**实时flink更好，批次spak更成熟**

**因为spark年代长一点，相对成熟一点；而flink的话是现在最好的实时计算框架**

对于以下场景，你可以选择 Spark。

1. 数据量非常大而且逻辑复杂的批数据处理，并且对计算效率有较高要求（比如用大数据分析来构建推荐系统进行个性化推荐、广告定点投放等）；
2. 基于历史数据的交互式查询，要求响应较快；
3. 基于实时数据流的数据处理，延迟性要求在在数百毫秒到数秒之间。

**都是基于内存计算的**

**微妙到毫秒级的选择flink**

**Sqoop要用于在Hadoop(Hive)与传统的数据库(mysql、postgresql...)间进行数据的传递，可以将一个[关系型数据库](https://baike.baidu.com/item/%E5%85%B3%E7%B3%BB%E5%9E%8B%E6%95%B0%E6%8D%AE%E5%BA%93/8999831" \t "https://baike.baidu.com/item/sqoop/_blank)（例如 ： MySQL ,Oracle ,Postgres等）中的数据导进到Hadoop的HDFS中，也可以将HDFS的数据导进到关系型数据库中。**

**Hdfs是针对TB级的数据量存储，太小的数据不适合用hdfs**

**Hbase不支持事务，只能单表查询**