



스케일링 플랜 사용 설명서

AWS Auto Scaling



AWS Auto Scaling: 스케일링 플랜 사용 설명서

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon의 상표 및 트레이드 드레스는 Amazon 외 제품 또는 서비스와 함께 사용하여 고객에게 혼란을 초래하거나 Amazon을 폄하 또는 브랜드 이미지에 악영향을 끼치는 목적으로 사용할 수 없습니다. Amazon이 소유하지 않은 기타 모든 상표는 Amazon과 제휴 관계이거나 관련이 있거나 후원 관계와 관계없이 해당 소유자의 자산입니다.

Table of Contents

크기 조정 계획이란 무엇입니까?	1
지원되는 리소스	1
크기 조정 계획 기능 및 이점	1
시작하는 방법	2
크기 조정 계획 사용	2
리전별 가용성	3
요금	3
크기 조정 계획 작동 방식	4
모범 사례	6
기타 고려 사항	6
ActiveWithProblems 오류 방지	7
시작하기	9
1단계: 확장 가능한 리소스 검색	10
사전 조건	10
새 크기 조정 계획에 Auto Scaling 그룹 추가	10
확장 가능한 리소스 검색에 대해 자세히 알아보기	12
2단계: 크기 조정 전략 지정	13
3단계: 고급 설정 구성(선택 사항)	16
일반 설정	16
동적 크기 조정 설정	19
예측 크기 조정 설정	19
4단계: 크기 조정 계획 생성	21
(선택 사항) 리소스에 대한 크기 조정 정보 보기	21
5단계: 정리	23
Auto Scaling 그룹 삭제	24
6단계: 다음 단계	24
보안	25
VPC 엔드포인트(AWS PrivateLink)	25
크기 조정 계획용 인터페이스 VPC 엔드포인트 생성	26
크기 조정 계획용 VPC 엔드포인트 정책 생성	26
엔드포인트 마이그레이션	27
데이터 보호	28
ID 및 액세스 관리	29
액세스 제어	29

크기 조정 계획에 IAM을 사용하는 방법	29
서비스 연결 역할	33
자격 증명 기반 정책 예시	34
규정 준수 확인	40
인프라 보안	41
할당량	43
관련 정보	44
문서 기록	46
.....	xlviii

크기 조정 계획이란 무엇입니까?

크기 조정 계획을 사용하여 관련 또는 연결된 크기 조정 가능한 리소스에 대한 Auto Scaling을 몇 분 만에 구성할 수 있습니다. 예를 들어 태그를 사용하여 프로덕션, 테스트 또는 개발과 같은 카테고리로 리소스를 그룹화할 수 있습니다. 그런 다음 각 카테고리에 속하는 크기 조정 가능한 리소스에 대한 크기 조정 계획을 검색 및 설정할 수 있습니다. 또는 클라우드 인프라에 포함된 AWS CloudFormation 경우 리소스 컬렉션을 만드는 데 사용할 스택 템플릿을 정의할 수 있습니다. 그런 다음 각 스택에 속하는 크기 조정 가능한 리소스에 대한 크기 조정 계획을 생성합니다.

지원되는 리소스

AWS Auto Scaling 다음 서비스 및 리소스에 대한 규모 조정 계획 사용을 지원합니다.

- Amazon Aurora - Aurora DB 클러스터에 프로비저닝된 Aurora 읽기 전용 복제본 수를 늘리거나 줄입니다.
- Amazon EC2 Auto Scaling - Auto Scaling 그룹의 용량을 원하는대로 늘리거나 줄여 EC2 인스턴스를 시작하거나 종료합니다.
- Amazon Elastic Container Service - Amazon ECS에서 원하는 태스크 수를 늘리거나 줄입니다.
- Amazon DynamoDB - DynamoDB 테이블 또는 전역 보조 인덱스의 프로비저닝된 읽기 및 쓰기 용량을 늘리거나 줄입니다.
- 스팟 플릿 - 스팟 플릿의 대상 용량을 늘리거나 줄여 EC2 인스턴스를 시작하거나 종료합니다.

크기 조정 계획 기능 및 이점

크기 조정 계획에서는 다음과 같은 기능을 제공합니다.

- 리소스 검색 — 애플리케이션에서 확장 가능한 리소스를 찾는 데 도움이 되는 자동 리소스 검색 AWS Auto Scaling 기능을 제공합니다.
- 동적 크기 조정 - 크기 조정 계획은 Amazon EC2 Auto Scaling 및 Application Auto Scaling 서비스를 사용하여 트래픽 또는 워크로드의 변화를 처리하기 위해 크기 조정 가능한 리소스의 용량을 조정합니다. 동적 크기 조정 지표는 표준 사용률이나 처리량 지표, 또는 사용자 지정 지표일 수 있습니다.
- 기본 제공 크기 조정 권장 사항 - AWS Auto Scaling 은 성능, 비용 또는 둘 사이의 균형을 최적화하는 데 사용할 수 있는 권장 사항이 포함된 크기 조정 전략을 제공합니다.
- 예측 크기 조정 - 크기 조정 계획은 Auto Scaling 그룹에 대한 예측 크기 조정도 지원합니다. 이를 통해 정기적 스파이크가 발생할 때 Amazon EC2 용량 크기를 더욱 빠르게 조정할 수 있습니다.

⚠ Important

예측 크기 조정을 위해서만 크기 조정 계획을 사용하는 경우 Auto Scaling 그룹에서 직접 예측 크기 조정 정책을 세우는 것이 좋습니다. 최근에 도입된 이 옵션은 지표 집계를 사용하여 새로운 사용자 지정 지표를 만들거나 블루/그린 배포에서 기록 지표 데이터를 보존하는 등의 더 많은 기능을 제공합니다. 자세한 설명은 Amazon EC2 Auto Scaling 사용자 가이드의 [Amazon EC2 Auto Scaling 예측 스케일링](#)을 참조하세요.

시작하는 방법

다음 리소스를 사용하여 크기 조정 계획을 생성하고 사용할 수 있습니다.

- [크기 조정 계획 작동 방식](#)
- [크기 조정 계획 모범 사례](#)
- [크기 조정 계획 시작하기](#)

크기 조정 계획 사용

다음 인터페이스 중 하나를 사용하여 크기 조정 계획을 생성, 액세스 및 관리할 수 있습니다.

- AWS Management Console - 크기 조정 계획에 액세스할 때 사용할 수 있는 웹 인터페이스를 제공합니다. 예 가입한 경우 예 로그인하고 탐색 표시줄의 AWS Management Console 검색 상자를 사용하여 AWS Auto Scaling 검색한 다음 선택하여 AWS Auto Scaling 규모 조정 플랜에 액세스할 수 있습니다. AWS 계정
- AWS Command Line Interface (AWS CLI) — 다양한 명령을 제공하며 Windows AWS 서비스, macOS 및 Linux에서 지원됩니다. 시작하려면 [AWS Command Line Interface 사용 설명서](#)를 참조하세요. 자세한 내용은 AWS CLI 명령 참조의 [autoscaling-plans](#)를 참조하세요.
- AWS Tools for Windows PowerShell— PowerShell 환경에서 스크립트를 작성하는 사용자를 위해 다양한 AWS 제품에 대한 명령을 제공합니다. 시작하려면 [AWS Tools for Windows PowerShell 사용자 가이드](#)를 참조하세요. 자세한 설명은 [AWS Tools for PowerShell Cmdlet 참조](#)를 참조하세요.
- AWS SDK — 언어별 API 작업을 제공하고 서명 계산, 요청 재시도 처리, 오류 처리와 같은 많은 연결 세부 정보를 처리합니다. 자세한 내용은 [AWS SDK](#)를 참조하십시오.
- HTTPS API - HTTPS 요청을 사용하여 호출하는 하위 수준의 API 작업을 제공합니다. 자세한 내용은 [AWS Auto Scaling API 참조](#)를 참조하십시오.

- AWS CloudFormation— 템플릿을 사용한 규모 조정 계획 생성을 지원합니다. CloudFormation 자세한 내용은 AWS CloudFormation 사용 설명서의 [AWS::AutoScalingPlans::ScalingPlan](#) 참조를 참조하십시오.

리전별 가용성

AWS Auto Scaling API는 여러 AWS 리전 가지로 제공되며 각 지역에 대한 엔드포인트를 제공합니다. 현재 API를 사용할 수 있는 모든 지역 및 엔드포인트 목록은 중국 엔드포인트 및 [할당량을](#) 참조하십시오. [AWS Auto Scaling](#) . [AWS AWS](#)

요금

모든 크기 조정 계획 기능을 사용할 수 있습니다. 기능은 사용하는 기타 AWS 클라우드 리소스에 대한 CloudWatch 서비스 요금 외에 추가 비용 없이 제공됩니다.

Note

예측 조정 기능은 CloudWatch [GetMetricData](#) 작업을 기반으로 용량 예측을 위해 기간별 지표 데이터를 수집하므로 비용이 발생합니다. 하지만 규모 조정 플랜 대신 Amazon EC2 Auto Scaling 조정 정책을 사용하여 예측 조정을 활성화하면 호출에 대한 요금이 부과되지 않습니다. [GetMetricData](#)

크기 조정 계획 작동 방식

AWS Auto Scaling에서는 크기 조정 계획을 사용하여 리소스 크기 조정을 위한 지침 집합을 구성할 수 있습니다. AWS CloudFormation을 사용하거나 확장 가능한 리소스에 태그를 추가하는 경우 애플리케이션별로 다양한 리소스 세트에 대한 크기 조정 계획을 설정할 수 있습니다. AWS Auto Scaling 콘솔은 각 리소스에 사용자 지정된 크기 조정 전략에 대한 권장 사항을 제공합니다. 크기 조정 계획을 생성하면 동적 크기 조정 및 예측 크기 조정 방법을 결합하여 크기 조정 전략을 지원합니다.

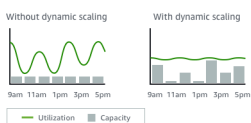
크기 조정 전략이란 무엇입니까?

AWS Auto Scaling에서는 크기 조정 전략에 따라 크기 조정 계획의 리소스 사용률을 최적화합니다. 가용성, 비용 또는 둘의 균형에 맞게 최적화할 수 있습니다. 또는 정의한 측정치 및 임계값에 따라 자체 사용자 지정 전략을 생성할 수도 있습니다. 리소스나 리소스 유형마다 별도의 전략을 설정할 수 있습니다.



동적 크기 조정이란 무엇입니까?

동적 크기 조정은 크기 조정 계획에 포함된 리소스의 대상 추적 크기 조정 정책을 생성합니다. 이 크기 조정 정책은 리소스 사용률이 달라질 때마다 리소스 용량을 크기 조정합니다. 충분한 용량을 제공하여 크기 조정 전략에 지정된 대상 값으로 사용률을 유지하기 위해서입니다. 이 과정은 온도 조절기를 사용하여 집안 온도를 유지하는 방법과 비슷합니다. 온도를 선택하면 온도 조절기에서 나머지 작업을 자동으로 수행합니다.

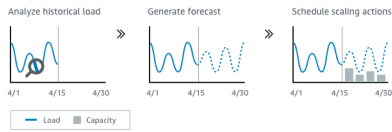


예를 들어 Amazon Elastic Container Service(Amazon ECS) 서비스가 실행하는 태스크 수를 CPU의 75%로 유지하도록 크기 조정 계획을 구성할 수 있습니다. 서비스의 CPU 사용률이 75%를 초과하면, 즉 서비스를 위해 예약된 CPU의 75% 이상이 사용될 경우 크기 조정 정책에서 서비스에 다른 태스크를 추가하여 로드 증가를 지원합니다.

예측 크기 조정이란 무엇입니까?

예측 크기 조정은 기계 학습을 사용하여 각 리소스의 과거 워크로드를 분석하고 향후 로드를 정기적으로 예상합니다. 이는 일기 예성 작동 방식과 유사합니다. 예상을 사용할 경우 예측 크기 조정은 크기 조정 작업을 예약함으로써 애플리케이션에 필요하게 될 리소스 용량이 사용 가능한지 미리 확인합니다.

동적 크기 조정과 마찬가지로 예측 크기 조정은 크기 조정 전략에 지정된 대상 값으로 사용률을 유지합니다.



예를 들어 예측 크기 조정을 활성화하고, Auto Scaling 그룹의 평균 CPU 사용률을 50퍼센트로 유지하도록 크기 조정 전략을 구성할 수 있습니다. 예측에 따르면 매일 8시에 트래픽이 급증합니다. 크기 조정 계획은 미래의 예약 크기 조정 작업을 생성하여 Auto Scaling 그룹이 트래픽을 처리할 수 있도록 준비시킵니다. 그러면 리소스 사용률을 항상 50퍼센트에 가깝게 유지하기 위한 용량 크기 조정을 목표로, 애플리케이션 성능을 일정하게 유지할 수 있습니다.

다음은 예측 크기 조정을 이해하기 위한 핵심 개념입니다.

- **로드 예측:** AWS Auto Scaling은 지정한 로드 지표에 대해 최대 14일 동안의 기록을 분석하여, 향후 2일 동안의 수요를 예측합니다. 이 데이터는 1시간 간격으로 제공되며 매일 업데이트됩니다.
- **예약 크기 조정 작업:** AWS Auto Scaling은 로드 예측과 일치하도록 용량을 늘리거나 줄이는 크기 조정 작업을 예약합니다. 예약된 시간에 AWS Auto Scaling은 최소 용량을 예약 크기 조정 작업에서 지정한 값으로 업데이트합니다. 이렇게 하는 목적은 크기 조정 전략에서 지정한 목표값으로 리소스 사용률을 유지하기 위한 것입니다. 예측보다 더 많은 용량이 애플리케이션에 필요하다면 동적 크기 조정에서 용량을 추가합니다.
- **최대 용량 동작:** 자동 크기 조정의 최소 및 최대 용량 제한이 각 리소스에 적용됩니다. 하지만 예측 용량이 최대 용량보다 높을 때 애플리케이션이 최대 용량을 초과하여 용량을 늘릴 수 있는지를 제어할 수 있습니다.

i Note

이제 Auto Scaling 그룹의 예측 크기 조정 정책을 대신 사용할 수 있습니다. 자세한 내용은 Amazon EC2 Auto Scaling 사용 설명서의 [Amazon EC2 Auto Scaling 예측 크기 조정](#)을 참조하세요.

크기 조정 계획 모범 사례

다음 모범 사례를 통해 크기 조정 계획을 최대한 활용할 수 있습니다.

- 시작 템플릿 또는 시작 구성을 생성할 때 세부 모니터링을 활성화하여 EC2 인스턴스에 대한 CloudWatch 지표 데이터를 1분 간격으로 가져오도록 하면 로드 변경 시 빠른 응답이 가능합니다. 간격을 5분으로 하면 응답 시간이 느려질 뿐만 아니라 오래된 지표 데이터를 기준으로 크기를 조정하게 됩니다. 기본적으로 EC2 인스턴스는 기본 모니터링, 즉 5분 간격으로 인스턴스의 지표 데이터를 사용하도록 설정됩니다. 추가 요금을 내면 세부 모니터링을 활성화하여 인스턴스의 지표 데이터를 가져오는 간격을 1분으로 바꿀 수 있습니다. 자세한 내용을 알아보려면 [Amazon EC2 Auto Scaling 사용 설명서의 Auto Scaling 인스턴스에 대한 모니터링 구성](#)을 참조하세요.
- Auto Scaling 그룹 지표도 활성화하는 것이 좋습니다. 그렇지 않으면 크기 조정 계획 생성 마법사 완료 시 제공되는 용량 예측 그래프에 실제 용량 데이터가 표시되지 않습니다. 자세한 내용은 [Amazon EC2 Auto Scaling 사용 설명서의 Auto Scaling 그룹 및 인스턴스에 대한 CloudWatch 지표 모니터링](#)을 참조하세요.
- Auto Scaling 그룹에서 사용하는 인스턴스 유형을 확인하고 버스트 가능 성능 인스턴스 유형 사용에 주의해야 합니다. T3 및 T2 인스턴스와 같은 성능 버스트 가능 Amazon EC2 인스턴스는 기본 수준의 CPU 성능과 함께 워크로드에서 필요한 만큼 성능을 높이는 버스트 기능을 제공하도록 설계되었습니다. 크기 조정 계획에서 지정한 목표 사용률에 따라, 기준선을 초과하게 실행하면 CPU 크레딧이 부족해져서 성능이 제한될 수 있습니다. 자세한 내용은 [성능 순간 확장 가능 인스턴스에 대한 CPU 크레딧 및 기준 성능](#)을 참조하세요. 이러한 인스턴스를 unlimited(으)로 구성하려면 Linux 인스턴스용 Amazon EC2 사용 설명서의 [Auto Scaling 그룹을 사용하여 성능 순간 확장 가능 인스턴스를 무제한으로 시작하기](#)를 참조하세요.

기타 고려 사항

Note

2021년 5월에 출시된 예측 크기 조정의 최신 버전이 있습니다. 이 버전에 도입된 일부 기능은 크기 조정 계획에서 사용할 수 없으며 이러한 기능에 액세스하려면 Auto Scaling 그룹에 직접 설정된 예측 크기 조정 정책을 사용해야 합니다. 자세한 내용은 [Amazon EC2 Auto Scaling 예측 크기 조정](#)을 참조하세요.

다음과 같은 사항들을 고려하세요.

- 예측 크기 조정은 로드 예측을 사용하여 미래의 용량을 예약합니다. 예측의 품질은 로드의 주기 및 훈련된 예측 모델의 적용 가능성에 따라 달라집니다. 예측 크기 조정은 예측 데이터의 품질과 예측 데이터를 통해 생성된 크기 조정 작업을 평가하기 위해 예측 전용 모드로 실행할 수 있습니다. 예측 크기 조정 모드는 크기 조정 계획을 생성할 때에만 Forecast only(예측 전용) 모드로 설정할 수 있으며, 예측 품질 평가를 완료하면 이 모드를 Forecast and scale(예측 및 크기 조정) 모드로 변경할 수 있습니다. 자세한 내용은 [예측 크기 조정 설정](#) 및 [예측 모니터링 및 평가](#) 단원을 참조하세요.
- 예측 크기 조정에 대해 다른 지표를 지정하도록 선택할 경우 크기 조정 지표와 로드 지표가 밀접한 상호 연관성이 있어야 합니다. 지표 값은 Auto Scaling 그룹의 인스턴스 수에 비례하여 증가하고 감소합니다. 이렇게 하면 지표 데이터를 사용하여 인스턴스 수를 비례적으로 확장하거나 축소할 수 있습니다. 예를 들어 로드 지표는 총 요청 수이고, 크기 조정 지표는 평균 CPU 사용률입니다. 총 요청 수가 50% 증가할 경우, 용량이 변경되지 않는다면 평균 CPU 사용률도 50% 증가해야 합니다.
- 크기 조정 계획을 생성하려면 먼저 크기 조정 작업을 생성했던 콘솔에 액세스하여 더 이상 필요 없는 이전 예약 크기 조정 작업을 삭제해야 합니다. AWS Auto Scaling은 기존에 예약된 크기 조정 작업과 겹치는 예측 크기 조정 작업을 생성하지 않습니다.
- 사용자 지정한 최소 및 최대 용량 설정과 동적 크기 조정에 사용되는 기타 설정은 다른 콘솔에 표시됩니다. 그러나 크기 조정 계획을 생성한 후 다른 콘솔에서 이러한 설정을 수정하지 않는 것이 좋습니다. 다른 콘솔로부터 크기 조정 계획에 이러한 업데이트가 제공되지 않기 때문입니다.
- 크기 조정 계획에는 다양한 서비스의 리소스가 포함될 수 있으며 다만 각 리소스는 한 번에 하나씩만 크기 조정 계획에 포함될 수 있습니다.

ActiveWithProblems 오류 방지

크기 조정 계획을 생성하거나 리소스를 크기 조정 계획에 추가할 때 "ActiveWithProblems" 오류가 발생할 수 있습니다. 크기 조정 계획이 활성화되어 있지만 하나 이상의 리소스에 대한 크기 조정 구성을 적용할 수 없는 경우 오류가 발생합니다.

이러한 현상은 일반적으로 리소스에 이미 크기 조정 정책이 있거나 Auto Scaling 그룹이 예측 크기 조정에 대한 최소 요구 사항을 충족하지 못하기 때문에 발생합니다.

리소스에 다양한 서비스 콘솔의 크기 조정 정책이 이미 있는 경우 AWS Auto Scaling은 기본적으로 이러한 다른 크기 조정 정책을 덮어쓰거나 새 크기 조정 정책을 생성하지 않습니다. 선택적으로 기존 크기 조정 정책을 삭제하고 AWS Auto Scaling 콘솔에서 생성된 목표 추적 크기 조정 정책으로 바꿀 수 있습니다. 이렇게 하려면 덮어쓸 크기 조정 정책이 있는 각 리소스에 대해 Replace external scaling policies(외부 크기 조정 정책 바꾸기) 설정을 활성화합니다.

예측 크기 조정을 사용하는 경우 새 Auto Scaling 그룹을 생성한 후 24시간 동안 기다렸다가 예측 크기 조정을 구성하는 것이 좋습니다. 초기 예측을 생성하려면 최소한 24시간의 과거 데이터가 있어야 합니다.

다. 그룹에 24시간 미만의 과거 데이터가 있고 예측 크기 조정이 활성화된 경우, 그룹에서 필요한 양의 데이터를 수집한 후 다음 예측 기간까지 크기 조정 계획에서 예측 데이터를 생성할 수 없는 문제가 발생합니다. 그러나 크기 조정 계획을 편집하고 저장하여 24시간 데이터를 사용할 수 있는 즉시 예측 프로세스를 재시작할 수도 있습니다.

크기 조정 계획 시작하기

애플리케이션에 사용할 크기 조정 계획을 생성하려면 먼저 AWS 클라우드에서 실행할 애플리케이션을 철저히 검토하세요. 다음에 유의하세요.

- 다른 콘솔에서 생성한 기존 크기 조정 정책이 있는지 여부 크기 조정 계획을 생성할 때 기존 크기 조정 정책을 교체하거나 (값의 변화 없이) 해당 정책을 유지할 수 있습니다.
- 전체 리소스를 기준으로 애플리케이션의 확장 가능한 각 리소스에 대해 적합한 목표 사용률입니다. 예를 들어 Auto Scaling 그룹의 EC2 인스턴스에서 사용 가능한 CPU와 비교하여 사용할 것으로 예상되는 CPU의 양입니다. 또는 프로비저닝된 처리량 모델을 사용하는 DynamoDB와 같은 서비스의 경우 테이블 또는 인덱스가 사용할 것으로 예상되는 읽기 및 쓰기 작업의 양을 사용 가능한 처리량과 비교하여 계산합니다. 즉, 프로비저닝된 용량 대 사용한 용량의 비율입니다. 크기 조정 계획을 생성한 후 언제든지 목표 사용률을 변경할 수 있습니다.
- 서버 시작 및 구성에 걸린 시간 이를 알면 각 EC2 인스턴스가 시작 후 워밍업되도록 시간을 구성하여 이전 서버가 계속 실행되는 동안 새 서버가 시작되지 않도록 하는 데 도움이 됩니다.
- (새로 만든 Auto Scaling 그룹을 사용할 경우) 지표 기록이 예측 크기 조정에 사용할 수 있을 만큼 긴지 여부를 나타냅니다. 일반적으로 14일의 기록 데이터가 있으면 더 정확한 예측이 생성됩니다. 최소 기간은 24시간입니다.

애플리케이션을 잘 이해할수록 크기 조정 계획을 더 효과적으로 구축할 수 있습니다.

다음 태스크는 크기 조정 계획에 익숙해지는 데 도움이 됩니다. 단일 Auto Scaling 그룹에 대한 크기 조정 계획을 생성하고 예측 크기 조정 및 동적 크기 조정을 활성화합니다.

작업

- [1단계: 확장 가능한 리소스 검색](#)
- [2단계: 크기 조정 전략 지정](#)
- [3단계: 고급 설정 구성\(선택 사항\)](#)
- [4단계: 크기 조정 계획 생성](#)
- [5단계: 정리](#)
- [6단계: 다음 단계](#)

1단계: 확장 가능한 리소스 검색

이 단원에서는 AWS Auto Scaling 콘솔에서 크기 조정 계획을 생성하는 방법에 대한 실습이 포함되어 있습니다. 크기 조정 계획을 처음 생성하는 경우, Amazon EC2 Auto Scaling 그룹을 사용하여 샘플 크기 조정 계획을 생성해서 시작하는 게 좋습니다.

사전 조건

크기 조정 계획 사용을 연습하려면 Auto Scaling 그룹을 생성합니다. Auto Scaling 그룹에서 하나 이상의 Amazon EC2 인스턴스를 시작합니다. 자세한 내용은 Amazon EC2 Auto Scaling 사용 설명서의 [Amazon EC2 Auto Scaling 시작하기](#)를 참조하세요.

CloudWatch 지표가 활성화된 Auto Scaling 그룹을 사용하여 크기 조정 계획 생성 마법사를 완료할 때 제공되는 용량 데이터 그래프를 확보할 수 있습니다. 자세한 내용은 Amazon EC2 Auto Scaling 사용 설명서의 [Auto Scaling 그룹 지표 활성화](#)를 참조하세요.

가능한 경우 예측 크기 조정 기능에 CloudWatch 지표 데이터를 사용할 수 있도록 며칠 이상 로드를 생성합니다.

크기 조정 계획을 사용하는 데 필요한 권한이 있는지 확인합니다. 자세한 내용은 [크기 조정 계획에 사용되는 Identity and Access Management](#) 섹션을 참조하세요.

새 크기 조정 계획에 Auto Scaling 그룹 추가

콘솔에서 크기 조정 계획을 생성하면 확장 가능한 리소스를 우선적으로 찾을 수 있습니다. 시작하기 전에 다음 요구 사항을 충족하는지 확인합니다.

- Auto Scaling 그룹을 생성하고 이전 단원에서 설명한 것과 같이 EC2 인스턴스를 하나 이상 시작했습니다.
- 생성한 Auto Scaling 그룹이 24시간 이상 존재했습니다.

크기 조정 계획을 생성하려면

1. <https://console.aws.amazon.com/autoscaling/>에서 AWS Auto Scaling 콘솔을 엽니다.
2. 화면 상단의 탐색 모음에서 Auto Scaling 그룹을 생성할 때 사용한 것과 동일한 리전을 선택합니다.
3. 시작 페이지에서 시작하기(Get started)를 선택합니다.

Management Tools

AWS Auto Scaling

helps you quickly and easily scale multiple resources

AWS Auto Scaling enables you to quickly discover all of the scalable resources underlying your application and set up application scaling in minutes using built-in scaling recommendations.

Create a scaling plan

Optimize your application in just a few steps


Get started

Pricing

AWS Auto Scaling is free.

AWS Auto Scaling is enabled by Amazon CloudWatch and carries no additional fees. Service fees for your application resources and Amazon CloudWatch apply.

How it works



- 확장 가능 리소스 찾기(Find scalable resources) 페이지에서 다음 중 하나를 수행합니다.
 - CloudFormation 스택으로 검색(Search by CloudFormation stack)을 선택하고, 사용할 AWS CloudFormation 스택을 선택합니다.
 - 태그로 검색(Search by tag)을 선택합니다. 이후 각 태그에 대해 키(Key)에서 태그 키를 선택하고, (값)Value에서 태그 값을 선택합니다. 태그를 추가하려면 Add another row(다른 행 추가)를 선택합니다. 태그를 제거하려면 Remove(제거)를 선택합니다.
 - EC2 Auto Scaling 그룹 선택(Choose EC2 Auto Scaling groups)을 선택하고, 하나 이상의 Auto Scaling 그룹을 선택합니다.

Note

입문용 자습서의 경우 EC2 Auto Scaling 그룹 선택(Choose EC2 Auto Scaling groups)을 선택하고, 생성한 Auto Scaling 그룹을 선택합니다.

Choose a method

Search by CloudFormation stack
Search for resources provisioned by an AWS CloudFormation stack.

Search by tag
Search for resources by tags applied to them.

Choose EC2 Auto Scaling groups
Choose one or more Auto Scaling groups to include in your scaling plan.

Choose Auto Scaling groups Info

Auto Scaling groups

Choose Auto Scaling groups ▼

my-auto-scaling-group ✕

5. 다음(Next)을 선택하여 크기 조정 계획 생성 프로세스를 계속합니다.

확장 가능한 리소스 검색에 대해 자세히 알아보기

샘플 크기 조정 계획을 이미 생성했으며 추가 계획을 생성하려는 경우, 다음 시나리오에서 CloudFormation 스택 또는 태그 집합을 사용하는 방법을 참조하세요. 콘솔을 사용하여 크기 조정 계획을 생성할 때 확장 가능한 리소스를 검색하기 위해 CloudFormation 스택으로 검색(Search by CloudFormation stack) 또는 태그로 검색(Search by tag) 옵션을 선택할지를 이 섹션을 이용해 결정할 수 있습니다.

크기 조정 계획 생성 마법사의 1단계에서 CloudFormation 스택으로 검색(Search by CloudFormation stack) 또는 태그로 검색(Search by tag) 옵션을 선택하면 확장 가능 리소스가 크기 조정 계획에 사용 가능한 스택 또는 태그 집합과 연결될 수 있습니다. 크기 조정 계획을 정의할 때 이들 리소스 중 포함하거나 제외할 리소스를 선택할 수 있습니다.

CloudFormation 스택을 사용하여 확장 가능한 리소스 검색

CloudFormation을 사용하면 스택으로 작업하여 리소스를 프로비저닝합니다. 스택의 모든 리소스는 스택의 템플릿으로 정의합니다. 크기 조정 계획은 여러 리소스에 대한 확장을 쉽게 구성할 수 있도록 스택 위에 오케스트레이션 계층을 추가합니다. 크기 조정 계획이 없으면 각 확장 가능한 리소스에 대해 개별적으로 크기 조정을 설정해야 합니다. 즉, 리소스 프로비저닝 및 크기 조정 정책에 대한 순서를 파악하고 이러한 종속성이 작동하는 방식에 대한 미세한 차이를 이해해야 합니다.

AWS Auto Scaling 콘솔에서, 기존 스택을 선택하여 자동 크기 조정을 위해 구성할 수 있는 리소스를 검색하도록 할 수 있습니다. AWS Auto Scaling은 선택한 스택에 정의된 리소스만 검색합니다. 중첩 스택은 탐색하지 않습니다.

CloudFormation 스택에서 ECS 서비스를 검색하려면, AWS Auto Scaling 콘솔은 어떤 ECS 클러스터에서 해당 서비스를 실행 중인지 알아야 합니다. 이를 위해서는 ECS 서비스가 서비스를 실행하는 ECS 클러스터와 동일한 CloudFormation 스택에 있어야 합니다. 그렇지 않으면 기본 클러스터의 일부여야 합니다. 또한 정확하게 식별할 수 있도록 ECS 서비스 이름이 각 ECS 클러스터에서 고유해야 합니다.

CloudFormation에 대한 자세한 내용은 AWS CloudFormation 사용 설명서의 [AWS CloudFormation이란 무엇입니까?](#)를 참조하세요.

태그를 사용하여 확장 가능한 리소스 검색

태그는 태그 필터를 사용하여 AWS Auto Scaling 콘솔에서 확장 가능한 관련 리소스를 검색하는 데 사용할 수 있는 메타데이터를 제공합니다.

태그를 사용하여 다음 리소스를 검색합니다.

- Aurora DB 클러스터
- 오토 스케일링
- DynamoDB 테이블 및 글로벌 보조 인덱스

둘 이상의 태그로 검색할 경우 검색할 각 리소스에 나열된 태그가 모두 있어야 합니다.

태그 지정에 대한 자세한 내용은 다음 문서를 참조하세요.

- Amazon Aurora 사용 설명서에서 [Aurora 클러스터 태그 지정](#) 방법을 알아봅니다.
- Amazon EC2 Auto Scaling 사용 설명서에서 [Auto Scaling 그룹 태그 지정](#) 방법을 알아봅니다.
- Amazon DynamoDB 개발자 가이드에서 [DynamoDB 리소스 태그 지정](#) 방법을 알아봅니다.
- [AWS 리소스에 태그를 지정](#)하는 모범 사례에 대한 자세한 내용은 AWS 일반 참조을(를) 참조하세요.

2단계: 크기 조정 전략 지정

이전 단계에서 발견된 리소스에 대해 다음 절차에 따라 크기 조정 전략을 지정합니다.

각 리소스 유형에 대해 AWS Auto Scaling은 주어진 시간에 리소스가 얼마나 사용되고 있는지를 판단하는 데 가장 일반적으로 사용되는 지표를 선택합니다. 이 지표를 기반으로 애플리케이션의 성능을 최적화하기 위해 가장 적합한 크기 조정 전략을 선택합니다. 동적 크기 조정 기능과 예측 크기 조정 기능을 활성화하면 크기 조정 전략을 서로 공유할 수 있습니다. 자세한 내용은 [크기 조정 계획 작동 방식](#) 섹션을 참조하세요.

다음과 같은 크기 조정 전략을 사용할 수 있습니다.

- 가용성 최적화 - AWS Auto Scaling은 리소스를 자동으로 확장 및 축소하여 리소스 사용률을 40%로 유지합니다. 이 옵션은 애플리케이션에 긴급하면서도 때로는 예측 불가능한 확장이 필요할 때 유용합니다.
- 밸런스 가용성 및 비용 - AWS Auto Scaling은 리소스를 자동으로 확장 및 축소하여 리소스 사용률을 50%로 유지합니다. 이 옵션을 적용하면 높은 가용성을 유지하면서 비용을 절감할 수 있습니다.
- 비용 최적화 - AWS Auto Scaling은 리소스를 자동으로 확장 및 축소하여 리소스 사용률을 70%로 유지합니다. 예상치 못한 수요 변화가 있을 때 사용 중인 애플리케이션에서 버퍼 용량 감소를 처리할 수 있는 경우, 이 옵션은 비용을 낮추는 데 유용합니다.

예를 들면, 크기 조정 계획은 그룹 내 모든 인스턴스에 대한 CPU의 평균 사용량을 기반으로 Amazon EC2 인스턴스를 추가하거나 제거하도록 Auto Scaling 그룹을 구성합니다. 크기 조정 전략을 변경하여 가용성, 비용 또는 이 2가지의 조합에 대한 사용률을 최적화할지 여부를 선택합니다.

또는 기존 전략이 사용자의 요구를 충족하지 못할 경우, 사용자 지정 전략을 구성할 수 있습니다. 사용자 지정 전략을 사용하면 목표 사용률 값을 변경하거나 다른 지표를 선택하거나 또는 이 두 가지를 모두 수행할 수 있습니다.

Important

입문용 자습서의 경우, 다음 절차의 첫 번째 단계만 완료한 후 다음(Next)을 선택하여 계속 진행합니다.

크기 조정 전략을 지정하려면

1. 크기 조정 전략 지정(Specify scaling strategy) 페이지에서 크기 조정 계획 세부 정보(Scaling plan details)의 이름(Name)에 크기 조정 계획의 이름을 입력합니다. 크기 조정 계획의 이름은 리전에 대한 크기 조정 계획 집합 내에서 고유해야 합니다. 128자까지 입력할 수 있으며, 파이프 "|", 슬래시 "/" 또는 콜론 ":"을 포함하면 안 됩니다.
2. 포함된 모든 리소스는 리소스 유형별로 나열됩니다. Auto Scaling 그룹의 경우 다음을 수행합니다.

Auto Scaling groups (1)

Specify a scaling strategy for 1 Auto Scaling group.

 Include in scaling plan**Scaling strategy**

The strategy defines the scaling metric and target value used to scale your resources.

 Optimize for availability

Keep the average CPU utilization of your Auto Scaling groups at 40% to provide high availability and ensure capacity to absorb spikes in demand.

 Balance availability and cost

Keep the average CPU utilization of your Auto Scaling groups at 50% to provide optimal availability and reduce costs.

 Optimize for cost

Keep the average CPU utilization of your Auto Scaling groups at 70% to ensure lower costs.

 Custom

Choose your own scaling metric, target value, and other settings.

 Enable predictive scalingSupport your scaling strategy by continually forecasting load and proactively scheduling capacity ahead of when you need it. [Info](#) **Enable dynamic scaling**Support your scaling strategy by creating target tracking scaling policies to monitor your scaling metric and increase or decrease capacity as you need it. [Info](#)**▶ Configuration details**

a. 이 단계를 건너뛰어 기본 크기 조정 전략 및 지표를 사용합니다. 다른 크기 조정 전략이나 지표를 대신 사용하려면 다음 단계를 진행합니다.

i. 크기 조정 전략(Scaling strategy)에서 원하는 크기 조정 전략을 선택합니다.

입문용 자습서의 경우 가용성 최적화(Optimize for availability)를 선택합니다. 이렇게 하면 Auto Scaling 그룹의 평균 CPU 사용률이 40%로 유지됩니다.

ii. 사용자 지정(Custom)을 선택한 경우 구성 세부 정보(Configuration details)를 확장하여 원하는 지표 및 목표 값을 선택합니다.

- 크기 조정 지표(Scaling metric)에서 원하는 크기 조정 지표를 선택합니다.
- 대상 값(Target value)에서 1분 간격 동안 목표 사용률 또는 목표 처리량과 같은 원하는 목표 값을 선택합니다.
- 로드 지표(Load metric)[Auto Scaling 그룹 전용]의 경우, 예측 크기 조정에 사용할 원하는 로드 지표를 선택합니다.
- 외부 크기 조정 정책 바꾸기(Replace external scaling policies)를 선택하여 AWS Auto Scaling에서 이전에 크기 조정 계획 외부(예: 다른 콘솔)에서 생성된 크기 조정 정책을 삭제하고 이를 크기 조정 계획에 의해 생성된 새로운 대상 추적 조정 정책으로 바꿀 수 있는지 여부를 지정합니다.

b. (선택 사항) 기본적으로 Auto Scaling 그룹에 대해 예측 크기 조정이 활성화됩니다. Auto Scaling 그룹에 대한 예측 크기 조정을 끄려면 예측 크기 조정 활성화(Enable predictive scaling)를 선택 취소합니다.

- c. (선택 사항) 각 리소스 유형에 대해 동적 크기 조정이 기본적으로 활성화됩니다. 리소스 유형에 대한 동적 크기 조정을 끄려면 동적 크기 조정 활성화(Enable dynamic scaling)를 선택 취소합니다.
 - d. (선택 사항) 확장 가능한 여러 리소스가 검색되는 출처에 해당하는 애플리케이션 소스를 지정하면 기본적으로 모든 리소스 유형이 자동으로 크기 조정 계획에 포함됩니다. 크기 조정 계획에서 리소스의 유형을 생략하려면 Include in scaling plan(크기 조정 계획에 포함)을 비워둡니다.
3. (선택 사항) 다른 리소스 유형에 대한 크기 조정 전략을 지정하려면 이전 단계를 반복합니다.
 4. 완료되면 다음(Next)을 선택하여 크기 조정 계획 생성 프로세스를 계속합니다.

3단계: 고급 설정 구성(선택 사항)

이제 각 리소스 유형에 사용할 크기 조정 전략을 지정했으므로 Configure advanced settings(고급 설정 구성) 단계를 사용하여 리소스별로 기본 설정을 사용자 지정하도록 선택할 수 있습니다. 각 리소스 유형마다 사용자 지정할 수 있는 여러 설정 그룹이 있습니다. 그러나 대부분의 경우, 크기를 신중하게 조정해야 하는 최소 용량 또는 최대 용량의 값 외에는 기본 설정이 효율적입니다.

기본 설정을 유지하려면 이 절차를 건너뛵니다. 이러한 설정은 언제든지 크기 조정 계획을 편집하여 변경할 수 있습니다.

Important

입문용 자습서에서는 Auto Scaling 그룹의 최대 용량을 업데이트하고 예측 크기 조정을 예측 전용 모드로 활성화하기 위해 몇 가지 사항들을 변경해 보겠습니다. 자습서의 모든 설정을 사용자 지정할 필요는 없지만 각 섹션의 설정을 간략하게 살펴보도록 하겠습니다.


일반 설정

이 절차를 사용하여 이전 단계에서 지정한 설정을 리소스별로 확인하고 사용자 지정합니다. 또한 각 리소스의 최소 용량 및 최대 용량을 사용자 지정할 수도 있습니다.

일반 설정을 확인하고 사용자 지정하려면

1. 고급 설정 구성(Configure advanced settings) 페이지에서 단원 머리글 왼쪽에 있는 화살표를 선택하여 섹션을 확장합니다. 자습서에서 Auto Scaling 그룹 섹션을 확장합니다.
2. 이 자습서에서 사용 중인 Auto Scaling 그룹을 화면에 표시된 테이블에서 선택하세요.

3. Include in scaling plan(크기 조정 계획에 포함) 옵션을 선택된 상태로 둡니다. 이 옵션을 선택하지 않을 경우, 리소스는 크기 조정 계획에서 생략됩니다. 리소스를 적어도 하나 이상 포함하지 않을 경우, 크기 조정 계획을 생성할 수 없습니다.
4. 보기를 확장하여 일반 설정 섹션의 세부 정보를 보려면 섹션 머리글의 왼쪽에 있는 화살표를 선택하세요.
5. 다음 항목 중에 대한 설정을 선택할 수 있습니다. 이 자습서의 경우, Maximum capacity(최대 용량) 설정을 찾아 현재 값 대신에 3 값을 입력하세요.
 - 크기 조정 전략(Scaling strategy) - 가용성, 비용 또는 둘 다 균형에 맞게 최적화하거나 사용자 지정 전략을 지정할 수 있습니다.
 - 동적 크기 조정 활성화(Enable dynamic scaling) - 이 설정의 선택을 취소하면 대상 추적 크기 조정 구성을 사용하여 선택한 리소스를 크기 조정할 수 없습니다.
 - 예측 크기 조정 활성화(Enable predictive scaling) - [Auto Scaling 그룹만 해당] 이 설정의 선택을 취소하면 선택한 그룹은 예측 크기 조정을 사용하여 크기 조정할 수 없습니다.
 - 크기 조정 지표(Scaling metric) - 사용할 크기 조정 지표를 지정합니다. 사용자 지정(Custom)을 선택할 경우, 콘솔에서 사용할 수 있는 사전 정의 지표 대신에 사용할 사용자 지정 지표를 지정할 수 있습니다. 자세한 내용은 이 섹션의 다음 주제를 참조하세요.
 - 목표값(Target value) - 사용할 목표 사용률 값을 지정합니다.
 - 로드 지표(Load metric) - [Auto Scaling 그룹 전용] 사용할 로드 지표를 지정합니다. 사용자 지정(Custom)을 선택할 경우, 콘솔에서 사용할 수 있는 사전 정의 지표 대신에 사용할 사용자 지정 지표를 지정할 수 있습니다. 자세한 내용은 이 섹션의 다음 주제를 참조하세요.
 - 최소 용량(Minimum capacity) - 리소스의 최소 용량을 지정합니다. AWS Auto Scaling은 리소스가 이 크기보다 작아지지 않게 합니다.
 - 최대 용량(Maximum capacity) - 리소스의 최대 용량을 지정합니다. AWS Auto Scaling은 리소스가 이 크기보다 커지지 않게 합니다.

 Note

예측 크기 조정을 사용하면 예측 용량을 기준으로 사용할 다른 최대 용량 동작을 선택할 수 있습니다. 이 설정은 Predictive scaling settings(예측 크기 조정 설정) 섹션에 있습니다.

사용자 지정 지표

AWS Auto Scaling은 자동 크기 조정에 가장 많이 사용되는 지표를 제공합니다. 하지만 필요에 따라 콘솔의 지표 대신에 다른 지표의 데이터를 가져올 수도 있습니다. Amazon CloudWatch에는 선택할 수 있는 다양한 지표가 있습니다. CloudWatch를 사용하여 자체 지표를 게시할 수도 있습니다.

JSON을 사용하여 CloudWatch 사용자 지정 지표를 지정합니다. 이 지침을 따르기 전에 [Amazon CloudWatch 사용 설명서](#)를 숙지하는 것이 좋습니다.

사용자 지정 지표를 지정하려면 템플릿의 필수 파라미터 세트를 사용하여 JSON 형식의 페이로드를 생성하세요. CloudWatch의 각 파라미터에 대한 값을 추가합니다. 스케일링 계획의 고급 설정에서 Scaling metric(크기 조정 지표) 및 Load metric(로드 지표)에 대한 사용자 지정 옵션의 일부로서 템플릿을 제공합니다.

JSON은 다음 두 가지 방식으로 데이터를 나타냅니다.

- 객체: 순서가 지정되지 않은 이름-값 쌍 모음. 객체는 여는 중괄호({)와 닫는 중괄호(}) 안에 정의됩니다. 각 이름-값 쌍은 이름으로 시작하고 뒤에 콜론과 값이 옵니다. 이름-값 쌍은 쉼표로 구분됩니다.
- 배열: 순서가 지정된 값 모음. 배열은 여는 대괄호([)와 닫는 대괄호(]) 안에 정의됩니다. 배열의 항목들은 쉼표로 구분됩니다.

각 파라미터에 대한 샘플 값을 포함하는 JSON 템플릿의 예는 다음과 같습니다.

```
{
  "MetricName": "MyBackendCPU",
  "Namespace": "MyNamespace",
  "Dimensions": [
    {
      "Name": "MyOptionalMetricDimensionName",
      "Value": "MyOptionalMetricDimensionValue"
    }
  ],
  "Statistic": "Sum"
}
```

자세한 내용은 AWS Auto Scaling API 참조의 [사용자 지정 크기 조정 지표 사양](#) 및 [사용자 지정 로드 지표 사양](#)을 참조하세요.

동적 크기 조정 설정

이 절차를 사용하여 AWS Auto Scaling가 생성하는 대상 추적 크기 조정 정책에 대한 설정을 확인하고 사용자 지정합니다.

동적 크기 조정을 위한 설정을 확인하고 사용자 지정하려면

1. 보기를 확장하여 동적 크기 조정 설정(Dynamic scaling settings) 섹션의 세부 정보를 보려면 섹션 머리글의 왼쪽에 있는 화살표를 선택하세요.
2. 다음 항목에 대한 설정을 선택할 수 있습니다. 다만 이 자습서에서는 기본 설정을 사용합니다.
 - 외부 크기 조정 정책 바꾸기(Replace external scaling policies) - 이 설정의 선택을 취소하면 이 크기 조정 계획의 외부에서 생성된 기존 크기 조정 정책을 유지하며 새 크기 조정 정책이 생성되지 않습니다.
 - 축소 비활성화(Disable scale-in) - 이 설정의 선택을 취소하면 지정된 지표가 목표 값보다 낮을 때 리소스의 현재 용량을 줄이기 위한 자동 축소가 허용됩니다.
 - 휴지 기간(Cooldown) - 확장 및 축소 휴지 기간을 생성합니다. 이전 크기 조정 활동이 적용될 때까지 기다리는 시간을 휴지 기간이라고 합니다. 자세한 내용은 Application Auto Scaling 사용 설명서의 [휴지 기간](#)을 참조하세요. (이 설정은 리소스가 Auto Scaling 그룹인 경우에는 표시되지 않습니다.)
 - 인스턴스 워밍업(Instance warmup) - [Auto Scaling 그룹 전용] 새로 시작한 인스턴스가 CloudWatch 지표에 기여하기 전에 경과하는 시간의 분량을 조절합니다. 자세한 내용은 Amazon EC2 Auto Scaling 사용 설명서의 [인스턴스 워밍업](#)을 참조하세요.

예측 크기 조정 설정

리소스가 Auto Scaling 그룹에 속하는 경우, 이 절차를 사용하여 예측 크기 조정을 위해 AWS Auto Scaling이 사용하는 설정을 확인하고 사용자 지정합니다.

예측 크기 조정을 위한 설정을 확인하고 사용자 지정하려면

1. 보기를 확장하여 예측 크기 조정 설정(Predictive scaling settings) 섹션의 세부 정보를 보려면 섹션 머리글의 왼쪽에 있는 화살표를 선택하세요.
2. 다음 항목에 대한 설정을 선택할 수 있습니다. 이 자습서의 경우, Predictive scaling mode(예측 크기 조정 모드)를 Forecast only(예측 전용)로 변경하세요.

- 예측 크기 조정 모드(Predictive scaling mode) - 크기 조정 모드를 지정합니다. 기본값은 Forecast and scale(예측 및 크기 조정)입니다. Forecast only(예측 전용)로 변경하면 크기 조정 계획에서는 향후 용량을 예측하지만 크기 조정 작업을 적용하지는 않습니다.
 - 사전 시작 인스턴스(Pre-launch instances) - 확장 시 초기에 실행할 크기 조정 작업을 크기 조정합니다. 예를 들면, 예측에서 오전 10시에 용량을 추가하라고 되어 있고 버퍼 시간은 5분(300초)입니다. 해당 크기 조정 작업의 실행 시간은 오전 9시 55분입니다. 이는 인스턴스가 시작될 때부터 서비스가 시작될 때까지 몇 분이 걸릴 수 있는 Auto Scaling 그룹에 유용합니다. 실제 시간은 인스턴스 크기 및 완료할 시작 스크립트가 있는지 여부 등의 여러 요인에 따라 다릅니다. 기본값은 300초입니다.
 - 최대 용량 동작(Max capacity behavior) - 예측 용량이 현재 지정된 최대 용량에 근접하거나 이를 초과할 때 선택된 리소스를 최대 용량 이상으로 확장할 수 있는지 여부를 제어합니다. 기본값은 최대 용량 설정 적용(Enforce the maximum capacity setting)입니다.
 - 최대 용량 설정 적용(Enforce the maximum capacity setting) - AWS Auto Scaling은 리소스 용량을 최대 용량보다 크게 확장할 수 없습니다. 최대 용량은 하드 제한으로 적용됩니다.
 - 최대 용량을 예측 용량과 동일하게 설정(Set the maximum capacity to equal forecast capacity) - AWS Auto Scaling은 리소스 용량을 최대 용량보다 높게 예측 용량 이하로 확장할 수 있습니다.
 - 예측 용량보다 크게 최대 용량 확장(Increase maximum capacity above forecast capacity) - AWS Auto Scaling은 최대 용량보다 지정된 버퍼 값만큼 크게 리소스 용량을 확장할 수 있습니다. 예기치 않은 트래픽이 발생하는 경우 대상 추적 크기 조정 정책 추가 용량을 제공하는 것이 목적입니다.
 - 최대 용량 동작 버퍼(Max capacity behavior buffer) - 예측 용량보다 크게 최대 용량 확장(Increase maximum capacity above forecast capacity)을 선택한 경우, 예측 용량이 최대 용량에 근접하거나 이를 초과할 때 사용할 용량 버퍼의 크기를 선택합니다. 값은 예측 용량에 상대적인 비율로 지정됩니다. 예를 들어 10% 버퍼를 적용 시 예측 용량이 50이고 최대 용량이 40이라면 최대 유효 용량은 55입니다.
3. 설정을 사용자 지정한 후 다음(Next)을 선택합니다.

Note

변경 사항을 되돌리려면 해당 리소스를 선택하고 Revert to original(원상태로 되돌리기)을 선택하세요. 그러면 선택한 리소스가 크기 조정 계획 내에서 마지막으로 알려진 상태로 재 설정됩니다.

4단계: 크기 조정 계획 생성

검토 및 생성(Review and create) 페이지에서 크기 조정 계획에 대한 세부 정보를 검토한 후 Create scaling plan(크기 조정 계획 생성)을 선택합니다. 크기 조정 계획의 상태를 보여주는 페이지로 이동합니다. 크기 조정 계획은 리소스가 업데이트되는 동안 생성이 완료될 때까지 잠시 시간이 걸릴 수 있습니다.

예측 크기 조정을 적용할 때 AWS Auto Scaling는 지난 14일간의 지정된 로드 지표 기록을 분석하여(최소 24시간의 데이터 필요) 2일 전에 예측 데이터를 생성합니다. 그런 다음, 크기 조정 작업을 예약하여 예측 기간의 시간별 예측과 일치하도록 리소스 용량을 크기 조정합니다.

크기 조정 계획의 생성이 완료된 후, 크기 조정 계획(Scaling plans) 화면에서 그 이름을 선택하여 크기 조정 계획 세부 정보를 확인하세요.

(선택 사항) 리소스에 대한 크기 조정 정보 보기

하나의 리소스에 대해 생성된 크기 조정 정보를 보려면 이 절차를 사용하세요.

데이터는 다음과 같은 방식으로 표시됩니다.

- CloudWatch의 최신 지표 기록 데이터를 보여 주는 그래프
- AWS Auto Scaling의 데이터를 기반으로 로드 예측 및 용량 예측을 보여주는 예측 크기 조정 그래프.
- 리소스에 대해 예약된 모든 예측 크기 조정 작업을 목록으로 나열하는 테이블.

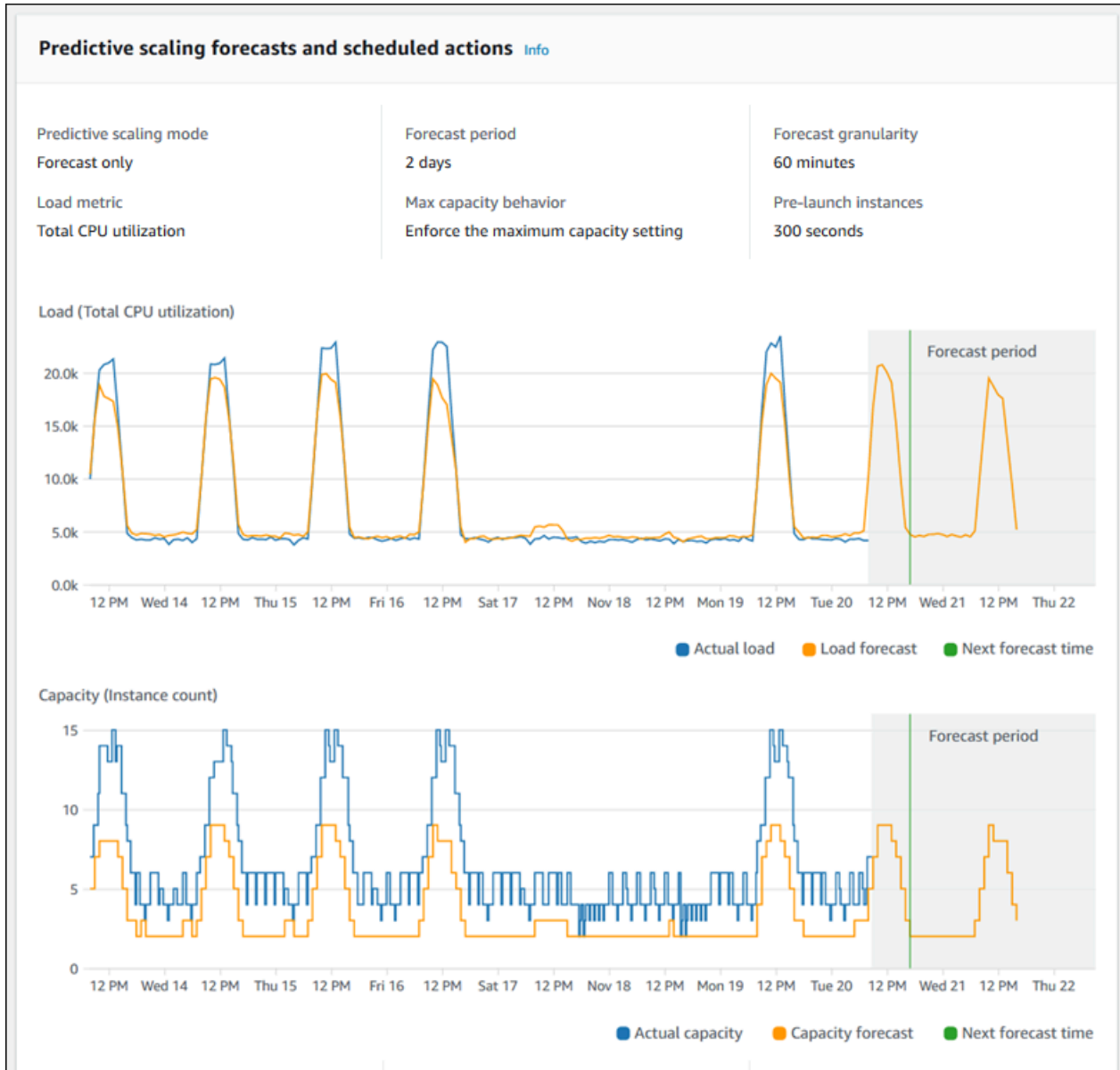
리소스에 대한 크기 조정 정보를 보려면

1. <https://console.aws.amazon.com/autoscaling/>에서 AWS Auto Scaling 콘솔을 엽니다.
2. 크기 조정 계획(Scaling plans) 페이지에서 크기 조정 계획을 선택합니다.
3. 크기 조정 계획 세부 정보(Scaling plan details) 페이지에서, 보려는 리소스를 선택합니다.

예측 모니터링 및 평가

크기 조정 계획이 실행되면 로드 예측, 용량 예측 및 크기 조정 작업을 모니터링하여 예측 크기 조정의 성능을 검사할 수 있습니다. 이 모든 데이터는 예측 크기 조정이 활성화된 모든 Auto Scaling 그룹의 AWS Auto Scaling 콘솔에 사용 가능합니다. 초기 예측을 수행하려면 크기 조정 계획 시 최소한 24시간 이상의 과거 로드 데이터가 필요하다는 점에 유의하세요.

다음 예제에선 각 그래프의 왼쪽에 과거 패턴이 표시됩니다. 오른쪽에는 예측 기간에 대한 크기 조정 계획에서 생성된 예측이 표시됩니다. 실제 값과 예측 값(파란색과 주황색으로 각각 표시됨)이 모두 나타나 있습니다.



AWS Auto Scaling은 데이터를 바탕으로 자동으로 학습합니다. 첫째, 로드를 예측합니다. 그 다음 용량 예측 계산을 통해 애플리케이션을 지원하기 위해 필요한 최소 인스턴스 수를 결정합니다. 용량 예측을 기반으로 AWS Auto Scaling은 예측되는 로드 변화에 앞서 Auto Scaling 그룹을 크기 조정하는 크기 조정 작업을 예약합니다. 동적 크기 조정이 활성화된 경우(권장 사항), Auto Scaling 그룹은 인스턴스 그룹의 현재 사용률을 기반으로 추가 용량을 확장(하거나 제거)할 수 있습니다.

예측 크기 조정이 얼마나 잘 수행되는지 평가할 때 실제 값과 예측값이 시간이 지남에 따라 얼마나 가깝게 일치하는지 모니터링합니다. 크기 조정 계획을 생성할 때 AWS Auto Scaling에서는 가장 최근의

실제 데이터를 기반으로 그래프를 제공합니다. 또한 다음 48시간 동안의 초기 예측도 제공합니다. 그러나 크기 조정 계획을 생성할 때 실제 데이터와 비교할 예측 데이터는 거의 없습니다. 과거 예측값과 실제 값을 비교하기 전에 몇 개 기간 동안 크기 조정 계획으로 예측값을 얻을 때까지 기다립니다. 며칠간의 일일 예측을 마친 후에는 실제 값과 비교할 수 있는 더 큰 예측값 샘플을 얻게 됩니다.

매일 발생하는 패턴의 경우 크기 조정 계획을 생성하고 예측 효과를 평가하는 시간 간격을 며칠로 단축할 수 있습니다. 그러나 이 시간은 최신 패턴 변경을 기반으로 한 예측 평가에는 부족합니다. 예를 들어 지난주 새 마케팅 캠페인을 시작한 Auto Scaling 그룹에 대한 예측을 살펴보고 있다고 가정해 보겠습니다. 이 캠페인으로 매주 같은 이틀 동안 웹 트래픽이 크게 증가합니다. 이러한 경우 예측의 효과를 평가하기 전에 이 그룹에서 1주일에 한두 가지 새 데이터를 수집하기를 기다리는 것이 좋습니다. 지표 데이터 수집만을 시작한 최신 Auto Scaling 그룹에 대해서도 동일한 권장 사항이 적용됩니다.

적절한 시간 동안 모니터링한 후에 실제 값과 예측값이 일치하지 않으면 로드 지표의 선택도 고려해야 합니다. 효과성을 위해 로드 지표는 이 Auto Scaling 그룹의 모든 인스턴스에 대한 전체 로드를 안정적이고 정확하게 나타내야 합니다. 로드 지표는 예측 크기 조정의 핵심입니다. 최적이지 아닌 로드 지표를 선택하면 예측 크기 조정으로 정확한 로드 및 용량 예측과 Auto Scaling 그룹에 대한 올바른 용량 조절을 예약하는 작업이 수행되지 않을 수 있습니다.

5단계: 정리

시작하기 자습서를 완료한 후 해당 크기 조정 계획을 유지할 수 있습니다. 하지만 이 크기 조정 계획을 적극적으로 활용하지 않는 경우에는 계정에 불필요한 요금이 발생하지 않도록 해당 계획을 삭제하는 것이 좋습니다.

크기 조정 계획을 삭제하면 대상 추적 크기 조정 정책 및 그와 연결된 CloudWatch 경보, 그리고 AWS Auto Scaling이 사용자를 대신하여 생성한 예측 크기 조정 작업이 삭제됩니다.

크기 조정 계획을 삭제하더라도 AWS CloudFormation 스택, Auto Scaling 그룹 또는 기타 확장 가능한 리소스는 삭제되지 않습니다.

크기 조정 계획을 삭제하려면

1. <https://console.aws.amazon.com/autoscaling/>에서 AWS Auto Scaling 콘솔을 엽니다.
2. Scaling plans(크기 조정 계획) 페이지에서 이 자습서에 대해 작성한 크기 조정 계획을 선택하고 삭제>Delete)를 선택하세요.
3. 확인 메시지가 나타나면 Delete(삭제)를 선택합니다.

크기 조정 정책을 삭제한 후에는 리소스를 원래의 용량으로 되돌릴 수 없습니다. 예를 들어 크기 조정 계획을 삭제하면서 Auto Scaling 그룹을 10개의 인스턴스로 크기 조정된 경우, 크기 조정 계획을 삭제한 후에도 그룹은 10개의 인스턴스로 크기 조정된 상태를 유지합니다. 개별 서비스마다 콘솔에 액세스하면 특정 리소스의 용량을 업데이트할 수 있습니다.

Auto Scaling 그룹 삭제

계정에서 Amazon EC2 요금이 발생하지 않도록 하려면 이 자습서를 위해 생성한 Auto Scaling 그룹도 삭제해야 합니다.

단계별 지침은 Amazon EC2 Auto Scaling 사용 설명서에서 [Auto Scaling 그룹 삭제](#)를 참조하세요.

6단계: 다음 단계

크기 조정 계획과 일부 기능에 익숙해졌으니 이제 AWS CloudFormation을 사용하여 고유한 크기 조정 계획 템플릿을 작성할 수 있습니다.

AWS CloudFormation 템플릿은 인프라 구성 요소 간의 상호 연결과 함께 애플리케이션 또는 서비스를 실행하는 데 필요한 Amazon Web Services 인프라를 설명하는 JSON 또는 YAML 형식의 텍스트 파일입니다. AWS CloudFormation을 사용하여 관련 리소스 모음을 스택으로 배포하고 관리할 수 있습니다. AWS CloudFormation에는 추가 요금이 없으며 애플리케이션 실행에 필요한 AWS 리소스 요금만 지불하면 됩니다. 템플릿에 정의된 모든 AWS 리소스를 사용할 수 있습니다. 자세한 내용은 AWS CloudFormation 사용 설명서의 [AWS CloudFormation 개념](#)을 참조하세요.

AWS CloudFormation 사용 설명서에서는 시작하는 데 도움이 되는 간단한 템플릿을 제공합니다. 샘플 템플릿은 AWS CloudFormation 템플릿 참조 문서의 [AWS::AutoScalingPlans::ScalingPlan](#) 섹션에서 예제로 이용 가능합니다. 샘플 템플릿은 단일 Auto Scaling 그룹에 대한 크기 조정 계획을 생성하며 예측 크기 조정 및 동적 크기 조정을 사용할 수 있도록 합니다.

자세한 내용은 AWS CloudFormation 사용 설명서에서 [AWS CloudFormation 시작하기](#)를 참조하세요.

크기 조정 계획 보안

클라우드 AWS 보안이 최우선 과제입니다. AWS 고객은 가장 보안에 민감한 조직의 요구 사항을 충족하도록 구축된 데이터 센터 및 네트워크 아키텍처의 혜택을 누릴 수 있습니다.

보안은 기업과 기업 간의 AWS 공동 책임입니다. [공동 책임 모델](#)은 이 사항을 클라우드의 보안 및 클라우드 내 보안으로 설명합니다.

- 클라우드 보안 - AWS 클라우드에서 AWS 서비스를 실행하는 인프라를 보호하는 역할을 합니다. AWS 또한 안전하게 사용할 수 있는 서비스를 제공합니다. 적용되는 규정 준수 프로그램에 대해 자세히 알아보려면 규정 준수 [프로그램별 범위 내 AWS 서비스 규정 준수 프로그램별](#) 참조하십시오. AWS Auto Scaling
- 클라우드에서의 보안 — 사용하는 AWS 서비스에 따라 책임이 결정됩니다. 또한 귀하는 데이터의 민감도, 회사 요구 사항, 관련 법률 및 규정을 비롯한 기타 요소에 대해서도 책임이 있습니다.

이 문서는 크기 조정 계획을 사용할 때 공동 책임 모델을 적용하는 방법을 이해하는 데 도움이 되며, 크기 조정 계획에 대한 액세스 권한을 관리하는 방법을 이해하는 데 도움이 됩니다.

주제

- [크기 조정 계획 및 인터페이스 VPC 엔드포인트\(AWS PrivateLink\)](#)
- [데이터 보호](#)
- [크기 조정 계획에 사용되는 Identity and Access Management](#)
- [규정 준수 확인](#)
- [인프라 보안](#)

크기 조정 계획 및 인터페이스 VPC 엔드포인트(AWS PrivateLink)

인터페이스 VPC 엔드포인트를 사용하도록 AWS Auto Scaling 구성하여 VPC의 보안 상태를 개선할 수 있습니다. 인터페이스 엔드포인트는 VPC와 네트워크 간의 모든 네트워크 트래픽을 제한하여 AWS Auto Scaling API에 비공개로 액세스할 수 있는 기술인 [에 의해 AWS PrivateLink](#)구동됩니다. AWS Auto Scaling AWS 인터페이스 엔드포인트를 사용하면 인터넷 게이트웨이, NAT 디바이스 또는 가상 프라이빗 게이트웨이가 필요 없습니다.

반드시 구성할 필요는 없지만 구성하는 AWS PrivateLink것이 좋습니다. AWS PrivateLink [및 VPC 엔드포인트에 대한 자세한 내용은 무엇입니까](#)를 참조하십시오. [AWS PrivateLink](#) 가이드에서.AWS PrivateLink

주제

- [크기 조정 계획용 인터페이스 VPC 엔드포인트 생성](#)
- [크기 조정 계획용 VPC 엔드포인트 정책 생성](#)
- [엔드포인트 마이그레이션](#)

크기 조정 계획용 인터페이스 VPC 엔드포인트 생성

다음 서비스 이름을 사용하여 AWS Auto Scaling 스케일링 플랜용 엔드포인트를 생성하십시오.

```
com.amazonaws.region.autoscaling-plans
```

자세한 내용은 가이드의 [인터페이스 VPC 엔드포인트를 사용한 AWS 서비스 액세스](#)를 참조하십시오. AWS PrivateLink

다른 설정은 변경할 필요가 없습니다. AWS Auto Scaling 서비스 엔드포인트 또는 프라이빗 인터페이스 VPC 엔드포인트 중 사용 중인 엔드포인트를 AWS 서비스 사용하여 다른 API를 호출합니다.

크기 조정 계획용 VPC 엔드포인트 정책 생성

정책을 VPC 엔드포인트에 연결하여 API에 대한 액세스를 제어할 수 있습니다. AWS Auto Scaling 이 정책은 다음을 지정합니다.

- 작업을 수행할 수 있는 보안 주체.
- 수행할 수 있는 작업.
- 작업을 수행할 수 있는 리소스.

다음 예에서는 엔드포인트를 통해 크기 조정 계획을 삭제할 수 있는 모든 사용자 권한을 거부하는 VPC 엔드포인트 정책을 보여 줍니다. 또한 이 정책 예에서는 모든 사용자에게 다른 모든 작업을 수행할 수 있는 권한을 부여합니다.

```
{
  "Statement": [
    {
      "Action": "*",
      "Effect": "Allow",
      "Resource": "*",
```

```

        "Principal": "*"
    },
    {
        "Action": "autoscaling-plans:DeleteScalingPlan",
        "Effect": "Deny",
        "Resource": "*",
        "Principal": "*"
    }
]
}

```

자세한 정보는 AWS PrivateLink 가이드의 [VPC 엔드포인트 정책](#)을 참조하세요.

엔드포인트 마이그레이션

2019년 11월 22일에 API 호출을 위한 새로운 기본 DNS 호스트 이름 및 엔드포인트가 도입되었습니다. `autoscaling-plans.region.amazonaws.com`. AWS Auto Scaling 새 엔드포인트는 최신 릴리스 AWS CLI 및 SDK와 호환됩니다. 아직 설치하지 않았다면 최신 AWS CLI 및 SDK를 설치하여 새 엔드포인트를 사용하십시오. 를 업데이트하려면 사용 AWS Command Line Interface 설명서의 [AWS CLI pip를 사용하여 설치하기](#)를 참조하십시오. AWS CLI AWS SDK에 대한 자세한 내용은 [Amazon Web Services용 도구](#)를 참조하십시오.

Important

이전 버전과의 호환성을 위해 기존 `autoscaling.region.amazonaws.com` 엔드포인트는 API 호출에 계속 지원됩니다. AWS Auto Scaling `autoscaling.region.amazonaws.com` 엔드포인트를 프라이빗 인터페이스 VPC 엔드포인트로 설정하려면 Amazon EC2 Auto Scaling 사용 설명서의 [Amazon EC2 Auto Scaling 및 인터페이스 VPC 엔드포인트](#)를 참조하세요.

CLI 또는 API 사용 시 호출할 엔드포인트 AWS Auto Scaling

현재 AWS Auto Scaling 릴리스에서는 AWS Auto Scaling API에 대한 호출이 대신 `autoscaling-plans.region.amazonaws.com` `autoscaling.region.amazonaws.com` 엔드포인트로 자동으로 전달됩니다.

각 명령에 엔드포인트를 지정할 때 `--endpoint-url https://autoscaling-plans.region.amazonaws.com` 파라미터를 사용하여 CLI에서 새 엔드포인트를 호출할 수 있습니다.

권장되지는 않지만, 각 명령에 엔드포인트를 지정할 때 `--endpoint-url https://autoscaling.region.amazonaws.com` 파라미터를 사용하여 CLI에서 이전 엔드포인트를 호출할 수도 있습니다.

API를 호출하는 데 사용되는 다양한 SDK에 대해서는 해당 SDK 설명서를 참조하여 요청을 특정 엔드포인트로 보내는 방법을 알아보세요. 자세한 내용은 [Amazon Web Services용 도구](#)를 참조하세요.

데이터 보호

AWS [공동 책임 모델](#)의 데이터 보호에 적용됩니다 AWS Auto Scaling. 이 모델에 설명된 대로 AWS는 모든 데이터를 실행하는 글로벌 인프라를 보호하는 역할을 AWS 클라우드합니다. 이 인프라에서 호스팅되는 콘텐츠에 대한 제어를 유지하는 것은 사용자의 책임입니다. 사용하는 AWS 서비스의 보안 구성과 관리 작업에 대한 책임도 사용자에게 있습니다. 데이터 프라이버시에 대한 자세한 내용은 [데이터 프라이버시 FAQ](#)를 참조하세요. 유럽의 데이터 보호에 대한 자세한 내용은 AWS 보안 블로그의 [AWS 공동 책임 모델 및 GDPR](#) 블로그 게시물을 참조하세요.

데이터 보호를 위해 AWS 계정 자격 증명을 보호하고 AWS IAM Identity Center OR AWS Identity and Access Management (IAM)을 사용하여 개별 사용자를 설정하는 것이 좋습니다. 이 방식을 사용하면 각 사용자에게 자신의 직무를 충실히 이행하는 데 필요한 권한만 부여됩니다. 또한 다음과 같은 방법으로 데이터를 보호하는 것이 좋습니다.

- 각 계정에 다중 인증(MFA)을 사용합니다.
- SSL/TLS를 사용하여 리소스와 통신하세요. AWS TLS 1.2는 필수이며 TLS 1.3을 권장합니다.
- 를 사용하여 API 및 사용자 활동 로깅을 설정합니다. AWS CloudTrail
- 포함된 모든 기본 보안 제어와 함께 AWS 암호화 솔루션을 사용하십시오 AWS 서비스.
- Amazon Macie와 같은 고급 관리형 보안 서비스를 사용하여 Amazon S3에 저장된 민감한 데이터를 검색하고 보호합니다.
- 명령줄 인터페이스 또는 API를 AWS 통해 액세스할 때 FIPS 140-2로 검증된 암호화 모듈이 필요한 경우 FIPS 엔드포인트를 사용하십시오. 사용 가능한 FIPS 엔드포인트에 대한 자세한 내용은 [Federal Information Processing Standard\(FIPS\) 140-2](#)를 참조하십시오.

고객의 이메일 주소와 같은 기밀 정보나 중요한 정보는 태그나 이름 필드와 같은 자유 양식 필드에 입력하지 않는 것이 좋습니다. 여기에는 콘솔, API AWS Auto Scaling 또는 AWS 서비스 SDK를 사용하거나 다른 방법으로 작업하는 경우가 포함됩니다. AWS CLI AWS 이름에 사용되는 태그 또는 자유 형식 텍스트 필드에 입력하는 모든 데이터는 청구 또는 진단 로그에 사용될 수 있습니다. 외부 서버에 URL을 제공할 때 해당 서버에 대한 요청을 검증하기 위해 보안 인증 정보를 URL에 포함시켜서는 안 됩니다.

크기 조정 계획에 사용되는 Identity and Access Management

AWS Identity and Access Management (IAM)은 관리자가 리소스에 대한 액세스를 안전하게 제어할 수 있도록 AWS 서비스 있도록 도와줍니다. IAM 관리자는 리소스를 사용할 수 있는 인증 (로그인) 및 권한 부여 (권한 보유)를 받을 수 있는 사용자를 제어합니다. AWS Auto Scaling IAM은 추가 AWS 서비스 비용 없이 사용할 수 있습니다.

IAM 설명서 전체 내용은 [IAM 사용 설명서](#)를 참조하세요.

액세스 제어

요청을 인증하는 데 유효한 자격 증명이 있더라도 권한이 없다면 크기 조정 계획을 생성하거나 액세스할 수 없습니다. 예를 들어 크기 조정 계획을 생성하고 예측 크기 조정을 구성하는 등의 권한이 있어야 합니다.

다음 단원에서는 IAM 관리자가 IAM을 사용해 크기 조정 계획 작업이 가능한 사람을 제어하여 크기 조정 계획을 보호하는 방법에 대한 세부 정보를 제공합니다.

주제

- [크기 조정 계획에 IAM을 사용하는 방법](#)
- [예측 크기 조정 서비스 연결 역할](#)
- [크기 조정 계획에 대한 자격 증명 기반 정책 예제](#)

크기 조정 계획에 IAM을 사용하는 방법

IAM을 사용하여 AWS Auto Scaling 규모 조정 계획을 만들고, 액세스하고, 관리할 수 있는 사람을 관리하기 전에 먼저 조정 계획에서 사용할 수 있는 IAM 기능이 무엇인지 이해해야 합니다.

주제

- [보안 인증 기반 정책](#)
- [리소스 기반 정책](#)
- [액세스 제어 목록\(ACL\)](#)
- [태그 기반 인증](#)
- [IAM 역할](#)

보안 인증 기반 정책

IAM 자격 증명 기반 정책을 사용하면 허용되거나 거부되는 작업과 리소스 및 작업이 허용되거나 거부되는 조건을 지정할 수 있습니다. 크기 조정 계획은 특정 작업, 리소스 및 조건 키를 지원합니다. JSON 정책에서 사용하는 모든 요소에 대해 알아보려면 IAM 사용 설명서의 [IAM JSON 정책 요소 참조](#)를 참조하십시오.

작업

관리자는 AWS JSON 정책을 사용하여 누가 무엇에 액세스할 수 있는지 지정할 수 있습니다. 즉, 어떤 보안 주체가 어떤 리소스와 어떤 조건에서 작업을 수행할 수 있는지를 지정할 수 있습니다.

JSON 정책의 Action 요소는 정책에서 액세스를 허용하거나 거부하는 데 사용할 수 있는 작업을 설명합니다. 정책 작업은 일반적으로 관련 AWS API 작업과 이름이 같습니다. 일치하는 API 작업이 없는 권한 전용 작업 같은 몇 가지 예외도 있습니다. 정책에서 여러 작업이 필요한 몇 가지 작업도 있습니다. 이러한 추가 작업을 종속 작업이라고 합니다.

연결된 작업을 수행할 수 있는 권한을 부여하기 위한 정책에 작업을 포함시킵니다.

IAM 정책 설명의 크기 조정 계획 작업은 작업 앞에 `autoscaling-plans:` 접두사를 사용합니다. 정책 문에는 Action 또는 NotAction 요소가 포함되어야 합니다. 크기 조정 계획에는 이 서비스로 수행할 수 있는 태스크를 설명하는 고유한 작업 집합이 있습니다.

단일 명령문에서 여러 작업을 지정하려면 다음 예제와 같이 쉼표로 구분합니다.

```
"Action": [
    "autoscaling-plans:DescribeScalingPlans",
    "autoscaling-plans:DescribeScalingPlanResources"
```

와일드카드(*)를 사용하여 여러 작업을 지정할 수 있습니다. 예를 들어, Describe라는 단어로 시작하는 모든 작업을 지정하려면 다음 작업을 포함합니다.

```
"Action": "autoscaling-plans:Describe*"
```

정책 설명에서 사용할 수 있는 크기 조정 계획 작업의 전체 목록을 보려면 서비스 승인 참조의 [AWS Auto Scaling에 사용되는 작업, 리소스 및 조건 키](#)를 참조하세요.

리소스

Resource 요소는 작업이 적용되는 객체를 지정합니다.

크기 조정 계획에는 IAM 정책 설명의 Resource 요소로 사용할 수 있는 서비스에 정의된 리소스가 없습니다. 따라서 IAM 정책에서 사용할 수 있는 Amazon 리소스 이름(ARN)이 없습니다. 크기 조정 계획 작업에 대한 액세스를 제어하려면 IAM 정책을 작성할 때 반드시 *(별표)를 리소스로 사용해야 합니다.

조건 키

Condition 요소(또는 Condition 블록)를 사용하면 정책이 발효되는 조건을 지정할 수 있습니다. 예를 들어, 특정 날짜 이후에만 정책을 적용할 수 있습니다. 조건을 표현하려면 미리 정의된 조건 키를 사용합니다.

크기 조정 계획은 서비스별 조건 키를 제공하지 않지만, 일부 전역 조건 키 사용은 지원합니다. 모든 AWS 글로벌 조건 키를 보려면 IAM 사용 설명서의 [AWS 글로벌 조건 컨텍스트 키](#)를 참조하십시오.

Condition 요소는 옵션입니다.

예

크기 조정 계획의 자격 증명 기반 정책 예를 보려면 [크기 조정 계획에 대한 자격 증명 기반 정책 예제](#) 단원을 참조하세요.

리소스 기반 정책

Amazon Simple Storage Service 등의 다른 Amazon Web Services도 리소스 기반 권한 정책을 지원합니다. 예를 들어, 권한 정책을 S3 버킷에 연결하여 해당 버킷에 대한 액세스 권한을 관리할 수 있습니다.

크기 조정 계획은 리소스 기반 정책을 지원하지 않습니다.

액세스 제어 목록(ACL)

크기 조정 계획은 액세스 제어 목록(ACL)을 지원하지 않습니다.

태그 기반 인증

크기 조정 계획은 태그를 지정할 수 없습니다. 또한 태그를 지정할 수 있는 서비스 정의 리소스가 없습니다. 따라서 리소스의 태그를 기반으로 하는 액세스 제어를 지원하지 않습니다.

크기 조정 계획에는 태그 기반의 액세스 제어를 지원하는 Auto Scaling 그룹과 같은 태그 지정 가능한 리소스가 포함될 수 있습니다. 자세한 내용은 AWS 서비스에 대한 IAM 문서를 참조하세요.

IAM 역할

[IAM 역할](#)은 특정 권한을 가지고 있는 AWS 계정 내 엔터티입니다.

임시 자격 증명 사용

임시 자격 증명을 사용하여 페더레이션을 통해 로그인하거나, IAM 역할을 수입하거나, 교차 계정 역할을 수입할 수 있습니다. AWS STS API 작업(예: [AssumeRole](#) 또는 [GetFederationToken](#))을 호출하여 임시 보안 인증 정보를 가져옵니다.

크기 조정 계획은 임시 자격 증명 사용을 지원합니다.

크기 조정 계획을 위한 서비스 연결 역할

AWS Auto Scaling 사용자를 대신하여 다른 서비스를 호출하는 데 필요한 권한으로 AWS 서비스 연결 역할을 사용합니다. 서비스 연결 역할을 사용하면 필요한 권한을 수동으로 추가할 필요가 없으므로 크기 조정 계획을 더 쉽게 설정할 수 있습니다. 자세한 내용은 IAM 사용 설명서의 [서비스 연결 역할 사용](#)을 참조하십시오.

AWS Auto Scaling 스케일링 플랜을 사용할 때 몇 가지 유형의 서비스 연결 역할을 사용하여 사용자를 AWS 서비스 대신하여 다른 역할을 호출합니다.

- 예측 규모 조정 서비스 연결 역할 — 에서 과거 지표 AWS Auto Scaling 데이터에 액세스할 수 있습니다. CloudWatch 또한 로드 예측 및 용량 예측을 기반으로 Auto Scaling 그룹에 대해 예약된 작업을 생성할 수 있습니다. 자세한 설명은 [예측 크기 조정 서비스 연결 역할](#) 섹션을 참조하세요.
- Amazon EC2 Auto Scaling 서비스 연결 역할 — AWS Auto Scaling Auto Scaling 그룹의 대상 추적 조정 정책에 액세스하고 관리할 수 있습니다. 자세한 내용은 Amazon EC2 Auto Scaling 사용 설명서의 [Amazon EC2 Auto Scaling 서비스 연결 역할](#)을 참조하세요.
- Application Auto Scaling 서비스 연결 역할 - 다른 확장 가능한 리소스에 대한 대상 추적 조정 정책에 액세스하고 관리할 수 AWS Auto Scaling 있습니다. 각 서비스에는 서비스 연결 역할이 하나씩 있습니다. 자세한 정보는 Application Auto Scaling 사용 설명서의 [Application Auto Scaling 서비스 연결 역할](#)을 참조하세요.

다음 절차를 사용하여 계정에 이미 서비스 연결 역할이 있는지 확인할 수 있습니다.

서비스 연결 역할이 이미 있는지 확인하려면

1. <https://console.aws.amazon.com/iam/>에서 IAM 콘솔을 엽니다.
2. 탐색 창에서 역할을 선택합니다.
3. 목록에서 AWSServiceRole을 검색하여 계정에 있는 서비스 연결 역할을 찾습니다. 확인하려는 서비스 연결 역할의 이름을 찾습니다.

서비스 역할

AWS Auto Scaling 규모 조정 계획에 대한 서비스 역할이 없습니다.

예측 크기 조정 서비스 연결 역할

AWS Auto Scaling 스케일링 플랜을 사용할 때 사용자 대신 다른 사람에게 호출해야 하는 AWS 권한에 대해 서비스 연결 역할을 사용합니다. 자세한 설명은 [크기 조정 계획을 위한 서비스 연결 역할](#) 섹션을 참조하세요.

다음 단원에서는 예측 크기 조정에 대한 서비스 연결 역할을 생성하고 관리하는 방법을 설명합니다. 먼저 IAM 엔터티(사용자, 그룹, 역할 등)가 서비스 연결 역할을 생성, 편집 또는 삭제할 수 있도록 권한을 구성해야 합니다.

서비스 연결 역할에 의해 부여된 권한

AWS Auto Scaling 예측 규모 조정을 활성화하면 이름이 `AWSServiceRoleForAutoScalingPlans_ECAutoScaling2`인 서비스 연결 역할을 사용하여 사용자를 대신하여 다음 작업을 호출합니다.

- `cloudwatch:GetMetricData`
- `autoscaling:DescribeAutoScalingGroups`
- `autoscaling:DescribeScheduledActions`
- `autoscaling:BatchPutScheduledUpdateGroupAction`
- `autoscaling:BatchDeleteScheduledAction`

`AWSServiceRoleForAutoScalingPlans_EC2`는 `autoscaling-plans.amazonaws.com` 서비스가 역할을 맡을 것으로 AutoScaling 신뢰합니다.

서비스 연결 역할 생성(자동)

AutoScaling 역할 `AWSServiceRoleForAutoScalingPlans_EC2`개를 수동으로 만들 필요는 없습니다. AWS 계정에서 규모 조정 계획을 생성하고 예측 규모 조정을 활성화할 때 이 역할을 자동으로 생성합니다.

사용자를 대신하여 서비스 연결 역할을 AWS 생성하려면 필요한 권한이 있어야 합니다. 자세한 내용은 IAM 사용 설명서의 [서비스 연결 역할 권한](#)을 참조하세요.

서비스 연결 역할 생성(수동)

IAM 콘솔, IAM CLI 또는 IAM API를 사용하여 수동으로 서비스 연결 역할을 생성할 수 있습니다. 자세한 내용은 IAM 사용 설명서의 [서비스 연결 역할 생성](#) 섹션을 참조하세요.

서비스 연결 역할을 생성하려면(AWS CLI)

다음 [create-service-linked-role](#) CLI 명령을 사용하여 서비스 연결 역할을 생성합니다.

```
aws iam create-service-linked-role --aws-service-name autoscaling-plans.amazonaws.com
```

서비스 연결 역할 편집

IAM을 사용하여 AWSServiceRoleForAutoScalingPlans_EC2의 AutoScaling 설명을 편집할 수 있습니다. 자세한 내용은 IAM 사용 설명서의 [서비스 연결 역할 편집](#)을 참조하세요.

서비스 연결 역할 삭제

스케일링 플랜을 더 이상 사용할 필요가 없는 경우 AWSServiceRoleForAutoScalingPlans_EC2를 AutoScaling 삭제하는 것이 좋습니다.

AWS 계정에서 예측 크기 조정을 활성화한 모든 크기 조정 계획을 삭제한 후에만 서비스 연결 역할을 삭제할 수 있습니다. 이렇게 하면 크기 조정 계획에 대한 권한을 실수로 제거하는 것을 방지할 수 있습니다.

IAM 콘솔, IAM CLI 또는 IAM API를 사용하여 서비스 연결 역할을 삭제할 수 있습니다. 자세한 내용은 [IAM 사용 설명서](#)의 서비스 연결 역할 삭제 단원을 참조하세요.

AWSServiceRoleForAutoScalingPlans_EC2개의 AutoScaling 서비스 연결 역할을 삭제한 후 예측 규모 조정이 활성화된 조정 계획을 생성하면 역할이 다시 AWS Auto Scaling 생성됩니다.

지원되는 리전

AWS Auto Scaling 스케일링 플랜이 제공되는 모든 지역에서 서비스 연결 역할을 사용할 수 있습니다. AWS 리전 크기 조정 계획의 리전별 가용성에 대한 자세한 내용은 AWS 일반 참조의 [AWS Auto Scaling 엔드포인트 및 할당량](#)을 참조하세요.

크기 조정 계획에 대한 자격 증명 기반 정책 예제

기본적으로 신규 IAM 사용자는 어떤 작업 권한도 없습니다. IAM 관리자는 크기 조정 계획으로 작업할 수 있는 IAM 자격 증명(예: 사용자 또는 역할) 권한을 부여하는 IAM 정책을 생성하고 할당해야 합니다.

이러한 예제 JSON 정책 문서를 사용하여 IAM 정책을 생성하는 방법을 알아보려면 IAM 사용 설명서의 [JSON 탭에서 정책 생성](#)을 참조하세요.

주제

- [정책 모범 사례](#)
- [사용자가 크기 조정 계획을 생성하도록 허용](#)
- [사용자가 예측 크기 조정을 활성화할 수 있도록 허용](#)
- [필요한 추가 권한](#)
- [서비스 연결 역할 생성에 필요한 권한](#)

정책 모범 사례

ID 기반 정책은 누군가가 계정에서 AWS Auto Scaling 리소스를 생성, 액세스 또는 삭제할 수 있는지 여부를 결정합니다. 이 작업으로 인해 AWS 계정에 비용이 발생할 수 있습니다. 자격 증명 기반 정책을 생성하거나 편집할 때는 다음 지침과 권장 사항을 따릅니다.

- AWS 관리형 정책으로 시작하여 최소 권한 권한으로 이동 — 사용자와 워크로드에 권한을 부여하려면 여러 일반적인 사용 사례에 권한을 부여하는 AWS 관리형 정책을 사용하세요. 해당 내용은 에서 사용할 수 있습니다. AWS 계정사용 사례에 맞는 AWS 고객 관리형 정책을 정의하여 권한을 더 줄이는 것이 좋습니다. 자세한 정보는 IAM 사용 설명서의 [AWS 관리형 정책](#) 또는 [직무에 관한AWS 관리형 정책](#)을 참조하세요.
- 최소 권한 적용 – IAM 정책을 사용하여 권한을 설정하는 경우 작업을 수행하는 데 필요한 권한만 부여합니다. 이렇게 하려면 최소 권한으로 알려진 특정 조건에서 특정 리소스에 대해 수행할 수 있는 작업을 정의합니다. IAM을 사용하여 권한을 적용하는 방법에 대한 자세한 정보는 IAM 사용자 설명서에 있는 [Policies and permissions in IAM](#)(IAM의 정책 및 권한)을 참조하세요.
- Use conditions in IAM policies to further restrict access(IAM 정책의 조건을 사용하여 액세스 추가 제한) – 정책에 조건을 추가하여 작업 및 리소스에 대한 액세스를 제한할 수 있습니다. 예를 들어 SSL 을 사용하여 모든 요청을 전송해야 한다고 지정하는 정책 조건을 생성할 수 있습니다. 예를 AWS 서비스들어 특정 작업을 통해 서비스 작업을 사용하는 경우 조건을 사용하여 서비스 작업에 대한 액세스 권한을 부여할 수도 AWS CloudFormation있습니다. 자세한 정보는 IAM 사용자 설명서의 [IAM JSON 정책 요소: 조건](#)을 참조하세요.
- IAM Access Analyzer를 통해 IAM 정책을 검증하여 안전하고 기능적인 권한 보장 – IAM Access Analyzer에서는 IAM 정책 언어(JSON)와 모범 사례가 정책에서 준수되도록 신규 및 기존 정책을 검증합니다. IAM Access Analyzer는 100개 이상의 정책 확인 항목과 실행 가능한 권장 사항을 제공하여 안전하고 기능적인 정책을 생성하도록 돕습니다. 자세한 정보는 IAM 사용자 설명서의 [IAM Access Analyzer policy validation](#)(IAM Access Analyzer 정책 검증)을 참조하세요.

- 멀티 팩터 인증 (MFA) 필요 - IAM 사용자 또는 루트 사용자가 필요한 시나리오가 있는 경우 추가 보안을 위해 AWS 계정 MFA를 활성화하십시오. API 작업을 직접적으로 호출할 때 MFA가 필요하다면 정책에 MFA 조건을 추가합니다. 자세한 정보는 IAM 사용자 설명서의 [Configuring MFA-protected API access](#)(MFA 보호 API 액세스 구성)를 참조하세요.

IAM의 모범 사례에 대한 자세한 내용은 IAM 사용자 설명서의 [IAM의 보안 모범 사례](#)를 참조하세요.

사용자가 크기 조정 계획을 생성하도록 허용

다음은 크기 조정 계획을 생성할 수 있는 권한을 부여하는 자격 증명 기반 정책의 예입니다.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "autoscaling-plans:*",
        "cloudwatch:PutMetricAlarm",
        "cloudwatch>DeleteAlarms",
        "cloudwatch:DescribeAlarms",
        "cloudformation:ListStackResources"
      ],
      "Resource": "*"
    }
  ]
}
```

크기 조정 계획에 대한 작업을 수행하려면 최종 사용자에게 본인 계정의 특정 리소스로 작업할 수 있는 추가 권한이 있어야 합니다. 이러한 권한은 [필요한 추가 권한](#)에 나와 있습니다.

또한 각 콘솔 사용자에게는 계정에서 확장 가능한 리소스를 검색하고 콘솔에서 CloudWatch 메트릭 데이터의 그래프를 볼 수 있는 권한이 필요합니다. AWS Auto Scaling 콘솔을 사용하는 데 필요한 추가 권한 집합은 다음과 같습니다.

- `cloudformation:ListStacks`: 스택 나열
- `tag:GetTagKeys`: 특정 태그 키가 포함된 크기 조정 가능한 리소스 찾기
- `tag:GetTagValues`: 특정 태그 값이 포함된 리소스 찾기
- `autoscaling:DescribeTags`: 특정 태그가 포함된 Auto Scaling 그룹 찾기

- `cloudwatch:GetMetricData`: 지표 그래프의 데이터 보기

사용자가 예측 크기 조정을 활성화할 수 있도록 허용

다음은 예측 크기 조정을 활성화할 수 있는 권한을 부여하는 자격 증명 기반 정책의 예입니다. 이러한 권한은 Auto Scaling 그룹을 크기 조정하도록 설정된 크기 조정 계획의 기능을 확장합니다.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "cloudwatch:GetMetricData",
        "autoscaling:DescribeAutoScalingGroups",
        "autoscaling:DescribeScheduledActions",
        "autoscaling:BatchPutScheduledUpdateGroupAction",
        "autoscaling:BatchDeleteScheduledAction"
      ],
      "Resource": "*"
    }
  ]
}
```

필요한 추가 권한

크기 조정 계획을 성공적으로 구성하려면 최종 사용자에게 크기 조정을 구성할 각 대상 서비스에 대한 권한이 부여되어야 합니다. 대상 서비스를 사용하는 데 필요한 최소 권한을 부여하려면 이 섹션의 정보를 읽고 IAM 정책 설명의 Action 요소에 관련 작업을 지정하세요.

Auto Scaling 그룹

크기 조정 계획에 Auto Scaling 그룹을 추가하려면 사용자는 Amazon EC2 Auto Scaling의 다음 권한이 있어야 합니다.

- `autoscaling:UpdateAutoScalingGroup`
- `autoscaling:DescribeAutoScalingGroups`
- `autoscaling:PutScalingPolicy`
- `autoscaling:DescribePolicies`
- `autoscaling>DeletePolicy`

ECS 서비스

크기 조정 계획에 ECS 서비스를 추가하려면 사용자는 Amazon ECS 및 Application Auto Scaling의 다음 권한이 있어야 합니다.

- `ecs:DescribeServices`
- `ecs:UpdateService`
- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling>DeleteScalingPolicy`

스팟 플릿

크기 조정 계획에 스팟 플릿을 추가하려면 사용자에게 Amazon EC2 및 Application Auto Scaling의 다음 권한이 있어야 합니다.

- `ec2:DescribeSpotFleetRequests`
- `ec2:ModifySpotFleetRequest`
- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling>DeleteScalingPolicy`

DynamoDB 테이블 또는 글로벌 인덱스

크기 조정 계획에 DynamoDB 테이블 또는 글로벌 인덱스를 추가하려면 사용자는 DynamoDB 및 Application Auto Scaling의 다음 권한이 있어야 합니다.

- `dynamodb:DescribeTable`
- `dynamodb:UpdateTable`

- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling>DeleteScalingPolicy`

Aurora DB 클러스터

크기 조정 계획에 Aurora DB 클러스터를 추가하려면 사용자는 Amazon Aurora 및 Application Auto Scaling의 다음 권한이 있어야 합니다.

- `rds:AddTagsToResource`
- `rds>CreateDBInstance`
- `rds>DeleteDBInstance`
- `rds:DescribeDBClusters`
- `rds:DescribeDBInstances`
- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling>DeleteScalingPolicy`

서비스 연결 역할 생성에 필요한 권한

AWS Auto Scaling 내 사용자 중 누구라도 예측 확장이 활성화된 규모 조정 계획을 처음 생성할 때 서비스 연결 역할을 AWS 계정 생성할 수 있는 권한이 필요합니다. 서비스 연결 역할이 아직 없는 경우 계정에서 AWS Auto Scaling 생성하십시오. 서비스 연결 역할은 사용자를 대신하여 AWS Auto Scaling 다른 서비스를 호출할 수 있는 권한을 부여합니다.

역할 자동 생성이 성공하려면 사용자가 `iam:CreateServiceLinkedRole` 작업에 대한 권한을 보유해야 합니다.

```
"Action": "iam:CreateServiceLinkedRole"
```

다음은 서비스 연결 역할을 생성하는 권한을 부여하는 자격 증명 기반 정책의 예입니다.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "iam:CreateServiceLinkedRole",
      "Resource": "arn:aws:iam::*:role/aws-service-role/autoscaling-plans.amazonaws.com/AWSServiceRoleForAutoScalingPlans_EC2AutoScaling",
      "Condition": {
        "StringLike": {
          "iam:AWSServiceName": "autoscaling-plans.amazonaws.com"
        }
      }
    }
  ]
}
```

자세한 설명은 [예측 크기 조정 서비스 연결 역할](#) 섹션을 참조하세요.

규정 준수 확인

특정 규정 준수 프로그램의 범위 내에 AWS 서비스 있는지 알아보려면 AWS 서비스 규정 준수 [프로그램의 AWS 서비스 범위별, 규정](#) 참조하여 관심 있는 규정 준수 프로그램을 선택하십시오. 일반 정보는 [AWS 규정 준수 프로그램 AWS 보증 프로그램 규정 AWS](#) 참조하십시오.

를 사용하여 AWS Artifact 타사 감사 보고서를 다운로드할 수 있습니다. 자세한 내용은 의 보고서 <https://docs.aws.amazon.com/artifact/latest/ug/downloading-documents.html> 참조하십시오 AWS Artifact.

사용 시 규정 준수 AWS 서비스 책임은 데이터의 민감도, 회사의 규정 준수 목표, 관련 법률 및 규정에 따라 결정됩니다. AWS 규정 준수에 도움이 되는 다음 리소스를 제공합니다.

- [보안 및 규정 준수 킷스타트 가이드](#) - 이 배포 가이드에서는 아키텍처 고려 사항을 설명하고 보안 및 규정 준수에 AWS 중점을 둔 기본 환경을 배포하기 위한 단계를 제공합니다.
- [Amazon Web Services의 HIPAA 보안 및 규정 준수를 위한 설계 — 이 백서에서는 기업이 HIPAA 적격 애플리케이션을 만드는 AWS 데 사용할 수 있는 방법을 설명합니다.](#)

Note

모든 AWS 서비스 사람이 HIPAA 자격을 갖춘 것은 아닙니다. 자세한 내용은 [HIPAA 적격 서비스 참조](#)를 참조하십시오.

- [AWS 규정 준수 리소스 AWS](#) — 이 워크북 및 가이드 모음은 해당 산업 및 지역에 적용될 수 있습니다.
- [AWS 고객 규정 준수 가이드](#) — 규정 준수의 관점에서 공동 책임 모델을 이해하십시오. 이 가이드에서는 보안을 유지하기 위한 모범 사례를 AWS 서비스 요약하고 여러 프레임워크 (미국 표준 기술 연구소 (NIST), 결제 카드 산업 보안 표준 위원회 (PCI), 국제 표준화기구 (ISO) 등) 에서 보안 제어에 대한 지침을 매핑합니다.
- AWS Config 개발자 안내서의 [규칙을 사용하여 리소스 평가](#) — 이 AWS Config 서비스는 리소스 구성이 내부 관행, 업계 지침 및 규정을 얼마나 잘 준수하는지 평가합니다.
- [AWS Security Hub](#) — 이를 AWS 서비스 통해 내부 AWS 보안 상태를 포괄적으로 파악할 수 있습니다. Security Hub는 보안 통제를 사용하여 AWS 리소스를 평가하고 보안 업계 표준 및 모범 사례에 대한 규정 준수를 확인합니다. 지원되는 서비스 및 제어 목록은 [Security Hub 제어 참조](#)를 참조하십시오.
- [AWS Audit Manager](#) — 이를 AWS 서비스 통해 AWS 사용량을 지속적으로 감사하여 위험을 관리하고 규정 및 업계 표준을 준수하는 방법을 단순화할 수 있습니다.

인프라 보안

관리형 서비스로서 AWS 글로벌 네트워크 보안으로 AWS Auto Scaling 보호됩니다. AWS 보안 서비스 및 인프라 AWS 보호 방법에 대한 자세한 내용은 [AWS 클라우드 보안을](#) 참조하십시오. 인프라 보안 모범 사례를 사용하여 AWS 환경을 설계하려면 Security Pillar AWS Well-Architected Framework의 [인프라 보호](#)를 참조하십시오.

AWS 게시된 API 호출을 사용하여 네트워크를 통해 액세스할 AWS Auto Scaling 수 있습니다. 고객은 다음을 지원해야 합니다.

- 전송 계층 보안(TLS). TLS 1.2는 필수이며 TLS 1.3을 권장합니다.
- DHE(Ephemeral Diffie-Hellman) 또는 ECDHE(Elliptic Curve Ephemeral Diffie-Hellman)와 같은 완전 전송 보안(PFS)이 포함된 암호 제품군. Java 7 이상의 최신 시스템은 대부분 이러한 모드를 지원합니다.

또한 요청은 액세스 키 ID 및 IAM 주체와 관련된 비밀 액세스 키를 사용하여 서명해야 합니다. 또는 [AWS Security Token Service](#)(AWS STS)를 사용하여 임시 보안 인증 정보를 생성하여 요청에 서명할 수 있습니다.

크기 조정 계획 할당량

AWS 계정에는 크기 조정 계획과 관련된 다음과 같은 할당량(이전에는 제한이라고 함)이 있습니다.

증가를 요청하려면 [Auto Scaling 제한 양식](#)을 사용하세요. 증가 요청에 리소스 유형(예: Amazon EC2 Auto Scaling, Amazon ECS 또는 DynamoDB)을 지정해야 합니다.

리전별 계정당 기본 할당량

항목	기본값
리소스 유형당 최대 확장 가능 리소스 수	할당량은 리소스 유형에 따라 다릅니다. Amazon DynamoDB: 3000 Amazon EC2 Auto Scaling 그 룹: 200 다른 모든 리소스 유형: 500
최대 크기 조정 계획 수	100
크기 조정 계획당 최대 크기 조정 지침 수	500
크기 조정 지침당 최대 대상 추적 구성 수	10

워크로드를 조정할 때 서비스 할당량을 염두에 두세요. 예를 들어, 서비스에서 허용되는 최대 용량 단위 수에 도달하면 확장이 중지됩니다. 수요가 감소하고 현재 용량이 감소하면 AWS Auto Scaling이 다시 확장할 수 있습니다. 이 서비스 할당량 제한에 다시 도달하지 않도록 증가를 요청할 수 있습니다. 각 서비스에는 리소스의 최대 용량에 대한 자체 기본 할당량이 있습니다. 다른 Amazon Web Services의 기본 할당량에 대한 자세한 내용은 Amazon Web Services 일반 참조의 [서비스 엔드포인트 및 할당량](#)을 참조하세요.

관련 정보

이 서비스로 작업할 때 다음과 같은 추가 리소스가 도움이 될 수 있습니다.

리소스	설명
AWS 일반 참조의 <u>AWS Auto Scaling 엔드포인트 및 할당량</u>	AWS Auto Scaling 리전 및 엔드포인트에 대한 정보입니다.
AWS Auto Scaling API 참조	각 API 작업에 대한 설명서에는 요청 구문, 요청 파라미터, 응답 요소가 표시되며 언어별 SDK 참조 항목에 대한 링크가 제공됩니다.
AWS CLI 명령 참조에서의 autoscaling-plans	크기 조정 계획에 사용할 수 있는 AWS CLI 명령에 대한 설명입니다.
AWS CloudFormation 사용 설명서	AWS CloudFormation 사용 설명서 리소스 AWS::AutoScalingPlans::ScalingPlan 을 참조하면 수동 작업 없이 크기 조정 계획을 구축하고 모델링하며 관리할 수 있습니다.
제품 페이지	AWS Auto Scaling에 대한 정보를 제공하는 기본 웹 페이지입니다.
AWS re:Post	클라우드 소식을 거쳐 전문가가 검토한 기술적 질문에 대한 답변을 제공하는 AWS 관리형 Q&A(질의응답) 서비스입니다.
AWS CloudTrail 사용 설명서	AWS CloudTrail 사용을 통해 AWS Auto Scaling API에 호출을 로그하고 로그를 Amazon S3에 저장하는 방법을 알아봅니다. 이러한 CloudTrail 로그를 사용하여 어떤 요청이 이루어졌는지, 어떤 소스 IP 주소에서 요청을 했는지, 누가 언제 요청했는지 등을 확인할 수 있습니다.
Amazon EC2 Auto Scaling 사용 설명서의 Amazon EC2 Auto Scaling을 위한 대상 추적 크기 조정 정책	대상 추적 크기 조정 정책을 사용하여 Auto Scaling 그룹의 크기를 조정하는 방법을 알아보세요.

리소스	설명
Application Auto Scaling 사용 설명서의 Application Auto Scaling을 위한 대상 추적 크기 조정 정책	DynamoDB 인덱스, 테이블, Amazon ECS 서비스 등 대상 추적 크기 조정 정책을 사용하는 기타 조정 가능한 리소스의 크기를 조정하는 방법을 알아보세요.
Amazon EC2 Auto Scaling 사용 설명서의 Amazon EC2 Auto Scaling을 위한 예측 크기 조정	이는 2021년 5월에 출시된 예측 크기 조정의 최신 버전입니다. 이 버전에 도입된 일부 기능은 크기 조정 계획에서 사용할 수 없으며 이러한 기능에 액세스하려면 Auto Scaling 그룹에 직접 설정된 예측 크기 조정 정책을 사용해야 합니다.

다음은 AWS에 대해 더 자세히 배울 수 있는 일반 리소스입니다.

- [교육 및 워크숍](#) – 역할 기반의 과정 및 전문 과정은 물론 자습형 실습에 대한 링크를 통해 AWS 기술을 연마하고 실무에 도움이 되는 경험을 쌓을 수 있습니다.
- [AWS 개발자 센터](#) – 자습서를 살펴보고, 도구를 다운로드하고, AWS 개발자 이벤트에 대해 알아보세요.
- [AWS 개발자 도구](#) – AWS 애플리케이션을 개발 및 관리하기 위한 개발자 도구, SDK, IDE 도구 키트 및 명령줄 도구 링크입니다.
- [시작하기 리소스 센터](#) - AWS 계정을(를) 설정하고 AWS 커뮤니티에 가입하고 첫 번째 애플리케이션을 시작하는 방법을 알아보세요.
- [실습 자습서](#) – 단계별 자습서에 따라 AWS에서 첫 번째 애플리케이션을 시작하세요.
- [AWS 백서](#) – AWS 솔루션 아키텍트 또는 기타 기술 전문가가 아키텍처, 보안 및 경제 등의 주제에 대해 작성한 포괄적 AWS 기술 백서 목록의 링크입니다.
- [AWS Support 센터](#) – AWS Support 사례를 생성하고 관리할 수 있는 허브입니다. 또한 포럼, 기술 FAQ, 서비스 상태 및 AWS Trusted Advisor 등의 기타 유용한 자료에 대한 링크가 있습니다.
- [AWS Support](#) – 클라우드에서 1대 1로 애플리케이션을 구축 및 실행하도록 지원하는 빠른 응답 지원 채널인 AWS Support에 대한 정보가 포함된 기본 웹 페이지입니다.
- [문의처](#) - AWS 결제, 계정, 이벤트, 침해 및 기타 문제에 대해 문의할 수 있는 중앙 연락 창구입니다.
- [AWS 사이트 약관](#) – 저작권 및 상표, 사용자 계정, 라이선스 및 사이트 액세스와 기타 주제에 대한 세부 정보입니다.

크기 조정 계획에 대한 문서 기록

다음 표에서는 AWS Auto Scaling 설명서에 대한 중요 추가 사항을 설명합니다. 이 설명서에 대한 업데이트 알림을 받으려면 RSS 피드를 구독하면 됩니다.

변경 사항	설명	날짜
새로운 “보안” 장	AWS Auto Scaling 사용 설명서의 새로운 보안 장은 AWS Auto Scaling 사용 시 공동 책임 모델 을 적용하는 법을 이해하는 데 도움이 됩니다. 이번 업데이트에서는 사용 설명서의 ‘Identity and Access Management’ 장이 새롭고 더욱 유용한 내용인 AWS Auto Scaling의 자격 증명 및 액세스 관리 단원으로 대체되었습니다.	2020년 3월 12일
Amazon VPC 엔드포인트에 대한 지원	이제 VPC와 AWS Auto Scaling 간에 프라이빗 연결을 설정할 수 있습니다. 마이그레이션 고려 사항 및 지침은 크기 조정 계획 및 인터페이스 VPC 엔드포인트 를 참조하세요.	2019년 11월 22일
예상 용량보다 큰 최대 용량 확장에 대한 지원 및 안내서 변경 사항	크기 조정 계획에서 지정된 버퍼 값만큼 최대 용량을 예상 용량보다 크게 확장할 수 있도록 콘솔 지원을 추가합니다. 자세한 내용은 AWS Auto Scaling 사용 설명서의 예측 크기 조정 설정 을 참조하세요. 이 릴리스에는 AWS Auto Scaling 시작하	2019년 3월 9일

	기 자습서에서 재작성된 단원도 몇몇 포함되어 있습니다.	
예측 크기 조정 및 개선	이제 예측 크기 조정을 사용하여 Amazon EC2 Auto Scaling 그룹을 사전 대응적으로 크기 조정할 수 있습니다. 또한 이 릴리스는 크기 조정 계획 외부(예: 다른 콘솔)에서 생성된 크기 조정 정책을 대체하고 계획의 동적 크기 조정 기능을 활성화할지 여부를 제어하기 위한 지원을 추가합니다. 자세한 내용은 AWS Auto Scaling 시작하기 를 참조하세요.	2018년 11월 20일
사용자 지정 리소스 설정에 대한 지원	각 개별 리소스 또는 동시에 여러 리소스에 대한 다양한 설정의 사용자 지정에 대한 지원을 추가했습니다. 자세한 내용은 AWS Auto Scaling 시작하기 를 참조하세요.	2018년 10월 9일
애플리케이션 소스로서의 태그	이 릴리스에는 애플리케이션 소스인 태그 세트 지정에 대한 지원이 추가됩니다.	2018년 4월 23일
새로운 서비스	AWS Auto Scaling의 최초 릴리스입니다.	2018년 1월 16일

기계 번역으로 제공되는 번역입니다. 제공된 번역과 원본 영어의 내용이 상충하는 경우에는 영어 버전이 우선합니다.