



“서버를 늘릴 것인가? 더 잘 쓸 것인가?”

메모리 대란 시대의 데이터센터 비용절감 전략

임관수 전무

브로드컴 (VMware by Broadcom)



Agenda

1. Cloud Market Trend와 Enterprise IT의 현실
2. VMware의 방향성과 솔루션 제시
3. Business Value > 비용 혁신 고객사례
4. Summary





Global Cloud Market Trend

인용 "Private Cloud Outlook, 2026"



OVERVIEW

Private Cloud Outlook 2026

2년 차를 맞은 이번 연구는 2025년부터 2026년까지 엔터프라이즈 IT의 클라우드 전략이 어떻게 변화해왔는지, 그리고 AI가 현대 IT를 어떻게 재편하고 있는지를 살펴봅니다.

Insight Objectives

AI와 진화하는 운영모델이 미치는 영향 :

- 🔍 인프라 스트럭처 현대화
- 🔔 애플리케이션 현대화
- ☁️ 보안 현대화
- 👤 IT 조직, 플랫폼 엔지니어링팀 역량



Geographical Reach

1,800
Global
surveys

AMER:600

US, Brazil

EMEA:600

UK, France, Germany

APJ:600

India, Japan, Australia



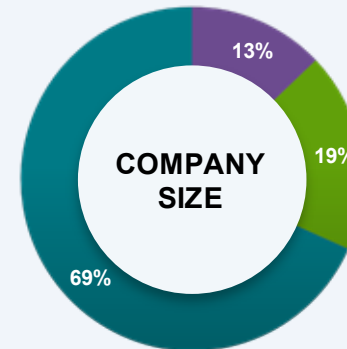
Methodology

Sample:

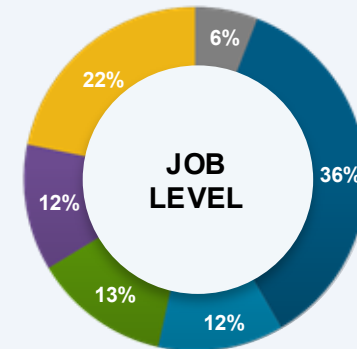
Web-based survey (20 mins)

Date Range:

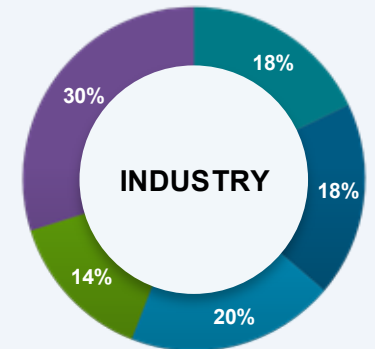
February 11 – March 13, 2026



- Small Enterprise (1,000 - 2,499)
- Medium Enterprise (2,500-4,999)
- Large Enterprise (5,000+)



- Chairman, President, Principal, Partner, Owner
- C-Level Executive
- Senior Vice President
- Vice President
- Department or Function Head
- Director



- Financial Services
- Public Sector
- Healthcare
- Life Sciences/Pharma
- Other



The AI Tipping Point

프라이빗 클라우드 도입을 이끄는 3가지 동력



The Inference Shift

56%

프라이빗 클라우드에서
추론(inference)/프로덕션 AI를
운영 중이거나 운영할 계획



AI 비용,
냉정한 정산의 시간

62%

Gen AI와 Agentic AI 비용에 대해
매우 또는 극도로 우려하는 IT 리더
비율



보안,
피할 수 없는 과제



AI 도입 시 가장 중요한 요구사항
#1 데이터 보호와 프라이버시
#2 보안과 통제



IT 인재,
부족의 시대

#1

IT 역량 격차의 핵심은 'AI 인프라 및
운영역량'
중앙 집중형 플랫폼 팀으로 운영되는
곳은 4곳 중 1곳(25%)뿐

AI는 비용·복잡성·통제라는 3C 문제유발 => 프라이빗 클라우드가 해법

Cost / Complexity / Control

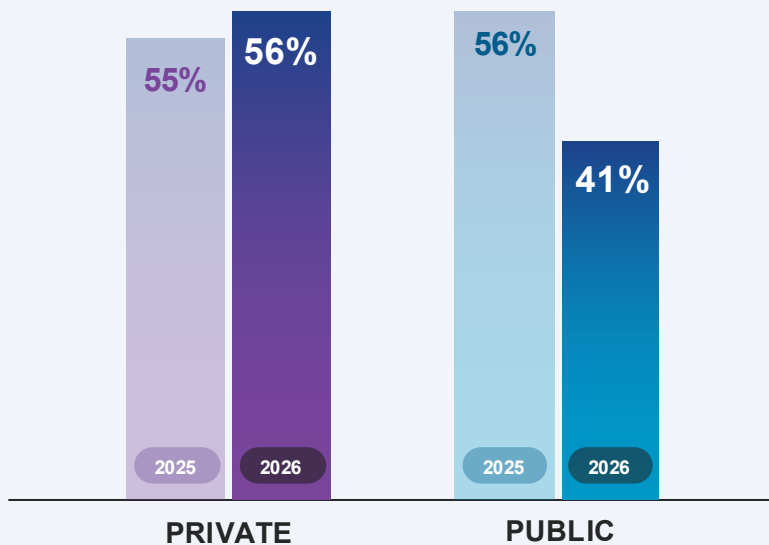




AI 추론 시대 : 이제는 ‘거버넌스’와 ‘비용 통제’가 필수

실제 운영(Production) 단계에서는 프라이빗 클라우드 선호

AI 추론 워크로드가
실제로 실행되는 위치



운영 단계 AI 추론 비중
퍼블릭 클라우드 사용 감소

> 15%

“AI 모델은 테스트 단계에서는 원활하게 작동했지만, 실제 배포 이후에는 상당한 수준의 최적화와 조정이 필요했다. 이 과정에서 비용이 많이 들었고, 투자 대비 효과(ROI)가 나타나기까지 비교적 오랜 시간이 걸렸다.”

Director of IT, Public Sector, United States

Prediction:

AI는 클라우드 경제학을 바꾸고 있다.

3C(비용·복잡성·통제)가 프라이빗 클라우드로의 전환을 가속한다.

Q: Where AI-based applications or workloads currently run n=1800

Data is based on a global survey conducted by Radius Tech in partnership with Broadcom. Radius Tech surveyed 1,800 senior IT decision-makers in February - March 2026 across 8 countries in North America, Europe, and Asia-Pacific.



프라이빗 클라우드 워크로드는 꾸준히 확대

IT 리더십의 투자 우선순위: 프라이빗 클라우드

클라우드 우선순위 1위:
프라이빗 클라우드 워크로드 확대



of IT leaders chose building new workloads in private cloud a top 3-year priority

53% in 2025

기업 클라우드 전략: 향후 3년 예산 계획

2025	2026	increase	
51%	72% ▲	+21 budget intent increase private cloud	 Private cloud
61%	71% ▲	+10 budget intent increase public cloud	 Public cloud

2x Faster Growth for Private Cloud

Q: Expected plans for cloud strategy over the next 3 years n=1800

Q: Expected cloud spend changes in the next 3 years n=1800





Enterprise IT : Reality & Challenges

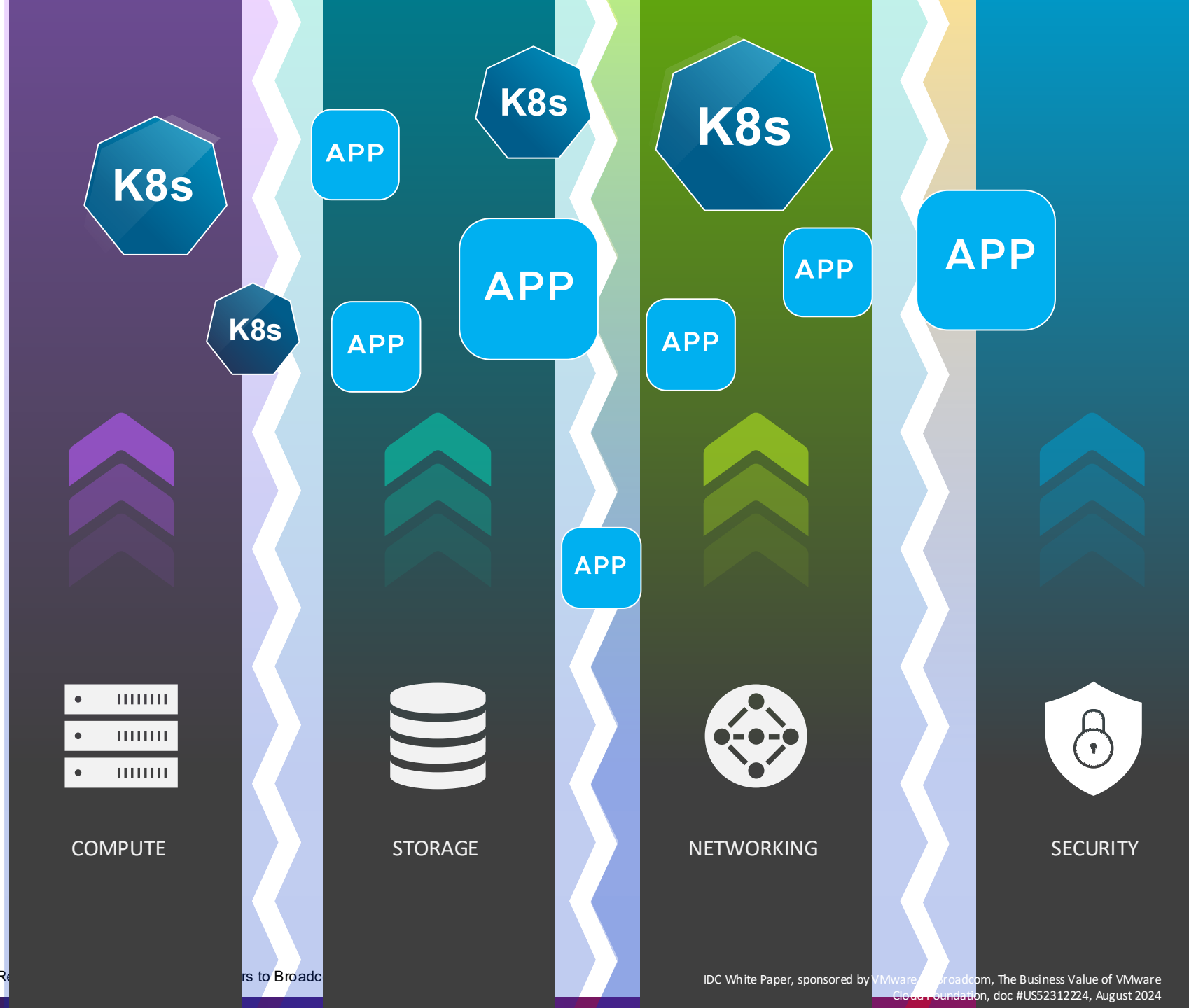


혁신을 낳추는 데이터센터 사일로(Silo)



34% higher
VM당 3년 비용(\$2,307)

복잡한 OS 및 App 생태계



플랫폼의 필연성

AI 준비는 하드웨어 도입이 아니라 데이터센터의 플랫폼화에서 시작



AI는 현대화의 최우선 과제

57%

Adding AI capabilities to modernize their applications



프라이빗 클라우드가
성숙도의 격차 해소

93%

Trust private cloud for security, reliability, and scalability platform



플랫폼엔지니어링과
IT인프라가 하나로 협력

76%

Collaborate across teams

현대화(Moernization)

플랫폼 성숙도(Platform Maturity)

운영 연계(Operational Alignment)





Broadcom's Strategy & Solutions



Frontier AI Models :

보안 위협 환경이 근본적으로 변화하고 있습니다.

April 20, 2026

WORLD
ECONOMIC
FORUM

"Risk of frontier AI is not just a cybersecurity issue, but a resilience issue for global stability"

[Web Economic Forum](#)

ISACA

April 17, 2026

"Time to exploit new vulnerabilities has been reduced from days or weeks to mere hours"

[ISACA](#)

REUTERS

April 13, 2026

Experts warn Mythos could expose legacy banking systems to unprecedented cyber risks

[Reuters](#)

SECURITYWEEK
INTERNET AND ENTERPRISE SECURITY NEWS, INSIGHTS & ANALYSIS

April 7, 2026

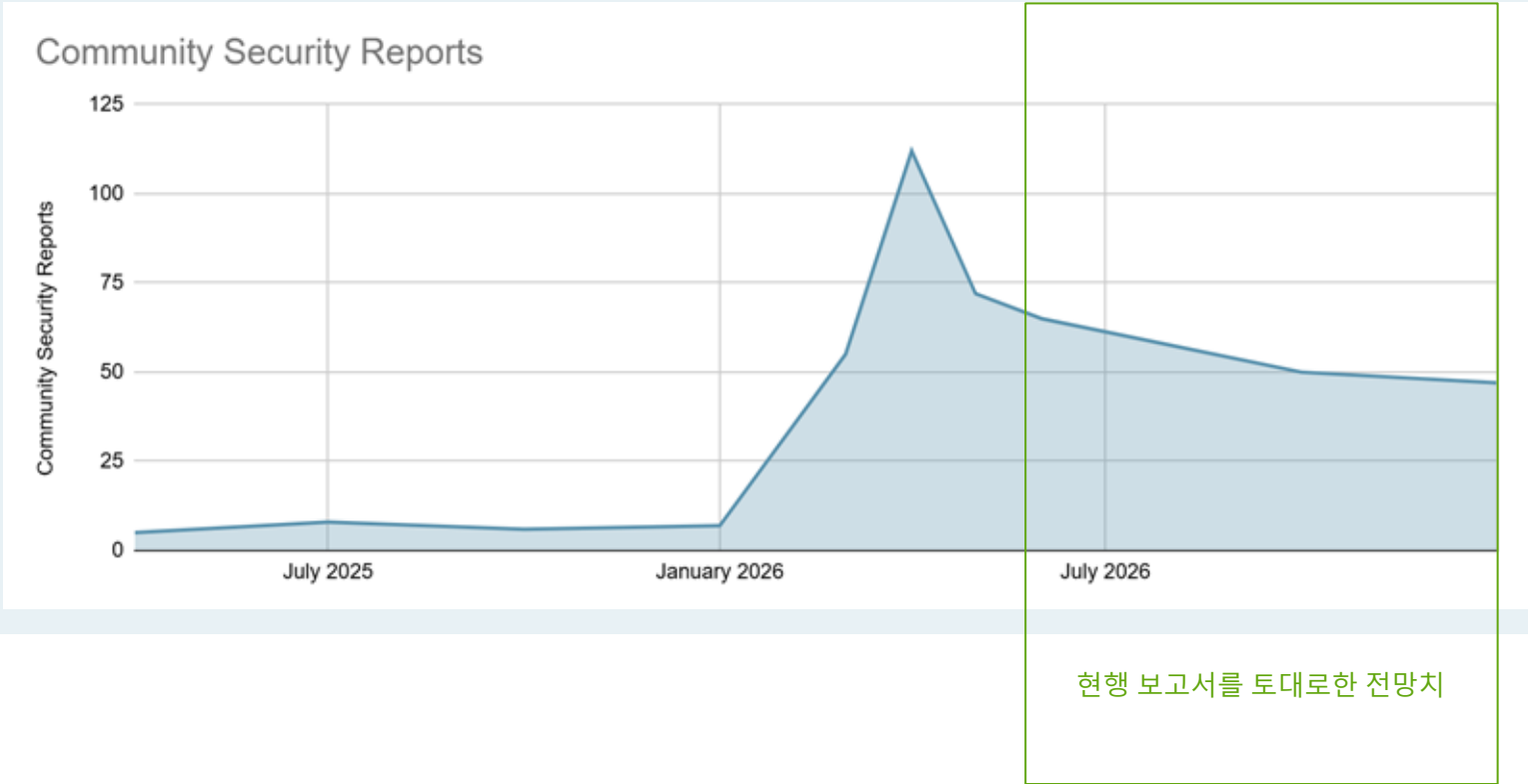
"Mythos Preview has identified thousands of zero-day critical vulnerabilities – the oldest found so far is a 27-year old bug in OpenBSD"

[SecurityWeek](#)



Spring 보안 환경은 급격히 변화 중

AI 프런티어 모델이 발전함에 따라 노출 위험의 지속 증가 예상(예: 취약점 체이닝)



1766%

2026년 4월, 커뮤니티 식별 보안 이슈가 전월 대비 증가

350+

2026년 4월, Broadcom이 AI 기반 분석을 통해 선제적으로 발견한 보안 문제

60%

Spring Boot 다운로드의 상당 부분이 OSS 지원 기간이 만료된 버전

Spring 프라이빗 리포지토리에 대한 접근 권한 사용 방법

- **Tanzu Platform**
- **Tanzu Spring Essentials**



“Broadcom은 Frontier AI Model을 보안 엔지니어링 전 과정에 통합하고 있습니다.

취약점 발견, 익스플로잇 검증, 패치 생성, 회귀 테스트까지 모두 포함됩니다.

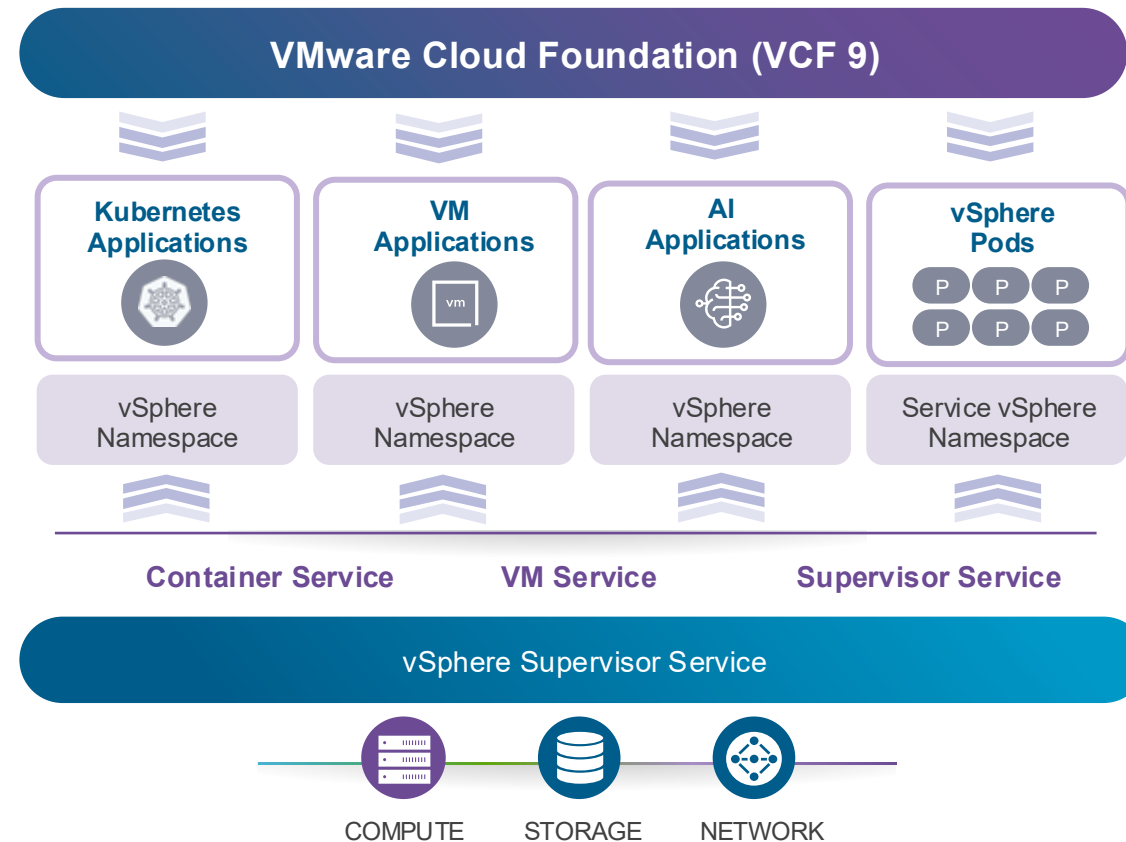
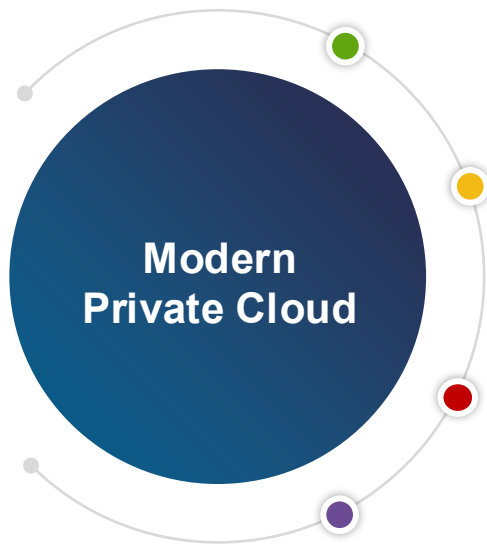
이러한 모델에서 가치를 끌어내기 위해 필요한 규율을 우리의 엔지니어링팀 운영 방식에 내재화하고 있습니다.”

[Broadcom ISG Blog](#) “What We Learned Testing Frontier AI Security Models Against Our Own Code” (April 23, 2026)



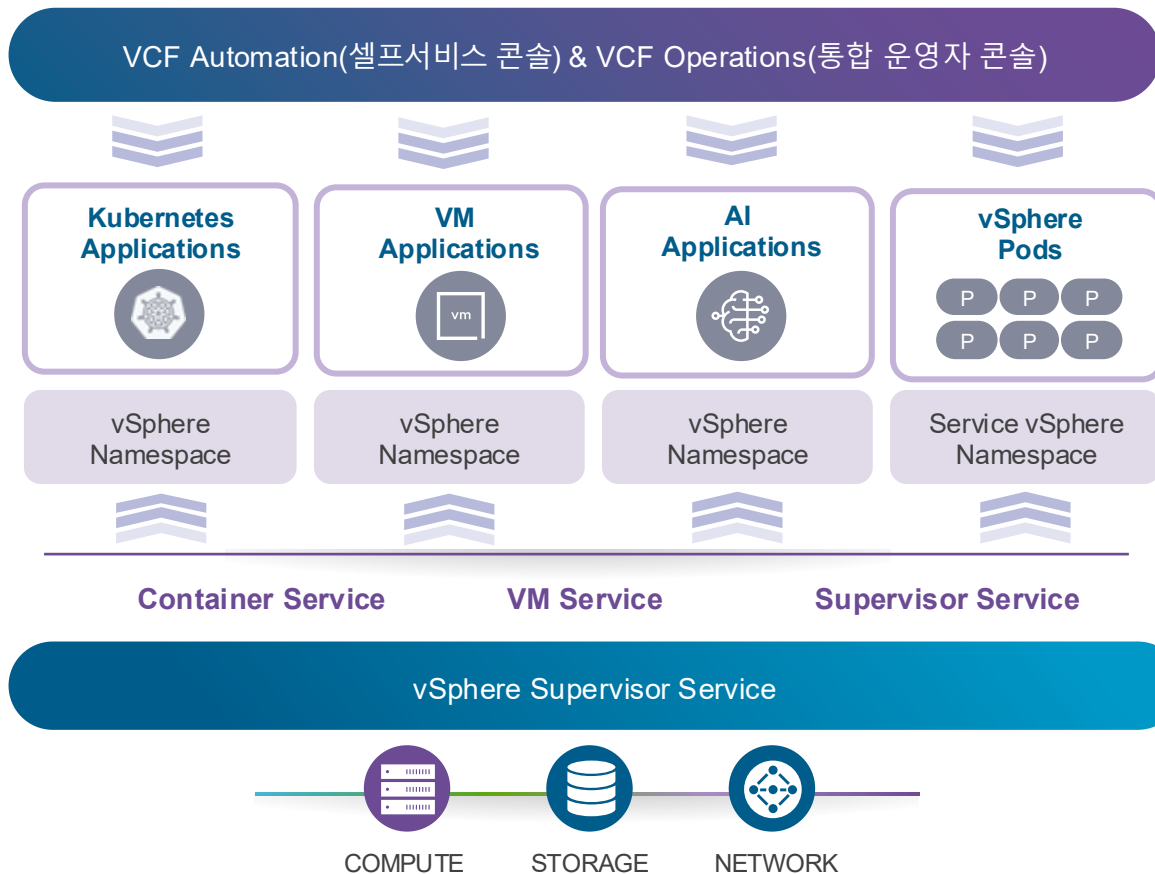
What & Why VCF 9?

VMware의 새로운 모던 프라이빗 클라우드 플랫폼으로 데이터센터 혁신을 위한 플랫폼



Platform for All Apps

VM, Container 및 AI 앱까지 하나의 플랫폼에서 모든 앱이 동일한 기준으로 운영될 수 있도록 보장하는 진정한 통합 플랫폼



VM과 컨테이너(k8s)를 위한 통합 플랫폼

- VM과 Pod를 vCenter에서 관리, 단일 스케줄링

플랫폼에 통합된 Private AI 서비스

- VCF에 기본 내장된 AI 서비스를 통해 On-Prem 환경에서 쉽게 AI 서비스 배포 및 관리

통합 보안 및 거버넌스

- 플랫폼에 내제된 End to End 보안관리 체계 & 라이브패치

운영효율성 및 자동화

- 인프라부터 애플리케이션까지 성능, 용량 등 통합 운영관리



최신 버전으로의 업그레이드

Why upgrade to VCF 9.1?

① 기능적 측면 :

완전히 새로워진 아키텍처와 차원이 다른 클라우드 서비스



② 보안적 측면 :

가장 안전한 VCF(or vSphere) 버전
- vSphere 7 (기술지원 종료)
- vSphere 8 (27년 10월 기술지원 종료 예정)



③ 비용적 측면 :

서버 비용상승 국면에서의 가장 현실적 대안
인프라 투자비용 획기적 절감



VCF 9 Offering

Advanced Services



Advanced Cyber Compliance³



VMware Site Recovery Manager



VMware Live Recovery cloud



Data Services Manager³



vDefend³



Avi Load Balancer



Tanzu Platform

VMware Cloud Foundation



vSphere
(includes vCenter)



vSAN¹



VCF Networking
(includes NSX)



VCF Operations²
(includes HCX)



VCF Automation
(Includes VKS cluster management)



VKS⁴



VCF Private AI Services⁵

Broadcom Support

¹ vSAN is licensed 1 TiB/Core for VCF and 0.25 TiB/Core for VVF. Additional vSAN can be purchased per TiB.

² VVF customer receive VCF Operations for VVF which does not include HCX

³ For VCF and VCF Edge only; not available for VVF.

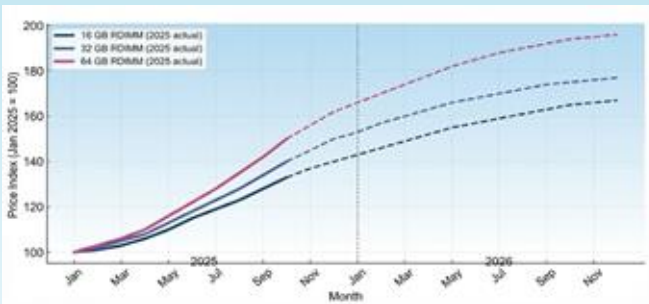
⁴ Limited experience for VKS with VVF



Challenge > 메모리 가격상승과 서버 공급부족



현상 : 서버 수급 부족



- **메모리 가격폭등** : 표준 DRAM 가격은 올해 두 배까지 상승할 것으로 예상¹
- **서버가격 동반 상승** : 표준 서버 구성품들도 같은 추세
- **재고 부족 심화** : 2026년까지 생산 가능한 물량이 이미 모두 소진²



원인 : 공급체계 문제



- **폭발적 수요(AI)** : 하이퍼스케일러의 대규모 AI 센터 구축 경쟁³
- **생산구조 변화(HBM)** : AI GPU용 HBM으로 생산라인 전환
- **DRAM 과점구조** : 삼성·SK하이닉스·마이크론이 전 세계 공급의 95%를 점유⁴



예측 : 장기화

IDC

"In late 2025, the global semiconductor ecosystem is experiencing an unprecedented memory chip shortage with knock-on effects for the device manufacturers and end users that could **persist well into 2027.**"

IEEE
Spectrum

Speaking of his sense of the DRAM market, Intel CEO Lip-Bu Tan told attendees at the Cisco AI Summit last week: "There's no relief until 2028."

- **IDC** : 이 상황은 디바이스 제조사 및 최종 사용자에게까지 미치고 있고 2027년까지 지속될 수 있다.
- **IEEE Spectrum** : 2028년까지 어떤 완화도 없을 것이다. @Cisco AI Summit



단기적 해소가 어려운 상황을 전제하고 IT 운영전략 재고 필요



관점의 전환 (Shift of Viewpoint)

기존 서버의
활용



신규 증설
최소화



인프라 전반의
효율성 향상



Strategy > 데이터센터 효율화 “Turning Point”

Resource
Utilization



Cost



- **Resource Pooling**

- 컴퓨팅 자원 사용률 향상 (CPU, RAM)

- **Storage Space 효율화**

- Software-Defined Storage를 통한 성능/비용 이점 구현

- **Flexible Infrastructure 구현**

- 탄력적 인프라 확장 및 자동화

Private Cloud

- **기존 서버 활용**

- 자원(CPU, RAM) 최적화를 통한 비용 절감

- **신규 서버 증설 최소화 및 지연**

- 기존 서버 활용 극대화로 신규 투자 절감 및 지연

- **운영 효율화를 통한 OPEX 절감**

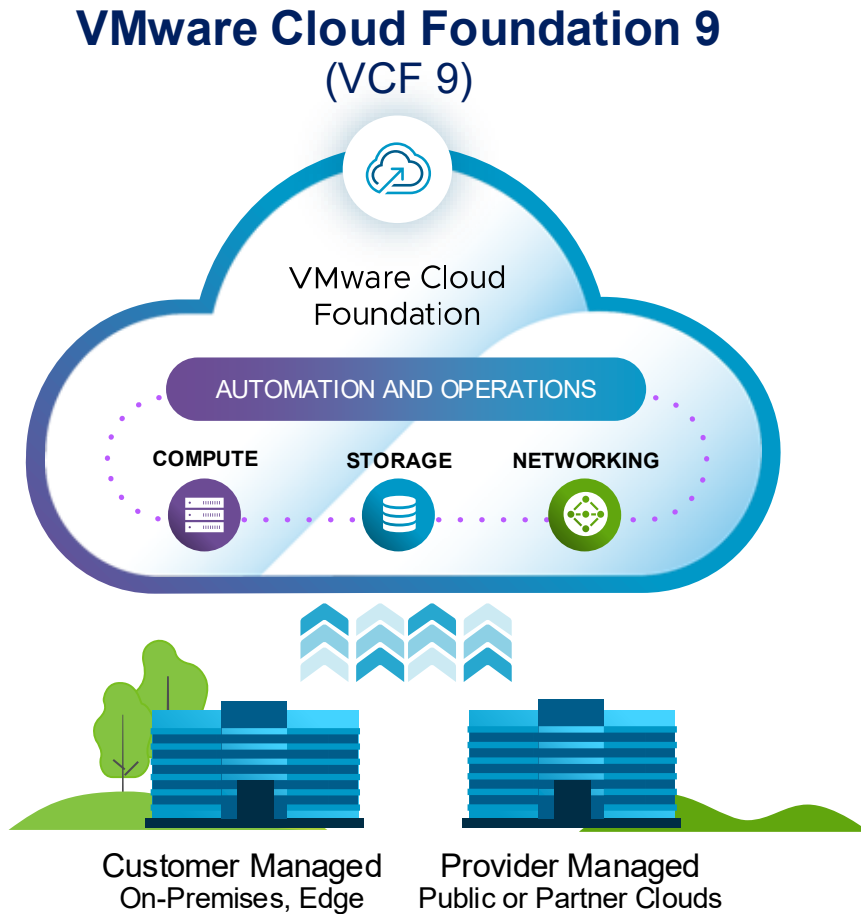
- Flexible Infra 및 자동화를 통한 운영 효율화 확보

Turning Point >

“이제는 서버를 늘리는 것보다 서버 활용도를 높이는 전략이 필요”



Solution > VCF 9을 통한 TCO 절감



01

기존 서버 재활용 및 신규 서버 도입 최소화

by 자원최적화 진단서비스 (Resource Optimization Assessment)

02

독점적 메모리 확장 기술로 집적도 극대화

by VCF 9 Memory Tiering Technology

03

신규 인프라 구성 시 비용절감 극대화

by VCF 9 Storage – New Architecture(ESA)

04

하이브리드 클라우드를 통한 유연한 인프라 전략

by 라이선스 이동성 (License Portability)



기존 인프라 다시 돌아보기

문제인식

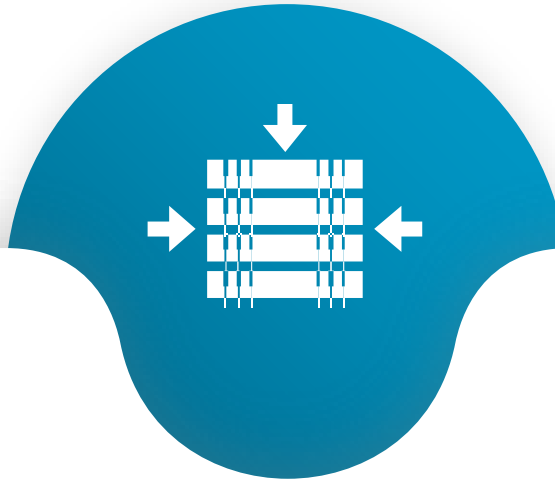
기존 데이터센터 컴퓨팅 리소스의 낭비와 낮은 자원 활용도

Legacy System



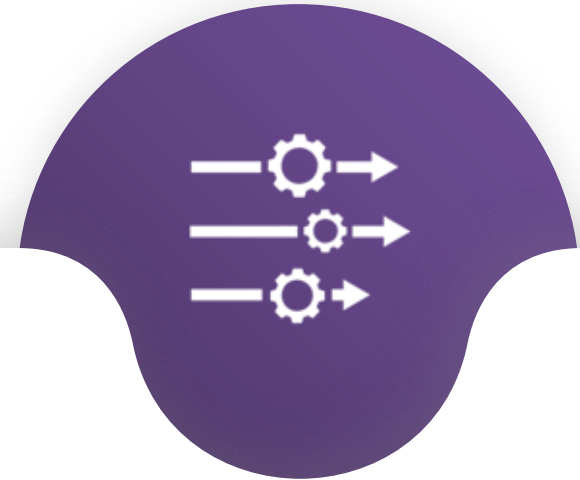
“오래된 워크로드가 현재적 가상화, 클라우드 환경으로 이전되지 않아 효율 저하”

Over-Sizing



“장애 및 Peak 대비로 서버, 스토리지를 과도하게 산정하여 실제 사용량 저하”

최적화 미이행



“효율성 관련 핵심 KPI 를 통한 자원 최적화 미이행.”



Solution > 자원 최적화 진단서비스를 통한 추가 리소스 확보

Solution

컴퓨팅 자원 최적화 진단

01

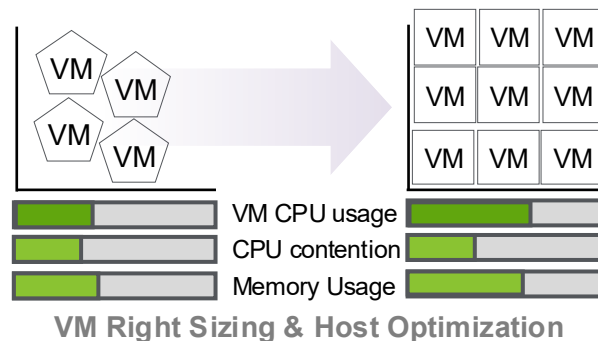
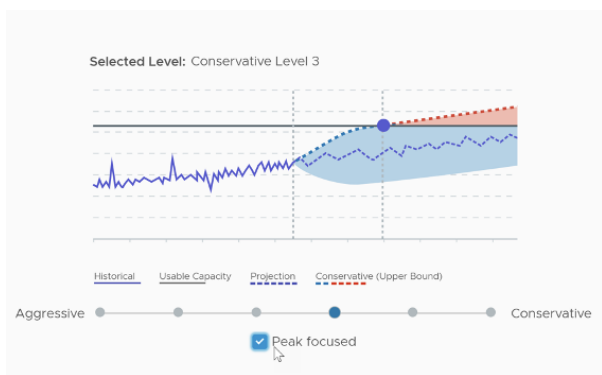
주요 지표 분석
Key Metric Analysis

02

최적화 사이징
Right Sizing & Optimization

03

리소스 통합 제언
Consolidation Simulation & Proposal



40% ↓ 30% ↑

구매비용

자원 활용률

Example



Solution > 자원최적화를 통한 인프라 추가 확보(사례)

컴퓨팅 자원 최적화 컨설팅

01

주요 지표 분석
Key Metric Analysis

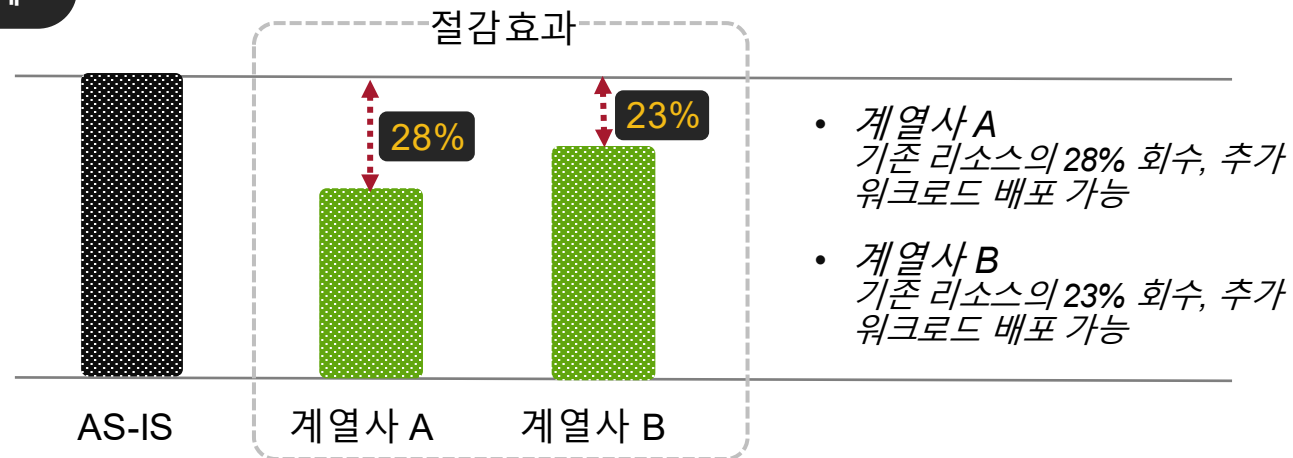
02

워크로드 최적 사이징
Right Sizing

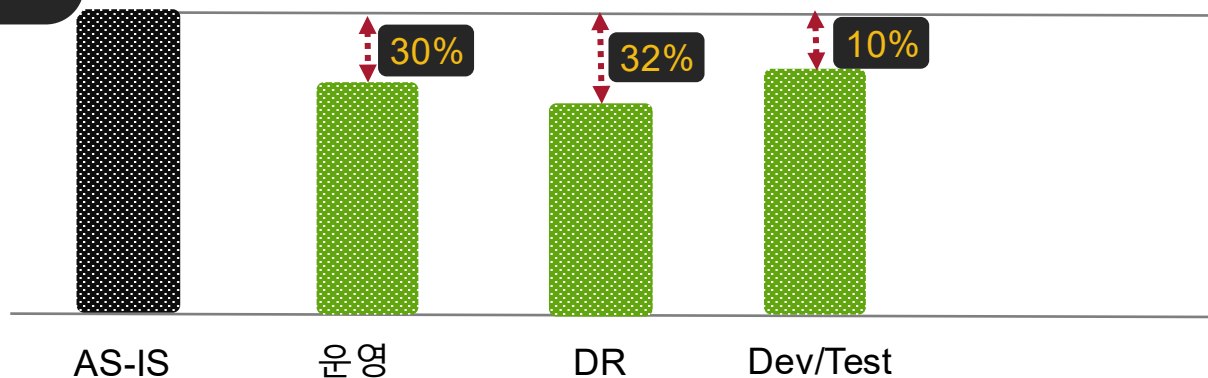
03

리소스 통합 제언
Consolidation Simulation & Proposal

DC 전체



계열사 A



서버 자원 활용의 불균형에 의한 리소스 낭비

Problem

애플리케이션의 변화와 자원 활용의 불균형으로 인한 리소스 낭비

CPU



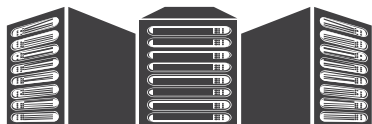
Memory



Storage



Cluster and Usage



Without Memory-Tiering of VCF

- 서버 추가 증설 필요
- CPU 자원 활용률 매우 저하



메모리 티어링 기술을
활용한 가용 메모리 확장
(NVMe 활용)

CPU



Memory



Storage



Cluster and Usage



With Memory-Tiering of VCF

- 서버 증설 없이 NVMe Device 추가로 Memory 확장
- 추가 워크로드(VM 등) 구동 가능 -> CPU 자원 활용률 향상



Solution > 메모리 티어링 기술을 통한 메모리 확장

Solution

Memory Tiering
in VCF9



NVMe를 2차 메모리
계층으로 활용

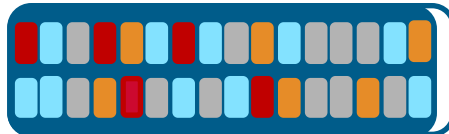
최대 **40%**
TCO 절감 효과

Server #1
with DRAM Only

CPU



Memory(DRAM)



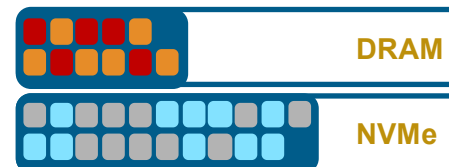
- 서버 증설 필요
- 메모리 병목(Memory Bound) 상황
- 추가 VM 배포 불가

Server #1
with MemoryTiering

CPU



Memory(DRAM)



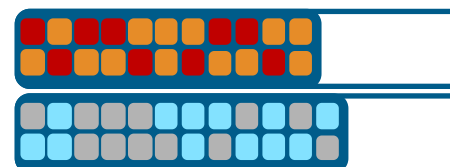
- 서버 증설 불필요
- 메모리티어링으로 메모리 병목(Memory Bound) 해소
- 추가 VM 배포 가능

Server #1
with More VMs

CPU



Memory(DRAM)



- 서버 증설 불필요
- 워크로드 추가 배포
- 메모리티어링으로 메모리 병목(Memory Bound) 해소

Hot Pages Warm Pages Cold Pages Very Cold Pages



Solution > 국내기업 데이터센터 진단 결과(사례)

Linux 클러스터

Total Memory 17 TB

Hot 1.1 TB (7%)

Cold 10.5 TB (68%)

WINDOWS 클러스터

Total Memory 19 TB

Hot 1.4 TB (7%)

Cold 12.4 TB (73%)

클러스터	호스트	전환가능 호스트	현재 VM	물리 메모리 합	VM 할당 합 (GB)	Active 합 (GB)	Cold 합 (GB)
EDGECLUSTER	1	1	13	1530.48 GB	656.3 GB	16.7	464.4
LINUXCLUSTER	15	15	521	17270.55 GB	12499.5 GB	1146.5	10545.4
LINUXCLUSTER02	5	5	317	7673.1 GB	5140.3 GB	536	4147.2
LINUXCLUSTER03	5	5	112	32763.68 GB	13396.3 GB	1318.5	10570.3
MCMTCLUSTER	1	1	31	1530.52 GB	937.3 GB	158.1	725.5
WINDOWSCLUSTER	15	15	551	19194.33 GB	16596.3 GB	1430.6	12489.6

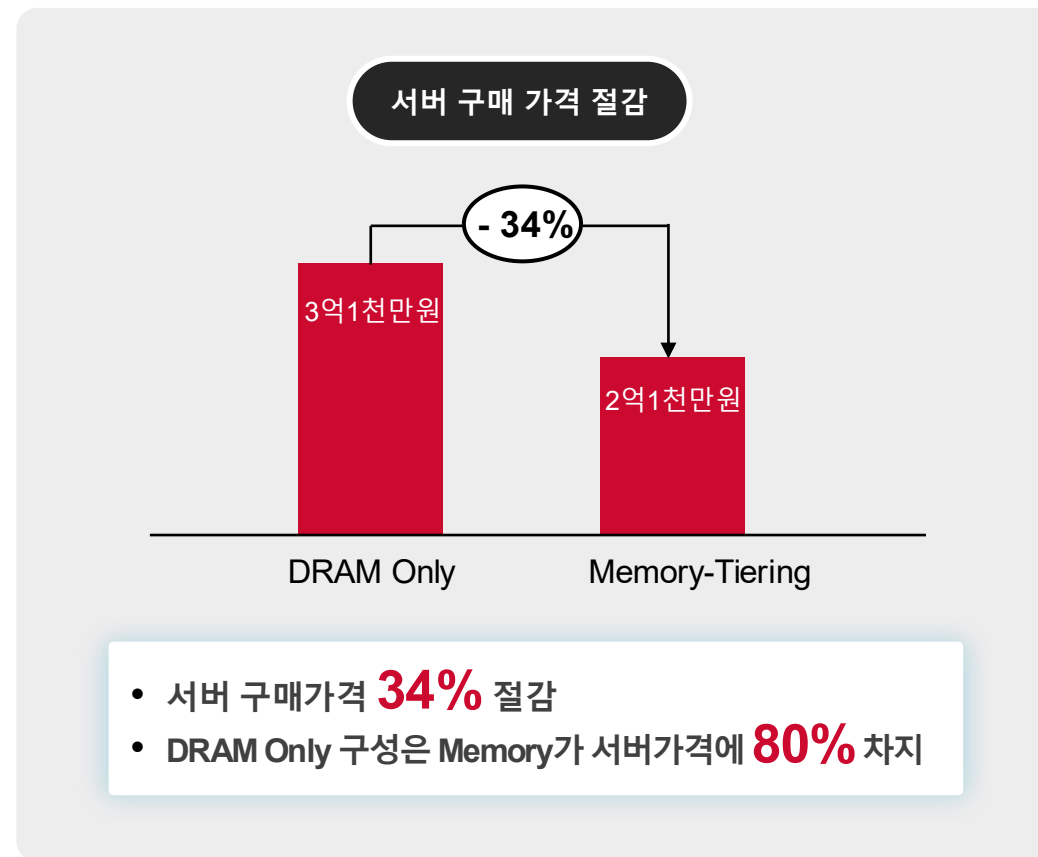


Solution > 메모리 티어링 기술을 통한 서버비용 절감(사례)

국내 A 카드사

“표준사양 기준으로 **서버 도입가격** 낮추는 전략”

Hardware	DRAM-Only	NVMe-Tiered
HPE DL380 G12 Intel Xeon 6730 P (2xCPU / 64C)	\$40,000	\$40,000
DRAM - HPE 64GB (1x64GB) DDR5-6400 (\$5,000/ea)	\$160,000(2TB)	\$80,000 (1TB)
NVMe Tier - 1.6TB - DWPD3 (\$7,000/ea)	-	\$14,000 (2 x 1.6TB)
Total Usable Memory	2 TB	2.6 TB
Total cores per server	64	64
Total	\$200,000	\$135,000 (34% Off)



Solution > 메모리 티어링 기술을 통한 서버비용 절감(사례)

국내 B은행

“표준사양 확장으로 **호스트 수량 & 비용**을 줄이는 전략”

① B사 표준서버
- DRAM Only

Dell PowerEdge R770	
Component	Capacity
CPU (Intel Xeon 6737P 32C *2)	64 Cores
MEM (96GB RDIMM, 6400MT/s)	1.5TB
Total Memory	1.5TB
Total Price	3.7억원

(① 표준서버)

② B사 표준서버 + 1.6TB NVMe 구성
- DRAM + NVMe(미러구성)

Dell PowerEdge R770	
Component	Capacity
CPU (Intel Xeon 6737P 32C *2)	64 Cores
MEM (96GB RDIMM, 6400MT/s)	1.5TB
NVMe (1.6TB SSD NVMe)	1.6TB*2 (3.2TB)
Total Memory	3.1TB
Total Price	4.1억원

① 표준서버 대비 **11%** 증가

③ DRAM Only 구성 대비 **40%** 절감

③ B사 표준서버 + 1.5TB Mem 증설
- DRAM Only

Dell PowerEdge R770	
Component	Capacity
CPU (Intel Xeon 6737P 32C *2)	64 Cores
MEM (96GB RDIMM, 6400MT/s)	3TB
Total Memory	3TB
Total Price	6.8억원

① 표준서버 대비 **84%** 증가

VM당 단가
(예 :50VMs per Host) “② 8백20만원” vs “③ 1천4백만원”



Solution > 신규 인프라 구성 시 비용 최소화(Storage 측면)

Solution

VCF Storage – ESA Technology

- ▶▶▶ 성능 저하 없이 용량 절감 효과 실현
- ▶▶▶ 최대 4배 더 높은 압축률을 제공
- ▶▶▶ 전체 장치를 활용해 사용 가능한 스토리지 확장
- ▶▶▶ 성능 오버헤드를 최소화해 더 많은 VM을 통합

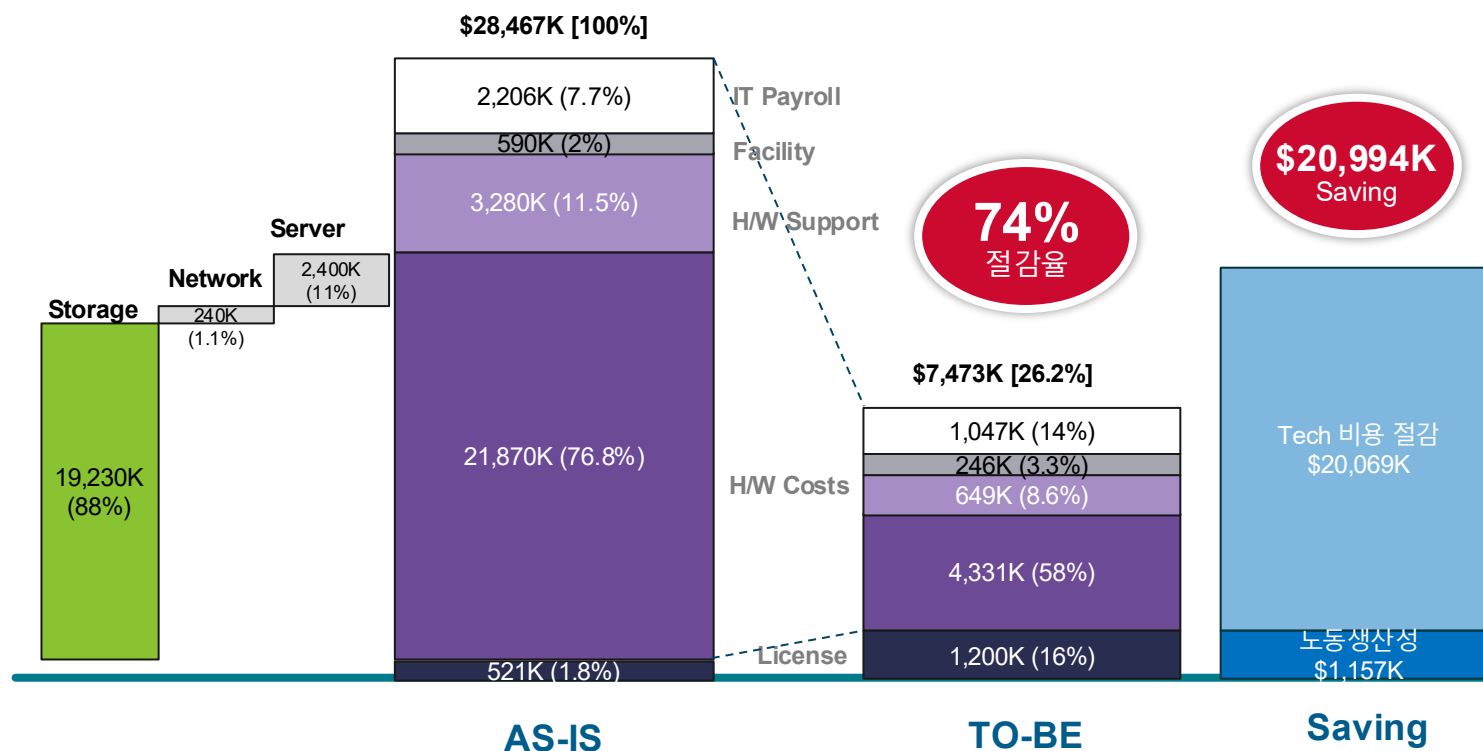


Up to
40%
lower TCO



Solution > 신규 인프라 구축 시 TCO 절감(사례)

VCF 전환 TCO 비교



효과

01

Tech Leverage 효과

- » 3Tier → HCI 구조 전환으로 Storage 구매 비용 80% 절감
- » HW 유지보수 비용 절감 (Server/Storage/Network)

02

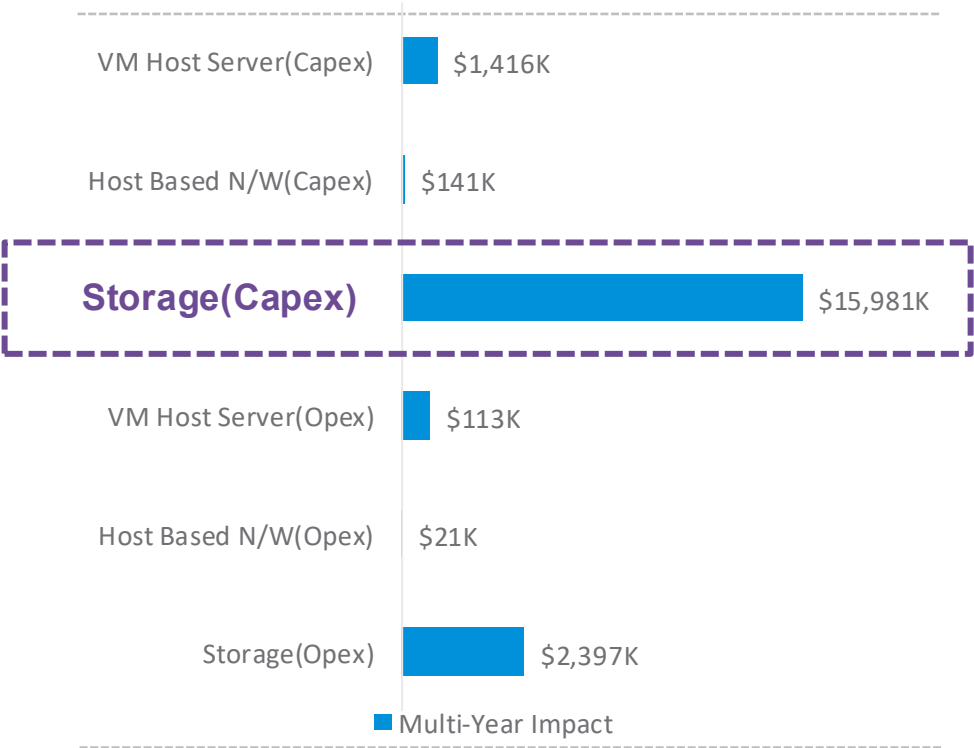
운영효율성 향상

- » HW 유지보수 인력규모 감소
- » 영역별 약 40 ~ 60% 이상의 노동생산성 향상

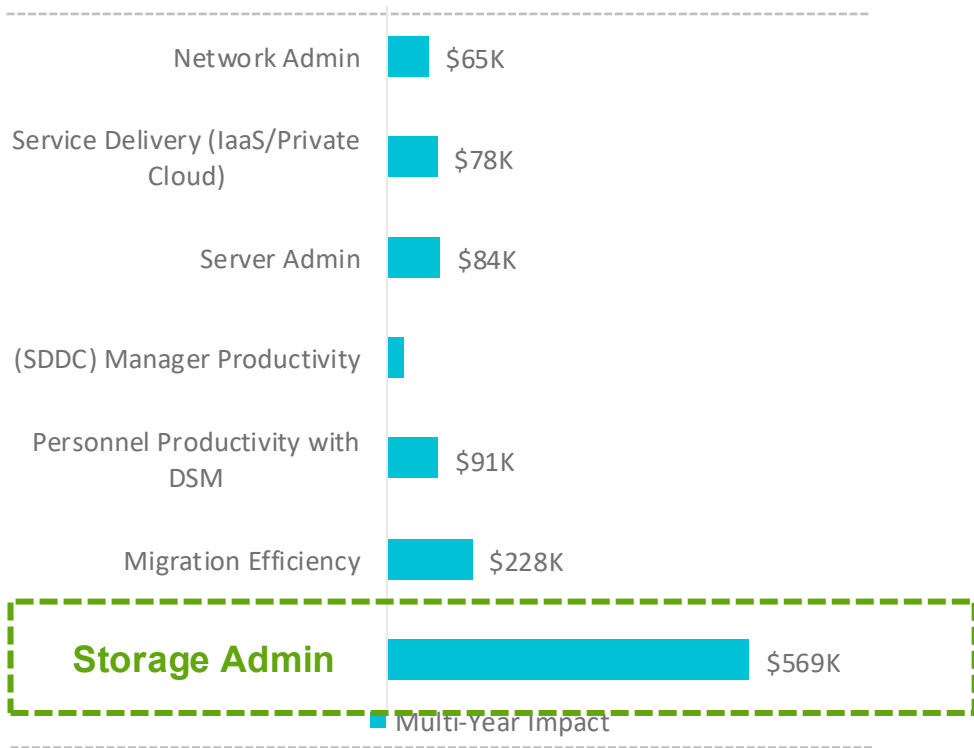


Solution > 신규 인프라 구축 시 TCO 절감(사례)

하드웨어 비용 절감, CPU 효율성 향상



하드웨어 비용 절감, CPU 효율성 향상



Standard Chartered Selects Broadcom to Deliver Secure, Always-On Banking Services at Global Scale

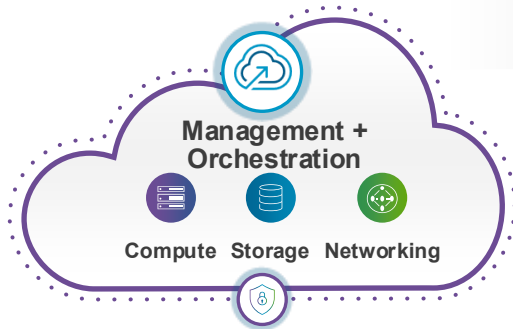
July 15, 2026



Solution > 하이브리드 클라우드 활용

라이선스 이동성 (Bring Your Own License)

VMware Cloud
Foundation™



To Public Cloud

VCF License + Hyperscaler Infra

On-Prem 라이선스를 활용해서 Public Cloud 인프라에 적용
Azure VMware Solution, Google Cloud VMware Engine,
Oracle Cloud VMware Solution, AWS EVS 등



Google Cloud



To Private Cloud

VCF License + Customer Owned Infra

서버가격 조정 국면에서 손쉽게 On-Prem으로 마이그레이션
(TCO 비교로 의사결정)



and more...



Summary — Call to Action



최신 버전으로 업그레이드(VCF 9.1) !!!



기존 서버를 활용한 인프라 확장 – “자원최적화 진단서비스”



기존 서버를 활용한 인프라 확장 및 신규 투자 최소화 – “메모리티어링 by VCF 9”



신규 인프라 구성간 TCO 절감 – “VCF 9 : 모던 프라이빗 클라우드 플랫폼”





Thanks