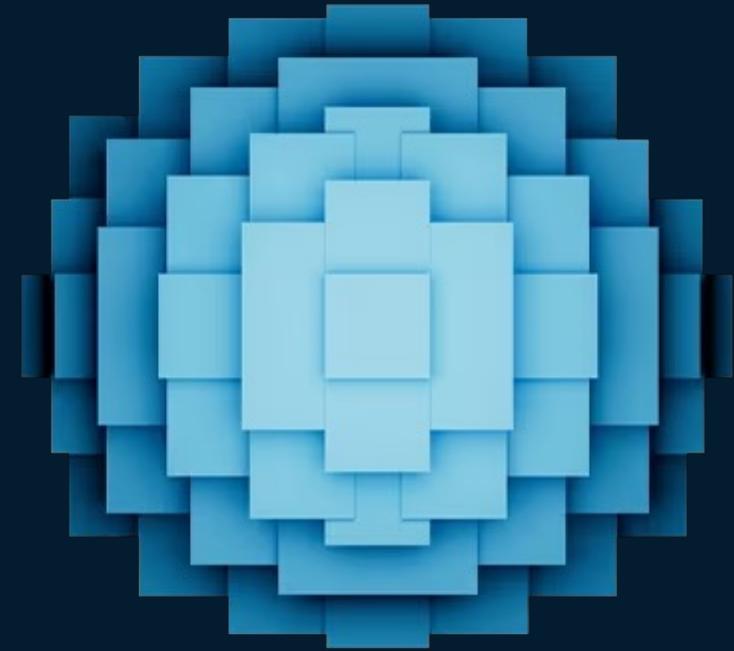


Less Cost More Cloud

NetApp CloudOps Solutions



 **NetApp**

We make the cloud work for you.

Hyunseok, Kang

NetApp Korea

클라우드의 높은 비용

모든 것은 클라우드로?... '비용 폭탄'의 역습

박신영 기자 ☆

입력 2022.07.11 15:26 수정 2022.07.11 15:26 지면 B1

혁신 위한 기본 인프라로 각광 받지만
최근 투자 대비 효과 있는지 논란 벌어져

스토리지 사용량 증가하면 비용 폭발적 증가
자체 서버 활용해 직접 데이터 관리하는
'온프레미스'로 전환하면 비용 절반 절감 가능

지난 10년간 업계의 화두는 'everything should go to the cloud(모든 것 클라우드로)'였습니다. 클라우드는 혁신을 위한 기본 인프라이며 정보 비용을 합리적으로 절감해 주는 업계의 총아로 찬양받았습니다. 팬데믹
인해 IT 지출은 전반적으로 감소했지만 클라우드에 대한 지출은 증가했
클라우드는 누구도 반론을 제기하지 않는 표준이 됐습니다.

환율 변동에 울고 웃는 기업들... 클라우드 시장에도 여파

이종현 2022.07.21 09:48:44

가+

가-

달러 강세가 지속되는 가운데 국내에 제품을 공급하고 있는 글로벌 기업의 고민이 깊어지고 있다. 환율 상승으로 국내에 들어오는 SW 및 하드웨어, 그리고 클라우드 서비스 가격이 상승하면서 기업의 IT투자에 영향을 미치고 있기 때문이다. <디지털데일리>는 3회에 걸쳐 달러 강세에 따른 국내 IT인프라 시장에 미치는 여파를 분석해본다. <편집자>

기업 > 컴퓨팅·IT서비스

“새는 돈 막아라”...클라우드 비용 절감 솔루션 도입 '급증'

김동진 기자

입력 2020.11.06 06:00

클라우드 비용 관리, 과요금 방지 등의 기능을 담은 '클라우드 비용 절감 솔루션'을 도입하는 기업이 급증하고 있다. 코로나 팬데믹으로 인한 디지털 전환 확산으로 기반 인프라인 클라우드를 도입하는 기업들이 늘면서, 운영 비용 절감에 관한 수요도 함께 증가한 것으로 풀이된다.

클라우드 운영 조직의 주요 과제

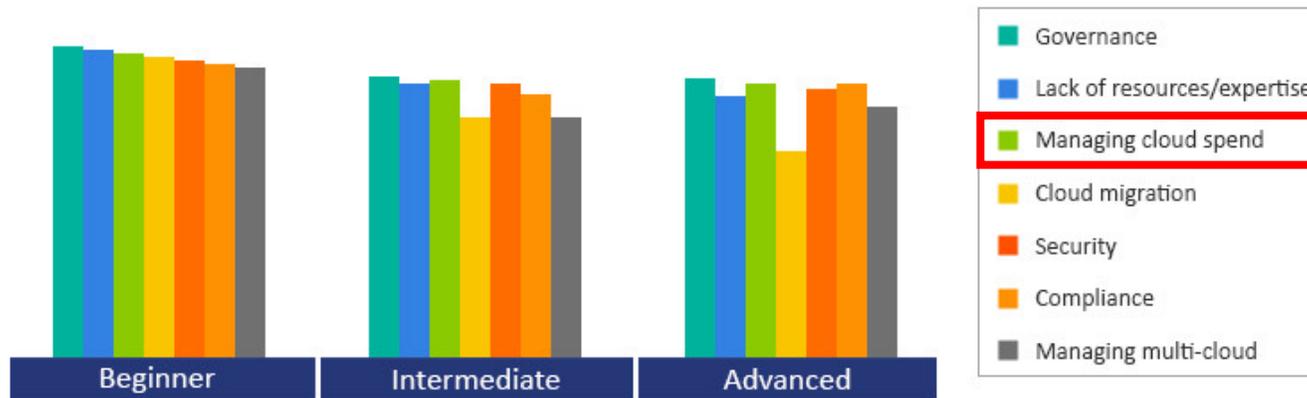
Top cloud initiatives for 2022 across all organizations

Source : Flexera 2022 State of the Cloud Report



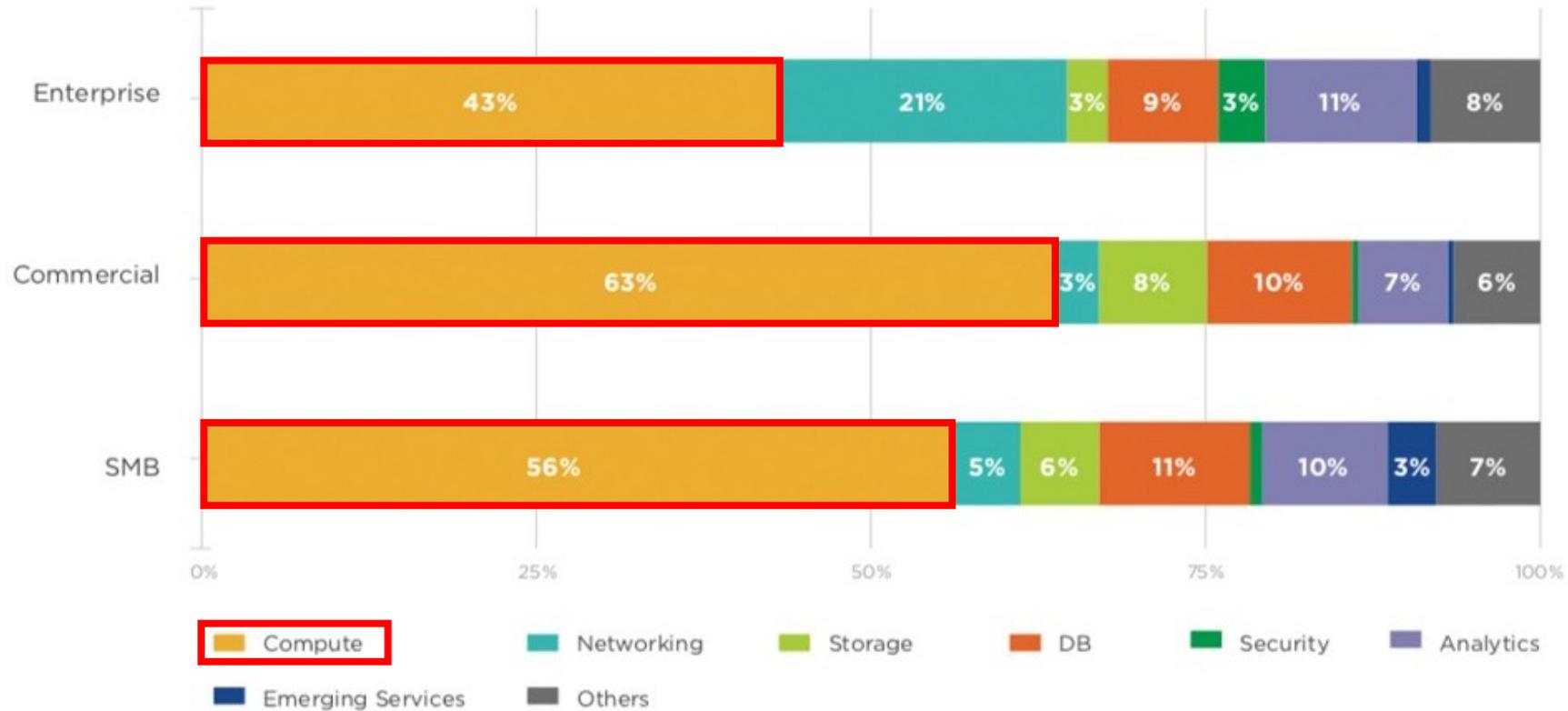
Cloud Challenges by Maturity

% of Respondents



주요 클라우드 서비스 항목

Cloud Spend by Service Category



클라우드 컴퓨팅 구매 옵션



On-demand

- **가장 높은 컴퓨팅 비용**
- 필요 시, 리소스 즉시 생성
- 높은 SLA



Reserved

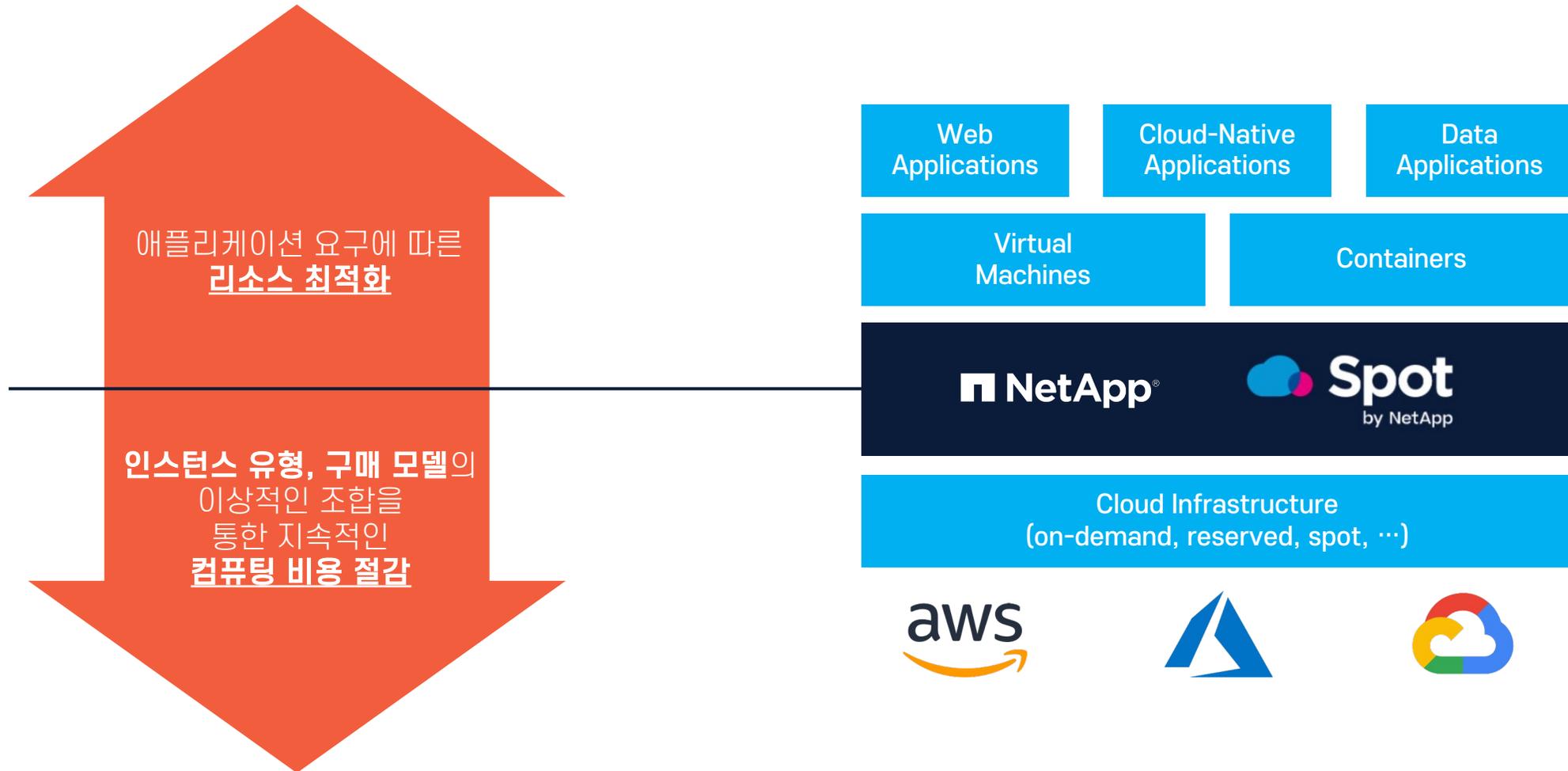
- On-demand 대비 낮은 컴퓨팅 비용 (최대 72%)
- **낮은 유연성**



Spot

- 가장 저렴한 컴퓨팅 비용
- **No SLA**

지속적인 클라우드 환경 최적화 : Spot by NetApp



독보적인 머신러닝과 AI 기반 예측 분석

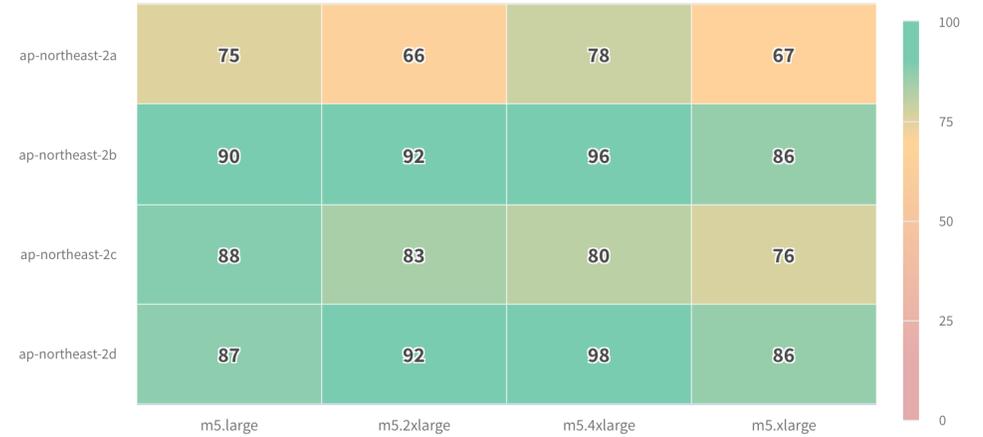
SPOT TYPES

	General Purpose (4)				Compute Optimized			Accelerated Computing			Memory Optimized		
	m4	m5	m5a	m5ad	m5d	m5zn	m6g	m6i	t2	t3	t3a	t4g	
medium							<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
large	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>										
xlarge	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>										
2xlarge	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>										
3xlarge						<input type="checkbox"/>							
4xlarge	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>					

워크로드 별 다양한 인스턴스 유형 지원



SPOT MARKET SCORING



선택한 인스턴스에 대한 Spot Market Scoring

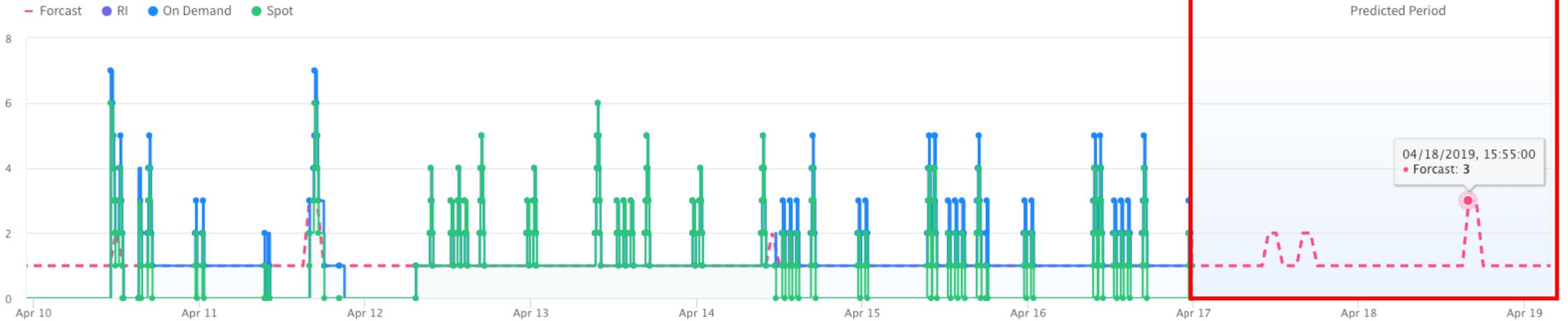
최적의 Spot 인스턴스를 선택

- 리전(Region)에 따라
- 가용 영역(AZ)에 따라
- 인스턴스 유형에 따라
- 인스턴스 크기에 따라

예측 기반 오토 스케일링

Activity & Prediction

6 hours 1 day **7 days**



Target Scaling Policy 1: Target Scaling

Policy Name * Policy Type *

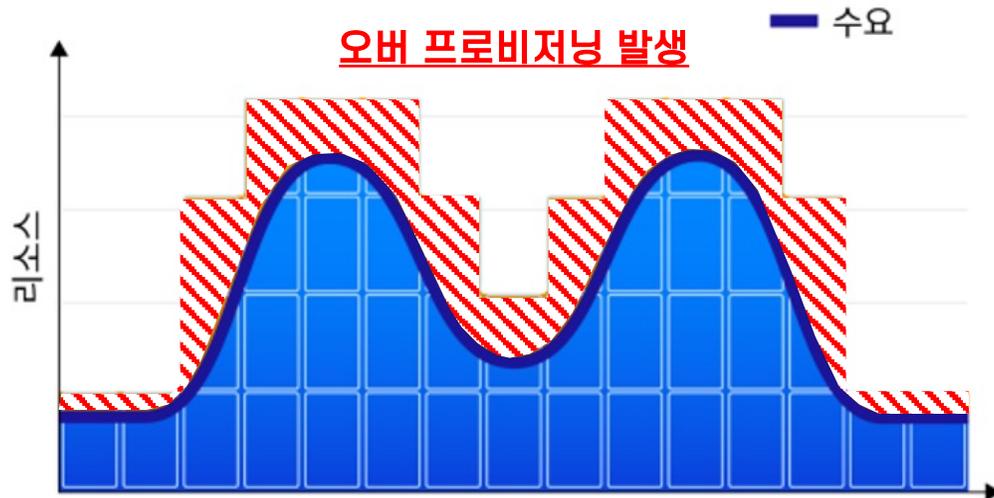
Use predictive autoscaling Predict and scale
Predict only

Metric Name * Target Value * Cooldown seconds

애플리케이션 요구 및 사용자 기반의 오토스케일링

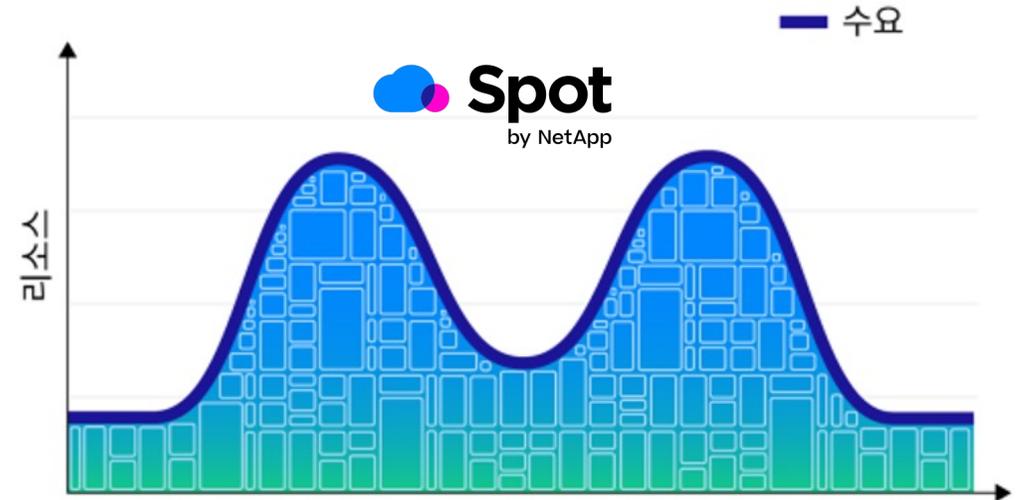
Without Spot by NetApp

- 동일한 유형/세대 별 리소스 지원
- 비효율적인 리소스 사용



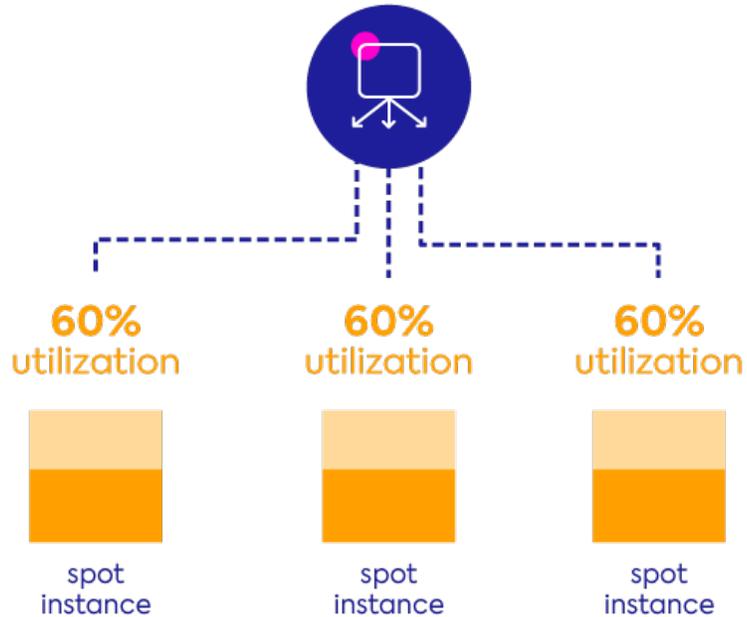
With Spot by NetApp

- 이기종 컴퓨팅 리소스 지원
- 효율적인 리소스 사용

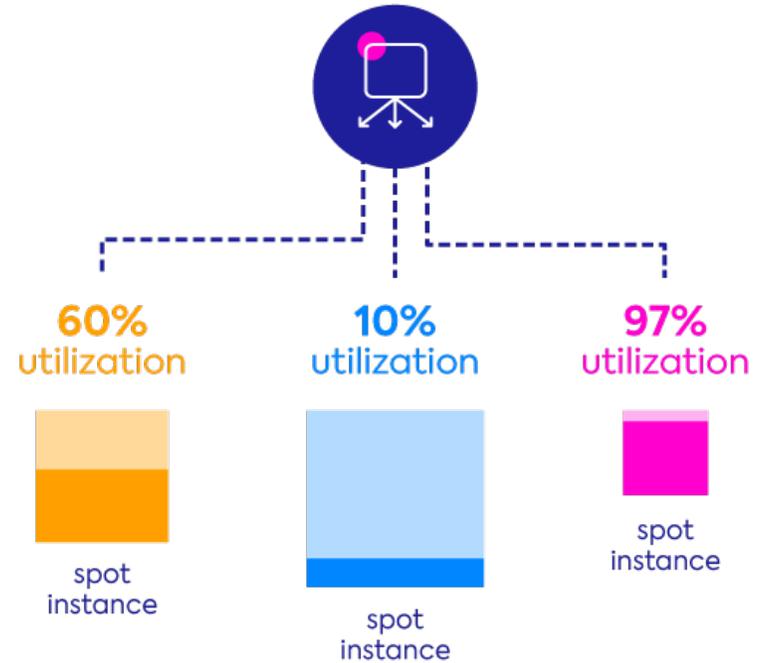


이기종 인스턴스 간 부하 분산을 지원 : ITF(Intelligent Traffic Flow)

× Limited spot selection and availability



× Unbalanced resource utilization



스테이트풀 워크로드를 위한 다양한 기능 지원



Spot Management Instance

기존 애플리케이션의 상태를 유지

루트 볼륨, 데이터 볼륨

네트워크 인터페이스 및 보안 그룹

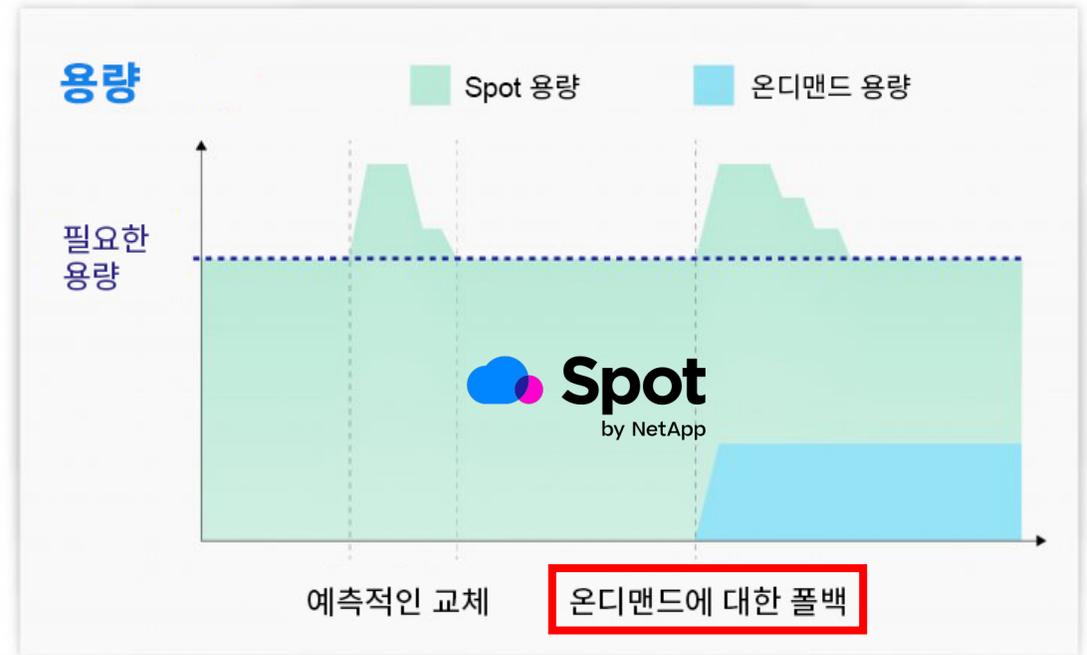
태그 및 키 페어

서비스 가용성을 보장하기 위한 추가 가용성 옵션 제공

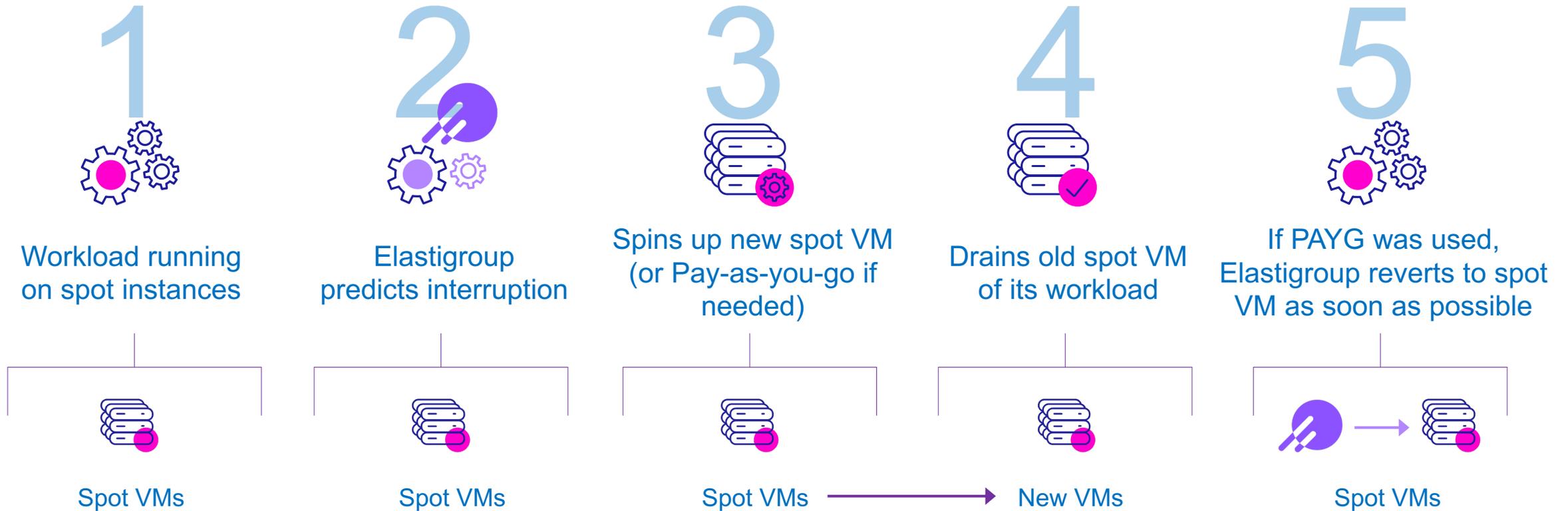
Without Spot by NetApp



With Spot by NetApp



Spot Elastigroup 동작 방식



Elastigroup 적용 워크로드



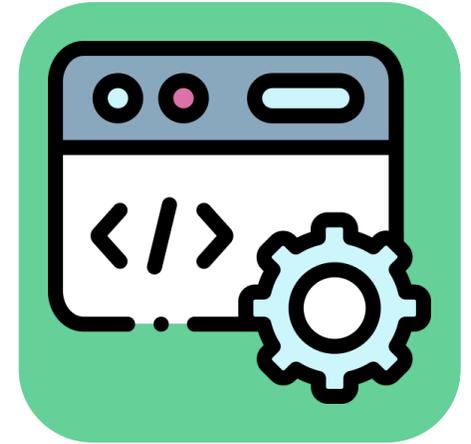
배치 업무



Fault-tolerance
아키텍처



빅 데이터



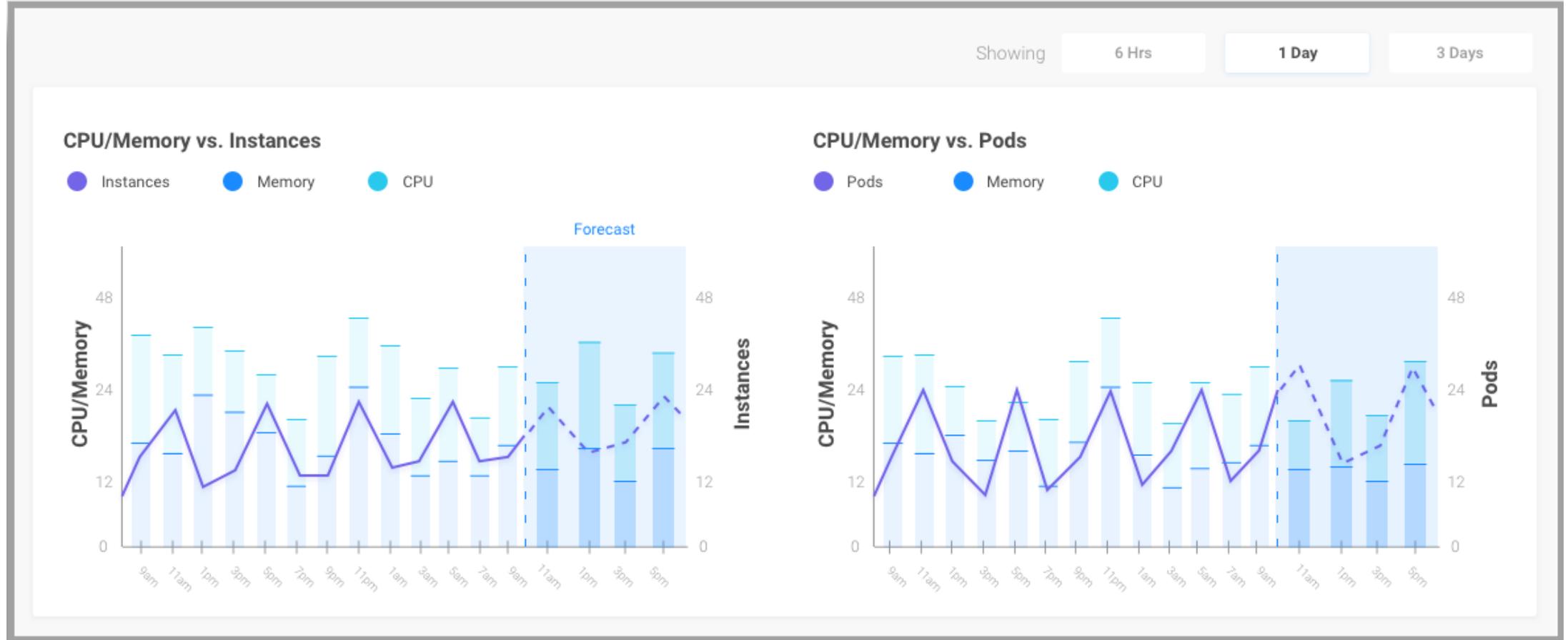
개발 환경

컨테이너 워크로드 최적화 : Spot Ocean



- 컨테이너 기반 auto-scaling
- 노드 리소스 활용 극대화
- Right-sizing을 위한 권장사항 제공/자동화

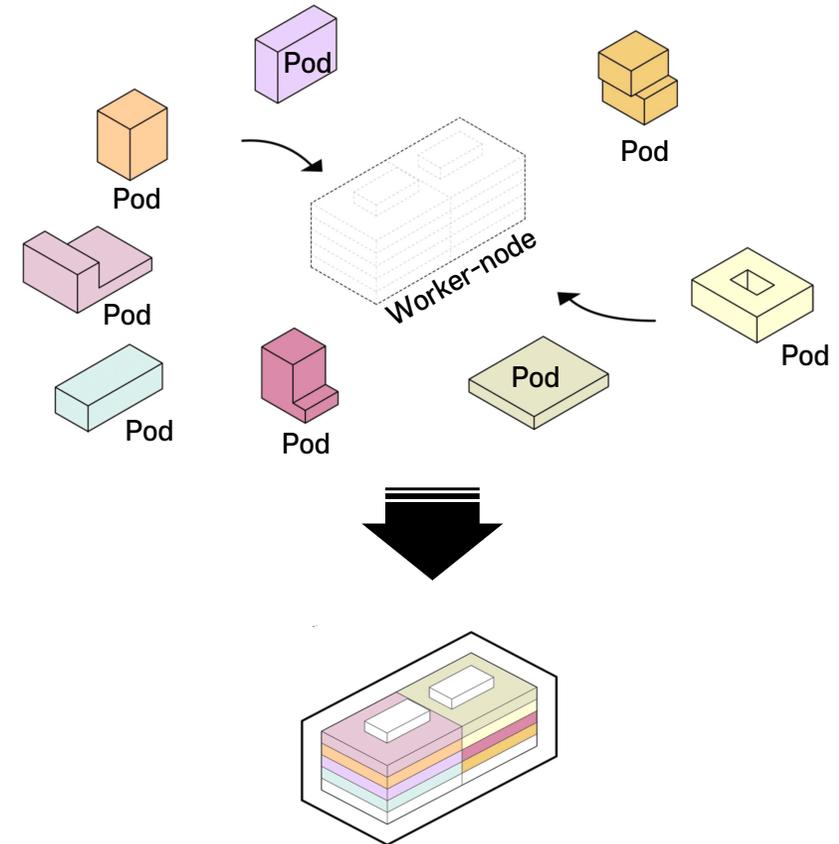
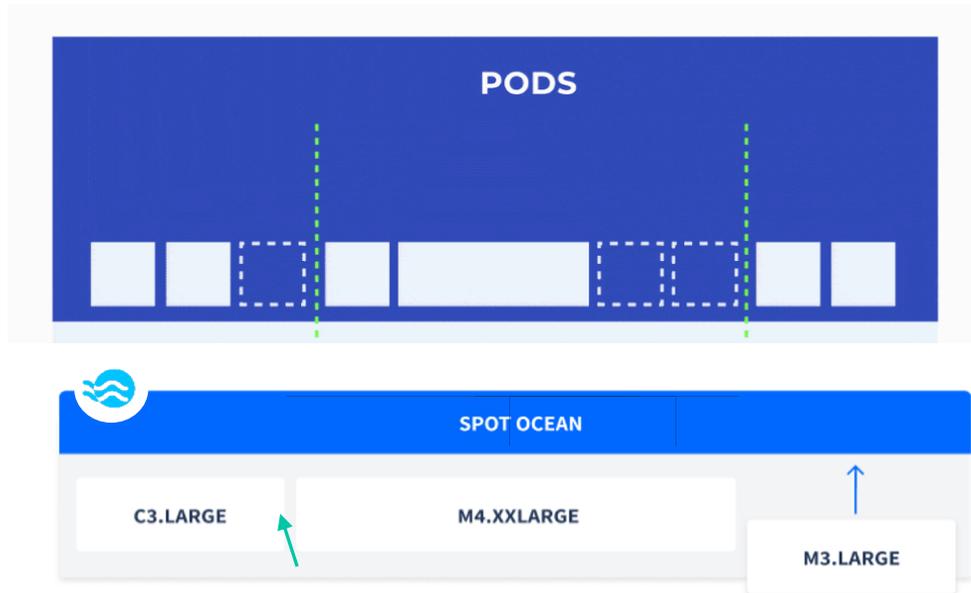
컨테이너 기반의 Autoscaling



컨테이너 사용량 기반의 최적화 : Right-sizing

Resource Name	Resource Type	Namespace	Desired Pod Count	Age	Impact ↓
meister-apiciqa-meister	deployment	apiciqa	1	7 days	HIGH
Container	CPU Request	Memory Request	Resizing Recommendations		
geoupdate	Current 0.2 vCPU ↓ Decrease to 0.002 vCPU	Current 500 MiB ↓ Decrease to 20 MiB	requests: cpu: 0.002 memory: 20Mi		
wm-server	Current 0.2 vCPU ↓ Decrease to 0.001 vCPU	Current 500 MiB ↓ Decrease to 372 MiB	requests: cpu: 0.001 memory: 372Mi		
meister-glogrotate	Current 0.2 vCPU 👍 Keep 0.2 vCPU	Current 500 MiB ↓ Decrease to 9 MiB	requests: cpu: 0.2 memory: 9Mi		
meister	Current 15 vCPU ↓ Decrease to 0.535 vCPU	Current 25600 MiB ↓ Decrease to 3046 MiB	requests: cpu: 0.535 memory: 3046Mi		

노드 리소스의 사용을 극대화 : Bin-packing



Worker-node 리소스 사용을 극대화를 위한 Bin-packing

빠른 컨테이너 배포를 위한 기능 제공 : Headroom

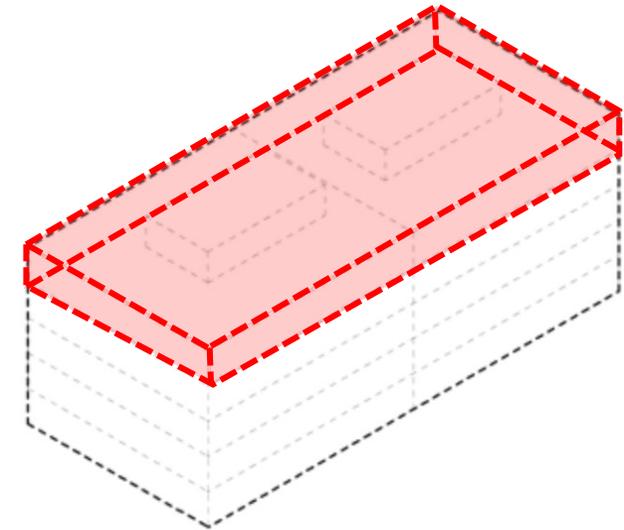
Headroom Configuration ⓘ

Automatically (according to the most frequent Pod size in the cluster)
Configure the percentage Automatic Headroom will take up, out of total cluster capacity

0% 50%

Manually

Reserve units of CPU (mCPU), Memory (MiB) and GPU (Cores)

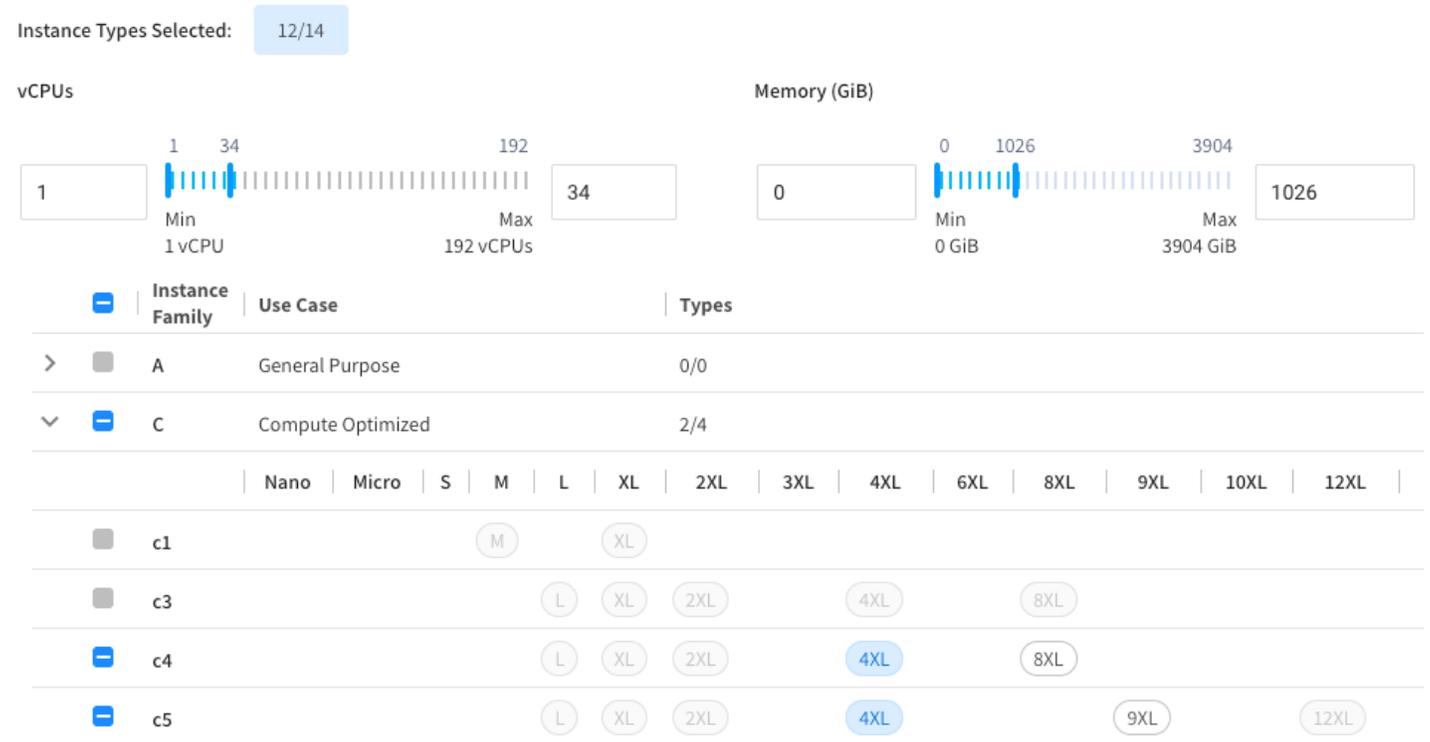


Total Cluster capacity

동일 클러스터 내 워크로드 분리 : VNG(Virtual Node Group)

특정 애플리케이션 요구사항 충족 :

- 워크로드 별 인스턴스 타입 지정
- 워크로드 별 용량 설정
- 워크로드 별 Spot/On-demand 혼용



스케줄 기반의 Scaling 기능 : Shutdown Hours

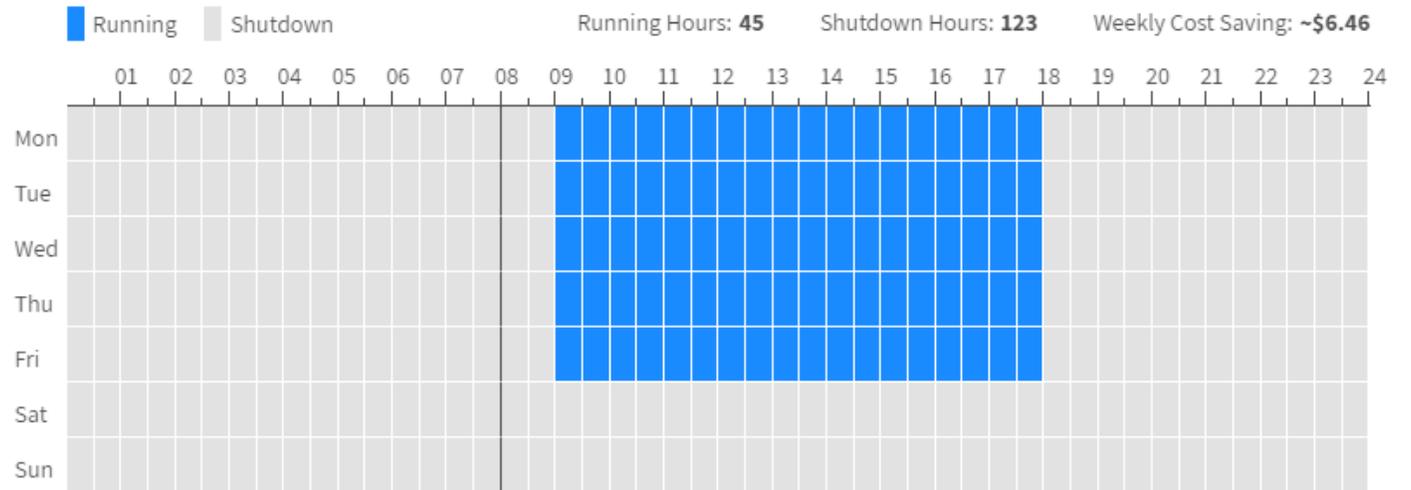
자원이 필요할 때만 손쉽게 사용

- 사용량 패턴에 따라 스케줄링
- 필요한 시간에 원하는 용량을 확장
- 개발 환경에서 업무시간에만 사용

Cluster Shutdown Hours ⓘ

Shutdown hours enabled

Times shown in local time zone (GMT +09:00)



Cancel Save

획기적인 라이선스 정책



	Free <i>No-cost way to try out Spot</i>	Pay-as-you-Go <i>Pay-as-you-save pricing, no annual commitment</i>	Subscription <i>Greater discounts, priority support, predictable bill</i>
Term	None	Month-to-month	Annual
Support Hours	9×5	24×7	24×7
Support Options	Email Only	Email & Live Chat	Email & Live Chat
Contracted SLA		✓	✓
24×7 Proactive NOC monitoring		✓	✓
Connected cloud accounts	1	Unlimited	Unlimited
Customer success package			✓
Training and remote workshops			✓
Onboarding consulting session			✓
Early access and preview			✓
Scale	Up to 20 virtual machines	Contact Spot - varies by scenario	Unlimited
Price	Free	% of savings realized	Negotiated based on usage and savings

Key Takeaway

Cloud 비용을 최대 90% 절감

On-demand 인스턴스와 동일한 SLA 제공

Spot instances의 최대 활용률

실행가능한 가시성을 제공하여 Cloud 비용 관리

인프라 운영 오버헤드 최대 85% 절감

인프라 프로비저닝 단순화

자동화 된 작업을 통해 지속적으로 비용 및 리소스 활용률 최적화

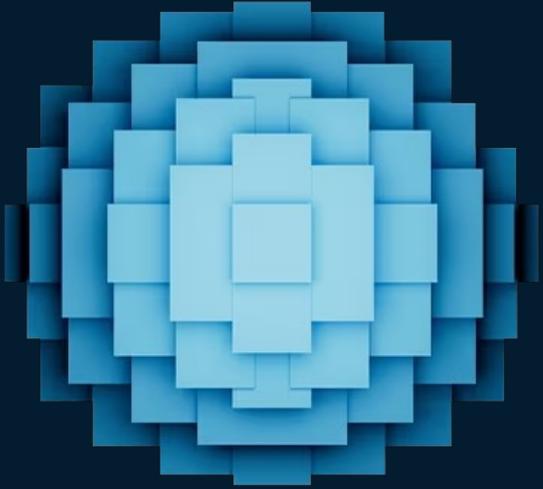
애플리케이션 요구사항에 따라 자동으로 확장

주요 Workload를 위한 인프라 단순화

Container 및 Kubernetes 인프라 관리의 복잡성 감소

Scale-out application을 위한 인프라 자동화





 NetApp®

Thank You.