***Practical 9:*** ***Demonstrate the use of Sqoop tool to transfer data between Hadoop & relational database servers***

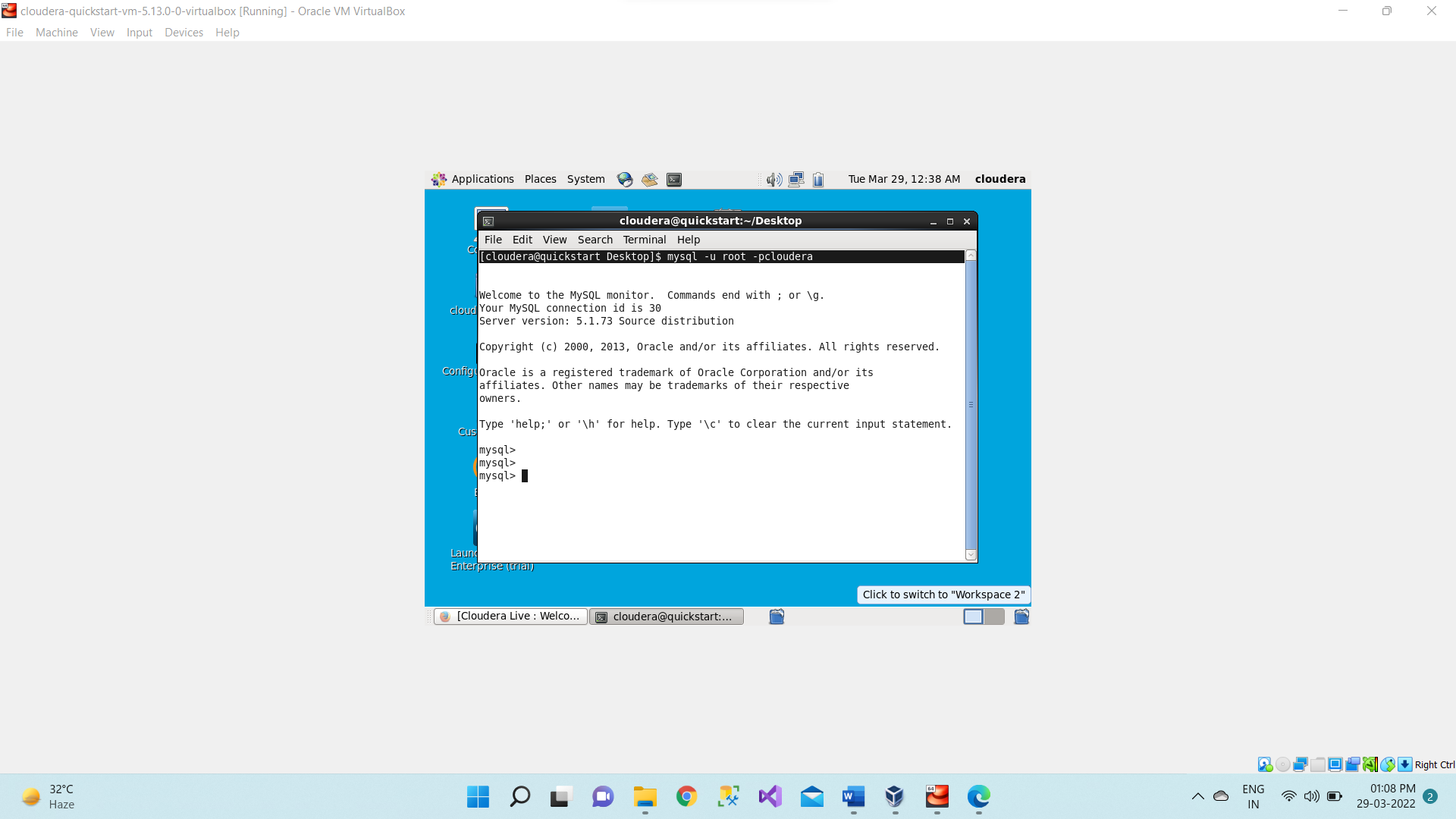
**Steps involved Demonstrate the use of Sqoop tool to transfer data between Hadoop & relational database servers below in the following order-**

**Sqoop:**

The traditional application management system, that is, the interaction of applications with relational database using RDBMS, is one of the sources that generate Big Data. Such Big Data, generated by RDBMS, is stored in Relational Database Servers in the relational database structure. When Big Data storages and analyzers such as MapReduce, Hive, HBase, Cassandra, Pig, etc. of the Hadoop ecosystem came into picture, they required a tool to interact with the relational database servers for importing and exporting the Big Data residing in them. Here, Sqoop occupies a place in the Hadoop ecosystem to provide feasible interaction between relational database server and Hadoop’s HDFS. Sqoop − “SQL to Hadoop and Hadoop to SQL” Sqoop is a tool designed to transfer data between Hadoop and external datastores suac as relational databases and enterprise data warehouses. It is used to import data from relational databases such as MySQL, Oracle to Hadoop HDFS, and export from Hadoop file system to relational databases. It imports data from external datastores into HDFS, Hive, and HBase.

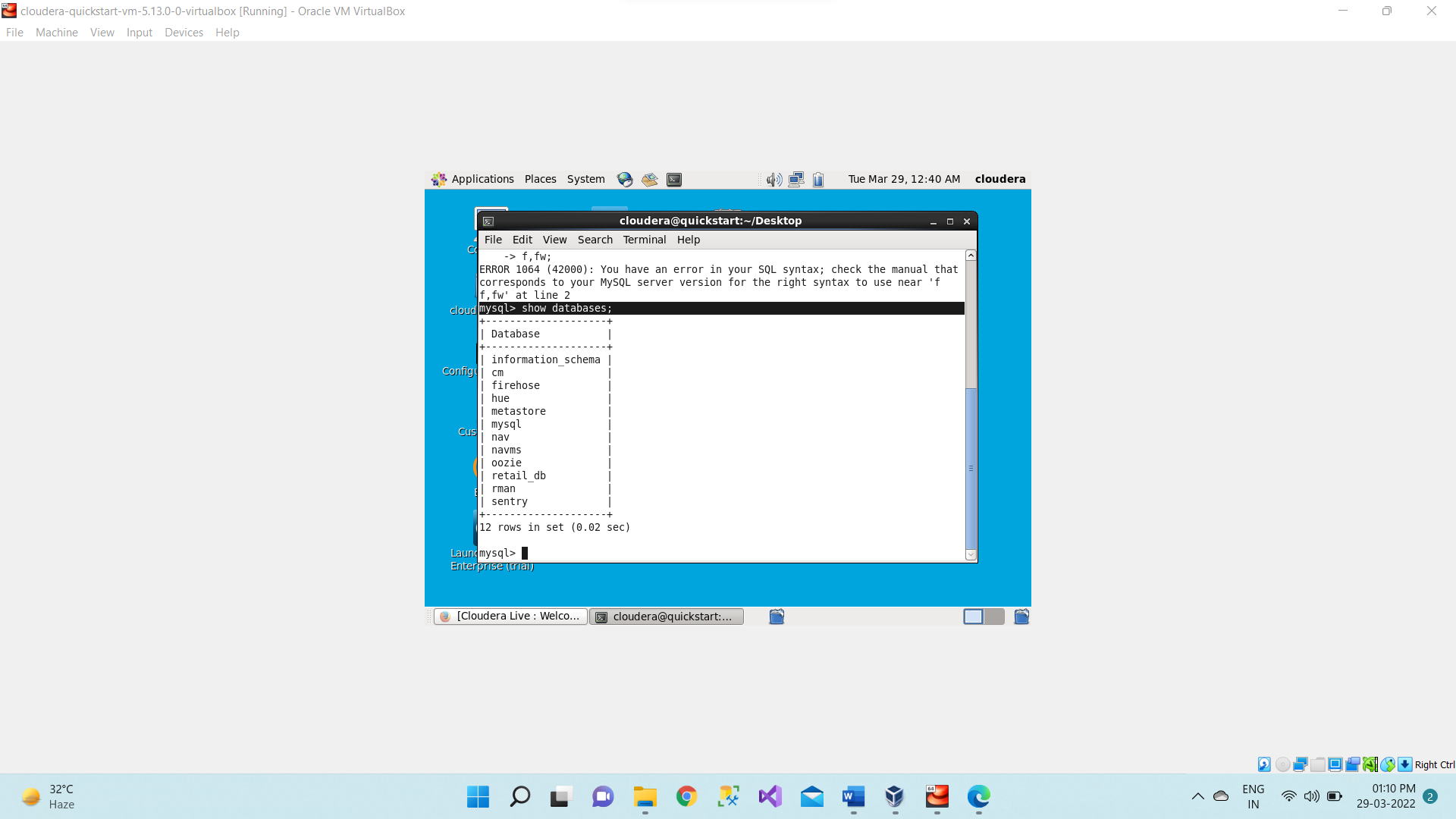
1. Starting the mysql by giving username as root and password as cloudera.

**mysql -u root -pcloudera**

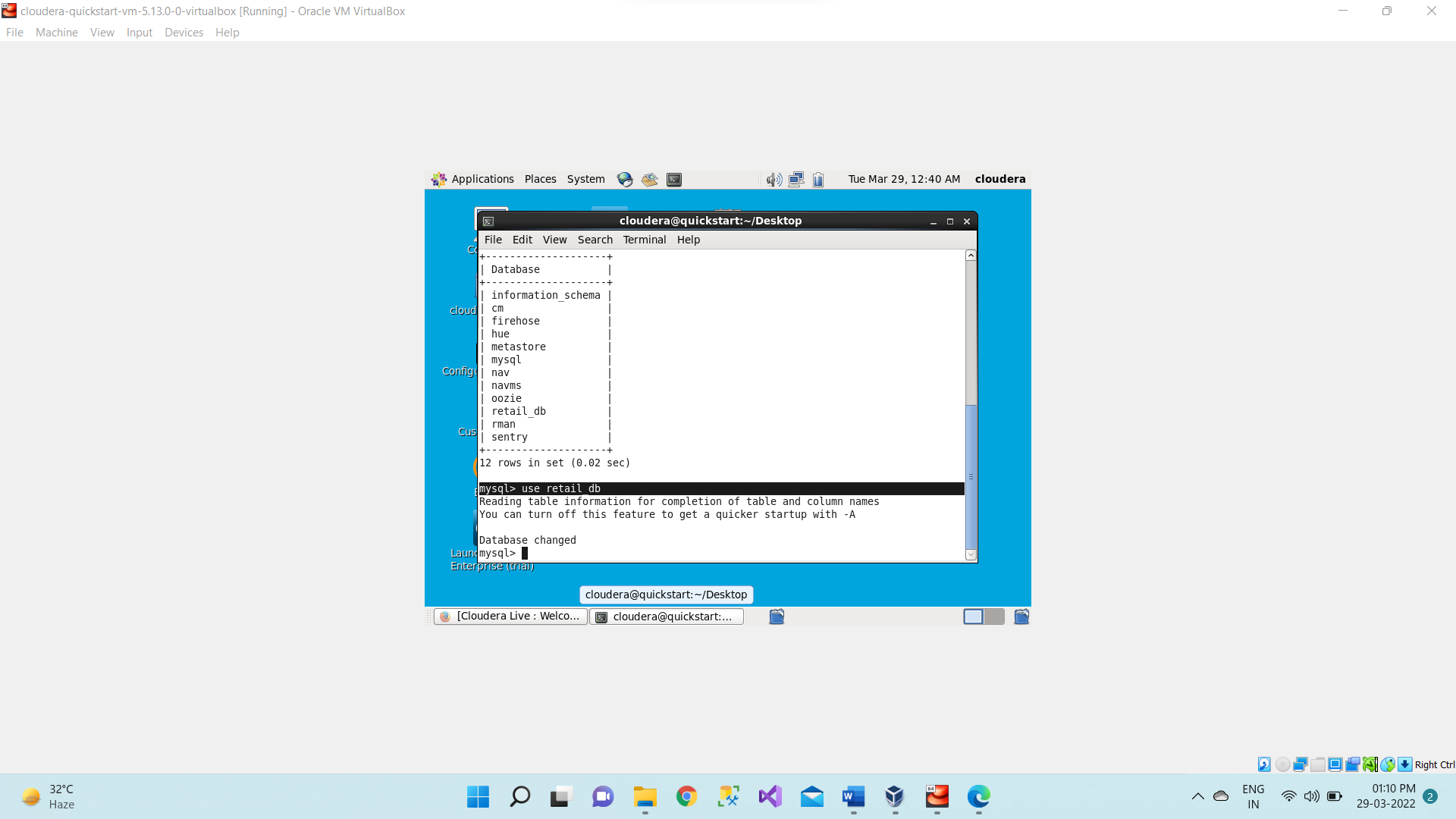


1. Now using below command it will displaying or give the list of databases which are already present or exist in mysql.

**show databases;**

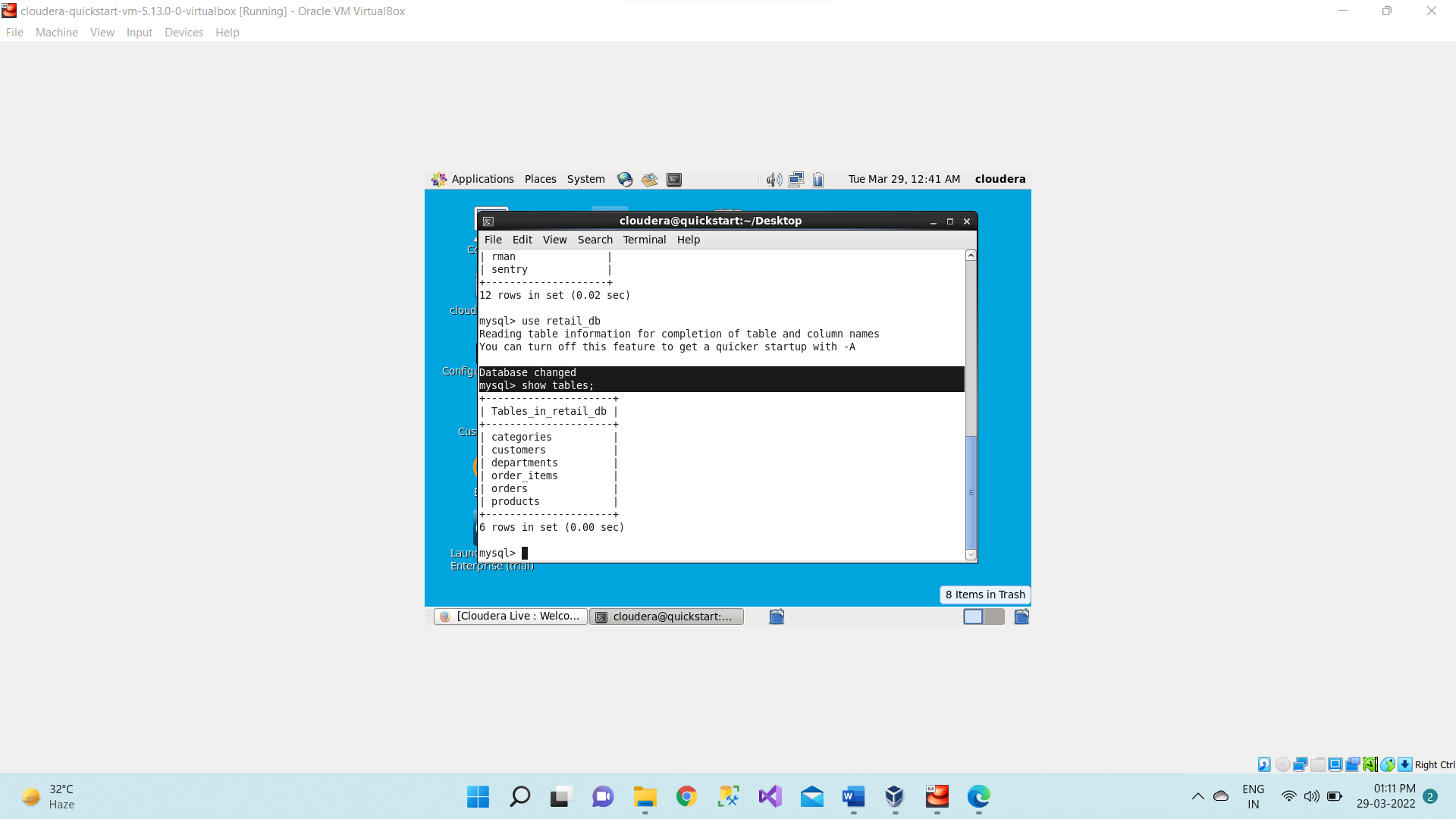


1. Now we are using the existing database i.e. retail\_db which are already present in mysql. **use retail\_db;**

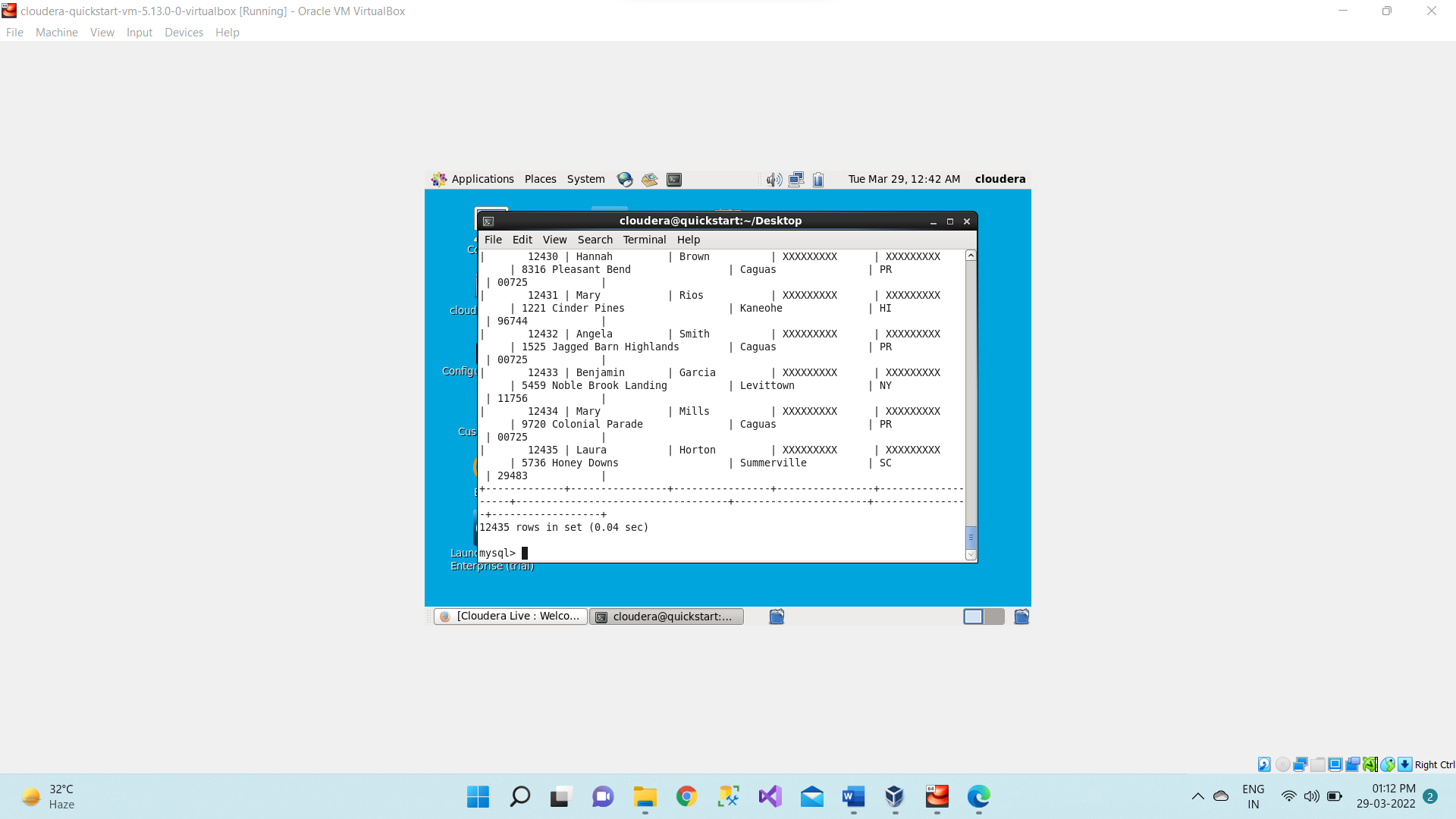


1. Now to see the tables under a specific database so we will be using the same command which is used to display the databases.

**show tables;**

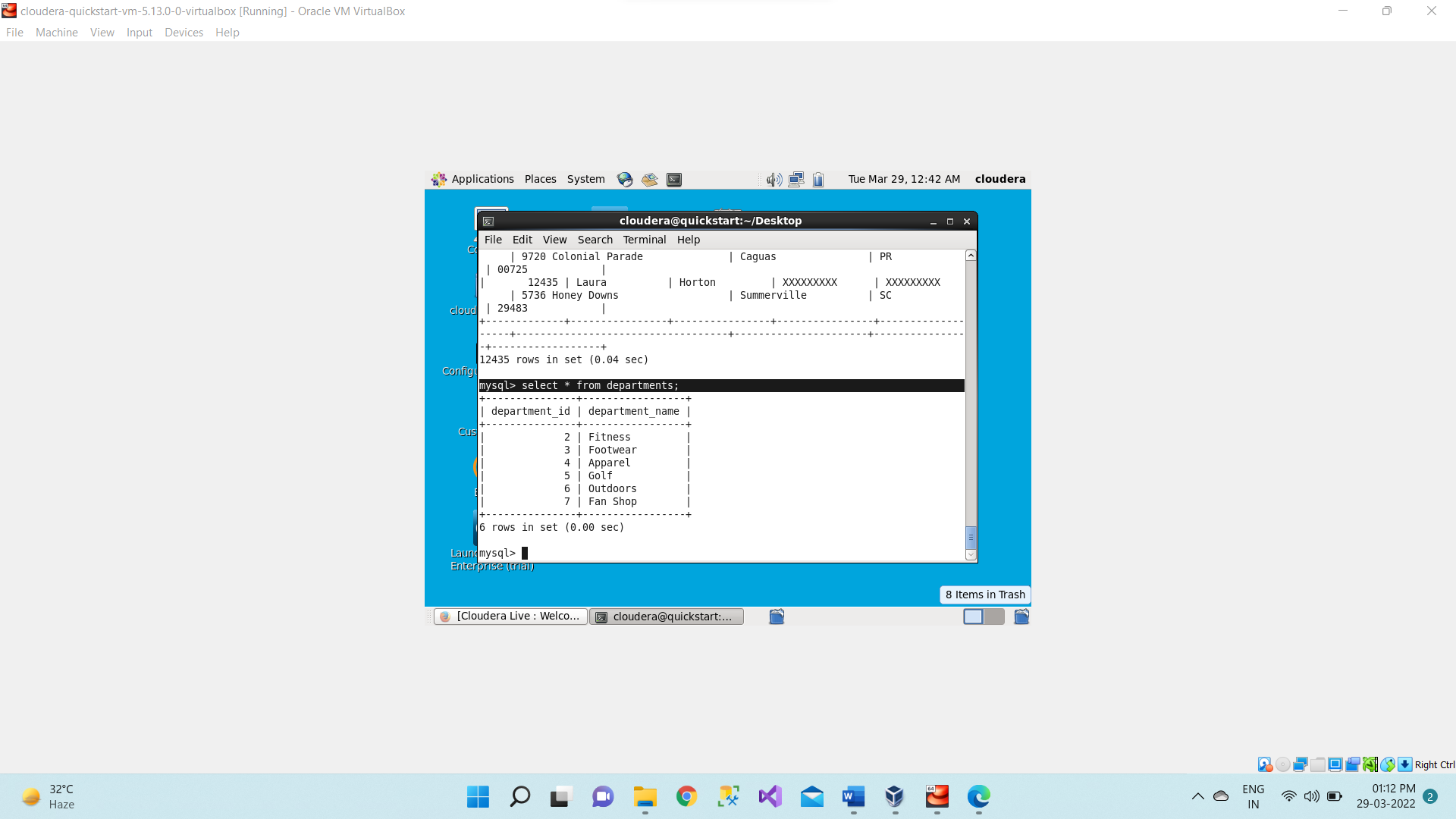


1. Here we are displaying all the records present in customers tables using below command. **select \* from customers;**

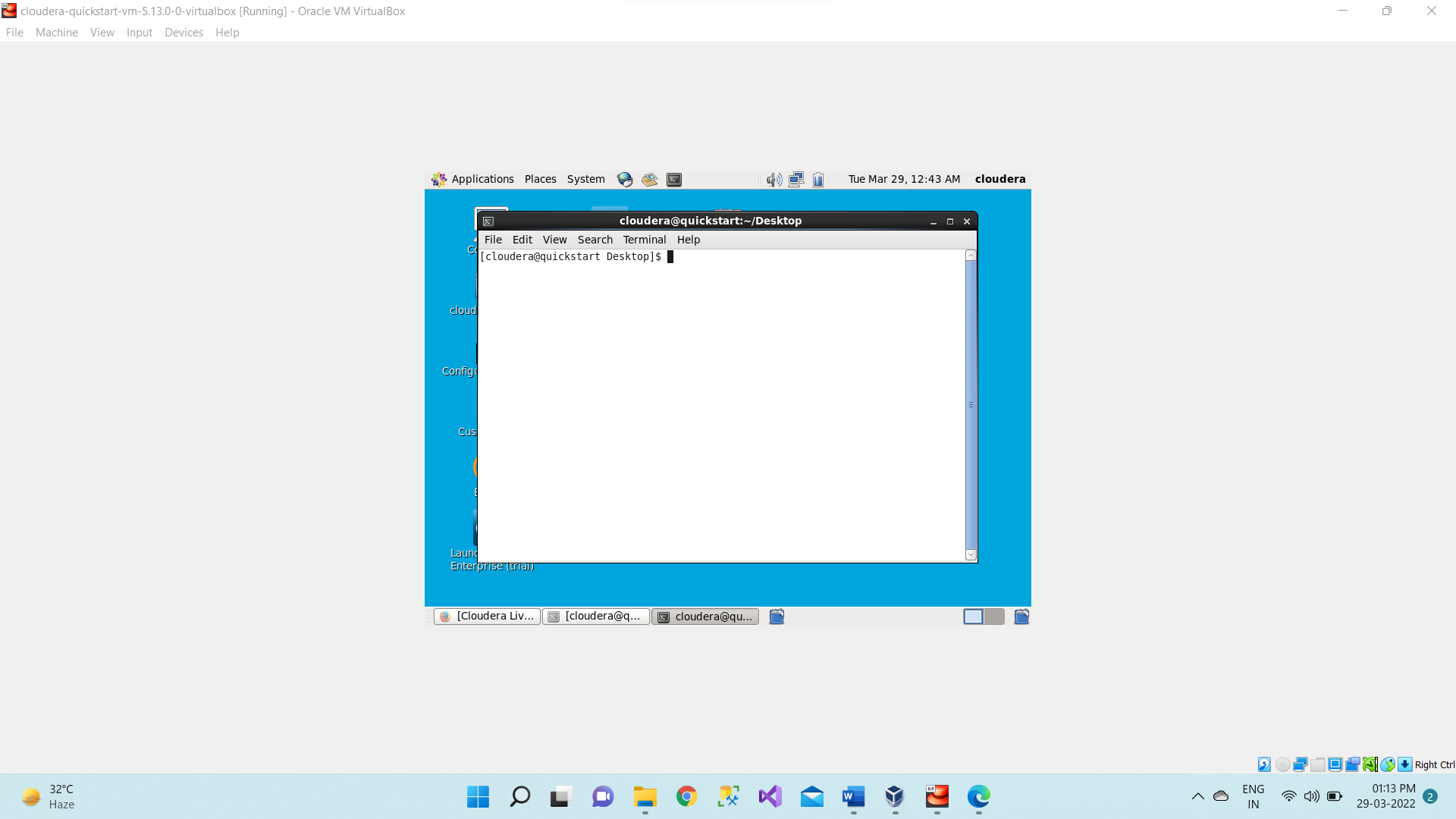


1. Let see any other table. Here we are displaying department table it has 6 rows in it.

**select \* from departments;**

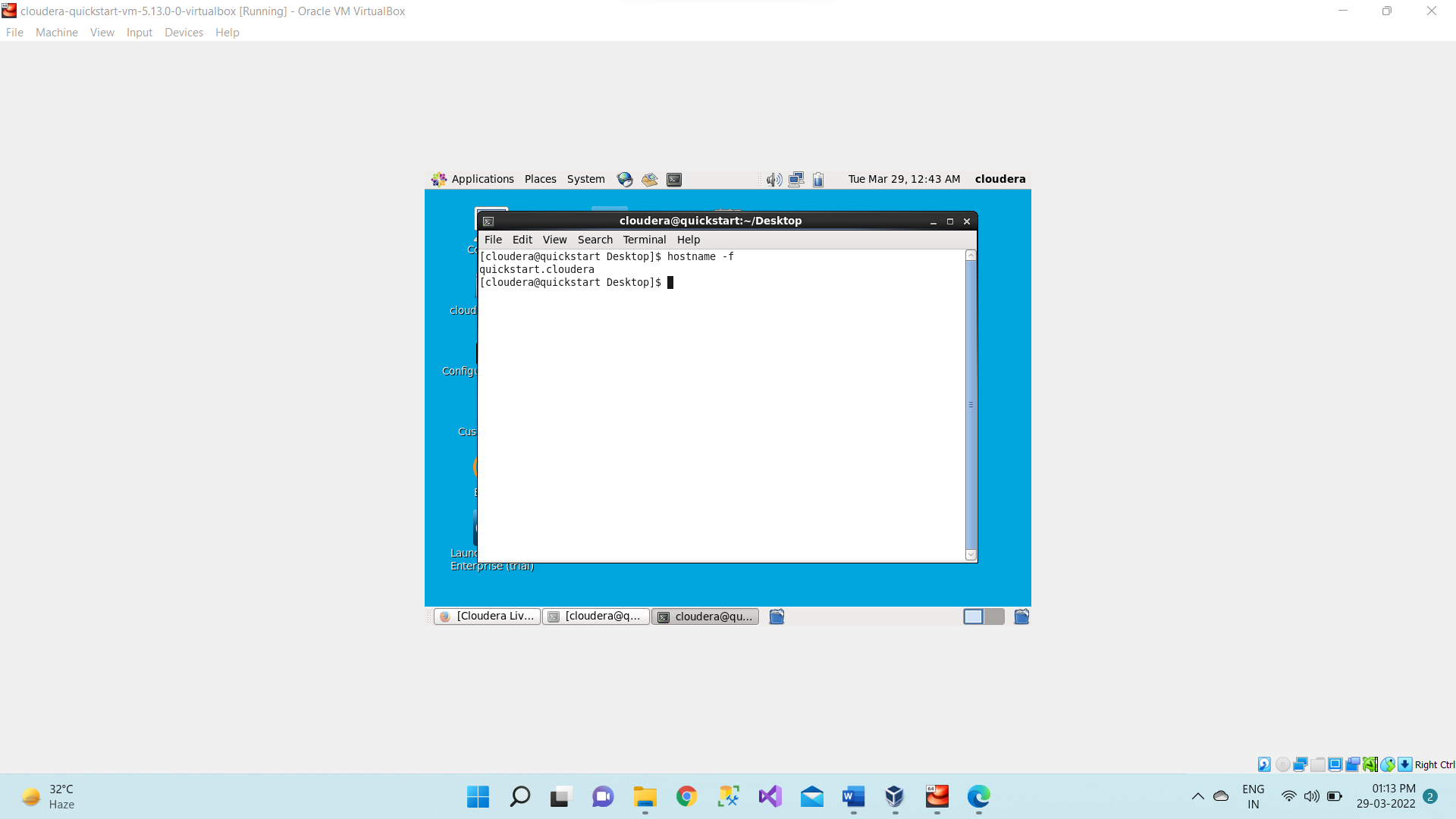


1. Open the new terminal for running command for sqoop.



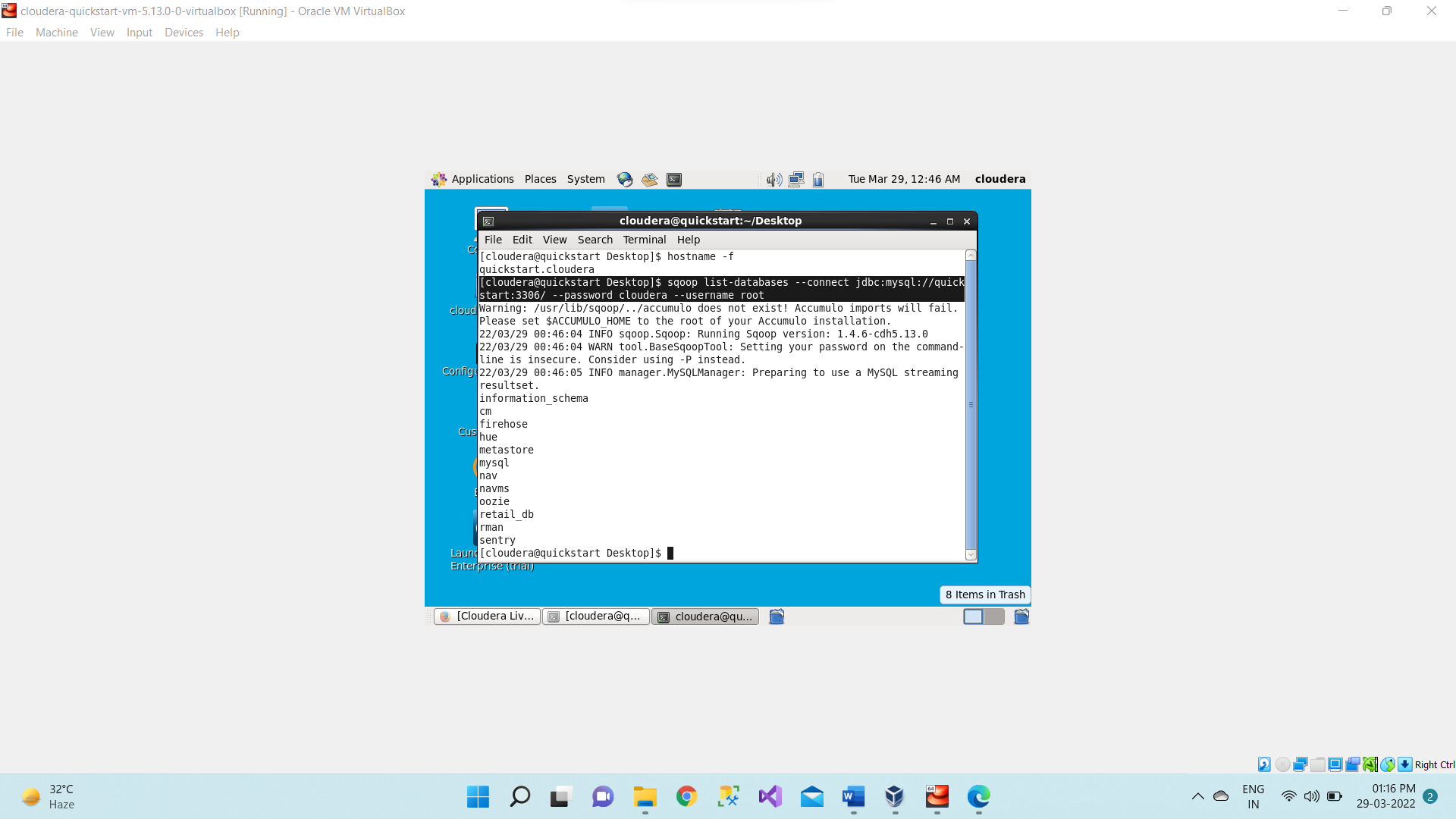
1. We will require hostname for this sqoop.

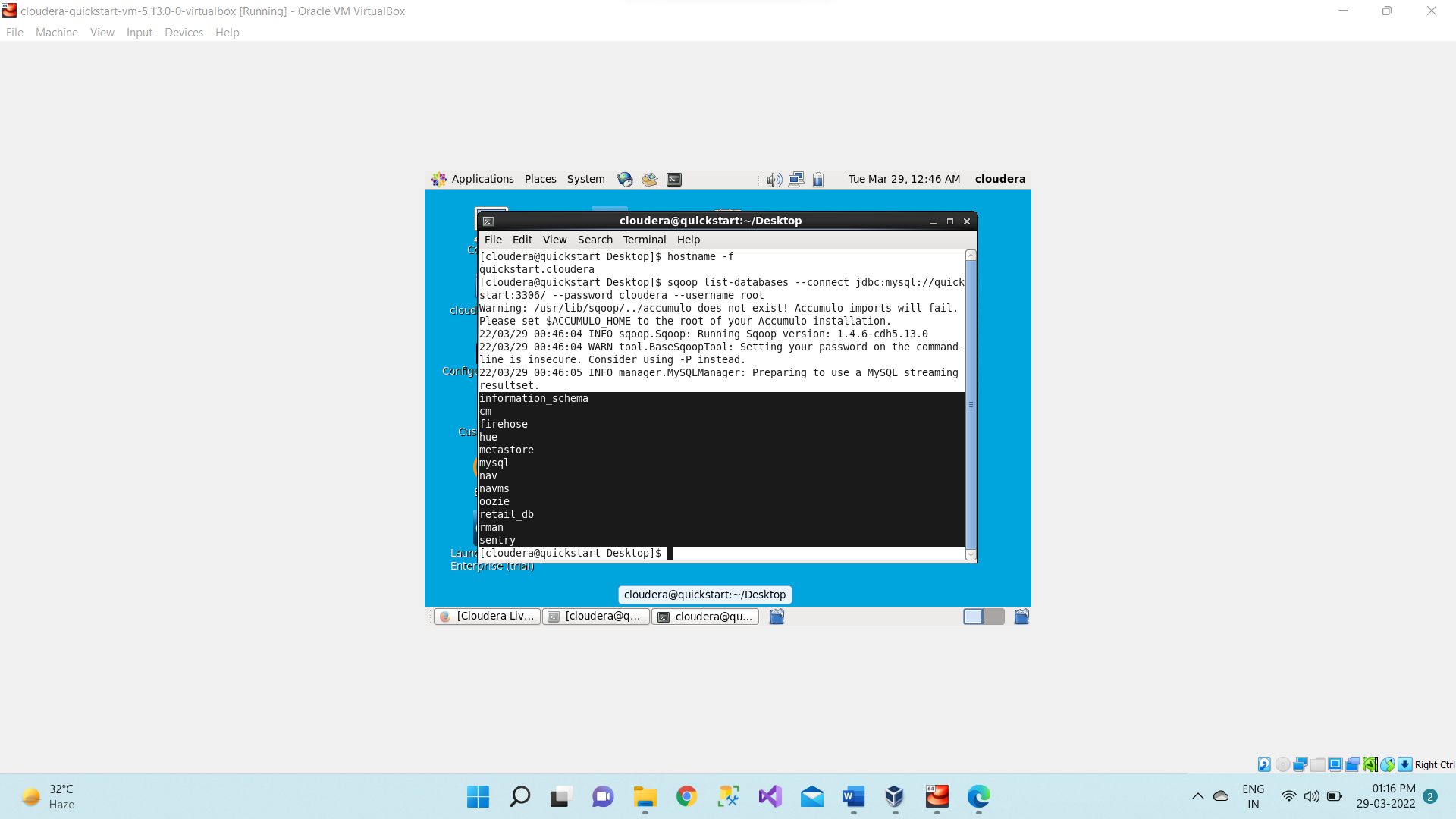
**hostname –f**



1. Then if we want to list down all the databases we will use below sqoop command.

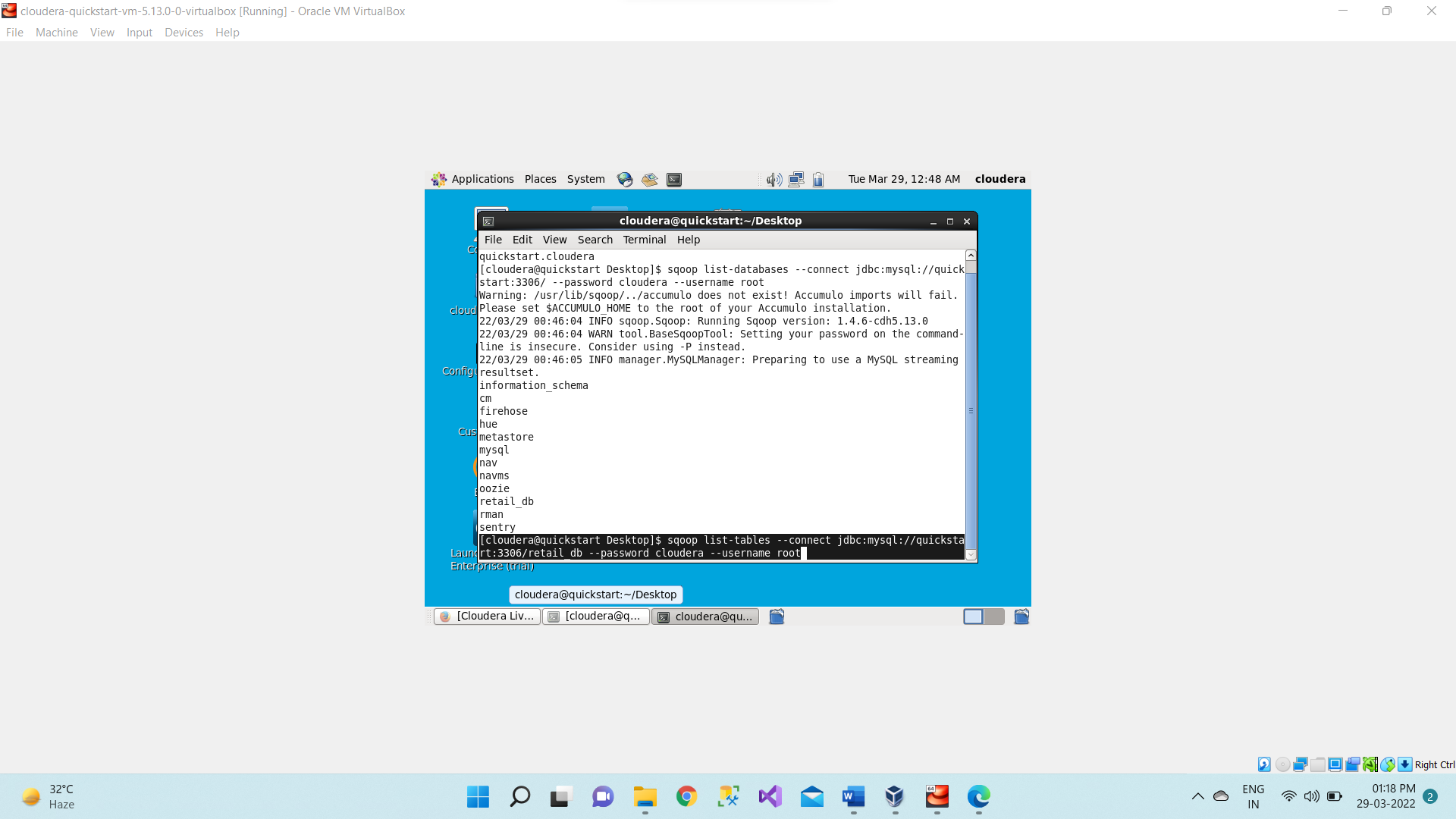
**sqoop list-databases --connect jdbc:mysql://quickstart:3306/ --password cloudera -- username root;**

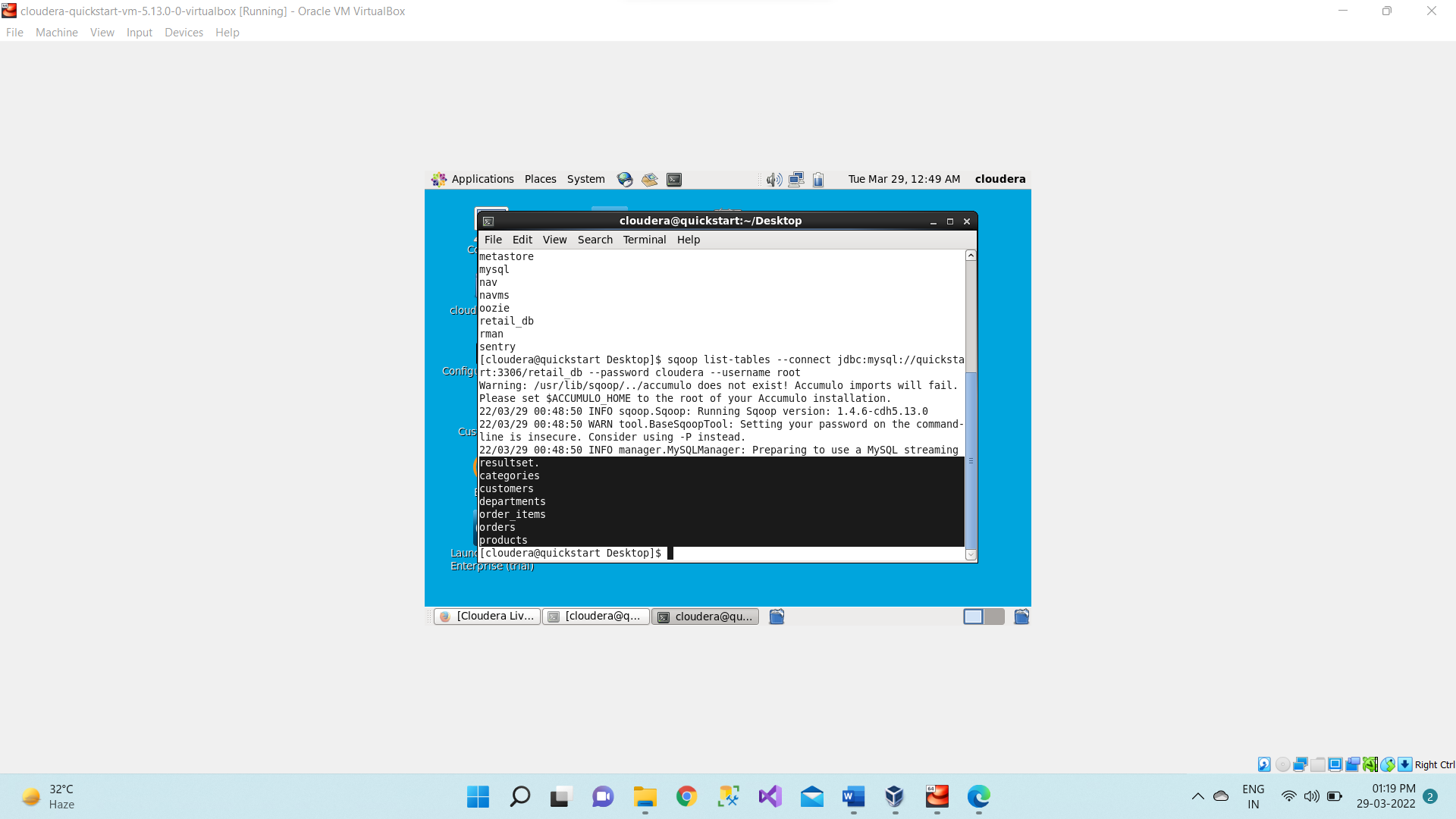




1. Now we will list out all the tables using below command.

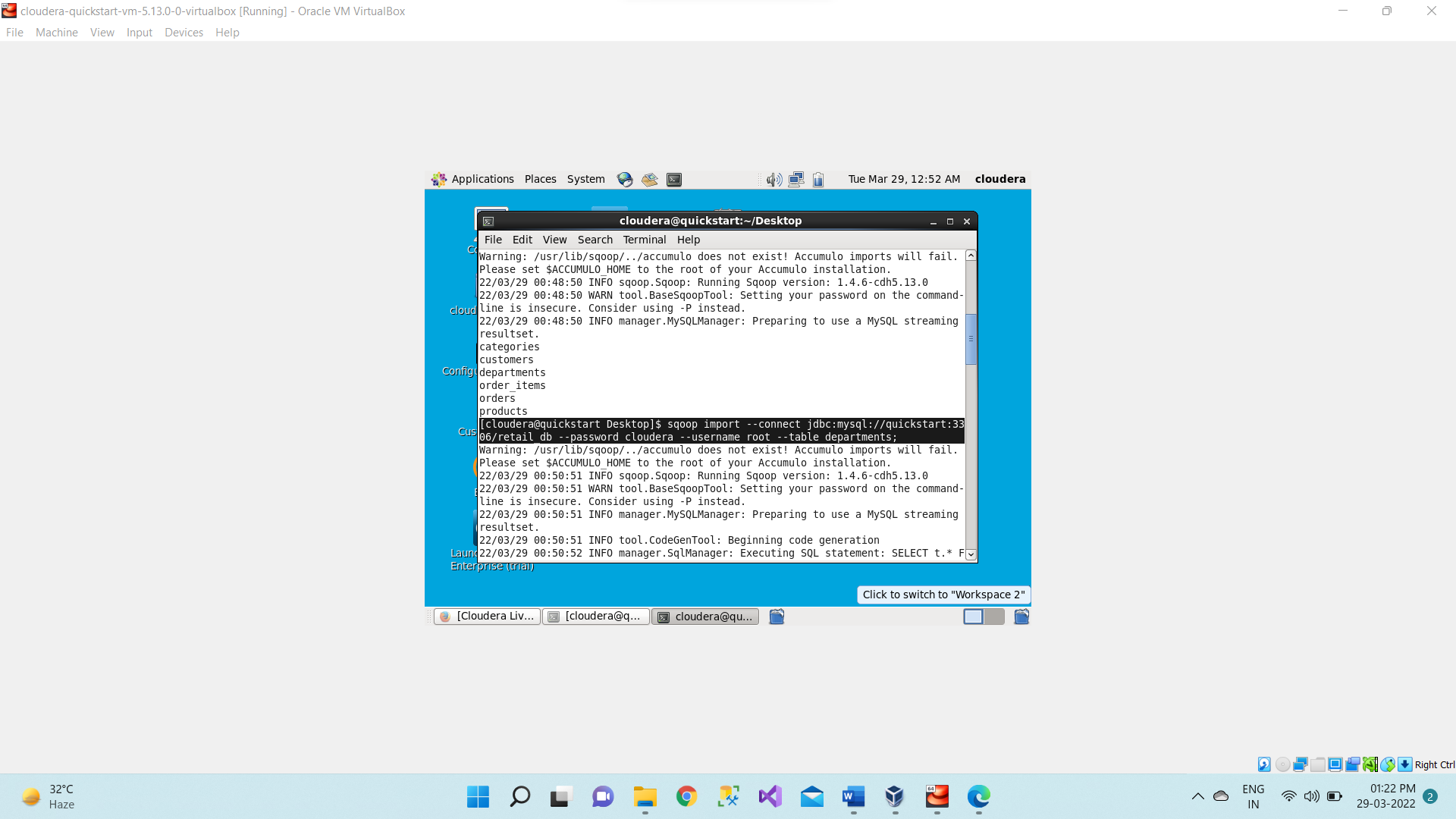
**sqoop list-tables --connect jdbc:mysql://quickstart:3306/retail\_db --password cloudera -- username root;**

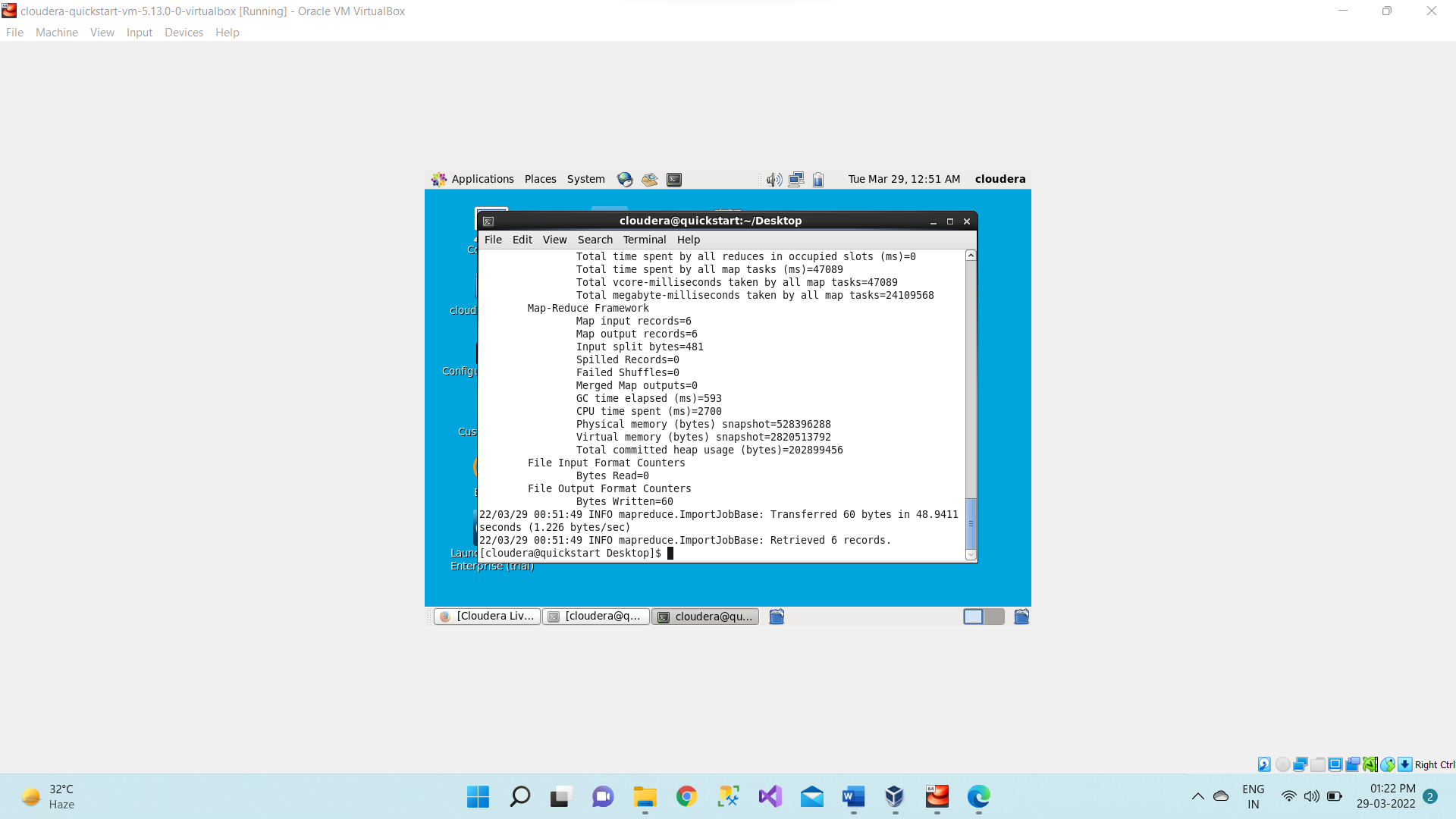




1. Now we will start with the import and export tools of the Hadoop. We want to Import table “departments” from reatail\_db database which are present inside in mysql.

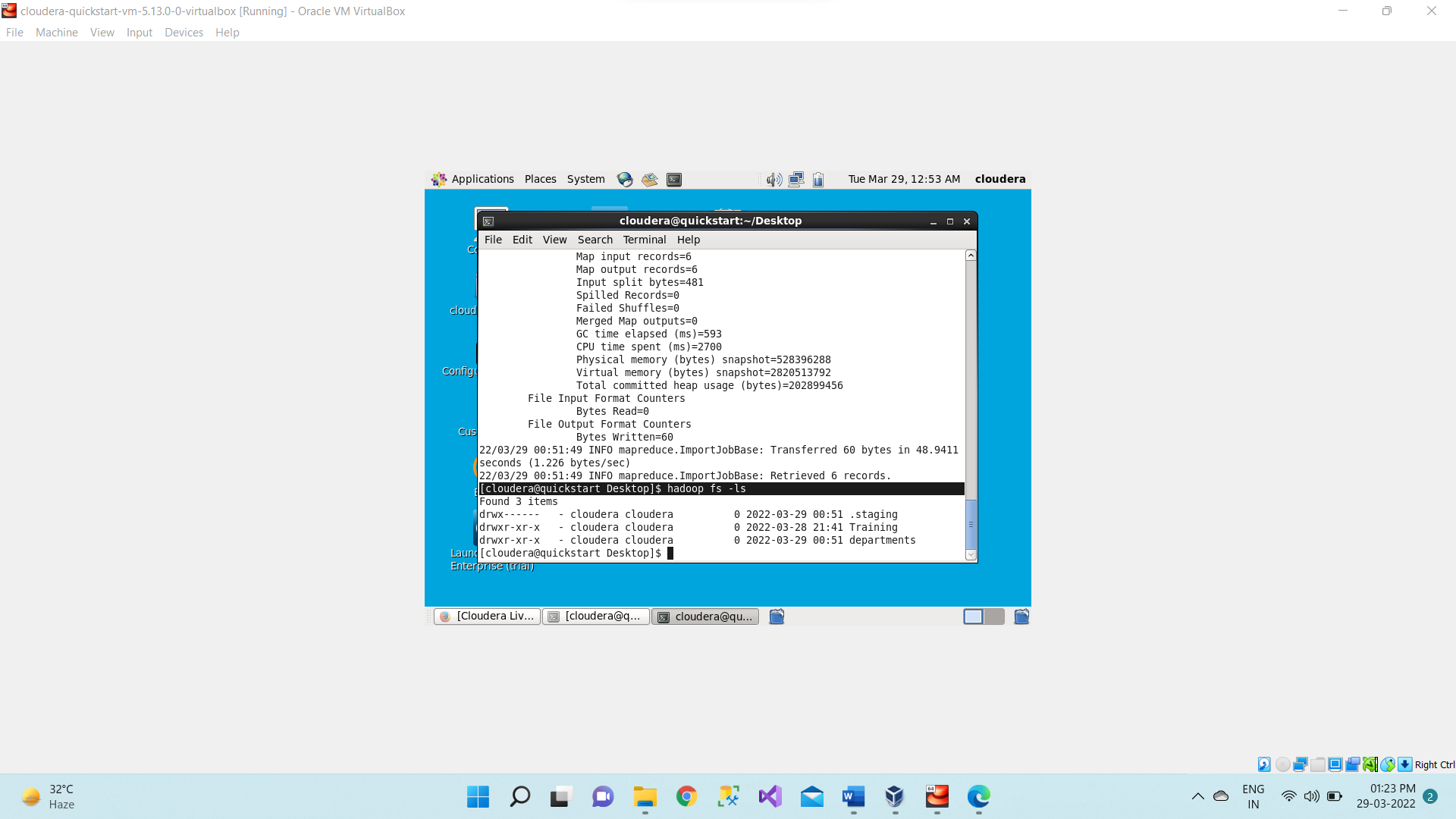
**sqoop import --connect jdbc:mysql://quickstart:3306/retail\_db --password cloudera -- username root --table departments**



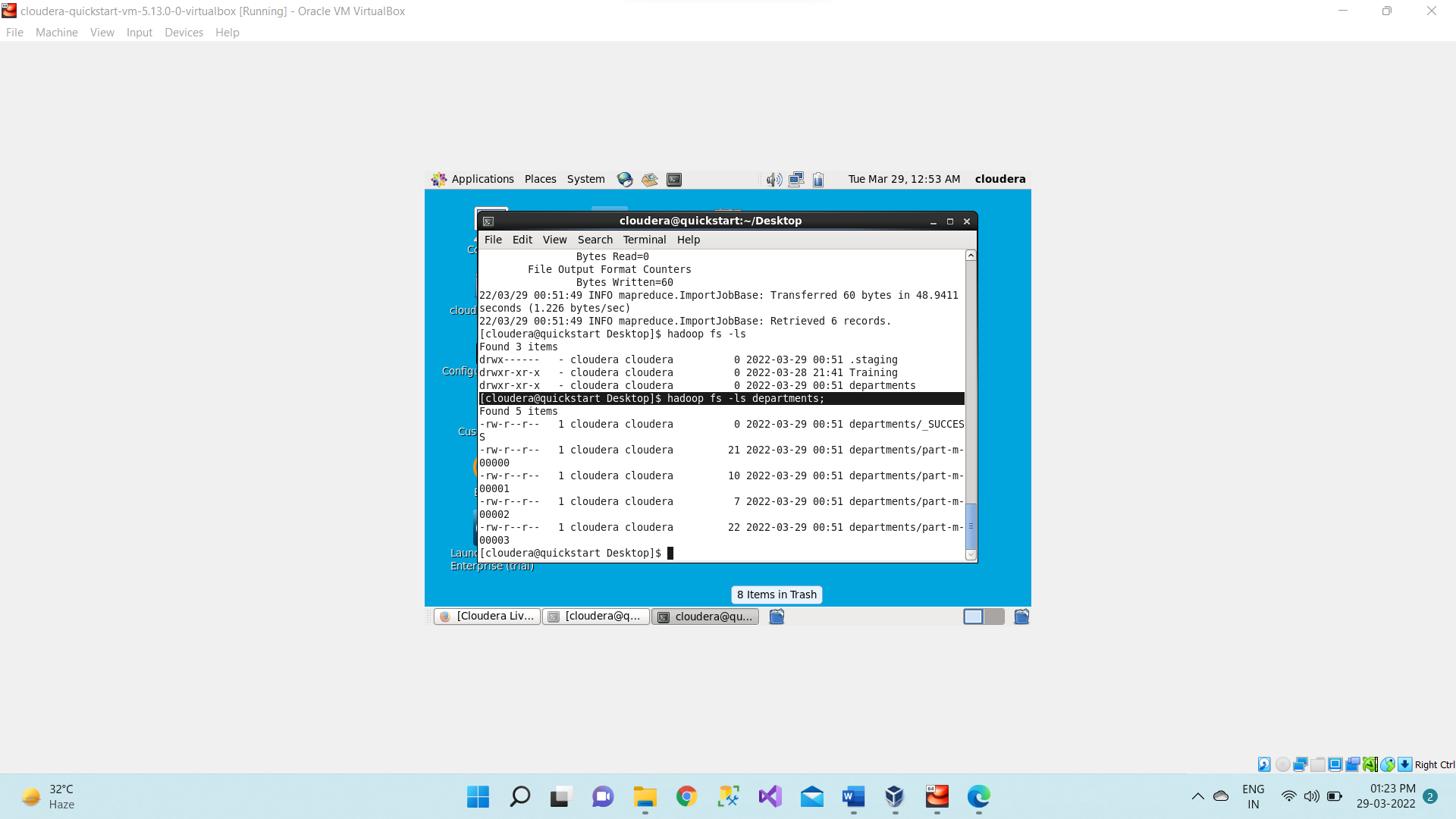


1. Now we will see whether all departments table successfully imported from mysql in Hadoop (hdfs) or not using below command.

**hadoop fs –ls**



1. Now we will see what inside this deparment using below command.

**hadoop fs –ls deparments;**

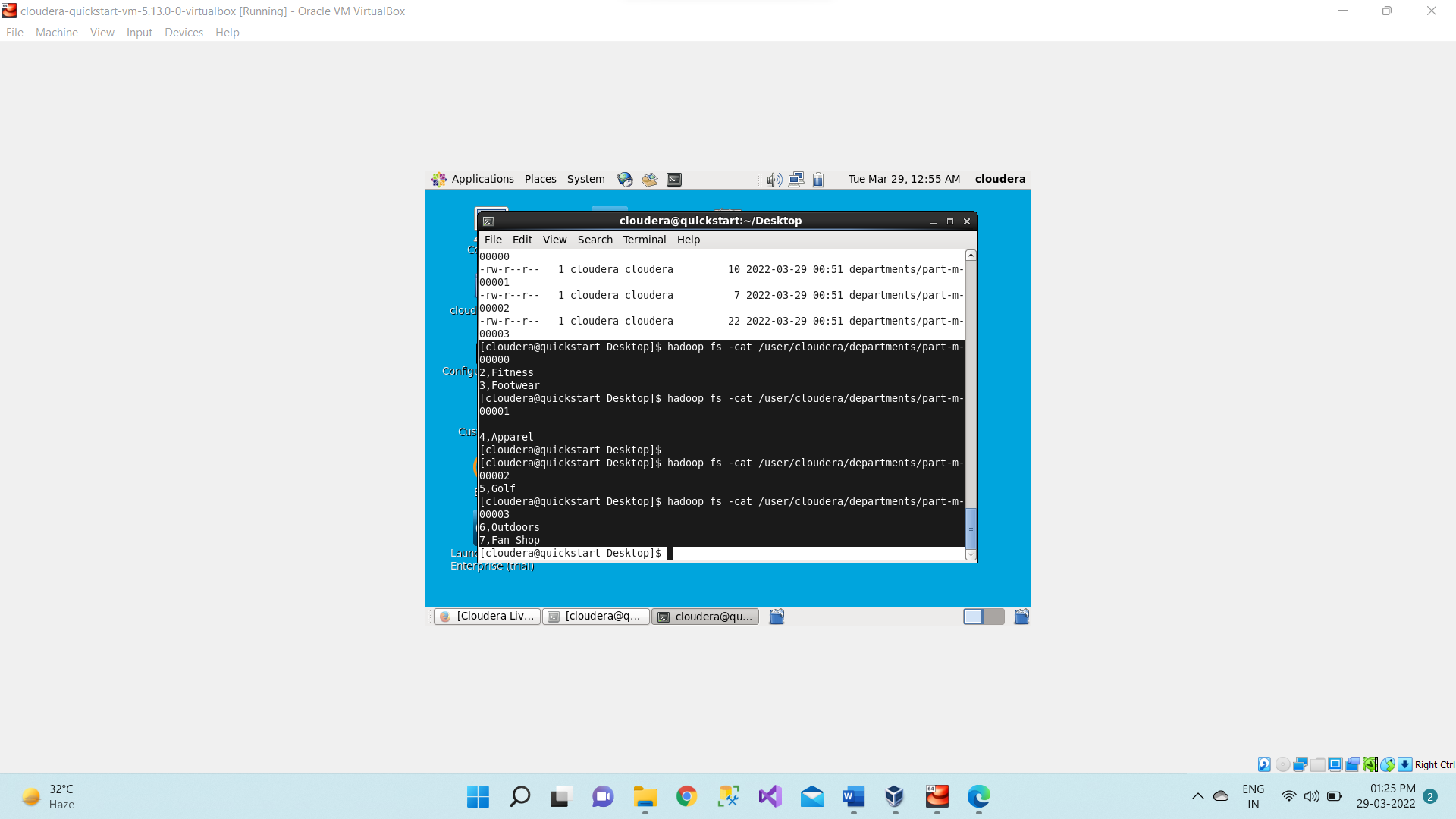
1. Now we will see what there inside this some of the part m files so that can be done with the help of below commands.

**hadoop fs -cat /user/cloudera/departments/part-m-00000**

**hadoop fs -cat /user/cloudera/departments/part-m-00001**

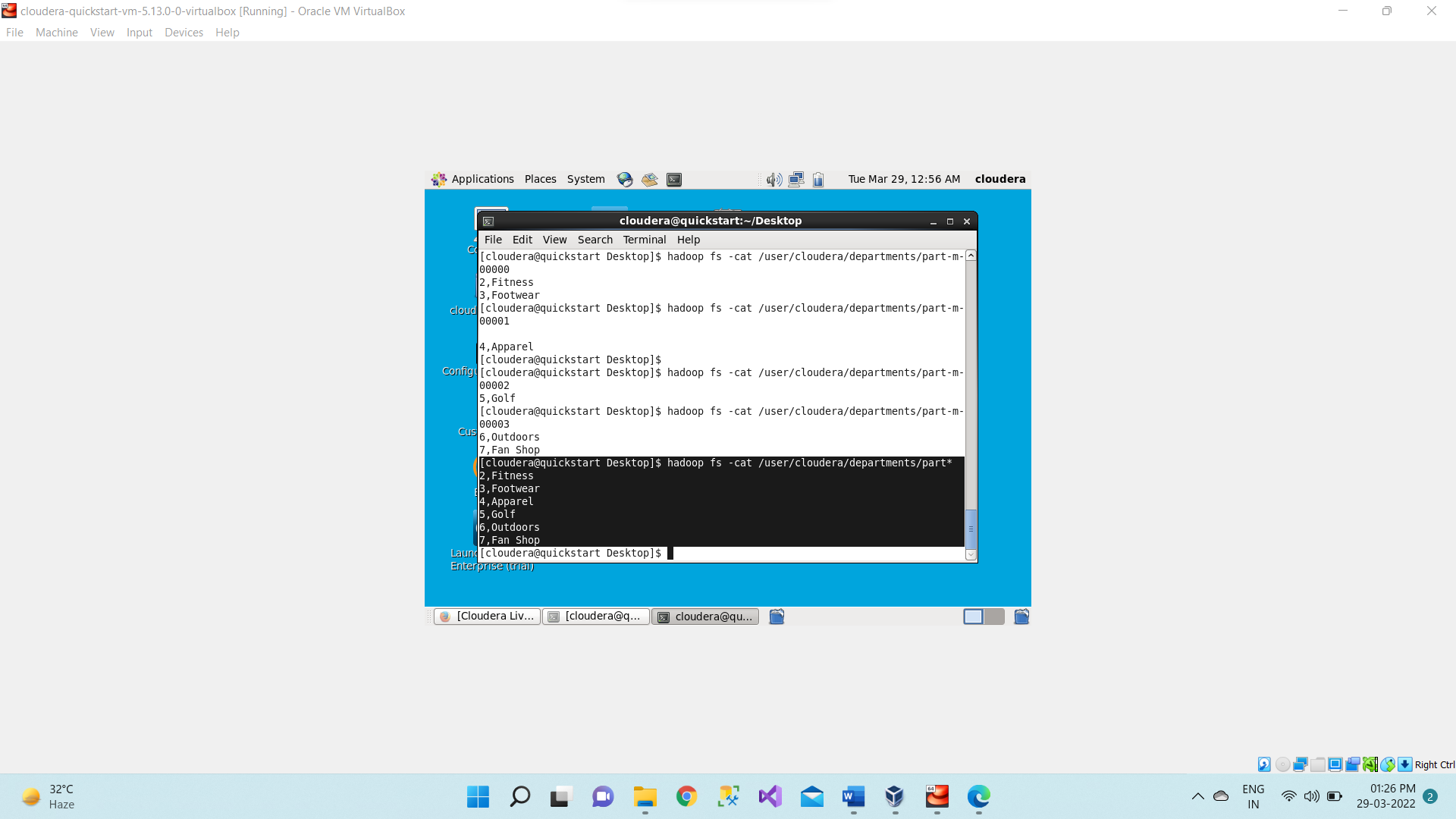
**hadoop fs -cat /user/cloudera/departments/part-m-00002**

**hadoop fs -cat /user/cloudera/departments/part-m-00003**



1. If want to display output of all part-m files together we will use below command.

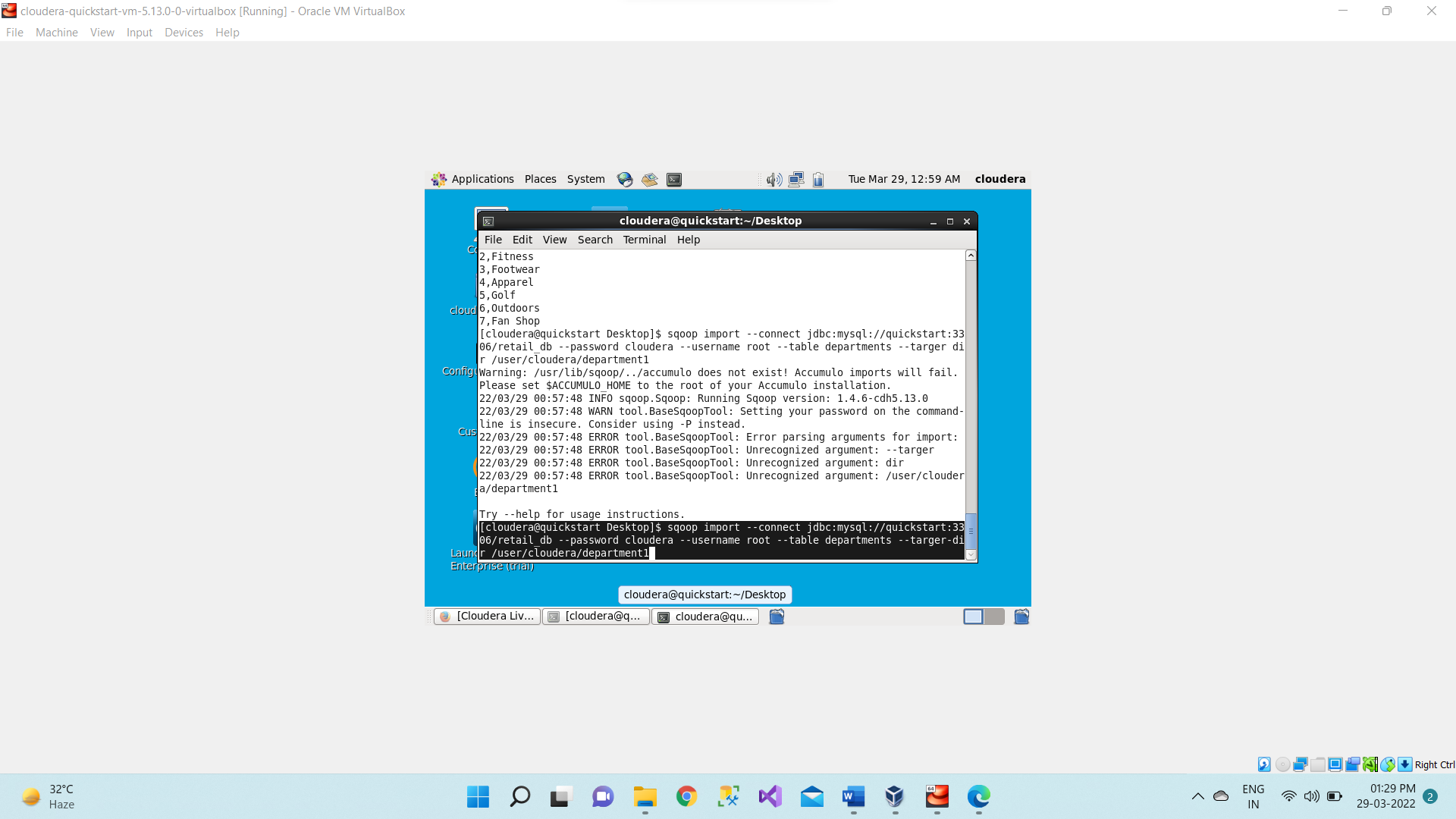
**hadoop fs -cat /user/cloudera/departments/part\***

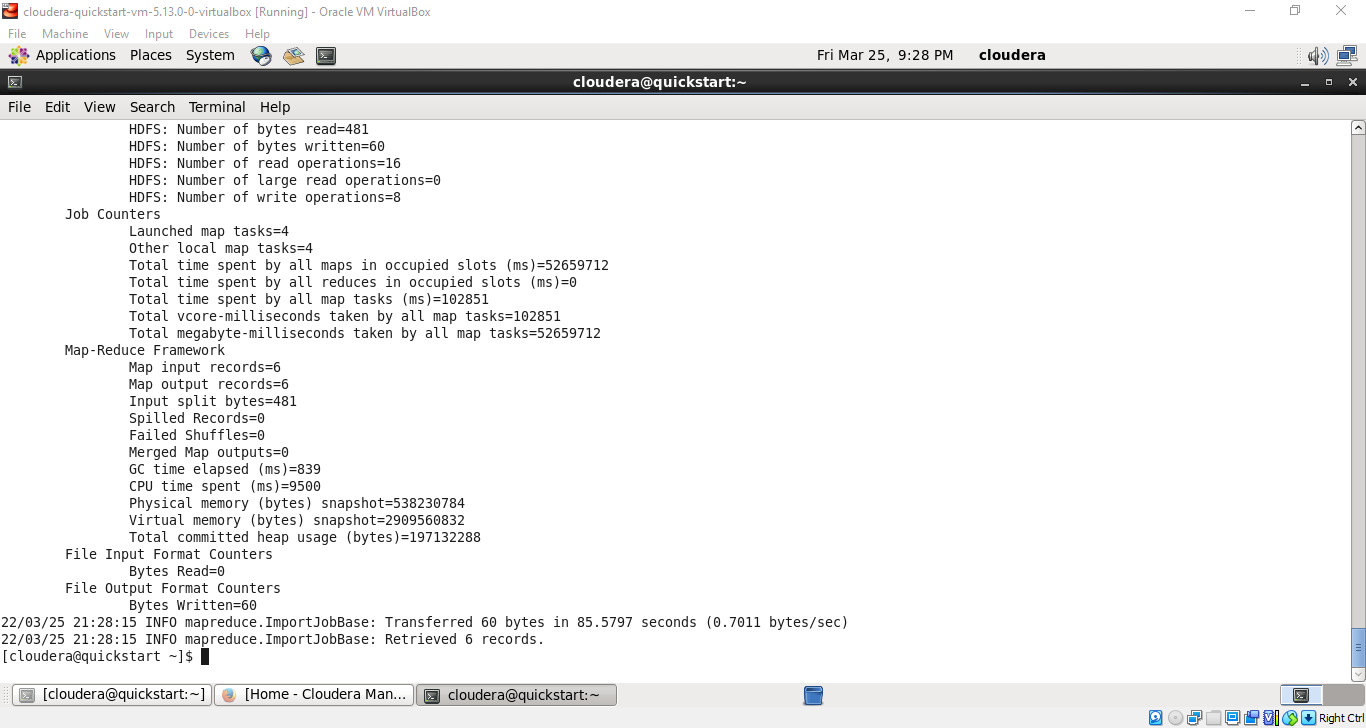


1. If we want to mention that where should we have our this output in the hdfs so for that we have to mention the target directory.

We will want to import my department table in –target directory as department1.

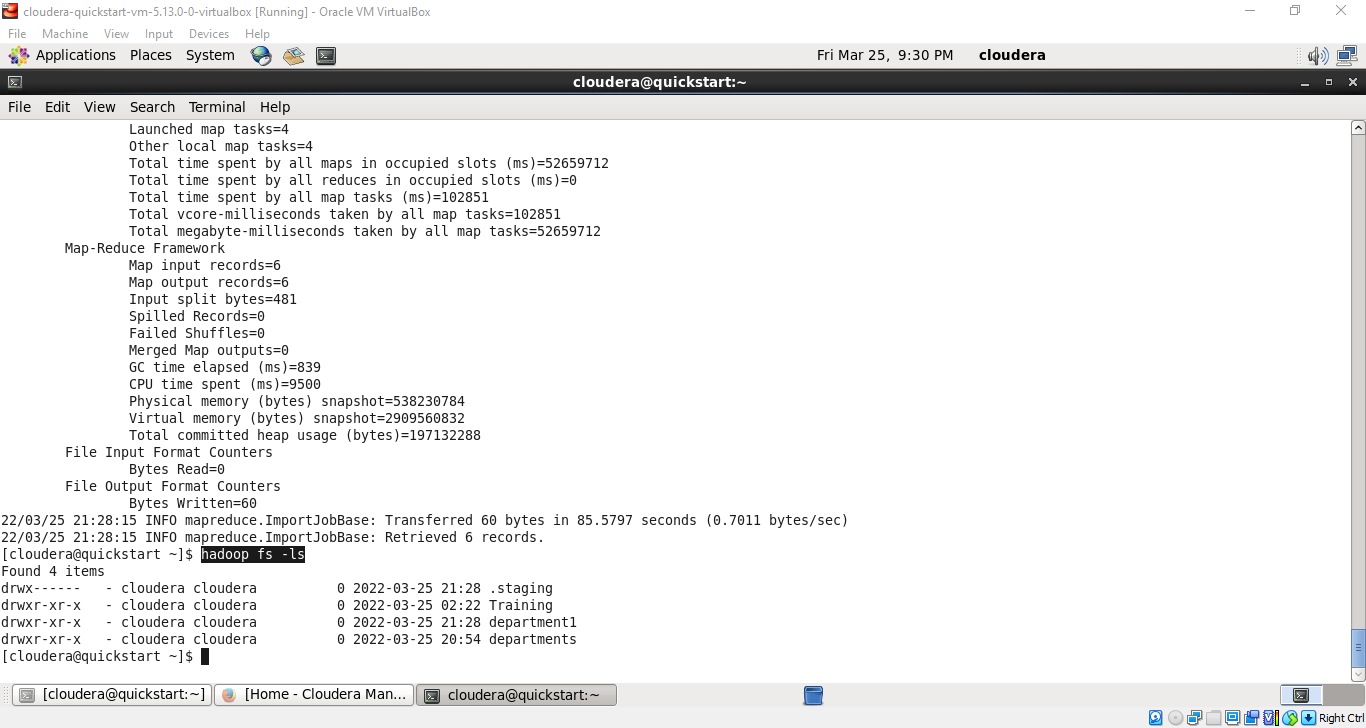
**sqoop import --connect jdbc:mysql://quickstart:3306/retail\_db --password cloudera -- username root --table departments --target-dir /user/cloudera/department1**





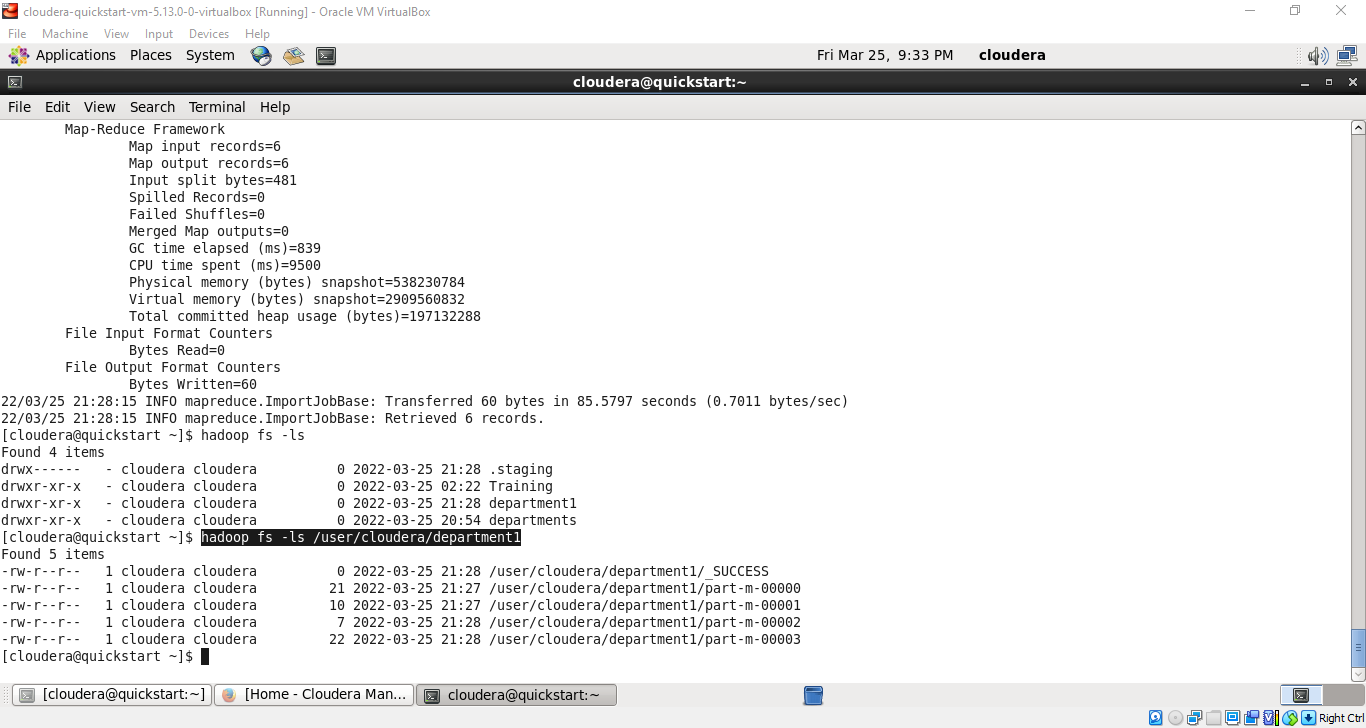
1. Now let’s check it using below command.

**hadoop fs -ls**



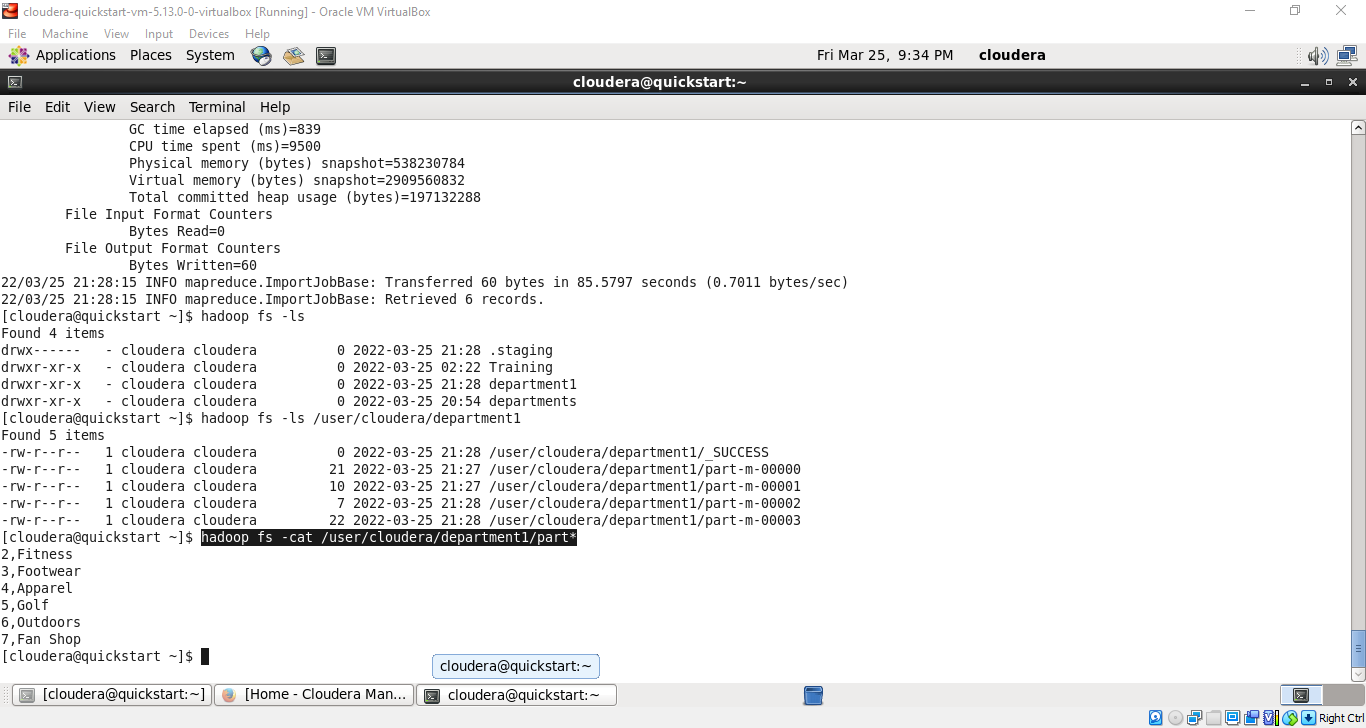
1. Now we will check what is present inside this department1 directory using below command.

**hadoop fs -ls /user/cloudera/department1**



1. Now we will read the content of these part-m files using below command.

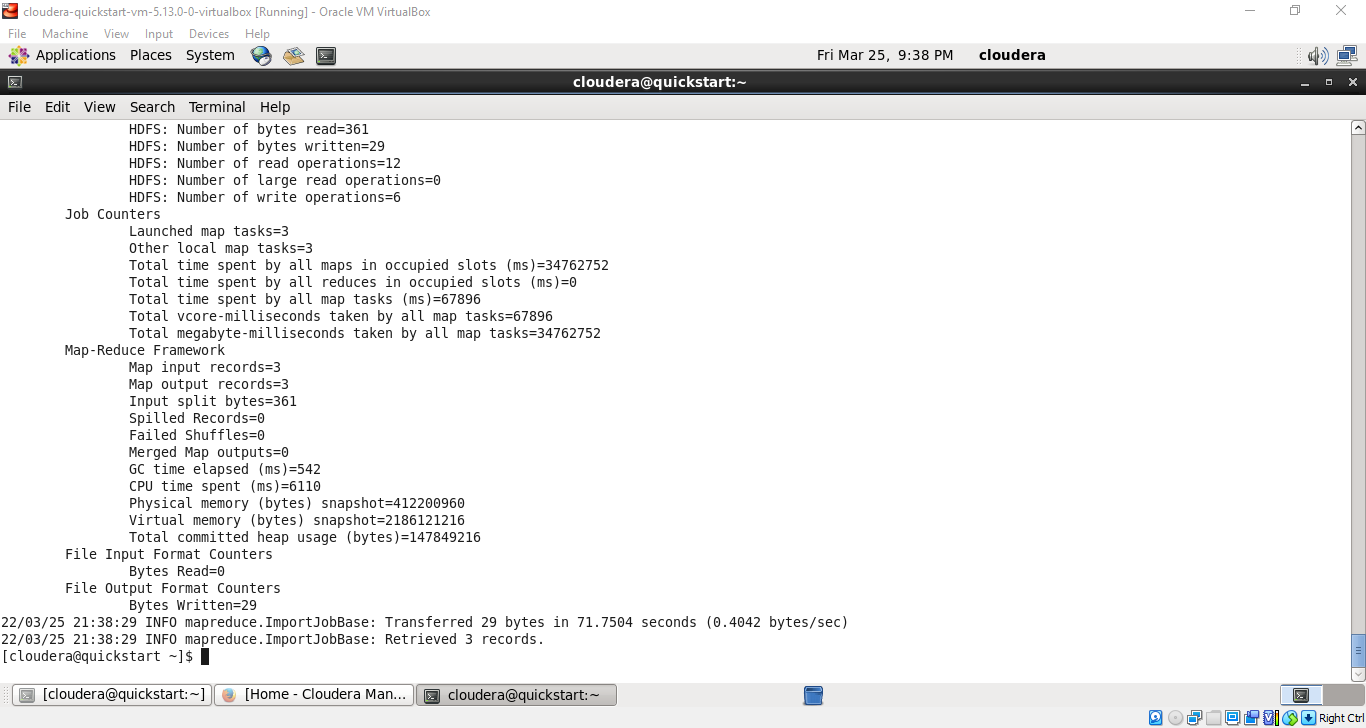
**hadoop fs -cat /user/cloudera/department1/part\***



1. Now we will filter out some or specific rows only from the departments table and have it in hdfs but before we will apply some conditions on the rows of the departments table and whichever rows will satisfy the condition only those rows are would be stored in the hdfs.

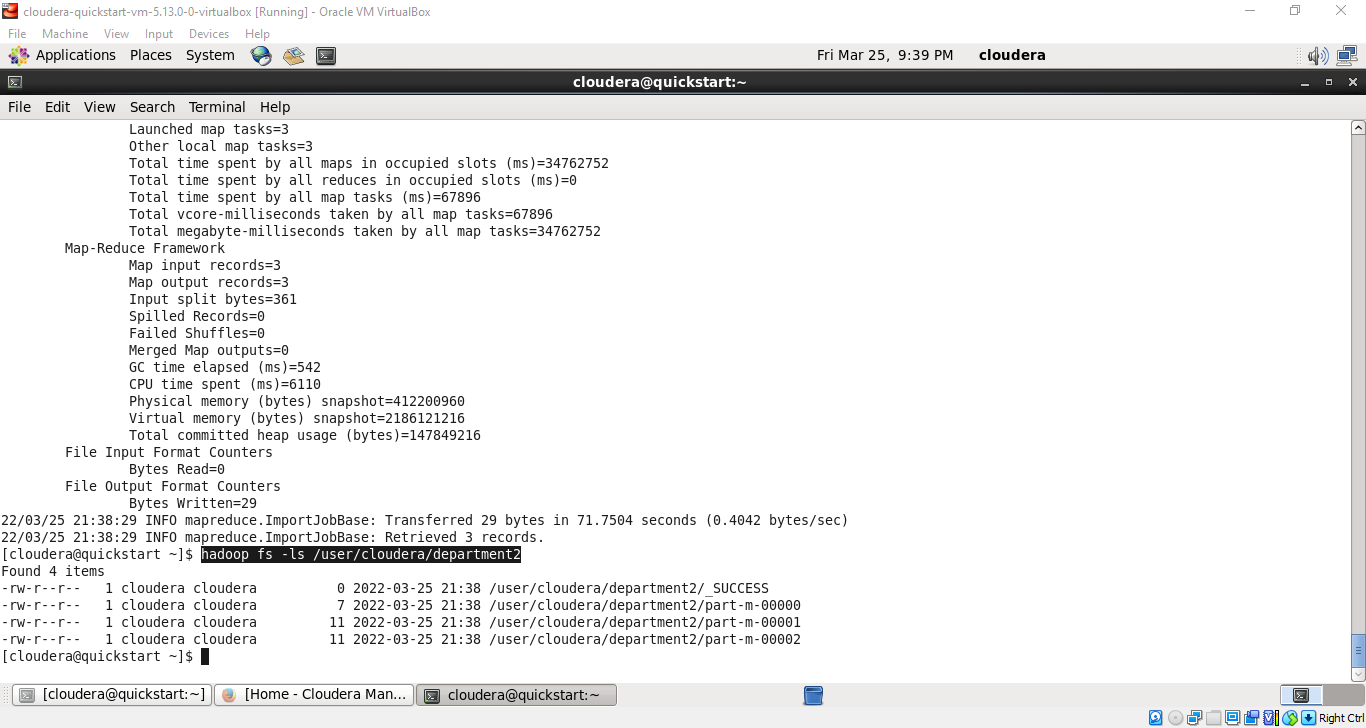
We want to fetch only those departments where department\_id is greater than 4.

**sqoop import --connect jdbc:mysql://quickstart:3306/retail\_db --password cloudera -- username root --table departments --where "department\_id>4" --target-dir /user/cloudera/department2;**



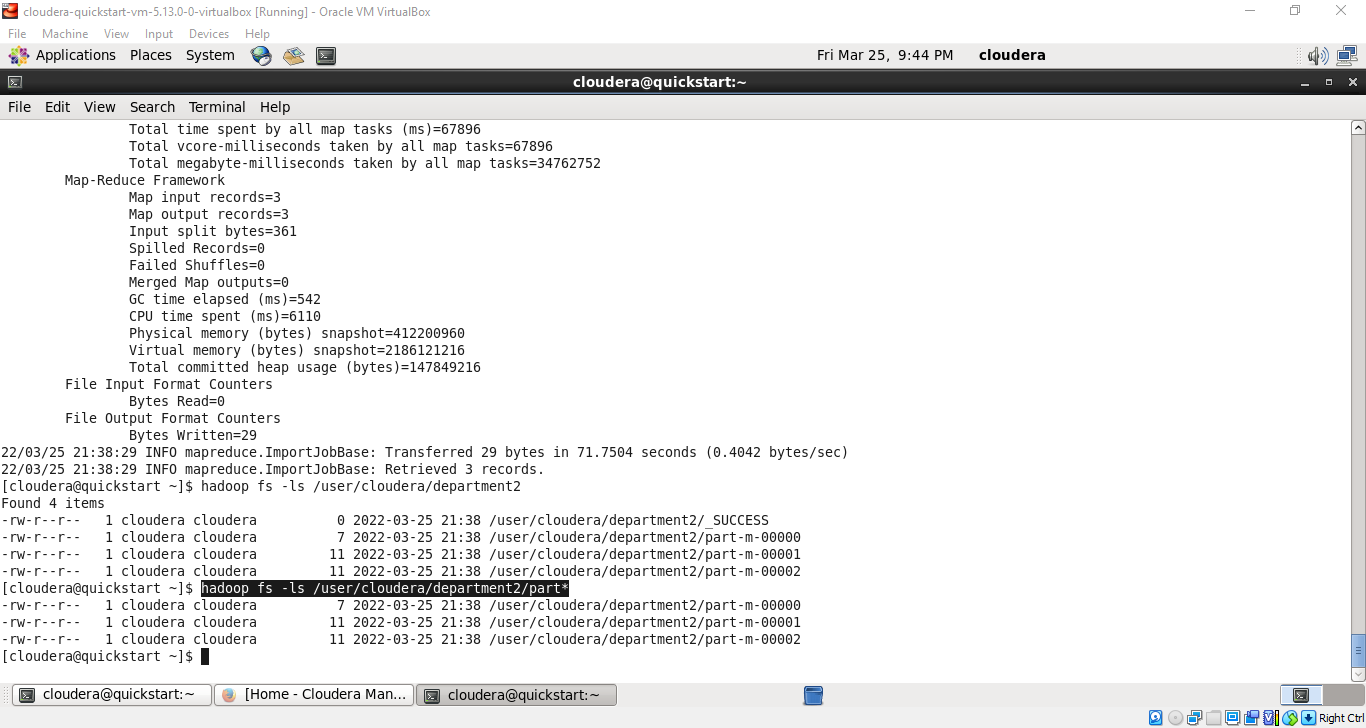
1. Now we will check it using below command.

**hadoop fs -ls /user/cloudera/department2**

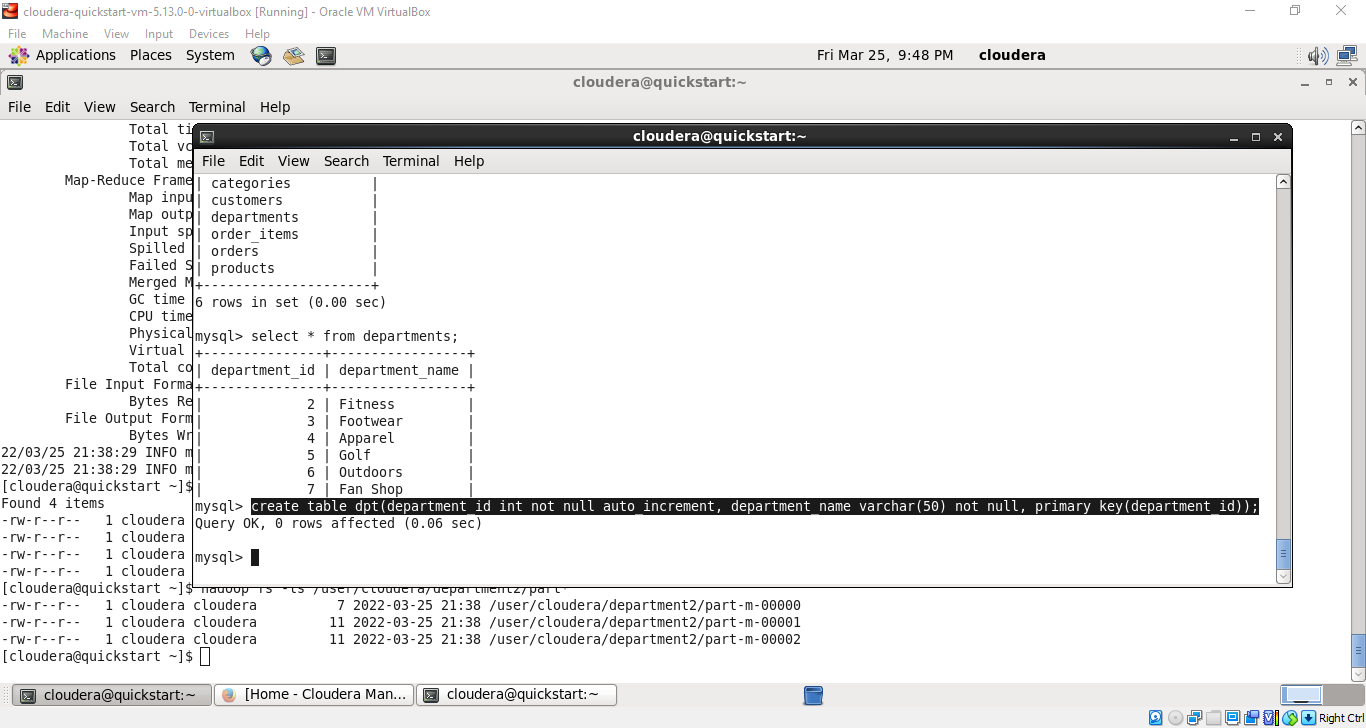


1. Now will read the content of these part-m files using cat command.

**hadoop fs -cat /user/cloudera/department2/part\***

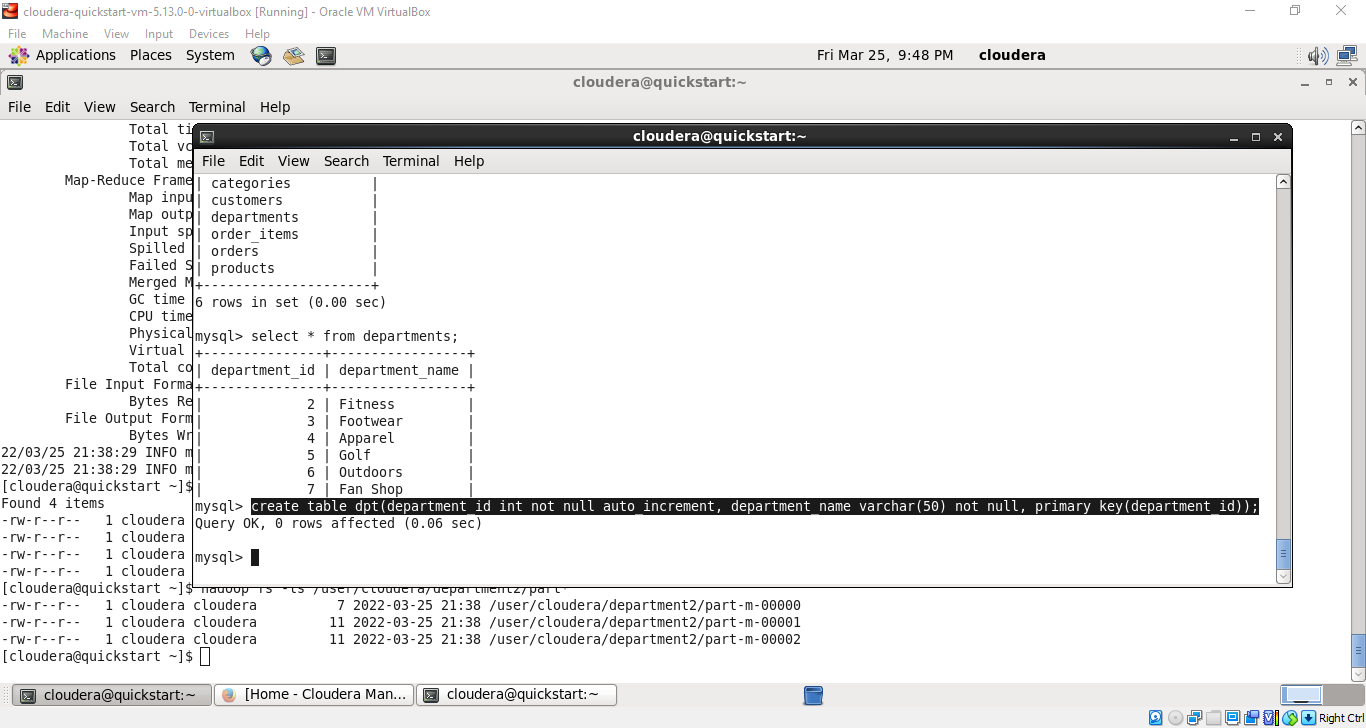


1. Now we will see the Export command. So what the export tool does is it will export the data from our hdfs to the RDBMS. So for that we need to have some table in mysql with some records so for that we will now move to mysql



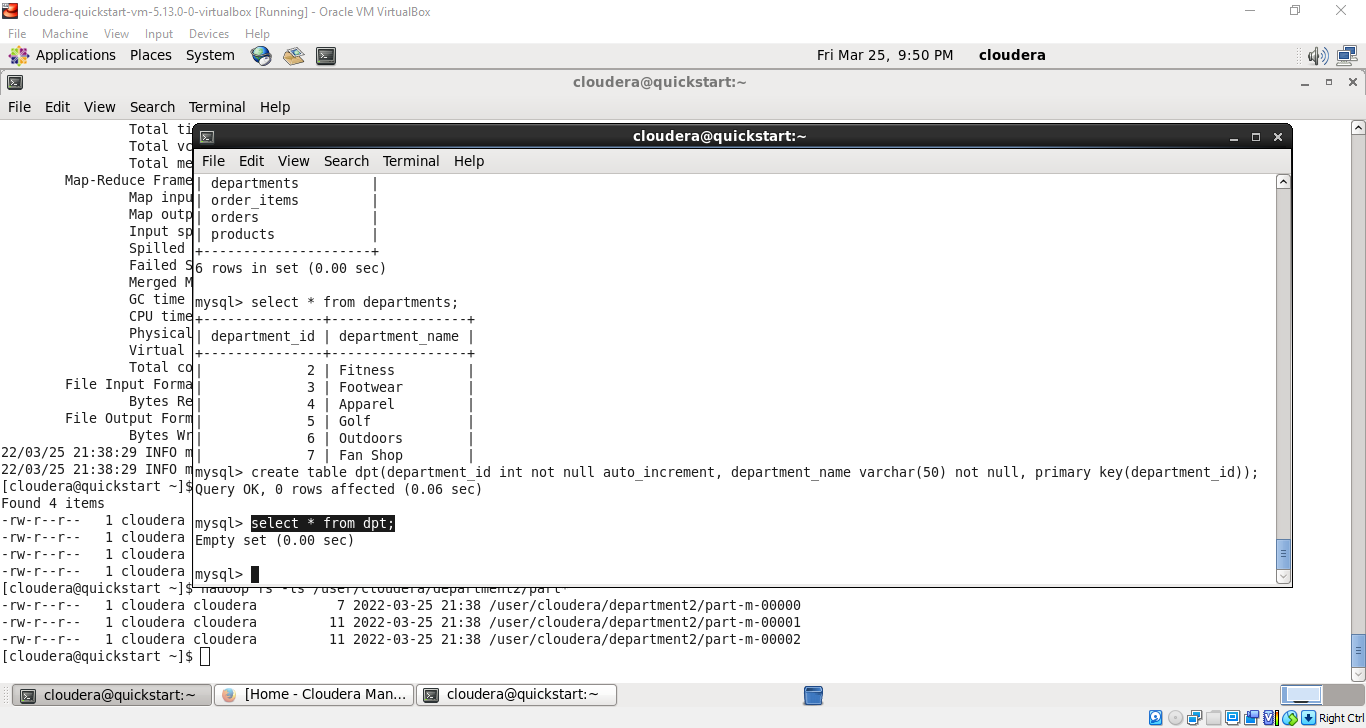
1. So here we will create the table “dpt” and it will be having two attributes as “department\_id” and “ department\_name”.

**create table dpt(department\_id int not null auto\_increment, department\_name varchar(50) not null, primary key(department\_id));**



1. Now we want to check what we have inside this dpt table.

**Select \* from dpt;**

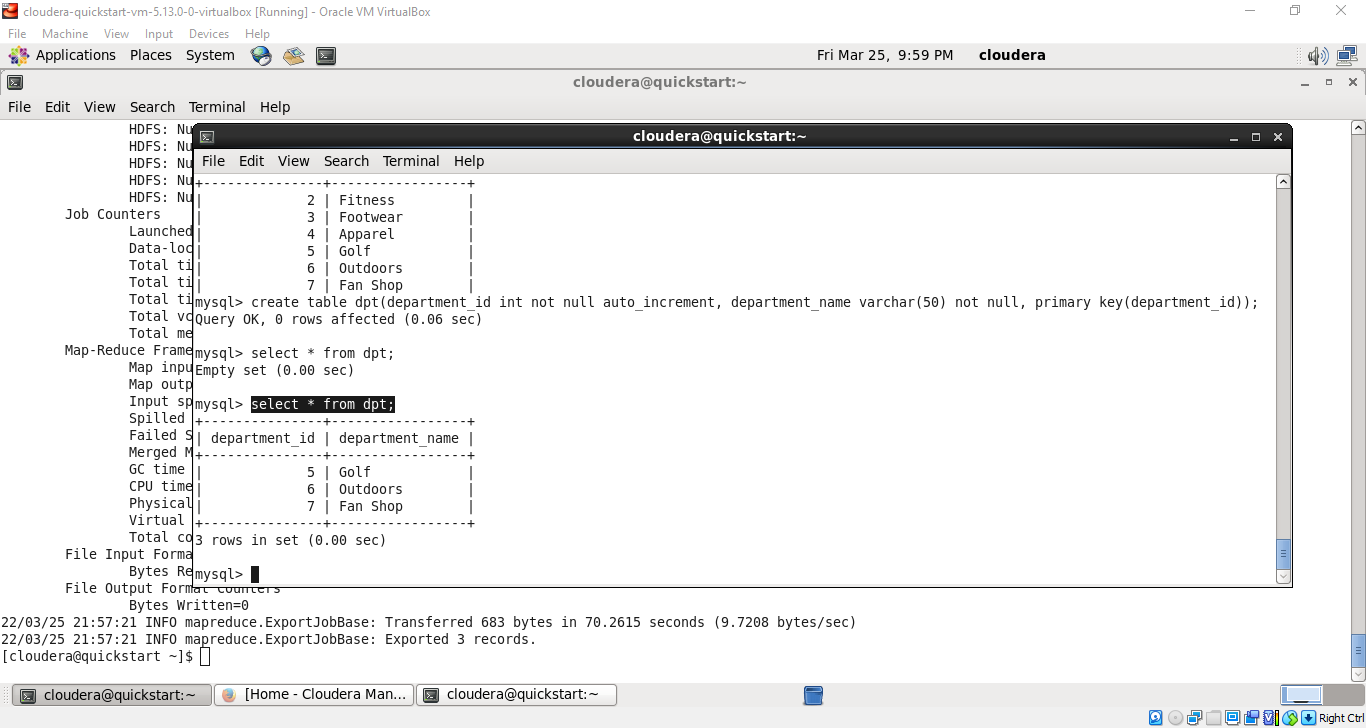


1. Now we will be exporting the data from the hdfs to dpt table of mysql. Now we will move to the sqoop terminal.

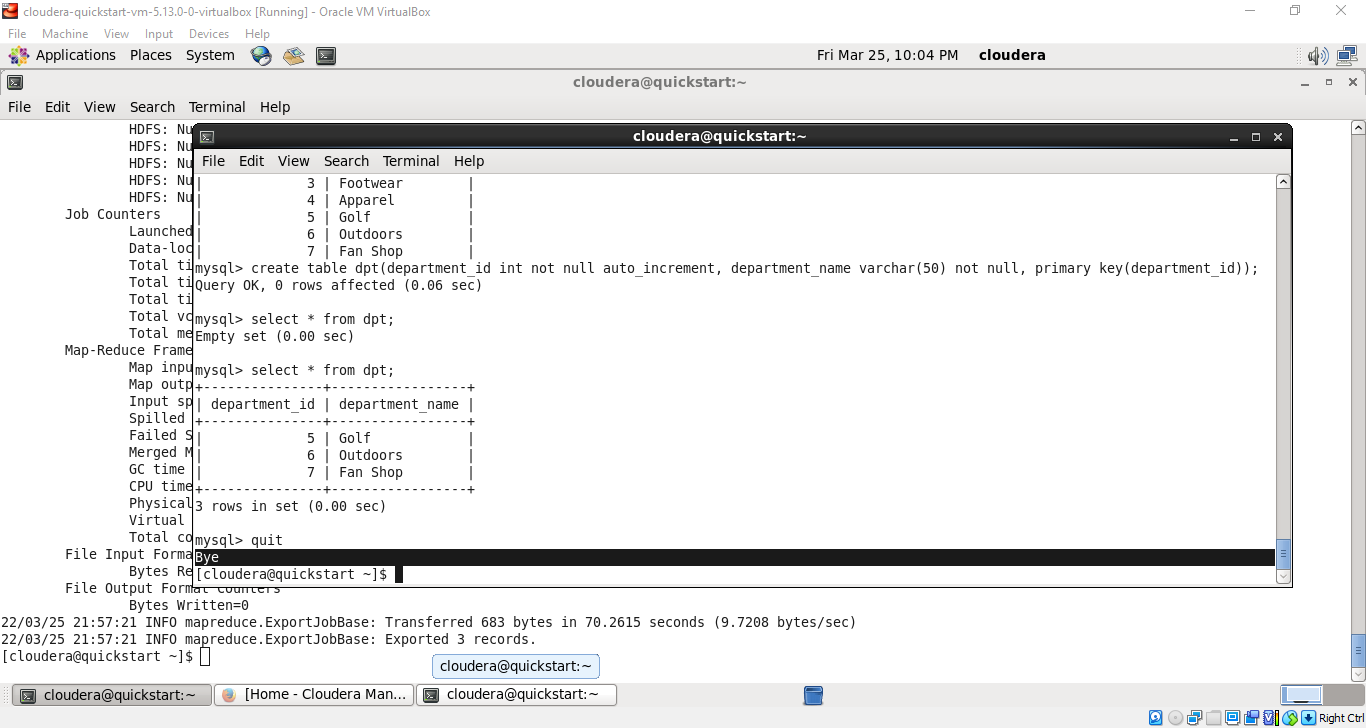


1. So now we are performing export operation using below command.

we are trying to export department2 which are present in cloudera to inside our dpt table which are present inside mysql. **sqoop export --connect jdbc:mysql://quickstart:3306/retail\_db --password cloudera -- username root --table dpt --export-dir /user/cloudera/department2;**



**Select \* from dpt;**



**As we can see these 3 records which are present in department2 table are successfully exported inside the dpt table of mysql.**