
Architect
SACC
2022 中国系统架构师大会
SYSTEM ARCHITECT CONFERENCE CHINA 2022
激发架构性能 点亮业务活力
云上会议 网络直播 | 2022年10月27-29日
ITPUB

小红书近线服务统一调度平台建设实践

容器架构负责人+高会军

大纲

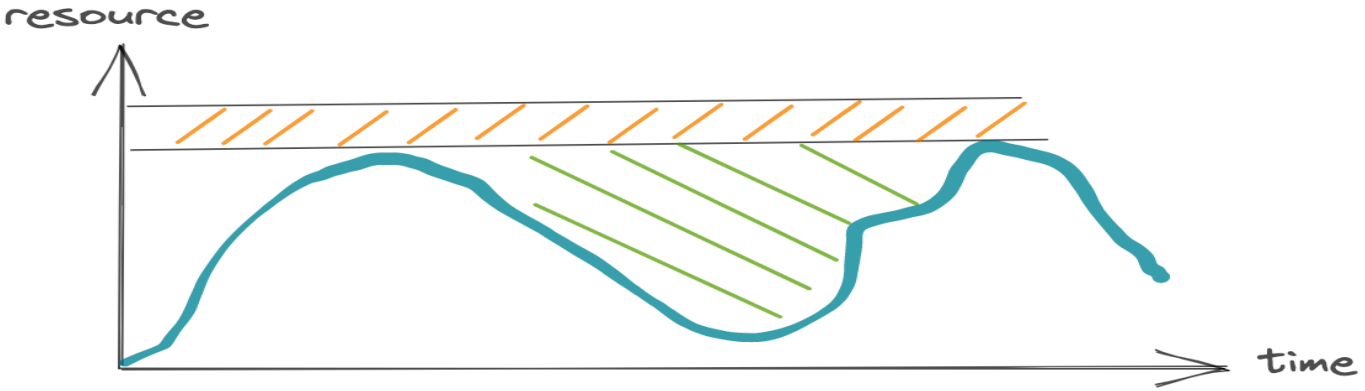
- 背景和思考：为什么要建设一个近线服务统一调度平台
- 解决方案与架构
  - 服务Qos资源保障模型
  - 整体架构
    - 统一入口
    - Virtual Kubelet
    - 调度
    - 弹性
- 收益
- 未来规划

背景和思考——什么是近线服务

服务类型	服务特征	常见服务	资源保障要求
在线	主要进行用户请求的实时处理，需要更快地响应最近的事件和用户交互，因此对于**延迟敏感**。	API服务	高
离线	**延迟不敏感**，更重视吞吐。	批处理、离线训练、回扫任务	低
近线	一般基于数据消息队列，进行数据的准实时处理。它居于离线和在线之间，既可以以**分钟级别甚至秒级的延迟**来准实时地处理数据，也有一定的数据批量处理能力。	抽帧、转码、实时索引构建、在线训练	中

背景和思考——面临的问题

- 近线服务和在线服务一样使用保障能力最高的计算资源，**成本不是最优**。
- 更多近线服务受限于成本控制，申请不到资源，无法上线，进而**影响业务收益**。
- 在线服务**冗余计算资源闲置**。



背景和思考——建设近线服务统一调度平台

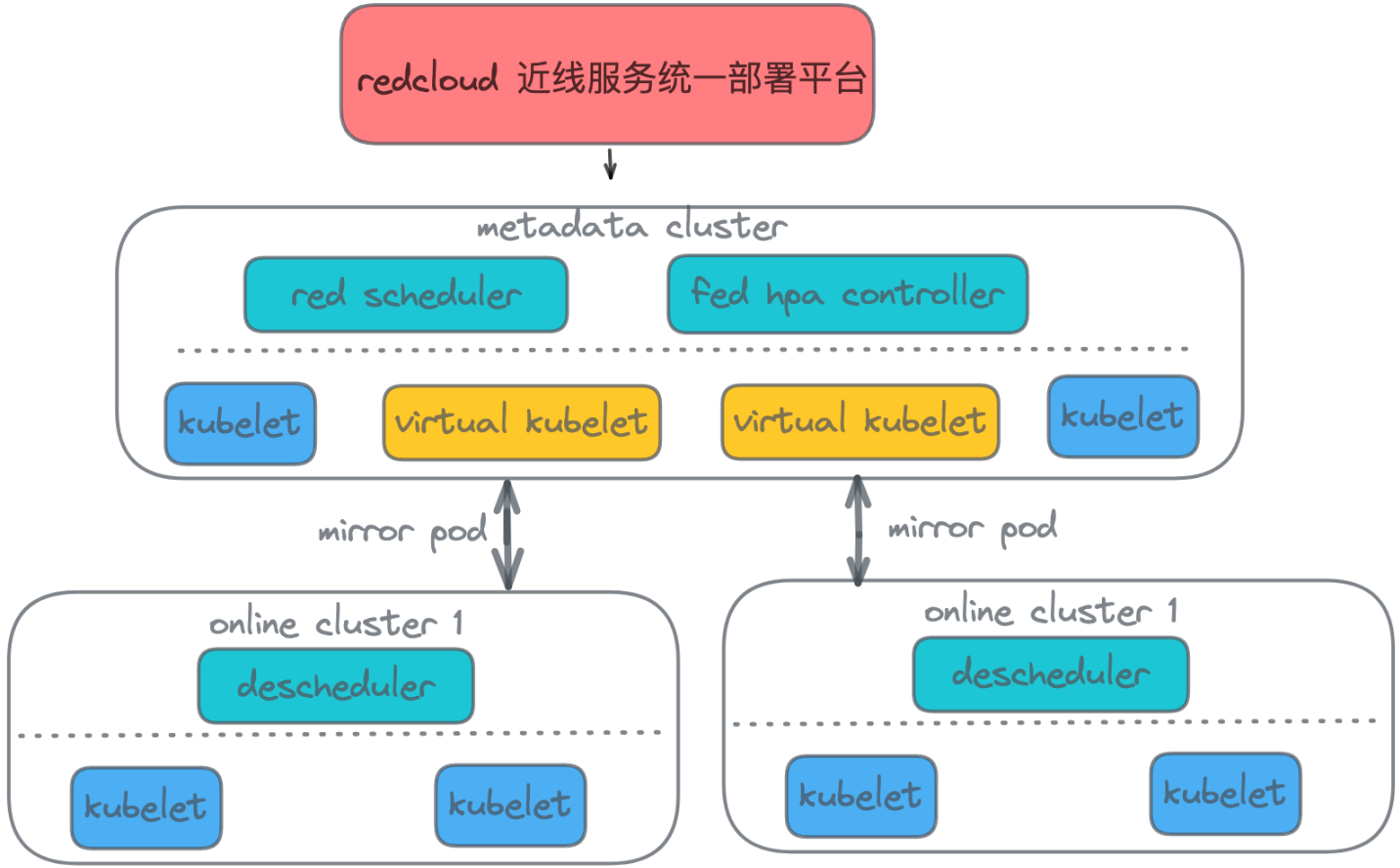
- 站在整个公司计算资源视角看待问题
- 统一的近线服务入口
- 统一调度和管理
- 差异化云原生能力支持
- 近线服务容错性改造

解决方案——算力来源和资源保障模型

资源类型	资源生命周期	成本	普适性
独占资源池机器	高	中	高
在线集群闲置算力（包含buffer池资源+分时算力）	中，天级别到周级别	免费	高
混部算力	低，小时级别到天级别	免费	低，尤其是针对大内存服务
公有云容器实例服务	高	高	中

综合以上三个维度，近线服务调度优先级为：**独占资源池机器 > 在线集群闲置算力 > 混部算力 >** 公有云容器实例服务

解决方案——方案与架构

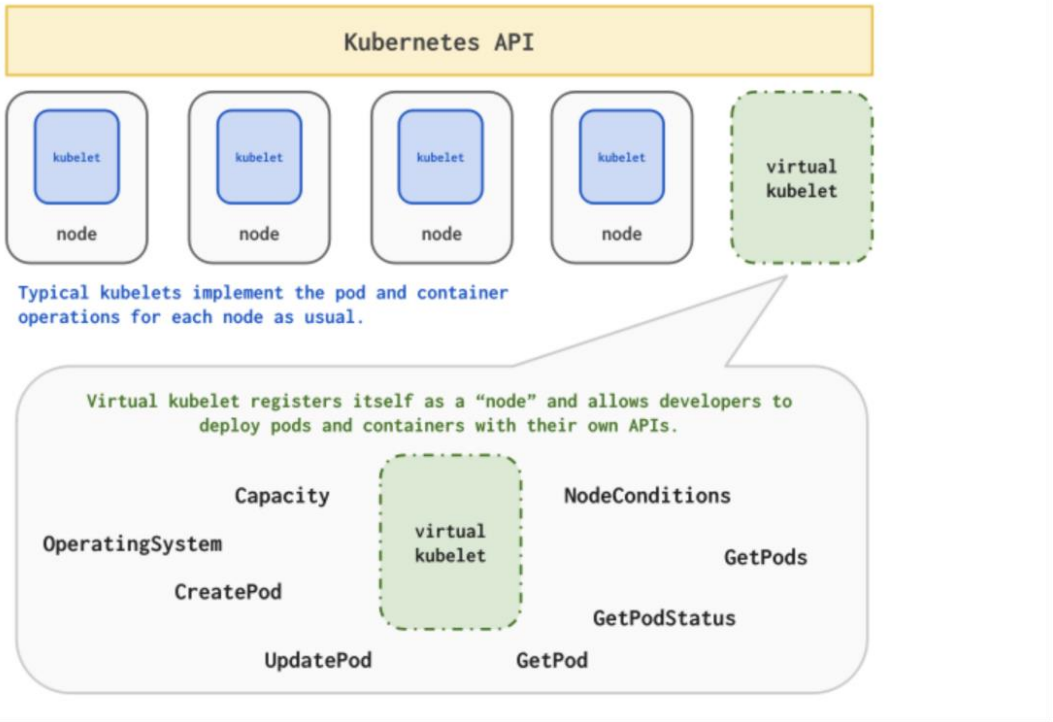

redcloud 近线服务统一部署平台
metadata cluster
red scheduler
fed hpa controller
kubelet
virtual kubelet
virtual kubelet
kubelet
mirror pod
mirror pod
online cluster 1
descheduler
kubelet
kubelet
online cluster 1
descheduler
kubelet
kubelet

解决方案——统一入口

- 抽象近线服务应用模型
- 发布管控
- 可视化云原生支持能力

解决方案——Virtual Kubelet

- 主要功能：1）属性转换 2）状态同步 3）资源管理

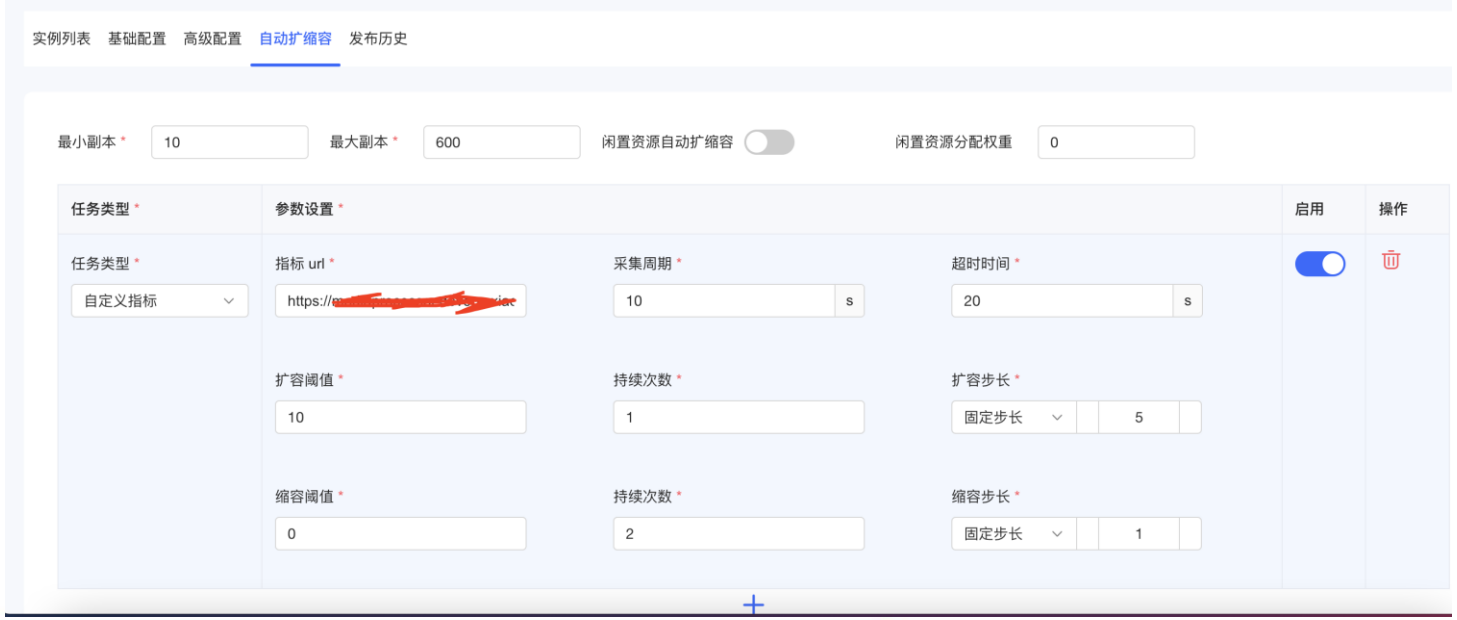


解决方案——调度

- red scheduler：优先级调度，抢占调度
- descheduler：1）对于独占资源池机器，当业务pod pending时，负责近线服务实例退场。2）对于buffer池机器，配合CA 做近线服务实例退场。

解决方案——弹性

- 丰富的弹性策略支持：支持 cron + 自定义指标 + 消息队列堆积数 + 闲置资源余量等多种策略
- 多集群联邦HPA



收益

- 成本优化上，小红书 **10 wc+** 的近线服务 **0 计算成本**运行
- 服务质量上，提升了近线服务的处理能力
- 沉淀了一套服务QOS资源保障能力模型

未来规划

- 服务QOS资源保障模型向智能化演进
- 联合消息中间件团队，建设资源和业务双友好的serverless事件平台
- 统一调度平台支持更多离线服务

欢迎加入小红书容器架构团队！

扫一扫上面的二维码图案，加我为朋友。

THANKS