1. Hive Data Definitions:

**Hive Data Definition** Language. **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.

Example:

CREATE TABLE IF NOT EXISTS mydb.employees (

name STRING COMMENT 'Employee name',

salary FLOAT COMMENT 'Employee salary',

subordinates ARRAY<STRING> COMMENT 'Names of subordinates',

deductions MAP<STRING, FLOAT>

COMMENT 'Keys are deductions names, values are percentages',

address STRUCT<street:STRING, city:STRING, state:STRING, zip:INT>

COMMENT 'Home address')

COMMENT 'Description of the table'

TBLPROPERTIES ('creator'='me', 'created\_at'='2012-01-02 10:00:00', ...)

LOCATION '/user/hive/warehouse/mydb.db/employees';

1. Hive Data Manipulations

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

* *filepath* can be:
  + a relative path, such as project/data1
  + an absolute path, such as /user/hive/project/data1
  + a full URI with scheme and (optionally) an authority, such as hdfs://namenode:9000/user/hive/project/data1
* The target being loaded to can be 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.
* *filepath* can refer to a file (in which case Hive will move the file into the table) or it can be a directory (in which case Hive will move all the files within that directory into the table). In either case, *filepath* addresses a set of files.
* If the keyword LOCAL is specified, then:
  + the load command will look for *filepath* in the local file system. If a relative path is specified, it will be interpreted relative to the user's current working directory. The user can specify a full URI for local files as well - for example: [file:///user/hive/project/data1](file:///\\user\hive\project\data1)
  + the load command will try to copy all the files addressed by *filepath* to the target filesystem. The target file system is inferred by looking at the location attribute of the table. The copied data files will then be moved to the table.
* If the keyword LOCAL is *not* specified, then Hive will either use the full URI of *filepath*, if one is specified, or will apply the following rules:
  + If scheme or authority are not specified, Hive will use the scheme and authority from the hadoop configuration variable fs.default.name that specifies the Namenode URI.
  + If the path is not absolute, then Hive will interpret it relative to /user/<username>
  + Hive will *move* the files addressed by *filepath* into the table (or partition)
* 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.

Example:

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

1. HiveQL Manipulations