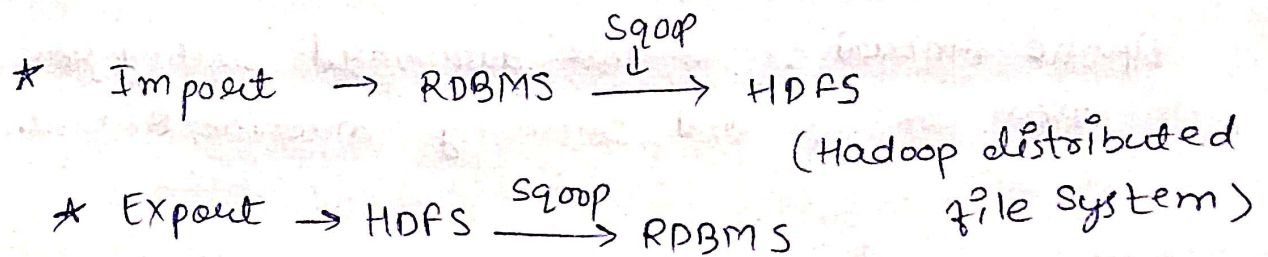① Installation of Java using Unix/Linux

② Hadoop is founded by Google file System (GFS).

> Symbol - yellow elephant.

> Since 2005 it's been evolving

> Hadoop is an open Source framework it is an interface or interlinking between database and HDFS.

> Aapache Sqope :- open Source framework used for efficient bulk data transfer.

* Import → RDBMS $\xrightarrow{\text{Sqoop}}$ HDFS

(Hadoop distributed file System)

* Export → HDFS $\xrightarrow{\text{Sqoop}}$ RDBMS

Use → Analyze the data.

Tools and Technology → Hive, pig, Spark.

Map Reduce — process large dataset in parallel processing

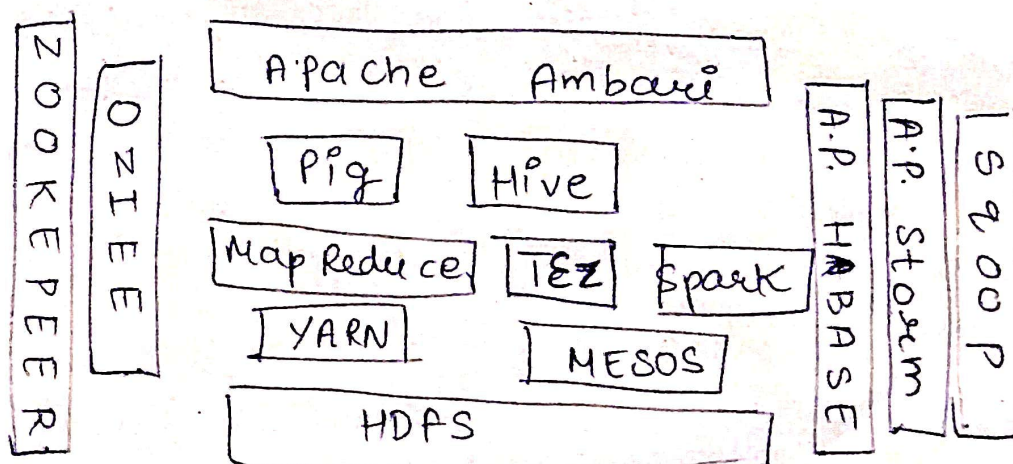Reduce — Aggregate or Summarize result

Map- process and breakdown data.

fig. Hadoop Architecture

cluster

distributed file system that stores data across multiple machines (nodes) but we can access and manage it it is on a single system.

YARN - yet another resource Negotiator

TEZ - graphical representation.

Spark - scripting language

A.P. HABASE - Storage process

A.P Storm - Real time operations

OZIEE - Scheduling purposes.

ZOOKEPEER - function coordination distributed system.

Apache Ambari :- control dashboard letting you switch them on/off and showing machine status.