1. Need of Hive :

* A data warehouse solution built on top of Hadoop - by Facebook.
* An essential tool in the Hadoop ecosystem that provides an SQL (Structured Query Language) dialect (called as Hive Query Language) for querying data stored in the Hadoop Distributed Filesystem (HDFS).
* Most data warehouse applications are implemented using relational databases that use SQL as the query language. Hive lowers the barrier for moving these applications to Hadoop. People who know SQL can learn Hive easily.
* Automatically uses HDFS for storage, but stores all the meta information about database and table in metadata DB locally to Hive.
* Hive is most suited for data warehouse applications, where relatively static data is analyzed, fast response times are not required, and when the data is not changing rapidly.

1. Hive for OLTP system :

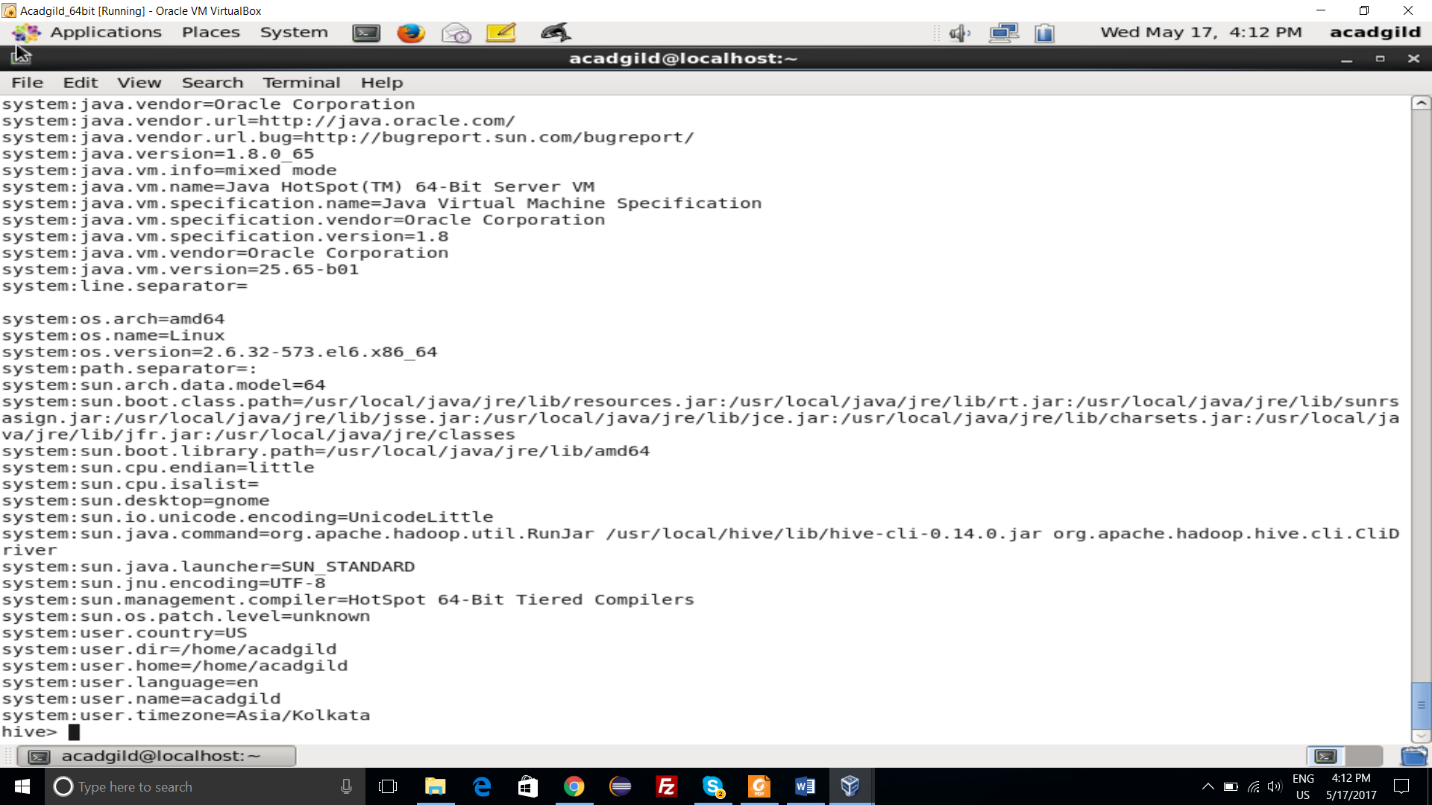
* No Hive does not provide insert and update at row level. So it is not suitable for OLTP system.
* SQL knowledge is wide spread and anyone who has decent knowledge would be able to use Hive effectively. Hive translates the query into Java Map Reduce code and runs the same on Hadoop cluster.
* Hive is best suited for Data Warehousing applications where data is structured, static and formatted. Hive is not a complete database. Design considerations of Hadoop and HDFS impose some constraints on what Hive can do.
* Hive does not provide row wise update and insert which is a biggest disadvantage of using it. But like we said earlier, Hive is not meant to be used for OTLP applications, it is meant to be used for Data Warehousing applications.

3) Metastore :

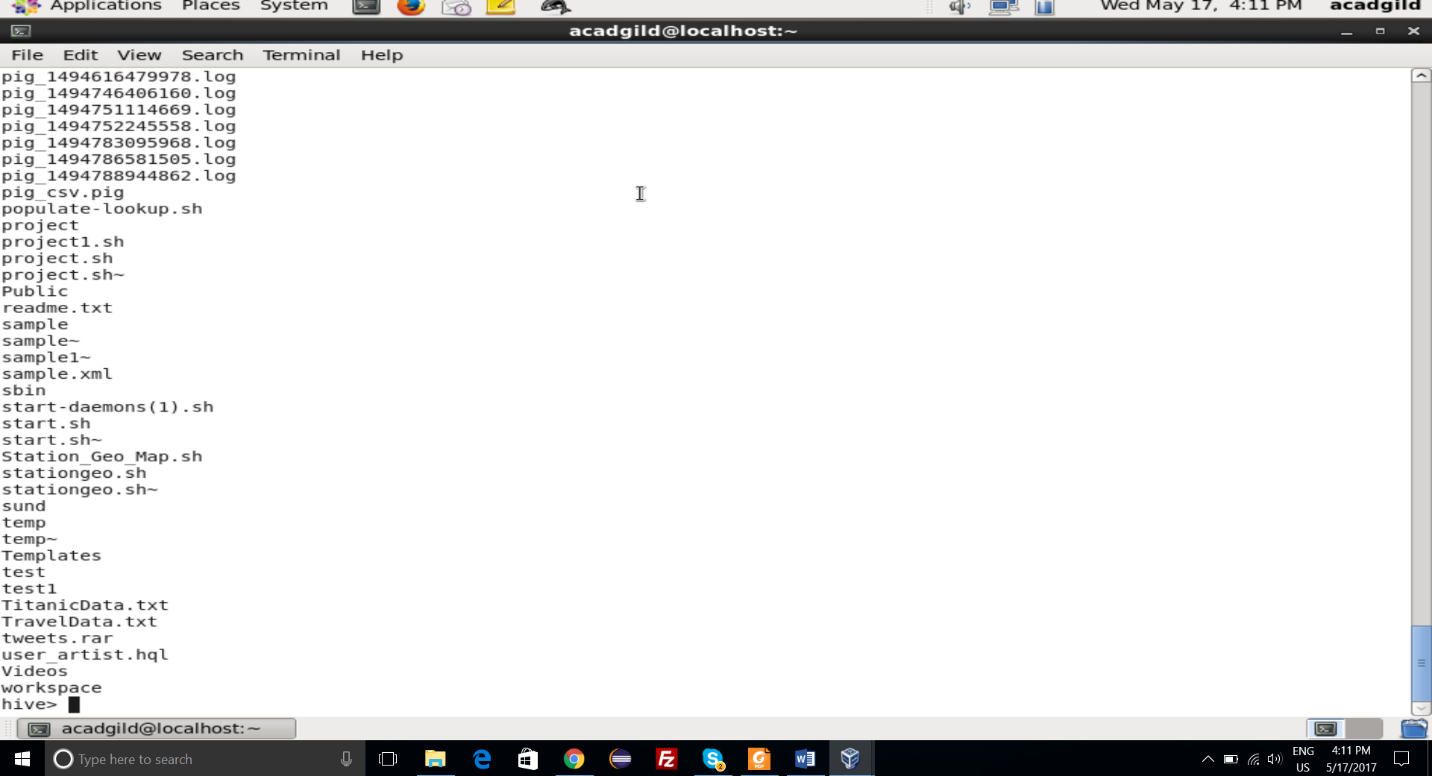
-> The Hive metastore service stores the metadata for Hive tables and partitions in a relational database, and provides clients (including Hive) access to this information via  
the metastore service API.  
 -> The metastore is the central repository of Hive metadata.  
 -> The metastore service runs in the same JVM as the Hive service and contains an embedded Derby \* database instance.  
 -> However, only one embedded Derby database can access the database files on disk at any one time, which means we can have only one Hive session open at a time.

1. Unix shell commands for Hive :

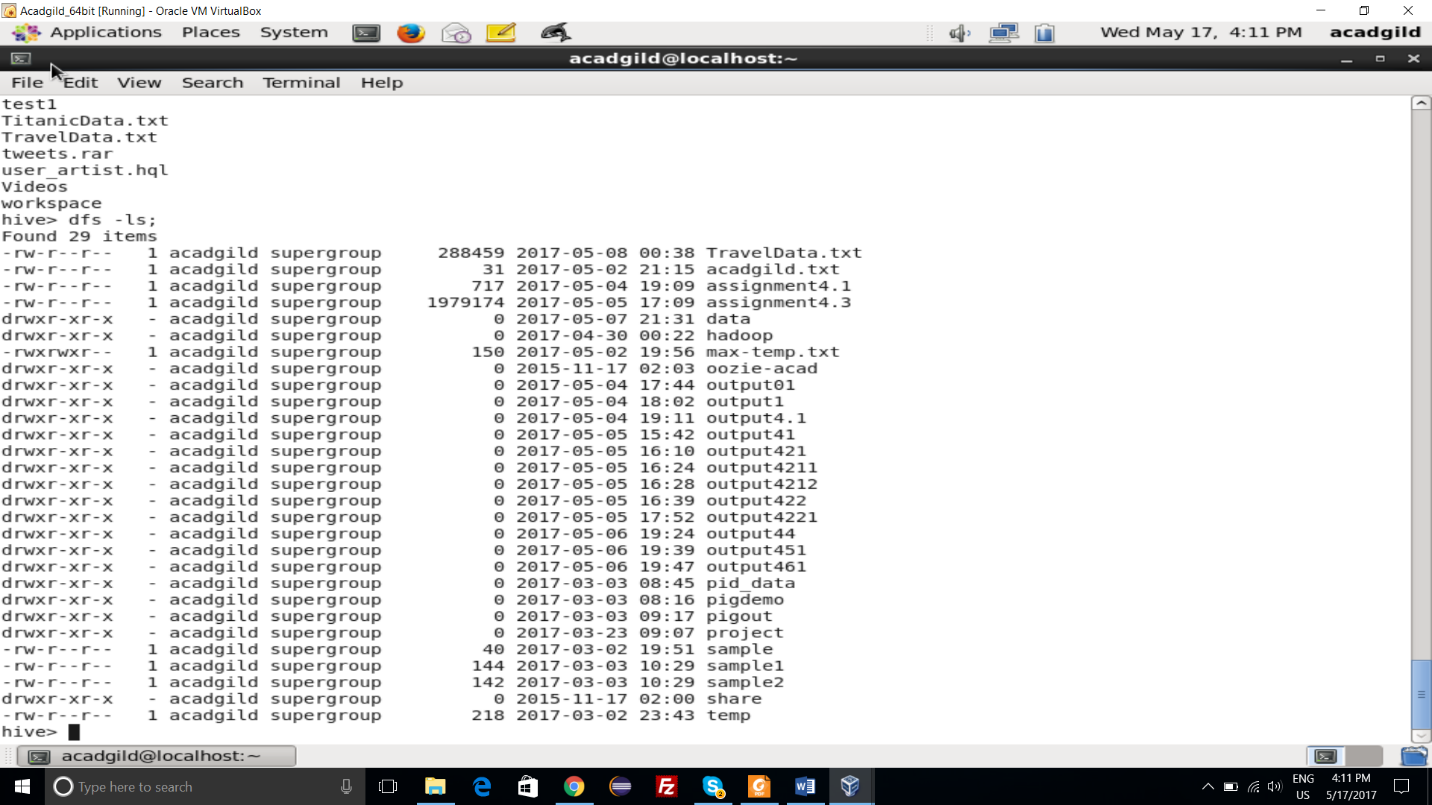
* Set : This will print a list of configuration variables that are overridden by user or hive.



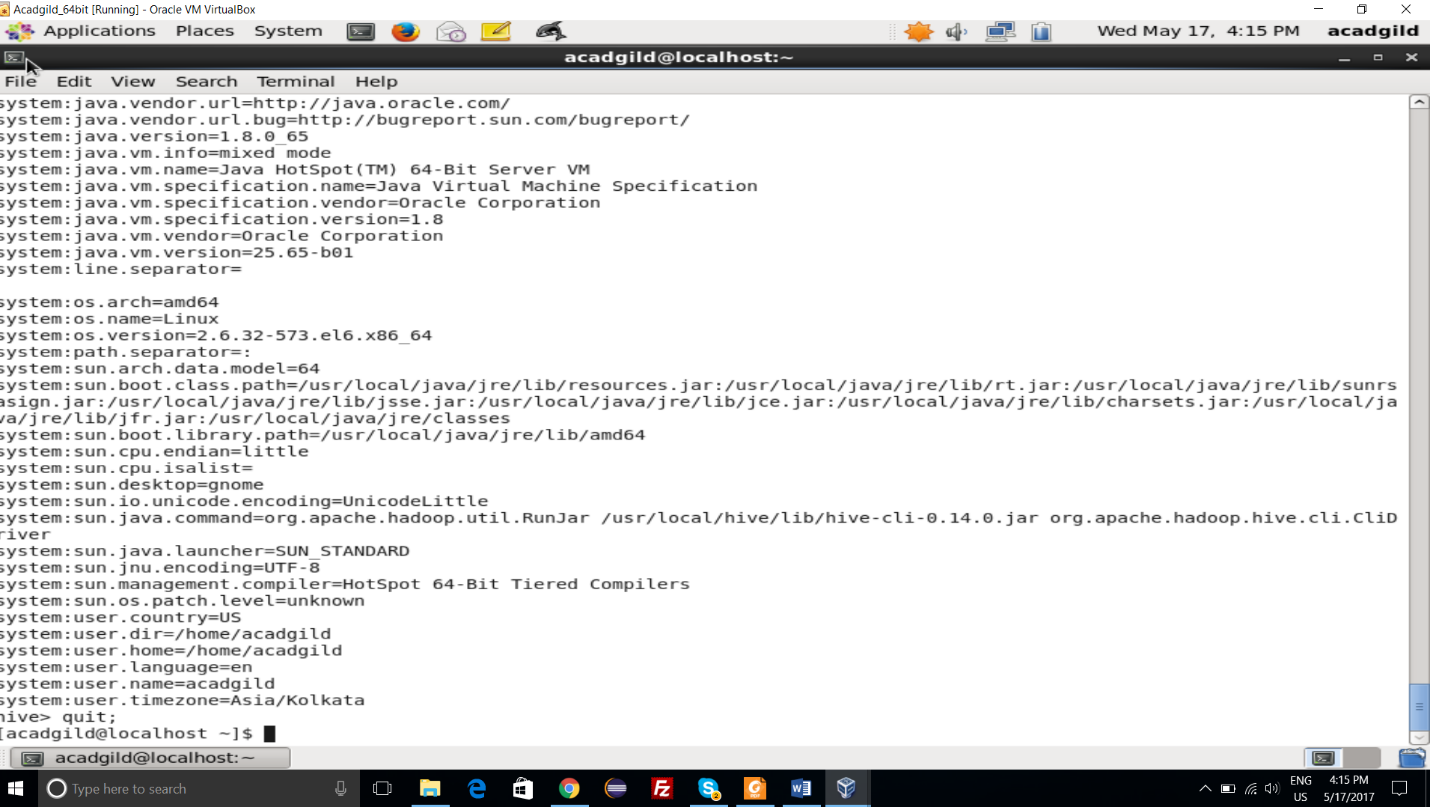
* !ls : lists all the files added to the distributed cache.



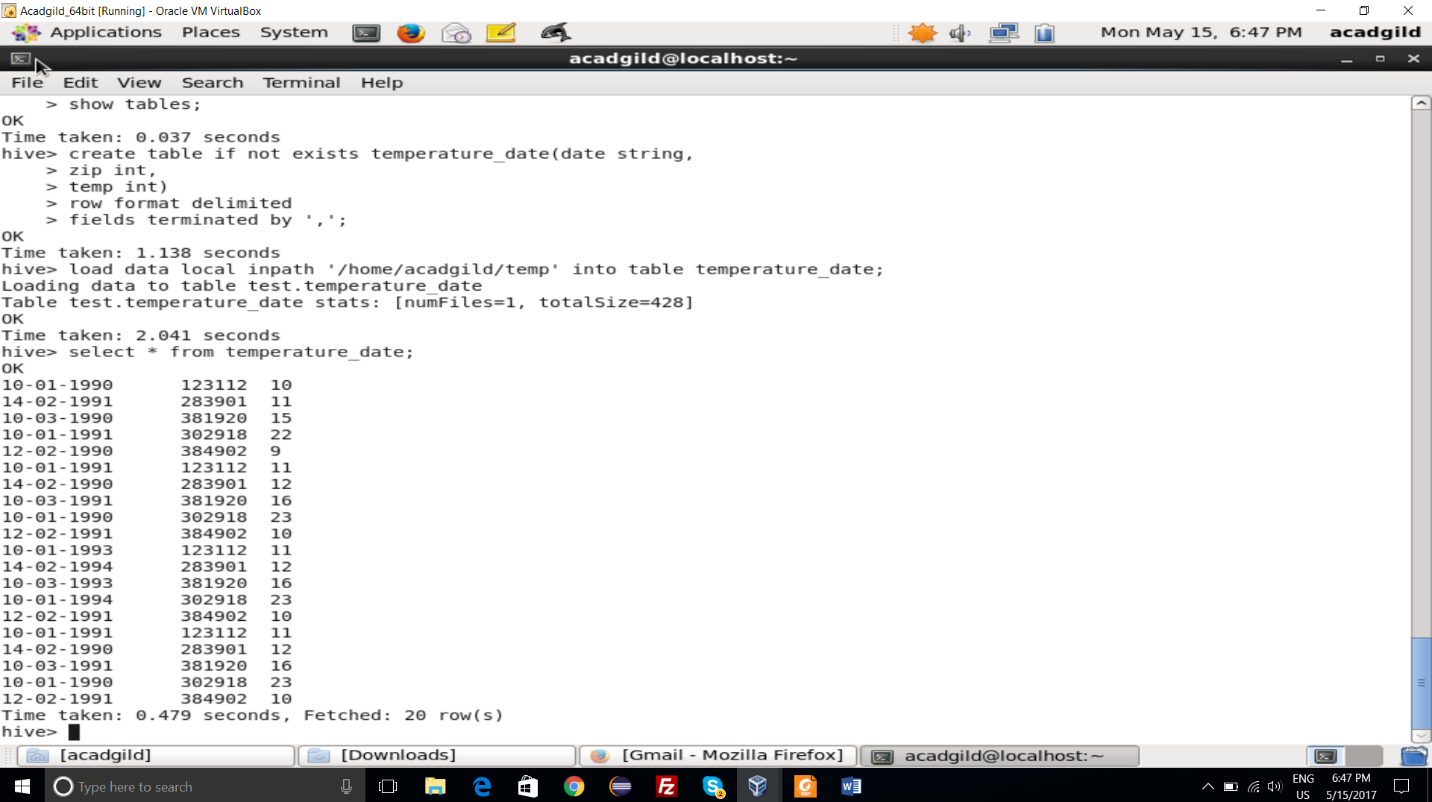
* dfs –ls : Executes a dfs command from hive shell.



* Quit : Use quit or exit to leave the interactive shell.



* Select : Used to select a particular command



1. Data Format Process in Hive :

* Yes, Hive uses the SerDe interface for IO operations.
* Different SerDe interfaces can read and write any type of data.
* If normal directly process the data where as different types of data is in the Hadoop.
* Hive use different SerDe interface to process such data.

Example : JsonSerde process Json data, AvroSerDe process Avro data