# 大数据在线分析处理架构

## 目前主要架构

Flume + kafka + storm/spark + hbase/redis

Flume 主要用于数据的收集和传输

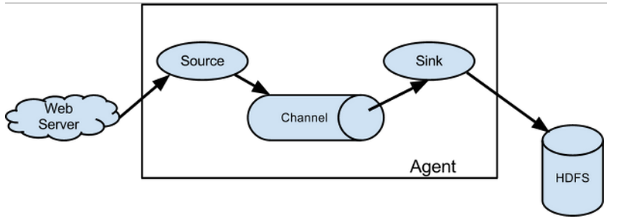
Kafka 作为消息队列缓存数据

Storm/spark 从kafka读取消息数据进行实时处理

Hbase 将处理后的数据保存

Flume简介

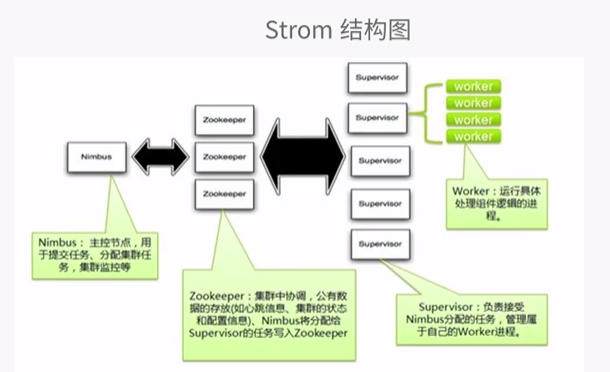
Flume通过在webserver上安装agent来实现日志的收集和传输，其架构图如下：



Storm简介

Storm是用来处理流式数据的，用户数据源源不断的流入，storm就永不停息的处理

其架构图如下



Nimbus：主控节点，用来分配任务到supervisor

Zookeeper：存放公共数据，nimbus将分配给superisor的任务写入zk

Supervisor：负责接收nimbus分配的任务，管理属于自己的worker进程

Worker：运行具体处理组件逻辑的进程

