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