**Session 24**

**Assignment 1**

**Problem Statement:** Explain with an example in brief.

**Part 1:** Hive Data Definition Languages.

Hive QL DDL statements are documented here, including:

* 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.

**Part 2:** Hive Data Manipulation Languages.

* [LOAD](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Loadingfilesintotables)- Loads a file into a given table. 'LOCAL' signifies that the input file is on the local file system. If 'LOCAL' is omitted then it looks for the file in HDFS.
* OVERWRITE- The keyword 'OVERWRITE' signifies that existing data in the table is deleted. If the 'OVERWRITE' keyword is omitted, data files are appended to existing data sets.
* INSERT- using insert overwrite . . . select command insert those rows into a new partition of your main Hive table.
  1. [Into Hive tables from queries](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-InsertingdataintoHiveTablesfromqueries)
  2. [Into directories from queries](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Writingdataintothefilesystemfromqueries)
  3. [Into Hive tables from SQL](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-InsertingintotablesfromSQL)
* [UPDATE](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Update)- Can be used to edit the data in the table.
* [DELETE](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Delete)- Can be used to delete the table.
* [MERGE](https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML#LanguageManualDML-Merge)- Can be used to merge two tables.

**Part 3:** Hive QL Manipulations

**Transcript of HiveQL: Data Manipulation**

* Loading Data into Managed Tables.
* Inserting Data into Tables from Queries.

Dynamic Partition Inserts.

* Creating Tables and Loading Them in One Query.
* Exporting Data.

The Hive Query Language (Hive QL) is a query language for Hive to process and analyze structured data in a Metastore.

* SELECT statement is used to retrieve the data from a table.
* WHERE clause works similar to a condition. It filters the data using the condition and gives you a finite result.
* The built-in operators and functions generate an expression, which fulfils the condition