

滴滴弹性云实践

谢丹博@滴滴出行

技术架构未来



About Me

- ✓ 先后就职于百度、小米、美团等公司
- ✓ 专注基础平台建设和运维自动化
- ✓ 滴滴弹性云项目负责人
- ✓ Open-Falcon社区核心成员

1

PART ONE

容器落地正当时

2

PART TWO

向左走向右左

3

PART THREE

全景展示

4

PART FOUR

总结&展望



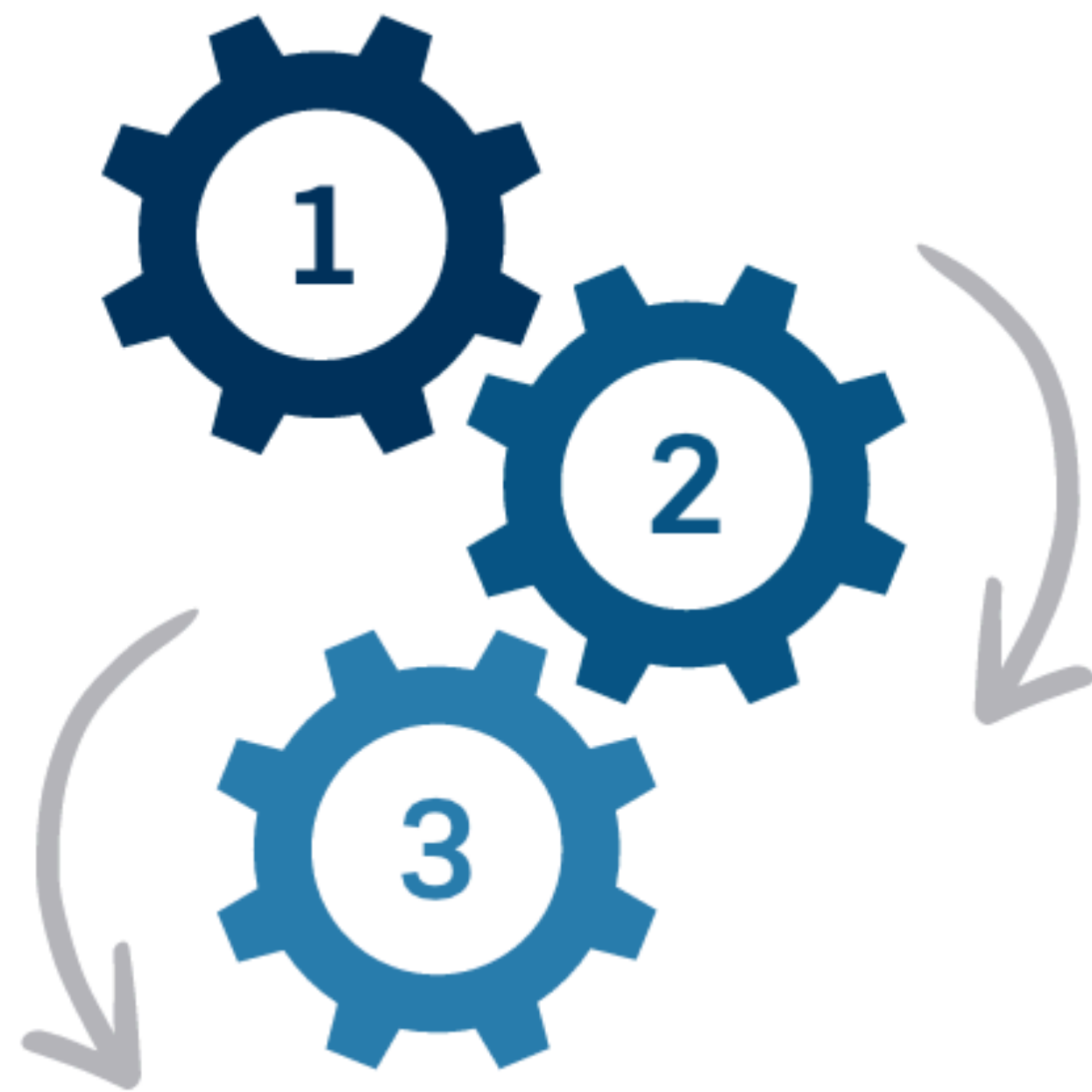
容器落地正当时

容器技术日趋成熟

容器技术正被越来越多的认可和接受

容器相关周边技术和实践日趋丰富

向左走向右走



变革

容器技术竞赛已经进入下半场，集群管理才是未来

价值

编排管理系统才是价值最大化的体现

选择

在mesos和Kubernetes之间该如何选择？



Why Kubernetes



沉淀

Kubernetes拥有google在大集群管理和微服务领域的沉淀



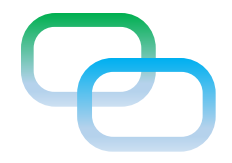
设计

Pod、RC、Service、Deployment、Label等优秀设计理念

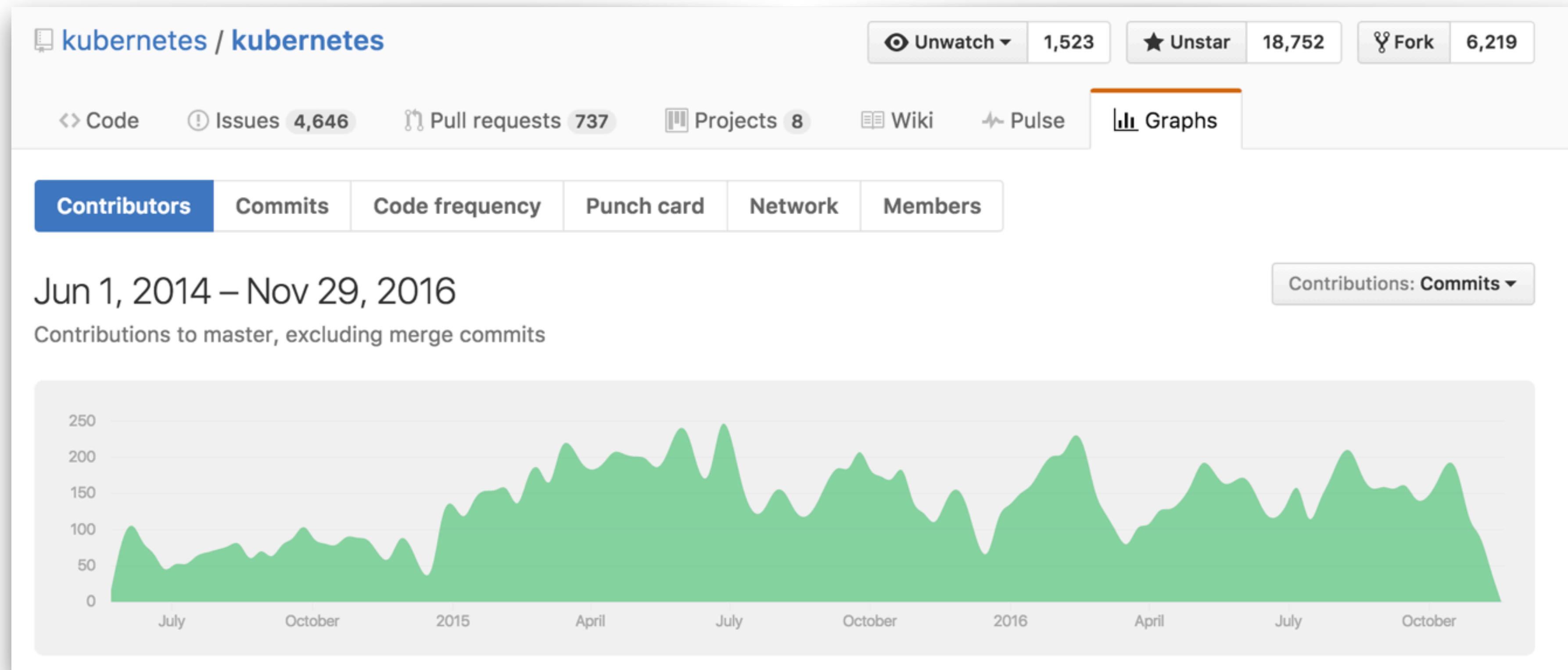


活跃

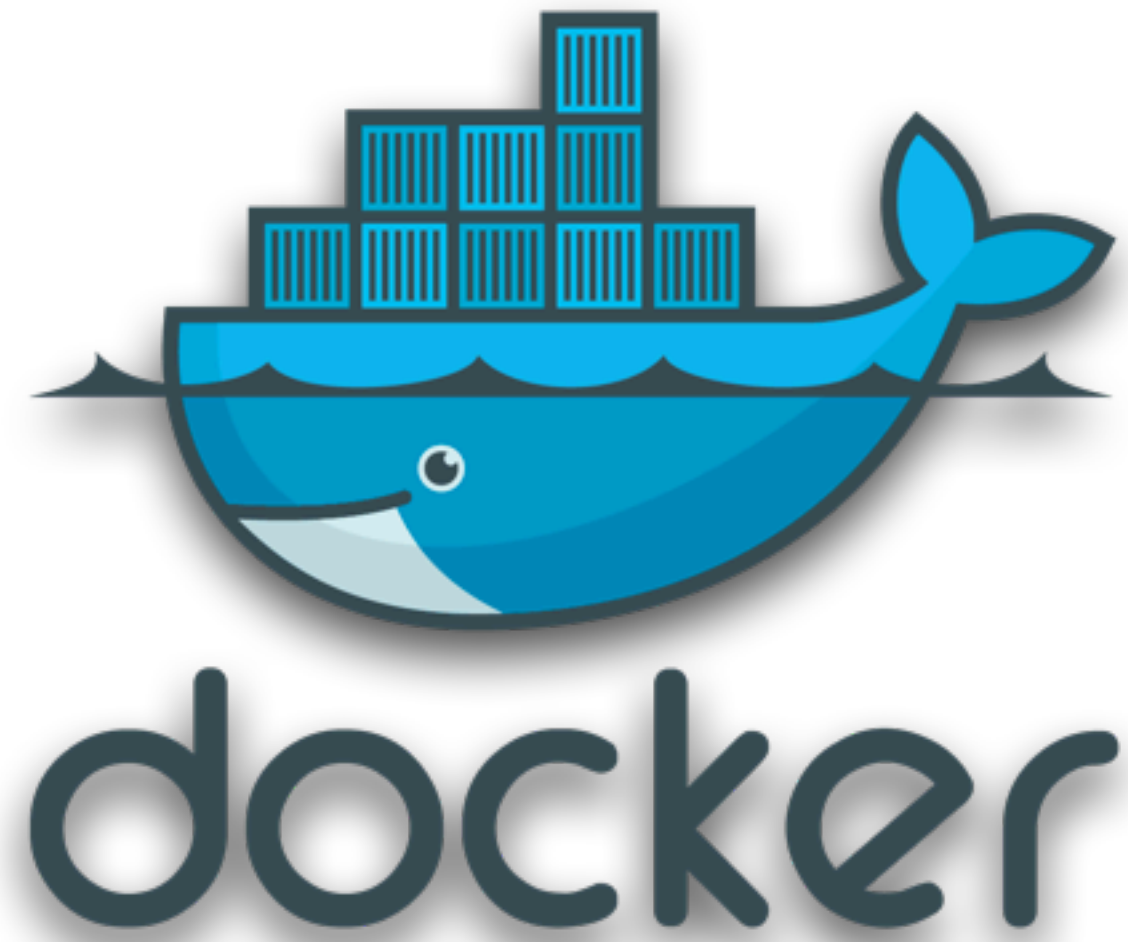
最活跃的开源社区和生态，版本迭代堪称神速



Why Kubernetes



项目概况





项目概况



项目历程

2016.7月启动，10月上线投入使用



硬件配置

2*E5-2670v3 | 16*16GB | 12*800G SSD | 10Gb网卡

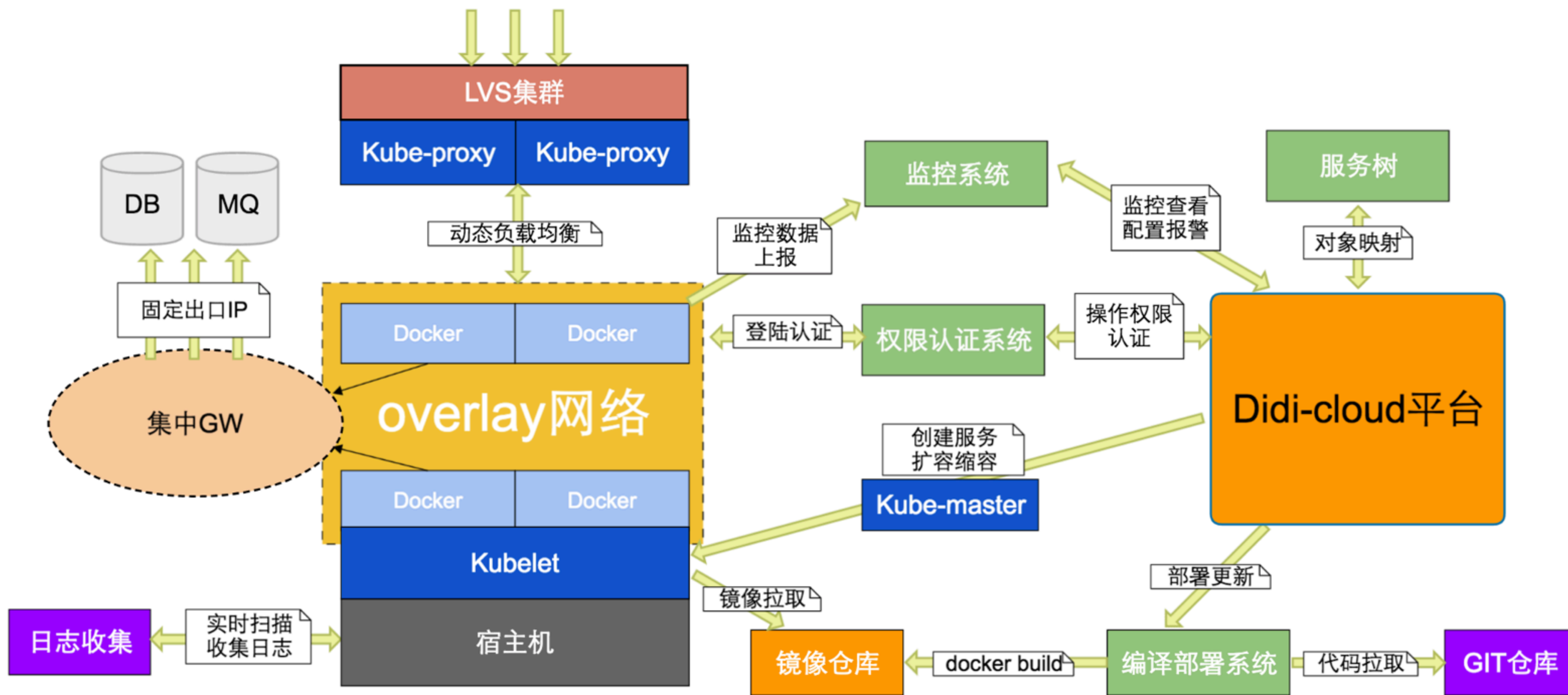


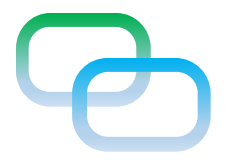
业务规模

宿主机100台，容器规模1000+，接入多条业务线

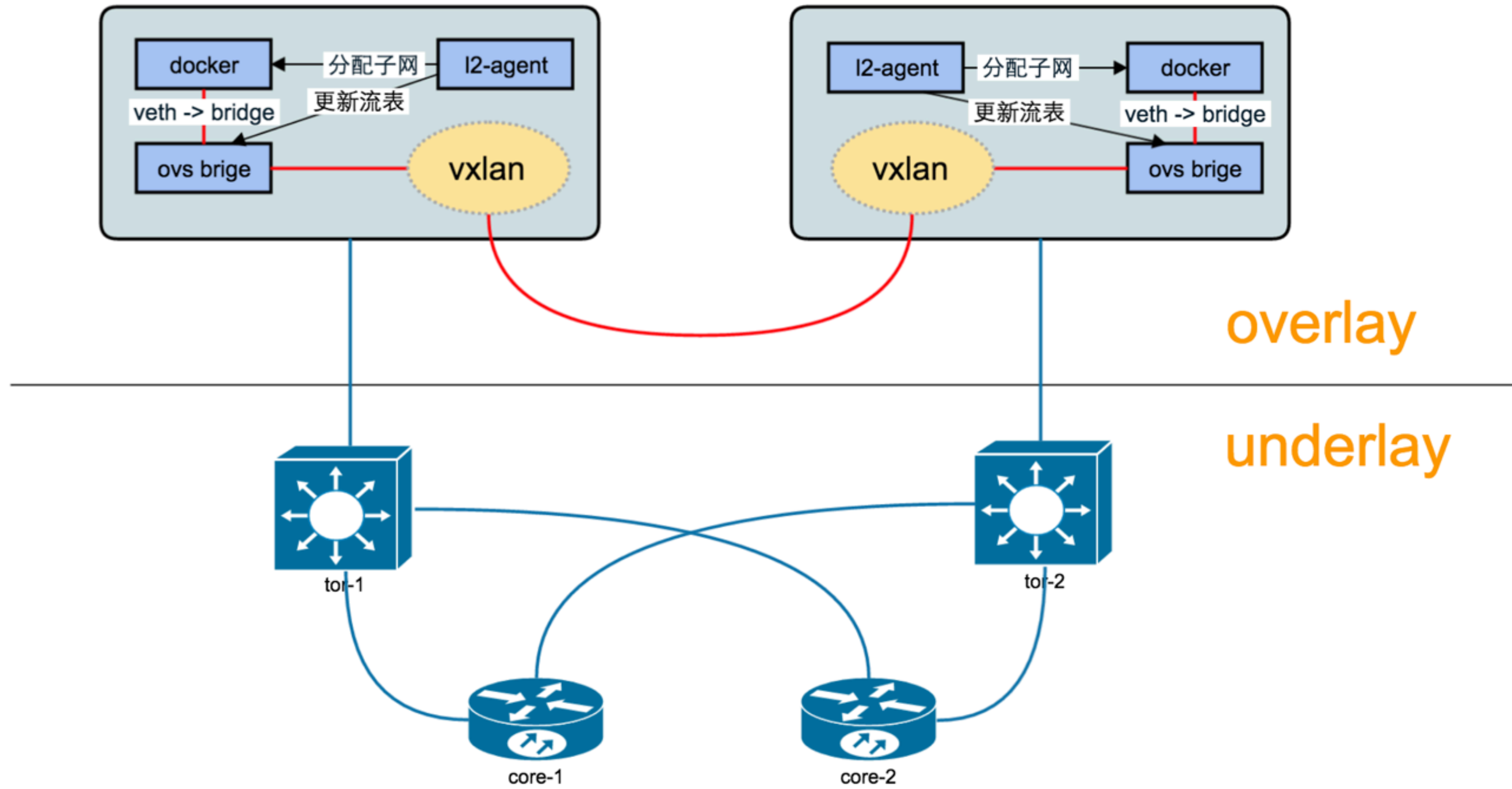


架构全景



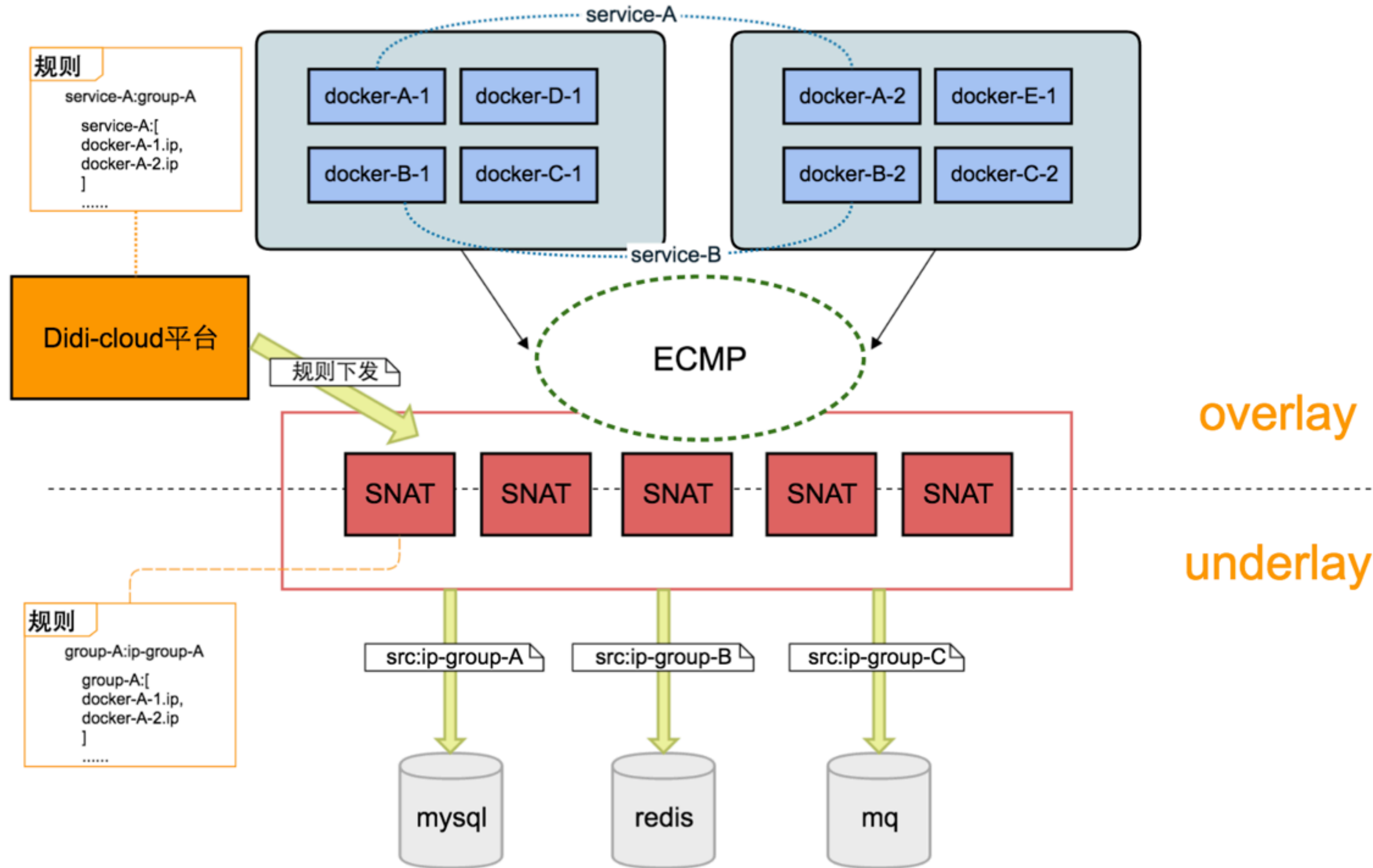


SDN网络

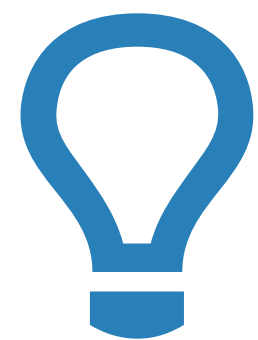




集中式GW



Kubernetes落地



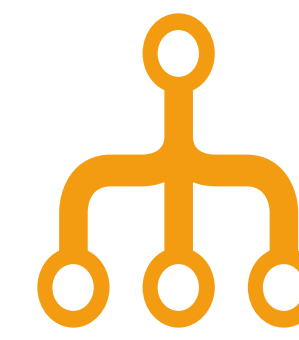
对用户透明



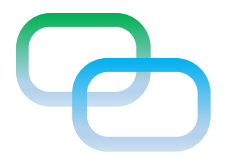
对接现有运维系统



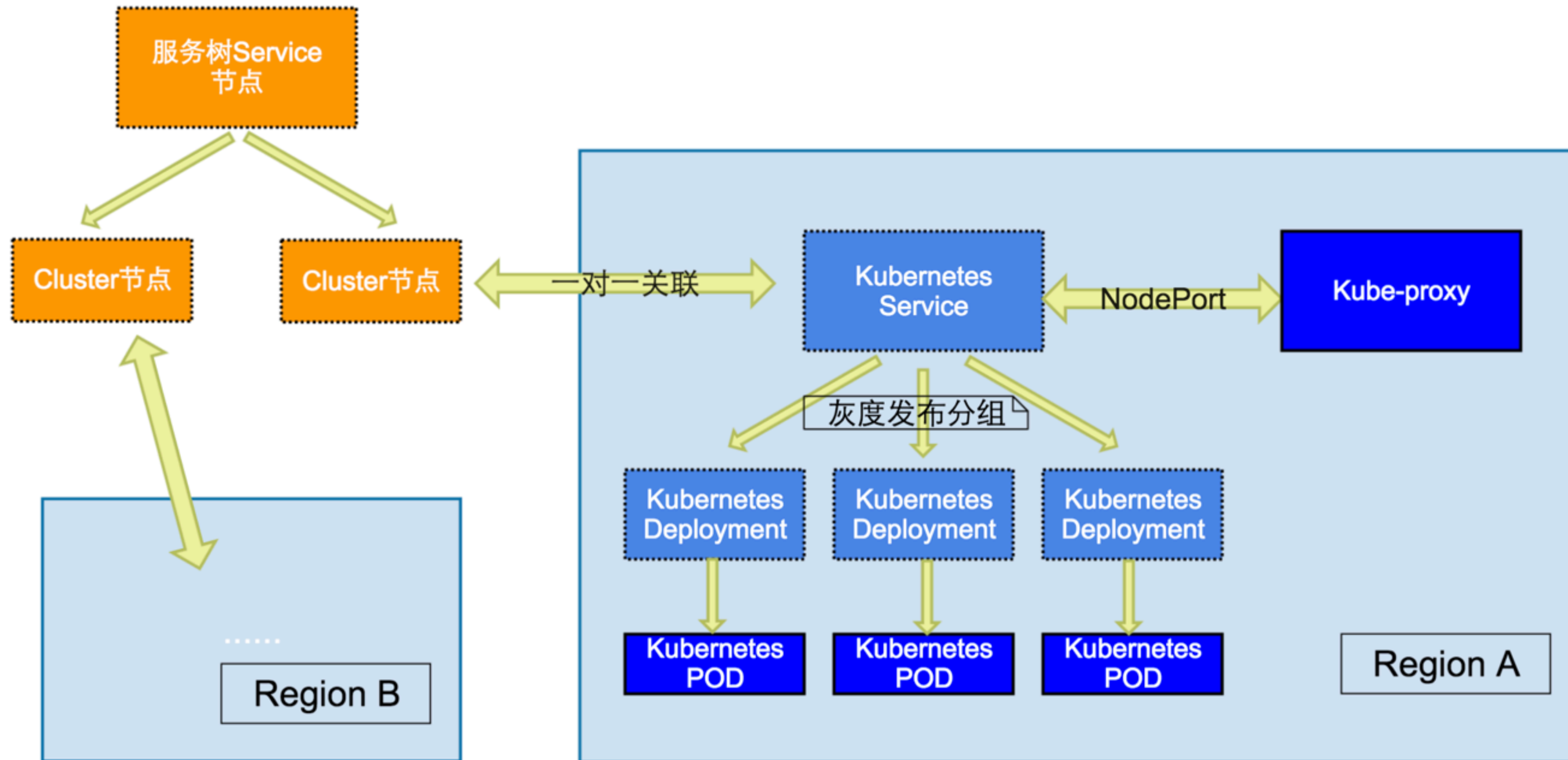
选择性开放功能参数



分**Region**部署架构



Kubernetes落地





集群管理

滴滴弹性云

服务管理 监控系统

搜索节点

gz01

新建部署 刷新

部署名称	当前状态	容器个数	创建人	创建时间	操作
test1	created	0	[REDACTED]	[REDACTED]	初始化 发布更新 弹性伸缩 更改配置
test2	pack	0	[REDACTED]	[REDACTED]	发布更新 弹性伸缩 更改配置



创建部署

mofang / 创建部署 (gz01)

创建部署

* 部署名称:

* 服务标签: 部门 - 产品 - 服务

* 容器规格: 普通型 - 2G

* 灰度发布 [?]: > (%) > (%) > (%)

* 服务端口 [?]: 容器服务端口, 目前一个容器仅支持一个服务端口, 如: 8080

启用全局端口 [?]:

服务可用探针 [?]:

服务存活探针 [?]:

[立刻创建](#)



Kubernetes实践

kube-scheduler异常bug

The screenshot shows a GitHub issue page for the Kubernetes project. The issue title is "`kubect delete node` make kube-scheduler enter failed state #31968". It is marked as "Closed" and was opened by user "workhardcc" on September 2nd. The issue has 7 comments. A comment from "workhardcc" on September 2nd describes a bug where deleting a node with pods running causes the kube-scheduler to enter a failed state. The comment includes a list of system logs showing the scheduler service failing and then restarting, followed by a panic error: "panic: runtime error: invalid memory address or nil pointer dereference [recovered]". The right sidebar shows the issue is categorized under "component/scheduler", "priority/P1", and "team/control-plane".

service 命名长度限制

```
4 pkg/api/validation/validation_test.go
@@ -3449,7 +3449,7 @@ func TestValidateService(t *testing.T) {
3449     {
3450         name: "too long name",
3451         tweakSvc: func(s *api.Service) {
3452 -             s.Name = strings.Repeat("a", 25)
3453         },
3454         numErrs: 1,
3455     },
@@ -3463,7 +3463,7 @@ func TestValidateService(t *testing.T) {
3463     {
3464         name: "too long generateName",
3465         tweakSvc: func(s *api.Service) {
3466 -             s.GenerateName = strings.Repeat("a", 25)
3467         },
3468         numErrs: 1,
3469     },

```

Allow service names up to 63 characters (RFC 1035) #29523

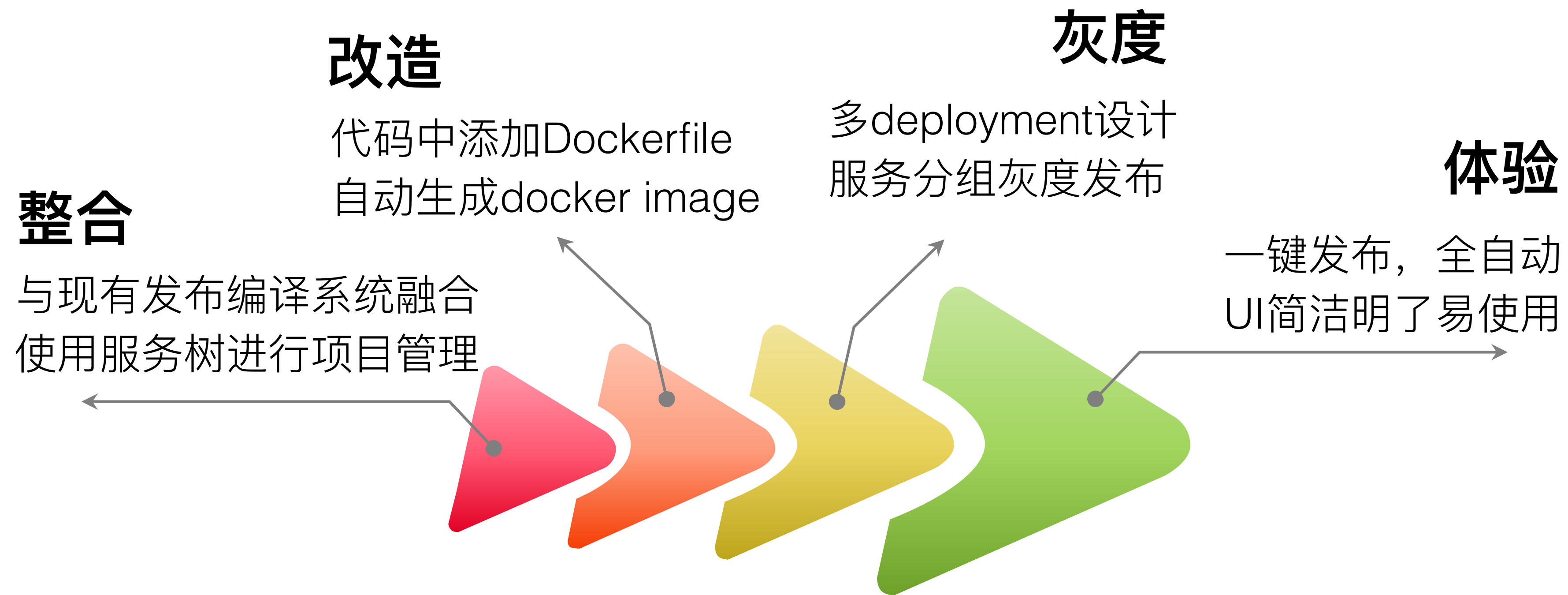
Merged k8s-merge-robot merged 1 commit into `kubernetes:master` from `fraenkel:service_names_rfc1035`

Conversation 18 Commits 1 Files changed 9

fraenkel commented on 25 Jul kubernetes contributor fixes #3752

googlebot added the `cla: yes` label on 25 Jul

代码发布



代码发布

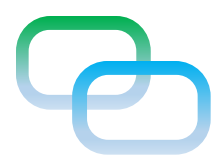
The screenshot shows the Didi Elastic Cloud (滴滴弹性云) console interface. The main area displays a deployment table for the environment 'gz01'. A modal dialog box titled '发布更新' (Release Update) is open, allowing for the configuration of a new deployment.

部署名称	当前状态	容器个数
test1	created	0
test2	pack	0

发布更新 (Release Update) Dialog Fields:

- 部署名称 (Deployment Name): name
- 代码地址 (Code Address): [Redacted]
- 代码路径 (Code Path): /
- 代码版本 (Code Version): master (selected), a3453tv (available)

Buttons: 取消 (Cancel), 提交 (Submit)



代码发布

滴滴弹性云 服务管理 监控系统

搜索节点

gz01

新建部署 刷新

部署名称	当前状态	容器个数
test1	created	0
test2	pack	0

发布更新

部署名称: preview

状态

序号	操作	发布代码	发布镜像	状态
1	发起编译	master/a9u090		
2	正在编译			
3	编译成功			
4	正在发布			
4.1	第一组状态		updated	rollbacking
4.2	第二组状态		updated	rollbacking
4.3	第三组状态		ready	ready
4.4	第四组状态		ready	ready
5	发布完成			

查看日志

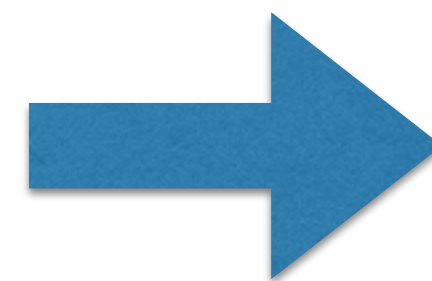
下一组 终止回滚

快速伸缩

当前容器数量:

目标容器数量  :

[保存](#) [查看操作记录](#)



-  尚未开始
-  更新集群
-  发起伸缩
-  正在伸缩 0/9
 - 
-  伸缩完成

[编译日志](#) [关闭任务](#)



快速伸缩

刷新

POD名称	pod状态	运行时间	ip	主容器名称	主容器状态	主容器详情
first-cc-mofang-gz01-2391188889-285up	Running	0.0h	172.16.1.5	first-cc-mofang-gz01	running	详情
first-cc-mofang-gz01-2391188889-3avg9	Running	0.0h	172.16.20.2	first-cc-mofang-gz01	running	详情
first-cc-mofang-gz01-2391188889-eyum0	Running	0.0h	172.16.137.3	first-cc-mofang-gz01	running	详情
first-cc-mofang-gz01-2391188889-gevz7	Running	4.7h	172.16.72.2	first-cc-mofang-gz01	running	详情
first-cc-mofang-gz01-2391188889-k0gzm	Running	0.0h	172.16.28.2	first-cc-mofang-gz01	running	详情
first-cc-mofang-gz01-2391188889-mb0h4	Running	0.0h	172.16.49.2	first-cc-mofang-gz01	running	详情
first-cc-mofang-gz01-2391188889-nz6be	Running	0.0h	172.16.75.4	first-cc-mofang-gz01	running	详情
first-cc-mofang-gz01-2391188889-pvseb	Running	0.0h	172.16.116.5	first-cc-mofang-gz01	running	详情
first-cc-mofang-gz01-2391188889-td5wv	Running	0.0h	172.16.95.2	first-cc-mofang-gz01	running	详情
first-cc-mofang-gz01-2391188889-yok0s	Running	0.0h	172.16.46.5	first-cc-mofang-gz01	running	详情

< 1 >

操作记录

操作时间	操作人员	操作概要	操作结果	编译日志
2016-12-06 15:46:45		伸缩 1->10	成功	编译日志

弹性伸缩

前提条件：服务可以在任意时刻稳定快速的伸缩

快速伸缩

+

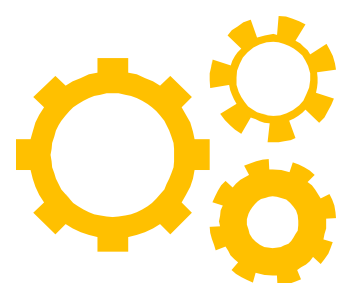
触发条件

=

弹性伸缩



监控实践



基础监控

cadvisor获取容器cpu、mem等基础信息
上报到监控系统



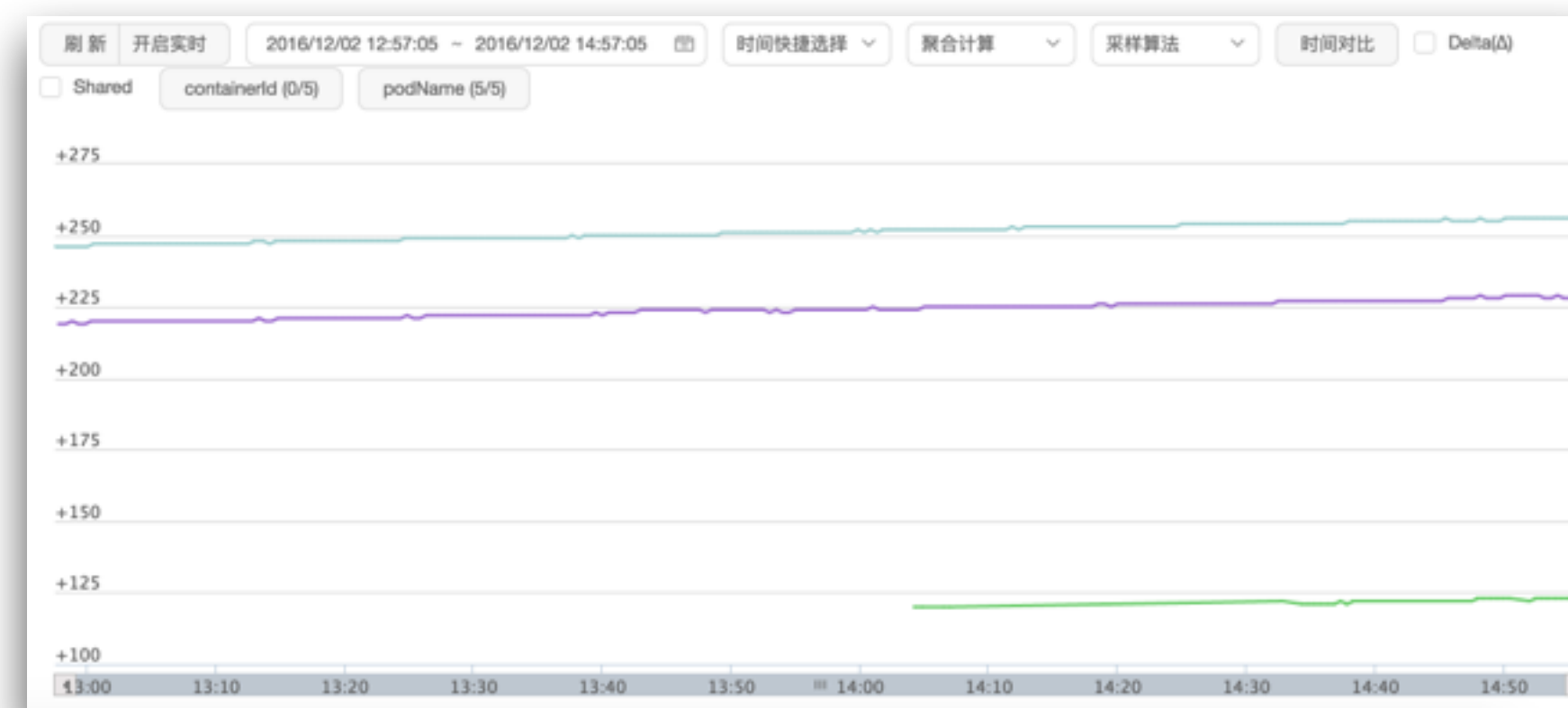
业务监控

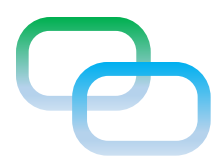
容器内置metriclib，业务主动上报数据



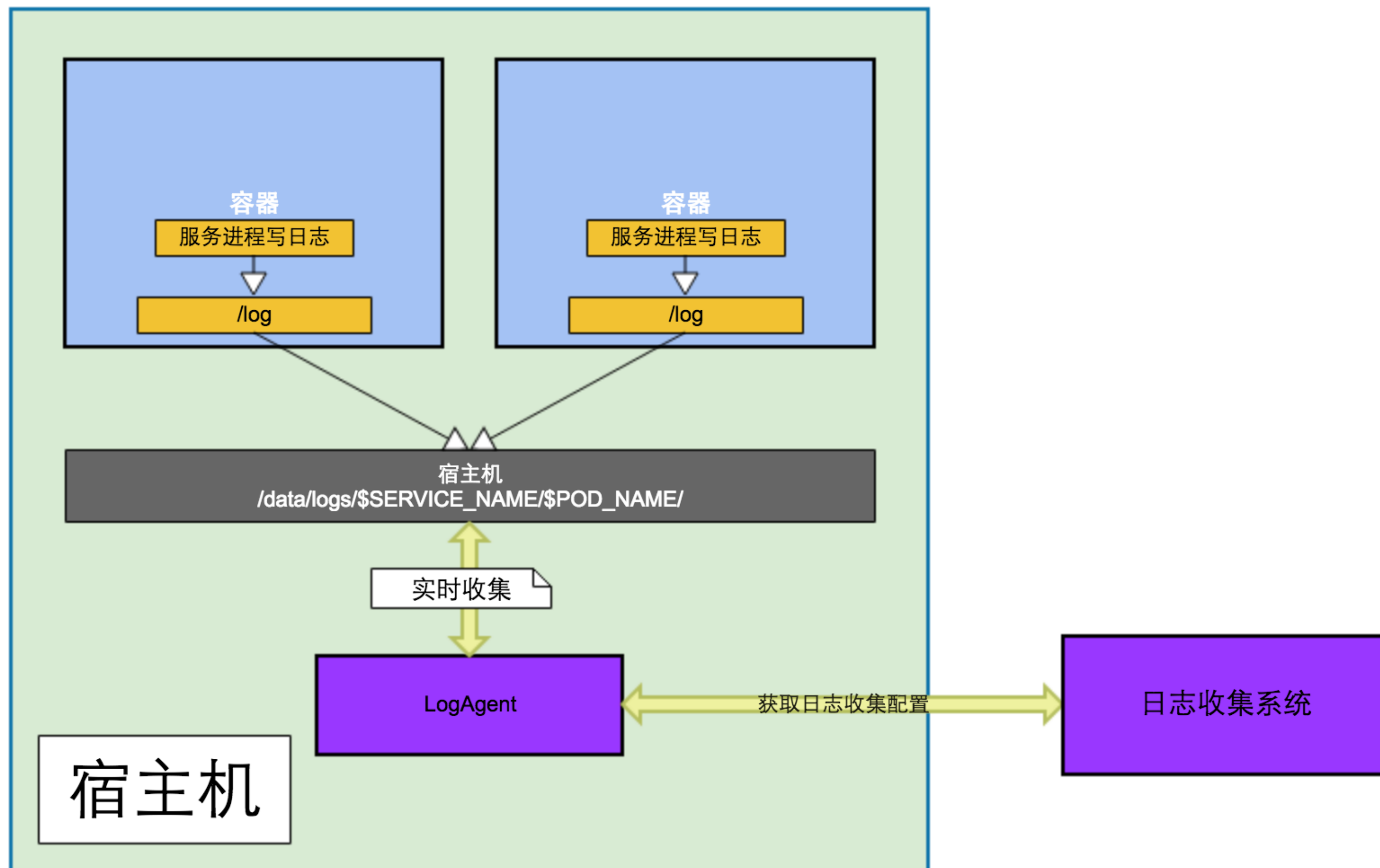
监控系统改造

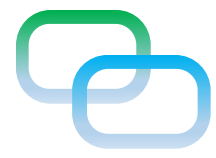
页面UI、图标改造适应容器需求



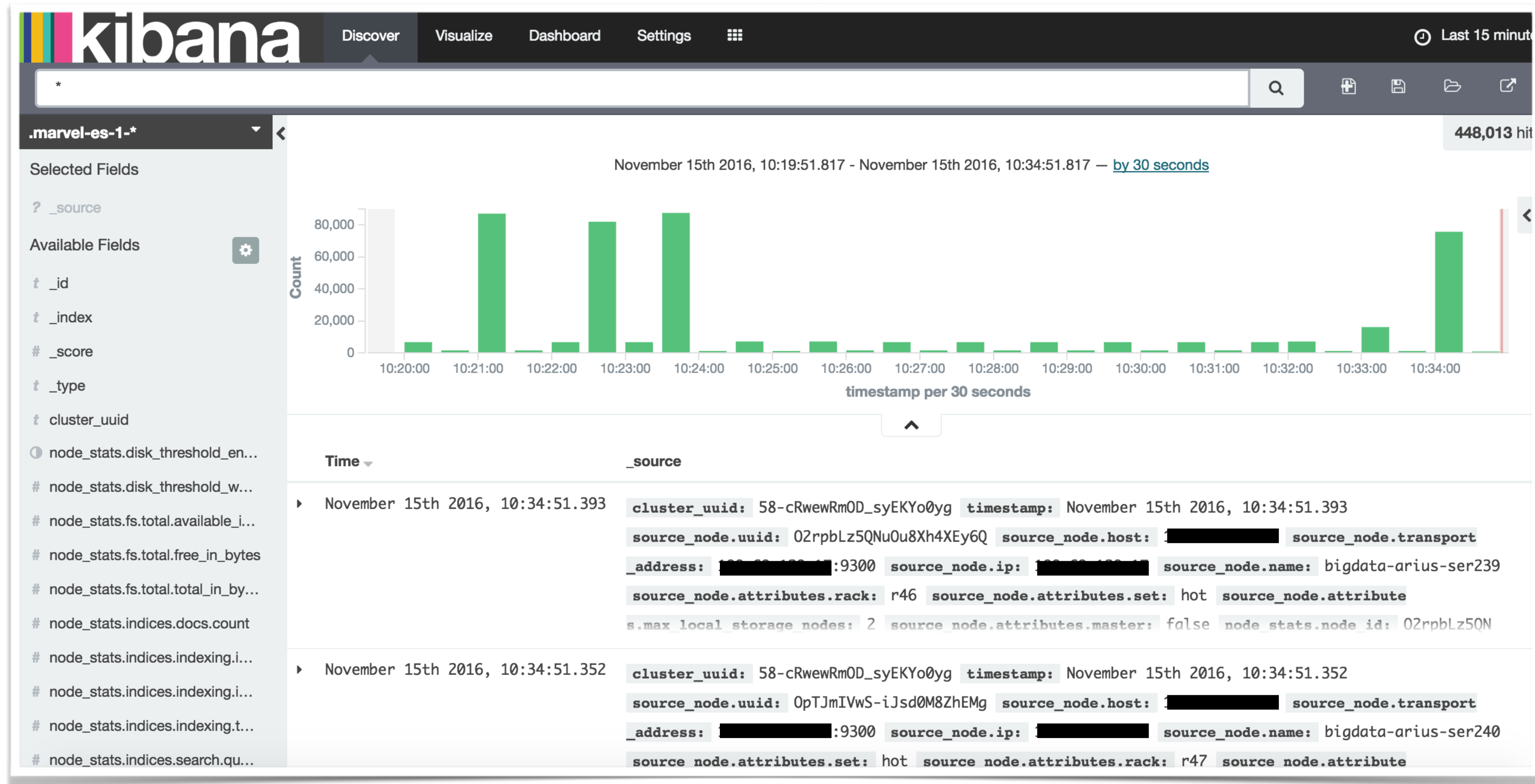


日志解决方案





日志解决方案



服务发现

1

容器物理和物理网络中的服务调用

使用Kube-proxy (nodePort) +LVS

2

容器集群内服务调用

Kube-proxy (clusterIP)

滴滴服务框架Disf, 动态发现和注册等功能



容器登陆

1 与滴滴现有机器登陆系统整合

2 镜像安装dssh客户端

3 用户登陆与物理机一致



POD

刷新

POD名称	pod状态	运行时间	ip	主容器名称	主容器状态	主容器详情
op-h5-h5-online-gz01-4079370962-azda2	Running	3days 4.0h	172.16.131.3	op-h5-h5-online-gz01	running	详情
op-h5-h5-online-gz01-4079370962-c41at	Running	3days 4.0h	172.16.1.3	op-h5-h5-online-gz01	running	详情
op-h5-h5-online-gz01-4079370962-h0mtz	Running	3days 4.0h	172.16.52.2	op-h5-h5-online-gz01	running	详情
op-h5-h5-online-gz01-4079370962-ltcgv	Running	3days 4.0h	172.16.69.2	op-h5-h5-online-gz01	running	详情
op-h5-h5-online-gz01-4079370962-oonbx	Running	3days 4.0h	172.16.88.2	op-h5-h5-online-gz01	running	详情

```

op-sven-opsec00.gz01:~$ dssh 172.16.1.3
Warning: Permanently added '[172.16.1.3]:22222' (RSA) to the list of known hosts.
Could not chdir to home directory /home/odin/xiedanbo: No such file or directory
-bash-4.1$ ip addr
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
41: eth0@if42: <BROADCAST,MULTICAST,UP,LOWER_UP,M-DOWN> mtu 1450 qdisc noqueue state UP
    link/ether 02:42:ac:10:01:03 brd ff:ff:ff:ff:ff:ff
    inet 172.16.1.3/24 scope global eth0
        valid_lft forever preferred_lft forever
  
```

总结



机遇

服务容器化收益巨大，重新定义运维



挑战

技术不是最大的挑战，用户习惯才是



展望



紧跟社区

确保新功能为我所用，Federation、StatefulSets等在代码、建议提交上，为社区做更多贡献

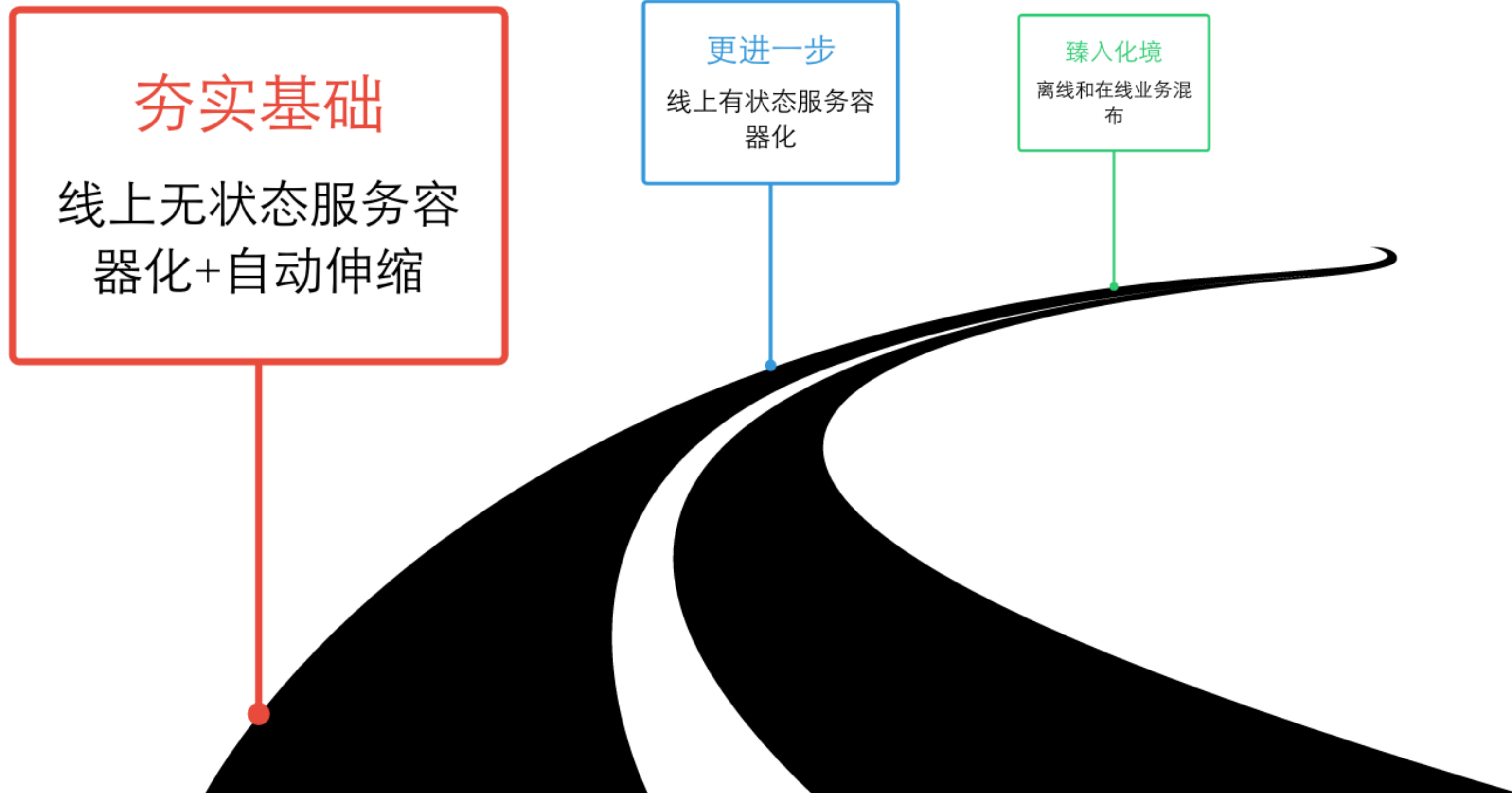


产品领先

紧跟业务，持续演进，保持敏锐
坚持用做产品的理念来推动新技术的落地



Roadmap





GIAC | BEIJING
Dec.12.16-17

技术架构未来

架構
ARCHNOTES
高 可 用 架 构



Thank you!

欢迎志同道合的小伙伴加入我们



简历砸来：xiedanbo@didichuxing.com