



Pure Orange Shot Webinar #2

디지털 혁신을 위한 스마트 데이터 플랫폼 구축 전략

김진환 이사

Enterprise / SE
Pure Storage Korea



디지털 혁신을 위한 동력



비즈니스
분석



사이버 보안



클라우드



애플리케이션
현대화 및
DevOps



고객 경험



인공지능(AI)
머신러닝(ML)



4.8배
매출 증대 가능성

3.2배
고객 만족도 향상

2.4배
업무 생산성

디지털 혁신 관련 기술 트렌드

S3: 진정한 개방형 데이터 플랫폼

- 빅 데이터에 대한 단일 출처 정보 (전사적 데이터 통합 플랫폼)
- 하이브리드/멀티 클라우드 데이터 플랫폼

Pluggable 데이터 레이크 하우스

- 동일한 데이터, 다른 엔진, 데이터 복사 없음
- 오픈 테이블 형식 및 스토리지 API
- 개방형 데이터 레이크하우스 선택

Kubernetes: as-a-service 아키텍처 핵심

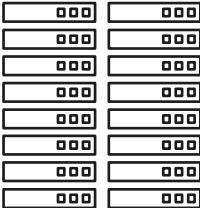
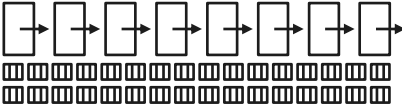
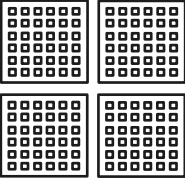
- 효율적인 클러스터 리소스 공유 및 격리
- 표준 API로 운영 자동화 및 단순화
- 애플리케이션 이동성 확보로 멀티/하이브리드 클라우드 구현 용이

분산 ML

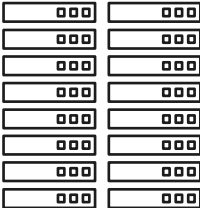
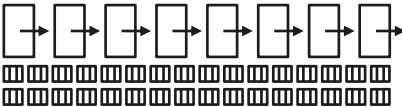
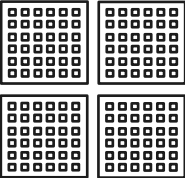
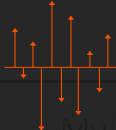
- AI 서비스 다양화
- 여러 GPU 서버에 걸친 ML 교육
- 당신만의 ChatGPT

데이터 플랫폼 도전과제

다양한 분석 요건으로 인한 플랫폼 사일로화로 복잡성 및 비용 증가

	BACKUP & DATA WAREHOUSE	DATA LAKE	STREAMING ANALYTICS	AI CLUSTER
	 APPLIANCES	 DAS	 DISAGGREGATED	 HPC SYSTEM W/ GPU
DATA TYPE	Structured	Unstructured	Unstructured	Unstructured
PROCESSING MODE	Batch	Batch	Micro-batch / real-time	Real-time
I/O TYPE	Sequential Write, Random Read	Sequential	Random	Sequential to random
ARCHITECTURE	Scale-up	Scale-out	Multi-dimensional	Massively parallel

통합 데이터 플랫폼을 위한 스토리지 요건

	BACKUP & DATA WAREHOUSE	DATA LAKE	STREAMING ANALYTICS	AI CLUSTER
	 APPLIANCES	 DAS	 DISAGGREGATED	 HPC SYSTEM W/ GPU
스토리지 필수 요건	Structured	Fast NFS Protocol	Fast Object Protocol	Unstructured
	Batch	Batch	Micro-batch / real-time	Real-time
	 일관된 고성능	 스케일아웃	 애플리케이션 멀티 프로토콜	 초대용량 병렬처리
	Scale-up	Scale-out	Multi-dimensional	Massively parallel

왜 오브젝트 (S3) 서비스인가?

API 기반으로 데이터 서비스 표준화가 용이한 오브젝트 스토리지(S3) 활용 확대

	S3 (on-prem)	HDFS
구현	Native 데이터 서비스 - 간단한 유지 관리 (구성, 업그레이드, 확장, 장애 복구 등) - 간단한 통합	유저레벨 데이터 서비스 - 복잡한 유지 관리 (구성, 업그레이드, 확장, 장애 복구 등) - 복잡한 통합 과정
비용	\$\$ (하드웨어)	\$\$\$ (소프트웨어 + 하드웨어)
가용성	스토리지 가용성에 의존 (99.9999% 이상)	서버 가용성에 의존 (99.99%), 클러스터 사이즈에 의존
성능	Transactional 데이터 쓰기 가능 (with DB) 대용량 읽기에 최적화 (Small, Mid 사이즈)	Transactional 데이터 쓰기 어려움 대용량 읽기에 최적화 (Large 사이즈)
기타	Cloud 및 데이터 서비스에 최적화	데이터 프로세싱에 최적화

35%

오브젝트 스토리지
연평균 성장률
(2017년~2022년)

100%

전년 대비 오브젝트 스토리
지에 대한 기업 내 우선순
위 및 이니셔티브 증가율

87%

2022년까지 비정형 데이터
의 대부분이 오브젝트 스토
리지에 저장

80%

오브젝트 스토리지가 주요
IT 이니셔티브를 지원

오브젝트 스토리지 지원 확대

Aggregated(DAS) 모델에서 Disaggregated(CS분리) 로의 전환 가속화



Tiered Storage

Frozen Tier

Eon mode

Disaggregated

Smart Store

FAST OBJECT(S3)

- 초고속 S3 연결을 통한 일관된 성능 제공
- 컴퓨팅과 데이터의 분리를 통한 높은 확장성
- 하드웨어 장애에 대한 리스크 최소화(초고속 복구, 안정적 업그레이드, 간편한 신규 배포)

데이터 레이크하우스를 위한 FlashBlade//S

단일 플랫폼에서 UFFO 데이터 서비스를 제공하여 다양한 데이터 관련 워크로드를 통합

BACKUP & RAPID RESTORE

COMMVAULT
VEEAM



DATA WAREHOUSE

VERTICA
ORACLE



DATA LAKE

SAS SPARK



STREAMING ANALYTICS

kafka elastic



AI CLUSTER

Caffe2 TensorFlow



HIGH THROUGHPUT
FILE & OBJECT



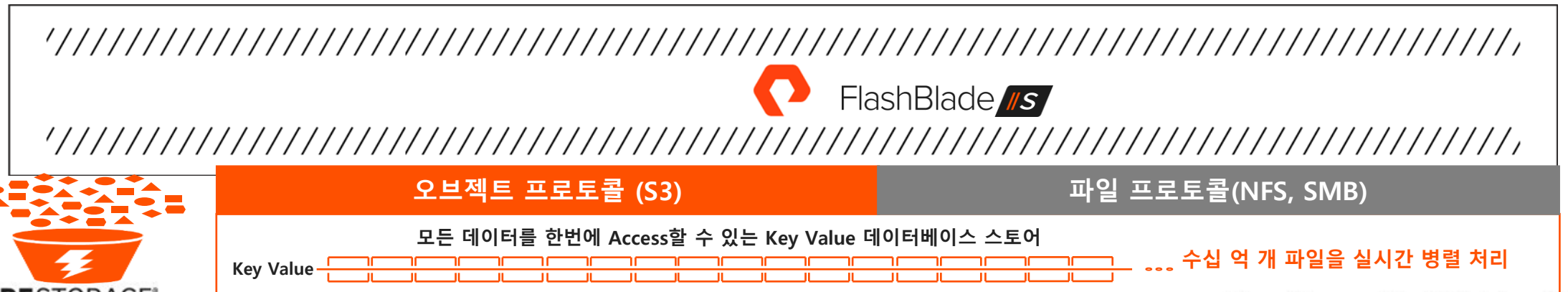
SIMPLE, NATIVE
SCALE-OUT



MULTI-DIMENSIONAL
PERFORMANCE

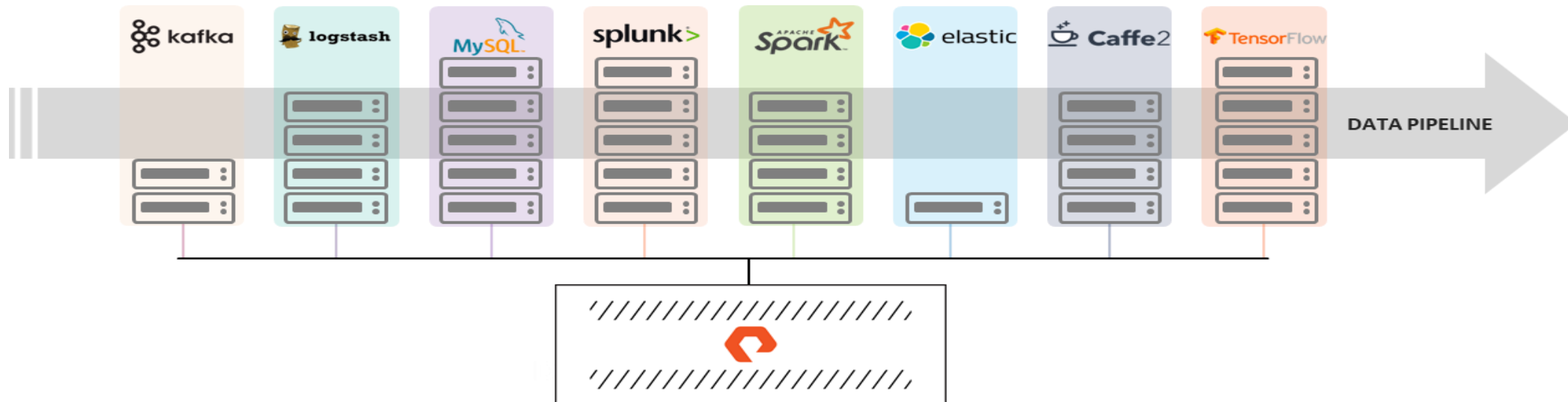


MASSIVELY
PARALLEL



FlashBlade//S 특징점

업계 유일의 다이렉트 플래시 기술이 적용되어 경쟁사 대비 높은 성능, 수명, 용량 집적도를 제공



1

일관된 고성능

- Flash 최적화 소프트웨어 및 하드웨어
- All NVMe 기반 (샤시당 60GB/s, 시스템당 48M IOPS)

2

높은 저장 효율

- 고집적 3D V-NAND 기술
- 고효율 압축 기능
- 52RU에서 20PB 구성

3

높은 제품 신뢰성

- 업계 선도 미디어 관리 기술로 Flash 수명 연장
- 선제적 대응 서비스로 장애 사전예방

RAPIDFILE TOOLKIT 2.0

시스템 관리자 생산성 향상

- Linux 유틸리티보다 수십~수백배 빠른 속도
- 파일 관리 및 분석 가속화
- pls, pdu, pfind, pchown, prn, pcp

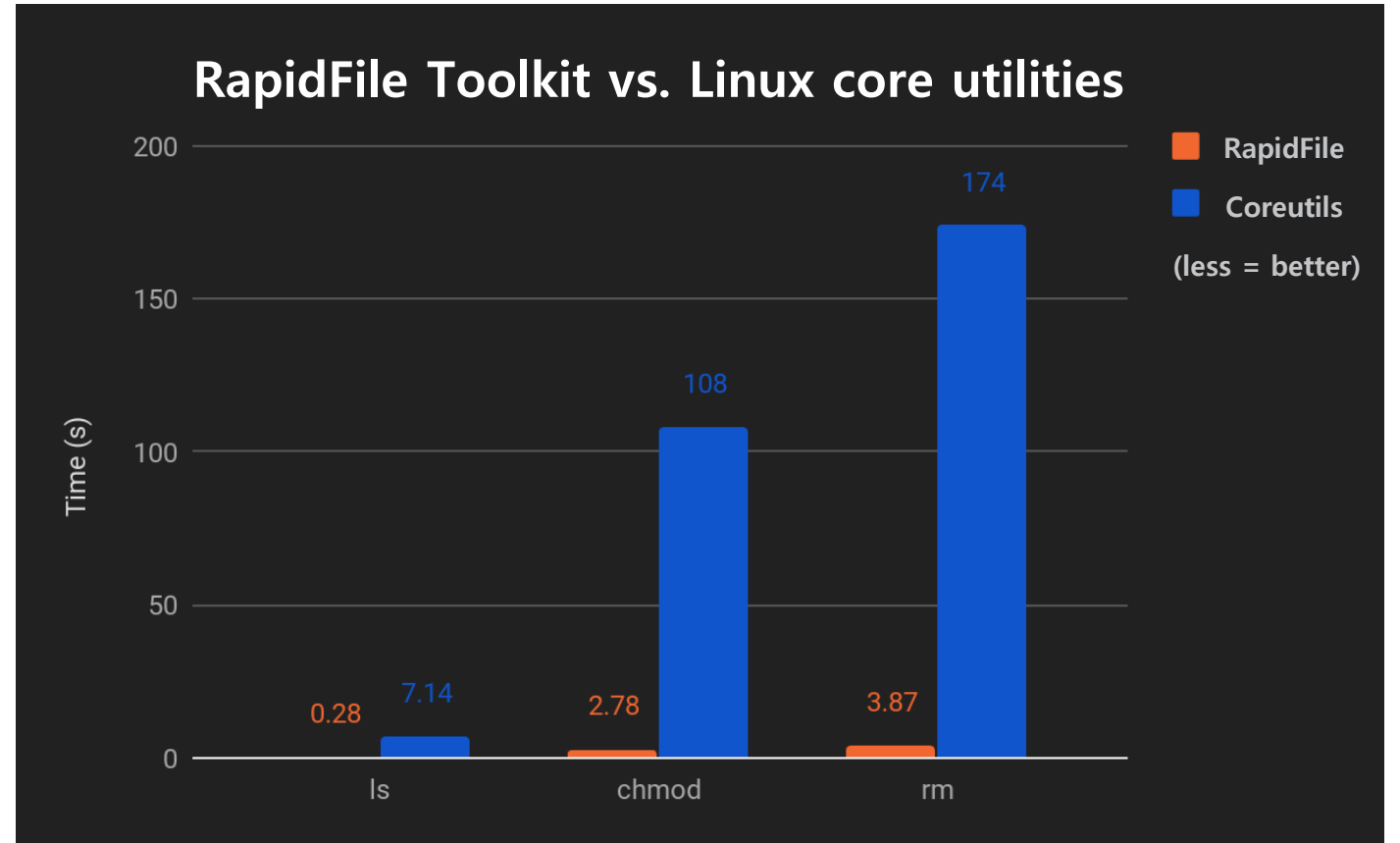
보다 빠른 데이터 이동 및 분석

- Perforce 체크아웃 20배 가속화
- 빠른 데이터 복사 from and to scratch 영역

데이터+ 파이프라인 가속화 & 단순화

- 20배 빠른 파일시스템 인덱싱, 메타데이터 캐싱 타임 감소
- EDA, Genomics, DevOps, HPC, Analytics & Apache Spark 및 AI/ML 지원

Analytics, AI/ML, EDA & DevOps 워크로드 가속화



Flashblade//S 전환 시 비용 절감

Usable 1PB, 5년 TCO 기준 전체 비용 13억원(1M \$) 절감

기존 소유 비용

HDFS with HDD	Number / USD
서버 수	50
랙 수	3.125
서버 비용	\$1,150,000
라이선스 (아파치)	\$0
DC 비용(5 년)	\$312,500
총 소유 비용	\$1,462,500



교체 이후 소유 비용

FB //S	Number / USD
서버 수	21
랙 수	1
서버 비용	\$420,000
라이선스 (아파치)	\$0
DC 비용(5 년)	\$60,000
총 소유 비용	\$480,000



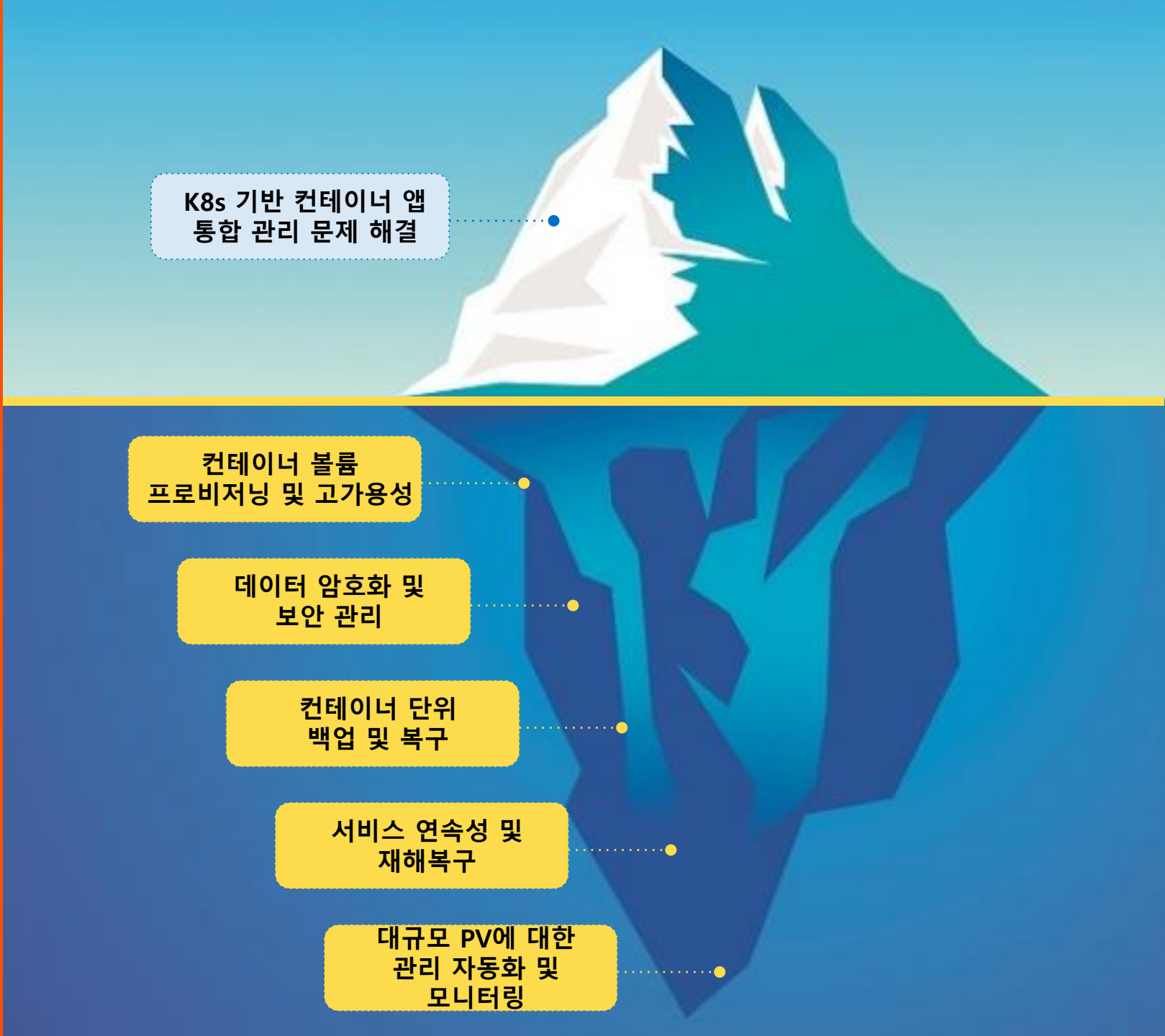
비용 분석

- HDFS 서버 : \$23,000
- 컴퓨트 서버 (HDD 제외) : \$20,000
- 노드 당 SW 라이선스 : \$0
- 랙 당 운영 비용 : \$1,000

Move to Stateful

컨테이너 데이터 고려사항

- 엔터프라이즈 데이터 서비스
- 멀티 클라우드 모빌리티
- 컨테이너 백업 복구
- 애플리케이션 단위 보안
- 수천 개 이상의 PV 관리 자동화



The diagram features a large iceberg floating in a blue ocean. The tip of the iceberg is above the water line, while the vast majority of the iceberg is submerged. A yellow horizontal line separates the water surface from the submerged part. A small blue dot on the tip is connected by a dotted line to a light blue box. Five yellow dots on the submerged part are each connected by a dotted line to a yellow box. The boxes contain text in Korean, listing various considerations for container data management.

K8s 기반 컨테이너 앱
통합 관리 문제 해결

컨테이너 볼륨
프로비저닝 및 고가용성

데이터 암호화 및
보안 관리

컨테이너 단위
백업 및 복구

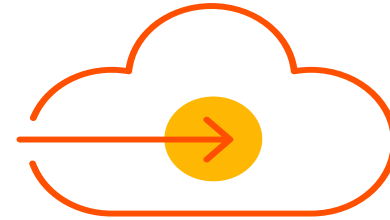
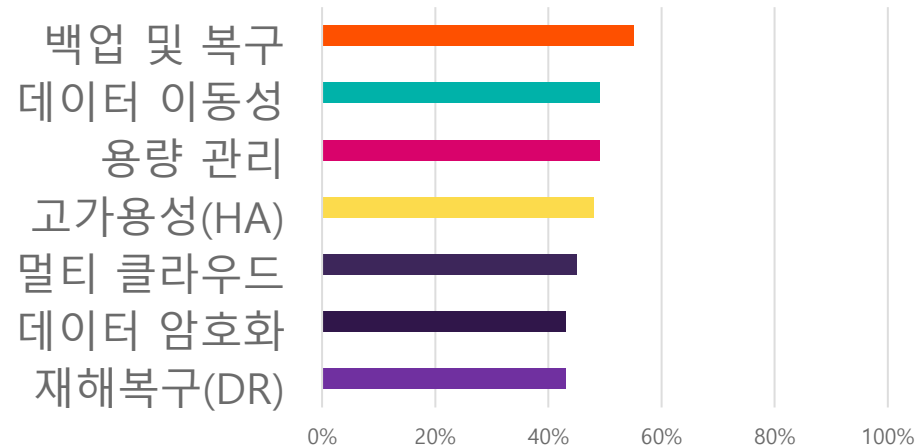
서비스 연속성 및
재해복구

대규모 PV에 대한
관리 자동화 및
모니터링

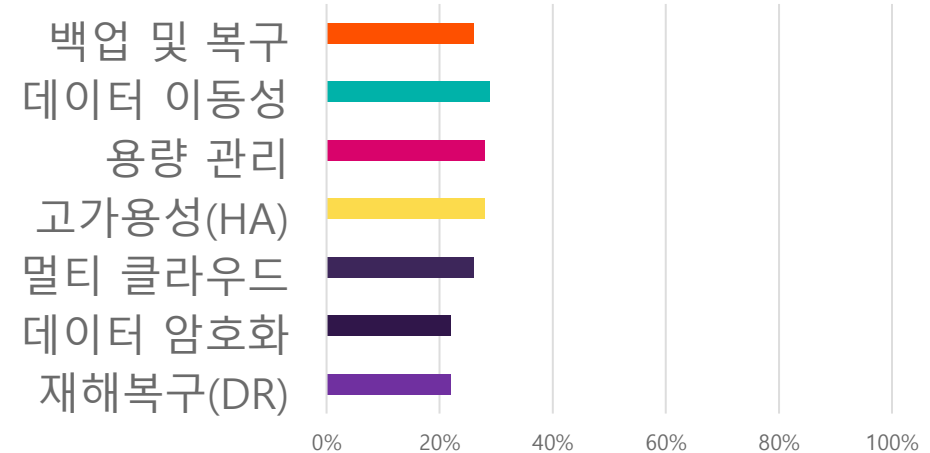
Kubernetes 플랫폼 도전 과제



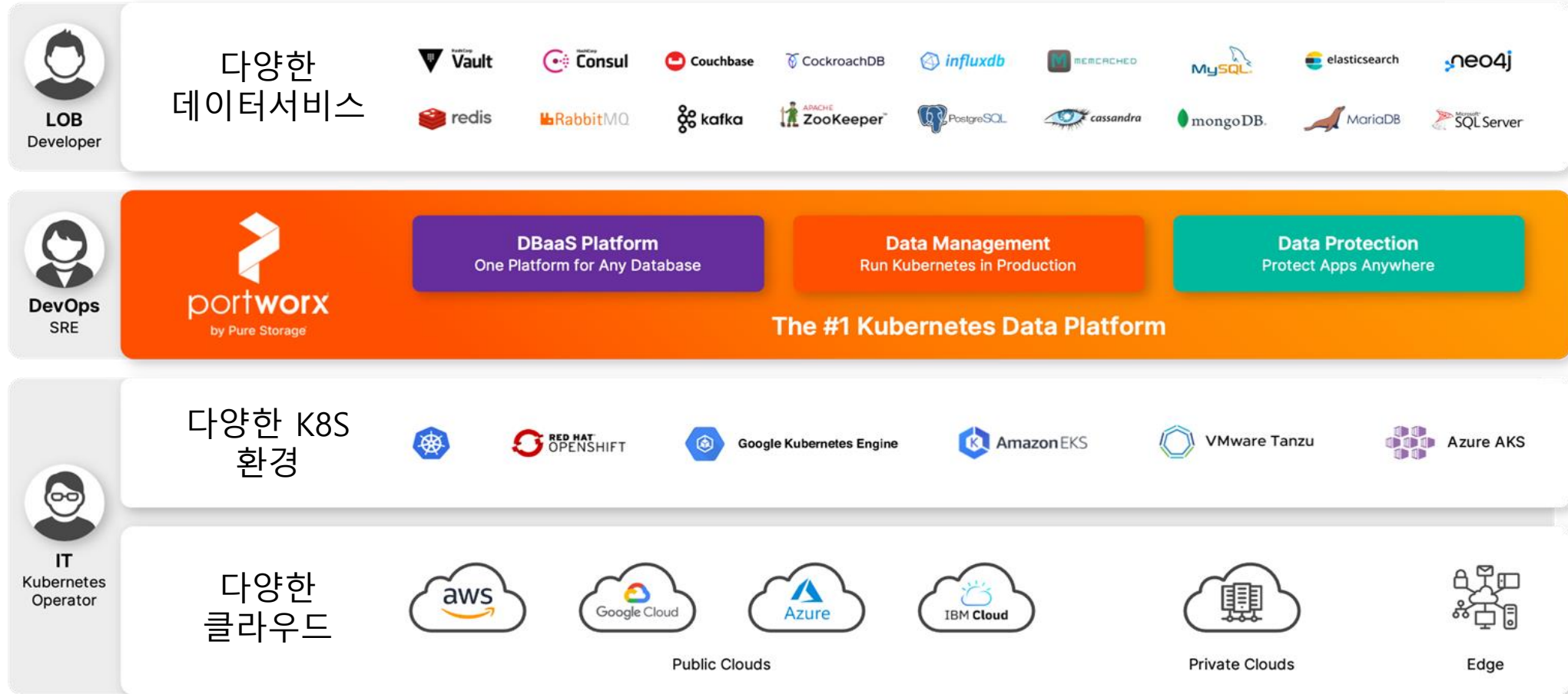
사용자 요구사항



가장 어려운 도전 과제



No.1 Kubernetes 데이터 플랫폼 Portworx



Portworx 특징

Portworx Enterprise

소프트웨어 기반 플랫폼 표준화

- ✓ 소프트웨어 정의 기반의 유연한 구성
- ✓ 다양한 클라우드, K8S 지원

성능 최적화

- ✓ 애플리케이션 인식으로 최적의 성능
- ✓ SLA 별 스토리지 그룹화

비즈니스 연속성 보장

- ✓ 복잡한 재해 복구 환경을 간편하게 구성
- ✓ 데이터 서비스에 대한 고가용성 및 복원성 구현

Portworx Data Services

단일 플랫폼에서 실행

- ✓ 다양한 종류의 데이터베이스를 벤더 락-인 없이 모든 인프라에서 쉽게 사용

단일 API를 통한 오퍼레이션

- ✓ 인프라에 상관없는 단일 워크플로우 및 API 제공

단일 솔루션을 통한 관리

- ✓ 엔터프라이즈 수준의 데이터 서비스의 모든 라이프사이클 관리

Portworx Backup

셀프 서비스 백업

- ✓ 역할 기반 액세스 제어를 통한 셀프 서비스 제공
- ✓ 애플리케이션 단위의 유연한 백업

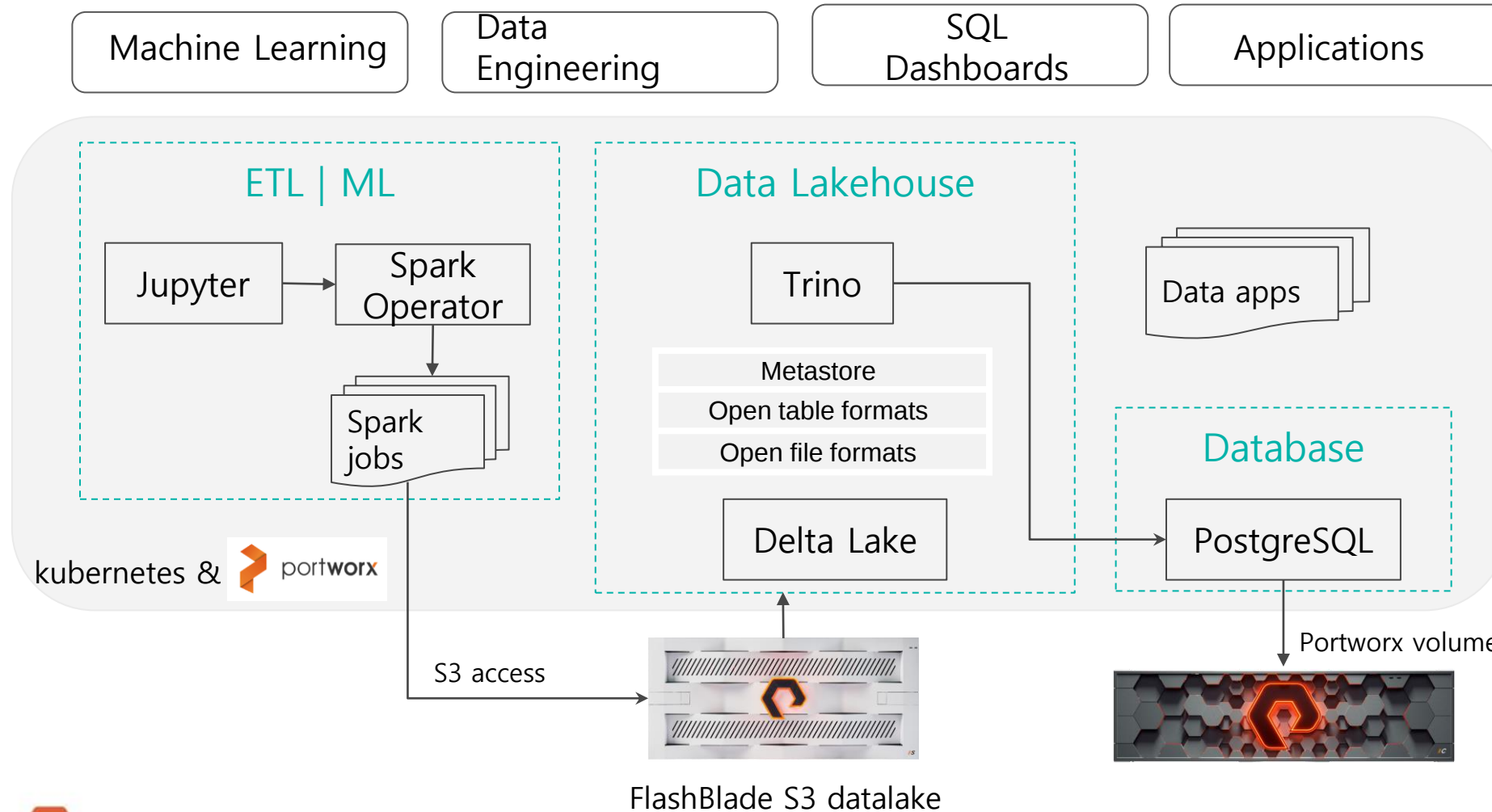
몇 분 만에 애플리케이션 복구

- ✓ 단 몇 번의 클릭만으로 애플리케이션의 백업/복구

사이버 공격 방어

- ✓ Object Lock을 통한 백업 데이터의 안전한 보호

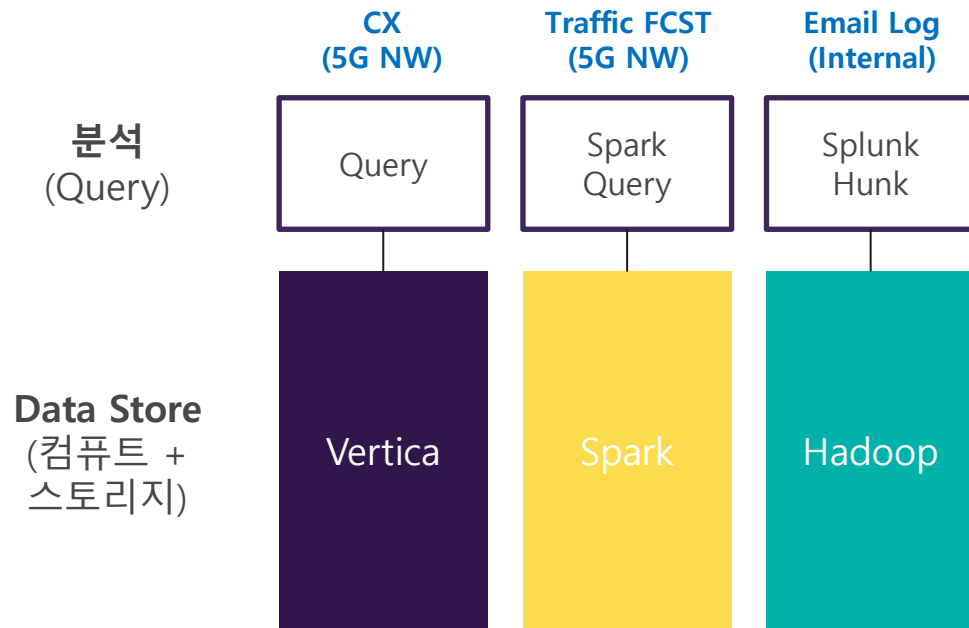
Pure storage를 이용한 Open Data Lakehouse 구성



- 벤더 비종속적
- 비용 효율적
- 데이터 확장 대비
- 단순한 구성
- 클라우드 레디

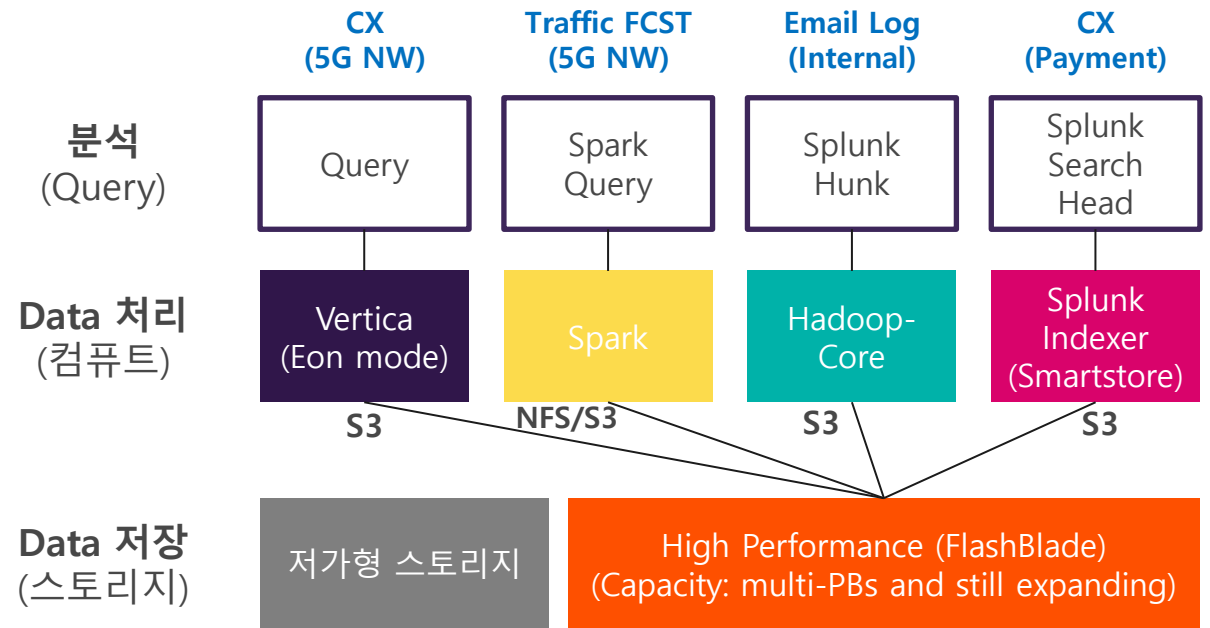
Pure storage 구축 사례

대형 통신사의 통합 데이터 레이크하우스



도전 과제

- DAS 아키텍처로 인해 성능과 확장성이 한계에 도달
- 복잡한 운영 및 장애 처리로 어려움
- 데이터 증가에 따라 고성능 및 확장 가능한 스토리지 필요



개선 효과

- 컴퓨팅과 스토리지를 분리하여 PB급 고성능 시스템 구축
- 여러 분석 애플리케이션을 단일 플랫폼으로 통합 및 사일로 제거
- Evergreen 프로그램을 통한 미래형 유니파이드 데이터 레이크

Pure storage 구축 사례

인공지능 학습 및 개발을 위한 슈퍼클러스터

고객 환경

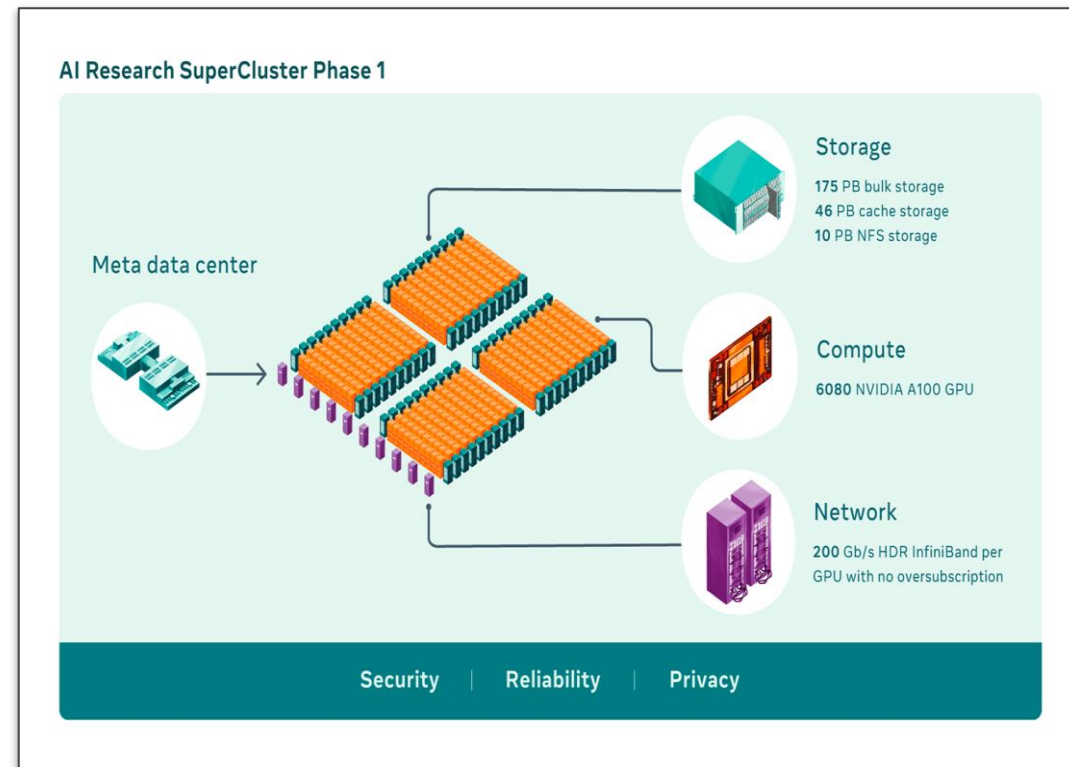
- 기존 AI 클러스터 성능 개선 (사물 인식, 음성 인식, 음성 분석 등)
- 760 NVIDIA DGX A100 시스템, 6080 GPUs
- 200 PB 이상의 블록/파일/오브젝트 스토리지 서비스

도전 과제

- 대량의 Write 에 대한 고성능 보장
- AI 모델링 시간 단축
- 운영 비용 절감 및 확장 유연성 확보

개선 효과

- 클라이언트에서의 메타데이터 처리 성능 향상
- GPU 사용율 2배 이상 향상
- NVIDIA collectives (NCCL) 성능 9배 향상
- 대용량 NLP 워크플로우 성능 3배 향상



Pure Storage 포트폴리오

모던 데이터 경험을 통한 인프라 및 애플리케이션 현대화를 구현하는 퓨어스토리지의 포트폴리오

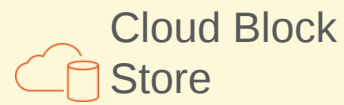
애플리케이션 현대화



K8s 애플리케이션을 위한
하이브리드 데이터 관리



K8S를 위한 DBaaS



하이브리드
모빌리티



AI / 분석 업무를
위한 초고속
파일 & 오브젝트

운영 관리 현대화



AI기반 스토리지
운영 관리 자동화



서브스크립션 기반
Storage as a Service



Storage as Code

인프라스트럭처 현대화



초 고성능 Tier-0
대 단위 업무 통합



미션 크리티컬
업무용
블록 스토리지



대용량 유니파이드
스토리지



디지털 혁신을 위한 스마트 데이터 플랫폼, Pure Storage

시장선도 기술력

- ✓ Flash 최적화 소프트웨어/하드웨어
- ✓ 9년 연속 Gartner 평가 1위
- ✓ 7년간 유일하게 성장하는 벤더

최고의 사용자 경험

- ✓ 엔터프라이즈 데이터 서비스
- ✓ 클라우드 네이티브 데이터 서비스
- ✓ 고객 만족도 1위 (NPS +85)

비용절감

- ✓ 전력/상면 절감
- ✓ 다양한 소비 모델 (구매/구독)
- ✓ 지속 가능 IT 실현



Gartner 9년 연속 1위
ESG 2022 award winner



홈페이지 : www.purestorage.com/kr

페이스북 : www.facebook.com/purestoragekorea

블로그 : www.blog.naver.com/purestorage_korea

유튜브 : www.youtube.com/@PureStoragekr



TECH WEBINAR

Pure Orange Shot

2023 퓨어스토리지 웨비나 시리즈