Batch / real time processing

HDFS: file manage

Map Reduce

YARN: resource manag

PIG: high-level platform for creating programs that run on [Apache Hadoop](https://en.wikipedia.org/wiki/Hadoop)

**HIV**: data query and analysis

convert sql to be used by mapreduce

**Impala**: MPP(massive product process

open source, native analytic database

Sqpoop: efficiently transferring bulk data

**Spark**: im-memory process db. [open-source](https://en.wikipedia.org/wiki/Open-source_software) [distributed](https://en.wikipedia.org/wiki/Distributed_computing) general-purpose [cluster-computing](https://en.wikipedia.org/wiki/Cluster_computing) [framework](https://en.wikipedia.org/wiki/Software_framework)