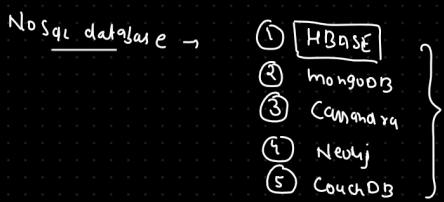


HBASE

Hadoop, Hive, NoSQL database, Streaming tools, Spark, Airflow, Cloud services

- ✓ Hadoop →
 - ① Storage (HDFS)
 - ② Processing (mr)
- ✓ HIVE → ③ SQL (HQL)



Data

- ① Batch data
- ② Streaming = minibatch data

Batch data
historical
1 week, 1 month, 6 months.

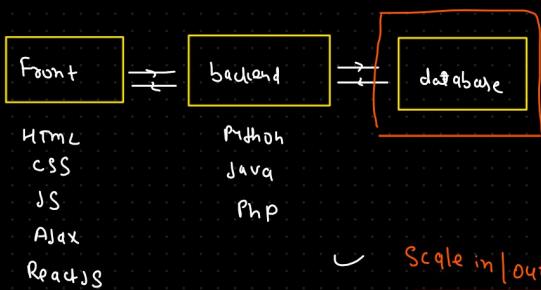
1 day, 1 hour,

Streaming → kafka
flink
storm

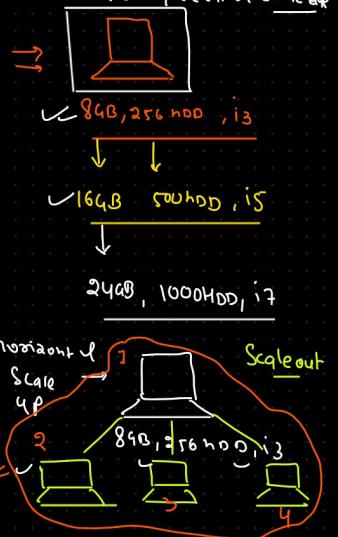
- Spark →
- ① machine learning → Python, Java, Scala, R
 - ② SQL →
 - ③ Graph →
 - ④ Streaming →

- = ① NoSQL vs SQL
- ② types of nosql ,
- ③ HBASE →
 - HBASE Data model
 - HBASE Architecture
- ④ hbase
- ⑤ HBASE vs HDFS
- HBASE vs HIVE

- 70-80
- ≤ 80
- 2000
- 1 flat file system ← XML, CSV, txt, TSV, JSON, PDF
- 2 RDBMS (OLTP) ← MySQL, Oracle, PostgreSQL, SQLite
- 3 OLAP (Warehouse) ← HIVE, Pig, Tez
- 4 NoSQL database



Scale in | out → Scale-in / Vertical Scaling



SQL

- ① SQL → S-A-L
- RDBMS (relation)

Non-SQL

- non-SQL (SQL)
- non relational
- Distributed

- ② table

document

key-value

graph based

- ③ ACID

CAP

- ④ Vertical

horizontal

Scaleout

Scaleup

⑤ Predefine Schema

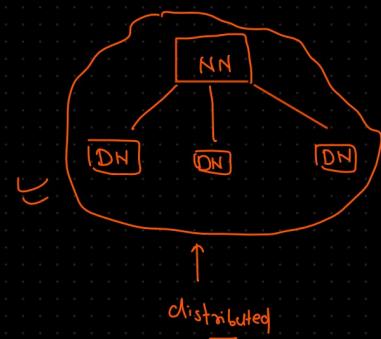
- MySQL, Oracle



⑥ Dynamic Schema

cluster ↑
Distributed

hadoop
spark
nosql

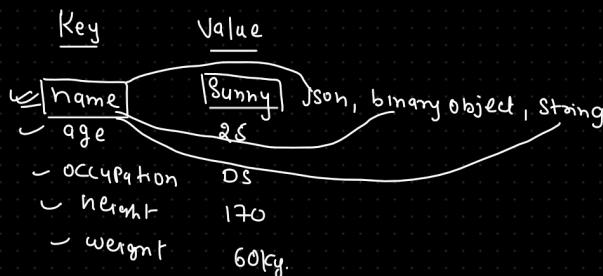
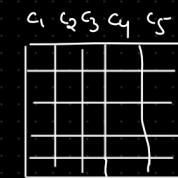


types of no-SQL DB

- ① key-value pair → Redis, dynamo, Riak,
- ② column-oriented → HBASE, cassandra, hbase
- ③ document based → CouchDB, mongoDB, Amazone Simple DB
- ④ graph based → Neo4j, OrientDB, flockDB

\approx  , \approx 100 TB

RDBMS



② Column Oriented DB

(unique key)

Rowkey	\approx	
	Col1	Col2
1	s1s2s3s4	s2s3s4
2		
3		
4		
5		

Column oriented Schema:
HBASE

③ Document → Cloud version (Atlas)

(mongoDB compass)

\approx JSON = { height: 170 ← JSON
 or weight: 80 ←
 XML BMI: 20 } }

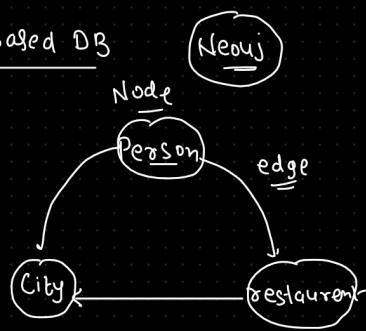
{ height: 180
weight: 90
BMI: 21 }

{
}
}
}

Height	Weight	BMI
170	80	20
180	90	21
190	100	22
160	120	23
165	70	19

XML
{
<head>
</head>
<div>
</div>

④ Graph based DB



NoSQL → ① non-relational

② Schemaless

③ Simple API (Python, Java, Server)

④ distributed

⑤ not following ACID

MySQL

Height	Weight	BMI
170	80	20
180	90	21
190	100	22
160	120	23
165	70	19

→ Select * from table name;

⑩ → 0 mongod

{ height : 170

weight : 80

BMI : 20 }

{ height : 180

weight : 90

BMI : 21 }

Big table (google) 2006

DB[0] DB[height]

2007 HBASE → hadoop Project.

2007 → HBASE → 0.15

2008 → 0.8, 0.19, 0.20

2010 → HBASE

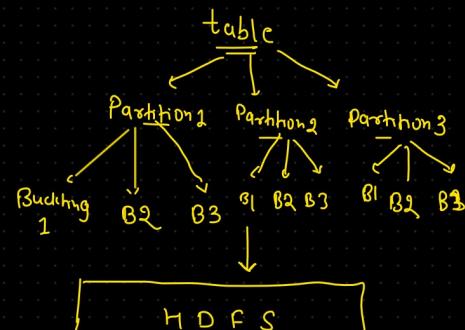
Data model

RDBMS

MySQL

Height	Weight	BMI
170	80	20
180	90	21
190	100	22
160	120	23
165	70	19

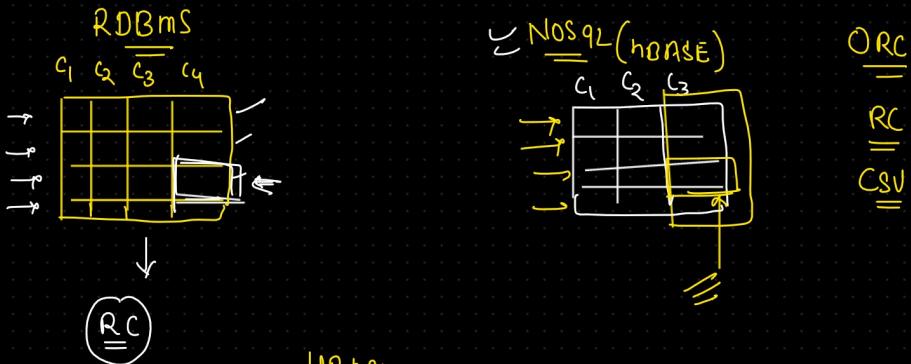
HIVE data model



HBASE Data model

Column Oriented DB

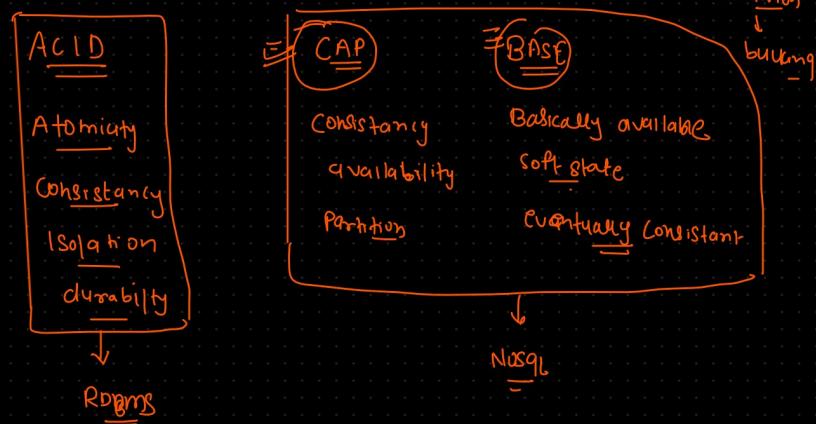
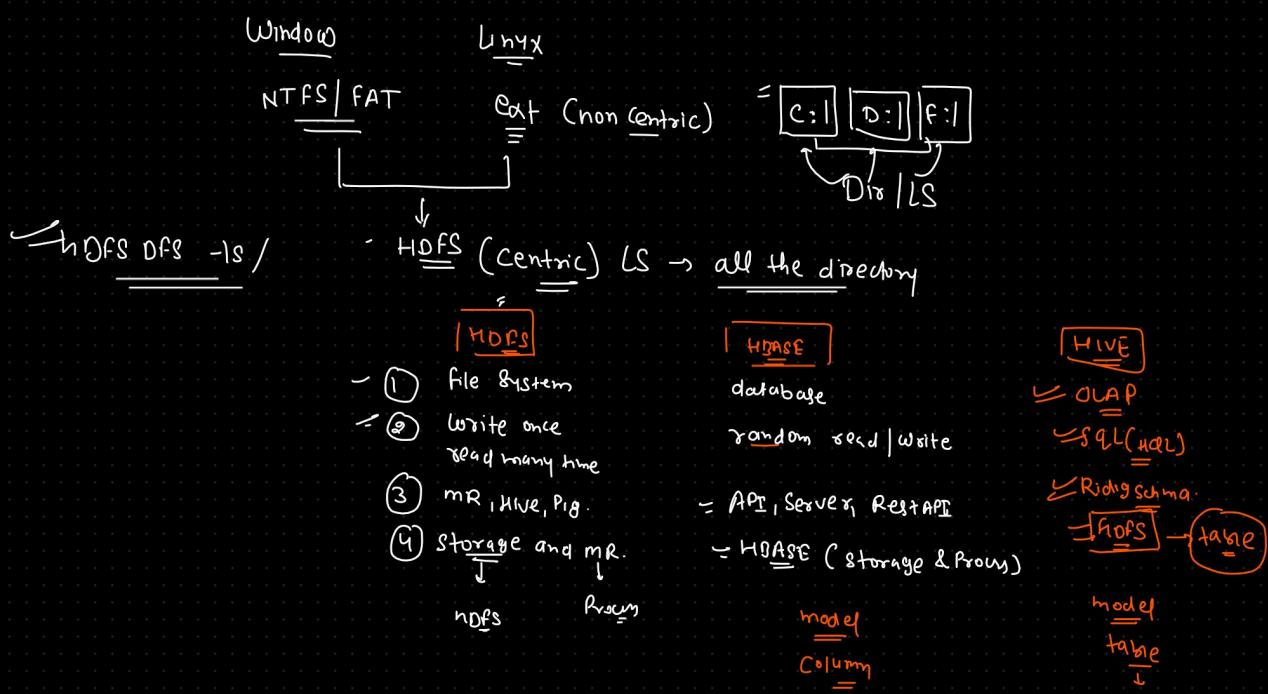
Rowid	Personal detail		Office detail	
	Name	Contact No.	Eid	EmailId
01	Sunny	8770xxxx	IN0075	SS@G
	Swapnil	70245678	IN0077	X@G



HBASE vis HDFS
HBASE vis HIVE

Architecture

data model mean how internal structure of the data storage of application/database ?



Architecture of HBASE

- ① HMASTER
- ② HRegion Server
- ③ Hregion
- ④ Zookeepers
- ⑤ HDFS

