

# 실습 3: 아키텍처 확장 및 로드 밸런싱

이 실습에서는 Elastic Load Balancing(ELB)과 Auto Scaling 서비스를 사용하여 인프라를 부하 분산하고 자동 조정하는 과정을 살펴봅니다.

**Elastic Load Balancing**은 수신되는 애플리케이션 트래픽을 여러 Amazon EC2 인스턴스에 자동으로 분산합니다. ELB를 사용하면 애플리케이션 트래픽을 라우팅하는 데 필요한 로드 밸런싱 용량을 원활하게 제공함으로써 애플리케이션의 내결함성을 달성할 수 있습니다.

**Auto Scaling**은 애플리케이션 가용성을 유지하는 데 도움이 되며, 정의한 조건에 따라 Amazon EC2 용량을 자동으로 확장 또는 축소할 수 있습니다. Auto Scaling을 활용하면 실행 중인 Amazon EC2 인스턴스의 수를 원하는 수준으로 유지할 수 있습니다. 또한 Auto Scaling은 수요가 급증할 경우 Amazon EC2 인스턴스의 수를 자동으로 증가시키기 때문에 성능을 그대로 유지할 수 있으며, 수요가 적을 경우 자동으로 용량을 감소시켜 비용 낭비를 막아줍니다. Auto Scaling은 수요 패턴이 안정적이거나 사용량이 시간, 일 또는 주 단위로 변경되는 애플리케이션에 적합합니다.

## 목표

본 실습을 완료하면 다음을 수행할 수 있습니다.

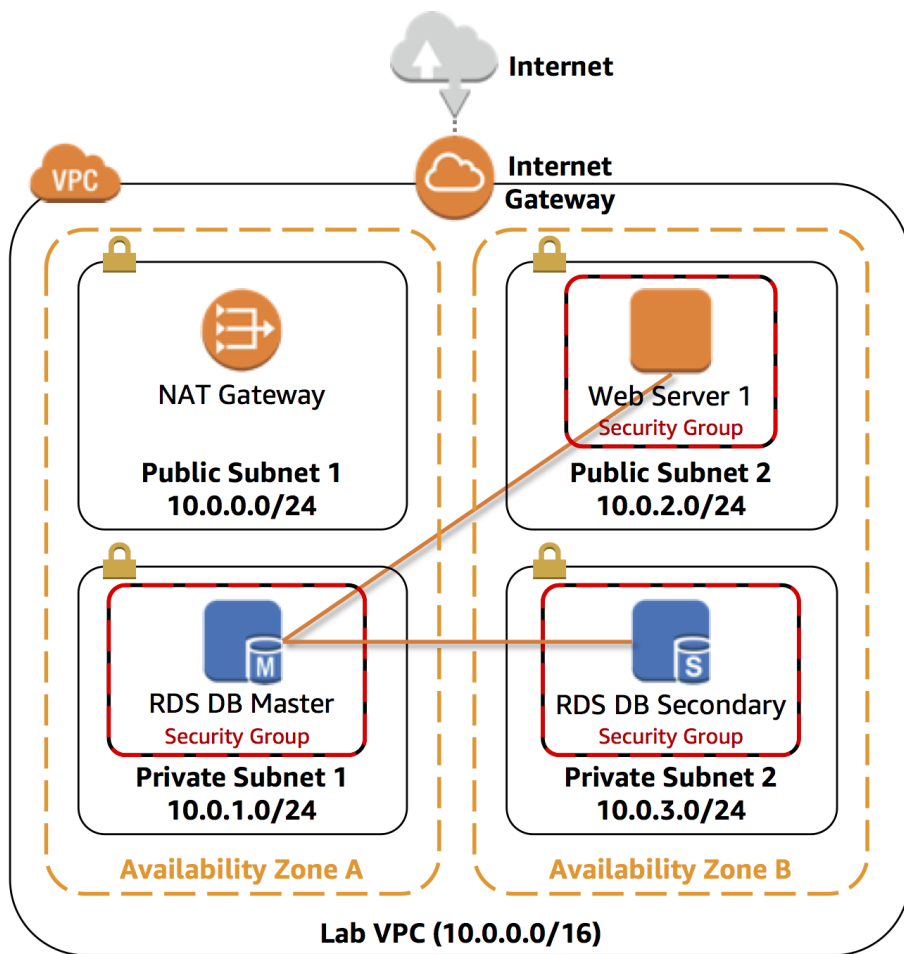
- 실행 중인 인스턴스에서 Amazon Machine Image(AMI) 생성
- 로드 밸런서 생성
- EC2 시작 템플릿 생성
- Auto Scaling 그룹 생성
- 프라이빗 서브넷 내에서 새 인스턴스를 자동으로 조정
- Amazon CloudWatch 경보 생성 및 인프라 성능 모니터링

## 소요 시간

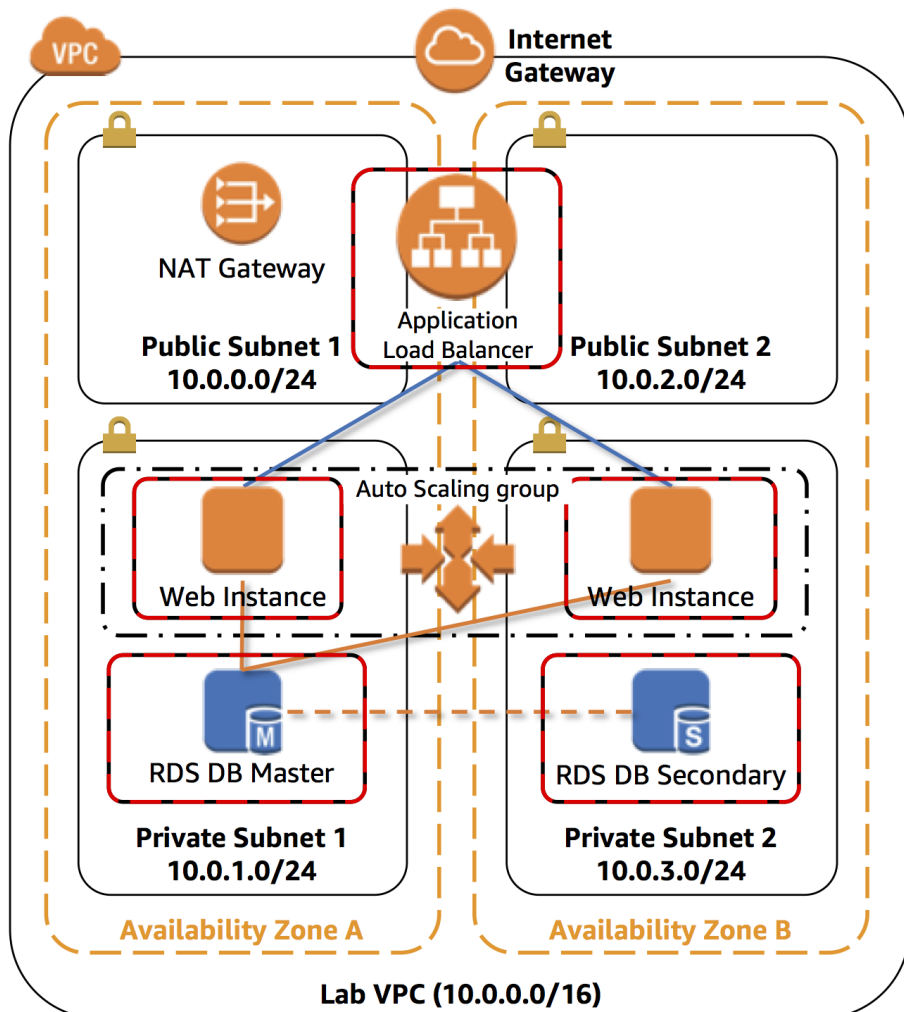
본 실습을 완료하는 데는 약 **45분** 이 소요됩니다.

## 시나리오

다음 인프라로 시작합니다.



최종 인프라 상태는 다음과 같습니다.



## 실습 시작

### 과제 1: Auto Scaling용 AMI 생성

이 과제에서는 기존의 *Web Server 1* 에서 AMI를 생성합니다. 이렇게 하면 부팅 디스크의 콘텐츠가 저장되므로 동일한 콘텐츠를 사용하여 새 인스턴스를 시작할 수 있습니다.

1. **AWS Management Console**의 **Services** 메뉴에서 **EC2**를 클릭합니다.
2. 왼쪽 탐색 창에서 **Instances**를 클릭합니다.

먼저 인스턴스가 실행 중인지 확인합니다.

3. **Web Server 1**의 **Status Checks**에 *2/2 checks passed* 가 표시될 때까지 기다립니다. 새로 고침 아이콘 을 클릭하여 업데이트합니다.

이제 이 인스턴스를 기반으로 AMI를 생성합니다.

4. **Web Server 1**을 선택합니다.
5. **Actions** 메뉴에서 **Image and templates > Create Image**를 클릭하고 다음을 구성합니다.
  - **Image name:** WebServerAMI
  - **Image description:** Lab AMI for Web Server
6. **Create Image** 를 클릭합니다.

이 AMI는 실습 후반부에서 Auto Scaling 그룹을 시작할 때 사용합니다.

### 과제 2: 로드 밸런서 생성

이 과제에서는 여러 EC2 인스턴스 및 가용 영역에 걸쳐 트래픽 균형을 조정할 수 있는 로드 밸런서를 생성합니다.

7. 왼쪽 탐색 창에서 **Load Balancers**를 클릭합니다.
8. **Create Load Balancer** 를 클릭합니다.

여러 유형의 로드 밸런서가 표시됩니다. 이 과제에서는 요청 수준(계층 7)에서 작동하는 *Application Load Balancer* 를 사용하여 요청 내용에 따라 트래픽을 대상(EC2 인스턴스, 컨테이너, IP 주소 및 Lambda 함수)으로 라우팅합니다. 자세한 내용은 로드 밸런서 비교를 참조하십시오.

9. **Application Load Balancer** 아래에서 **Create** 를 클릭하고 다음을 구성합니다.
  - **Name:** LabELB
  - **VPC:** *Lab VPC* (**Availability Zones** 섹션)
  - **Availability Zones:** 둘 다를 선택하여 사용 가능한 서브넷을 표시합니다.
  - **Public Subnet 1**과 **Public Subnet 2**를 선택합니다.

이렇게 하면 로드 밸런서가 여러 가용 영역에서 작동하도록 구성됩니다.

10. **Next: Configure Security Settings** 를 클릭합니다.

"Improve your load balancer's security." 라는 경고는 무시해도 됩니다.

11. **Next: Configure Security Groups** 를 클릭합니다.

HTTP 액세스를 허용하는 *Web Security Group* 이 이미 생성되어 있습니다.

12. **Web Security Group**을 선택하고 **default**를 선택 취소합니다.

13. **Next: Configure Routing** 을 클릭합니다.

라우팅은 로드 밸런서로 전송되는 요청을 보낼 위치를 구성합니다. Auto Scaling에 사용할 *Target Group* 을 생성해야 합니다.

14. **Name**에 LabGroup을 입력합니다.

15. **Next: Register Targets** 를 클릭합니다.

실습 후반부에서 Auto Scaling이 자동으로 인스턴스를 대상으로 등록합니다.

16. **Next: Review** 를 클릭합니다.

17. **Create** 을 클릭하고 **Close** 를 클릭합니다.

로드 밸런서의 상태가 *provisioning* 으로 표시됩니다. 준비가 될 때까지 기다리지 않아도 됩니다. 다음 과제를 진행하십시오.

### 과제 3: EC2 시작 템플릿 생성

시작 템플릿을 사용하여 Auto Scaling 그룹을 생성하려면 먼저 Amazon 머신 이미지(AMI)의 ID 및 인스턴스 유형 등 EC2 인스턴스를 시작하는 데 필요한 파라미터를 포함하는 시작 템플릿을 생성해야 합니다.

이 작업에서는 시작 템플릿을 생성합니다.

18. **Services** 메뉴에서 **EC2**를 선택합니다.

19. 왼쪽 탐색 창의 **Instances** 아래에서 **Launch Templates**를 선택합니다.

20. **Create launch template** 을 선택합니다.

21. **Launch template name and description** 섹션에서 다음을 구성합니다.

- **Launch template name:** Lab-template-NUMBER

주의 - *NUMBER* 값은 임의의 숫자를 사용하십시오.

예를 들어) *Lab-template-98469549*

만일 템플릿의 이름이 이미 존재한다고 나오면 다른 숫자를 사용하시고 다시 시도해 보십시오.

- **Template version description:** version 1

인스턴스의 루트 볼륨에 대한 템플릿이며 운영 체제, 애플리케이션 서버 및 애플리케이션을 포함할 수 있는 **Amazon Machine Image(AMI)** 를 선택하라는 메시지가 표시됩니다. AMI에서 **인스턴스**를 바로 시작할 수 있는데, 이 인스턴스는 AMI의 사본으로, 클라우드에서 실행되는 가상 서버입니다.

AMI는 다양한 버전의 Windows 및 Linux에서 사용할 수 있습니다. 이 실습에서는 *Amazon Linux* 를 실행하는 인스턴스를 시작합니다.

22. **AMI** 에서 다음을 구성합니다.

- WebServerAMI를 입력합니다.
- **WebServerAMI**를 클릭합니다.

Web Server AMI는 *My AMIs* 아래에 있습니다.

23. **Instance type**에서 *t3.micro* 를 선택합니다.

인스턴스를 시작할 때 **인스턴스 유형**에 따라 인스턴스에 할당된 하드웨어가 결정됩니다. 각 인스턴스 유형은 서로 다른 컴퓨팅, 메모리, 스토리지 용량을 제공하는데, 이 용량에 따라 서로 다른 **인스턴스 패밀리**로 분류됩니다.

24. **Security groups** 에서 *Web Security Group* 을 선택합니다.

이미 생성되어 있는 *Web Security Group* 을 사용하도록 시작 템플릿을 구성합니다.

25. **Create launch template** 을 클릭합니다.

26. **View launch templates** 를 클릭합니다.

## 과제 4: Auto Scaling 그룹 생성

Amazon EC2 Auto Scaling은 사용자가 정의한 정책, 일정 및 상태 확인에 따라 자동으로 Amazon EC2 인스턴스를 *시작* 또는 *종료* 하도록 설계된 서비스입니다. 또한 여러 가용 영역에 인스턴스를 자동으로 분산하여 애플리케이션이 고가용성을 갖도록 합니다.

이 작업에서는 Auto Scaling 그룹을 생성합니다.

27. 왼쪽 탐색 창의 **Auto Scaling** 아래에서 **Auto Scaling Groups**를 클릭합니다.

28. **Create an Auto Scaling group** 을 클릭하고 다음을 구성합니다.

- **Name:** Lab Auto Scaling Group
- **Launch template:** 작업3에서 생성한 시작 템플릿을 선택합니다.
- **Version:** 시작 템플릿 버전을 선택합니다.
- **Next** 를 클릭합니다.

29. **Configure settings** 페이지에서 다음과 같이 구성합니다.

- **Purchase options and instance types** 섹션에서 **Adhere to launch template** 를 선택합니다.
- **Network** 섹션에서 다음과 같이 구성합니다.
  - **VPC:** *Lab VPC*
  - **Subnets:** *Private Subnet 1* 과 *Private Subnet 2* 를 선택합니다.
  - **Next** 를 클릭합니다.

30. **Configure advanced options** 페이지에서 다음을 구성합니다.

- **Load - Balancing - optional** 세션에서 다음을 선택합니다.
  - **Attach to an existing load balancer.**
- **Attach to an existing load balancer** 세션에서 다음을 선택합니다.
  - **Choose from your load balancer target groups.**
  - 그리고 **Existing load balancer target groups** 에서 *LabGroup | HTTP* 을 선택 합니다.

그러면 Auto Scaling 그룹은 *\_LabGroup\_*의 대상 그룹으로 신규 EC2 인스턴스를 등록합니다. 로드 밸런서는 이 대상 그룹에 있는 인스턴스로 트래픽을 전송합니다.

- **Health checks - optional** 세션 에서 다음을 구성합니다.

- **Health check grace period: 200**

기본적으로 상태 확인 유예 기간은 300으로 설정됩니다. 이 환경은 실습 환경이므로 Auto Scaling이 첫 번째 상태 확인을 수행하는데 오래 기다릴 필요가 없도록 200으로 설정했습니다.

- **Additional settings - optional** 세션 에서 다음을 구성합니다.
  - **Monitoring:** *Enable group metrics collection within CloudWatch* 를 선택합니다.
- **Next** 를 클릭합니다.

31. **Configure group size and scaling policies** 페이지에서 다음을 구성합니다.

- **Desired capacity: 2**
- **Minimum capacity: 2**
- **Maximum capacity: 4**

이렇게 하면 Auto Scaling에서 인스턴스를 자동으로 추가/제거하여 항상 2~4개의 인스턴스를 계속해서 실행할 수 있습니다.

32. **Scaling policies** 섹션에서 다음을 수행합니다.

- **Target tracking scaling policy** 를 선택합니다.
- **Target value: 60**
- **Next** 를 클릭합니다.

이렇게 하면 Auto Scaling에서 모든 인스턴스의 평균 CPU 사용률을 60% 로 유지할 수 있습니다. Auto Scaling은 필요에 따라 용량을 자동으로 추가 또는 제거하여 지표를 지정된 대상 값으로 유지하거나 이 값에 가까운 값으로 유지합니다. 로드 패턴 변동으로 인한 지표의 변동에 맞춰 조정합니다.

33. **Add notifications** 페이지에서 **Next** 를 클릭합니다.

34. **Add Tags** 페이지에서 태그를 설정합니다.

태그를 설정하면 Auto Scaling 그룹에 이름으로 태그가 지정되고 Auto Scaling 그룹에서 시작한 EC2 인스턴스에도 표시됩니다. 그러면 어떤 애플리케이션에 어떤 EC2 인스턴스가 연결되어 있는지 더 쉽게 식별할 수 있습니다. 또한 비용 센터 같은 태그를 추가하면 결제 파일에서 애플리케이션 비용을 쉽게 지정할 수도 있습니다.

35. **Add tag** 를 클릭하고 다음을 구성합니다.

- **Key: Name**
- **Value: Lab Instance**
- **Next** 를 클릭합니다.

36. Auto Scaling 그룹의 세부 정보를 확인한 다음 **Create Auto Scaling group** 을 클릭합니다.

애플리케이션이 곧 2개의 가용 영역에서 실행될 것이며 Auto Scaling이 인스턴스 또는 가용 영역에서 장애가 발생하는 경우에도 해당 구성을 유지합니다.

이제 Auto Scaling 그룹을 생성했으며 해당 그룹에서 EC2 인스턴스를 시작했는지 확인할 준비가 완료되었습니다.

37. Auto Scaling 그룹을 클릭합니다.

**Group Details**를 검토하여 Auto Scaling 그룹에 대한 정보를 확인합니다.

38. **Activity** 탭을 클릭합니다.

Status 열에는 인스턴스의 현재 상태가 포함됩니다. 인스턴스를 시작하면 상태 열에 *PreInService* 가 표시됩니다. 인스턴스가 시작되면 상태가 *Successful* 로 변경됩니다. 새로 고침 버튼을 클릭하여 인스턴스의 현재 상태를 볼 수도 있습니다.

39. **Instance management** 탭을 클릭합니다.

Auto Scaling 그룹에서 EC2 인스턴스를 시작했으며 해당 인스턴스가 *InService* 수명 주기 상태에 있음을 알 수 있습니다. Health Status 열에 해당 인스턴스에 대한 EC2 인스턴스 상태 확인 결과가 표시됩니다.

40. **Monitoring** 탭을 클릭합니다. 여기에서 Autoscaling 그룹에 대한 모니터링 관련 정보를 볼 수 있습니다.

## 과제 5: 로드 밸런싱 작동 확인

이 과제에서는 로드 밸런싱이 올바르게 작동하는지 확인합니다.

41. 왼쪽 탐색 창에서 **Instances**를 클릭합니다.

이름이 **Lab Instance**인 새 인스턴스 2개가 표시됩니다. 이 두 인스턴스는 Auto Scaling에 의해 시작되었습니다.

인스턴스 또는 이름이 표시되지 않으면 30초간 기다린 후 오른쪽 위에서 새로 고침 을 클릭합니다.

먼저 새 인스턴스가 상태 확인을 통과했는지 확인합니다.

42. 왼쪽 탐색 창에서 **Target Groups**(*Load Balancing* 섹션)를 클릭합니다.

43. **LabGroup**을 클릭합니다.

44. **Targets** 탭을 클릭합니다.

이 대상 그룹에는 두 개의 **Lab Instance**가 대상으로 나열되어야 합니다.

45. 두 개의 인스턴스 **Status**가 모두 *healthy* 로 전환될 때까지 기다립니다. 오른쪽 위에서 새로 고침 아이콘 을 클릭하여 업데이트를 확인합니다.

*Healthy* 는 인스턴스가 로드 밸런서의 상태 확인을 통과했음을 나타냅니다. 즉, 로드 밸런서가 인스턴스로 트래픽을 전송합니다.

이제 로드 밸런서를 통해 Auto Scaling 그룹에 액세스할 수 있습니다.

46. 왼쪽 탐색 창에서 **Load Balancers**를 클릭합니다.

47. **LabELB** 를 선택하지 않은 경우 클릭해서 선택합니다.

48. 하단 창에서 로드 밸런서의 **DNS 이름**을 복사합니다. "(A Record)"를 생략해야 합니다.

이름은 [LabELB-1998580470.us-west-2.elb.amazonaws.com](https://LabELB-1998580470.us-west-2.elb.amazonaws.com) 과 유사해야 합니다.

49. 새 웹 브라우저 탭을 열고 방금 복사한 DNS 이름을 붙여 넣은 다음 Enter 키를 누릅니다.

애플리케이션이 브라우저에 나타납니다. 이는 로드 밸런서가 요청을 수신하여 EC2 인스턴스 중 하나로 전송한 후 결과를 되돌려 준 것을 말합니다.

## 과제 6: Auto Scaling 테스트

인스턴스가 최소 2개, 최대 4개인 Auto Scaling 그룹을 생성했습니다. 최소 크기는 2이고 부하가 전혀 없으므로 현재 2개의 인스턴스가 실행 중입니다. 이제 부하를 늘려 Auto Scaling이 인스턴스를 추가하도록 합니다.

50. 곧 다시 돌아올 것이므로 애플리케이션 탭을 닫지 않은 상태에서 AWS Management Console로 돌아갑니다.

51. **Services** 메뉴에서 **CloudWatch**를 클릭합니다.

52. 왼쪽 탐색 창에서 **Alarms(ALARM 아람)**를 클릭합니다.

두 경보가 표시됩니다. 이러한 경보는 Auto Scaling 그룹에 의해 자동으로 생성되었습니다. 이 두 경보는 2~4개의 인스턴스를 보유해야 하는 제한 범위를 벗어나지 않으면서 자동으로 평균 CPU 부하를 60%에 가깝게 유지합니다.

53. 이름에 *AlarmHigh* 가 포함된 **OK** 경보를 클릭합니다.

**OK**가 표시된 경보가 없으면 1분 정도 기다린 다음 경보 상태가 변경될 때까지 오른쪽 상단의 새로 고침 을 클릭합니다.

**OK**는 경보가 트리거되지 않았음을 나타냅니다. 이 경보는 평균 CPU가 높을 경우 인스턴스를 추가하는 **CPU Utilization > 60**에 대한 경보입니다. 현재 차트에 표시된 CPU 사용률은 매우 낮아야 합니다.

이제 CPU 사용률을 높이는 계산을 수행하도록 애플리케이션에 지시합니다.

54. 웹 애플리케이션이 있는 브라우저 탭으로 돌아갑니다.

55. AWS 로고 옆의 **Load Test**를 클릭합니다.

그러면 애플리케이션에서 높은 로드가 생성됩니다. 브라우저 페이지가 자동으로 새로 고쳐지고 Auto Scaling 그룹의 모든 인스턴스에서 로드가 생성됩니다. 이 탭을 닫지 마십시오.

56. **CloudWatch** 콘솔이 있는 브라우저 탭으로 돌아갑니다.

5분 안에 **AlarmLow** 경보가 **OK** 로 변경되고 **AlarmHigh** 경보 상태가 *In alarm* 으로 변경됩니다.

오른쪽 위에서 60초 간격으로 새로 고침 을 클릭하여 디스플레이를 업데이트할 수 있습니다.

CPU 비율 증가를 나타내는 **AlarmHigh** 차트가 표시됩니다. 인스턴스가 3분 이상 60% 선을 넘으면 Auto Scaling이 트리거되어 추가 인스턴스가 추가됩니다.

57. **AlarmHigh** 경보의 상태가 *In alarm* 상태가 될 때까지 기다립니다.

이제 시작된 추가 인스턴스를 볼 수 있습니다.

58. **Services** 메뉴에서 **EC2**를 클릭합니다.

59. 왼쪽 탐색 창에서 **Instances**를 클릭합니다.

**Lab Instance** 레이블이 지정된 3개 이상의 인스턴스가 실행 중인 것으로 표시됩니다. 새 인스턴스는 경보에 대한 응답으로 Auto Scaling에 의해 생성되었습니다.

## 과제 7: Web Server 1 종료

이 과제에서는 *Web Server 1* 을 종료합니다. 이 인스턴스는 Auto Scaling 그룹에서 사용할 AMI를 생성하는 데 사용되었지만 더 이상 필요하지 않습니다.

60. **Web Server 1**을 선택하고 이 인스턴스만 선택되었는지 확인합니다.

61. **Instance state** 메뉴에서 **Terminate Instance**를 클릭합니다.

62. **Terminate** 를 클릭합니다.

실습 완료



축하합니다! 실습을 마치셨습니다.