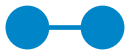
公有云里的容器

邱剑 架构师

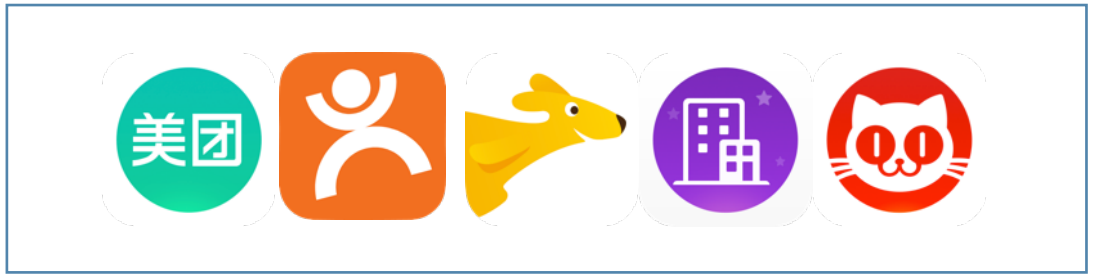
GIAC 2016



美团云
Meituan Open Services

目录

关于美团云(MOS)

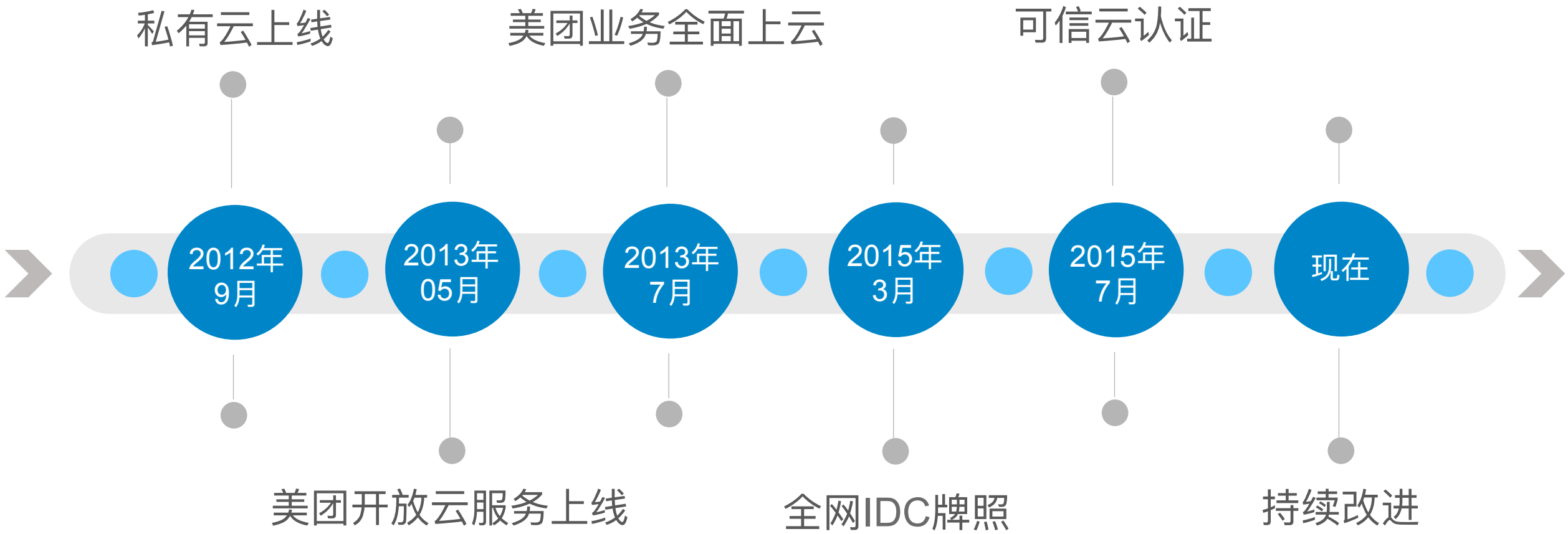


私有云

公有云

美团云
Meituan Open Services

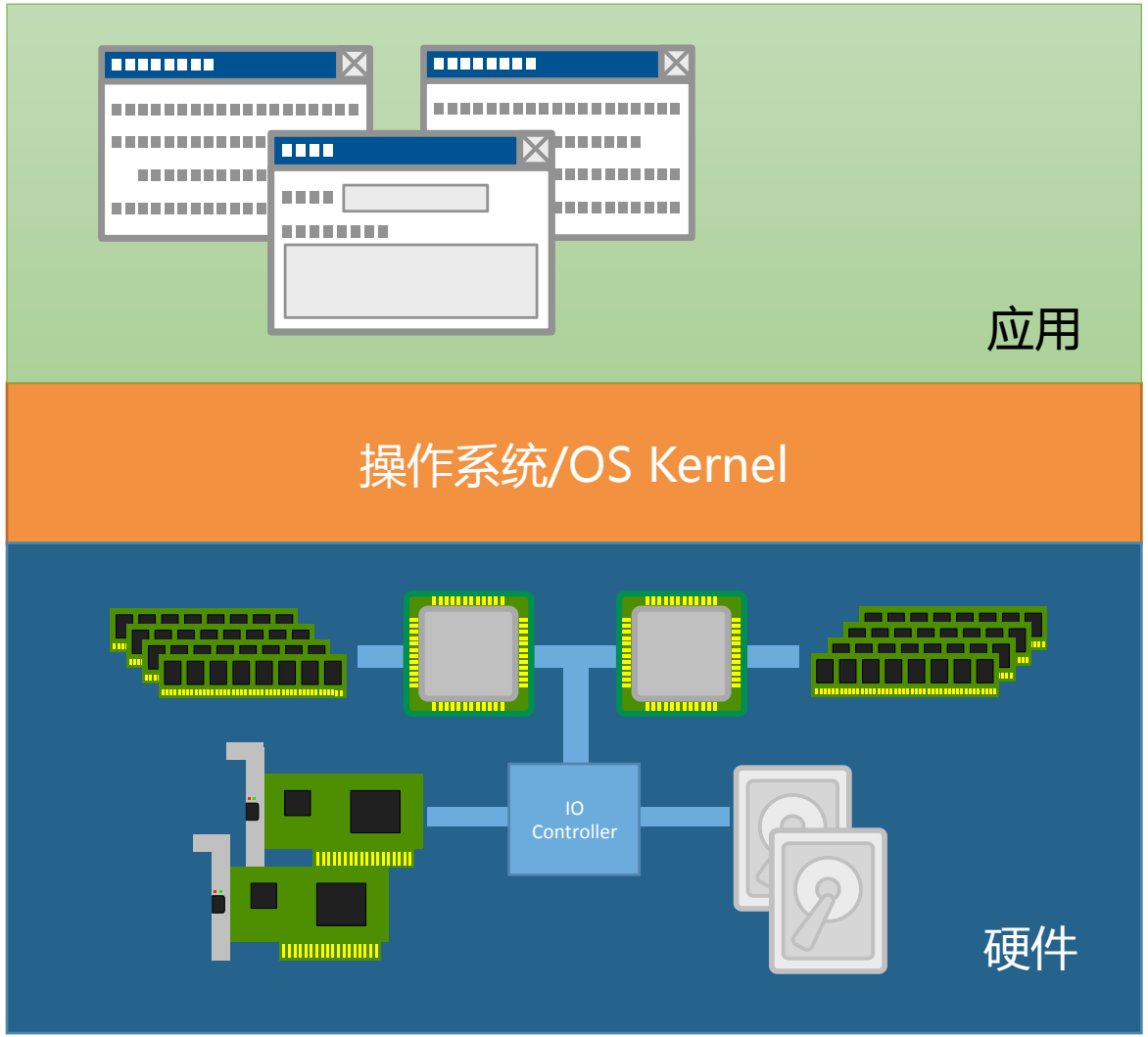
美团云发展历程



美团云
Meituan Open Services

目录

主机

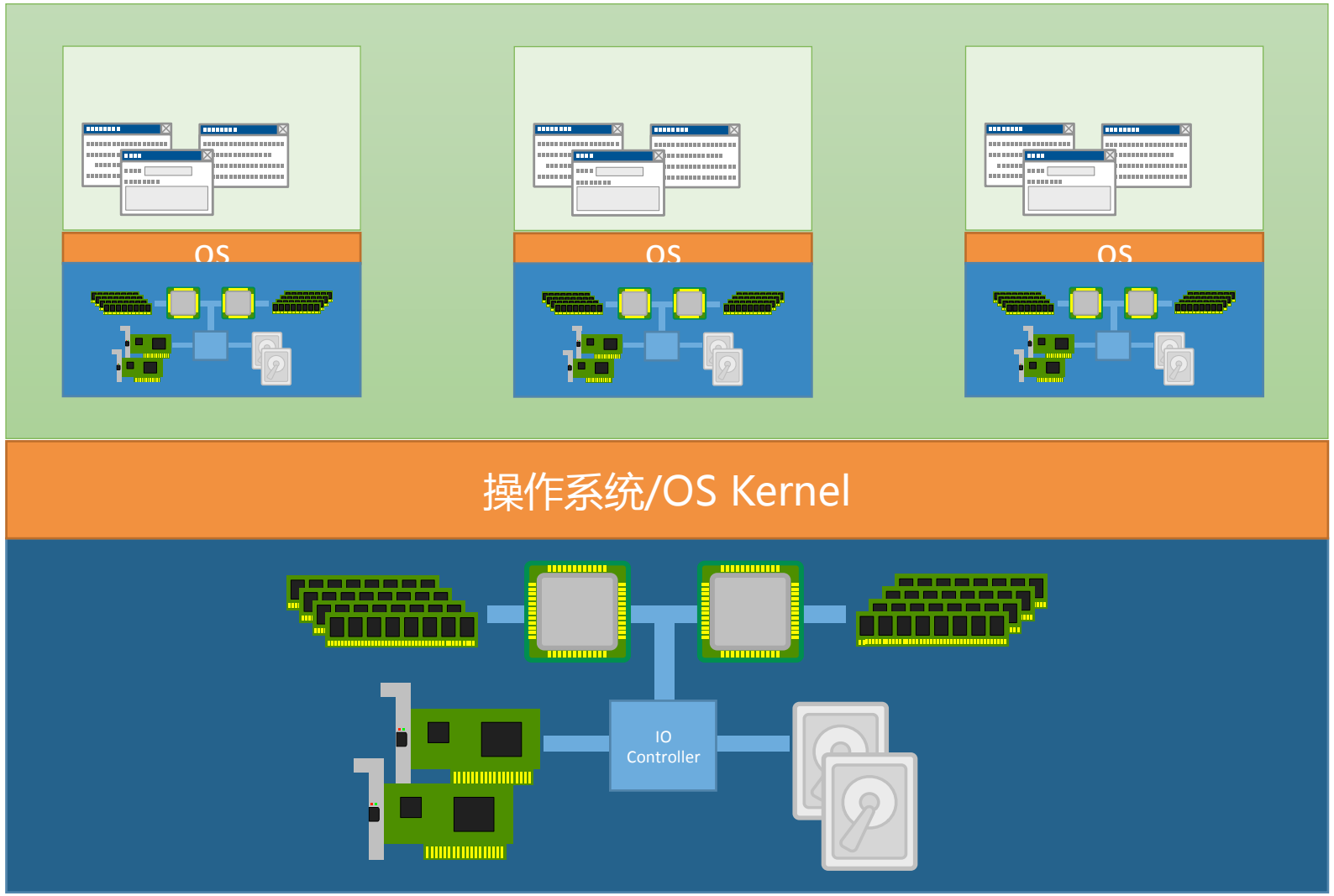


系统调用
System call

硬件接口
HW Interfaces

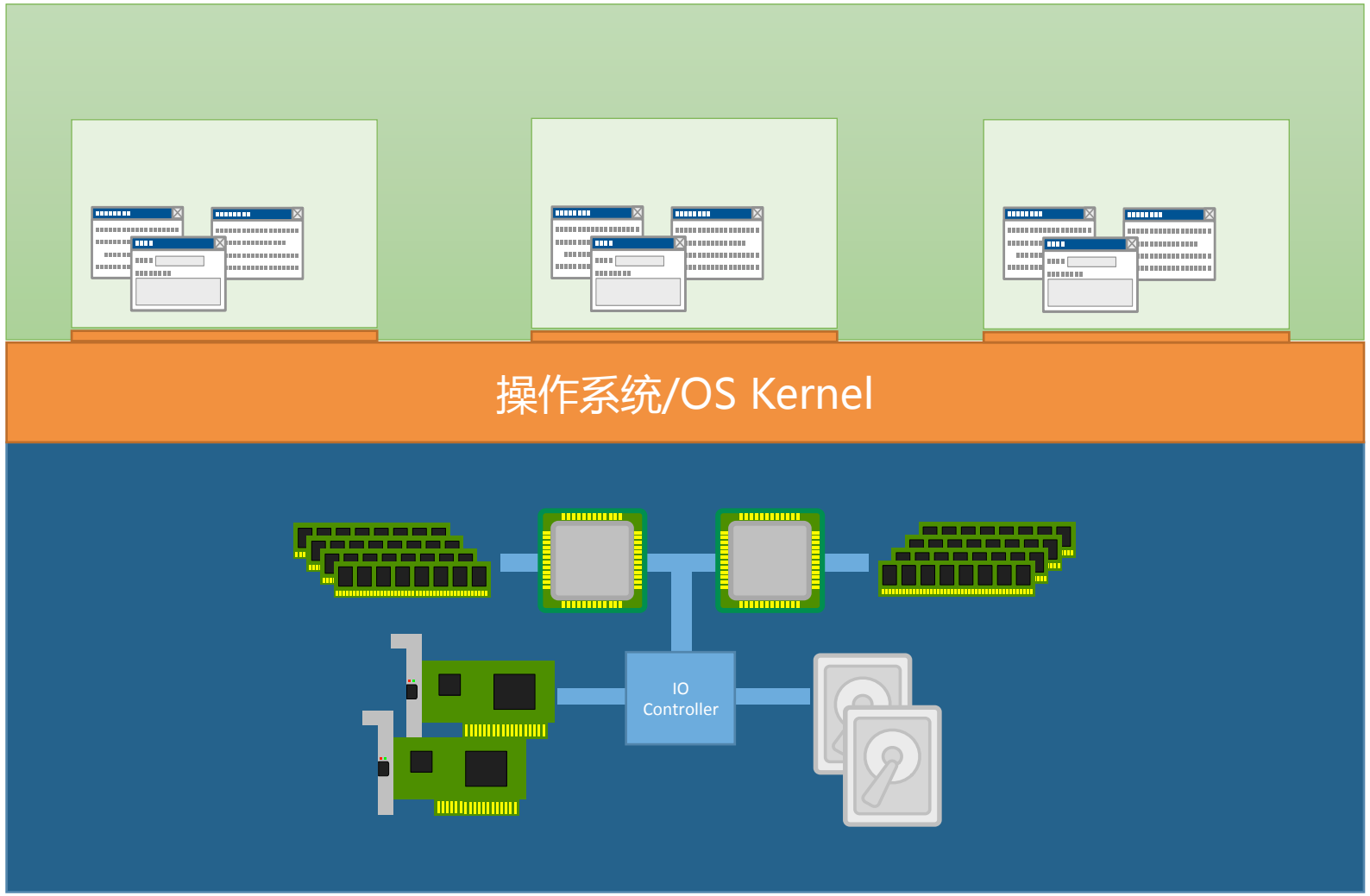
美团云
Meituan Open Services

虚拟机



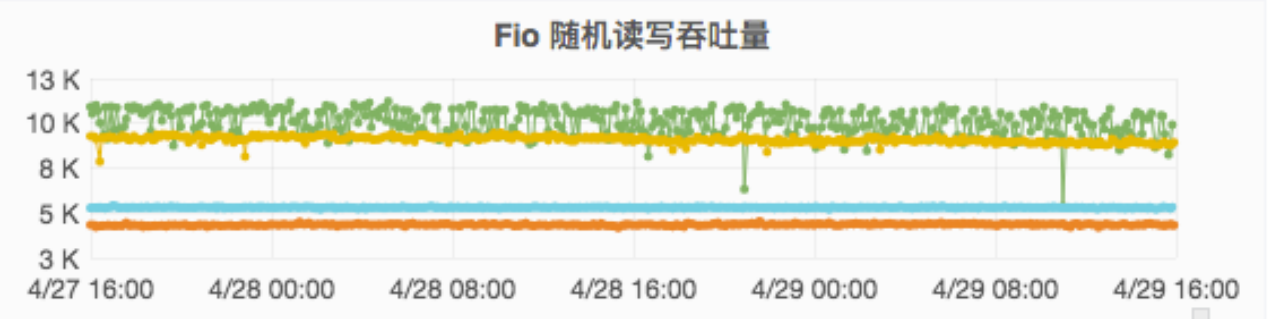
美团云
Meituan Open Services

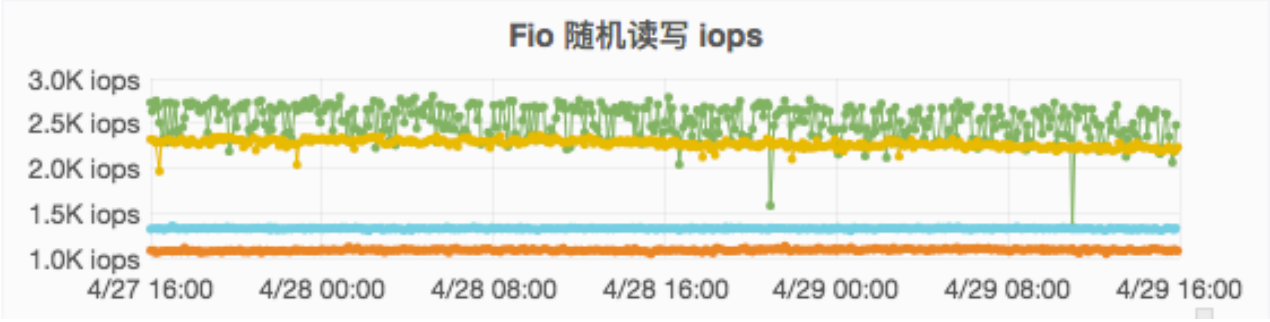
容器



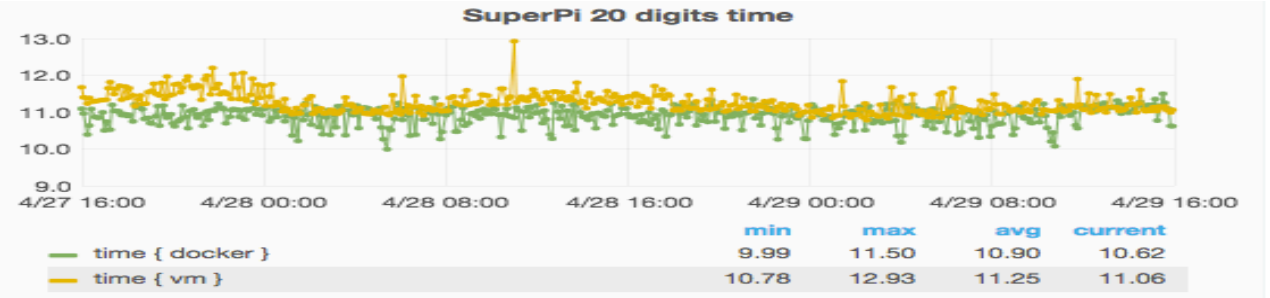
美团云
Meituan Open Services

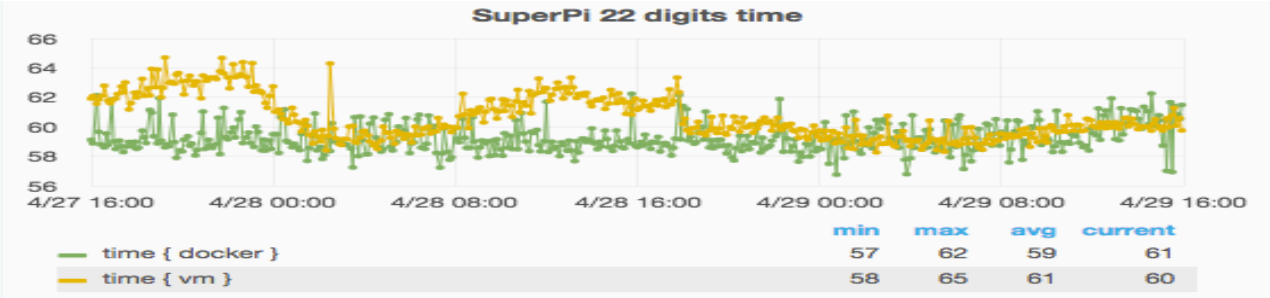
虚拟机 vs 容器 —— 性能对比



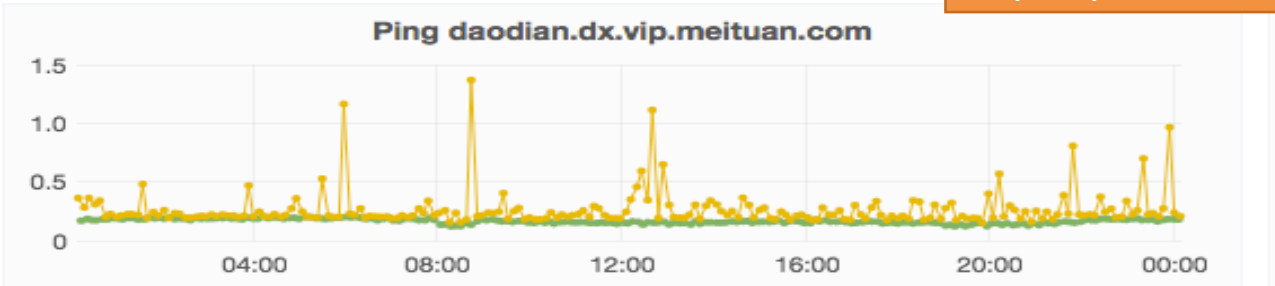


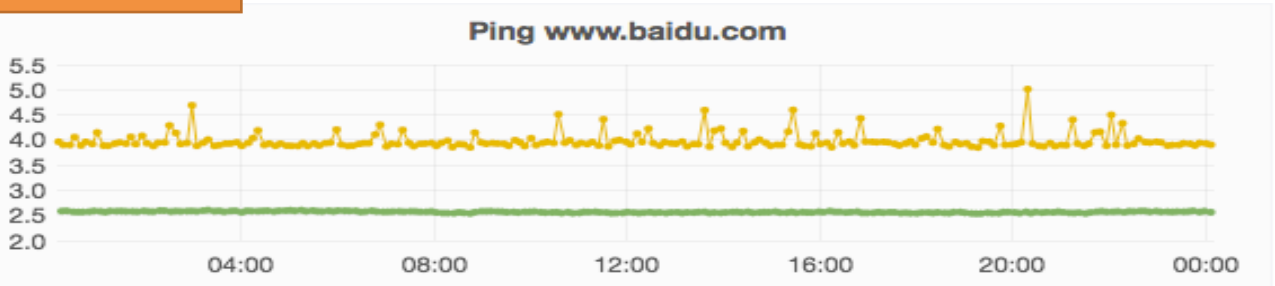
Fio random write: VM/Docker = 82% read: VM/Docker = 90%





Super pi: VM/Docker = 97%

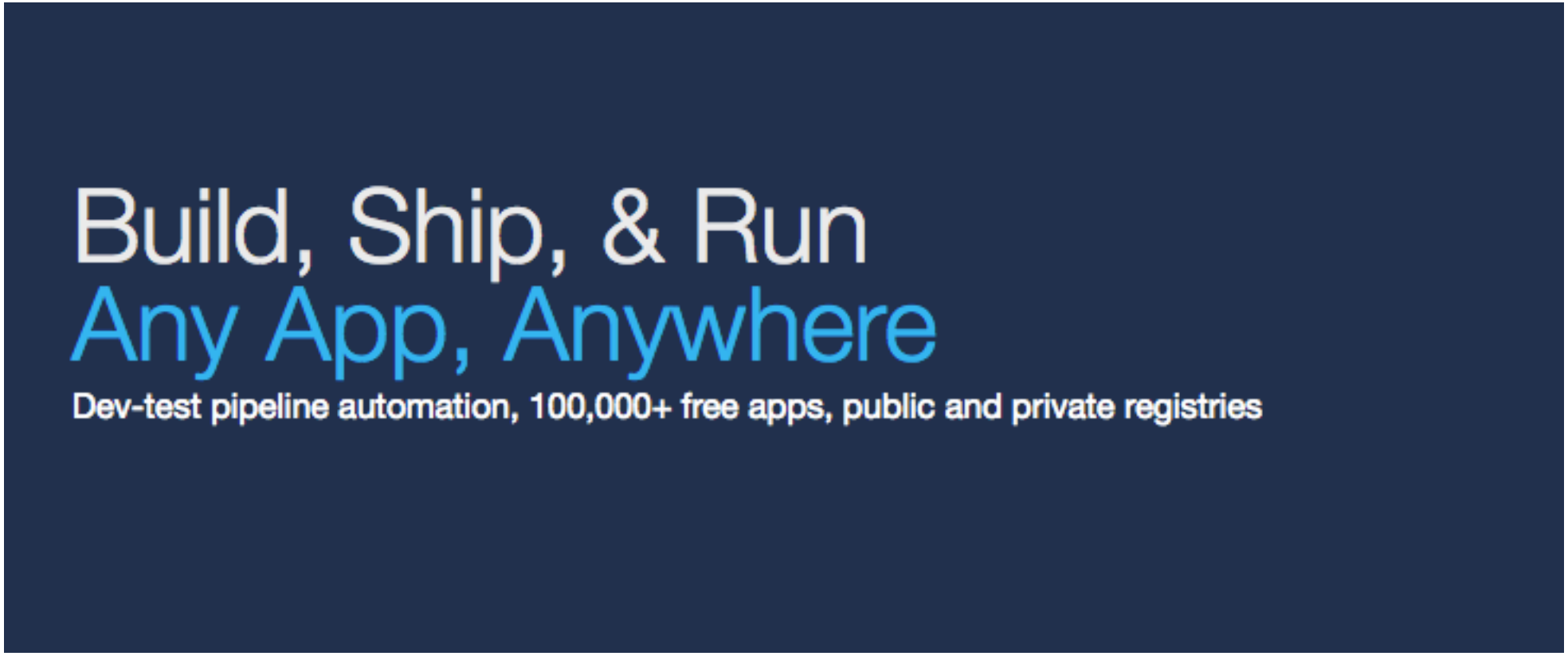




美团云
Meituan Open Services

容器应用镜像

- Docker Hub

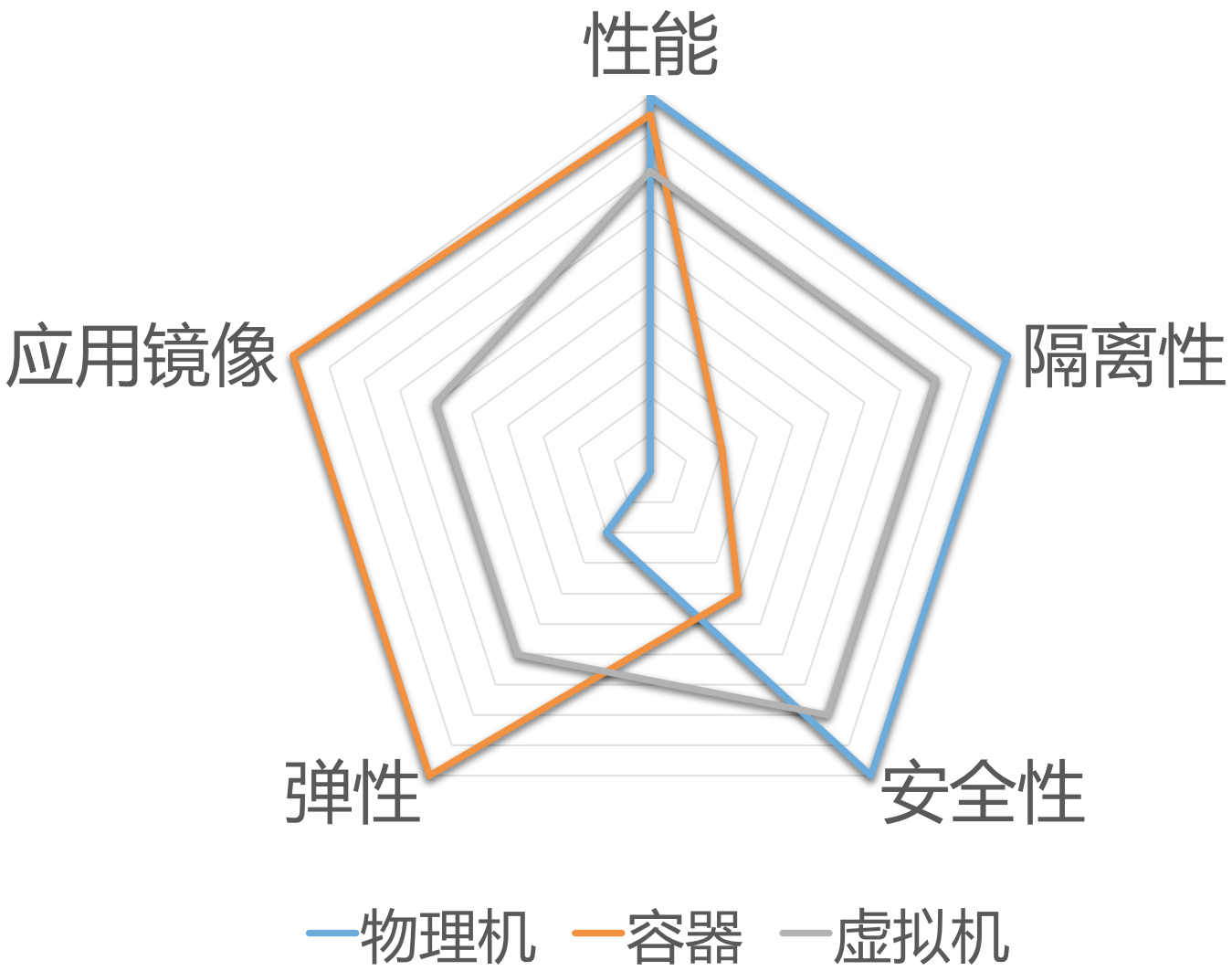


美团云
Meituan Open Services

容器的安全性

- Dirty COW - (CVE-2016-5195) - Docker Container Escape
  - https://blog.paranoidsoftware.com/dirty-cow-cve-2016-5195-docker-container-escape/
- 进程炸弹
  - 各个容器的进程共享宿主机调度器
  - pids cgroup
- Buffered IO隔离
  - Cgroup blkio 无法限制buffered IO (pre kernel 4.2)

虚拟机 vs 容器



美团云
Meituan Open Services

目录

- 关于美团云(MOS)
- 虚拟机 vs 容器(Container)
- MOS Container
- Container@MOS

美团云
Meituan Open Services

MOS Container

虚拟机

物理机

容器

美团云平台

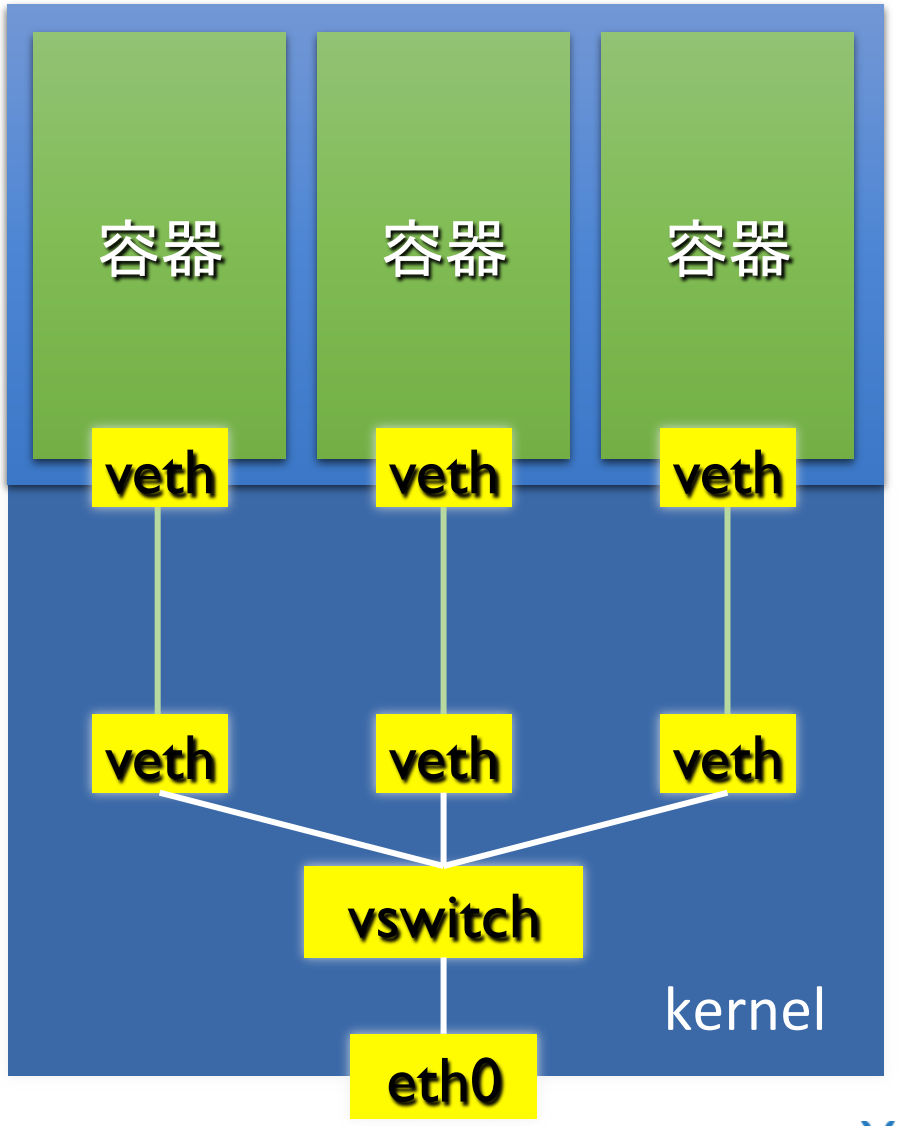
美团云
Meituan Open Services

MOS Container

- MosDocker
  - 基于Docker 1.11.0
  - 主要特性：
    - 解决Docker Daemon退出导致容器退出的问题
    - MosBridge：支持美团云网络
    - 支持挂载基于LVM的Volume
    - 支持监控，保存镜像，迁移，更改cgroup配置等管理功能
    - 支持容器和VM的混合部署
    - 若干BUG fixes

MOS Container

- mosBridge
  - 支持OVS和MVS
  - 持久化网络配置
  - 支持容器重启后的自动配置



美团云
Meituan Open Services

MOS Container

- 基于LVM的Volume
  - Volume over direct-LVM
  - 良好的本地IO性能
  - 支持Volume限容！

MOS Container

- 本地加速Docker registry mirror
  - 预加载Docker Hub的官方镜像
  - 本地透明拦截Docker pull请求

目录

- 关于美团云(MOS)
- 虚拟机 vs 容器(Container)
- MOS Container
- Container@MOS
  - PaaS in Container
  - Container in KVM

美团云
Meituan Open Services

Container@MOS

- 博采众长
- 对用户透明

PaaS in Container

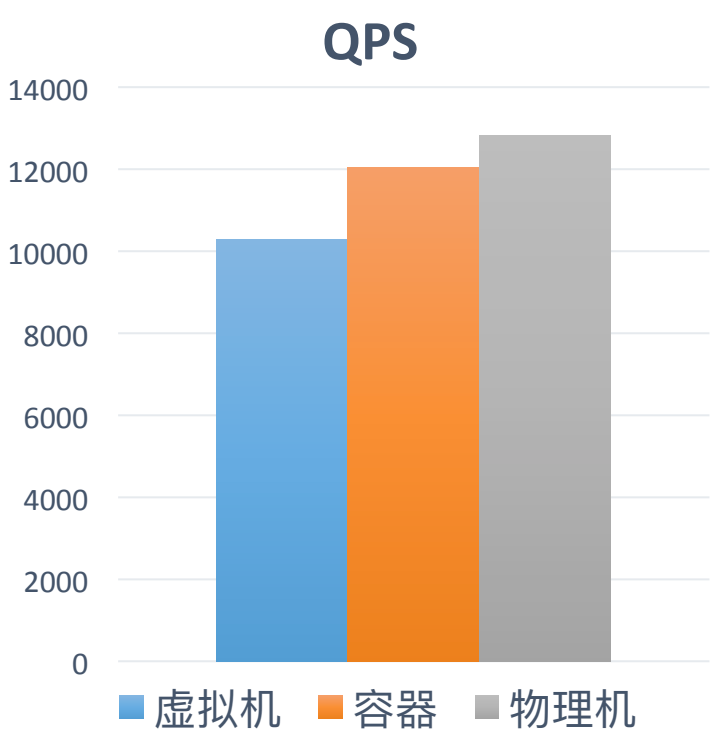
- 容器提供基础的隔离性
  - 网络隔离
  - IO隔离
  - CPU、内存隔离
  - 权限隔离
- 容器保障高性能和快速启动
- 目前支持PaaS应用：MySQL，MongoDB，Redis，Memcache

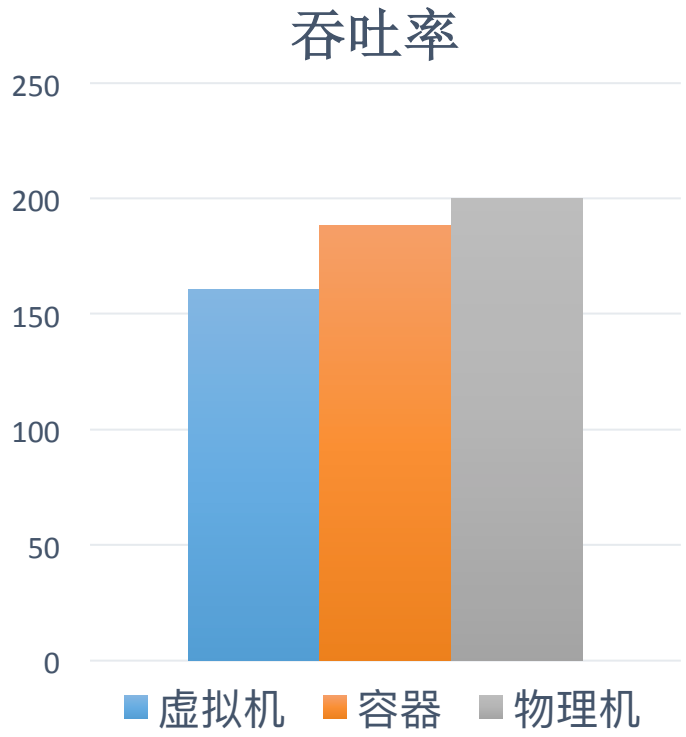
PaaS in Container

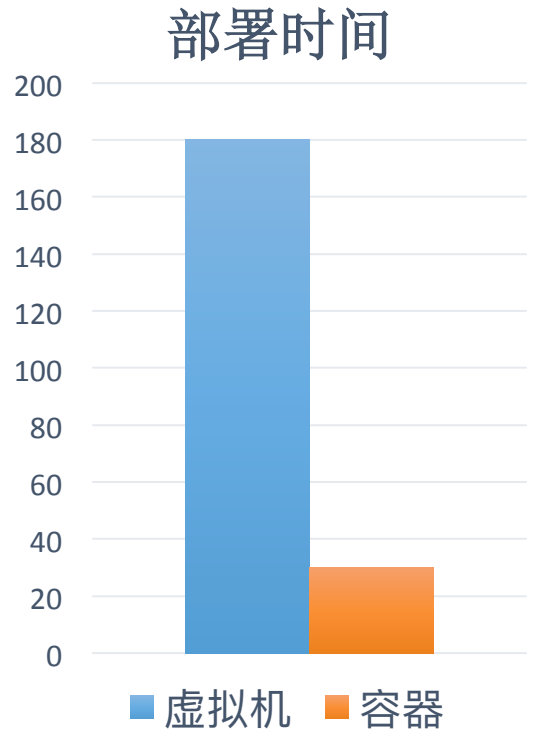
- 安全和隔离性加固
  - 自定义镜像
    - 修复漏洞
      - CVE-2016-6662
      - CVE-2015-4335
      - CVE-2016-8339
    - 稳定版本
  - 禁止执行shell命令和访问本地文件
  - UDF白名单
  - 保留系统账户，禁用高危权限，提供低权限普通账户
  - 强制Direct IO

PaaS in Container

Xeon E5-2650 v2 32C 128G 7200转SATA RAID10

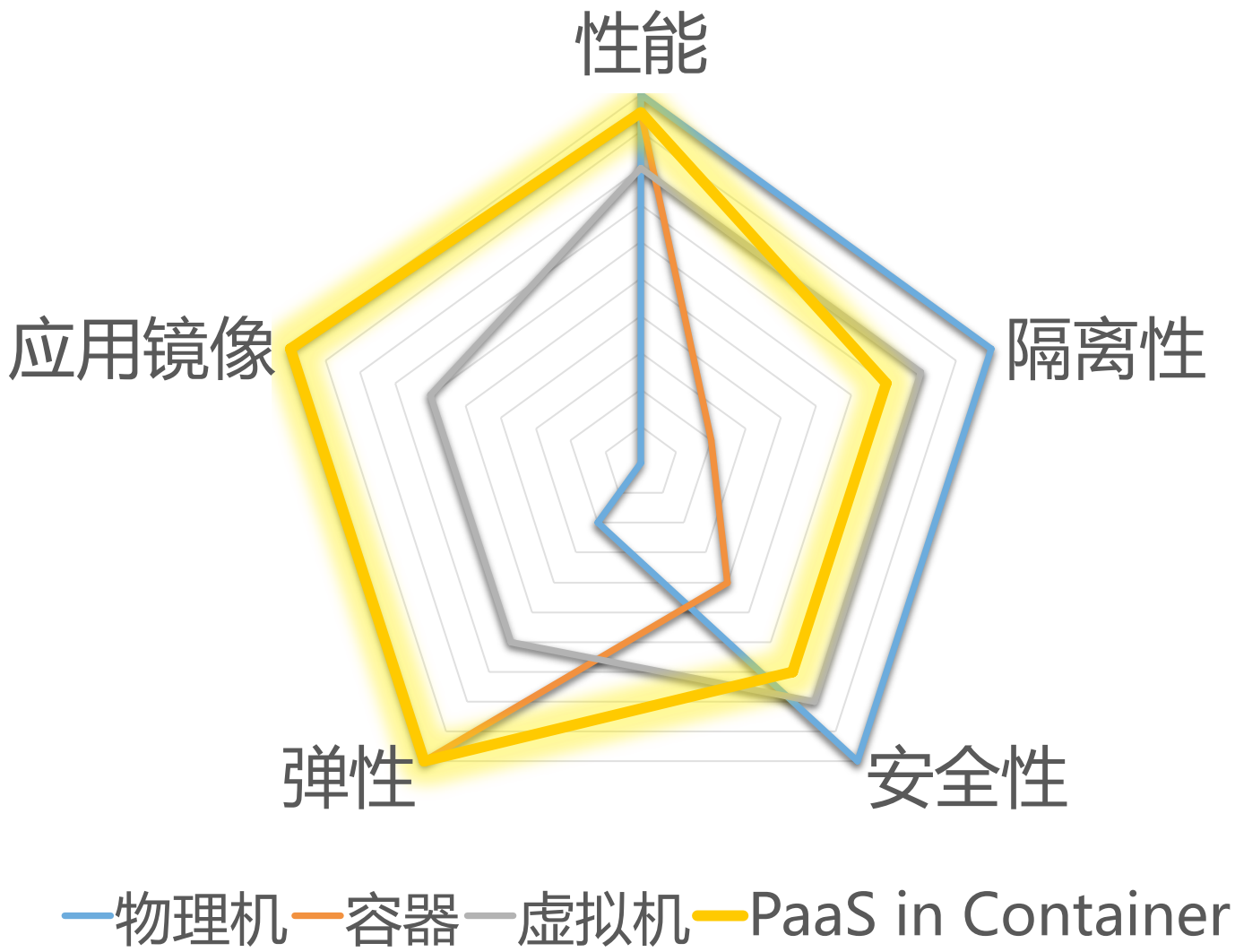






美团云
Meituan Open Services

PaaS in Container



美团云
Meituan Open Services

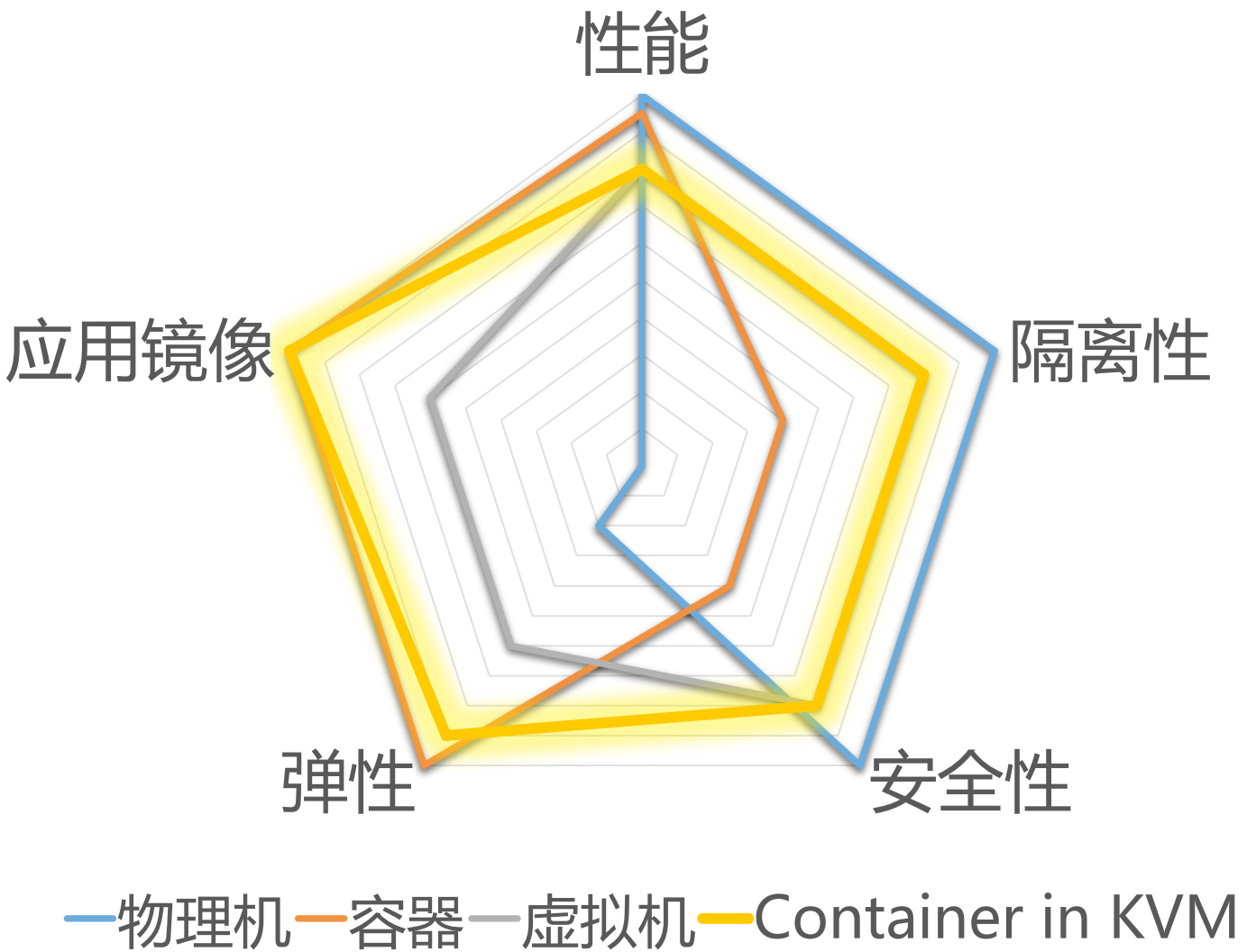
Container in KVM

- 普通模式
  - Run container in VM
  - 优点：
    - VM的隔离性
    - 方便易上手
  - 缺点
    - VM操作系统浪费资源
    - 容器启动慢
    - 两层网络虚拟化开销

Container in KVM

- 轻量模式
  - 裁剪CentOS 7系统，2s启动时间
  - 挂载容器的LVM分区
  - VM内核启动后自动拉起容器
  - 优点：
    - KVM的隔离性
    - 直接使用容器镜像
    - Native network stack

Container in KVM



美团云
Meituan Open Services

小结

- 虚拟机 vs 容器 —— 各有所长
- 美团云对容器的改进工作
- 在公有云中使用容器
  - PaaS in Container
  - Container in KVM

美团云
Meituan Open Services

专业提升效率 稳定创造价值

https://mos.meituan.com

美团云
Meituan Open Services