Group A: Assignment No: 3

# Creating, Dropping, and altering Database tables.

student@student22:~$ jps 4724 NameNode

4852 DataNode

7017 Jps

5261 ResourceManager

5054 SecondaryNameNode

5390 NodeManager

student@student22:~$ cd /usr/local/hive/bin student@student22:/usr/local/hive/bin$ ./hive SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/home/edureka/apache-hive-2.1.0-bin/lib/log4j-slf4j-impl- 2.4.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in [jar:file:/usr/local/hadoop/hadoop- 2.7.0/share/hadoop/common/lib/slf4j-log4j12- 1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See <http://www.slf4j.org/codes.html#multiple_bindings> for an explanation. SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/home/edureka/apache-hive-2.1.0- bin/lib/hive-common-2.1.0.jar!/hive-log4j2.properties Async: true

Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases. hive> set hive.cli.print.current.db=true;

hive (default)> show databases; OK

Aatif aishwarya\_desai default

Time taken: 0.291 seconds, Fetched: 4 row(s) hive (default)> use Aatif;

OK

Time taken: 0.01 seconds hive (Aatif)> show tables; OK

Aatif

Time taken: 0.039 seconds, Fetched: 1 row(s)

hive (Aatif)> Create Database TEIT; OK

Time taken: 0.172 seconds

hive (Aatif)> use TEIT; OK

Time taken: 0.007 seconds

hive (TEIT)> Create Table Student(FirstName STRING, LastName STRING, StudentId INT);

OK

Time taken: 0.361 seconds

hive (TEIT)> ALTER DATABASE TEIT SET DBPROPERTIES('creator' = 'Ansari Aatif',

'created\_for' = 'Learning Hive'); OK

Time taken: 0.103 seconds

hive (TEIT)> DESCRIBE DATABASE EXTENDED TEIT; OK

teit

hdfs://localhost:9000/user/hive/warehouse/teit.db studentUSER {creator=Ansari Aatif, created\_for=Learning Hive}

Time taken: 0.009 seconds, Fetched: 1 row(s)

hive (TEIT)> DROP DATABASE TEIT CASCADE; OK

Time taken: 1.6 seconds

hive (TEIT)> show databases; OK

Aatif aishwarya\_desai default

Time taken: 0.005 seconds, Fetched: 4 row(s) hive (TEIT)>

# Creating an external Hive table.

hive (TEIT)> DROP DATABASE TEIT CASCADE; OK

Time taken: 1.6 seconds

hive (TEIT)> show databases; OK

Aatif aishwarya\_desai default

Time taken: 0.005 seconds, Fetched: 4 row(s)

hive (TEIT)> create database TEIT; OK

Time taken: 0.145 seconds hive (TEIT)> use TEIT; OK

Time taken: 0.006 seconds

hive (TEIT)> Create Table Student\_Data(FirstName STRING, LastName STRING, StudentId INT);

OK

Time taken: 0.211 seconds

**Download data set of 2007 & 2008 from** [**http://stat-computing.org/dataexpo/2009/the-**](http://stat-computing.org/dataexpo/2009/the-data.html)[**data.html**](http://stat-computing.org/dataexpo/2009/the-data.html)

1. **Load table with data,insert new values and fields in the table :**

hive> set hive.cli.print.current.db=true; hive (default)> show databases;

OK

Aatif aishwarya\_desai default

first flight te111

Time taken: 0.883 seconds, Fetched: 6 row(s)

hive (default)> create database TEIT; OK

Time taken: 0.259 seconds hive (default)> use TEIT; OK

Time taken: 0.006 seconds

hive (TEIT)> CREATE TABLE IF NOT EXISTS Flight(Year SMALLINT, Month

TINYINT, DayofMonth TINYINT, DayOfWeek TINYINT, DepTime SMALLINT) COMMENT 'Flight' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE TBLPROPERTIES ('creator'= 'Ansari Aatif ');

OK

Time taken: 0.495 seconds

hive (TEIT)> load data local inpath '/home/student/Desktop/2007.csv' into table Flight; Loading data to table teit.flight

OK

Time taken: 1.381 seconds

hive (TEIT)> SELECT \* FROM Flight; OK

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| NULL | NULL | NULL | NULL | NULL |
| 1987 | 10 | 14 | 3 | 741 |
| 1987 | 10 | 15 | 4 | 729 |
| 1987 | 10 | 17 | 6 | 741 |
| 1987 | 10 | 18 | 7 | 729 |
| 1987 | 10 | 19 | 1 | 749 |
| 1987 | 10 | 21 | 3 | 728 |
| 1987 | 10 | 22 | 4 | 728 |
| 1987 | 10 | 23 | 5 | 731 |

Time taken: 0.893 seconds, Fetched: 200 row(s) hive (TEIT)>

hive (TEIT)> CREATE TABLE IF NOT EXISTS Flight2009(Year SMALLINT, Month

TINYINT, DayofMonth TINYINT, DayOfWeek TINYINT, DepTime SMALLINT) COMMENT 'Flight' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE TBLPROPERTIES ('creator'= 'Ansari Aatif ');

OK

Time taken: 0.083 seconds

hive (TEIT)> load data local inpath '/home/student/Desktop/2009dataset.csv' into table Flight2009;

Loading data to table teit.flight2009 OK

Time taken: 0.583 seconds

hive (TEIT)> SELECT \* FROM Flight2009; OK

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| NULL | NULL | NULL | NULL | NULL |
| 1987 | 10 | 14 | 3 | 741 |
| 1987 | 10 | 15 | 4 | 729 |
| 1987 | 10 | 17 | 6 | 741 |
| 1987 | 10 | 18 | 7 | 729 |
| 1987 | 10 | 19 | 1 | 749 |
| 1987 | 10 | 21 | 3 | 728 |
| 1987 | 10 | 22 | 4 | 728 |
| 1987 | 10 | 23 | 5 | 731 |

Time taken: 0.104 seconds, Fetched: 200 row(s) hive (TEIT)>

# JOIN

hive (TEIT)> SELECT m8.Year, m8.Month, m7.Year, m7.Month FROM Flight2009 m8 JOIN Flight m7;

|  |  |  |  |
| --- | --- | --- | --- |
| NULL | NULL | NULL | NULL |
| 1987 | 10 | 1987 | 10 |
| 1987 | 10 | 1987 | 10 |
| 1987 | 10 | 1987 | 10 |
| 1987 | 10 | 1987 | 10 |
| 1987 | 10 | 1987 | 10 |
| 1987 | 10 | 1987 | 10 |
| 1987 | 10 | 1987 | 10 |
| 1987 | 10 | 1987 | 10 |

# Find the average departure delay per day in 2008.

**Calculate average delay**

hive> select sum(delay) from hbase\_flight\_new;

Day 5 under the results in Step (B) — had the highest number of delays.

**CREATE TABLE:**

1. ***CREATE TABLE student (Name STRING, rollNo int, Dept STRING)***

***ROW FORMAT DELIMITED***

***FIELDS TERMINATED BY ‘,’***

1. *CREATE TABLE IF NOT EXISTS Flight(Year SMALLINT, Month TINYINT, DayofMonth TINYINT, DayOfWeek TINYINT, DepTime SMALLINT) COMMENT 'Flight' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','*

**ALTER TABLE:**

ALTER TABLE name RENAME TO new\_name

ALTER TABLE name ADD COLUMNS (col\_spec[, col\_spec ...])

ALTER TABLE name DROP [COLUMN] column\_name

ALTER TABLE FLIGHT RENAME TO FLIGHT1

ALTER TABLE FLIGHT ADD COLUMNS (age int)

ALTER TABLE FLIGHT DROP age

**LOADING FILES INTO TABLE:**

***LOAD DATA LOCAL INPATH ‘path’ INTO TABLE FLIGHT***

**VIEWING TABLE DATA:**

***SELECT \* FROM FLIGHT***

**CREATING INDEX:**

CREATE INDEX flight\_index ON TABLE FLIGHT(Year)

AS 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler';

DROP INDEX: DROP flight\_index on FLIGHT

**JOIN: TO HVE A JOIN CREATE FIRST FLIGHT TABLE AND NEW TABLE FLIGHT 2009**

CREATE TABLE IF NOT EXISTS Flight2009(Year SMALLINT, Month TINYINT, DayofMonth TINYINT, DayOfWeek TINYINT, DepTime SMALLINT) COMMENT 'Flight' ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

> load data local inpath '/home/student/Desktop/2009dataset.csv' into table Flight2009;

JOIN QUERY:   
SELECT m8.Year, m8.Month, m7.Year, m7.Month FROM Flight2009 m8 JOIN Flight m7;