GOPS2016 Shanghai

GOPS

2016 全球运维大会

重新定义运维

上海站

会议时间： 9月23日-9月24日

会议地点： 上海·雅悦新天地大酒店

主办单位： 开放运维联盟 OOPSA Open OPS Alliance　高效运维社区 GreatOPS Community

指导单位： 数据中心联盟 Data Center Alliance

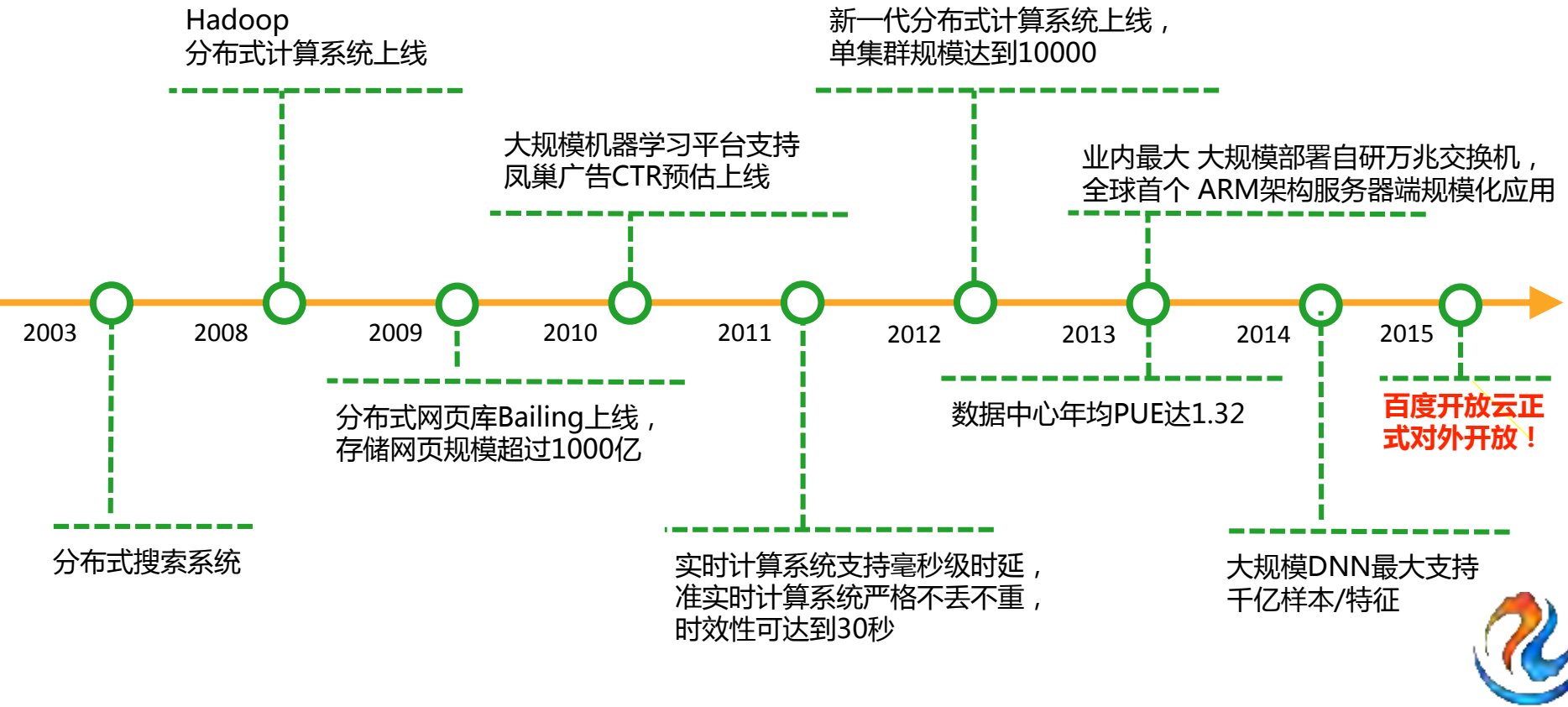
Openstack在百度开放云的系统改进

杨一 百度开放云

目录

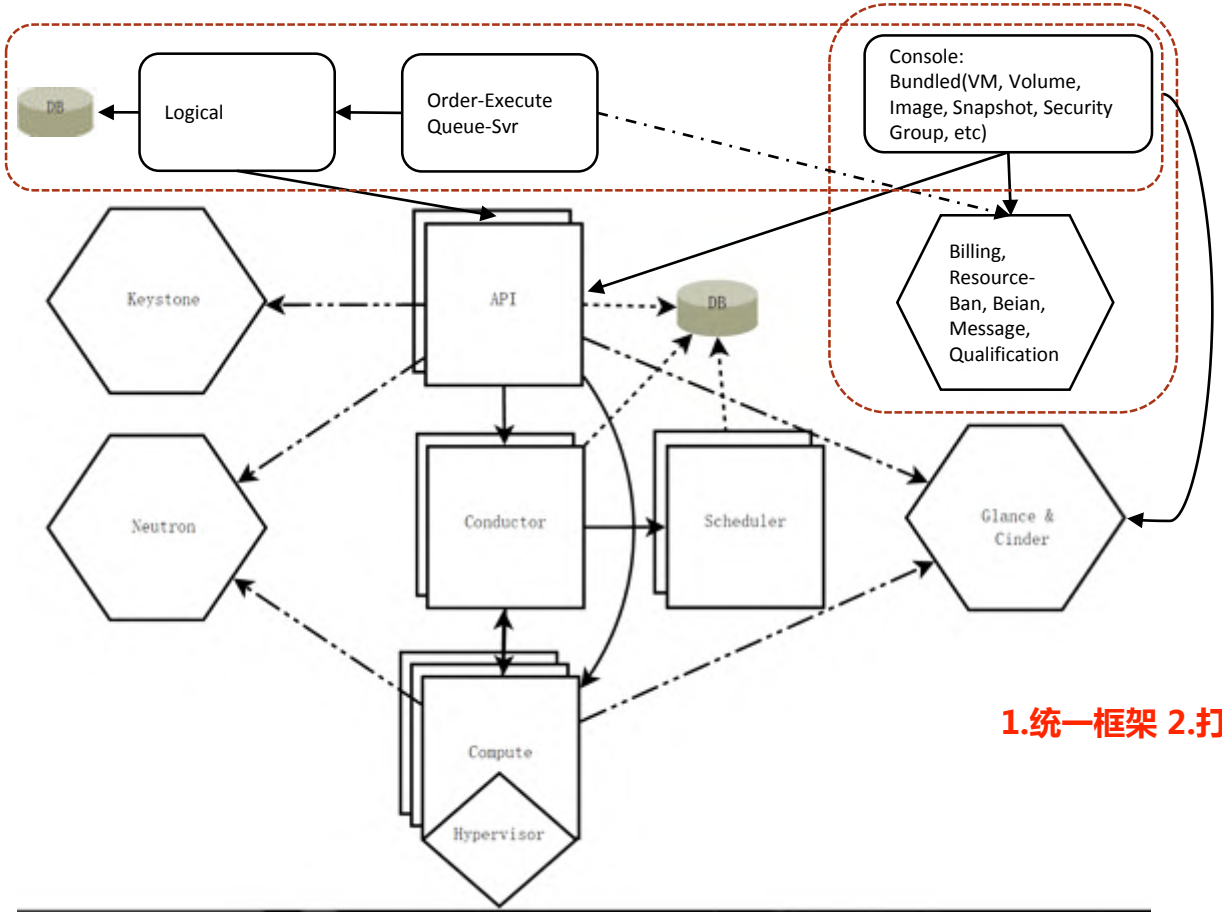
缘起 - 百度开放云的技术积累

天生的云计算公司：技术积淀 & 开放资源



目录

画皮 - 开放云初期的计算平台系统



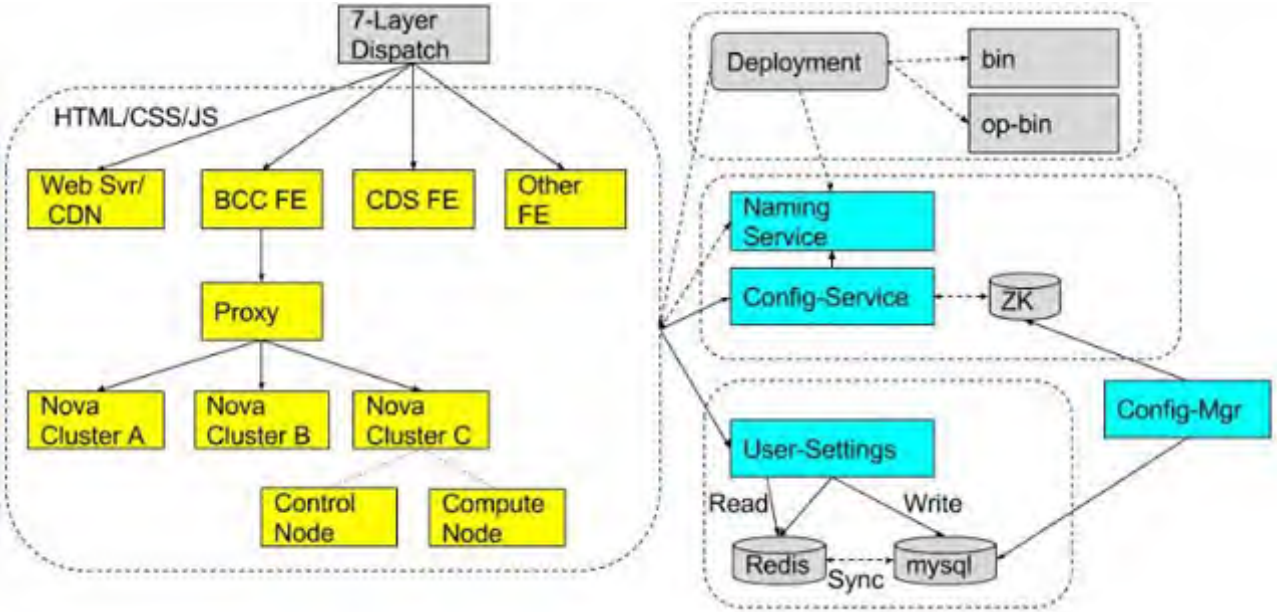
- 快速发布产品，了解市场
- 拥抱开源技术，坚定自信
- 提升可运维性，稳定系统
- 面向服务设计，便于重构

从0到1，生存是第一位

1.统一框架 2.打通客户使用场景 3.构建业务支撑体系

目录

换骨 – API改造之服务拆分，管理和配置



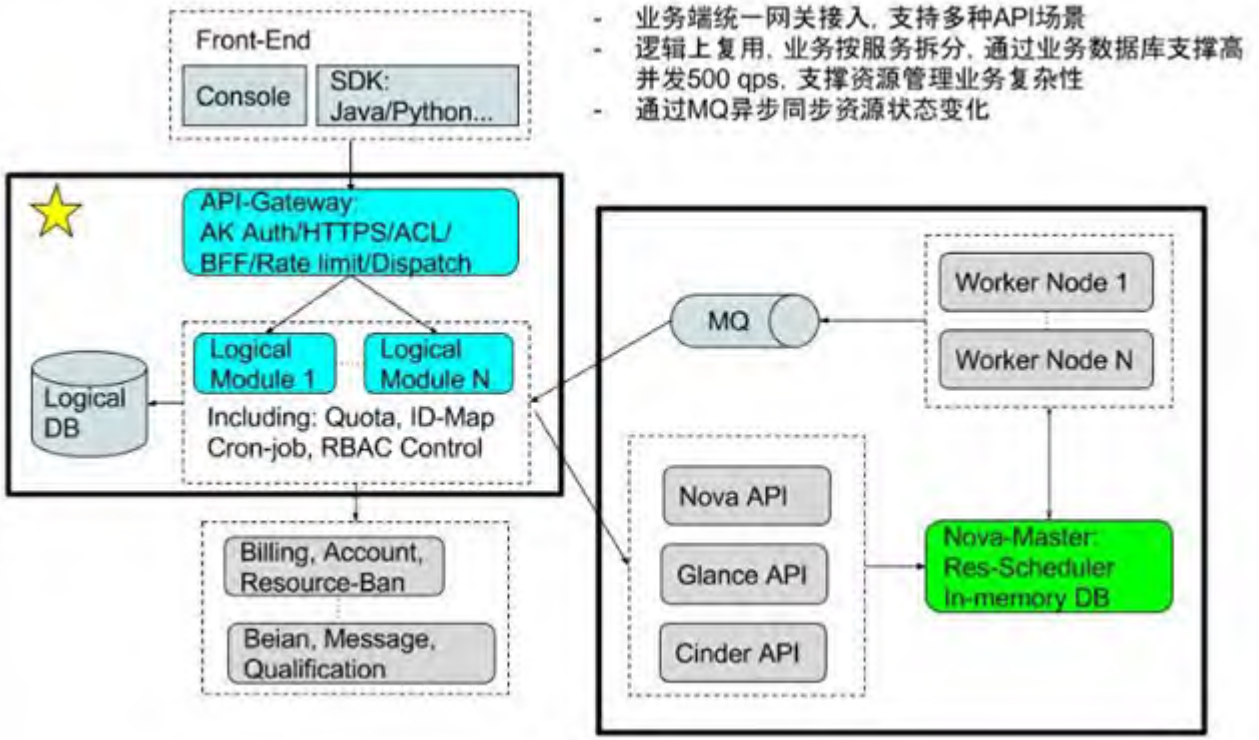
问题

- 多Feature，多团队协同低效
- 功能耦合度高
- 配额/白名单/小流量/灰度发布
- 配置管理困难

解决

- 前端模块7层协议拆分
- 命名服务和服务配置中心
- 用户配置服务
- 一键上线平台

换骨 - API改造之微服务构建逻辑层



问题

- 性能差，接口不符合业务需求
- API版本多，代码重复
- 部分功能实现成本高昂

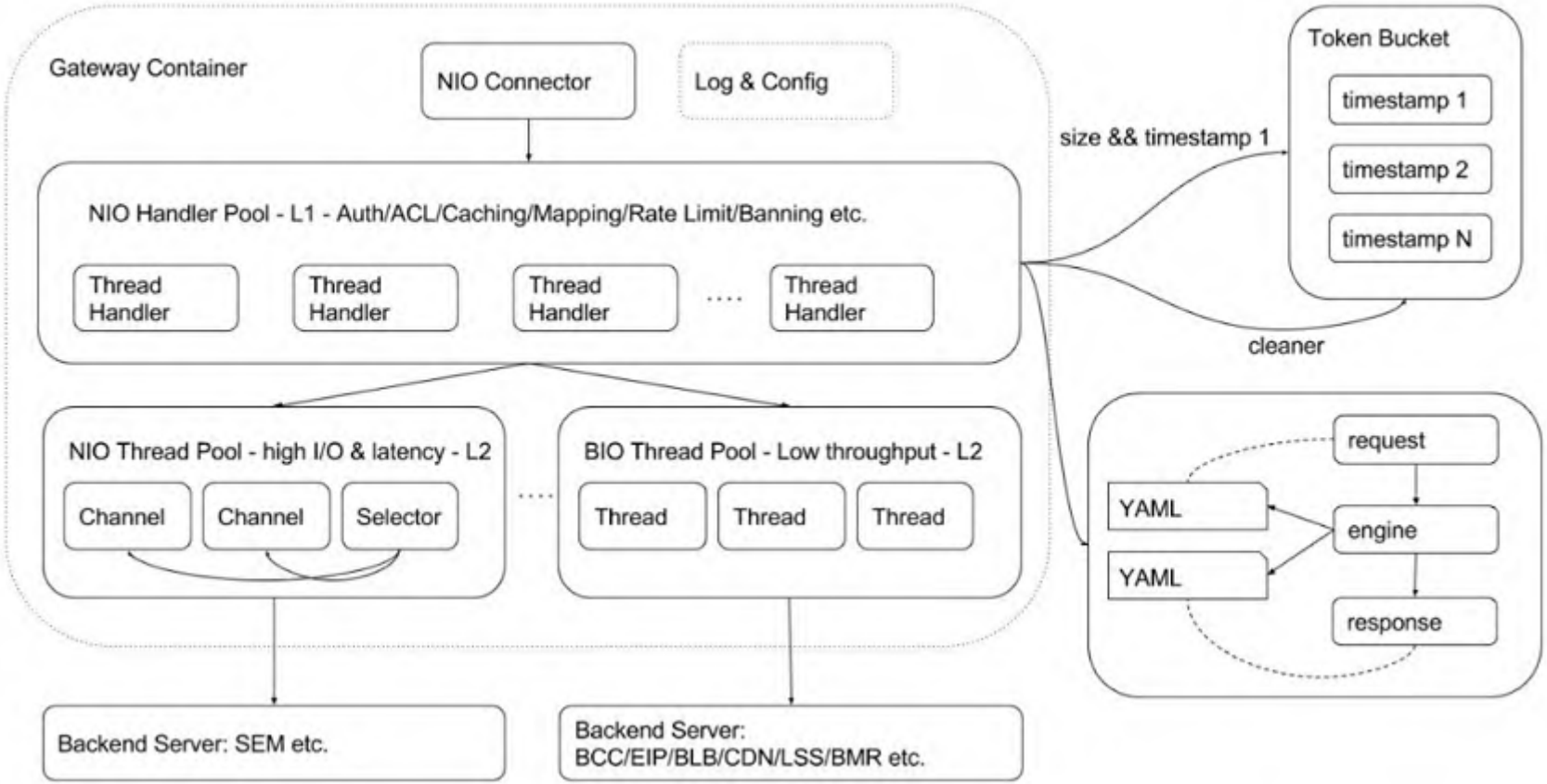
解决

- 构建业务逻辑层扛Openstack压力
- 加入服务网关 & BFF
- 空间换时间，用户视角组织数据和逻辑 & MQ同步状态

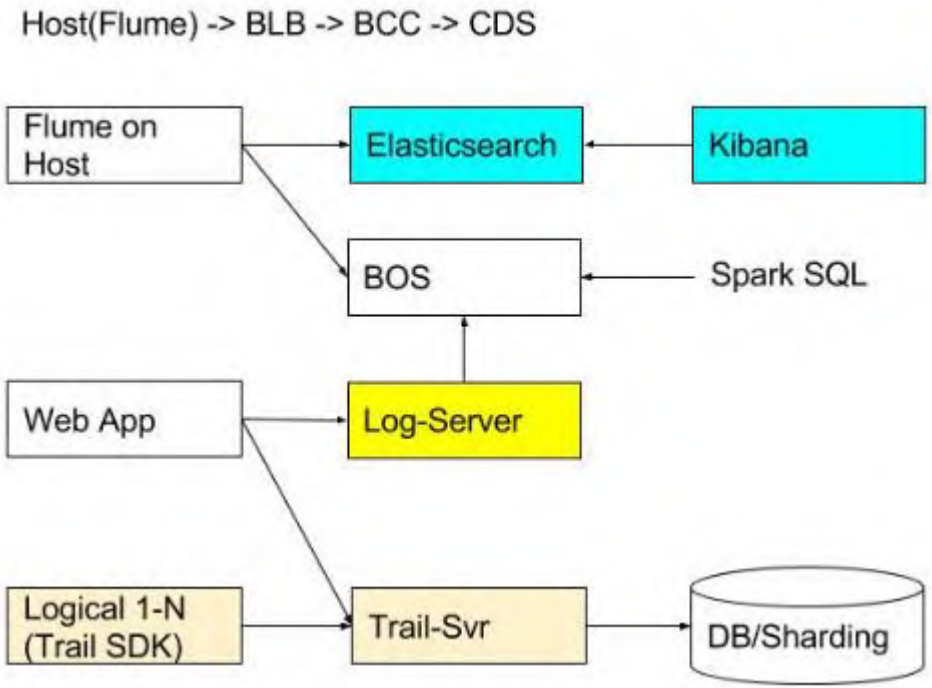
1.权限 2.定时任务 3.Quota 4.统一ID-Mapping 等多种公共服务

换骨 – API改造之统一服务网关

1.多通道异步转发 2.基于YAML的BFF 3.令牌桶:平滑限速/封禁

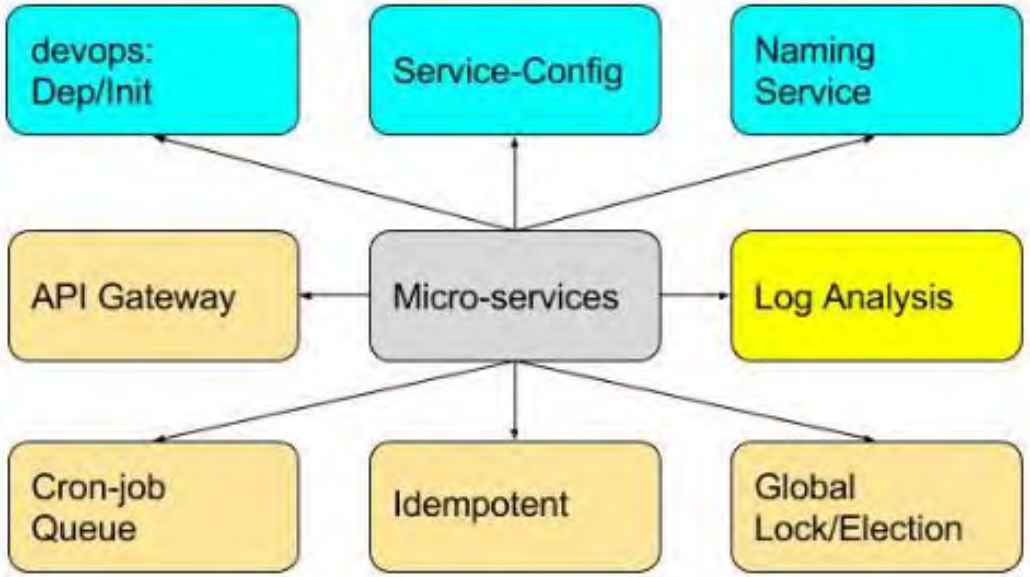


换骨 – 4份日志：Trace日志，分析日志，行为日志，操作日志



- **Tracing日志**，统一request-ID，检索系统日志，类EFK
- **分析日志**：API QPS，异常统计，访问量统计，功能使用率；通过Baidu Spark集群 + Spark SQL完成
- **行为日志**：支持记录用户行为到后端
- **操作日志**：支持客户自我查看和追踪操作状态

换骨 – 开放云API微服务构建组件概览

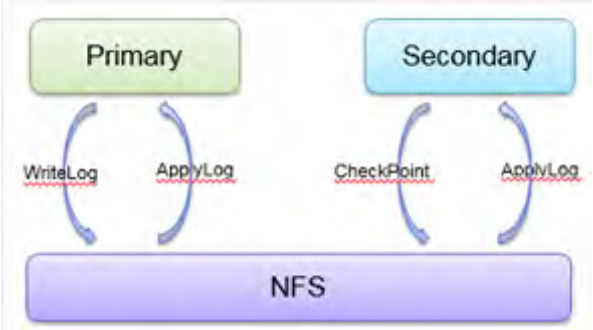


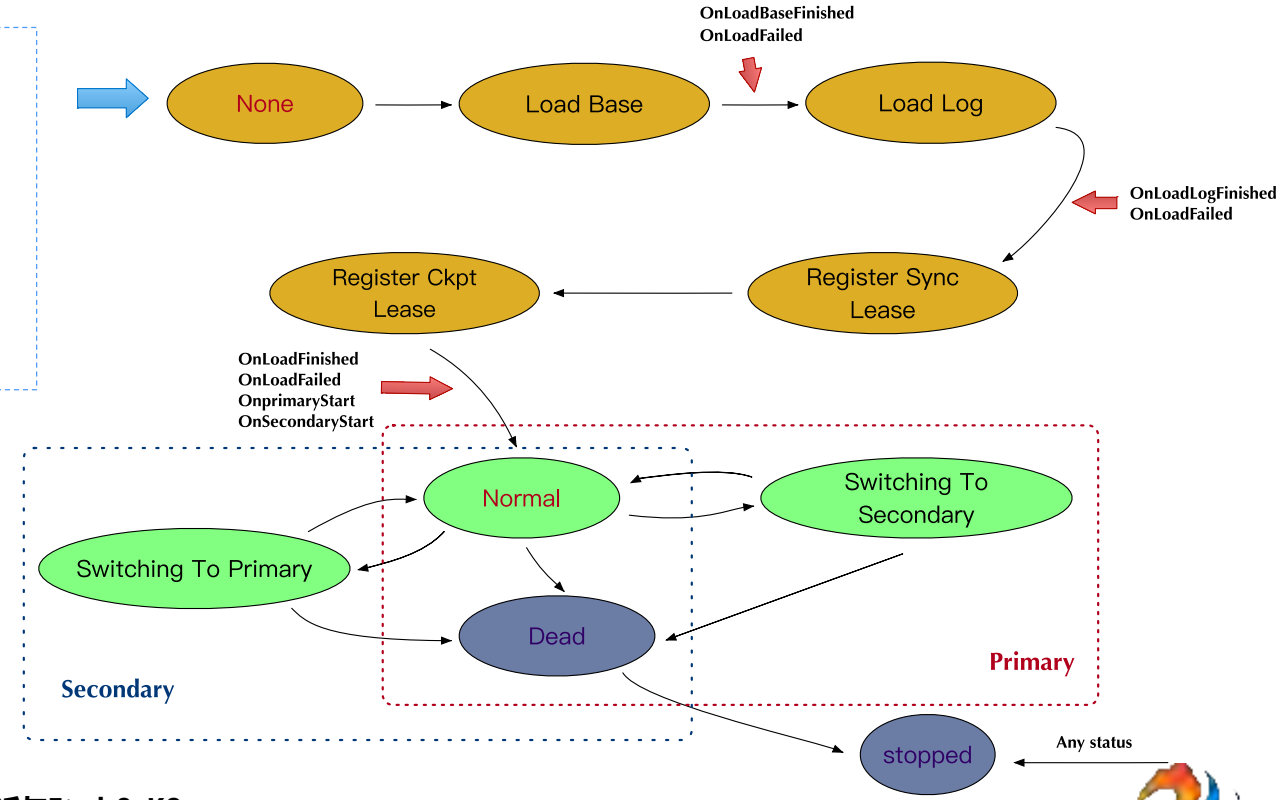
- 根据团队场景和需求裁剪和选择
- 可运维性和团队组织和架构相匹配
- 需要进行规划和服务梳理

目录

筑心 - Nova Master内存数据库

- 调度&资源状态数据内存管理
- 去中心化 -> 中心化
- 支持单集群10000以上物理机调度与管理。多通道LOG：并发写入3000 qps以上
- 强一致选主，主写，从备
- 内存数据基于红黑树组织和自建索引维护



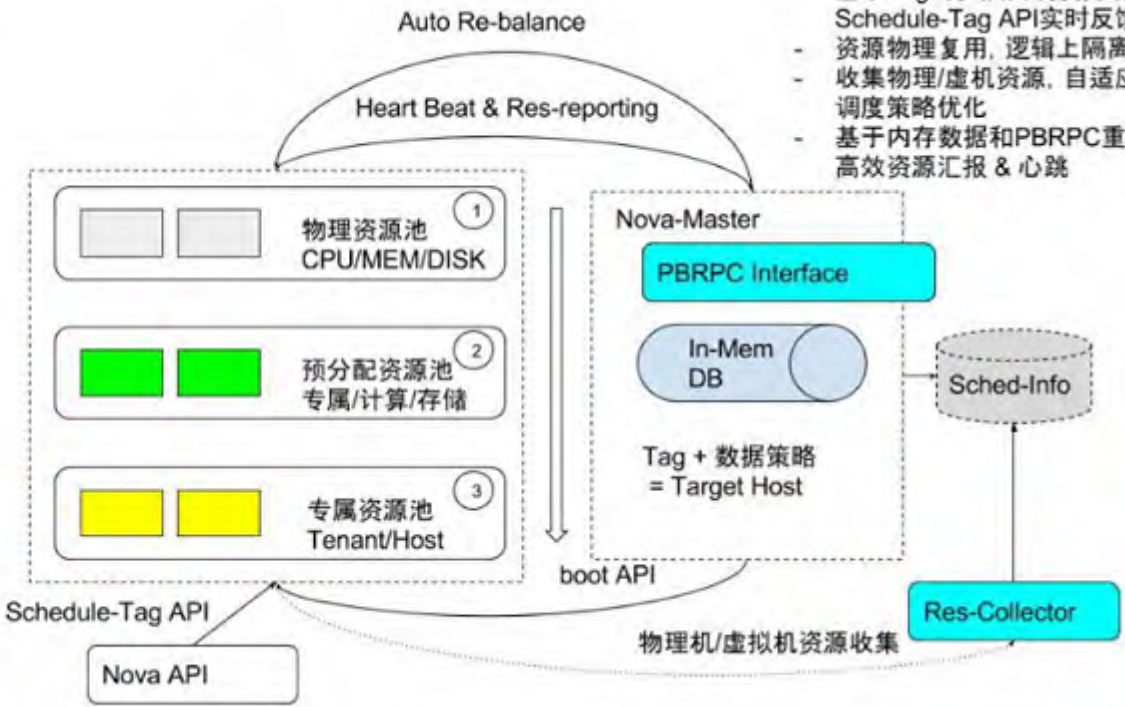


相较于Mysql & Openstack，更接近与Etcd & K8s

筑心 – 重写Scheduler

对比:Openstack支持过滤调度等三种方式;无法做到自适应和逻辑资源池分配

- 基于Tag的多队列调度，支持Schedule-Tag API实时反馈
- 资源物理复用，逻辑上隔离
- 收集物理/虚机资源，自适应调度策略优化
- 基于内存数据和PBRPC重写高效资源汇报 & 心跳

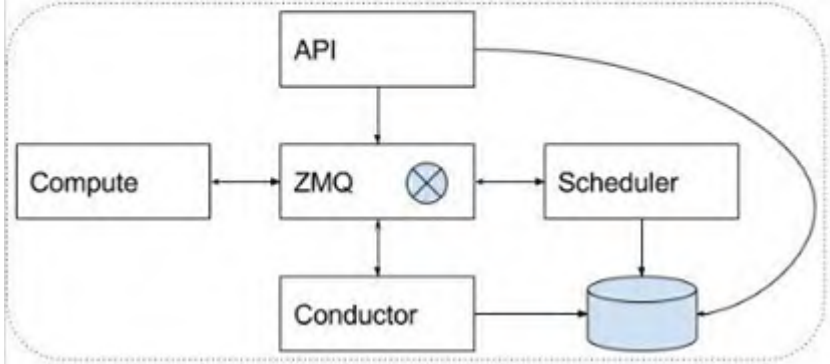


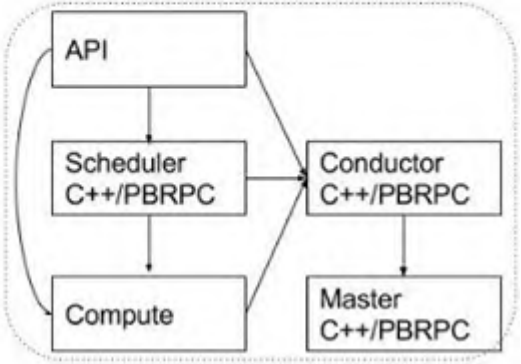
为什么需要持久化两份数据作为调度依据？ **一份是热数据：核心资源汇报参考，一份是冷数据：离线的历史趋势和数据汇总**
逻辑资源池：根据Tag标记倒排索引，根据需求归并
Re-balance：自动根据负载热迁移 – peek time monitoring

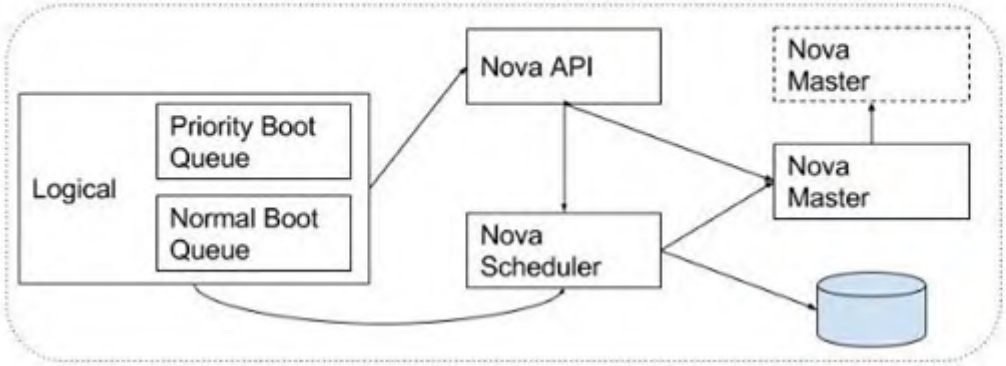
筑心 – RPC模型改造 & 资源预感知

2大问题：

- Eventlet协程切换在高并发下卡死
- ZMQ吞吐量严重不足，扩容耗费控制节点资源







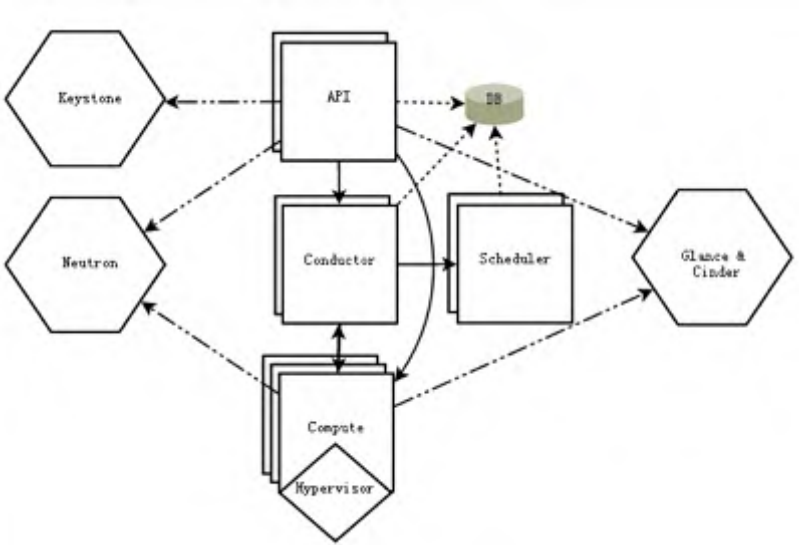
- 建立订单优先队列
- 预分配套餐资源

目录

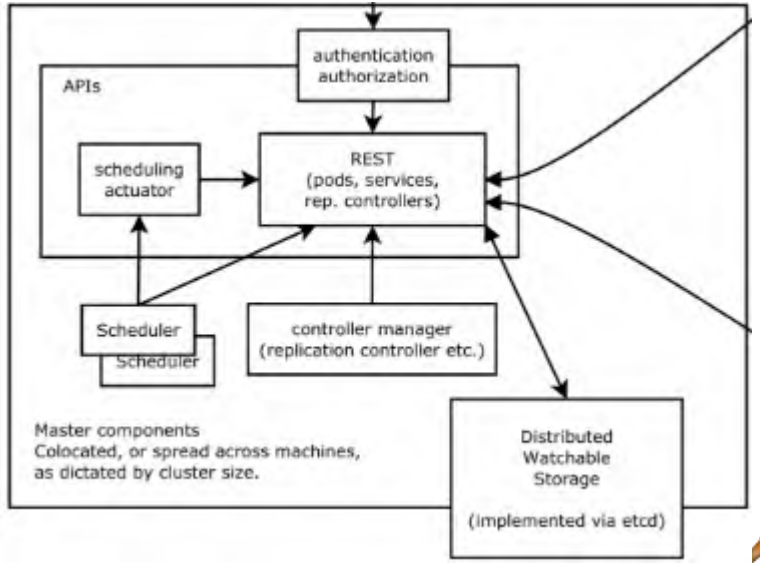
化龙 – 从Openstack-Nova到Kubernetes

- 云计算资源管理平台本身就是一个微服务和分布式的问题
- 相对于Nova，Kubernetes的管理方式更值得借鉴
- Openstack也即将基于K8s重构

从业务出发，不自我设限



OR



化龙 – 展望和总结

只有笨蛋才能征服上甘岭，折腾起来

- 高效的研发和迭代（200+ Feature，BCC/DCC/BBC/GPU四大系列产品）
- 完备的API贯穿整个虚拟机使用场景，控制台基于API构建，大客户/代理商基于API自建控制台；所有开放接口性能达到100 qps, 查询类 qps均值从8 qps -> 500 qps以上
- 客户资源的分配/扩缩容/操作成功率稳步提升，线上运维无故障时间保持在99.95%以上
- 3台控制节点，支撑起10000物理机，100K虚机规模的集群
- 团队获得2016可信云大会颁发的计算资源管理技术创新奖
- 申请10+技术专利

Thanks

高效运维社区

开放运维联盟

荣誉出品

想第一时间看到高效运维公众号的好文章么？

请打开高效运维公众号，点击右上角小人，并如右侧所示设置即可：

GOPS2016 全球运维大会更多精彩