**Practical 6**

**Aim: Using Hive Tool**

**What is Apache Hive?**

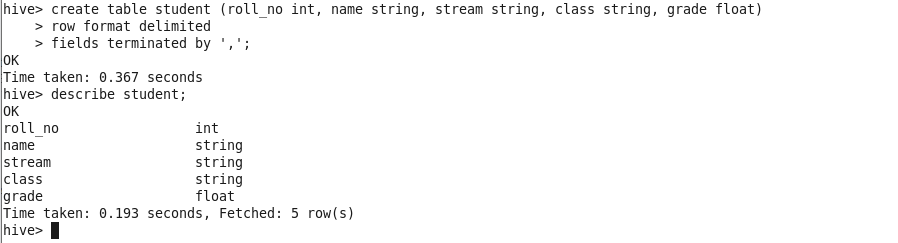
Apache Hive is a data warehouse software project built on top of Apache Hadoop for providing data query and analysis. Hive gives an SQL-like interface to query data stored in various databases and file systems that integrate with Hadoop.

Example 1:

1. Create a data warehouse database named 'RJCollege' using Hive.
2. Check the creation of a data warehouse database.



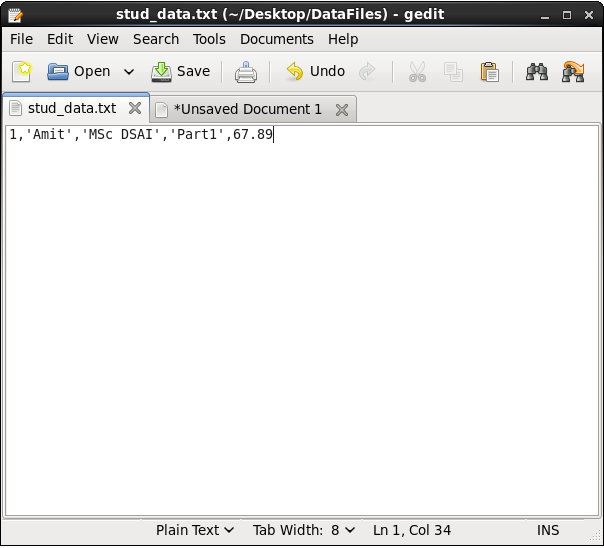
1. Create a table named 'student' in the RJCollege warehouse.



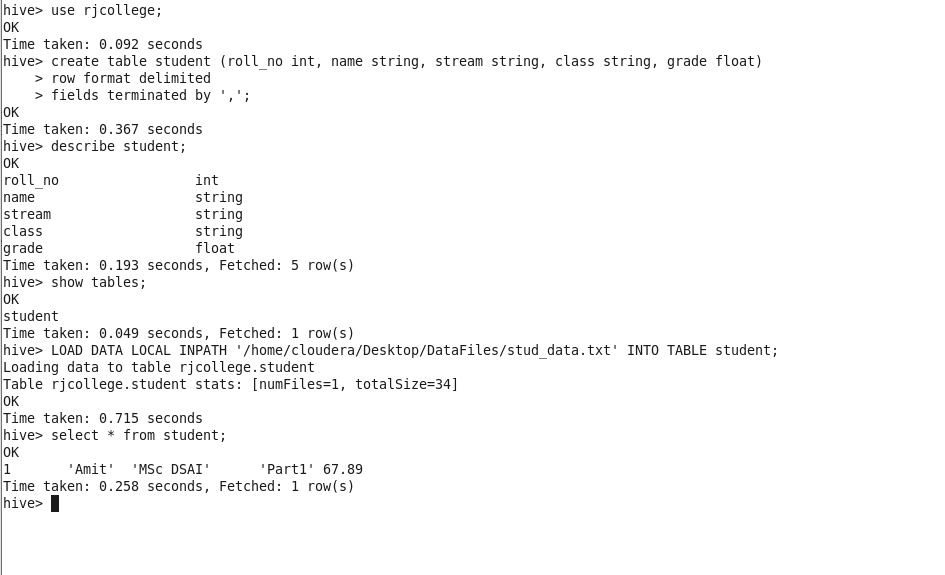
1. Create the data file named 'studData.txt' with students data and enter any 5 students data and copy it to HDFS.

Example Data:

1,'Amit','MSc DSAI','Part1',67.89



1. Display the schema of the student table.
2. Display the list of all tables or confirm the creation of a student table.
3. Load data of studData.txt into the Hive table.
4. Display all students information/results.



Example 2:

1. Create the csv file to store the data of the LIC Insurance policy customers and enter some data.

CustDetails

CustId,CustName,DOB,Gendar,CustAddress,Contact,CustmialId

PolicySaleDetails

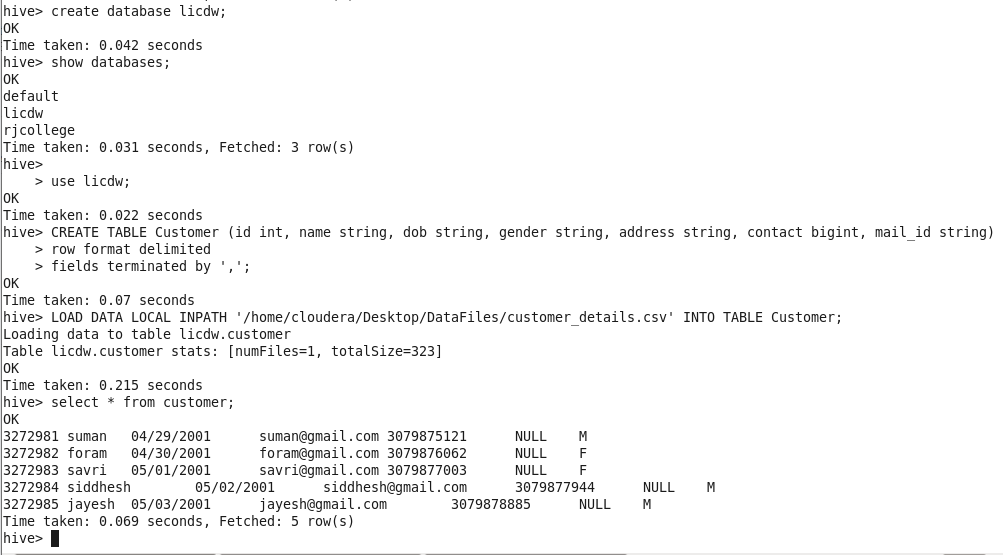
CustID,PolicyID,PurchaseDate,PremiumType,PremiumPaid

PolicyDetails

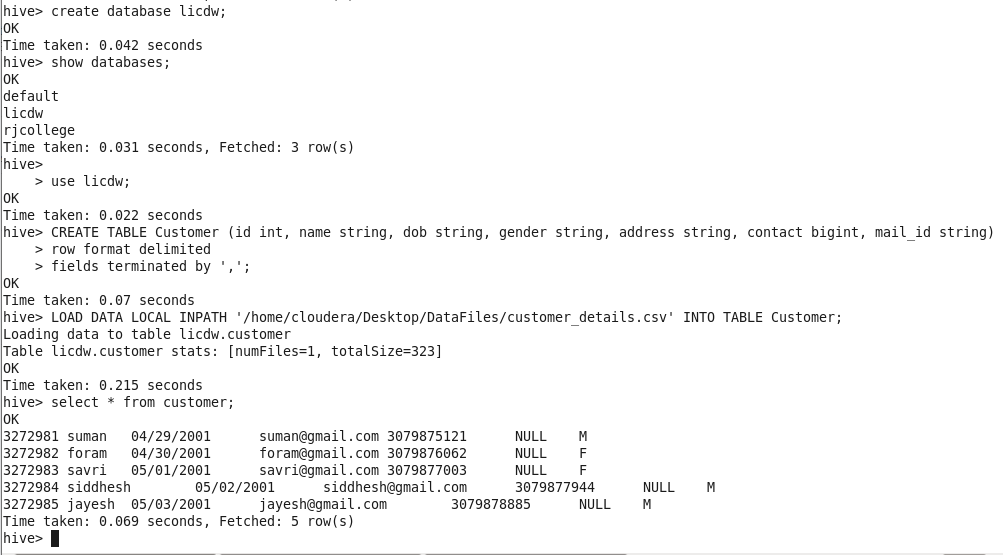
PolicyId,PolicyName,PolicyType,StartingAgeCriteria,Tenure,MaturityBenefits

1. Create a data warehouse database named ‘LICDW’ using Hive.

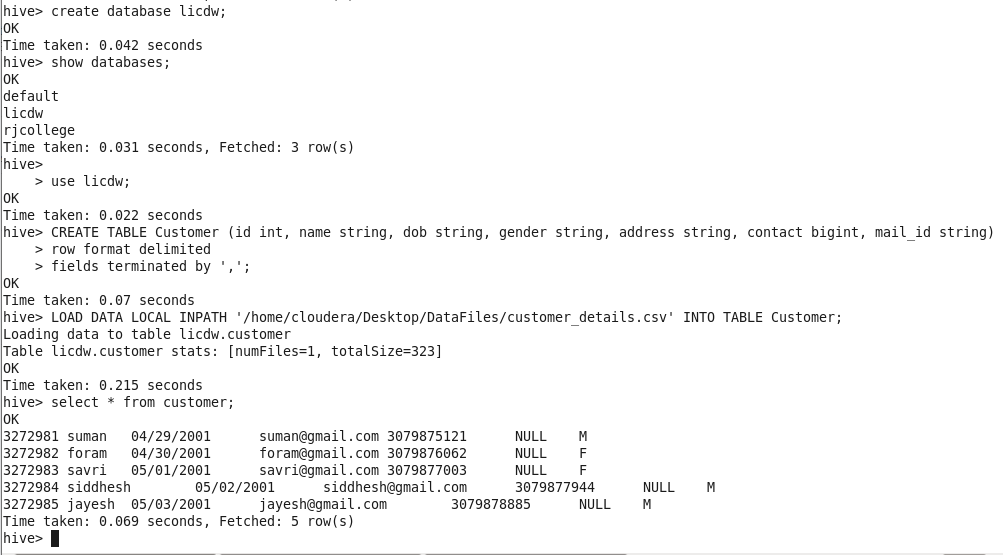
hive>create database licdw;



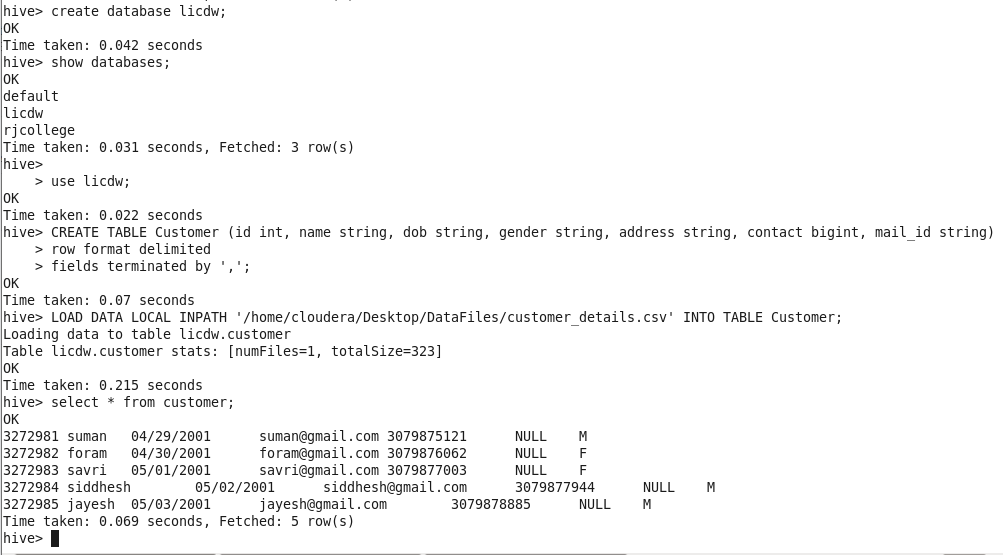
1. Create an Internal/managed table for CustDetails using Hive.



1. Load the data of custDetailsData.csv into the CustDetails table.



1. Display all records of the customer table.



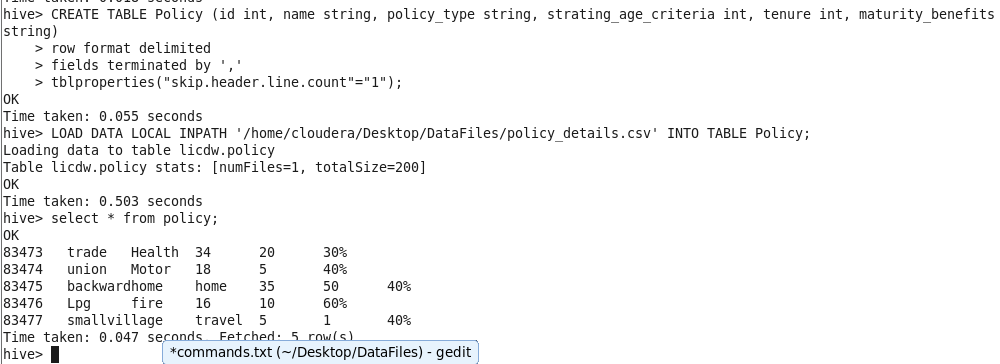
1. Create a policySaleDetails table as an external table using Hive.

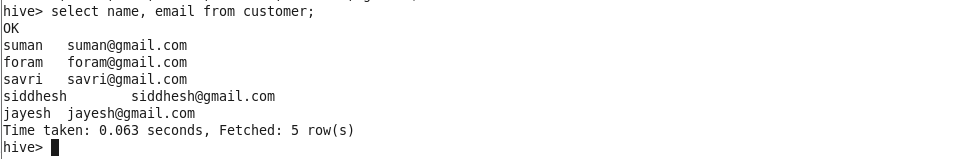
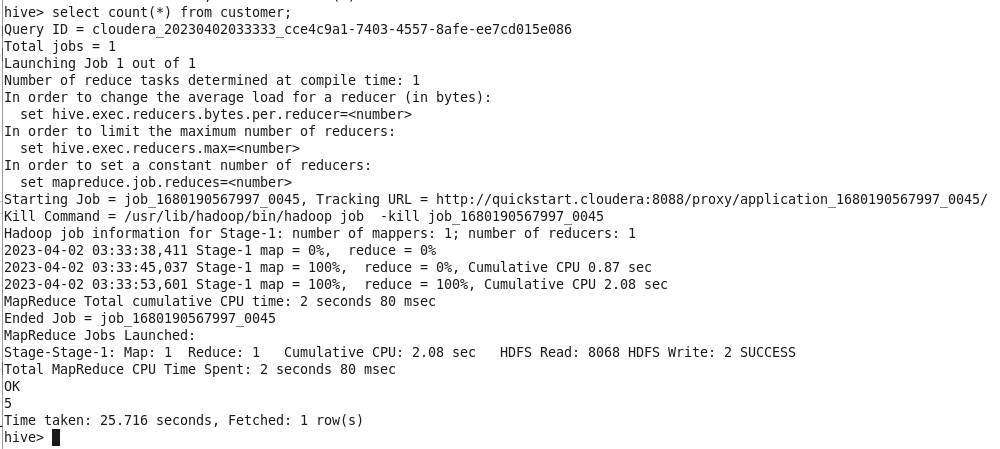
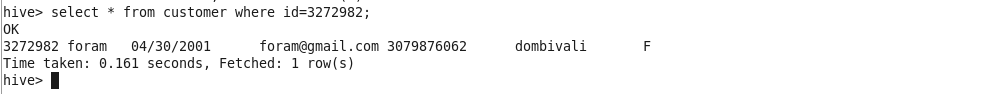
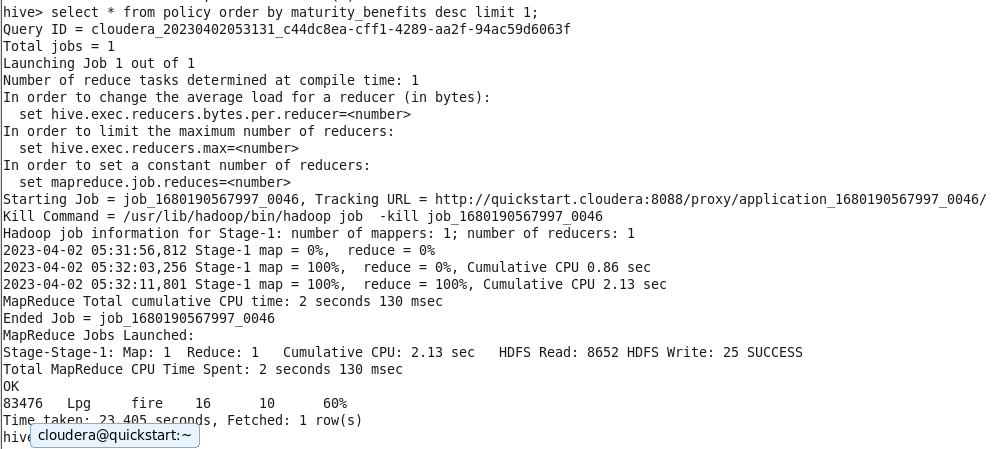
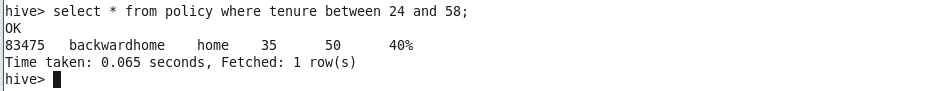
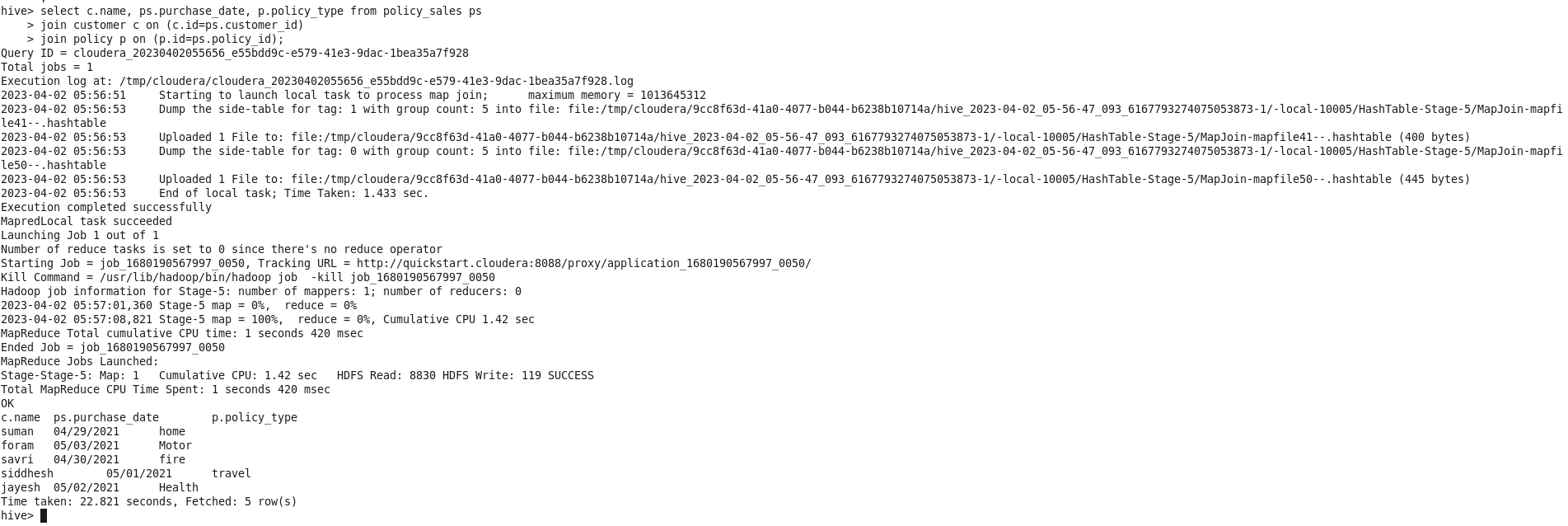


1. Load the data of PolicySaleDetailsData.csv file to PolicySaleDetails table.
2. Display the schema and data details of the PolicySaleDetails table.



1. Loading data to table licdw.policy\_detail
2. Skip header line of dataset file while loading data in Hive table.



1. Insert some records in both internal and external Hive tables using the Insert command.
2. Create the PolicyDetails temporary table.
3. Insert 2 records in the PolicyDetails table using insert command.
4. Check the existing hive version, if it is 4.x then try transaction table creation.
5. Create PolicyDetailsDup table using PolicyDetails table, using CTAS statement.
6. Execute describe and select statements for PolicyDetailsDup table.
7. Create a PolicyDetailsLike table using the existing PolicyDetails table.
8. Execute describe and select statements for PolicyDetailsLike table.
9. Display the list of customers and their mail ids from the custDetails table.
10. Get the count of the total number of customers.
11. Display the premium paid details of customer having id 3272982.
12. Display the policy with maximum benefit 
13. Display the details of policy having tuners in the range of 24 to 58 months.
14. Get each customer's name, policy purchased and its type. 
15. Display all policy types using distinct clause.
16. Display the policy wise list of customers using group by clause.