

Apache Pulsar在批流混合场景的 稳定性实践

背景介绍

场景介绍

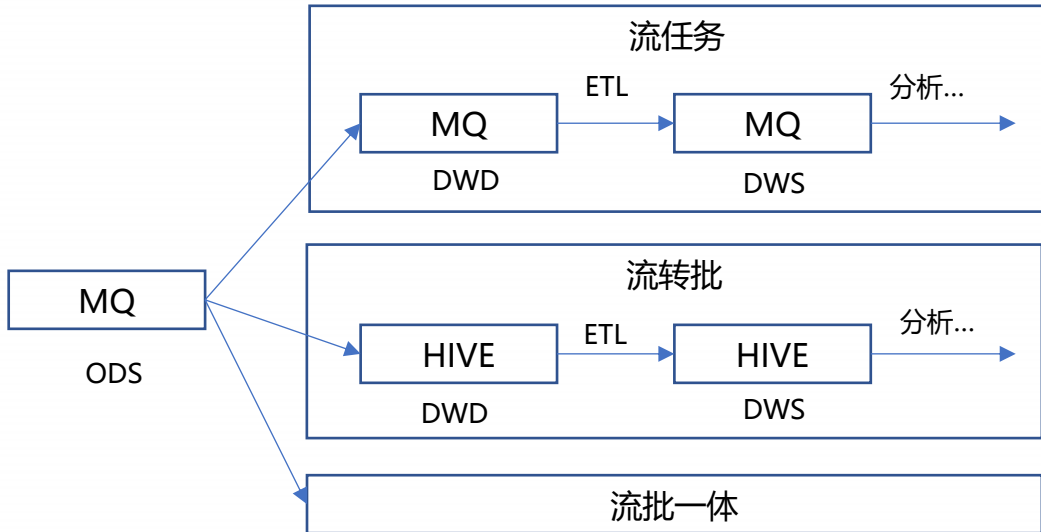
1. 要求实时性与准确性并重
2. 流任务保证实时更新
3. 定时全量批任务保证模型准确

常见场景

1. AI模型训练
2. 实时分析与统计报表

场景特点

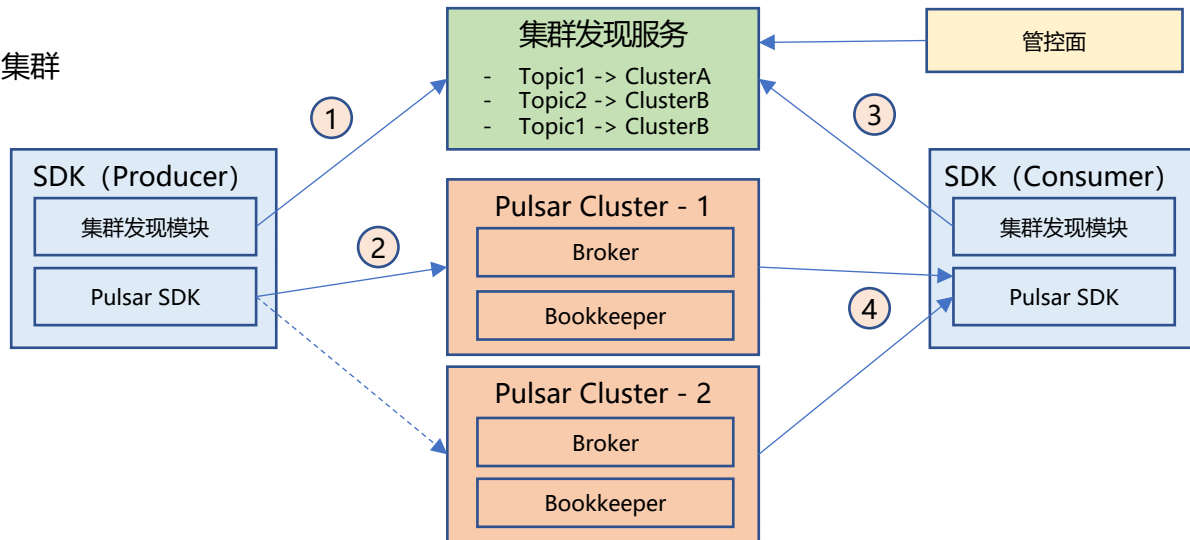
1. 大小包混合，集群CPU资源难以均衡
2. 缓存容易失效，经常冷读导致磁盘瓶颈
3. 场景互斥，配置难以统一



交互流程介绍

交互过程

1. 生产者通过Topic名称获得需要连接的集群
2. 生产者根据“集群发现服务”动态调整连接的集群
3. 消费者通过Topic名称获得需要连接的集群
4. 消费者与所有集群的同名Topic建立连接



混合调度实践

现象

- 实时场景CPU较高，离线场景带宽较高但CPU空闲
- 集群间资源不均，CPU/IO分别成为瓶颈

最佳实践

基于集群发现服务，跨集群动态调度Topic

效果

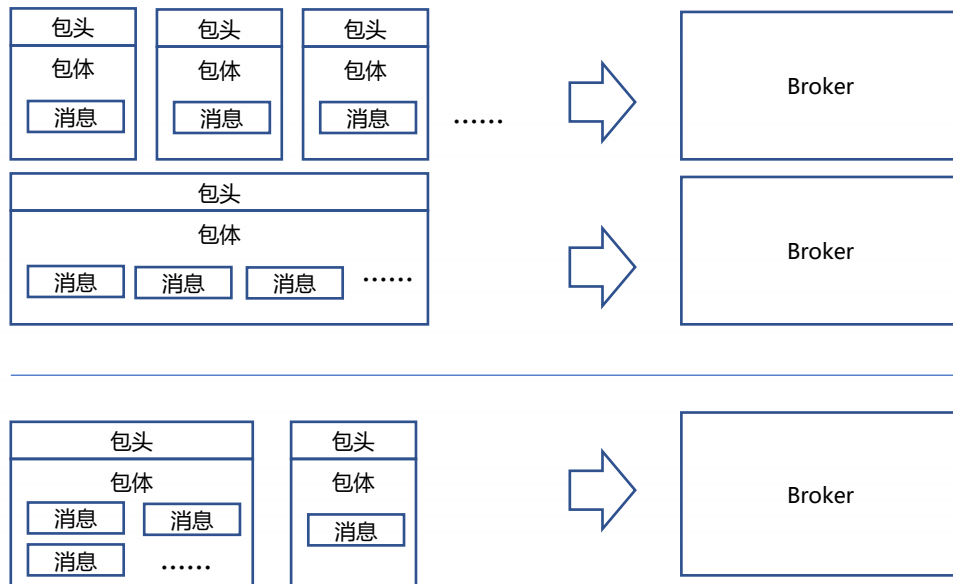
- 流量上涨，CPU反降
- 跨集群资源均衡

集群1

- 32C机器
- 流量200MB/s
- 单节点TPS 8w+
- CPU 80%+

集群2

- 32C机器
- 流量8GB/s
- 单节点TPS 8K
- CPU 30%+



预读实践

现象

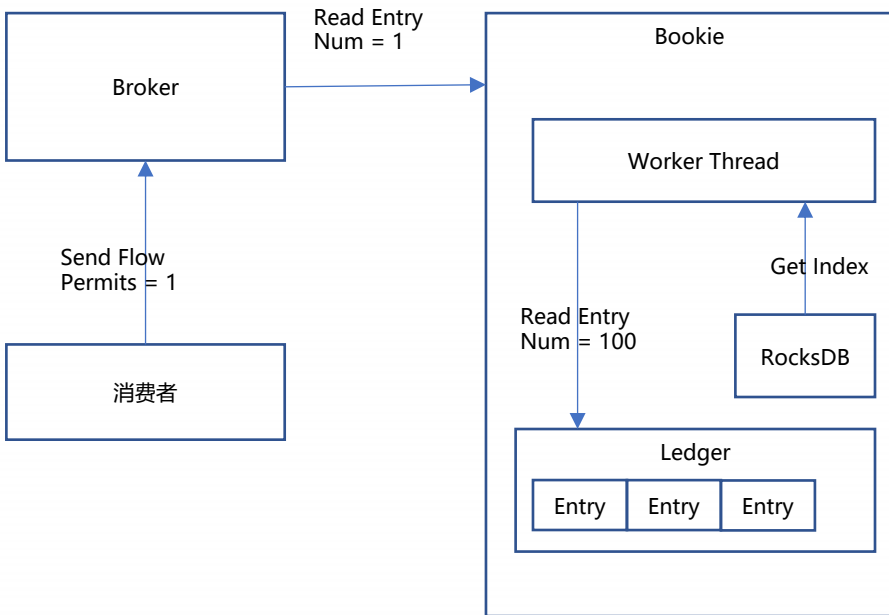
1. 消费小消息容易导致BK RocksDB 线程CPU高
2. 消费大消息容易导致IO瓶颈
3. 混合场景配置不好设

原因

1. BK ReadAheadBatchBytesSize 默认不开启
2. Broker预读Entry默认100条、5MB

最佳实践

- 500KB/批能兼顾大多数大小包场景
- 实现自适应预读大小



冷读实践

现象

1. 离线类、模型类任务容易出现从头消费并引起冷读
2. 在/离线任务重启, 缓存容易失效
3. 即使多Ledger盘, 冷读依然容易有磁盘热点

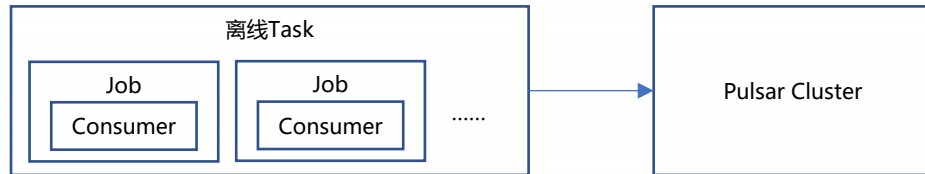
原因

文件无法保证完全打散, 有概率同时读一块盘

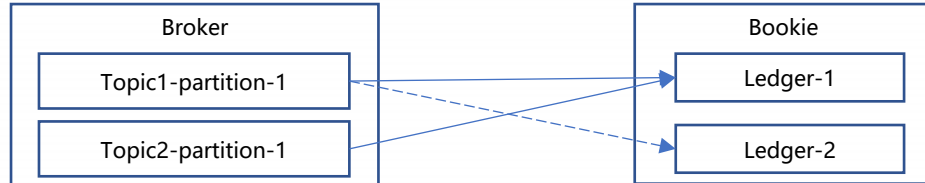
最佳实践

- 基于时延的背压权重
- 丢弃超时读请求

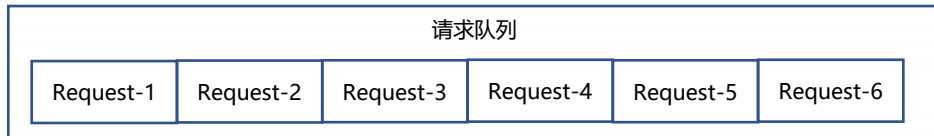
任务重调度

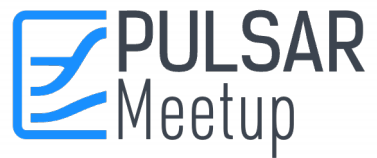


热点冲突



BK丢弃超时读请求





Thanks