**Assignment-3**

|  |  |  |
| --- | --- | --- |
| **Student ID** | **Name** | **Email** |
| **8800149** | **Basavraj Jaliminche** | **Bjaliminche0149@econestogac.on.ca** |
| **8800060** | **Zarana Gohil** | **Zgohil0060@conestogac.on.ca** |
| **8807575** | **Dharti Patel** | **Dpatel757@conestogac.on.ca** |

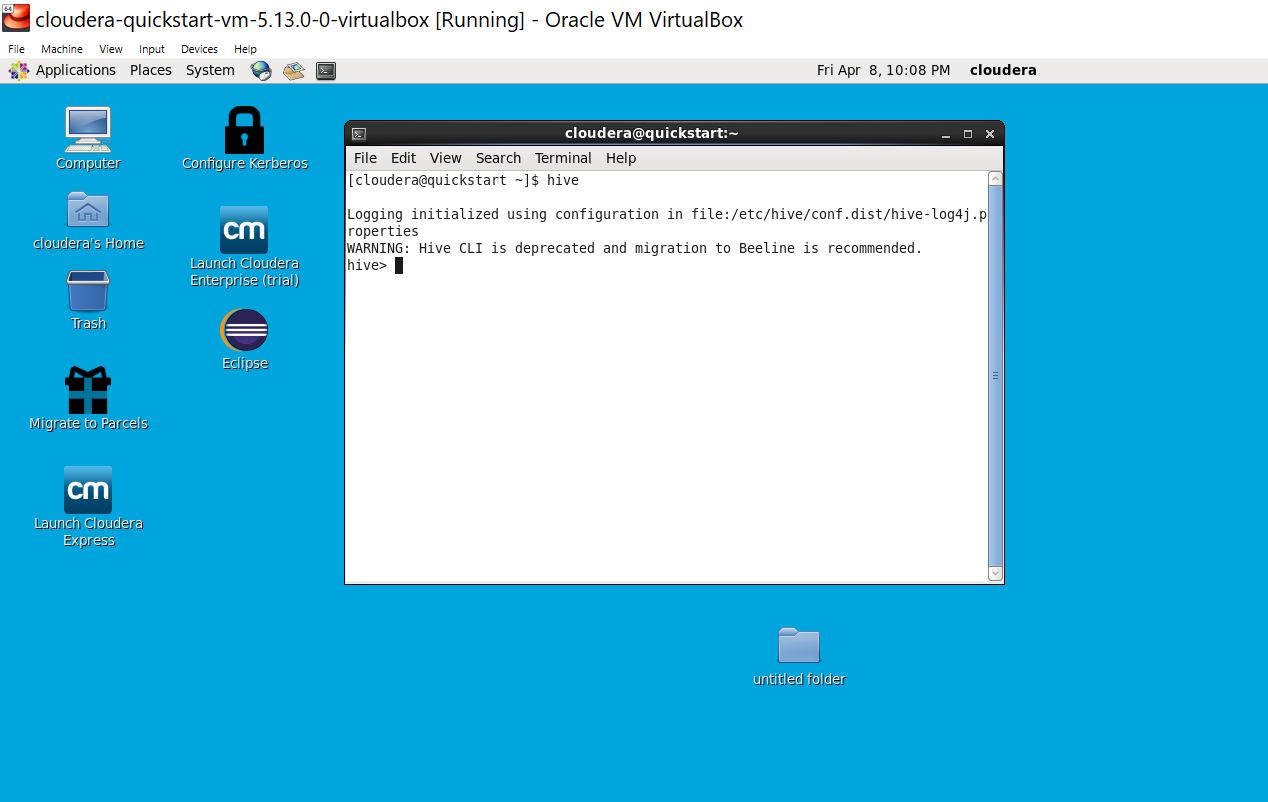
**Assignment-3**

**Laboratory Exercise-05**

**1.Install Software of cloudera**

**2. Run Hive from the command line.**

---open command line and type as $ hive

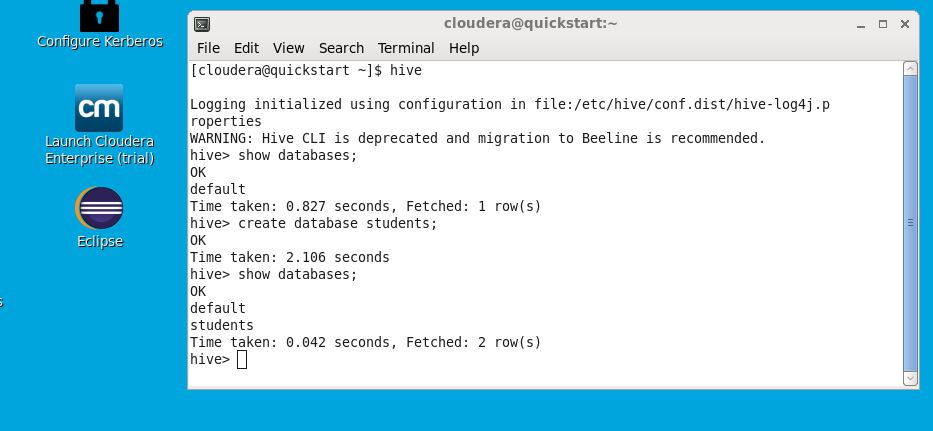


**3. Display all databases:**

-- On command line type as ‘show databases’;

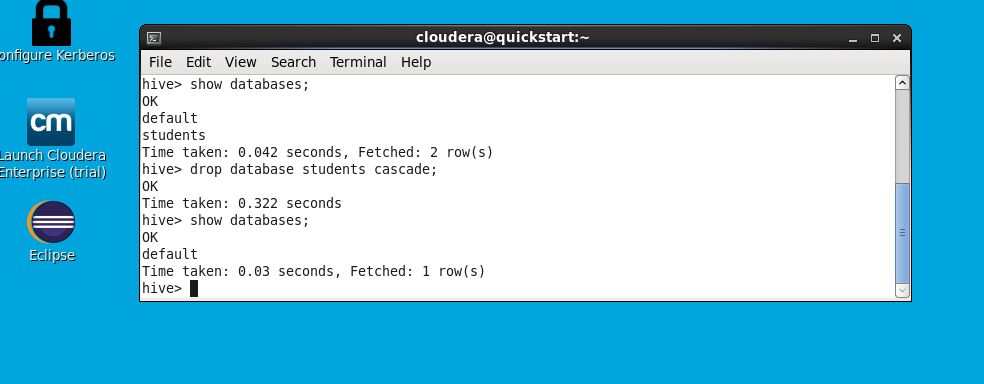


**4. Create a DB named students**

* Type on command line as ‘create database students’
* then you will get message as ‘ok’.
* For showing databases use command as ‘show database’

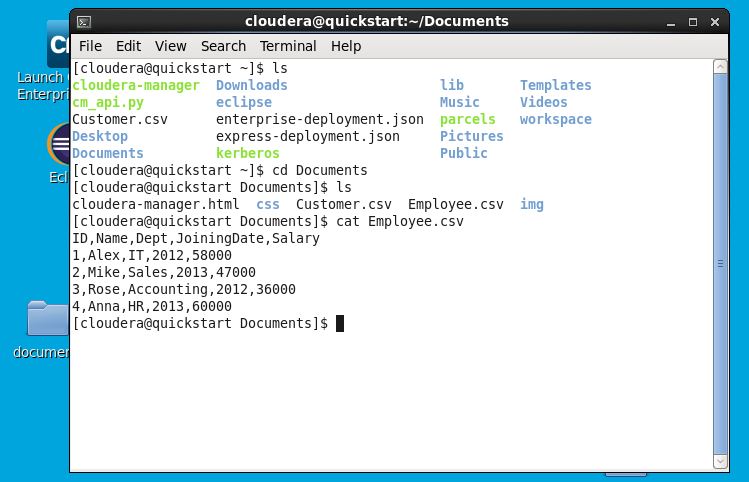
**5. Dropping the DB students**

* Type command as ‘drop database students’. Then student’s database is dropped.
* To obtain the result we will type command as ‘show databases’.



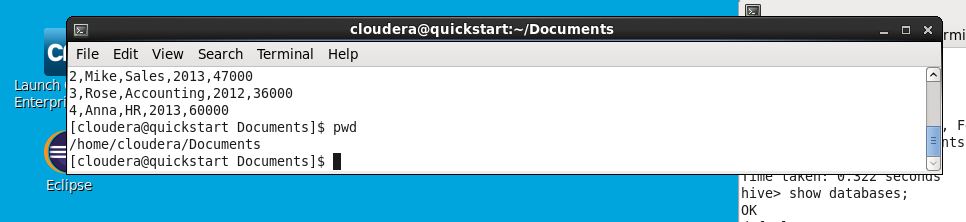
**6. Display the Employee.csv**

* We have a header line
* The data is comma-separated
* In Hadoop, you write the data once, and you read it many times. You must clean the data before it gets into the DB
* Do the Extract Transform Load (ETL)
* Schema: Integer, String, string, Integer, Integer



**7. Display the full path of the source of data**

* Type command as ‘$pwd’
* Here u get the full path of the source of data

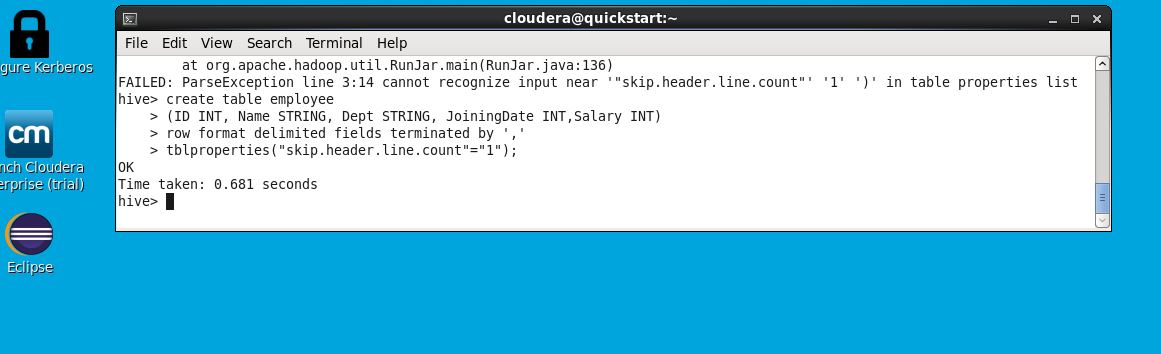


**8. Create a table named employee**

* Write the commands as

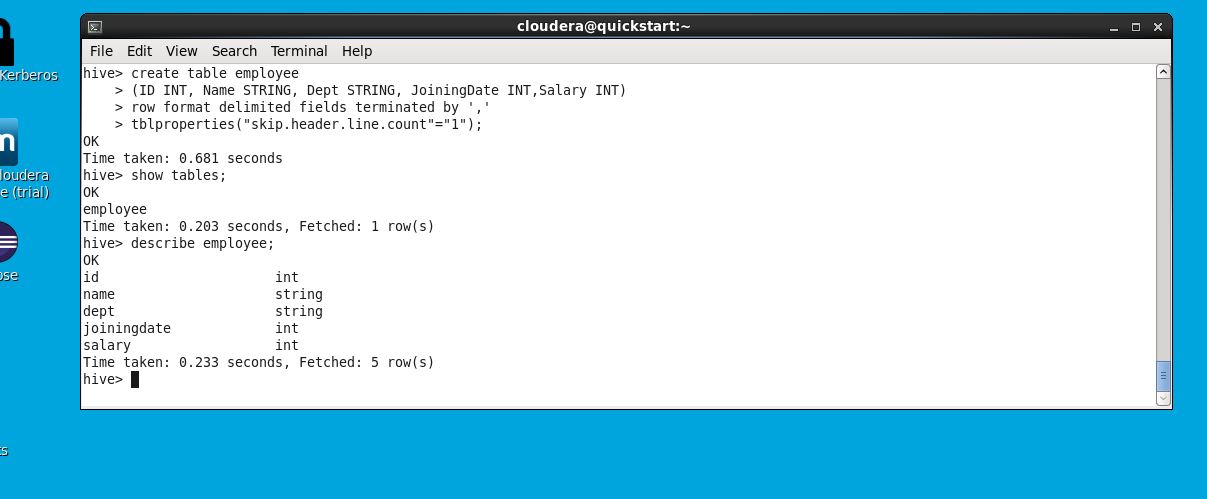
1. Create table employee
2. (ID INT, Name STRING, JoiningDate INT, Salary INT)
3. row format delimited fields terminated by ‘,’
4. tblproperties(“skip.header.line.count”=”1”);

* Then the table is created



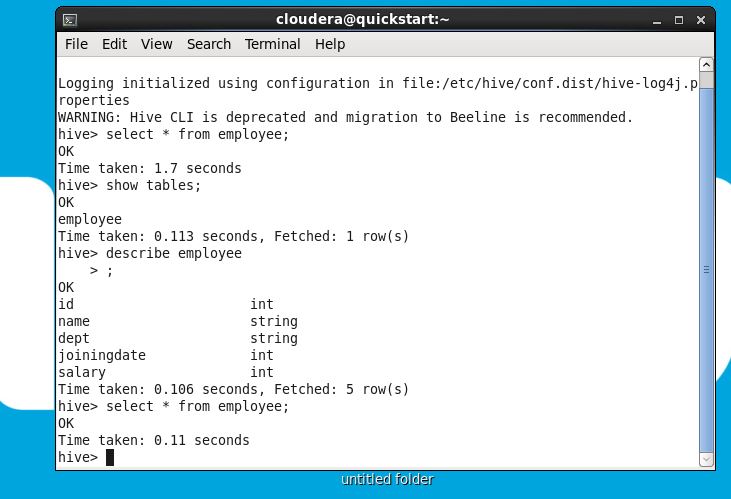
**9. Verify the table employee**

* For verifying the tables type command as ‘show tables ‘.
* For showing the data from table employee – ‘describe employee’.



**10. Display the data in the table**

* For displaying all data from table type command as ‘select \* from employee;’.

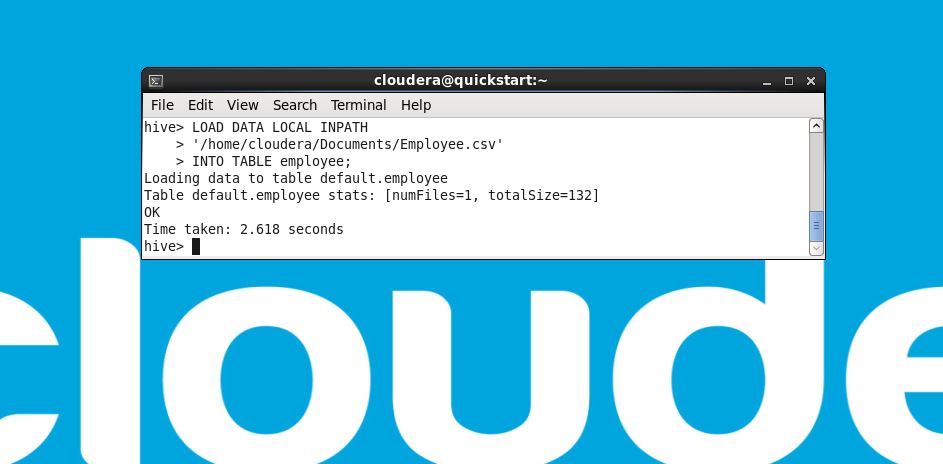
****

**11. Loading the data to Hive**

* Type command as ‘LOAD DATA LOACAL INPATH

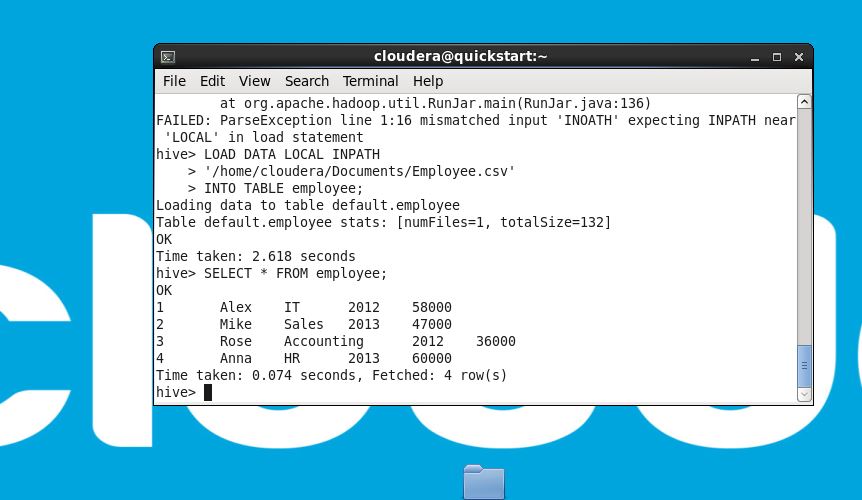
‘/HOME/CLODERA/Documents/Employee.csv’

INTO TABLE employee;’



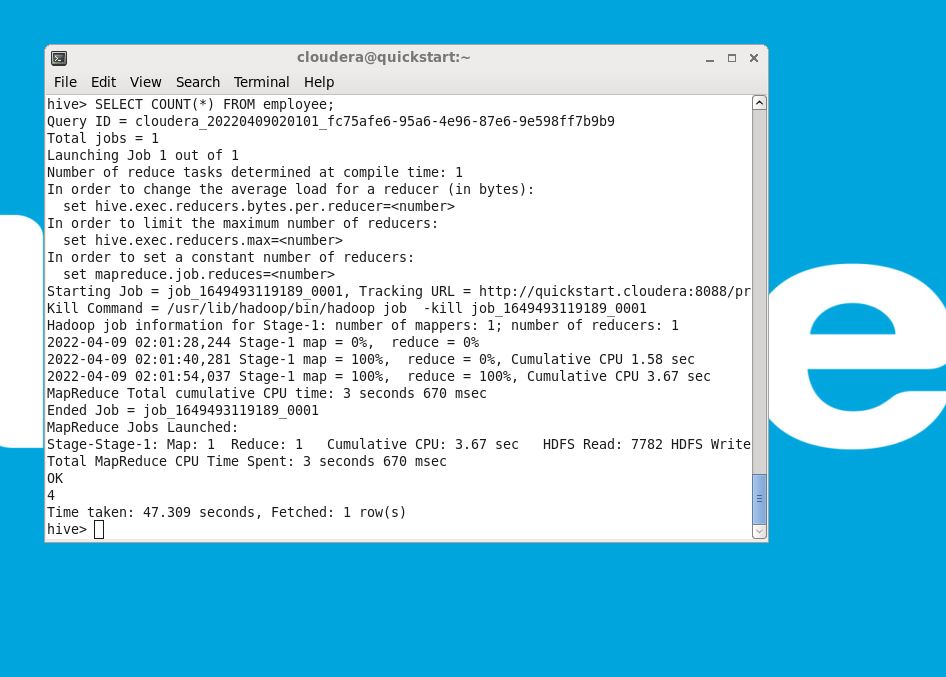
**12. Display the data**

* Select \* from employee;



**13. Try the full Map Reduce Phase by the Count(\*)**

Select count(\*) from employee;

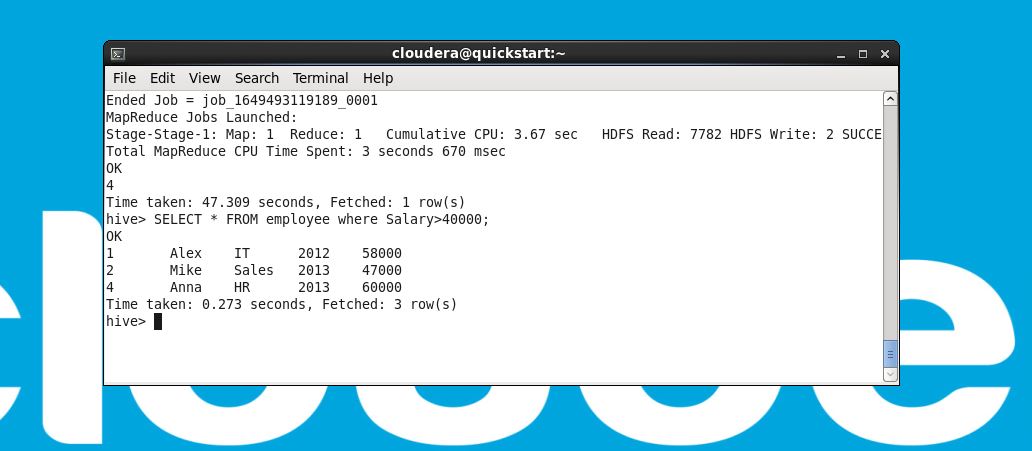


**14. More Enhanced query**

select \*

from employee

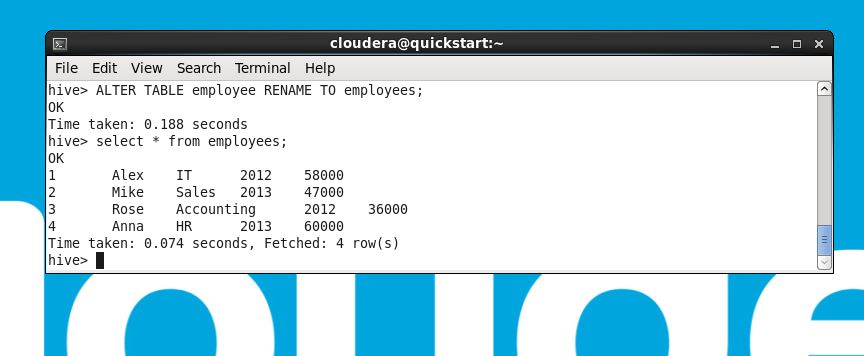
where salary>40000;



**15. Renaming the table**

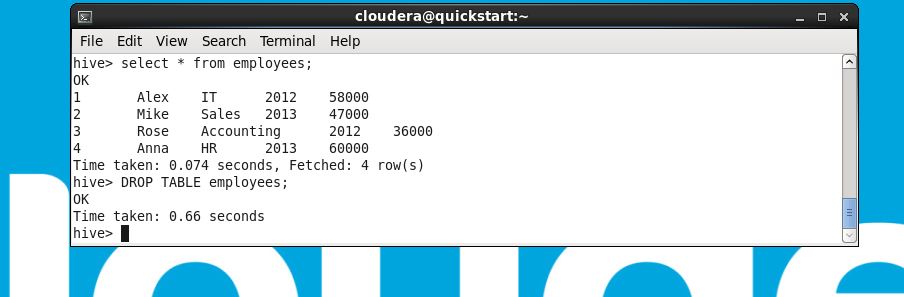
ALTER TABLE employee

RENAME TO employees;



**16. Drop the table employees**

Drop table employees;



**17. Create the table order**

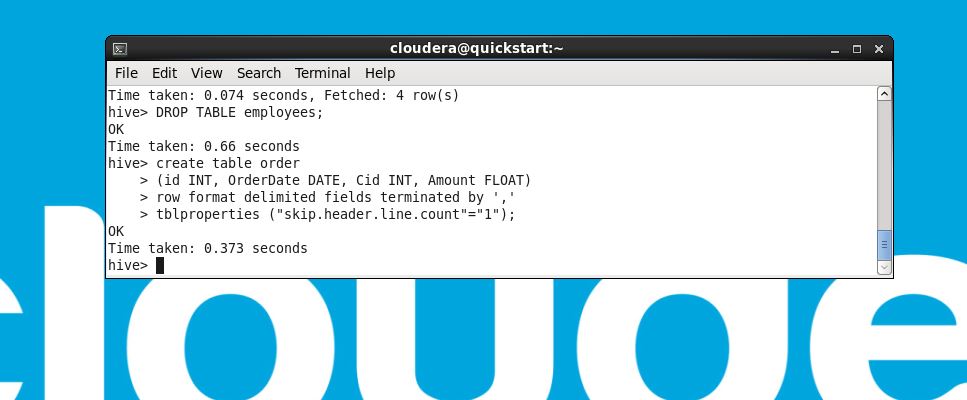
* We have to create the table called order.

Create table order

(id INT, orderdateDate DATE, Cid INT,Amount Float)

Row format delimited fields terminated by ‘,’

Tbl properties(“skip.header.line.count”=”1”);



To load data in the table we have to use command

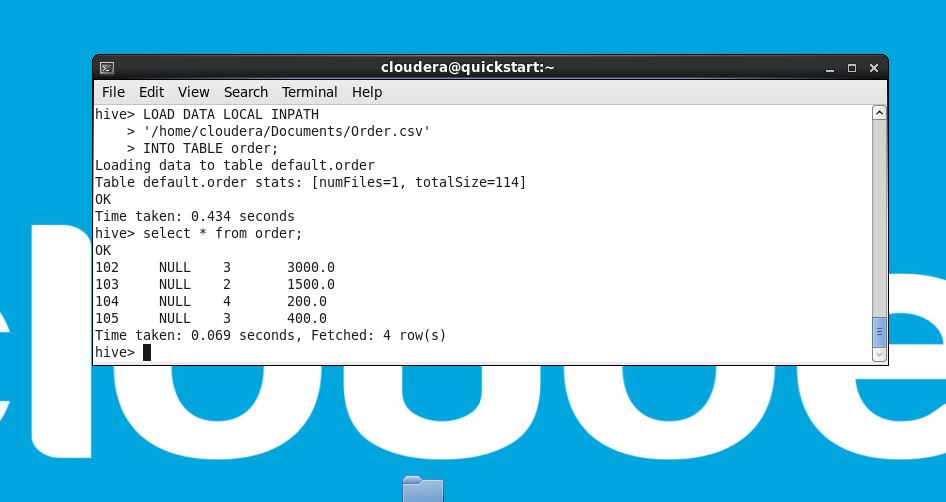
LOAD DATA LOCAL INPATH

‘/home/cloudera/Documents/Order.csv’

INTO TABLE order;

Then data is loaded in file and the for showing the data we have to use

Select \* from order;



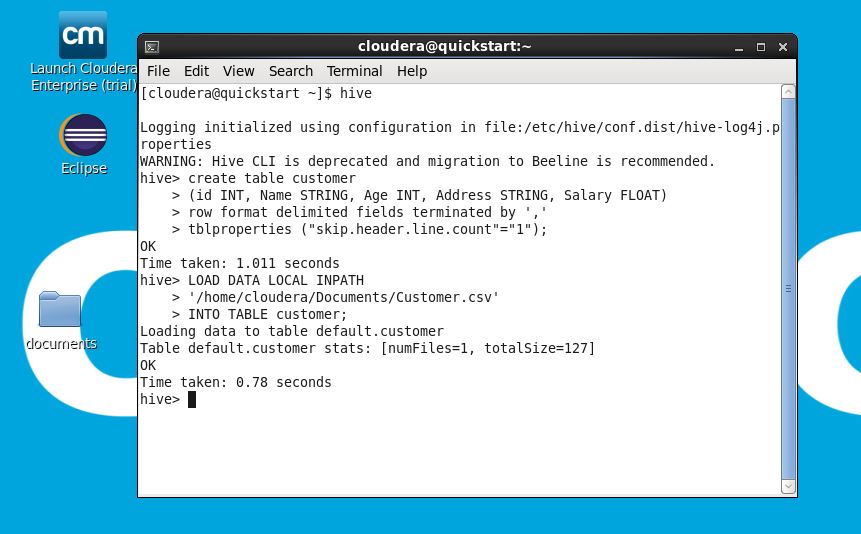
**18. Same for table Customer**

Create table order

(id INT, orderdateDate DATE, Cid INT,Amount Float)

Row format delimited fields terminated by ‘,’

Tbl properties(“skip.header.line.count”=”1”);

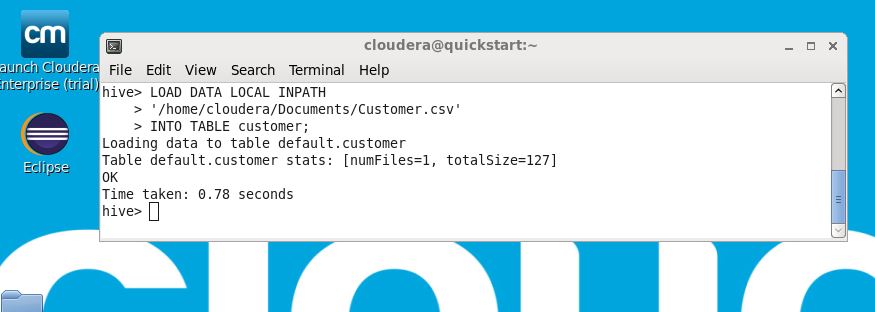


For loading data into the table from file path:

LOAD DATA LOCAL INPATH

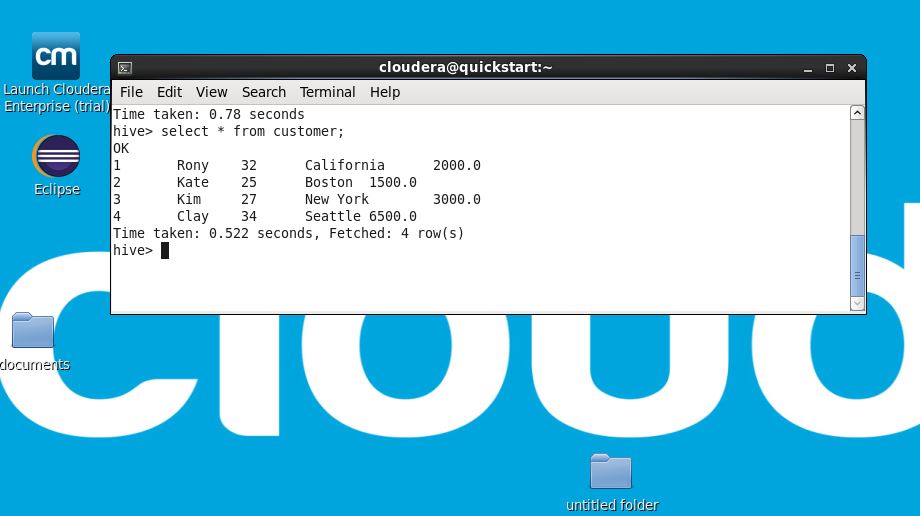
‘/home/clodera/Documents/Customer.csv’

INTO TABLE customer;



For displaying data, we have to use:

Select \* from customer;



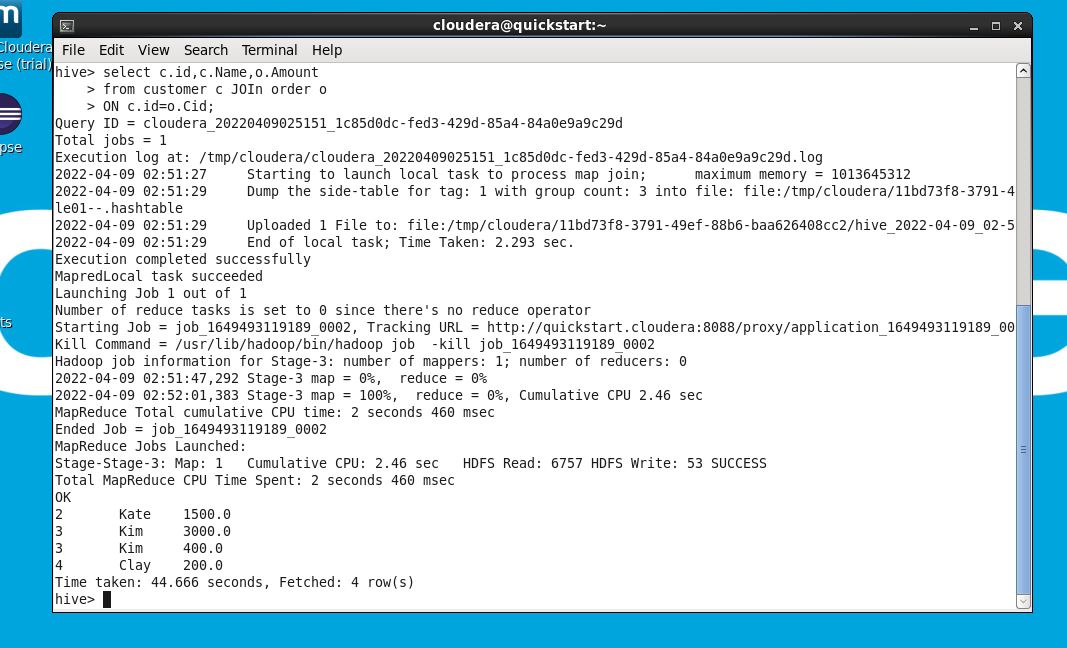
**19. A Complex Query**

**Joining table query:**

Select c.id, c.Name, o.Amount

From customer c JOIN order o

ON c.id = o.Cid;



**20. Drop DB**

For dropping table we have to use command as:

Drop database students cascade;

