



利用 Auto Scaling 实现高可用性

自主进度动手实验室

版本 2.0c

版权归 © 2019 Amazon Web Services, Inc. 及其附属公司所有。保留所有权利。未经 Amazon Web Services, Inc. 事先书面许可，不得复制或再分发本文的部分或全部内容。禁止商业性质的复印、出借或销售。

错误或更正？请发送电子邮件至 aws-course-feedback@amazon.com，联系我们。

还有其他问题？请通过 <https://aws.amazon.com/contact-us/aws-training/> 联系我们。

目录

实验室概述.....	3
概述	3
本实验将逐步指导您使用 Elastic Load Balancing (ELB) 和 Auto Scaling 服务对基础设施进行负载均衡和自动扩展。	3
涵盖的主题	3
什么是 Elastic Load Balancer (ELB) 与 Auto Scaling?.....	3
场景	4
您将从以下基础设施开始:	4
基础设施的最终状态为:	4
登录 AWS 管理控制台	5
登录 AWS 管理控制台	5
通过 CloudFormation 模板自动创建实验环境	5
为 Auto Scaling 创建 AMI.....	6
创建负载均衡器.....	7
创建启动配置和 Auto Scaling 组.....	8
验证 Auto Scaling 是否运行.....	12
测试 Auto Scaling.....	14
实验完成.....	15
结论.....	16
结束您的实验室课程.....	16
其他资源.....	16

实验室概述

概述

本实验将逐步指导您使用 Elastic Load Balancing (ELB) 和 Auto Scaling 服务对基础设施进行负载均衡和自动扩展。

涵盖的主题

本实验室课程将演示：

- 创建负载均衡器 (Elastic Load Balancer)。
- 创建启动配置和 Auto Scaling 组。
- 自动扩展私有子网内的新实例
- 创建 Amazon CloudWatch 警报并监控基础设施的性能。

什么是 Elastic Load Balancer (ELB) 与 Auto Scaling?

Elastic Load Balancer 在多个 Amazon EC2 实例间自动分配入站应用程序流量。它可以让您实现应用程序容错能力，从而无缝提供路由应用程序流量所需的负载均衡容量。Elastic Load Balancing 提供两种负载均衡器，这两种负载均衡器均具备高可用性、自动扩展功能和可靠的安全性。两种类型分别是：

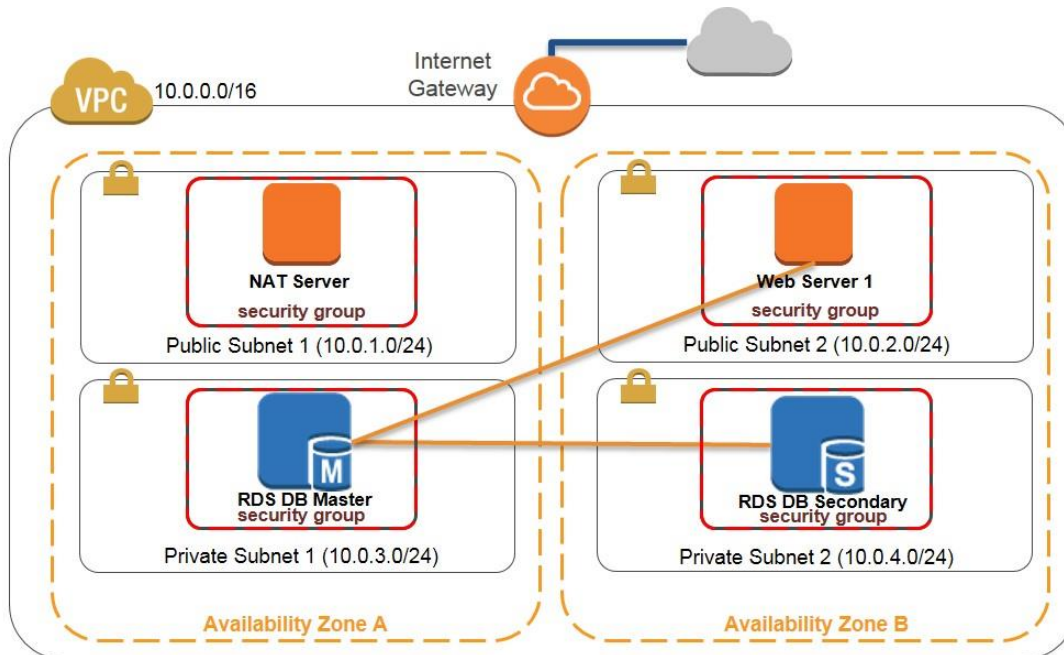
- Classic Load Balancer (<https://aws.amazon.com/elasticloadbalancing/classicloadbalancer/>)（可根据应用程序或网络级信息路由流量）。
- Application Load Balancer (<https://aws.amazon.com/elasticloadbalancing/applicationloadbalancer/>)（可根据包括请求内容的高级应用程序级信息路由流量）。

Classic Load Balancer 适用于在多个 EC2 实例之间进行简单的流量负载均衡，而 Application Load Balancer 则适用于需要高级路由功能、微服务和基于容器的架构的应用程序。Application Load Balancer 让您能够跨同一 EC2 实例的多个端口将流量路由到多项服务或执行负载均衡。

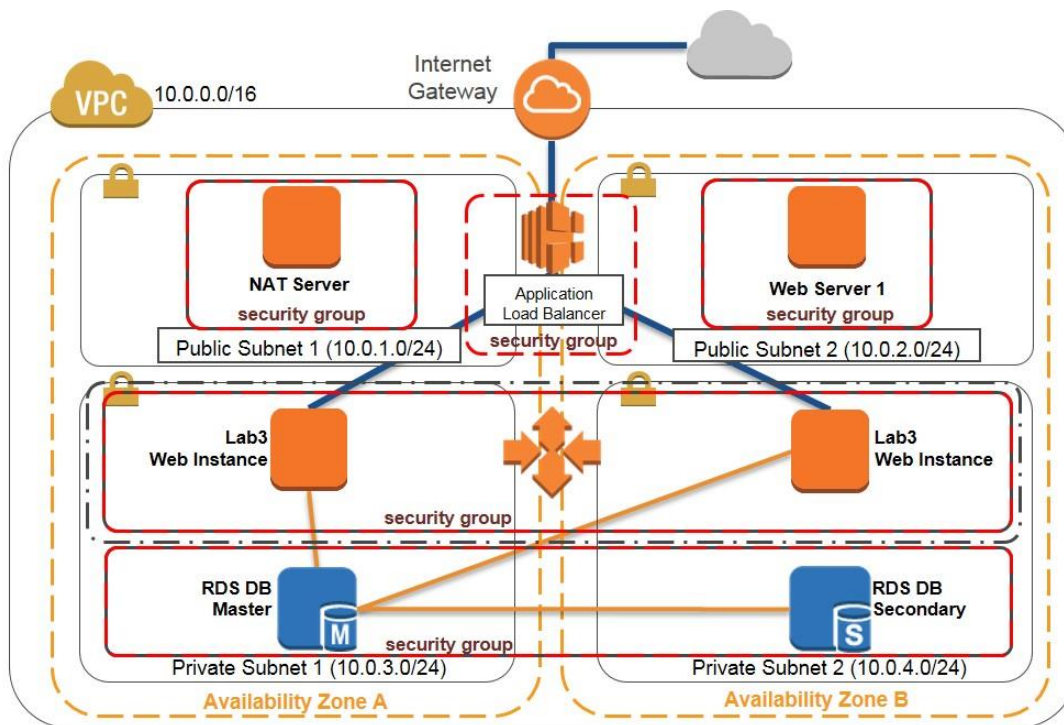
借助 **Auto Scaling**，您可以保持应用程序的可用性，并可根据您定义的条件自动扩展或缩减 [Amazon EC2](<https://aws.amazon.com/ec2/>) 容量。您可以使用 Auto Scaling 来帮助确保您运行的 Amazon EC2 实例数量符合所需数量。此外，Auto Scaling 还可以在需求峰值期自动增加 Amazon EC2 实例数量以保持性能，并在需求平淡期自动减少实例数量以降低成本。Auto Scaling 不仅非常适合需求模式稳定的应用程序，也适合使用模式每小时、每天或每周都不同的应用程序。

场景

您将从以下基础设施开始：



基础设施的最终状态为：



登录 AWS 管理控制台

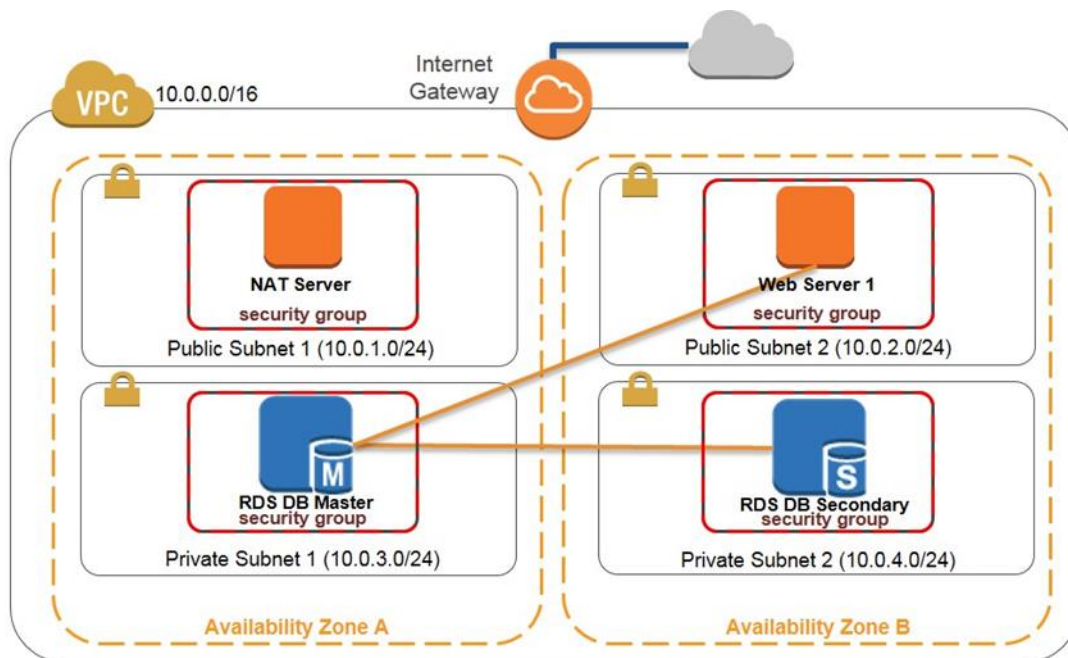
登录 AWS 管理控制台

欢迎使用自学实验室！您要做的第一步是登录 Amazon Web Services。

1. 单击“启动实验”开始实验。
2. 单击“登录网址”，到达 AWS 管理控制台登录界面
3. 使用这些证书登录控制台：
 - a. 在登录界面的“用户名”框中，输入“账号”。
 - b. 在“密码”框中，输入“密码”。
4. 单击“登录”。

通过 CloudFormation 模板自动创建实验环境

在此任务中，您将使用 CloudFormation 来快速部署实验所需要的环境。如先前提到，此实验的开始会需要先预设以下资源：



5. 在 AWS 管理控制台的服务菜单上，单击“Cloud Formation”。
6. 点击“创建新堆栈”的蓝色按钮
7. 选择指定Amazon S3 模板URL，并将以下链接粘贴上：

http://s3.cn-north-1.amazonaws.com.cn/awssomeday/GreenField_Lab5_CloudFormation_NoDB

8. 点击 “下一步”。
9. 在堆栈名称上输入 “MyStack”，接着点击 “下一步”。
10. 保留所有的默认设定，直接下滑到底端，点击 “下一步”。
11. 点击 “创建”。

此堆栈的创建过程大约需要五分钟，在等待的同时您可以阅读以下文档以得到更多CloudFormation的相关讯息：

- 产品详情：<https://www.amazonaws.cn/cloudformation/>
- 常见问题：<https://aws.amazon.com/tw/cloudformation/faqs/>

12. 确认状态转变成 “CREATE_COMPLETE”之后，再继续下一个阶段的任务。您可以随时点击右上角的更新按钮来检视最新的创建状态。

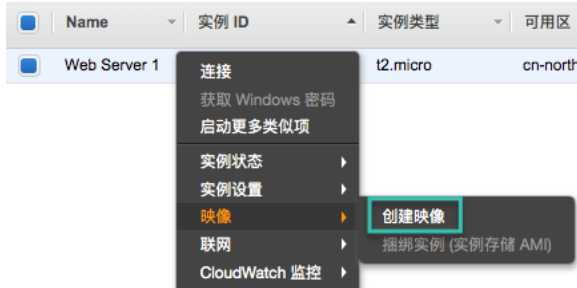


堆栈名称	创建时间	状态	描述
akh	2019-04-16 10:13:43 UTC+0800	CREATE_COMPLETE	

为 Auto Scaling 创建 AMI

在此任务中，您将先创建一个 AMI，以便启动要与 Auto Scaling 搭配使用的新实例。

13. 在 AWS 管理控制台的服务菜单上，单击 EC2。
14. 在左侧导航窗格中，单击 “实例”。
15. 确认 **Web Server 1** 的状态检查结果是否显示 “2/2 的检查已通过”。如果没有，请等到显示该状态，然后再继续下一步。使用右上角的刷新图标来检查更新。
16. 右键单击 **Web Server 1**，然后单击**映像**, 创建映像。



17. 配置以下设置（并忽略未列出的所有设置）：
 - 映像名称：输入 “Web Server AMI”。
 - 映像描述：输入 “Lab AMI for Web Server”。
18. 单击 “创建映像”。

19. 确认屏幕将显示新 AMI 的 AMI ID，单击关闭。
20. 在左侧导航窗格中，单击“AMI”可以看到映像创建的状态。

创建负载均衡器

在此任务中，您将创建一个负载均衡器，用于均衡两个可用区内多个 EC2 实例之间的流量。

21. 一样在EC的左侧导航窗格中，单击“负载均衡器”（在导航窗下方的位置）。
22. 单击“创建负载均衡器”。
23. 选择“应用程序负载均衡器”，单击 **创建**。
24. 配置以下设置（并忽略未列出的所有设置）：

- 名称：键入“LabELB”。
- VPC：单击“MyVPC”。
- 可用区：选择两个可用区以查看可用子网。
- 然后选择“Public Subnet 1”和“Public Subnet 2”。

可用区

指定要为负载均衡器后用的可用区。负载均衡器仅将流量路由至这些可用区中的目标。每个可用区您只能指定-

25. 单击 **下一步：配置安全设置**。
26. 若出现警告信息，请忽略以下警告：“加强您的负载均衡器安全。您的负载均衡器未在使用任何安全侦听器”，然后单击**下一步：配置安全组**。
27. 点击**选择一个现有的安全组**。
28. 选择 **名称** 中包含 **Web-Security-Group**，且**描述** 为 **Enable HTTP access** 的安全组，然后清除默认复选框（表示采用默认安全组）。

步骤 3: 配置安全组

安全组是一组防火墙规则，控制至您的负载均衡器的流量。在此页上，您可以添加规则来允许到达您的负载均衡器的特定流量。首先，确定是创建一个新安全组还是选择一个现有安全组

分配安全组: ☐ 创建一个新安全组

☒ 选择一个现有的安全组

安全组 ID	名称	描述
<input type="checkbox"/> sg-020725f07da07eef4	default	default VPC security group
<input type="checkbox"/> sg-0be97001851cf57cd	GreenField-DBSecurityGroup-7AEV7KVFE74D	DB Instance Security group
<input type="checkbox"/> sg-02f61bfef602195b2	GreenField-NATSecurityGroup-160MBNG880CYK	Enable internal access to the NAT device
<input checked="" type="checkbox"/> sg-05868f464b7bb2bcb	GreenField-WebSecurityGroup-9WL2RGGFRJAM	Enable HTTP access

29. 单击 **下一步：配置路由**。

30. 在**目标组**下，对于**名称**，输入“LabGroup”。

步骤 4: 配置路由

您的负载均衡器的路由对目标组中的目标请求使用您指定的协议和端口，然后使用运行。

目标组

目标组	新建目标组
名称	LabGroup
目标类型	<input checked="" type="radio"/> 实例 <input type="radio"/> IP <input type="radio"/> Lambda 函数
协议	HTTP
端口	80

31. 展开 **高级运行状况检查设置**，然后配置以下设置（并忽略未列出的所有设置）：

- 正常阈值：输入 2
- 不正常阈值：输入 3
- 超时：输入 10

高级运行状况检查设置

端口	<input checked="" type="radio"/> 流量端口 <input type="radio"/> 覆盖
正常阈值	2
不正常阈值	3
超时	10 秒
间隔	30 秒
成功代码	200

32. 单击 **下一步：注册目标**。Auto Scaling 稍后将自动添加实例。

33. 单击 **下一步：审核**。

34. 单击 **创建**。

35. 在“已成功创建负载均衡器”消息上，单击“关闭”。

创建启动配置和 Auto Scaling 组

在此任务中，您将为 Auto Scaling 组创建一项启动配置。启动配置是 Auto Scaling 组在启动 EC2 实例时使用的模板。创建启动配置时，您可以指定实例的信息，例如 AMI、实例类型、密钥对、一个或多个安全组和块存储设备映射。Auto Scaling 组包含一系列具有相似特征的 EC2 实例，这些实例被视为逻辑组以便进行实例扩展和管理。

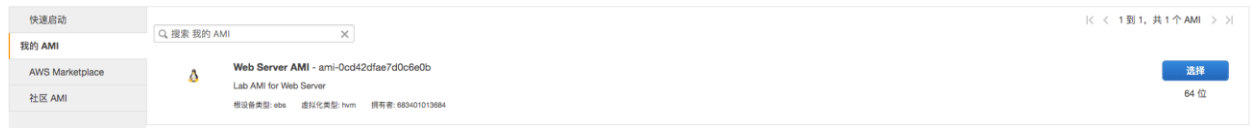
36. 一样在EC2的左侧导航窗格中，单击 **Auto Scaling 组**。

37. 单击 **创建 Auto Scaling 组** 的蓝色按钮。

38. 单击 **开始使用**。

39. 在左侧导航窗格中，单击 **我的 AMI**。

40. 在“Web Server AMI”对应的行中，单击 **选择**。



41. 接受 t2.micro 选项，然后单击 **“下一步：配置详细信息”**。

42. 配置以下设置（并忽略未列出的所有设置）：

- 名称：键入“LabConfig”
- 监控：单击 **“启用 CloudWatch 详细监控”**。

创建启动配置

名称	LabConfig
购买选项	<input type="checkbox"/> 请求竞价实例
IAM 角色	无
监控	<input checked="" type="checkbox"/> 启用 CloudWatch 详细监控 了解更多

43. 单击 **“下一步：添加存储”**。

44. 单击 **“下一步：配置安全组”**。

45. 单击 **“选择一个现有安全组”**，然后选择名称中包含“Web-Security-Group”，并且描述为“Enable HTTP Access”的安全组。

46. 单击 **“审核”**。

47. 查看启动配置的详细信息，然后单击 **创建启动配置**。

忽略“加强安全...”警告；这属于正常现象。

48. 单击**“在没有密钥对的情况下继续”**。

49. 勾选视窗下方的声明**“我确认我无法连接到此实例...”**

50. 单击 **“创建启动配置”**。

截至目前为止您设定的是 Auto Scaling 的启动配置，尚未实际部署任何资源。接下来要设定的 Auto Scaling 的扩展策略。

51. 配置以下设置（并忽略未列出的所有设置）：

组名称：输入“Lab AS Group”

组开始大小：键入“2”（个实例）

1. 配置 Auto Scaling 组详细信息	2. 配置扩展策略	3. 配置通知	4. 配置标签	5. 审核
创建 Auto Scaling 组				
组名	Lab AS Group			
启动配置	LabConfig			
组大小	从 2 个实例开始			

网络：单击“10.0.0.0/16的VPC”。忽略有关“无公共 IP”的消息；这属于正常现象。

子网：单击 “Private Subnet 1 (10.0.3.0/24)”，并单击 “Private Subnet 2 (10.0.4.0/24)”。

组名 *i* Lab AS Group

启动配置 *i* LabConfig

组大小 *i* 从 2 个实例开始

网络 *i* vpc-0df942a686e94223b (10.0.0.0/16) 新建 VPC

子网 *i* subnet-0d35d723893fcbf8b(10.0.3.0/24) | cn-north-1a x
subnet-0fd73927dd0d6354e(10.0.4.0/24) | cn-north-1b x 新建子网

52. 展开 **高级详细信息**，配置以下设置（并忽略未列出的所有设置）：

负载均衡：单击 **从一个或多个负载均衡器接收流量**。

目标组：单击 “LabGroup”。

运行状况检查类型：单击 “ELB”。

监控：单击 **启用 CloudWatch 详细监控**。

高级详细信息

负载均衡 *i* ☒ 从一个或多个负载均衡器接收流量

Classic 负载均衡器 *i*

目标组 *i* LabGroup x

运行状况检查类型 *i* ☒ ELB ☐ EC2

运行状况检查宽限期 *i* 300 秒

监控 *i* ☒ 启用 CloudWatch 详细监控 [了解更多](#)

实例保护 *i*

服务相关角色 *i* AWSServiceRoleForAutoScaling

53. 单击 **下一步：配置扩展策略**。

54. 选择 **“使用扩展策略调整此组容量”**。

55. 修改扩展范围文本框，将扩展范围设为 2 到 6 个实例。

创建 Auto Scaling 组

如果您要自动调整组的大小（实例数），可以选择调整扩展策略。扩展策略是指一组说明，地调整组大小。[了解更多](#)有关扩展策略的信息。

- ☐ 将此组保持在其初始大小
- ☒ 使用扩展策略调整此组的容量

在 2 和 6 个实例之间进行扩展。这两个值是组大小的最大和最小值。

56. 单击 **“使用分步或简单扩展策略扩展 Auto Scaling 组”**。



57. 在“增加组大小”中，对于执行策略的时间，单击“添加新警报”。



58. 清除“发送通知到：”。

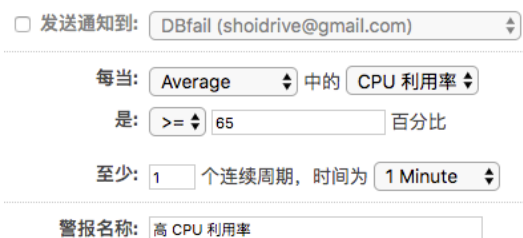
59. 配置以下设置（并忽略未列出的所有设置）：

每当：Average，然后单击“CPU 利用率”。

是：单击 \geq ，然后键入 65（表示百分比）。

至少：输入 1，然后单击 1 分钟。

警报名称：用“高 CPU 利用率”替换现有条目



60. 单击 **创建警报**。

61. 在增加组大小中，配置以下设置（并忽略未列出的所有设置）：

执行以下操作：输入 1，单击 **实例**，然后输入 65

实例需要：输入 60（每个步骤后的预热时间，单位为秒）



62. 在下方的 **减少组大小**，对于“执行策略的时间”，单击 **添加新警报**。

63. 清除 **发送通知到：**。

64. 配置以下设置（并忽略未列出的所有设置）：

每当：Average，然后单击“CPU 利用率”。

是：单击 \leq ，然后输入 20

至少：输入 1，然后单击 1 分钟。

警报名称：用“低 CPU 利用率”替换现有条目”

☐ 发送通知到: DBfail (shoidrive@gmail.com)

每当: Average 中的 CPU 利用率

是: >= 20 百分比

至少: 1 个连续周期, 时间为 1 Minute

警报名称: 低 CPU 利用率

65. 单击 **创建警报**。

66. 在 Decrease Group Size 中，对于执行以下操作：单击 **移除**，输入 1，单击 **实例**，然后输入 20

名称: Decrease Group Size

执行策略的时间: 低 CPU 利用率 [编辑](#) [删除](#)

达到警报阈值: CPUUtilization >= 20, 持续 60 秒

用于指标维度 AutoScalingGroupName = Lab AS Group

请执行以下操作: [删除](#) 1 个实例 当 20 <= CPUUtilization < +无限

[添加步骤](#) ⓘ

67. 单击 **下一步：配置通知**。

68. 单击 **下一步：配置标签**。

69. 配置以下设置（并忽略未列出的所有设置）：

- 密钥：输入“Name”
- 值：输入“Lab Web Instance”

70. 单击 **审核**。

71. 审核 Auto Scaling 组的详细信息，然后单击 **创建 Auto Scaling 组**。

72. Auto Scaling 组创建完成之后，单击 **关闭**。

验证 Auto Scaling 是否运行

在此任务中，您将验证 Auto Scaling 是否正常运行。

73. 在左侧导航窗格中，单击 **实例**。

系统将显示四个实例：Web Server 1、NAT Server，以及两个标记为 Lab Web Instance 的新实例。

注意：新实例应在几分钟后显示为正在运行。

74. 在左侧导航窗格中，单击 **负载均衡** 项目下的 **目标群组**

75. 选择 LabGroup，然后单击 **目标** 选项卡。两个 Lab Web Instance 应列入该目标组。

创建目标组 操作

按标签和属性筛选，或者按关键字搜索

名称	端口	协议	目标类型	负载均衡器	VPC ID
LabGroup	80	HTTP	instance	LabELB	vpc-0dff0c9f1f6e2fe9e7

目标组: LabGroup

描述 目标 运行状况检查 监控 标签

只要注册过程完成且新注册的目标通过初始运行状况检查，负载均衡器就会开始将请求路由至此目标。如果您的目标需求增加，您可注册其他目标。

编辑

这些可用区中没有包含任何正常目标。正在将请求路由至所有目标。

已注册目标

实例 ID	名称
i-0908a0e84c29c4d69	Lab Web Instance
i-012dd2d6fd1ee90f3	Lab Web Instance

76. 等待两个实例的 **状态** 转换为 **healthy**。使用右上角的刷新图标来检查更新。

77. 在左侧导航窗格中，单击 **负载均衡器**。

78. 选择 LabELB，然后在下方窗格的 **描述** 选项卡上，复制负载均衡器的 **DNS 名称**，并确保其中不含“(A 记录)”。

AMI

ELASTIC BLOCK STORE

卷

快照

网络与安全

安全组

弹性 IP

个置放群组

密钥对

网络接口

负载均衡

负载均衡器

目标群组

名称 DNS 名称 状态

LabELB	LabELB-1519403146.cn-nort...	active
--------	------------------------------	--------

负载均衡器: LabELB

描述 侦听器 监控 集成的服务 标签

基本配置

名称	LabELB
ARN	arn:aws-cn:elasticloadbalancing:cn-north-1:343456005339:load
DNS 名称	LabELB-1519403146.cn-north-1.elb.amazonaws.com.cn (A 记录)
状态	active

79. 将 DNS 网址粘贴在浏览器上，您应该能看到以下网页：

aws Load Test RDS

Meta-Data	Value
InstanceId	i-0c50915aa3685ca4c
Availability Zone	cn-north-1a

Current CPU Load: 0%

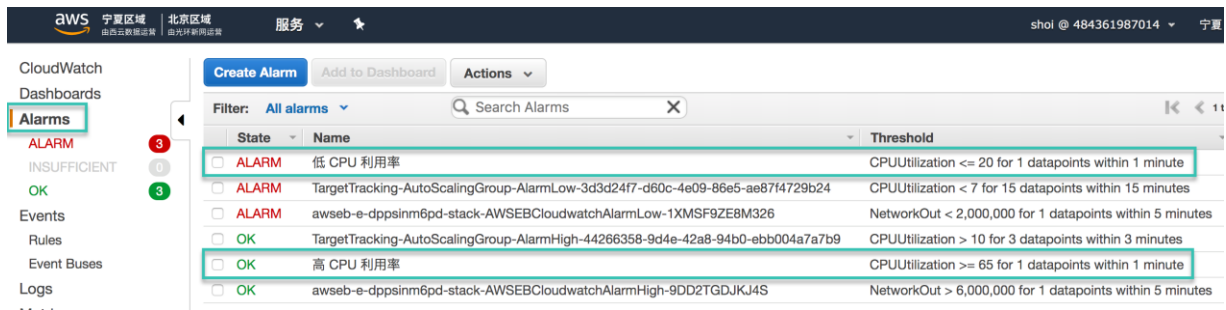
测试 Auto Scaling

您创建了一个最小规模为两个实例、最大规模为六个实例的 Auto Scaling 组。您创建了 Auto Scaling 策略，可以逐个增加或减少组中的实例。您创建了 Amazon CloudWatch 警报，可以在组的总体平均 CPU 使用率大于等于 65% 和小于等于 20% 时触发这些策略。目前，两个实例都在运行，因为最小规模为两个实例，该组当前没有任何负载。现在，您将使用创建的 CloudWatch 警报来监控该基础设施。

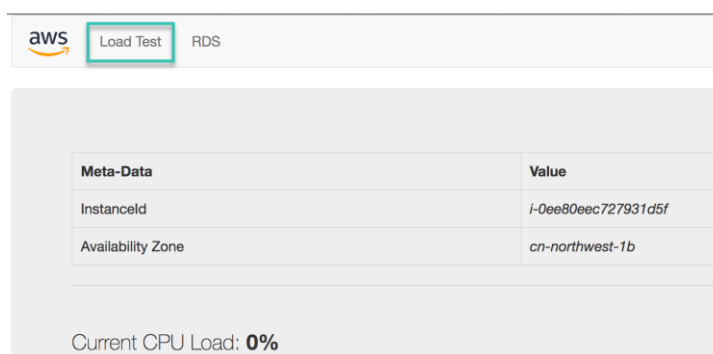
在此任务中，您测试了刚才实施的 Auto Scaling 配置。

80. 在 **服务** 菜单上，单击 CloudWatch。

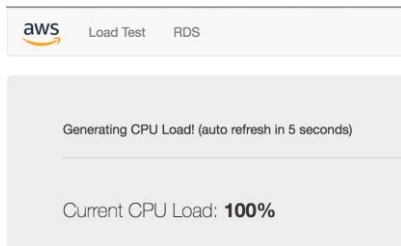
81. 在左侧导航窗格中，单击 **Alarms**（不是 ALARM）。系统将显示 **高 CPU 利用率** 和 **低 CPU 利用率** 两个警报。低 CPU 利用率的状态为 **ALARM**，而高 CPU 利用率的状态为 **OK**。这是因为组的当前 CPU 使用率小于 20%。Auto Scaling 没有删除任何实例，原因是组的当前大小处于其最小规模 (2)。



82. 将您在前一个任务中复制的负载均衡器的 DNS 名称粘贴到新的浏览器窗口或选项卡中，然后单击 AWS 徽标右方的 **Load Test**。



应用程序负载会测试实例，并每隔 5 秒自动刷新一次。当前 CPU 负载升至 100%。



83. 在 AWS CloudWatch 控制台上返回至窗口或选项卡。在 **5 分钟内**，低 CPU 警报状态变为 OK，而高 CPU 警报状态变为 ALARM。单击刷新图标以查看更改。

State	Name
<input type="checkbox"/> ALARM	高 CPU 利用率
<input type="checkbox"/> ALARM	TargetTracking-AutoScalingGroup-AlarmLow-3d3d24f7-d60c-4e09-86e5-ae87f4729b24
<input type="checkbox"/> ALARM	awseb-e-dppsinm6pd-stack-AWSEBCloudwatchAlarmLow-1XMSF9ZE8M326
<input type="checkbox"/> OK	TargetTracking-AutoScalingGroup-AlarmHigh-44266358-9d4e-42a8-94b0-ebb004a7a7b9
<input type="checkbox"/> OK	低 CPU 利用率
<input type="checkbox"/> OK	awseb-e-dppsinm6pd-stack-AWSEBCloudwatchAlarmHigh-9DD2TGDJKJ4S

84. 在服务菜单上，单击 EC2。

85. 在左侧导航窗格中，单击实例。现在，应该要看到有**两个以上**标记为 Lab Web Instance 的实例正在运行。它们可能处于创建期间，标签可能不会立即显示。新实例由 Auto Scaling 基于您在之前步骤中创建的 CloudWatch 警报创建。

实例	按标签和属性筛选，或者按关键字搜索
标签	
限制	
实例	
实例	
启动模板	
Spot 请求	
预留实例	

Name	实例 ID	实例类型	可用区	实例状态	状态检查
Default-Environment	i-0bf5fee7d2e20d77d	t2.micro	cn-northwest-1b	running	2/2 的检查...
Lab Web Instance	i-040b3ccf7cb5f6c77	t2.micro	cn-northwest-1a	running	2/2 的检查...
Lab Web Instance	i-07ab1728da8ed48c7	t2.micro	cn-northwest-1a	running	2/2 的检查...
Lab Web Instance	i-0ee80eec727931d5f	t2.micro	cn-northwest-1b	running	2/2 的检查...

86. 以上我们通过模拟负载测试，触发了 CloudWatch 警报，让原本只有两个 Lab Web Instance 的实例经由 Auto Scaling 自动响应来增加实例。

实验完成

恭喜您！您已成功使用 Auto Scaling 和 Elastic Load Balancing 对架构进行了管理

结论

恭喜您！现在，您已成功地：

- 从正在运行的实例中创建 Amazon 系统映像 (AMI)。
- 创建负载均衡器。
- 创建启动配置和 Auto Scaling 组。
- 自动扩展私有子网内的新实例
- 创建 Amazon CloudWatch 警报并监控基础设施的性能。

结束您的实验室课程

按照下列步骤关闭控制台、结束实验室。

1. 在 AWS 管理控制台的导航栏中，单击 “**UPT15xxxxxxxxxx @<AccountNumber>**”，再单击 “**注销**”。
2. 关闭所有活动的 SSH 客户端会话或远程桌面会话。
3. 在云平台的实验页面上，单击 “**结束实验**”。
4. 在确认消息中，单击**确定**。

其他资源

- 更多有关 AWS 培训和认证的信息，请访问 <http://aws.amazon.com/training/>。

如有反馈、建议或发现有任何需要修正之处，请通过 aws-course-feedback@amazon.com 给我们发送电子邮件。