

Auto Scaling 1

Hyunchan, Park

<http://oslab.jbnu.ac.kr>




Division of Computer Science and Engineering

Jeonbuk National University

Auto Scaling

- Scale in-out 을 자동으로 수행해주는 서비스
 - 유동적으로 변동하는 애플리케이션의 로드를 처리할 수 있는 정확한 수의 Amazon EC2 인스턴스를 보유하도록 보장
 - 내결함성 향상
 - 비정상 인스턴스를 검출하고, 대체 인스턴스 시작
 - 복수의 가용 영역을 사용하도록 구성 가능
 - 가용성 향상
 - 현재의 트래픽 요구를 처리할 수 있는 적절한 용량을 항상 갖출 수 있음
 - 비용 관리 개선
- 기타
 - EC2 Console 에서 제어
 - 비용: 무료
 - 참고: [AWS Auto Scaling Documentation](#)

Auto Scaling 구성 요소

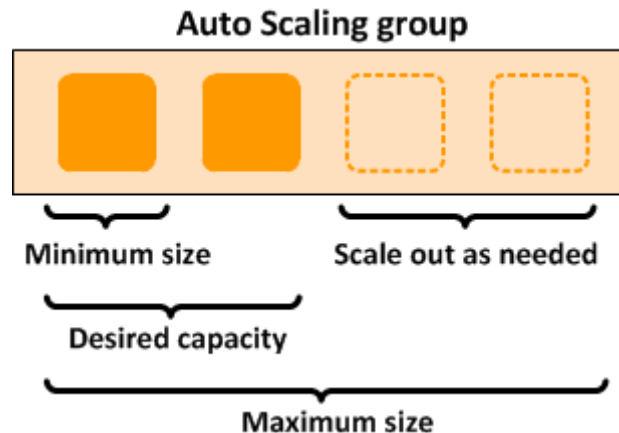
	<h2>그룹</h2> <p>EC2 인스턴스는 그룹에 정리되어 조정 및 관리 목적의 논리적 단위로 처리할 수 있습니다. 그룹을 생성할 때 EC2 인스턴스의 최소 및 최대 인스턴스 수와 원하는 인스턴스 수를 지정할 수 있습니다. 자세한 내용은 Auto Scaling 그룹 단원을 참조하십시오.</p>
	<h2>시작 구성</h2> <p>그룹에서는 시작 구성을 그룹의 EC2 인스턴스용 템플릿으로 사용합니다. 시작 구성을 생성할 때 인스턴스의 AMI ID, 인스턴스 유형, 키 페어, 보안 그룹, 블록 디바이스 매핑 등의 정보를 지정할 수 있습니다. 자세한 내용은 시작 구성 단원을 참조하십시오.</p>
	<h2>확장 계획</h2> <p>확장 계획은 Auto Scaling에 확장을 수행하는 시기와 방식을 전달합니다. 예를 들어, 지정된 조건의 발생(동적 확장) 또는 일정에 따른 확장 계획을 수립할 수 있습니다. 자세한 내용은 조정 계획 단원을 참조하십시오.</p>

시작 구성, 시작 템플릿 및 AS 그룹

- 시작 구성
 - EC2 인스턴스를 시작하는 데 사용하는 템플릿
- 시작 템플릿
 - 시작 구성의 Renewal: 새로운 기능들에 대한 지원
 - 스팟 인스턴스를 자동으로 선택해 구성해주는 기능이 추가됨
- AS 그룹
 - 유사한 특성을 공유하고,
인스턴스 조정 및 관리 목적의 논리적 그룹으로 취급되는,
EC2 인스턴스들의 모음
 - 조정 정책: 인스턴스의 수를 자동으로 조절하는데 사용되는 정책
- 특징
 - 한 시작 구성 혹은 템플릿은 여러 AS 그룹에서 사용될 수 있음
 - 한 AS 그룹은 하나의 시작 구성 혹은 템플릿만 사용할 수 있음

AS 그룹

- Auto scaling 그룹의 제어 파라미터
 - 최소 인스턴스 수: 그룹의 크기가 이 값 아래로 내려가지 않음
 - 목표 인스턴스 수: 확보하는 것이 언제나 보장되는 인스턴스 수
 - 실제 콘솔에서는 현재 Auto scaling 그룹이 정책에 따라 유지하려는 숫자가 표시
 - 최대 인스턴스 수: 그룹의 크기가 이 값을 넘지 않음



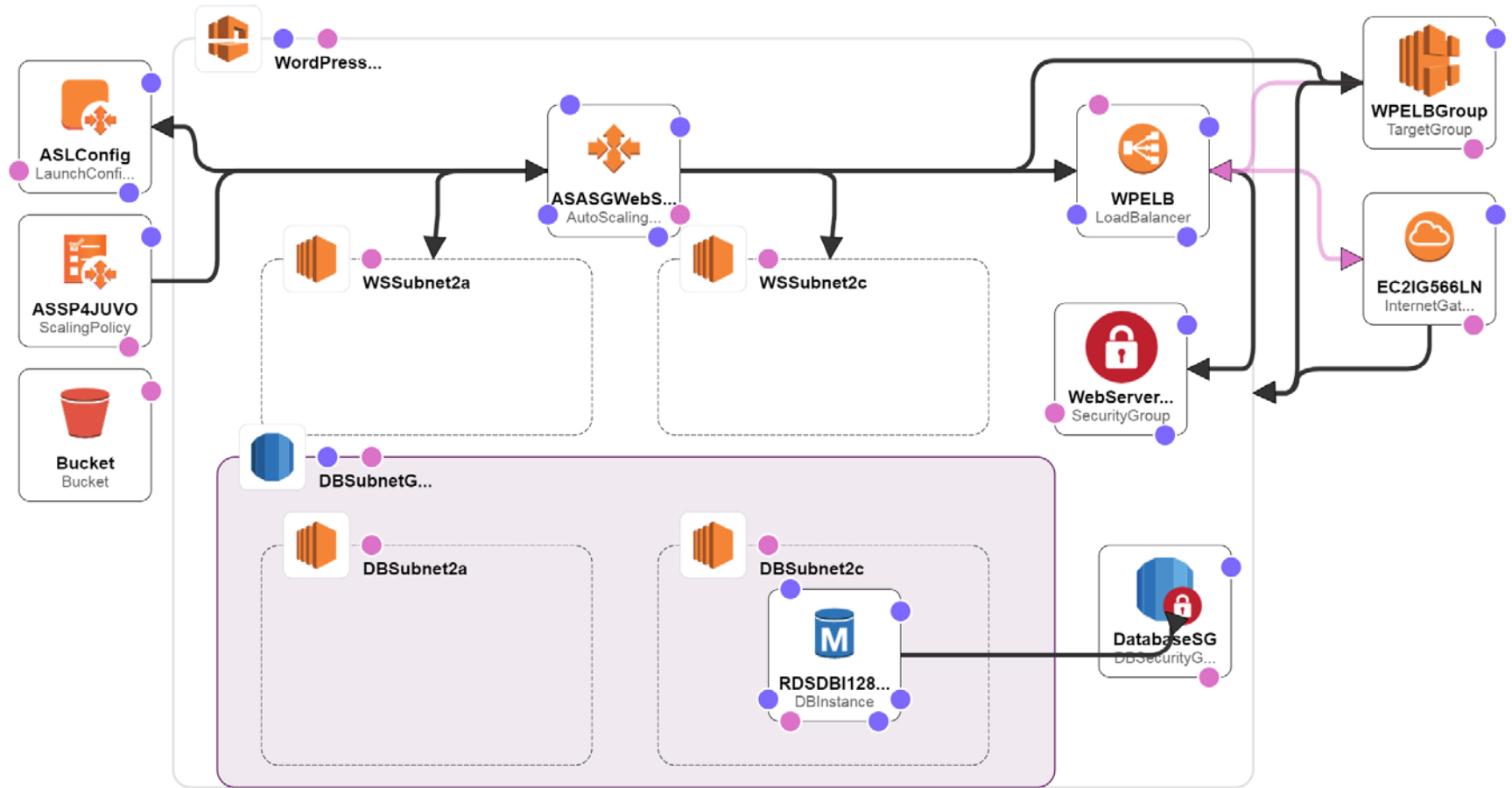
AS 그룹

- 구성 시, 미리 필요한 정보
 - 서버 시작 및 구성에 걸린 시간 (생성 시작부터 사이트 접속 가능까지)
 - 애플리케이션의 성능과 가장 연관성이 높은 지표 (예. 홈페이지 반응 속도)
 - Auto Scaling 그룹에 포함하려는 가용 영역 수
 - Auto Scaling에서 수행하려는 역할
 - Auto Scaling의 조정 기능을 사용하여 용량을 늘리거나 줄일지,
그저 일정한 서버 수가 항상 실행되도록 하는 데만 Auto Scaling을 사용할지 여부
(Auto Scaling은 두 기능을 동시에 수행할 수 있음)
 - 사용할 수 있는 기존 리소스 (예: EC2 인스턴스 또는 AMI)

실습

- ELB 대상그룹의 기존 인스턴스 제거
 - 오토스케일링으로 관리되는 인스턴스와 섞여 혼란스럽기 때문.
삭제하지 않아도 오토스케일링 동작에는 무방함
- Auto Scaling 설정
 - 시작 템플릿 생성: 사용할 인스턴스 설정
 - Auto Scaling 그룹 생성: Auto scaling 설정
- 부하 테스트를 통한 AS 동작 확인
- (주의!)
 - 교육용 계정에서는 인스턴스를 최대 3개까지만 설정 가능
 - 1개->2개->3개 로 증가하는 과정과,
 - 3개->2개->1개 로 감소하는 과정을 살펴볼 것

실습 구성



ELB 대상그룹 편집 및 기존 인스턴스 종료

- DB 인스턴스 상태 확인
 - 중지된 경우, 재시작
- 각자 도메인 주소로 접근하여 서비스 상태 확인
- ELB 대상그룹에서 모든 인스턴스를 등록 취소



- 기존 인스턴스들은 Terminate (종료)
 - 더 이상 필요하지 않음
 - (AB Test 를 수행할 인스턴스는 하나 필요함)

Auto Scaling 그룹 생성: 마법사 시작

aws

서비스 ▾

전용 호스트 New

정기 인스턴스

용량 예약

▼ 이미지

AMI

▼ Elastic Block Store

볼륨

스냅샷

수명 주기 관리자

▼ 네트워크 및 보안

보안 그룹 New

탄력적 IP New

배치 그룹 New

키 페어 New

네트워크 인터페이스 New

▼ 로드 밸런싱

로드밸런서

대상 그룹 New

▼ Auto Scaling

시작 구성

Auto Scaling 그룹

① 새로운 EC2 콘솔을 시작합니다.
AWS는 사용 편의성을 높이고 성능을 개선하기 위해 EC2 콘솔을 재설계하고 있습니다. 주기적으로 새 화면을 릴리스할 예정입니다. 새로운 화면을 사용해 보고 개선할 부분을 알려주십시오. 이전 콘솔과 새 콘솔 간에 전환하려면 [New EC2 Experience] 토글을 사용하십시오.

리소스

미국 동부 (버지니아 북부) 리전에서 다음 Amazon EC2 리소스를 사용하고 있음:

실행 중인 인스턴스	3	로드 밸런서	1
배치 그룹	0	보안 그룹	6
볼륨	5	스냅샷	1
인스턴스(모든 상태)	5	전용 호스트	0
키 페어	1	탄력적 IP	0

② AWS Launch Wizard for SQL Server를 사용하여 AWS에서 Microsoft SQL Server Always On 가용성 그룹을 손쉽게 크기 조정, 구성 및 배포할 수 있습니다. 자세히 알아보기

계정 속성

지원되는 플랫폼

- VPC

기본 VPC

vpc-375ea64a

설정

EBS 암호화

영역

기본 크레딧 사양

콘솔 실험

추가 정보

시작 안내서

설명서

모든 EC2 리소스

포럼

요금

문의처

인스턴스 시작

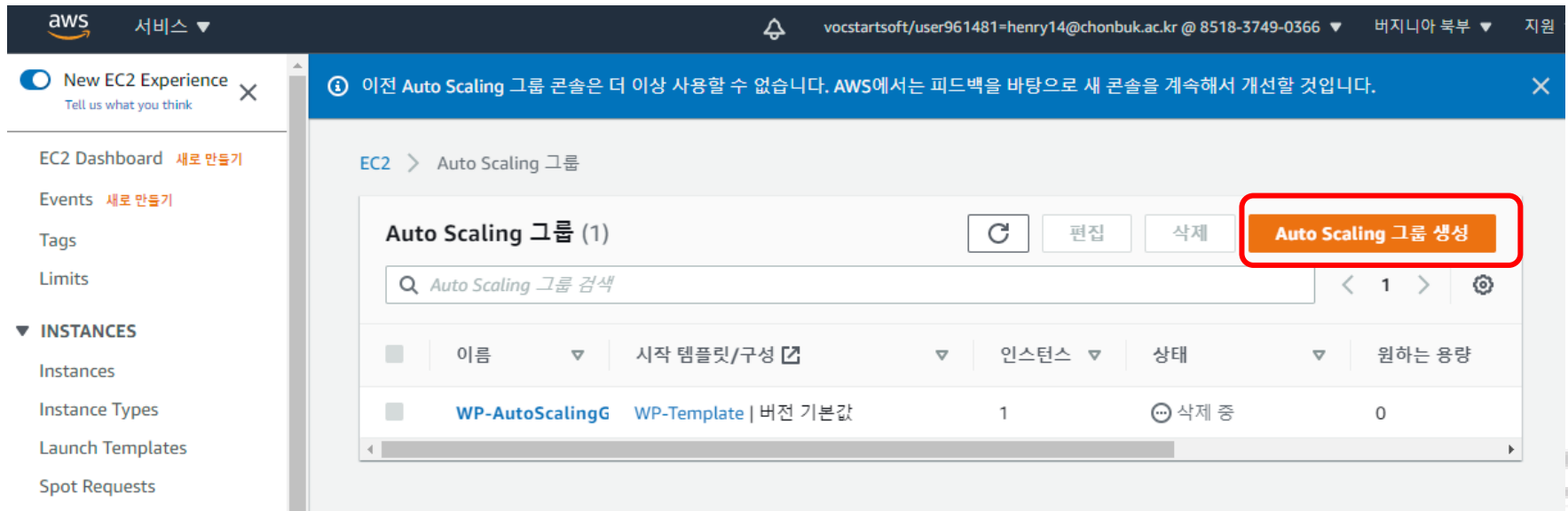
시작하려면 클라우드의 가상 서버인 Amazon EC2 인스턴스를 시작하십시오.

인스턴스 시작 ▾

참고: 인스턴스는 미국 동부 (버지니아 북부) 리전에서 시작됩니다



Auto Scaling 그룹 생성: 마법사 시작



The screenshot shows the AWS Management Console interface. At the top, the AWS logo and user information are visible. A blue banner at the top of the console area contains a message: "이전 Auto Scaling 그룹 콘솔은 더 이상 사용할 수 없습니다. AWS에서는 피드백을 바탕으로 새 콘솔을 계속해서 개선할 것입니다." Below this, the breadcrumb navigation shows "EC2 > Auto Scaling 그룹". The main content area is titled "Auto Scaling 그룹 (1)". It includes a search bar with the placeholder text "Auto Scaling 그룹 검색". To the right of the search bar are buttons for "Refresh", "Edit", "Delete", and "Auto Scaling 그룹 생성" (highlighted with a red box). Below the search bar is a table with columns: "이름", "시작 템플릿/구성", "인스턴스", "상태", and "원하는 용량". The table contains one entry: "WP-AutoScalingG" with a link to "WP-Template | 버전 기본값", 1 instance, and a status of "삭제 중".

이름	시작 템플릿/구성	인스턴스	상태	원하는 용량
WP-AutoScalingG	WP-Template 버전 기본값	1	삭제 중	0

시작 템플릿 부터 생성

EC2 > Auto Scaling 그룹 > Auto Scaling 그룹 생성

1단계
시작 템플릿 또는 구성 선택

2단계
설정 구성

3단계 (선택 사항)
고급 옵션 구성

4단계 (선택 사항)
그룹 크기 및 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 구성을 사용하는 경우 시작 템플릿으로 마이그레이션할 수 있습니다.

이름

Auto Scaling 그룹 이름
그룹을 식별할 이름을 입력합니다.

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

시작 템플릿 [Info](#)

[시작 구성으로 전환](#)

시작 템플릿

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



[시작 템플릿 생성](#)

[취소](#)

[다음](#)



시작 템플릿 아래 내용 수정

시작 템플릿 콘텐츠

아래에서 시작 템플릿의 세부 정보를 지정합니다. 비워 둔 필드는 시작 템플릿에 포함되지 않습니다.

Amazon Machine Image(AMI) - 필수 정보

AMI - 필수

WP-base-image

ami-07deb5efba2dda268

카탈로그: 내 AMI 아키텍처: 64비트(x86) 가상화: hvm

네트워크 설정

네트워킹 플랫폼 정보



Virtual Private Cloud(VPC)

AWS 클라우드 내에서 논리적으로 격리된 자체 영역에 있는 가상 네트워크로 시작



EC2-Classic

다른 고객과 공유하는 단일 플랫폼 네트워크에서 시작합니다.

보안 그룹 정보

보안 그룹 선택

WP-WebServerSG sg-016725a400dbffc2f ✕

VPC: vpc-0b6e7a53b2ce6eb0f

시작 템플릿 아래 내용 수정하고 완성!

▼ 고급 세부 정보 정보

구매 옵션 정보

☐ 스팟 인스턴스 요청

온디맨드 가격이 최대 가격으로 제한된 스팟 가격으로 스팟 인스턴스 요청

IAM 인스턴스 프로파일 정보

시작 템플릿에 포함하지 않음 ▼



새 IAM 프로파일 생성



종료 방식 정보

시작 템플릿에 포함하지 않음 ▼

최대 절전 중지 방식 정보

시작 템플릿에 포함하지 않음 ▼

종료 방지 정보

시작 템플릿에 포함하지 않음 ▼

세부 CloudWatch 모니터링 정보

활성화

탄력적 GPU 정보

시작 템플릿에 포함하지 않음

- 기본 모니터링: 5분 단위
- 세부 모니터링: 1분 단위 (유료)
상태 검사를 보다 자주 수행하여,
부하 변화에 보다 빠르게 대응하게 설정함



다시 AS 그룹 생성

EC2 > 시작 템플릿

시작 템플릿 (1/1) 정보



작업 ▲

시작 템플릿 생성

🔍 태그 또는 속성을 기준으로 필터링하거나 키워드로 검색

시작 템플릿 ID

시작 템플릿 이름

기본 버전 설정

태그 관리

스팟 플릿 생성

Auto Scaling 그룹 생성

세부 정보 보기



lt-0953012ea708733f8

WP-Template



AS그룹 생성: 1단계

EC2 > Auto Scaling 그룹 > Auto Scaling 그룹 생성

1단계
시작 템플릿 또는 구성 선택

2단계
설정 구성

3단계 (선택 사항)
고급 옵션 구성

4단계 (선택 사항)
그룹 크기 및 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 구성을 사용하는 경우 시작 템플릿으로 마이그레이션할 수 있습니다.

이름

Auto Scaling 그룹 이름
그룹을 식별할 이름을 입력합니다.

WP-AS-GROUP

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

시작 템플릿 [Info](#)

[시작 구성으로 전환](#)

시작 템플릿

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

WP-Template ▼



[시작 템플릿 생성](#)

버전

Default (1) ▼



[시작 템플릿 버전 생성](#)



AS그룹 생성: 2단계

- 자동으로 생성되는 인스턴스들을 on-demand 로 만들지, spot 으로 만들지 결정
- Spot 으로 생성하여 비용을 아끼기 위해 100% spot 으로 설정
- (사실 t2.micro 는 비용이 동일하거나 거의 차이가 나지 않음)

설정 구성

3단계 (선택 사항)
고급 옵션 구성

4단계 (선택 사항)
그룹 크기 및 조정 정책 구성

5단계 (선택 사항)
알림 추가

6단계 (선택 사항)
태그 추가

7단계
검토

인스턴스 구매 옵션 [Info](#)

시작 템플릿을 사용하여 그룹의 모든 인스턴스 사이에서 균일한 구성을 생성합니다. 또는 스팟 인스턴스 및 온디맨드 인스턴스 시작과 같은 다양한 요구 사항을 수용하도록 옵션을 정의합니다.

시작 템플릿 준수

시작 템플릿에 따라 구매 옵션(온디맨드 또는 스팟)과 인스턴스 유형이 결정됩니다.

구매 옵션과 인스턴스 유형 조합

시작할 온디맨드 및 스팟 용량과 선택적으로 여러 인스턴스 유형을 지정합니다. 이 옵션은 인스턴스 플릿의 규모와 비용을 최적화하는 데 매우 유용합니다.

인스턴스 배포

온디맨드 기본 용량 - 선택 사항

Auto Scaling 그룹이 기본적으로 보유해야 하는 온디맨드 용량을 지정합니다. 최대 그룹 크기는 이 값까지 증가합니다(단, 감소하지는 않음).

0

온디맨드 인스턴스

기본 용량을 초과하는 온디맨드 비율(%)

기본 용량 부분 이외의 추가 용량을 온디맨드 인스턴스와 스팟 인스턴스로 분할한 비율(%)입니다.

100

% 온디맨드

0

% 스팟

전북대학교 컴퓨터공학부
Division of Computer Science and Engineering
Chonbuk National University

17

AS그룹 생성: 2단계

인스턴스 유형 [Info](#)

애플리케이션의 요구 사항에 가장 적합한 인스턴스 유형을 선택합니다.

기본 인스턴스 유형

가중치 [Info](#)

1.

t2.micro
1vCPU 1 Gib Memory

▼

^

▼

×

 시작 템플릿에서 인스턴스 유형을 지정하지 않습니다. 따라서 [시작 템플릿 준수]를 선택할 수 없습니다. 위에서 인스턴스 유형을 추가하여 계속할 수 있습니다.

추가 인스턴스 유형

[다시 실행 권장 사항](#)

인스턴스 유형 추가



AS그룹 생성: 2단계

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 시작하는데 적합합니다.

VPC

vpc-0b6e7a53b2ce6eb0f (WP-VPC)
10.0.0.0/16

↻

[VPC 생성](#)

서브넷

서브넷 선택

↻

us-east-1a | subnet-0aa136b565d0eb1e6 (WebSubnet-Public-1a)
10.0.0.0/24

×

us-east-1b | subnet-0aeee7c00fe08751d (WebSubnet-Public-1b)
10.0.1.0/24

×

us-east-1c | subnet-00e5999ca1366b4ea (WebSubnet-Public-1c)
10.0.2.0/24

×

[서브넷 생성](#)



AS그룹 생성: 3단계

1단계
시작 템플릿 또는 구성 선택

2단계
설정 구성

3단계 (선택 사항)
고급 옵션 구성

4단계 (선택 사항)
그룹 크기 및 조정 정책 구성

5단계 (선택 사항)
알림 추가

6단계 (선택 사항)
태그 추가

7단계
검토

고급 옵션 구성 [Info](#)

안정성을 높이고 쉽게 확장 가능하도록 애플리케이션에 대한 수신 트래픽을 여러 인스턴스에 분산할 로드 밸런서를 선택합니다. 상태 확인 교체 및 모니터링을 더 자세히 제어할 수 있는 옵션을 설정할 수도 있습니다.

로드 밸런싱 - 선택 사항 [Info](#)

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

☐ 로드 밸런서 없음
Auto Scaling 그룹에 대한 트래픽은 로드 밸런서에 의해 front되지 않습니다.

☒ Attach to an existing load balancer
Choose from your existing load balancers.

☐ Attach to a new load balancer
Quickly create a basic load balancer to attach to your Auto Scaling group.

Attach to an existing load balancer

Select the load balancers that you want to attach to your Auto Scaling group.

☒ 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

기존 로드 밸런서 대상 그룹

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

대상 그룹 선택



WordpressELBGroup | HTTP
Application Load Balancer: WordpressELB



AS그룹 생성: 3단계

상태 확인 - 선택 사항

상태 확인 유형 [Info](#)

EC2 Auto Scaling은 상태 확인에 실패한 인스턴스를 자동으로 교체합니다. 로드 밸런싱을 활성화한 경우 항상 활성화된 EC2 상태 확인 외에도 ELB 상태 확인을 활성화할 수 있습니다.

☒ EC2 ☐ ELB

상태 확인 유예 기간

EC2 Auto Scaling이 새 인스턴스가 서비스 상태가 된 후 새 인스턴스에 대한 첫 번째 상태 확인을 수행할 때까지의 시간입니다.

초

추가 설정 - 선택 사항

모니터링 [Info](#)

☒ CloudWatch 내에서 그룹 지표 수집 활성화



AS그룹 생성: 4단계

EC2 > Auto Scaling 그룹 > Auto Scaling 그룹 생성

1단계

[시작 템플릿 또는 구성 선택](#)

2단계

[설정 구성](#)

3단계 (선택 사항)

[고급 옵션 구성](#)

4단계 (선택 사항)

[그룹 크기 및 조정 정책 구성](#)

5단계 (선택 사항)

[알림 추가](#)

6단계 (선택 사항)

[태그 추가](#)

7단계

[검토](#)

그룹 크기 및 조정 정책 구성 [Info](#)

Auto Scaling 그룹의 원하는 용량, 최소 용량 및 최대 용량을 설정합니다. 선택적으로 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

최소 용량

최대 용량



AS그룹 생성: 4단계

조정 정책 - 선택 사항

조정 정책을 사용하여 수요의 변화를 충족하도록 Auto Scaling 그룹의 크기를 동적으로 조정할지 여부를 선택합니다. [Info](#)



대상 추적 조정 정책

원하는 결과를 선택하고 조정 정책에 따라 필요한 경우 용량을 추가 및 제거하여 해당 결과를 달성합니다.



없음

조정 정책 이름

ScaleOutPolicy-CPU70

지표 유형

평균 CPU 사용률 ▼

대상 값

70

인스턴스 요구 사항

180

지표에 포함하기 전 워밍업 시간(초)



확대 정책만 생성하려면 축소 비활성화

- 조건을 만족하여 조정이 한 번 이루어진 후, 검사를 유예하는 시간



AS그룹 생성: 7단계 검토 후 완료. 5,6 생략

❗ 이전 Auto Scaling 그룹 콘솔은 더 이상 사용할 수 없습니다. AWS에서는 피드백을 바탕으로 새 콘솔을 계속해서 개선할 것입니다.

✅ WP-AS-GROUP, 1 조정 정책(를) 생성함. 그룹 지표 수집이 활성화되었습니다.

EC2 > Auto Scaling 그룹

Auto Scaling 그룹 (1)



편집

삭제

Auto Scaling 그룹 생성

🔍 Auto Scaling 그룹 검색



1



이름



시작 템플릿/구성



인스턴스



상태



원하는 용량



WP-AS-GROUP...

WP-Template | 버전 기본값

0



용량 업데이트 중

1


* 목표 용량 1에 맞춰 인스턴스 구성 중.
180초 이내이므로, 상태 검사는 수행 안 됨



잠시 후...AS 그룹 상태 확인

EC2 > Auto Scaling 그룹

Auto Scaling 그룹 (1/1) 🔄 편집 삭제 Auto Scaling 그룹 생성

<input checked="" type="checkbox"/>	이름 ▾	시작 템플릿/구성  ▾	인스턴스 ▾	상태 ▾	원하는 용량 ▾	최소 ▾	최대 ▾	가용 영역
<input checked="" type="checkbox"/>	WP-AS-GROUP	WP-Template 버전 기본값	1	-	1	1	3	us-east-1a, us-east-1b

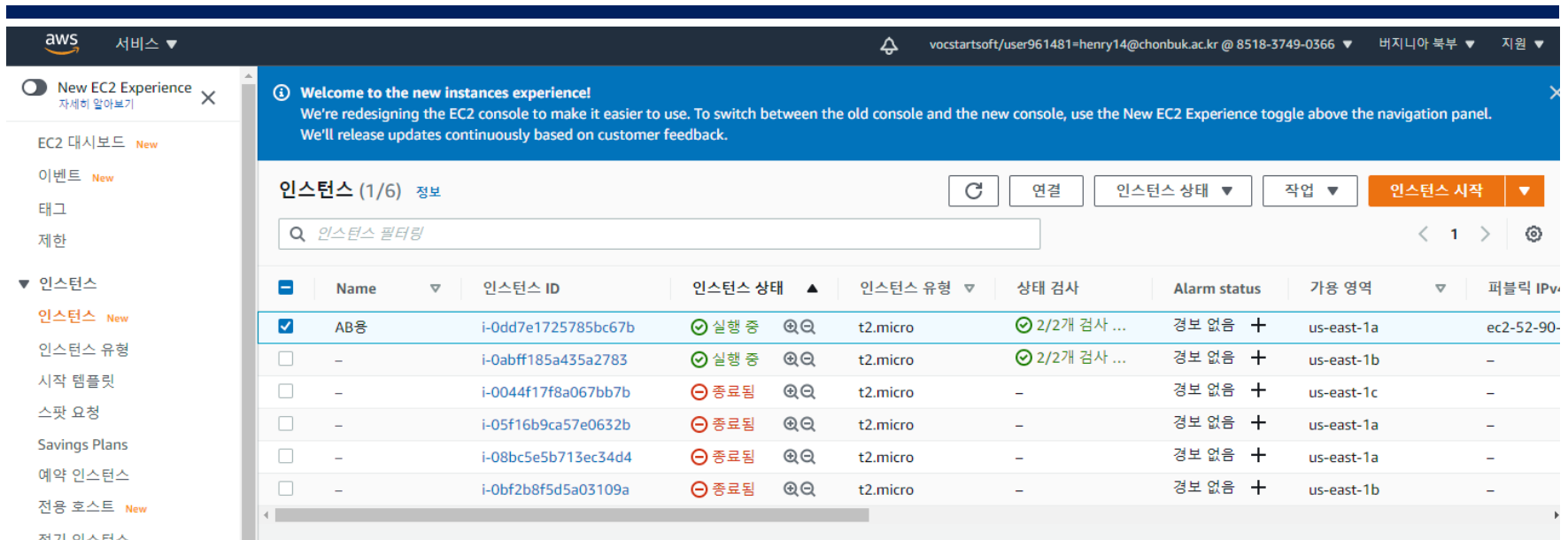
그룹 세부 정보 편집

원하는 용량 1	Auto Scaling 그룹 이름 WP-AS-GROUP
최소 용량 1	생성된 날짜 Fri Nov 27 2020 13:02:50 GMT+0900 (대한민국 표준시)
최대 용량 3	Amazon 리소스 이름(ARN) arn:aws:autoscaling:us-east-1:851837490366:autoScalingGroup:0fe0516d-3262-4ed2-a137-279dfd2394f5:autoScalingGroupName/WP-AS-GROUP

* 참고: 원하는 용량을 수정하면, 그에 맞춰 즉각 scale in, out 이 이루어짐



인스턴스 화면에서 확인



인스턴스 (1/6) 정보

인스턴스 필터링

	Name	인스턴스 ID	인스턴스 상태	인스턴스 유형	상태 검사	Alarm status	가용 영역	퍼블릭 IP
<input checked="" type="checkbox"/>	AB용	i-0dd7e1725785bc67b	실행 중	t2.micro	2/2개 검사 ...	경보 없음	us-east-1a	ec2-52-90-
<input type="checkbox"/>	-	i-0abff185a435a2783	실행 중	t2.micro	2/2개 검사 ...	경보 없음	us-east-1b	-
<input type="checkbox"/>	-	i-0044f17f8a067bb7b	종료됨	t2.micro	-	경보 없음	us-east-1c	-
<input type="checkbox"/>	-	i-05f16b9ca57e0632b	종료됨	t2.micro	-	경보 없음	us-east-1a	-
<input type="checkbox"/>	-	i-08bc5e5b713ec34d4	종료됨	t2.micro	-	경보 없음	us-east-1a	-
<input type="checkbox"/>	-	i-0bf2b8f5d5a03109a	종료됨	t2.micro	-	경보 없음	us-east-1b	-

- 맨 위 인스턴스는 AB 테스트용
- 퍼블릭 IP는 없음 (ELB를 통하므로 불필요)
- 보안 그룹 등 다른 설정들도 확인



대상 그룹 확인

EC2 > Target groups > WordpressELBGroup

WordpressELBGroup

Delete

arn:aws:elasticloadbalancing:us-east-1:851837490366:targetgroup/WordpressELBGroup/9ef03ceb14279534

Basic configuration

Target type
instance

Protocol : Port
HTTP : 80

Protocol version
HTTP1

VPC
vpc-0b6e7a53b2ce6eb0f

Load balancer
WordpressELB

Group details

Targets

Monitoring

Tags

Registered targets (1)



Deregister

Register targets

Filter resources by property or value

< 1 > ⚙

<input type="checkbox"/>	Instance ID	Name	Port	Zone	Status	Status details
<input type="checkbox"/>	i-0abff185a435a2783		80	us-east-1b	✓ healthy	



The screenshot shows a web browser window. The address bar contains the text '주의 요함 | chansmusic.tk'. The page content features a large, bold heading 'USER'S BLOG!' and a subtitle 'Just another WordPress site'. A small downward-pointing arrow is visible on the right side of the page.

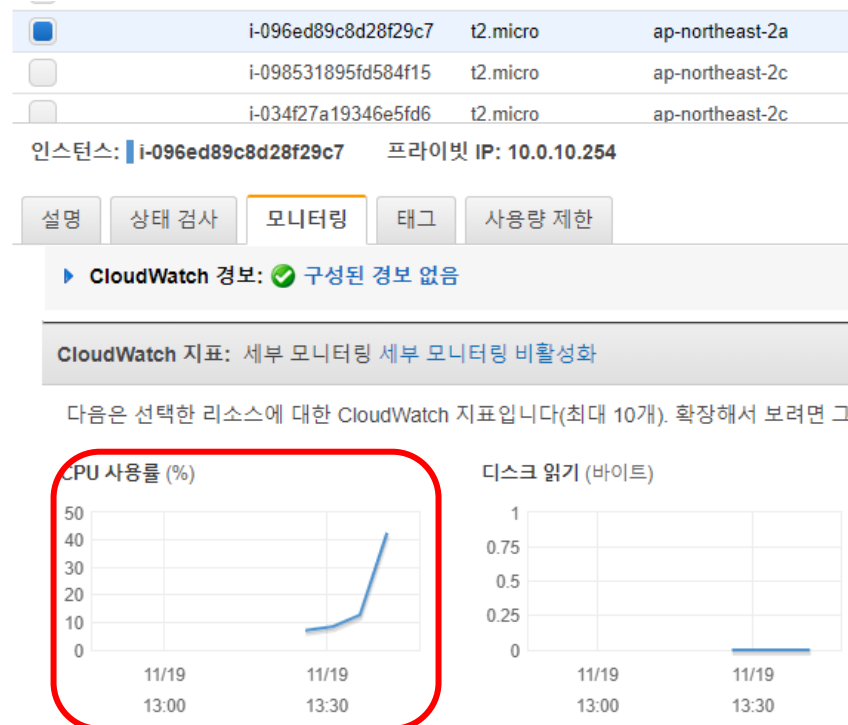


부하 테스트를 통한 Auto Scaling 동작 확인



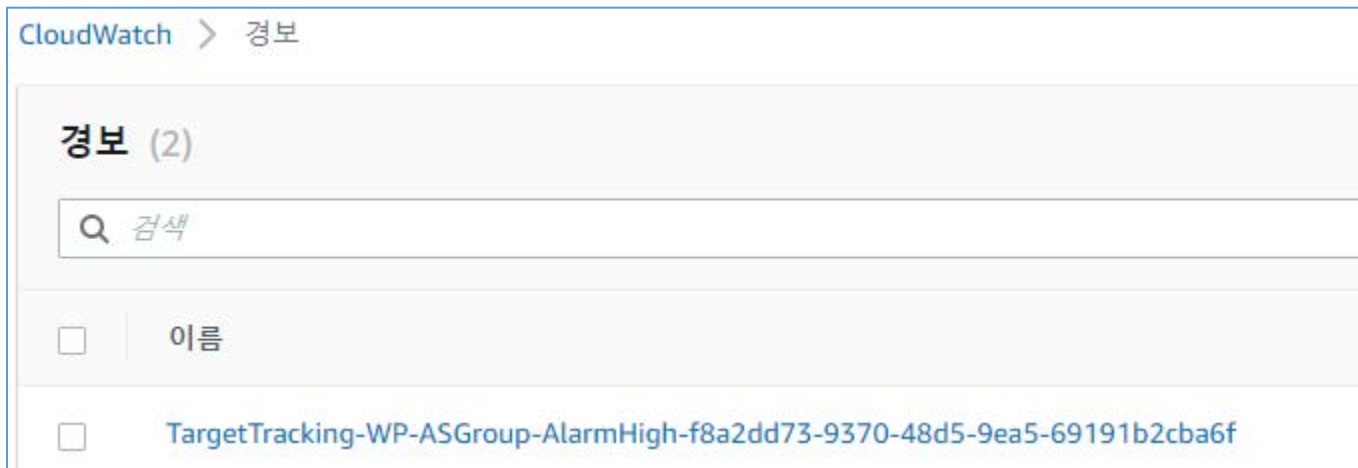
1. CPU 사용률 확인

- AB 로 로드를 걸었을 때, CPU 사용률이 설정한 70%를 넘어가는지 확인
 - \$ ab -t 1500 -c 200 <http://chansmusic.tk/>
 - 중요한 옵션은 Concurrency! 동시 요청 수를 높여야 서비스에 부하를 많이 줄 수 있음
 - Fail 나는지 여부는 계속해서 확인할 것. Fail 이 나면 측정이 잘못 이루어질 수 있음
- EC2 인스턴스 화면에서 모니터링 확인



2. CloudWatch 확인

- CloudWatch
 - AWS 내에서 발생하는 거의 모든 데이터를 모니터링하고 경보하는 서비스
 - 비용, Auto scaling 그룹의 설정값 위반 등 우리가 사용한 alarm 서비스는 모두 CloudWatch 가 제공함
 - 대시보드 경보 아래에 경보/확인 등 2개의 알람이 생성되어 있을 것
 - 증가/감소 정책을 위한 알람



CloudWatch 확인

CPUUtilization

3 분 내 3개의 데이터 포인트에 대한 CPUUtilization > 70

Percent

77.5

38.8

0.167

11:05 11:10 11:15 11:20 11:25 11:30 11:35 11:40 11:45 11:50 11:55 12:00 12:05 12:10 12:15 12:20 12:25

■ CPUUtilization

세부 정보

이름

TargetTracking-WP-ASGroup-AlarmHigh-f8a2dd73-9370-48d5-9ea5-69191b2cba6f

설명

DO NOT EDIT OR DELETE. For TargetTrackingScaling policy arn:aws:autoscaling:ap-northeast-2:458235646241:scalingPolicy:2f7328c2-262a-40a0-9c6a-631eb8f821be:autoScalingGroupName/WP-ASGroup:policyName/Scale Group Size.

임계값

3 분 내 3개의 데이터 포인트에 대한 CPUUtilization > 70

마지막 변경

2019-11-19 13:56:05

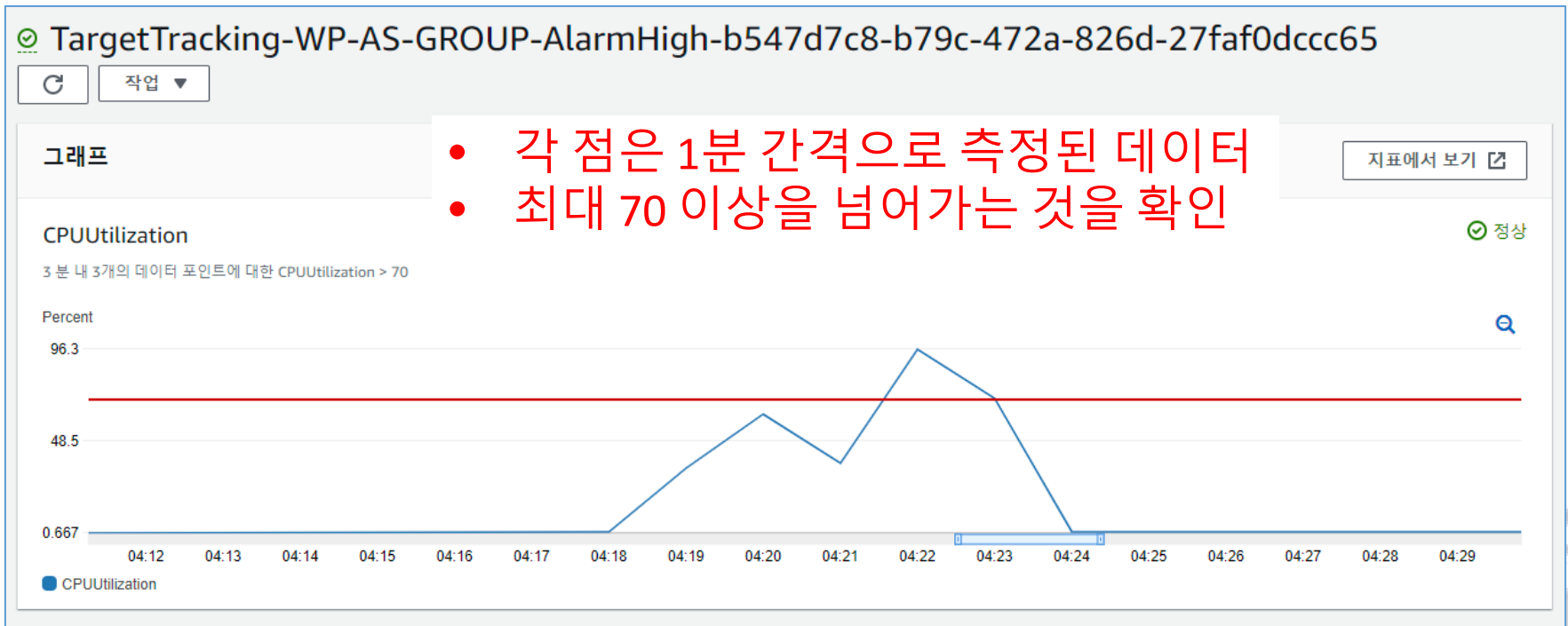
ARN

arn:aws:cloudwatch:ap-northeast-2:458235646241:alarm:TargetTracking-WP-ASGroup-AlarmHigh-f8a2dd73-9370-48d5-9ea5-69191b2cba6f



2. CloudWatch 증가 옵션의 경보 확인

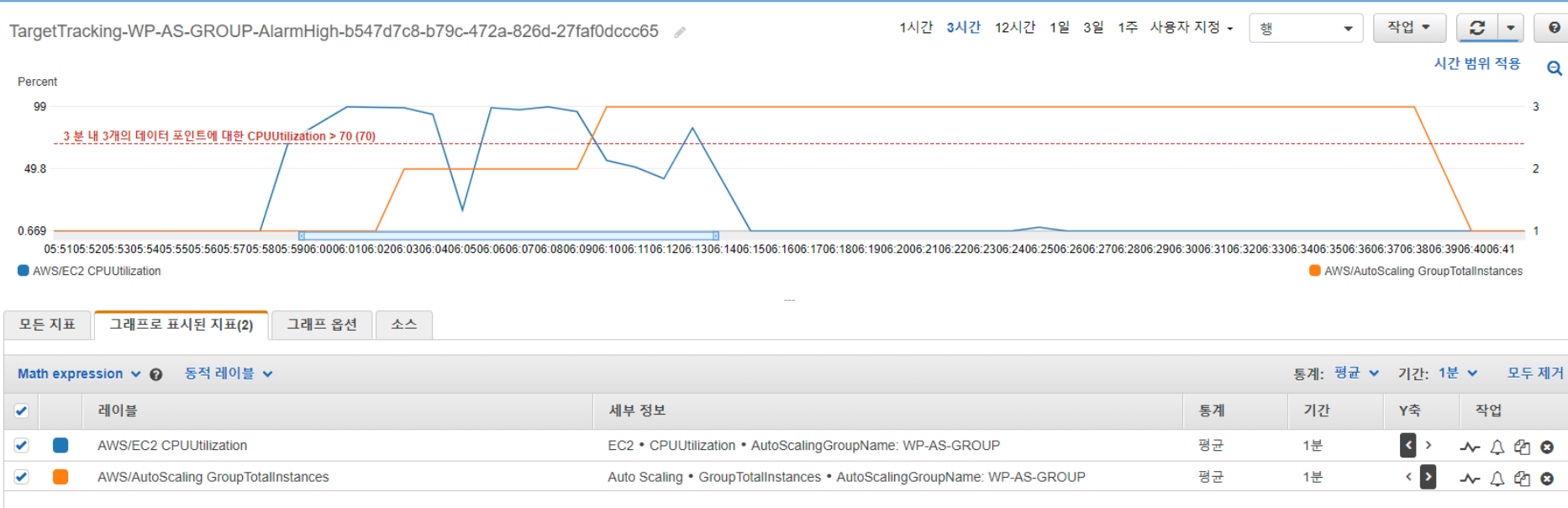
- Red: Threshold 70 라인
- Blue: 실제 측정된 CPU 사용률



- 좀 더 길고 강한 부하를 주자: `$ ab -t 900 -c 200 http://chansmusic.tk/`

3. 지표 추가해서 보기

- “지표에서 보기”
 - 모든 지표 메뉴에서 AutoScaling Group TotalInstances 확인
 - Y축을 오른쪽에 표시
- CPU 사용률에 따라 알람이 발생하고, 그에 따라 인스턴스가 증가
 - 1 -> 2 -> 3
 - 나중에 어떻게 감소했을까? -> 감소 정책 확인



CloudWatch 알람 확인

TargetTracking-WP-ASGroup-AlarmLow-4e6e7f20-1417-43d0-93e4-96c1724ea8d8

편집

삭제

복사

- 자동으로 설정된 축소 조건
- 15분간 CPU 사용률 52.5% 이하

그래프

지표에서 보기 [↗](#)

CPUUtilization

🟢 정상

15 분 내 15개의 데이터 포인트에 대한 CPUUtilization < 52.5

Percent

77.5

38.8

0.165

11:15 11:30 11:45 12:00 12:15 12:30 12:45 13:00 13:15 13:30 13:45 14:00

■ CPUUtilization



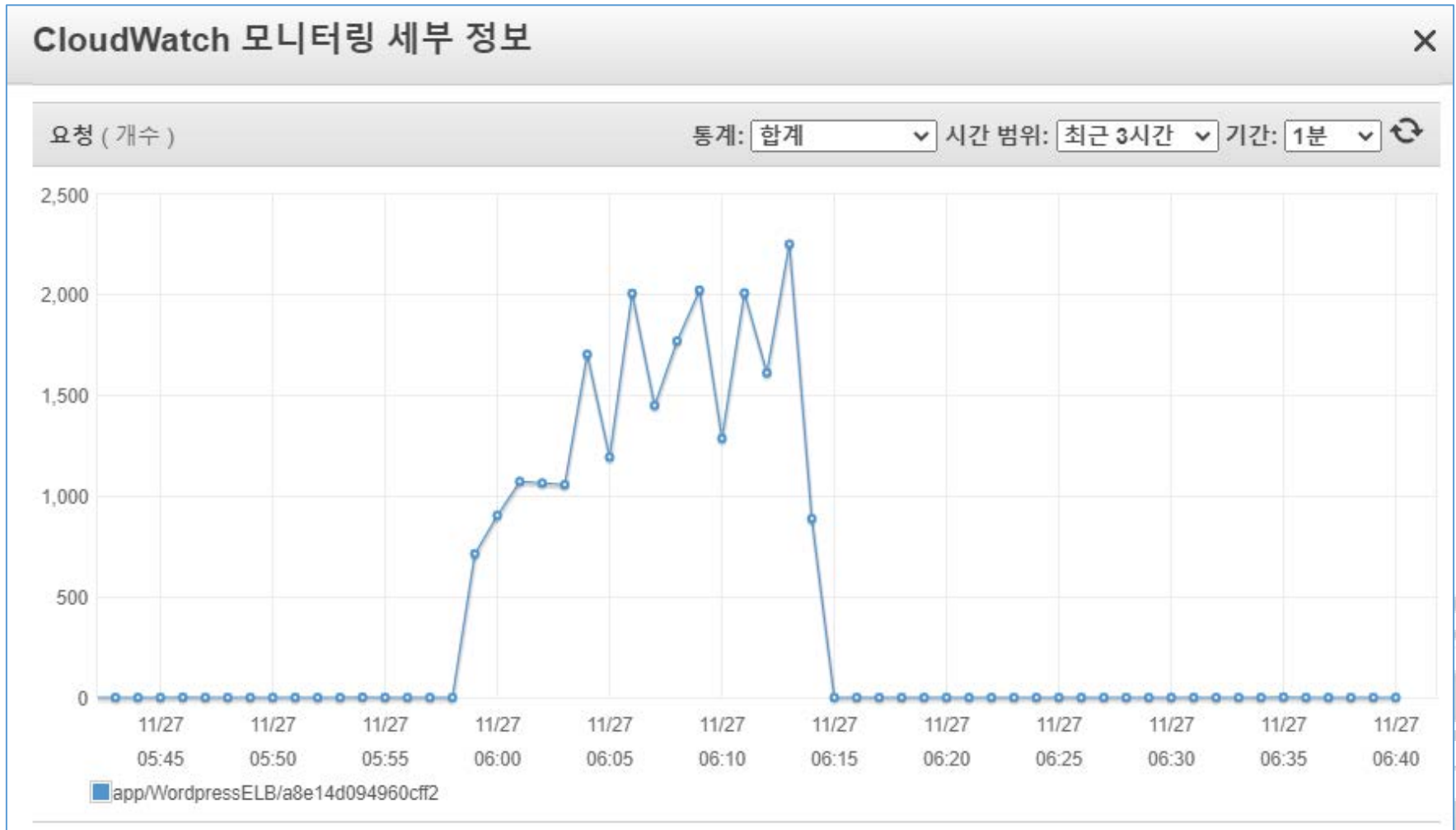
3.AS Group 에서 확인: 작업 기록

- 위 그래프에서 나타난 인스턴스의 증감이 기록되어 있음

작업 기록 (7)					
<input type="text" value="활동 기록 필터링"/>					
상태	설명	원인	시작 시간	종료 시간	
Successful	Terminating EC2 instance: i-05511261df2b03ba7	At 2020-11-27T06:34:50Z a monitor alarm TargetTracking-WP-AS-GROUP-AlarmLow-7afc984a-d181-45e1-850d-ee86ecd59bb8 in state ALARM triggered policy ScaleOutPolicy-CPU70 changing the desired capacity from 2 to 1. At 2020-11-27T06:34:52Z an instance was taken out of service in response to a difference between desired and actual capacity, shrinking the capacity from 2 to 1. At 2020-11-27T06:34:52Z instance i-05511261df2b03ba7 was selected for termination.	2020 11월 27, 03:34:52 오후 +09:00	2020 11월 27, 03:40:40 오후 +09:00	
Successful	Terminating EC2 instance: i-069741b12d4ff2ef6	At 2020-11-27T06:33:50Z a monitor alarm TargetTracking-WP-AS-GROUP-AlarmLow-7afc984a-d181-45e1-850d-ee86ecd59bb8 in state ALARM triggered policy ScaleOutPolicy-CPU70 changing the desired capacity from 3 to 2. At 2020-11-27T06:33:53Z an instance was taken out of service in response to a difference between desired and actual capacity, shrinking the capacity from 3 to 2. At 2020-11-27T06:33:53Z instance i-069741b12d4ff2ef6 was selected for termination.	2020 11월 27, 03:33:53 오후 +09:00	2020 11월 27, 03:39:38 오후 +09:00	
Successful	Launching a new EC2 instance: i-05511261df2b03ba7	At 2020-11-27T06:09:46Z a monitor alarm TargetTracking-WP-AS-GROUP-AlarmHigh-b547d7c8-b79c-472a-826d-27faf0dccc65 in state ALARM triggered policy ScaleOutPolicy-CPU70 changing the desired capacity from 2 to 3. At 2020-11-27T06:10:08Z an instance was started in response to a difference between desired and actual capacity, increasing the capacity from 2 to 3.	2020 11월 27, 03:10:11 오후 +09:00	2020 11월 27, 03:13:43 오후 +09:00	
Successful	Launching a new EC2 instance: i-069741b12d4ff2ef6	At 2020-11-27T06:02:46Z a monitor alarm TargetTracking-WP-AS-GROUP-AlarmHigh-b547d7c8-b79c-472a-826d-27faf0dccc65 in state ALARM triggered policy ScaleOutPolicy-CPU70 changing the desired capacity from 1 to 2. At 2020-11-27T06:02:58Z an instance was started in response to a difference between desired and actual capacity, increasing the capacity from 1 to 2.	2020 11월 27, 03:03:01 오후 +09:00	2020 11월 27, 03:07:03 오후 +09:00	

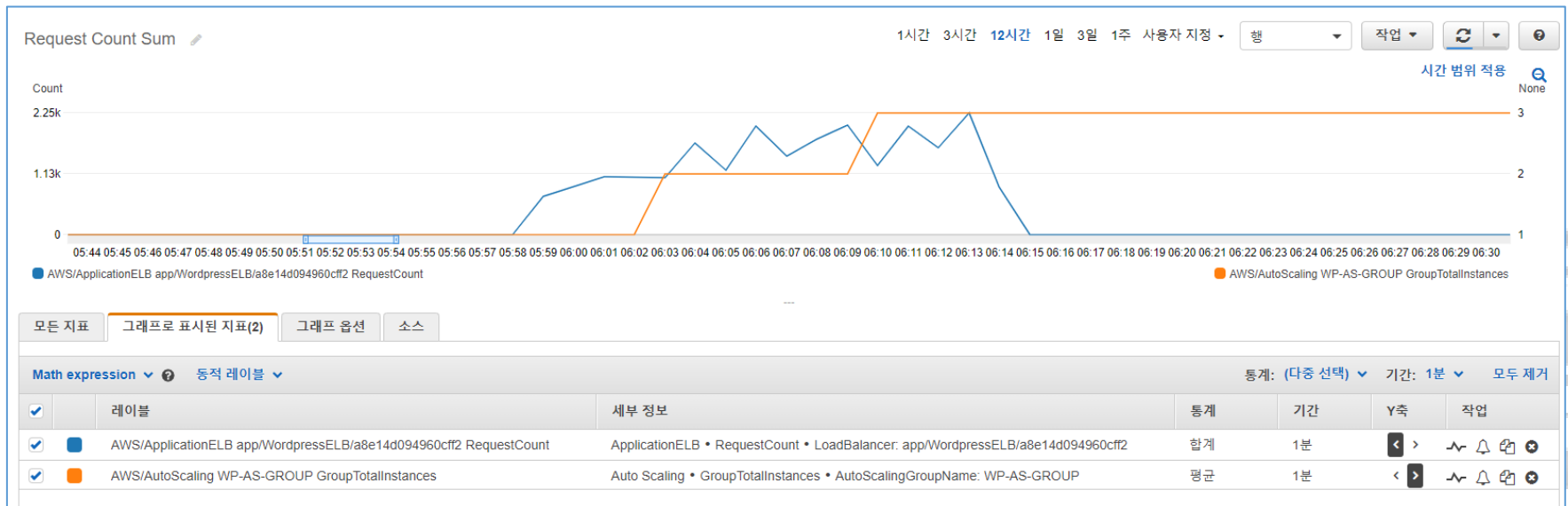


4. ELB 에서 확인



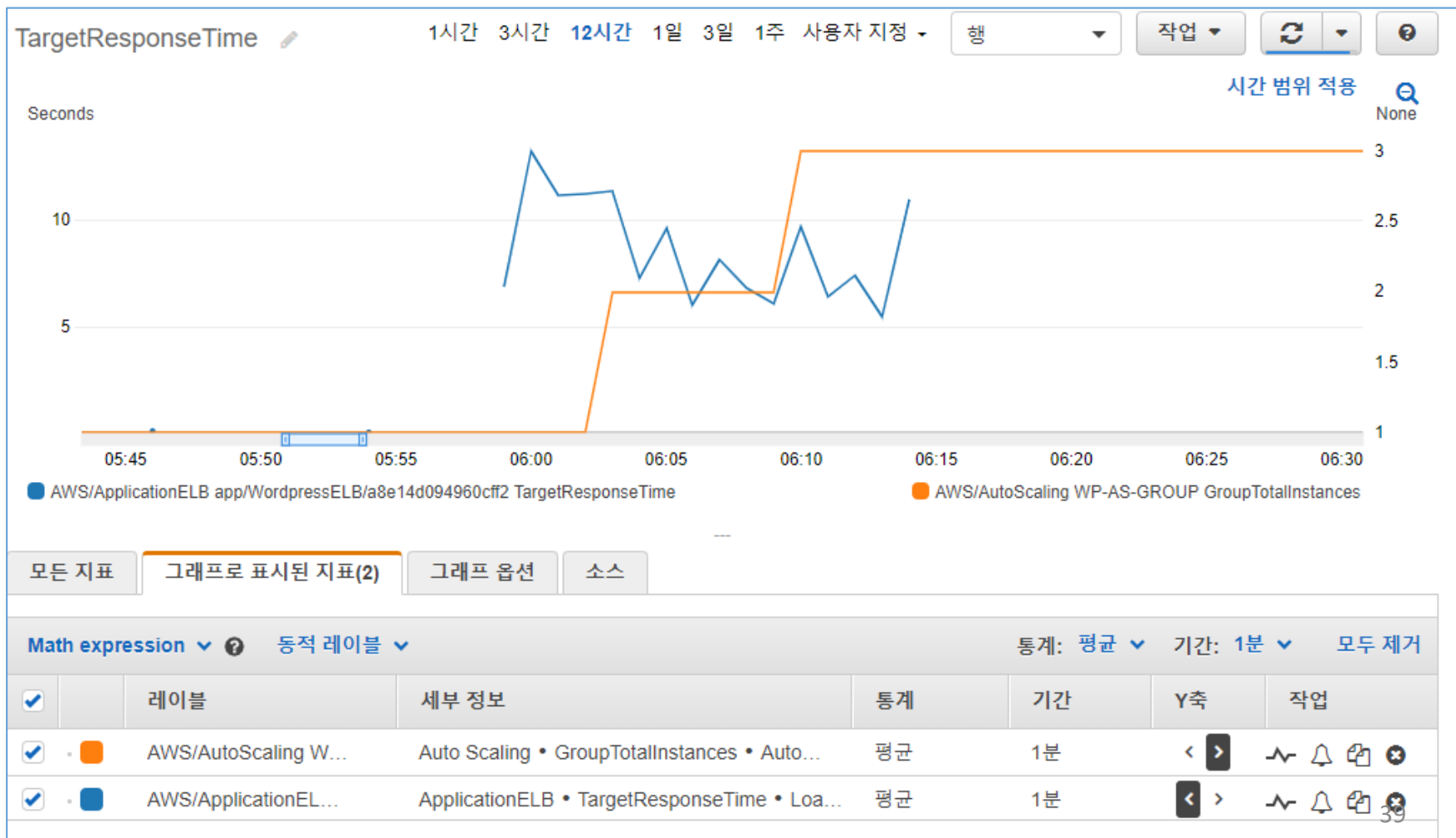
5. 다시 지표에서 확인

- Instance 개수와 함께 확인
 - 인스턴스 개수가 2개로 늘어나면서 요청 처리수가 늘어났는데,
 - 3개로 늘어난 다음에 AB 테스트가 종료되어버려서,
그 이상 늘어나는 것을 확인할 수 없었음
 - 앞서 180 으로 주었던 유휴 시간 간격을 좀더 줄여도 될 것 같고,
AB 테스트를 보다 길게 수행해보아도 좋을 것



6. ELB 모니터링: 대상 응답 시간

- 평균 응답 시간이 10초 이상으로 높다가 3번째 인스턴스가 생성되고 5s 이하로 줄어들었음
- 전체 서비스 성능이 향상되었고, 사용자의 만족도가 높아졌으리라 예상 가능



개인 과제 #9 (1/2)

- 제출: 하나의 ppt에 각 캡처 파일을 넣은 후, PDF 파일로 변환해 LMS 제출
 - 파일 이름: 학번.pdf
 - Page #1: 제목, 학번, 이름
 - Page #2~: 아래 캡처 화면 하나씩
- Auto Scaling 실습 내용 진행 (1/2)
 - 시작 템플릿 세부 정보 (#2)
 - Auto Scaling 그룹 세부 정보 (#3)
 - 부하 테스트 결과 (#4~~)
 - 인스턴스 용량이 3개까지 늘어나는 것을 확인할 것
 - AB 수행 결과 화면, AS 그룹 활동 기록, 인스턴스 CPU 사용량 그래프 (CloudWatch), ELB 대상 응답 시간
- 기한: 12/7 (월) 23:59
 - 지각 감점: 5%p / 12H
 - 1주 이후 제출 차단

