**Data Analytics Laboratory**

**Task 12**

**Functions and Indexes –Hive**

## Aim

To create functions, indexes using HIVE.

## Introduction to Hive

* Views are generated based on user requirements which save any result set data as a view.
* The usage of view in Hive is same as that of the view in SQL. It is a standard RDBMS concept.
* All DML operations can be executed on a view.
* The built-in functions available in Hive.

|  |  |  |
| --- | --- | --- |
| Return Type | Signature | Description |
| BIGINT | round(double a) | It returns the rounded BIGINT value of the double. |
| BIGINT | floor(double a) | It returns the maximum BIGINT value that is equal or less than the double. |
| BIGINT | ceil(double a) | It returns the minimum BIGINT value that is equal or greater than the double. |
| double | rand(), rand(int seed) | It returns a random number that changes from row to row. |
| string | concat(string A, string B,...) | It returns the string resulting from concatenating B after A. |
| string | substr(string A, int start) | It returns the substring of A starting from start position till the end of string A. |
| string | trim(string A) | It returns the string resulting from trimming spaces from both ends of A. |
| string | regexp\_replace(string A, string B, string C) | It returns the string resulting from replacing all substrings in B that match the Java regular expression syntax with C. |
| int | size(Map<K.V>) | It returns the number of elements in the map type. |
| int | size(Array<T>) | It returns the number of elements in the array type. |
| value of <type> | cast(<expr> as <type>) | It converts the results of the expression expr to <type> e.g. cast('1' as BIGINT) converts the string '1' to it integral representation. A NULL is returned if the conversion does not succeed. |
| string | from\_unixtime(int unixtime) | convert the number of seconds from Unix epoch (1970-01-01 00:00:00 UTC) to a string representing the timestamp of that moment in the current system time zone in the format of "1970-01-01 00:00:00" |
| string | to\_date(string timestamp) | It returns the date part of a timestamp string: to\_date("1970-01-01 00:00:00") = "1970-01-01" |
| int | year(string date) | It returns the year part of a date or a timestamp string: |
| int | month(string date) | It returns the month part of a date or a timestamp string. |
| int | day(string date) | It returns the day part of a date or a timestamp string: |
| string | get\_json\_object(string json\_string, string path) | It extracts json object from a json string based on json path specified, and returns json string of the extracted json object. It returns NULL if the input json string is invalid. |

## In-Lab Tasks

**LOAD DATA:**

hive**>** LOAD DATA LOCAL INPATH './usr/Desktop/kv1.txt' OVERWRITE INTO TABLE Employee;

**PROGRAM:**

**[cloudera@localhost ~]$ gedit a.txt**

**[cloudera@localhost ~]$ hdfs dfs -mkdir /tmp/hiv1**

**[cloudera@localhost ~]$ hdfs dfs -put a.txt /tmp/hiv1**

**[cloudera@localhost ~]$ hive**

Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-common-0.10.0-cdh4.7.0.jar!/hive-log4j.properties

Hive history file=/tmp/cloudera/hive\_job\_log\_f21d992b-046f-40e9-acc7-7919479a61b8\_1904851889.txt

**hive> create database employeedetails4;**

OK

Time taken: 1.437 seconds

**hive> show databases;**

OK

default

emp

employee

employeedetail

employeedetails

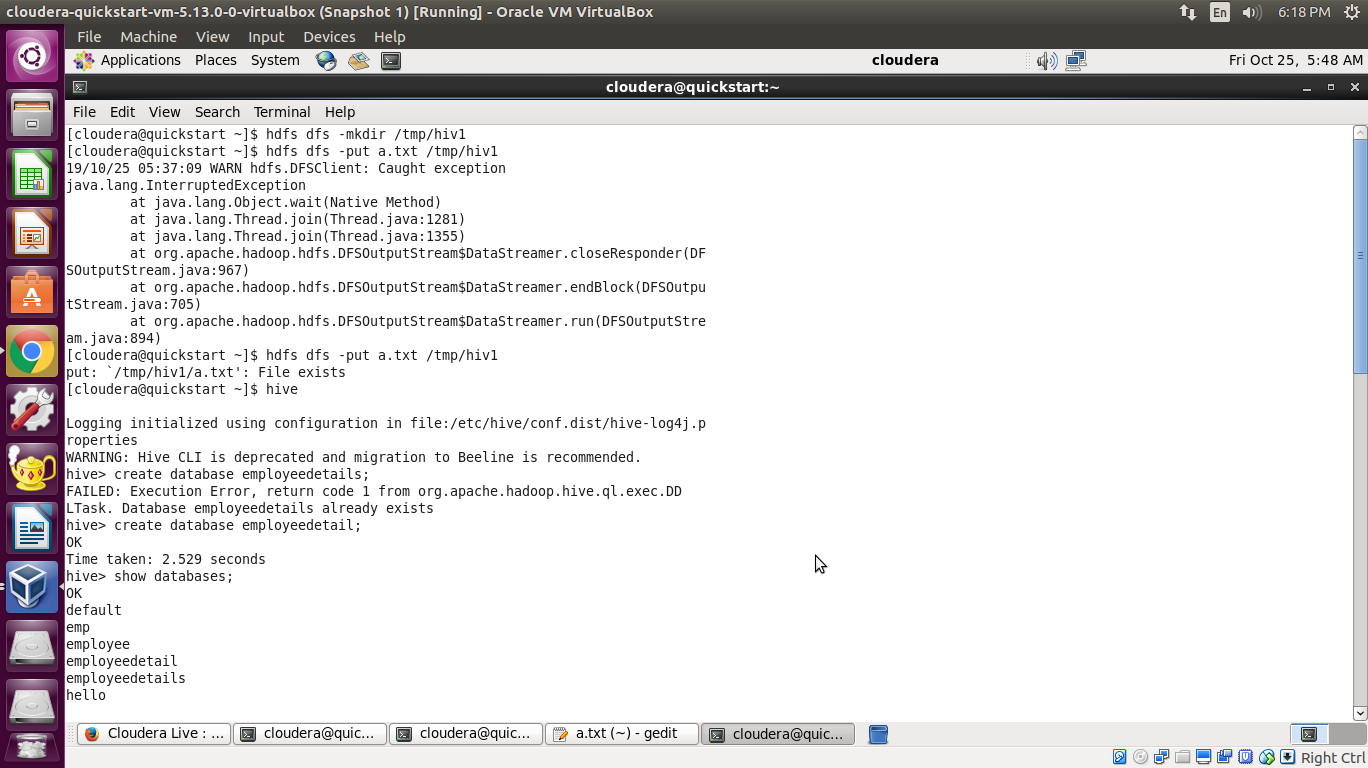
hello

Time taken: 0.543 seconds

**hive> use employeedetails4;**

OK

Time taken: 0.049 seconds



**hive> create table employee(empid int,name String,age int,dept String) row format delimited fields terminated by ‘,';**

OK

Time taken: 0.676 seconds

**hive> describe employee;**

OK

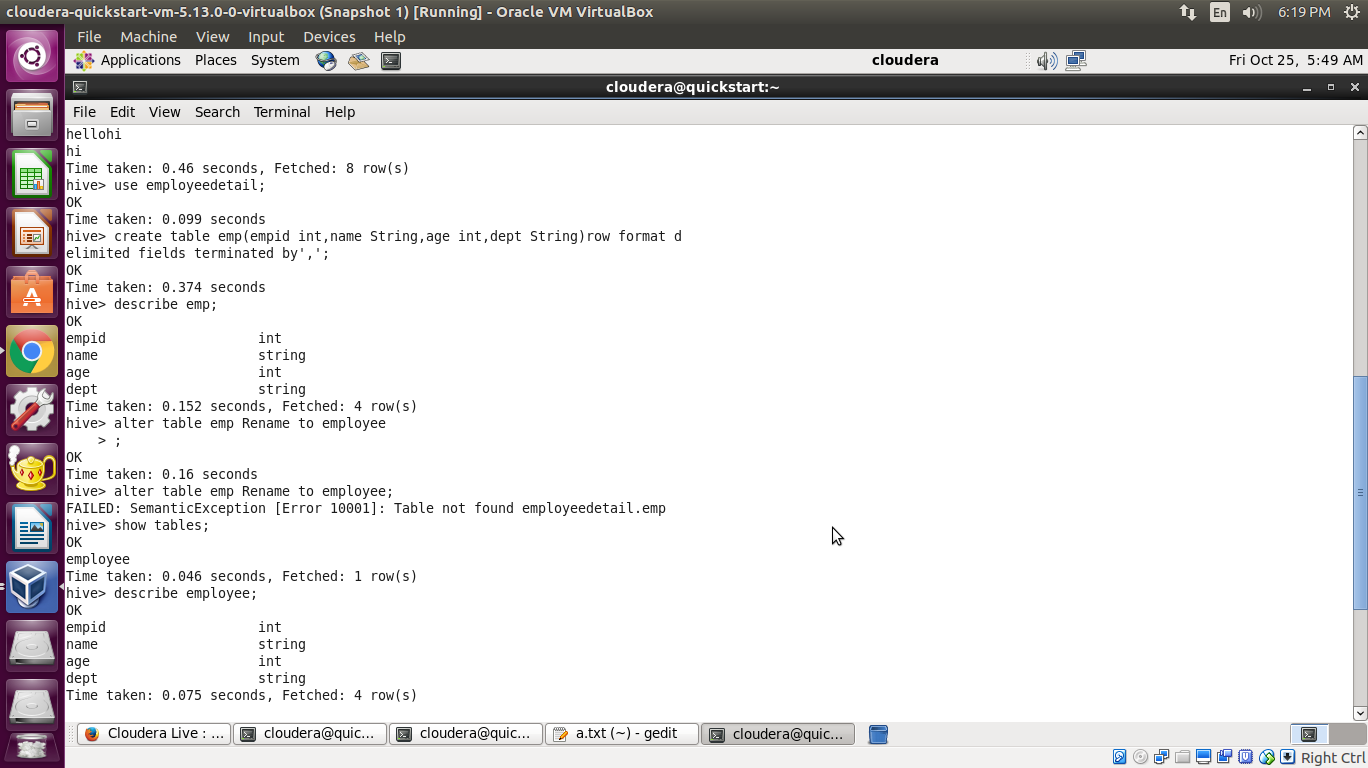
empid int

name string

age int

dept string

Time taken: 0.268 seconds



**hive> load data inpath '/tmp/hiv1/a.txt' overwrite into table employee;**

Loading data to table employeedetails4.employee

rmr: DEPRECATED: Please use 'rm -r' instead.

Moved: 'hdfs://localhost.localdomain:8020/user/hive/warehouse/employeedetails4.db/employee' to trash at: hdfs://localhost.localdomain:8020/user/cloudera/.Trash/Current

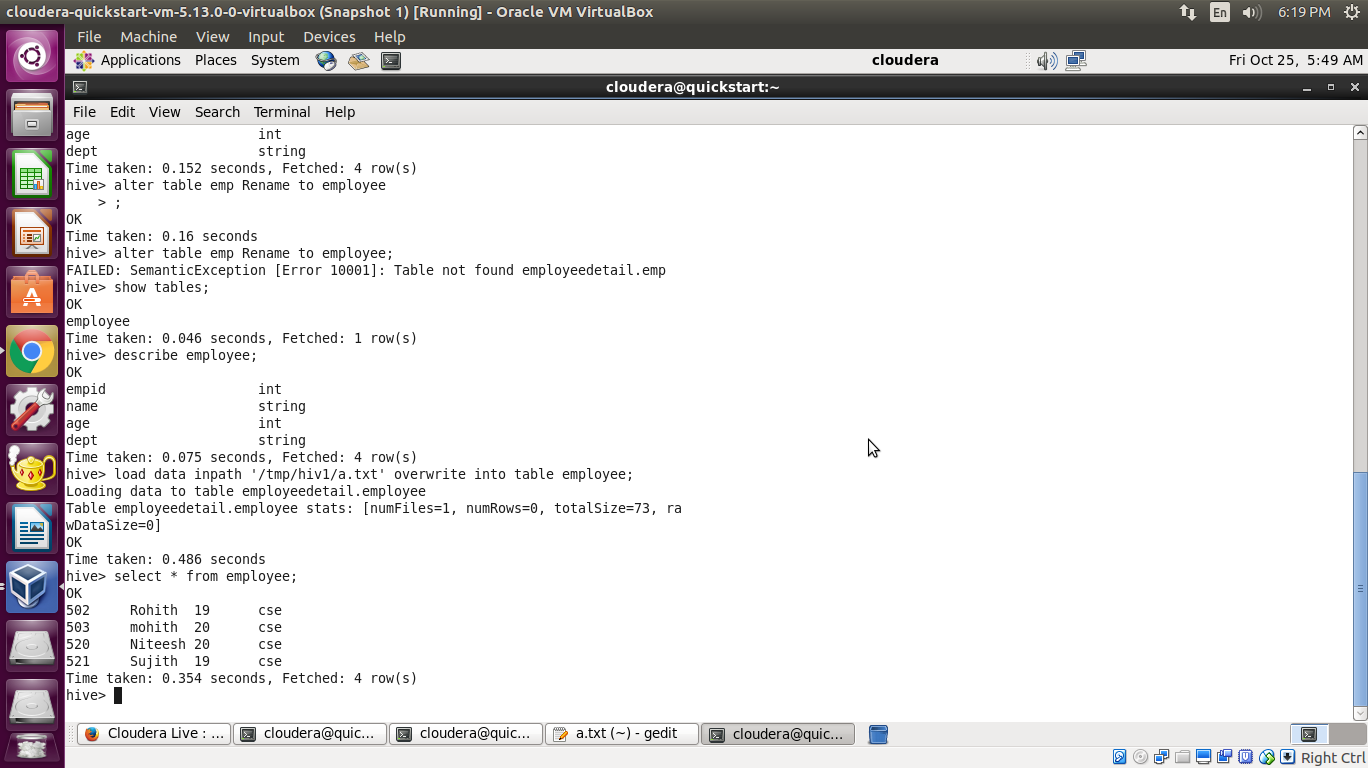
chgrp: changing ownership of '/user/hive/warehouse/employeedetails4.db/employee/a.txt': User does not belong to hive

Table employeedetails4.employee stats: [num\_partitions: 0, num\_files: 1, num\_rows: 0, total\_size: 77, raw\_data\_size: 0]

OK

Time taken: 1.227 seconds

**hive> Select \* from employee;**



**Results**

The program to use built in functions and create views is implemented.

**Faculty Signature**