

6 August 2025

Big Data Lab

__/__/__

- * Hadoop was founded by GFS
Google File System
- * Yellow elephant - Symbol
- * Since, 2005 it's been evolving.

Hadoop: Hadoop is an open source platform.
It is an interface or interlinking
between database and hdfs

Apache Sqoop: This is also an open source
framework used for efficient bulk
data transfer.

1. Import → ~~need~~ database → Database create →
Schema → Records Enter

Apache Sqoop

Import → RDBMS → HDFS

Export → HDFS → RDBMS

* Use :- Analyzing the data

* Tools & Technology - Hive, Pig, Spark, Map Reducer

↓ ↓
Hadoop processing Aggregate the info

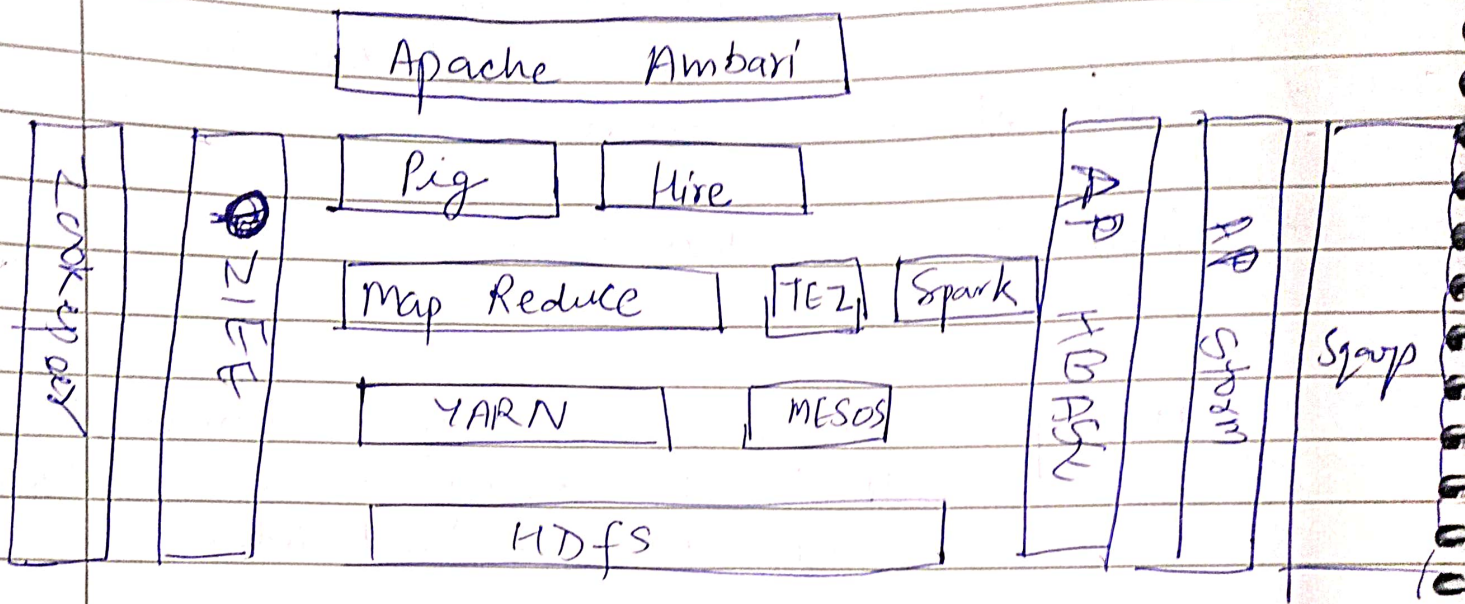


Fig. Hadoop Architecture

Distribute file System:- Systems are connected
If data is lost in
one system we can access it on another.

YARN- Yet Another resource Negotiator
Function- Tells us from which resource
we should get data
Same as MESOS

Map Reducer → Alternatives TEZ @ SPARK
[Can make graphical representation]

SPARK → Scripting language

//_

Apache HBASE → Storage work usage

Apache storm → real time processing usage

module update can be done without shutdown it .

OZIEE → scheduling purpose

Zookeeper → state of machine
up → function run
down → function stop

Apache Ambari → Access of Everything.