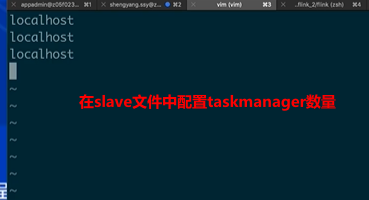
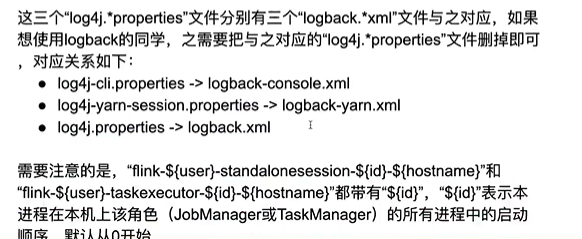
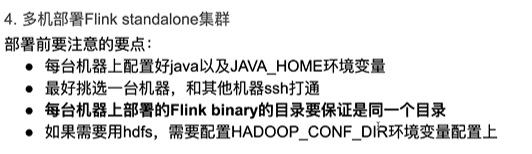
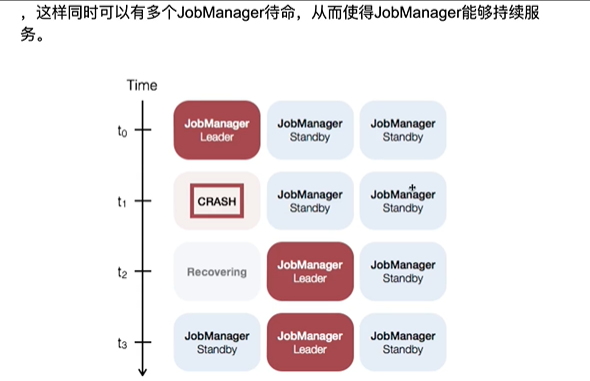
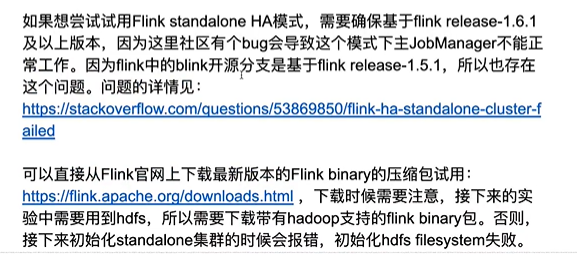
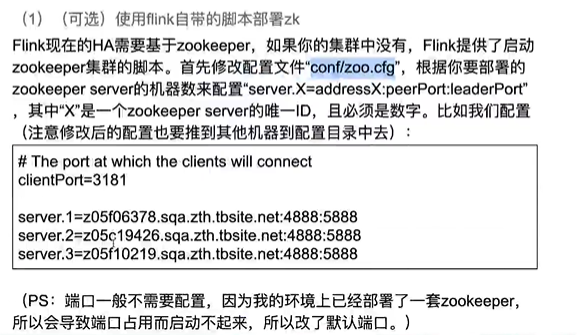
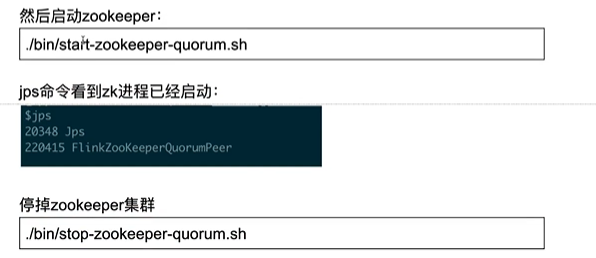
客户端 将客户提交的代码 优化 编译成图 然后提交给 JobManager

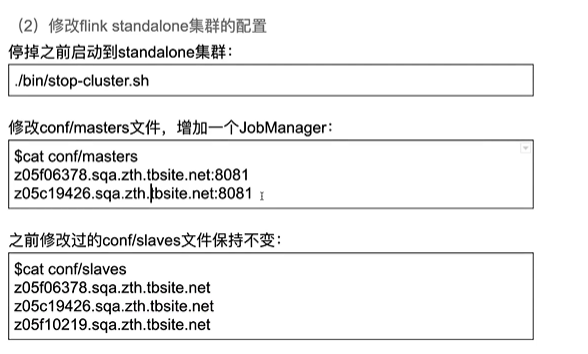
JobManager 负责调度，让taskmanager 执行任务

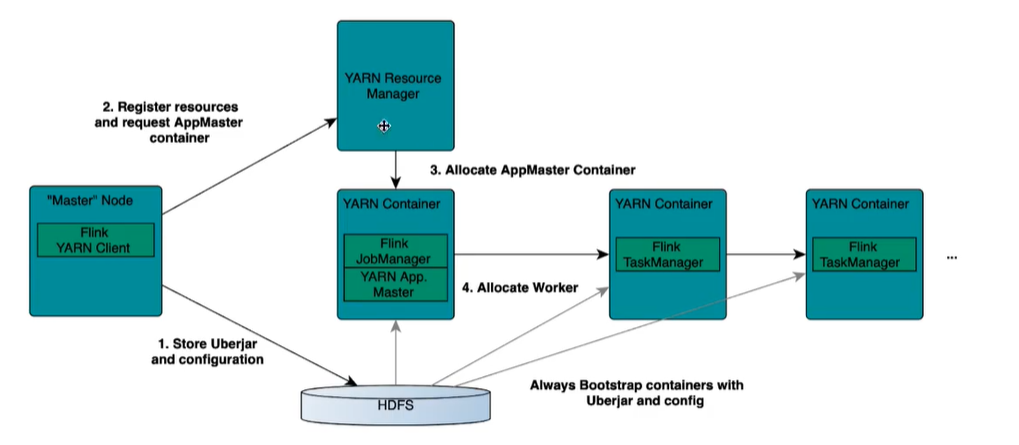
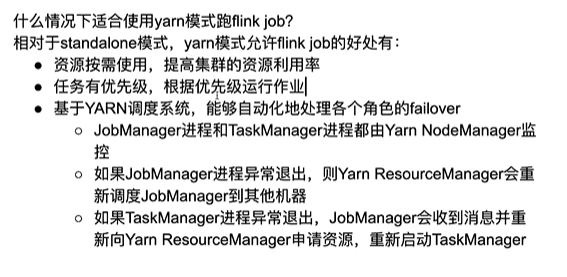




* 如果有两个flink，会产生一些冲突，启动flink会绑定6123端口
* flink启动后，会在/tmp目录下，会产生一个记录pid的文件，日志的命名会有pid
* zoo.cfg 可以用它启动zookeeper集群
* 
* 
* 
* 8081端口是master 或者worker配置的端口
* 
* 提交hdfs时，yarn已经存在，需要删除tmp文件下的一个文件
* 高可用HA
* 
* 
* 
* 

需要将zoo.cfg同步到各个节点



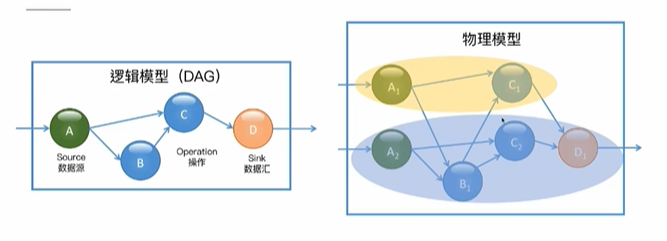
* yarn 模式paoflink job
* 
* Jobmanager找yarn的Resourcemanager申请资源
* 
* -n 是在yarn session模式下指定taskmamager的数量，-yn是在 yarn single job模式下指定taskmanager数量，-p是实际资源

# DataStream API

## 什么是流处理

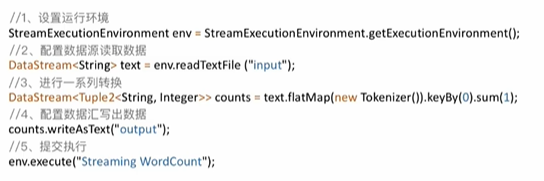
Flink是分布式流处理引擎，

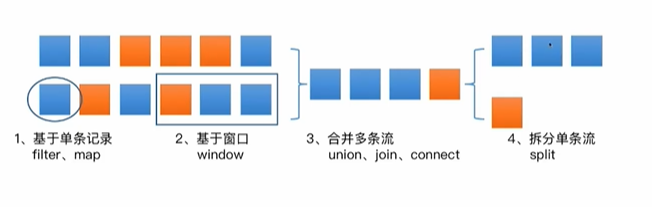
DAG，有向无环图



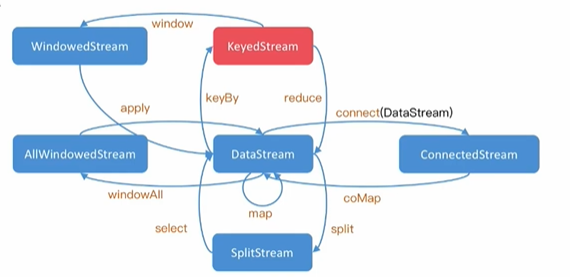


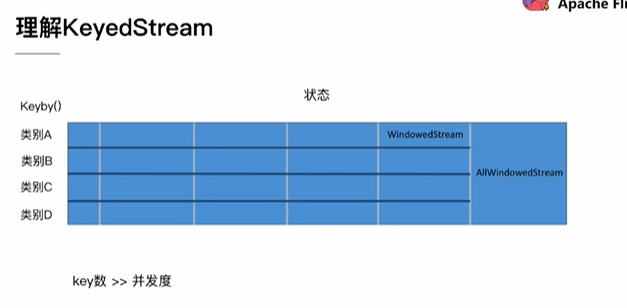
## Flink DataStream 程序结构

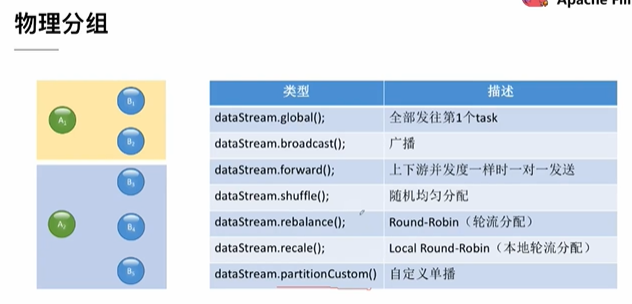




## DataStream基本转换

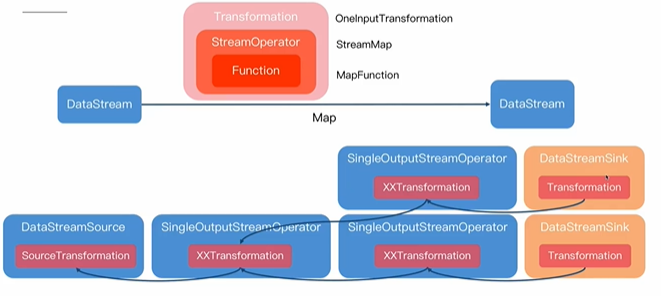








## API原理



# 客户端操作