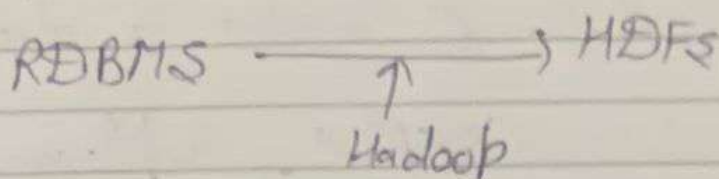07/Aug/25

Hadoop:

$$RDBMS \xrightarrow{\quad\text{Hadoop}\quad} HDFS$$

Hadoop, is an open-source framework which forms interface and linking between data sets 2 HDFs.

Technology & Tools: Hive, spark, Pig
                        MaPReduce

Symbol - Yellow Elephant
founded by → GFS
                Google file System

~~Write code for~~ python

```
import java.util.Scanner;
public class To_print_right_angle_triangle {
public static void main (String [ ] args) {
     Scanner sc = new Scanner (System.in);
System.out.println (" To print the shape of
          right_angle triangle );

System.out.println ("*");
  //      //      //     ("* *");
  //      //      //     ("* * *");
  //      //      //     ("* * * *");
  //      //      //     ("* * * * *");
  }
}
```

1.) Cicating Database
        — ① Create a database

       CREATE Employee.db;

      — ② Define Table Schema
        — emp_id
        — PRIMARY KEY (Unique identy)

      ③ Populate the Records
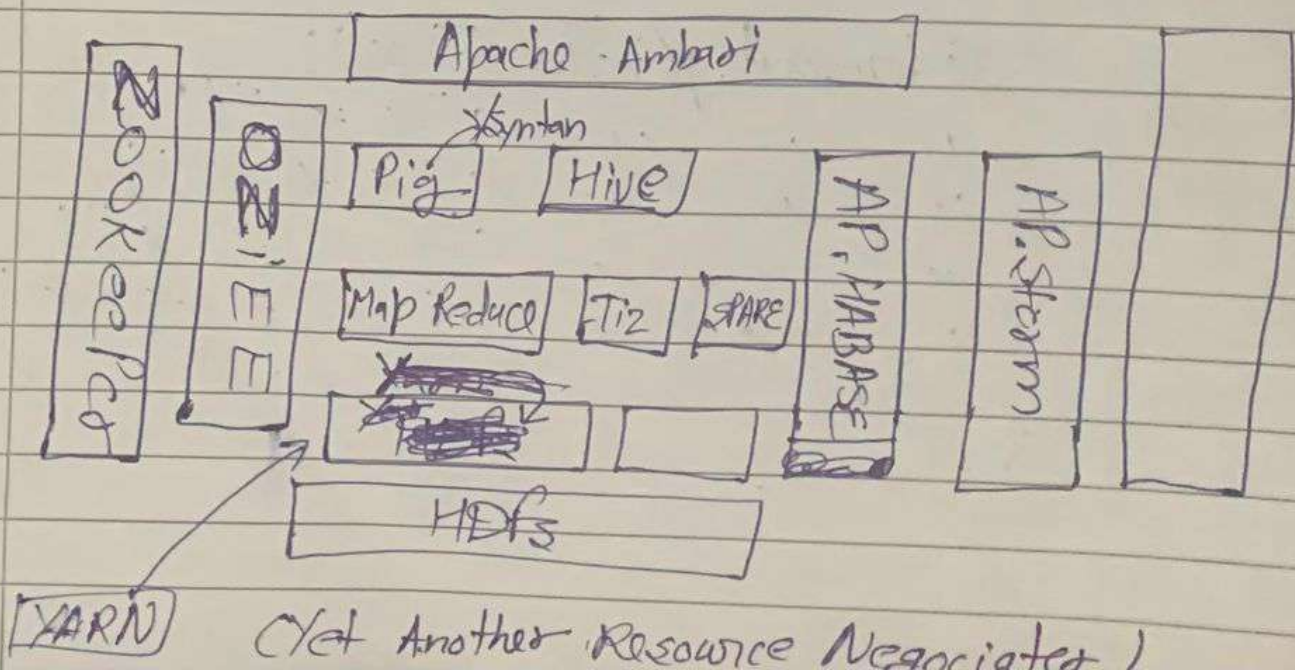   — Insertion of data

① Import operation
      Import
        ↓
  REBMS ——→ HDfs

② Analyze ——→ Result
③ Export op.

    HDfs ——————→ REBMS

Hadoop Architecture:-

ZooKeeper	ONZIEE				Ap. HABASE	Ap. Storm	

Apache · Ambari

syntan

Pig | Hive

Map Reduce | Tiz | SPARE

HDfs

[YARN] (Yet Another Resource Negociater)

Map - Distribution of dataset / spliting
Reduce - Aggreting together all the data
Hive - Sql queries Run
Tiz - Graphical Representation
Spare - Scaler programming Language
AP.HABASE - Storage Mostly
AP.Storm - Real - time procesing
Sql - HDfs
OZIEE - Scheduling
ZookeePer - ZookeePer Keeps the record
all of the state present in the
System