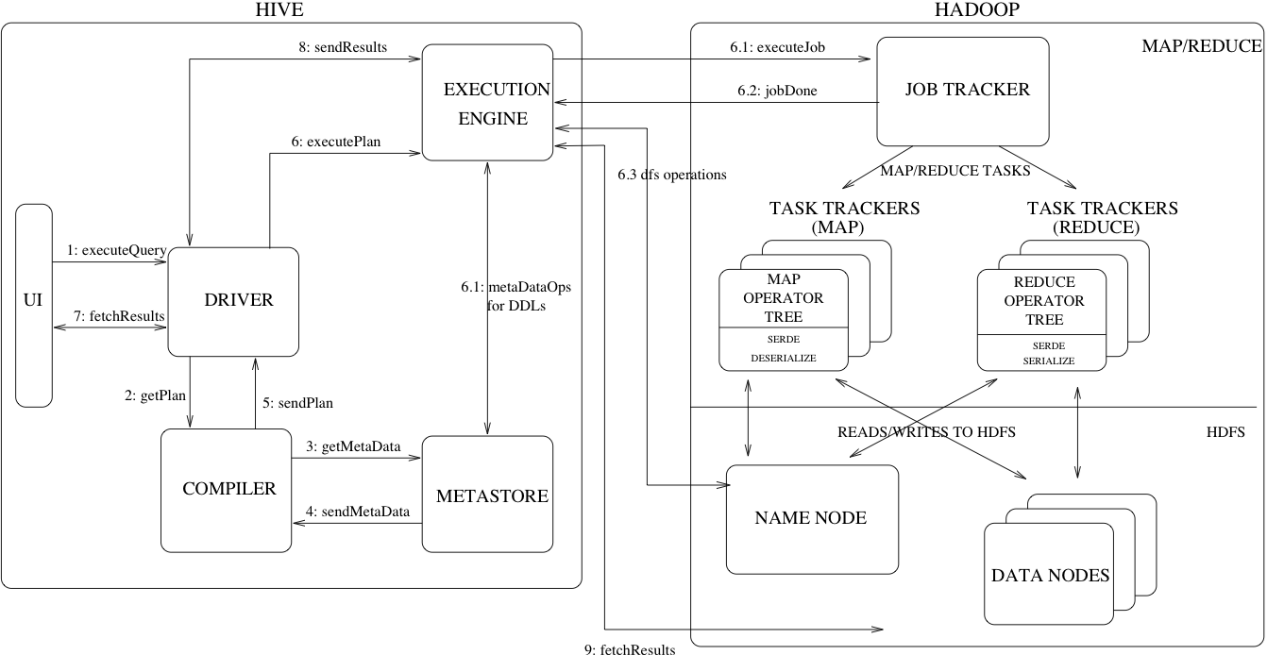
Hive is a data warehousing infrastructure based on Apache Hadoop



* **Databases**: Namespaces function to avoid naming conflicts for tables, views, partitions, columns, and so on.
* **Tables**: Homogeneous units of data which have the same schema.
* **Partitions** 分区: Each Table can have one or more partition Keys which determines how the data is stored. Partitions—apart from being storage units—also allow the user to efficiently identify the rows that satisfy a specified criteria.
* **Buckets** 分桶: Data in each partition may in turn be divided into Buckets based on the value of a hash function of some column of the Table.
* **Metastore** is a central repository of metadata for Hive tables and partitions in a relational database.