




F1



- ① Create Database
- ② Create Table
- ③ Load data in HiveTable



→ Installed on some machines

- ① Create DB
- ② Create Table
- ③ Load Data

↳ CPU
↳ RAM

↳ external Memory



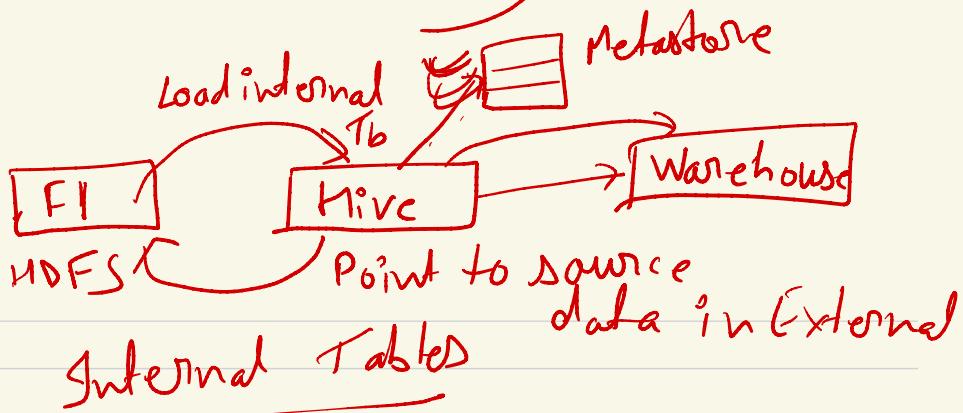
Table

Internal
(Managed Table)

External

↳ Metadata will be maintained

Metadata will be maintained by Hive
 Actual data will be at source location
 Actual data will be stored in Hive Warehouse

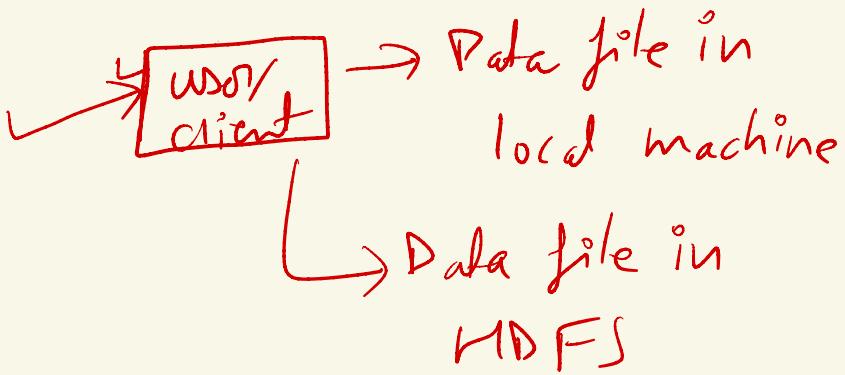


HDFS → load in Hive → Pumpin' Warehouse

Load Data in Hive

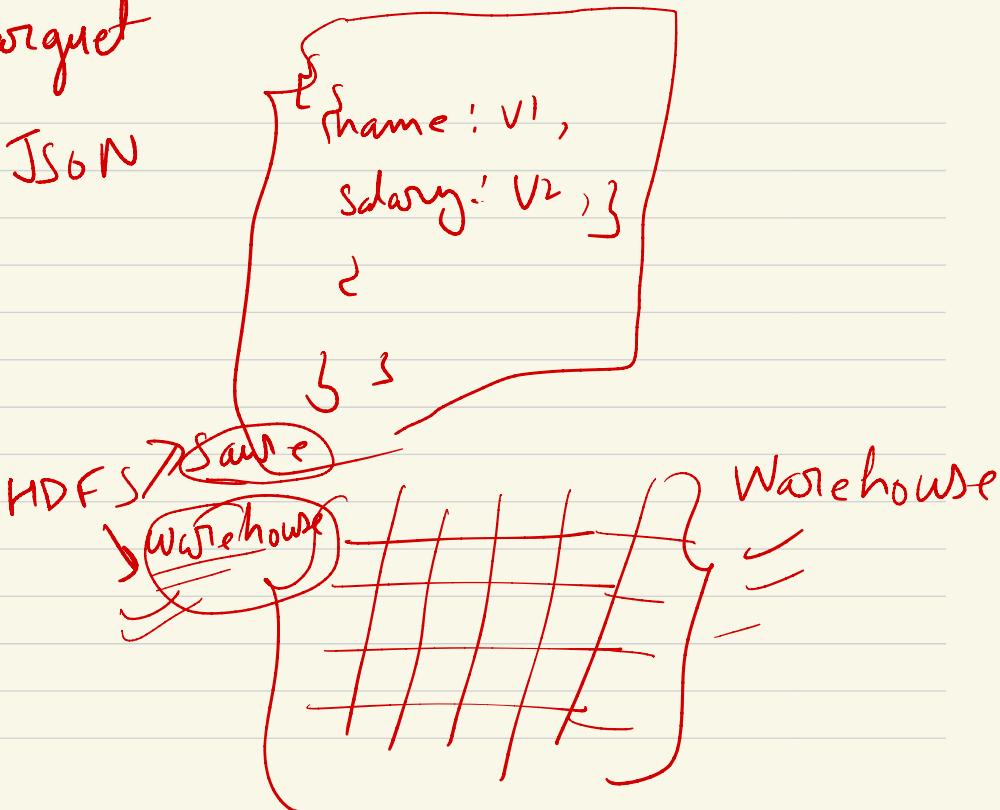
- ① Load from local
- ② Load from HDFS

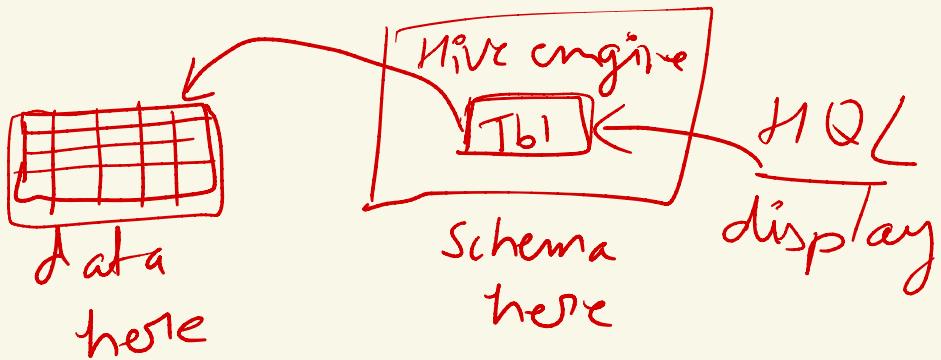
MapReduce
Cluster



✓ Parquet

✓ JSON





Internal Tables \rightarrow Metadata in hive
 \hookrightarrow Original/Kaw data in hive

External Tables \rightarrow Metadata in hive

~~arr~~ Array $\rightarrow [1, 10, 20, 40]$

map \rightarrow Key : Value

~~m~~ = {
~~"age"~~ : 21,
~~"gender"~~ : 'M'

arr[9 index]

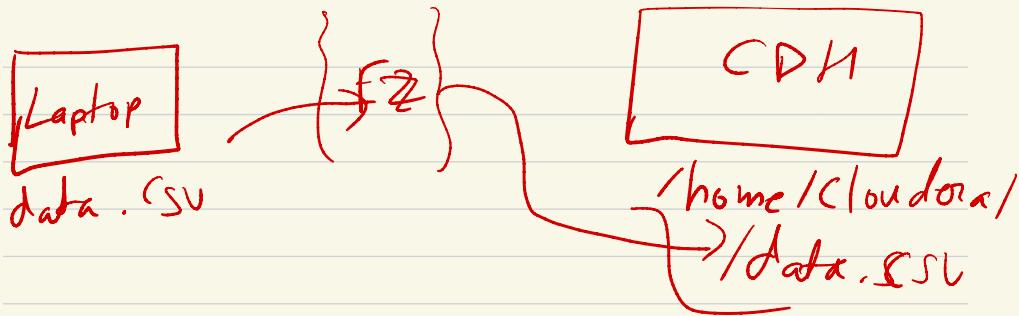
10 elements arr

arr[0] → 1st

arr[8] → 9th element

→ m1["age"] = 21

Output



load - - - "

file:///home/CloudData/data.csv

cp file:///home/CloudData/data.csv /tmp/

/tmp/hive.dbs/data.csv

file:///tmp/hive-class1/data.csv