**Flink**

1. 分布式、高性能、高可用、准确的流处理框架（支持实时流处理和批处理）
2. 特点 流式优先 连续处理 有状态计算 可伸缩 高性能
3. 数据来源回写 应用数据、传感器和设备、文件系统和存储、日志消息
4. 部署 local本地 集群cluster 云cloud 实际部署集群yarn部署
5. 提供的api有两种 stream data api 和 batch dataapi
6. 组件datasource 数据源 transformations 算子 datasink数据存储
7. 流处理模型，节点处理完数据，存入缓存，通知下个节点从缓存中获取数据继续计算
8. 批处理模型，节点处理完数据，存入缓存，并不会通过网络传输到下个节点，当缓存存满或者结果处理完后，存储到磁盘，当所有的数据被处理完后才会进入下个节点
9. 框架选择、
   * 1. 数据状态管理
     2. 数据处理次数；仅处理一次，保证最少处理一次
     3. 延迟低 如果是小型项目可以选择storm
     4. 如果已经使用sprak，并且仅满足秒级业务可以使用sparkstreaming
     5. 仅处理一次，数据量大，高吞吐，低延迟，需要进行状态管理或者窗口操作可以使用flink
     6. Flink重型 storm轻型（低延迟，低吞吐） sparkstreaming（小批量） 家族成员较多可以方便集成