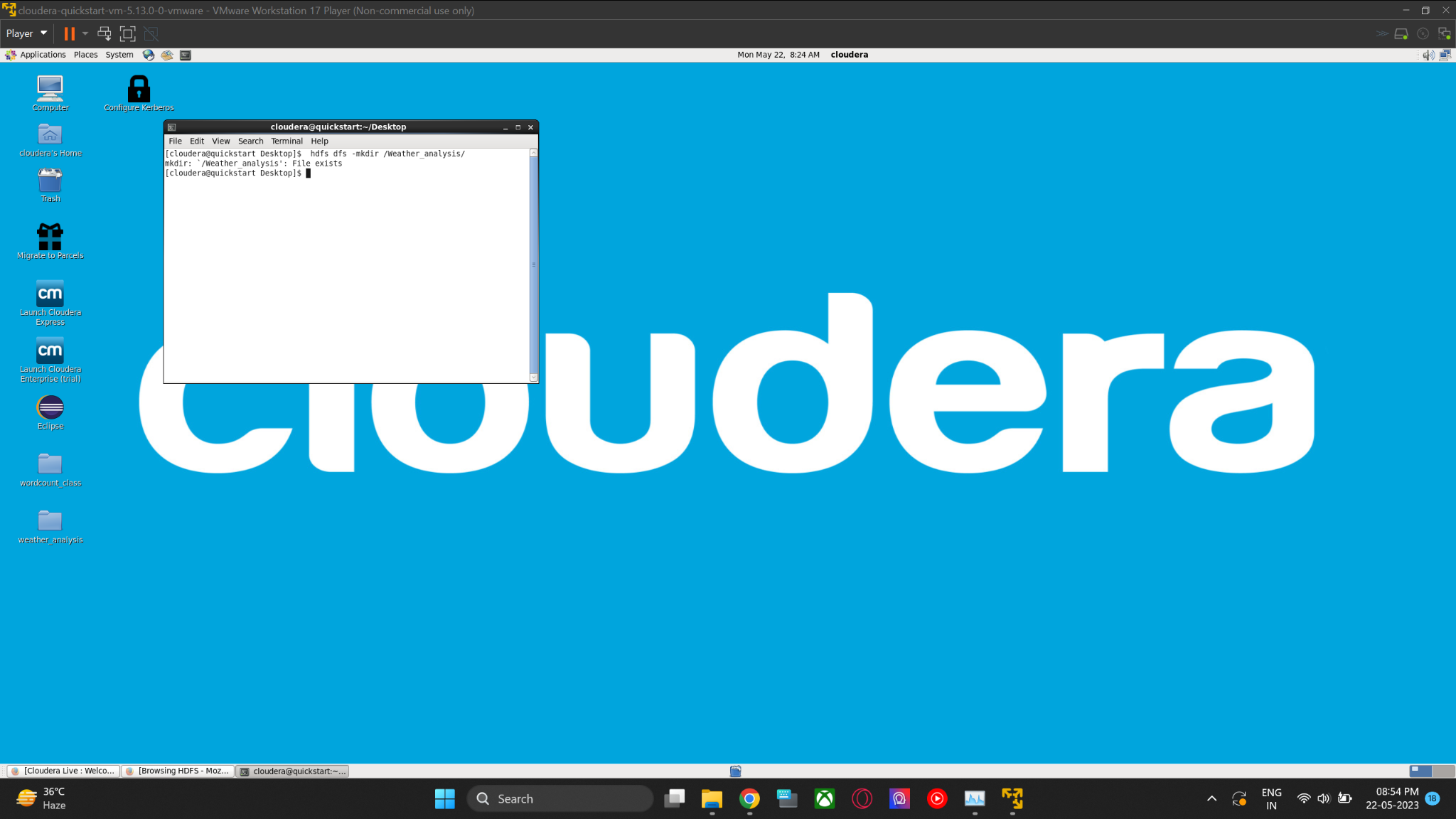
directory and then access it



Step 4: Run Cloudera VM



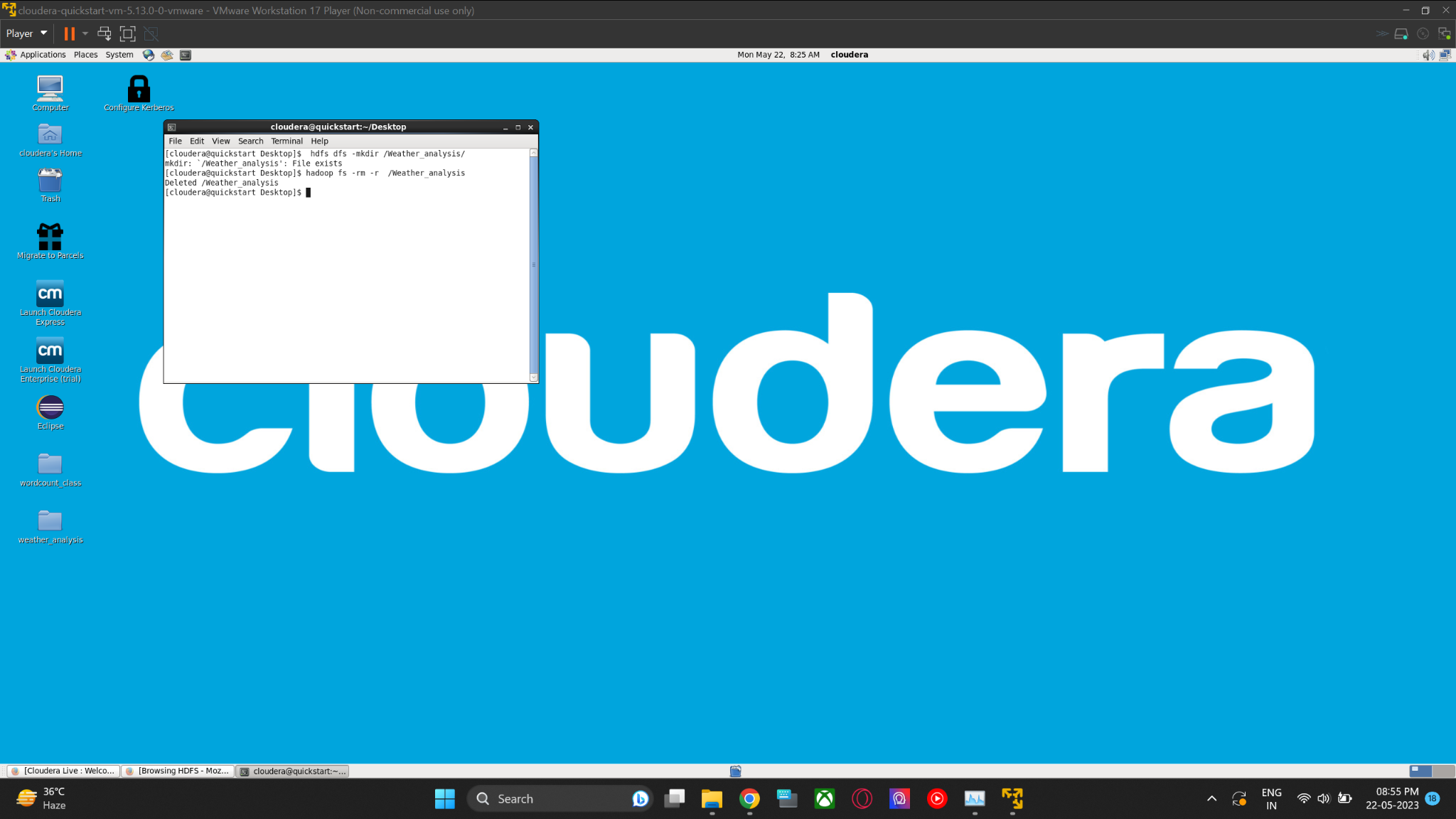
Step 5: If we encounter a database already existing error…



Step 6: To resolve this issue, execute the following command to delete the file from Hadoop.

**$ hadoop fs -rm -r /word\_count**

**$ hadoop fs -rm -r /Weather\_analysis**



Step 7: Create a folder with name word count in a virtual machine by using following

Command.

**$ mkdir wordcount\_class**

Step 8: Using the following command, navigate to the directory of the name word

count in a virtual machine.

**$ cd wordcount\_class/**

Step 9: Write programs for sum-reducer, map-reducer, and wordcount using the given command, and utilize a text file to count the number of occurrences of each word.

**$ gedit SumReducer.java**

**$ gedit WordMapper.java**

**$ gedit WordCount.java**

**$ gedit UN.txt**

Step 10: Create a Hadoop directory for storing the TXT file and then put the TXT file

into that directory by using the given command.

**$ hdfs dfs -mkdir /word\_count/**

**$ hdfs dfs -put UN.txt /word\_count/**

Step 11: Compile the Java files using the following commands, which include the

Hadoop libraries.

**$ javac -cp $(hadoop classpath) \*.java**

Step 12:Create a Java archive for distribution using the following command

**$ jar -cvf wordcount.jar \*.class**

Step 13:To run a JAR file in Hadoop and see the word count in the TXT file, type the

following command:

**$ hadoop jar wordcount.jar WordCount /word\_count/UN.txt /word\_count/Output**

**$ hadoop fs -cat /word\_count/Output/part-r-00000**