ASSIGNMENT 10.1

1. Explain in brief the need of hive.

SOLUTION

* It is used for querying data stored in the Hadoop Distributed Filesystem (HDFS) by writing SQL like queries.
* It is suitable for accessing and analyzing data in Hadoop using SQL syntax known as HiveQL (Hive Query Language).
* These HiveQL will automatically translate these queries into MapReduce jobs so that it can be executed on Hadoop.

2. Is Hive suitable to be used for OLTP systems? Explain in brief.

SOLUTION

* Hive does not provide INSERT and UPDATE at row level whereas OLTP is characterized by a large number of short on-line transactions (INSERT, UPDATE, DELETE)
* So, Hive is not suitable for OLTP systems

3. What is a metastore in Hive?

SOLUTION

* Metastore is a central repository in Hive.  It is used for storing schema information or metadata in the external database.
* It is a relational database storing the metadata of hive tables, partitions, Hive databases including column and column type information, the serializers and deserializers etc.
* By default Hive store metadata in derby database.

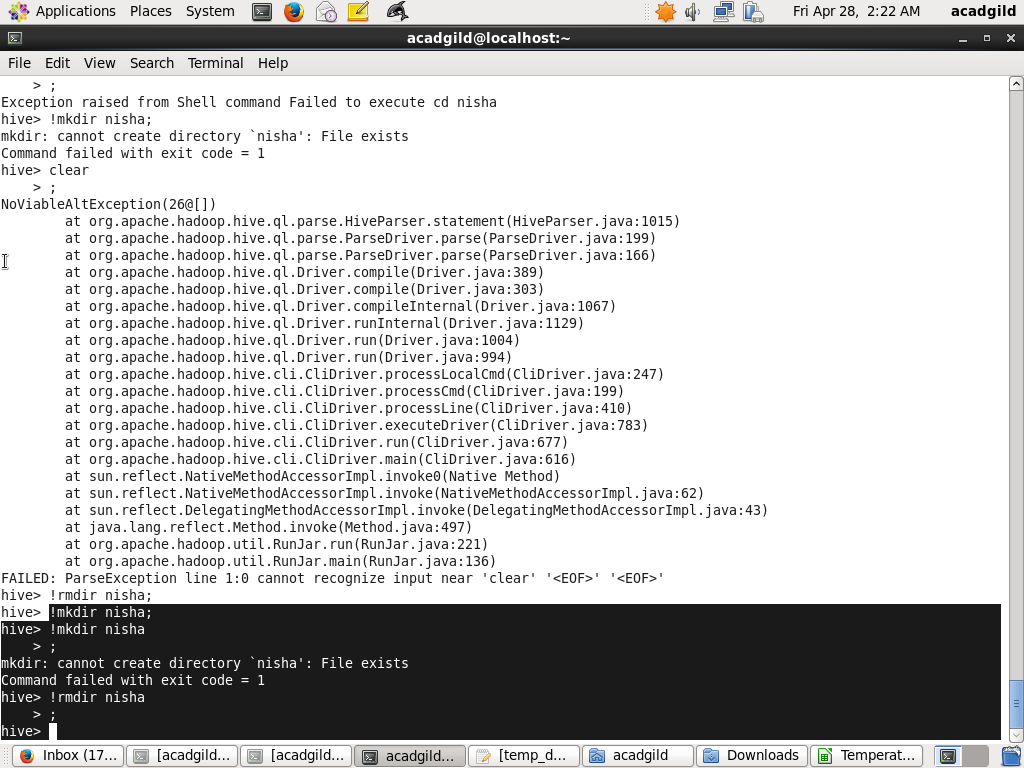
4. Can we run unix shell commands from hive? Give 5 examples and share the screenshot.

SOLUTION

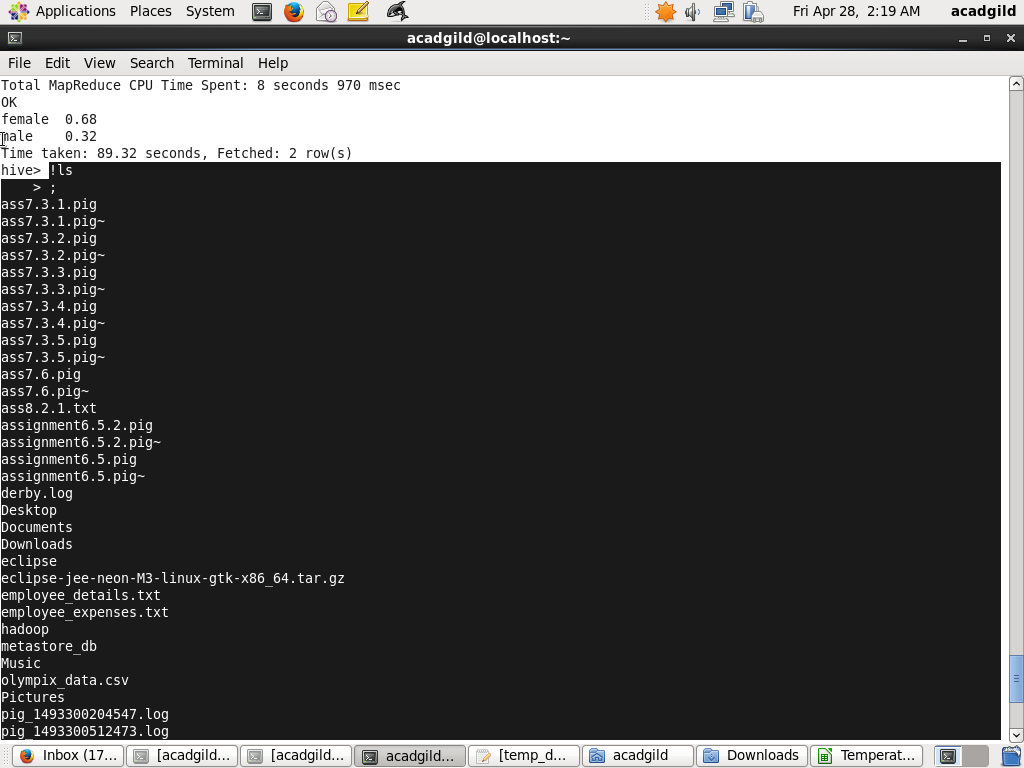
Yes we can run unix shell commands from hive by including ! symbol before unix shell commands

EXAMPLES

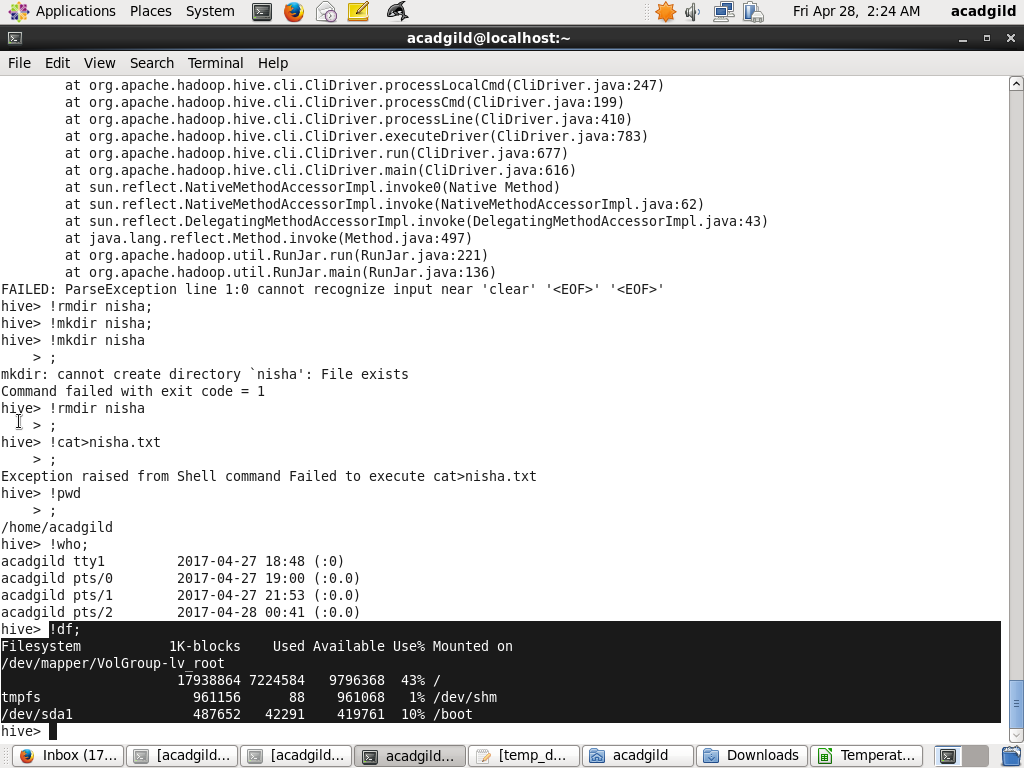
mkdir and rmdir COMMANDS



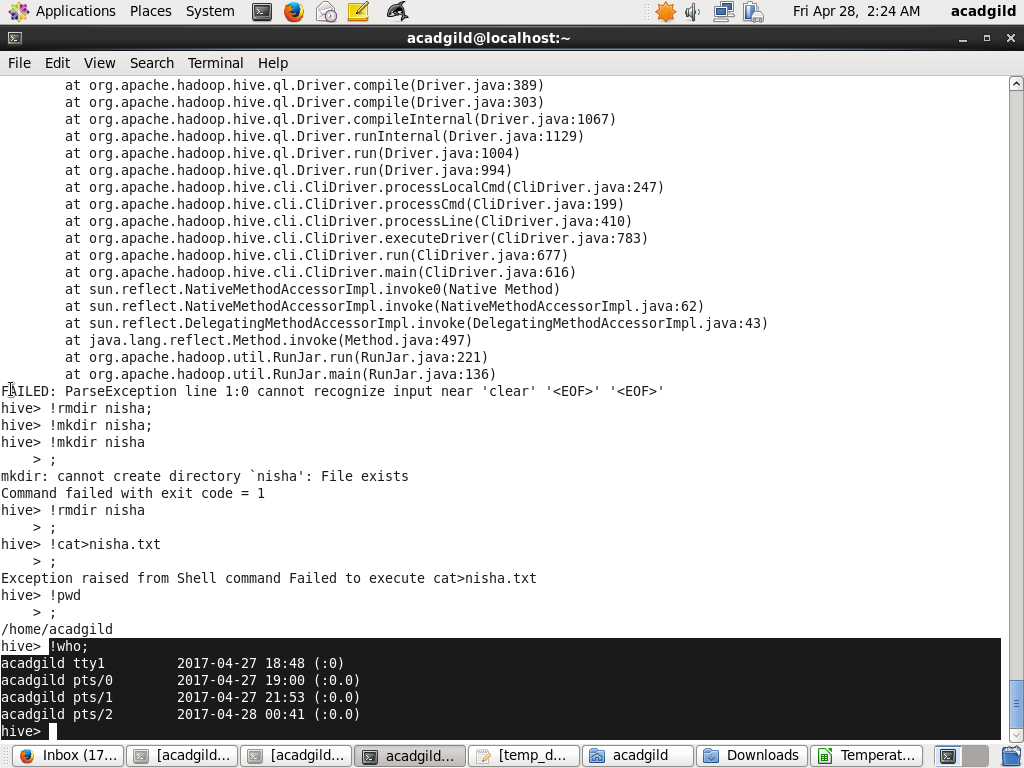
ls COMMAND



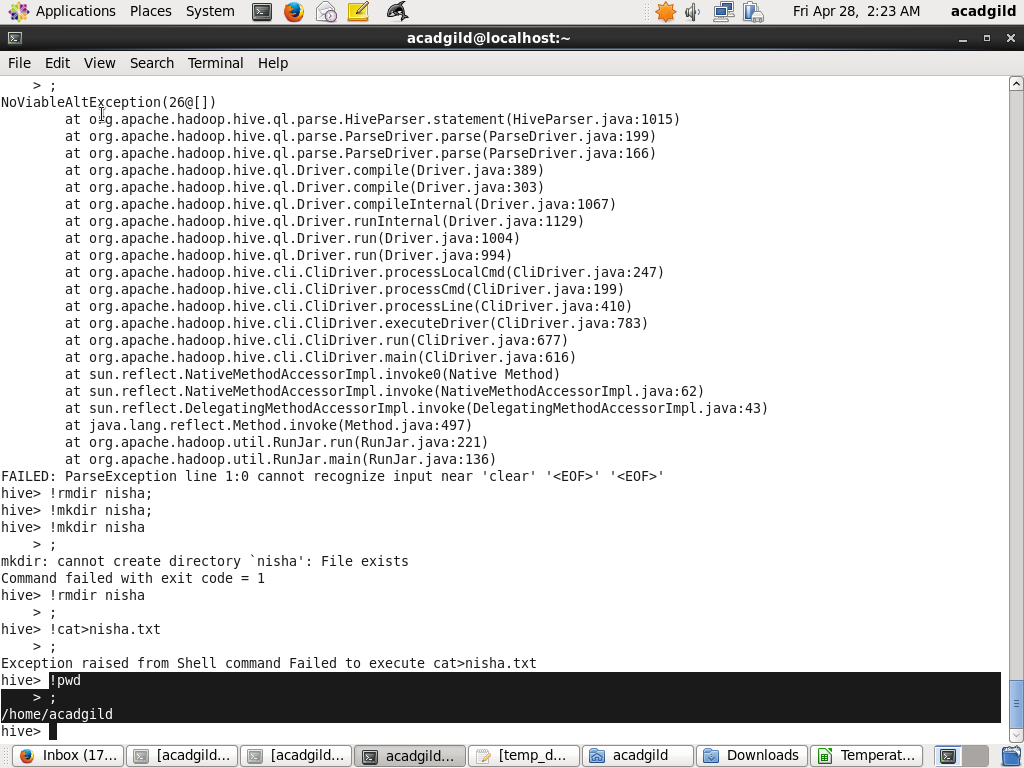
df COMMAND



who COMMAND



pwd COMMAND



5. Hive can process any type of data formats? Explain in brief?

SOLUTION

* Yes hive can process different types of data format because it uses SerDe interface for IO operations, different SerDe can be used to read and write different data types.
* Hive uses different SerDe to process different types of data formats

Example

JsonSerDe: process Json data.

RejexSerDe: process weblog data.

AvroSerde: Avro format data.