



# 携程容器实践

吴毅挺

2016.6.4

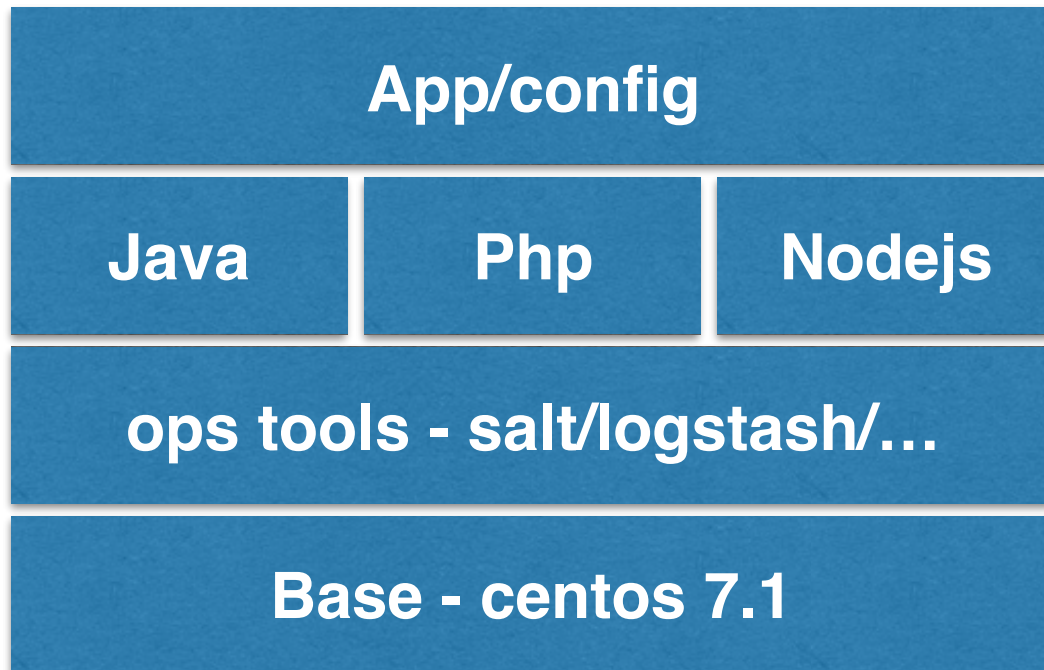
# Agenda

- 容器应用场景
  - 基于代码编译包的持续交付
  - 基于容器镜像的持续交付
- 容器监控
- Trouble shooting
- Q&A

# 容器应用场景

- 基于代码编译包的持续交付 - pets
  - Docker as VM
  - Docker & OpenStack
  - 测试环境管理
- 基于容器镜像的持续交付 - cattle
  - Job
  - website/service?

# Docker as VM



优势:

基于现有OpenStack扩展

兼容现有运维系统, eg 监控&发布

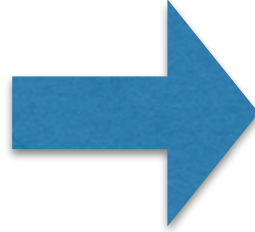
简化镜像维护

快速交付

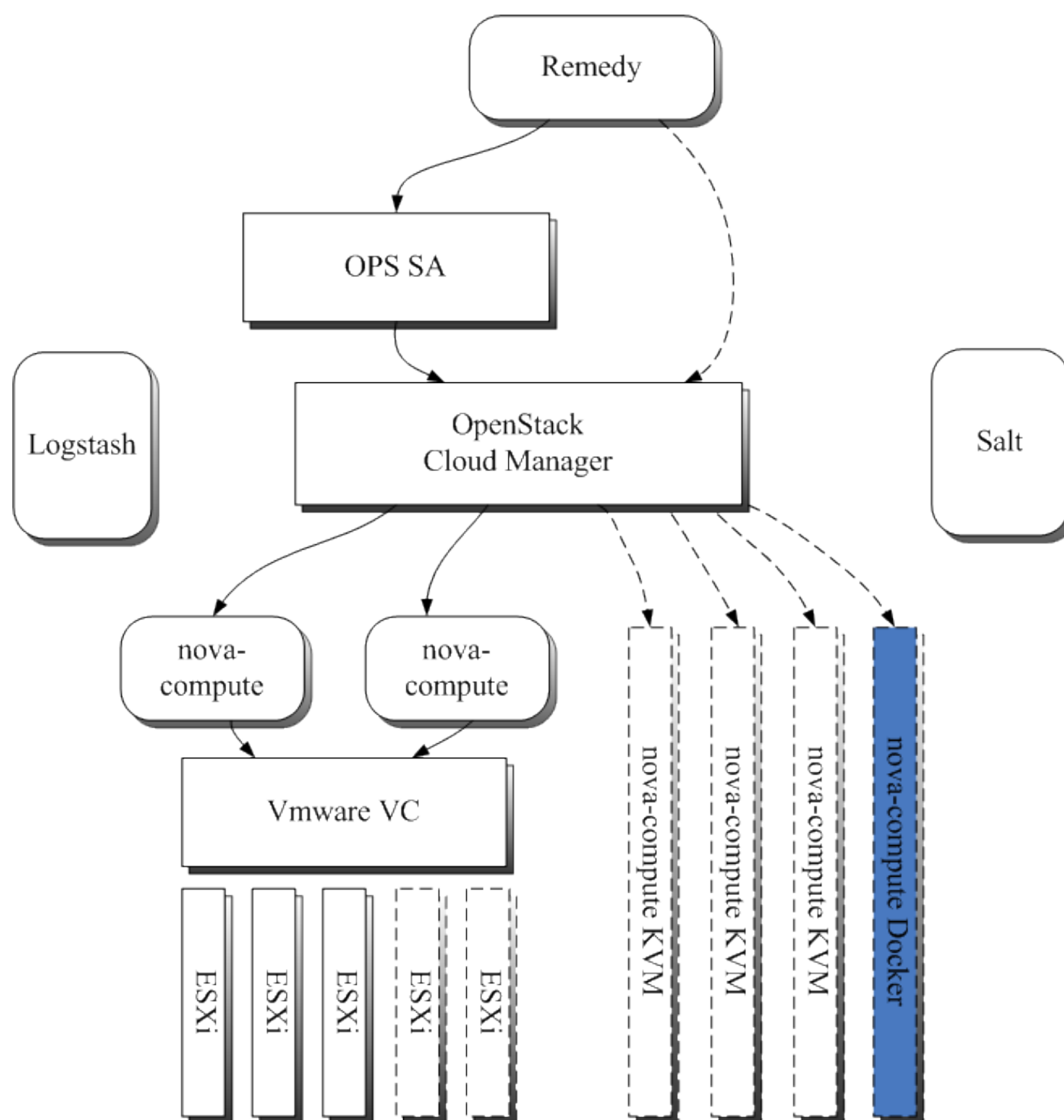
缺点:

镜像升级如何发布?

# 测试环境构建

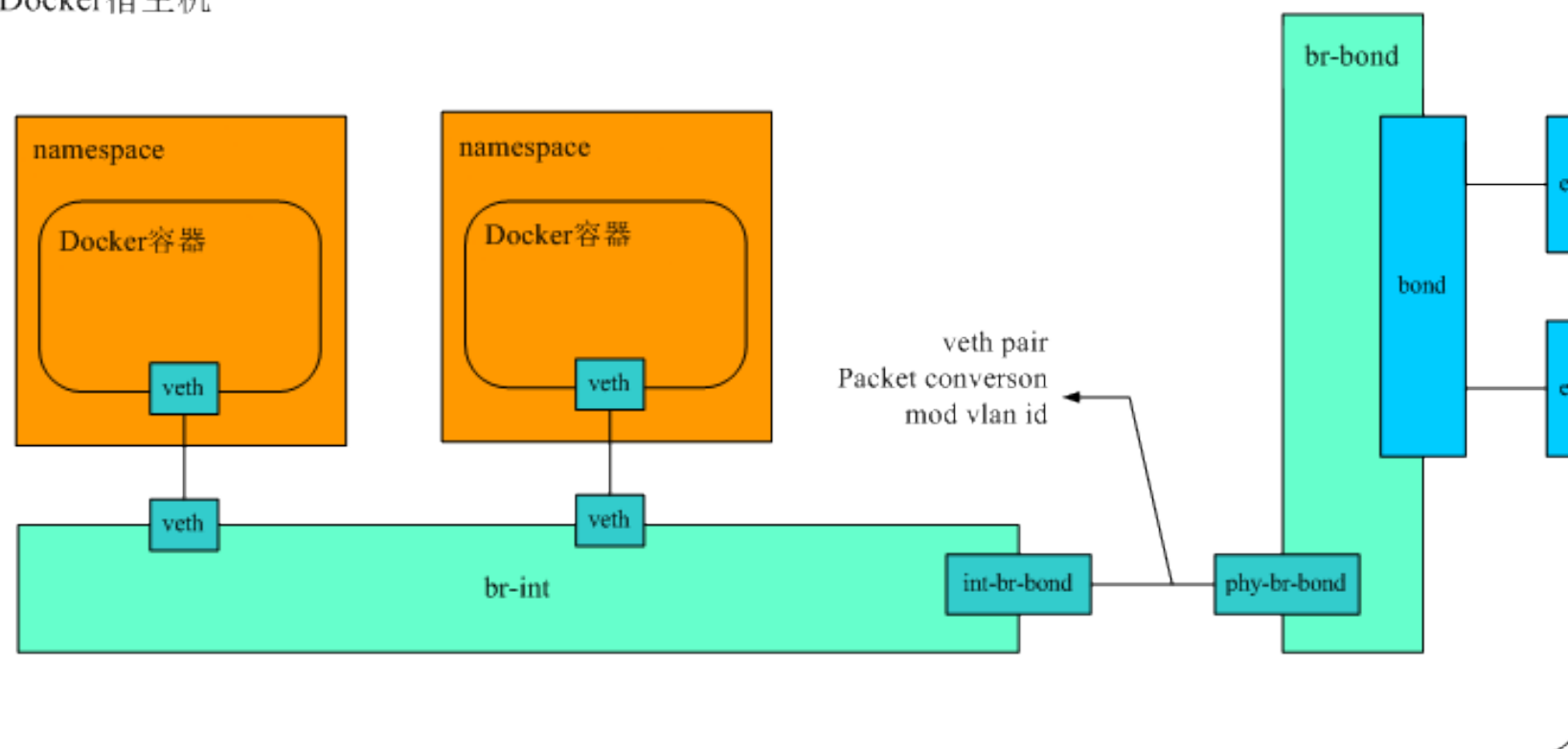
- 申请 vm
  - 关联 app -> vm
  - 部署App
  - 镜像维护代价大
- 
- 按需自动分配，无需申请
  - 单容器单应用，无需关联
  - 兼容现有发布模式
  - 功能测试环境 持续发布

# Docker & OpenStack集成



# OpenStack-Docker网络模型

Docker宿主机



ovs + vlan

容器独立ip

容器启动后再配置网络

docker缺乏网络信息

libnetwork

openstack on docker

# Job现状

- 部署不规范
- 手工维护crontab
- 环境配置不一致
- 运行历史情况
- 单点

```
[root@~]# crontab -l
##/1 * * * * /bin/bash /opt/app/bus_data_2.0/web/cron/shell/u mmend.sh
##/1 * * * * /bin/bash /opt/app/bus_data_2.0/web/cron/shell/u mmendByQueue.sh
##/1 * * * * /bin/bash /opt/app/bus_data_2.0/web/cron/shell/e pt.sh

*/1 * * * * /usr/bin/php /opt/app/t cron/n endSms.php
*/1 * * * * /usr/bin/php /opt/app/t cron/fe _logToDB.php
*/1 * * * * /usr/bin/php /opt/app/t cron/fe ToDB.php
*/1 * * * * /usr/bin/php /opt/app/t /cron/a php
*/1 * * * * /usr/bin/php /opt/app/t /cron/t elationTactics/b actics.php
*/1 * * * * /usr/bin/php /opt/app/t cron/ba php
*/1 * * * * /usr/bin/php /opt/app/t /cron/t elationTactics/b ics.php
*/1 * * * * /usr/bin/php /opt/app/t /cron/b onTactics/b ket_tactics.php

*/1 * * * * /usr/bin/php /opt/app/tieyou_clog/cJob.php

*/1 * * * * /usr/bin/php /opt/app/tieyou_bus_x/cron/log/t php
*/1 * * * * /usr/bin/php /opt/app/tieyou_bus_x/cron/log/t 
*/1 * * * * /usr/bin/php /opt/app/tieyou_bus_order/cron/ elTicket.php
*/60 0 * * * /usr/bin/php /opt/app/tieyou_bus_order/cron/ eOut.php
*/1 * * * * /usr/bin/php /opt/app/tieyou_bus_admin/cron/ ctics/build_bus_line_deductible_tactics.php
##/60 * * * * /usr/bin/php /opt/app/tieyou_bus_admin/ p
*/1 * * * * /usr/bin/php /opt/app/tieyou_bus_admin/cron/ php

*/30 * * * * /usr/bin/php /opt/app/tieyou_bus_data/cron/ dle.php

*/1 * * * * /usr/bin/php /opt/app/tieyou_bus_order/cron/ g...msResult.php
```



# Job持续交付目标

- 基于容器镜像发布
- 动态调度
- 容器/镜像等细节对用户透明
- 监控
- 配置管理
  - SLA
  - 版本控制

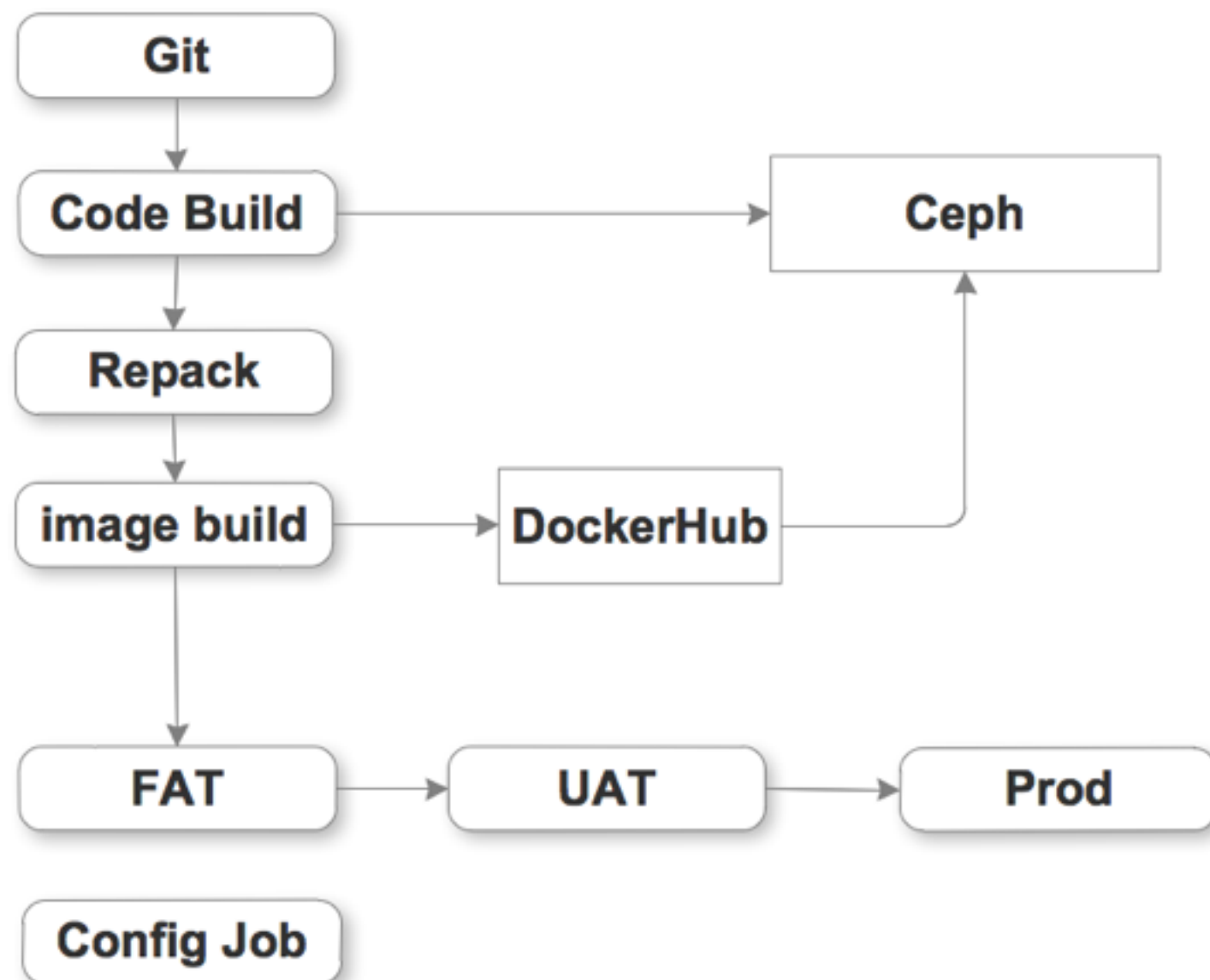
**仅仅使用Chronos/Mesos  
还不能满足需求**

# Job模型

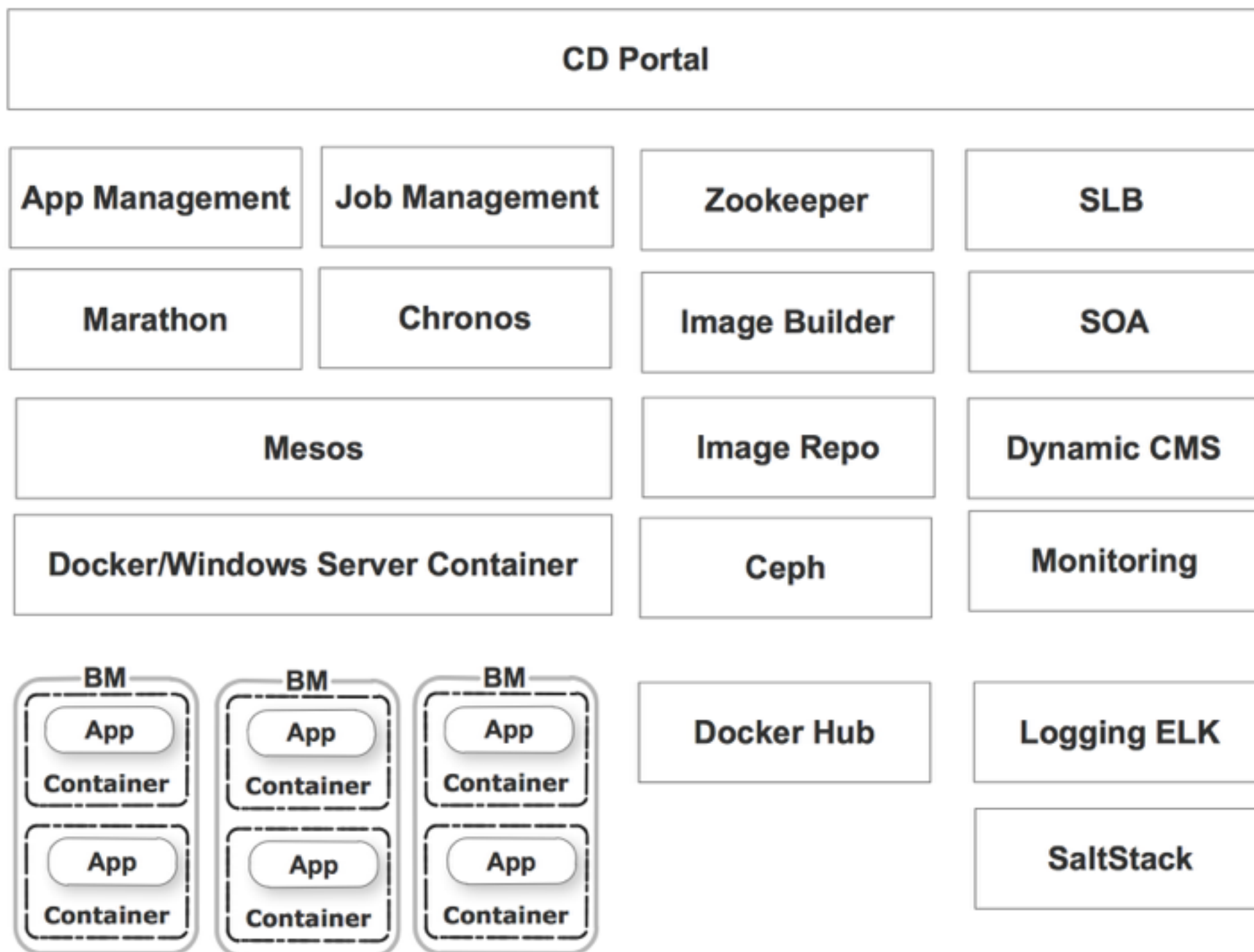
- 1 App -> N Job -> N task
- App 可发布单元/代码repo
- FAT/UAT/Prod Job配置一致性，除状态之外
- 新增Job 默认在所有环境都disable
- 同一环境所有的Job保持同一代码版本

```
App/  
├─ Job-A  
│   ├── task-20160529183806  
│   ├── task-20160529183807  
│   └─ task-20160529183808  
├─ Job-B  
│   ├── task-20160529183806  
│   └─ task-20160529183807  
└─ Job-C
```

# 交付流程



# 架构



• n

# Demo

← 首页/ 应用列表/ Job管理

is\_admin (100%)























FAT [45733] UAT [45734] PROD [未发布]

Branch\_release [45733]

发布

+ 创建Job

批量导入Jobs

Job名称[ID]	运行命令	计划时间 ⇅	最近一次运行	下一次运行	最近一次修改 ⇅	当前状态 ⇅	操作
tes	timeout 10 /bin/b	R/2016-05-26T14:25:00Z/PT2M	N/A	N/A	暂无修改记录	ENABLED	     
tes	timeout 10 /bin/b	R/2016-05-26T14:25:00Z/PT1M	N/A	N/A	暂无修改记录	ENABLED	     
tes	timeout 10 /bin/b	R/2016-05-26T14:25:00Z/PT1M	N/A	N/A	暂无修改记录	ENABLED	     
tes	timeout 10 /bin/b	R/2016-05-26T14:25:00Z/PT1M	N/A	N/A	暂无修改记录	ENABLED	     
tes	timeout 10 /bin/b	R/2016-05-26T14:25:00Z/PT1M	N/A	N/A	暂无修改记录	ENABLED	     

« 1 2 »

5 10 20

- Job Dashboard
- SLA monitoring

# 容器监控方式

- Host监控
  - docker stats
  - cAdvisor
  - zabbix-docker-monitoring
- 容器内监控
  - 内核修改: /proc
  - 修改工具: free/top/..
  - lxcfs:利用fuse提供修改过的/proc

**需要扩展当前的监控系统**

**升级、维护代价大**

# 容器内监控

- 通过LD\_PRELOAD劫持对/proc文件的访问
- 参考lxcfs重新计算/proc文件内容
- CPU Load 计算尚未支持

<https://github.com/CtripCloud/docker-monitor-injector>

```
[root@FATC1008 /]# free
              total        used        free      shared  buff/cache   available
Mem:          2148352      581012      691636      4680308      875704      691636
Swap:              0              0              0
[root@FATC1008 /]# unset LD_PRELOAD
[root@FATC1008 /]# free
              total        used        free      shared  buff/cache   available
Mem:          197881448      6580840      164832196      4680308      26468412      186118872
Swap:           8388604           32408           8356196
[root@FATC1008 /]#
```

# Trouble shooting

- systemd
  - docker run disable tty
  - yum version lock
- cpu 100%, iowait 高, Host 有空闲物理内存, 但容器使用大量swap
  - memory.limit\_in\_bytes
  - memory.memsw.limit\_in\_bytes
  - memory.swapiness
  - App 内存配置&监控, eg Java xmax



# Trouble shooting

- `systemctl error: too many open files`
  - `ulimit`
  - `fs.inotify.max_user_instances 8192`



# Q&A

- We are hiring!
- CD/PaaS Engineer
- [ytwu@ctrip.com](mailto:ytwu@ctrip.com)