

## Experiment No. 04

**Aim:** To create Hive Database and descriptive analytics basic statistics.

**Theory:**

### HIVE

Hadoop is an open-source framework to store and process Big Data in a distributed environment. It contains two modules, one is MapReduce and another is Hadoop Distributed File System (HDFS). The Hadoop ecosystem contains different sub-projects (tools) such as Sqoop, Pig, and Hive that are used to help Hadoop modules.

Hive is a data warehouse infrastructure tool to process structured data in Hadoop. It resides on top of Hadoop to summarize Big Data and makes querying and analyzing easy.

Initially, Hive was developed by Facebook, later the Apache Software Foundation took it up and developed it further as an open source under the name Apache Hive.

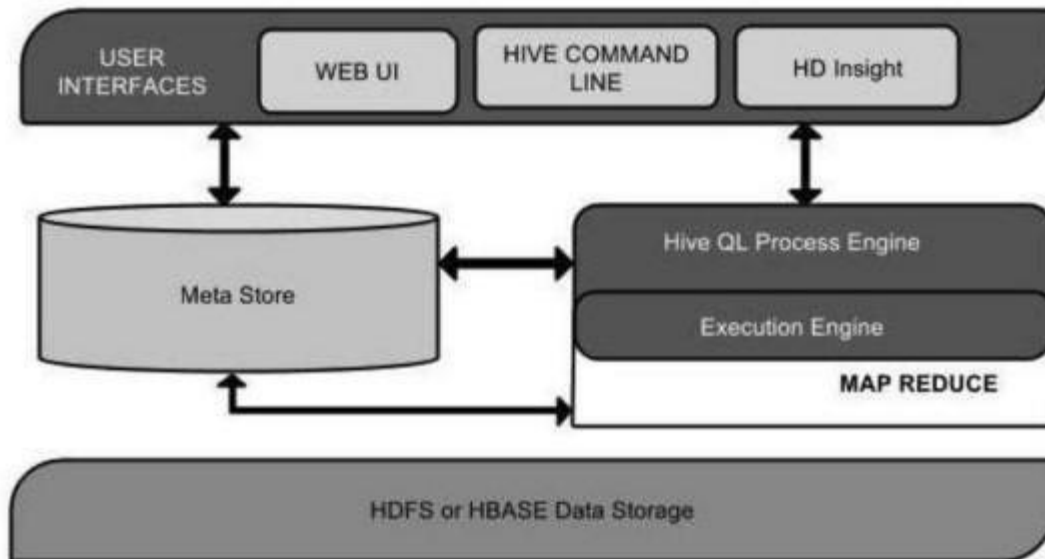
It is used by different companies. For example, Amazon uses it in Amazon Elastic MapReduce. Hive provides the functionality of reading, writing, and managing large datasets residing in distributed storage. It runs SQL-like queries called HQL (Hive query language) which get internally converted to MapReduce jobs.

Using Hive, we can skip the requirement of the traditional approach of writing complex MapReduce programs. Hive supports Data Definition Language (DDL), Data Manipulation Language (DML), and User Defined Functions (UDF).

### Features of Hive

- ➔ It stores schema in a database and processes data into HDFS
- ➔ It is designed for OLAP.
- ➔ It provides SQL-type language for querying called HiveQL or HQL.
- ➔ It is familiar, fast, scalable, and extensible

## Architecture of Hive



Limitations of Hive commands:

- 1.Hive doesn't support sub queries.
- 2.Hive surely supports over-writing, but unfortunately, it doesn't support deletion and updates.
- 3.Hive is not designed for OLTP, but it is used for it.

### Basic Hive Commands

- 1.Create: This will create the new database in the Hive.

**create database database\_name;**

- 2.Show: show command will show all the databases residing in the Hive.

**show databases;**

- 3.Use: The command to use the database.

**use database\_name;**

- 4.Drop: The drop will remove a table from Hive

**Drop table\_name;**

- 5.Create table: This command creates the table.

**create table table\_name (column names and types);**

- 6.Alter: Alter command will help you rename the table or table columns.

**alter table table\_name rename to new\_table\_name;**

For renaming column name, replace table name with column name.

7. Describe: Describe command will help you with the information about the schema of the table.

**describe table name;**

8.LOAD, INSERT: The Load operation is used to move the data into corresponding Hive table.

**LOAD data inpath into table [tablename];**

Sr. No.	Description	Commands
1	Start HIVE	\$ hive
	o/p	Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties  WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
2	Create database	hive> create database retail;
	o/p	OK  Time taken: 0.709 seconds
3	Show databases	hive> show databases
	o/p	OK  Time taken: 0.284 seconds, Fetched: 2 row(s)
4	Use database	hive> use retail;
	o/p	OK  Time taken: 0.087 seconds
5	Create table	hive> create table txnrecords(txnno INT, txndate STRING, custno INT,amount DOUBLE,category STRING, product STRING, state STRING,spendby STRING)row format  delimited fields terminated by ',' stored as textfile;



	o/p	OK	name
		NULL	
		1	XYZABC
		2	LMN
		Time taken: 0.516 seconds, Fetched: 4 row(s)	

## Output:

```

Applications  Places  System
Browse and run installed applications cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.p
roperties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> create database job;
OK
Time taken: 2.001 seconds
hive> show databases
> ;
OK
default
job
Time taken: 0.296 seconds, Fetched: 2 row(s)
hive> use job;
OK
Time taken: 0.083 seconds
hive> create table employee(id INT, name STRING, salary DOUBLE)row format delim
ited fields terminated by ',' stored as textfile;
OK
Time taken: 0.34 seconds
hive> describe employee;
OK
id                int
name              string
salary            double
Time taken: 0.188 seconds, Fetched: 3 row(s)
hive> select * from employee;
OK
Time taken: 0.514 seconds
hive> load data local inpath '/home/cloudera/Desktop/file.csv' overwrite into table employee;
Loading data to table job.employee
Table job.employee stats: [numFiles=1, numRows=0, totalSize=94, rawDataSize=0]
[Cloudera Live : Welco... cloudera@quickstart:~

```

```
Applications Places System cloudera@quickstart:~
Change desktop appearance and behavior, get help, or log out
File Edit View Search Terminal Help
Table job.employee stats: [numFiles=1, numRows=0, totalSize=94, rawDataSize=0]
OK
Time taken: 0.868 seconds
hive> select * from employee;
OK
115      'Shraddha'      40000.0
101      'Antima'       450000.0
106      'Kajal'        45000.0
108      'Neha'         50000.0
117      'Omkar'        40000.0
Time taken: 0.089 seconds, Fetched: 5 row(s)
hive> alter employee rename to empinfo;
NoViableAltException(31@[])
    at org.apache.hadoop.hive.ql.parse.HiveParser.alterStatement(HiveParser.java:7153)
    at org.apache.hadoop.hive.ql.parse.HiveParser.ddlStatement(HiveParser.java:2602)
    at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.java:1589)
    at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.java:1065)
    at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:201)
    at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:166)
    at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:522)
    at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1356)
    at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1473)
    at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1285)
    at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1275)
    at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:220)
    at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:172)
    at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:383)
    at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:775)
    at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:693)
    at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:628)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
[Cloudera Live : Welco... cloudera@quickstart:~
Applications Places System Wed Sep 14,
Access documents, folders and network places
File Edit View Search Terminal Help
tement | truncateTableStatement | alterStatement | descStatement | showStatement | metastoreCheck | createViewSta
ement | createMacroStatement | createIndexStatement | dropIndexStatement | dropFunctionStatement | reloadFunction
nt | lockStatement | unlockStatement | lockDatabase | unlockDatabase | createRoleStatement | dropRoleStatement |
| showRoleGrants | showRolePrincipals | showRoles | grantRole | revokeRole | setRole | showCurrentRole );})
    at org.apache.hadoop.hive.ql.parse.HiveParser.ddlStatement(HiveParser.java:2503)
    at org.apache.hadoop.hive.ql.parse.HiveParser.execStatement(HiveParser.java:1589)
    at org.apache.hadoop.hive.ql.parse.HiveParser.statement(HiveParser.java:1065)
    at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:201)
    at org.apache.hadoop.hive.ql.parse.ParseDriver.parse(ParseDriver.java:166)
    at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:522)
    at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1356)
    at org.apache.hadoop.hive.ql.Driver.runInternal(Driver.java:1473)
    at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1285)
    at org.apache.hadoop.hive.ql.Driver.run(Driver.java:1275)
    at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:220)
    at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:172)
    at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:383)
    at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:775)
    at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:693)
    at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:628)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:606)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:221)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
FAILED: ParseException line 1:5 cannot recognize input near 'Drop' 'empinfo' '<EOF>' in ddl statement
hive> Drop table empinfo;
OK
Time taken: 0.611 seconds
hive> █
[Cloudera Live : Welco... cloudera@quickstart:~
```

## Conclusion:

Thus, the hive database was created and basic descriptive statistical analysis was performed on it.