Green House
Date 08/07/23
Page No.

Big Data lab

Data → Raw facts & figures.

Data	Big Data
Small	
> GB, KB, MB.	> PB, TB.
> Speed ↓.	> Speed ↑.
> Less complex.	> More Complex.
> Structured.	> Unstructured.

Hadoop → GFS
Google File System.

Technology used :-

Language → Java.

Use/Application → Data analysis.

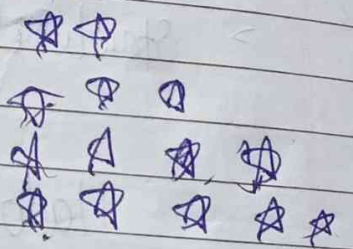
Rasp. Calculation.

Cluster Formation.

HDFS
Speed is fast.

```
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++){
                sout ("*");
            }
        }
    }
```

Output:-



SQL QUERIES :-

Create database employee;

CREATE - Employee.dbs;

Map use parallel pro

① Import

RDB

② Analysis

make

into

③ Export

HD

Had

2 QUARTER MP

1.

2.

Map use parallel processing.

1. Import

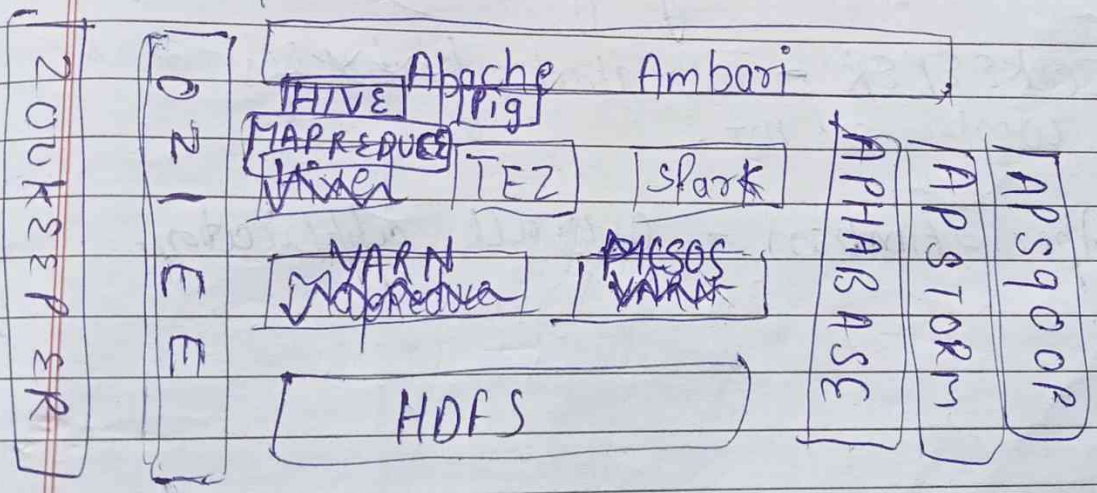
RDBMS → HDFS.

2. Analysis - Map Reduce :- Map is used to make data sets and the reduces all the data into one

3. Export

HDFS → RDBMS.

Hadoop Architecture :-



1. Clusters are formed

2. YARN → Yet Another Resource
→ Manages Resources.
→ Heartbeat of Hadoop.

PICSOS → Alternate.

3. Pig → Scripting language, API generating.
inbuilt

Green House
Date
Page No.

4. Hive - Access of SQL Queries & Run.
5. TEZ - Graphical Representation.
6. SPARK - write syntax.
7. Apache HBASE - storage
8. Apache Storm - real time functioning.
9. Apache Sqoop - Connection b/w SQL & HDFS.
10. OZIEE - Schedulling of Tasks.
11. ZOOKEEPER - Tells about system is working or not.
12. Apache Ambari - Overall application.