spark是一个基于内存计算的开源的集群计算系统，是一种与 Hadoop 相似的开源集群计算环境；Spark 是在 Scala 语言中实现的，它将 Scala 用作其应用程序框架。与 Hadoop 不同，Spark 和 Scala 能够紧密集成，其中的 Scala 可以像操作本地集合对象一样轻松地操作分布式数据集。

storm实时分布式处理系统

·基于Clojure语言开发(运行在JVM上)

·实时处理作业(一次一处理)

spark准实时分布式处理系统

·基于scala语言开发(运行在JVM，可用java类库)

·根据时间间隔，批处理作业

·mapReduce能将Job中间结果存入内存

·基于HDFS，内存无命中时同hadoop

hadoop分布式处理系统

Kettle 是一款国外开源的ETL工具，纯java编写