

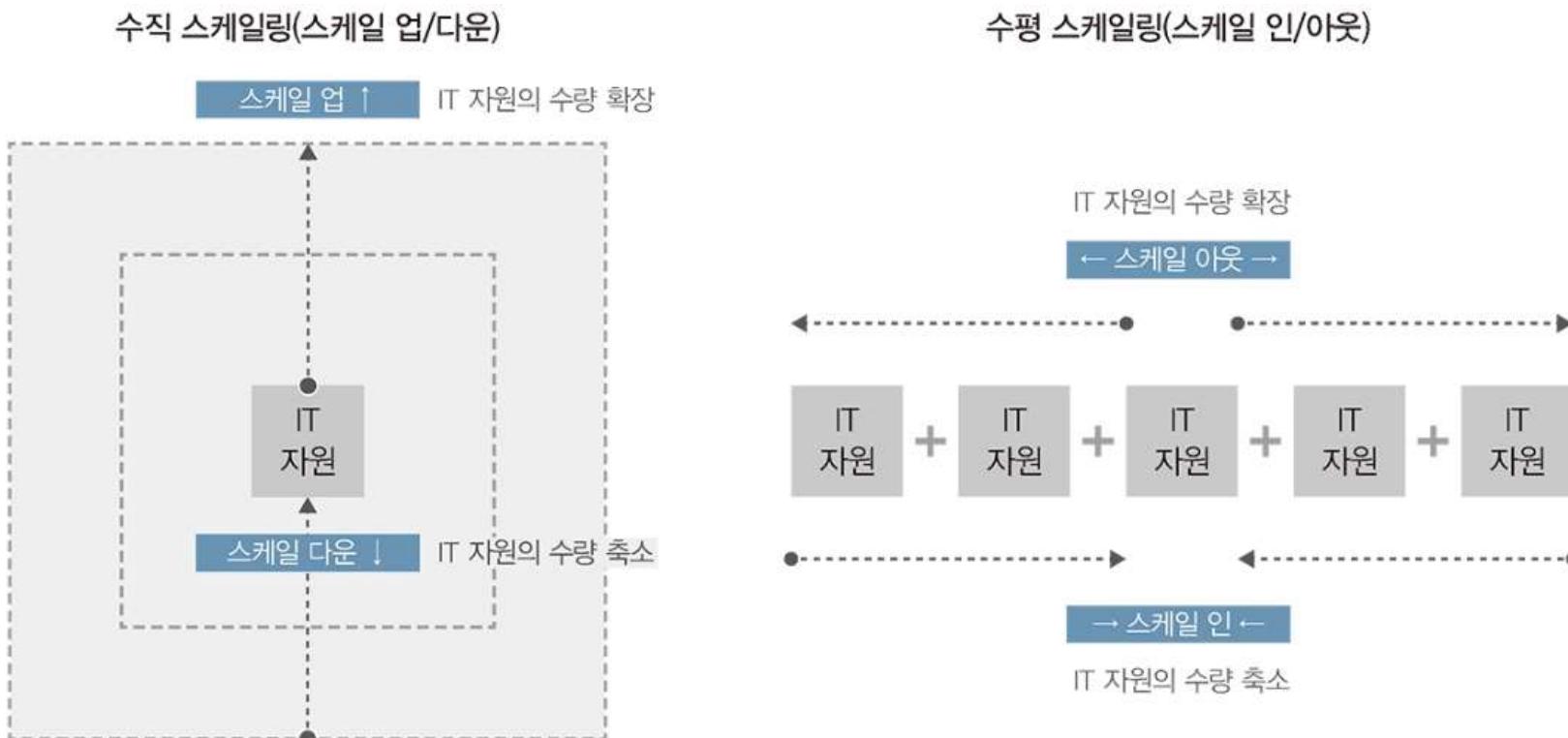
클라우드컴퓨팅



Amazon AWS 활용: 오토스케일링 서비스

스케일링(Scaling)

- 스케일링(Scaling)
 - IT 자원을 확장하거나 축소하는 기능
 - IT 자원의 사용 요구가 가변적일수록 자원의 성능 및 수량을 확장/축소하는 스케일링은 기능은 필수
 - 스케일링의 종류 ; 수직(vertical) vs 수평(horizontal)



AWS 오토 스케일링

- AWS 오토 스케일링

- AWS 리소스를 동적으로 확장하거나 축소하여 클라우드 자원의 규모를 자동으로 조정하는 기능
- AWS에서 제공하는 리소스 중 상당수는 오토 스케일링 기능을 통해 동적으로 규모를 조정할 수 있음



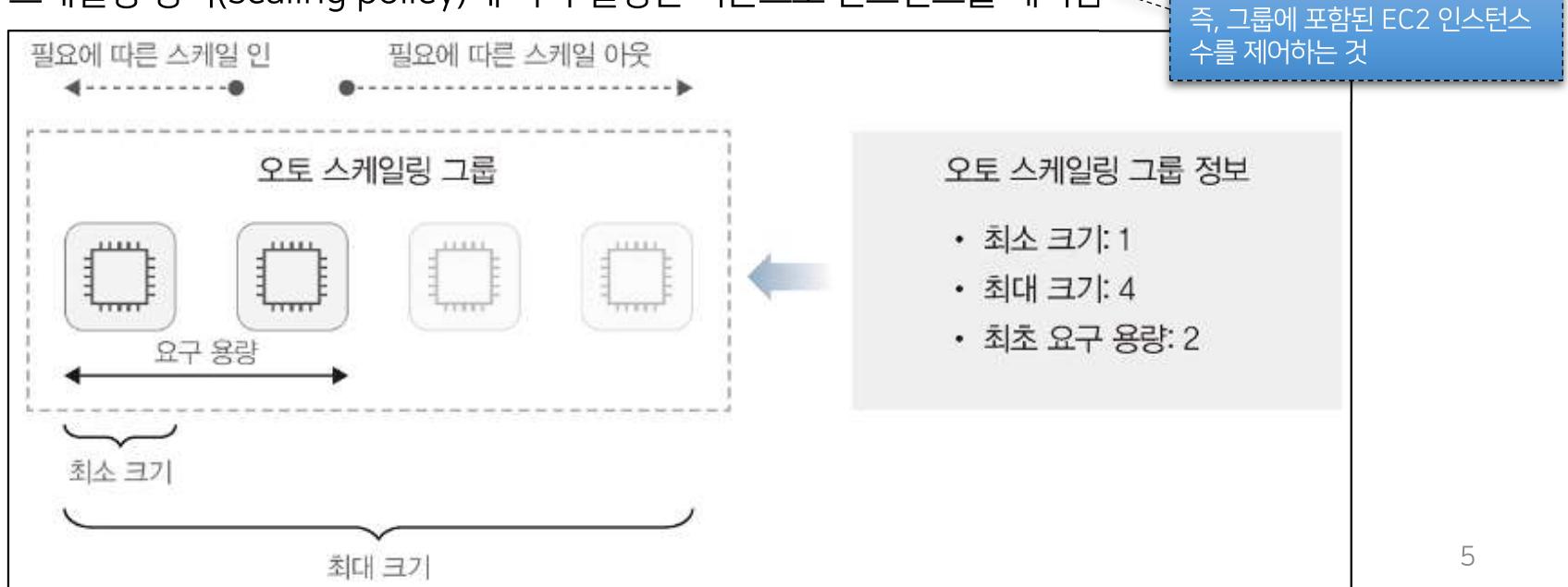
AWS 오토 스케일링

- EC2 오토 스케일링

- EC2 인스턴스에 설치된 어플리케이션 워크로드를 수용하도록 동적으로 EC2 인스턴스 수를 확장하거나 축소하여 워크로드를 유지하는 서비스
 - 즉, EC2 인스턴스에 대한 수평 스케일링(OUT or IN)을 동적으로 수행함
 - 참고) 워크로드 ; 일반적으로, 주어진 시간 안에 컴퓨터 시스템이 처리해야 하는 작업량을 의미

- 동작 방식

- '오토 스케일링 그룹'이라는 EC2 인스턴스 모음(그룹)을 생성
- 최소 및 최대 인스턴스 수의 범위를 지정하고, 범위 내에서 인스턴스 수를 동적으로 조절
- 스케일링 정책(scaling policy)에 따라 설정한 기준으로 인스턴스를 제어함

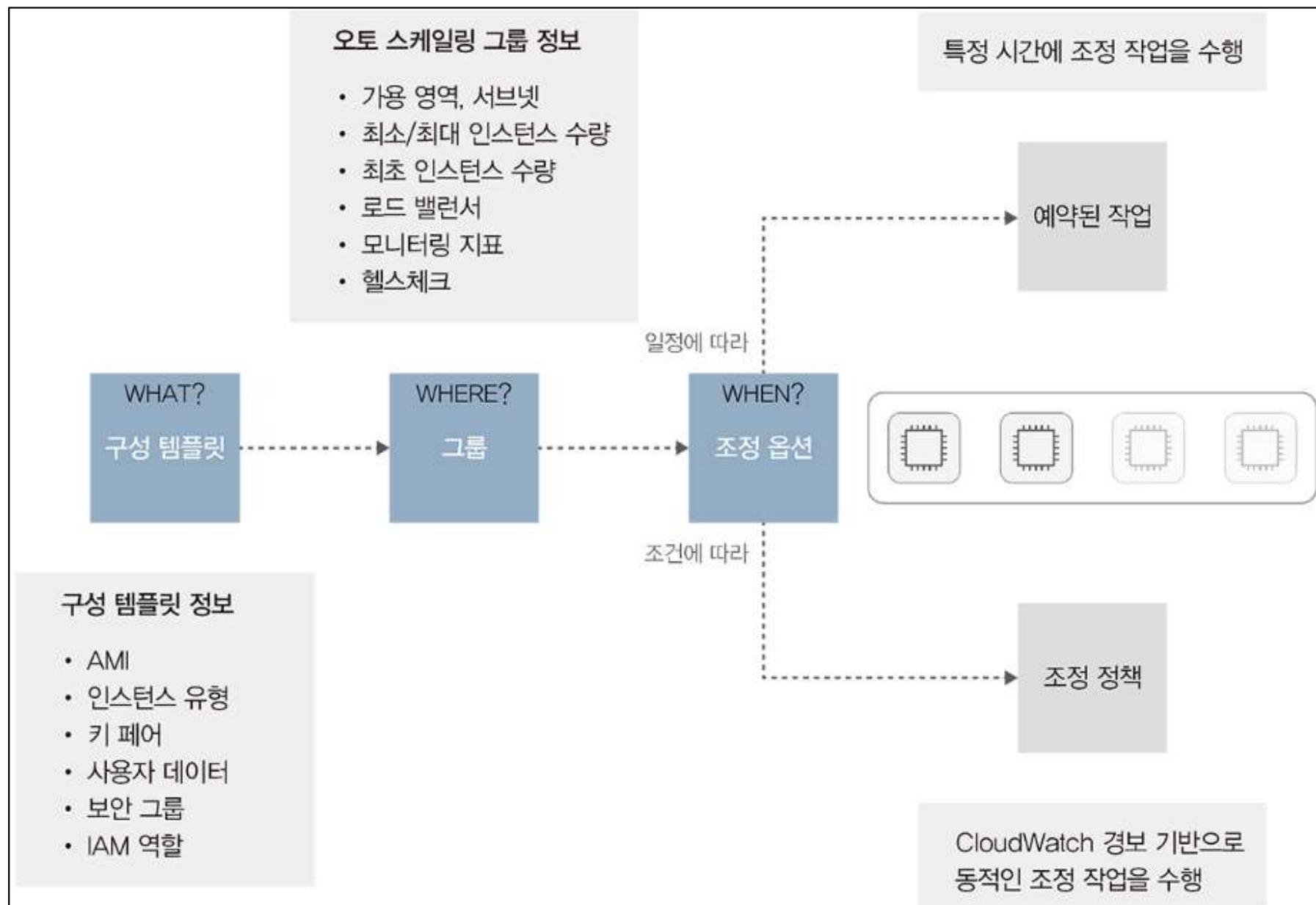


AWS 오토 스케일링

- EC2 오토 스케일링 구성 요소
 - 그룹
 - EC2 인스턴스를 오토 스케일링으로 조정하고 관리하기 위해 EC2 인스턴스를 논리적으로 구분하는 그룹을 생성
 - 그룹을 생성할 때, EC2 인스턴스의 최소 및 최대 인스턴스 수와, 최초 인스턴스 수 등을 지정
 - 구성 템플릿
 - EC2 인스턴스를 구성하는 템플릿으로 인스턴스 AMI, 인스턴스 유형, 키 페어, 보안 그룹 등을 지정
 - 즉, 오토 스케일링을 통해 생성될 EC2 인스턴스의 구성을 정의하는 것 (설치될 OS, SW 등…)
 - 조정 옵션
 - 오토 스케일링 그룹을 조정하는 방법을 정의하는 것 (= Scaling Rule)
 - 지정한 조건이나 일정에 따라 오토 스케일링 그룹을 조정할 수 있음

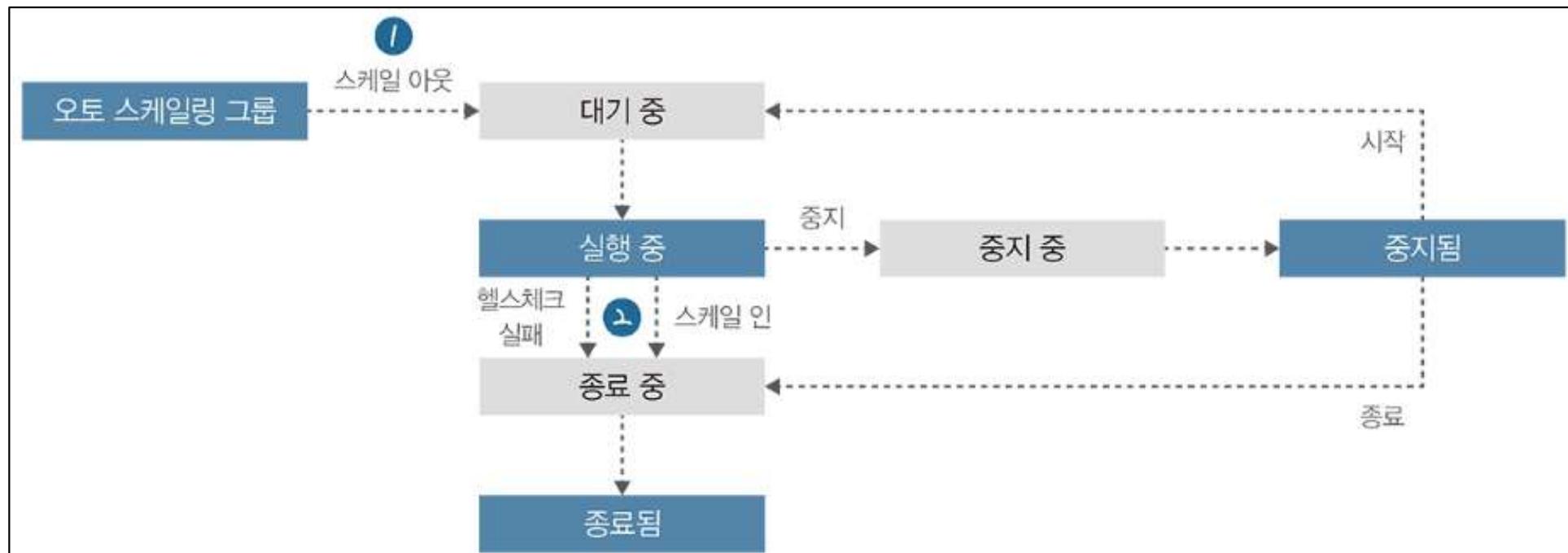
AWS 오토 스케일링

- EC2 오토 스케일링 구성 요소 ; 구성 템플릿, 그룹, 조정 옵션



AWS 오토 스케일링

- EC2 오토 스케일링 ; 인스턴스 수명 주기(life cycle)
 - EC2 오토 스케일링 그룹에 속한 EC2 인스턴스의 수명 주기
 - 아래 그림에서
 - (1)번은 인스턴스 확장 이벤트이고, (다음 페이지…)
 - (2)번은 인스턴스 축소 이벤트에 해당함 (다음 다음 페이지…)

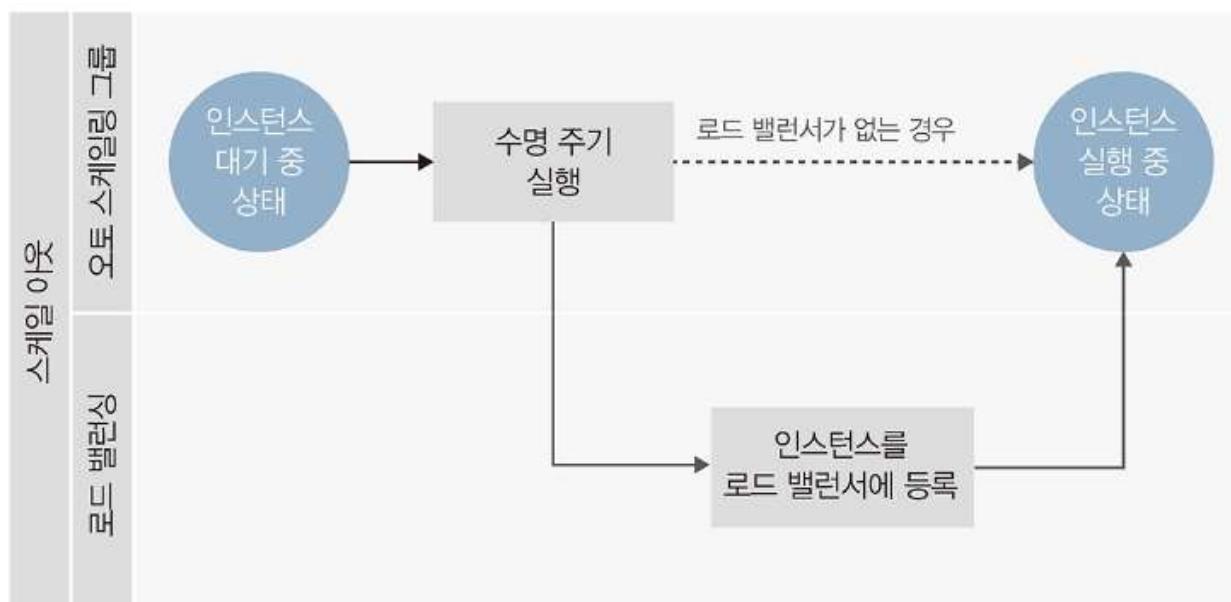


AWS 오토 스케일링

- EC2 오토 스케일링 ; 인스턴스 수명 주기(life cycle)

- (1) 인스턴스 확장 이벤트

- 인스턴스 확장 이벤트(= 인스턴스 수를 늘리는 경우)가 발생하는 경우:
 - 오토 스케일링 그룹의 크기를 수동으로 늘리는 경우
 - 스케일 아웃(Scale-Out) 정책 사용 & 임계치 이상으로 부하가 증가한 경우 (실시간 확장)
 - 특정 시간에 오토 스케일링 그룹의 크기를 자동으로 늘리는 예약된 작업을 수행한 경우
 - 인스턴스 확장 이벤트가 발생하면,
 - 오토 스케일링은 할당된 구성 템플릿을 사용하여 필요한 수의 인스턴스를 시작함
 - 이때, 인스턴스는 [대기 중] 상태로 시작됨.
 - 인스턴스가 완전히 구성되면 로드 밸런서(LB)에 등록되고, [실행 중] 상태로 전환됨



AWS 오토 스케일링

- EC2 오토 스케일링 ; 인스턴스 수명 주기(life cycle)

- (2) 인스턴스 축소 이벤트

- 인스턴스 축소(= 인스턴스 수를 줄이기) 이벤트가 발생하는 경우:
 - 오토 스케일링 그룹의 크기를 수동으로 줄이는 경우
 - 스케일 인(Scale-In) 정책 사용 & 임계치 이하로 부하 감소 (실시간 축소)
 - 특정 시간에 오토 스케일링 그룹의 크기를 줄이는 예약된 작업을 수행하는 경우
 - 인스턴스 축소 이벤트가 발생하면,
 - 오토 스케일링은 하나 이상의 인스턴스를 종료함
 - 이 때, 오토 스케일링은 종료 정책을 이용하여 종료할 인스턴스를 결정함(예: 가장 최근에 만들어진 EC2 우선 종료 등)
 - 선택된 인스턴스는 종료되고, 오토 스케일링에서 삭제되며, 로드 밸런서의 연결이 해제됨
 - 단, 미처 완료하지 못한 작업을 완료하기 위해, 즉시 종료되지 않고 일정 시간을 대기한 후 종료되기도 하는데 이러한 상태를 드레이닝(draining)이라고 함



AWS 오토 스케일링

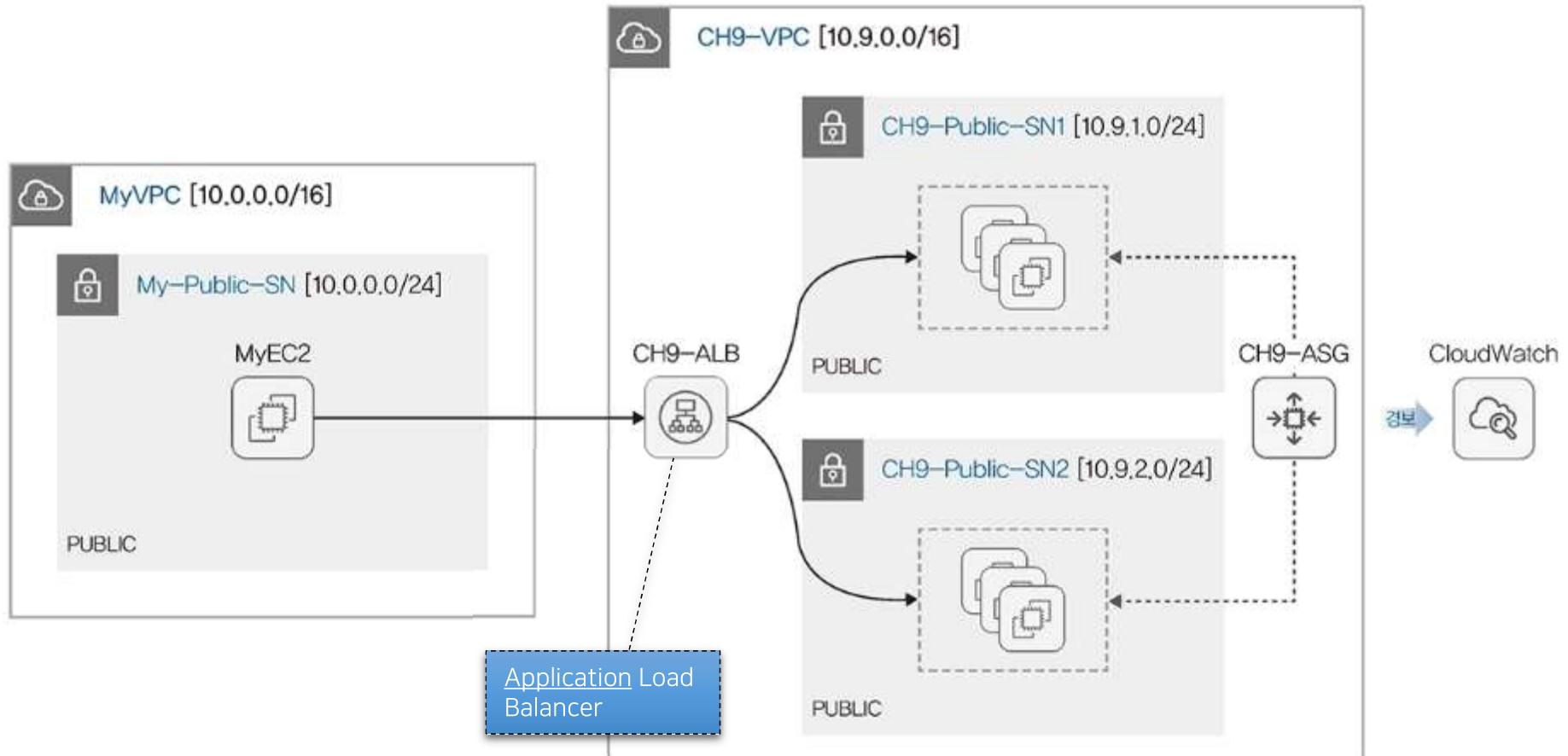
- AWS EC2 오토 스케일링 조정 옵션
 - 인스턴스를 일정 수로 유지
 - (일반적인 오토 스케일링과는 다름)
 - 오토 스케일링 그룹의 인스턴스 수를 지정된 수로 유지하는 옵션 (확장/축소의 개념이 아님)
 - 실행 중인 인스턴스의 상태를 주기적으로 확인하고, 비정상/장애 상태의 인스턴스가 발견되면 해당 인스턴스를 종료하고 새로운 인스턴스를 실행하여 인스턴스 수를 일정하게 유지
 - 수동 조정
 - 오토 스케일링 그룹의 용량을 수동으로 변경
 - 동적 조정
 - 실시간 트래픽/부하 변화에 따라 오토 스케일링 그룹의 용량을 동적으로 조정하는 것으로, 아래의 동적 조정 정책을 지원함
 - 대상 추적 조정 ;
 - 특정 대상 지표의 목표 값을 유지하기 위해 오토 스케일링을 적용
 - 예) CPU 사용률 50% 유지를 목표로 인스턴스 수를 조정
 - 단계 조정 (다음 페이지...)

AWS 오토 스케일링

- AWS EC2 오토 스케일링 조정 옵션
 - 동적 조정
 - 오토 스케일링 그룹의 용량을 단계별로 조정하여 오토 스케일링을 적용
 - 임곗값을 기준으로 임곗값을 넘어서는 정도에 따라 단계적으로 조정
 - 단순 조정
 - 오토 스케일링 그룹의 용량을 단일하게 조정하여 오토 스케일링을 적용
 - 임곗값을 기준으로 임곗값을 넘어서는 정도와 무관하게 단일 규칙으로 조정(예: 임곗값 초과 시 인스턴스 +2)
 - 일정을 기반으로 예약된 조정
 - 시간/날짜에 따라 오토 스케일링을 적용
 - 예) 일과시간에는 인스턴스 수를 늘리고, 그 외의 시간에는 인스턴스 수를 줄임
 - 사전 예방적 조정(패턴 기반의 예측된 조정)
 - 과거 기록을 기반으로 트래픽/부하의 패턴을 예측하고, 이를 기반으로 인스턴스 수를 미리 조정

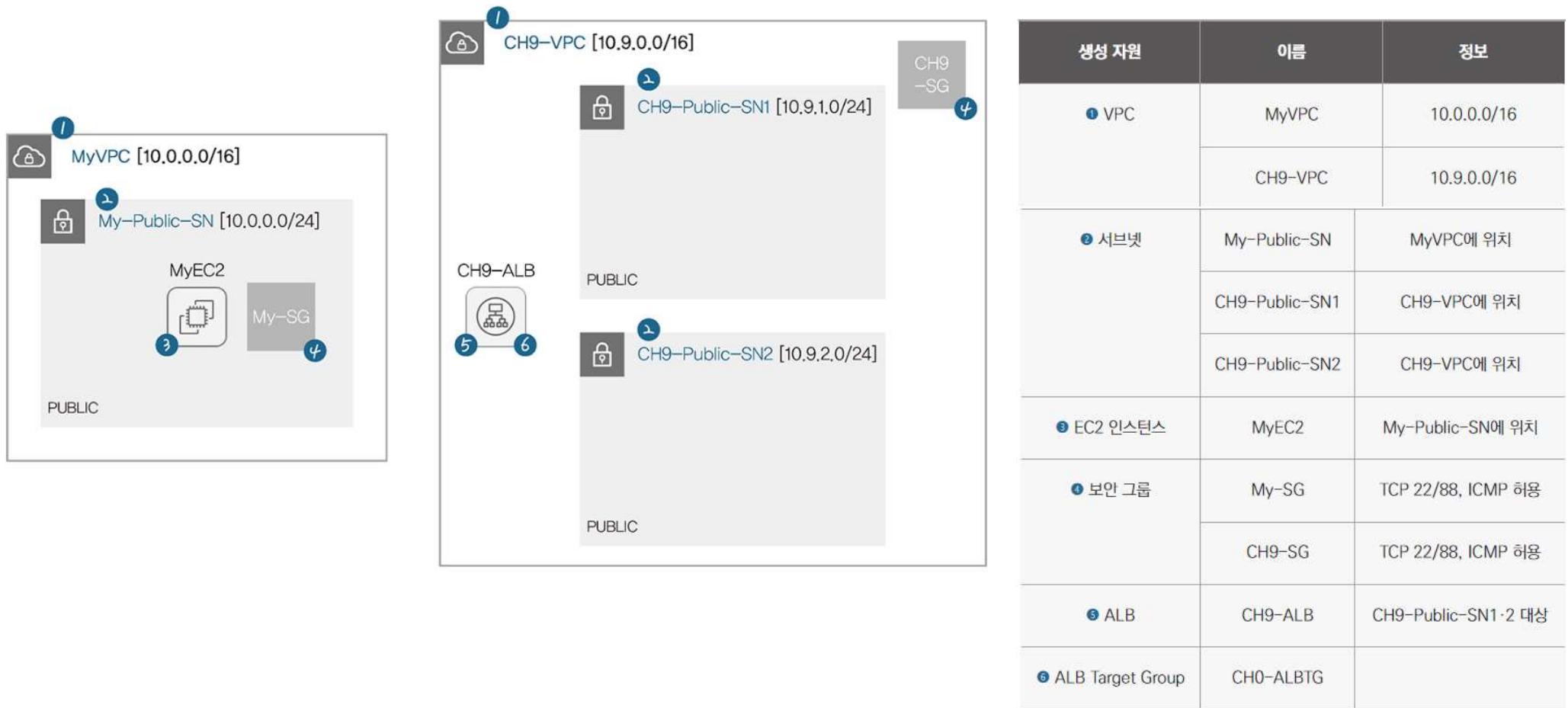
AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - 목표 구성도



AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - CloudFormation을 이용한 자원 배포



AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - AWS CLI 도구를 사용해서 MyEC2 인스턴스 정보 조회
 - 먼저, CloudFormation에서 정의한 MyEC2 리소스 설정 값 확인

MyEC2:

```
Type: AWS::EC2::Instance
Properties:
  InstanceType: !Ref InstanceType
  IamInstanceProfile : ASLabInstanceProfile
  ImageId: !Ref LatestAmiId
  KeyName: !Ref KeyName
Tags:
  - Key: Name
    Value: MyEC2
NetworkInterfaces:
  - DeviceIndex: 0
    SubnetId: !Ref MyPublicSN
    GroupSet:
      - !Ref MySG
    AssociatePublicIpAddress: true
UserData:
  Fn::Base64:
    !Sub |
      #!/bin/bash
      hostnamectl --static set-hostname MyEC2
      yum install httpd tmux -y
      curl "https://awscli.amazonaws.com/awscli-exe-linux-x86\_64.zip" -o "awscliv2.zip"
      unzip awscliv2.zip
      ./aws/install --bin-dir /usr/bin --install-dir /usr/bin --update
      sed -i "s/UTC/Asia\Seoul/g" /etc/sysconfig/clock
      ln -sf /usr/share/zoneinfo/Asia/Seoul /etc/localtime
```

AWS CLI 다운로드
및 설치

AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - AWS CLI 도구를 사용해서 MyEC2 인스턴스 정보 조회
 - AWS CLI 도구 사용하기

```
# AWS CLI 툴 버전 확인  
  
aws --version  
  
aws-cli/2.10.3 Python/3.9.11 Linux/4.14.304-226.531.amzn2.x86_64 exe/x86_64.amzn.  
2 prompt/off
```

```
# AWS CLI로 생성된 인스턴스 정보 확인(연속)  
  
while true; do aws ec2 describe-instances --query 'Reservations[*].Instances[*].[InstanceId, State.Name, PrivateIpAddress]' --output text; date; sleep 1; done  
  
i-0a88fc7beed1a855f    running 10.0.0.163  
  
Tue Feb 28 01:49:37 KST 2023  
  
...
```

AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [1] EC2 인스턴스 시작 템플릿 생성하기
 - Amazon EC2 오토 스케일링 서비스를 통해서 추가로 생성될 EC2 인스턴스의 구성을 정의하기 위해, 인스턴스에 대한 시작 템플릿을 구성



AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [1] EC2 인스턴스 시작 템플릿 생성하기

시작 템플릿 이름 및 설명

시작 템플릿 이름 - 필수

EC2Template — **입력 1**

이 계정에 대해 고유해야 합니다. 최대 128자입니다. '&', '*', '@' 등의 특수 문자나 공백은 사용할 수 없습니다.

템플릿 버전 설명

EC2 Auto Scaling — **입력 2**

최대 255자

3 체크 Scaling 지침 정보
EC2 Auto Scaling에 이 템플릿을 사용하려면 이 항목을 선택합니다.
 EC2 Auto Scaling에 사용할 수 있는 템플릿을 설정하는 데 도움이 되는 지침 제공

AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [1] EC2 인스턴스 시작 템플릿 생성하기

▼ 애플리케이션 및 OS 이미지(Amazon Machine Image) - 필수 정보

AMI는 인스턴스를 시작하는데 필요한 소프트웨어 구성(운영 체제, 애플리케이션 서버 및 애플리케이션)이 포함된 템플릿입니다. 아래에서 찾고 있는 항목이 보이지 않으면 AMI를 검색하거나 찾아보십시오.

수천 개의 애플리케이션 및 OS 이미지를 포함하는 전체 카탈로그 검색

④ 클릭

⑤ 선택

Quick Start

Amazon Linux

macOS

Ubuntu

Windows

Red Hat

S

Amazon Machine Image(AMI)

⑥ 선택

Browse more AMIs

Including AMIs from AWS, Marketplace and the Community

Amazon Linux 2 AMI (HVM) - Kernel 5.10, SSD Volume Type

ami-0f6e451b865011317 (64비트(x86)) / ami-09e04c9fc45f2e3b (64비트(Arm))

가상화: hvm ENA 활성화됨: true 루트 디바이스 유형: ebs

프리 티어 사용 가능

19

AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [1] EC2 인스턴스 시작 템플릿 생성하기

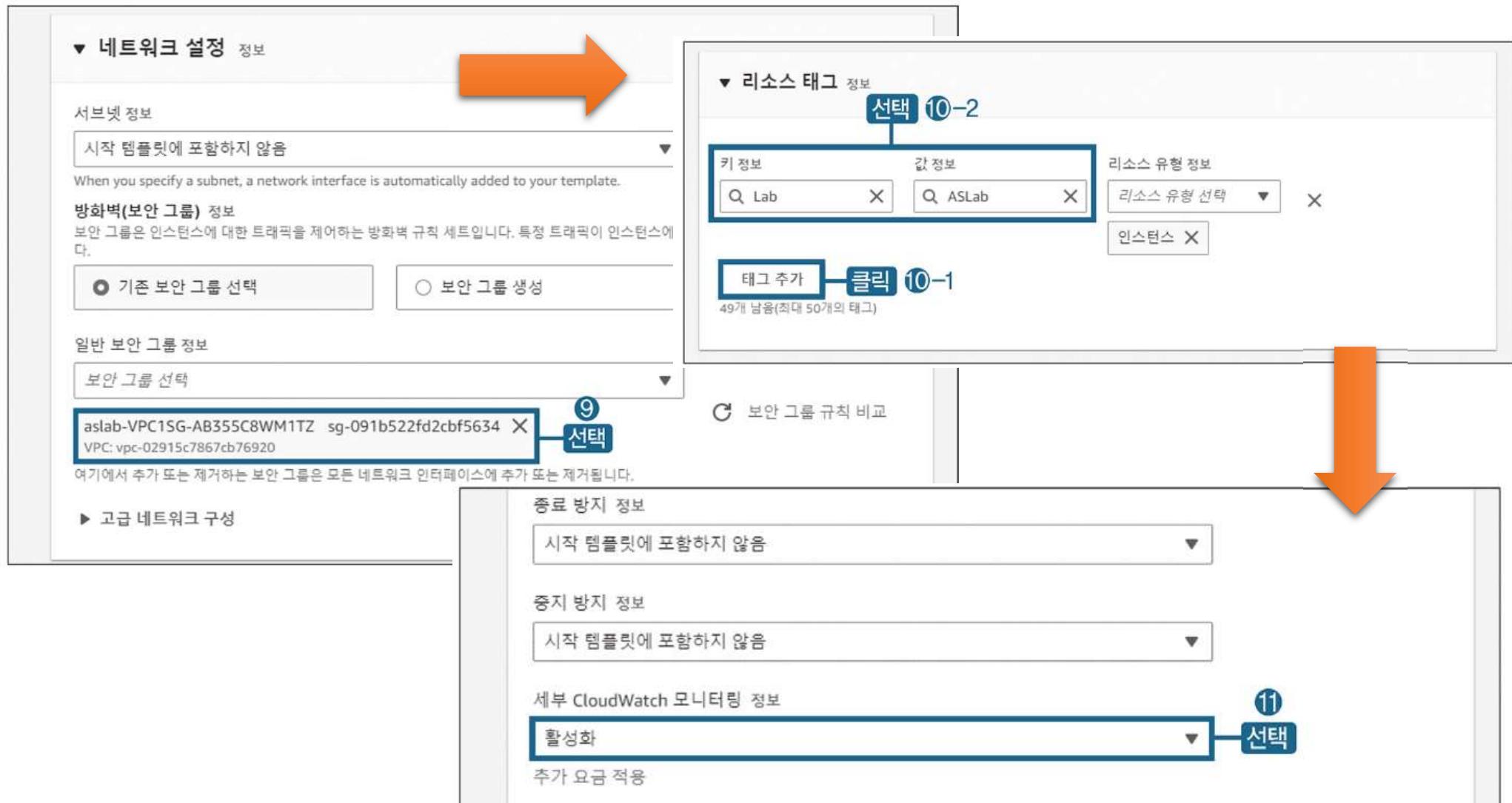
The screenshot shows the AWS CloudFormation console interface for creating a new stack. The top section is titled 'AWS 템플릿' (AWS Template) and shows the 'aws:CloudFormation:Stack' template. Below it, the 'AWS 템플릿' section displays the CloudFormation template code.

7 선택 (Select) is highlighted over the 't2.micro' instance type in the '인스턴스 유형' (Instance Type) section. This section also includes details about the instance: 1 vCPU, 1 GiB 메모리, and pricing information for RHEL, Linux, SUSE, and Windows operating systems.

8 선택 (Select) is highlighted over the 'ongja_key' key pair name in the '키 페어(로그인)' (Key Pair (Login)) section. This section also includes a '새 키 페어 생성' (Create New Key Pair) button.

AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [1] EC2 인스턴스 시작 템플릿 생성하기



AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [1] EC2 인스턴스 시작 템플릿 생성하기

사용자 데이터 - *optional* 정보

필드에 사용자 데이터를 입력합니다.

SHELL SCRIPT 와 같은 방식으로 EC2 인스턴스 이미지를 설정

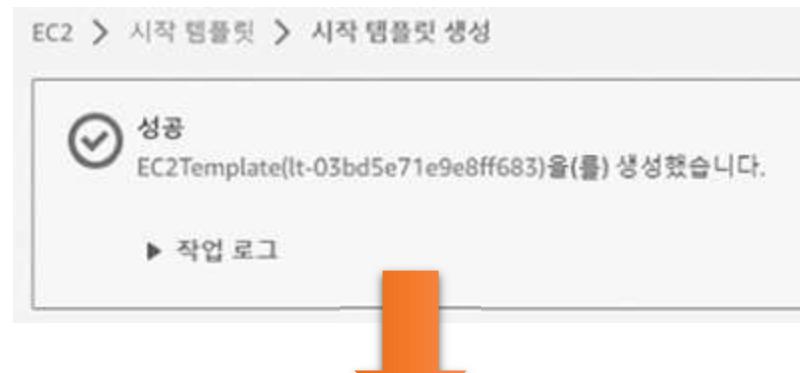
```
#!/bin/bash
RZAZ=`curl http://169.254.169.254/latest/meta-data/placement/availability-
zone-id`
IID=`curl 169.254.169.254/latest/meta-data/instance-id`
LIP=`curl 169.254.169.254/latest/meta-data/local-ipv4`
amazon-linux-extras install -y php8.0
yum install httpd htop tmux -y
systemctl start httpd && systemctl enable httpd
echo "<h1>RegionAz($RZAZ) : Instance ID($IID) : Private IP($LIP) : Web
Server</h1>" > /var/www/html/index.html
echo "1" > /var/www/html/HealthCheck.txt
curl -o /var/www/html/load.php https://cloudneta-book.s3.ap-northeast-
2.amazonaws.com/chapter5/load.php --silent
```

12

입력

AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [1] EC2 인스턴스 시작 템플릿 생성하기
 - 생성 결과 확인



The screenshot displays the EC2 Start Template list and a detailed view of a specific template.

EC2 > 시작 템플릿

시작 템플릿 (1/1) 정보

시작 템플릿 ID	시작 템플릿 이름	기본 버전	최신 버전	생성 시간
lt-0771d5ab9e6868379	EC2Template	1	1	2023-02-27T18:11:13.000Z

EC2Template (lt-0771d5ab9e6868379)

시작 템플릿 세부 정보

시작 템플릿 ID lt-0771d5ab9e6868379	시작 템플릿 이름 EC2Template	기본 버전 1	소유자 arn:aws:iam::831684716714:root
-----------------------------------	--------------------------	------------	---------------------------------------

AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [2] 오토 스케일링 그룹 생성하기



AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [2] 오토 스케일링 그룹 생성하기

EC2 > Auto Scaling 그룹 > Auto Scaling 그룹 생성

1단계
시작 템플릿 또는 구성 선택

2단계
인스턴스 시작 옵션 선택

3단계 - 선택 사항
고급 옵션 구성

4단계 - 선택 사항
그룹 크기 및 크기 조정 정책 구성

5단계 - 선택 사항
알림 추가

6단계 - 선택 사항
태그 추가

7단계
검토

시작 템플릿 또는 구성 선택 Info

이 Auto Scaling 그룹에서 시작하는 모든 EC2 인스턴스에 공통된 설정이 포함된 시작 템플릿을 지정합니다. 현재 시작 구성 사용하는 경우 시작 템플릿으로 마이그레이션할 수 있습니다.

이름

Auto Scaling 그룹 이름
그룹을 식별할 이름을 입력합니다.

FirstEC2ASG **입력 1**

현재 리전에서 이 계정에 대해 고유해야 하며 255자를 넘지 않아야 합니다.

시작 템플릿 Info

시작 템플릿

Amazon Machine Image(AMI), 인스턴스 유형, 키 페어 및 보안 그룹과 같은 인스턴스 수준 설정이 포함된 시작 템플릿을 선택합니다.

EC2Template **선택 2**

시작 템플릿 생성 □

버전

직전에 생성한 템플릿 AMI를 사용

시작 구성으로 전환

AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [2] 오토 스케일링 그룹 생성하기 ; 오토 스케일링이 적용될 네트워크 설정

EC2 > Auto Scaling 그룹 > Auto Scaling 그룹 생성

1단계
시작 템플릿 또는 구성 선택

2단계
인스턴스 시작 옵션 선택

3단계 - 선택 사항
고급 옵션 구성

4단계 - 선택 사항
그룹 크기 및 크기 조정 정책 구성

5단계 - 선택 사항
알림 추가

6단계 - 선택 사항
태그 추가

7단계
검토

인스턴스 시작 옵션 선택 Info

인스턴스가 시작되는 VPC 네트워크 환경을 선택하고 인스턴스 유형과 구매 옵션을 사용자 지정합니다.

네트워크 Info

대부분의 애플리케이션에서는 여러 가용 영역을 사용할 수 있으며 EC2 Auto Scaling이 여러 영역 간에 인스턴스를 균일하게 분산할 수 있습니다. 기본 VPC와 기본 서브넷은 빠르게 시작하는 데 적합합니다.

VPC

Auto Scaling 그룹의 가상 네트워크를 정의하는 VPC를 선택합니다.

vpc-02915c7867cb76920 (CH9-VPC)
10.9.0.0/16

선택

VPC 생성

가용 영역 및 서브넷

선택한 VPC에서 Auto Scaling 그룹이 사용할 수 있는 가용 영역과 서브넷을 정의합니다.

가용 영역 및 서브넷 선택	X
ap-northeast-2a subnet-060c3248600730a69	X
<input checked="" type="checkbox"/> (CH9-Public-SN1)	
10.9.1.0/24	
ap-northeast-2c subnet-0e5e021894fc0f93f	X
<input checked="" type="checkbox"/> (CH9-Public-SN2)	
10.9.2.0/24	
10.9.2.0/24	

선택

서브넷 생성

AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [2] 오토 스케일링 그룹 생성하기 ; 로드 밸런서에 연결 (분산처리 기능 활성화)

The screenshot shows the AWS Auto Scaling 'Create New Auto Scaling Group' wizard, Step 2: Load Balancing - Selection. The left sidebar lists steps 1 through 7. Step 2 is currently selected, showing the 'Load Balancing - Selection' page.

고급 옵션 구성 - 선택 사항

안정성을 높이고 쉽게 크기 조정할 수 있도록 애플리케이션에 대한 수신 트래픽을 여러 인스턴스에 분산하려면 로드 밸런서를 선택합니다. 상태 확인 교체 및 모니터링을 더 자세히 제어할 수 있는 옵션을 설정할 수도 있습니다.

로드 밸런싱 - 선택 사항

1 선택

아래 옵션을 사용하여 Auto Scaling 그룹을 기준 로드 밸런서 또는 사용자가 정의한 새 로드 밸런서에 연결합니다.

기준 로드 밸런서에 연결
기준 로드 밸런서 중에서 선택합니다.

새 로드 밸런서에 연결
Auto Scaling 그룹에 연결될 기본 로드 밸런서를 빠르게 생성합니다.

기준 로드 밸런서에 연결

Auto Scaling 그룹에 연결할 로드 밸런서를 선택합니다.

로드 밸런서 대상 그룹에서 선택
이 옵션을 사용하면 Application Load Balancer, Network Load Balancer 또는 Gateway Load Balancer를 연결할 수 있습니다.

Classic Load Balancer에서 선택

기준 로드 밸런서 대상 그룹

Auto Scaling 그룹과 동일한 VPC에 속하는 인스턴스 대상 그룹만 선택할 수 있습니다.

대상 그룹 선택 ▾

ALB-TG | HTTP
Application Load Balancer: CH9-ALB

선택

AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [2] 오토 스케일링 그룹 생성하기 ; 로드 밸런서를 통한 인스턴스 모니터링(상태 확인) 활성화



AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [2] 오토 스케일링 그룹 생성하기 ; 그룹 크기 조정

EC2 > Auto Scaling 그룹 > Auto Scaling 그룹 생성

1단계 시작 템플릿 또는 구성 선택

2단계 인스턴스 시작 옵션 선택

3단계 - 선택 사항 고급 옵션 구성

4단계 - 선택 사항 그룹 크기 및 크기 조정 정책 구성

5단계 - 선택 사항 알림 추가

6단계 - 선택 사항 태그 추가

7단계

그룹 크기 및 크기 조정 정책 구성 - 선택 사항 Info

Auto Scaling 그룹의 원하는 용량, 최소 용량 및 최대 용량을 설정합니다. 선택적으로 크기 조정 정책을 추가하여 그룹의 인스턴스 수를 동적으로 조정할 수 있습니다.

그룹 크기 - 선택 사항 Info

원하는 용량을 변경하여 Auto Scaling 그룹의 크기를 지정합니다. 최소 및 최대 용량 한도를 지정할 수도 있습니다. 원하는 용량은 한도 범위 내에 있어야 합니다.

원하는 용량 초기 그룹 용량 (= 초기 EC2 인스턴스 수)

최소 용량 **1** **입력**

최대 용량

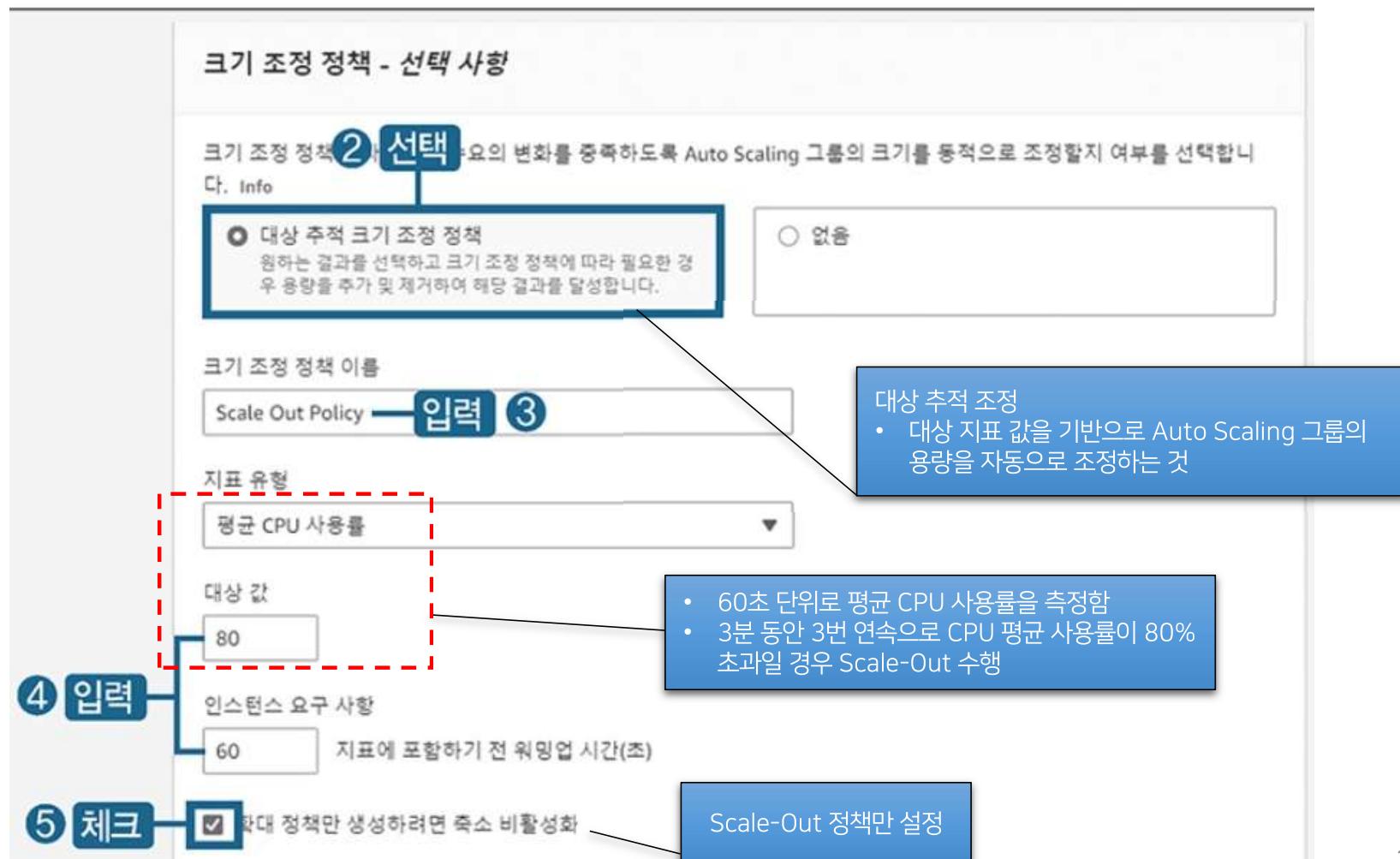
원하는 용량 초기 그룹 용량 (= 초기 EC2 인스턴스 수)

최소 용량 **1** **입력**

최대 용량

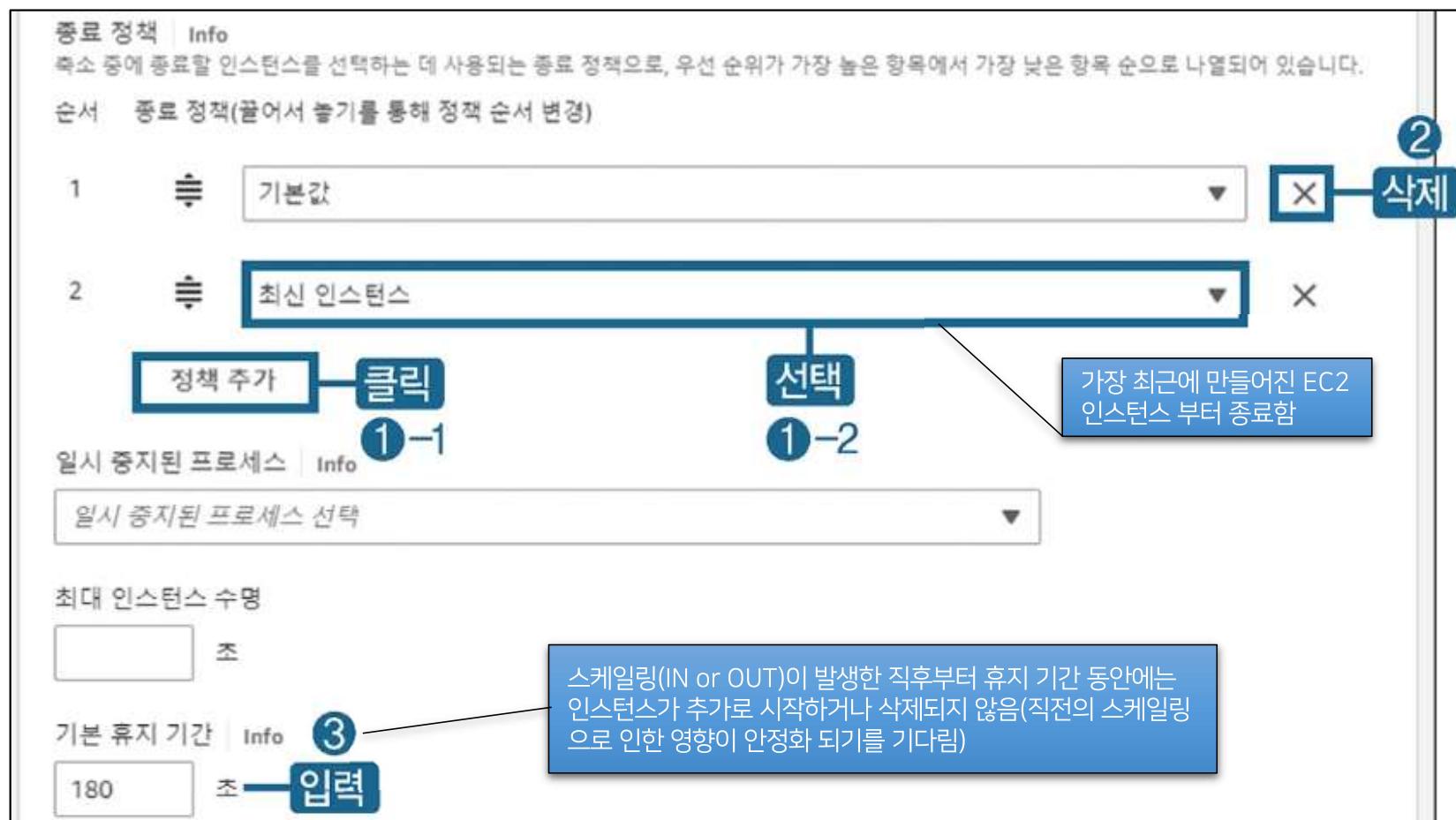
AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [2] 오토 스케일링 그룹 생성하기 ; 정책 설정



AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [3] 인스턴스 종료 정책 수립 ; 최신 인스턴스가 먼저 삭제되도록 종료 정책 수립
 - 인스턴스 축소 정책을 정의하기 전에, 인스턴스 종료 정책을 먼저 설정해야 함
 - 인스턴스 축소 시, 어느 인스턴스가 먼저 종료될 것인지를 결정



AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [4] 인스턴스 축소 정책 수립 ; Scale-In

The screenshot shows the AWS Auto Scaling Groups console. A blue box highlights the 'FirstEC2ASG' group. A blue callout '1 체크' points to the checkbox next to the group name. A blue box highlights the '자동 크기 조정' tab. A blue callout '2 클릭' points to the '작업' button. A blue box highlights the '등적 크기 조정 정책 생성' button. A blue callout '3 클릭' points to it. A large blue box at the bottom contains the 'Scale Out Policy' configuration, which includes a checkbox for '대상 주적 크기 조정' (Targeted scaling), '활성' (Active), '평균 CPU 사용률(율) 80(으)로 유지하는 데 필요한 경우' (When needed to maintain an average CPU usage rate of 80%), '필요에 따라 용량 단위 추가 또는 제거' (Add or remove capacity units as needed), '60초(지표에 포함하기 전 워밍업 시간)' (60 seconds (warm-up time before metrics are included)), and '비활성' (Inactive). A blue callout '아직은 Scale-Out 정책만 있는 상황' (Currently only a Scale-Out policy exists) points to the bottom right of the policy details.

Auto Scaling 그룹 (1/1) Info

Auto Scaling 그룹 검색

1 체크

2 클릭

3 클릭

작업

등적 크기 조정 정책 생성

Scale Out Policy

대상 주적 크기 조정

활성

평균 CPU 사용률(율) 80(으)로 유지하는 데 필요한 경우

필요에 따라 용량 단위 추가 또는 제거

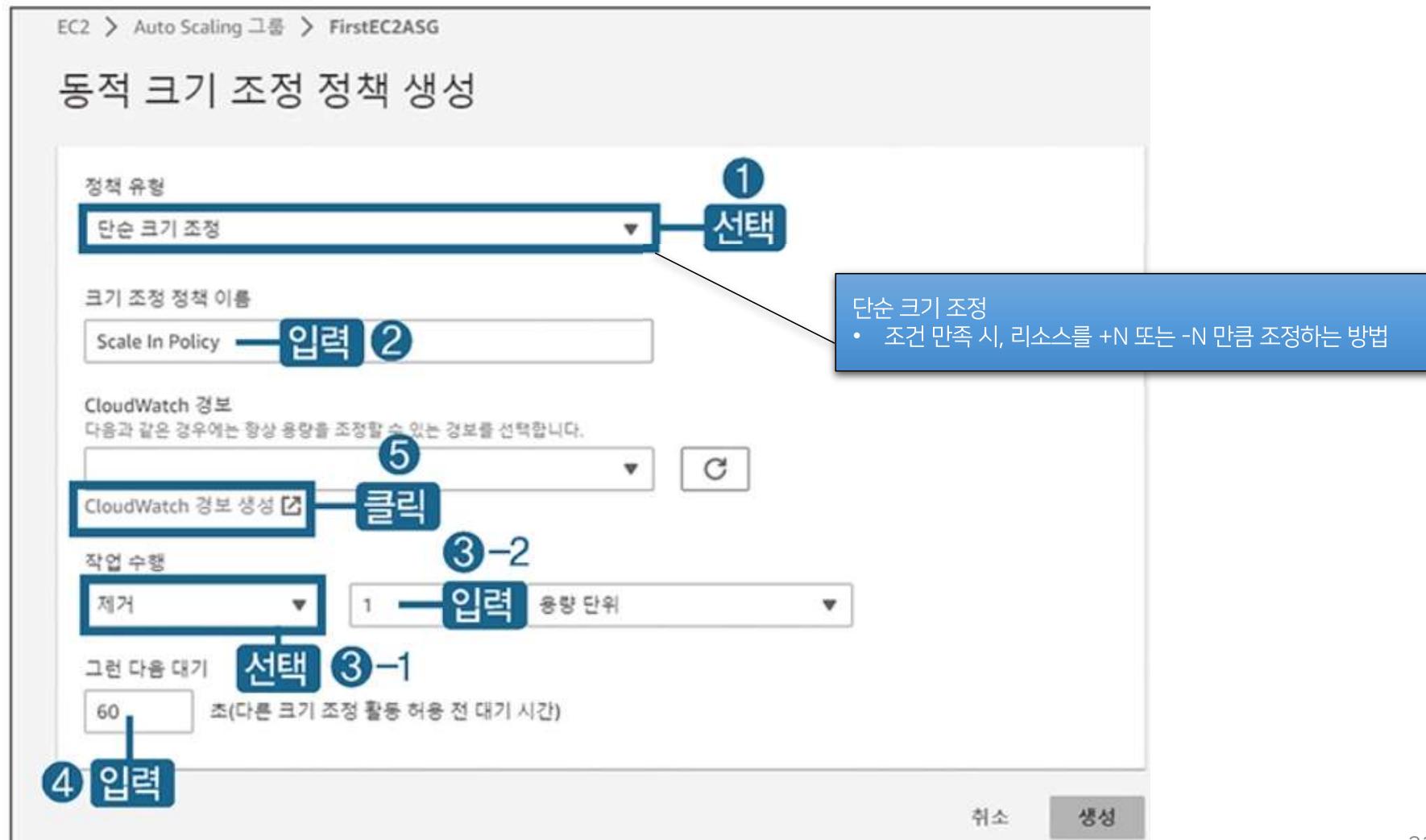
60초(지표에 포함하기 전 워밍업 시간)

비활성

아직은 Scale-Out 정책만 있는 상황

AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - [4] 인스턴스 축소 정책 수립 ; Scale-In (단순 크기 조정)



AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)
 - Scale-IN/OUT 규칙 조회하기

The screenshot shows the AWS Auto Scaling Groups console for the group 'FirstEC2ASG'. The 'Scale In Policy' section is highlighted with a blue border. It contains the following details:

- 단순 크기 조정
- 활성
- 선택한 경보 없음
- 제거 1 용량 단위
- 60초(다른 크기 조정 활동 허용 전 대기 시간)

The 'Scale Out Policy' section is also visible but not highlighted:

- 대상 주적 크기 조정
- 활성
- 평균 CPU 사용률을(를) 80(으)로 유지하는 데 필요한 경우
- 필요에 따라 용량 단위 추가 또는 제거
- 60초(지표에 포함하기 전 워밍업 시간)
- 비활성

AWS 오토 스케일링

- AWS EC2 오토 스케일링 구성하기 (예)

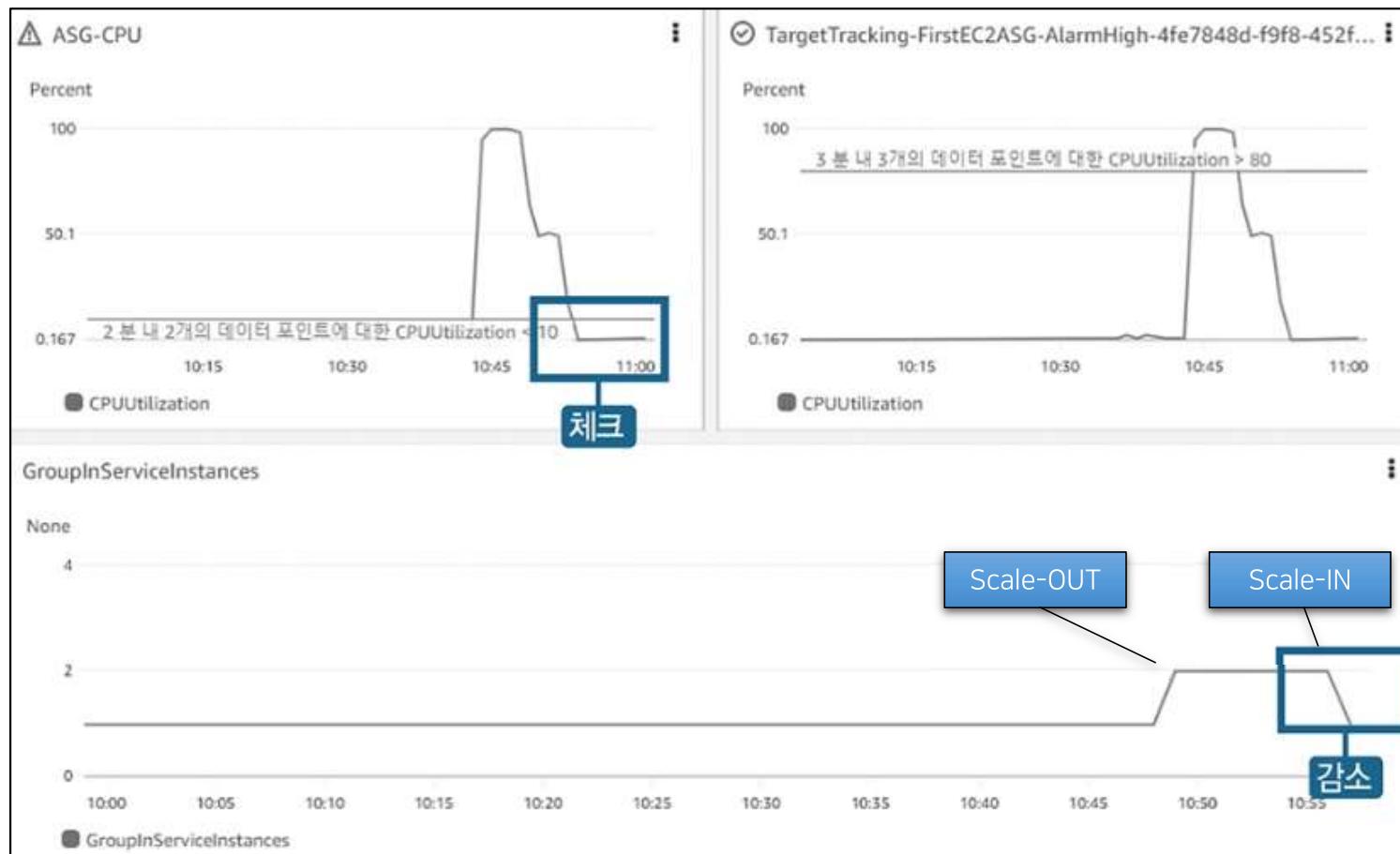
- Scale-IN/OUT 동작 확인하기

- 최초로 생성된 EC2 인스턴스에서 Apache Bench로 CPU 부하 생성 후, EC2 인스턴스 수 확인
 - 부하 생성이 종료된 이후, EC2 인스턴스 수가 감소하는지 확인
 - AWS CloudWatch를 활용하여 모니터링

아파치벤치 (ApacheBench)

웹 프로그램 :

아파치벤치는 HTTP 웹 서버의 성능 측정을 위한 싱글 스레드 명령 줄 컴퓨터 프로그램이다. 원래 아파치 HTTP 서버를 테스트할 목적으로 설계되었으며 현재는 모든 웹 서버를 테스트할 수 있을 만큼 충분히 일반화되어 있다. [위키백과](#)



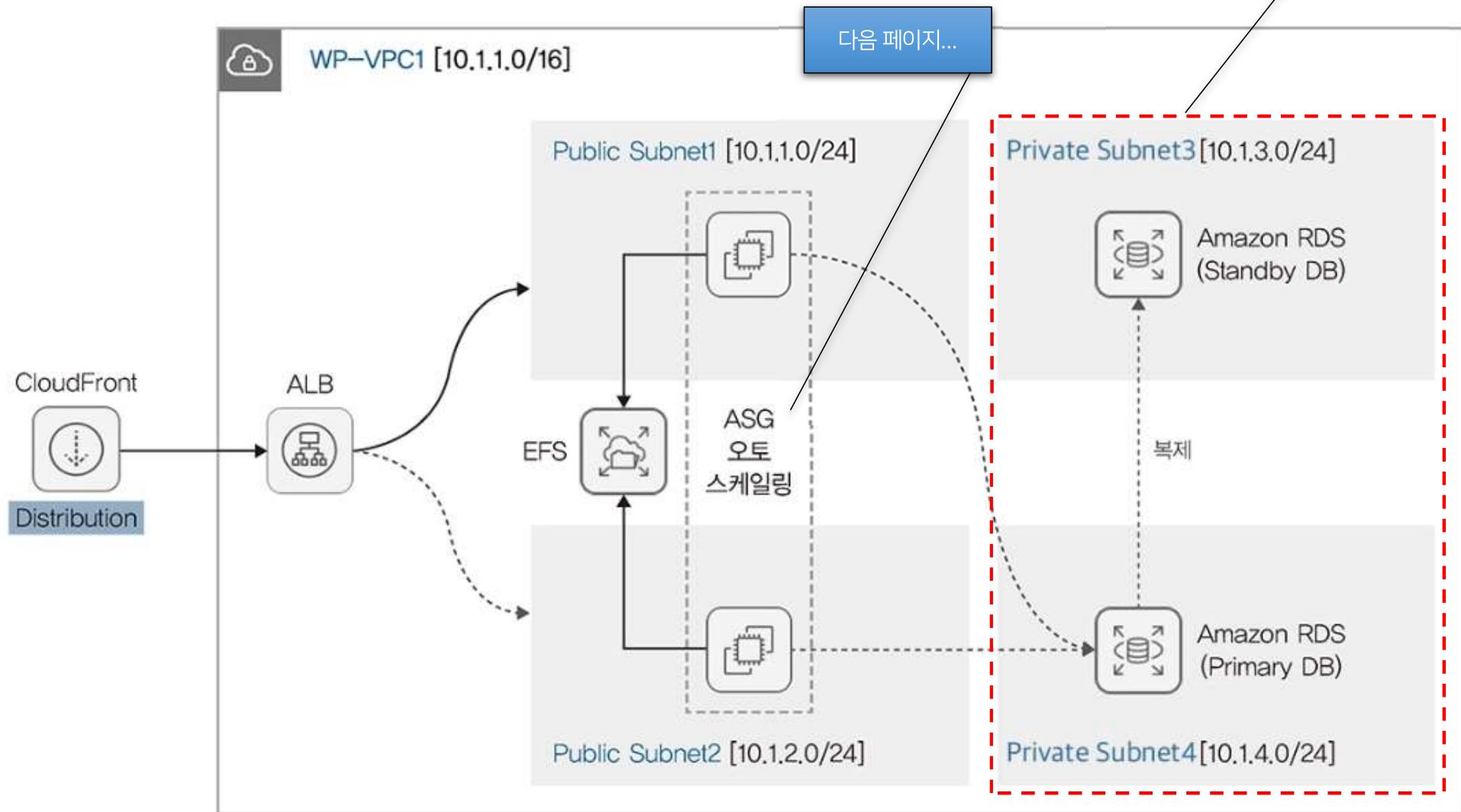


Amazon AWS 활용: 시스템 안정성

Amazon AWS 활용: 시스템 안정성 개선

- AWS 기능을 활용하여 시스템 안정성을 개선하는 방법
 - 예) 웹+DB로 구성된 서비스의 안정성을 개선하기

(참고) 강의 16(3) : Multi-AZ 복제방식을 사용한 Amazon DB 서비스의 고가용성 확보



Amazon AWS 활용: 시스템 안정성 개선

- AWS 기능을 활용하여 시스템 안정성을 개선하는 방법
 - 예) 웹+DB로 구성된 서비스의 안정성을 개선하기
 - 오토 스케일링은 시시각각 변화하는 부하량에 대응하기 위한 효과적인 솔루션임
 - 이에 더해, Amazon AWS는 가용성 보장을 위한 오토 스케일링 기능을 제공함

The screenshot shows a portion of the Amazon EC2 Auto Scaling User Guide. At the top, there's a navigation bar with links to AWS Documentation, Auto Scaling, and User Guide, along with a Feedback button. Below the navigation, the page title is "Amazon EC2 Auto Scaling" and the sub-page title is "Choose your scaling method". There are PDF and RSS links. The main content area starts with a statement: "Amazon EC2 Auto Scaling provides several ways to scale your instances based on the needs of your application. You can define scaling policies or scheduled actions to automatically add or remove capacity from your Auto Scaling group." A callout box highlights the first method: "Maintain a fixed number of instances". The text for this method explains that it's the default and describes how instances are launched to meet desired capacity and maintained even if they become unhealthy. A blue callout box contains the Korean translation of this text.

Feedback

Amazon EC2 Auto Scaling

User Guide

What is Amazon EC2 Auto Scaling?

Set up

Get started

Launch templates

Launch configurations

Auto Scaling groups

Recycle your instances

Scale your group

Choose your scaling method

Set scaling limits

Choose your scaling method

Feedback

Amazon EC2 Auto Scaling provides several ways to scale your instances based on the needs of your application. You can define scaling policies or scheduled actions to automatically add or remove capacity from your Auto Scaling group.

Maintain a fixed number of instances

The default for an Auto Scaling group is to not have any attached scaling policies or scheduled actions, which causes it to maintain a fixed size. After you create your Auto Scaling group, it starts by launching enough instances to meet its desired capacity. If there are no scaling conditions attached to the group, it continues to maintain its desired capacity even if an instance becomes unhealthy. Amazon EC2 Auto Scaling monitors the health of each instance in your Auto Scaling group. When it finds that an instance has become unhealthy, it replaces it with a new instance. You can read a more in-depth description of this process in [Health checks for instances in an Auto Scaling group](#).

별도의 성능 확장/축소 정책이 연결되지 않은 경우로, 리소스에 장애 발생 시 새로운 자원으로 대체하여 항상 고정된 수의 자원이 ACTIVE 상태인 것을 보장하는 오토 스케일링 기법

Amazon AWS 활용: 시스템 안정성 개선

- AWS 기능을 활용하여 시스템 안정성을 개선하는 방법
 - 예) 웹+DB로 구성된 서비스의 안정성을 개선하기
 - Amazon AWS는 고가용성 보장을 위한 오토 스케일링 기능을 제공함

The screenshot shows the configuration steps for an Auto Scaling group. The current step is '그룹 크기 및 크기 조정 정책 구성 - 선택 사항' (Group Size and Capacity Adjustment Policy Configuration - Selection Step). The steps are as follows:

- 1단계 시작 템플릿 또는 구성 선택
- 2단계 인스턴스 시작 옵션 선택
- 3단계 - 선택 사항 고급 옵션 구성
- 4단계 - 선택 사항 그룹 크기 및 크기 조정 정책 구성
- 5단계 - 선택 사항 알림 추가
- 6단계 - 선택 사항 태그 추가
- 7단계 검토

The main configuration area is titled '그룹 크기 - 선택 사항' (Group Size - Selection Step). It asks for the desired capacity, minimum capacity, and maximum capacity. The values entered are:

- 원하는 용량: 2 (Input 1)
- 최소 용량: 2 (Input 2)
- 최대 용량: 4 (Input 3)

A callout box highlights the maximum capacity input field with the note: '인스턴스의 수는 항상 2개로 유지되며 4개가 되지는 않음' (The number of instances will always be 2, not 4).

At the bottom, there is a section titled '크기 조정 정책 - 선택 사항' (Capacity Adjustment Policy - Selection Step) with the instruction: '크기 조정 정책을 사용하여 수요의 변화를 충족하도록 Auto Scaling 그룹의 크기를 정할지 여부를 선택합니다.' (Select whether to adjust the Auto Scaling group's size based on demand changes). A blue arrow points to the '선택' (Select) button, which is highlighted with a blue box.

A callout box to the right of the '선택' button contains the following points:

- 동적으로 인스턴스의 수를 조정하지 않음
- 단, 2개의 활성 인스턴스에서 장애 발생 시, 신규 인스턴스를 실행해서 교체하는 방식으로 2개 활성 인스턴스 수를 유지함

At the bottom of the configuration screen, there are two radio button options:

- 대상 주적 크기 조정 정책 (Targeted Capacity Adjustment Policy): '원하는 결과를 선택하고 크기 조정 정책에 따라 필요한 경우 용량을 추가 및 제거하여 해당 결과를 달성합니다.' (Select the desired result and add or remove capacity as needed based on the adjustment policy.)
- 없음 (None): '선택' (Selected) is highlighted with a blue box.

THE END

- 참고

- AWS 교과서, 김원일, 서종호, 김석필 지음, 길벗, 2023.10.20