QCon 全球软件开发大会
INTERNATIONAL SOFTWARE DEVELOPMENT CONFERENCE

BEIJING 2018

《新一代数据中心对传统基础软件架构的挑战》

演讲者 / 王华夏

主办方 Geekbang. 极客邦科技 InfoQ

业务需求驱动技术创新

01 一个典型案例.

02 服务发现和域名解析.

03 存储和镜像替换.

04 新的挑战.

05 终极目标.

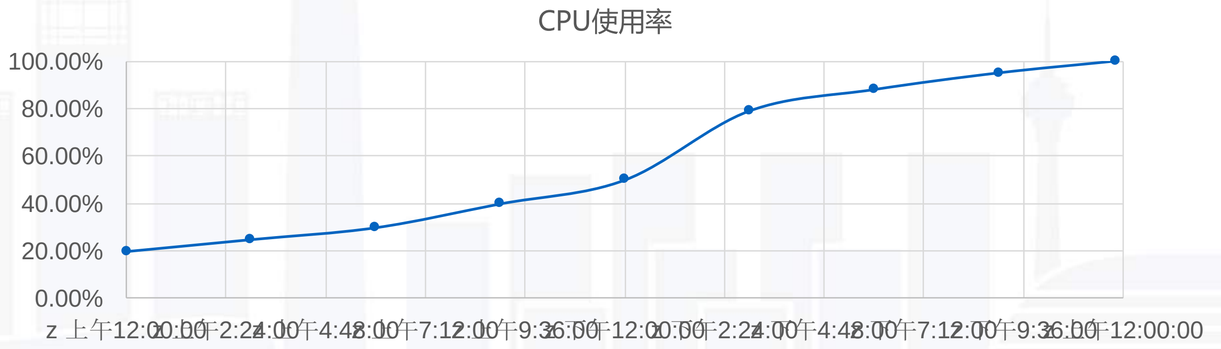
真实案例

关键词 ：11.11

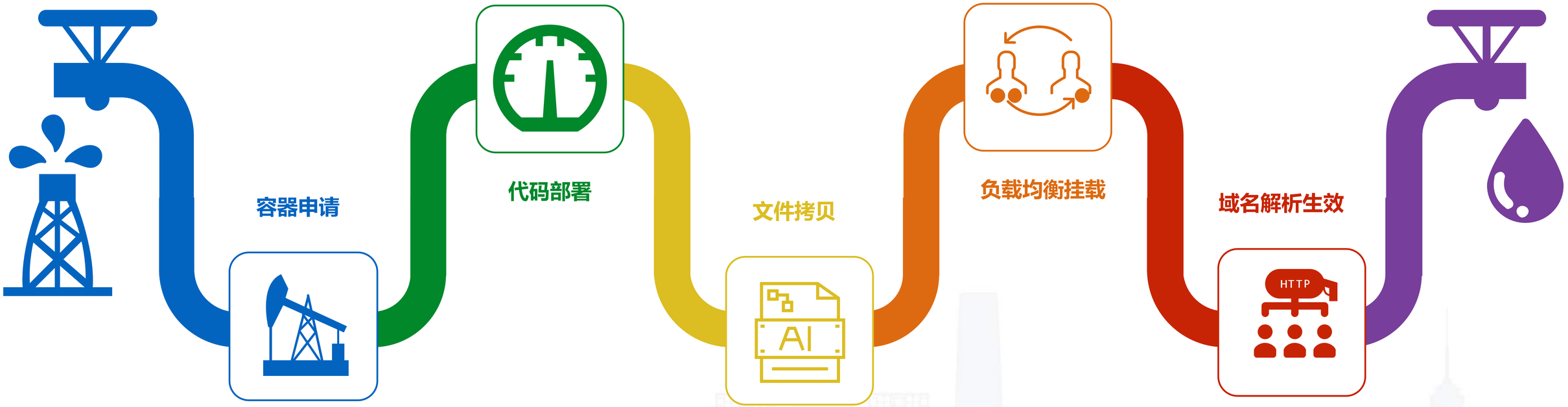
关键词 ：23:50

关键词 ：BUG

要死了！

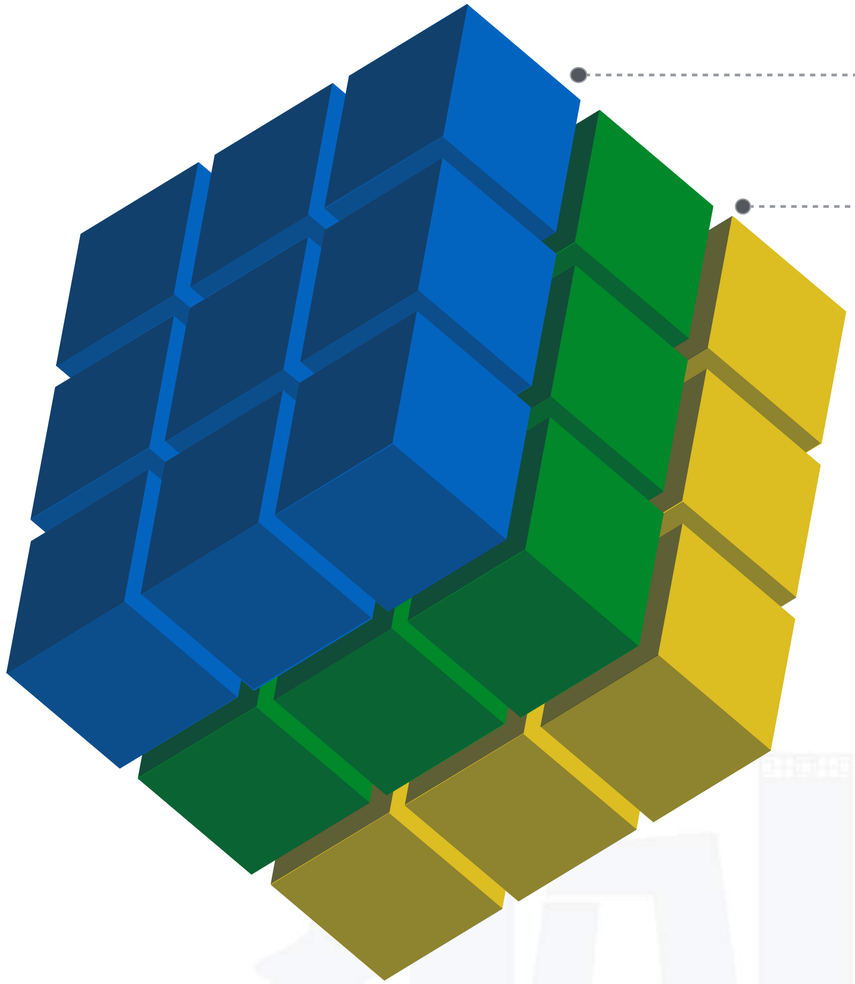


如何活过11.11？



JDOS1.0容器集群发布方式

紧急扩容过程中的挑战

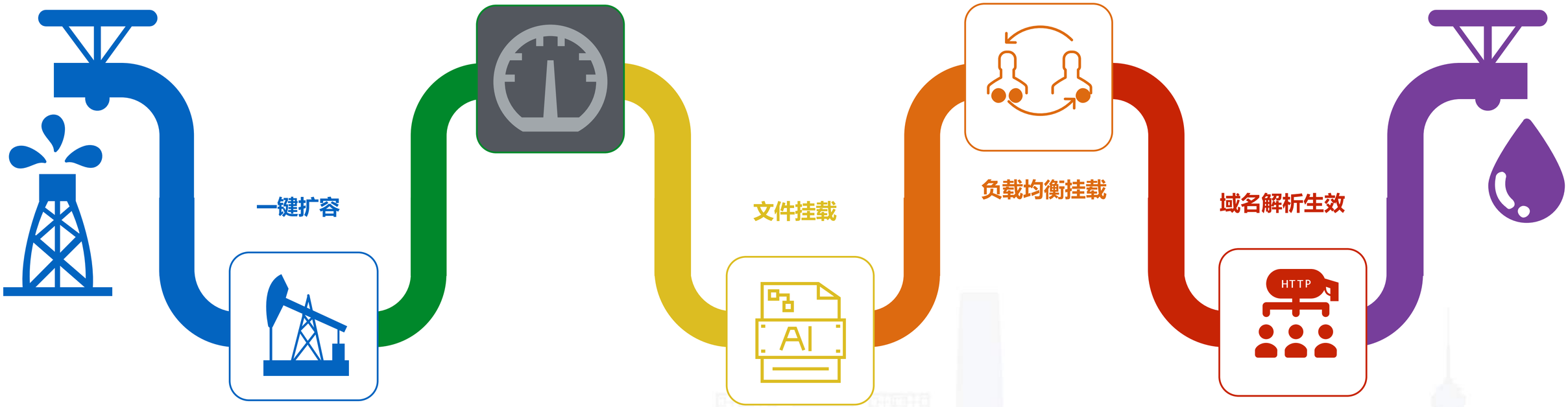


负载均衡服务发现

域名服务快速生效

Severing 文件快速拷贝

问题如何解决？



JDOS2.0容器集群发布方式

业务需求驱动技术创新

01 一个典型案例.

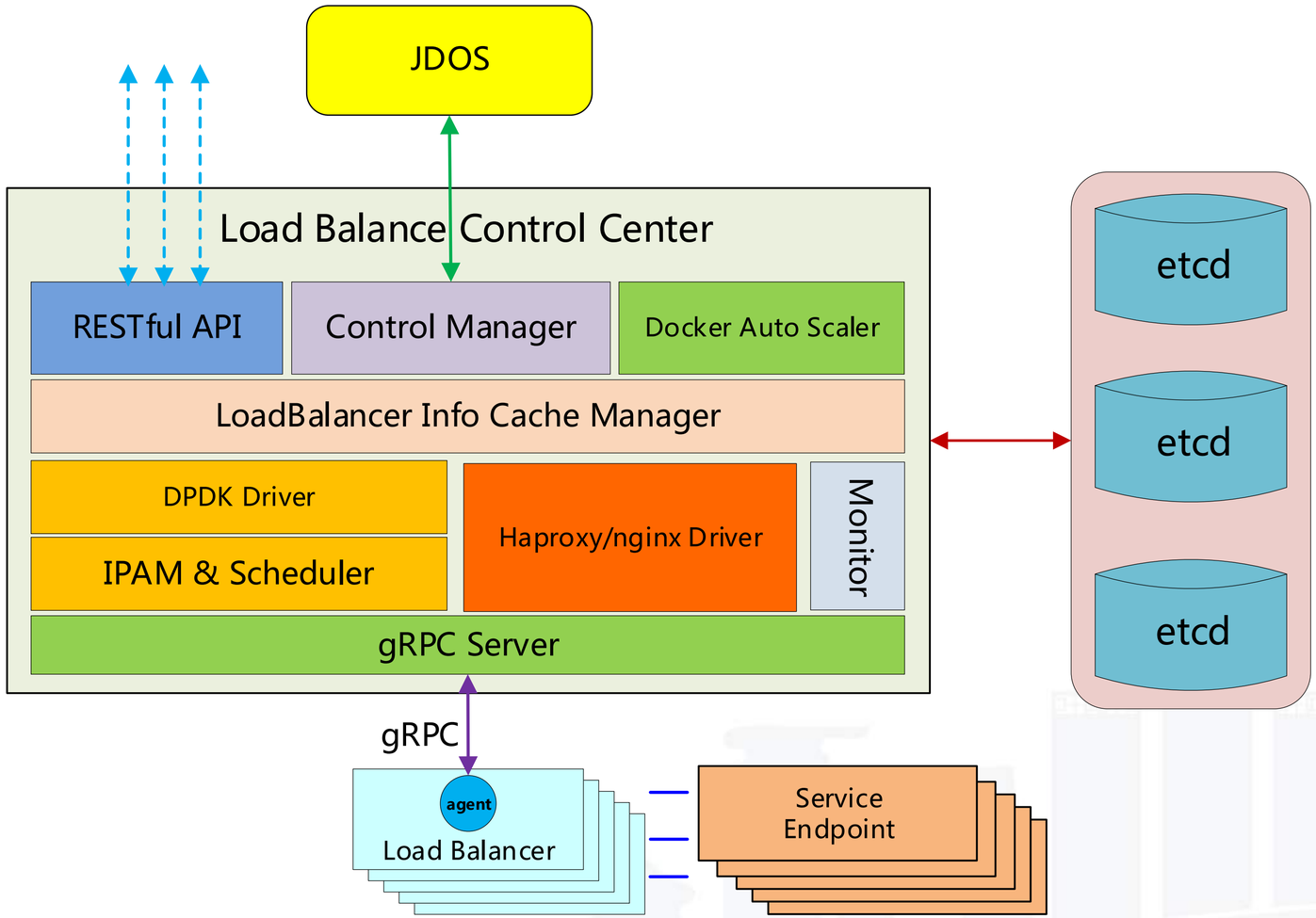
02 服务发现和域名解析.

03 存储和镜像替换.

04 新的挑战.

05 终极目标.

负载均衡-1



01

容器化部署

- 全部组件支持容器化部署.

02

健康检查

- 健康检查支持服务自动发现.

03

自动扩容

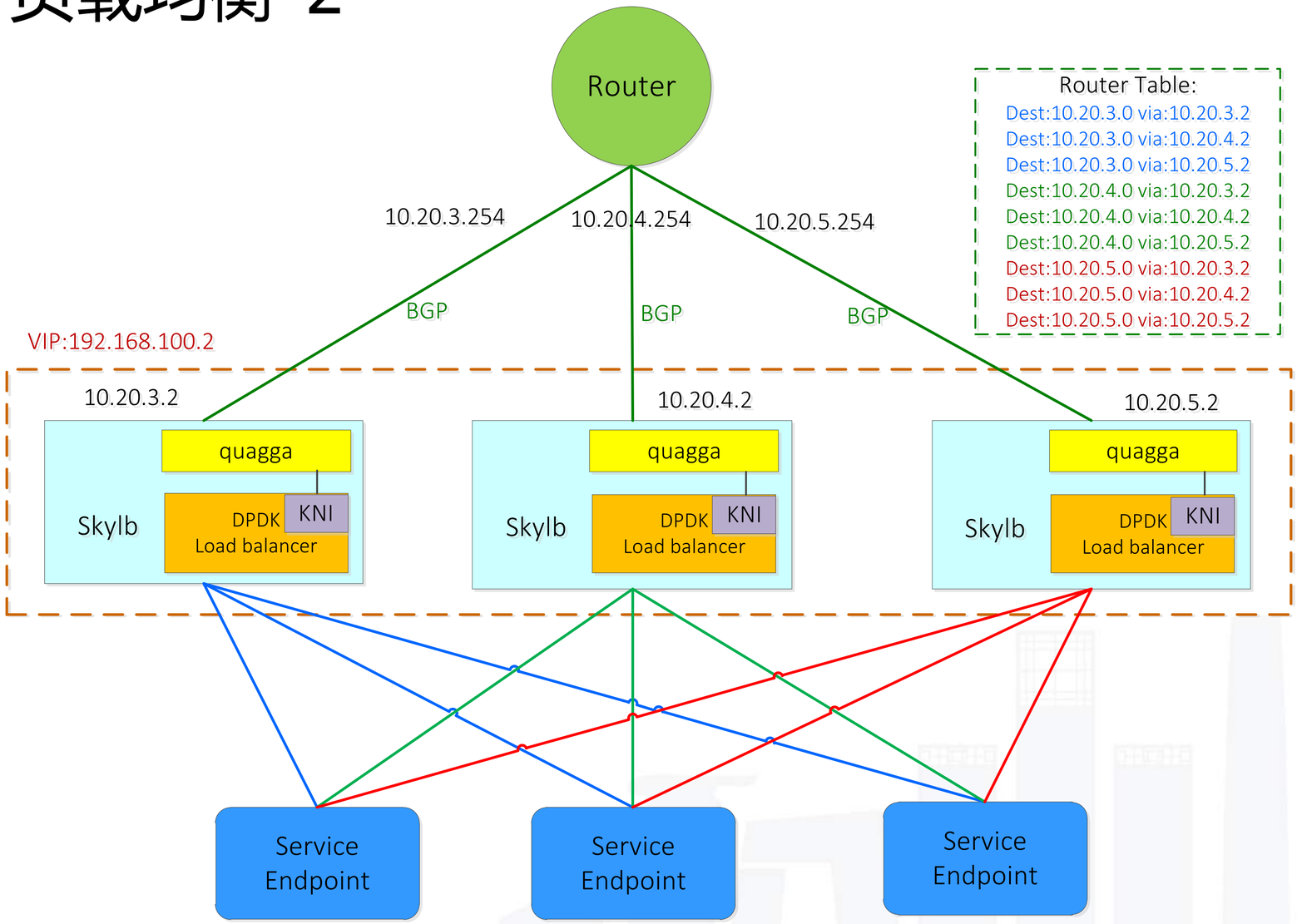
- 根据实际流量自动扩容集群

04

L4&L7

- HA和Nginx实现7层负载均衡
- DPDK实现4层负载均衡.

负载均衡-2



https://github.com/tiglabs/jupiter

01

L4软负载均衡实现

- 非LVS软负载均衡

02

高可用

- 集群模式

03

高性能

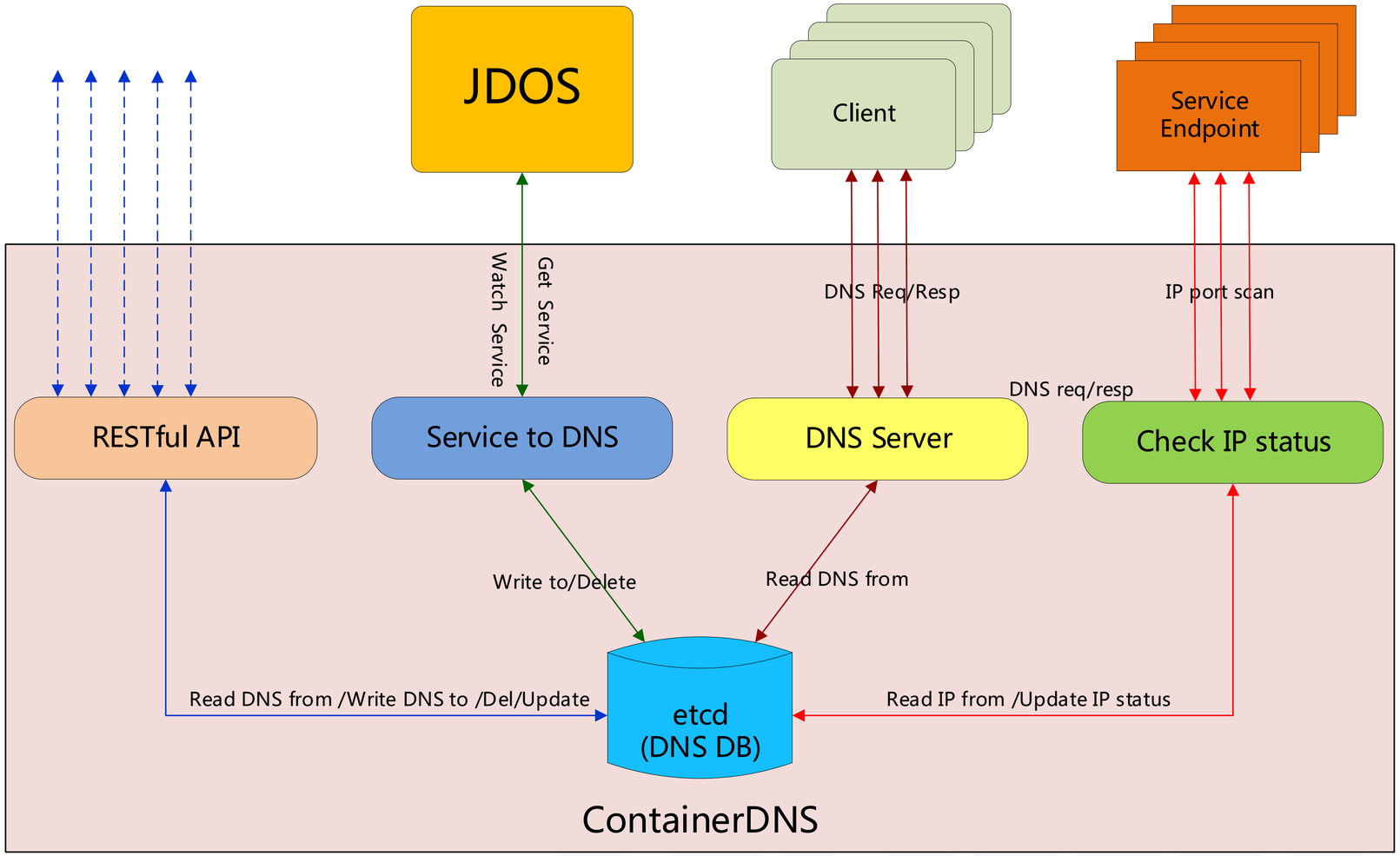
- 210W QPS

04

VIP实时生效

- 秒级生效

域名解析服务-1



01 容器化部署

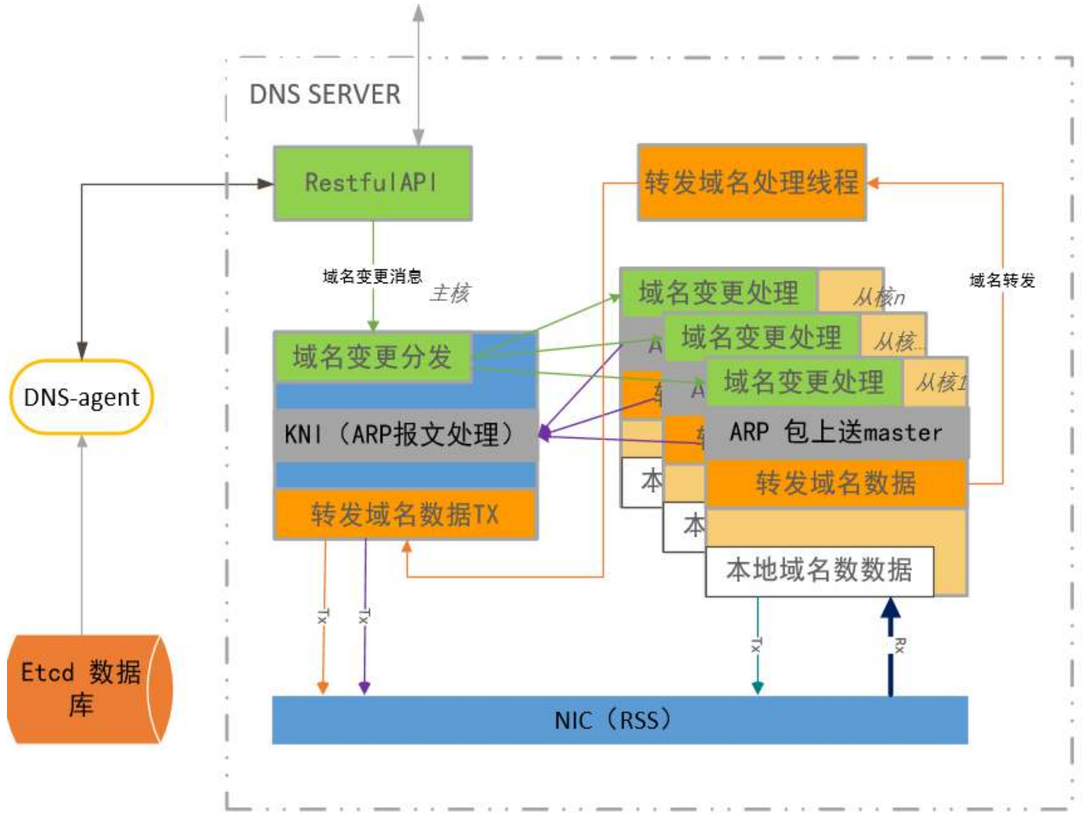
02 实时生效

03 LB功能

04 域名信息记录ETCD

https://github.com/tiglabs/containerdns

域名解析服务-2



01

DNS SERVER高可用

- 集群模式
- 方便扩容

02

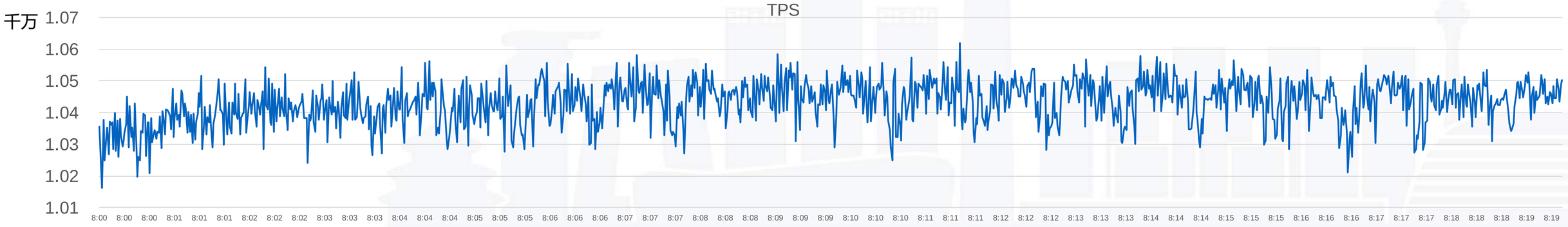
DPDK技术加速

- 网卡加速

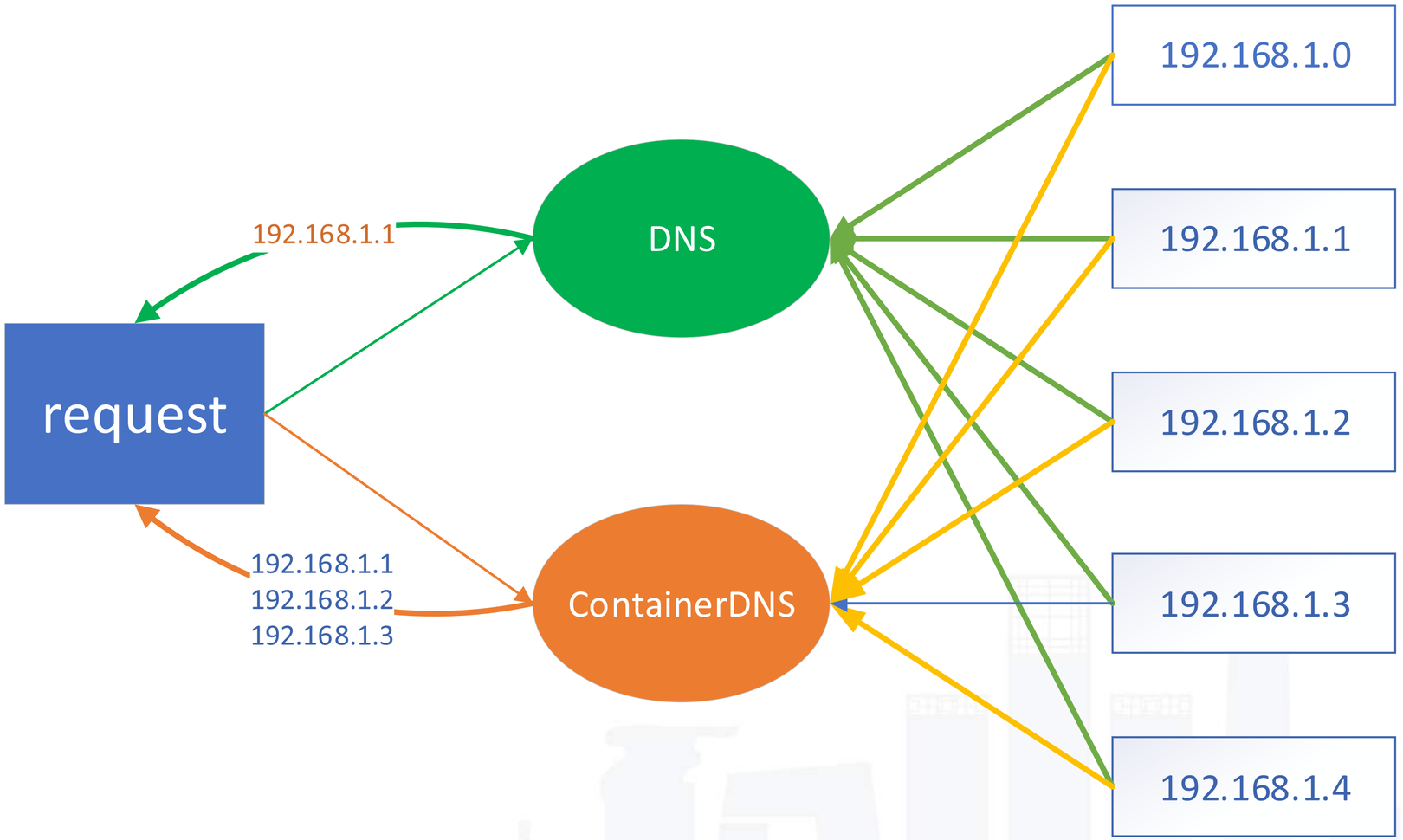
03

超高性能

- TPS超过10000w.



域名解析服务-3



域名来代替一系列IP

业务需求驱动技术创新

01 一个典型案例.

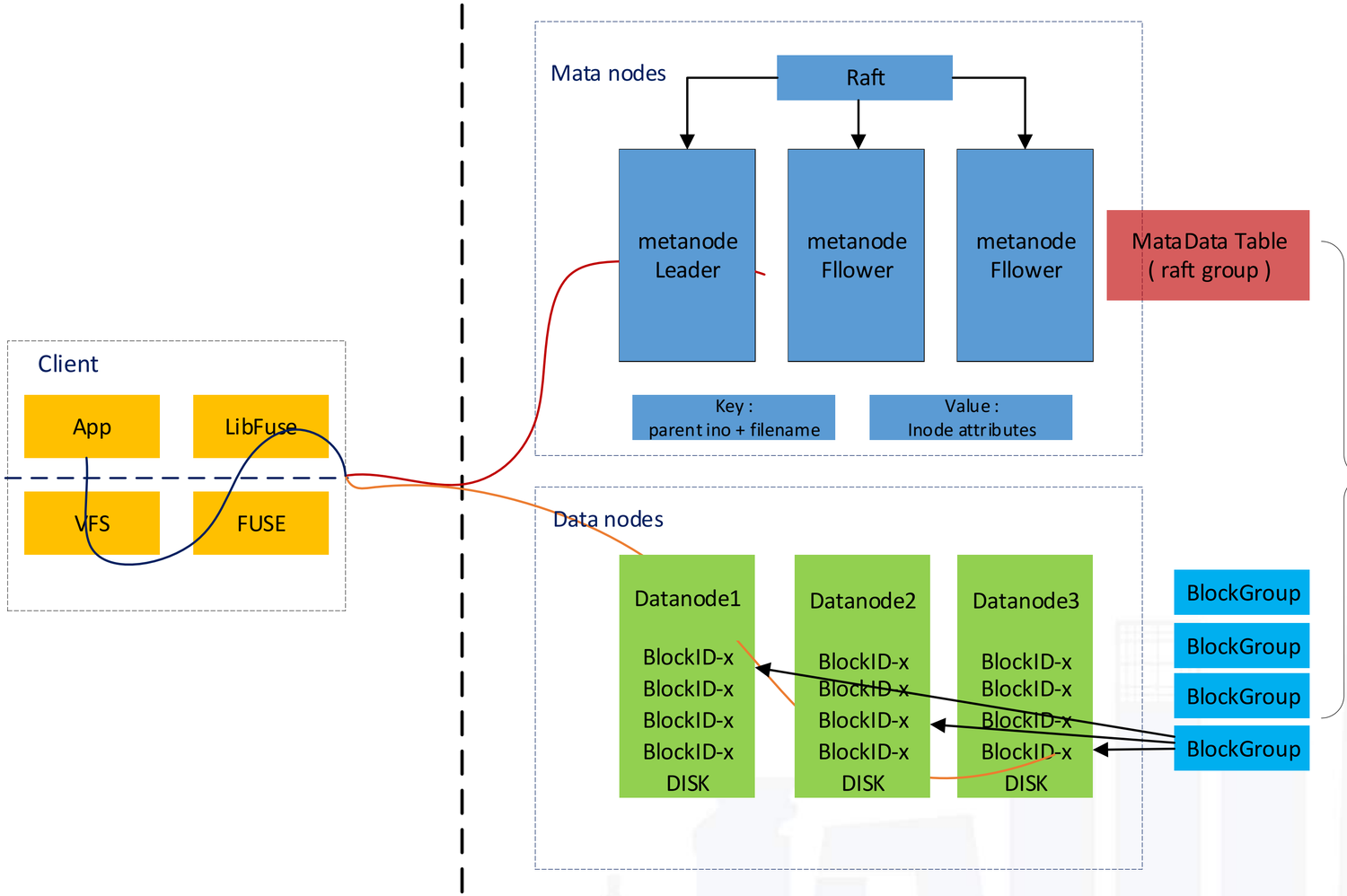
02 服务发现和域名解析.

03 存储和镜像替换.

04 新的挑战.

05 终极目标.

ContainerFs-1



大容量高性能

- 满足业务增长对文件存储的容量需求.
- 采用定制存储硬件高性能 非常适合数据吞吐型应用.

共享访问

- Fuse标准接口实现共享挂载访问 业务无需任何修改即可无缝使用.
- 帮助多业务多实例应用获得相同的数据源.

云原生

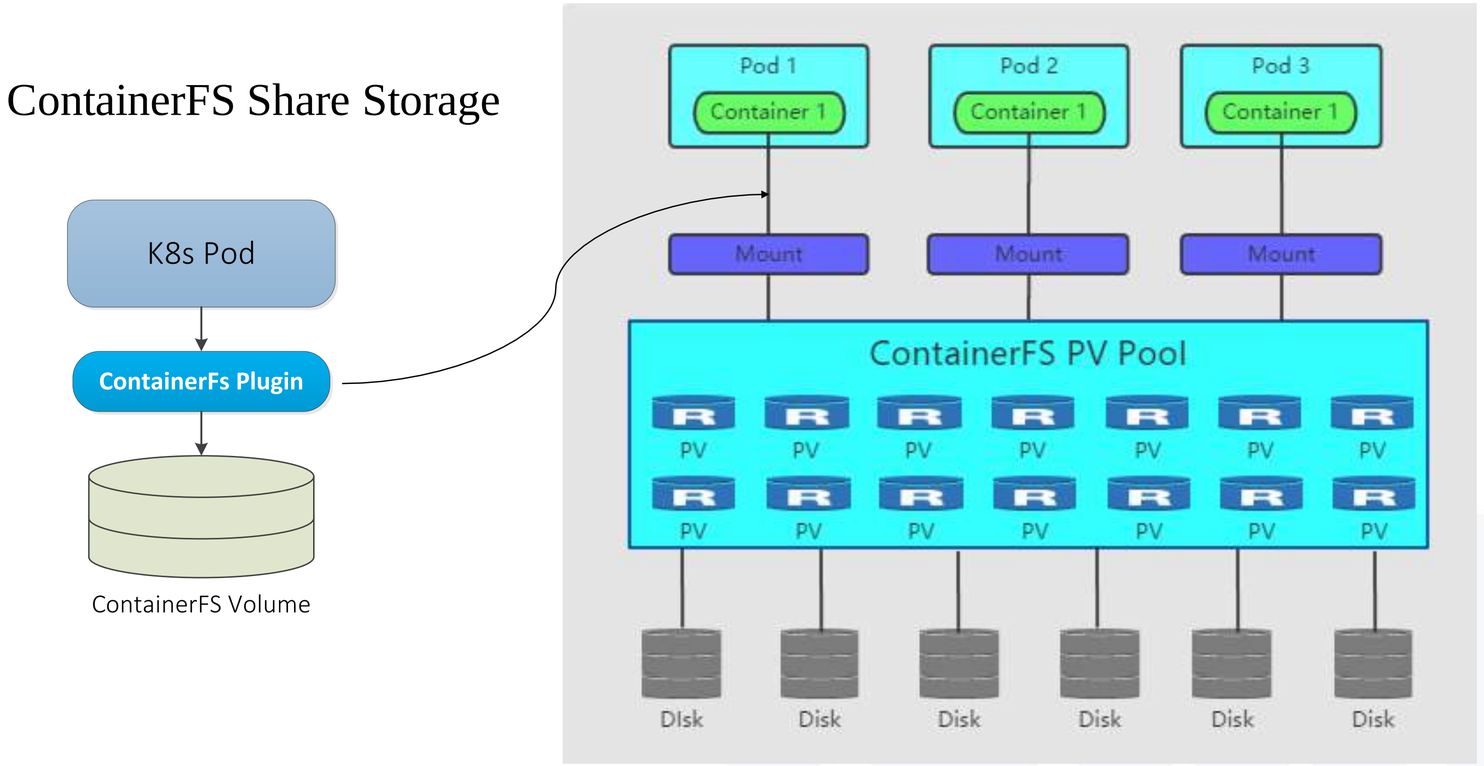
- ContainerFS部署不依赖特殊硬件 可以在kubernetes集群编排部署.
- 支持kubernetes标准CSI接入 是行业内最早支持CSI接入标准.

低成本

- 借助大量计算节点的硬盘位 实现复用 节省采购成本.
- 副本可配 EC等功能进一步降低存储空间 降低成本.

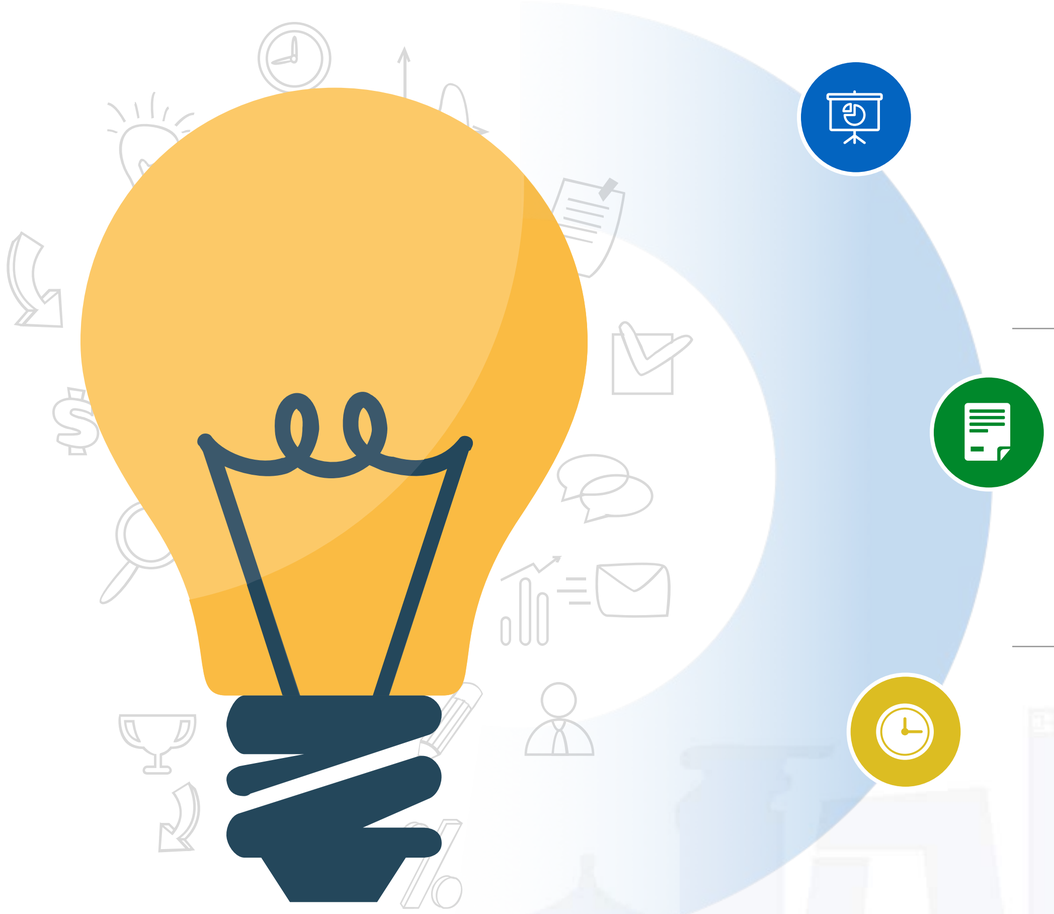
https://github.com/tiglabs/containerfs

ContainerFs-2



- 通过K8s PV卷共享
- 同一服务共享文件
- 扩容自动挂载
- K8s无缝集成

共享存储够了吗？

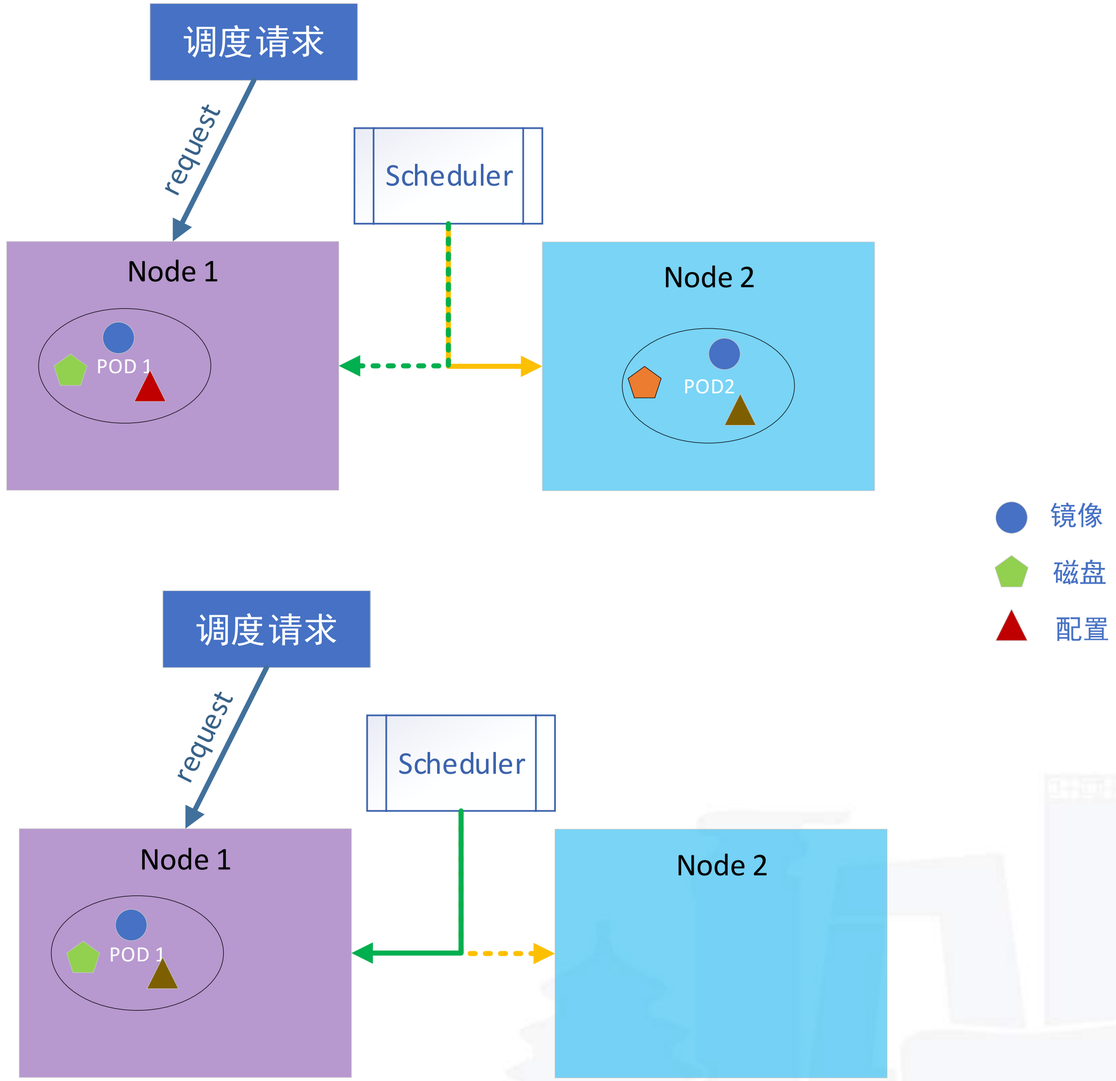


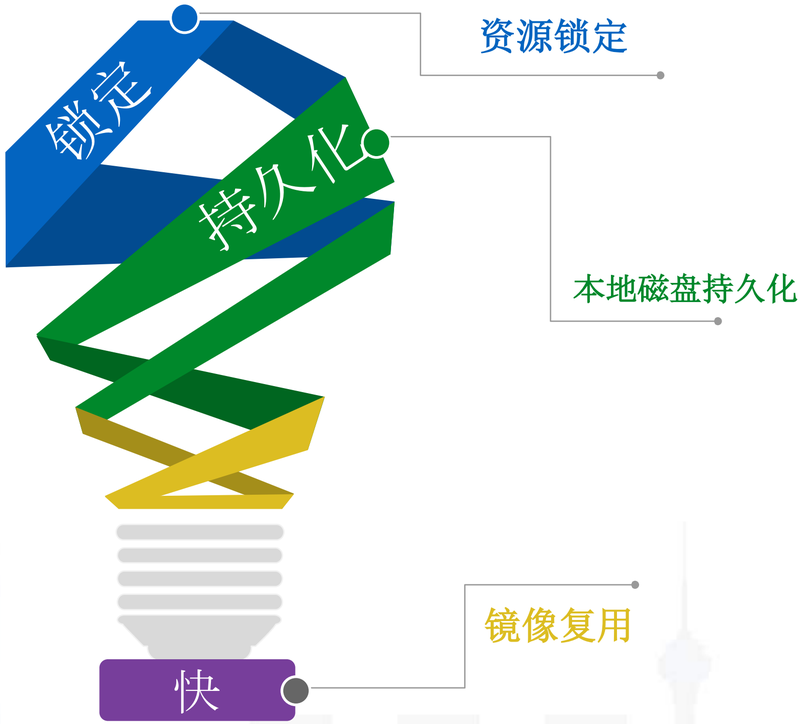
小文件

同一个集群中需保留文件有差异

同一个集群中文件生命周期不统一

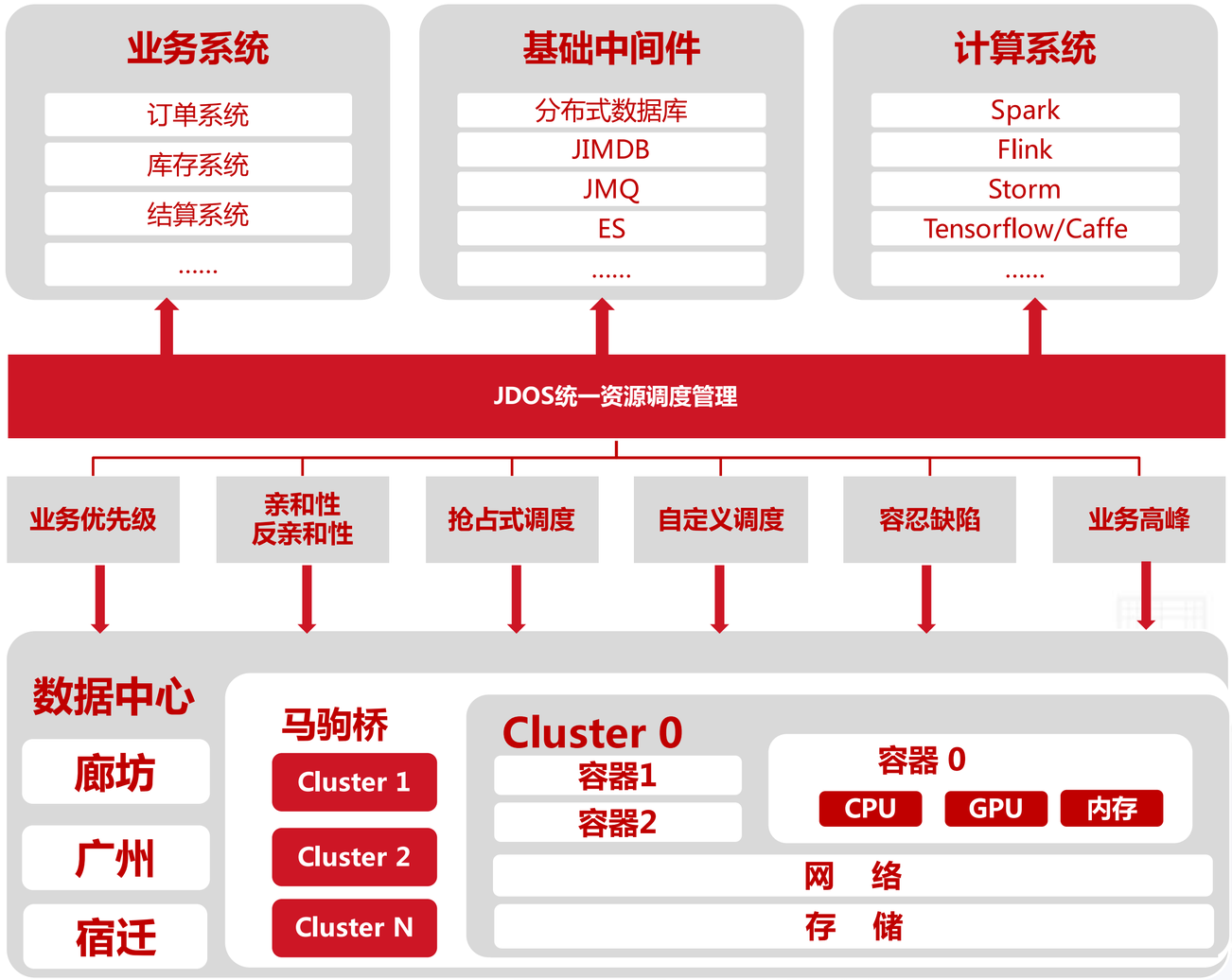
一种并不优雅但有效的解决方案





牺牲了调度的灵活性

JDOS生态



秒级扩缩容

- 秒级实现集群的扩容和缩容.
- 秒级的负载均衡、DNS服务发现

故障自动恢复

- 服务器故障容器自动迁移
- 容器故障，自动恢复.

DevOps支持

- 代码编译、镜像构建
- 从代码到测试上线全链路

全链路监控

- 系统状态秒级监控
- 业务进程函数级监控

服务器：20000
容器数：450000

业务需求驱动技术创新

01 一个典型案例.

02 服务发现和域名解析

03 存储和镜像替换.

04 新的挑战.

05 终极目标.

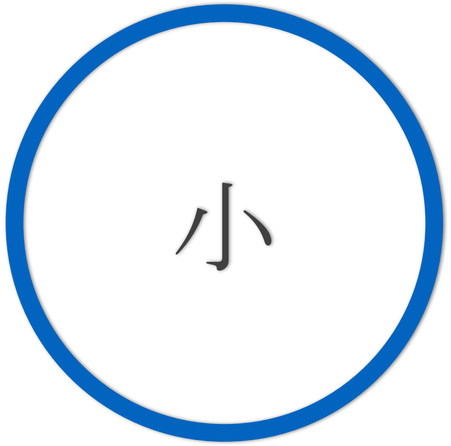
新的挑战-无界零售

部署场所分散

部署场所分散全国各地网络分散需要公网连接

部署场所资源有限

部署场所资源有限，服务器数量有限，网络带宽有限，交换机规格也有限



部署场景规模小

每个部署场所规模都比较小，需要的网络资源少，内存，cpu，磁盘需求都比较小

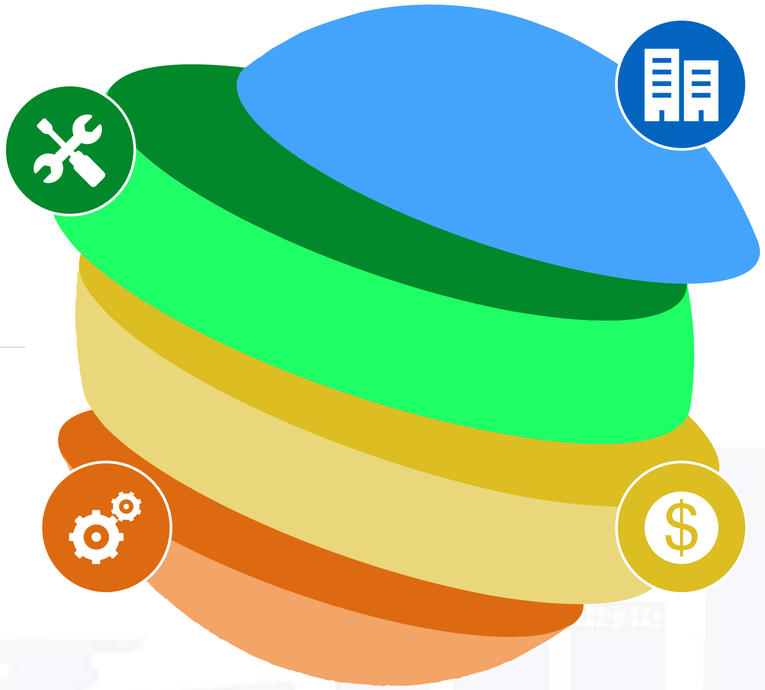
部署场所多

部署场所非常多，业务程序需要批量快速部署全部场所

新零售场景中的新问题

线上的基础设施服务如何延伸到线下？

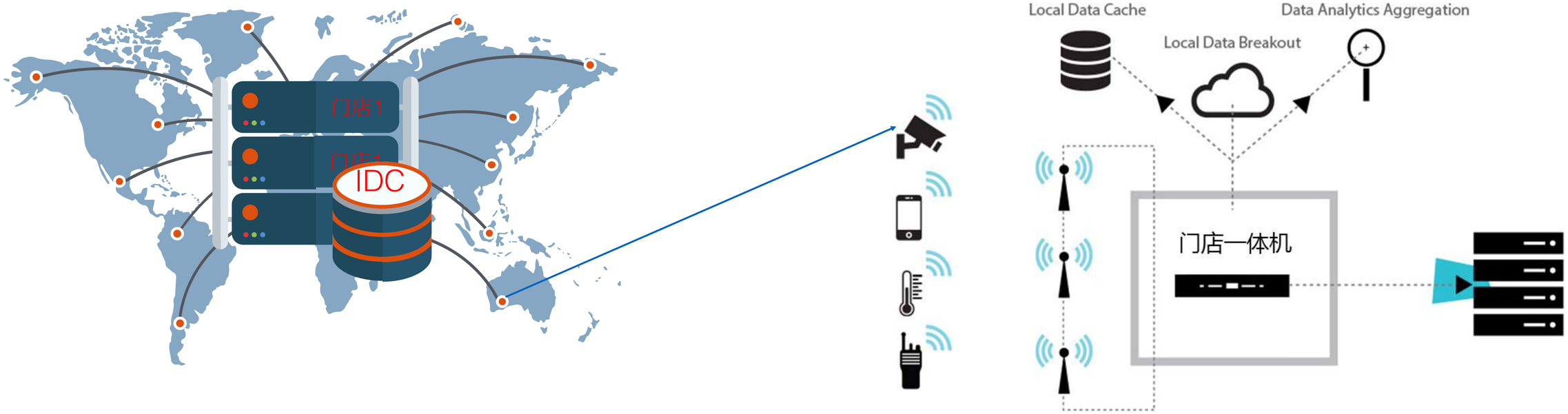
海量同质化的门店，如何快速部署？



门店如何共享数据中心资源？

门店IT设施如何低成本管理？

第四代零售革命中的IT基础设施生态



充分利用JD机房资源来部署基础服务

不同规模的场景，采用不同的部署方案

实现一次构建，批量部署的方案

边缘计算降低网络消耗

业务需求驱动技术创新

01 一个典型案例.

02 服务发现和域名解析

03 存储和镜像替换.

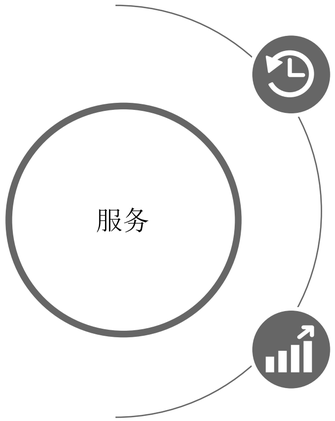
04 新的挑战.

05 终极目标.

我们的核心价值是什么？

方便研发人员进行业务部署

支持线上服务的稳定

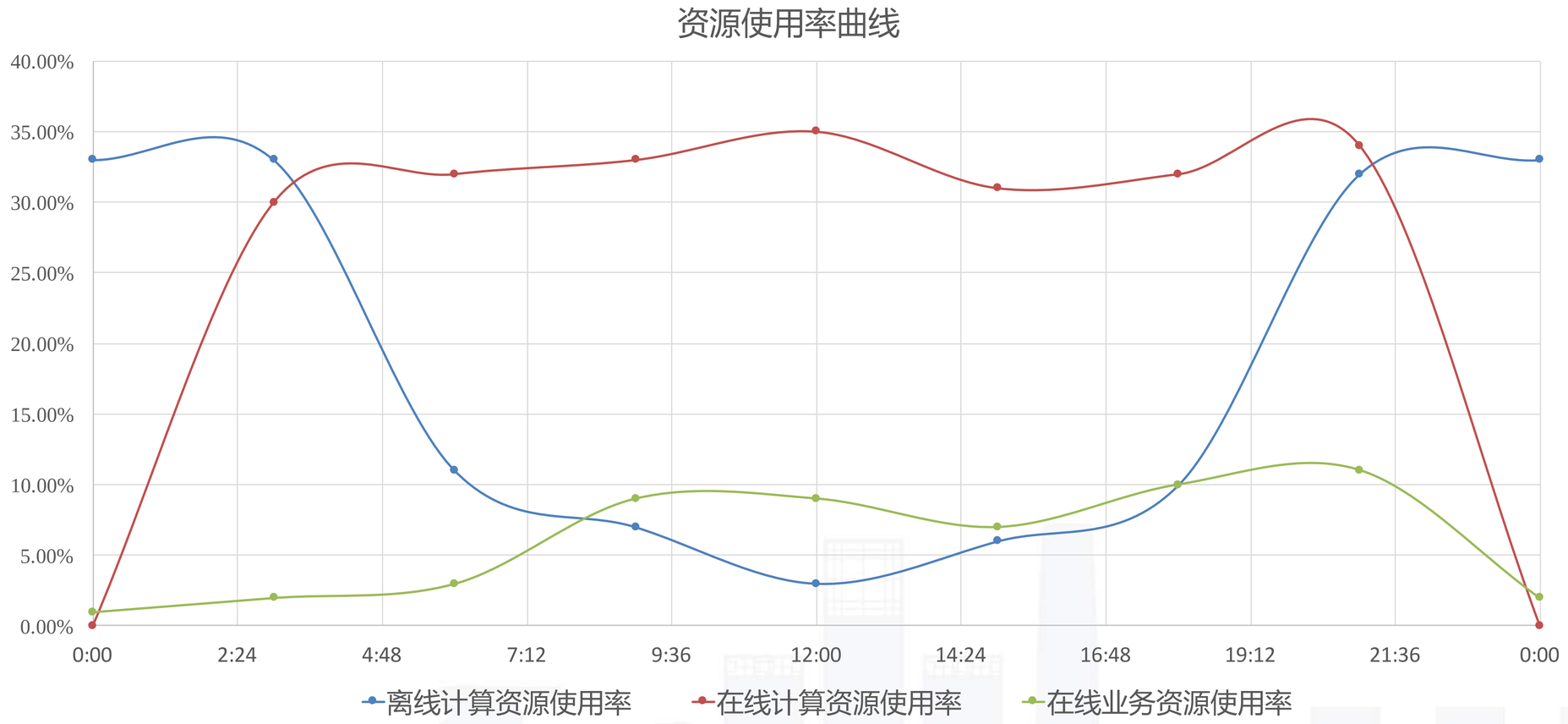


公司任何部门的核心价值都是为公司省钱，赚钱

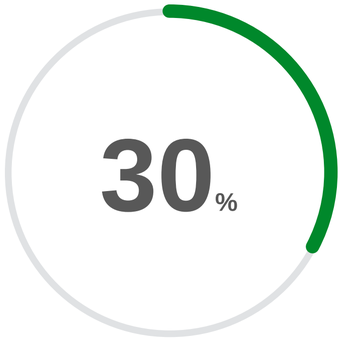
为公司赚钱

为公司省钱

资源使用率-现实情况

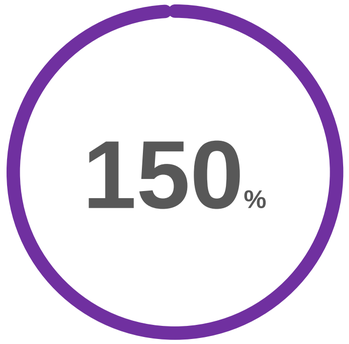

资源使用率曲线
40.00%
35.00%
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
0:00
2:24
4:48
7:12
9:36
12:00
14:24
16:48
19:12
21:36
0:00
离线计算资源使用率
在线计算资源使用率
在线业务资源使用率

部署层面的解决方案



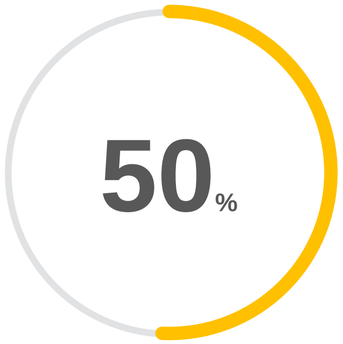
离线和在线业务混部

将离线和在线业务进行隔离部署相较于混合部署，需要额外增加20~30%的机器。



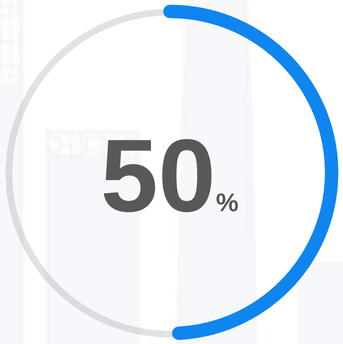
不同用户的任务混部

不同用户的任务隔离部署较之混合部署，需要额外增加20~150%的机器。



大型集群部署

将大型集群划分为多个小型集群相较于整合的大型集群，需要更多的额外机器。将一个大型集群划分为两个小型集群，需要25~50%的额外机器。



细粒度资源请求

粗粒度的资源请求相较于细粒度的资源请求，需要额外的30~50%的资源。

阿基米德调度-JD实践

调度的核心

灵活超卖

资源分配算法（Requests / Limits ）

针对业务进行分级打分

亲和性和反亲和性

资源回收

阿基米德调度-安全问题

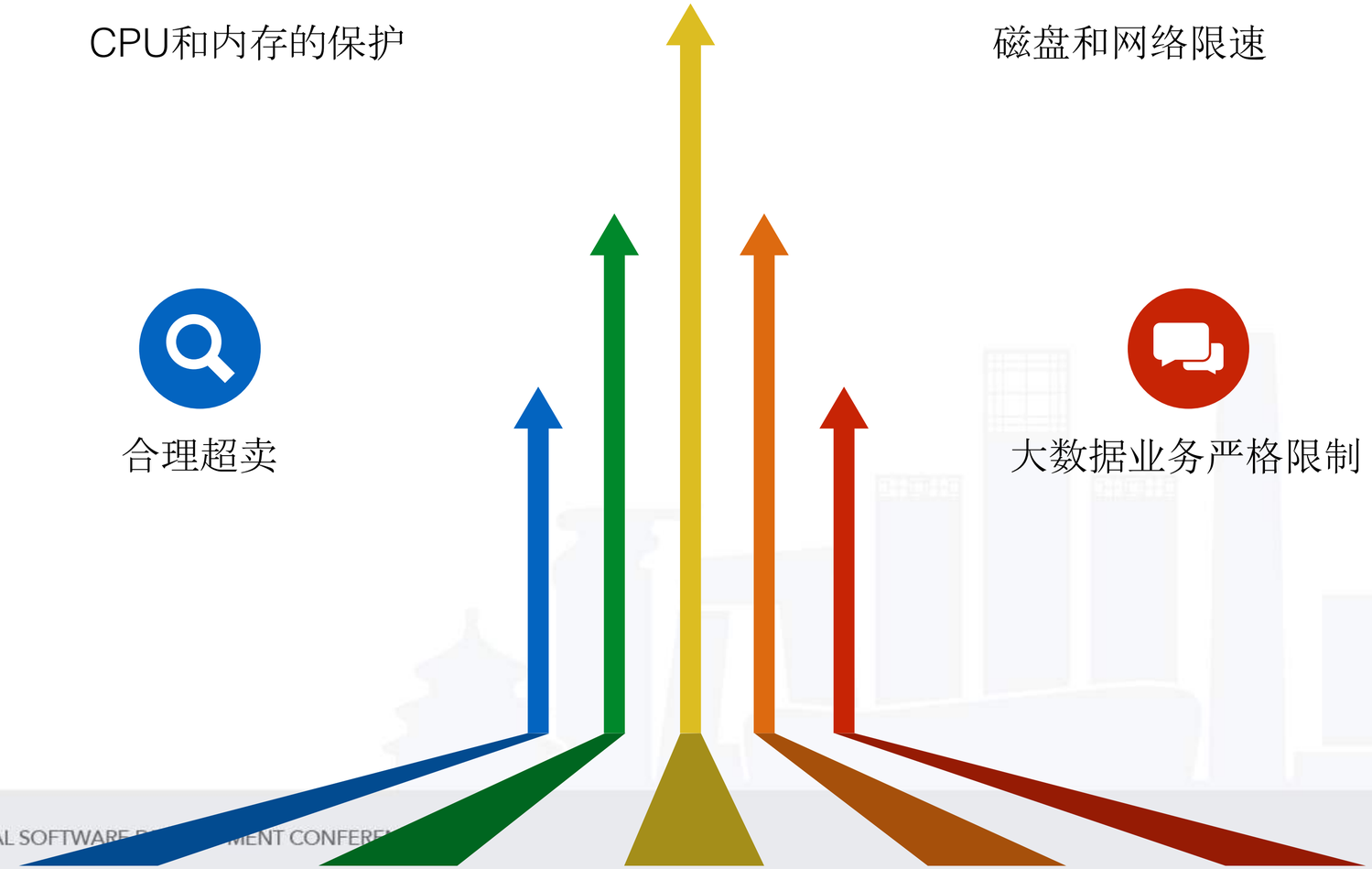
核心业务分布均匀

CPU和内存的保护

磁盘和网络限速

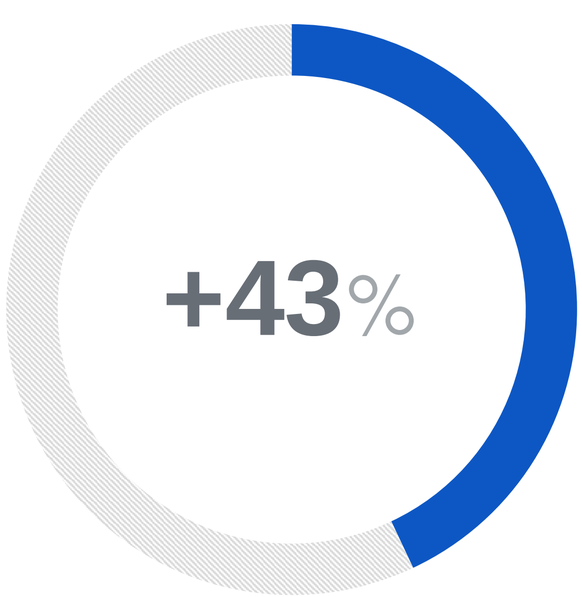
合理超卖

大数据业务严格限制

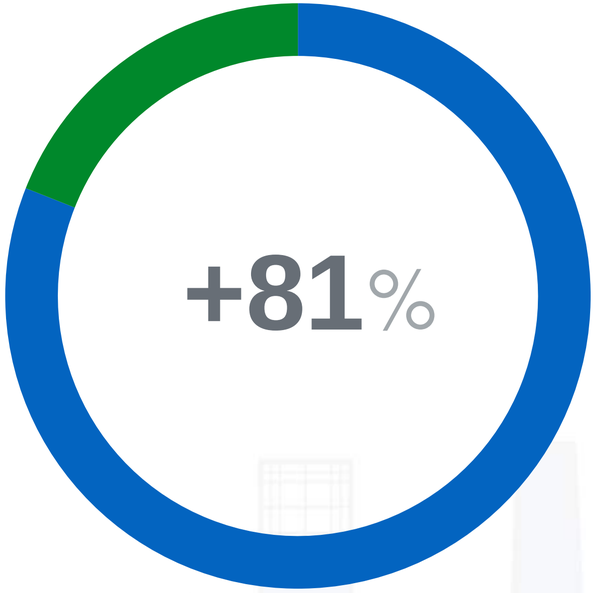


安全也是一种边界

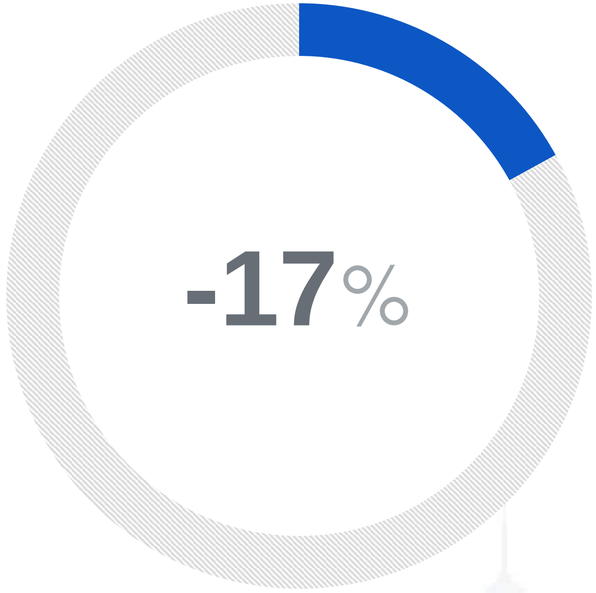
阿基米德调度-成果



CPU平均使用率



内存平均使用率



数据中心能耗

弹性调度所有计算资源，节省服务器采购成本**数亿元**

IT基础设施技术创新的源动力

关注QCon微信公众号，
获得更多干货！

Thanks!

INTERNATIONAL SOFTWARE DEVELOPMENT CONFERENCE

主办方 Geekbang. 极客邦科技 InfoQ