

대학생을 위한 AWS 강의(13)

EC2 인스턴스 유형

서진호

제 13 강 목표

- EC2 인스턴스 유형
- 적합한 아마존 EC2 인스턴스 선택
- 컴퓨팅 최적화 인스턴스
- 차세대 범용 인스턴스
- 인공지능 전용 인스턴스

1. 적합한 아마존 EC2 인스턴스 선택

- EC2 인스턴스 유형은 다양한 사용 사례 및 워크로드에 최적화 되었고 여러 크기로 제공한다. 이를 통해 워크로드 요구 사항에 따라 리소스를 최적으로 조정할 수 있다.
- AWS에서는 EC2 인스턴스에 대해 인텔 제온 프로세서를 활용하여 고객에게 뛰어난 성능과 가치를 제공한다.
- 인스턴스를 선택할 때 다음을 고려하라. 예) 코어 수, 메모리 크기, 스토리 크기 및 유형, 네트워크 성능, I/O 요구 사항, CPU 기술
- 빨리 처리하고 쉬기(HUGI) - 컴퓨팅 인스턴스가 클수록 시간과 비용을 절약할 수 있다. 즉, 더 짧은 시간 동안 시간당 비용이 높은 인스턴스를 실행하는 것이 더 경제적일 수 있다.

2. 인스턴스 유형

패밀리	설명	사용 사례 예시
t2, m3, m4, m5	범용 균형 잡힌 성능	웹 사이트, 웹 애플리케이션, 개발, 코드 리포지토리, 마이크로 서비스, 비즈니스 앱
c3, c4, c5	컴퓨팅 최적화 뛰어난 CPU 성능	프런트 엔드 플릿, 웹 서버, 배치 처리, 분산 분석, 과학 및 엔지니어링 앱, 광고 제공, MMO 게임, 비디오 인코딩
g3, p2, p3	GPU 최적화 고성능 GPU	Amazon AppStream 2.0, 비디오 인코딩, 기계 학습, 고성능 데이터베이스, 과학
r5, r5a, r4, x1e, x1, 대용량 메모리 , z1d	메모리 최적화 대규모 RAM 공간	인 메모리 데이터베이스, 데이터 마이닝
i3, i3en, d2, h1	스토리지 최적화 높은 I/O, 고밀도	NAS, 데이터 웨어하우징, NoSQL

aws.amazon.com/ec2/instance-types

3. 컴퓨팅 최적화 인스턴스 – C5n

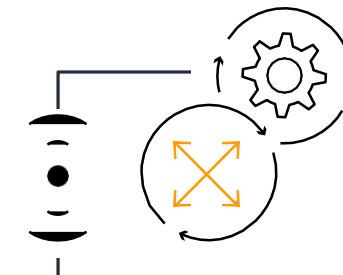
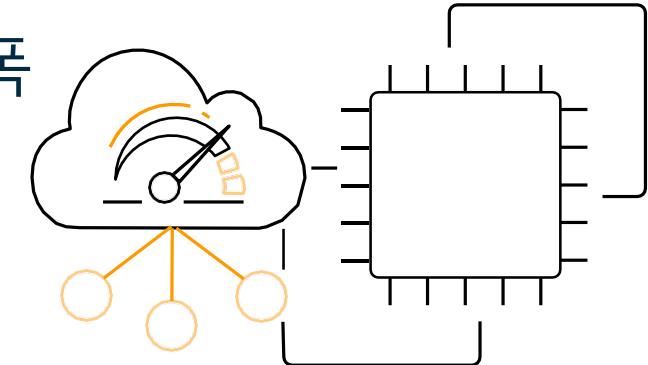
- 인텔 제온 확장 가능 프로세서 탑재
- 최대 인스턴스 크기에서 100 Gbps의 네트워크 대역폭
- 작은 인스턴스 크기에서 25 Gbps의 최대 대역폭
- C5 인스턴스 대비 33% 증가된 메모리 공간 제공



보다 빠른 분석과
빅데이터 워크로드



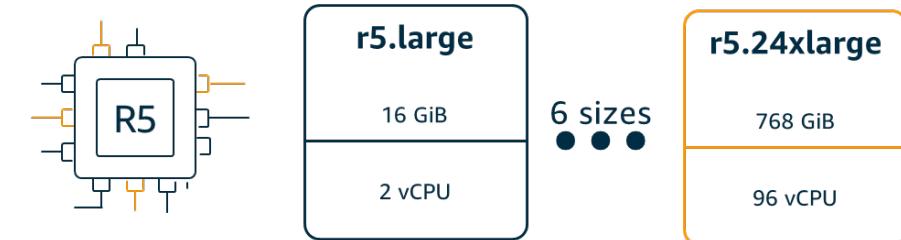
네트워크 바인딩 워크로드에
대한 보다 낮은 비용



AWS의 모든 보안, 확장성 및
안정성을 계속해서 활용 가능

4. 메모리 최적화 인스턴스 – R5

- 2.5 GHz 인텔 제온 확장형 프로세서 (Skylake)
- 메모리 대 vCPU 비율이 8:1인 메모리 최적화 인스턴스
- 최대 25 Gbps 네트워크 대역폭
- R5d 인스턴스는 최대 3.6 TB의 로컬 NVMe SSD 제공



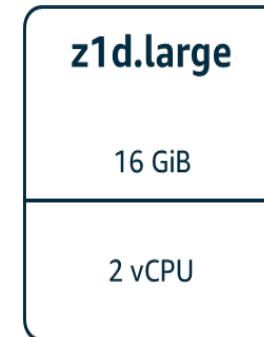
R5a: Now available with
AMD EPYC 7000 processor

NEW

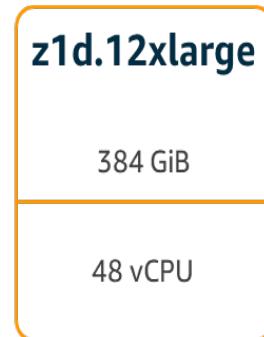
R5.metal Bare Metal instances
coming soon on Intel Xeon
Scalable processors

5. 특별한 워크로드를 위한 고성능 인스턴스 - z1d

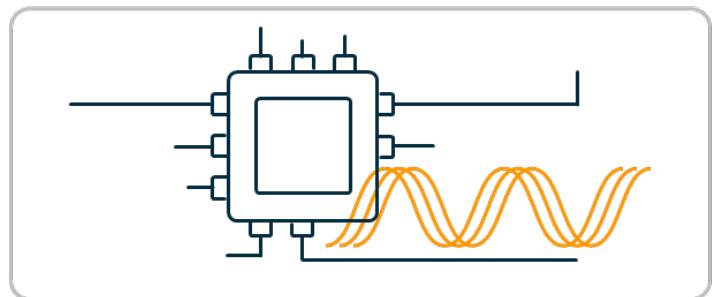
- 4 GHz의 일관된 올코어 터보로 동작하는 커스텀 인텔 제온 확장형 프로세서의 고성능 인스턴스
- 8:1 메모리 대 vCPU 비율
- 최대 25 Gbp의 네트워크 대역폭과 최대 1.8 TB의 로컬 NVMe 스토리지



6 sizes
• • •



Electronic Design Automation



Relational databases



Gaming



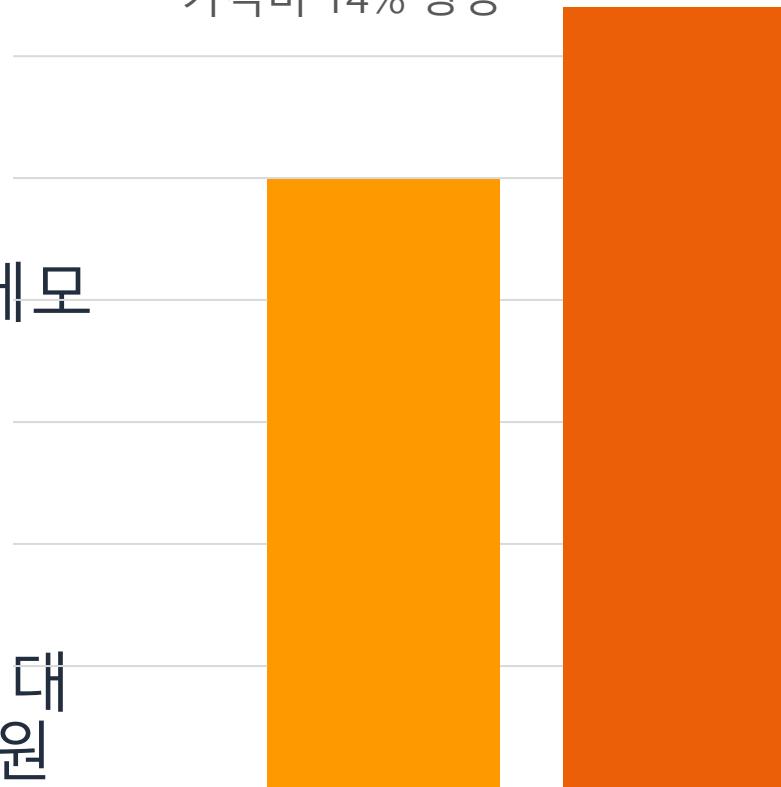
z1d.metal Bare Metal instances coming soon

6. 차세대 범용 인스턴스 – M5

- 2.5 GHz 인텔 제온 확장형 프로세스(Skylake) 탑재
- 더 큰 크기의 새로운 인스턴스-m5.24xlarge
최대 vCPU 96개 및 메모리 384GiB (4:1 메모리 대 vCPU 비율)
- 더 작은 크기의 인스턴스보다 향상된 네트워크 및 EBS 성능
- 벡터 및 부동 소수점 워크로드에 대해 최대 2배의 성능을 제공하는 인텔 AVX-512 지원

M5에서 성능

가격비 14% 향상



■ M4 ■ M5

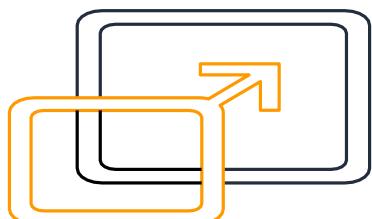
7. 차세대 범용 인스턴스 – T3

- 컴퓨팅, 메모리, 네트워크 리소스의 균형
- 필요한 경우 언제든지 기준선 수준의 CPU 성능에 CPU 사용량을 순간 확장할 수 있는 기능 제공
- 시간당 \$0.0052로 가장 낮은 비용의 인스턴스로, 인텔 제온 확장형 프로세서를 사용하여 T2 대비 최대 30% 개선된 가격 대비 성능 제공

t3.nano
0.5 GiB
2 vCPU
Base perf 5%

7 sizes

t3.2xlarge
32 GiB
8 vCPU
Base perf 40%



T3는 기본 성능을 초과하는 무제한 버스팅이 24시간 평균, vCPU-시간당 \$0.05 요금



8. 인공지능/머신러닝 특정 인스턴스 유형

패밀리	설명
T 인스턴스 타입	일관되게 높은 수준의 CPU가 필요하지 않지만 필요할 때 빠른 CPU를 사용하여 장점을 얻을 수 있는 범용, 버스트 가능한 성능 인스턴스
M 인스턴스 타입	컴퓨트, 메모리, 네트워크 대역폭이 균형을 이루는 범용 인스턴스
C 인스턴스 타입	높은 CPU 요구 사항들이 있는 컴퓨팅 바인딩 워크로드에 이상적인 컴퓨팅 최적화 인스턴스
R 인스턴스 타입	아파치 스파크와 같은 대용량 데이터셋을 메모리에 저장하여 분산 처리의 장점을 얻는 워크로드에 최적화된 메모리 최적화 인스턴스
P, G, 인퍼런시아 및 트레이니엄 인스턴스 타입	하드웨어 가속기 또는 GPU와 같은 보조 프로세서가 있는 고성능 컴퓨트 인스턴스들 또는 추론을 위한 AWS 인퍼런시아(inferencia) 및 훈련 워크로드들을 위한 AWS 트레이니엄(trainium)과 같은 아마존 맞춤형 하드웨어
일래스틱 인퍼런시아 액셀레이터	배치 변환 및 추론과 같은 특정 워크로드들에 추가 컴퓨팅 성능이 필요할 때 다른 인스턴스 유형에서 사용하는 네트워크 불인 보조

실습 – EC2 인스턴스 타입

