BIG DATA
DEVELOPMENT

ACADGILD

Session 07: Hive Operations

Assignment 3 Question

Explain with an example in brief.

1) Hive Data Definitions

Solution:

- It includes commands that define the different structures in a database.
- HDL is a subset of Hive SQL statements that describe the data structure in Hive by creating, deleting, or altering schema objects such as databases, tables, views.
- The commands are:

CREATE, DROP, TRUNCATE, DESCRIBE, AND ALTER.

Syntax:

CREATE TABLE HIVE_TABLE (abc INT, xyz STRING); // creating tables

DROP 'TABLE NAME'; // deleting the table or database

TRUNCATE – used to delete the structure of the table.

ALTER - used to change the structure of the table.

2) Hive Data Manipulations

Solution:

- It includes commands used to modify the values in the table or to extract the data from the table.
- The commands are:
- LOAD, INSERT, UPDATE, DELETE
- LOAD:- load the data into a table.

Syntax :- `LOAD DATA [LOCAL] INPATH 'filepath' [OVERWRITE] INTO TABLE tablename [PARTITION (partcol1=val1, partcol2=val2 ...)]`

- Insert: -to insert value in table

Syntax:-
```
INSERT OVERWRITE TABLE tablename1 [PARTITION
(partcol1=val1, partcol2=val2 ...) [IF NOT EXISTS]]
select_statement1 FROM from_statement;
```

```
INSERT INTO TABLE tablename1 [PARTITION (partcol1=val1,
partcol2=val2 ...)] select_statement1 FROM from_statement;
```

- Update: - change the entries in the table.

```
Standard Syntax:
```

```
UPDATE tablename SET column = value [, column = value ...]
[WHERE expression]
```

3) HiveQL Manipulations

Solution:

HiveQL Manipulations are of three types:

a) Select where

b) Select order by /sort by

c) Select group by

- Select where :- used to filter the dataset with the where clause and apply the conditions.
- Select order by clause :- used to arrange the dataset in the ascending or descending order by one field or multiple fields.
- Select group by :- used to form subsets of the database using different fields of the data.

Syntax:-

```
    [WITH CommonTableExpression (,
CommonTableExpression)*]    (Note: Only available starting with
Hive 0.13.0)

SELECT [ALL | DISTINCT] select_expr, select_expr, ...
  FROM table_reference
  [WHERE where_condition]
  [GROUP BY col_list]
  [ORDER BY col_list]
  [CLUSTER BY col_list
    | [DISTRIBUTE BY col_list] [SORT BY col_list]
  ]
 [LIMIT [offset,] rows]
```