第二次作业： MapReduce Spark Storm

姓名：张帅豪　　学号：18030100101

MapReduce

> MapReduce 模块：开源分布式计算的第一个流行的框架
>
> 主要适用于大批量的集群任务，由于是批量执行，故时效性偏低。原生支持 Java 语言开发 MapReduce ，其它语言需要使用到 Hadoop Streaming 来开发。

Spark

> Spark 提出了内存计算的概念，核心思想就是中间结果尽量不落盘。
>
> Spark 相对 MapReduce 获得了千百倍的性能提升，代价是更高的内存开销以及 OOM 风险。Spark可以被用于处理多种作业类型，比如实时数据分析、机器学习与图形处理。多用于能容忍小延时的推荐与计算系统。

Storm

> Storm是一个分布式的、可靠的、容错的流式计算框架。 通过产生数据流的源头和消费数据流的管道来抽象流计算的世界，一开始就是为实时处理设计，因此在实时分析/性能监测等需要高时效性的领域广泛采用。Storm可应用于--数据流处理、持续计算（持续地向客户端发送数据，它们可以实时的更新以及展现数据，比如网站指标）、分布式远程过程调用（轻松地并行化CPU密集型操作）。

相同点

- 都是用于大数据的计算。
- Spark与Hadoop MapReduce均为开源集群计算系统

差异点

名称	处理方式	特点
MapReduce	批处理	每次启动任务后，需要等待较长时间才能获得结果
Spark	批处理	基于内存计算实现，可以以内存速度进行计算，优化工作负载迭代过程，加快数据分析处理速度
Storm	流处理	Twitter的流式处理大数据分析方案