美团云docker实践

郑坤@美团云

Jan. 25 2016

Agenda

- Why and How?
- 美团云Docker实践
  - ✦美团云Docker框架
  - ✦工程方法论：复用已有的轮子
  - ✦美团云Docker相关组件和流程
  - ✦改造我们的镜像仓库
  - ✦自定义容器网络
  - ✦set - 多容器管理单元
  - ✦内部推广和规划

美团云

Why Docker

- 更轻量：基于容器的虚拟化，仅包含业务运行所需的runtime环境，CentOS/Ubuntu基础镜像仅170M；宿主机可部署100～1000个容器
- 更高效：无操作系统虚拟化开销
  - ✦ 计算：轻量，无额外开销
  - ✦ 存储：系统盘aufs/dm/overlayfs；数据盘volume
  - ✦ 网络：宿主机网络，NS隔离
- 更敏捷、更灵活：
  - ✦ 分层的存储和包管理，devops理念
  - ✦ 支持多种网络配置

How to use docker in MOS

Docker资源

仓库：Docker registry

编排：Swarm，Kubernetes …

网络：?

Host：?

美团云

仓库：VM Image仓库

计算与控制：美团云控制节点 …

网络：美团云网络

Host：VM宿主机

我们的目标

- 为公司业务提供Docker基础服务
- 公司业务向Docker的低成本迁移
- 快速实现原型、摸索积累经验
- 复用美团云已有架构、模块和基础设施
- 发挥Docker轻量、弹性、高性能的优势
- 合理的运维成本

在现有云平台架构基础上，通过扩展一些功能模块，实现Docker服务

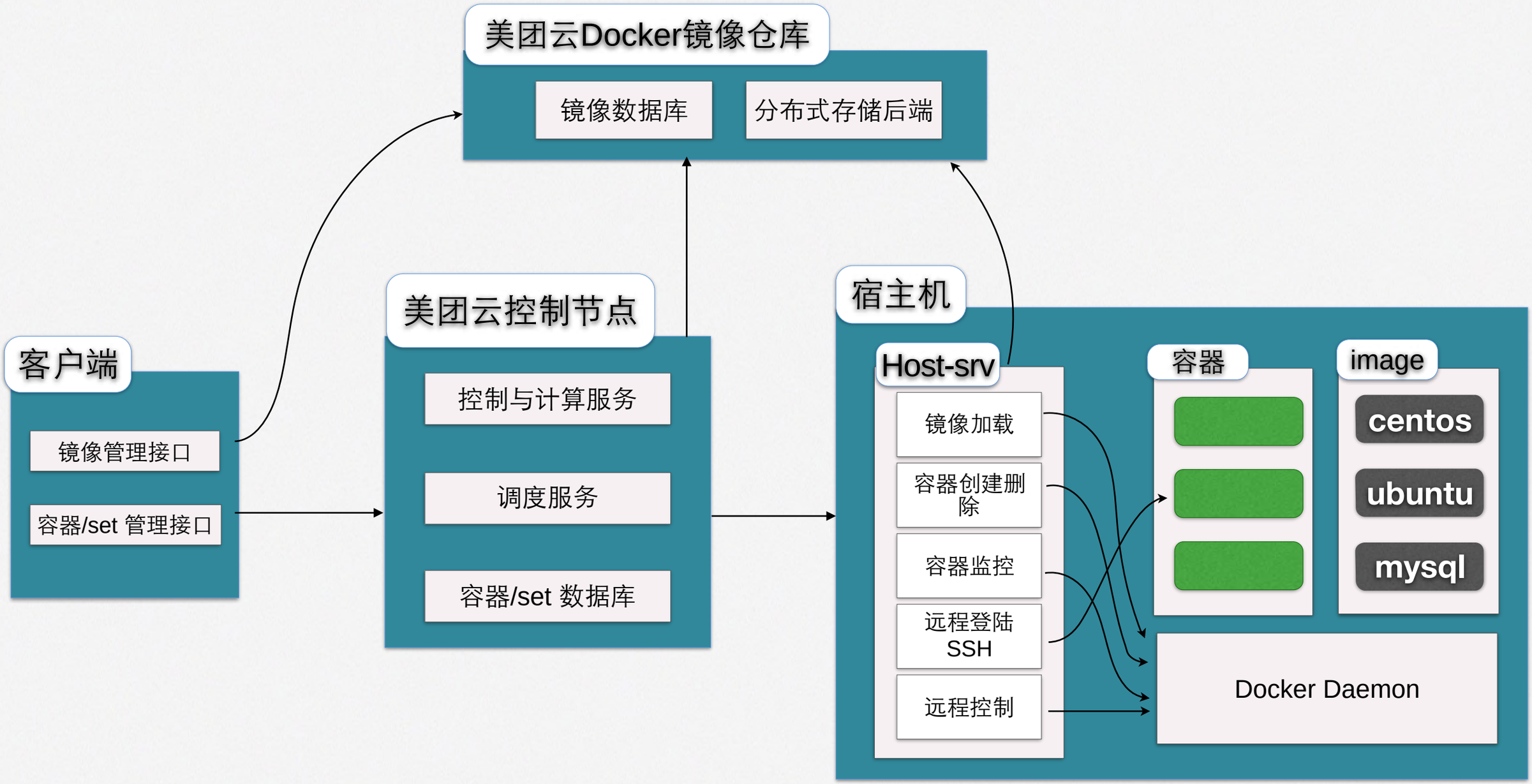
美团云

设计原则&实施方案

- 原则：
  - ✦ 复用美团云现有架构，快速开发、部署和上线
  - ✦ 优化云平台，体现Docker优势
- 方案
  - ✦ 控制、计算、调度框架：基于云平台VM框架扩展
  - ✦ 镜像仓库：基于云平台glance和swift服务，自建Docker image仓库
  - ✦ 宿主机：Host-SRV管理VM和Docker 容器
  - ✦ 网络：容器和VM一样通过宿主机OVS连接外部网络
  - ✦ 隔离：基于glance/region实现用户的image/container
  - ✦ 鉴权：基于token做image/container流程

美团云

美团云Docker框架

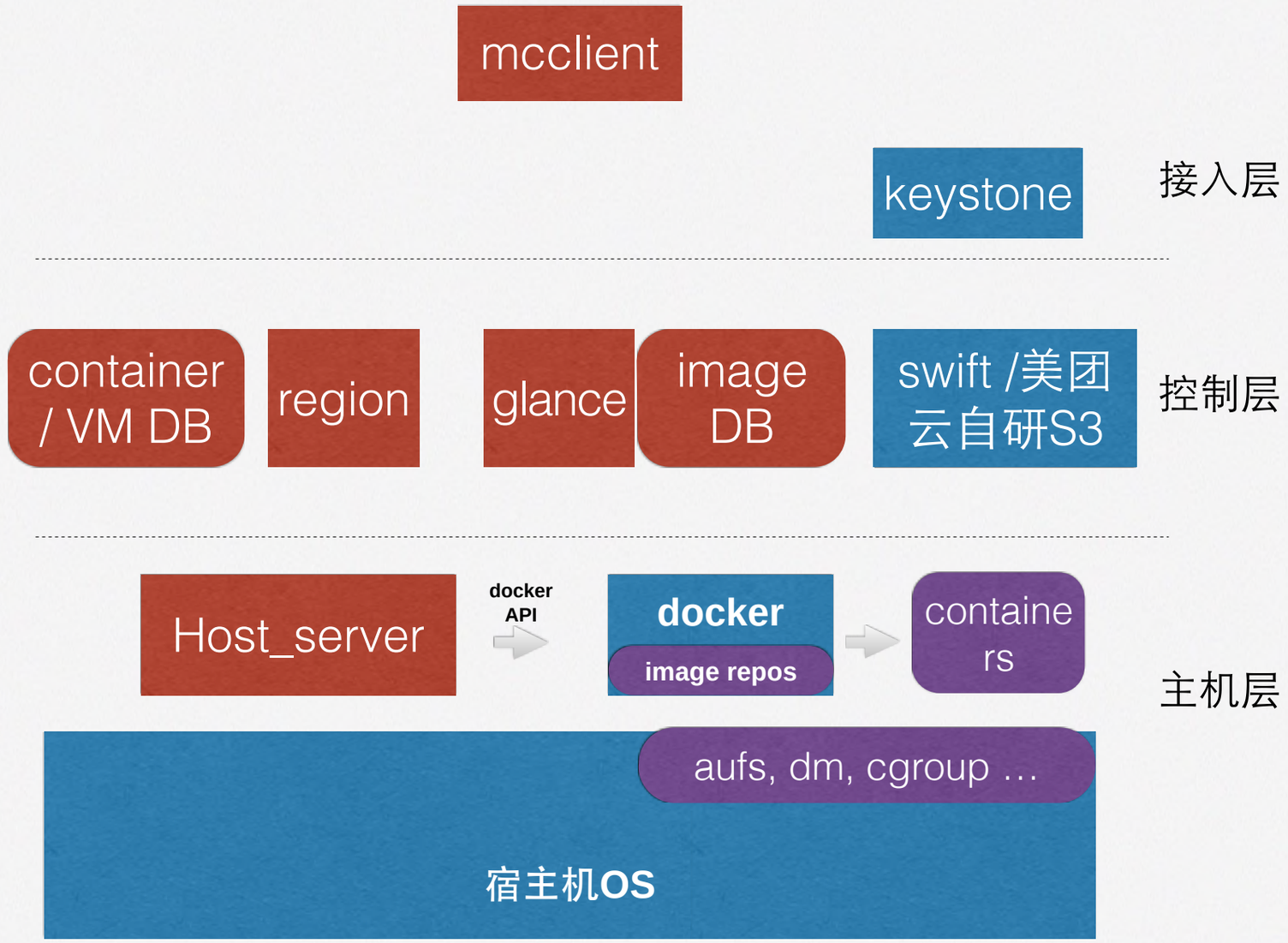


工程方法论 - 复用现有架构、模块和基础设施

- API复用
  - ✦create, start, stop, suspend, sync-status, delete, update, deploy, rebuild_disk, snapshot, attach-network, …
- Task复用
- 数据结构复用:
  - ✦Guests, Disks, Images, Storages, Network, …
- 宿主机基础设施复用
  - ✦Image缓存
  - ✦网络
  - ✦本地存储管理

美团云Docker服务相关组件

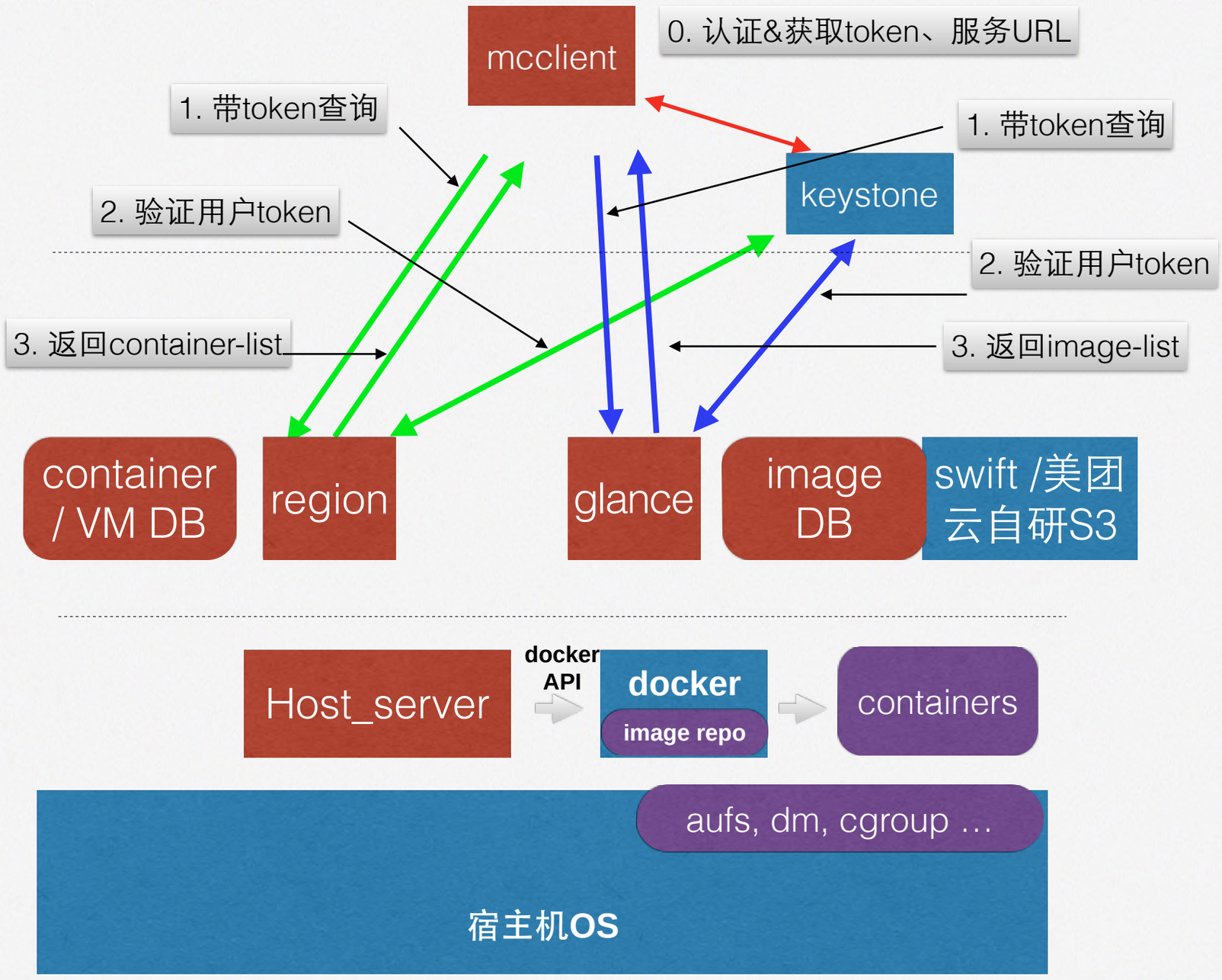
模块	来源	功能
mcclient	自主开发	CLI Access
identity	keystone	用户管理、目录服务、API
region	自主开发	资源管理、任务管理、API、日志
object storage	swift / 自主研发对象存储	对象存储
host	自主开发	network、storage、hypervisor
image	glance	镜像管理



Docker镜像/容器查询流程

镜像查询

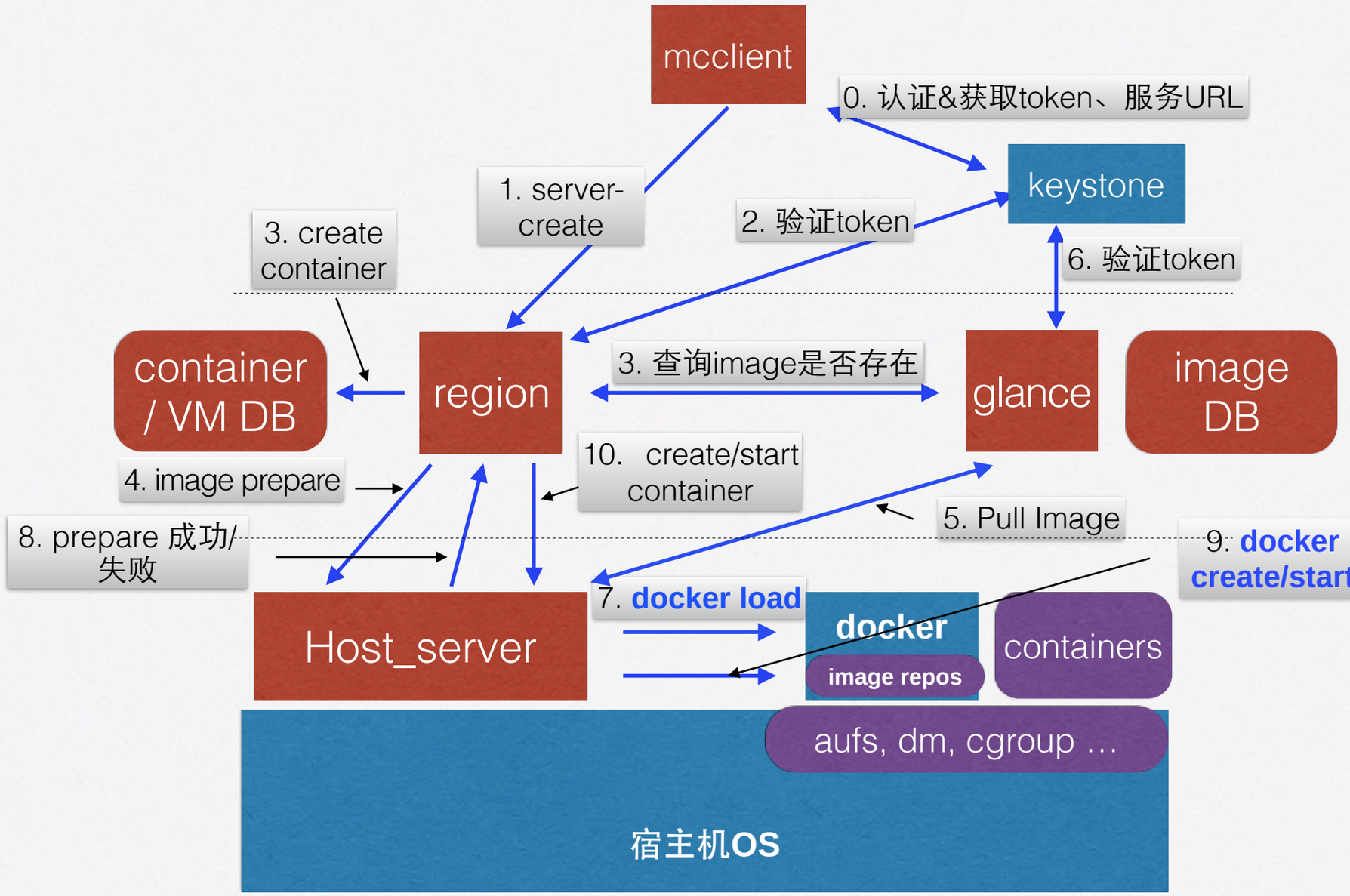
容器查询



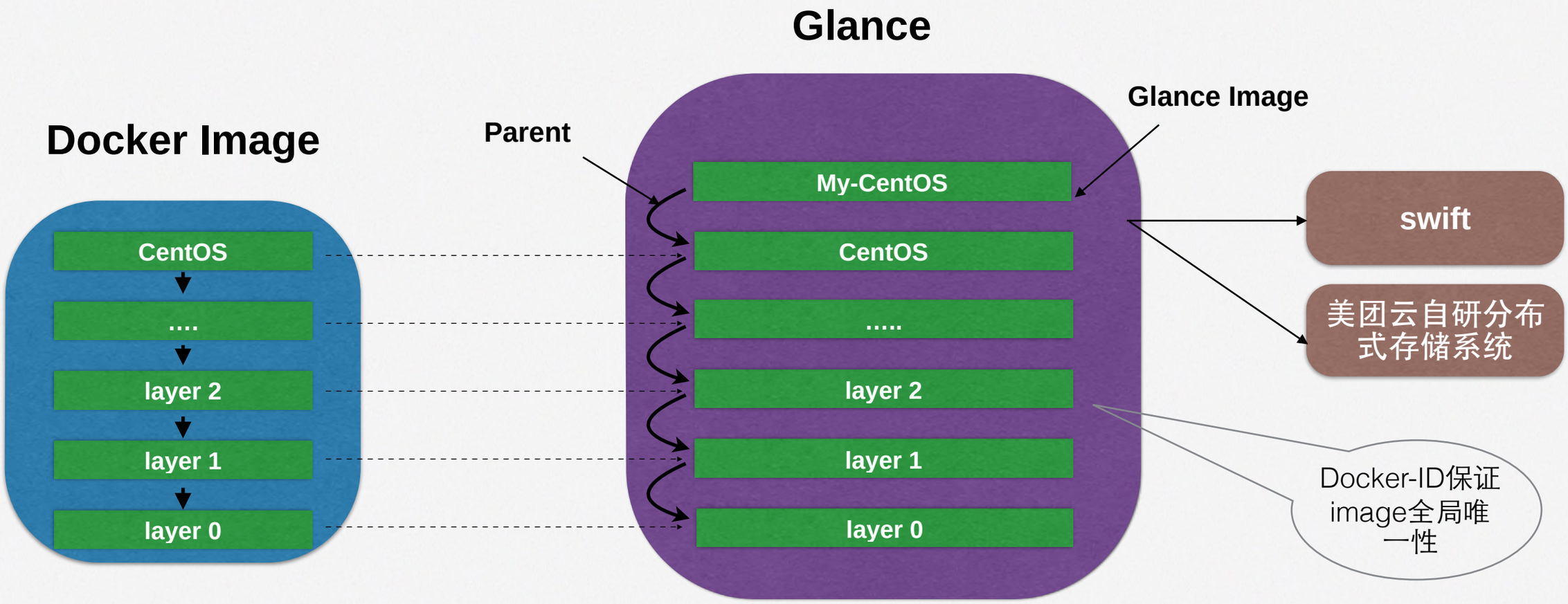
容器创建流程

主要阶段：

- 控制、调度和计算
  - ✦容器调度
  - ✦数据库更新
- image准备
  - ✦逐个layer pull到宿主机
  - ✦缓存避免重复pull
- 容器创建与启动
  - ✦create和start 分开
  - ✦deploy file
- 启动后配置
  - ✦网络
  - ✦sshd/密码设置



Docker镜像仓库：改造Glance




```
zhengkun@ubuntu-2:~$ climc image-list --history
+--------------------------------------+-----------------------------+-------------+----------+-----------+----------+---------+--------+------------+--------------------------------------+
| ID                                   | Name                        | Disk_format | Size     | Is_public | Min_disk | Min_ram | Status | Preference | Parent_id                            |
+--------------------------------------+-----------------------------+-------------+----------+-----------+----------+---------+--------+------------+--------------------------------------+
| fb9205ac-ee9d-4a93-a8cf-dc284725a62e | docker-centos               | docker      | 10240    | False     | 0        | 0       | active | 0          | 4dd5113f-64c0-4b12-9651-fce231aa2eb6 |
| 4dd5113f-64c0-4b12-9651-fce231aa2eb6 |                             | docker      | 10240    | False     | 0        | 0       | active | 0          | feabbbf4-2860-4f9a-b7e4-497472107483 |
| feabbbf4-2860-4f9a-b7e4-497472107483 |                             | docker      | 10240    | False     | 0        | 0       | active | 0          | 4a7bc8ed-f855-487e-91bf-b6b42f03e8e2 |
| 4a7bc8ed-f855-487e-91bf-b6b42f03e8e2 |                             | docker      | 10240    | False     | 0        | 0       | active | 0          | f71c1bc2-cd34-4515-af7e-9d94e2acd468 |
| f71c1bc2-cd34-4515-af7e-9d94e2acd468 |                             | docker      | 202004480| False     | 0        | 0       | active | 0          | 69240924-bd2f-4406-962e-53dfc3ab8100 |
| 69240924-bd2f-4406-962e-53dfc3ab8100 |                             | docker      | 10240    | False     | 0        | 0       | active | 0          |                                      |
| 666e690a-a5ff-464a-9276-a719755940bf | cirros-0.3.2-x86_64-disk.img | qcow2       | 13167616 | True      | 0        | 0       | active | 0          |                                      |
+--------------------------------------+-----------------------------+-------------+----------+-----------+----------+---------+--------+------------+--------------------------------------+
**** Total: 7 ****
```

美团云

自定义容器网络

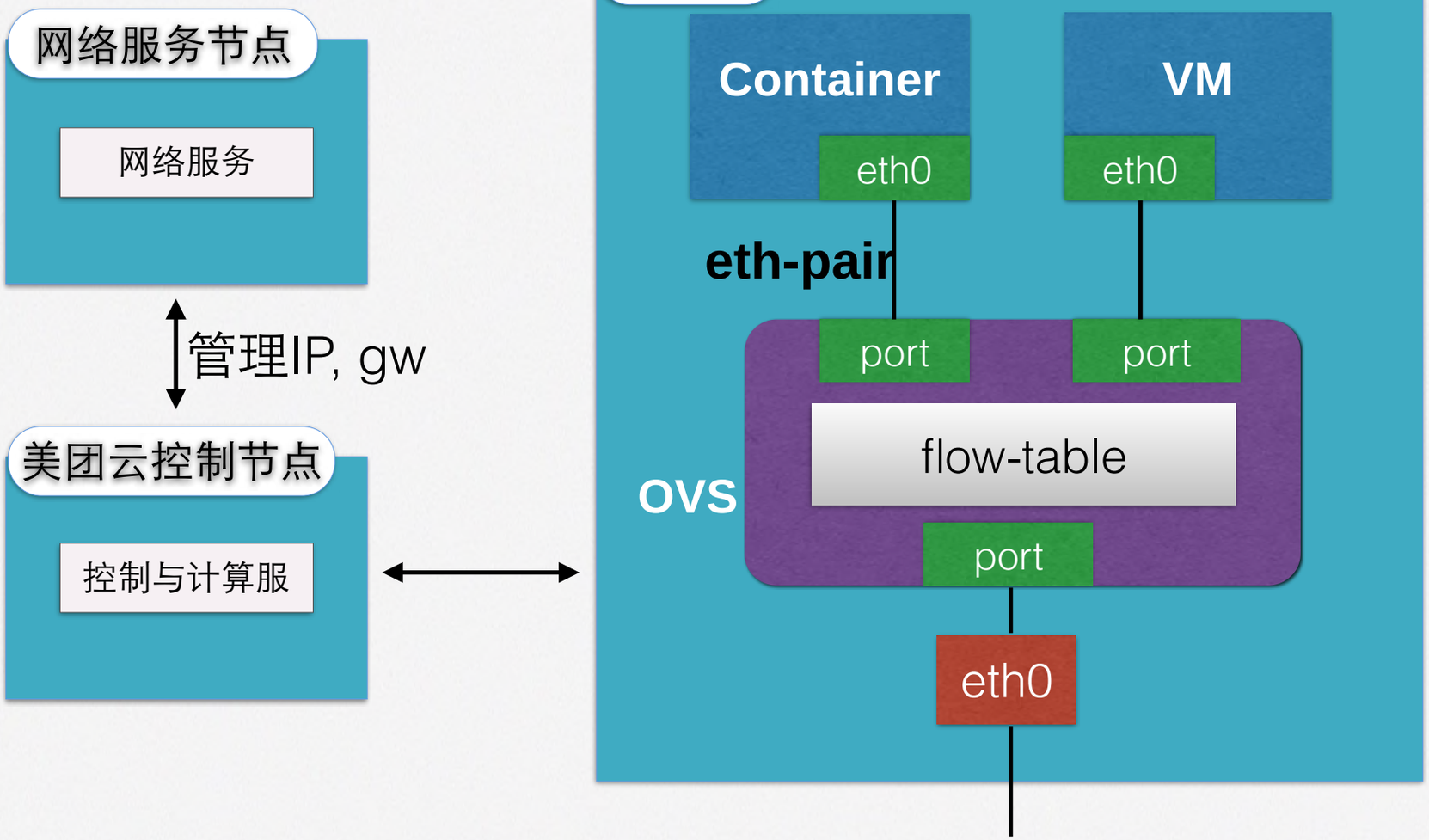
创建容器
(network_mode = none)

启动容器

创建eth-pair

分别接入容器NS和OVS

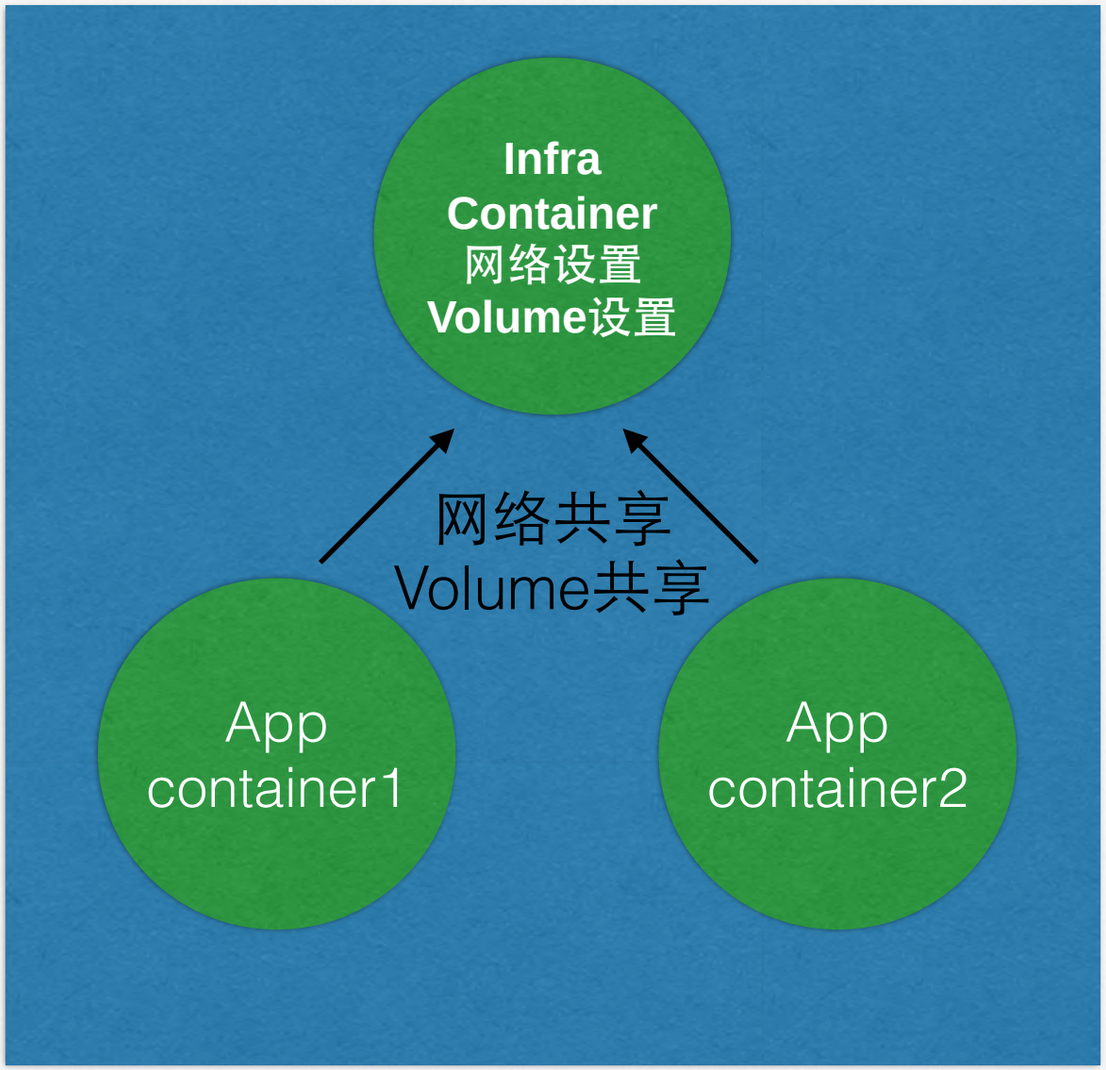
对容器的eth配置IP、路由， 配置OVS流表



美团云

set — 多容器管理单元

- set:
  - 1到多个容器组成
  - 调度的基本单位
  - 一个业务实例
  - 弹性伸缩的基本单位
  - 多容器资源共享
  - 原则上单容器单进程(but sshd)

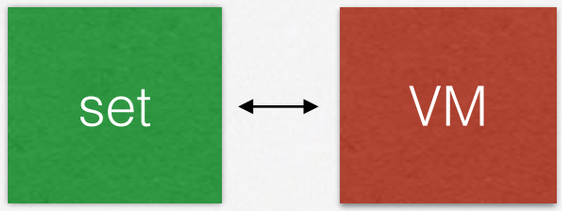


美团云

set — 多容器管理单元 Cont.

```
{
   "version": "v2",
   "id": 1,
   "appkey": "com.sankuai.inf.hulk.test",
   "containers": [
      {
         "index": 0,
         "image": “hulk.test-prod”,
         "options":  {
               "name": “test",
               "cpu": 80,
               "mem": 20,
               "volumes": [
                     {
                        "path": "/opt/logs",
                        "quota": 100
                     },
               "command": {
               "cmd": "/bin/bash",
               "args": ["-c", "run.sh"]
               }
      },
   ]
}
```

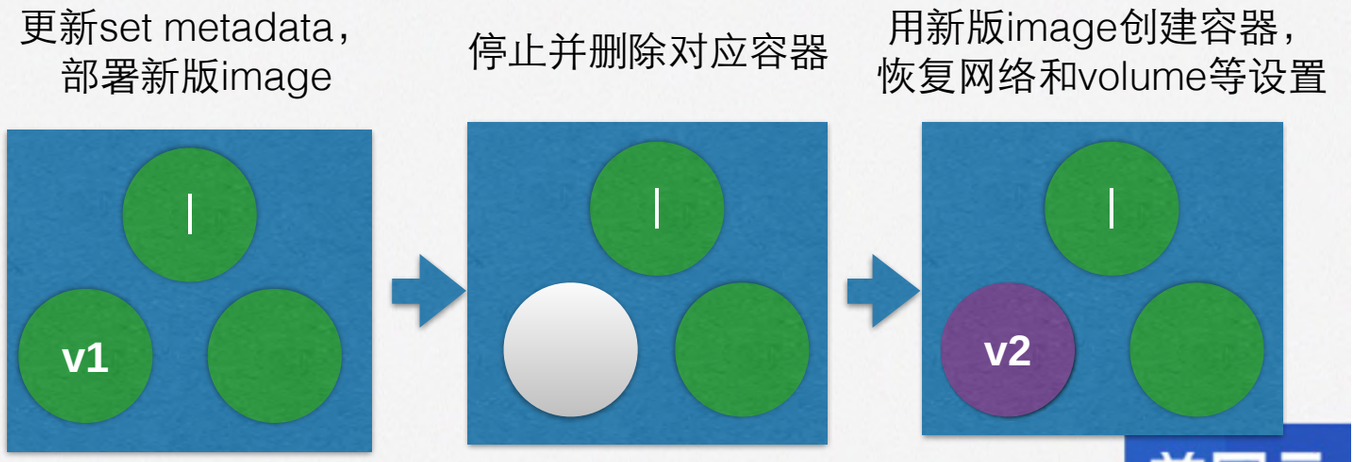
实现方式



支持特性

- 按指定顺序启动容器；
- infra-container，构建网络，设置volume，privileged权限
- 运行时灰度更新set内的container image
- 运行时ScaleUp
- 支持SSH登陆业务容器，密码设置

Example - 运行时灰度更新



美团云

推广实践

- 私有云：
  - ✦ 对接美团业务弹性伸缩服务
- 公有云：
  - ✦ 开始支持美团云RDS/Redis等PaaS业务

美团云

工作规划

2015 下半年

- 技术调研
- 系统设计、模块开发和联调测试
- 容器弹性扩容/缩容上线私有云

2016 上半年

- 私有云业务VM/Docker迁移
- 承载美团云PaaS业务
- 公有云容器服务架构与关键技术预研

2016 下半年

- 公有云容器服务系统设计、模块开发

美团云

Thank You！