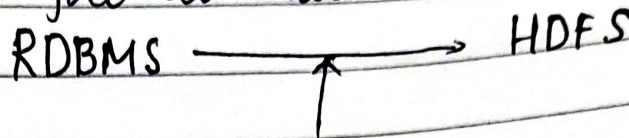


① Hadoop: It is an open source framework used for storage and processing of large datasets which is free to use and modify.



Technology & Tools - Hadoop, Hive, Spark, Pig, MapReduce

Hadoop forms interface linking between Database and HDFS.

Symbol = Yellow Elephant.

Founded By = GFS - Google File System

Creating a Database

① CREATE DATABASE Employee;

② CREATE TABLE EMPLOYEE;

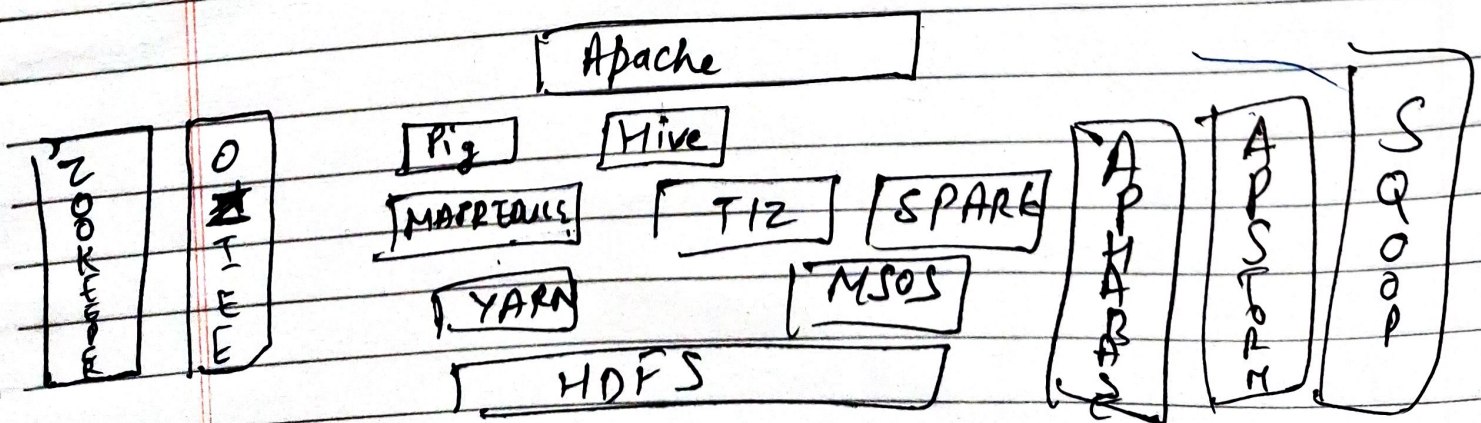
→ Table Schema defines no. of rows and columns

→ Primary key is the unique identity
emp-id

PRIMARY KEY

③ Populate the Records

Hadoop Architecture



YARN - Yet Another Resource Negotiator.
To manage resources in a PC.

MAP - Distribution of dataset / splitting

Reducer - Aggregating together all the data.

Pig - ① Scripting languages
② can add syntax

Hive - All SQL theories are run here

Graphical representation

All graphical rep. is done here

For scalar prog. language - Spark.

Apache HBase - Storage function

Apache Storm - Real-time processor.

Sqoop - SQL Relation with HDFS

OZIEE - Scheduling.

ZOOKEEPER - Keeps the record about the steps
preventing the system.
eg - State is a

Apache Ambari