**Assignment 7.3**

Explain the below concepts with an example in brief:

1. **Hive Data Definitions:**

Hive DDL statements includes below:

* CREATE DATABASE/SCHEMA, TABLE, VIEW, FUNCTION, INDEX
* DROP DATABASE/SCHEMA, TABLE, VIEW, INDEX
* TRUNCATE TABLE
* ALTER DATABASE/SCHEMA, TABLE, VIEW
* MSCK REPAIR TABLE (or ALTER TABLE RECOVER PARTITIONS)
* SHOW DATABASES/SCHEMAS, TABLES, TBLPROPERTIES, VIEWS, PARTITIONS, FUNCTIONS, INDEX[ES], COLUMNS, CREATE TABLE
* DESCRIBE DATABASE/SCHEMA, Table\_name, View\_name

PARTITION statements are usually options of TABLE statements, except for SHOW PARTITIONS.

1. **Hive Data Manipulations and 3. HiveQL Manipulations:**

Below are the ways to modify data in Hive:

1. [**LOAD**](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Loadingfilesintotables)

* Hive does not do any transformation while loading data into tables. Load operations are currently pure copy/move operations that move data files into locations corresponding to Hive tables.

Syntax:

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

1. **INSERT**

* Insert [into Hive tables from queries](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-InsertingdataintoHiveTablesfromqueries):

Query Results can be inserted into tables by using the insert clause.

Inserts can be done to a table or a partition. If the table is partitioned, then one must specify a specific partition of the table by specifying values for all of the partitioning columns. If [hive.typecheck.on.insert](https://cwiki.apache.org/confluence/display/Hive/Configuration+Properties#ConfigurationProperties-hive.typecheck.on.insert) is set to true, these values are validated, converted and normalized to conform to their column types

Syntax:

INSERT OVERWRITE will overwrite any existing data in the table or partition

INSERT OVERWRITE TABLE tablename1 [PARTITION (partcol1=val1, partcol2=val2 ...) [IF NOT EXISTS]] select\_statement1 FROM from\_statement;

INSERT INTO will append to the table or partition, keeping the existing data intact

(INSERT INTO syntax is only available starting in version 0.8)

INSERT INTO TABLE tablename1 [PARTITION (partcol1=val1, partcol2=val2 ...)] select\_statement1 FROM from\_statement;

* Insert [into directories from queries](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Writingdataintothefilesystemfromqueries):

Query results can be inserted into filesystem directories:

Syntax:

INSERT OVERWRITE [LOCAL] DIRECTORY directory1

  [ROW FORMAT row\_format] [STORED AS file\_format]

  SELECT ... FROM ...

Hive extension (multiple inserts):

FROM from\_statement

INSERT OVERWRITE [LOCAL] DIRECTORY directory1 select\_statement1

[INSERT OVERWRITE [LOCAL] DIRECTORY directory2 select\_statement2] ...

* Insert [into Hive tables from SQL](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-InsertingintotablesfromSQL)

The INSERT...VALUES statement can be used to insert data into tables directly from SQL.

Syntax:

INSERT INTO TABLE tablename [PARTITION (partcol1[=val1], partcol2[=val2] ...)] VALUES values\_row [, values\_row ...]

Where values\_row is:

( value [, value ...] )

where a value is either null or any valid SQL literal

1. [**UPDATE**](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Update)**:**

Updates can only be performed on tables that support ACID

Syntax:

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

* The referenced column must be a column of the table being updated.
* The value assigned must be an expression that Hive supports in the select clause.  Thus arithmetic operators, UDFs, casts, literals, etc. are supported.  Subqueries are not supported.
* Only rows that match the WHERE clause will be updated.
* Partitioning columns cannot be updated.
* Bucketing columns cannot be updated.

1. [**DELETE**](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Delete)**:**

Deletes can only be performed on tables that support ACID (Atomicity, Consistency, Isolation, and Durability)

Only rows that match the WHERE clause will be deleted.

Syntax:

DELETE FROM tablename [WHERE expression]

1. [**MERGE**](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Merge)

Merge can only be performed on tables that support ACID (Atomicity, Consistency, Isolation, and Durability)

[Merge](https://en.wikipedia.org/wiki/Merge_(SQL)) allows actions to be performed on a target table based on the results of a join with a source table.

Syntax:

MERGE INTO <target table> AS T USING <source expression/table> AS S

ON <boolean expression1>

WHEN MATCHED [AND <boolean expression2>] THEN UPDATE SET <set clause list>

WHEN MATCHED [AND <boolean expression3>] THEN DELETE

WHEN NOT MATCHED [AND <boolean expression4>] THEN INSERT VALUES<value list>