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。

hadoop 分布式批处理计算，强调批处理，常用于数据挖掘、分析

Kettle 是一款国外开源的ETL工具，纯java编写