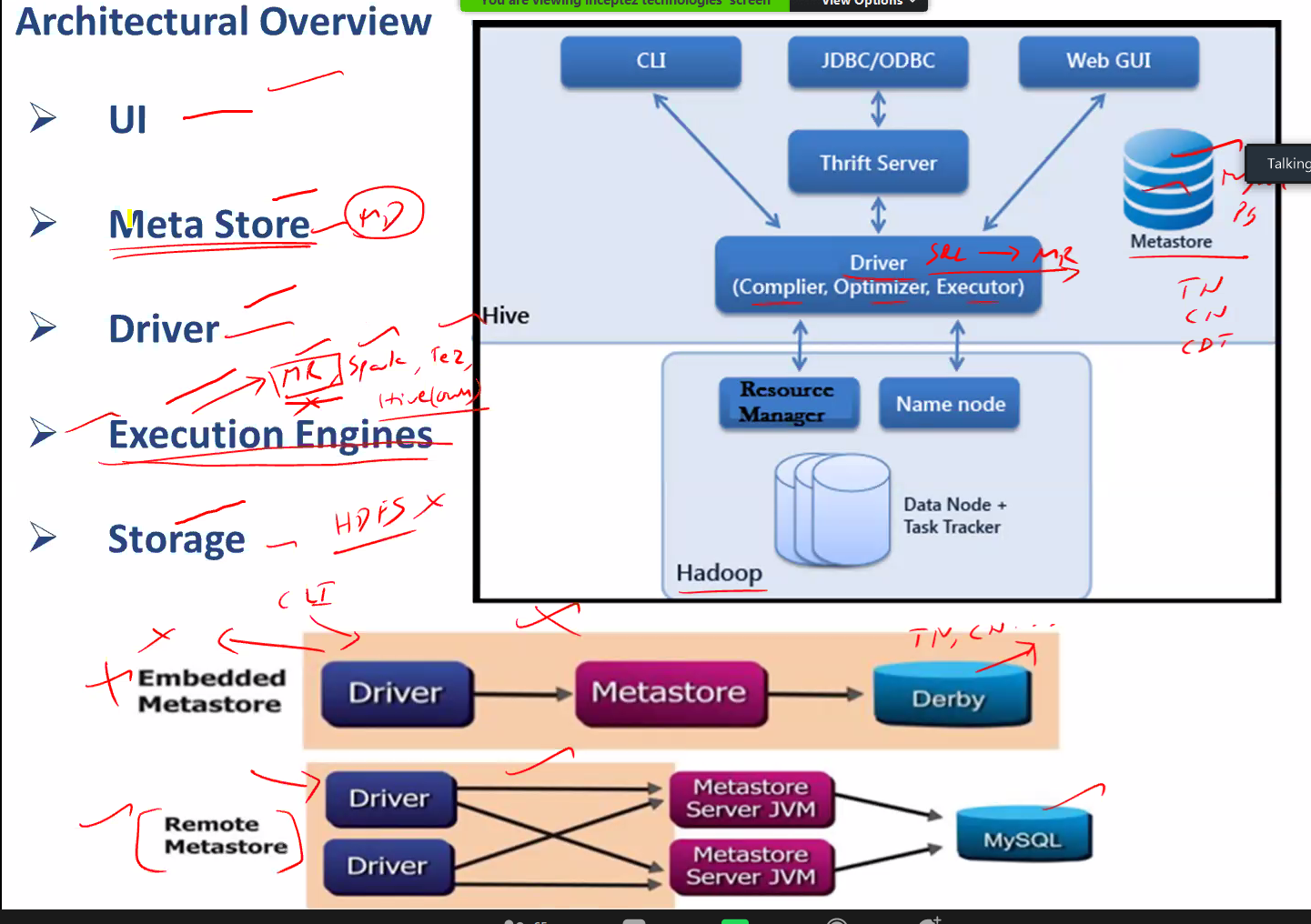
# HIVE

Hive is SQL layer on top of Hadoop. Its called HQL, Hive Query Language.

Hive Architecture:



### Metastore:

Metastore is required to store the Hadoop table structure information. It cannot be saved in Hadoop itself.

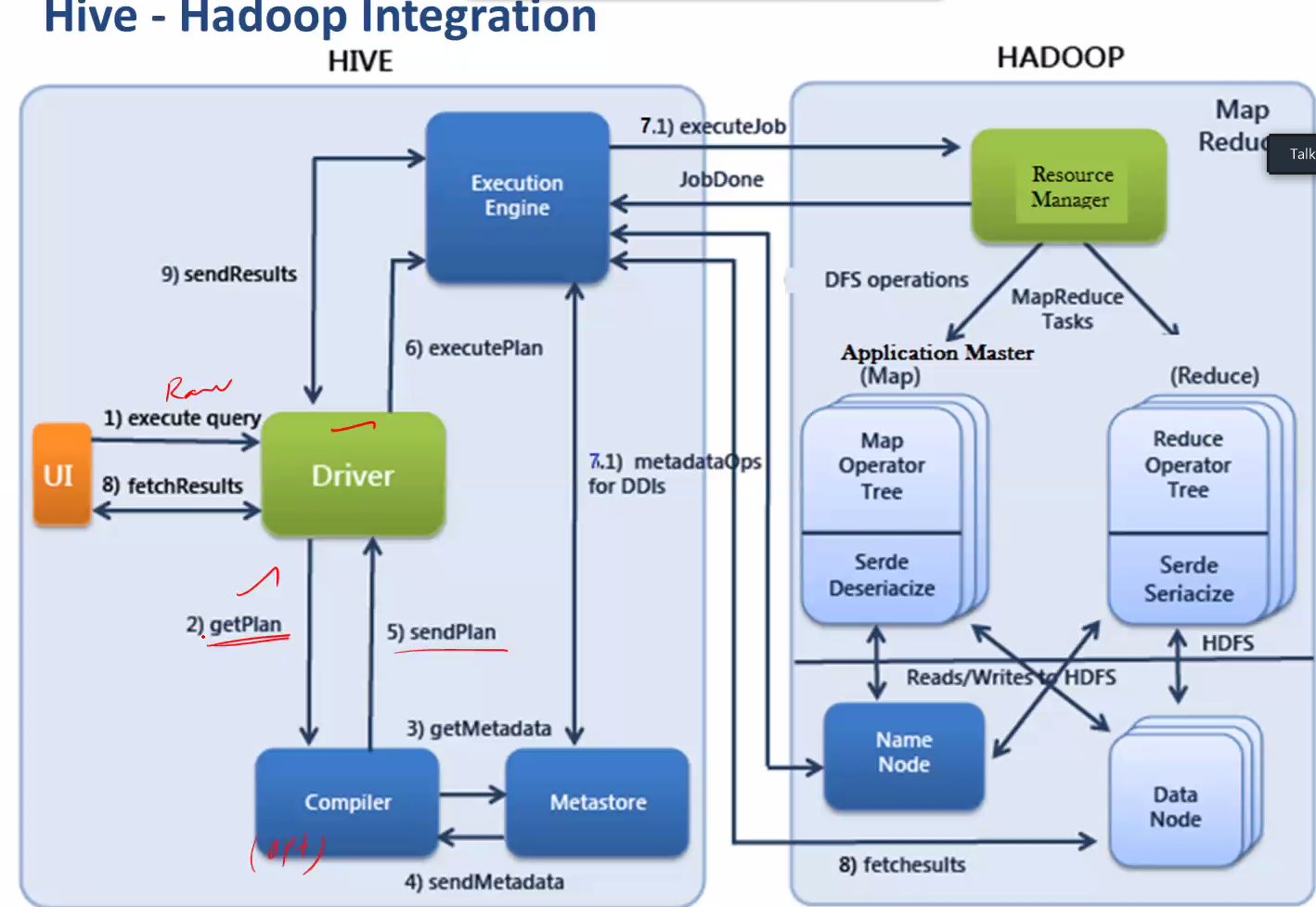
#### Embedded metastore:

Comes default with Hadoop, which contains Derby. Derby has limitation which can be used for Development and testing alone. No two sessions can be started at same time.

#### Metastore:

Hive supports many databases to store the metadata of table information. The metastore server jvm allows to start session with Hive

## Hive – Hadoop integration



5. Execution plan is generated, the best way to reach the database

6. Submit the plan to execution engire

7. Execute the job converting the job in to RM or Spark or Tez depending on the execution engine

NOTE: Some of the Queries can be directly fetched from Hadoop without involving any Hadoop execution engine. Hive to Hadoop directly. No use of RM, Spark or Tez

# SerDes - Serialization and Deserialization

Serialization: Converting the original data to byte code

Deserialization: Converting the byte code to original data

What is a SerDe?

* SerDe is a short name for "Serializer and Deserializer."
* Hive uses SerDe (and FileFormat) to read and write table rows.
* HDFS files --> InputFileFormat --> <key, value> --> Deserializer --> Row object
* Row object --> Serializer --> <key, value> --> OutputFileFormat --> HDFS files

# Hive Workout:

Create database:

CREATE (DATABASE|SCHEMA) [IF NOT EXISTS] database\_name

  [COMMENT database\_comment]

  [LOCATION hdfs\_path]

  [MANAGEDLOCATION hdfs\_path]

  [WITH DBPROPERTIES (property\_name=property\_value, ...)];

From above ( ) – is a mandatory parameter

From above [] – is a optional parameter

DATABASE|SCHEMA: These are the same thing. These words can be used interchangeably.

[IF NOT EXISTS]: This is an optional clause. If not used, an error is thrown when there is an attempt to create a database that already exists.

[COMMENT]: This is an optional clause. This is used to place a comment for the database. This comment clause can be used to add a description about the database. The comment must be in single quotes.

[LOCATION]: This is an optional clause. This is used to override the default location with the preferred one.

[WITH DBPROPERTIES]: This is an optional clause. This clause is used to set properties for the database. These properties are key-value pairs that can be associated with the database to attach additional information with the database.

To get the database details:

DESCRIBE DATABASE [EXTENDED] db\_name;