

이상희 교사

AWS를 활용한 웹 서비스 구성하기

5단원

▶ 1차수: AWS ELB

2차수: AWS Auto Scaling

3차수: AWS trusted Advisor

1차수

학습목표

- AWS의 핵심적인 서비스와 기능을 학습하기 위한 클라우드 개념을 이해 할 수 있다

1차수

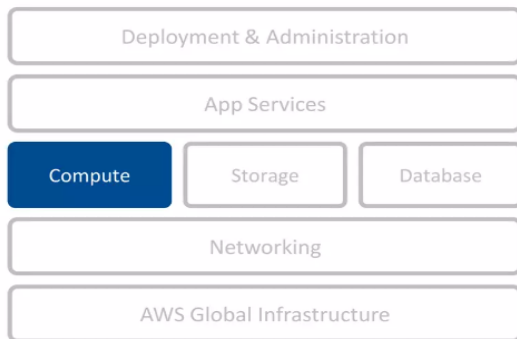
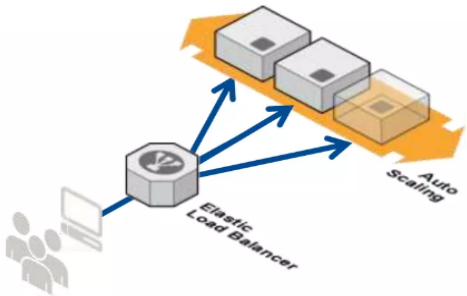
학습내용

- Elastic Load Balancing
- 전통적인 로드밸런서의 기능
- 계층별 로드밸런싱 개념과 특징
- 로드밸런서 유형
- ELB 용어/기능
- 복수의 가용영역을 이용하여 고가용성을 보장

- Elastic Load Balancing 은 AWS EC2 인스턴스에서 운영중인 어플리케이션, 마이크로 서비스 또는 컨테이너 서비스로 유입되는 트래픽을 자동으로 분산 처리하는 기능

Elastic Load Balancing

- Create highly scalable applications
(Traffic에 따른 확장성을 만족하는 애플리케이션 구축 가능)
- Distribute load across EC2 instances in multiple availability zones
(여러 가용 영역에 속해있는 EC2 인스턴스 전체에 로드 분산)



Feature	Details
Auto-scaling	Automatically scales to handle request volume
Available	Load balance across instances in multiple availability zones
Health checks	Automatically checks health of instances and takes them in or out of service
Session stickiness	Route requests to the same instance
Secure sockets layer	Supports SSL offload from web and application servers with flexible cipher support
Monitoring	Publishes metrics to Cloud Watch

- 단일지점을 통해 서버에 연결
- 애플리케이션 환경을 분리
- 고가용성과 내결함성을 제공
- 탄력성과 확장성을 향상

- 로드밸런싱 개념은 기존 온프레미스에서부터 존재
- 이른바 L4 스위치, L7 스위치라고 했는데, L이라는 뜻은 OSI 7 Layer 의 'Layer'에서 유래

L4 (네트워크)스위치

- TCP와 SSL지원
- 클라이언트와 서버가 연결 중계
- 헤더 변경 없음
- 프록시 프로토콜로 요청에 대해 소스 및 목적지 IP주소, 포트 추가

L7(어플리케이션)스위치

- HTTP와 HTTPS지원
- 클라이언트 연결은 로드밸런서에서 종료되고 로드밸런서와 서버는 별도로 연결
- 헤더변경가능
- 요청이 어디서부터 건너왔는지 알려주는 **헤더인 X-Forwarded-For**를 통해 클라이언트 IP를 백엔드 인스턴스로 전달가능



Application Load Balancer

HTTP 및 HTTPS

유연한 애플리케이션 관리
트래픽의 고급 로드 밸런싱
요청 수준(7 계층)에서 작동



Network Load Balancer

TCP 및 UDP

탁월한 성능 및 고정 IP
TCP 트래픽의 로드 밸런싱
연결 수준(4 계층)에서 작동



Gateway Load Balancer

IP

유연한 애플리케이션 관리
트래픽의 고급 로드 밸런싱
요청 수준(3 계층)에서 작동



Classic Load Balancer

HTTP, HTTPS 및 TCP를 위한 이전 세대

EC2-Classic 네트워크를 사용하는 애플리케이션용
요청 및 연결 수준에서 작동

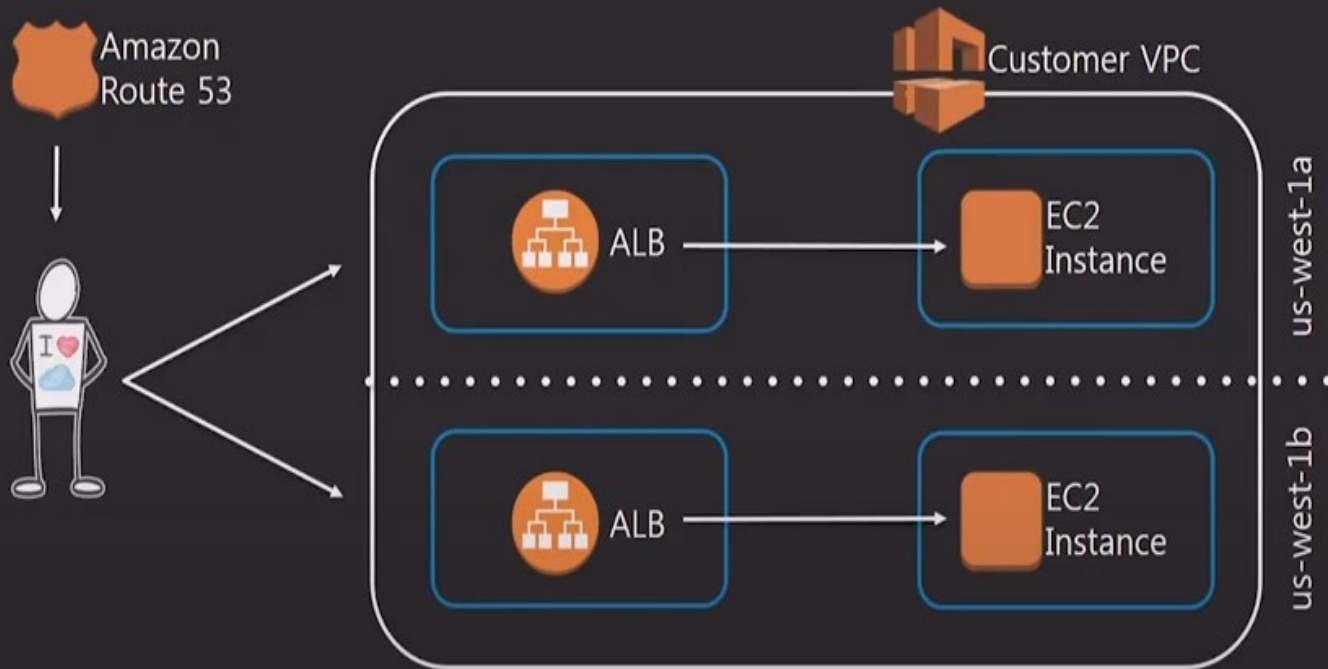
- L4로드밸런싱
 - IP주소와 포트번호를 기준으로 트래픽을 분산
- L7로드밸런싱
 - OSI 7계층의 최상위 계층 프로토콜들의 통신시 프로토콜 헤더의 내용을 기준으로 트래픽을 분산
- 로드밸런싱 알고리즘
 - ELB는 우선순위를 구분하지 않고 순서대로 분배하는 라운드로빈 알고리즘을 사용
- 상태점검(Health check)
 - EC2인스턴스가 정상인가를 점검하는 기능 인스턴스가 기동중이 아니라고 판단되면 해당 인스턴스는 트래픽 배분작업에서 제외됨

- 리스너(listener)
 - 서비스를 위한 유입트래픽 처리를 위한 프로토콜과 포트를 정의 모든 로드밸런서는 최소 1개의 리스너가 필요하고 최대10개까지 설정가능 라우팅 규칙들은 리스너에 정의
- Connection Draining
 - Auto scaling 이 사용자 요청을 처리중인 EC2인스턴스를 바로 삭제 못하게 하는 기능
 - 예를 들면 사용자수가 줄어들면 Auto Scaling이 인스턴스를 삭제 할 수 도 있는데 사용자가 해당 인스턴스에 연결하여 작업 시 완료되기전에 인스턴스가 삭제되면 작업수행에 문제가 되므로 삭제 전 사용자의 작업이 완료될 수 있도록 일정시간 대기를 합니다.또한 대기하는 동안에는 새로운 연결을 수용하지 않음

- Sticky Session
 - L7로드 밸런싱을 의미하는 용어로 사용자의 세션을 확인하여 적절한 인스턴스로 트래픽을 분배하는 기능 예를 들면 동일한 사용자가 서비스에 계속 접속할 때 처음 접속했던 인스턴스에 연결시킴 이 기능을 사용하지 않으면 라운드로빈 알고리즘에 따라 매번 다른 인스턴스에 연결됨
- 지연(Latency)
 - ELB로드밸런서와 EC2인스턴스간의 지연시간
- 요금

기능	Application Load Balancer	Network Load Balancer	Gateway Load Balancer	Classic Load Balancer
상태 확인	✓	✓	✓	✓
CloudWatch 지표	✓	✓	✓	✓
로깅	✓	✓	✓	✓
보안 소켓 계층(SSL) 오프로딩	✓	✓		✓
Connection Draining	✓	✓	✓	✓
소스 IP 주소 유지	✓	✓	✓	
고정 IP 주소		✓		
대상으로서 Lambda 함수	✓			
리디렉션	✓			
고정 응답 작업	✓			

복수의 가용 영역

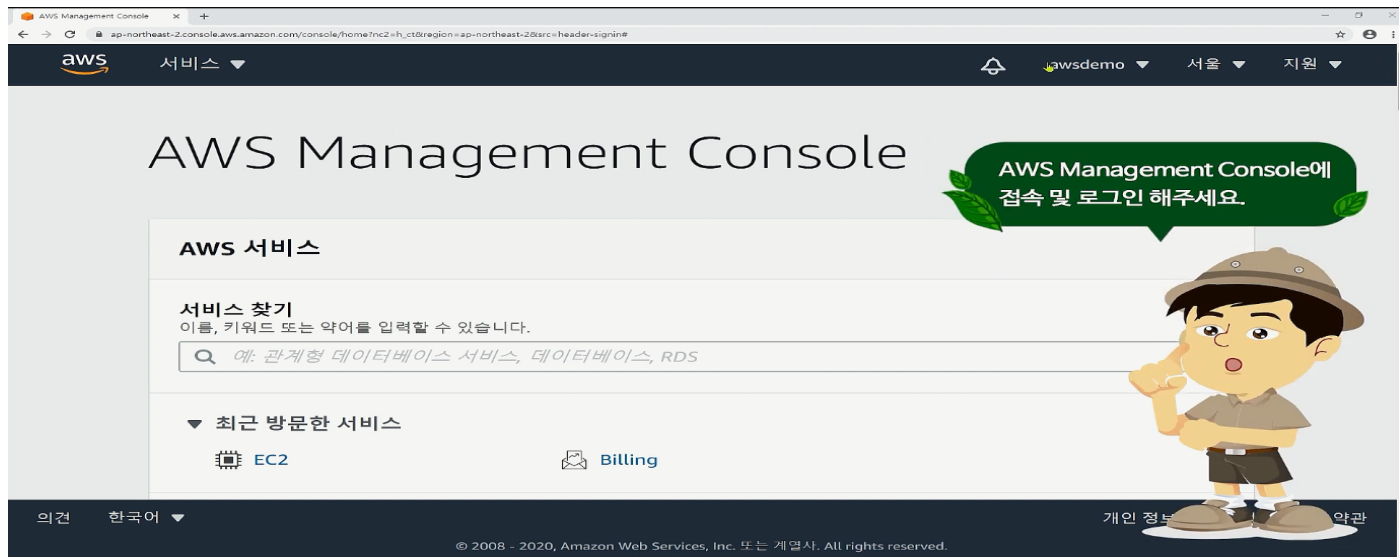


■ AWS ELB실습2-1

<https://rumble.com/v35cxze-elb2-1.html>

■ AWS ELB실습2-2

<https://rumble.com/v35cy3m-elb2-2.html>



Q1

다음중 전통적 L7(어플리케이션)스위치에서 제공하는 기능을 설명한 것 중 잘못된 것 하나는?

가.HTTP와 HTTPS지원

나.클라이언트 연결은 로드밸런서에서 종료되고 로드밸런서와 서버는 별도로 연결

다.헤더변경은 불가능

라.요청이 어디서부터 건너왔는지 알려주는 헤더인 X-Forwarded-For 를 통해 클라이언트 IP 를 백엔드 인스턴스로 전달가능

A

답:

다(모든 L7스위치는 패킷 헤더변경기능이 내장되 있음)

Q2

다음 ELB 개념에 관한 용어는 무엇인가?

L7로드 밸런싱을 의미하는 용어로 사용자의 세션을 확인하여 적절한 인스턴스로 트래픽을 분배하는 기능

예를 들면 동일한 사용자가 서비스에 계속 접속할때 처음 접속했던 인스턴스에 연결시켜줌 이 기능을 사용하지 않으면 라운드로비 알고리즘에 따라 매번 다른 인스턴스에 연결된다

A

답:

Sticky Session(스티키세션)

이상희 교사

AWS를 활용한 웹 서비스 구성하기

5주차

1차수: AWS ELB

▶ 2차수: AWS Auto Scaling

3차수: AWS trusted Advisor

2차수

학습목표

- AWS 의 Autoscaling 서비스를 이해 할 수 있다

2차수

학습내용

- Autoscaling이란?
 - Auto Scaling 시작 템플릿
- 주요 기능
 - Auto Scaling 그룹 구성 요소
- Scaling의 종류(수직적&수평적)
 - 그룹 용량
- 전통적 IT 리소스 사용패턴
 - EC2 Auto Scaling 호출
- 클라우드 IT 리소스 사용패턴
 - EC2 Auto Scaling 을 사용하는 스케일링 방법
- 오토스케일링 사용 고려사항
 - EC2 Auto Scaling 을 사용하여 비용 최적화
- Auto Scaling 구성요소
 - 오토스케일링 모니터링,확장,로드밸런싱ELB + CloudWatch

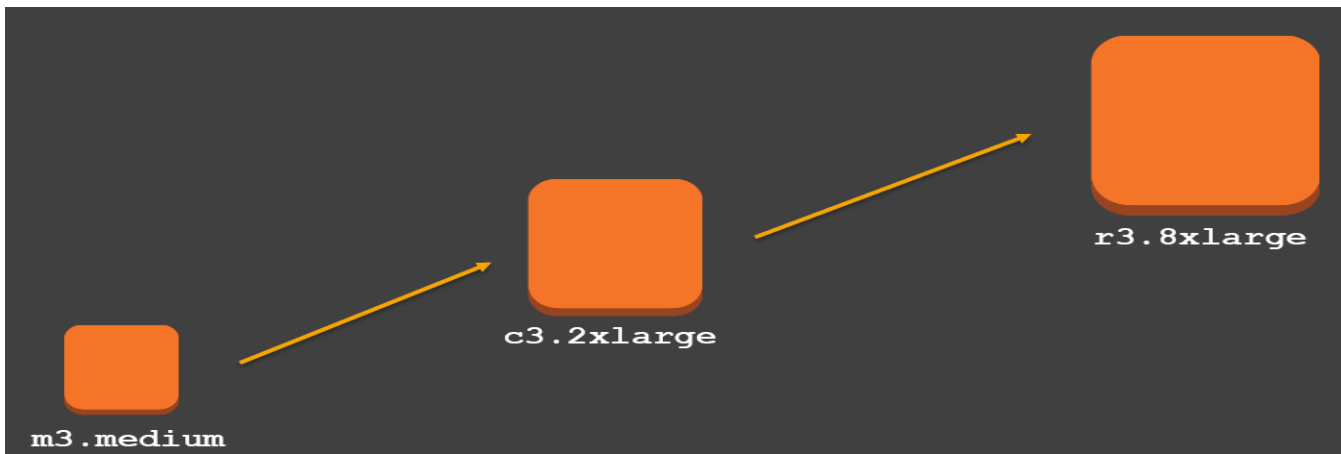
- Auto scaling이란 Amazon EC2를 포함 개별 AWS 서비스에 대한 확장 가능 리소스를 자동 조정하기 위한 솔루션이며 정책에 따라서 시스템을 자동으로 늘리고 줄여주는 서비스
- 사용자가 정의한 주기(스케줄링)나 이벤트(모니터링 알람)에 따라 가상서버를 자동으로 생성하거나 삭제한다

- 수요가 증가하면 서버를 자동으로 생성하여 늘림
- 수요가 감소하면 불필요한 서버규모를 줄이거나 자동 삭제
- 모니터링 서비스의 알람에 의해 동적으로 서버를 조절
- 사용자가 정의한 스케줄에 따라 서버를 조절
- 비정상적인 서버를 자동으로 삭제하고, 사용자가 정의한 용량을 유지
- 서비스제공 품질에 일관성을 부여
- 가상서버이용에 따르는 요금 절감

수직적Scaling(Scale Up/Down)

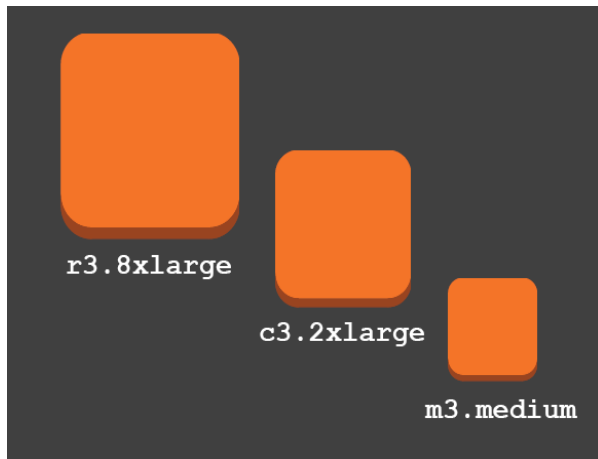
인스턴스 종류를 변경

EBS 용량증가





- 다른 EC2 인스턴스 종류
 - High memory instances
 - High CPU instances
 - High I/O instances
 - High storage instances
- 인스턴스의 크기변경이 용이
- 결국 한계점에 도달
- 인스턴스 Stop/Start가 필요

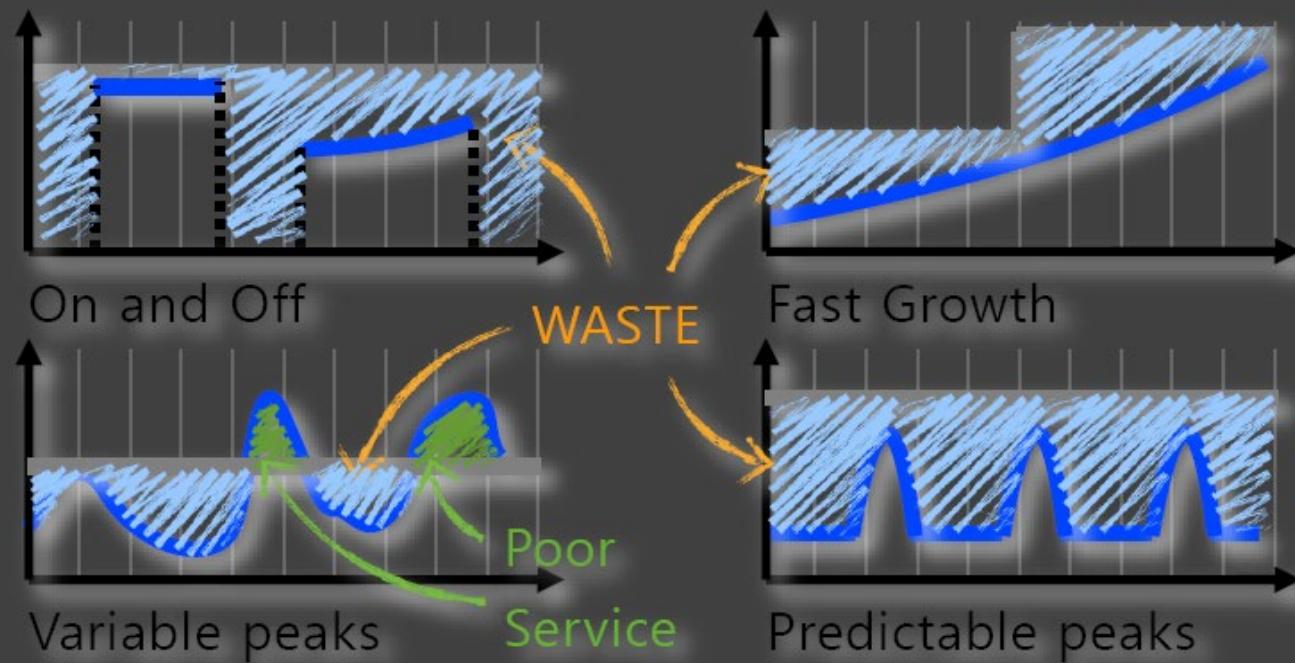


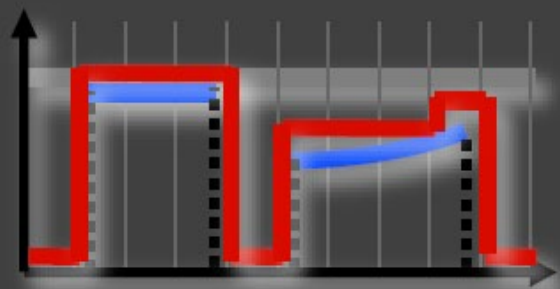


수평적Scaling(Scale Out/In)

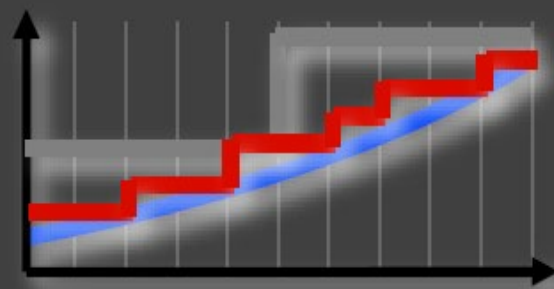
- 인스턴스 추가/제거
- ELB
- 오토스케일링(Autoscaling)



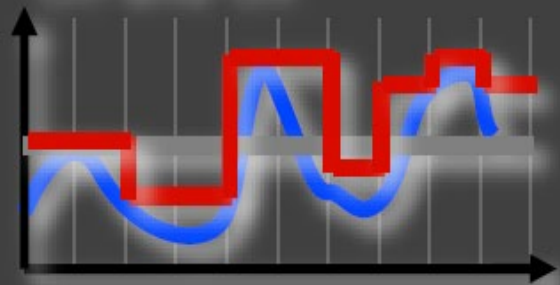




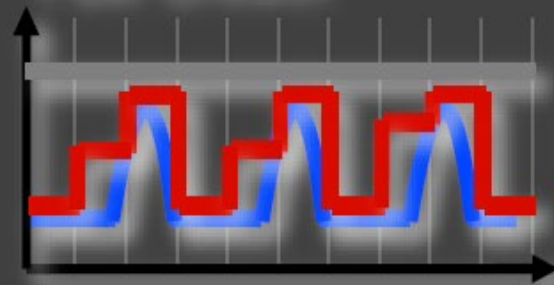
On and Off



Fast Growth

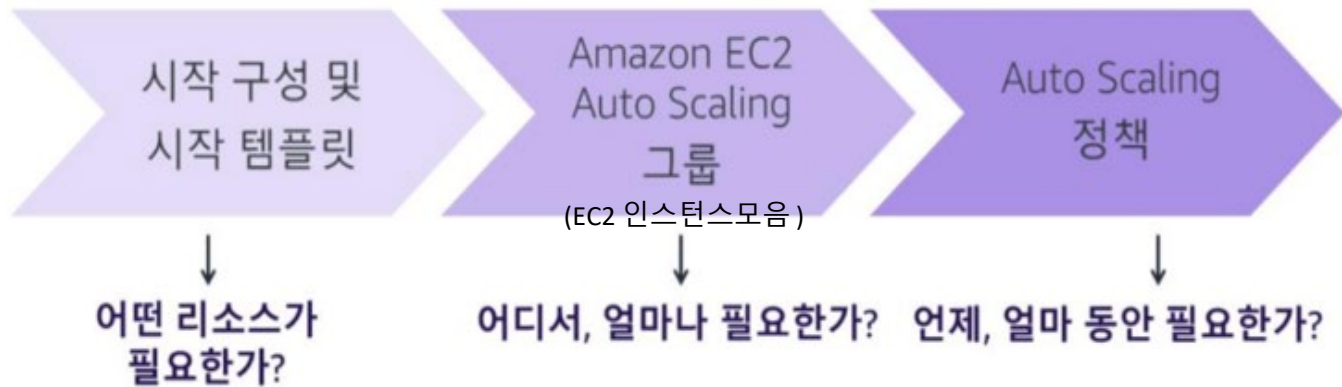


Variable peaks



Predictable peaks

- 워크로드의요구에 맞게 컴퓨터클러스터의 크기를 조정한다
- 인스턴스의 최대/최소숫자를 정의한다
- 언제 스케일아웃 및 스케일인이 발생할지 정의한다
- 스케일링 구동을위해 Amazon CloudWatch를 통해 메트릭 정보수집을 진행
- 온디맨드 및 스팟인스턴스유형에 대한 오토스케일링 실행
- 오토스케일링기능은 무료!!



시작 구성 및 시작
템플릿

어떤 리소스가
필요한가?

- AMI
- 인스턴스 유형
- 보안 그룹
- 역할


Amazon EC2
Auto Scaling
그룹

어디서, 얼마나 필요한가?

- VPC 및 서브넷
- 로드 밸런서

정의:

- 최소 인스턴스
- 최대 인스턴스
- 원하는 용량



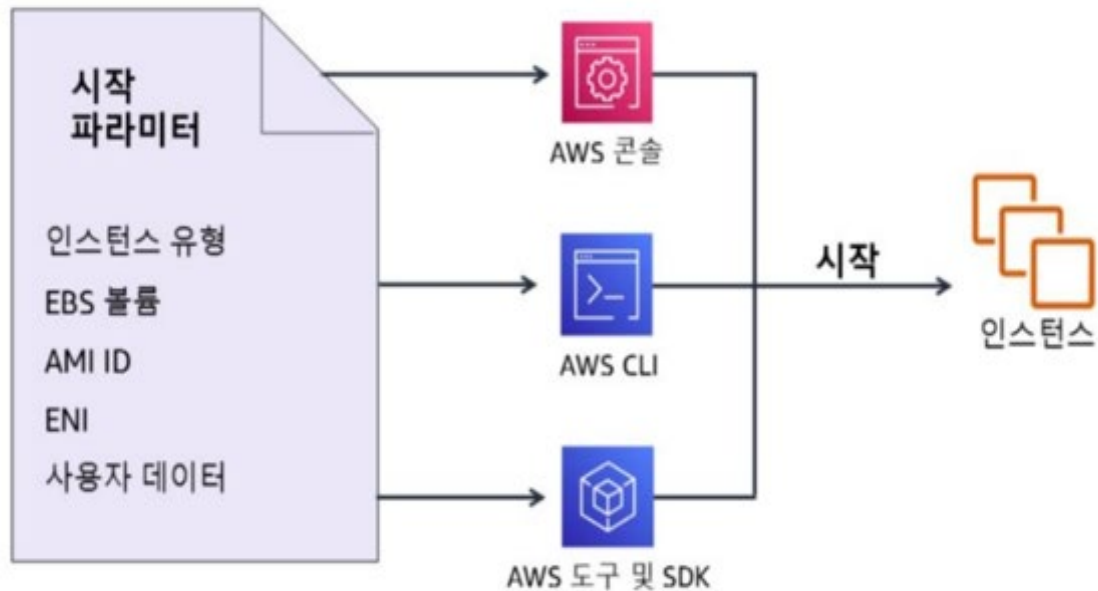
Auto Scaling 정책

언제, 얼마 동안 필요한가?

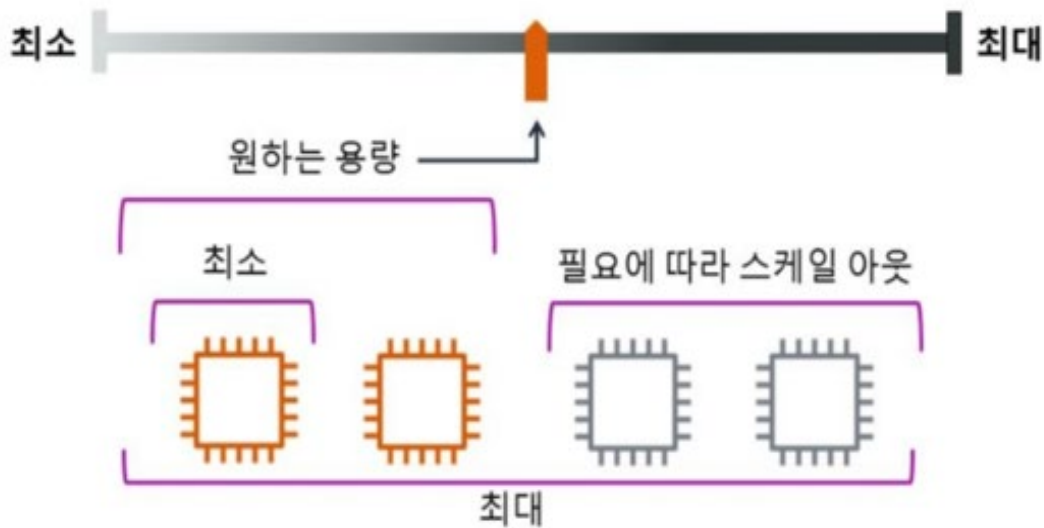
-
- 예약
 - 온디맨드
 - 예측 가능 오토 스케일링
 - 스케일 아웃 정책
 - 스케일 인 정책

시작

- 일관된 환경
- 단순한 권한
- 거버넌스 및 모범 사례
- 생산성 향상



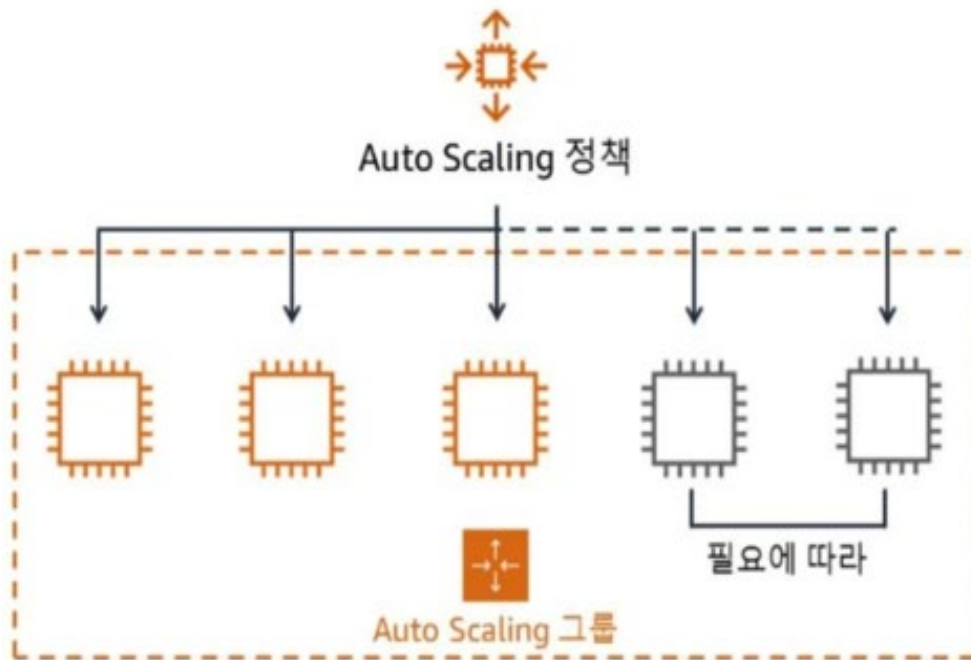
Amazon EC2 Auto Scaling 그룹



- 인스턴스의 논리적 그룹화
- 허용되는 최소 및 최대 인스턴스 수를 지정
- 인스턴스를 시작 또는 종료하여 용량 충족

다음을 사용하여
스케일링 호출

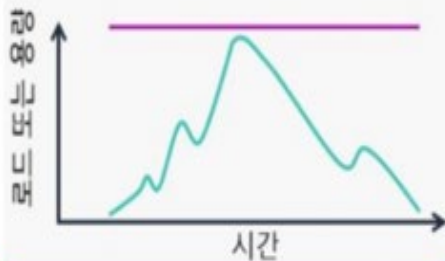
- 상태 확인
- CloudWatch 경보
- 일정
- 수동 스케일링



Auto Scaling 정책

예약

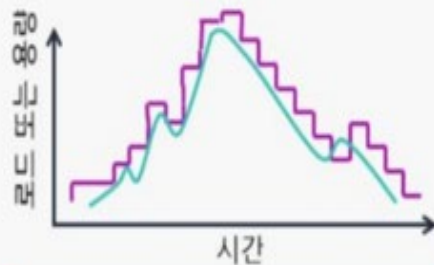
예측 가능한 워크로드



— 프로비저닝된 용량

동적

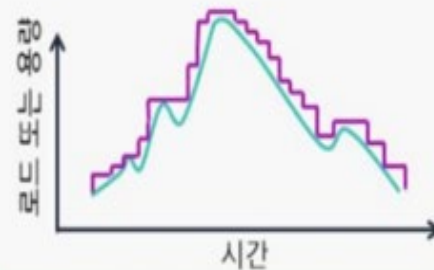
일반적 스케일링



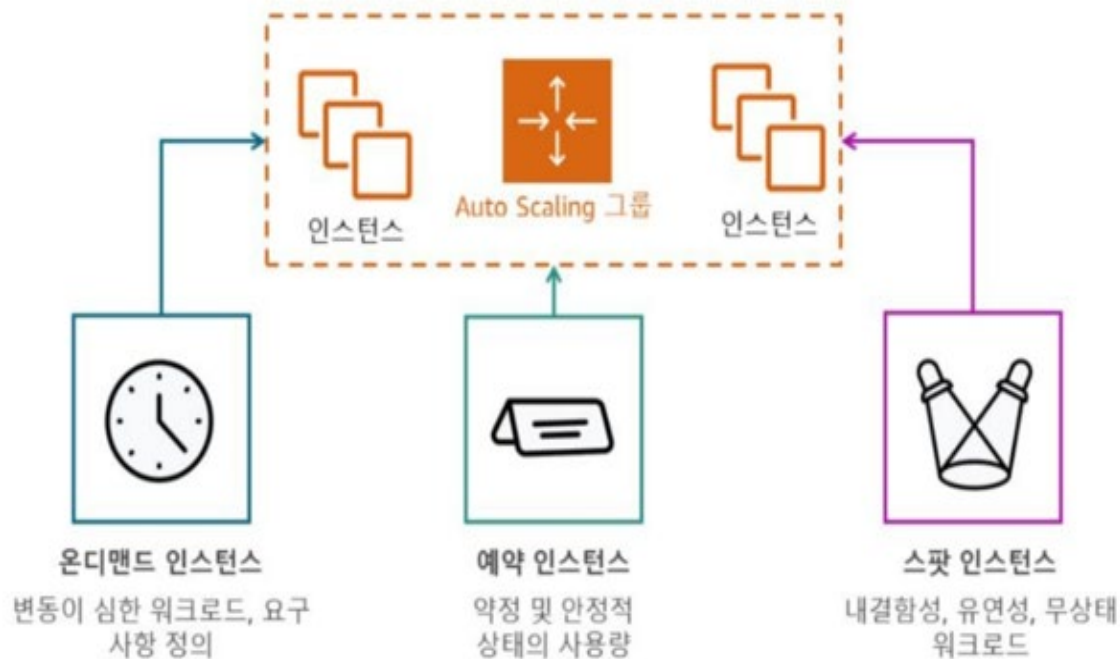
— 실제 용량 수요

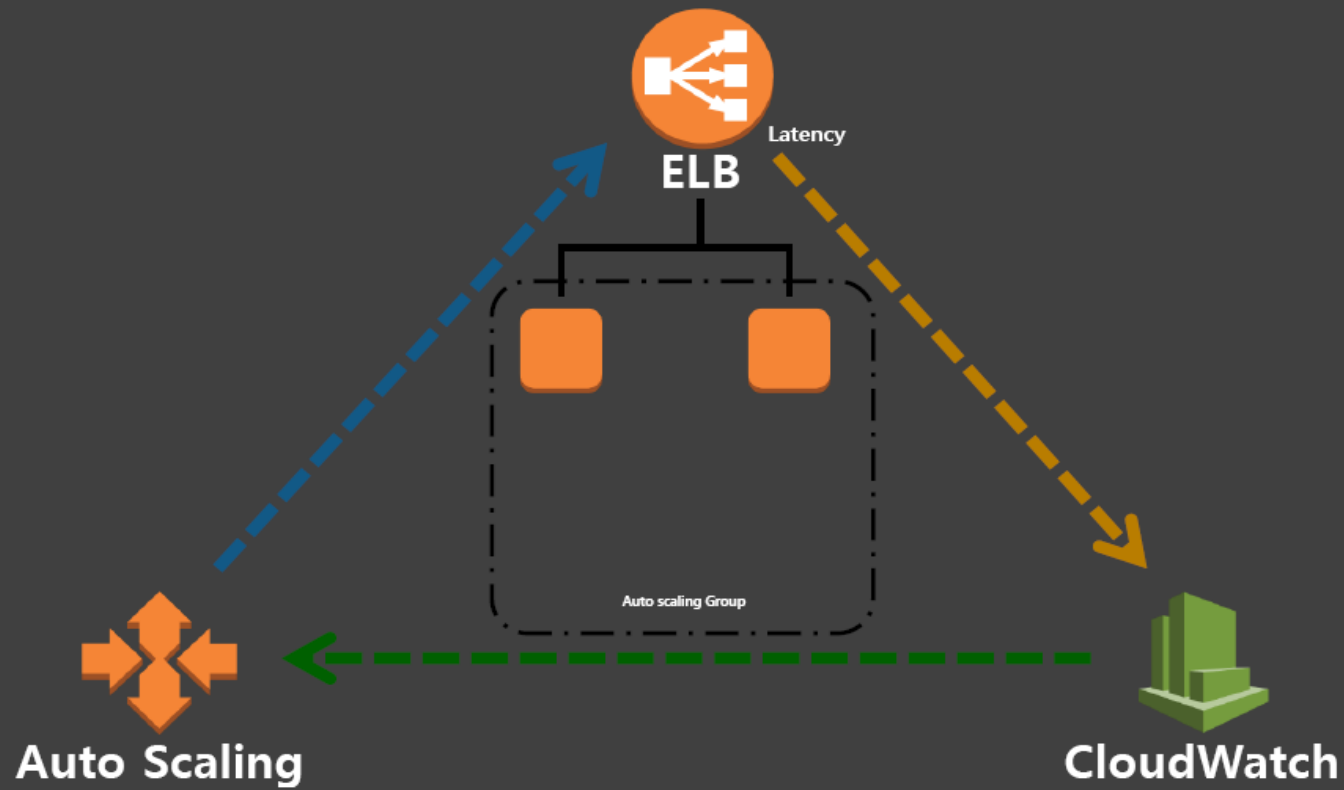
예측

가장 사용이 간편

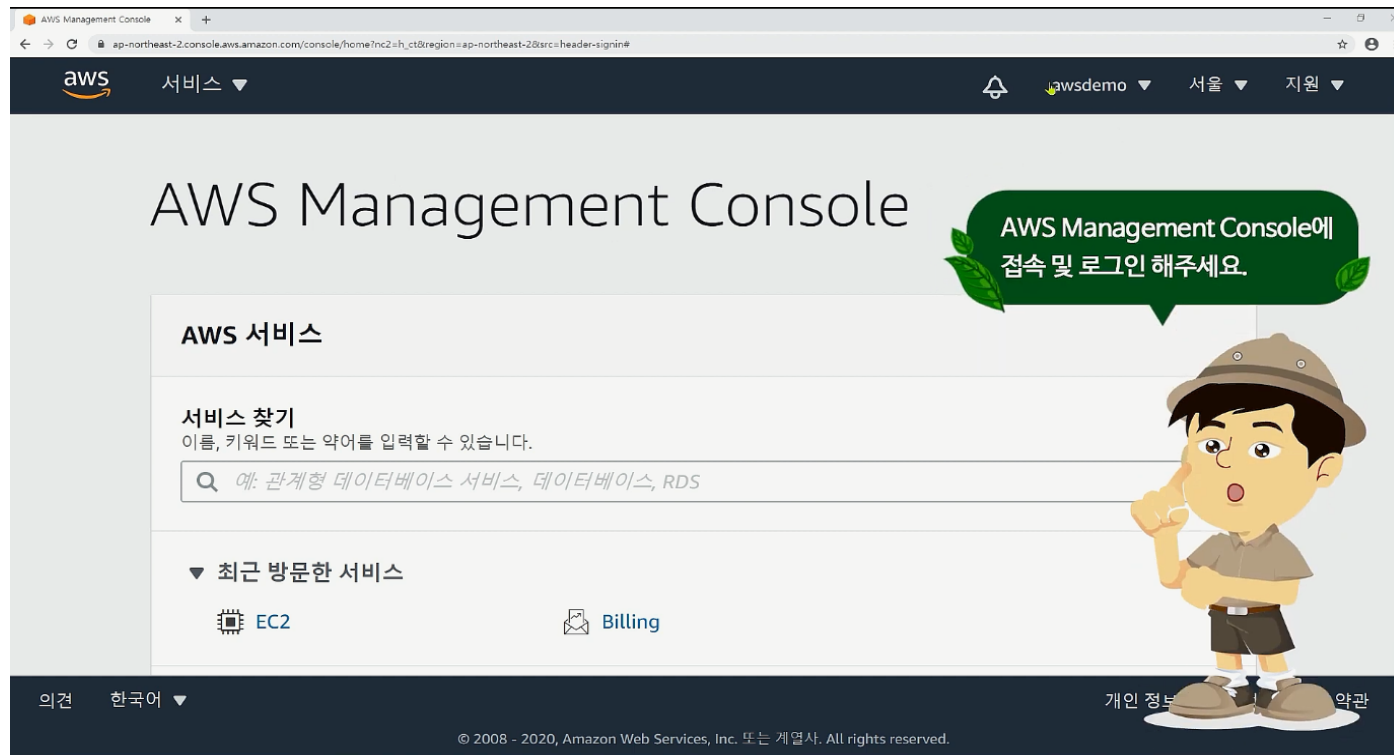


규칙을 수동으로
조정할 필요가 없음





- <https://rumble.com/v35e48c-autoscaling-update.html>



Q
1

다음 중 오토스케일링의 기능에 대해 잘못 서술한 것은?

- 가.수요가 증가하면 서버를 자동으로 생성하여 늘림
- 나.수요가 감소하면 불필요한 서버규모를 줄이거나 자동삭제
- 다.모니터링 서비스의 알람에 의해 동적으로 서버를 조절
- 라.늘어나거나 확장된 가상서버비용에 따르는 요금 부담 가중

A

답:
라(가상서버비용에 따르는 요금이 절감된다)

Q
2

다음 설명은 어떤 스케일링 방법을 설명한 것인지 답하시오

규모는 동일한 규모의 인스턴스의 숫자를 늘림으로서 늘어나는 부하를 늘어난 인스턴스로 분산시킴으로 균일한 서비스 제공과 성능을 개선하는 방법이다

A

답: 수평적Scaling (Scale Out/In을 이용하는 기술이다)

이상희 교사

AWS를 활용한 웹 서비스 구성하기

5주차

▶ 1차수: AWS ELB

2차수: AWS Auto Scaling

3차수: AWS trusted Advisor

3차수

학습목표

- AWS 를 효율적으로 이용하기 위한 Trusted Advisor를 이해 할 수 있다

3차수

학습내용

- AWS Trusted Advisor 소개/ 정의/ 분석항목
- Trusted Advisor 의 기능 / 대시보드의 기능
- Trusted Advisor 알람 기능/ 접속 컨트롤
- 누구나 사용가능한 AWS Trusted Advisor
- Trusted Advisor 는 AWS 관리 콘솔 이외에도 사용 가능

Security Group 이
잘못 설정되지는
않았는지.
특정 포트가 외부에
개방 되지는
않았는지..

AWS 에 구축한
서비스
아키텍처가 AWS
에서 권고하는
권장사항에
부합하는지....

AWS 계정 담당자 or 재무 담당자

현재 활용되고 있지않은
리소스를 삭제하여 비용을
줄
일 수 있는 여지가 있는지..?

AWS Trusted
Advisor 를
활용하면 편리!!

- AWS 모범 사례에 따라 AWS를 효율적으로 쓰고 있는건지, 비용을 절감할 여지는 없는지, 보안 설정은 잘 구성되어 있는지 정기적으로 점검하여 이러한 사항을 분야별로 나눠 실시간 지침을 제공하는 온라인 도구




Trusted Advisor > 권장 사항

Trusted Advisor 권장 사항


모든 검사 새로 고침 | 모든 검사 다운로드

이 페이지를 사용하여 AWS 계정에 대한 검사 결과의 개요를 확인할 수 있습니다. 검사 이름 또는 검사 카테고리를 선택하면 Trusted Advisor가 찾아낸 권장 조치 또는 잠재적인 문제를 확인할 수 있습니다. 각 검사는 문제를 해결하는 방법에 대한 추가 정보를 제공합니다. 모든 검사 결과의 요약물 다운로드할 수도 있습니다. [자세히 알아보기](#)

검사 요약

 2	 1	 0
조치 권고 정보	조사 권고 정보	제외 항목이 있는 검사 정보
보안 2	보안 1	

권장 조치

▶  루트 계정의 MFA

최종 업데이트: 36분 전

AWS Trusted Advisor 컨셉 - 출처:

<https://aws.amazon.com/premiumsupport/trustedadvisor/>

- 비용 최적화 비용을 절약할 수 있는 포인트를 찾아서 권고
- 보안 보안성에 문제가 될만한 부분을 찾아서 변경구성을 권고
- 고가용성 애플리케이션의 가용성을 높이기 위한 권고 사항제시
- 애플리케이션의 실행속도 및 응답성 개선에 도움이 되는 권고 사항제시
- 서비스 한도 점검

- 대시 보드
- Alarming
- Trusted Advisor 접속 컨트롤
- AWS Support API 를 사용한 검사 결과

- AWS 의 자원과 비용에 대한 개요 표시
- AWS 관리 콘솔내에서 표시 가능
- 각 카테고리 별로 아래그림과 같이 표시됨

Cost Optimization



0 7 0

\$2,715.60

Potential monthly savings

Performance



6 2 0

Security



3 5 3

Fault Tolerance



11 2 2



초록 : 이슈 사항 미발견



노랑 : 조사가 권장되는 항목



빨강 : 즉시 대응을 권장하는 항목

- Trusted Advisor 의 체크 상태와 개선의 윤곽을 전자메일로 통보
 - 변경된 점검 상태는 강조 표시
- 이용 방법
 - 대시 보드에서 알람을 활성화
 - 재무, 오퍼레이션, 보안 등 총 3개의 수신 주소에 대해 지정 가능
 - 통상 -매주 목요일 or 금요일에 알람 이메일이 발송됨
- 무료!!



- IAM 을 사용하여 Trusted Advisor 로의 접속제어가능
- 재무 전용 IAM account 생성 가능

Trusted Advisor 접속 컨트롤

Trusted Advisor 의
권한을 허용하는 사례

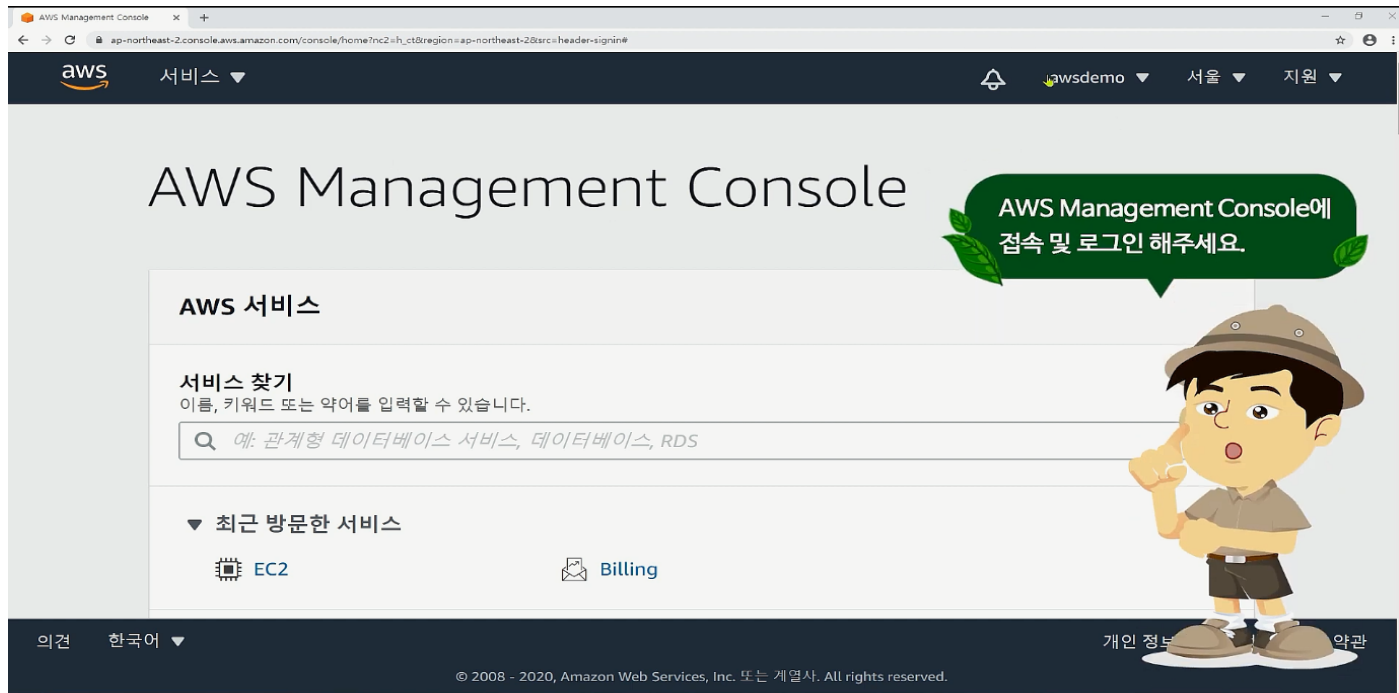


```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "trustedadvisor:*",
      "Resource": "*"
    }
  ]
}
```

- 4가지 Best Practice 를 무료이용 가능
 - 서비스 한도 현황 점검 (EC2 갯수, EIP 갯수..)
 - 퍼블릭하게 개방된 포트 확인
 - IAM 의 사용 체크
 - 루트 계정의 다중요소인증 (MFA) 사용 체크
- 다른 Best Practice 는 비즈니스/엔터프라이즈 서포트 이상의 등급에서 사용 가능
- Trusted Advisor는 AWS 관리 콘솔에서 이용 가능

- AWS Support API 를 사용 및 자동화 가능
 - AWS SDK / AWS CLI 로 이용 가능
 - AWS 비즈니스/엔터프라이즈 서포트 등급 이상에 해당
 - Trusted Advisor 사용 이외에도, AWS Support Program으로도 이용이 가능
- 상세사항
<https://docs.aws.amazon.com/awssupport/latest/user/getting-started.html>

- <https://rumble.com/v35fzaw-trusted-advisor-202308.html>



Q1

다음중 AWS Trusted Advisor의 분석항목이 아닌 것을 고르시오

가.비용 최적화

나.성능

다.고가용성 내결함성

라.사용자의 수에 따른 비용

A

답:

라 (사용자의 수에 따른 비용은 분석 항목에 없다)

Q2

다음과 같은 정보를 표시 확인 하는 Trusted Advisor의 기능은 무엇인가?

AWS 의 자원과 비용에 대한 개요 표시

AWS 관리 콘솔내에서 표시 가능

각 카테고리 별로 초록,황색,적색으로 각각 이슈사항 미발견,조사가 권장되는 항목,즉시 대응을 권하는 항목을 표시함

A

답:대시보드(Dash Board)