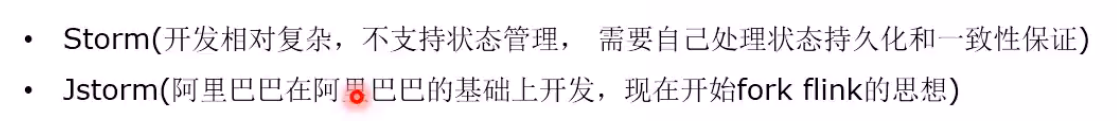
网易云<https://blog.csdn.net/wangyiyungw/article/details/80237270>

# Flink基本概念与部署

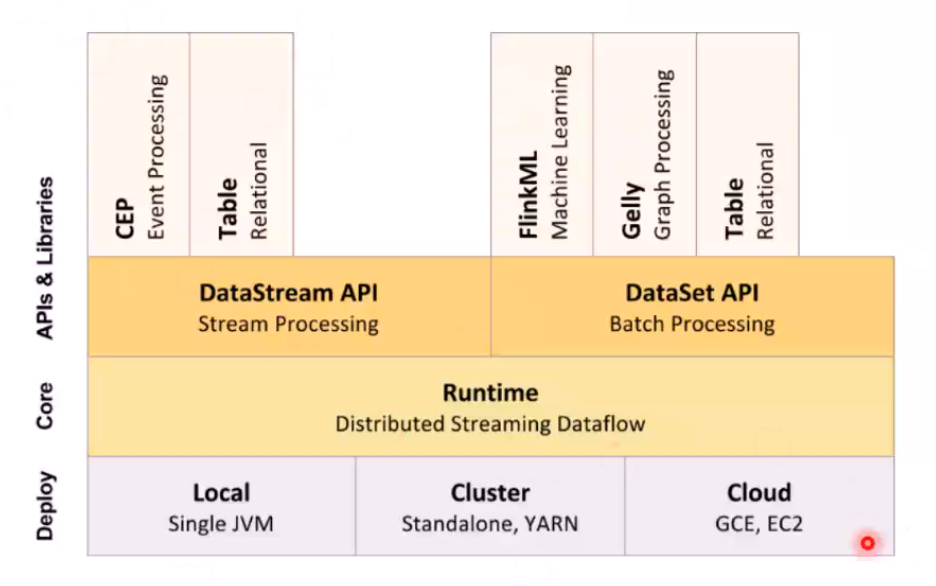
## 第一部分介绍

<https://www.oreilly.com/ideas/the-world-beyond-batch-streaming-101>

<https://www.oreilly.com/ideas/the-world-beyond-batch-streaming-1012>



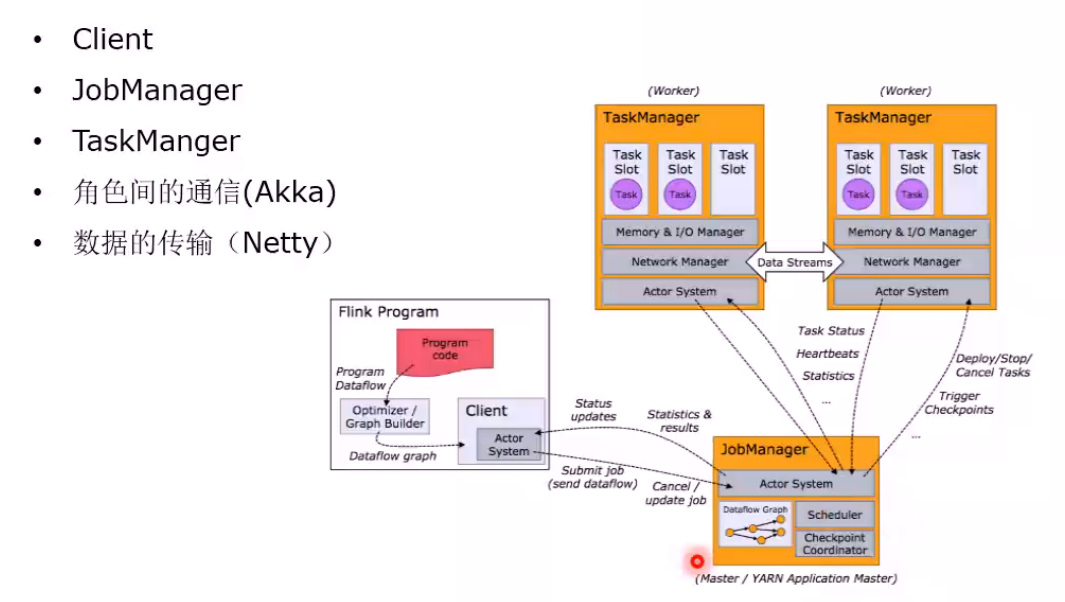
### Flink分层架构:



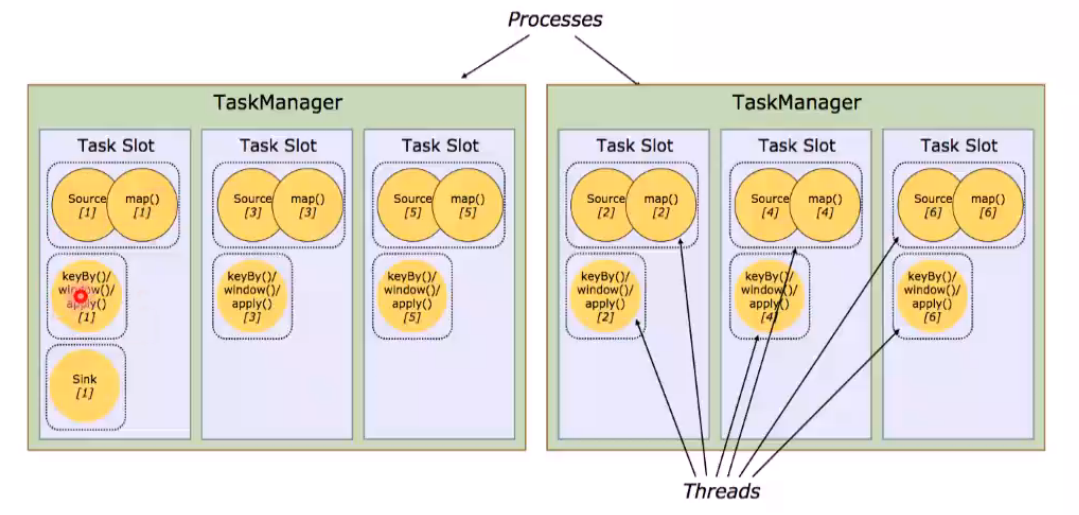
Flink支持流式处理以及离线处理，使用在流式处理比较多。

Cep全称复杂事件处理

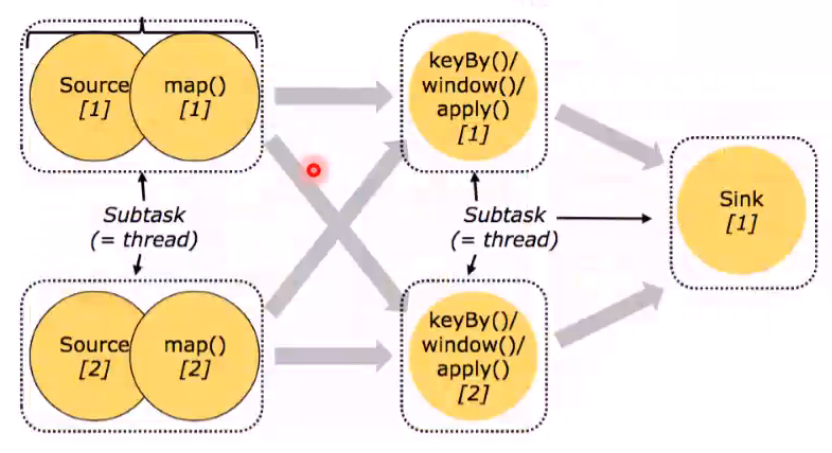
### Flink运行时架构



### TaskManger Slot



### OperatorChain && Task



### 运行时的几个概念理解

* JobManager

任务分发，checkpoint

* TaskManager

正真执行的

* TaskManagerSlots

TaskManager资源隔离

* Operator
* Task
* Parallelism

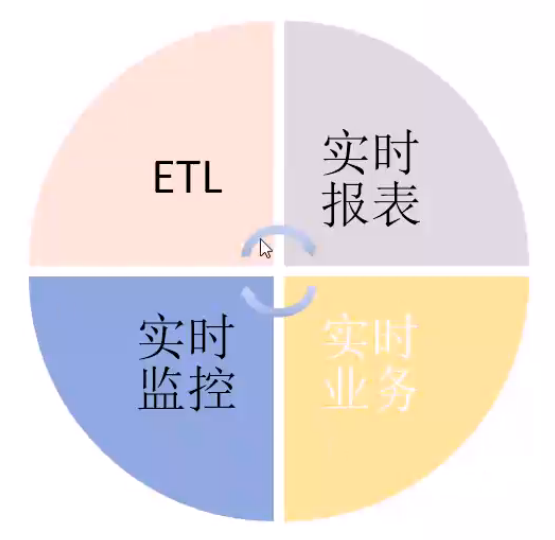
### EQS：

Spark streaming 吞吐量可能高点，微批次

Flink纯流式处理

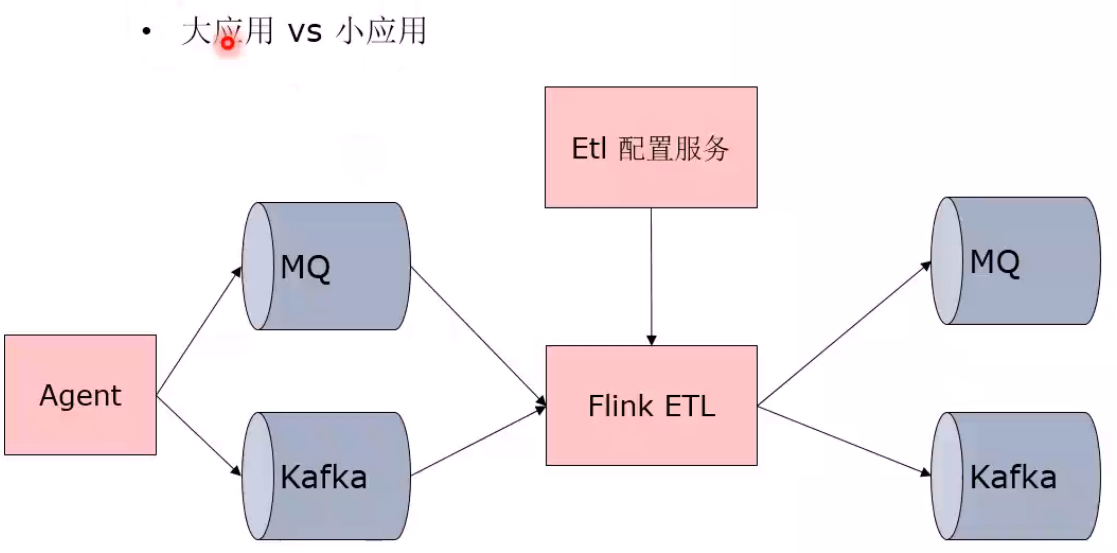
## 第二部分：Flink实践生产和学习方法

### 生产应用场景

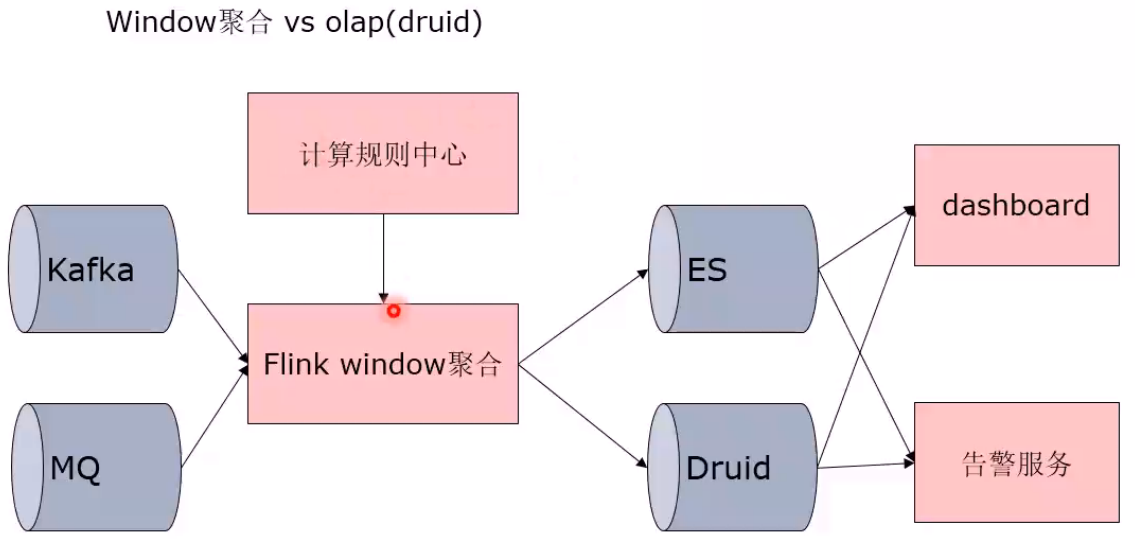


## ETL生产场景

### 大应用VS小应用



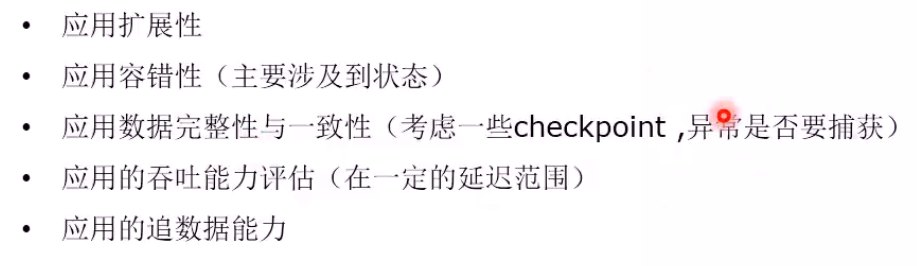
## 实时报表与监控场景

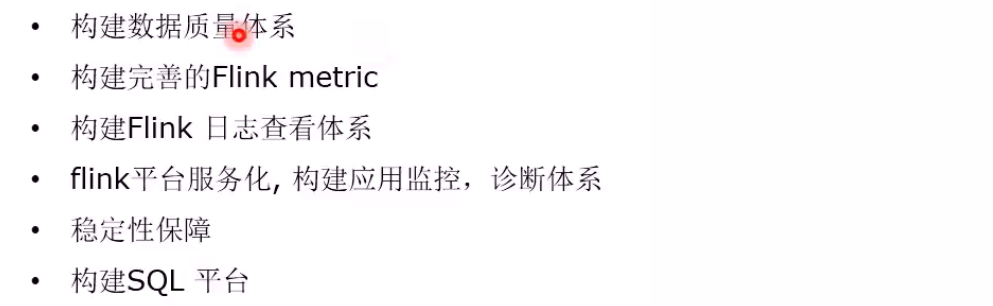


## 实时业务场景

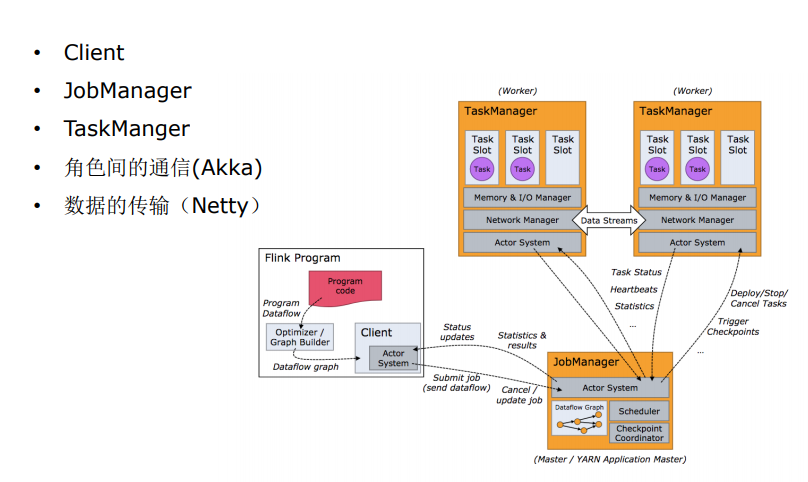
* CEP (Complex Event Process) 行为分析，风控领域
* 实时机器学习

## Flink生产经验

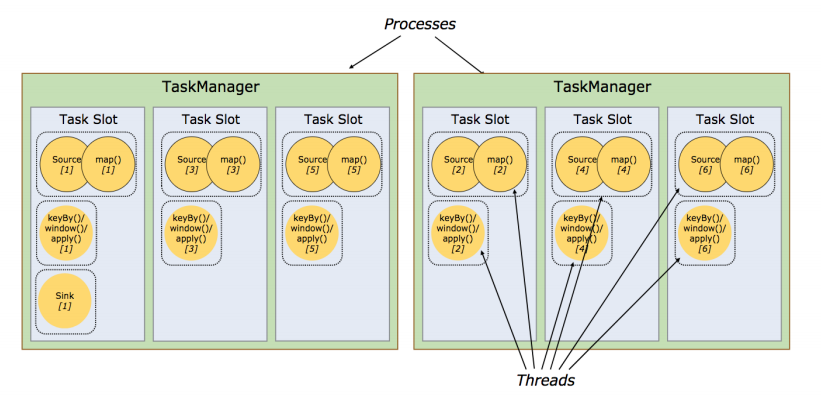




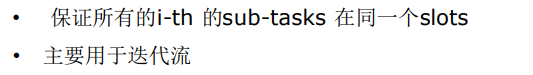
## Flink运行时架构



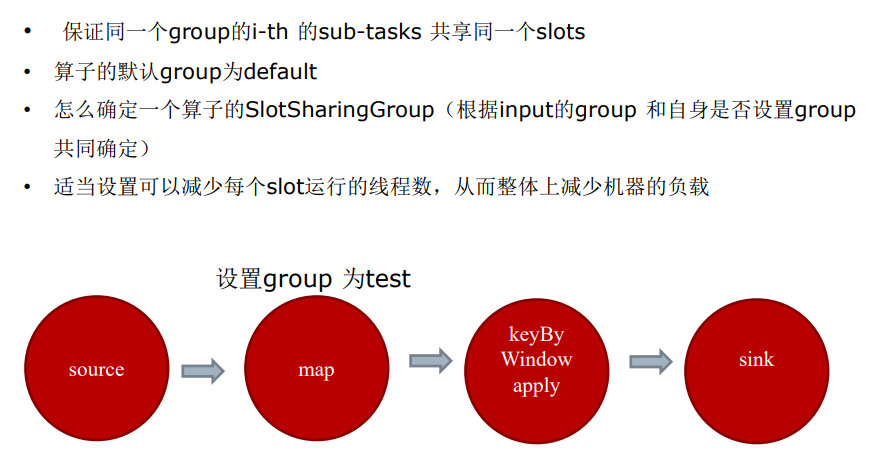
## TaskManager Slot



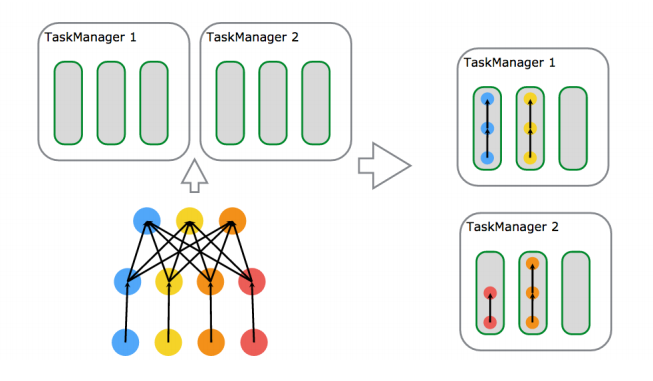
## CoLocationGroup

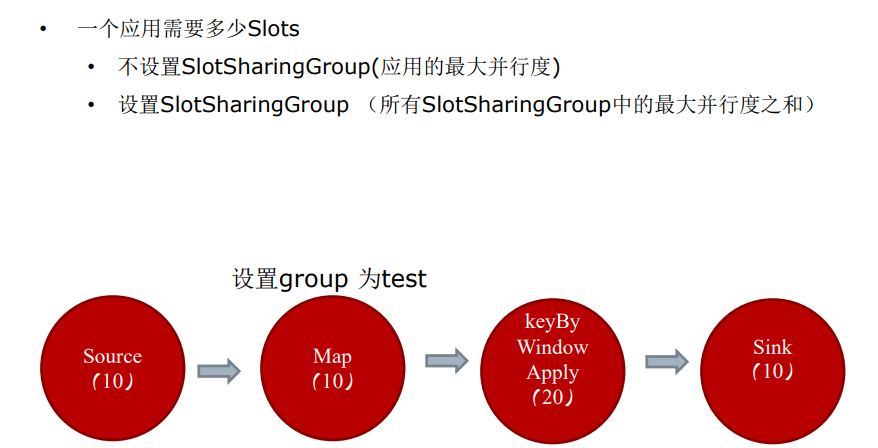


## SlotSharingGroup

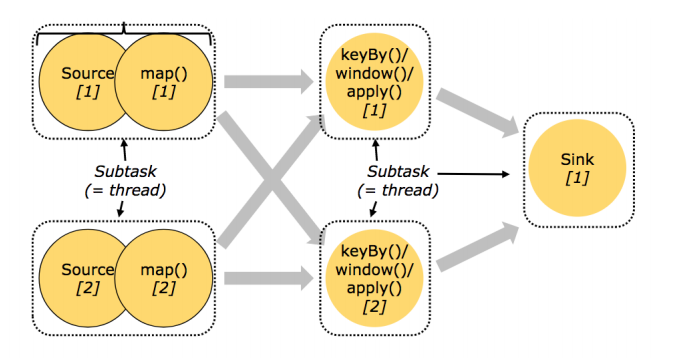


## Slots && parallelism





## OperatorChain && Task



## OperatorChain

* OperatorChain的优点

减少线程切换

减少序列化与反序列化

减少延迟并且提供吞吐能力

* OperatorChain组成条件

没有禁用Chain

上下游算子并行度一致

下游算子的入度为1 （input为1）

上下游算子在同一个slot group

上下游算子之间数据没有shuffle

## 运行时的几个概念

* JobManager

任务分发，checkpoint

* TaskManager

正真执行的

* TaskManagerSlots

TaskManager资源隔离

* Operator
* Task
* Parallelism

槽和并行度的关系

## 第三部分:Flink on yarn原理，生产和部署

### Flink部署方式

Local

Standalone

Yarn

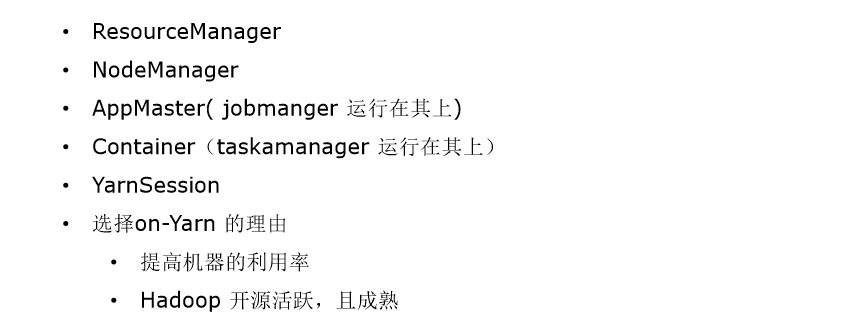
Mesos

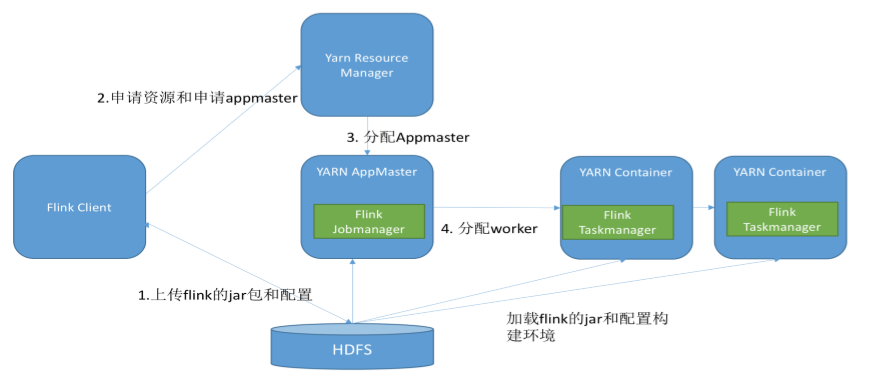
Docker

Kubernetes

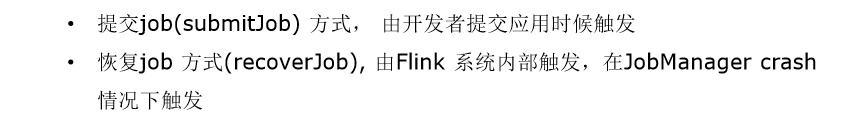
AWS

### Flink On Yarn

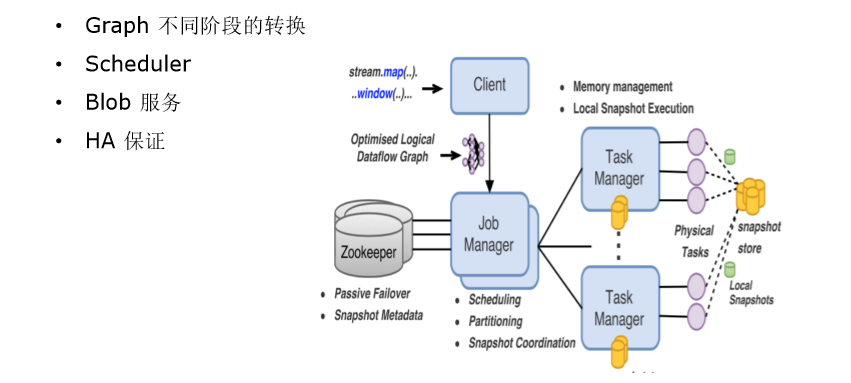




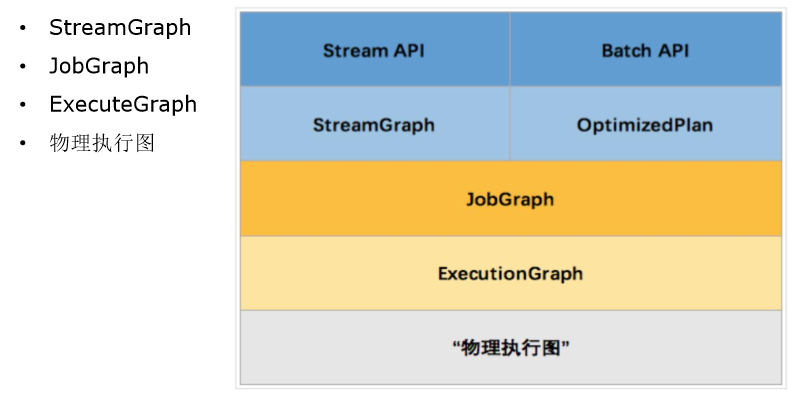
### Flinkjob启动方式



### 理解Job的启动过程

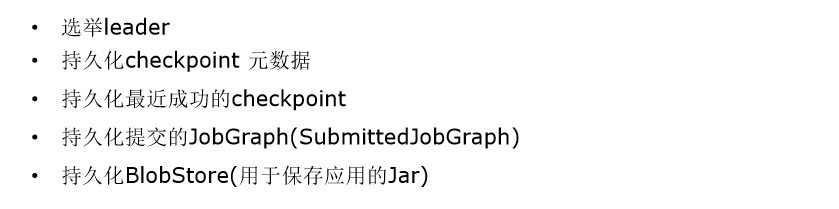


### Graph



SteamGraph在client生成dag有向无环图

### HA服务



### Flink On Yarn



### Flink-conf配置

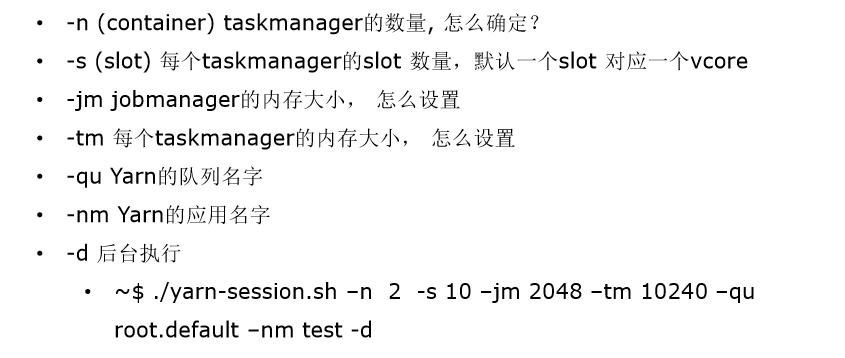


Flink配置：

State.backend：rocksdb #checkpoint先写入rocksdb，再异步写到hdfs

State.backend.fs.checkpointdir=hdfs://XX

### YarnSession启动命令



说明：

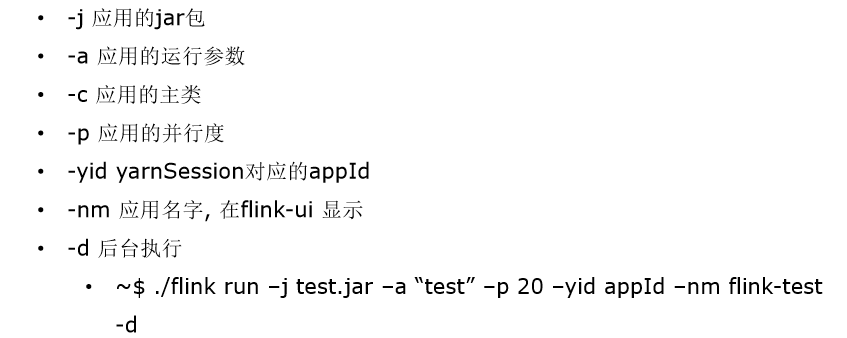
每个taskmanager的slot数据推荐6-10个。

container数据根据DAG来设置，根据应用所需要的Slot数，除以每个taskmanager设置的slot数量得到-n的数量，做冗余稍微多一点

jm在1.4以上需要2G以上

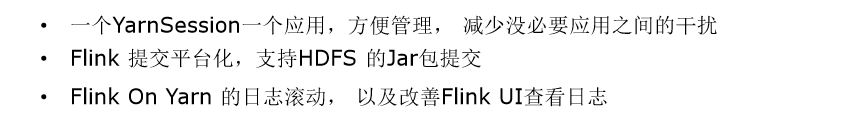
tm设置为slot的数量乘以1G或者2G，一个vcore对应多少内存。

### 应用启动命令



说明：-p是全局的并行度

### Flink On Yarn部署经验



### 小结：

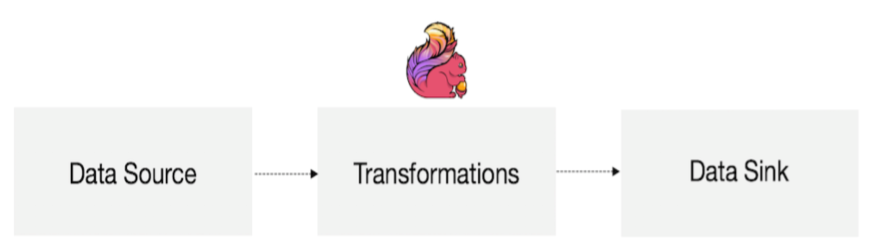


## EQS：

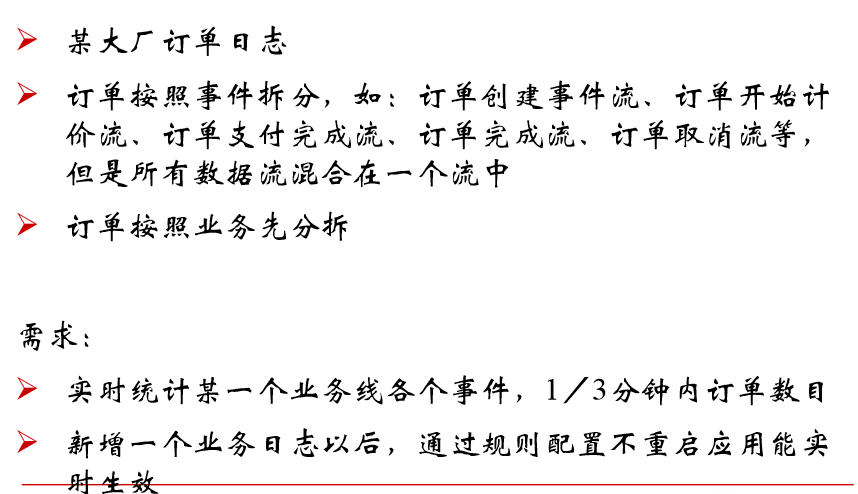
1.3.1有bug，hostname首字母大写无法正确安装

# DateStreamAPI介绍与实践

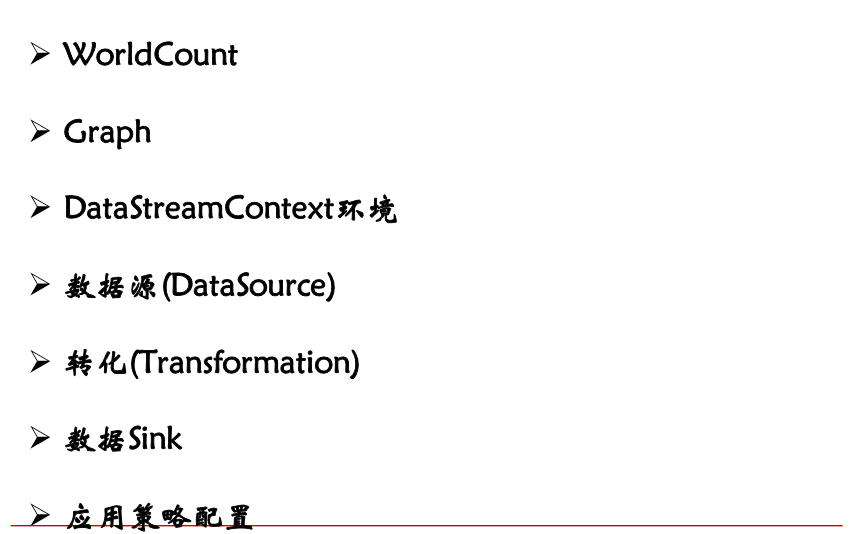
## 运行模型



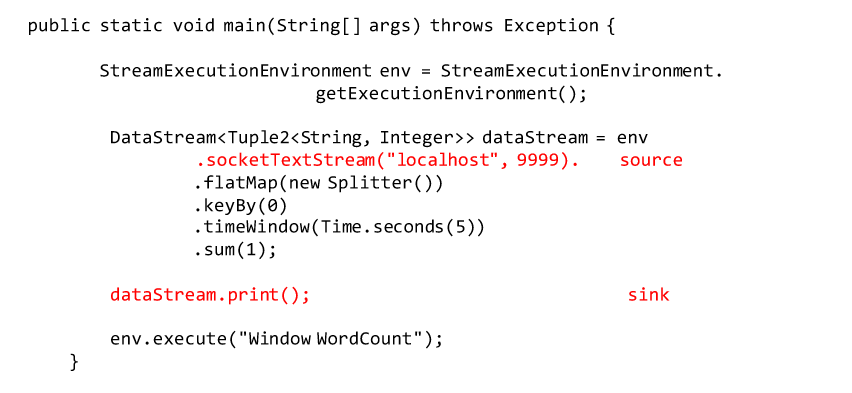
## 场景练习



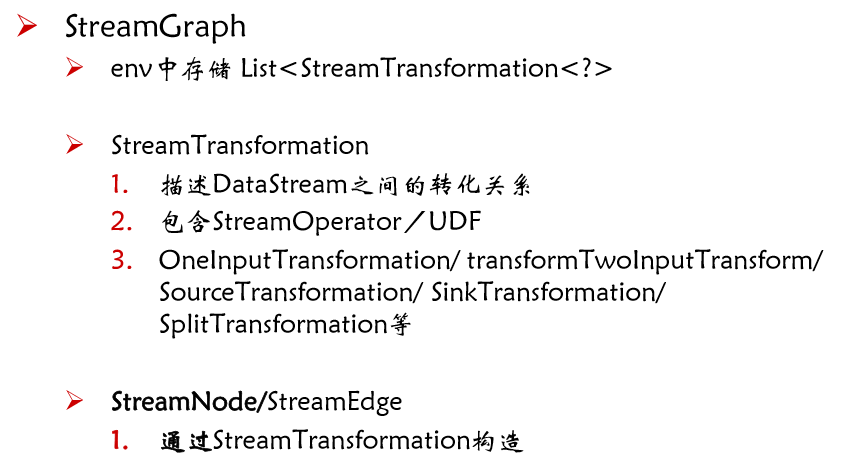
## 大纲

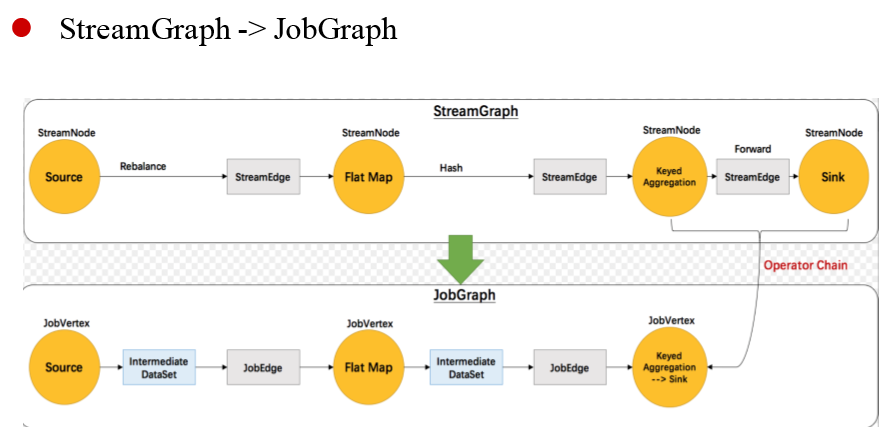


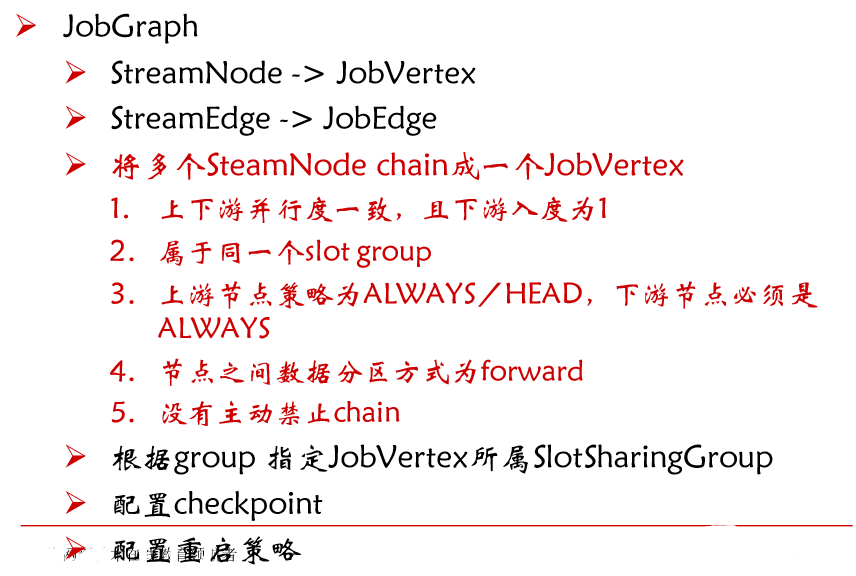
## WorldCount

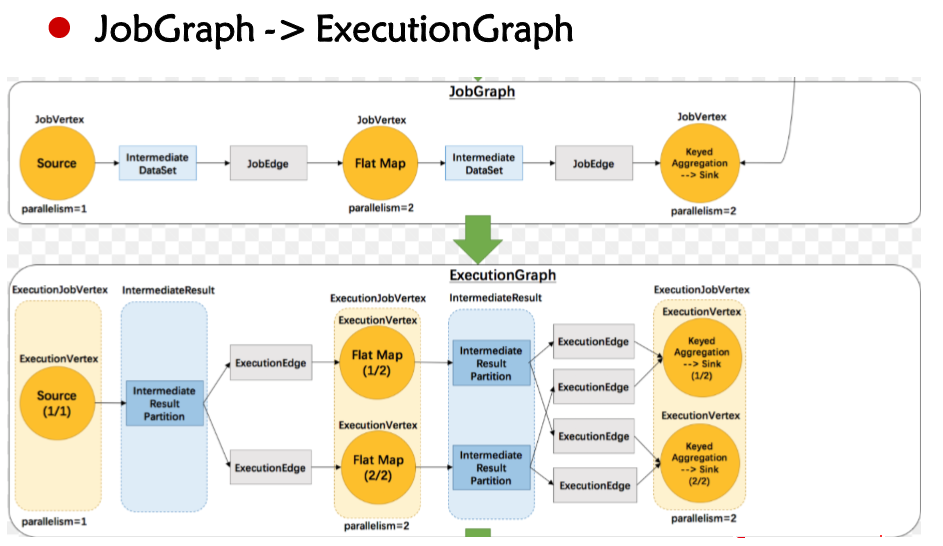


## Graph



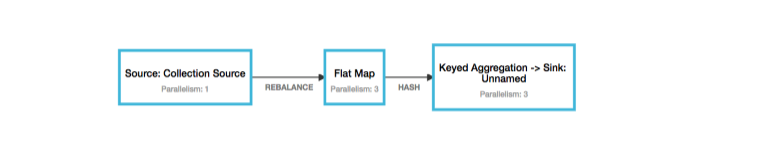




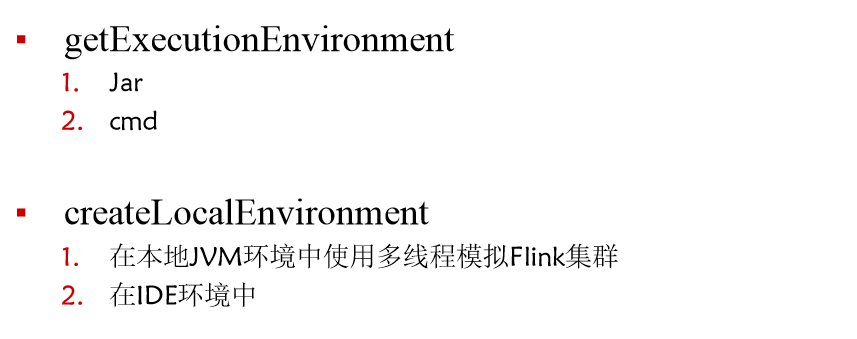




## 图DAG



## DataStreamContext环境



# Window与Time

<https://github.com/streaming-olap/training>

已经fork

# Connector

# 状态管理与恢复机制

# Metrics与监控

# Flink应用案列介绍

# Druid基本概念与架构设计

# Druid数九存储与写入

# F10