

INTRODUCTION TO BIG DATA

ECAP456

Dr. Rajni Bhalla
Associate Professor

Learning Outcomes



After this lecture, you will be able to

- explore the concepts of HIVE.
- learn features of HIVE.
- understand architecture of HIVE.
- learn components and functionalities of HIVE.
- explore working of HIVE.

Introduction



Introduction



Introduction



Introduction



Developed
By

Introduction



Developed
By



Introduction



Developed
By



Introduction

NETFLIX

amazon

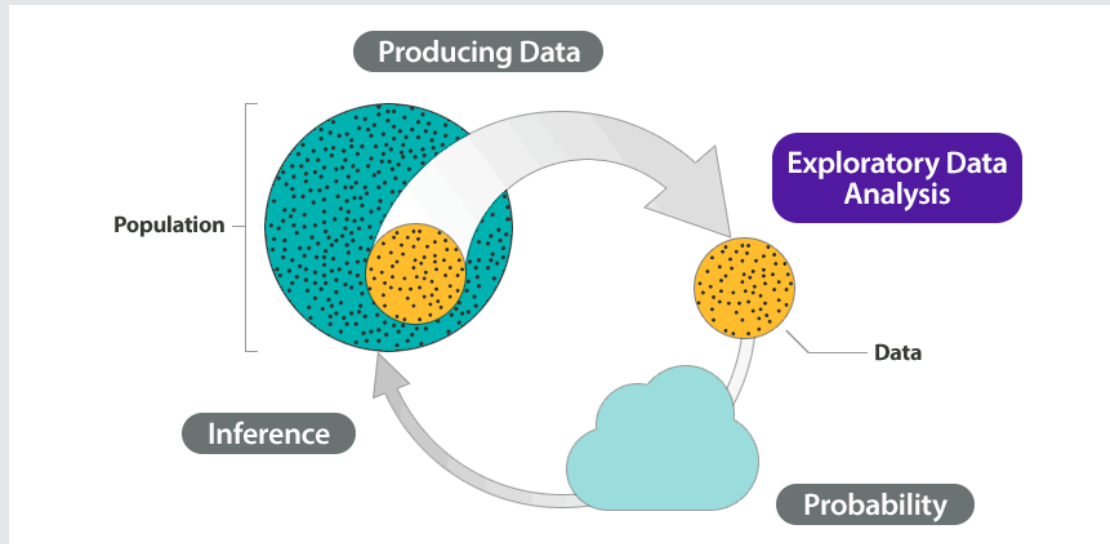
The Amazon logo, featuring a curved orange arrow pointing from the letter 'a' to the letter 'z'.

Introduction



Process Structured Data in
Hadoop

Introduction



Summarize Big Data

Introduction



Introduction



Amazon elastic
MapReduce

Introduction

Hive is not

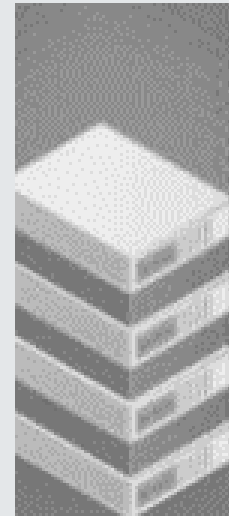


RDBMS

Relational Databases

Introduction

Hive is not



Designed for OLTP(Online
Transaction Processing)

Introduction



Hive is not

- A language for real-time queries
- and row-level updates

Hive

Features of HIVE

1. Store Schemas



Database

Hive

Features of HIVE

1. Store Schemas



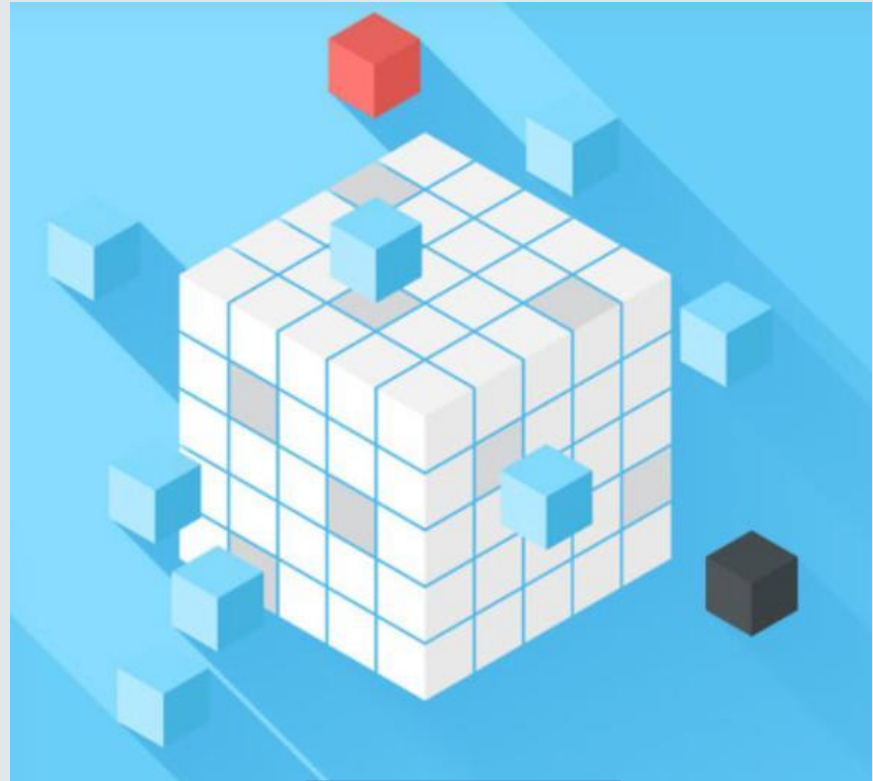
Database



Hive

Features of HIVE

1. Store Schemas
2. Designed for

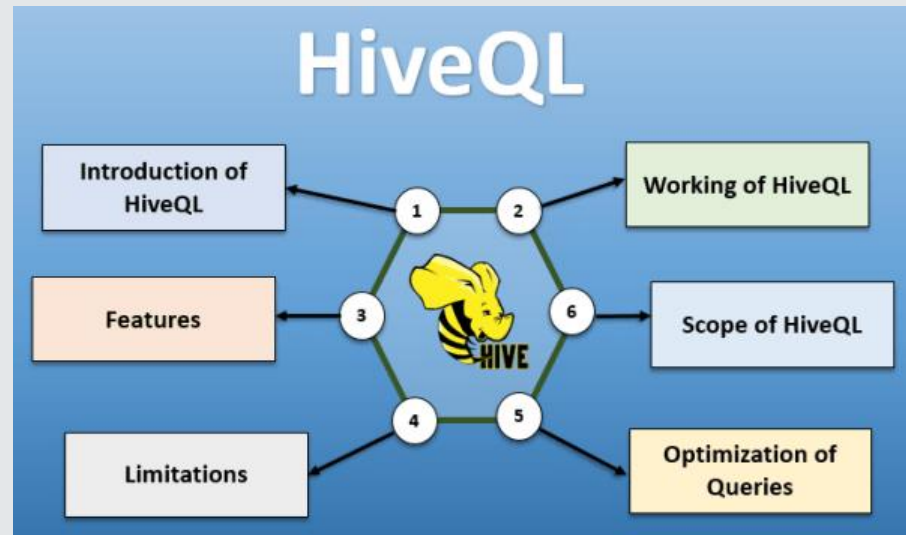


OLAP

Hive

Features of HIVE

1. Store Schemas
2. Designed for
3. provides SQL type language for querying called



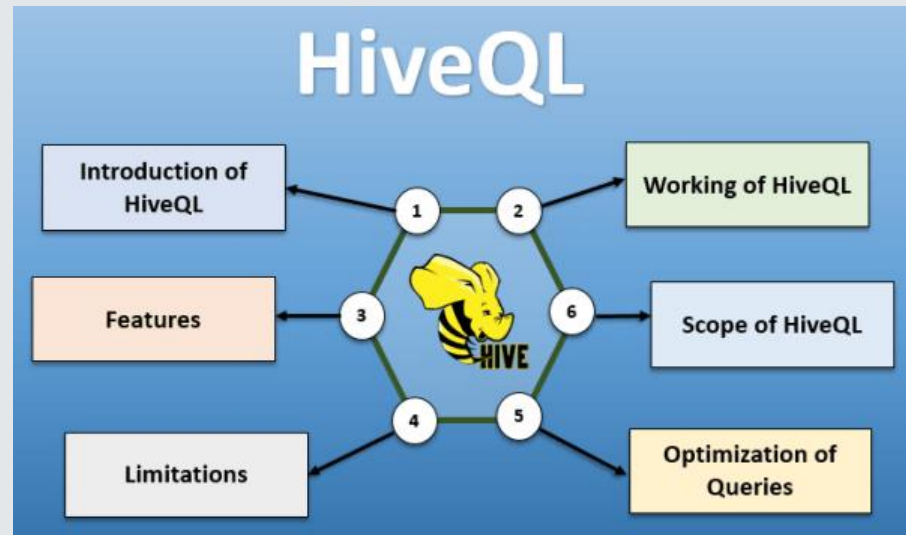
Hive

Features of HIVE

1. Store Schemas
2. Designed for
3. provides SQL type

language for querying called

4. Familiar, fast, scalable, and extensible.





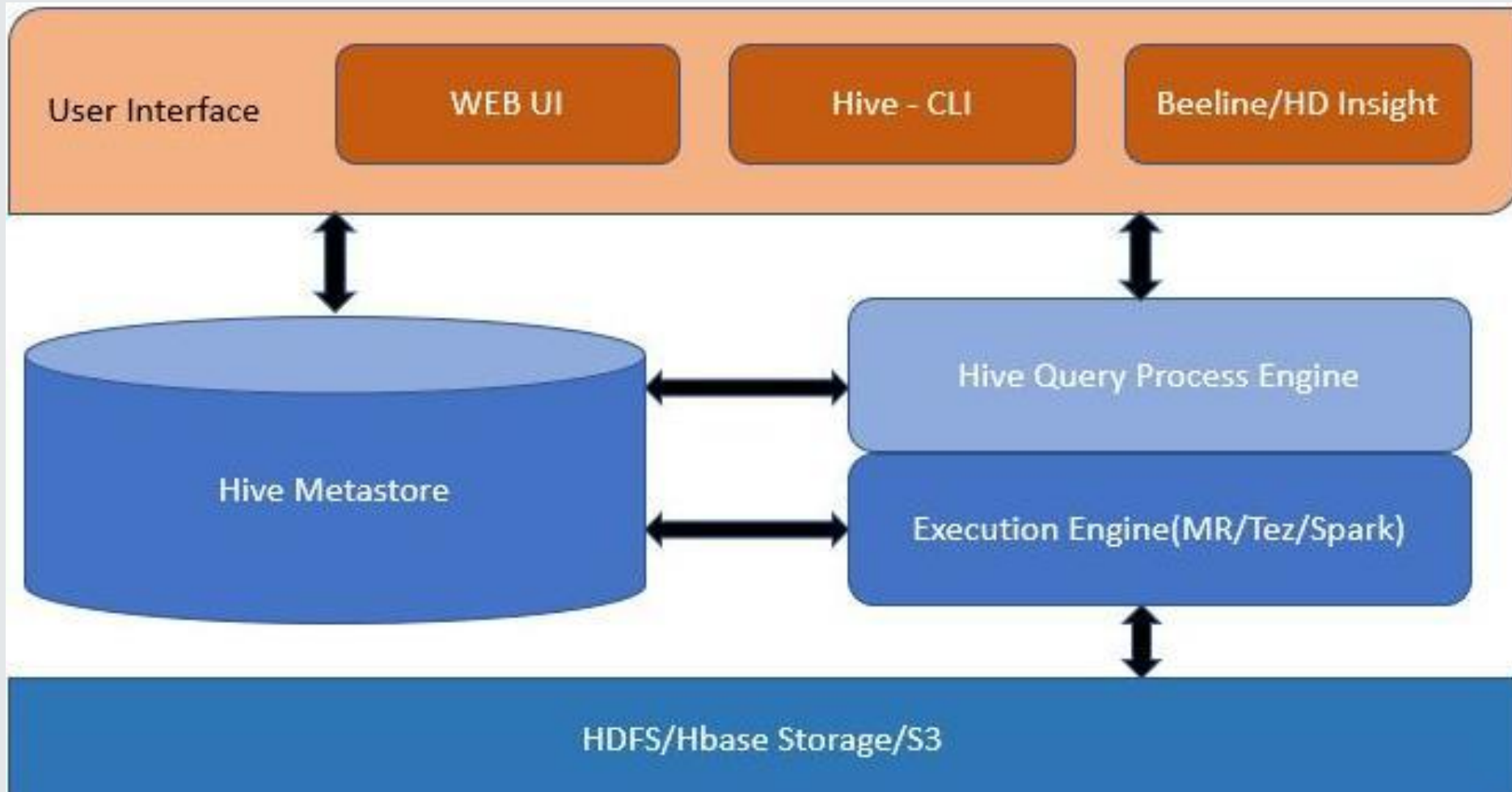
Architecture of HIVE



Architecture of HIVE

Next Figure shows the overall architecture of Hive. If you can understand this architecture diagram you can easily remember how Hive works.

Architecture of Hive



Architecture of Hive



Architecture of Hive

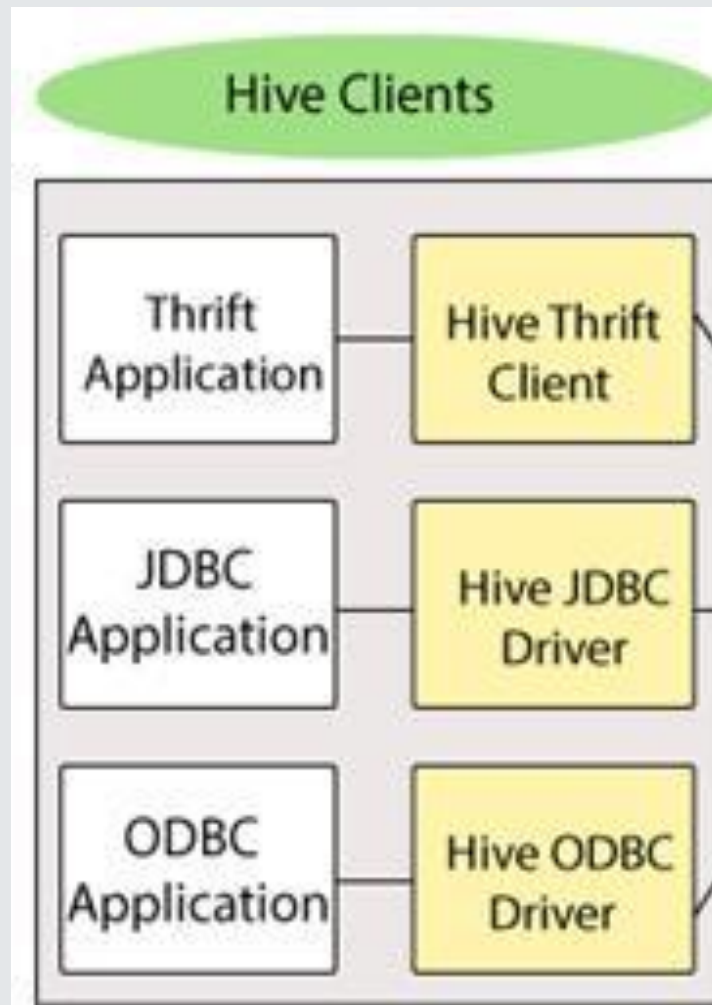


End User

Interact



Architecture of Hive



Architecture of Hive

Beeline

```
beeline> !connect jdbc:hive2://
Connecting to jdbc:hive2://
Enter username for jdbc:hive2://: hive
Enter password for jdbc:hive2://: *****
18/01/08 14:18:32 [main]: WARN service.CompositeSe:
Connected to: Apache Hive (version 1.2.1000.2.5.0.0-1245)
Driver: Hive JDBC (version 1.2.1000.2.5.0.0-1245)
Transaction isolation: TRANSACTION_REPEATABLE_READ
0: jdbc:hive2://> !q
Closing: 0: jdbc:hive2://
```

Hive CLI

```
hive$ hive
Logging initialized using configuration
hive> █
```

Architecture of Hive

Hive Query Process Engine

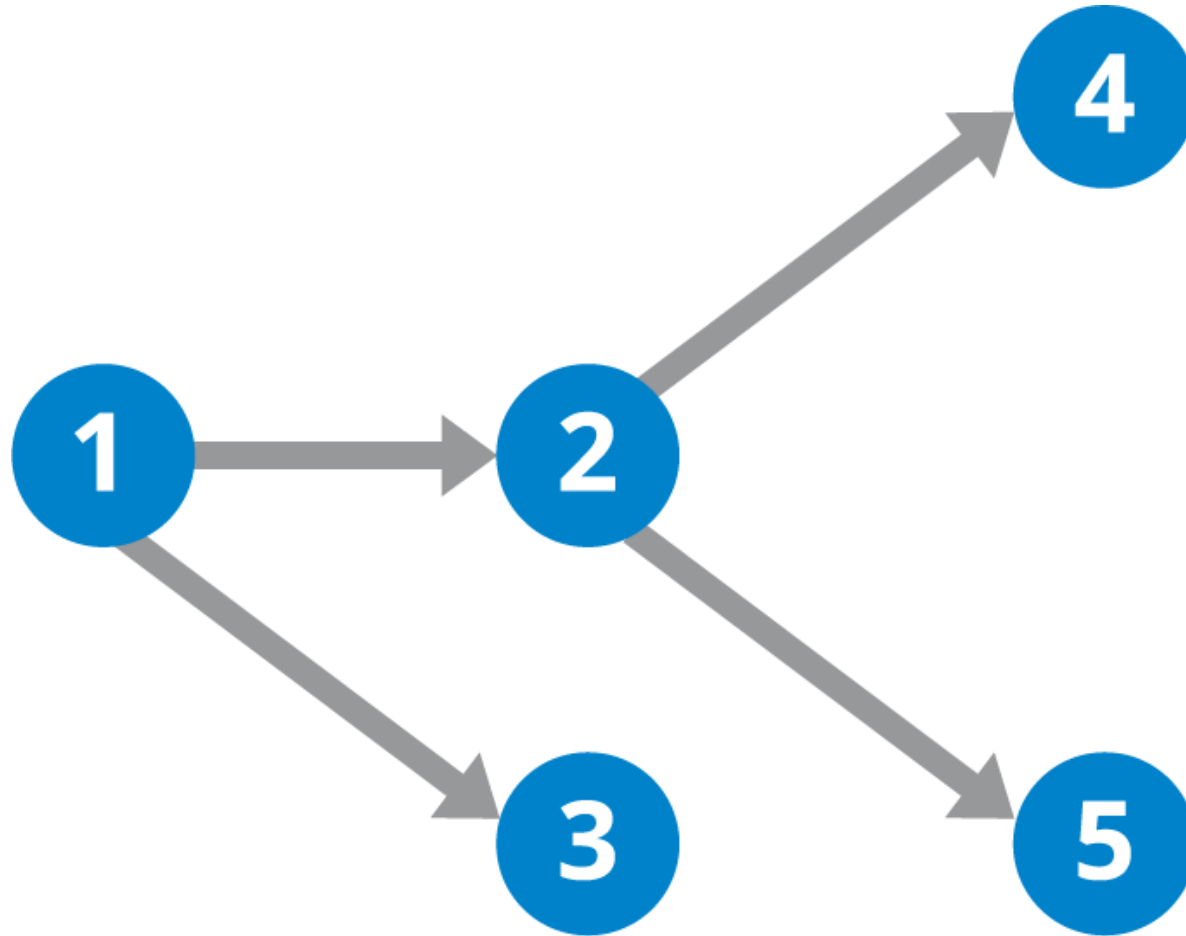
Architecture of Hive



Syntax

Semantics

Architecture of Hive



Directed Acyclic Graph

Architecture of Hive

Execution Engine(MR/Tez/Spark)

Architecture of Hive

- It is where the actual processing of the data will start.
- After compiler checking the syntax, performs the optimizations of the execution.
- Finally this execution plan will be given to Execution Engine. We have several execution engines that can be used with Hive.
- MapReduce is one of the execution engines.

Architecture of Hive



Architecture of Hive

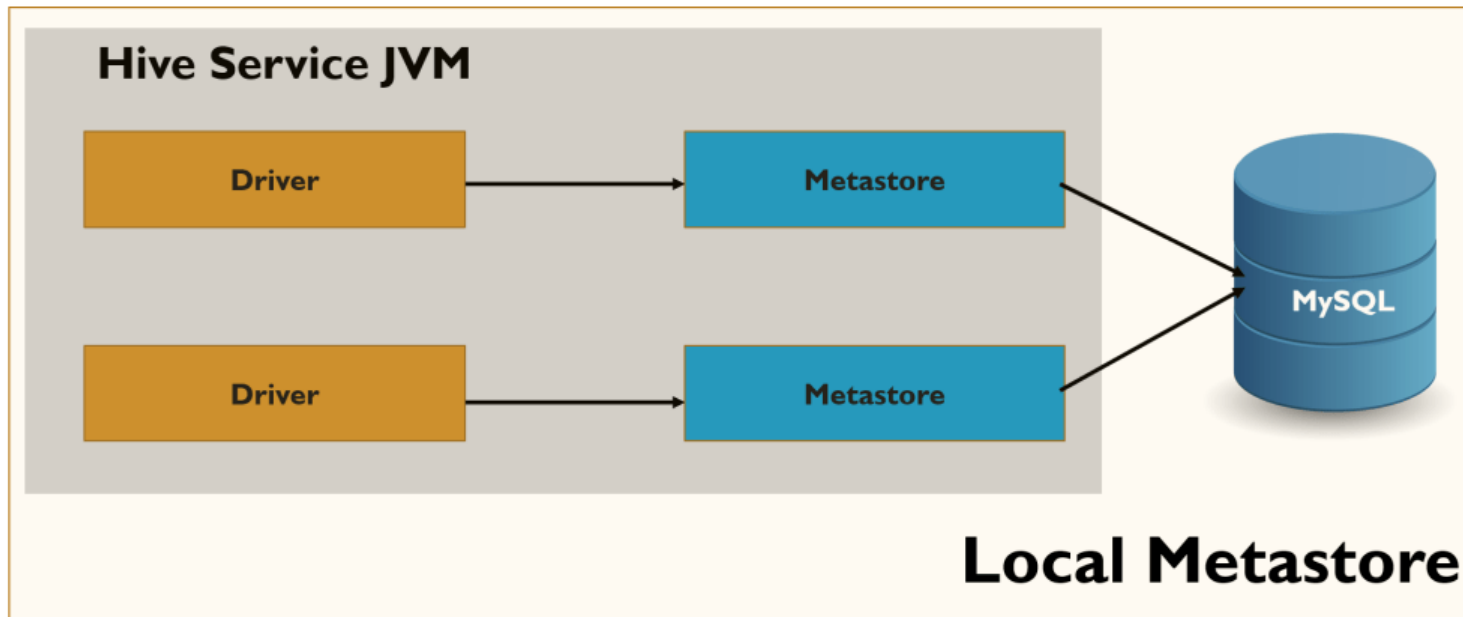
```
set hive.execution.engine=spark;
```

```
set hive.execution.engine=tez;
```

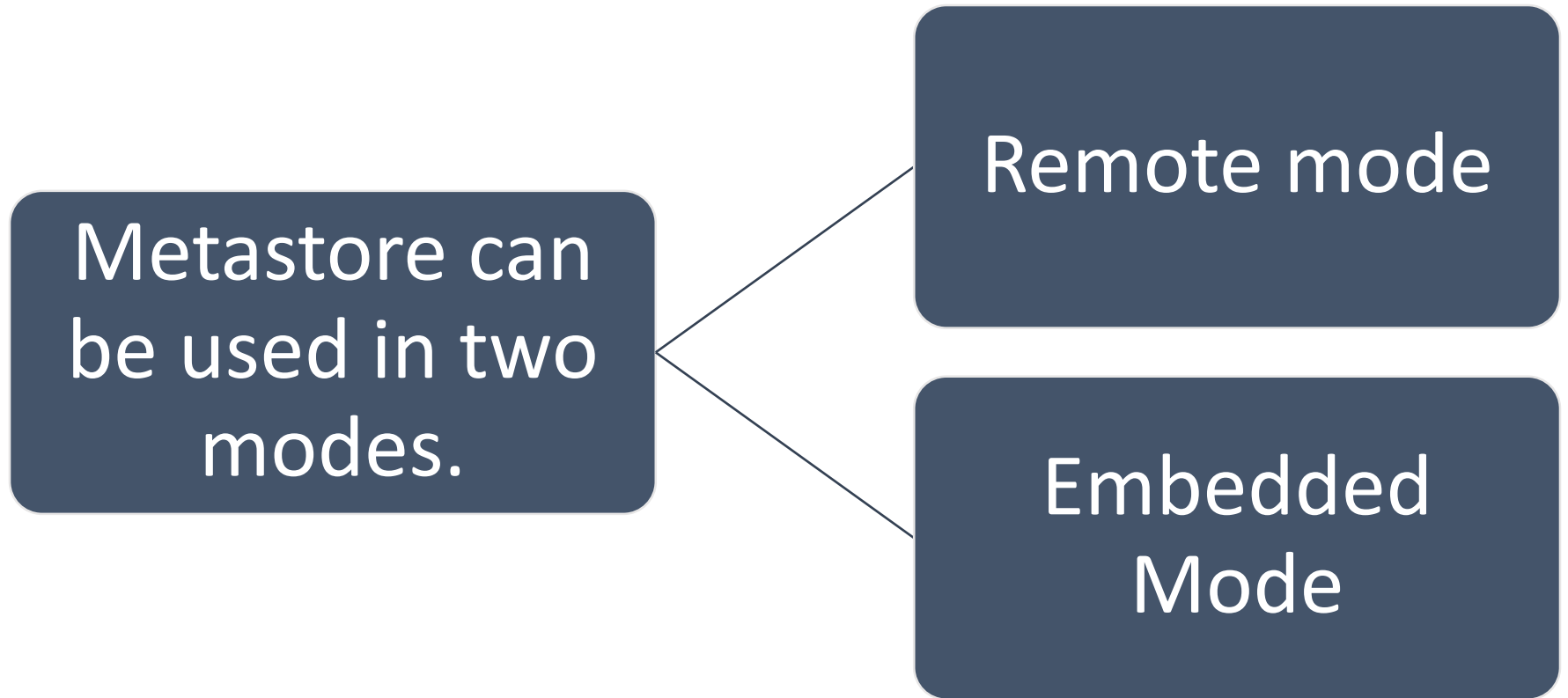
```
set hive.execution.engine=mr;
```

Architecture of Hive

Metastore



Architecture of Hive



Architecture of Hive

HDFS or Hbase
Storage

Architecture of Hive

HDFS or Hbase
Storage



Architecture of Hive

HDFS or Hbase
Storage



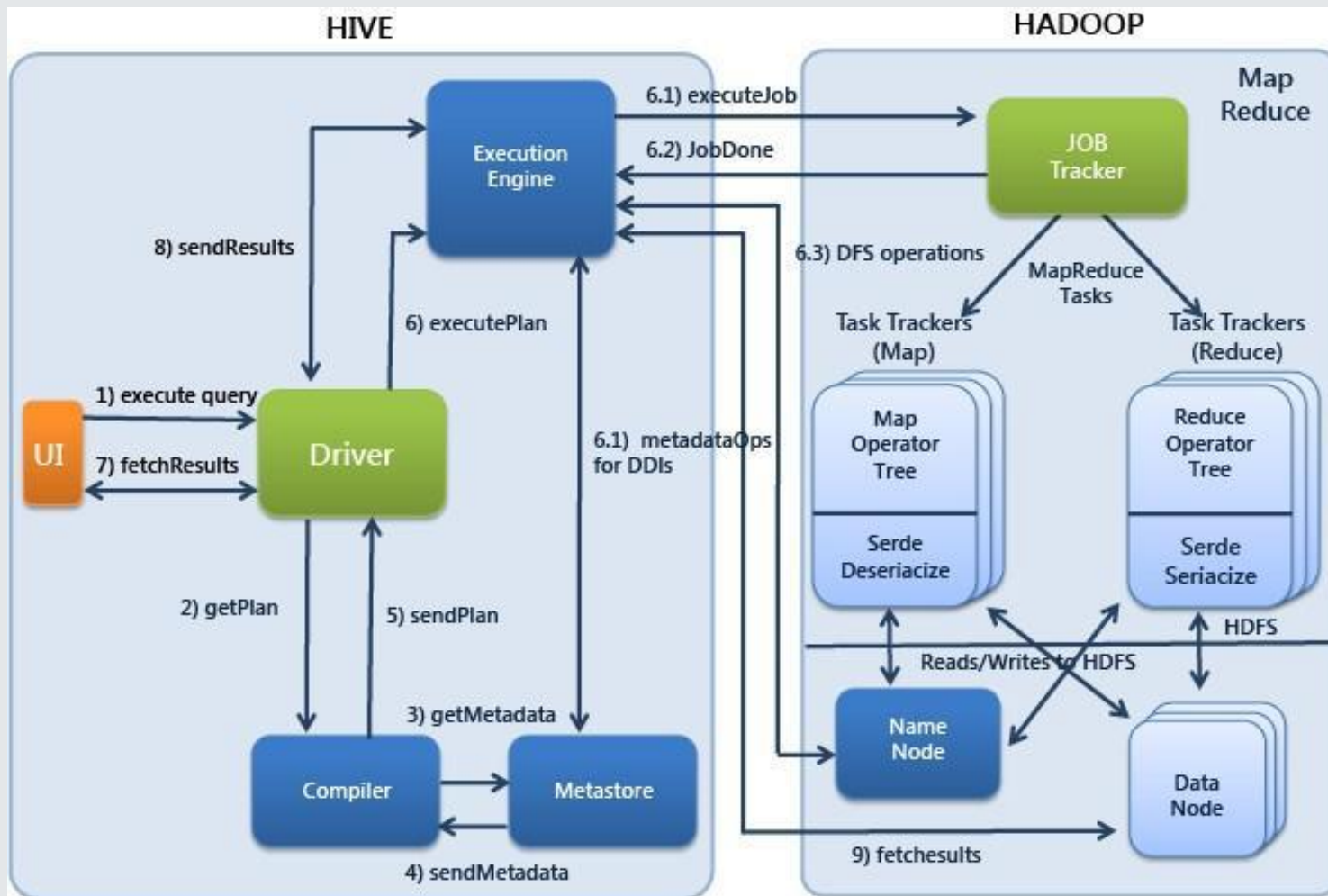
Architecture of Hive

HDFS or Hbase
Storage



Amazon S3

Components of HIVE





Working of HIVE

Working of HIVE

User
interface

Step 1

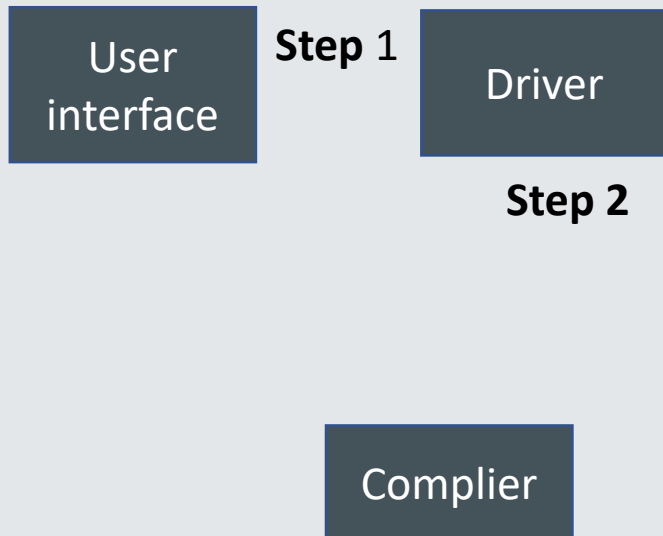
Working of HIVE

User
interface

Step 1

Driver

Working of HIVE



Working of HIVE

User
interface

Step 1

Driver

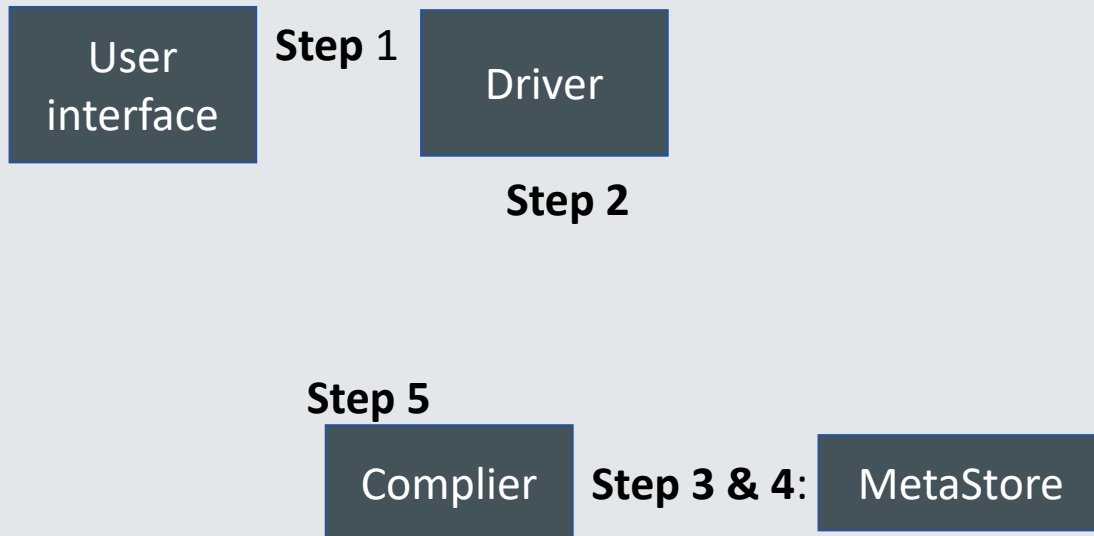
Step 2

Compiler

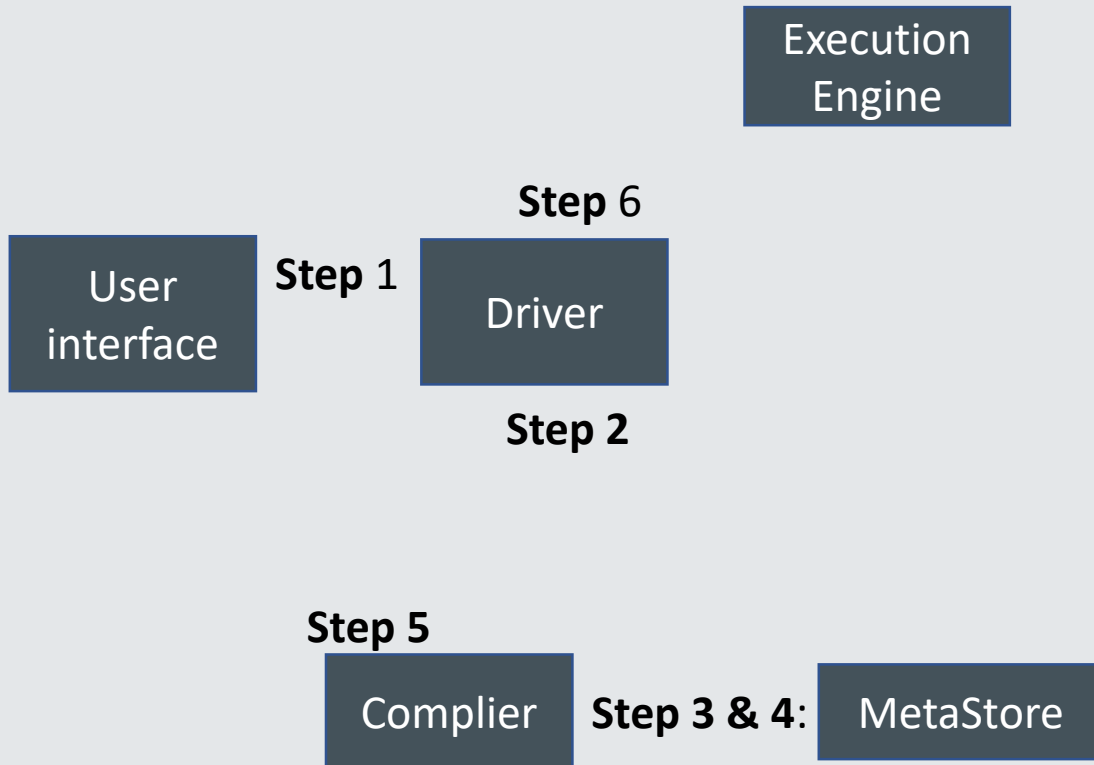
Step 3 & 4:

MetaStore

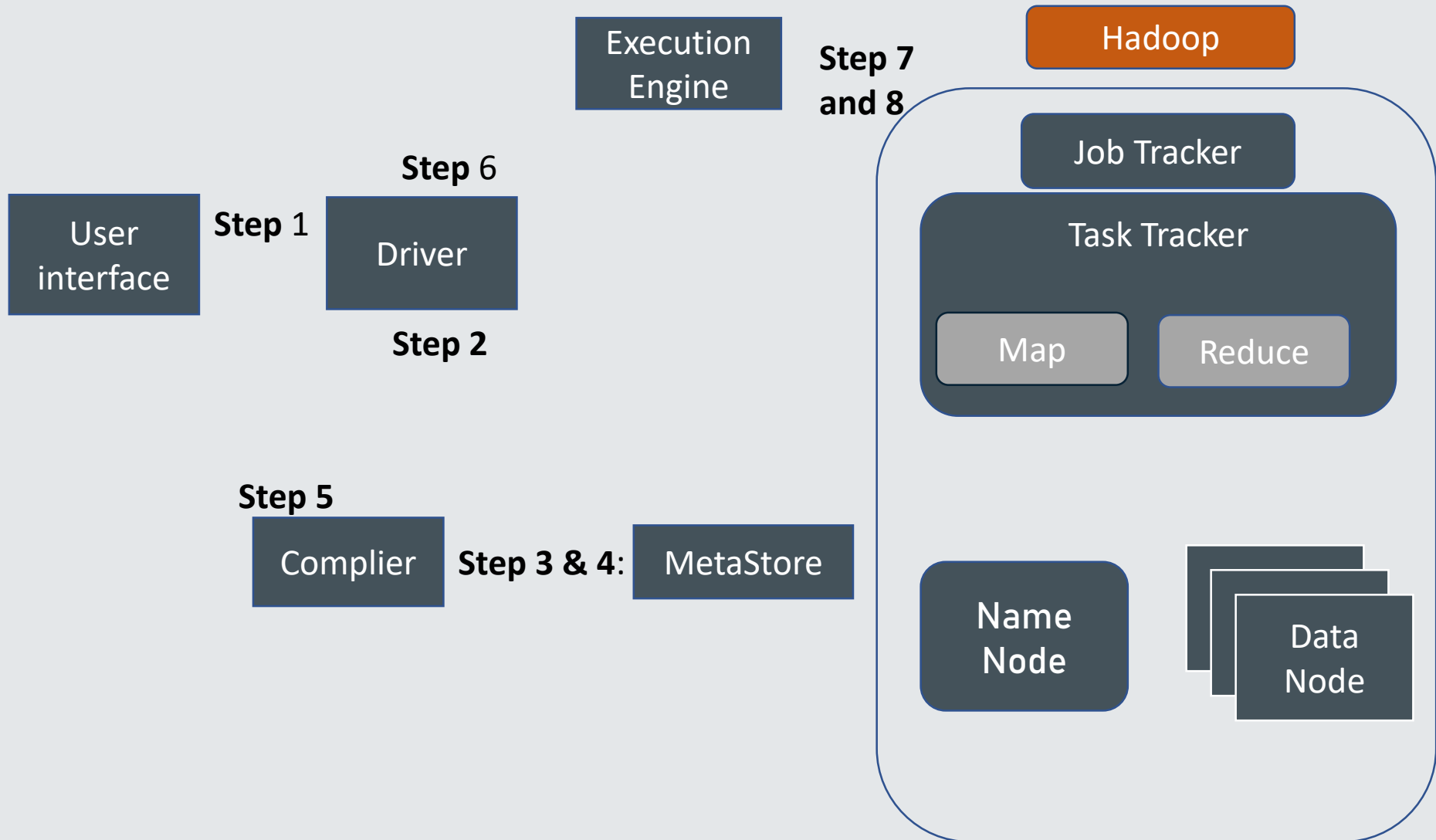
Working of HIVE



Working of HIVE



Working of HIVE



What is Hive in Hadoop?

- Submit SQL queries and perform MapReduce jobs.
- Hive has its own language, called HiveQL (HQL).
- It is similar to SQL.
- It translates SQL-like queries into MapReduce jobs.
- Don't need to learn Java to work with Hadoop Hive.



That's all for now...