Assignment 7.3

# Hive Data Definitions

Hive Data Definition Language (DDL) 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, partitions, and buckets.

e.g.

CREATE DATABASE IF NOT EXISTS acadgild\_db;

DROP DATABASE IF EXISTS acadgild\_db CASCADE;

create table if not exists emp\_details

(

emp\_name string,

unit string,

exp int,

location string

)

row format delimited

fields terminated by ',';

# Hive Data Manipulations

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

Load operations are currently pure copy/move operations that move datafiles into locations corresponding to Hive tables

## Syntax

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

If the keyword LOCAL is specified, then the load command will look for filepath in the local file system

If the OVERWRITE keyword is used then the contents of the target table (or partition) will be deleted and replaced by the files referred to by filepath; otherwise the files referred by filepath will be added to the table.

e.g.

LOAD DATA LOCAL INPATH '/home/acadgild/hive/emp\_details.txt' OVERWRITE INTO TABLE emp\_details;

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

## 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;

### Writing data into the filesystem from queries

Query results can be inserted into filesystem directories

## Syntax

INSERT OVERWRITE [LOCAL] DIRECTORY directory1

[ROW FORMAT row\_format] [STORED AS file\_format] (Note: Only available starting with Hive 0.11.0)

SELECT ... FROM ...

# HiveQL Manipulations

Hive provides a CLI to write Hive queries using Hive Query Language (HiveQL) which focuses on the *data manipulation language* parts that are used to put data into tables and to extract data from tables to the filesystem. Generally HQL syntax is similar to the SQL syntax. It uses SELECT ... WHERE clauses extensively.

Hive supports four file formats those are TEXTFILE, SEQUENCEFILE, ORC and RCFILE (Record Columnar File).

**Built-in operators**

Hive provides Built-in operators for Data operations to be implemented on the tables present inside Hive warehouse. These operators are used for mathematical operations on operands, and it will return specific value as per the logic applied.

Types of Built-in Operators in HIVE are:

Relational Operators (=, !=, <, >, etc.)

Arithmetic Operators (+, -, \*,/, etc.)

Logical Operators (AND, &&, OR, ||, NOT)

Operators on Complex types (A[n], M[key])

Complex type Constructors (array, create\_union, etc)