

腾讯云Kafka自动化运营实践

杨原
(ryanyyang)

基础架构部 – 高级工程师

SPEAKER



杨 原

基础架构部 – 高级工程师


腾讯云 Kafka 已经达到千个节点，数百集群的规模。日消息量达到数万亿级别，日流量总合PB级别。如何保证集群的正常服务以及如何处理异常情况成为日益重要问题。本次分享我们在 Kafka 运营中遇到的问题以及解决方式。同时介绍我们是如何实现自动化运营我们的 Kafka 集群。

腾讯云Kafka概述

- 基于Apache Kafka的分布式、高可扩展、高吞吐的云端服务
- 无需部署，封装所有集群细节，无需用户运维
- 按实例售卖，直接使用Kafka所有功能，提供多纬度监控
- 支持动态升降实例配置，按照需求付费
- 和腾讯云存储、大数据等无缝打通，使用方便

腾讯云Kafka现状

- 日消息量超万亿条，总流量达数十PB级别，单集群每分钟消息峰值十亿
- Broker节点过千个，集群达数百个
- 服务付费实例超千个，Topic数千



万亿级

日消息量

PB级

日吞吐量

十亿级

集群分钟吐量

面临哪些挑战？

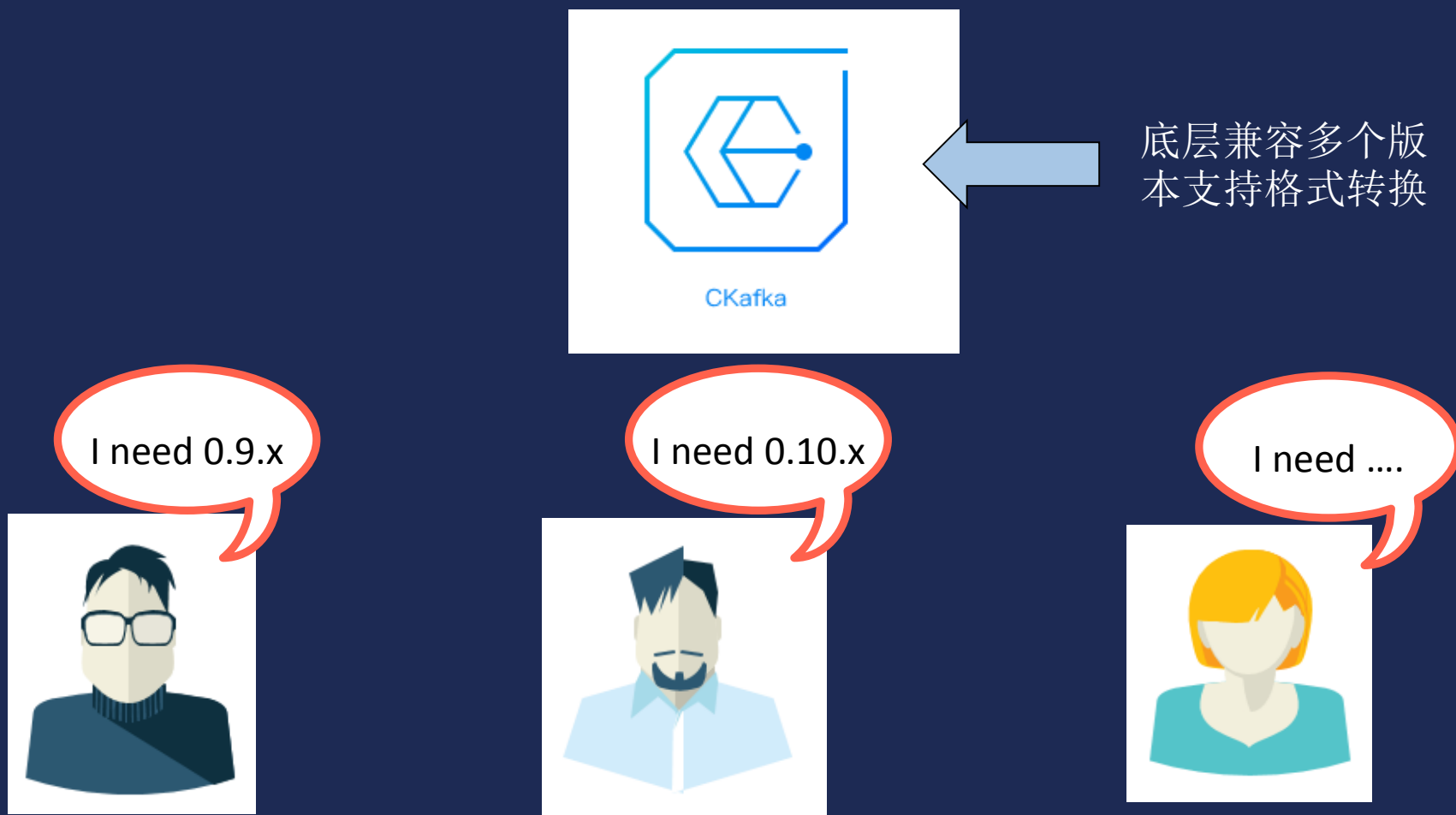
- 用户不同版本的选择？
- 实例的新建如何选择合适的Broker？
- 实例动态升降配-如何选择迁移分区？
- 节点的加入、移除-集群的负载均衡？
- 分区的创建、新增、迁移如何均衡？

面临哪些挑战？

- 用户不同版本的选择？
- 实例的新建如何选择合适的Broker？
- 实例动态升降配-如何选择迁移分区？
- 节点的加入、移除-集群的负载均衡？
- 分区的创建、新增、迁移如何均衡？

不同版本的选择

问题：如何满足用户使用的Kafka版本？



面临哪些挑战？

- 用户不同版本的选择？
- 实例的新建如何选择合适的Broker？
- 实例动态升降配-如何选择迁移分区？
- 节点的加入、移除-集群的负载均衡？
- 分区的创建、新增、迁移如何均衡？

实例选择合适的Broker

问题：新建实例，如何选择Broker 保证资源的充分利用？

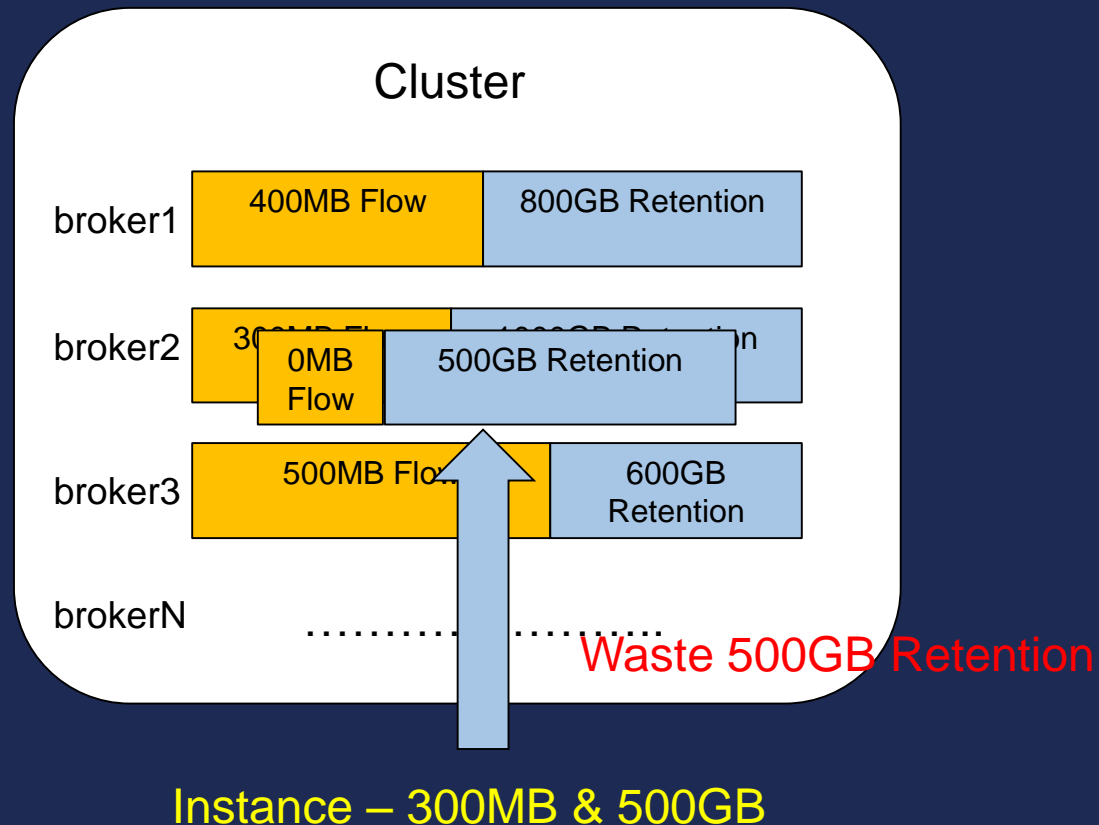
■ 选择集群中的那个Broker？

● 带宽(Flow)和磁盘容量(Retention)

型号	峰值带宽(MB/s)	磁盘容量(GB) ②
<input checked="" type="radio"/> 入门型	40	300
<input type="radio"/> 标准型	100	1000
<input type="radio"/> 进阶型	150	2500

■ 如何保证资源的最大利用率？

- 类似装箱算法
- 带宽售卖和磁盘售卖尽力1：1
- 奖励/惩罚机制



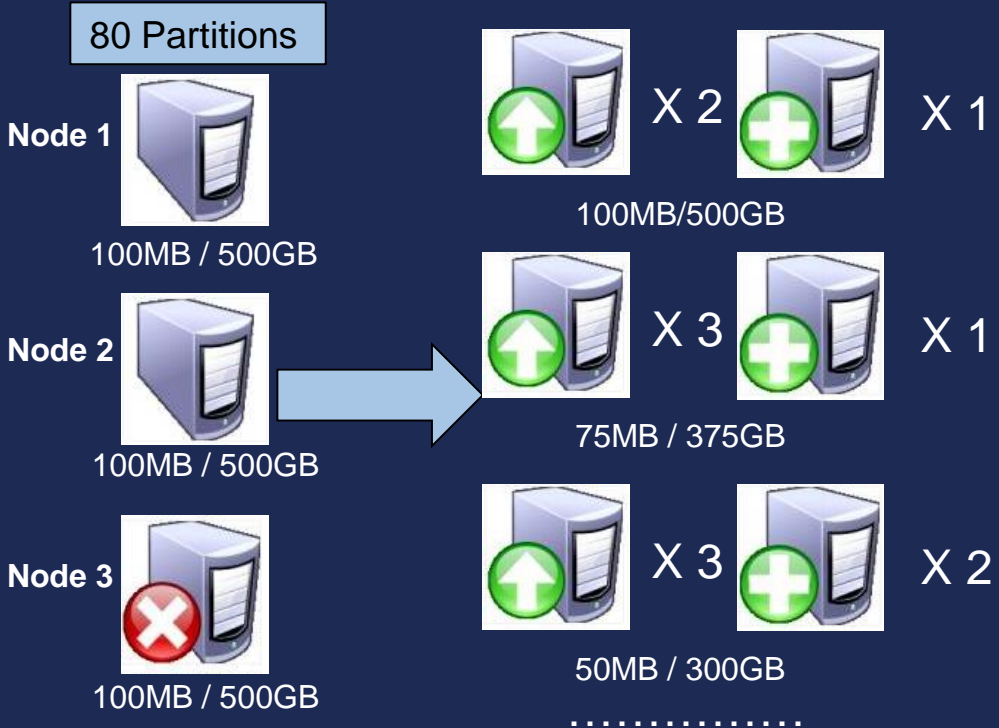
面临哪些挑战？

- 用户不同版本的选择？
- 实例的新建如何选择合适的Broker？
- 实例动态升降配-如何选择迁移分区？
- 节点的加入、移除-集群的负载均衡？
- 分区的创建、新增、迁移如何均衡？

实例的升配

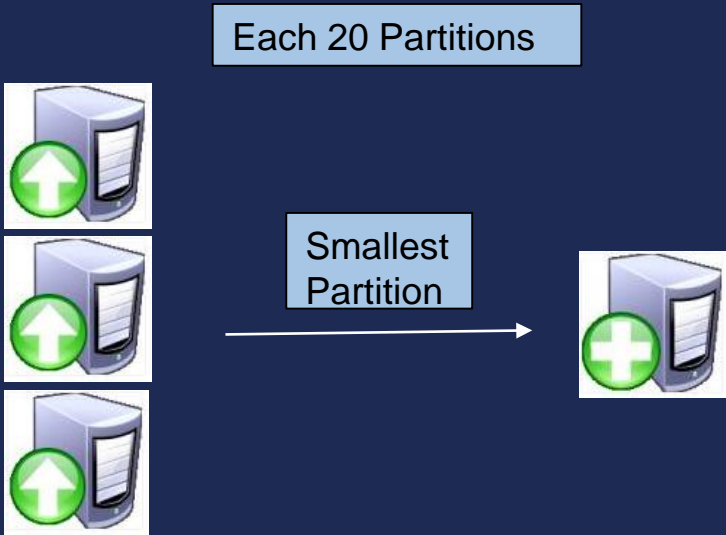
- 预估的容量不够用了？
 - 分区所在节点容量有冗余，直接升级，扣除资源
 - 所在Broker 资源不足，需要如何重新分资源以及分区迁移

Step1. 计算节点变更可能性及资源利用率



Step2. 计算迁移代价

addNum	reassignBrokerList	migrationSize	newBrokerList
1	Node 3	23195	Node 1 ~ 4
2	Node 3	23195	Node 1 ~ 5
3	Node 3	23195	Not enough
0	Node 3	40593	Node 1 Node 2 Node 4



面临哪些挑战？

- 用户不同版本的选择？
- 实例的新建如何选择合适的Broker？
- 实例动态升降配-如何选择迁移分区？
- 节点的加入、移除-集群的负载均衡？
- 分区的创建、新增、迁移如何均衡？

节点的加入、移除-Broker的负载均衡

■ 新增节点

- 实例的能力扩展
- 节点资源碎片整理
- 机器负载均衡

■ 移除节点

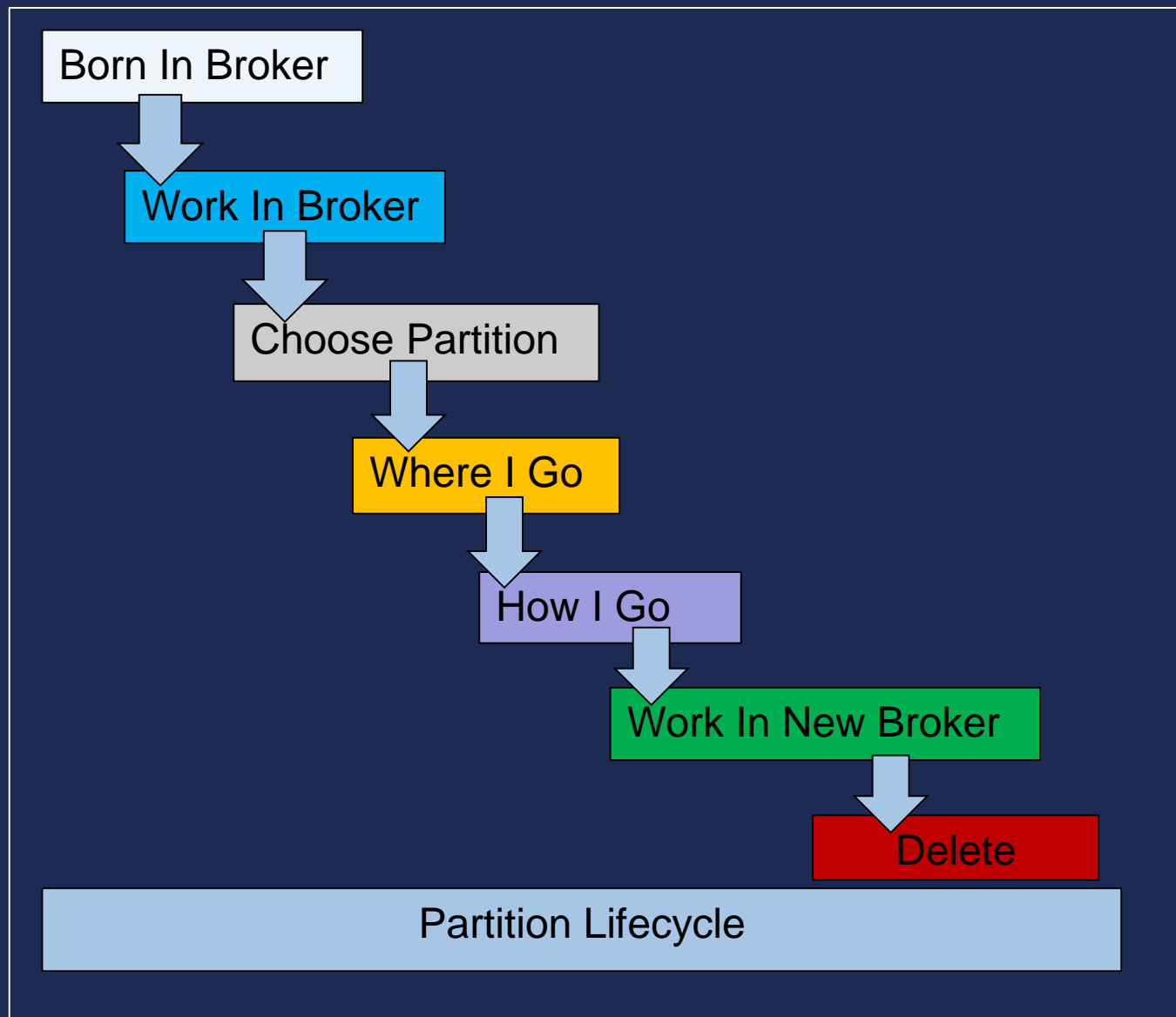
- 实例收缩
- 节点故障

面临哪些挑战？

- 用户不同版本的选择？
- 实例的新建如何选择合适的Broker？
- 实例动态升降配-如何选择迁移分区？
- 节点的加入、移除-集群的负载均衡？
- 分区的创建、新增、迁移如何均衡？

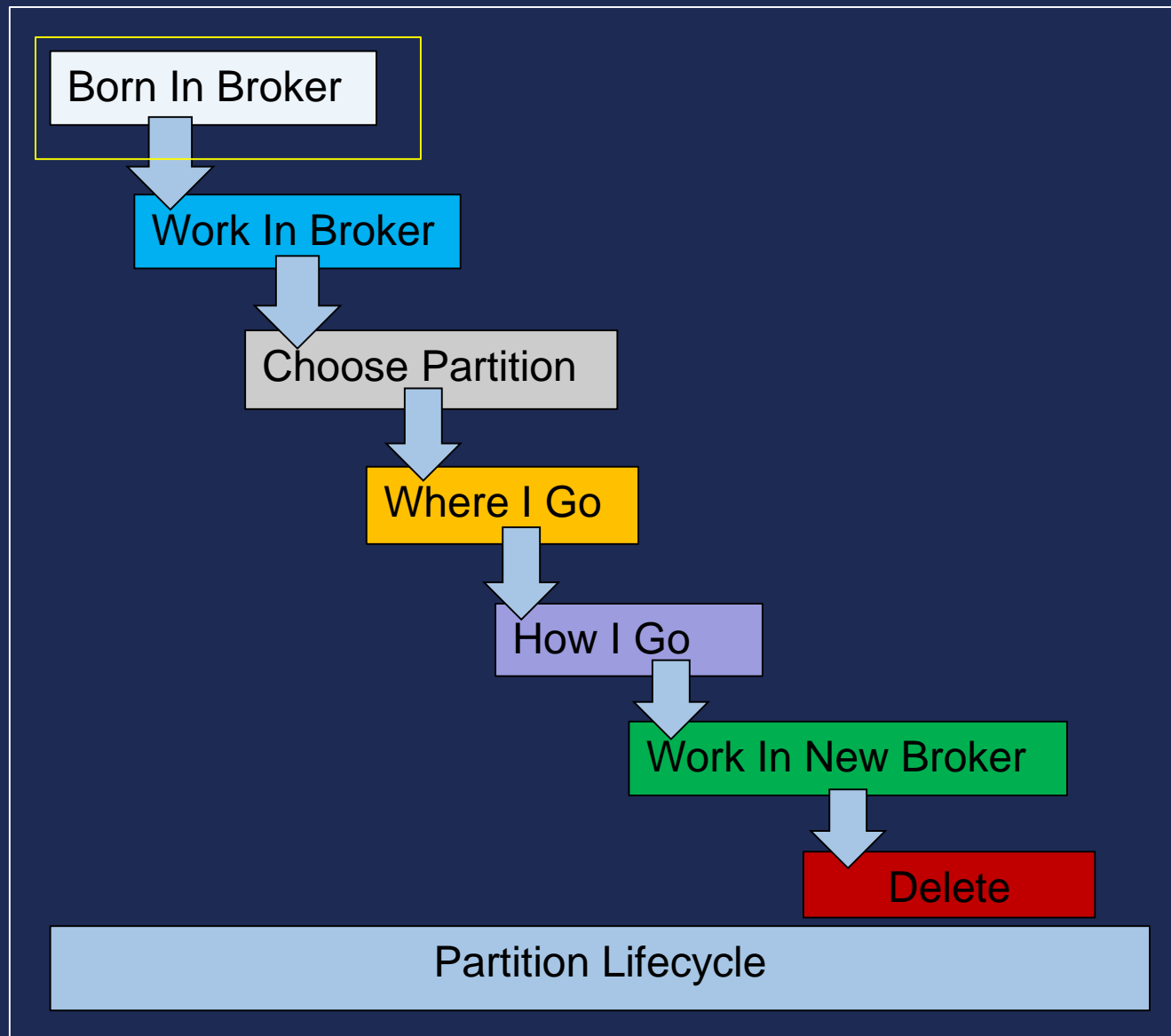
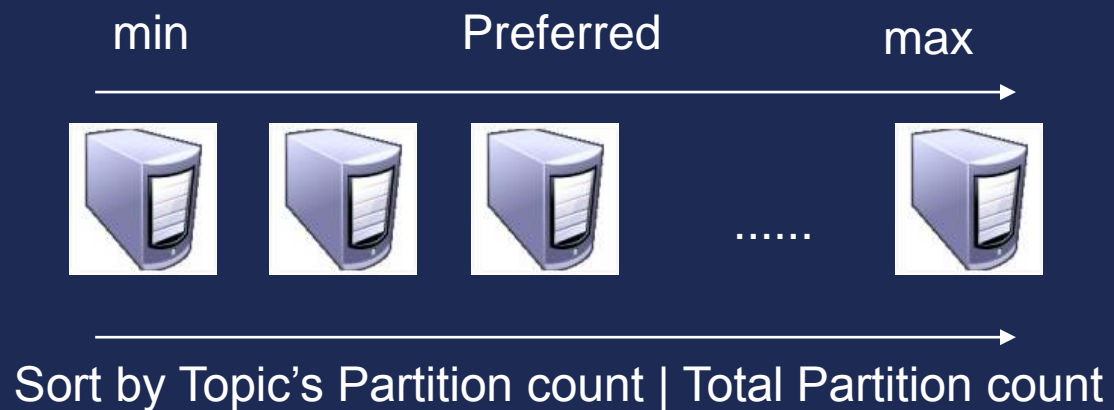
分区的创建、新增和迁移

- 选择哪个节点创建、新增？
- 哪些Partition应该迁移？
- 迁移多少Partition？
- 迁移到哪些节点？
- 迁移Leader还是Replica？



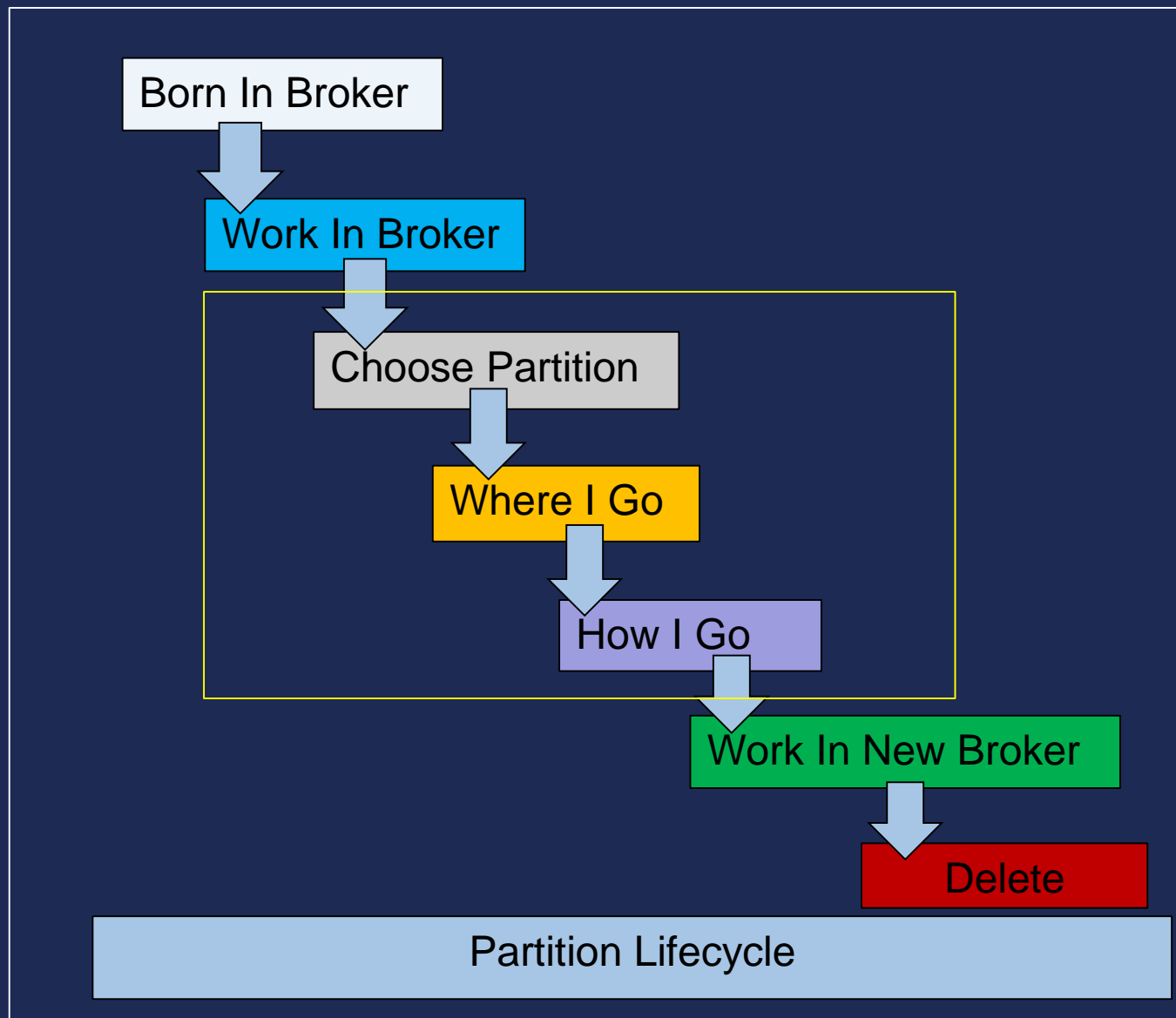
分区的创建、新增和迁移

■ 创建和新增在哪个节点上？



分区的创建、新增和迁移

- 为什么要迁移
 - 服务异常
 - 实例扩缩容
 - 负载均衡
- Leader迁移
 - 无数据迁移代价
 - CPU负载以及网卡出流量
- Replica迁移
 - 数据迁移代价大
 - 消耗机器资源
- 迁移对象和目的地
 - 数据大小
 - 生产/消费速率
 - 资源利用率

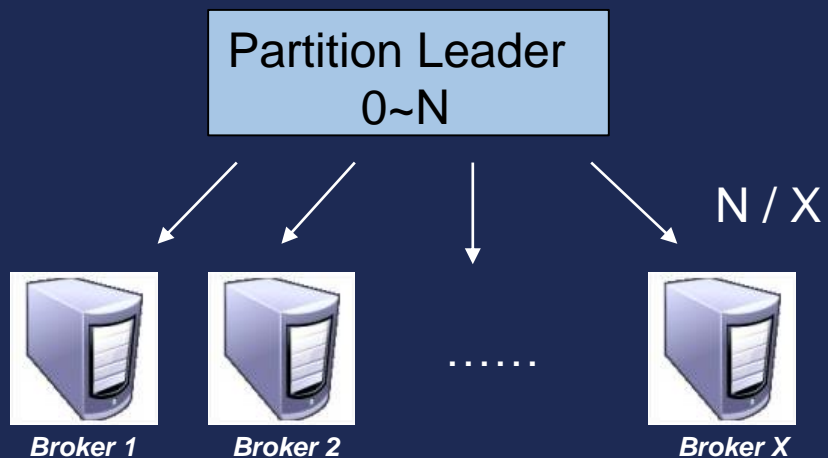


分区的创建、新增和迁移

■ Leader迁移

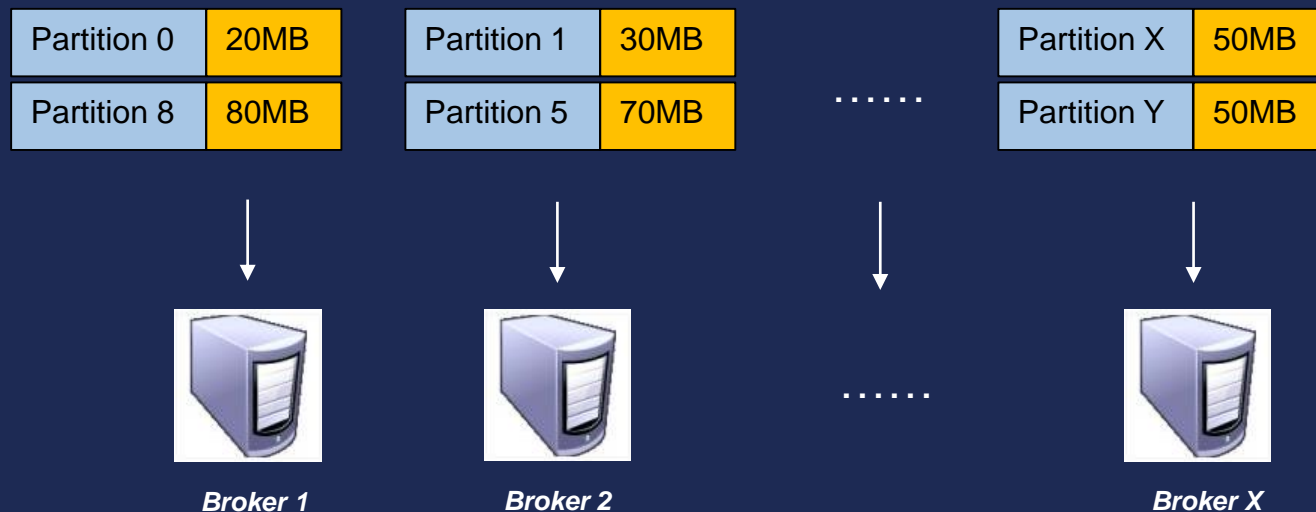
- Leader 分布
- 网络出流量

Rebalance Count



Broker for this instance

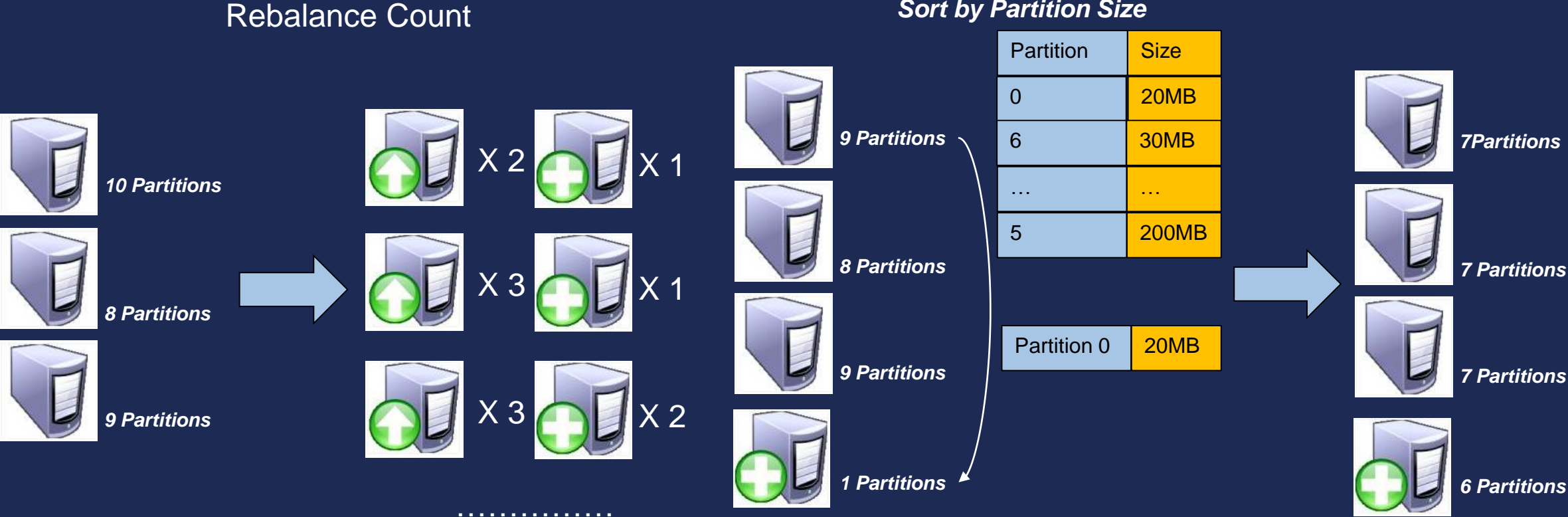
Rebalance Flow



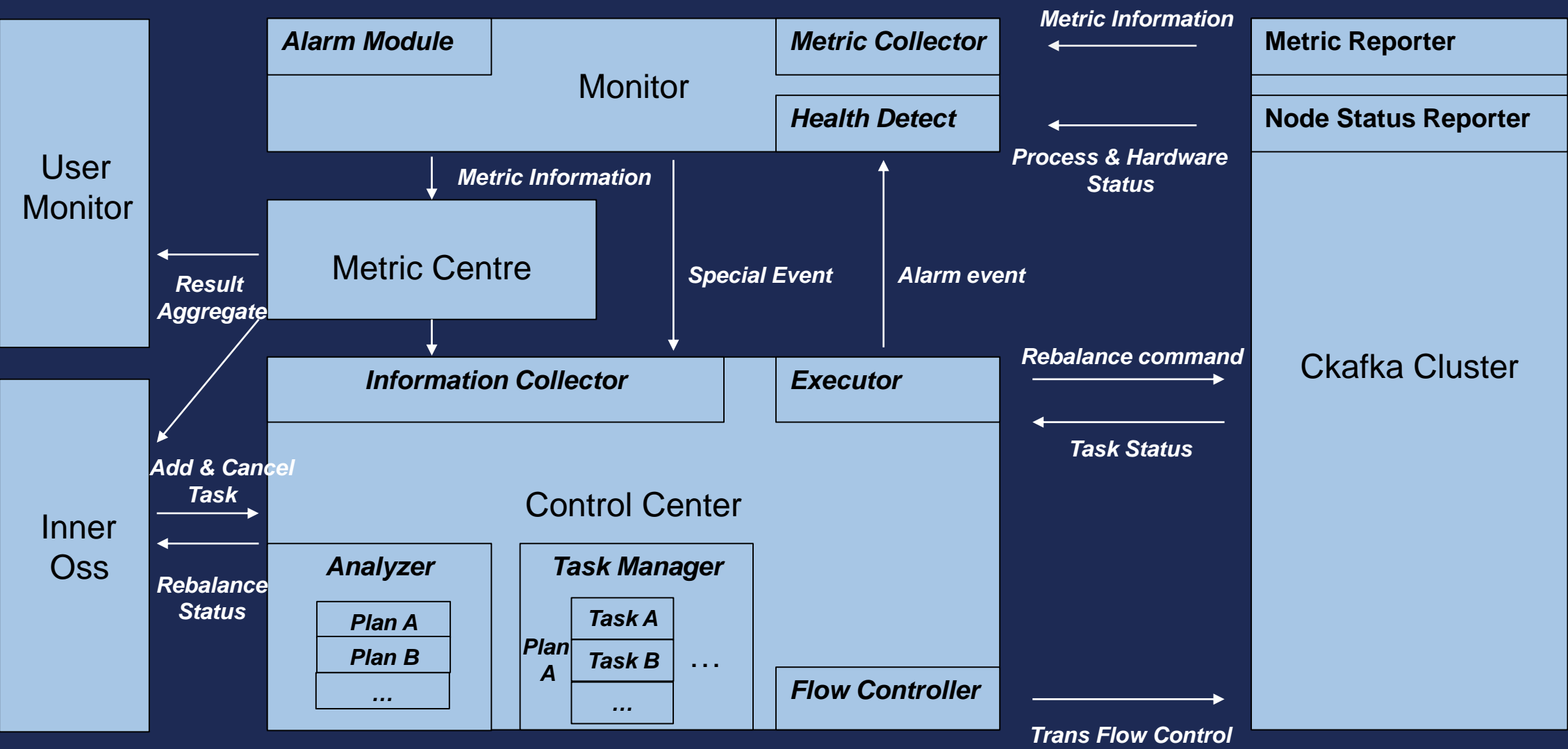
Broker for this instance

分区的创建、新增和迁移

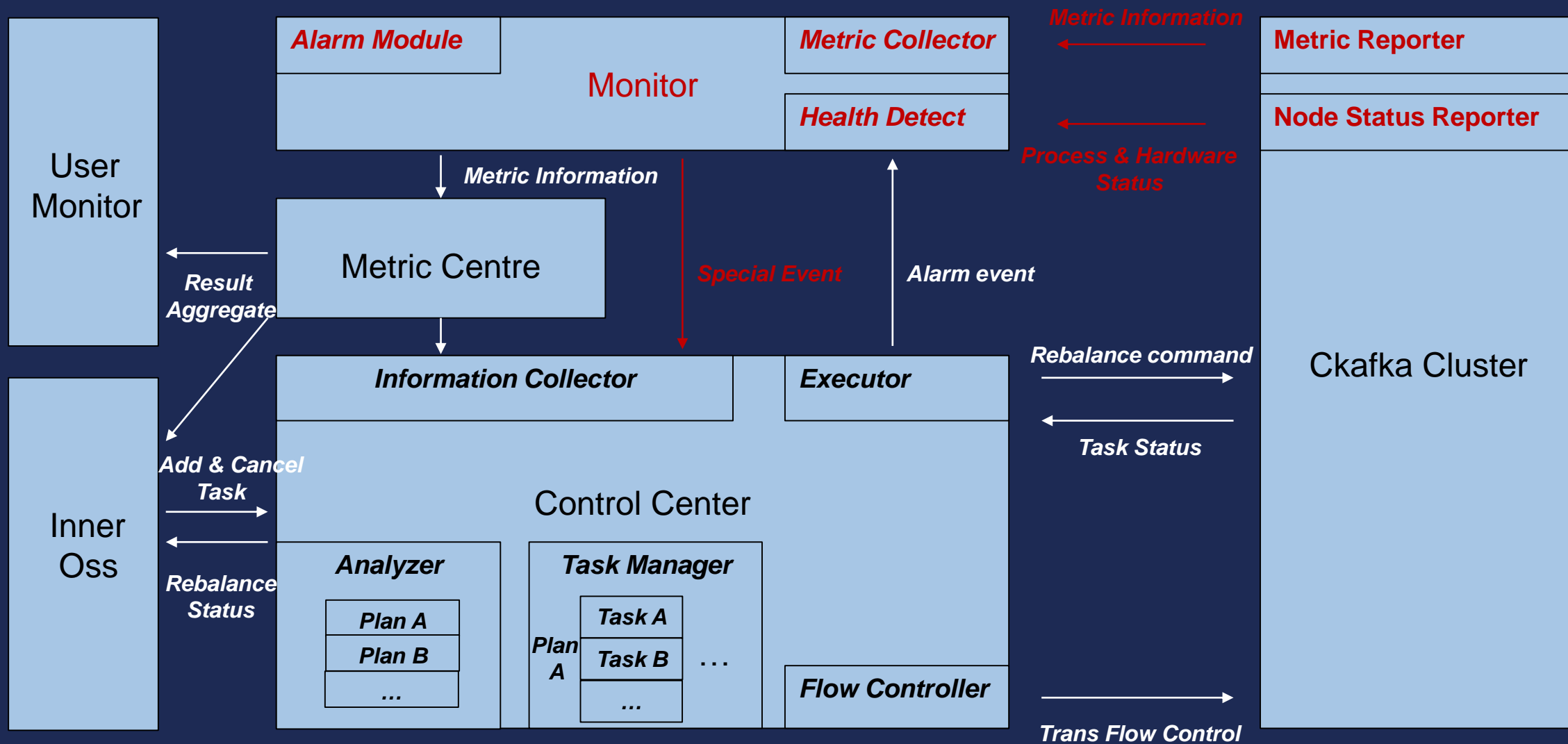
- Replica迁移
 - 资源利用率
 - 数据迁移代价



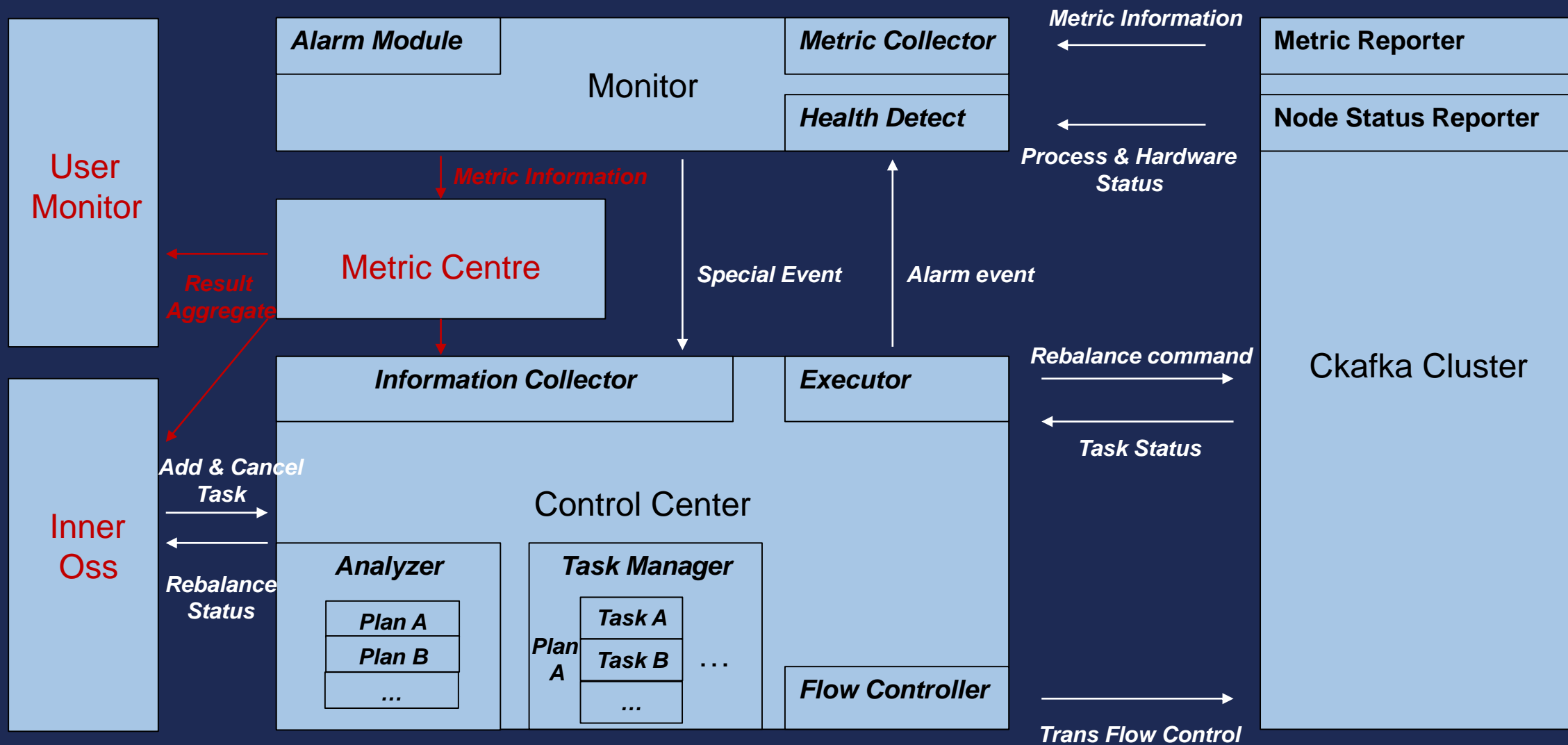
控制中心以及监控架构



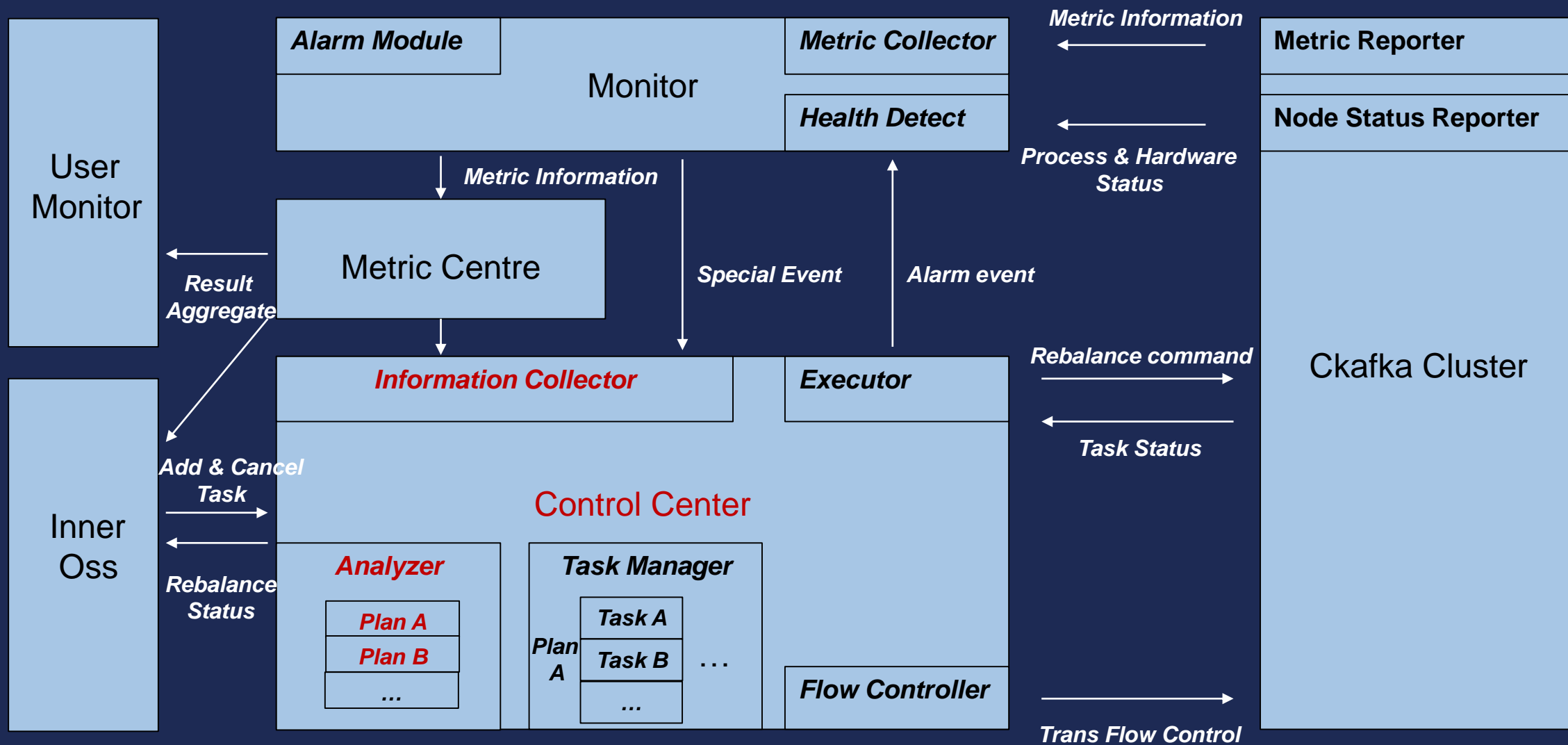
控制中心以及监控架构



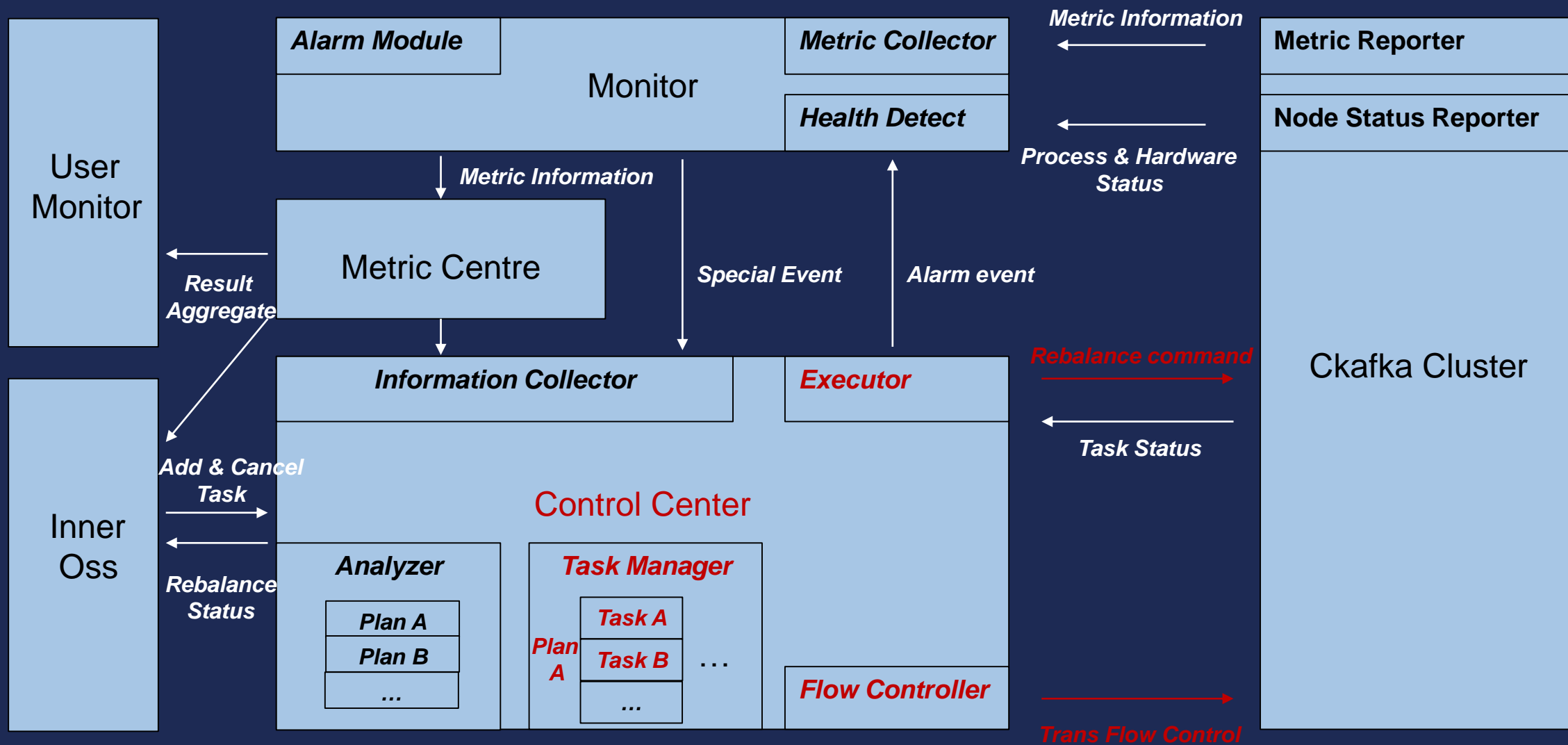
控制中心以及监控架构



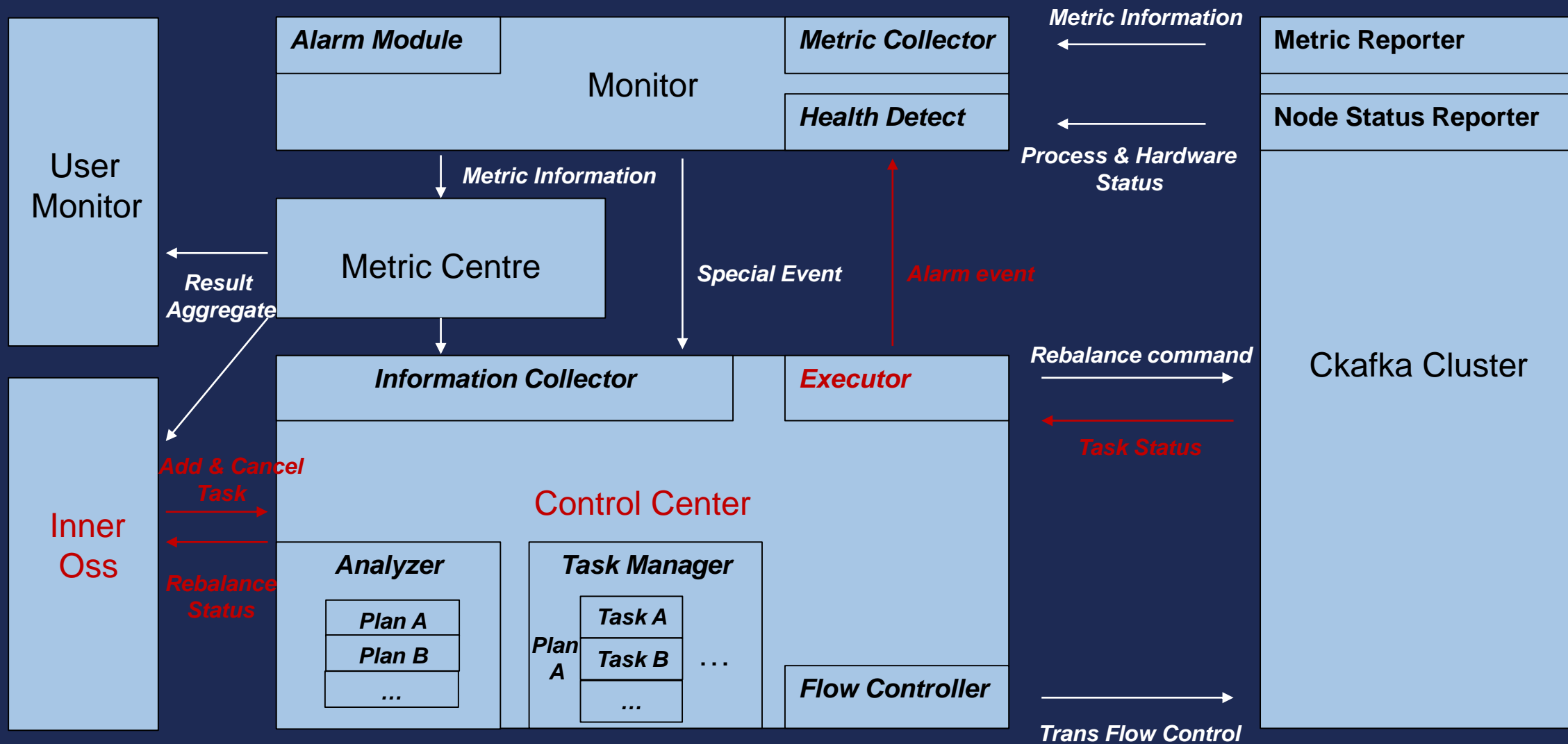
控制中心以及监控架构



控制中心以及监控架构



控制中心以及监控架构



未来展望

- 更多维度的调度决策
 - CPU \ MEN \ IO ...
 - 任务执行时间的衡量
- 预测调度
 - 负载增长、资源消耗
- 更高资源利用率
 - 实际售卖情况，策略优化

Thanks!



联合主办方:  腾讯云

|  开源中国
oschina.net

|  **kafka**
A distributed streaming platform

直播支持:  腾讯课堂
KE.QQ.COM 学习成就梦想