

腾讯云弹性伸缩

最佳实践

产品文档



腾讯云

【版权声明】

©2013-2017 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传播全部或部分本文档内容。

【商标声明】



及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。

【服务声明】

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或模式的承诺或保证。

文档目录

文档声明.....	2
最佳实践.....	4
为伸缩组设置固定出口IP.....	4
搭建可伸缩的微信小程序.....	10
五分钟节省30%成本（集群版）.....	21
设置服务开机启动.....	30

最佳实践

为伸缩组设置固定出口IP

本节向您介绍，如果集群需要主动外访，那么如何设置固定外访IP。

需求场景

如果您伸缩组中的集群，同时存在这三个需求：

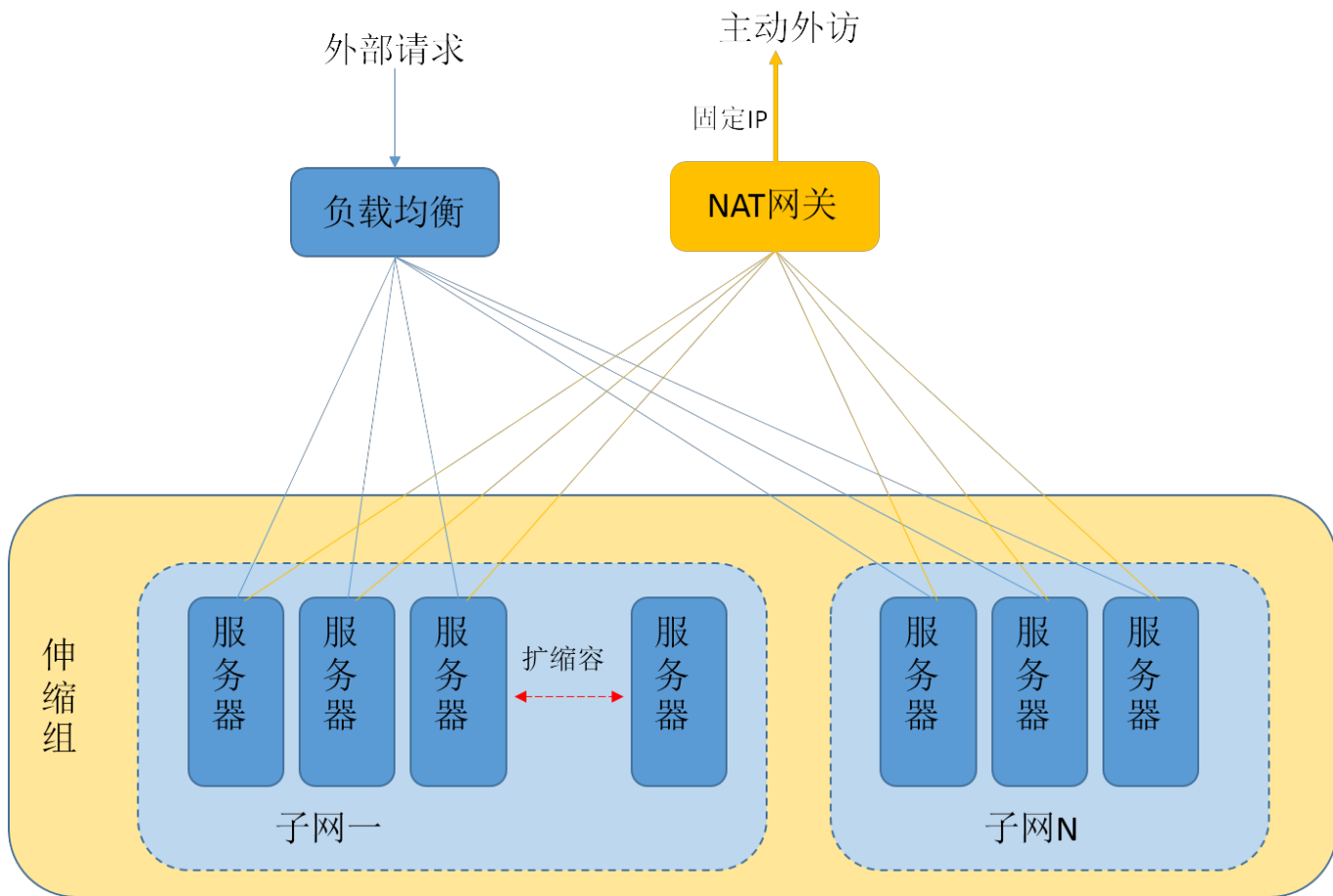
- 从负载均衡 CLB 接受请求
- 集群机器需要主动外访
- 外访时希望用固定的外网IP

那么您可以按如下方案进行设置。

方案简述

1. 通过负载均衡 CLB 接收和响应外部请求
2. 将机器放到 VPC 的子网中，将路由表指向 NAT 网关，主动外访请求统一经 NAT 网关的外网 IP 发出
3. 伸缩组的网络属性设为该子网，这样扩容出来的机器都会统一用 NAT 网关主动外访

如下图：



设置方法

第一步：创建VPC和子网

1. 创建VPC

1. 登录[腾讯云控制台](#)

，点击导航条【私有网络】，或者进入腾讯云[私有网络介绍页](#)中的【立即体验】按钮，进入[私有网络控制台](#)。

2. 选择列表上方下拉框中的地域，点击【新建】创建私有网络，例如，选择地域“华北地区（北京）”。

3. 填写私有网络和子网的名称和 CIDR，并选择子网的可用区。

4. 点击【创建】。

2. 创建子网

1. 登录[腾讯云控制台](#)

，点击导航条【私有网络】，点击左导航栏中的【子网】。选择下拉框中的地域和私有网络。



2. 点击【新建】，填写子网络名称、CIDR、可用区和关联路由表。然后点击【创建】确认。

3. 新建完成后，您就可以购买机器到这个子网中了。

第二步：创建 NAT 网关

1. 购买

1. 登录[腾讯云控制台](#)，选择【私有网络】选项卡，选择【NAT网关】。

2. 点击左上角【新建】按钮，在弹出框中依次输入或确定以下参数：

- 网关名称
- 网关类型（网关类型创建后可更改）
- NAT 网关服务的私有网络（就是刚才您创建的私有网络）
- 为 NAT 网关分配弹性 IP，这个 IP 就是以后您机器外访的固定 IP

3. 选择结束后点击【确认】按钮，即可完成 NAT 网关的创建。
4. 创建完 NAT 网关，您需要在私有网络控制台路由表页配置路由规则，以将子网流量指向 NAT 网关。

2. 设置路由表(重点)

1. 登录[腾讯云控制台](#)点击导航条【私有网络】，进入[私有网络控制台](#)，选择【路由表】。
2. 在路由表列表中，点击需要访问 Internet 的子网所关联的路由表 ID 进入路由表详情页，在路由策略中点击【编辑】按钮。
3. 点击新增一行，填入目的端（比如这种场景下可以填“0.0.0.0/0”，下一跳类型选择【NAT网关】，并选择已创建的 NAT 网关 ID，然后点击确定。

如果该路由表所绑定子网内的云主机需要通过公网网关访问公网，请不要选择该路由表绑定子网内的公网网关，点击查看详情。

目的端	下一跳类型	下一跳	备注
Local	Local	Local	系统默认下发，表示 VPC 内云主机网络互通
0.0.0.0/0	NAT 网关	nat-ope1wzme (as_...)	子网机器经 NAT 网关外访

+ 新增一行

确定 取消

选择刚才购买的 NAT 网关

至此，您在这个子网中的机器，即使没有公网 IP，也可以经过 NAT 网关主动外访了，对外而言还是固定的 IP。

如下图，即使我购买的是没有公网 IP 且带宽为 0 的主机，也可以主动外访：

```
[root@UM_1_162_centos ~]# ping www.baidu.com
PING www.a.shifen.com (220.181.111.188) 56(84) bytes of data.
64 bytes from 220.181.111.188: icmp_seq=1 ttl=54 time=45.7 ms
64 bytes from 220.181.111.188: icmp_seq=2 ttl=54 time=45.8 ms
```

但伸缩组需要识别这个子网，并确保机器都在这个子网上创建。

第三步：设置伸缩组

这个步骤的目的是将子网信息指向伸缩组，伸缩组就会把新扩容的机器开在这个子网中。
这样，扩容的机器会自动地用NAT网关的IP地址外访，达到固定出口IP的效果。

在[弹性伸缩控制台](#)，点击新建：

- 填写好伸缩组名字、启动配置（启动配置需事先设置好）、最大伸缩数、最小伸缩数、起始实例数等信息
- 选择“网络”和“子网”，指向刚才设置好的VPC和子网（重要）。

如下图：

名称 *

as_nat

名称不超过25个字符，名称仅支持中文、英文、数字、下划线、分隔符“-”、小数点。

最小伸缩数 *

0

最小实例数为0，可能会造成伸缩组中没有实例支撑业务

起始实例数

0

最大伸缩数 *

1

启动配置 *

网站应用服务器

新建启动配置 ?

支持网络 *

vpc-pajw5aw9 (ja...)

选择对应的VPC

支持子网 *

☒ subnet-hyqqysl6 (as_jayco)

选择刚才设置好的子网

☐ subnet-5fljio0s (as_for_nat)

您可选择多个子网，自动扩容的机器随机地从您勾选的子网创建，达到跨子网容灾的效果。

设置建议

移出策略 *

移出最旧的云主机

策略说明 ?

负载均衡

请选择负载均衡

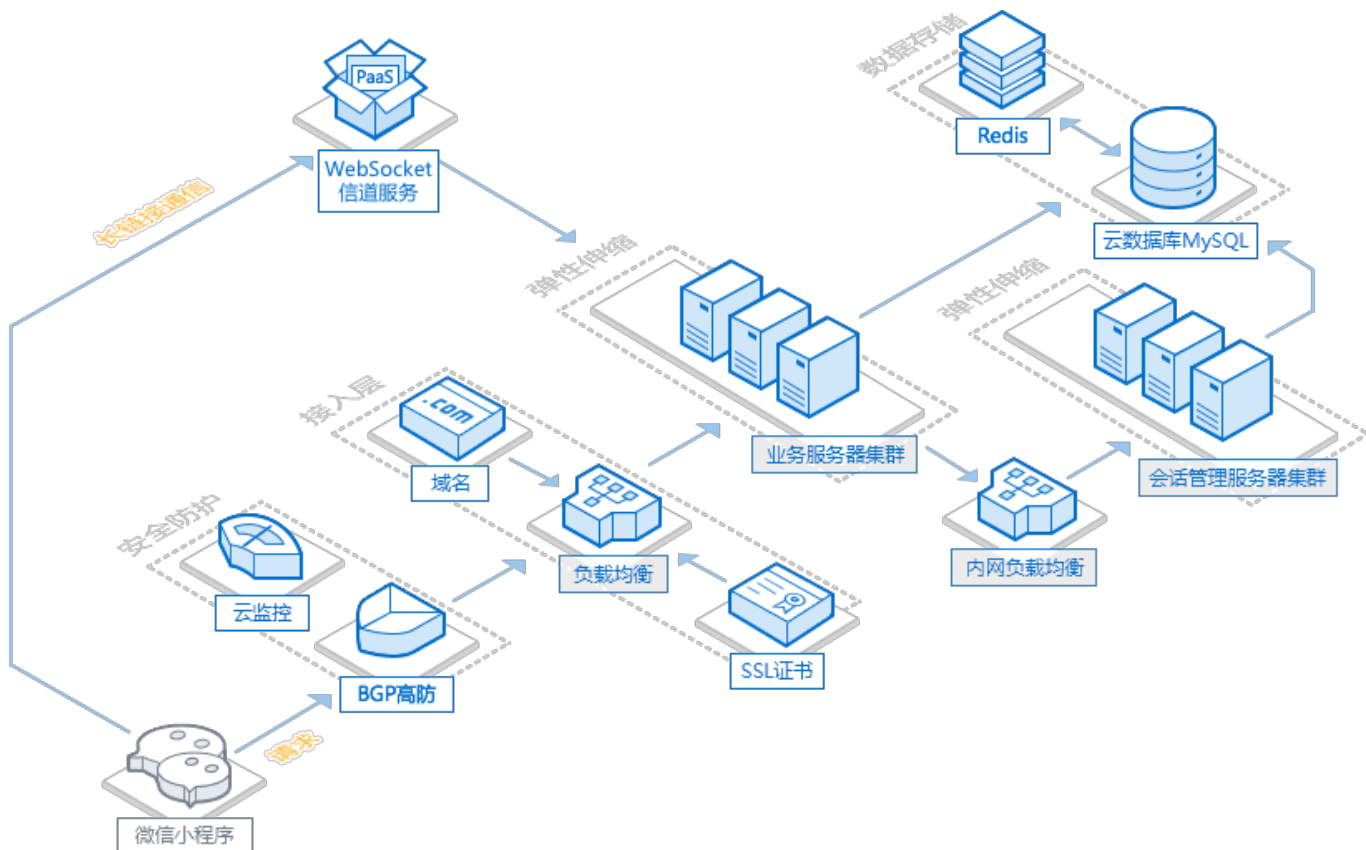
刷新 ?

扩容出来的机器会自动挂载到您关联的负载均衡下，您可以选择已有负载均衡或 新建。详细说明

至此设置完成。

搭建可伸缩的微信小程序 部署架构

我们可以按以下架构创建微信小程序，其中，业务服务器集群和会话管理服务器集群需要具备弹性伸缩能力。



为什么配置弹性伸缩？

以下两个场景下弹性伸缩能降低成本和提高业务连续性：

1. 小程序访问有明显的高峰低谷：根据测算，如果业务服务器集群和会话服务器集群需要不只一台机器，且高峰短于8个小时，采用 按闲时保留固定服务器 + 高峰时期增加临时服务器 的方式，能节约30%左右成本。您可以通过 AS 设置定时扩缩容任务，让腾讯云在忙时扩容临时服务器，闲时回收终止未充分利用的服务器。
2. 小程序访问量稳定的预期下，配置基于监控告警的伸缩策略可应对意外高负载，保障服务的持续可用，给问题解决争取时间。异常高负载包括 [CC攻击](#) 以及意外流量（例如，“脸萌”刚上线的远超预期的传播速度，或特定事件带来的突发访问）。可参看公益网站 [宝贝回家](#) 的案例。

注：弹性伸缩能力免费，扩容的 CVM 按秒正常计费。

弹性伸缩会做哪些事情？

弹性伸缩可以帮助您：

1. 定时给集群增加机器或减少机器；
2. 根据集群服务器的负载情况自适应地增加机器或减少机器；
3. 加的机器会自动注册到负载均衡中，实现全自动扩容。

配置前提

将微信小程序的能力搭建起来，可参考[详细步骤](#)。

为会话服务器配置弹性伸缩策略

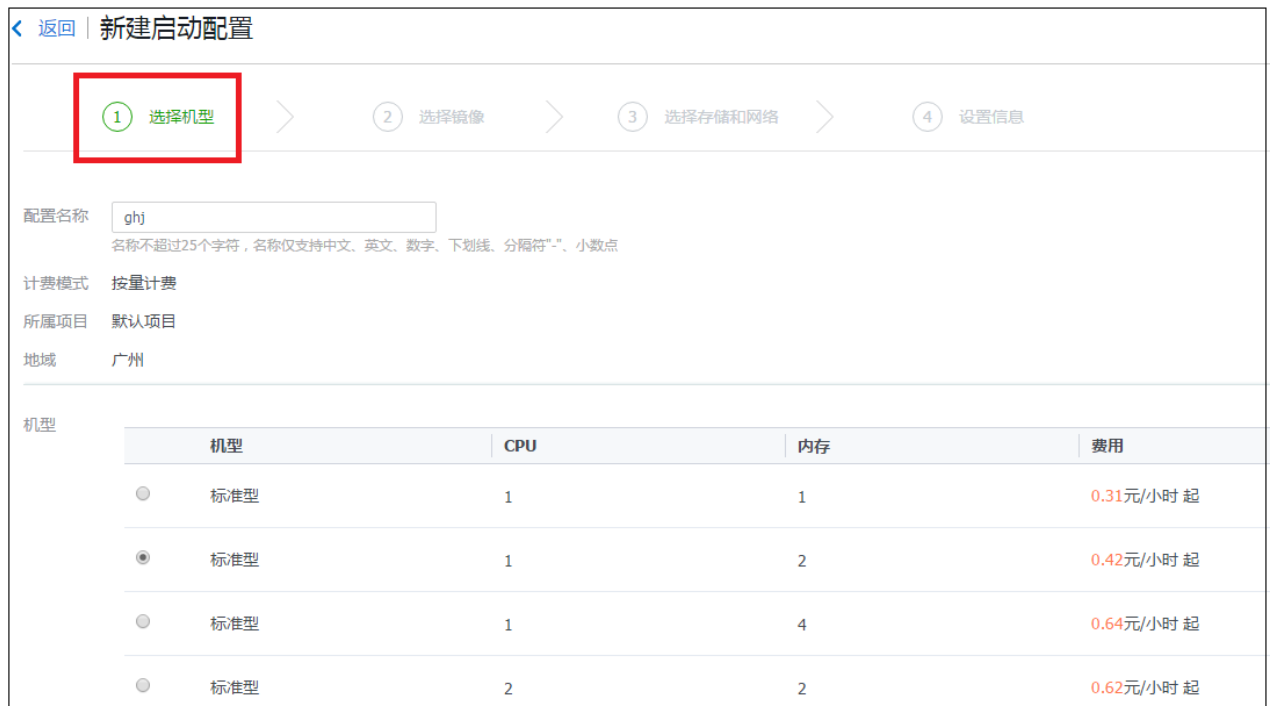
1. 创建启动配置

扩容时以 启动配置 为模板创建机器，因此我们事先通过启动配置指定地域、机型、镜像。

1. 登录[弹性伸缩控制台](#)，点击导航条中的【启动配置】。
2. 选择项目和地域，这里要注意必须选择小程序所在的项目和地域。

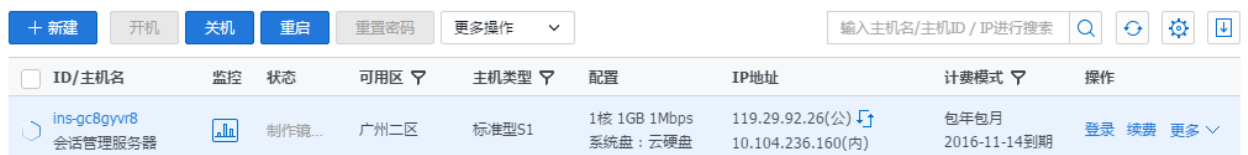


3. 接下来的操作与购买机器类似，您可跟着指引完成启动配置创建。



注：

此前您需要制作好镜像，这里需确保镜像里的应用能随操作系统启动，这样扩容出来的机器就能直接工作，无需人工介入。



2. 创建伸缩组

在弹性伸缩的 [控制台](#)，点击【新建】，按如下填写集群的管理信息：

- 名称：按需起一个名字，这里填“会话服务器集群”。
- 最小伸缩数：集群服务器数量的下限，这里填 0 即可。
- 起始实例数：伸缩组刚创建时，自动创建的机器数量。这里填 0 即可。
- 最大伸缩数：集群服务器数量的上限，这里按需填写。
- 启动配置：选择刚才您创建的启动配置。
- 支持网络：会话服务器的网络环境，一般选“基础网络”即可。
- 支持可用区：即选择机扩容器落在哪个可用区里，此处按会话服务器所在的可用区勾选即可。
- 移出策略：选择默认即可。
- 负载均衡：选择会话服务器的负载均衡。

新建伸缩组

通过工单提问

名称 *

会话服务器集群

名称不超过25个字符，名称仅支持中文、英文、数字、下划线、分隔符“-”、小数点

最小伸缩数 *

0

最小实例数为0，可能会造成伸缩组中没有实例支撑业务

起始实例数

0

最大伸缩数 *

5

启动配置 *

微信小程序会话服务器

新建启动配置

支持网络 *

基础网络

支持可用区 *

☐ 广州一区

☒ 广州二区

☒ 广州三区

您可选择多个可用区，自动扩容的机器会随机地从您勾选的可用区创建，达到跨可用区容灾的效果。
[设置建议](#)

移出策略 *

移出最旧的云主机

策略说明

负载均衡

256f71-0

刷新

扩容出来的机器会自动挂载到您关联的负载均衡下，您可以选择已有负载均衡或 [新建](#)。[详细说明](#)

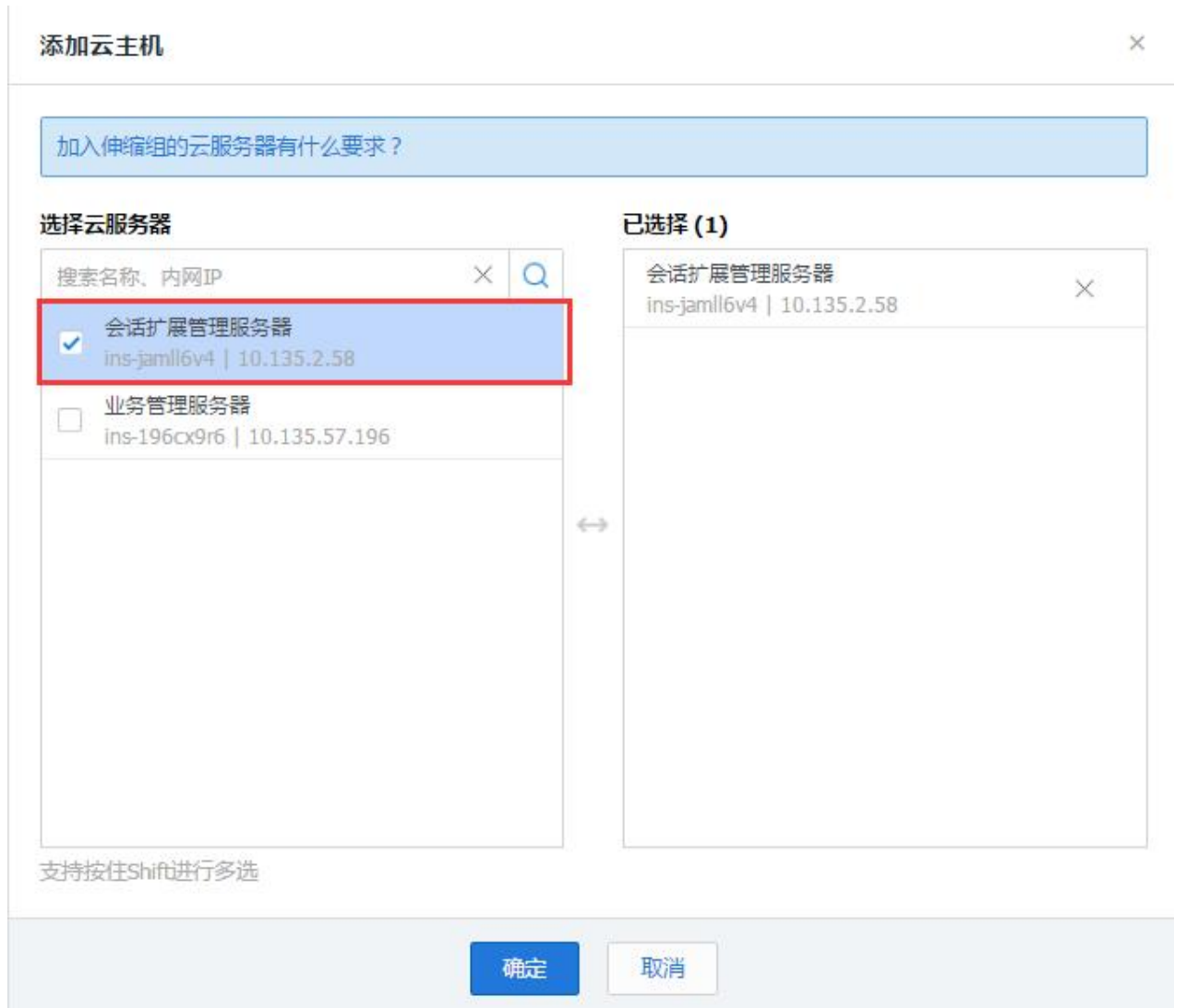
最后点击【确定】，完成创建。

3. 添加现有机器进伸缩组

1. 在 [控制台](#) 点击伸缩组名字，进入管理页，在页面下方点击“添加云主机”。



2. 在弹出的对话框中选择已有的会话服务器加入伸缩组。



3. 加入后对服务器设置“免于缩容”，这样在缩容活动中，伸缩组不会选择这台服务器缩容。

云主机列表

+添加云主机							
请输入云主机ID							
云主机ID/名称	监控状态	生命周期	移出保护	加入方式	启动配置	加入时间	操作
ins-jaml6v4 会话扩展管理服务器	健康	运行中	未启用	手动	-	2017-03-08 11:35:30	移出 设置移出保护

子机不会在缩容及替换活动中被移除

4. 设置扩容策略

您可以选择定时扩容、或者基于告警动态扩容（通常扩容任务和缩容任务成对出现）。



a. 定时扩缩容：

比如一个点餐小程序，可预期午饭时间的负载比其他时间高，您可以设置每天11：00-13：00扩容两台额外服务器支撑负载。

- 先设置一个定时扩容任务：

新建定时任务

名称 *

每日高峰前扩容

执行开始时间 *

2017-03-08

11时

:

00分

☒ 重复

重复周期

按天

每 1 天, 执行一次

执行结束时间

2018-12-31

13时

:

00分

伸缩组活动 *

更改最小实例数为: 0

更改期望实例数为: 10

更改最大实例数为: 10

确定

取消

勾选重复

从3月8日开始到12月31日，每天11点把集群的机器数增加到10台。

- 再设置一个定时缩容任务：

伸缩组详情

+新建

名称

每日高峰前扩容

新建定时任务

名称 *

每日高峰后缩容

执行开始时间 *

2017-03-08

13时

:

00分

☒ 重复

重复周期

按天

每 1 天, 执行一次

执行结束时间

2018-12-31

13时

:

00分

伸缩组活动 *

更改最小实例数为: 0

更改期望实例数为: 3

更改最大实例数为: 5

确定

取消

每天13点，把集群的服务器数量降低到3台

b. 基于告警扩容：

预期不明确的扩容，但需要防范出现意料之外的流量/攻击：

- 先设置一个告警扩容策略，用于应对异常流量：

新建告警触发策略

名称 *

防止意外流量

复制策略(选填)

请选择伸缩组

请选择

设置条件

if *

伸缩组内所有云主机:

CPU利用率

1分钟内

最大值

>

80

%

连续1次

详细统计规则

then

发送告警通知:

test

vickiitest

VirgilGroup

伸缩组活动 *

增加

2

台

云主机, 冷却

0

秒

伸缩活动

确定

取消

- 再设置一个告警缩容策略，用于清退未充分利用的服务器：

新建告警触发策略

名称 * 缩容策略

复制策略(选填) 请选择伸缩组 请选择

if * 伸缩组内所有云主机:

CPU利用率 1分钟内 最大值 < 20 % , 连续1次

详细统计规则

then 发送告警通知:

☐ vickiitest

☒ VirgilGroup

☐ 121

伸缩组活动 * 减少 1 台 云主机, 冷却 0 秒

确定 取消

利用率小于某个阈值

减少机器

为业务服务器配置弹性伸缩策略

这个过程与会话服务器类似：

- 为业务服务器创建一个启动配置；
- 为业务服务器创建一个伸缩组“业务服务器集群”，指向业务服务器的负载均衡；
- 为业务服务器设置伸缩策略。

验证伸缩性和查看伸缩活动

将伸缩组的期望实例数+1，伸缩组会自动扩容一台服务器到集群中。如果新扩的机器能正常处理请求，说明伸缩组已正常工作。

[返回](#) | 会话服务器集群

伸缩组详情

告警触发策略

定时任务

通知

伸缩活动

基本信息

[编辑](#)

← 点击编辑

名称 会话服务器集群

ID asg-qlmw9hlg

所在地域 广州 - 广州二区,广州三区

最小伸缩数 0

期望实例数 1 [+](#) ← 将期望实例数+1，伸缩组会自动扩容一台服务器

当前实例数 1

最大伸缩数 5

移出策略 移出最旧云主机

启动配置 [微信小程序会话服务器](#)

支持网络 基础网络

负载均衡 [lb-09mdgqi5](#)

创建时间 2017-03-08 11:23:37

伸缩组还支持[查询历史伸缩活动](#)，确保您完全掌控伸缩活动情况。

至此，您的小程序已经具备了智能扩容的能力。您无需再为扩容缩容的问题操心，只需留意伸缩组通知或者不定期查看历史伸缩活动即可。

五分钟节省30%成本（集群版）

本方案适合集群式部署的网站/APP。

如果您的业务满足以下条件，花5分钟配置这个方案，可节省30%成本：

- 网站使用集群的方式，且集群超过1台以上的服务器；
- 网站有较长时间的空闲。根据腾讯云的统计，90%的集群在凌晨 00：00-早上 09：00 这9个小时的负载低于30%。

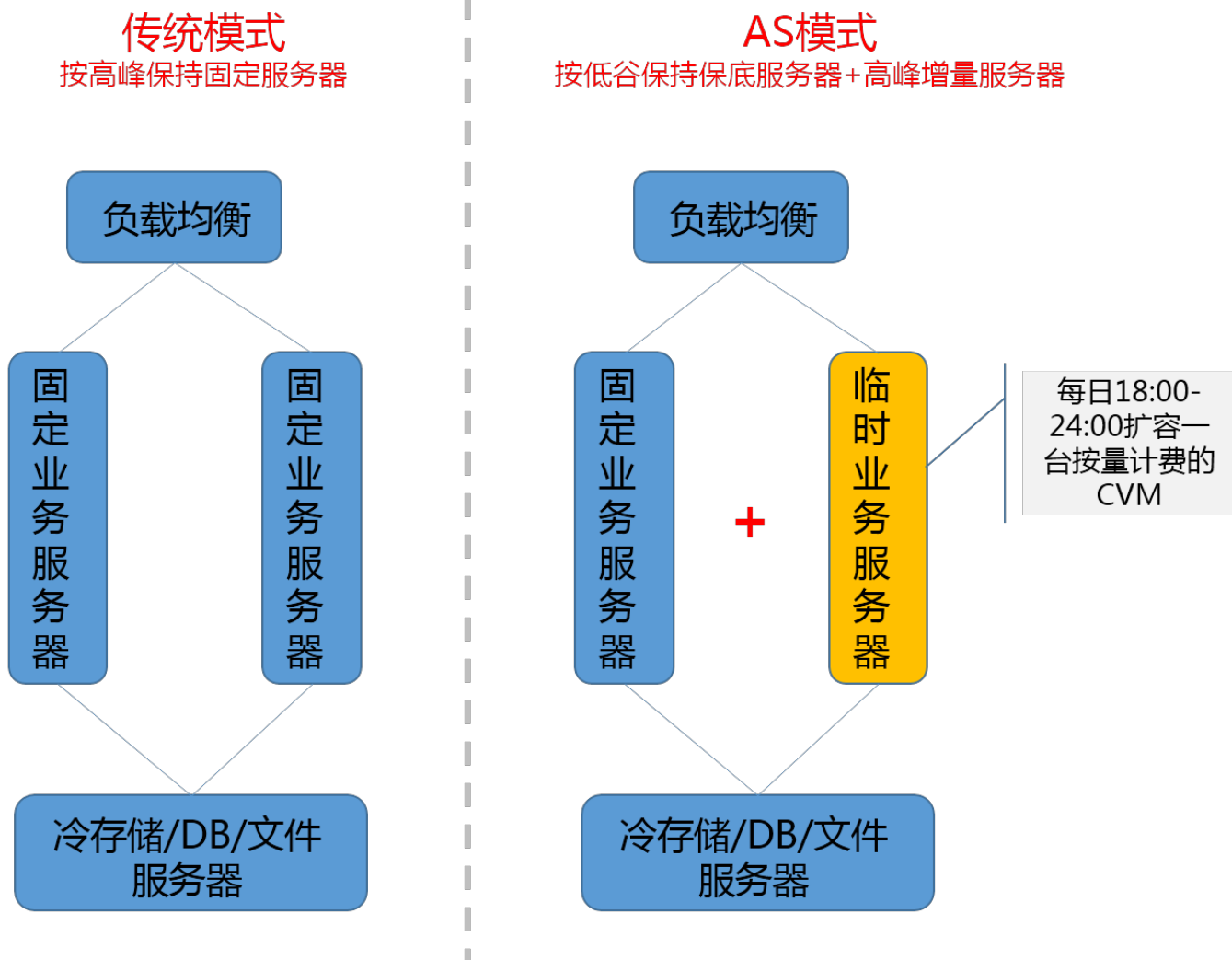
大部分网站的高峰时间不超过 8 个小时，剩下的 16 个小时的时间，完全可以把闲置的服务器作缩容处理，这种方式能帮助您节约大量成本。

本文以某休闲类网站为例，该网站 20：00-24：00 是访问高峰时段。

方案简述

- 按非高峰时段的负载部署固定资源，可采用包年包月 CVM；
- 高峰时段的不足部分采用按量计费的CVM。通过定时任务在 20：00 扩容1台，24：00 缩容回去。

新旧方案的对比：



收益

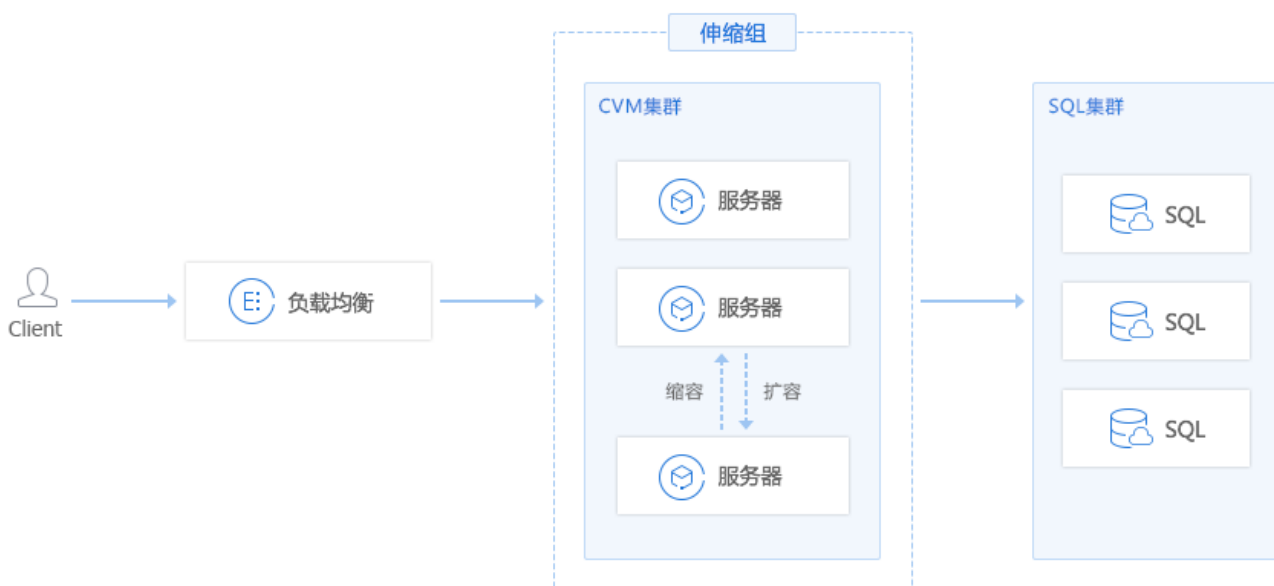
假设原方案需要两台 4核4G 的 CVM，改成一台 4核4G 的 CVM + 每天4个小时临时CVM，能节省30%左右开支。

示例中的小网站每年可以节省1500元：

	传统方案			弹性伸缩方案		
	资源	数量	开销	资源	数量	开销
所需资源	4核4G的包月CVM (固定资源)	2	592	4核4G的包月CVM (固定资源)	1	296
				120小时的临时CVM (4核4G)	1	148.8
	负载均衡	1	30	负载均衡	1	30
总费用			622			474.8
节省情况	每月节省147元，25%的幅度					

具体操作

实例的网站结构比较简单，只有应用服务器一个集群。如果复杂的网站，会有应用服务器集群、前端服务器集群、缓存服务器集群等，每个集群都可进行类似操作，每个集群对应一个伸缩组。



step 1. 创建集群机器的自定义镜像

这步非常简单，基于一台现成的集群机器中制作即可。如有疑问可查看 [制作自定义镜像 >>](#)

注：

您需要提前部署好镜像中的环境，保证镜像里的应用能随操作系统启动，这样扩容出来的机器就能直接工作，无需人工介入。

step 2. 创建启动配置

扩容时 AS 以启动配置为模板创建机器，因此我们事先通过启动配置指定地域、机型、镜像。

1. 登录 [弹性伸缩控制台](#)，点击导航条中的【启动配置】。
2. 选择项目和地域，这里要注意选择 Web 应用 所在的项目和地域。

ID/名称	已绑定	内存	带宽	
asc-l823tzpo yinpantest	1个	2GB	1Mbps	
asc-hm3gg09y 123	0个	1GB	1Mbps	
asc-l0e3bbqg H建议	0个	1核	2GB	1Mbps

3. 接下来的操作与购买机器类似，您可跟着指引完成启动配置创建。注意自定义镜像中，指定刚才您创建的镜像。

[返回](#) | 新建启动配置

1 选择机型

2 选择镜像

3 选择存储和网络

4 设置信息

配置名称

ghj

名称不超过25个字符，名称仅支持中文、英文、数字、下划线、分隔符“-”、小数点

计费模式

按量计费

所属项目

默认项目

地域

广州

机型

	机型	CPU	内存	费用
☐	标准型	1	1	0.31元/小时 起
☒	标准型	1	2	0.42元/小时 起
☐	标准型	1	4	0.64元/小时 起
☐	标准型	2	2	0.62元/小时 起

step 3. 为机器创建伸缩组

在 [弹性伸缩控制台](#)，点击【新建】，按如下填写集群的管理信息：

- 名称：按需起一个名字。比如这里填“应用服务器集群”
- 最小伸缩数：集群服务器数量的下限。示例这里填 0 即可。
- 起始实例数：伸缩组刚创建时，自动创建的机器数量。一般不会刚创建伸缩组就自动创建机器，建议这里填 0。
- 最大伸缩数：集群服务器数量的上限，这里按需填写。这里以 5 为例，即伸缩组最多有 5 台机器。
- 启动配置：选择刚才您创建的启动配置。
- 支持网络：会话服务器的网络环境，一般选“基础网络”即可。
- 支持可用区：即选择机扩容器落在哪个可用区里，此处按会话服务器所在的可用区勾选即可。
- 移出策略：选择默认。
- 负载均衡：选择集群的负载均衡。

新建伸缩组

通过工单提问

名称 *

应用服务器集群

名称不超过25个字符，名称仅支持中文、英文、数字、下划线、分隔符“-”、小数点

最小伸缩数 *

0

最小实例数为0，可能会造成伸缩组中没有实例支撑业务

起始实例数

0

最大伸缩数 *

5

启动配置 *

网站应用服务器

新建启动配置

支持网络 *

基础网络

支持可用区 *

广州一区

☒ 广州二区

广州三区

您可选择多个可用区，自动扩容的机器会随机地从您勾选的可用区创建，达到跨可用区容灾的效果。

设置建议

移出策略 *

移出最旧的云主机

策略说明

负载均衡

2b09d6-0

刷新

选择集群所使用的负载均衡

最后点击【确定】，完成创建。

step 4. 添加现有机器进伸缩组

1. 在 [控制台](#) 点击伸缩组名字，进入管理页，在页面下方点击【添加云主机】。

弹性伸缩

返回 | 应用服务器集群

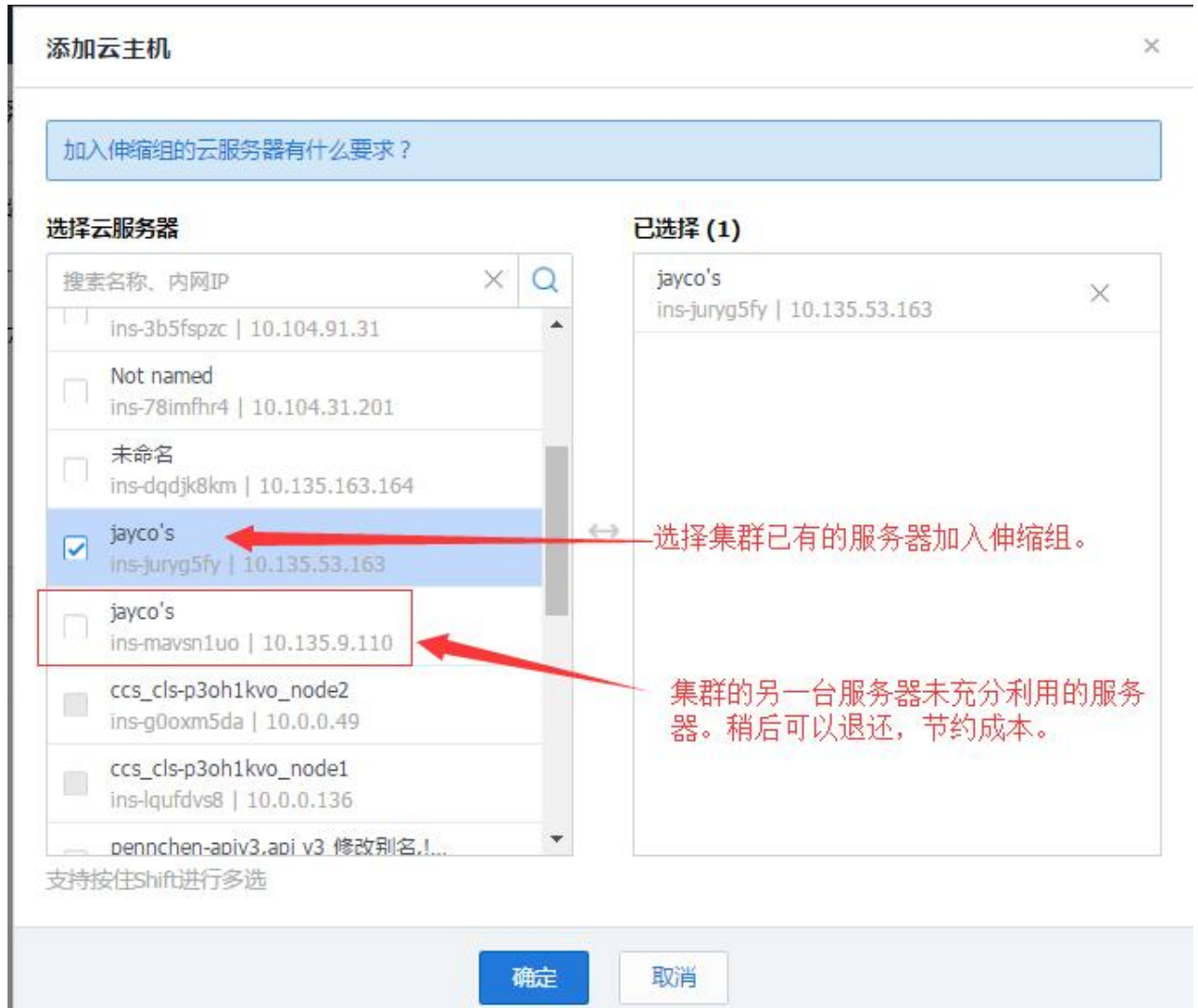
伸缩组

云主机列表

+添加云主机

云主机ID/名称	监控状态	生命周期	移出保护	加入方式	启动配置
列表为空					

- 在弹出的对话框中，选择集群已有的服务器加入伸缩组。如果现在是非高峰时期，集群中未充分利用的服务器可以退还，节约成本。



- 加入后对服务器设置“免于缩容”，这样在缩容活动中，伸缩组不会选择这台服务器缩容。这样集群中这台机器永远在服务，AS 不会更改它。

云主机列表

+添加云主机							
请输入云主机ID							
云主机ID/名称	监控状态	生命周期	移出保护	加入方式	启动配置	加入时间	操作
ins-juryg5fy jayco's	健康	运行中	未启用	手动	-	2017-03-09 15:42:52	移出 设置移出保护

step 5. 设置扩缩容策略（重点！）

AS 支持定时扩容或者基于告警动态扩容，也支持您接收扩缩容通知，以及翻看历史扩缩容详情。一切尽在您的

掌控中。



- 先设置一个20:00的定时扩容任务

新建定时任务

名称 * 24:00扩容

执行开始时间 * 2017-03-09 20时 : 00分 ☒ 重复

重复周期 按天 每 1 天, 执行一次

执行结束时间 2019-03-09 20时 : 00分

伸缩组活动 * 更改最小实例数为: 0 更改期望实例数为: 2 更改最大实例数为: 5

确定 取消

勾选“重复”，这样会每天执行

2017-03-09的20:00，伸缩组的实例数变为2（目前是1，即增加了1台）
每隔1天操作一次（即每日20:00）都调整到2台服务器
2019-03-09的20:00后，不再重复

注：

腾讯云的CVM需要1分钟左右创建，如果自定义镜像较大，可能需要更多时间。您可以将执行开始时间提早5分钟。

- 然后再设置一个24:00的定时缩容任务

新建定时任务

名称 *

24:00扩容

执行开始时间 *

2017-03-09

0时

:

05分

☒ 重复

重复周期

按天

每 1 天, 执行一次

执行结束时间

2019-03-09

0时

:

00分

伸缩组活动 *

更改最小实例数为: 0

更改期望实例数为: 1

更改最大实例数为: 5

确定

取消

24:00集群机器数量变为1台

至此大功告成！

网站的后台集群变为“1台固定应用服务器+1台高峰时定时创建的应用服务器”。
没加入伸缩组的其他集群机器，大部分时间未充分利用，可以退还掉节约成本。

设置服务开机启动

自动扩容的机器设置服务开机启动

使用场景

用auto

scaling扩容时，我们希望全程无人工介入。因此强烈为自动扩容的机器设置开机后服务自动启动，比如：

- httpd服务
- mysqld服务
- php-fpm服务
- tomcat服务
- 等等

其实花一分钟设置即可完成——修改 /etc/rc.d/rc.local文件！

设置方法（以centos为例）：

step 1：打开rc.local文件

输入

```
vim /etc/rc.d/rc.local
```

已有内容不动，在文件后面添加内容。

操作TIPS（资深用户可跳过）：

输入 “i”进入vim的insert模式，就可以输入内容，此时按方向键 “↓” 即可到达文件后面。

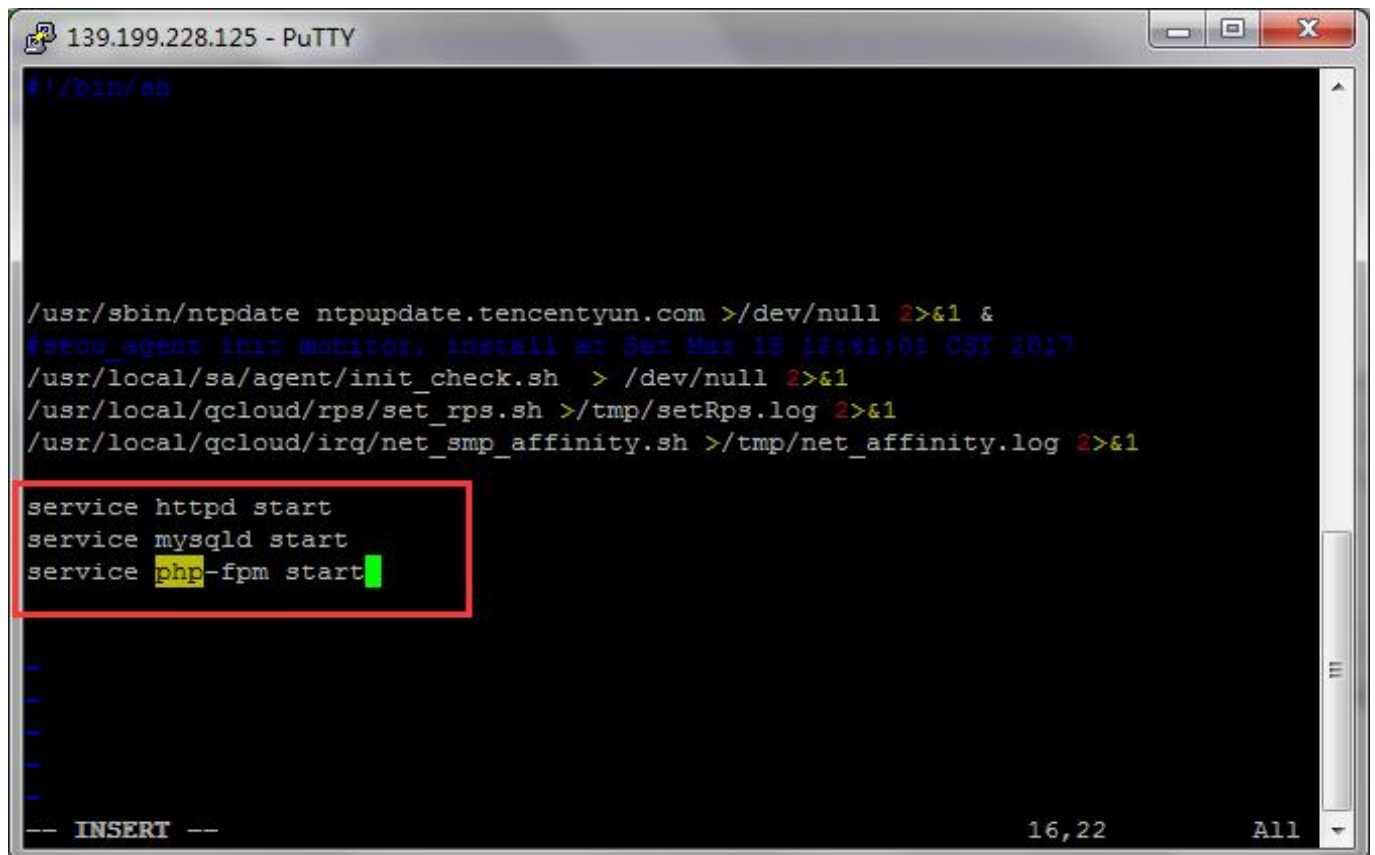
step 2：写好要启动的服务

本例子是希望搭建的网站开机自启动httpd、mysqld、php-fpm服务。因此我在rc.local后面加入这几句：

```
service httpd start
```

```
service mysqld start
```

```
service php-fpm start
```



```
139.199.228.125 - PuTTY
# /bin/sh

/usr/sbin/ntpdate ntpupdate.tencentyun.com >/dev/null 2>&1 &
#seou_agent init monitor, install at Sat Mar 18 12:41:01 CST 2017
/usr/local/sa/agent/init_check.sh > /dev/null 2>&1
/usr/local/qcloud/rps/set_rps.sh >/tmp/setRps.log 2>&1
/usr/local/qcloud/irq/net_smp_affinity.sh >/tmp/net_affinity.log 2>&1

service httpd start
service mysqld start
service php-fpm start

-- INSERT --
16,22 All
```

保存并退出。之后这台server开机后，网站就自动可访问了。注意不同网站需要的服务不同，此步骤根据您的需要设置即可。

操作TIPS（资深用户可跳过）：

内容输入完成后，按esc键，然后按住shift键+两下z就可退出了。即输入ZZ。

step 3:验证（可选）

重启服务器（输入reboot即可重启，或者在控制台上重启），服务器重启完成后不进入服务器，直接刷新网站的网页看是否有响应。如有即设置成功。

step 4：基于这台的机器制作镜像，在制作启动配置时使用该镜像

此步骤比较简单，如果操作遇到困难可再参考以下教程：

[制作自定义镜像](#)

[制作启动配置](#)