**--hadoop framework**

YARN --processing --admin/dev

HIVE

PIG -- structure data handling dataware housing

SqooP --database --->HDFS

apache storm --real time data processing

ambari --web portal

----------

hortonworks

cloudera --free /paid

**--------------apache spark------------------**

1) strorage --own storage --HDFS use

2) processing --own processing --YARN

3) SQL data --HIVE -HQL --apache SQL

4) live streaming --apache storm --by default

5) AI / ML ---apache mahout ---spark MLIB

6) analytics /visual --spark GraphX

**\*\*spark is 10 times more faster than YARN. \*\***

**\*\*it support R, JAVA, python, SCALA (build lang)**

**\*\*\*databricks is a plateform having -HDFS , S3 , live, file**

it contain -DBFS