Need of hive

* **HIVE** was created for people with **SQL** background.
* The queries written is similar to SQL named as **HIVEQL**. HIVE was developed to process completely **structured data**. It is not used for ustructured data.
* Hive gives an SQL-like interface to query data stored in various databases and file systems that integrate with Hadoop.
* Traditional SQL queries must be implemented in the [MapReduce](https://en.wikipedia.org/wiki/MapReduce) Java API to execute SQL applications and queries over distributed data. Hive provides the necessary SQL abstraction to integrate SQL-like queries ([HiveQL](https://en.wikipedia.org/wiki/HiveQL)) into the underlying Java without the need to implement queries in the low-level Java API.

Is Hive suitable to be used for OLTP systems

No Hive does not provide insert and update at row level. So it is not suitable for OLTP system.

Metastore in hive

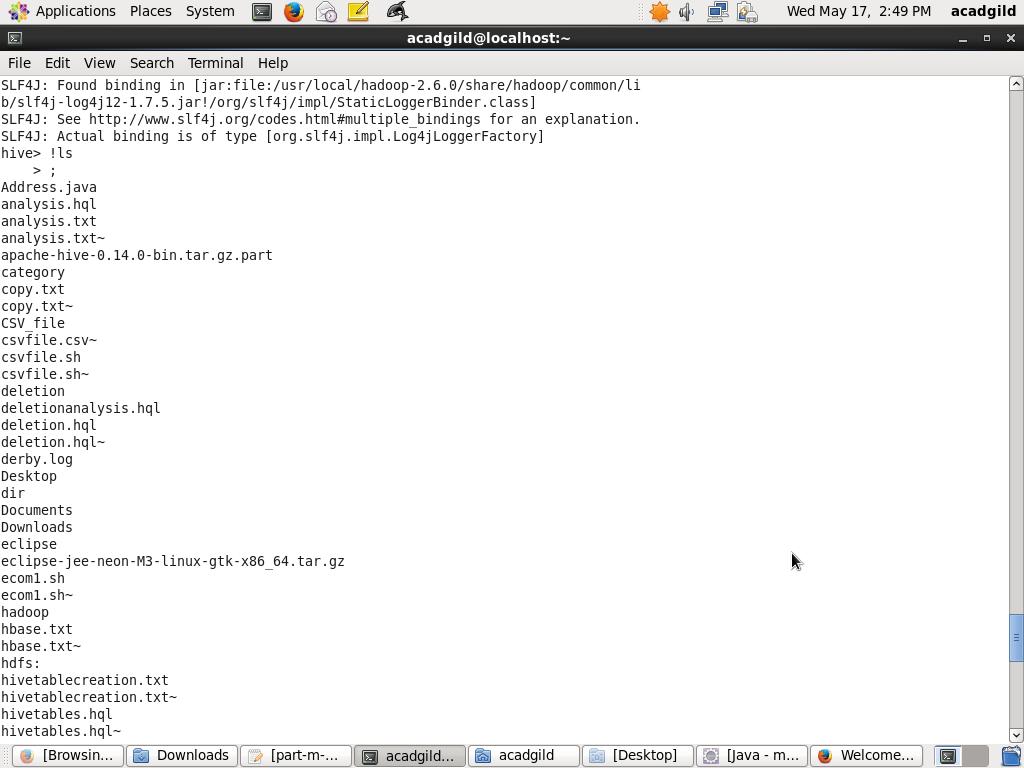
* The Hive metastore service stores the metadata for Hive tables and partitions in a relational database, and provides clients (including Hive) access to this information using the metastore service API.
* All Hive implementations need a metastore service, where it stores metadata. It is implemented using tables in a relational database. By default, Hive uses a built-in Derby SQL server.

Run unix shell commands in hive

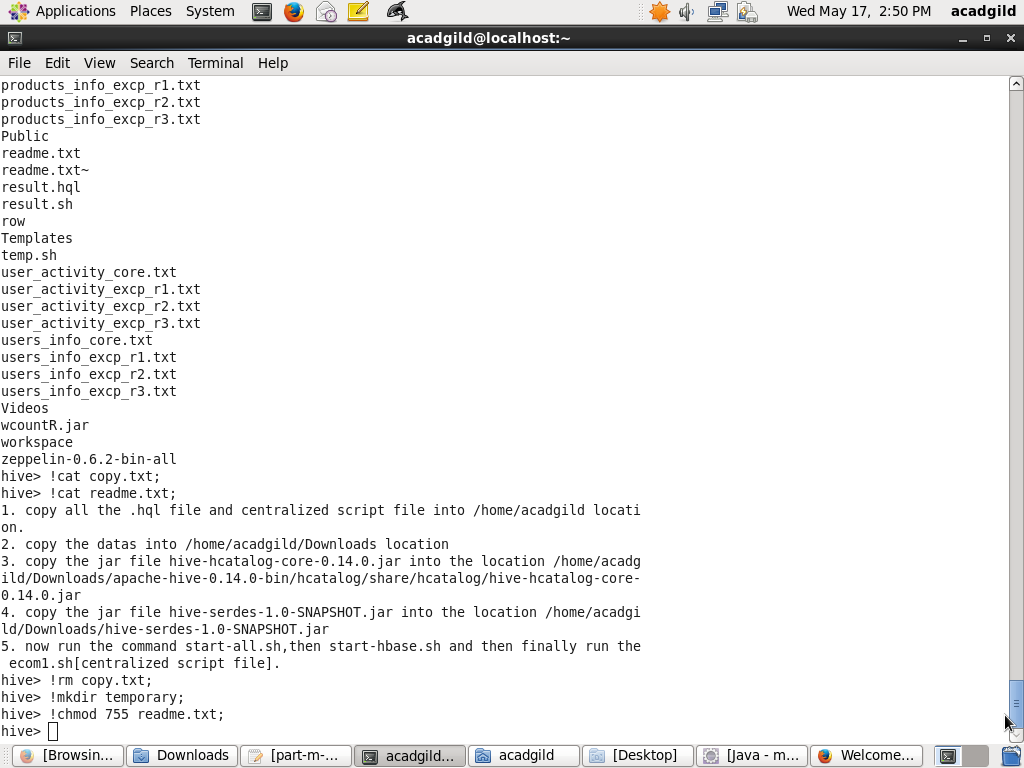
To run a unix shell command from Hive shell just append the unix shell command by !

For example:

Ls command in hive



Rm,cat,chmod,mkdir command in hive



Can we process any data format using 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 type of data is in the Hadoop, Hive use different SerDe interface to process such data.

Example:

MetadataTypedColumnsetSerDe : used to read/write CSV format data.

JsonSerDe : process Json data format