

Kubewharf: 打造更稳定高效云原生 底座的工具集

唐鹏程 Katalyst maintainer



字节云原生技术栈



集群联邦管理节点数

210,000+

最大集群节点数上万，实现大规模集群落地

平均每日变更数高达

30,000

次，云原生为前线业务提供稳定性保障

拥有

100,000+

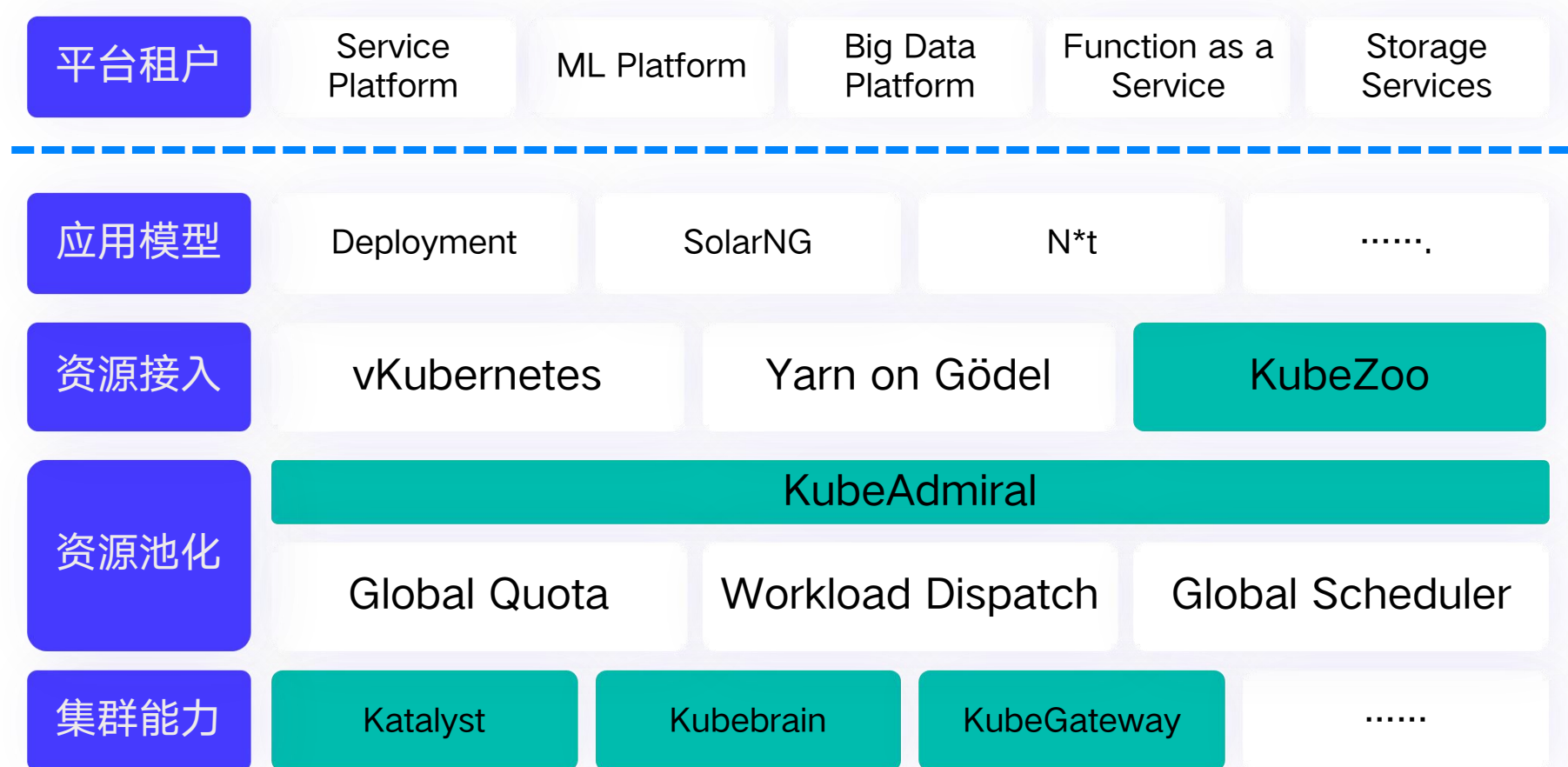
在线微服务，敏捷化构建能力持续增强

离线任务数

240 M+

通过云原生混部，大规模节省企业资源成本

Kubewharf

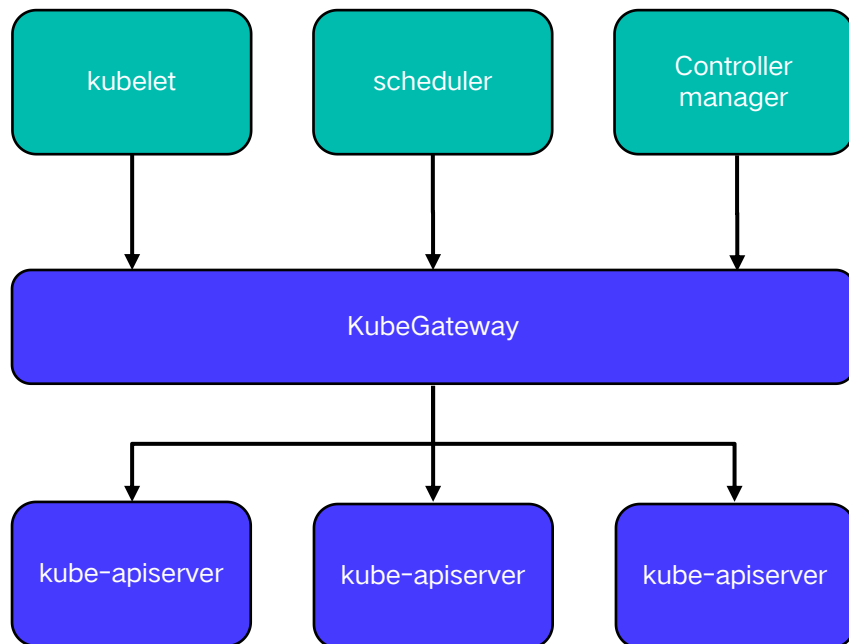


大规模集群稳定性

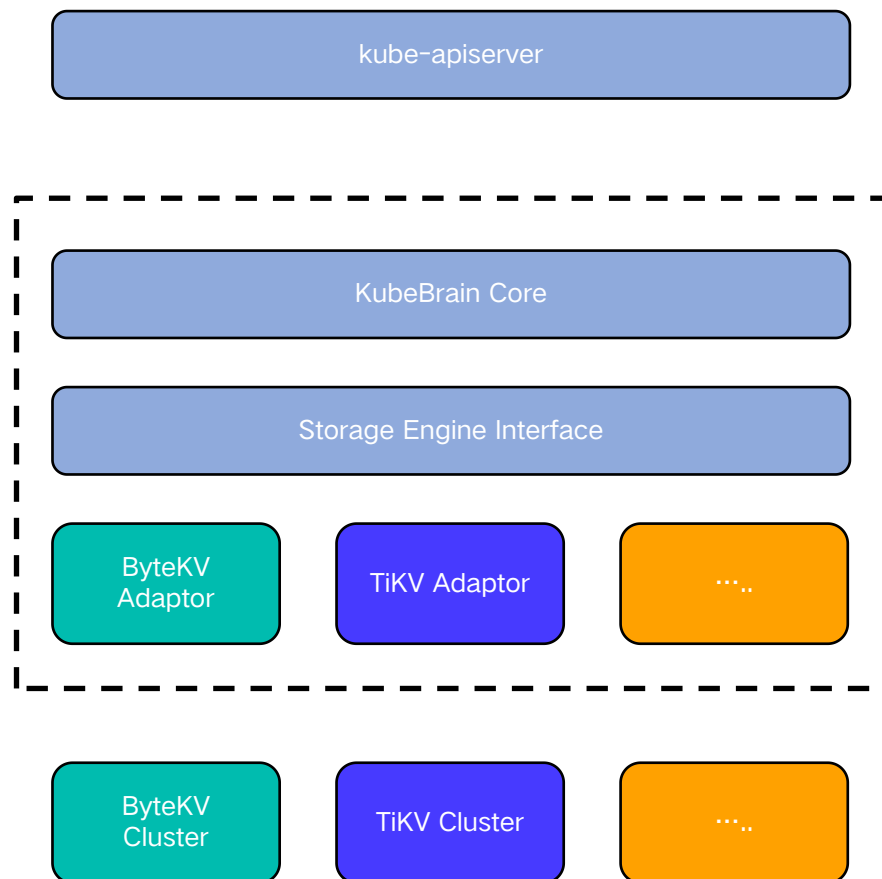


KubeGateway

专为 kube-apiserver 设计及定制的七层 负载均衡代理



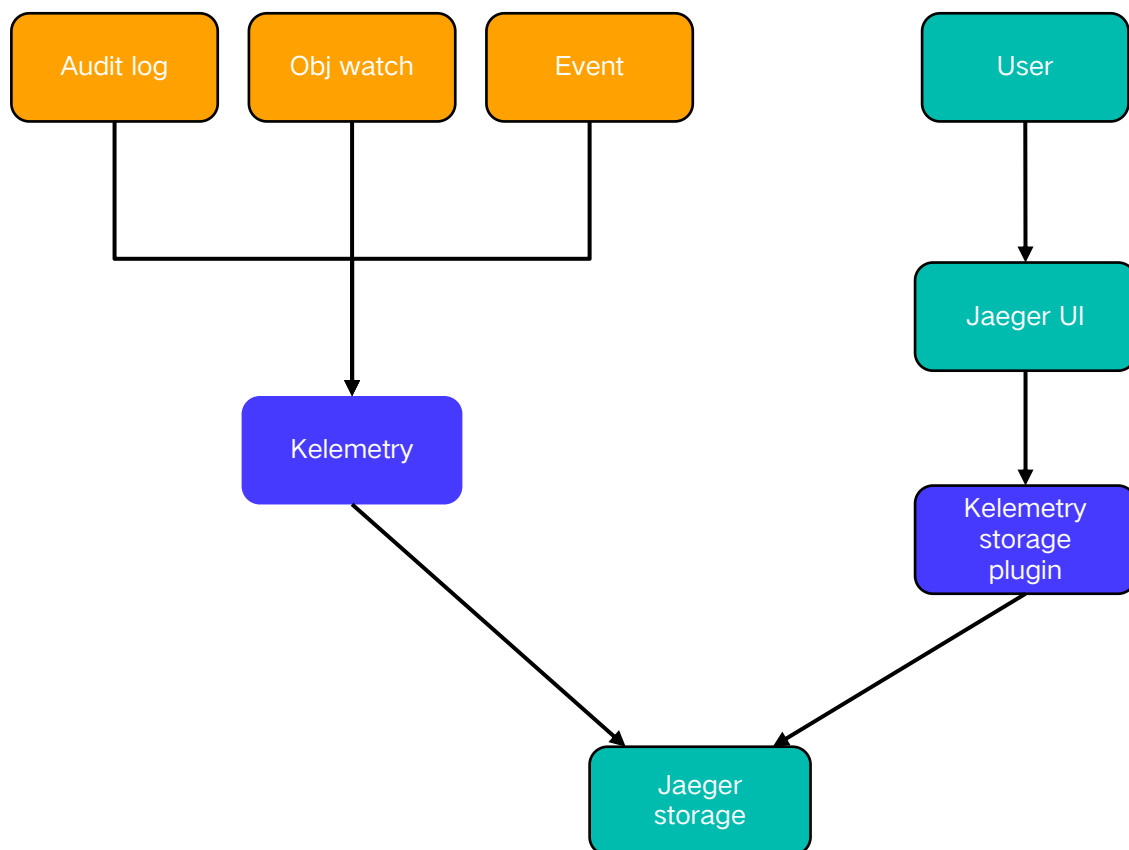
- HA k8s 集群中，L4 LB 无法均衡部分 k8s 中常见的请求流量（e.g. HTTP/2.0）
- KubeGateway 是一个透明的 L7 LB，专门用来均衡 kube-apiserver 的 HTTP/2.0 请求
- 支持基于 resource, verb, user, namespace, apigroup 等信息的路由策略及流量控制手段，如限流，降级等
- 在字节跳动，KubeGateway 已经平滑的接管了 K8s 集群，总 QPS 20w+



高性能元信息存储系统

- 在 k8s 集群规模到 10k+ 节点时, etcd 很容易成为瓶颈
- KubeBrain 可以替代 etcd 成为 k8s 的元信息存储系统
 - 向上实现了 kube-apiserver 需要的所有 storage server interface
 - 向下抽象了一套存储引擎接口, 可以对接各类 kv 存储, e.g. ByteKV, TiKV etc
- 通过 KubeBrain, 字节单集群规模可达 20k+ 节点, 百万 pod

Kelemetry



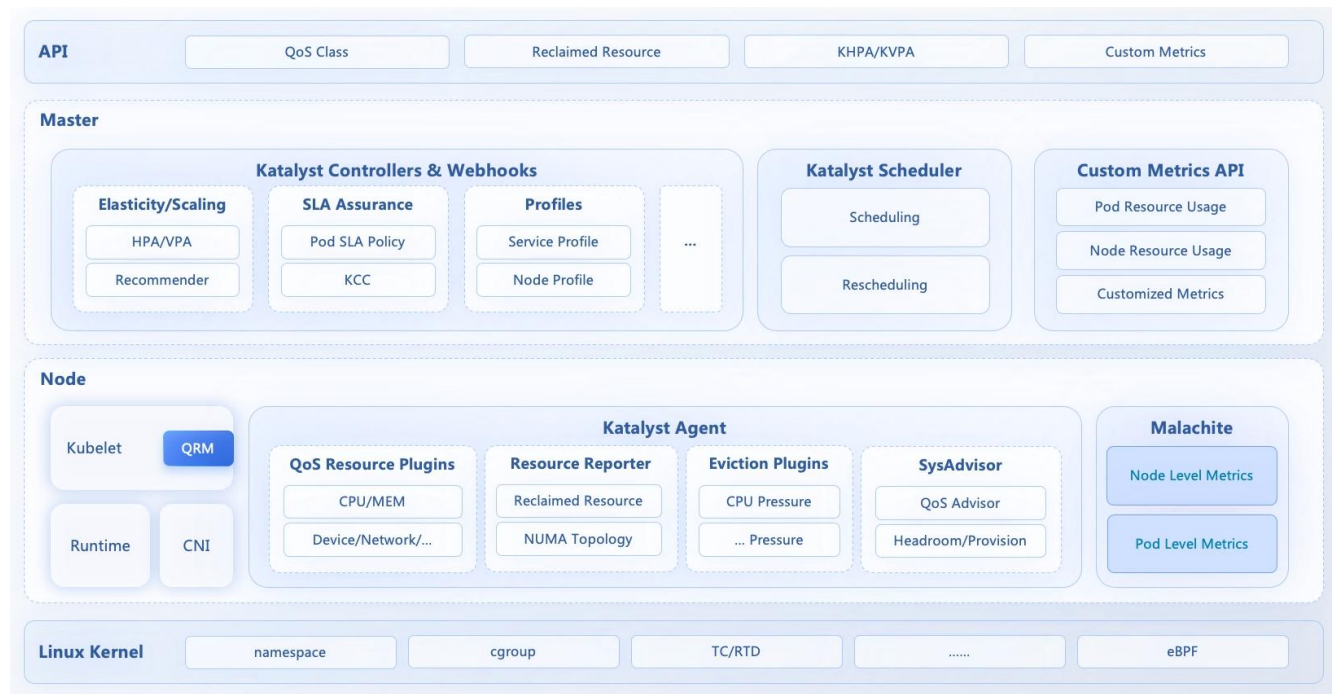
面向 Kubernetes 控制面的 全局追踪系统

- k8s 提供声明式、异步、分布式的 API, 事件之间的因果关系难以分析
- Kelemetry 可以追踪单个 k8s 对象的完整生命周期以及不同对象之间的相互影响
 - 根据对象的从属关系来构建 span 的父子关系
 - 支持从 audit log, controller event 等多种数据源生成 trace span
 - 支持收集来自多个集群的事件, 实现对跨集群组件的事件追踪

资源管理及成本优化



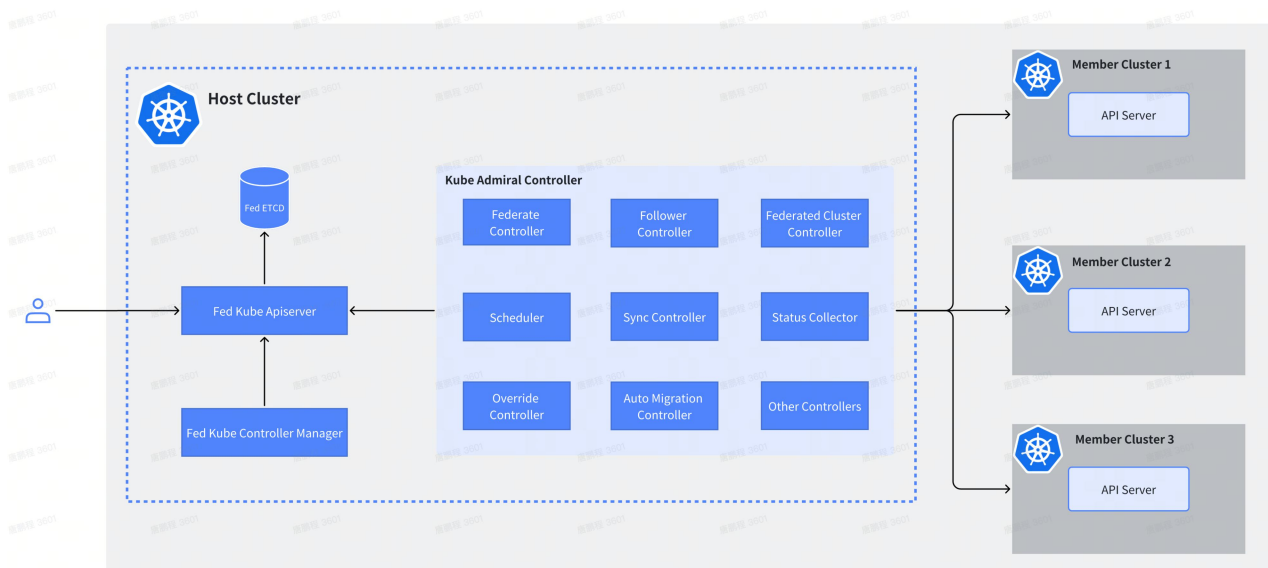
在离线混部



- K8s 正成为各类业务的统一底座，可以实现更精细化的资源管理，优化整体资源成本
- Katalyst 包含了
 - 基于 QoS 的资源模型抽象
 - 精细化资源分配、隔离能力
 - 异构设备及微拓扑调度管理
 - 其他资源弹性管理能力
- 核心业务集群实现了整机天级利用率从 23% 到 60% 的提升

KubeAdmiral

多云多集群调度管理



- 随着云原生发展，多云、多集群的模式成为资源管理的常态
- KubeAdmiral 为多集群管理提供统一入口
 - 提供兼容原生 k8s API 的接口
 - 支持公有云、私有云、用户自建等多种 k8s 集群管理
 - 提供灵活的调度框架和丰富的调度策略
- 支撑字节集团业务平台，管理超过 21 万台机器、1000 万+ pod

加入社区





Thanks.

