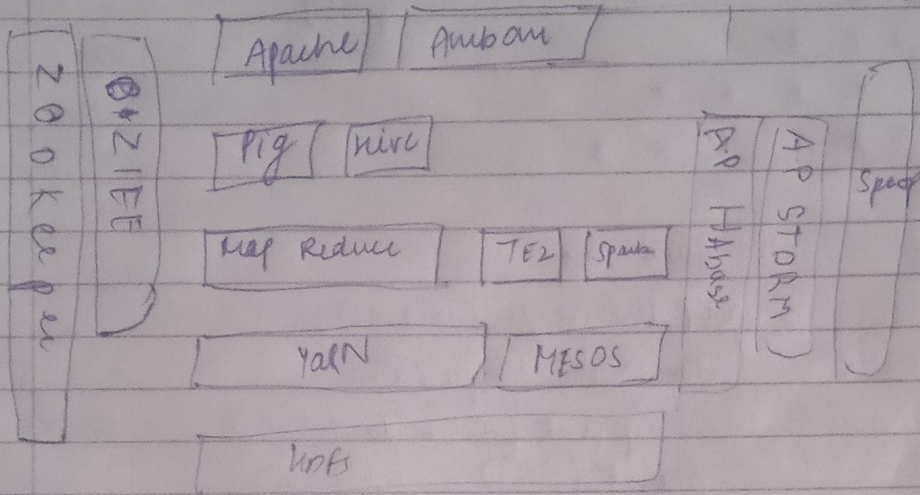


Use → Analyze the data

Tools & Techniques → Hive, Pig, Spark, Map Reduce

parallel processing

Aggregate the information



Hadoop Architecture

Cluster.

distributed file

YARN - Yet Another Resource Negotiator

cluster

TEZ → graphical setup

Spark → Scripted language

AP Hadoop Storage works

AP Storm → for real time processing

OZONE → Scheduling purposes

Zookeeper → just koi shai mai hai

ip → on

do → off

Ambari → we can enter the field which we want to access

1. Installation of java using Unix/Linux

Hadoop was founded by GFS Google File System  
Symbol: Below Elephant  
Since 2005 its been evolving

Hadoop is an open source platform it is an  
interference or interlinking b/w database and hdfs

Apache open source framework use for efficient data bulk data  
transfer Hadoop

Import → RDBMS → HDFS

Export → HDFS → RDBMS