

## 2015 Container 技术峰会

# 容器技术在SNG中间层的应用

funnychen ( 陈芳录 )

2015@OpenCloud

# 关于我

- funnynchen ( 陈芳录 )
- @腾讯社交网络运营部平台技术运营中心
- 擅长大规模组件运维、运维自动化建设

# 提纲

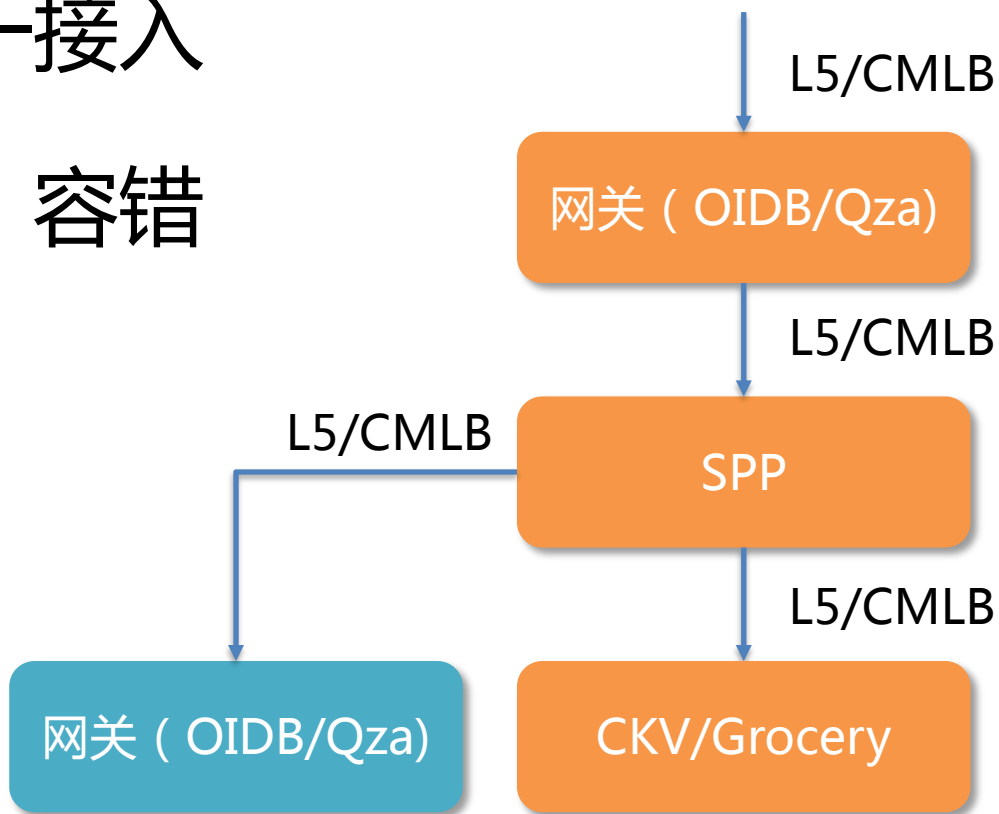
- 中间层介绍及其运维体系
- LXC虚拟化实践
- 蜂巢集群和Docker尝试

# 中间层在社交架构中的位置



# 典型业务架构

- 网关：鉴权、统一接入
- 路由组件：寻址、容错



# 现状

20000+服务器      30000+服务实例

1000+业务      三地部署



QQ群  
QUN.QQ.COM



QQ空间

广点通  
社交效果营销平台



QQ音乐



QQ Show  
SHOW.QQ.COM



QQ相册  
PHOTO.QQ.COM

# 中间层运维体系

虚拟化

容量管理

性能监控

组件框架  
版本管理

成本

告警预处理

集群管理

标准模块管理

效率

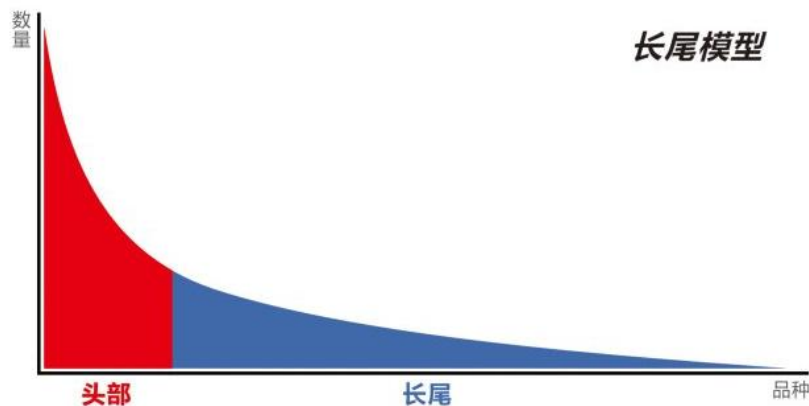
业务监控

组件监控

质量

# 引入虚拟化技术

- 提供细粒度资源单位，解决长尾问题
- 统一机型，提升资源流转效率





# 虚拟化选型

- 成本最低：海量服务对性能的要求
- 尽量不增加监控、运维成本
- 业务运行环境可控：自研业务

	实体机	XEN	LXC
网络	-	veth + bridge	veth + bridge
资源隔离	完全隔离	强	弱
性能损耗	无	有	极小
镜像与迁移	无	有	无

# 隔离方案

- 资源隔离
  - namespace ( 隔离uts,ipc,mnt,pid,user,net )
  - cgroup ( 限制进程组资源 )
    - 子系统：  
cpuset,cpu,cpuacct,memory,devices,freezer,net\_cls,blkio,tstat

# CPU和内存的隔离细节

- CPU：通过cpuset给子机指定CPU
- 内存：
  - 根据子机CPU核心数比例等额分配
  - 禁用swap：`memory.limit_in_bytes = memory.memsw.limit_in_bytes`



# IO限速

检测avgqu-sz

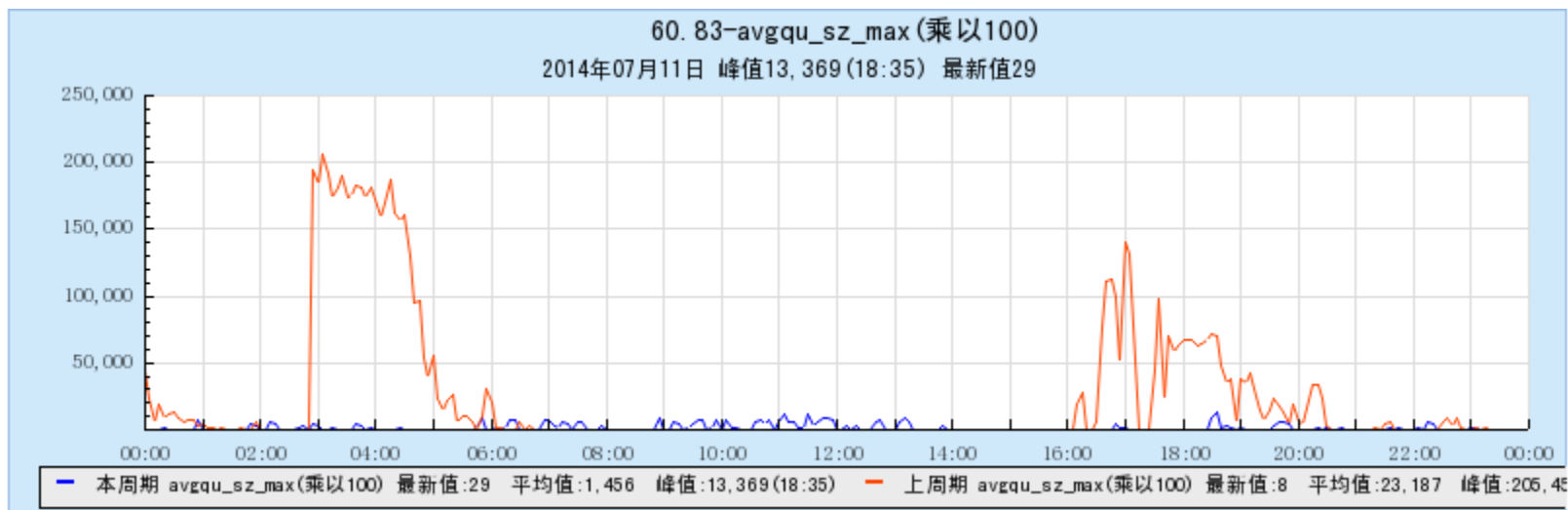
检测子机  
iops、r/w bps

动态调整  
blkio.throttle.\*

检测avgqu-sz

连续N次空闲

恢复



# 磁盘方案

LVM

Logical Volume  
vg-container\_100

Logical Volume  
vg-container\_101

Volume Group  
data

Physical Volume  
/dev/sda

Physical Volume  
/dev/sdb

Overlayfs

子机 A  
私有目录

upper dir

基础环境

lower dir

lower dir

子机 B  
私有目录

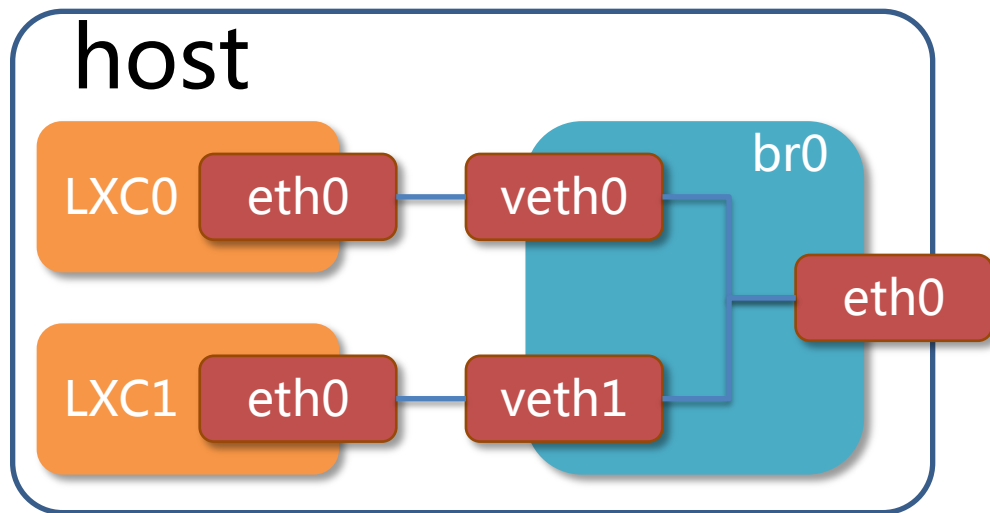
upper dir

子机A

子机B

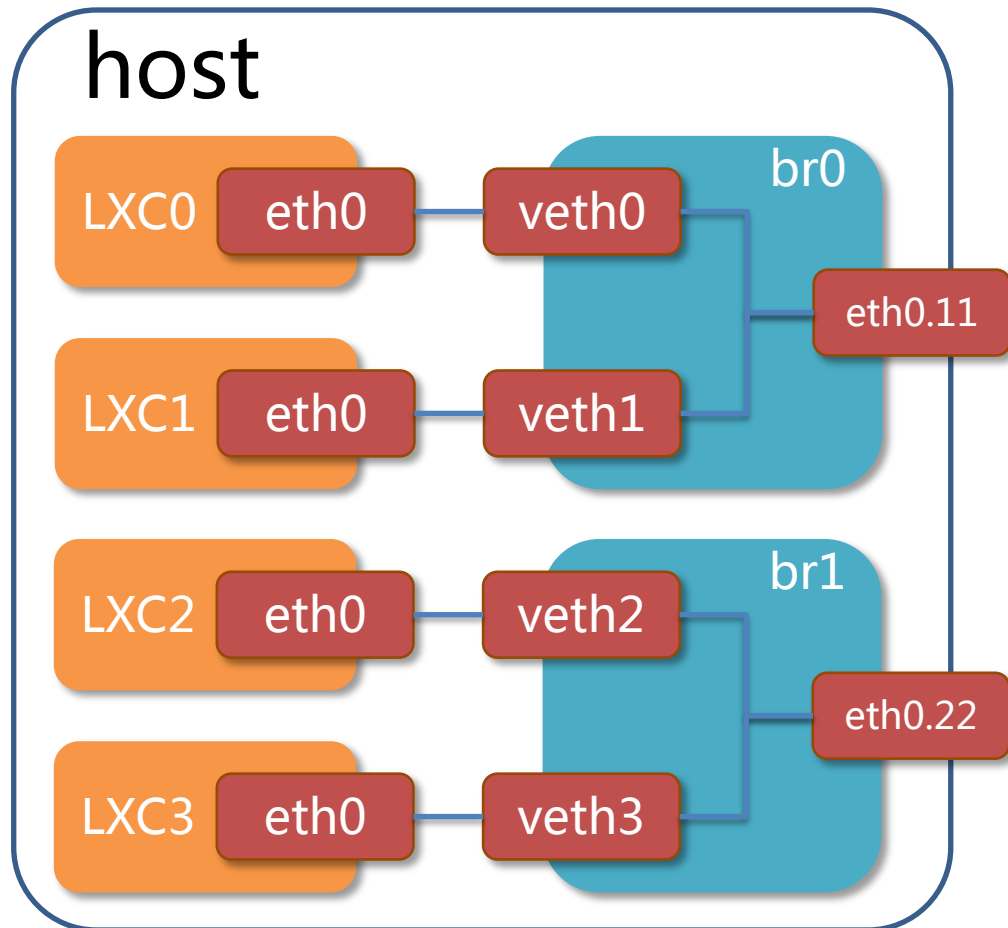
# 网络方案

- veth+bridge
  - 可直接ssh登录
  - 对业务程序透明
  - 兼容网管监控
- net\_cls + tc ( htb )
  - 保证子机最小带宽



# 网络方案升级

- VLAN太小！



# 内核增强

- /proc 虚拟化
  - 使子机只看到自己的CPU和内存情况，兼容网管监控
- 移植Task counter
  - 限制子机进程数量

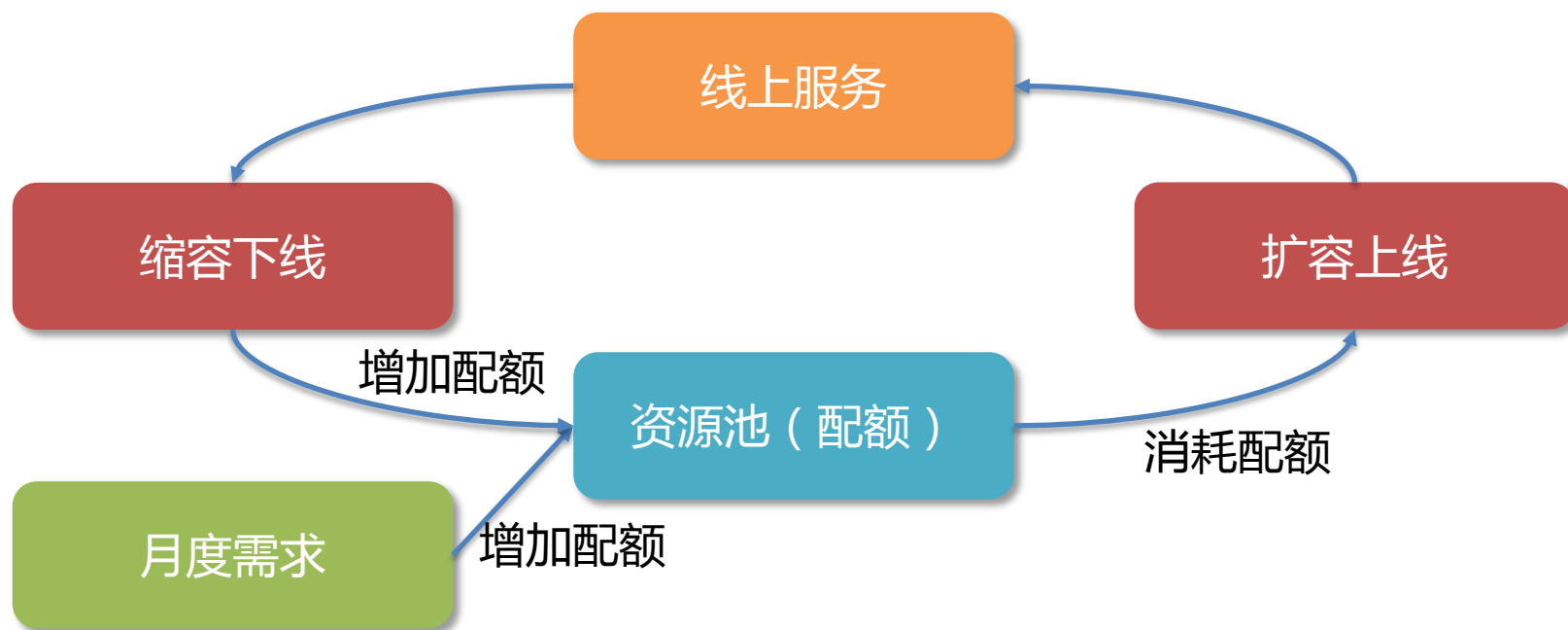


# 生产管理

- 固定机型：2核/4核/8核
- 独占CPU：子机间无CPU竞争
- 打散：同个业务的子机分散到不同母机
- CPU动态调整：母机内CPU使用率均衡
- veth / VG 命名规范化，易于辨认缩绑定子机

# 配额制度

- 配额关联业务低负载，推动持续成本优化



# 织云

- 内部云管理平台
- 分享：
  - [《梁定安：解密腾讯SNG云运维平台“织云”》](#)
  - [《腾讯SNG织云自动化运维体系》](#)

# 织云虚拟机自动扩容流程

[N][会员增值业务] > [特权与产品项目] > [游戏中心] > [手Q游戏中心礼包系统逻辑][逻辑SPP] CAE上云中

共计 16 个资源 其中 7 个资源不一致 [处理一致性](#) [进入旧版资源管理](#)

现网机器存在配置中没有的包: **youtu\_crosspython2.6**, 可以通过 [现网扫描包](#) 录入

**包 9** [+ 添加](#)

**业务包**

- ☒ I5\_protocol\_3.2os
- ☒ GameCenter\_GameGiftServer

**基础包**

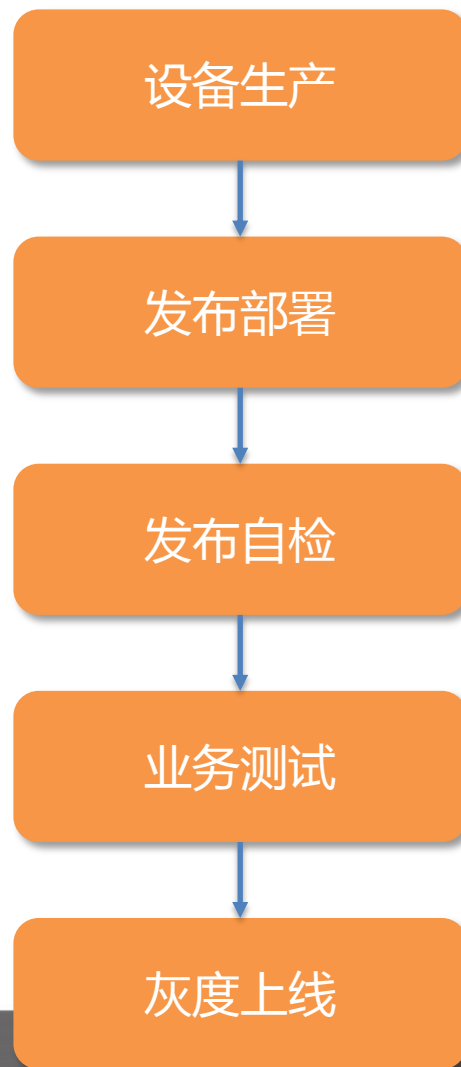
- ☒ clpAccess
- ☒ clear\_disk
- ☒ CloudDCAgent\_L5
- ☒ core\_check
- ☒ uniq\_client
- ☒ vidc\_static\_agent

**配置 6** [+ 添加](#) [导入CC下发配置](#)

- ☒ gamecenter\_log\_conf.ini
- ☒ gamegift\_service.ini
- ☒ gamecenter\_service.ini
- ☒ gamecenter\_bill\_conf.ini
- ☒ gamecenter\_handle.xml

[全部任务] 执行中任务 异常任务 已完成任务 [刷新](#)

流程名称	操作人	启动时间	状态
[全量][cae]变更流程		2014-09-26 17:27:19	1 2 3 4 5 6 7 8 9
[全量][cae]变更流程		2014-09-26 17:27:18	1 2 3 4 5 6 7 8 9
[全量][cae]变更流程		2014-09-26 17:24:20	1 2 3 4 5 6 7 8 9
[正式]部署一致性Agent		2014-09-26 17:20:17	1 2 3 4 5 6
[cae]容量调度-自动扩容		2014-09-20 12:39:23	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
[cae]容量调度-自动扩容		2014-09-20 12:22:14	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
[cae]容量调度-自动扩容		2014-09-19 20:42:06	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
[cae]容量调度-自动扩容		2014-09-13 12:15:42	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
[全量][cae]变更流程		2014-07-15 09:41:37	1 2 3 4 5 6 7 8 9
[web+logic][cae]缩容下线		2014-05-05 09:48:50	1 2 3 4 5 6 7 8
[逻辑层][cae]扩容		2014-05-04 14:45:24	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

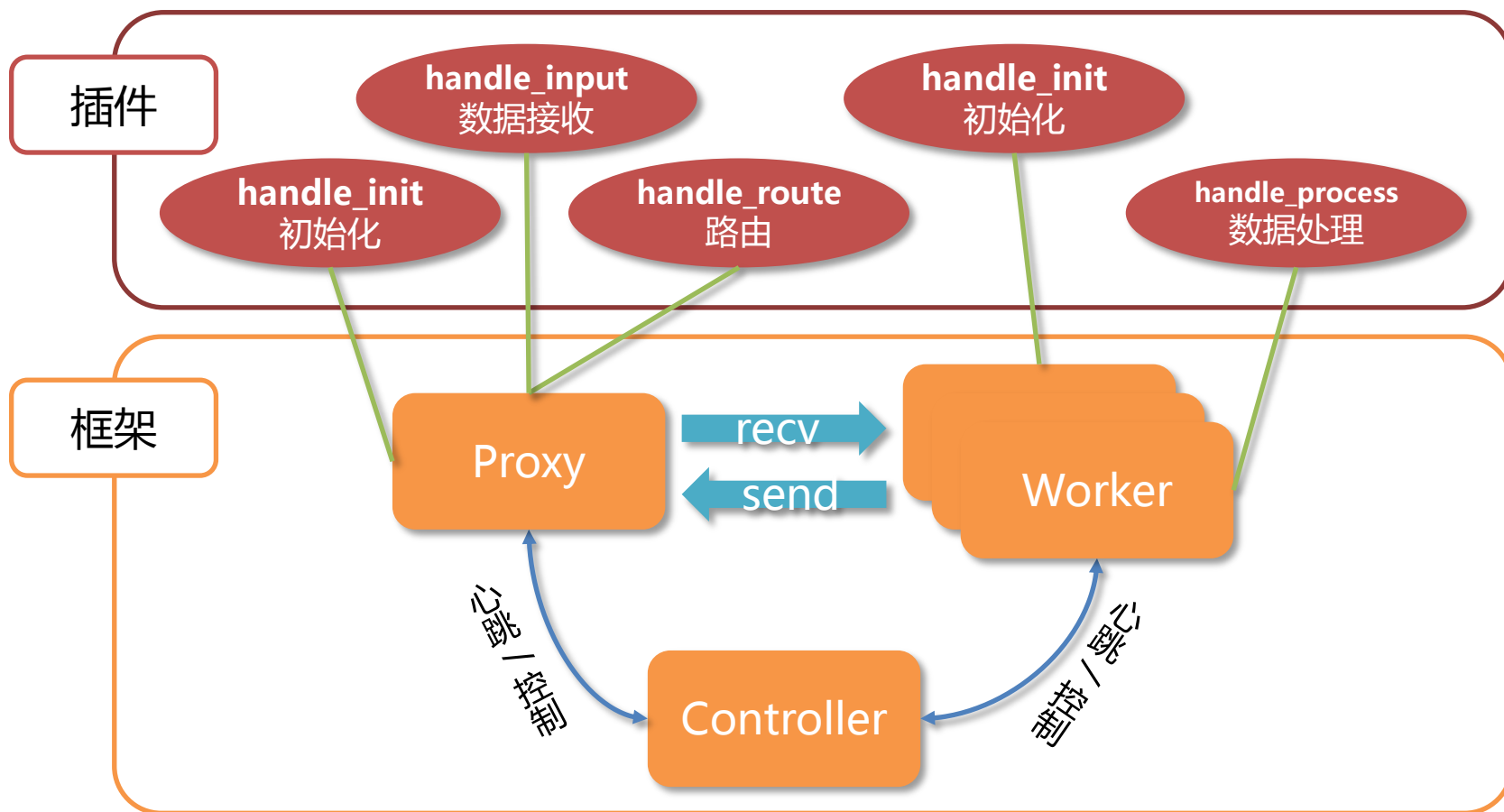


# 效率的痛

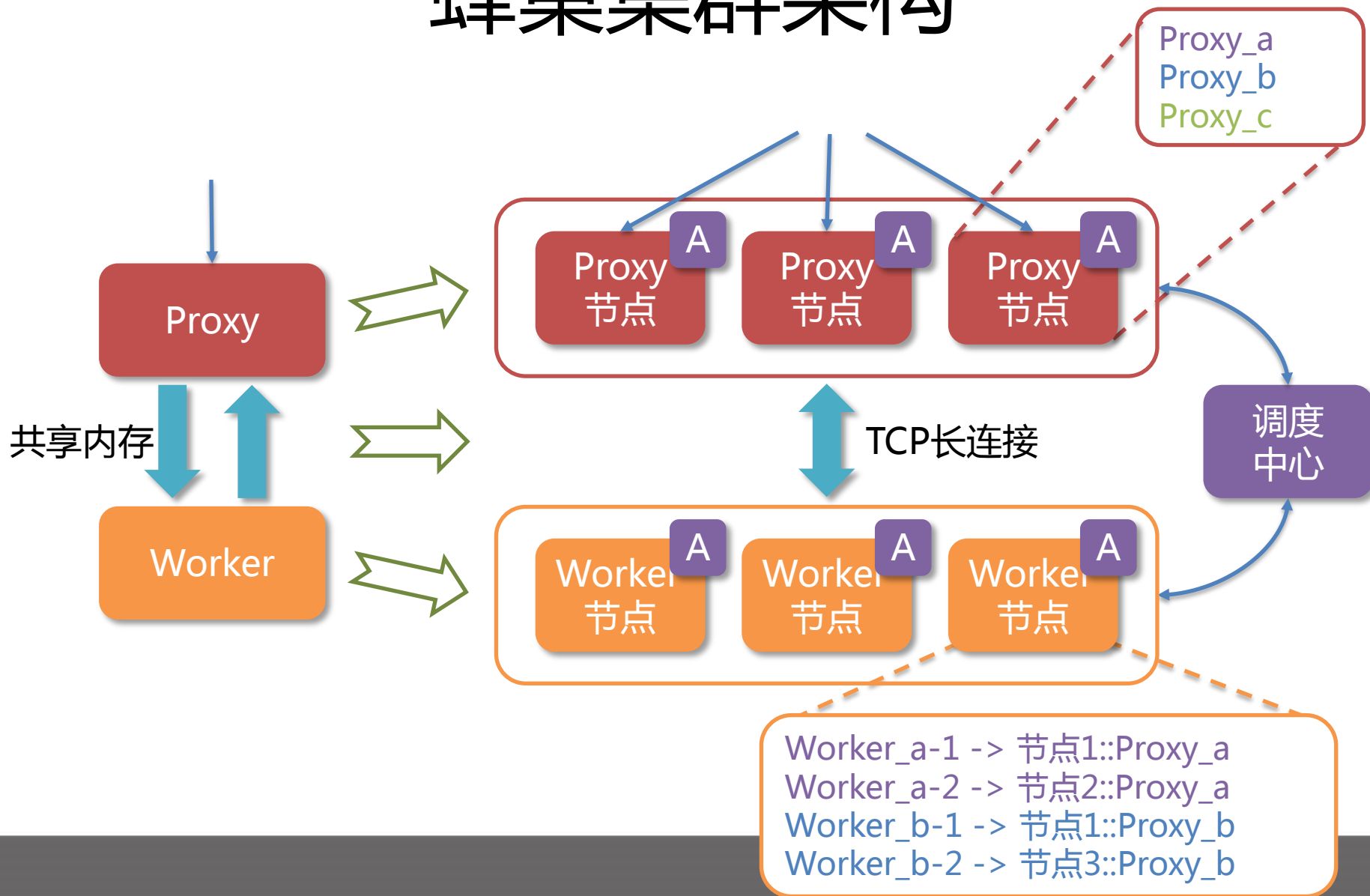
- 人均近万台设备、数百个模块
- 高强度扩容：春节、元宵、清明、五一.....



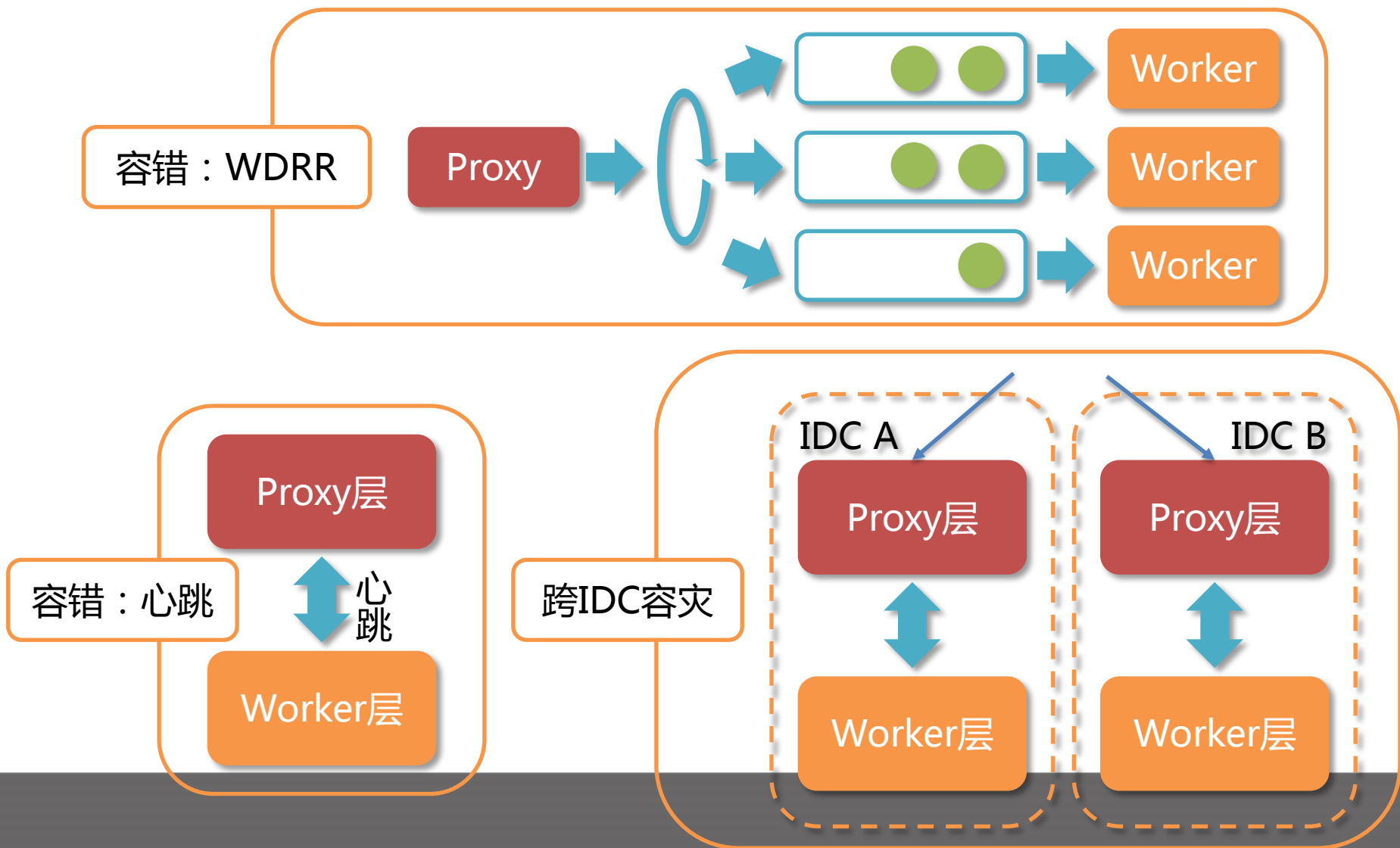
# SPP组件架构



# 蜂巢集群架构



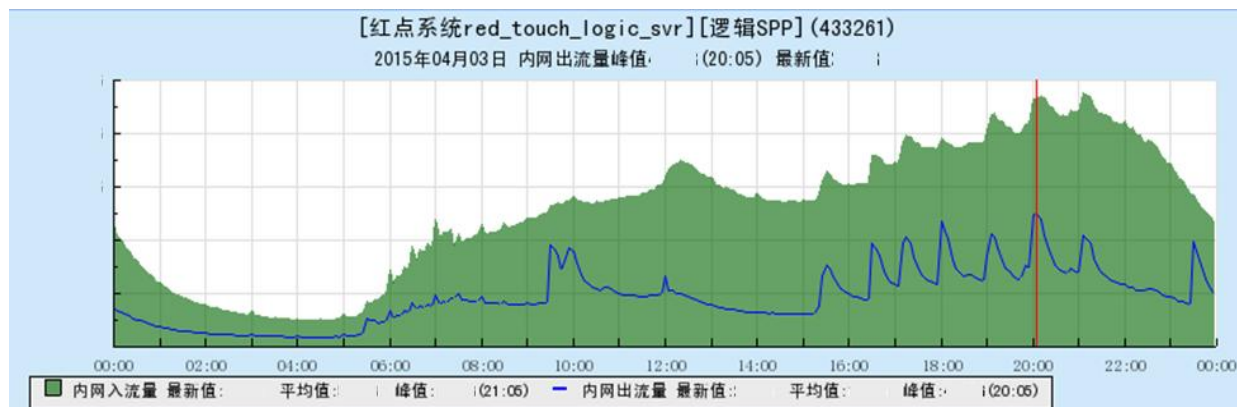
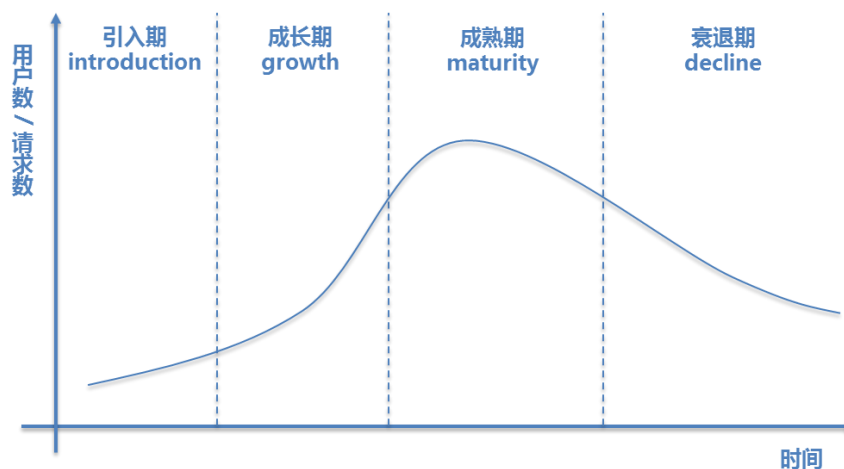
# 蜂巢高可用





# 典型场景

- 衰退期业务
- 业务活动

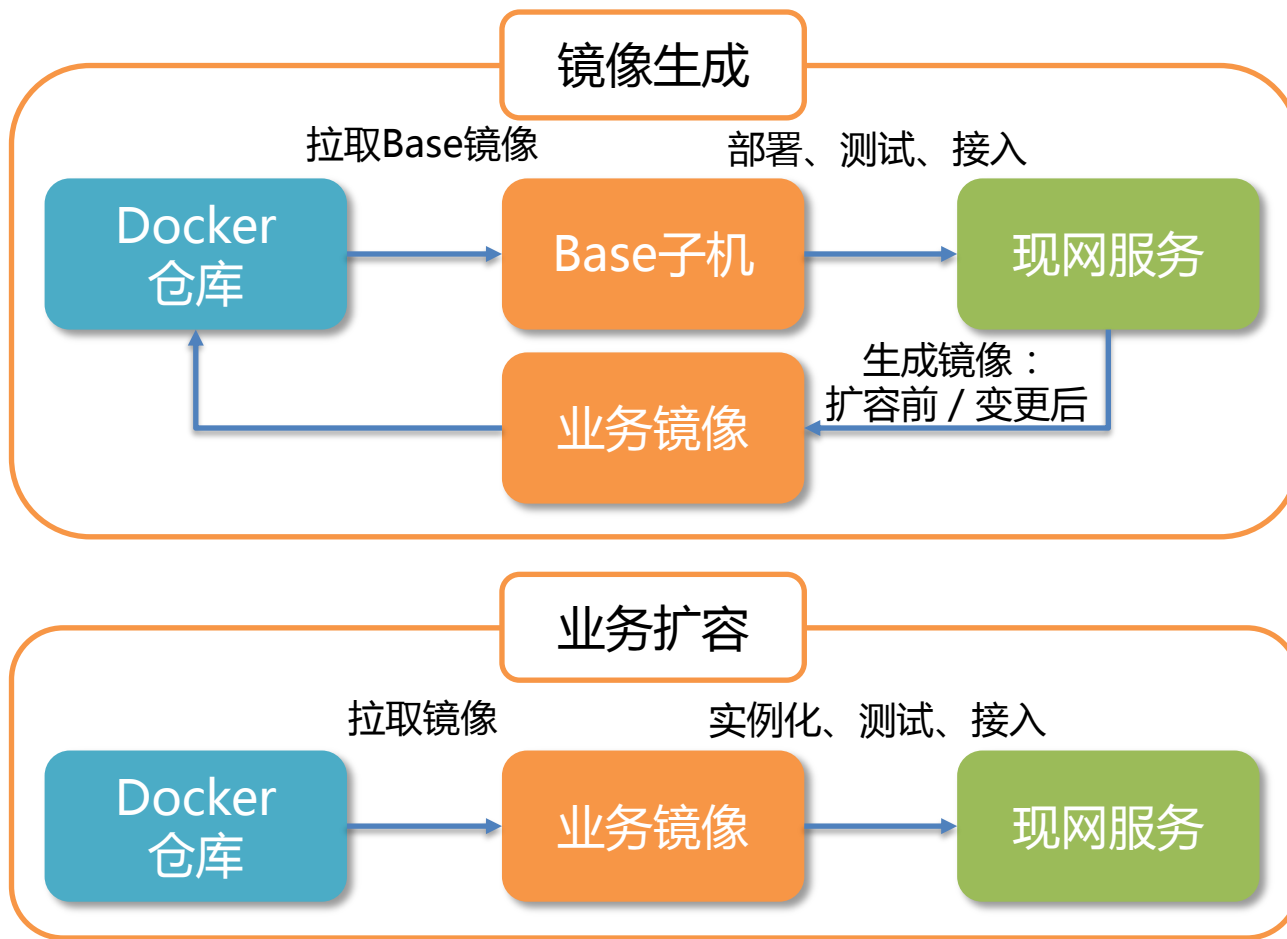


# 业务迁移绊脚石

- 包、配置、文件，如何关联管理？

The screenshot displays a cloud management console interface. On the left is a sidebar with navigation options: 设备列表, 资源管理 (selected), 包, 配置, 文件中心, 权限, 测试工具, 脚本, 流程, 模块访问, CronTab, 模调, 调度策略, and M指数 beta. The main content area shows a breadcrumb path: [N][会员增值业务] > [特权与产品项目] > [游戏中心] > [手Q游戏中心礼包系统逻辑][逻辑SPP] CAE上云中. Below this, it states '共计 16 个资源' with a warning that 7 resources are inconsistent and a button to '处理一致性'. A yellow banner indicates that the current machine has packages not in the configuration, specifically 'youtu\_cross:python2.6', and provides a '现网扫描包' button. The '包' (Packages) section shows 9 packages, categorized into '业务包' (Business Packages) and '基础包' (Basic Packages). Business packages include 'I5\_protocol\_32os' and 'GameCenter\_GameGiftServer'. Basic packages include 'clpAccess', 'core\_check', 'clear\_disk', 'uniq\_client', 'CloudDCAgent\_L5', and 'vidc\_static\_agent'. The '配置' (Configurations) section shows 6 configurations: 'gamecenter\_log\_conf.ini', 'gamecenter\_bill\_conf.ini', 'gamegift\_service.ini', 'gamecenter\_handle.xml', 'gamecenter\_service.ini', and 'gamecenter\_service.ini'. The '文件中心' (File Center) section shows 1 file, '/data/release/'.

# 引入 Docker



# 未来挑战：中间层遇见Docker

- Carrier：Docker平台
- 全路径Docker化
  - 开发 -> 测试 -> 预发布 -> 发布
- 蜂巢调度Docker实例
- 蜂巢Docker定制化
  - 不隔离pid、mnt、uts、user

# 回顾

- 中间层介绍及其运维体系
  - 万级设备和服务、千级业务
  - 成本、效率、质量
- LXC虚拟化实践
  - 隔离方案、磁盘方案、网络方案、内核增强、生产管理
- 蜂巢集群和Docker尝试
  - 集群架构、服务打包、Docker应用方案



2015 Container  
技术峰会

# THANKS



加入我们：

做一流的互联网运维团队，  
为用户创造优质的服务体验。

[funnychen@tencent.com](mailto:funnychen@tencent.com)