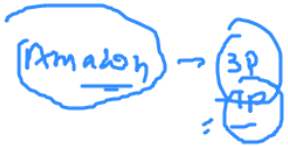
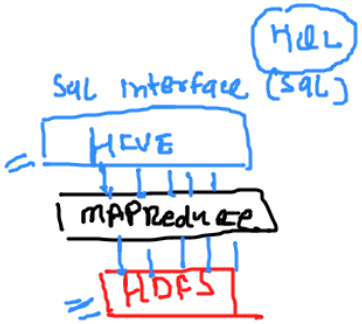


Hadoop 2.x

1. HDFS → Storage
2. YARN → Resource Managment
3. MAP Reduce → Proming of the data

.jar



ACID

RDBMS.

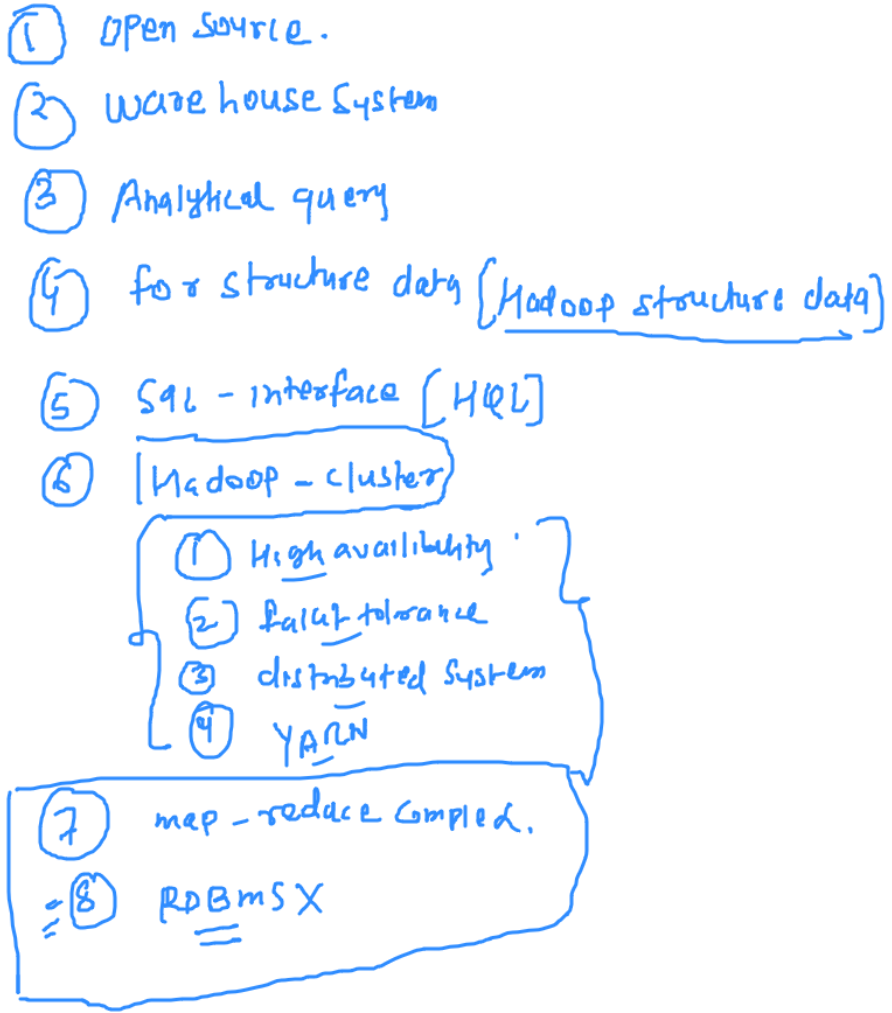
CURD

Hadoop HDFS

= map reducing

Java

Select * table name;


1 Open source.
2 Ware house System
3 Analytical query
4 for structure data [Hadoop structure data]
5 SQL - interface [HQL]
6 Hadoop - cluster
1 High availability
2 fault tolerance
3 distributed System
4 YARN
7 map - reduce Compiled.
8 RDBMSX

RDBMS

Schema

table structure

RKC

Hive property

1) Schema flexibility

2) table we can divide into bucket and partition

3) Hive tables are directly define into HDFS

directory → demo.csv

Create our schema

4) diff driver support is here.

HIVE ↔ Python

RDBMS

million = 10 lakh

latency

N. A S H

RDBMS

= indexing

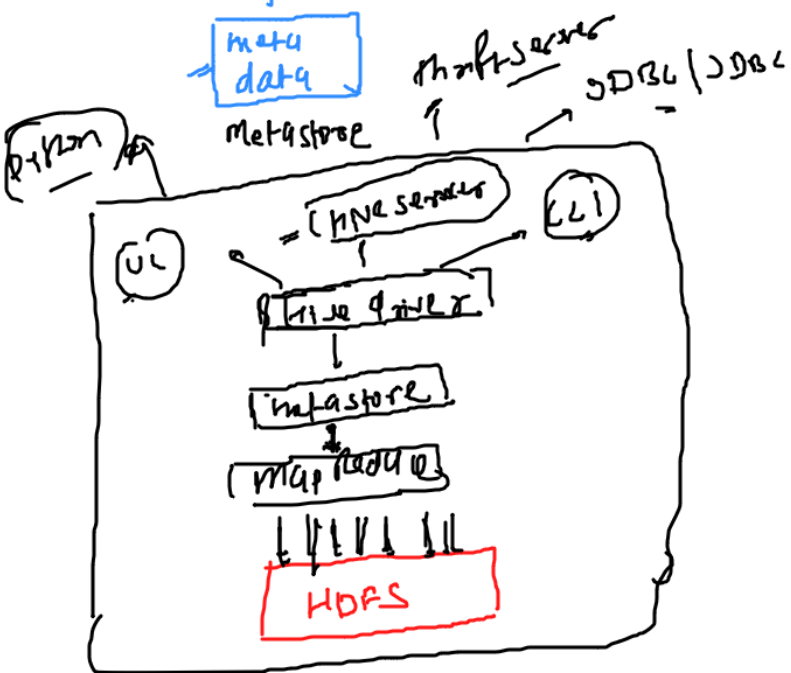
H → 5, 6, 7

Groups

HIVE Architecture.

(1) High level.

(2)



1. metastore → Store metadata [table → schema, location]
2. driver → HIVE → Hdoop (Hal) drive
3. Compiler - hive query to mr jobs.
4. Executor -
5. CLI | UI | thrift server

Datatype

Data type in Hive

Premetive datatype

1. Numeric → int, float
2. Datatime → any time/date stamp
3. String → 'c' 'sunny' → Char.

Complex datatype

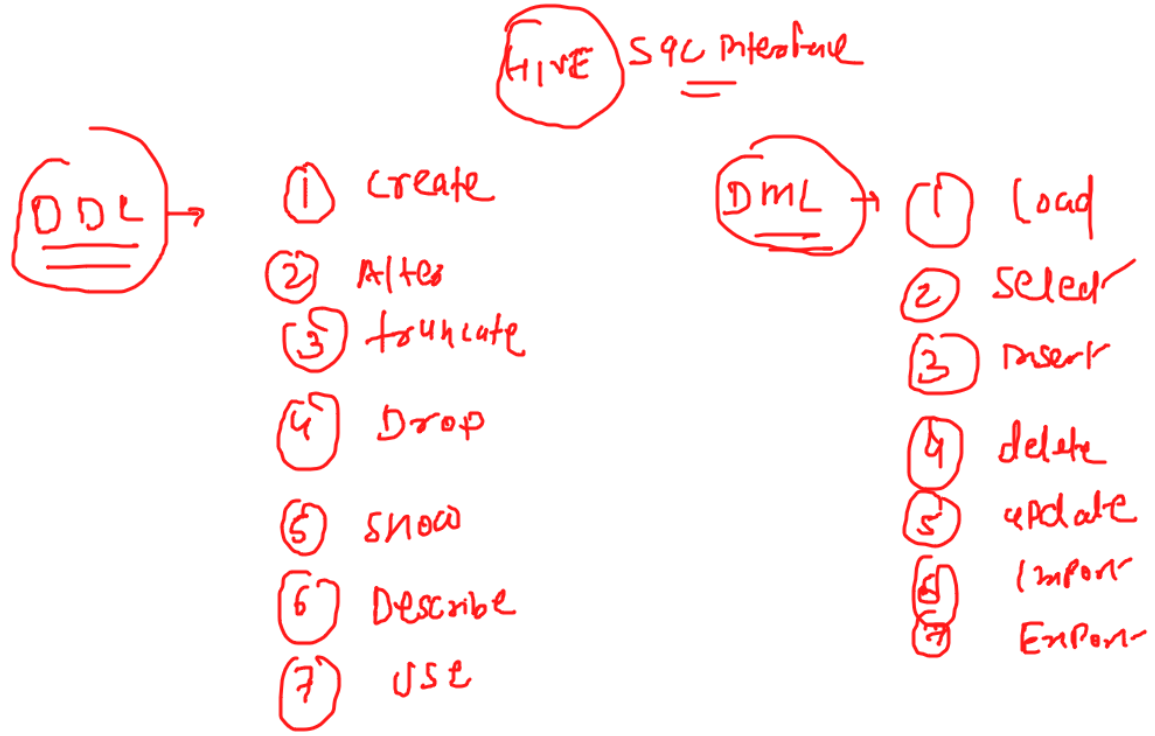
1. array [ ]
2. map
3. union
4. struct

operator

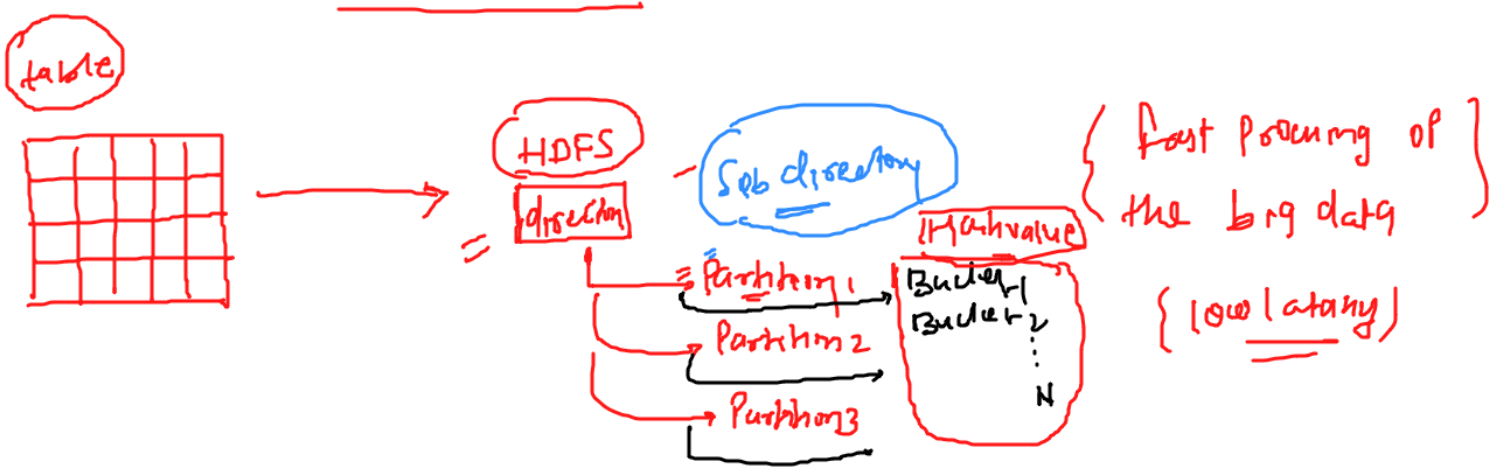
1. Relation operator
2. arithmatic op.
3. logical op.
4. string oper and op. on complex type

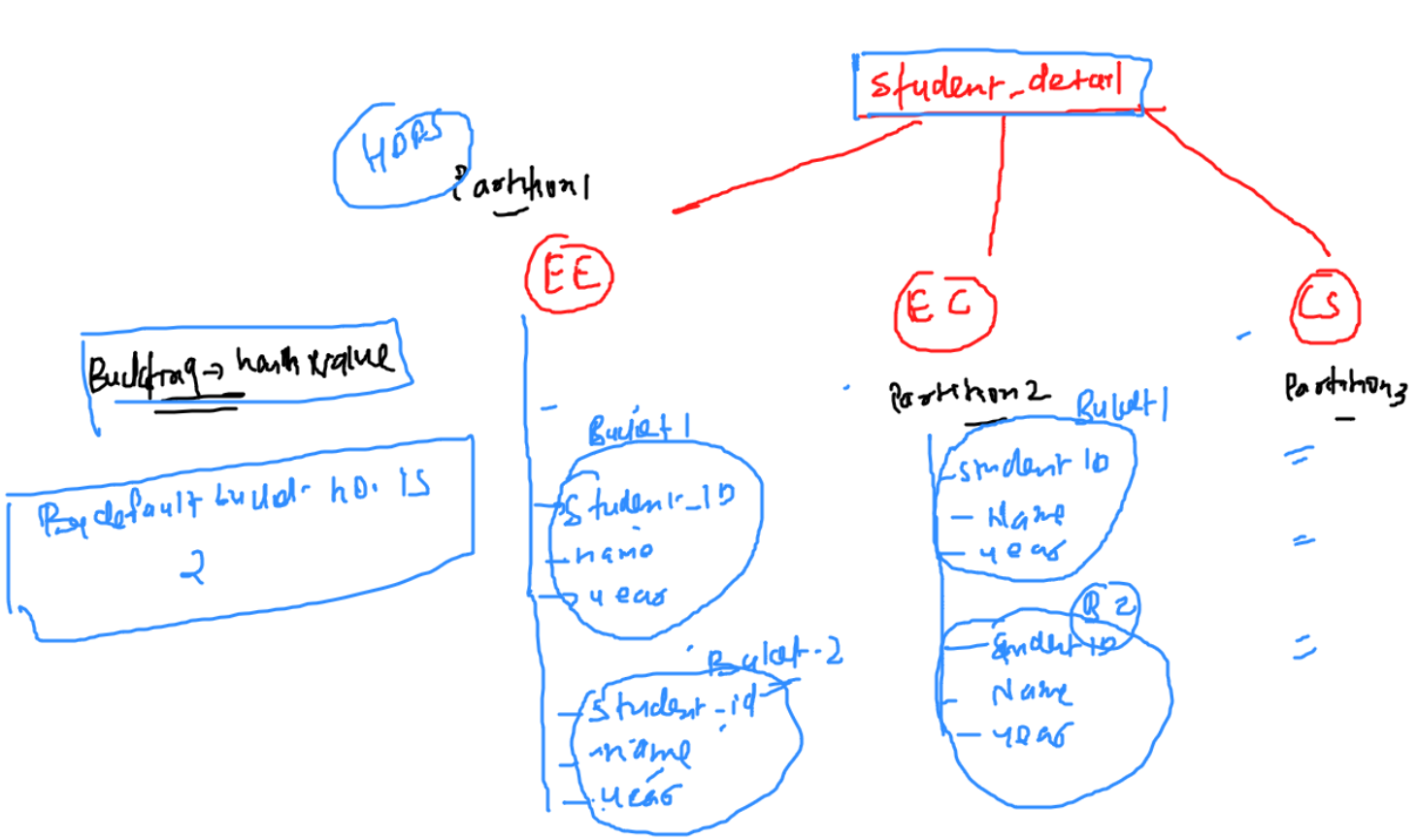
function

1. Builtin function
2. UDF


HIVE
SQL Interface
DDL
1 create
2 Alter
3 truncate
4 Drop
5 show
6 Describe
7 use
DML
1 load
2 select
3 insert
4 delete
5 update
6 Import
7 Export

Hive data model




student_detail
HDFS
Partition1
EE
EC
CS
Partition2
Partition3
Bucketing → hash value
By default bucket no. is 2
Bucket1
Student_ID
name
year
Bucket-2
Student_id
name
year
Bucket1
student ID
Name
year
R2
Student ID
Name
year
500
60
60
60
60
60

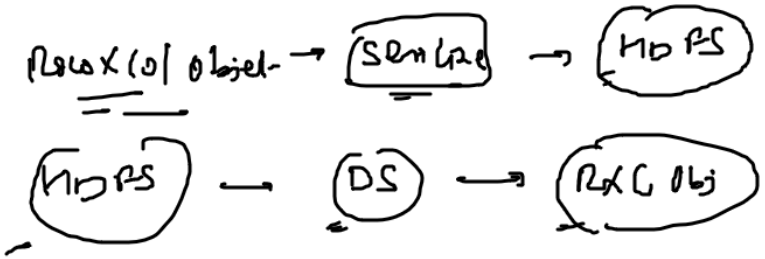
Joins

1. Inner Join
2. Full Join
3. Left Join
4. Right Join
5. Map Join
6. Bucket Map Join
7. Skew Join
8. Sort merge bucket Join

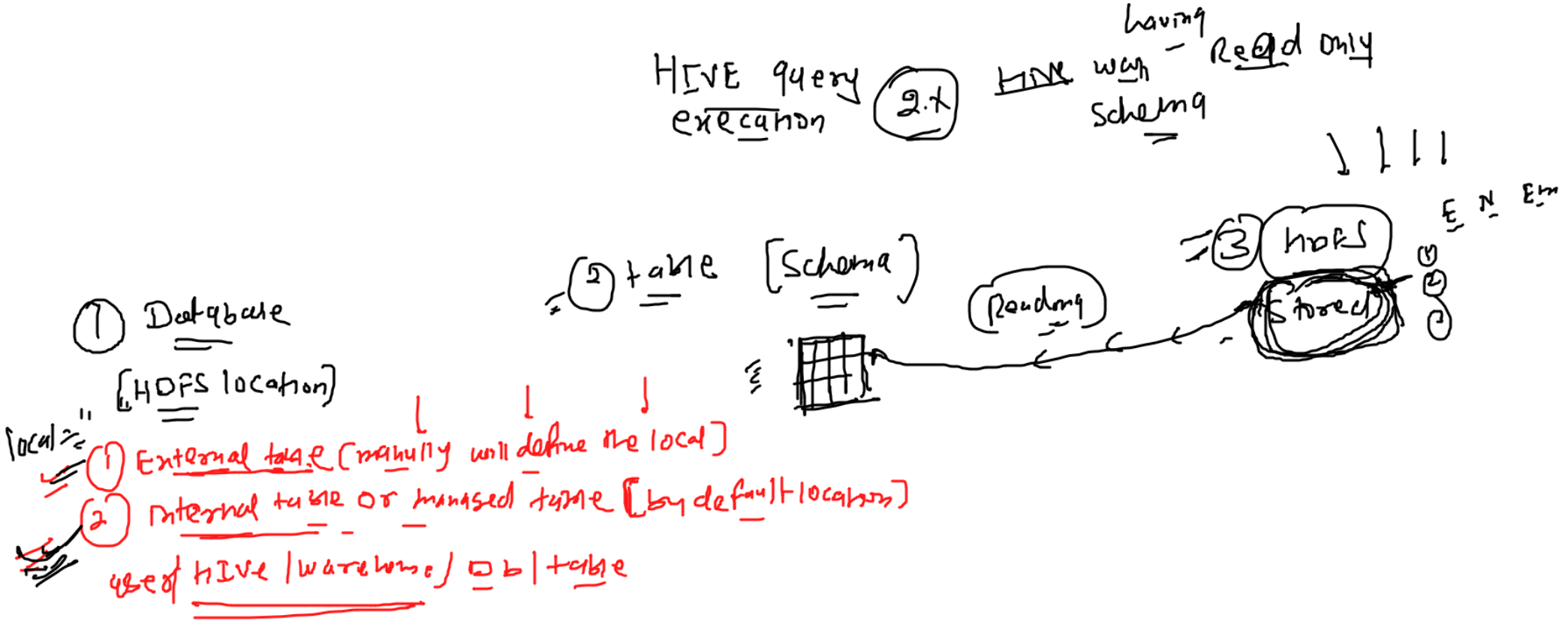
Serde of:

Serialization

de-serialization



1. textbroad serde
2. JSON_bane serde


HIVE query execution
2.x
HIVE with Schema
having Read only
3
HDFS
Stored
Reading
table [Schema]
1 Database
[HDFS location]
local
1 External table [manually will define the local]
2 Internal table or managed table [by default location]
use of HIVE / warehouse / db / table