

# 데이터 분석 및 AI, 성공적인 미래 전략을 위한 차세대 데이터 플랫폼



김진환

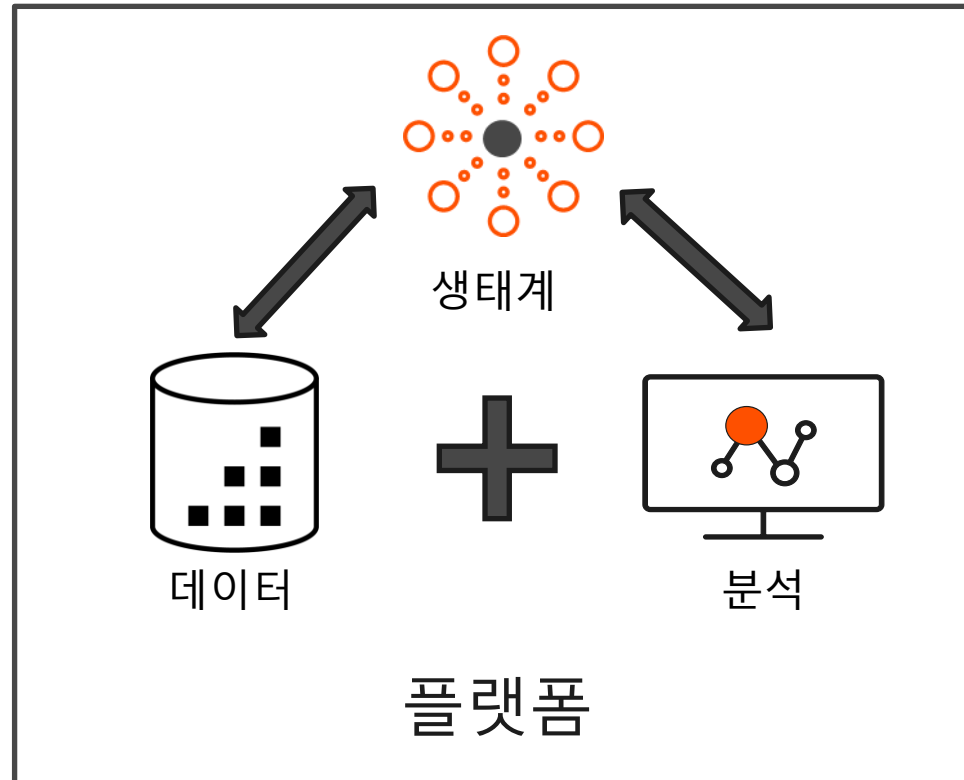
System Engineer, Pure Storage

2022.12.08

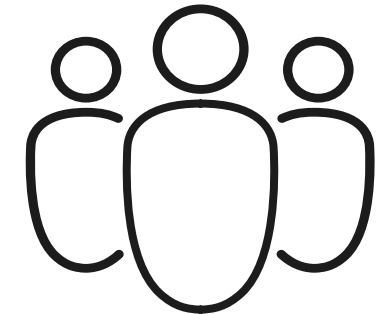
# 금융 환경의 변화



금융기관



Banking  
As a  
Service



소비자



# 금융 기업 도전 과제

## Digital 금융 전환

- 빅테크, 핀테크 경쟁  
우위 확보
- 실시간 지급결제 확대

## 디지털 화폐 대응

- 가상 화폐 연동
- 블록 체인 기술 내재화

## 에코시스템 확보

- 플랫폼 구축 및 파트너  
협력
- 새로운 매출 기회 확보

## 고객 경험 향상

- 애플리케이션 개발
- 개인화된 경험 제공
- 고객 충성도 향상

## 사이버 보안 강화

- 고객 데이터 보호
- 랜섬웨어 대응

## 지속 가능 경영

- ESG 이니셔티브
- 리스크 관리 연동



# 디지털 혁신을 위한 동력



비즈니스  
분석



사이버 보안



클라우드



애플리케이션  
현대화 및  
DevOps



고객 경험



인공지능(AI)  
머신러닝(ML)



**4.8배**  
매출 증대 가능성


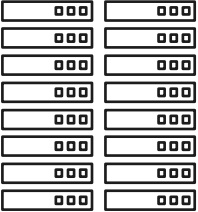
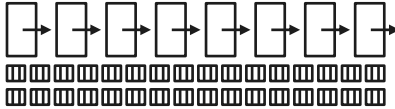
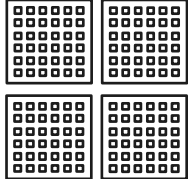
**3.2배**  
고객 만족도 향상

**2.4배**  
업무 생산성



# 데이터 분석 플랫폼 통합의 어려움

다양하고 상이한 워크로드와 프로토콜 제약으로 플랫폼의 사일로화 및 많은 리소스 낭비

	데이터웨어하우스	데이터레이크	스트리밍분석	AI 클러스터
인프라 특징	 <p>어플라이언스</p>	 <p>DAS</p>	 <p>마이크로 세분화(컨테이너, VM)</p>	 <p>HPC SYSTEM W/ GPU</p>
데이터 타입	정형데이터	비정형데이터	비정형데이터	비정형데이터
분석 작업 타입	배치처리	배치처리	마이크로 배치처리 / 실시간	실시간
I/O 타입	Sequential Write, Random Read	Sequential	Random	Sequential to random
아키텍처	스케일업 / 아웃	스케일아웃	다차원 성능	대규모 병렬분산처리

# 데이터 분석 플랫폼 도전 과제

BACKUP &  
DATA WAREHOUSE

DATA LAKE

STREAMING ANALYTICS

AI CLUSTER

## 가장 중요한 것은 데이터

- 다양하고 빠르게 변화하는 어플리케이션에 어떻게 서비스할 것인가?
- 빠르게 증가하는 데이터를 어떻게 저장할 것인가?
- 대용량 데이터를 어떻게 빠르게 분석할 것인가?
- 다양한 데이터 형태의 데이터를 어떻게 관리할 것인가?

KEY  
REQUIREMENT  
FOR  
STORAGE



대량처리



스케일아웃

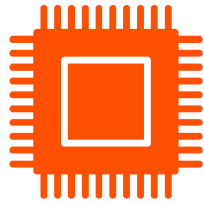


다양한 워크로드에  
고성능



초대용량  
병렬처리

# PureStorage 접근 전략



## Fast Object with File

대용량 데이터에 최적  
고성능 병렬 처리  
유연한 서비스 확대



## 컴퓨트/데이터 분리

자원 사용률 향상  
워크로드에 유연하게 대응  
비용 절감



## Cloud형 관리

Storage as a Service  
Storage as a Code  
쉽고 단순한 인터페이스

# 데이터 분석 및 AI를 위한 차세대 데이터 플랫폼

퓨어스토리지의 데이터 허브 플랫폼 - FlashBlade

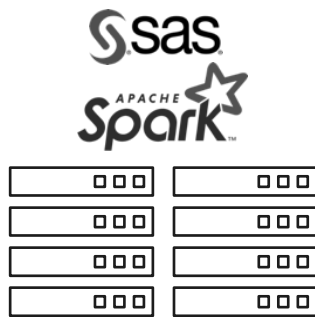
BACKUP



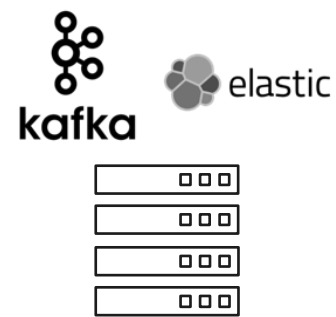
DATA WAREHOUSE



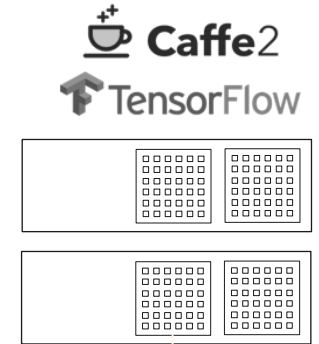
DATA LAKE



STREAMING ANALYTICS



AI CLUSTER



HIGH THROUGHPUT FILE & OBJECT

SIMPLE, NATIVE SCALE-OUT

MULTI-DIMENSIONAL PERFORMANCE

MASSIVELY PARALLEL



오브젝트 프로토콜 (S3)

파일 프로토콜(NFS, SMB)

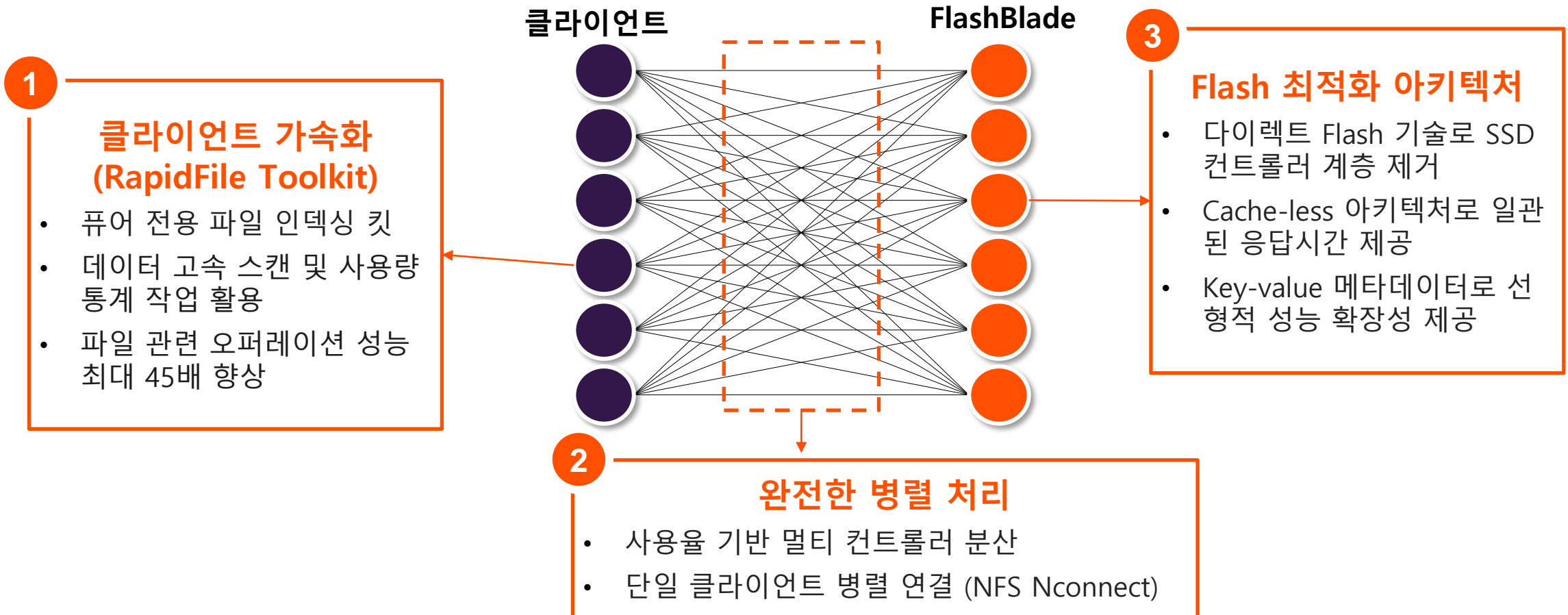
모든 데이터를 한번에 Access할 수 있는 Key Value 데이터베이스 스토어 ... 수십 억 개 파일을 실시간 병렬 처리





# PureStorage의 차별성

AI, 분석 환경에서 필요한 병목 없는 고성능의 데이터 조회 및 처리 성능 제공



# PureStorage의 차별성

차별화된 Flash 기술(소프트웨어, 하드웨어)로 더 높은 성능, 용량 집적도, 안정성 제공

PureStorage Direct Flash Module



1

## All NVMe (QLC)

- 업계 선도 NVMe 기술
- End-to-End NVMe
- 고집적 3D NAND

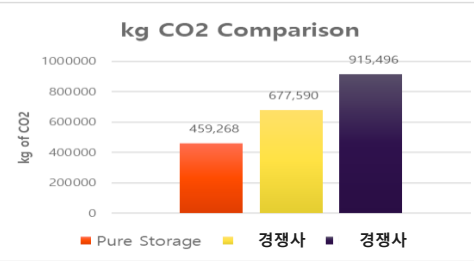
2

## Direct Flash

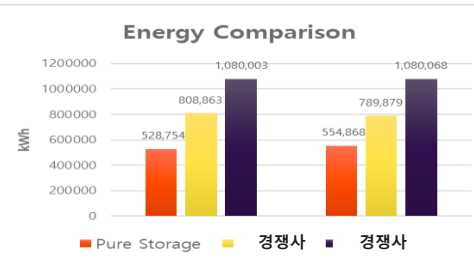
- 변환 계층 없는 Flash 직접 관리 기술
- 더 높은 중복 처리율
- 저전력 설계

더 높은 성능, 용량 집적도, 안정성 제공

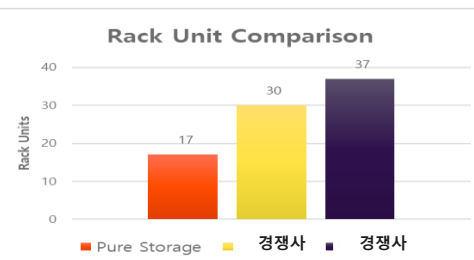
## 타사 SSD 스토리지 대비



**54%**  
DC Rack 상면 절감



**51%**  
에너지 소비 감소



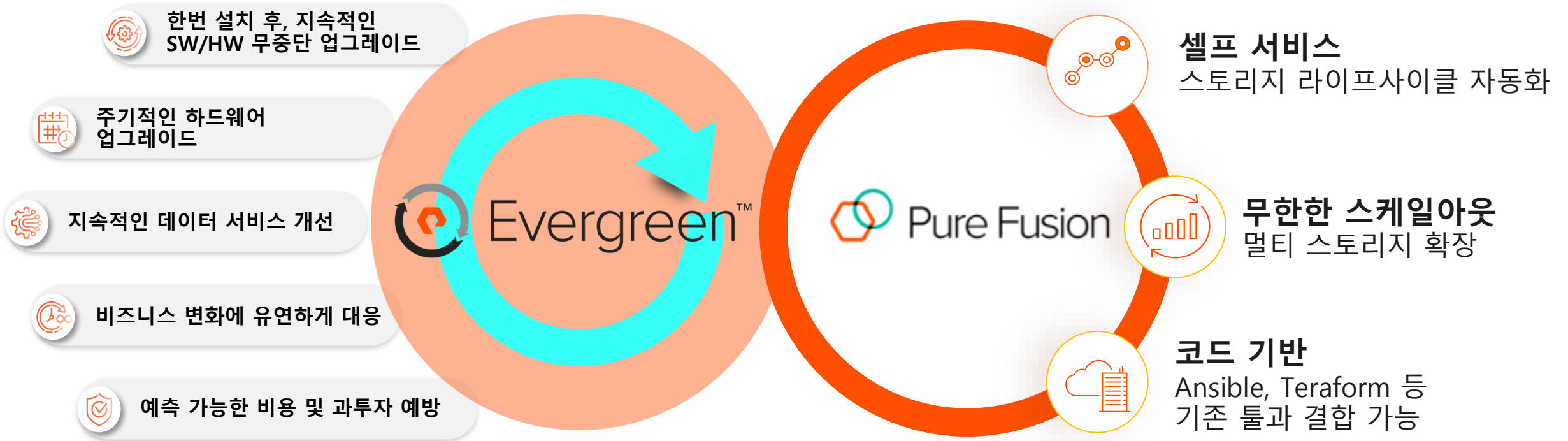
**50%**  
CO2 배출량 감소



# PureStorage의 차별성

클라우드 방식의 비용 모델 및 Code 기반의 셀프서비스로 퍼블릭 클라우드와 유사한 사용자 경험 제공

## 최신 소프트웨어와 하드웨어를 항상 유지



## 퍼블릭 클라우드의 사용자 경험

# PureStorage FlashBlade 적용 사례

데이터 분석 플랫폼을 통합하고 컴퓨팅, 스토리지 분리하여 데이터 준비 시간 단축 및 인프라 비용 절감

## ➡ 데이터 웨어하우스 고도화

### 고객 환경

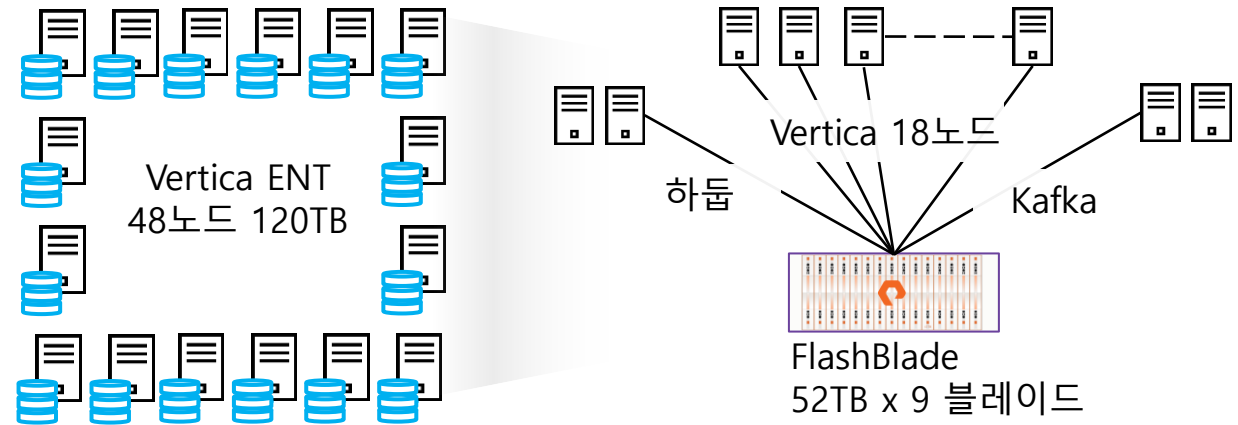
- 대형 금융사 데이터 웨어하우스
- Vertica 48 노드 (120TB) 구성

### 도전 과제

- 하드웨어 자원 사용률 효율화
- 하드웨어 장애 시 일관된 성능 보장 및 빠른 복구
- 유연하고 민첩한 노드 확장
- 카프카 및 하둡과 연계 분석

### 개선 효과

- 컴퓨팅 리소스 활용률 65% 개선(기존 48노드 → 18노드로 절감)
- 상면, 전력, 냉각 비용 절감 2배 절감
- 컴퓨팅, 스토리지 분리로 유연한 확장성 확보
- 데이터 플랫폼 통합으로 사전 작업 감소



# PureStorage FlashBlade 적용 사례

고성능, 고확장성의 분석 플랫폼 확보로 새로운 투자 모델 출시 시간 단축

## 실시간 분석 플랫폼

### 고객 환경

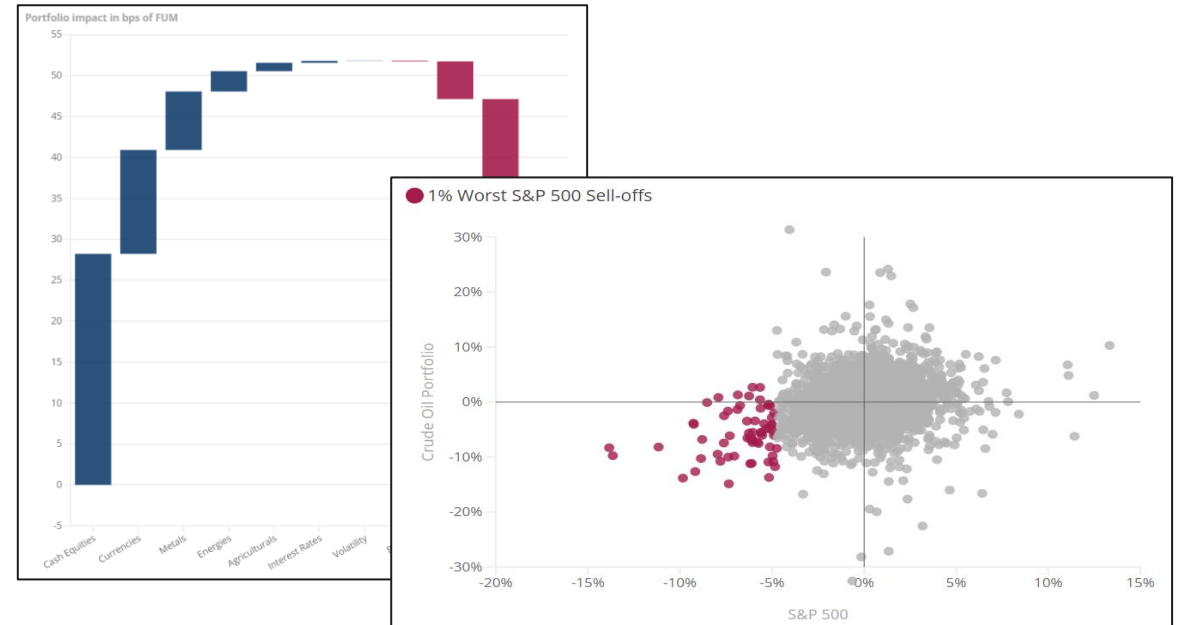
- 글로벌 투자회사 (총 자산 규모 1400조)
- Python, Java, Apache Spark 등을 이용한 투자모델 구축 및 실시간 분석

### 도전 과제

- 데이터 증가(2배/18개월)에 따른 성능/용량 확장성 확보
- 빠른 모델링을 위한 고성능 데이터 서비스 확보 (600K IOPS, 6GB/s 이상의 성능)

### 개선 효과

- 기존 플랫폼 대비 최대 10배 이상의 성능 향상
- 새로운 투자모델의 출시 시간 단축
- 투자자 수익 개선



**“10배에서 20배의 성능 향상은 당사의 시장 출시 기간 경쟁 우위를 창출하는 데 있어 판도를 바꿀 수 있습니다.”**

- Gary Collier  
CTO, Man Group

# PureStorage FlashBlade 적용 사례

백엔드의 데이터 처리 성능 향상과 클라이언트의 인덱싱 성능 개선으로 분석 성능 3배 이상 향상

## ➡ AI 리서치 클러스터 고도화

### 고객 환경

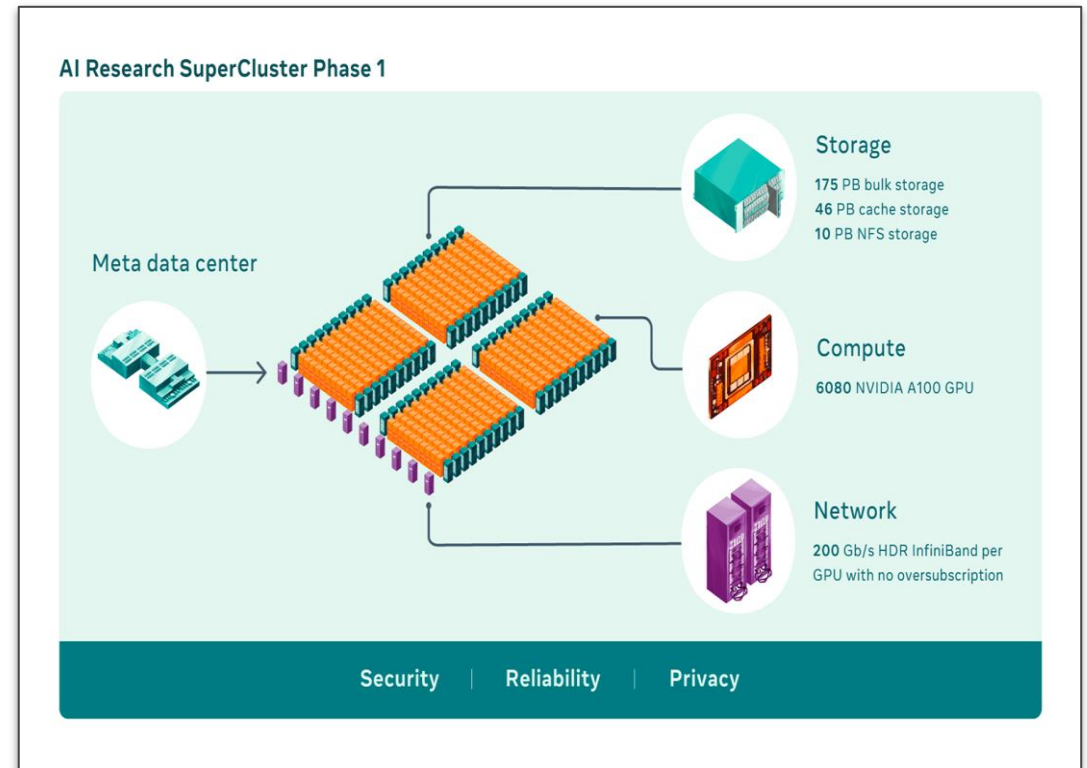
- 기존 AI 클러스터 성능 개선 (사물 인식, 음성 인식, 음성 분석 등)
- 760 NVIDIA DGX A100 시스템, 6080 GPUs
- 200 PB 이상의 블록/파일/오브젝트 스토리지 서비스

### 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



<https://ai.facebook.com/blog/ai-rsc>

# AI을 위한 차세대 데이터 플랫폼, PureStorage

Flash 최적화 기술

전용 하드웨어



FILE OBJECT



PURESTORAGE® FlashBlade® FB

데이터 분석, AI를 위한  
고성능 Unified File and Object 플랫폼



## 통합 데이터 플랫폼

데이터 서비스 확산이 용이한 API 기반 오브젝트 서비스와 멀티 프로토콜 기능 제공



## 고성능 데이터 서비스

초고속 병렬 처리와 전용 클라이언트 툴킷을 이용하여 빠른 데이터 학습 성능 등 대용량 데이터 처리에 최적화



## 유연한 확장 및 비용 절감

워크로드 변화에 따라 컴퓨팅과 스토리지를 탄력적으로 확장하고 비용 효율적으로 저장



## 데이터 보호 레벨 향상

고속의 백업/복구 성능으로 복구 시간 단축 및 SafeMode (Immutable Snap)를 이용한 랜섬웨어 대응





Uncomplicate Data Storage, Forever