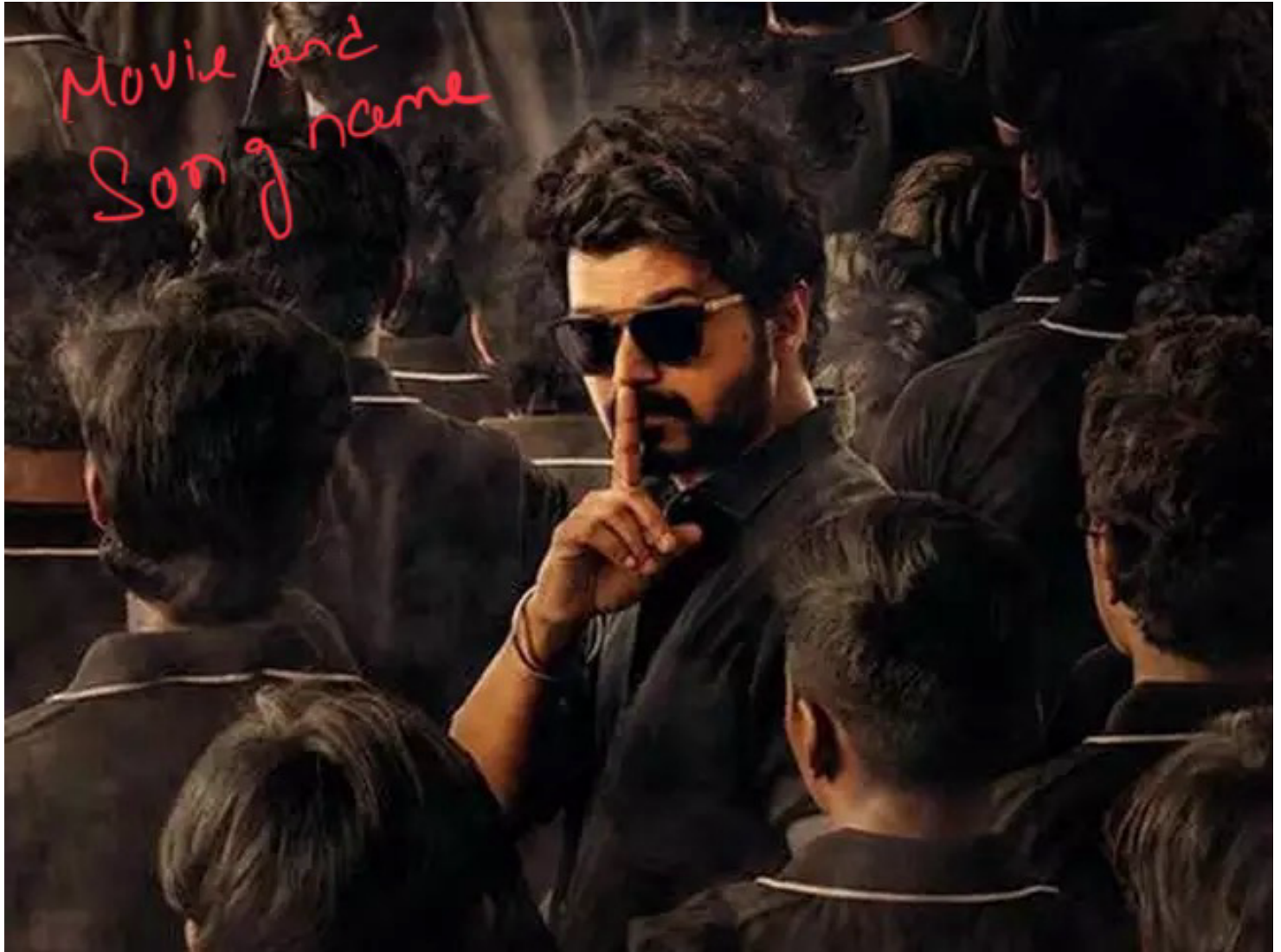
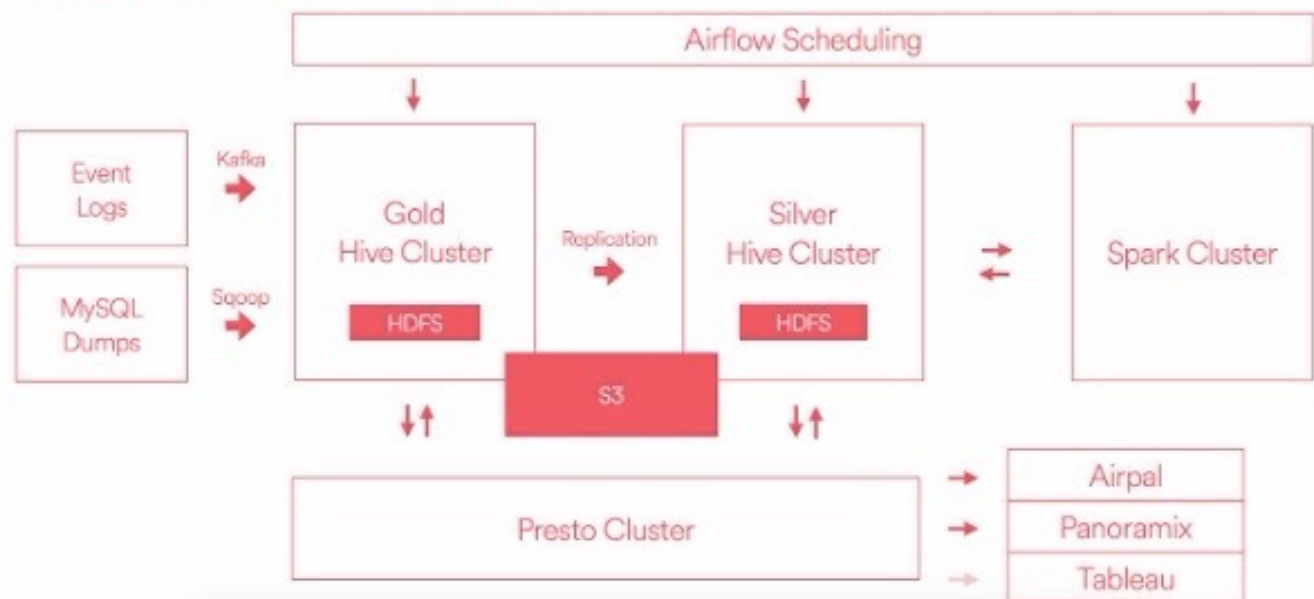


Agenda

- Problem statement
- Existing Solution
- What are data warehouses and what is OLAP?
- How does Apache hive act as a data warehouse?
- What makes hive as Hive? (Architecture)
- How to interact with Hive?
- What are different HQL commands that are used commonly?
- What are the optimization techniques in Hive ?
- Use cases of hive/ where to use hive/why and why not



AIRBNB DATA INFRA



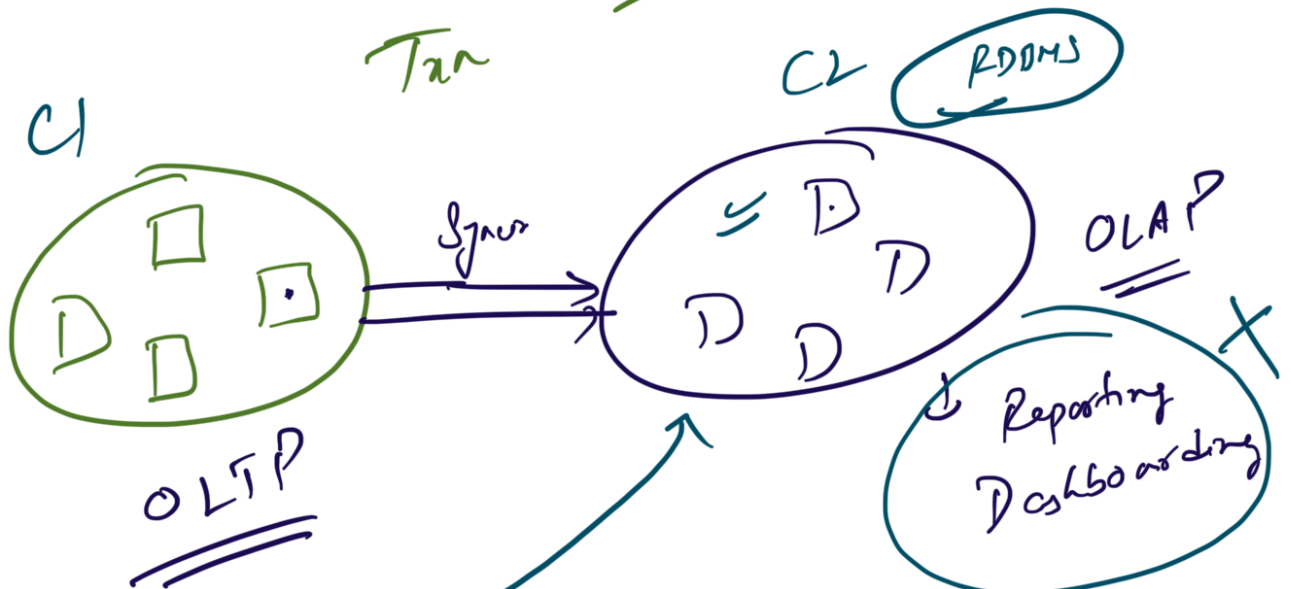
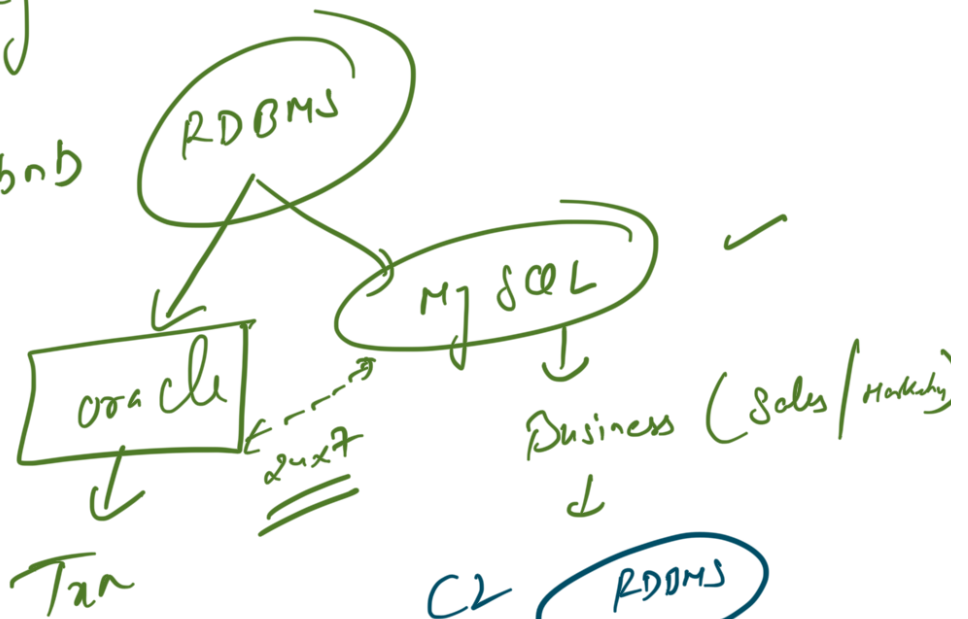
Problem statement:-

⇒ what kind of airbnb warehouse?

⇒ why they choose Apache Hadoop as DWH?

In 2014,

Airbnb



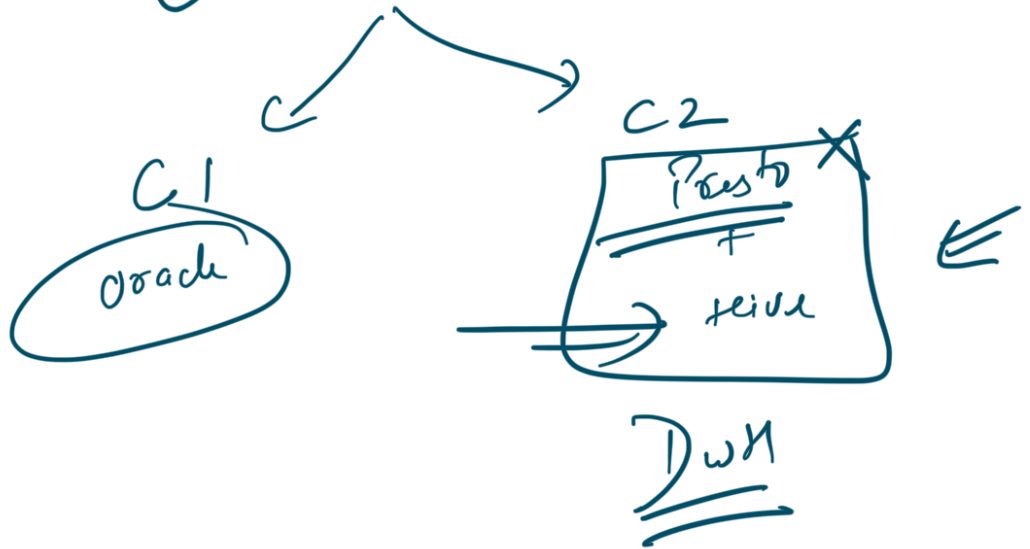
Issues:-

- (a) Reporting Jobs becomes too long.
- (b) RTO (Request Timeout) ... to be deployed.

✓ (c) ML Models

Solution :- ① Oracle in place of MySQL X

② cluster bound



Apache HIVE

Open Source World

Spark

ASF (Apache S/W Foundation)

Code + Product for free

Product + Services

DB

Cassandra

DP

Spark

Hadoop

→ Databricks

→ Databricks

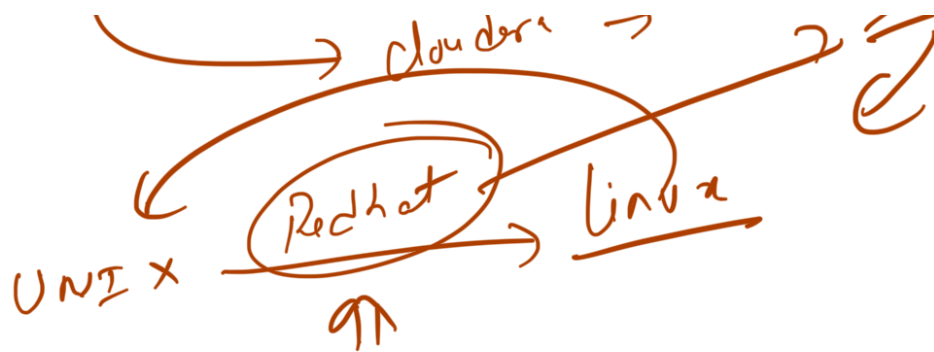
→ EMR

Data pool
HDI in Sigla

AWS
GCP
Azure

IBM

(+M)



Hive (DWH)

- ↳ Facebook around 2007
 - ↳ top level Apache Project
 - ↳ Hive QL to query structured data
- ↓
- S QL

Architecture

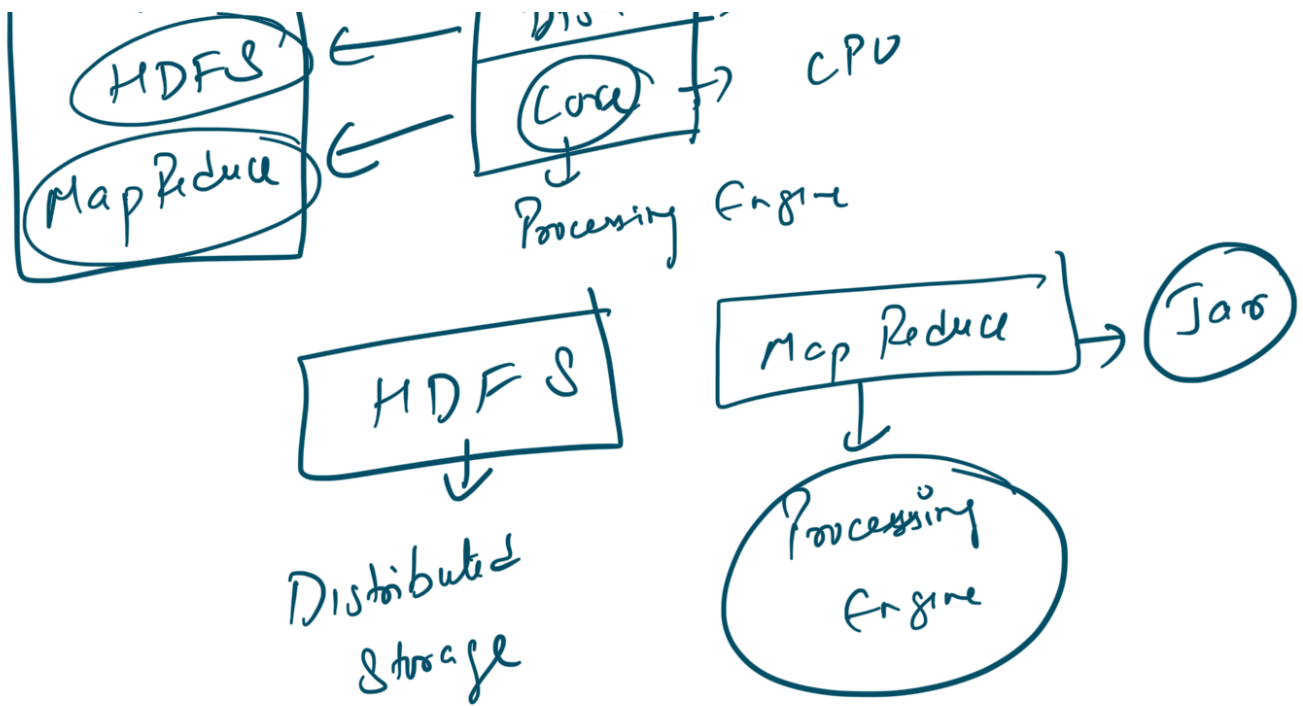
① Hadoop Core Components ✓

② MetaStore

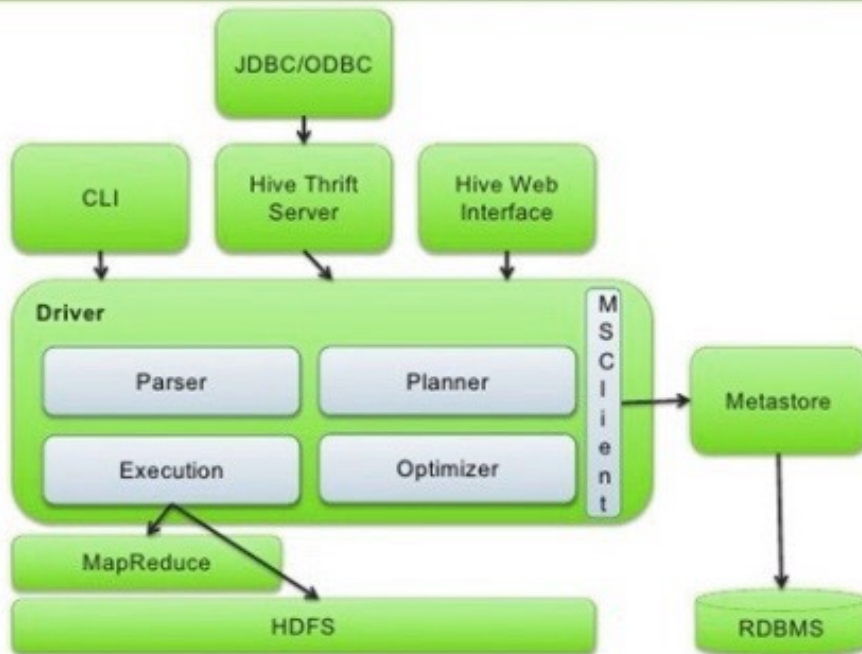
③ Driver

④ clients





Apache Hive Architecture



Steps to get answer

- ① Create a database and create a table inside it.
 - ② Use database and load the data
 - ③ Create a query and execute it.
 - ④
- ⇒

