



网易蜂巢

基于Kubernetes的公有云运维实践

刘超 网易蜂巢解决方案总架构师

关于我



我是谁

我是刘超，爱代码，爱开源

<http://blog.csdn.net/popsuper1982>

略懂Lucene, OpenStack, Docker, Mesos

Open DC/OS 社区贡献者

从哪里来

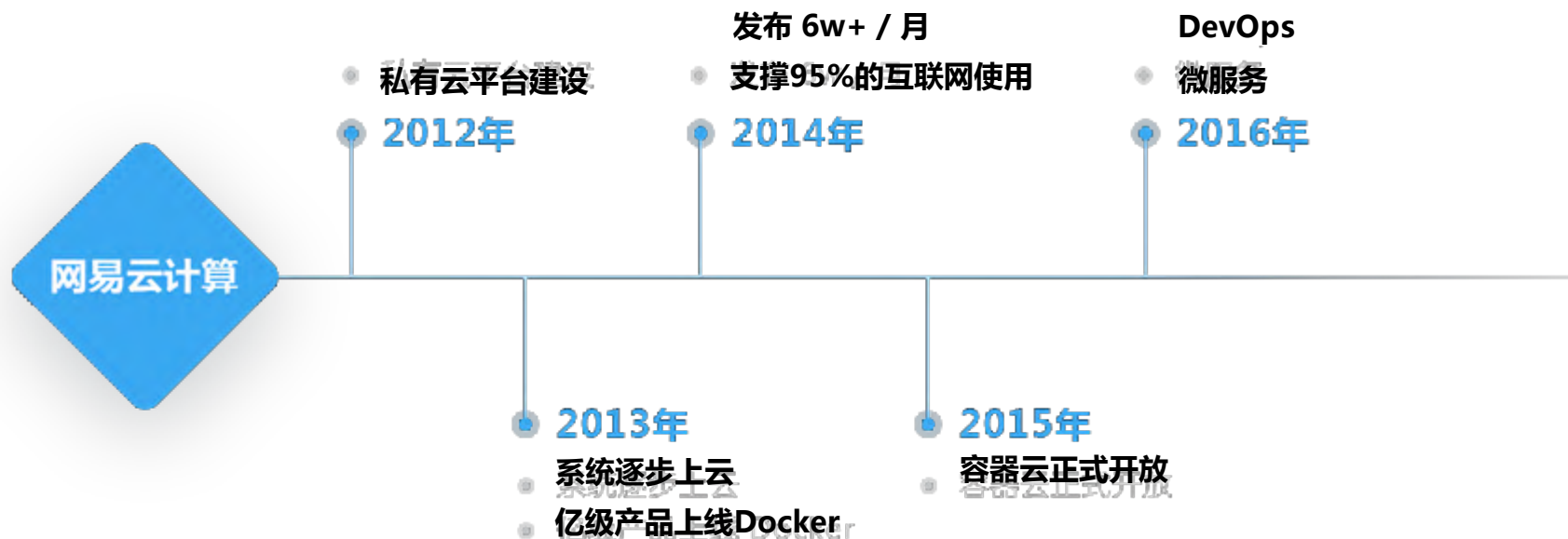
网易蜂巢解决方案总架构师

目标：写代码中最懂解决方案的，懂解决方案中最会写代码的

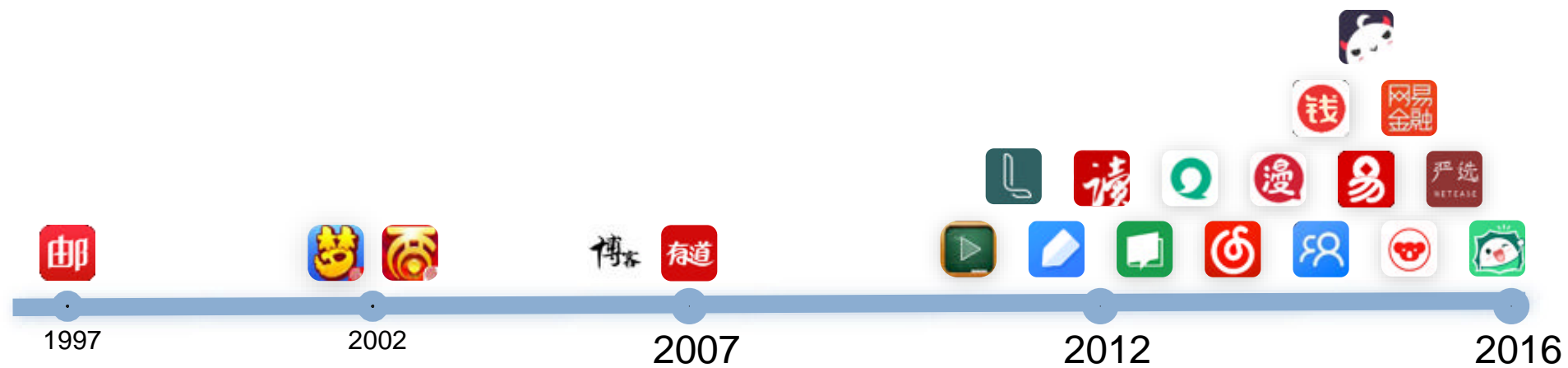
到哪里去

相信趋势：互联网+，容器，公有云，微服务，DevOps

网易蜂巢历程



网易蜂巢上的大规模应用



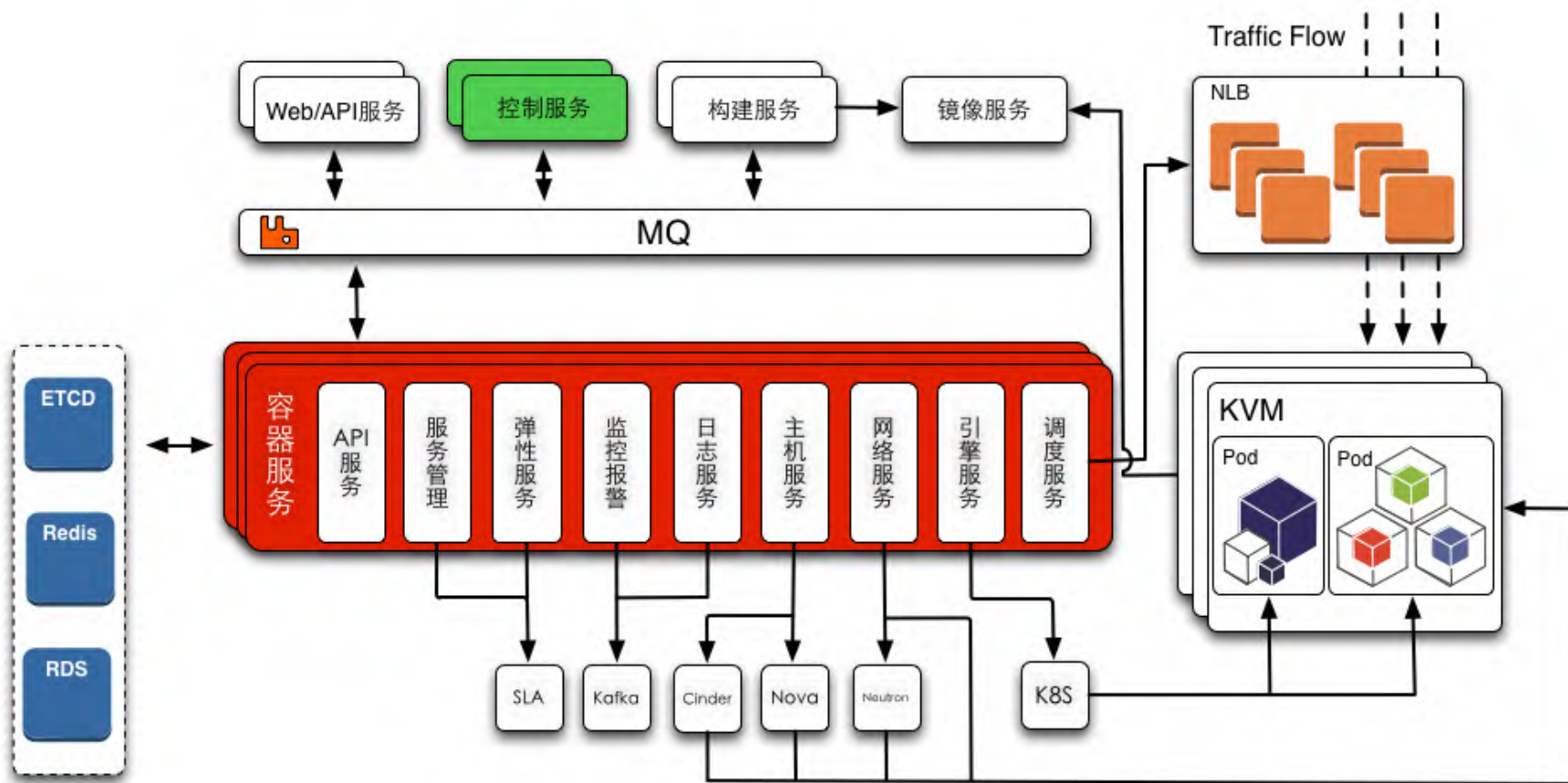
6.18

11.11



200,000,000用户

网易蜂巢的大规模容器平台





PART 02

私有云

私有云平台实现资源弹性

数据库

对象存储

负载均衡

缓存服务

CDN

安全服务

OpenStack

计算虚拟化

KVM

启动优化

镜像剪裁

网络虚拟化

OpenVswitch

Filter

VXLAN

存储虚拟化

Ceph

IO优化

X86 服务器

全万兆网络

全SSD存储

私有云平台优化

高可用，高性能PaaS

- 数据库：网易定制的MySQL内核分支，主从切换数据零丢失，提供健康检查和SQL优化工具
- 缓存服务：主从热备、跨可用域部署，自动容灾，高性能单笔延时毫秒级
- 对象存储：高可用性为99.99%，高可靠性三备份8个9，基于自研分布式非结构化存储系统

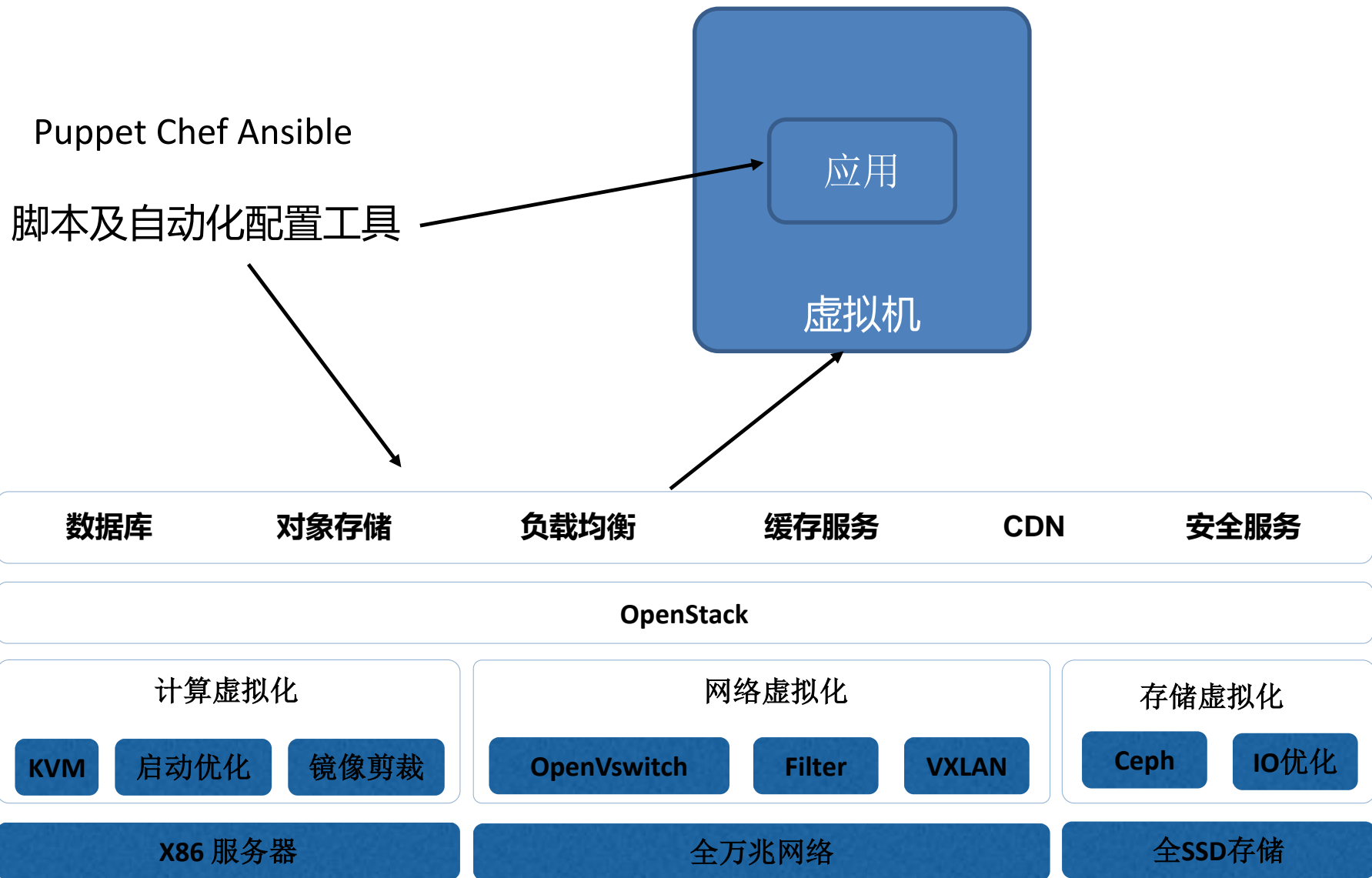
基于OpenStack自研IaaS

- 计算：定制KVM系统镜像，实现云主机IP静态化，优化OpenStack创建云主机流程
- 网络：二层至四层网络过滤防止MAC/IP欺骗，基于Linux TC修改OVS实现网络QoS
- 存储：云硬盘架构基于iscsi和Ceph实现，优化Ceph核心模块OSD

高规格硬件设备

- 多线BGP网络接入，万兆网络互联，全SSD存储

应用层虚拟机部署





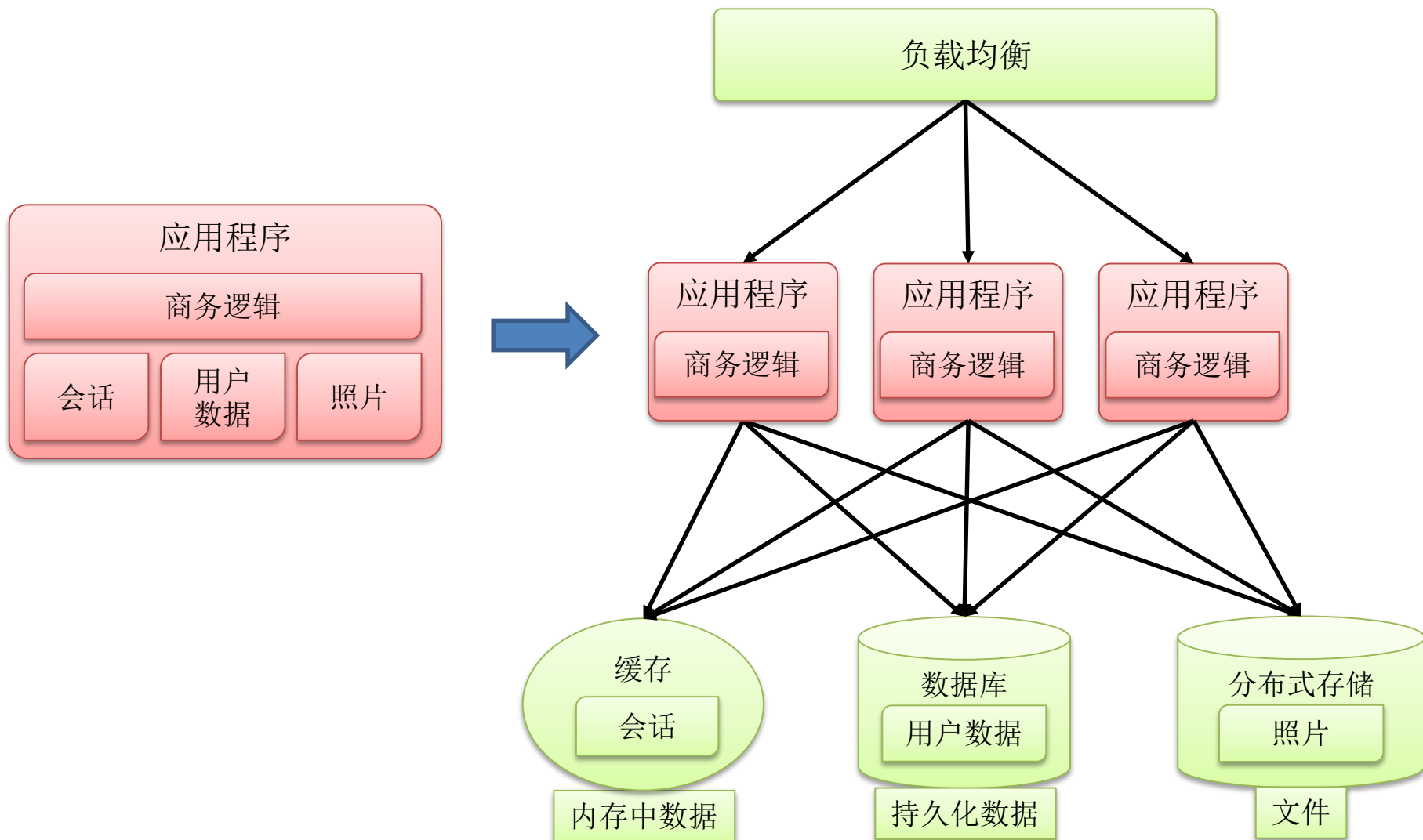
PART 03

容器云

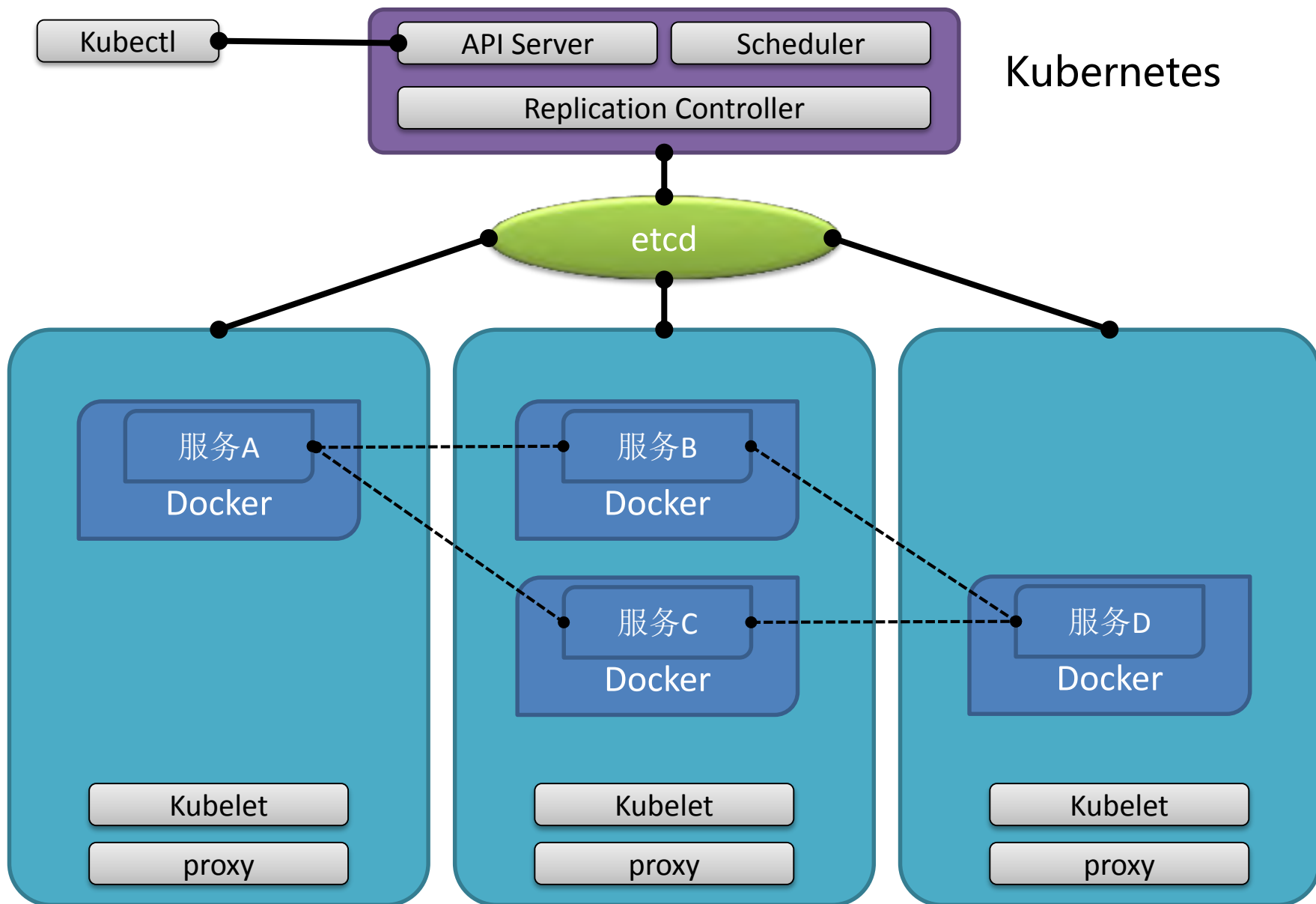
从虚拟机到容器

- 以资源为核心 -> 以应用为核心
- 有状态容器
- 容器跨主机互联
- 容器使用云盘存储

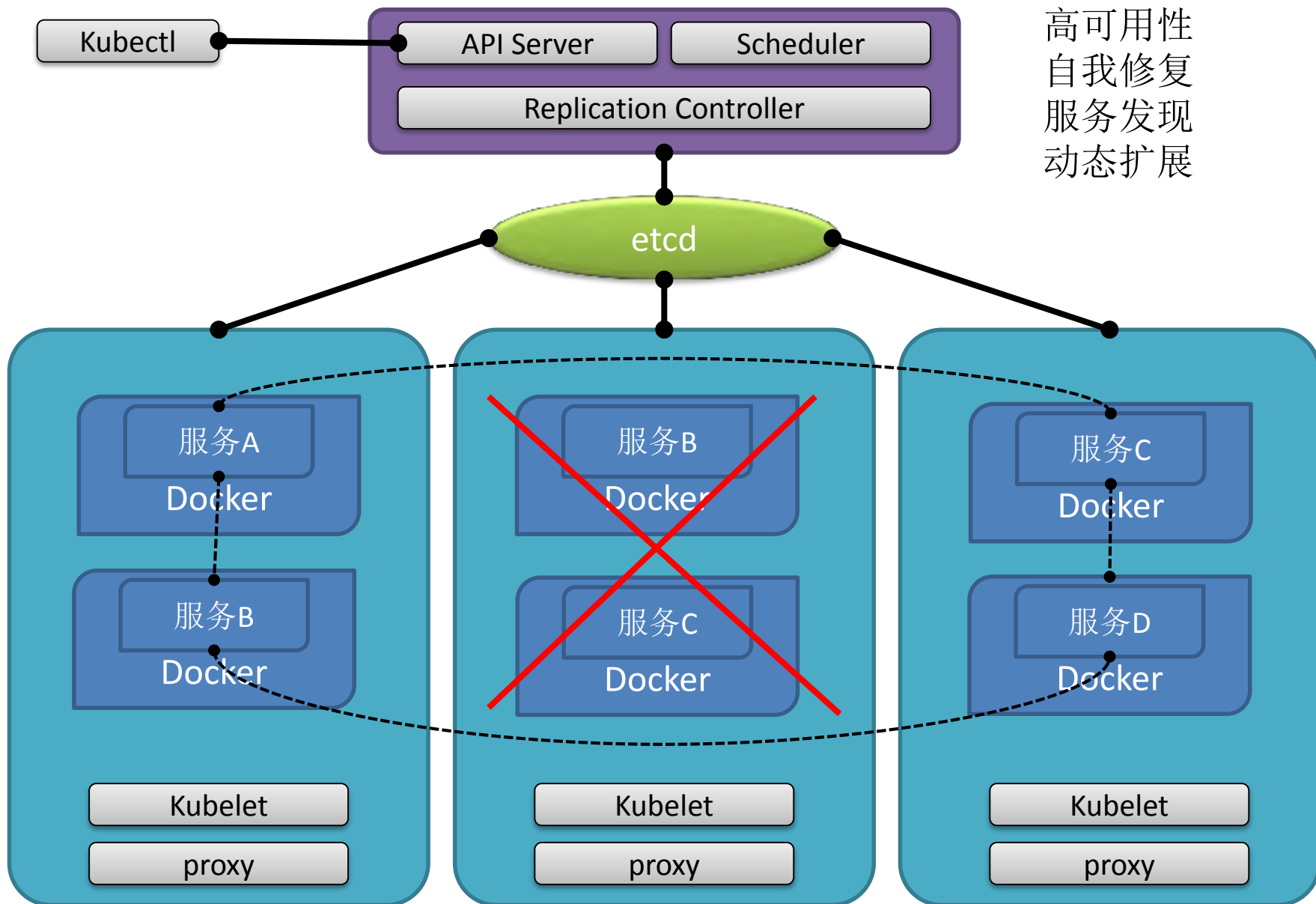
一板斧：去状态、可扩展



二板斧：容器化、可编排



二板斧：容器化、可编排



高可用性
自我修复
服务发现
动态扩展

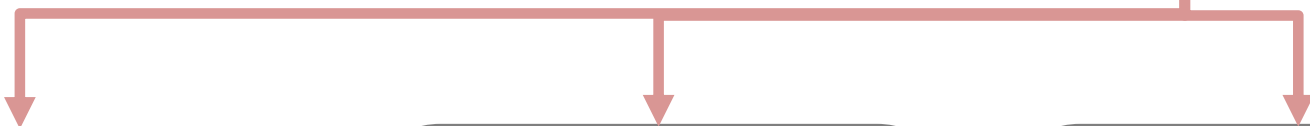
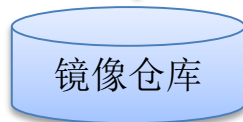
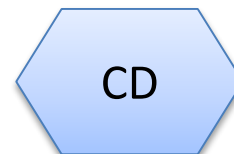
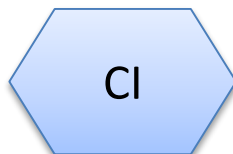
三板斧：DevOps、可迭代



提交



hook



测试环境 注入开发环境配置



Linux OS

Docker 容器

业务代码

运行环境：
应用容器、程序库、系统库、
目录结构、文件权限

联调环境 注入测试环境配置



Linux OS

Docker 容器

业务代码

运行环境：
应用容器、程序库、系统库、
目录结构、文件权限

生产环境 注入生产环境配置



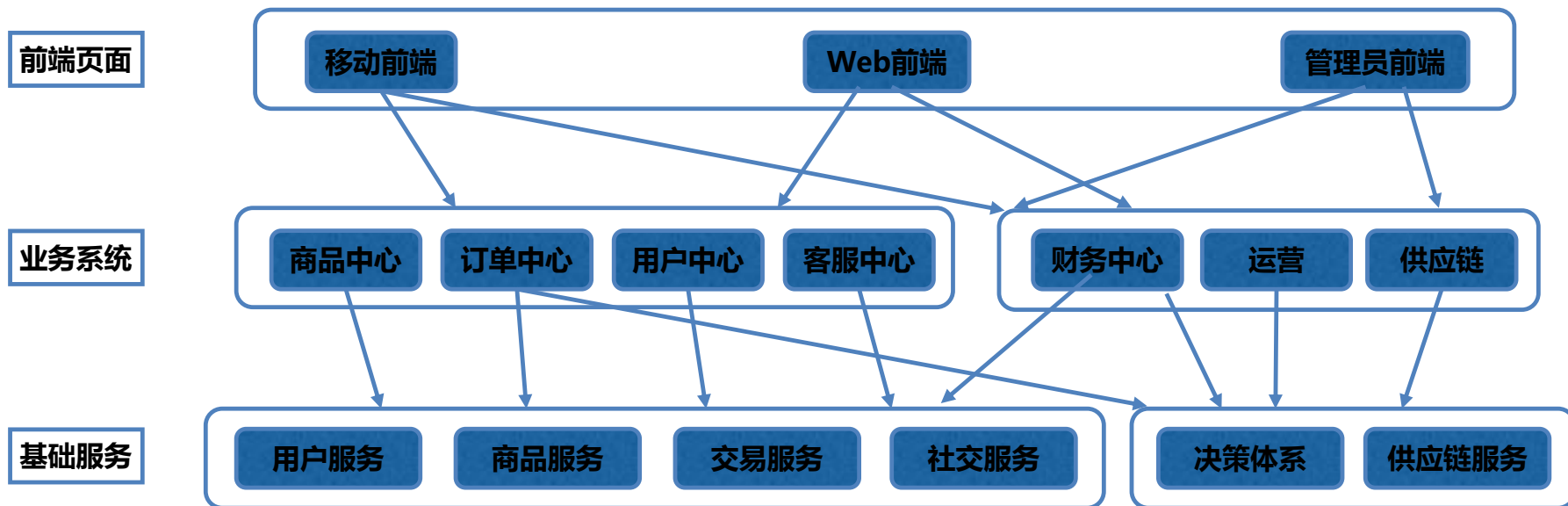
Linux OS

Docker 容器

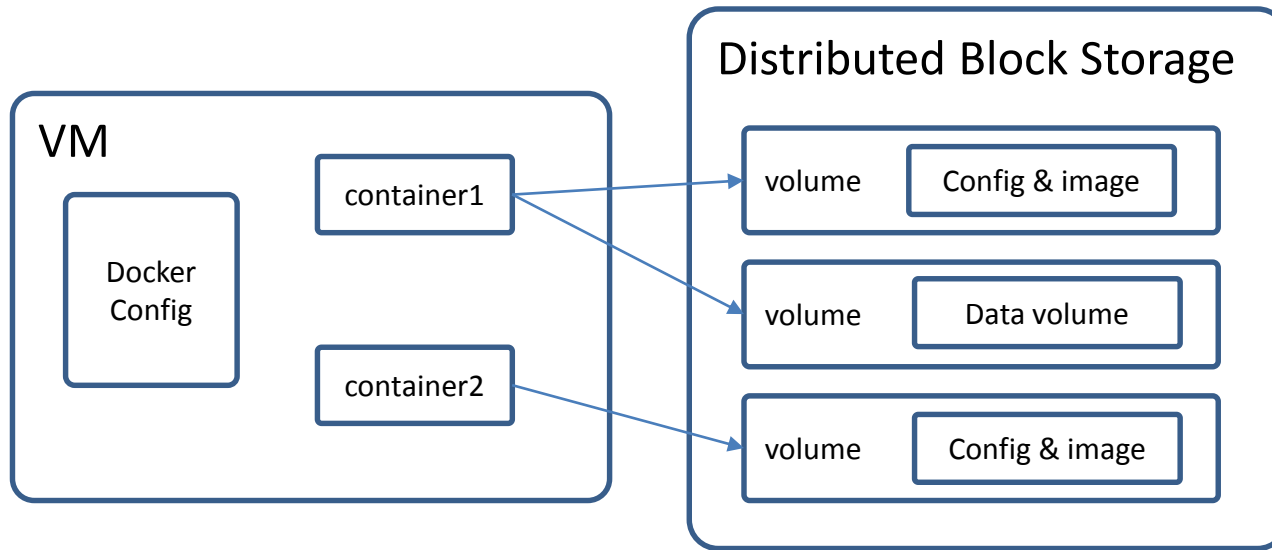
业务代码

运行环境：
应用容器、程序库、系统库、
目录结构、文件权限

微服务架构



有状态容器



- Node 故障时将远程盘挂载到新 Node，Copy 容器配置信息到 Containers 目录，只有重启 Docker Daemon 才能加载配置启动容器。解决方法：增加 reload 指令
- Docker daemon 启动时可以通过 `--graph=` 指定 docker 运行时根目录，但当一个 node 上运行多个容器时，所有容器的配置信息，文件系统相关数据，数据卷都存放在一个根目录下，导致容器无法独立迁移。解决方法：docker run 增加 `container-home=dir` 将容器数据保存在 dir 目录

从私有云到公有云

- 容器的安全
- 容器的启动速度
- 容器的规模
- 容器的租户隔离

网易蜂巢平台

编排

有状态容器

调度优化

多租户

弹性伸缩

服务发现

容器

镜像仓库

镜像构建

日志服务

监控服务

持续集成

数据库

对象存储

负载均衡

缓存服务

CDN

安全服务

OpenStack

计算虚拟化

网络虚拟化

存储虚拟化

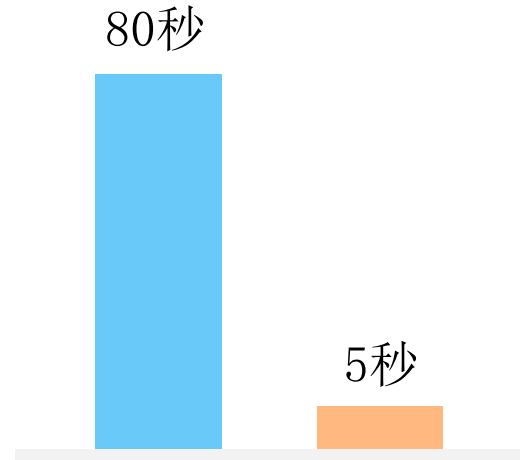
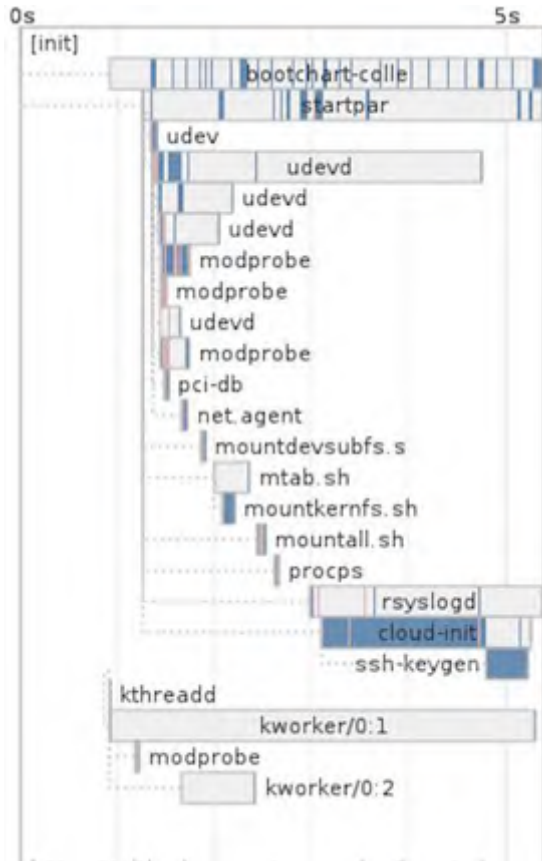
编排的优化

- **支持多租户**: 默认kubernetes的namespace只隔离replication controller , pod 等资源，网易实现节点，存储、网络的租户隔离
- **调度性能优化**：kubernetes调度优化，任务串行队列改为多个优先级队列
- **集群扩展性**：根据Pod/Node/Replication Controller等资源到拆分不同的etcd集群

容器的优化

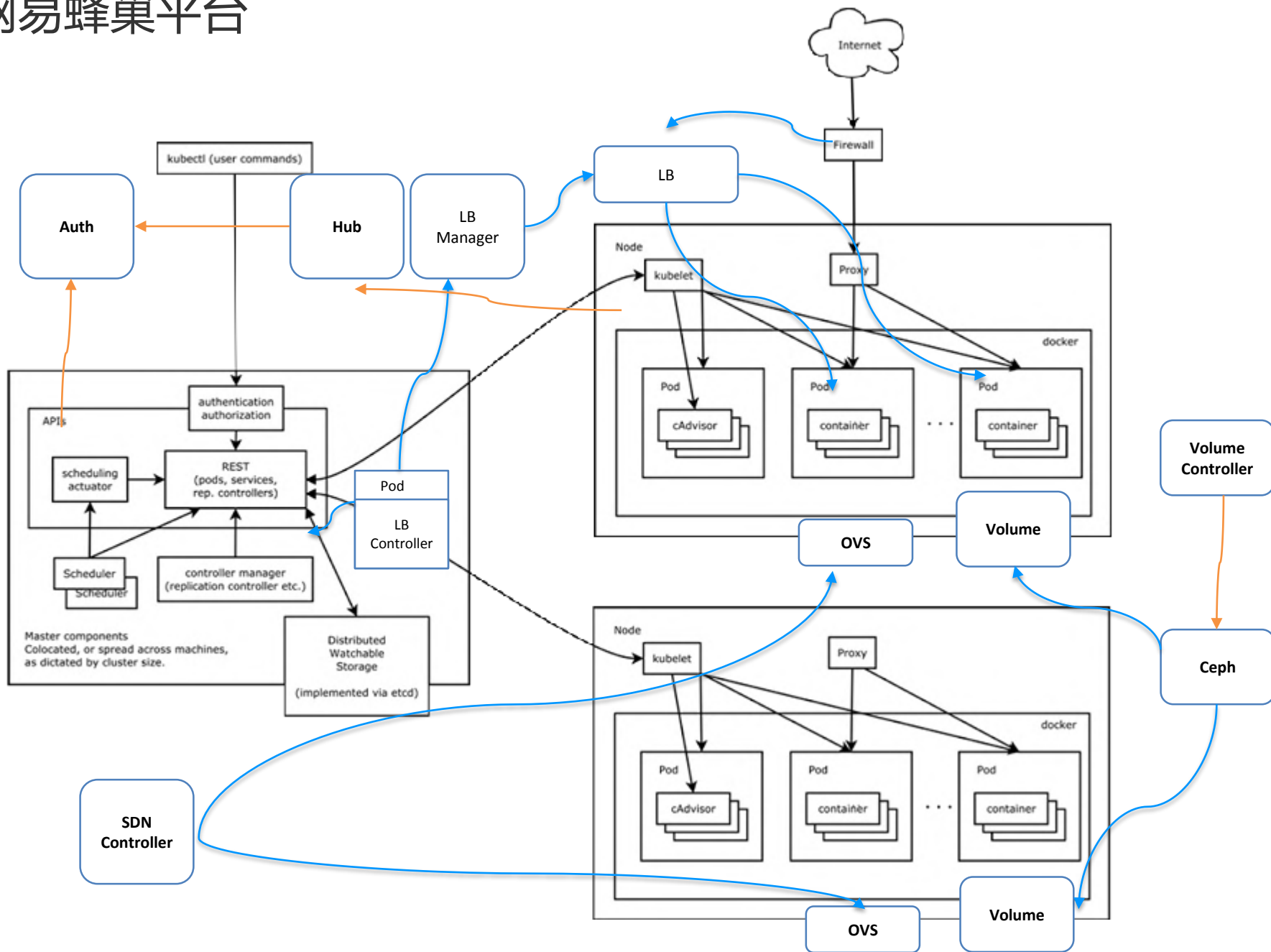
- **虚拟化扁平二层网络**，基于VXLAN实现租户隔离，外网网卡直接挂载到容器内部
- **有状态容器挂载云盘**，可实现跨主机迁移
- **提供统一的日志收集，分析，搜索服务**，利于分布式架构问题定位
- **引入服务端 APM 解决细粒度性能分析**，迅速发掘性能瓶颈

虚拟机启动优化

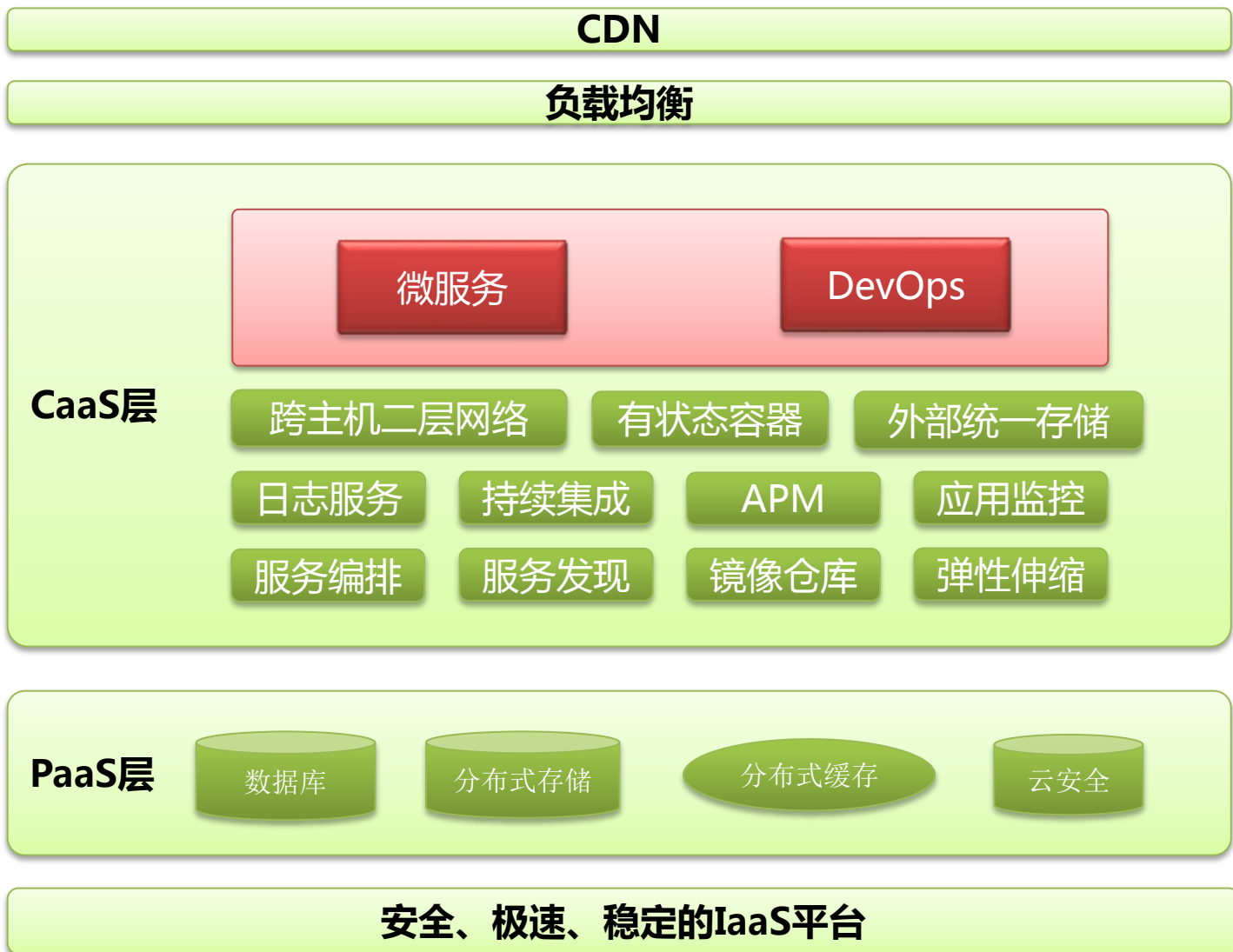


1. 网卡IP初始化
2. 网络路由注入
3. DNS服务IP配置
4. 网卡udev规则

网易蜂巢平台



蜂巢特色：聚焦应用



蜂巢特色：全开源平台



最流行的开源数据库
活跃的社区和日趋完善的功能
网易RDS提供标准接口和稳定保障



产品理念完善历史15年
社区热度高Github Star 18200+
企业级应用案例多，最佳实践经验足
透明开源，技术标准

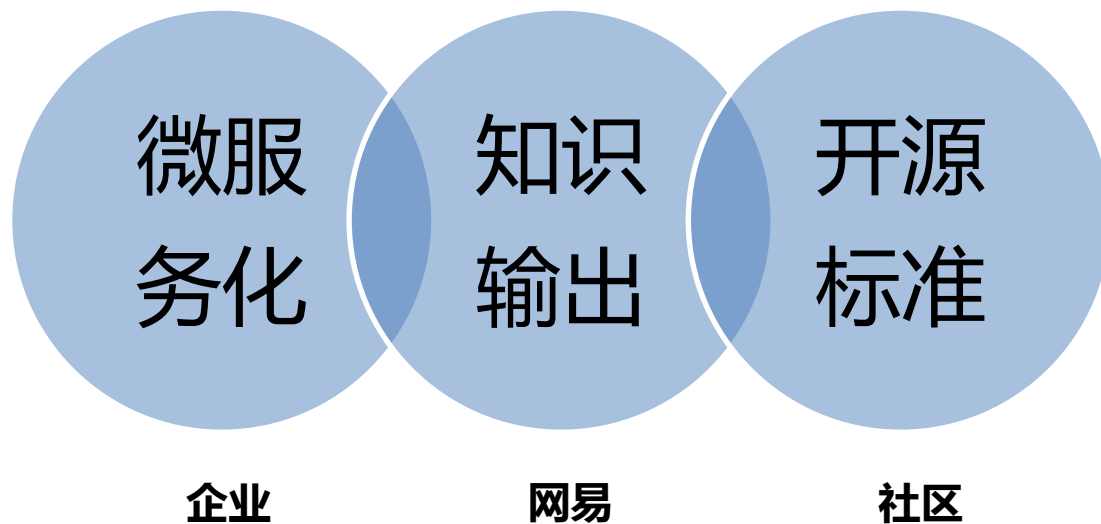


最流行的缓存服务
性能最快的Key-Value DB
网易NCR提供标准接口和稳定保障



最流行的开放IaaS平台
100%支持标准API
紧跟社区，新版本发布后3个月平台更新

网易蜂巢助力企业微服务化





PART 04

联系我们

Contact Us

C.163.COM

网易蜂巢，欢迎大家使用！



关注“网易蜂巢”微信公共账号 获取网易蜂巢最新动态！

EMAIL: cloudcomb@188.com