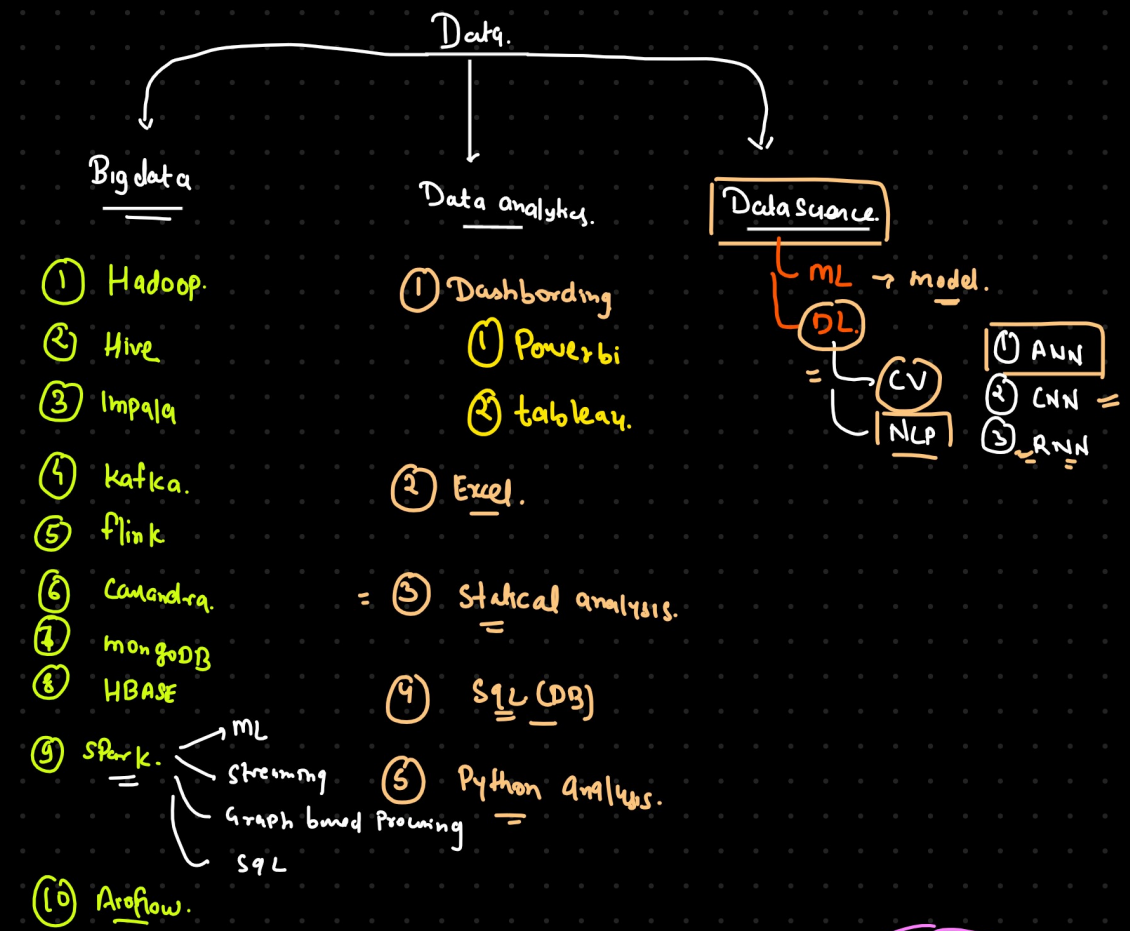
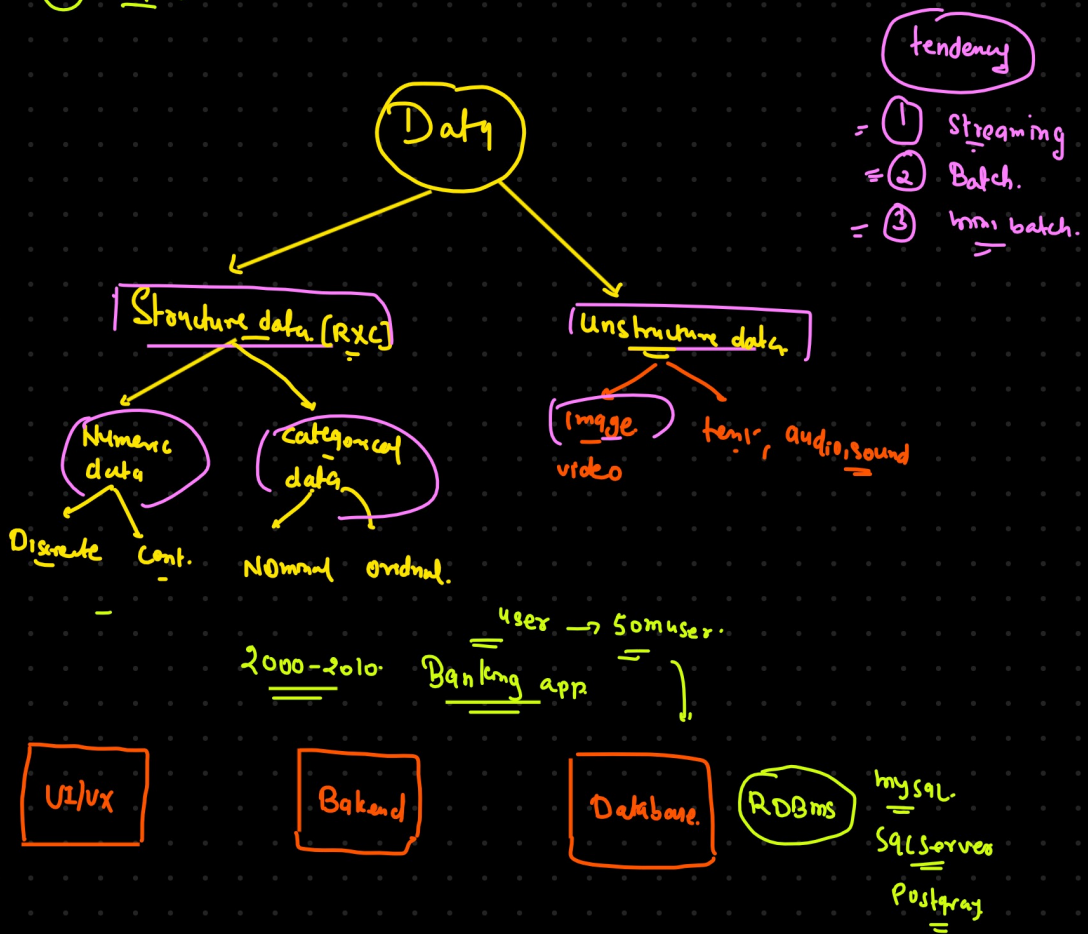
Big - data.

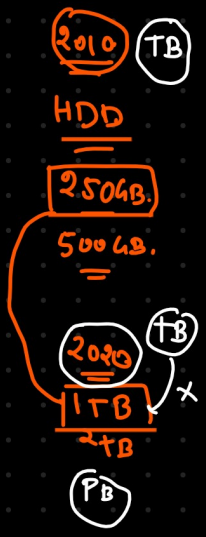

Data.
Big data
Data analytics.
Data science.
1 Hadoop.
2 Hive
3 Impala
4 kafka.
5 flink
6 Cassandra.
7 mongoDB
8 HBASE
9 Spark.
ML
Streaming
Graph based Processing
SQL
10 Airflow.
1 Dashbording
1 Powerbi
2 tableau.
2 Excel.
3 Statical analysis.
4 SQL (DB)
5 Python analysis.
ML → model.
DL
CV
NLP
1 ANN
2 CNN
3 RNN


tendency
1 Streaming
2 Batch.
3 mini batch.
Data
Structure data (RXC)
Unstructure data
Numeric data
Categorical data
Discrete
Cont.
Nominal
Ordinal
Image
video
text, audio, sound
user → 50 user.
2000-2010.
Banking app
UI/UX
Backend
Database.
RDBms
mysql.
SQL Server
Postgray

Big data.

1. Massive collection
2. Increasing exponentially.
3. Complicated.
4. Store, manage and processing is very difficult.

KB
MB
GB
TB
PB
ZB
EB
YB



5V →

1. Variety. → Structure, Semi-structure, Unstructure. (70-80)
2. Velocity. → Speed / rate at which data being generated.
3. Volume → massive volume of data
4. Veracity. → trustworthy.
5. value → valuable and reliable.

Batch data → Periodic data. [1 week, 1 month, 6 month, 1 year]

mini batch data. → less frequent [1 week, 10 day, 5 day, 2 day].

Streaming data. → very frequent data [less than second (live) within second]

1. Big data def
2. 5 'V of big data.
3. type of data
4. tendency of data.

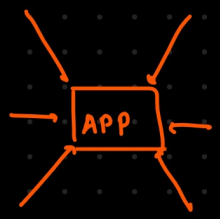
Evaluations of the tools.

1. file system.
2. DBMS (RDBMS)
3. Data ware house
4. Data lake

DFS (Hadoop)

File System	Database.
Entire data	specific data
non secure. (No Role)	secure (Role)
not able to handle.	handle.
not able to handle	handle

1. Size.
2. Security.
3. Concurrency.
4. Redundancy



OLTP

Database.

table: ice ✓ table: cust ✓ table: sales ✓

(SQL). transactional.

1. update
2. Drop

ETL

1. Extract
2. transformation.
3. Load

OLAP

Data ware house [tools]

(repository.)

= Analytical.

Insight

Data+LAKE

ETL → Extract, transform, load.

RDBMS → HDFS → Spark → HBASE

= E T L

1. Data industry.
2. introduction of Big data.
3. file system.
4. Database
5. DW and DL

Hadoop → DFS

or

HDFS

1. History of hadoop
2. HDFS Architecture.
3. HDFS → storing the data.
4. mapreduce → Procuing of data.
5. component of hadoop, HDFS
6. HDFS/LFS
7. YARN
8. Hadoop 1.x vls 2.x
9. all the component of YARN
10. Read and write
11. Rack awarnes.
12. HDFS commands.

1. HDFS.
2. HIVE
3. kafka
4. SPark
5. Airflow

Project

6-7 demo