

Big data \Rightarrow

data \Rightarrow raw format \rightarrow audio, video.

Big data \Rightarrow mini Tb $\&$ even like PB, TB

① Small data

GB, KB, MB

speed \downarrow

less complex

structured

Big data.

PB, TB

speed \uparrow

High complex.

Unstructured.

Hadoop \rightarrow developed at MIT \rightarrow GPS.

Google file system

Technologies
and \rightarrow

language used \rightarrow java.

Application of Hadoop \rightarrow data analysis.

data cleaning \rightarrow

Hadoop Tools \rightarrow map reduce

cluster formation \rightarrow

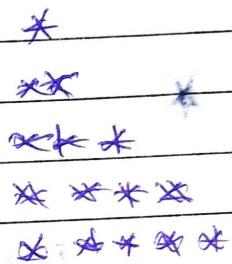
HDFS

```

public class A {
    public static void main (String [] args) {
        int n = 5;
        for (int i = 1; i <= n; i++) {
            for (int j = 1; j <= i; j++) {
                System.out.print ("*");
            }
            System.out.println ();
        }
    }
}

```

PQL queries →
 DDL ✓
 DML ✓ **group**



Delete @ Drop

Particular

row or
column wise

delete query

Pure Table wise database wise etc

Create database employee →

CREATE employee. DB

① Create database

Rows, Column define

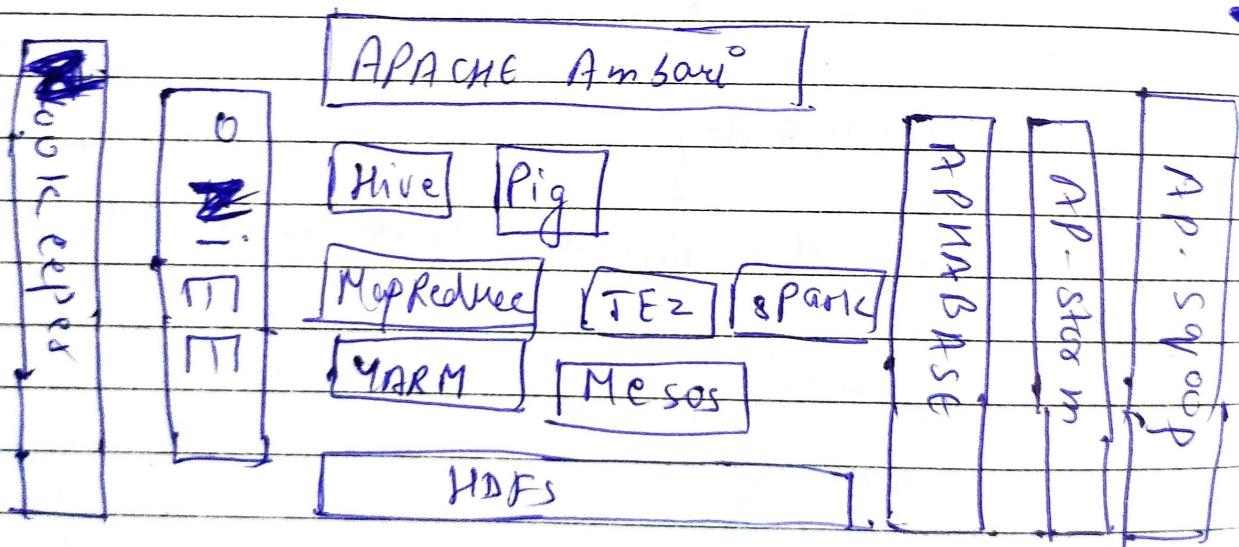
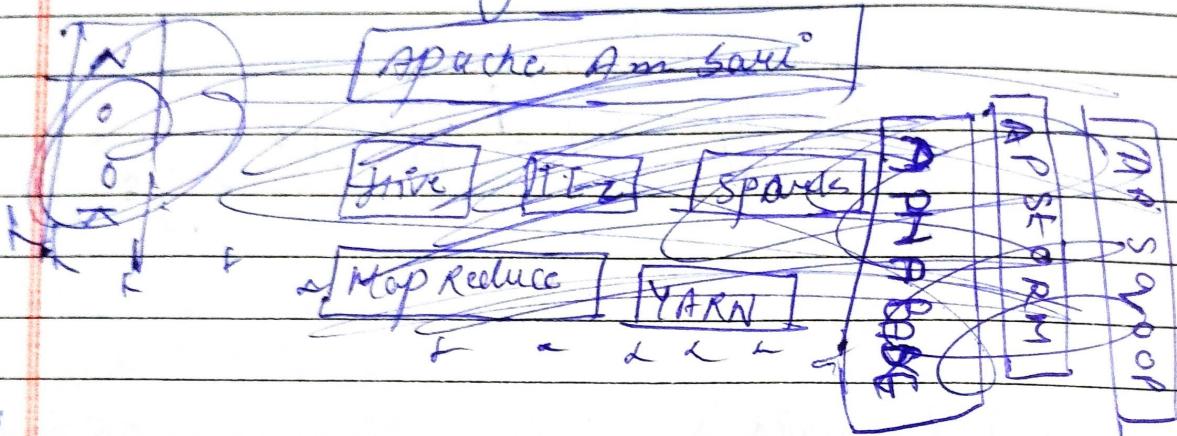
② Import operation : data set or import or export of RDBMS → ~~DBMS~~

UDF's

② Perform Analysis → MapReduce
 ③ Export
 HDFS → RDDMS (F)

Page No.:
 Date: / /

Architecture of Hadoop



① HDFS → formed clusters to store database.

② YARN → Yet Another Resource Manager

→ manage all the resources.
 → heartbeat of hadoop system.

③ HDFS → Alternative of %

④ MapReduce → firstly map $\xrightarrow{\text{data}}$ then $\xrightarrow{\text{process}}$ data

Page No.	Date
1	1

Combining $\xrightarrow{\text{data}}$ $\xrightarrow{\text{data}}$,
reduce.

⑤ Pig → scripting languages
syntax / code $\xrightarrow{\text{white space}}$ $\xrightarrow{\text{Javadoc}}$
 $\xrightarrow{\text{script}}$, automatically
code write $\xrightarrow{\text{script}}$

⑥ Hive → SQL queries with built-in
functions & HiveQL $\xrightarrow{\text{script}}$
et al. $\xrightarrow{\text{script}}$ isme.
342)

⑦ TEZ → graphical representation $\xrightarrow{\text{script}}$
and $\xrightarrow{\text{script}}$

Spark → $\xrightarrow{\text{script}}$ kind of like $\xrightarrow{\text{script}}$
syntax / code $\xrightarrow{\text{script}}$

⑧ APACHE BASE → Apache HABASE →
Storage of data $\xrightarrow{\text{script}}$ activities
process $\xrightarrow{\text{script}}$ use of different data stores $\xrightarrow{\text{script}}$

⑨ APACHE Storm → ML module $\xrightarrow{\text{script}}$
Run $\xrightarrow{\text{script}}$ $\xrightarrow{\text{script}}$ $\xrightarrow{\text{script}}$ update
real time functionality.

⑤ Apache Sqoop \rightarrow Data connector between SQL Server and HDFS.

Page No.:	5/29
Date:	1/1/2024

⑥ Oozie \rightarrow Scheduling framework for Hadoop tasks. It defines a workflow of tasks and their dependencies. A scheduler monitors the workflow and triggers tasks as needed. It can run on various systems like Linux or Windows.

⑦ Zookeeper \rightarrow State up, state down. It's a distributed system that provides coordination and synchronization services for distributed applications. It's used for maintaining configuration information, providing distributed synchronization, and managing distributed consistency.

⑧ Apache Ambari \rightarrow Overall application management. It's a platform for managing Apache Hadoop clusters. It provides a central interface for monitoring, configuring, and managing multiple Hadoop components like YARN, HDFS, and MapReduce. It supports both Windows and Linux environments.

Security concern: Data loss,