

PURE// ACCELERATE DIGITAL KOREA 2021

**BREAKTHROUGH**

상상을 초월하는 혁신

# 디지털 혁신(DT) 성공을 위한 데이터 파이프라인 현대화

송성환 부장  
Data Architect  
sesong@purestorage.com  
Pure Storage Korea

#PUREACCELERATE

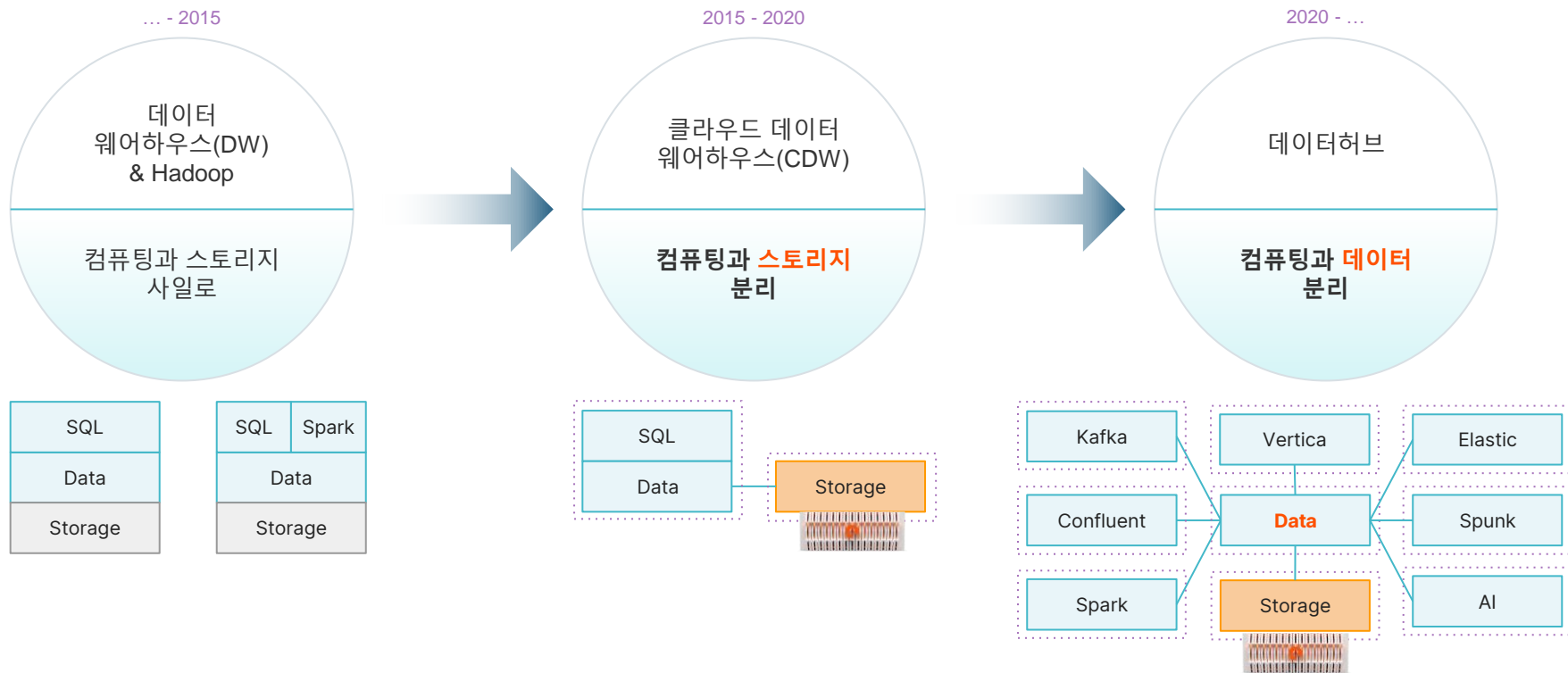


# Session Agenda

1. Pure Storage와 함께하는 디지털 혁신(DT)
2. UFFO를 통한 다양한 애플리케이션 개선 효과
3. 요약

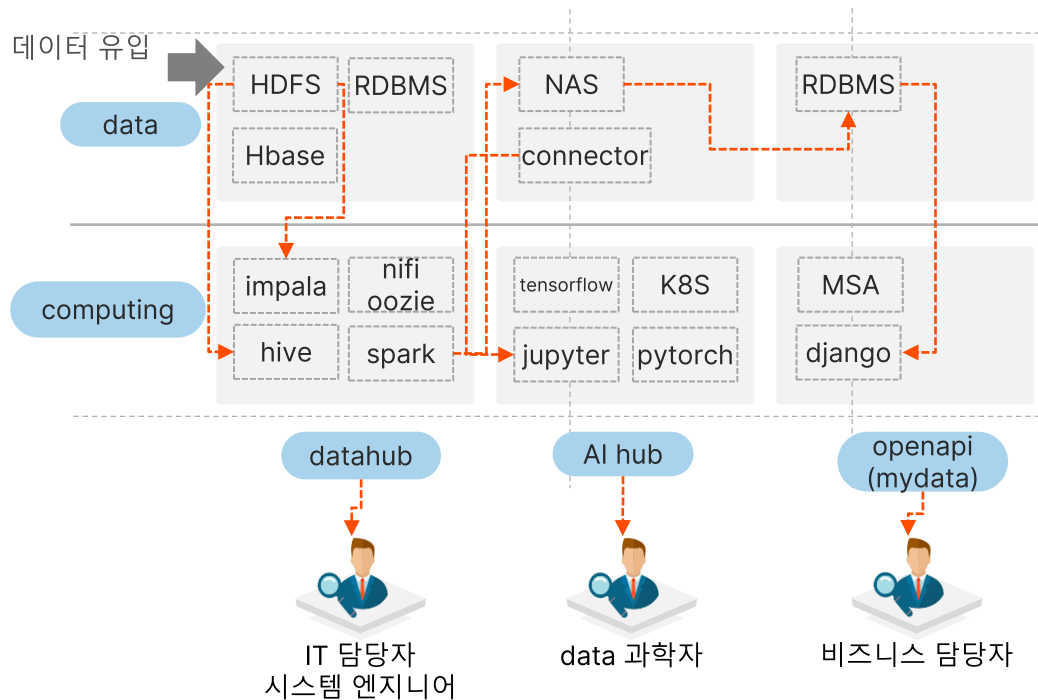
# Pure Storage와 함께하는 디지털 혁신

# 데이터 생태계의 변화



# 디지털 혁신(DT)의 어려움

현재의 조직과 인프라 환경으로는 기업의 비즈니스 전환 불가

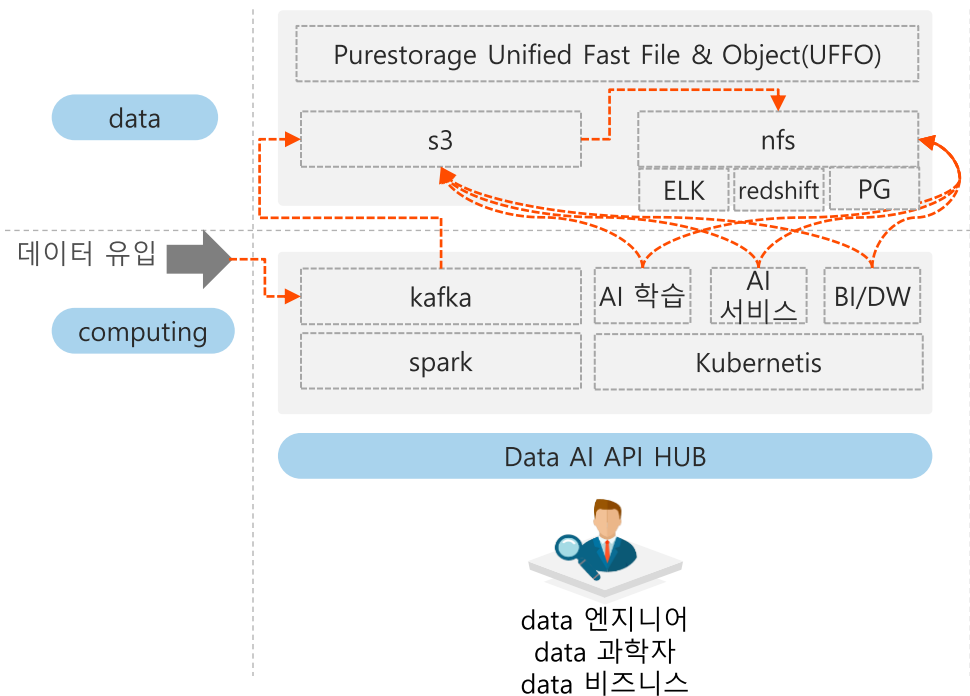


## 기업 DT의 어려움

- 연계된 많은 소프트웨어를 제대로 지원할 기술 인력 부족
- 늘어나는 하드웨어 인프라로 인해 하드웨어 에러 발생 증가
- 여전히 존재하는 부서간 데이터 사일로
- 끝나지 않는 데이터 처리 및 트레이닝 시간
- 불분명한 조직의 R&R로 협업의 어려움
- DT의 목적이 비즈니스 전환이라는 인식 부족
- 인력을 현재 필요한 기술력으로 전화하는 비용과 시간이 너무 큼

# 아키텍처의 단순화가 DT의 핵심

대량의 데이터를 처리할 수 있는 간단한 아키텍처 설계로 DT 고민 해결

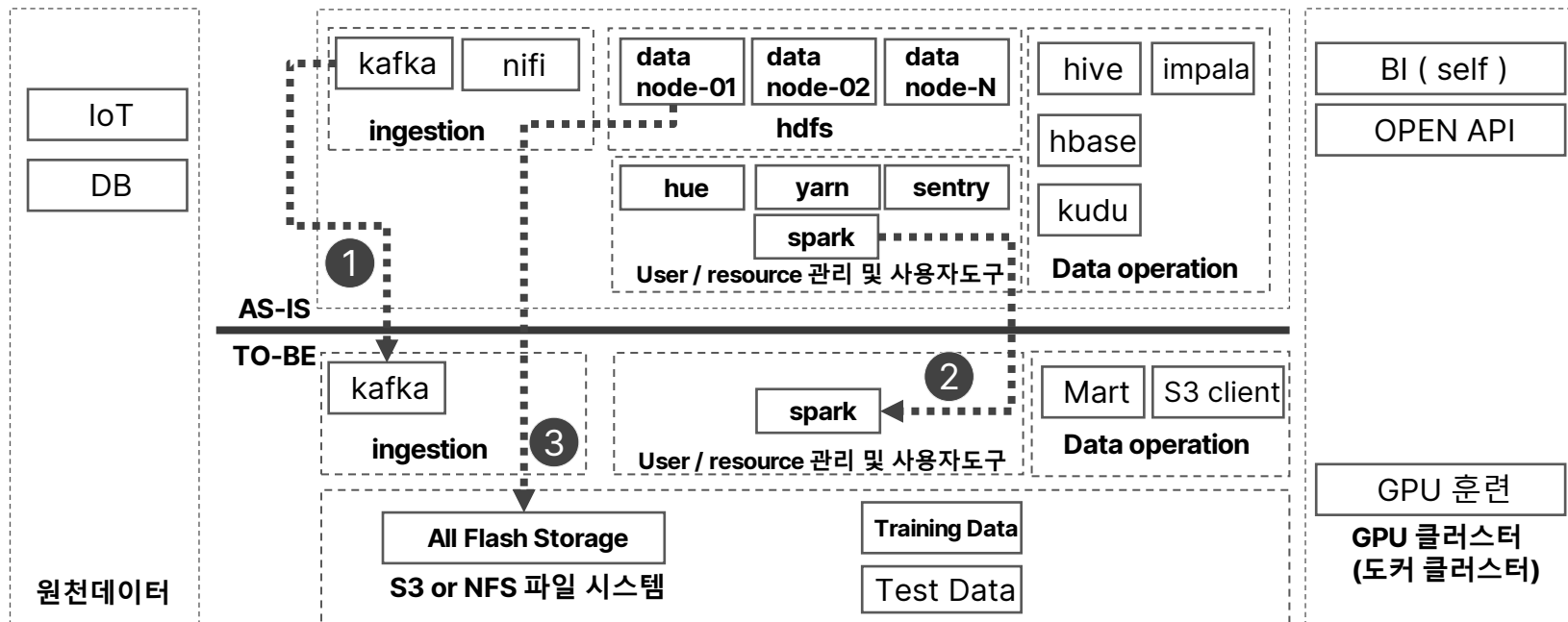


## DT을 위한 새로운 아키텍처

- 필요한 소프트웨어만 완성형으로 사용  
→ 소프트웨어 유지보수 및 인력 최소화
- 하드웨어 간소화  
→ 오류 발생률 감소, 간편한 관리, 상면 절감
- 데이터 처리 및 인공 지능 학습 속도 향상  
→ 실시간 데이터 처리, 실시간 인공지능 모델
- 스트리밍 아키텍처  
→ 불필요한 데이터 저장 최소화
- 비즈니스에 집중된 조직 구성 가능

# DT를 위한 새로운 아키텍처

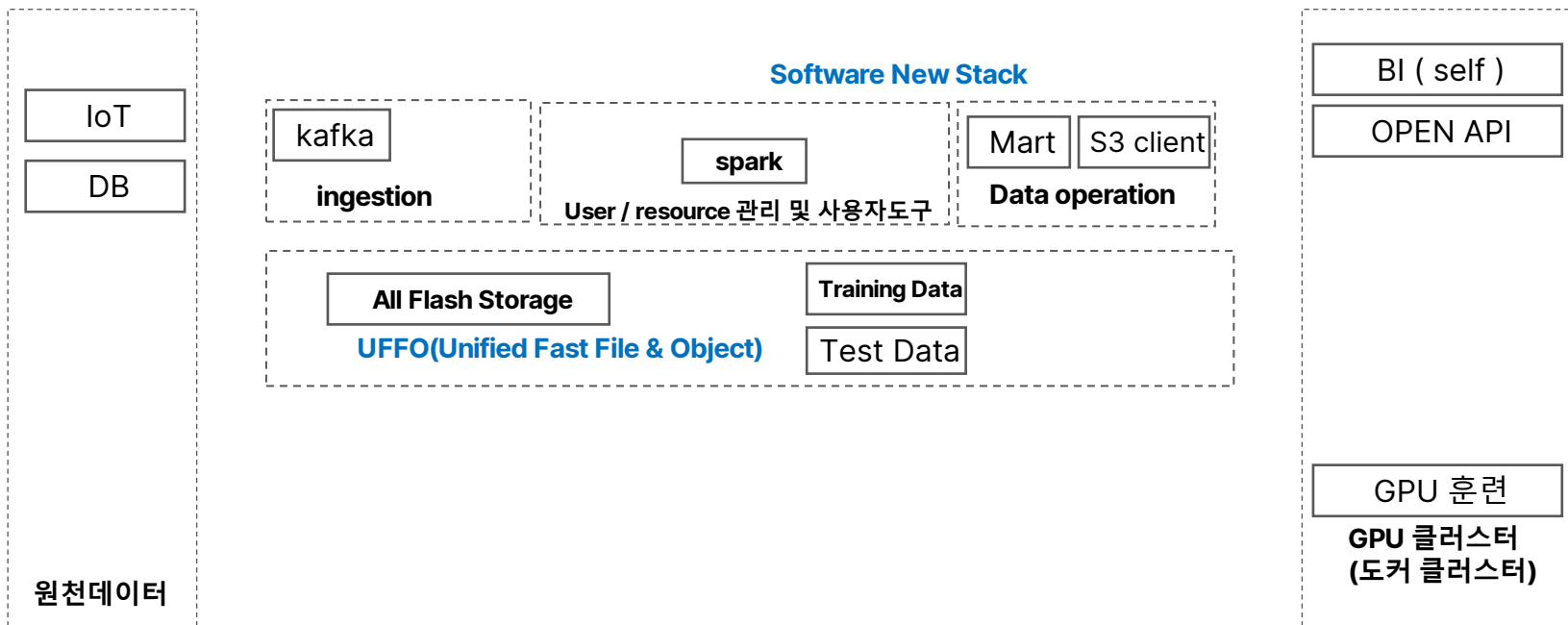
복잡하고 비효율적인 하둡 에코시스템에서 벗어난 새로운 소프트웨어 스택과 전용 하드웨어





# 데이터 중심 아키텍처의 중요성

소프트웨어를 간소화하고 과거 소프트웨어 기능은 발전된 하드웨어에 이관하여 비즈니스 집중도 향상



# 데이터 활용이 기업의 미래

현대화된 데이터 활용은 비즈니스 성공의 원동력

## TODAY

자유로운 데이터 접근과 활용의 어려움



업무별 데이터 분리, 관리의 어려움



최신 데이터 및 애플리케이션의 성능  
요건 충족 불가



제한된 데이터 공유 및 재활용 어려움



비효율적인 스토리지 및 컴퓨팅 자원  
사용률

## FUTURE

UFFO (Unified Fast File and Object)  
플랫폼은 디지털 시대 필수 인프라

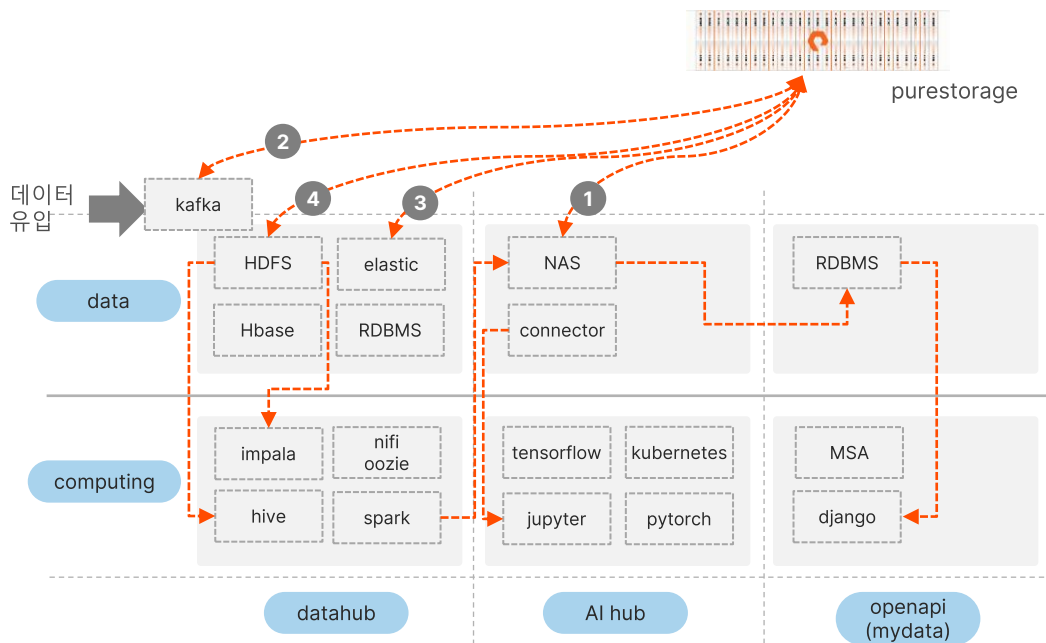
**4.8x** 수익향상  
- 신제품 수익에서 경쟁사를 능가할  
가능성이 4.8배 이상 높음

**3.2x** 고객 만족도 향상  
- 사전 분석투자 보고 받을 가능성이  
3.2배 이상 높음

**2.4x** 운영 우수성 향상  
- 직원 당 매출 증가 가능성이 2.4배  
이상 높음

# 신기술 적용을 위한 접근 방법

불필요한 데이터 이동 및 로딩으로 소비되는 시간 단축, 탁월한 성능 개선



## 단계적 접근 방법

1. 느린 인공지능 플랫폼 개선을 위해 고성능 분산 병렬 스토리지 고려

→ 업계 최초 AI 전용플랫폼(AIRI, FlashStackAI)

2. 카프카 등의 스트림 데이터 처리 및 영구 데이터 보존

→ 업계 유일의 Confluent 인증 플랫폼 FlashBlade

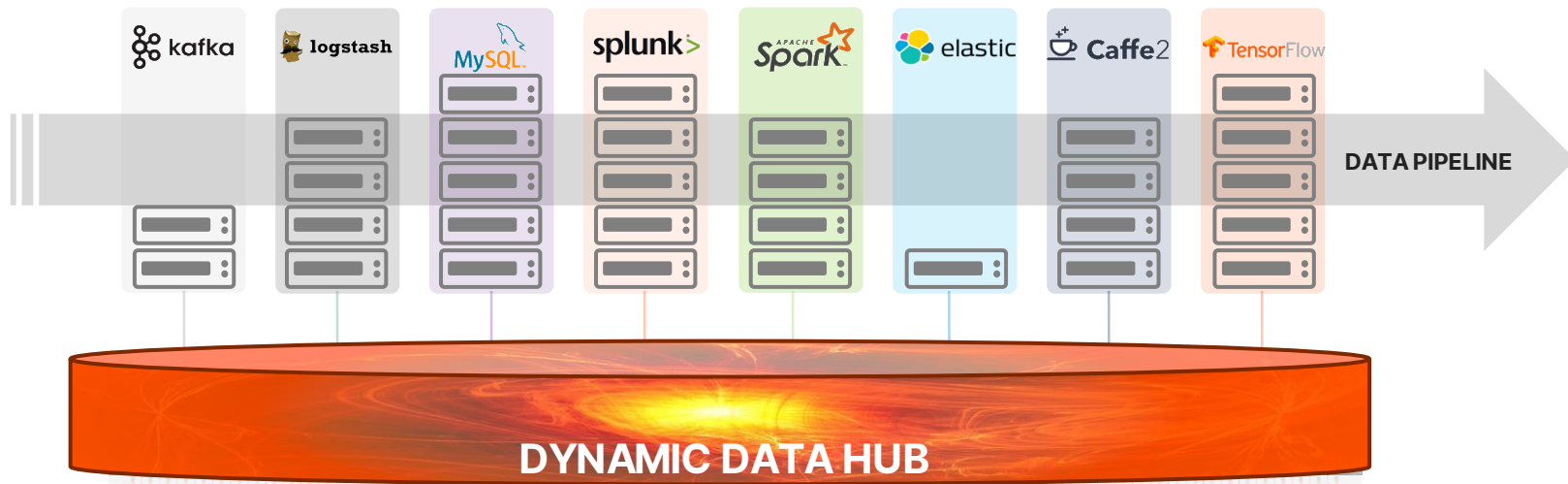
3. Elastic을 이용한 모니터링 데이터 확대 및 성능 개선

4. 비효율적인 HDFS를 S3로 대체

→ 자유로운 CS 독립 확장, 자원 활용률 향상

# 차세대 데이터 파이프라인 아키텍처

빠른 성능, 간편한 데이터 통합 플랫폼 구현



## 모든 데이터 워크로드에 최적화

Small, Random to Large, Seq.  
Architected for the Unknown

## 실시간 처리

Low Latency Performance  
for Instant Response

## 초병렬분산

No Serial Bottlenecks  
for Max Throughput

## CS분리

Grow Non-Disruptively  
by Disaggregating Compute & Storage

## 인프라 단순화

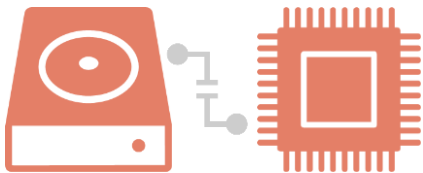
Focus More on Insights,  
Not Infrastructure

# UFFO 플랫폼을 활용한 애플리케이션 개선 효과

- Kafka
- Elastic
- Hadoop
- Spark
- Vertica
- Splunk

# FlashBlade를 활용한 현대적 아키텍처 구현

컴퓨팅과 데이터의 분리를 통한 최적화된 애플리케이션 환경 구축



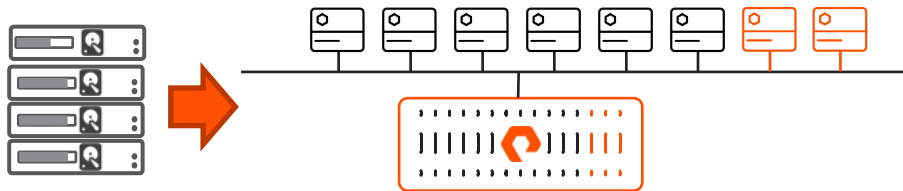
- Scalability
- Agility
- Simplicity



- **컴퓨팅 & 스토리지 분리**  
가용성 향상, 데이터 효율성 향상
- **멀티 워크로드 통합**  
워크로드 간의 간섭 및 장애 영향도 제거
- **다양한 업무 통합**  
개별 작업 성능을 위한 개별 클러스터 운영  
클러스터간 병목 제거
- **빠르고 자유로운 확장**  
다양한 워크로드를 위한 추가 확장 용이
- **하이브리드 클라우드**  
S3 복제 기반 퍼블릭 클라우드로 이관 및 확장

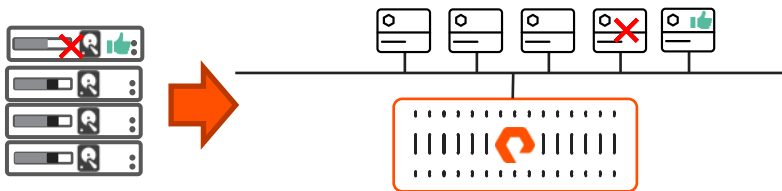
# 컴퓨팅과 데이터 분리 효과

유연한 비즈니스 확장, 빠른 장애 복구, 손쉬운 업그레이드



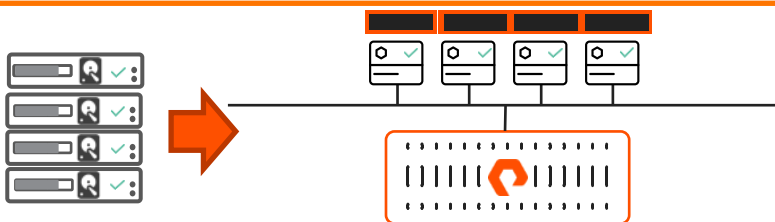
유연한  
확장

- 노드 추가 시 데이터 재배포 불필요
- 요구에 따라 독립적으로 컴퓨트와 스토리지 확장
- 빠른 무중단 확장
- 엔터프라이즈 레벨 데이터 보호



빠른  
장애  
복구

- 장애 노드 데이터 회수 불필요
- 노드 복구 시 데이터 재배포 불필요
- 노드 복구 시간 단축



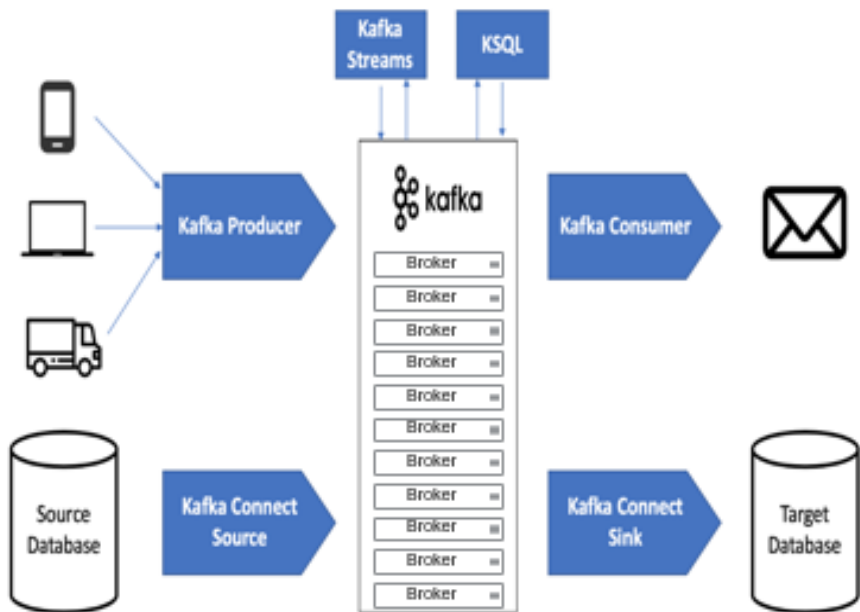
빠른  
업그레이드

- 업그레이드 시 노드별 데이터 이동 및 회수 불필요
- 컴퓨트 노드의 비활성/활성의 변경만으로 빠르게 애플리케이션 업그레이드

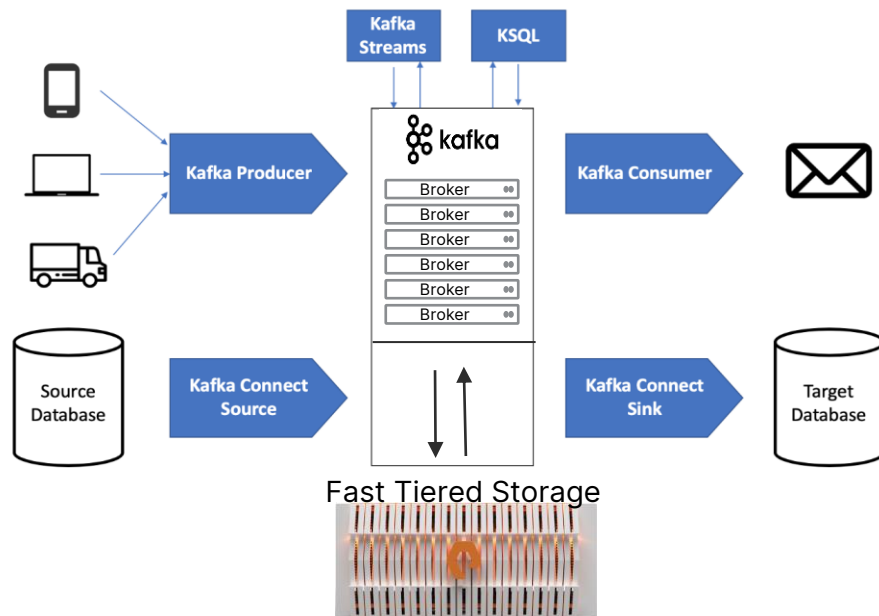
# Kafka 환경 개선

고성능의 S3 오브젝트 스토리지를 통해 쿼리 성능 개선 및 데이터 정책 규정 준수

## TODAY



## FUTURE





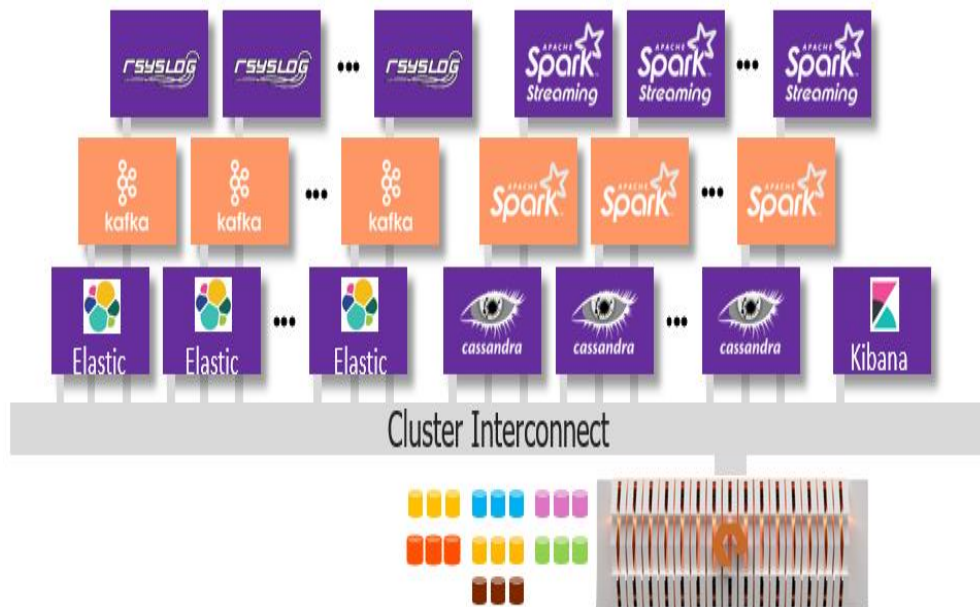
# Elastic 환경 개선

데이터허브로 데이터 활용성을 보장하고 디지털 혁신을 위한 데이터 파이프라인 완성

## TODAY



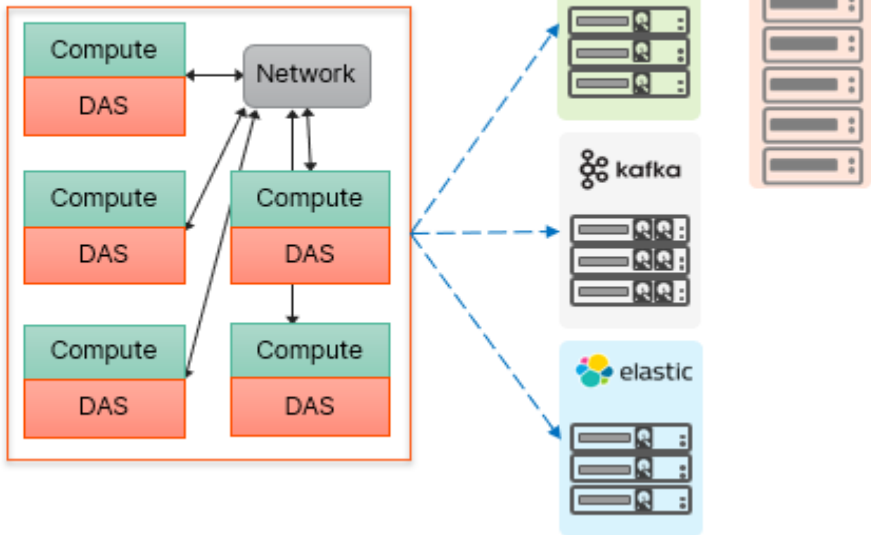
## FUTURE



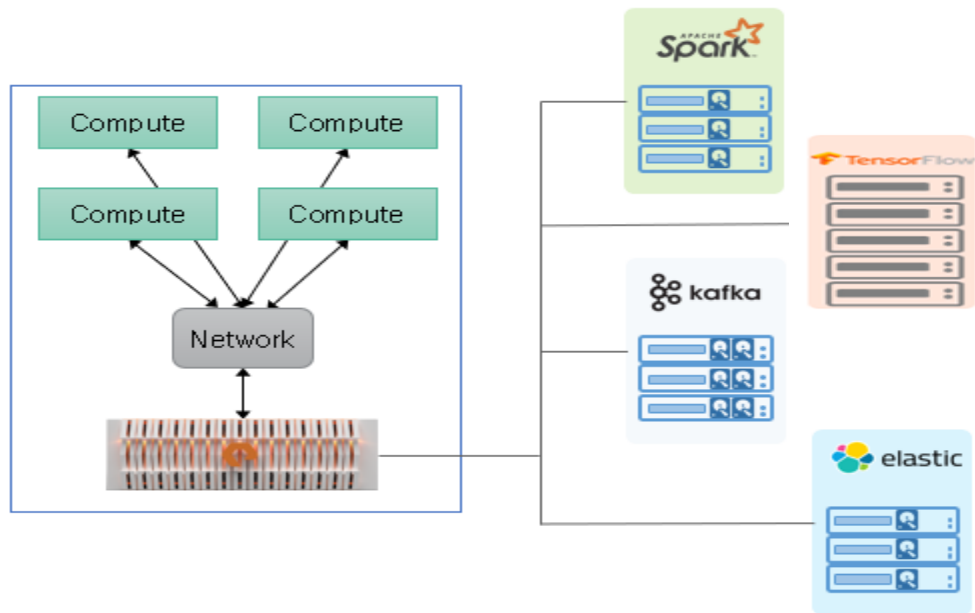
# Hadoop 환경 개선

컴퓨팅과 데이터의 분리를 통한 자원 활용률 향상 및 고성능 유지

## TODAY

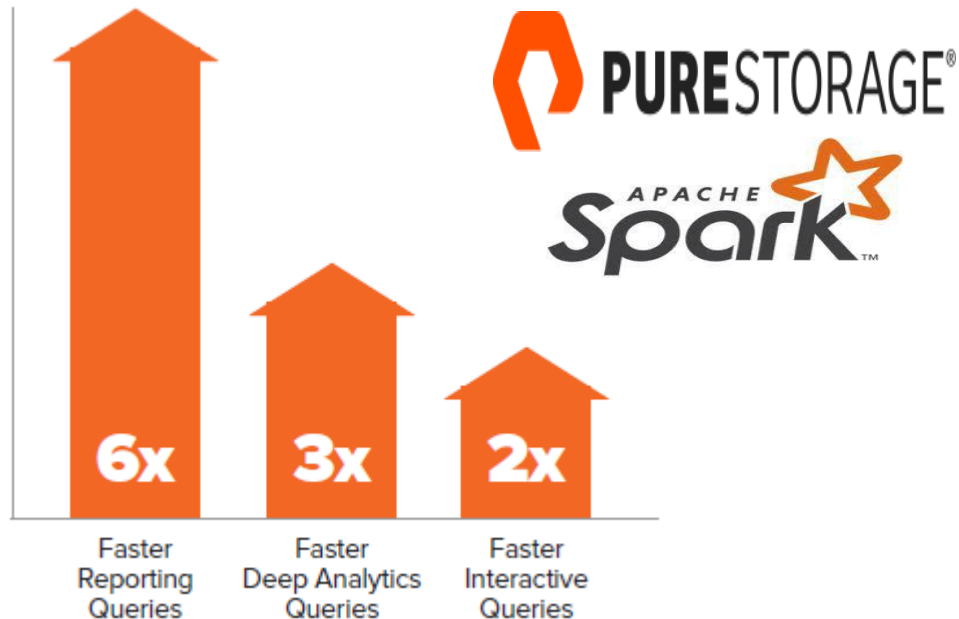


## FUTURE



# SPARK 성능 개선

대규모 데이터 처리환경에서 MapReduce보다 100배 이상 빠른 성능 제공

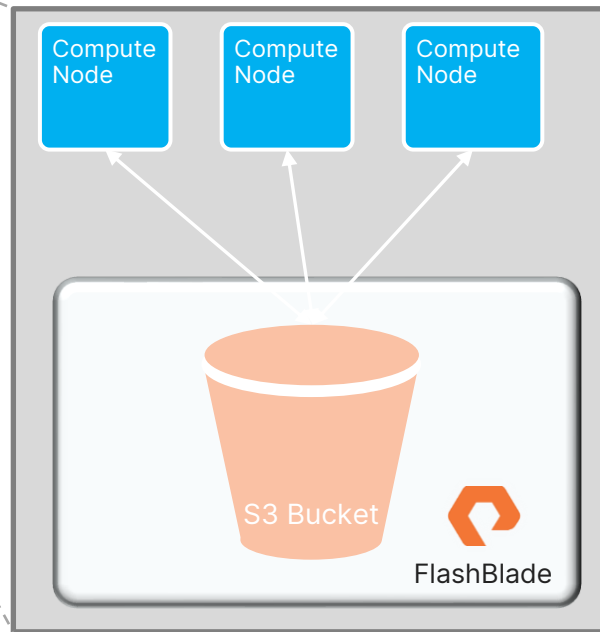
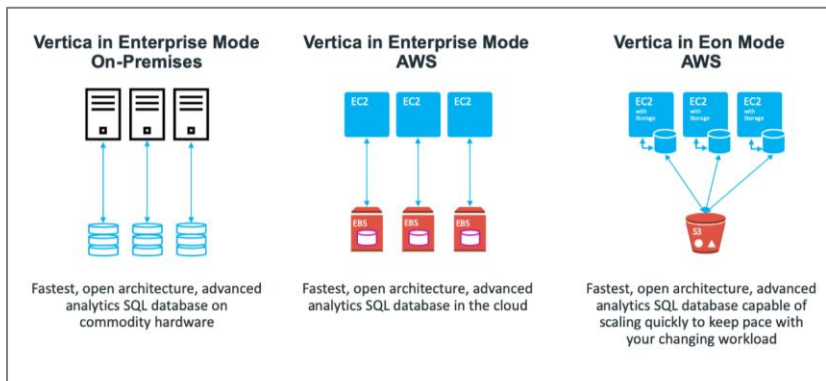


## Spark with FlashBlade 효과

- 성능 가속화
  - 타사 대비 6배 빠른 쿼리 성능
  - 3배 빠른 심층분석 쿼리, 2배 빠른 대화형 쿼리
- 유연한 확장 및 간편한 통합
  - 컴퓨팅과 스토리지 독립 확장
  - 전체 데이터 파이프라인 통합
- 다양한 시스템 및 파일 포맷 통합
  - Mesos, Kubernetes, Parquet, Hadoop Yarn 등과의 자유로운 연계
  - Rest API를 활용한 시스템 도구 통합

# Vertica Eon Mode with FlashBlade

업계 최초로 Vertica Eon 모드 인증



**VERTICA + PURE STORAGE**

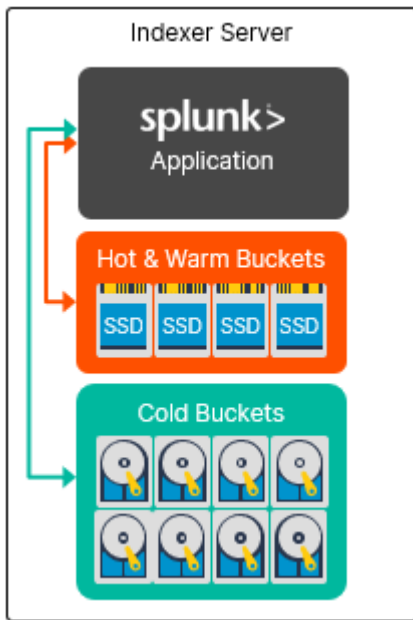
온프레미스 Vertica 구축을 위한

**업계 최초 C/S 분리 데이터 분석 플랫폼**

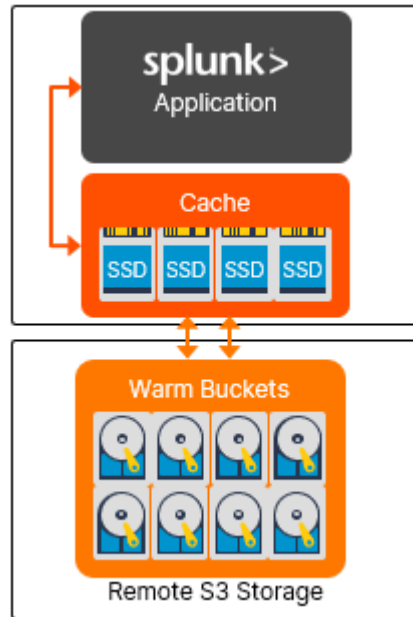
# Splunk 환경 개선

계층화된 데이터 레이어를 제거하여 검색 성능 및 자원 활용률 개선

## TODAY



## FUTURE



# AI/ML/DL 환경 개선

AI 설계부터 구축까지 해결한 올인원 패키지

UCS ML & FlashBlade

RA with 4xDGX & FlashBlade

RA with 2xDGX & FlashBlade



RA with Mellanox scalable to 64 racks of DGX w/ FlashBlades

RA A100 w/FlashBlade

AIRI Standard

AIRI "MINI"

FLASHSTACK AI

HYPERSCALE AIRI

# 요약

# Pure Storage = 디지털 혁신의 지름길

UFFO(Unified Fast File & Object)

FILE

OBJECT

FlashBlade<sup>®</sup> FB



## 데이터 허브

- 업계 최초 네이티브 파일 & 오브젝트 동시 제공
- 자유로운 접근성 및 확장성
- 모든 비정형 데이터 워크로드 통합
- 완벽한 데이터 보호(Safe Mode)

## 초고속 분석 플랫폼

- 비효율적인 하둡(Hadoop) 인프라를 고성능 S3로 현대화
- 다양한 분석 어플리케이션과의 완벽한 호환성
  - Confluent Tiered Storage
  - Vertica Eon Mode
  - Splunk SmartStore
  - ElasticSearch
- 클라우드와의 자유로운 연동

## AI 전용 플랫폼

- 업계 최초 AI 레퍼런스 플랫폼(AIRI, FlashStackAI)
- GPU와 스토리지 자원 통합 관리 모듈(Backend.AI)
- 업계 유일의 고성능 데이터 인덱싱 모듈 제공 (RapidFile)
- 완벽한 튜닝 패키지(Scaling Toolkit)



# Thank You!

---

## Contact us to continue the conversation

송성환 부장

Data Architect, APJ

[sesong@purestorage.com](mailto:sesong@purestorage.com)

공식 웹사이트

[www.purestorage.com/kr](http://www.purestorage.com/kr)

공식 유튜브

[www.youtube.com/c/PureStoragekr](http://www.youtube.com/c/PureStoragekr)

공식 페이스북

[www.facebook.com/purestoragekorea](http://www.facebook.com/purestoragekorea)

네이버 블로그

[blog.naver.com/purestorage\\_korea](http://blog.naver.com/purestorage_korea)

#PUREACCELERATE



