Big Data Lab

founder of Hadoop → Google File System

Small Data	Big Data
1) GB, KB, MB.	PB, TB
2) Speed ↓	Speed ↑
3) Complex ↓	Complex ↑
4) Structured	Unstructured

Symbol of Hadoop - Elephant

Technologies used in Hadoop —

Language used in 4 — Java.

Main fun of Hadoop - Data Analysis

Advantages →

1) Speed of Data Processing ↑
2) Good Fault Tolerance.

```
n = int (Enter no.of rows)
for i range (n):
    for j range (i).  _  .  _  .
        print (*)
```

SQL queries -

```
Create database employee;
      or
CREATE - Employee.db;
```

Hadoop Operations -

1) Import -   RDBMS ⟶ HDFS.

2) Analysis - Map Reduce

3) Export - HDFS → RDBMS

Apache Ambari

Zookeeper	OZIE			Ap. HBASE	Ap. Sharn	Ap. Szoop
		Hive	Pig			
		mapReduce	TEZ	Spark		
		YARN	Mesos			
		HDFS				
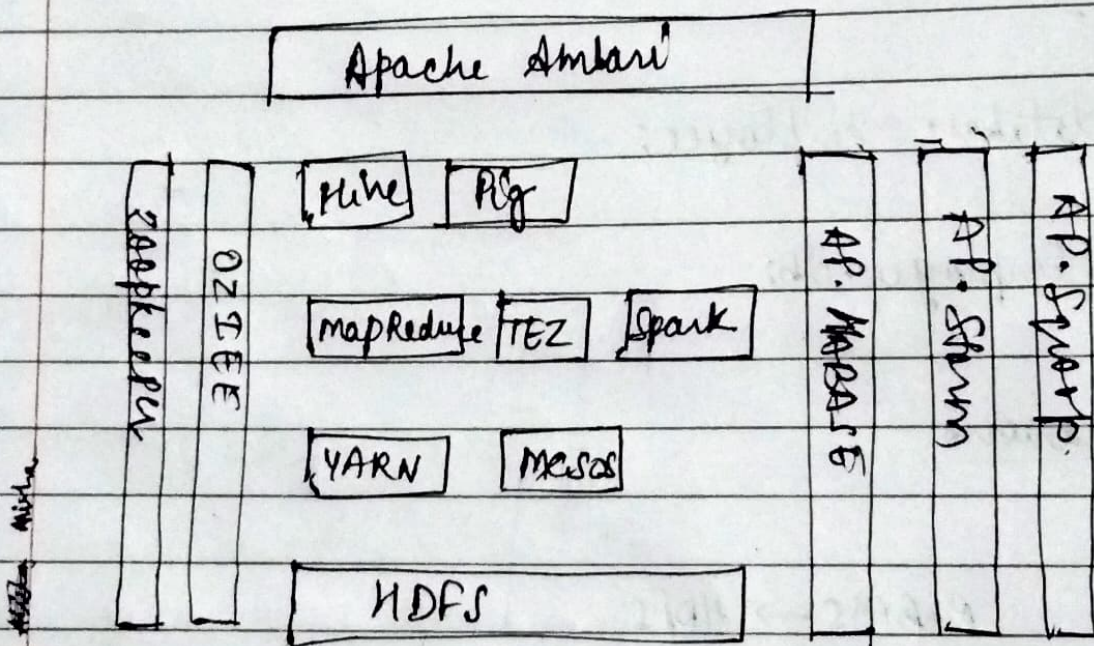
HDFS→ Cluster formation ~~due to~~ .

YARN→ Yet Another Resource negotiator
(manages all resources) (Heartbeat of Hadoop system)

Mesos→ Alternative of YARN

SR.

MapReduce → Analysis of data

Pig → inbuilt Scripting languages or API generating.

Hive → SQL queries can be runned in this.

TEZ → Graphical representation

Spark → we can write Scripting languages or generate APIs in it.

Ap. HABASE → Used for storage

Ap. Storm → Real time processing

Ap. Sqoop → Direct Connection b/w SQL and HDFS.

Ozie → Scheduling of tasks

Zookeeper → Tells the state of the system.

Apache Ambari → Overall Information of the application.