

알아두면 쓸모 있는 **EC2** 이야기

이우진

솔루션즈 아키텍트, AWS



Agenda

Amazon EC2 기본 개념

워크로드에 적합한 EC2 인스턴스 선택하기

EC2 인스턴스 최적화

스팟 인스턴스를 통한 유연한 확장

EC2 데이터 보호 관리

폭넓은 컴퓨팅 사용 고객

모든 워크로드를 위한 컴퓨팅 플랫폼

STARTUP



ENTERPRISE



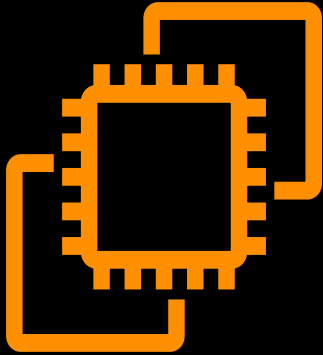
PUBLIC SECTOR



SI & ISV

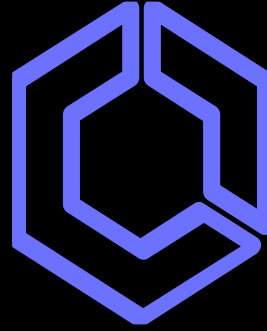


모든 워크로드를 위한 컴퓨팅 플랫폼



Amazon EC2

클라우드의 가상 서버
인스턴스



Amazon ECS, EKS, and Fargate

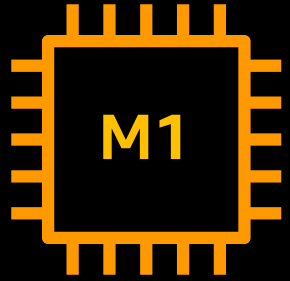
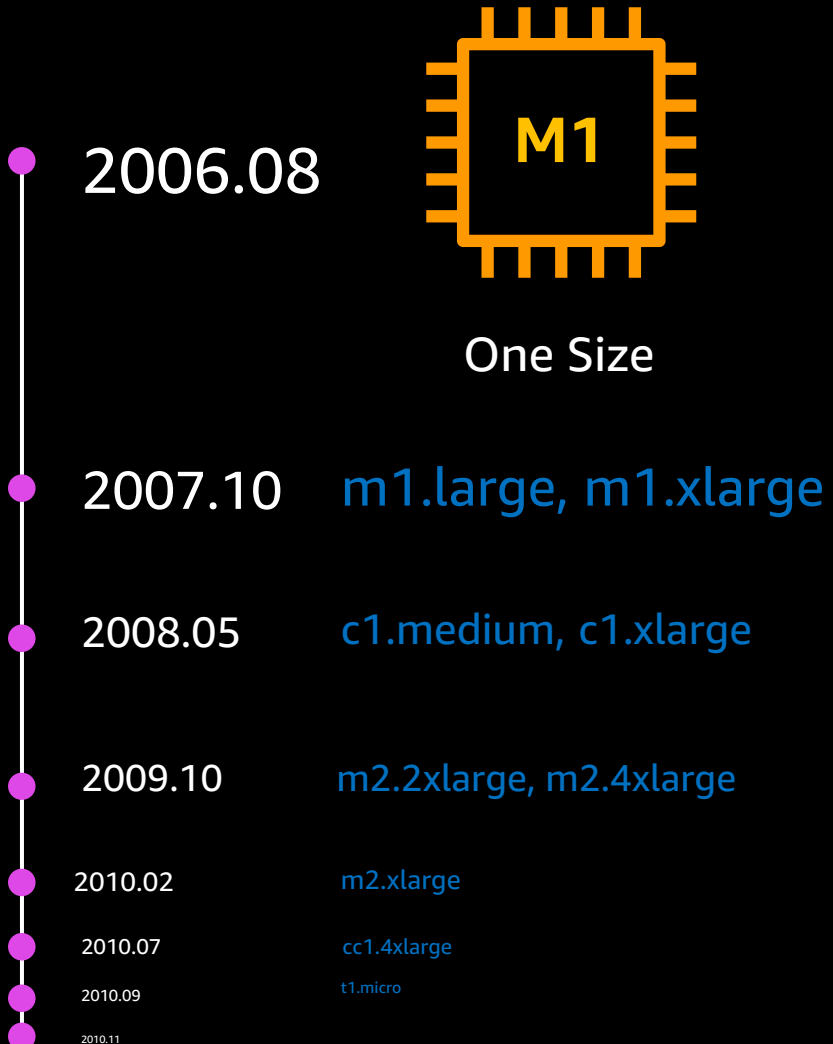
EC2의 관리 클러스터에서
Docker를 실행하기 위한
컨테이너 관리 서비스



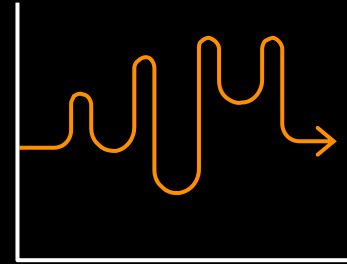
AWS Lambda

트리거에 대한 응답으로
Stateless 코드 실행을 위한
서버리스 컴퓨팅

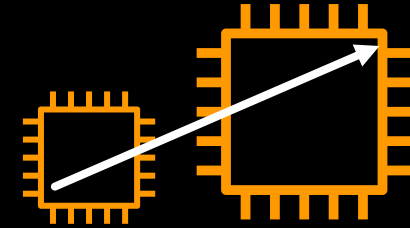
Amazon EC2@2006



One Size

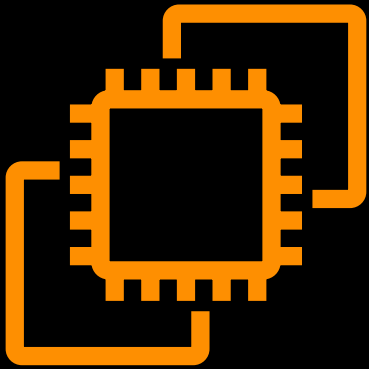


One Pay option
On-demand



Scale Up or Down

Amazon Elastic Compute Cloud(EC2)



Amazon EC2

Linux | Windows | MAC

Arm | x86 | AMD 아키텍처

범용 및 특정 워크로드용

베어메탈, 디스크, 네트워크 성능

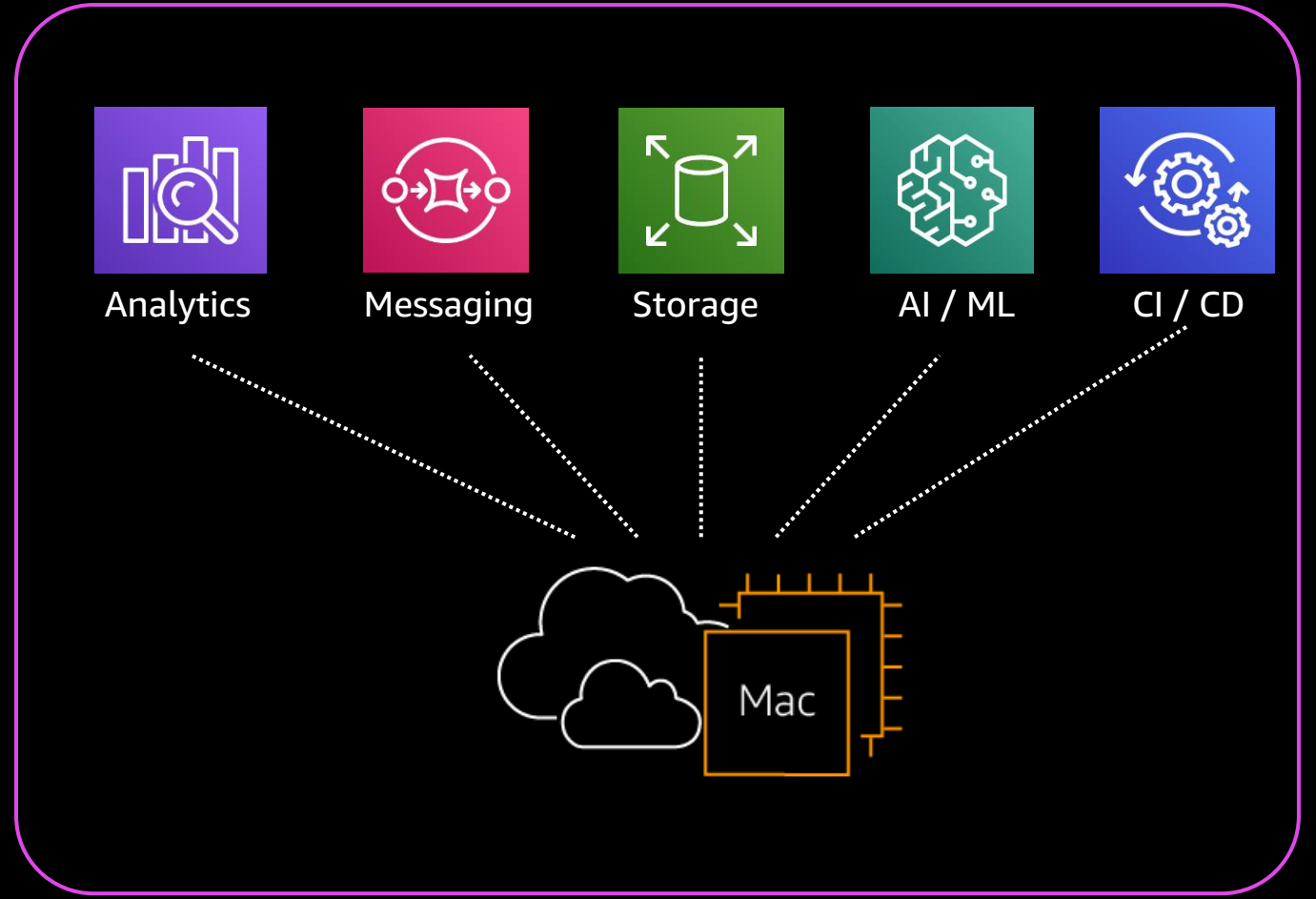
Packaged | Custom | Community AMIs

여러가지 구매 옵션: 온디맨드, 세이빙플랜, 스팟 인스턴스

Amazon EC2 Mac instances

APPLE 개발자를 위한 인스턴스 환경 제공

- Desktop 구입 또는 별도 계약 없이 Mac 환경을 제공
- 인프라 대신 코드에 집중
- DevOps 파이프라인 통합
- 더 높은 민첩성, 탄력성 및 보안
- AWS services 와 연동
(e.g., EBS, VPC, SSM, IAM, CloudWatch)



AWS License Manager

온프레미스 및 클라우드를 위한 간소화된 라이선스 관리

- 소프트웨어 공급 업체의 라이선스를보다 쉽게 관리
- 라이선스 규칙 정의, 사용량 검색, 접근 관리
- AWS 및 온-프레미스에 대해 단일 라이선스 관리
- 비 준수 소프트웨어 발견 및 오용 방지
- AWS Systems Manager 및 AWS Organizations와 완벽한 통합
- 모든 고객을 위한 무료 서비스

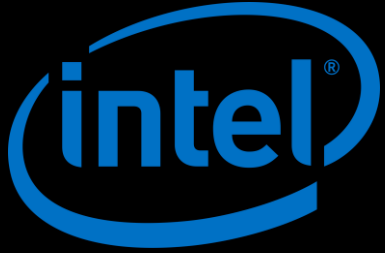


Microsoft
SQL Server



Processors and Accelerators

프로세서나 아키텍처의 폭넓은 선택



Intel Xeon Scalable
Processors



AMD EPYC
Processors



AWS Graviton3
Processors



Tensor Core
GPUs



FPGAs for custom
Hardware acceleration

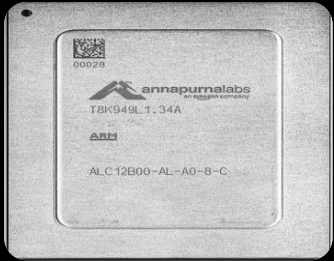
애플리케이션 및 워크로드에 맞는 적합한 프로세서

AWS Graviton Processor

AMAZON EC2 최고의 가격 대비 성능 제공

AWS Graviton 프로세서는 Amazon EC2에서 실행하는 클라우드 워크로드에 최고의 가성비를 제공하기 위해 64비트 Arm Neoverse 코어를 사용하여 Amazon Web Services에서 맞춤형으로 구축한 제품입니다.

AWS Graviton2 vs First generation



64-bit Arm Neoverse
cores

~30B Transistors

7nm technology

7x
performance

4x
compute cores

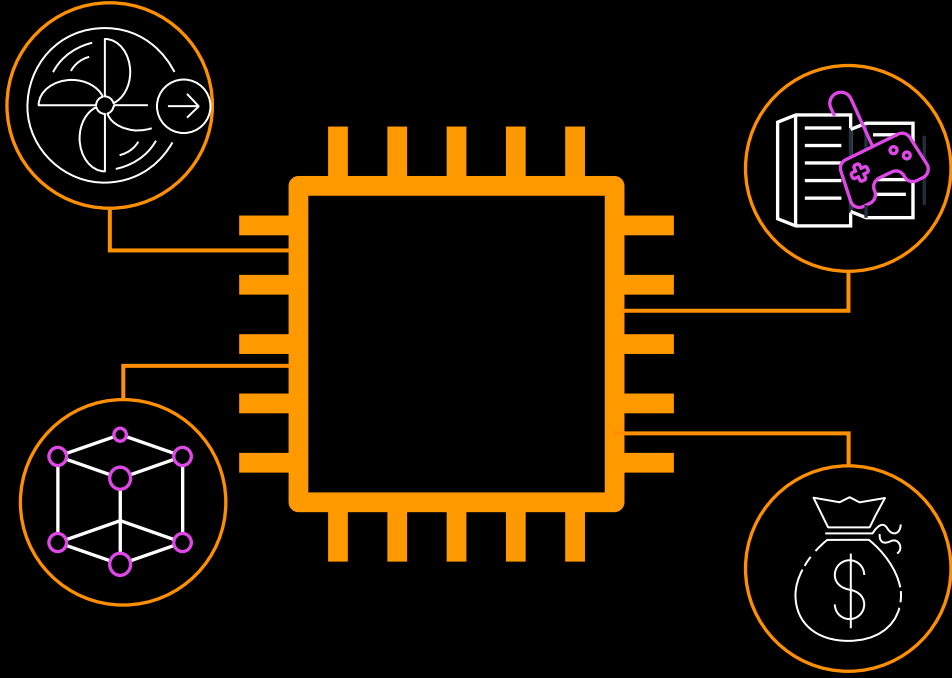
5x
faster memory

2x
cache

- ✓ AWS Linux2 | RHEL 8.2+ | SUSE | ubuntu
- ✓ Cloud-native workload에 대한 최적화
- ✓ Rapidly innovate, build, and iterate on behalf of customers
- ✓ Amazon RDS | EMR | ElastiCache | Elasticsearch 서비스에도 사용
- ✓ M6g, T4g, C6g, R6g, C6gn, X2gd instance

AWS Graviton3 *NEW*

GRAVITON2 대비 25% 더 높은 성능 향상



REINVENTING COMPUTING WITH
AWS Graviton3

- ✓ Cryptographic 및 Scientific 워크로드에 2x 성능 향상
- ✓ ML 워크로드에서 3x 성능 향상
- ✓ 최대 60% 에너지 사용 절감
- ✓ Preview : Amazon C7g , G5g instance

Amazon EC2@2021

광범위하고 세분화된 컴퓨팅 자원 선택

분류

CAPABILITIES

OPTIONS

450+

인스턴스 타입

for virtually every
workload and
business need

범용 M

버스팅용 T

컴퓨팅 C

메모리 R

고 I/O 스토리지 I

고밀도 스토리지 D

범용 GPU P

그래픽 G

성능 가속 F

프로세서
(Intel, AMD a, AWS g)

프로세서 속도
(up to 4.5 GHz)

메모리 확장 x
(up to 24 TiB)

스토리지 확장 d
(HDD and SSD)

네트워킹 n
(up to 400 Gbps)

베어 메탈 m
다양한 사이즈 s/m/l/xl
(Nano to 32xlarge)

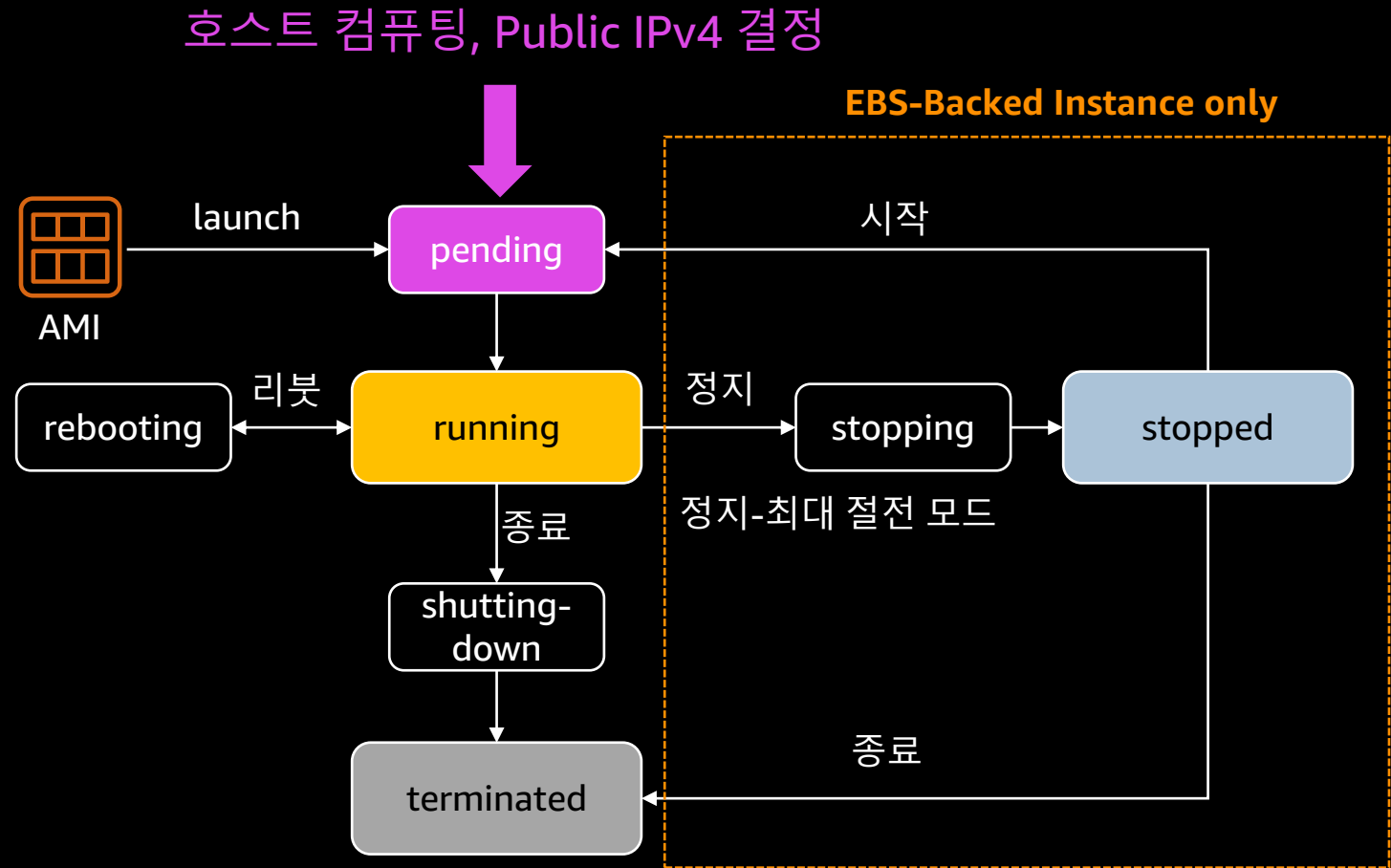
Amazon EBS

Amazon Elastic Inference

Elastic Fabric Adapter

Amazon EC2 수명 주기

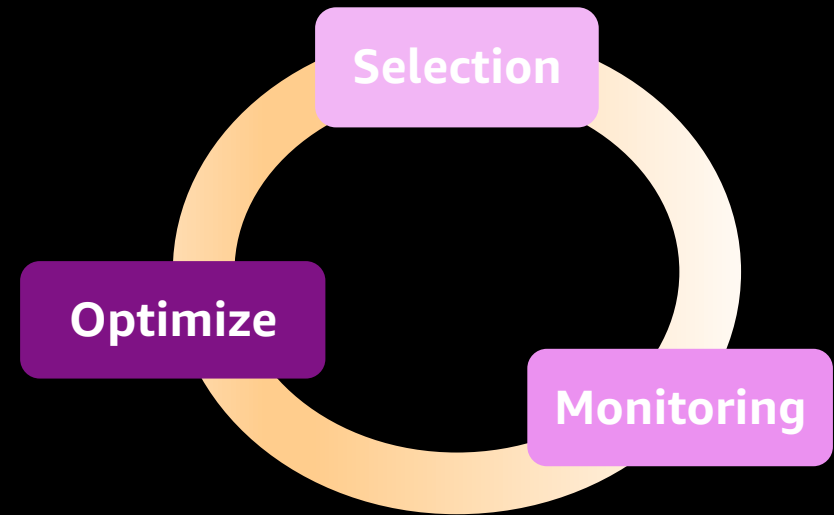
- **실행중(Running)**
 - 인스턴스 동작 중 상태
 - 과금 발생
 - 정지/종료/리붓 명령으로 상태 전이 가능
- **정지됨(Stopped)**
 - 중지된 상태
 - **EBS 볼륨을 루트로 사용하는 인스턴스만 가능**
 - 과금 안 됨
 - 시작/종료 명령으로 상태 전이 가능
 - **탄력적 IP 주소 유지**
- **종료됨(Terminated)**
 - 인스턴스가 완전히 제거된 상태
 - 시작이나 정지 불가능
 - 과금 안 됨



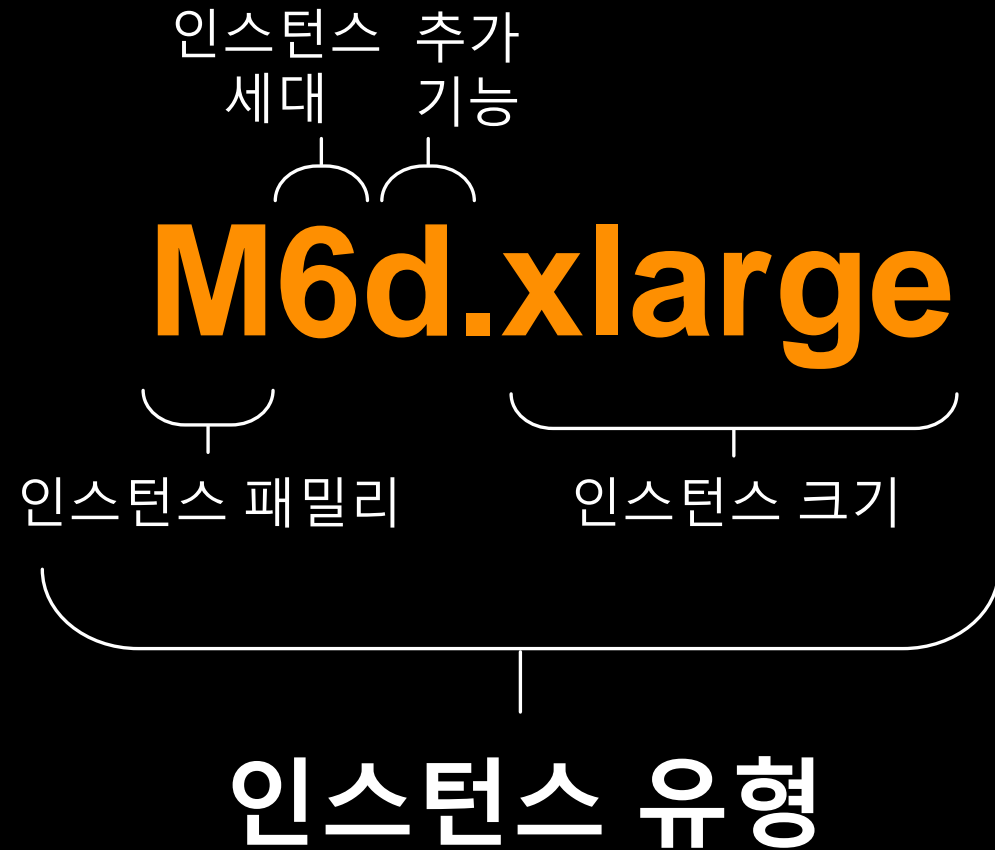
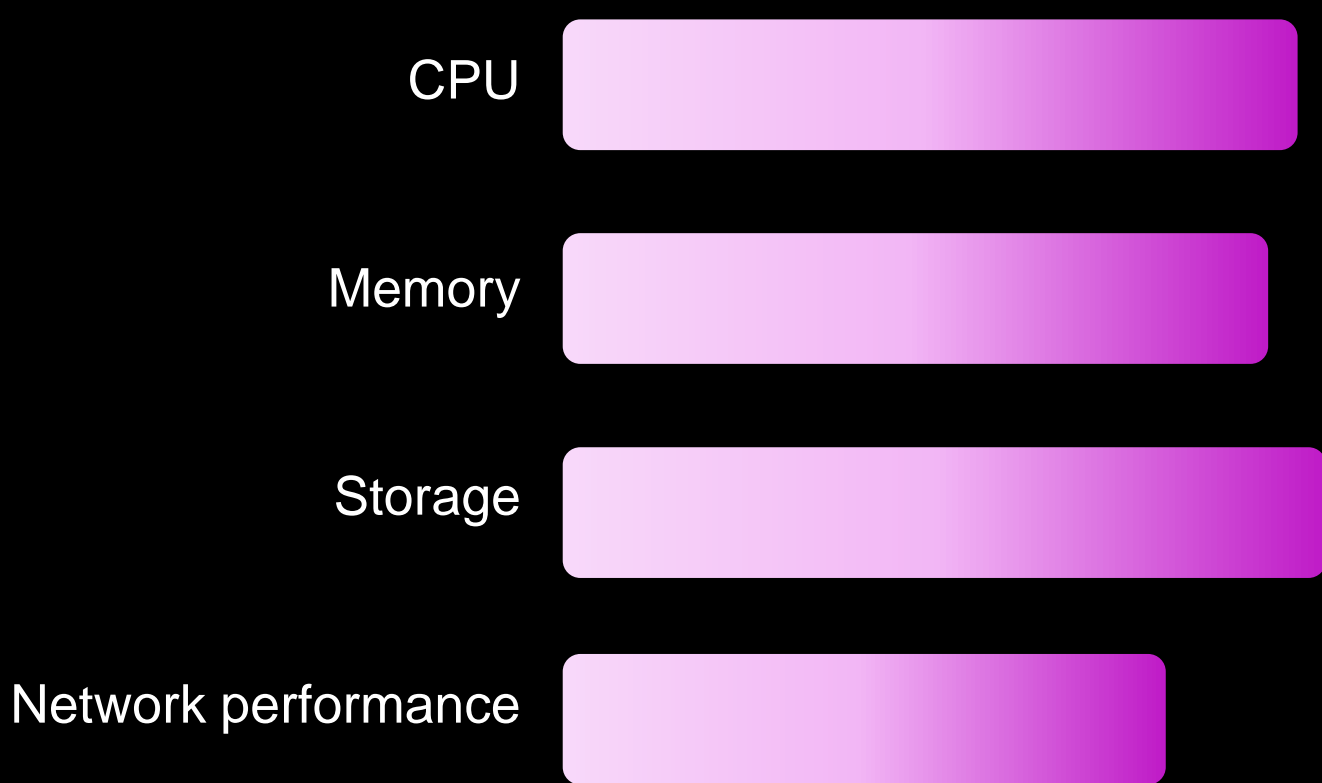
자세한 내용 : <https://docs.aws.amazon.com/ko-kr/AWSEC2/latest/UserGuide/ec2-instance-lifecycle.html>

나의 워크로드에 적합한 **EC2** 인스턴스 선택하기

**“Instance type
selection is a process,
but don’t overthink it.”**



Amazon EC2 인스턴스 표기법



Amazon EC2 인스턴스 추가 기능 표시

프로세서

g

AWS Graviton
프로세서

a

AMD EPYC
프로세서

i

Intel Xeon
프로세서

스토리지 및 네트워크

d

빠른 로컬
NVMe 스토리지

n

최대 100 Gbps
네트워크 성능

성능

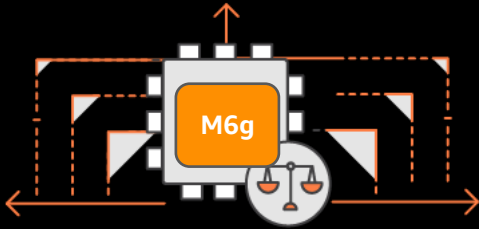
e

Enhanced performance

인스턴스는 하나 이상의 추가 기능을 가질 수 있습니다.

예) P3dn : 인스턴스 스토어가 NVMe SSD로 되어있고 네트워크 Bandwidth가 100Gbps인 인스턴스

Amazon EC2 범용 인스턴스



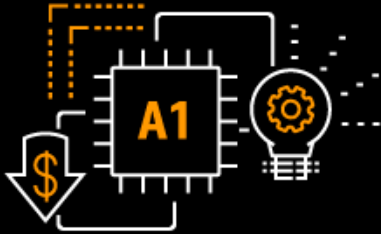
M6g 인스턴스

- ✓ 4:1의 메모리:vCPU 비율로 균형있는 컴퓨터, 메모리, 네트워크 지원
- ✓ 최신의 프로세서(M6g, M6I, M6a) 지원



T4g 인스턴스

- ✓ 지속적인 성능이 필요하지 않은 워크로드에 대한 버스트(Burst) 성능을 제공
- ✓ T4g, T3, T3a, T2 제공

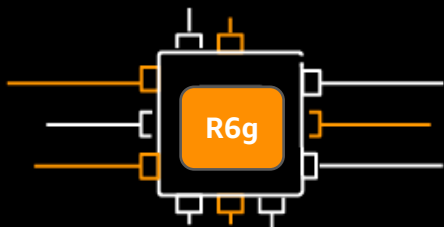


A1 인스턴스

- ✓ 적절한 메모리에 여러 코어로 확장하고 ARM 명령어를 실행할 수 있는 워크로드

Amazon EC2 메모리 집약적 인스턴스

인메모리 캐쉬, 고성능 데이터베이스, 빅데이터 분석 워크로드



R6g 인스턴스

- ✓ 메모리에서 큰 데이터 세트를 처리하기 위한 워크로드
- ✓ 최신의 프로세서(R6g, R6i, R5, R5a) 지원



X2 인스턴스

- ✓ 16:1 및 32:1의 메모리:vCPU 비율로 인메모리 워크로드
- ✓ 다양한 추가 기능을 제공

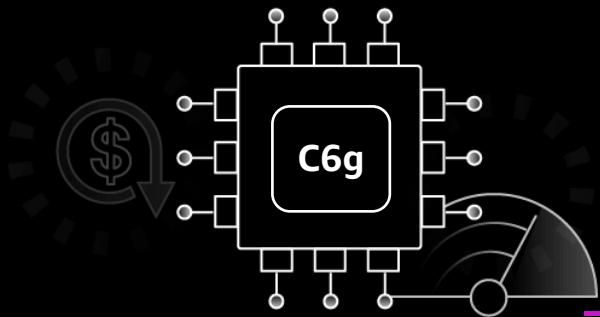


대용량 메모리 인스턴스

- ✓ 6~24TB의 대용량 인메모리 워크로드
- ✓ SAP HANA 실행 인증
- ✓ 베어 메탈 또는 가상 메모리 기반 선택 구성

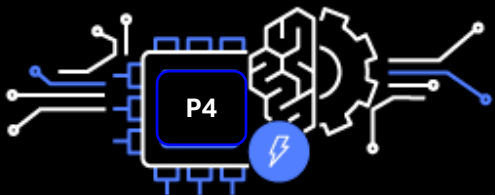
Amazon EC2 컴퓨팅과 가속화 인스턴스

분산 컴퓨팅, HPC, 멀티플레이어 게임, 비디오 인코딩, AI/ML, 3D 모델링 및 렌더링, 가상 그래픽, 가상현실/증강현실



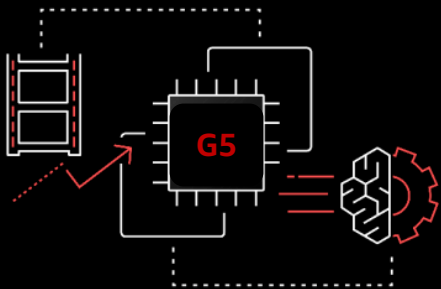
C6g 인스턴스

- ✓ 저렴한 가격에 vCPU당 높은 성능 제공
- ✓ 2:1의 메모리 : vCPU 비율
- ✓ 최신 Graviton3 프로세서 장착의 C7g 출시



P4 인스턴스

- ✓ 딥러닝 학습, HPC 시뮬레이션, 재무 컴퓨팅 및 배치 렌더링을 포함한 워크로드를 위한 GPU 컴퓨팅 인스턴스
- ✓ NVIDIA A100 Tensor Core GPU

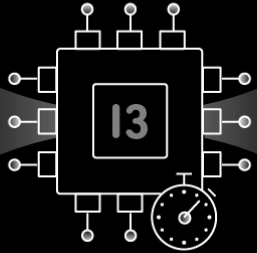


G5 인스턴스

- ✓ 3D 렌더링, 비디오 인코딩 및 AR/VR과 같은 워크로드를 위해 설계된 GPU 그래픽 인스턴스
- ✓ AMD EPYC Processors, NVIDIA A10G Tensor Core GPU

Amazon EC2 스토리지 집약적 인스턴스

높은 IO 트랜잭션, 빅데이터, 데이터웨어하우스, HDFS, MAPREDUCE, 로그분석



I3 인스턴스

- ✓ 트랜잭션이 많은 워크로드, 대기 시간이 짧은 워크로드에 최적화된 인스턴스
- ✓ Graviton2 기반의 Im4gn, Is4gen 타입 제공



D3 인스턴스

- ✓ 스토리지당 최저 비용
- ✓ 높은 순차적 디스크 처리량 지원 (580MiB/s ~ 4,600MiB/s)

자세한 내용 : <https://aws.amazon.com/ko/ec2/instance-types/>

EC2 인스턴스 타입에 대한 지속적인 최적화

EC2 인스턴스 최적화



비용효율적으로
EC2자원을
사용하고 있는가



사용하고 있는
EC2자원이 워크로드
요구사항에 적합한가

AWS Compute Optimizer

비용 절감 및 워크로드 성능 향상을 위한 AWS 리소스 권장

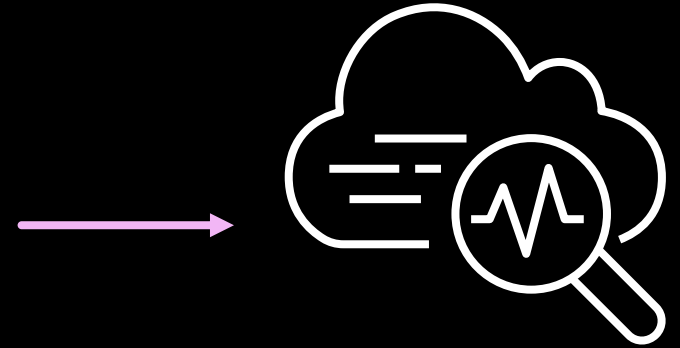
AWS Compute Optimizer는 자원의 사용률 기록 지표에 대한 기계 학습을 통하여 현재 워크로드에 대한 최적의 AWS 리소스(EC2, EBS & Lambda)를 권장합니다.



AWS Compute Optimizer

AWS 리소스를 확인하고 비용 및 성능 향상을 위한 권장 사항 제공

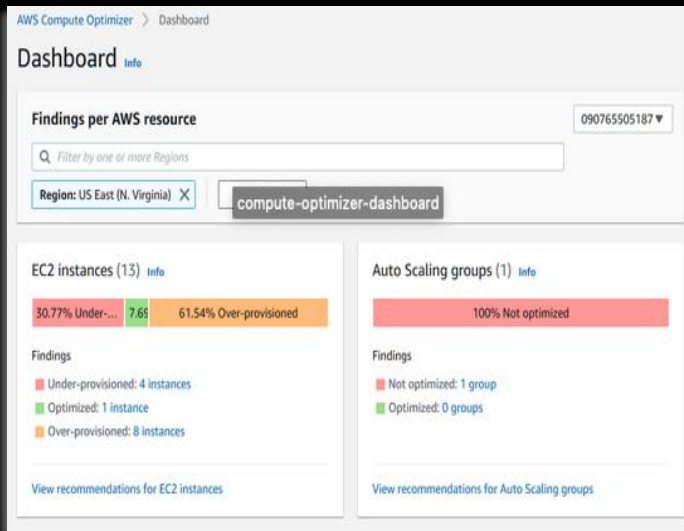
- ✓ AWS 리소스 구성 데이터와 Cloudwatch 지표 기반
- ✓ 최적의 AWS 리소스를 권장하는 머신러닝 기반 서비스
- ✓ 최적의 EC2 인스턴스 및 Amazon EC2 Auto Scaling 그룹 구성 권장
- ✓ 권장 인스턴스 유형에 따른 예상 성능 시나리오



권장 사항은 Amazon S3로 내보낼 수 있으며 AWS Cost Explorer 및 AWS Systems Manager와 통합됨

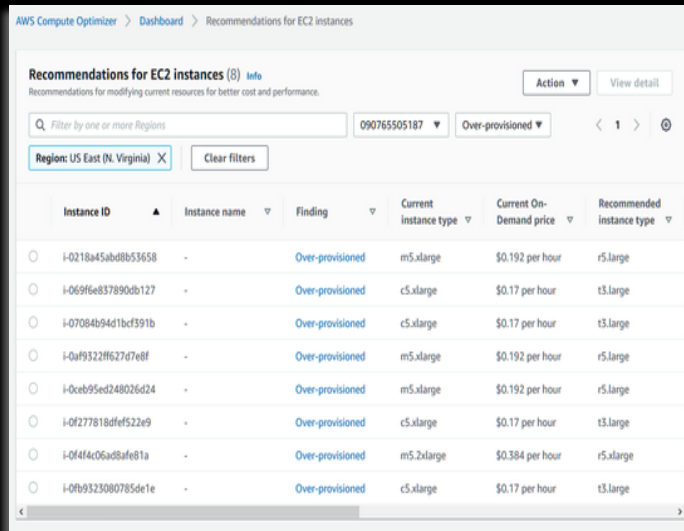
AWS Compute Optimizer 진행 절차

1. 권장 사항 자동 생성



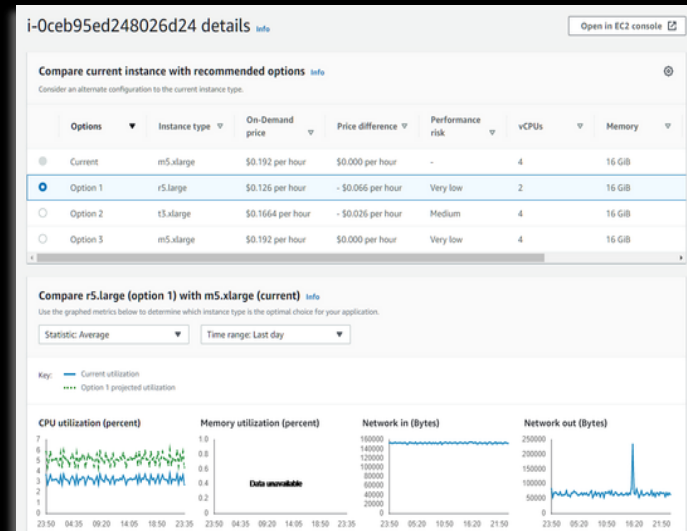
AWS 인프라를 검사하고 권장 사항을 생성합니다. 권장 사항을 제공하는데 최대 12시간이 소요될 수 있습니다.

2. 권장 사항 검토



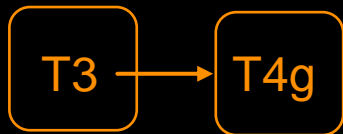
EC2 인스턴스, Auto Scaling 그룹, EBS 볼륨 또는 Lambda함수에 대한 권장 사항을 검토합니다.

3. 예상 시나리오 확인

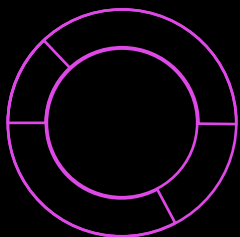


권장되는 인스턴스 유형에 대한 3가지 옵션을 확인하고 성능 및 비용을 비교 합니다.

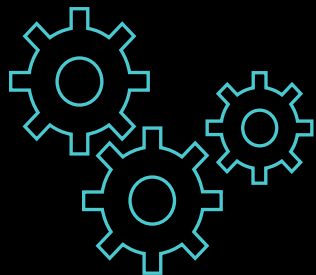
AWS Compute Optimizer **NEW**



X86 기반 인스턴스에서 **AWS Graviton 기반 인스턴스**로
변경 시 비용 및 성능 영향도 제공



Resource efficiency Metrics을 통하여 가장 큰 비용 및
성능 개선 방안을 빠르고 쉽게 파악



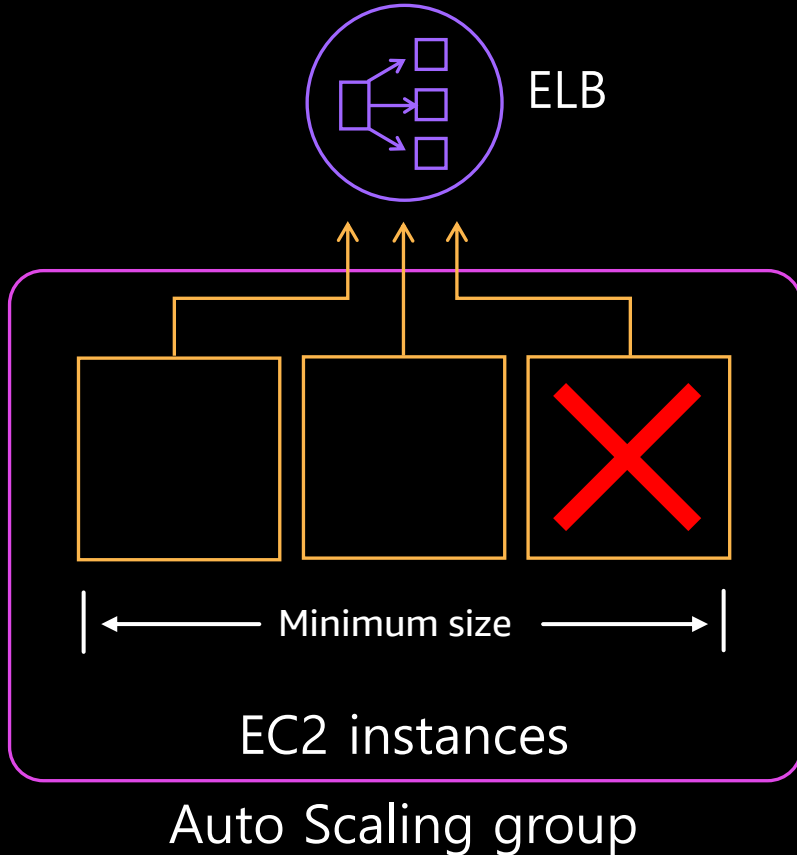
Enhanced Infrastructure Metrics로 최대 3개월간의 Cloudwatch
Metrics를 통하여 EC2 및 Auto Scaling group 권장 확인

스팟 인스턴스를 통한 유연한 확장

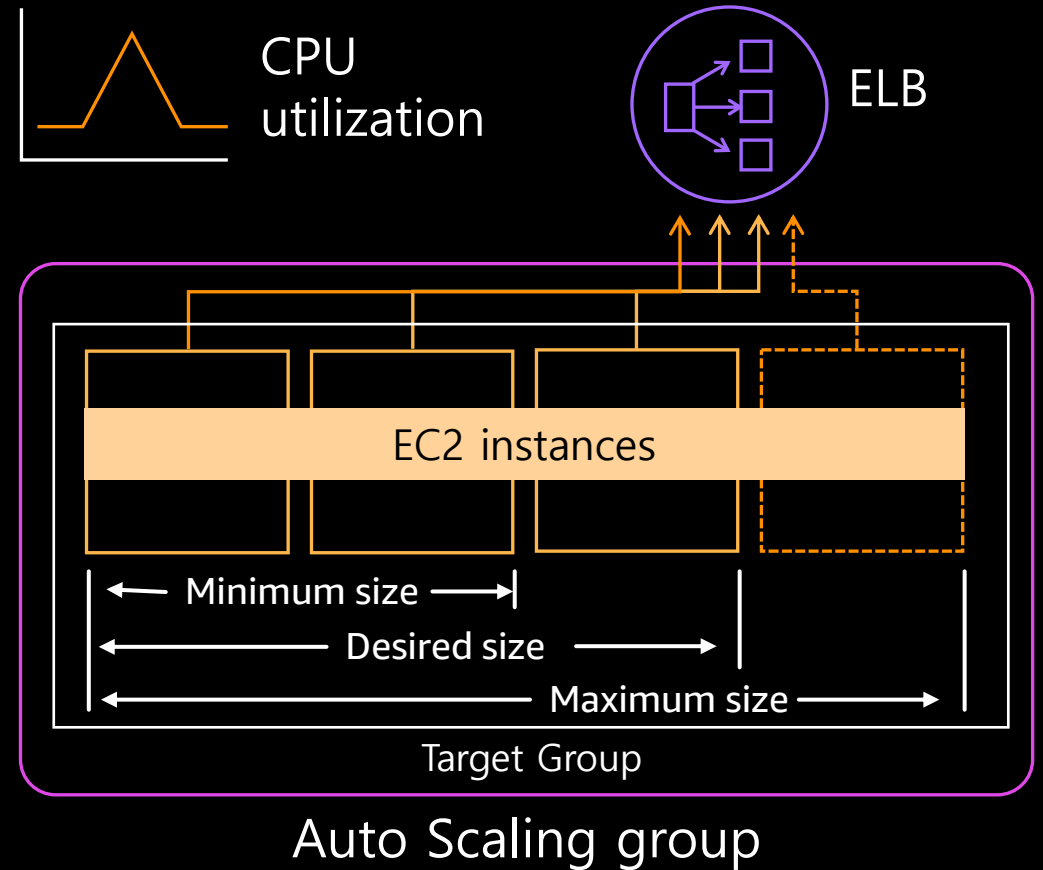
Auto Scaling Group

변화하는 수요에 동적으로 대응하고 비용을 최적화

Fleet management
비정상 인스턴스 교체

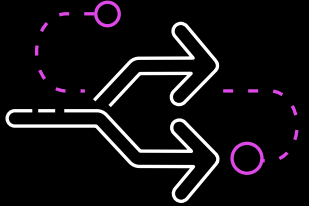


Dynamic scaling
수요에 맞게 확장

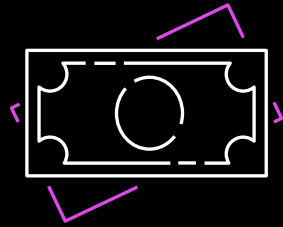


Amazon EC2 Spot Instance

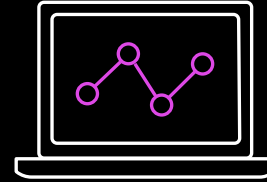
내결함성 및 상태 비 저장 워크로드를 위한 EC2 인스턴스



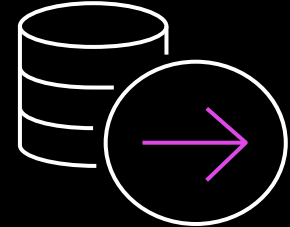
사용하지 않는 EC2
용량을 사용
온디맨드와 동일



온디맨드 요금 대비
최대 90%할인
스팟 가격

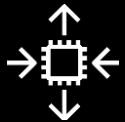


다른 인스턴스 타입, 사이즈,
위치를 선택하여 auto
scaling group이나 spot
fleet와 같이 사용

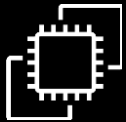


스팟 인스턴스는 중단될
수 있음
중단 2분에 경고 알람

EC2 Spot 인스턴스는 많은 AWS service에서 사용될 수 있습니다.



Amazon EC2
Auto Scaling



EC2 Fleet



Amazon ECS



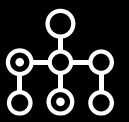
Amazon EKS



Amazon
EMR



AWS
CloudFormation



AWS Batch

스팟 배치 점수(Spot placement score)

스팟 요청의 성공 가능성을 사전에 확인

- 스팟 워크로드를 사용할 수 있는 최적의 위치 추천
- 스팟 용량, 인스턴스 유형 요구 사항 등을 기반으로 권장 사항 지정
- 과거 및 현재의 스팟 사용 추세, 요청 시간 요소를 기반으로 10~1까지의 범위에서 리전 및 가용 영역에 대한 점수 표시
- 가용 영역 권장 사항에 따라 인스턴스를 요청 시, EC2 Fleet 이나 Auto scaling 요청에 적용해야 함

Spot placement score

Spot placement score helps you to select optimal Regions or Availability Zones to run workloads that can use multiple instance types.

Target capacity and instance type requirements Edit

Target capacity	Instance types	요구 사항 정의
2000 instances	r5.8xlarge, c5.9xlarge, m5.8xlarge	

Placement scores Calculate placement scores

We calculate placement scores based on factors such as the number and composition of the instance types, the target capacity, the Spot usage trends, and the time of the request. Scores serve as a guideline, and no score guarantees that your Spot request will be fully or partially fulfilled. A score of 10 means that your Spot capacity request is highly likely to succeed in that Region or Availability Zone at the time of the request. A score of 1 means that your Spot capacity request is not likely to succeed.

Regions to evaluate

☐ Provide placement scores per Availability Zone

Clear filters

Region	Placement score
US East (N. Virginia)	8
us-east-1	
Europe (Ireland)	5
eu-west-1	
US West (Oregon)	3
us-west-2	

배치 점수

Auto Scaling Group Multi instance

AutoScalingGroupName: my-asg

CapacityRebalance: true

MixedInstancesPolicy:

LaunchTemplate:

LaunchTemplateSpecification:

LaunchTemplateName: my-launch-template

Version: \$Default

Overrides:

- InstanceType: c5.large
- InstanceType: c5a.large
- InstanceType: m5.large
- InstanceType: m5a.large
- InstanceType: c4.large
- InstanceType: m4.large
- InstanceType: c3.large
- InstanceType: m3.large

InstancesDistribution:

OnDemandBaseCapacity: 1

OnDemandPercentageAboveBaseCapacity: 0

SpotAllocationStrategy: # "capacity-optimized-prioritized", "capacity-optimized", "lowest-price"

MinSize: 1

MaxSize: 5

DesiredCapacity: 3

VPCZoneIdentifier: subnet-5ea0c127, subnet-6194ea3b, subnet-c934b782

Auto scaling 시에 운영 중단이 높은 스팟 인스턴스에 대하여 사전에 재조정 신호를 보내어 애플리케이션의 가용성 향상을 제공함.

여러 인스턴스 유형에 걸쳐 온디맨드 와 스팟 인스턴스 조합을 프로비저닝함.

인스턴스 확장에 대한 전략. Capacity-optimized-prioritized 또는 capacity-optimized 설정 시, 사용 가능한 스팟 용량을 기반으로 최적으로 선택된 스팟 풀을 사용하여 auto scaling을 적용함



워크로드에 따른 스팟 인스턴스 사용 예

세이빙 플랜 :

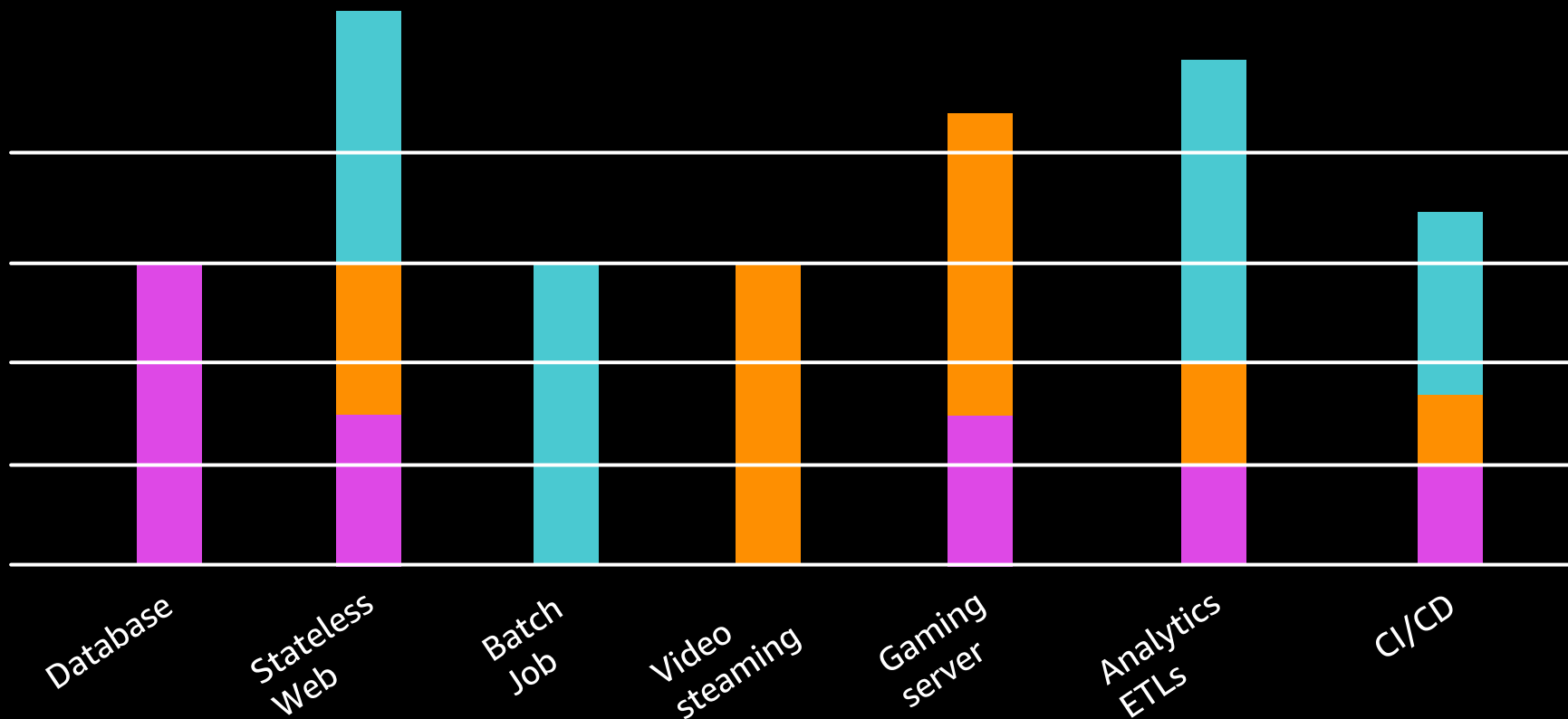
장기적으로 일정한 트래픽이 유지되는 워크로드

온디맨드 :

Scaling out이나 예측 불가능한 워크로드

스팟 인스턴스 :

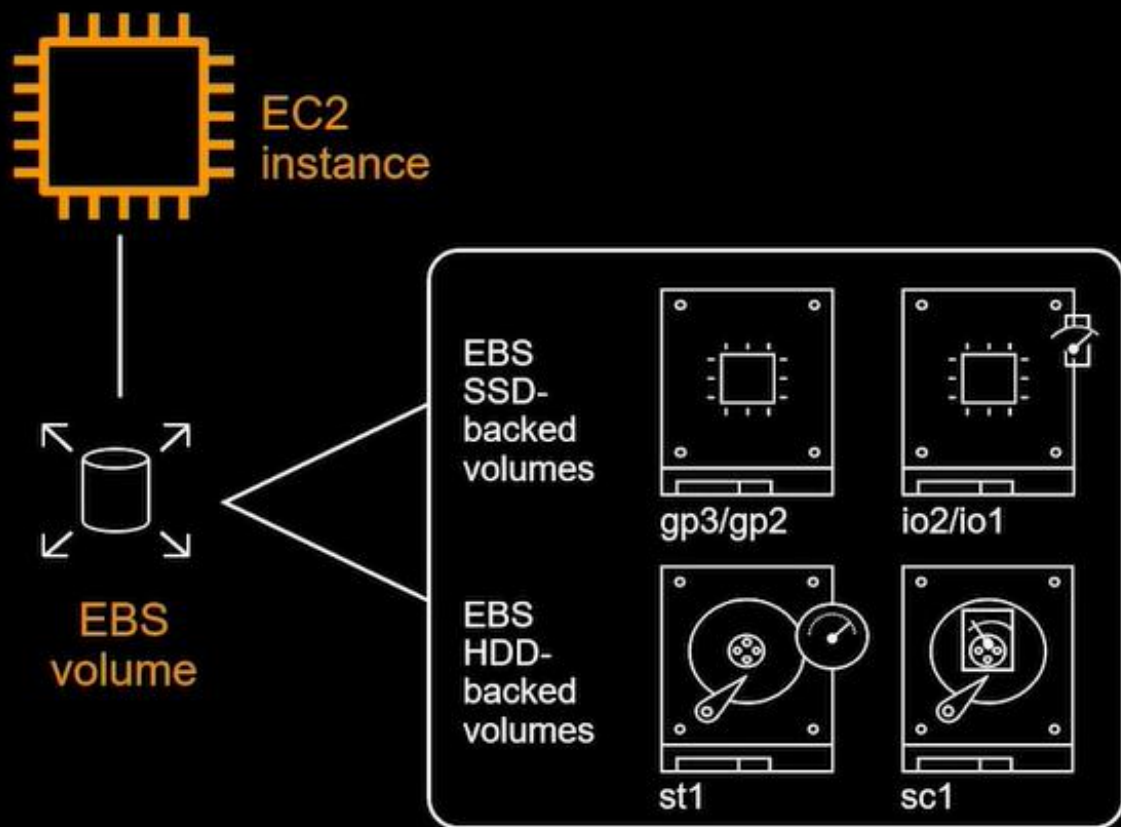
유연하고 내결합성을 갖춘 워크로드



Amazon EC2의 데이터 보호 관리

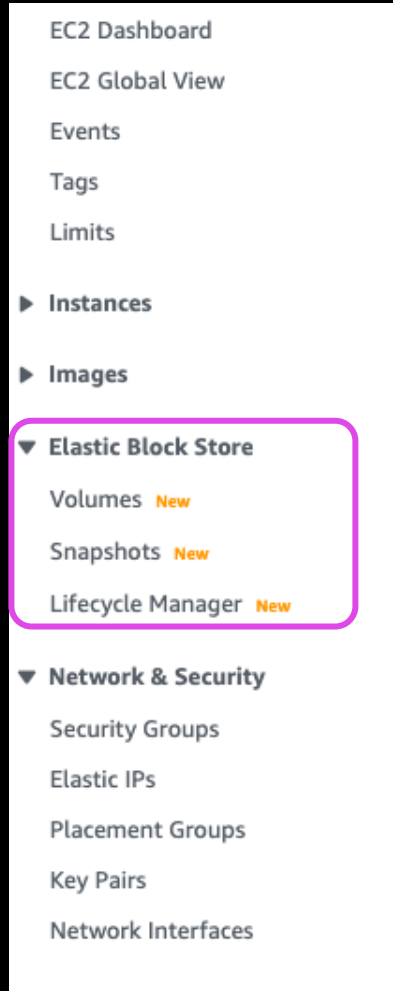


Amazon EBS

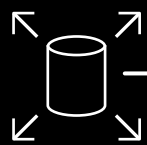
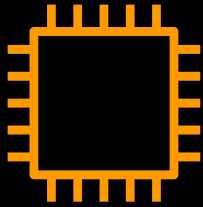


- 블록 스토리지
- API를 이용하여 생성, 연결, 수정
- 워크로드에 따라 스토리지 및 컴퓨팅 선택
- 하나의 EBS 볼륨은 하나의 인스턴스에만 연결
- 마그네틱 및 SSD 기반 볼륨 유형 선택
- 스냅 샷 지원 : 특정 시점 백업

Amazon EBS 스냅샷 관리



EC2 instance



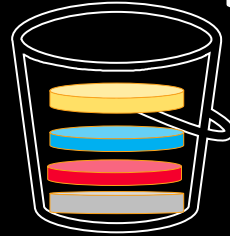
EBS
volume



Policy



EBS
스냅샷



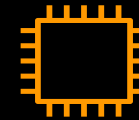
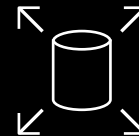
Amazon S3

Amazon Data Lifecycle Manager

- 스냅샷을 통하여 정기적으로 데이터 백업
- 정기적으로 표준화된 EBS-backed AMI 생성
- 정책 기반의 스냅샷 보존 및 삭제 자동화
- 단일/ 다중 볼륨 스냅샷
- 계정간 또는 리전간에 스냅샷 공유 및 복제 기능 제공
- Cloudwatch와 Cloudtrail를 통한 모니터링 및 알람



Amazon CloudWatch



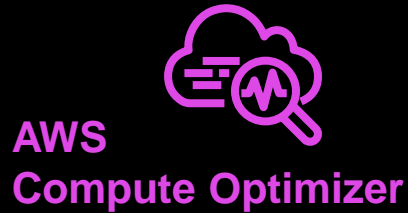
Take away

#1

M5d.xlarge

인스턴스 표기법으로
워크로드에 적합한 인스턴스
선택

#2



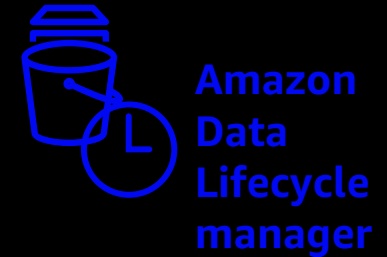
초기 선택한 인스턴스에 대한
모니터링을 통하여 최적의
인스턴스 구성

#3



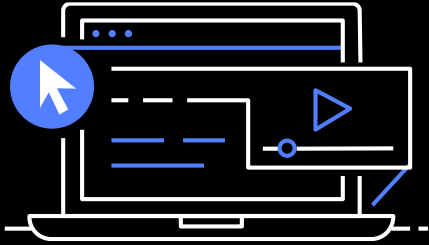
유연한 확장을 위한
Mixed 인스턴스 auto
scaling 구성

#4



EBS 스냅샷을 통한 데이터
보호 및 관리 정책 자동화 수립

AWS 디지털 교육



Flexibility to learn
your way

550개 이상의
무료 디지털 교육 및
심층적 강의실 교육을 통해
클라우드 기술 역량을
업그레이드 하세요!

추천 과정 (한글 자막)

- [AWS Cloud Practitioner Essentials](#)
AWS 클라우드 기초에 대해 학습하고, 기초 자격증인 AWS Certified Cloud Practitioner 시험을 준비할 수 있습니다.
- [Amazon DynamoDB for Serverless Architectures](#)
Amazon DynamoDB의 전반적인 소개와 Amazon DynamoDB가 서버리스 아키텍처 구축에 어떻게 활용되는지 알아봅니다.
- [AWS Security Fundamentals](#)
AWS 액세스 제어 및 관리, 거버넌스, 로깅, 그리고 암호화 방법을 포함한 기본적인 클라우드 컴퓨팅 및 AWS 보안 개념에 대해 알아봅니다.
- [AWS Database Offerings](#)
다양한 데이터베이스 기술 및 아키텍처에 대한 기본 개요와 AWS 데이터베이스 서비스를 소개합니다.

AWS Builders Online Series에 참석해주셔서 대단히 감사합니다.

저희가 준비한 내용, 어떻게 보셨나요?
더 나은 세미나를 위하여 **설문을 꼭 작성해 주시기 바랍니다.**



aws-korea-marketing@amazon.com



twitter.com/AWSKorea



facebook.com/amazonwebservices.ko



youtube.com/user/AWSKorea



linkedin.com/company/amazon-web-services



twitch.tv/aws

Thank you!

이우진

AWS Solutions architect
wjlee@amazon.com

