[TalkIT]

AI/HPC를 위한 데이터 레이크의 진화와 활용

효성인포메이션시스템
데이터사업팀 강준범 컨설턴트 / 김학균 부장

Keypoint

I. 데이터 발전에 따른 기업 분석 환경의 발전

II. AI 분석 환경을 위한 데이터레이크 필수 고려사항

III. HCSF (Hitachi Contents Software for File) 소개

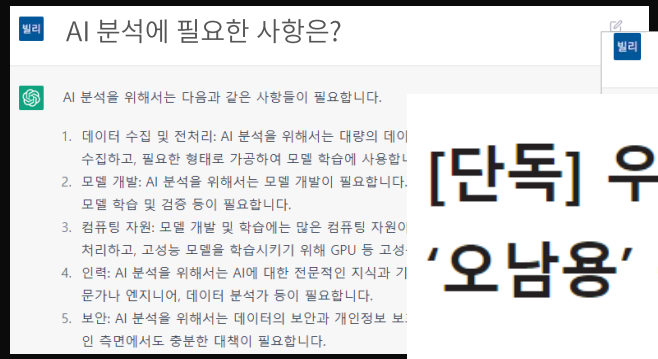
IV. 산업별 데이터 레이크 활용사례 : 바이오, 제조/자율주행, 금융, 미디어/엔터

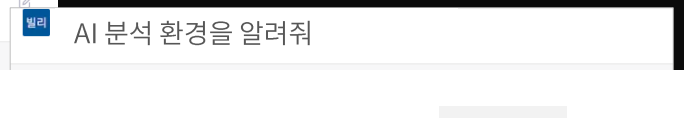
Dall-E & ChatGPT

- AI 알고리즘을 기반으로 한 서비스 → 모델 학습을 위한 방대한 양의 데이터와 분석 환경 필요

서비스

Insight



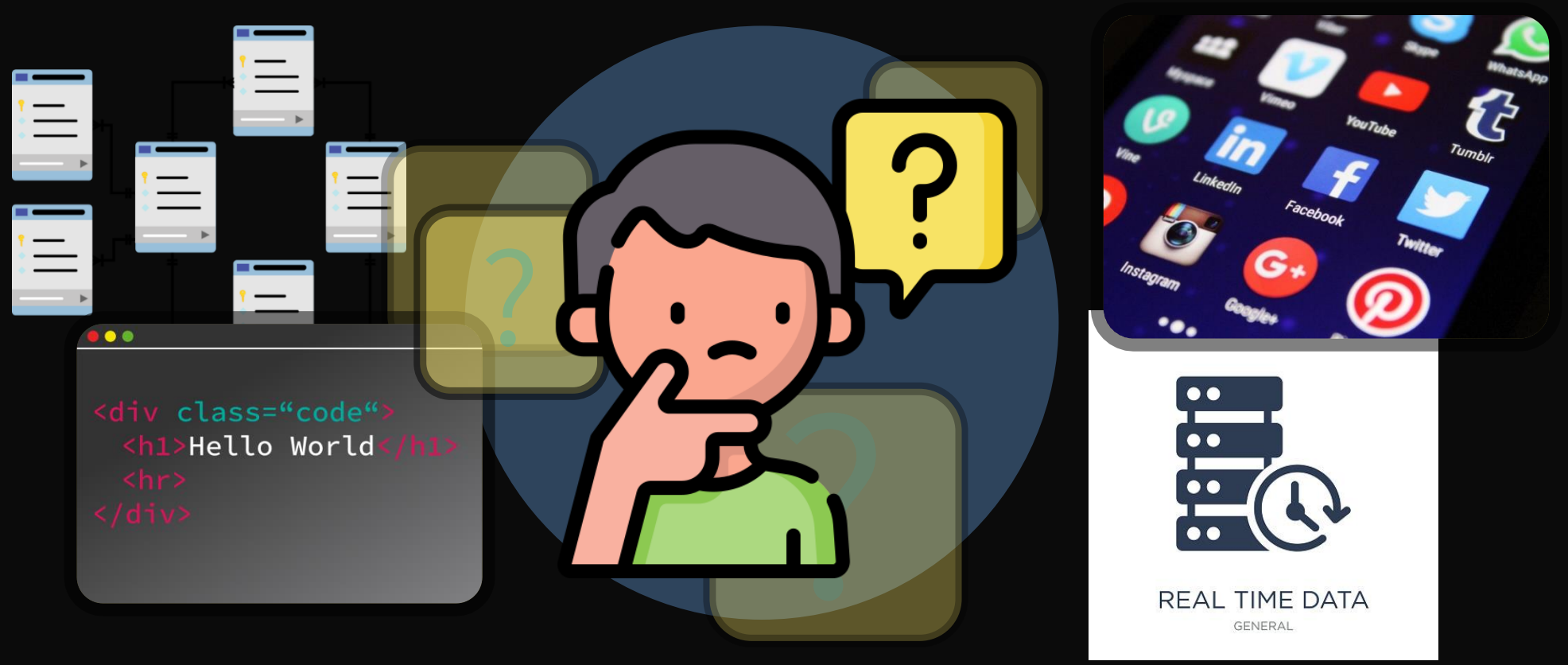


[단독] 우려가 현실로... 전자, 챗GPT 빗장 풀자마자 '오남용' 속출

> 반도체 사업장 챗GPT 허용 20일, 정보 유출 사고 3건 발생
> 설비 계측·수율 데이터, 미국 기업에 고스란히 전송...회수 불가
> 필요시 해당 임직원 징계.. 사내 전용 AI 서비스 구축 검토

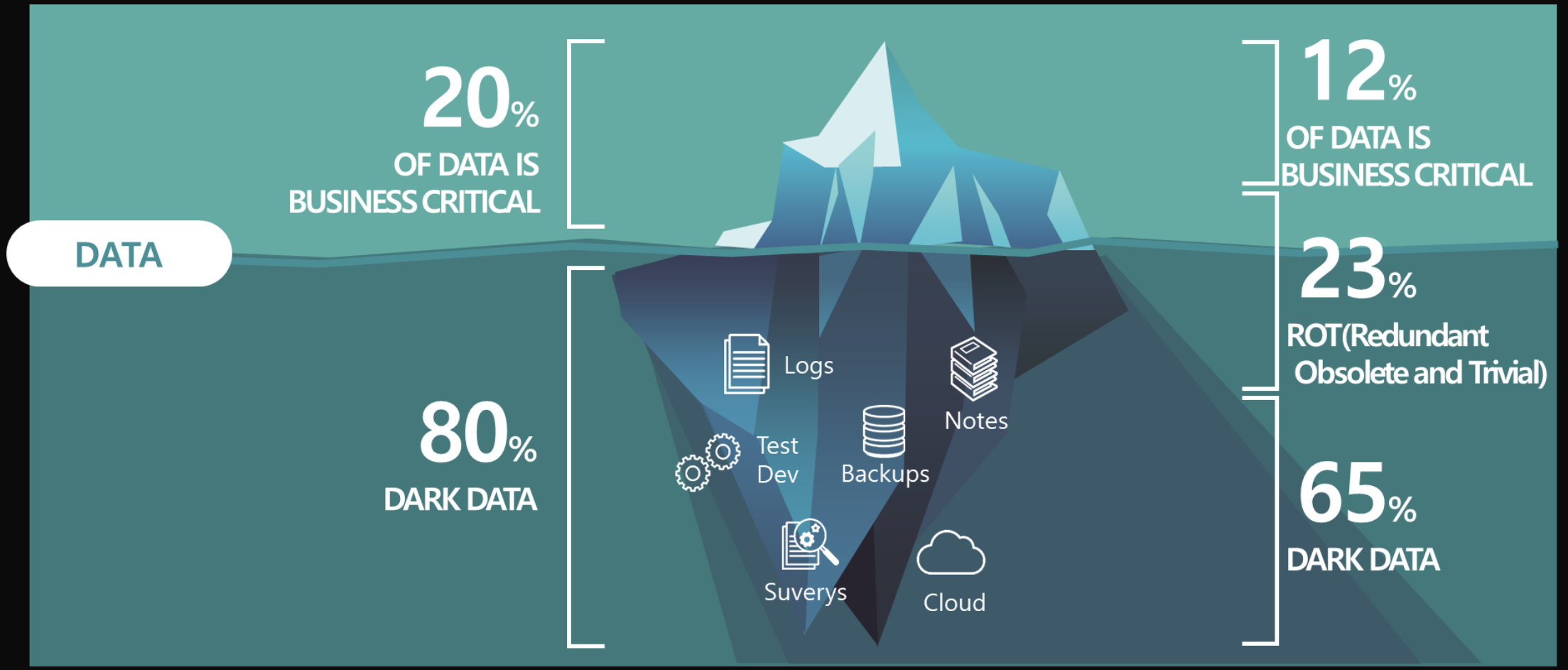
데이터의 발전

- 데이터의 형태 발전과 기하급수적 용량 증가



Dark Data

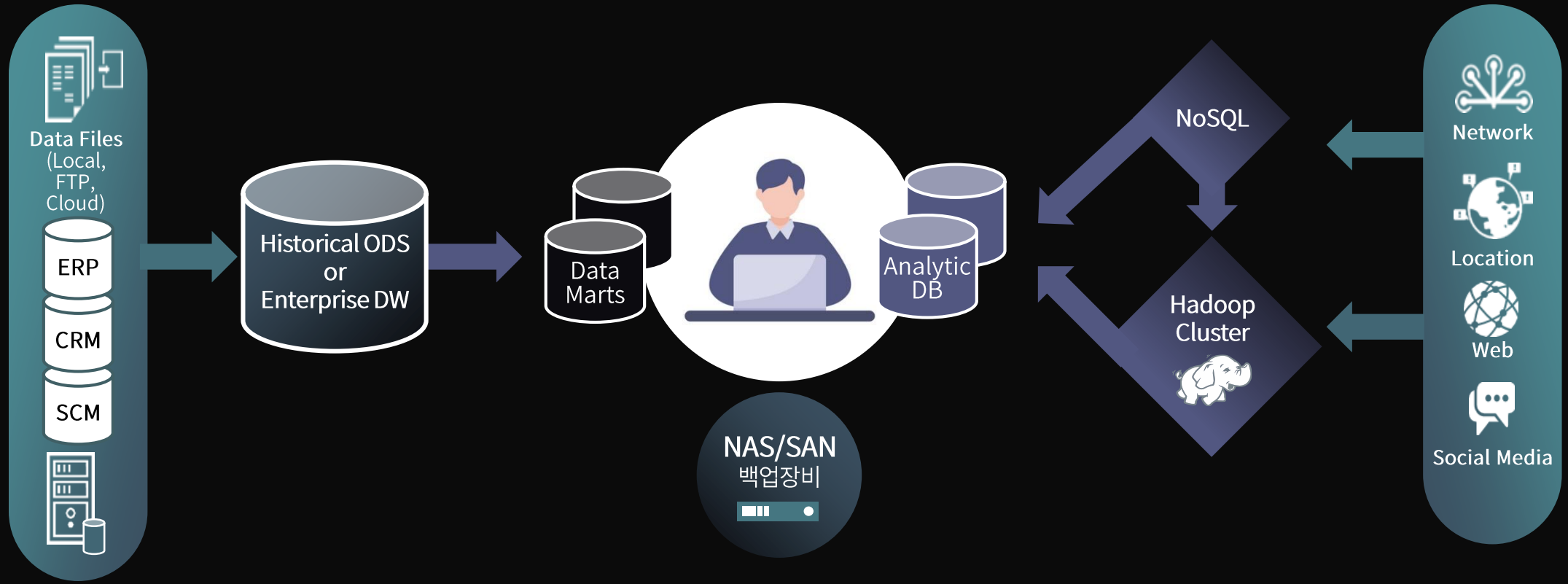
- 다크 데이터, 수집 및 분석 가능한 도구 부재! 너무 많은 데이터! 불완전한 데이터!



Source : Gartner, IBM, Datumize

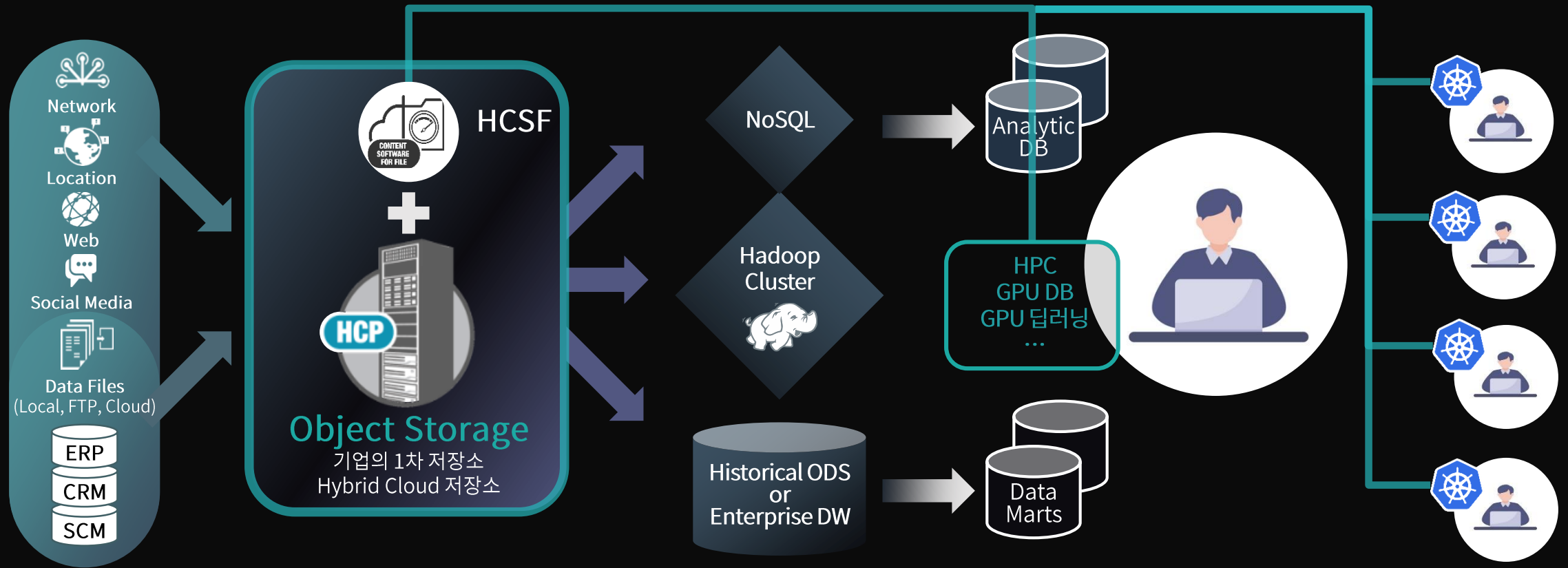
Hadoop기반 초창기 데이터 레이크 아키텍처

- DW 정보계 시스템, 초창기 Data Lake 시스템
  - ✓ 데이터를 모으기 위한 다양한 저장소 발생, 데이터 중복 등 사일로 및 연관관계 도출의 어려움
  - ✓ 모든 데이터를 데이터 타입에 상관없이 저장하기 위해 데이터 레이크 필요 (1차 저장소)
  - ✓ 분석 환경 및 분석 엔진 추가 도입 시 추가 저장소 발생

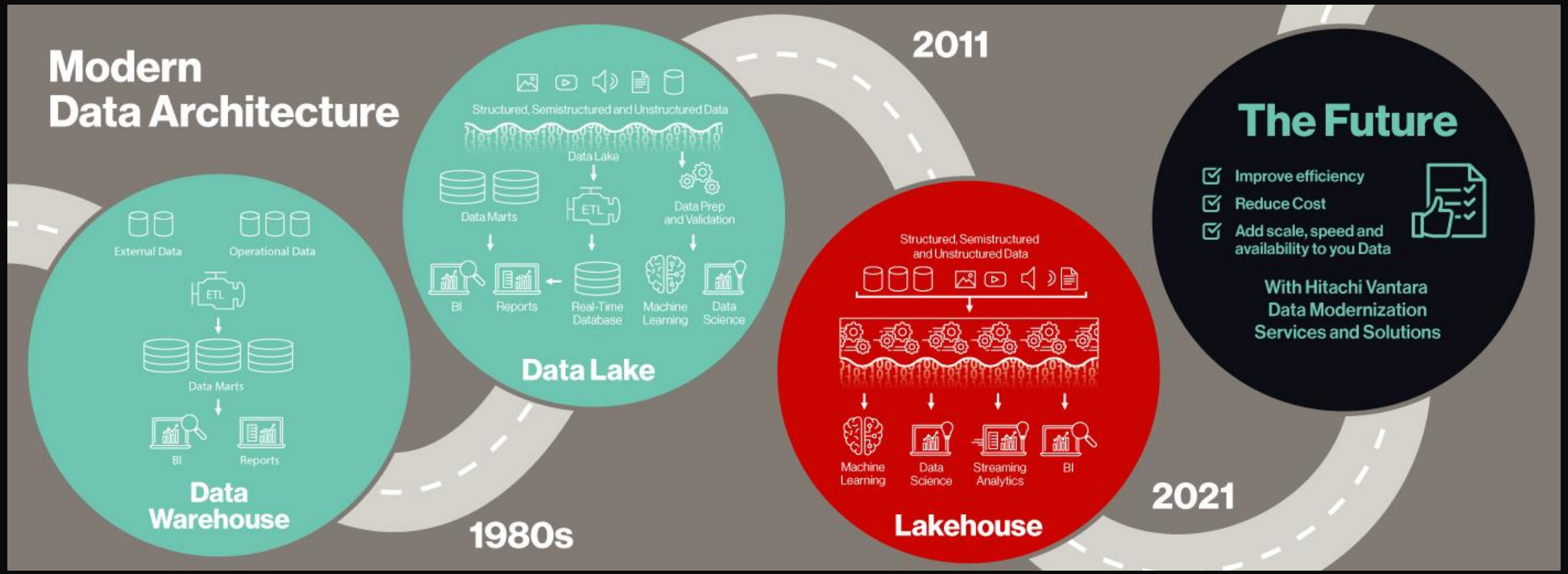


차세대 고성능 데이터 레이크 아키텍처

- 단일저장소 기반 Data Lake / 고성능 분석 도구(GPU) / 고성능 스토리지
  - ✓ 새로운 저장소가 발생할 때 마다 레거시 시스템에 연결하지 않고 오브젝트 스토리지로 연결 - 진정한 데이터 레이크 활용 가능
  - ✓ 오브젝트 스토리지와 오토 티어링 가능한 고성능 분산 스토리지 – 고성능 분석 도구의 스토리지 병목 현상(bottleneck) 해결



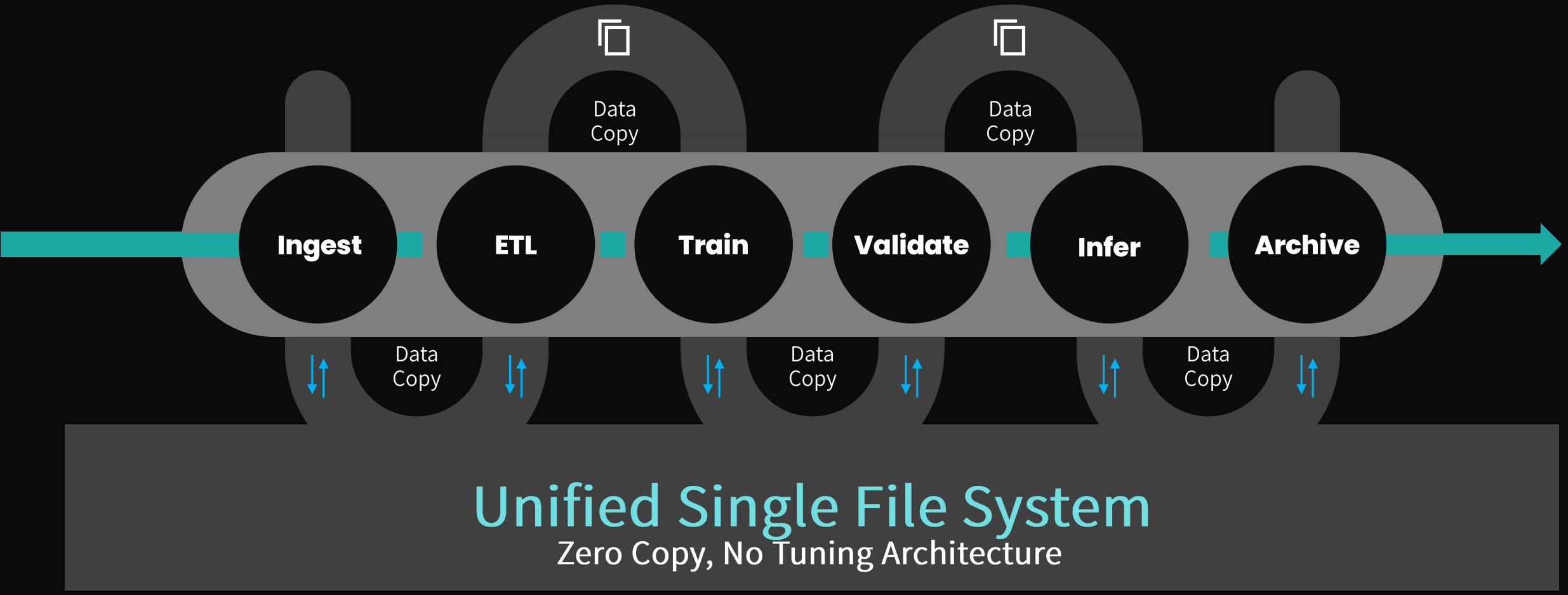
DataWarehouse > DataLake > Data Lakehouse



출처: www.hitachivantara.com

AI Data Pipelines의 저장소

- ✓ 별도의 튜닝 없이 높은 IOPs, Throughput 보장
- ✓ 멀티 프로토콜 지원을 통한 다양한 분석엔진과의 유기적 연동
- ✓ 무제한 확장성, 노드 증가에 따른 선형적 성능 향상



Deep Learning IO

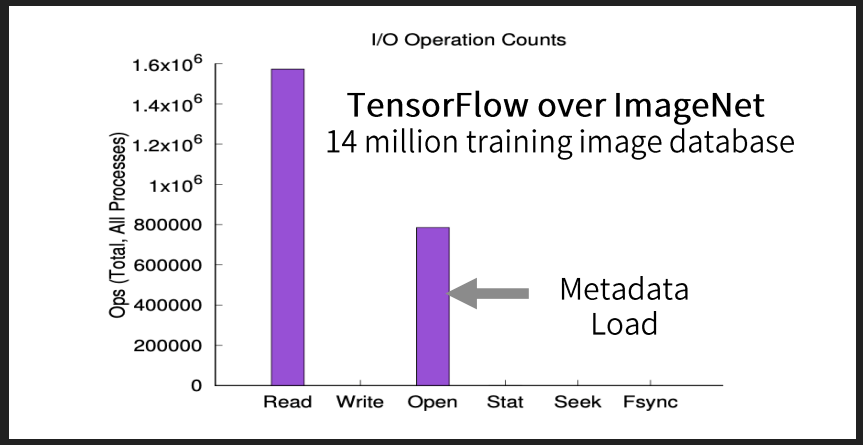
- 많은 Small Files에 대한 빈번한 Random Read 요구

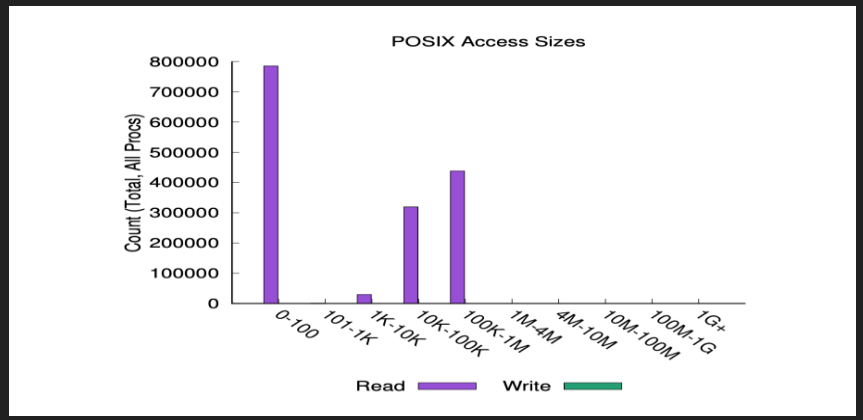
Deep Learning IO 프로세스

- **Mini batch** - 학습 데이터에서 임의의 하위 집합 반복
  각 mini batch에서 확률적 경사 하강법 수행
  (Stochastic Gradient Descent: SGD)
- **Epoch (반복)** – 전체 데이터 세트에 대해 Random 처리

Deep Learning IO 특성

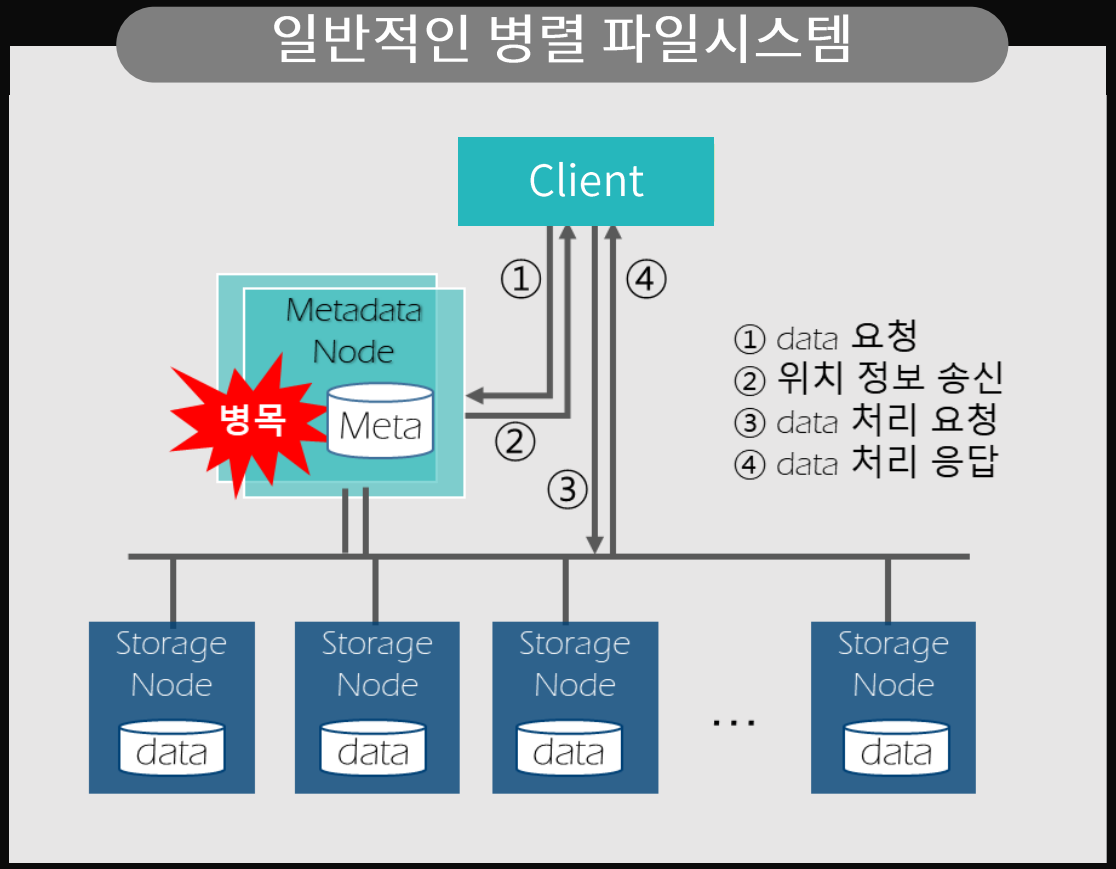
- 매우 빈번한 small IO 요청
- File을 찾기 위한 분산 파일 시스템의 Metadata overhead
- IOPs와 Throughput이 중요

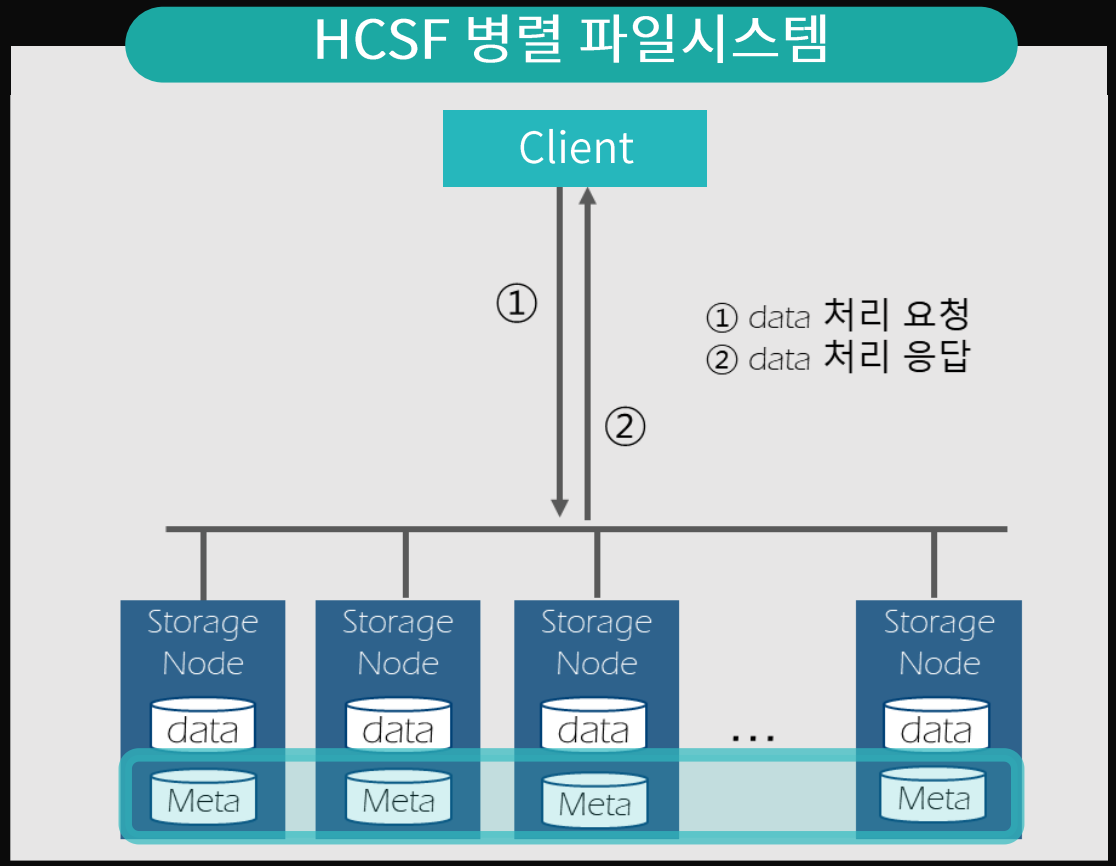




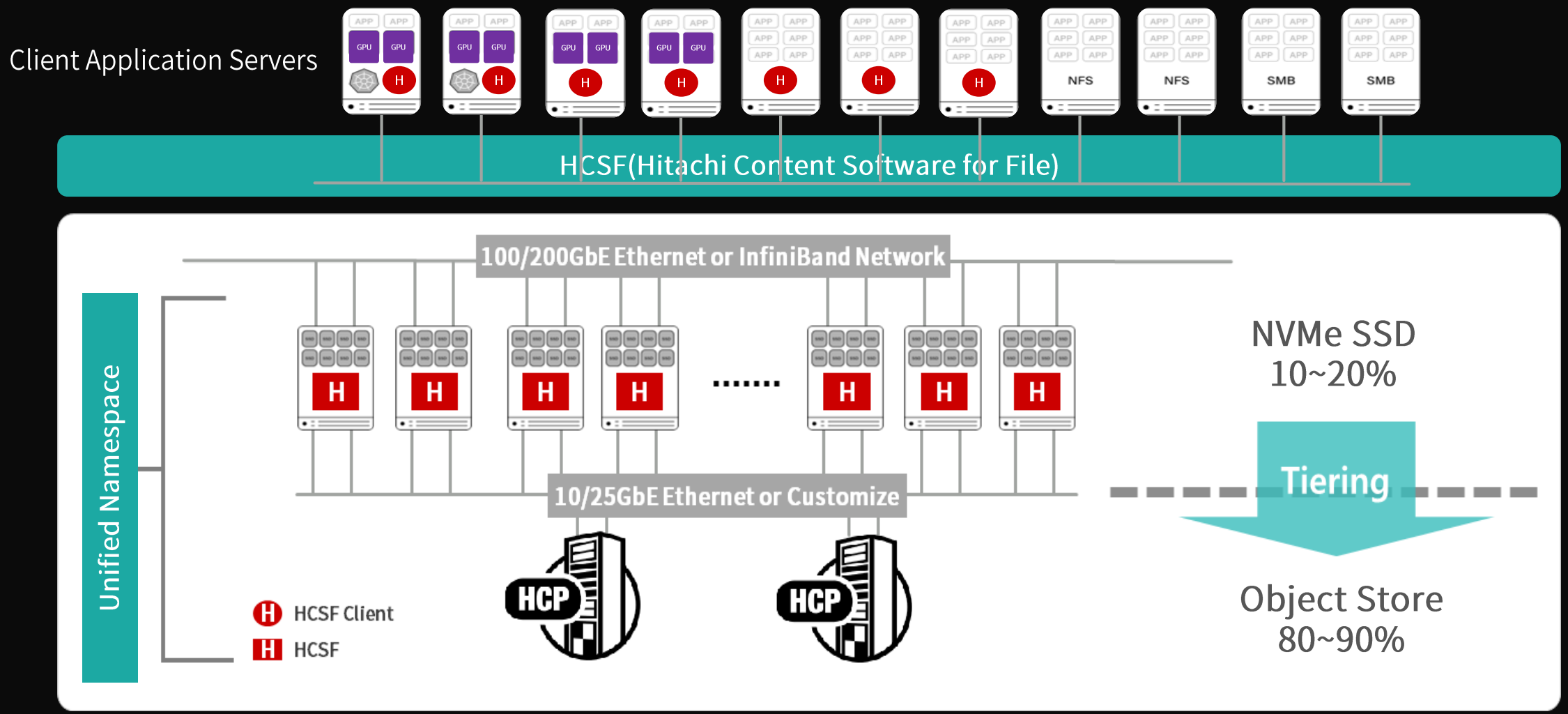
AI / HPC 환경을 위한 고성능 데이터 레이크 저장소 - MetaData 관리

- AI/ML 환경에서의 스토리지



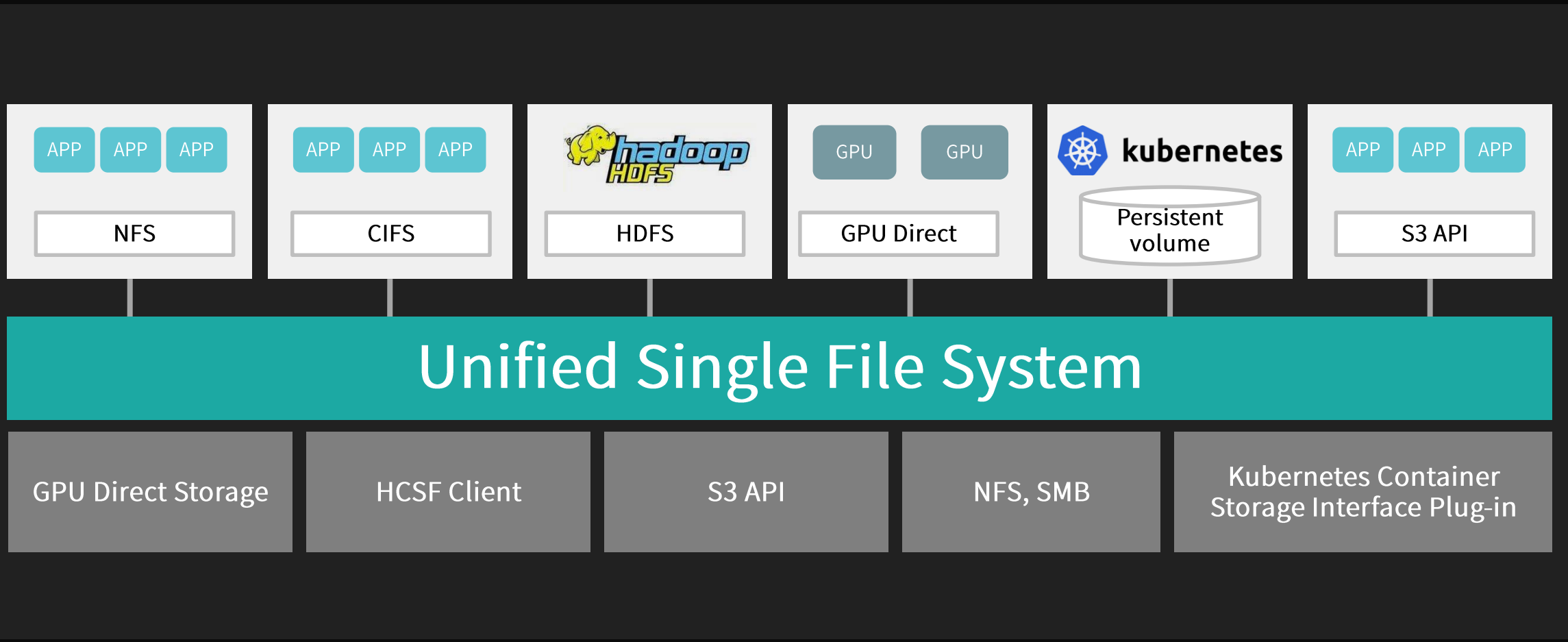


AI / HPC 환경을 위한 고성능 데이터 레이크 저장소 - Architecture



AI / HPC 환경을 위한 고성능 데이터 레이크 저장소 - Multi Protocol

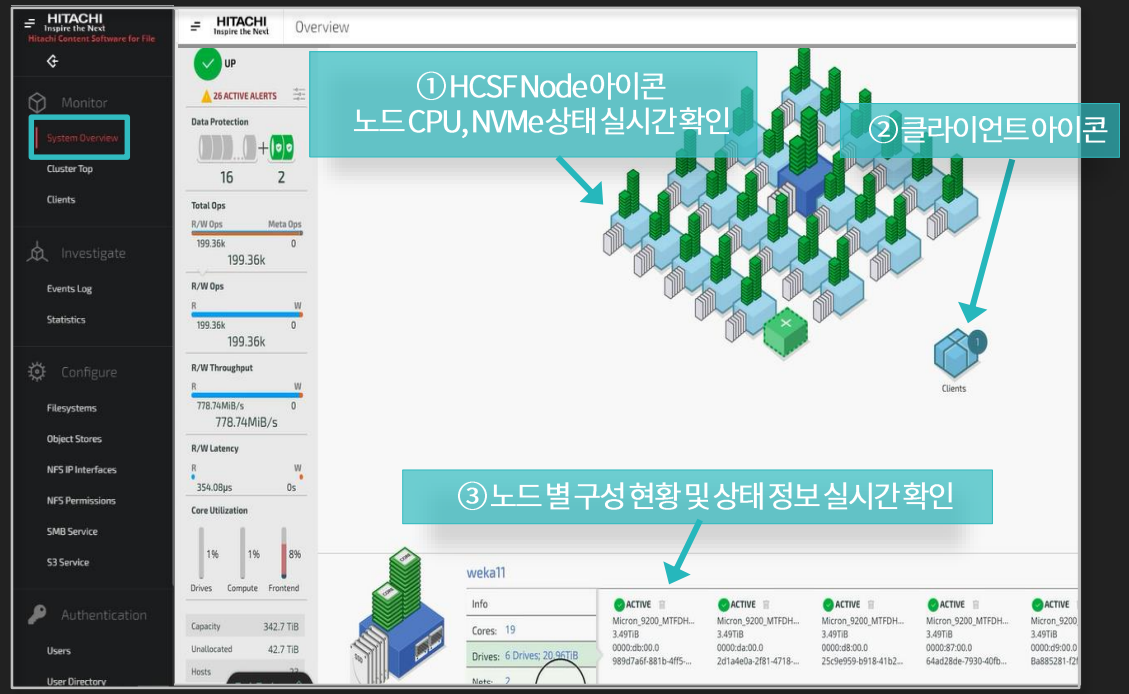
- ✓ HCSF Client (전용 Client)
- ✓ GPU Direct Storage
- ✓ NFS, SMB, S3
- ✓ Kubernetes CSI 지원



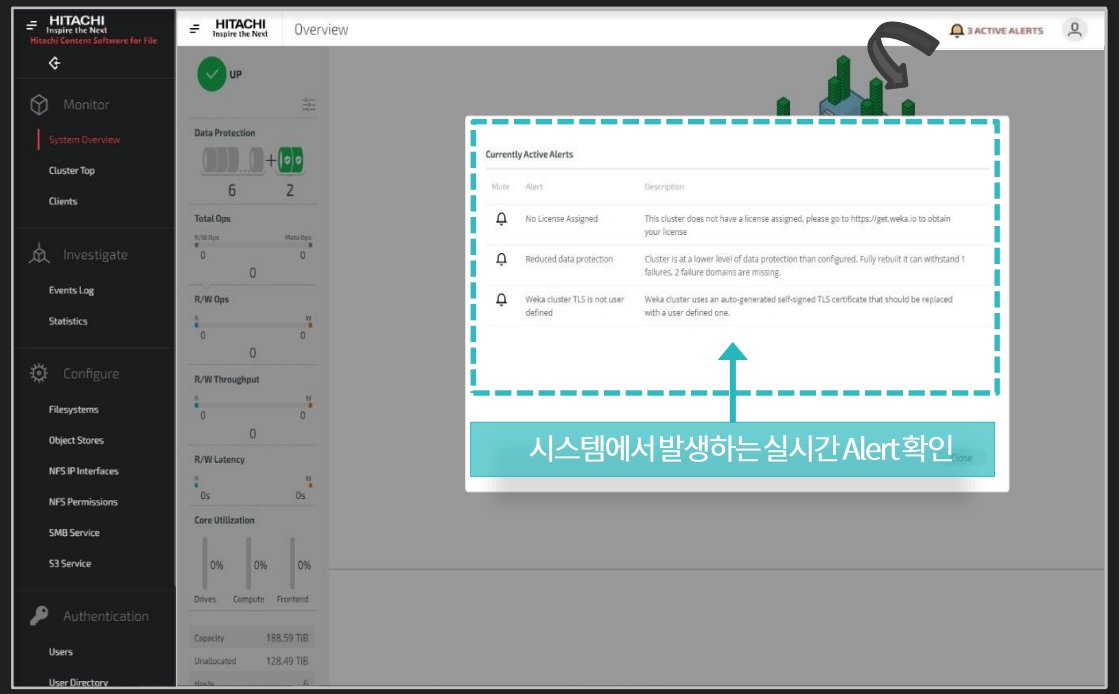
AI / HPC 환경을 위한 고성능 데이터 레이크 저장소 - 시스템 관리 모니터링

- CLI와 GUI 기반 관리를 기본으로 지원
  - ✓ 하드웨어 구성, 장애 및 모니터링, 리소스 사용률, 데이터 처리량, 스토리지 전체 성능 실시간 모니터링

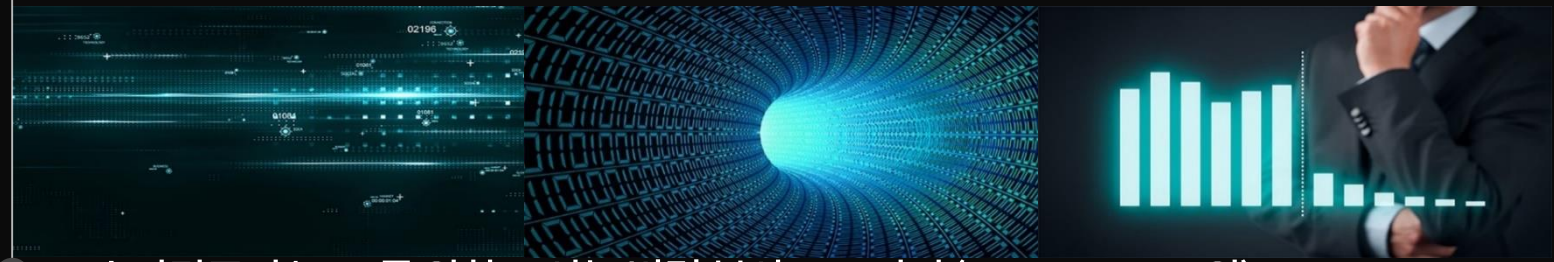
HCSF Overview 화면



HCSF Currently Active Alerts 화면



AI / HPC 환경을 위한 고성능 데이터레이크 저장소 - HCSF



HCSF?

Hitachi
Content
Software For
File

1. 슈퍼컴퓨터/HPC를 위한 고성능 병렬 분산 스토리지 (SC19 IO500 1위)

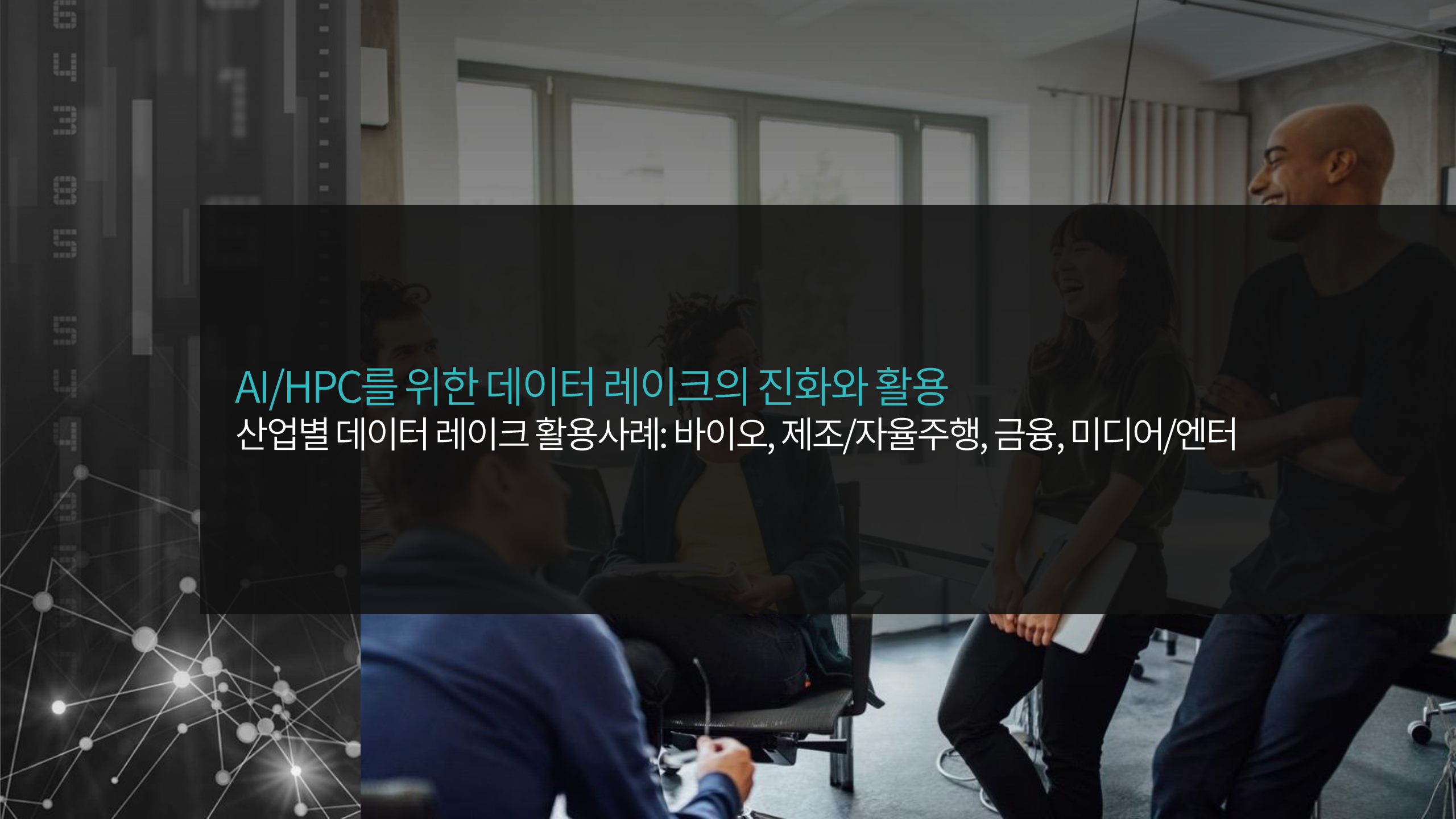
2. AI/ML 처럼 IO 집약적인 워크로드에 적합
   - ✓ 높은 처리량, 높은 IOPs, 매우 짧은 대기 시간이 동시에 필요한 혼합 워크로드에 특화
   - ✓ DPDK, GDS(네트워크 패킷 처리 기술)
   - ✓ EB규모의 확장 및 선형적 성능 향상 및 파일 크기에 따른 성능 제약 없음

3. zero-copy (진정한 Data lake storage)
   - ✓ POSIX, NFS, SMB, S3, CSI 및 GPUDirect® Storage(GDS) 지원, 데이터는 모든 프로토콜 간에 완전히 공유
   - ✓ Auto tiering은 정책 기반으로 별도 솔루션 없이, 성능 영향 없이 자동 동작

4. MSA 기반 아키텍처로 손쉬운 노드 확장과 Metadata 성능저하 방지
   - ✓ 타사 대비 확장 간편성
   - ✓ 타 분산파일시스템 대비 데이터 증가에 따른 Metadata 성능 병목 현상 없음
   - ✓ 통합관리 포탈 및 손쉬운 CLI를 통해 관리의 편리성 제공

5. 혼합 워크로드에서 초고성능 분석 환경을 위한 스토리지 통합 솔루션

❖ GDS : NVIDIA GPU Direct Storage 의 약어. GPU – Storage 병목 구간 고속 처리 기술
❖ DPDK : Data Plane Development Kit의 약어. 데이터 패킷 네트워크 고속 처리 기술



AI/HPC를 위한 데이터 레이크의 진화와 활용

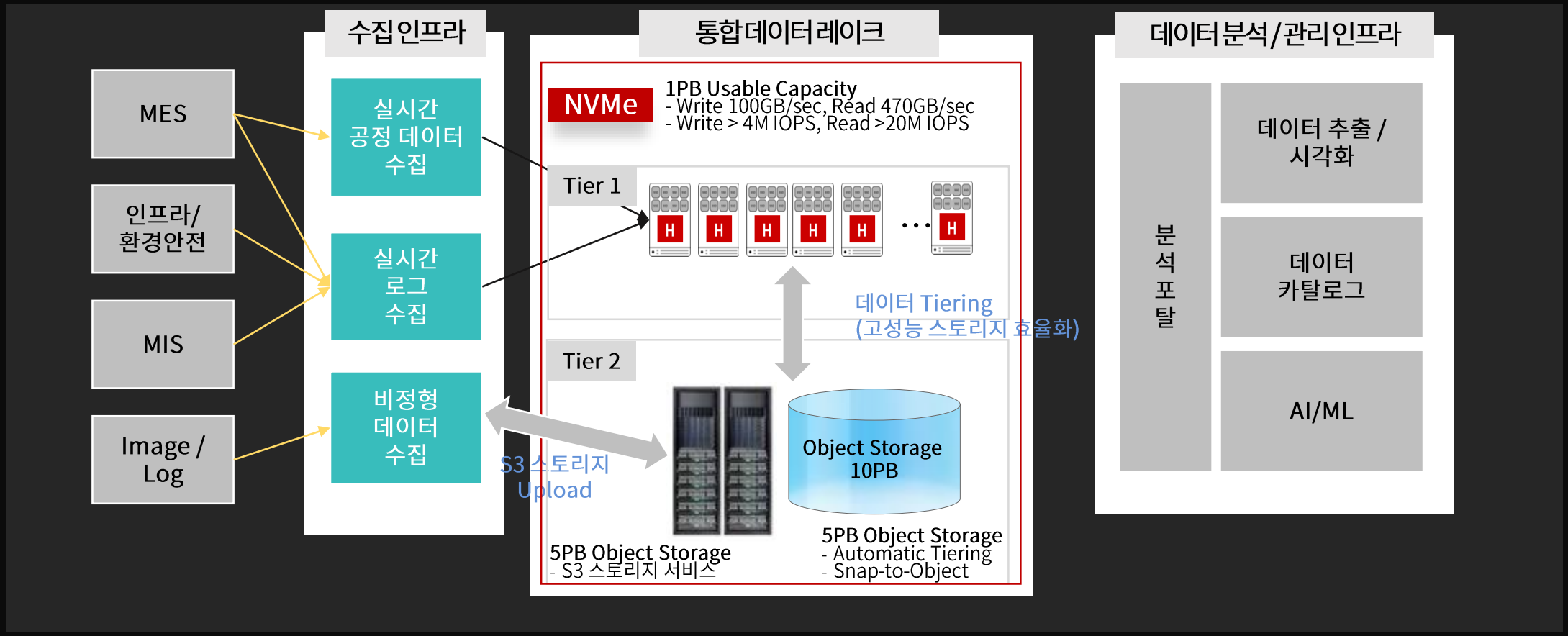
산업별 데이터 레이크 활용사례: 바이오, 제조/자율주행, 금융, 미디어/엔터

국내 제조기업 사례 – 데이터 분석 플랫폼

Category : 제조/전사 데이터 분석 체계

필수 고려 사항 : Tiering / 대용량 Query 플랫폼

- 기존 DW / Hadoop 데이터 분석 체계에서 차세대 전사 데이터 분석 체계를 위하 전사 통합 저장소 구축
- 고성능 데이터 분석 기반의 대용량 쿼리와 향후 AI/ML 을 위한 전사 분석 체계 마련
- 2nd Tier 오브젝트 스토리지는 NVMe 티어링 용도와 비정형 데이터 서비스 용도 두 가지로 나누어 동시에 운영



자율주행 인식 사례 : AV Lidar 제조

Category : 자율주행, 이미지 딥러닝

필수 고려 사항 : 모델링 성능, IOPs

- 자율주행 플랫폼은 지속적인 대용량 학습이 요구되며, 이를 위한 GPU 서버, 데이터 처리 규모의 확장이 필수입니다.

자동차, 드론, 로봇 공학 등의 분야에서 Lidar 센서 / 인식 소프트웨어 분야의 Global 선두 제조업체

고객사 현황 및 요구사항

- AI 기반의 주변 인식에 대한 실시간 분석 및 판단 요건이 필수
- 물체의 식별, 분류 및 추적을 위한 이미지의 판단 초고속 인프라가 핵심
- AI 효율 향상을 위한 대규모 데이터 세트의 지속적인 학습

솔루션 구성 및 구축 효과(ROI)

- BMT 결과
  - 스토리지 확장 시 선형적인 성능 향상 확인
  - GPU 클러스터에 대한 고대역폭 IOPs 제공 확인
- NVMe 92TB + 660TB object storage 구축
- Much faster Epoch time :
  - GPU 환경에서 HPC 워크로드 요건 충족
  - NFS보다 10배, 로컬 NVMe SSD 대비 3배
  - 반복적인 테스트를 위한 운영 환경 제공

글로벌 전기차 생산 업체 : 자율 주행차 사례

Category : 자율주행, 이미지 딥러닝

필수 고려 사항 : 모델링 성능, IOPs

- 자율주행 데이터 파이프라인 플랫폼

Business Challenge	GPU 서버 증설 효과 미비
• 10개의 GPU cluster에서 ML 모델 개발 (All-Flash NAS) • 단일 AV 당 40TB 이상의 데이터 생성 (8 시간) • Deep Learning training 을 위한 수 Petabytes 동시 처리 필요 ➔ 성능 미충족 / ML 모델링 어려움	• Petabytes 규모의 데이터 처리를 위한 지속적 GPU 서버 증설 • DL Data Set : 수백만 개의 4K 이미지 파일 • All-Flash NAS는 1.5GBytes/s 달성 • Data sets을 로컬 NVMe에 복사 ➔ GPU 遊休 상태 발생(GPU 활용률이 10%)

신규 솔루션 채택 (since 2018)

- 3개의 학습 clusters (6PB flash / 27PB object storage)
- epoch (Ingest->ETL->Train->Validation) 단축 : 2주-> 4시간
- 단일 네임스페이스에서 PB 단위, 수십억개의 파일 학습 활용
- 선형적인 성능 확장을 통한 10억 단위의 IOPs 성능 확보
- All Flash NAS 대비 7배, Local NVMe 대비 2배 이상 성능 확보 및 GPU Utilization은 극대화 효과

차세대 방사광가속기 사례

Category : 연구소, BIO

필수 고려 사항 : 운영성, 고대역폭

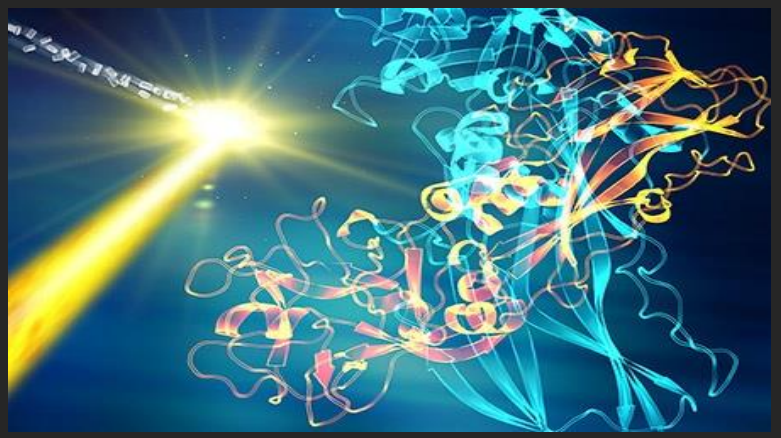
- 미국 에너지부 산하의 입자 가속기 연구소로 입자 물리학, 천체 물리학, 재료, 화학, 생물 및 에너지 과학을 위한 Science Computing 수행 연구소

Business Challenge

- 가속기에서 원자 해상도 자료에 대한 분석 인프라 요건 (1TB/s Throughput)
  - 전례 없는 대역폭의 데이터 입출력 플랫폼 필요
  - 엄청난 속도의 X선을 활용한 원자 구조에 대한 실시간 해석 인프라
  - 극단적으로 짧은 Beam time에 반복적인 실험 지속을 위한 최적화된 플랫폼

고성능 데이터 분석 플랫폼

- 방사광에 의한 세포 소멸 전 초대용량 데이터 분석을 위한 고성능, 고신뢰성 시스템 제공
- 많은 연구원들이 하나의 공간에 연구할 수 있는 데이터 레이크 환경 제공 (단일 NameSpace)
- 지속적인 성능 / 용량 증대를 위한 유연한 확장성과 사용 편리성 제공



생물 정보학 워크로드 지원 사례

Category : 연구소, BIO

필수 고려 사항 : 제로튜닝, 단일 NameSpace

- 극저온 전자 현미경 활용 사례 [Cryo-electron microscopy]

Business 환경

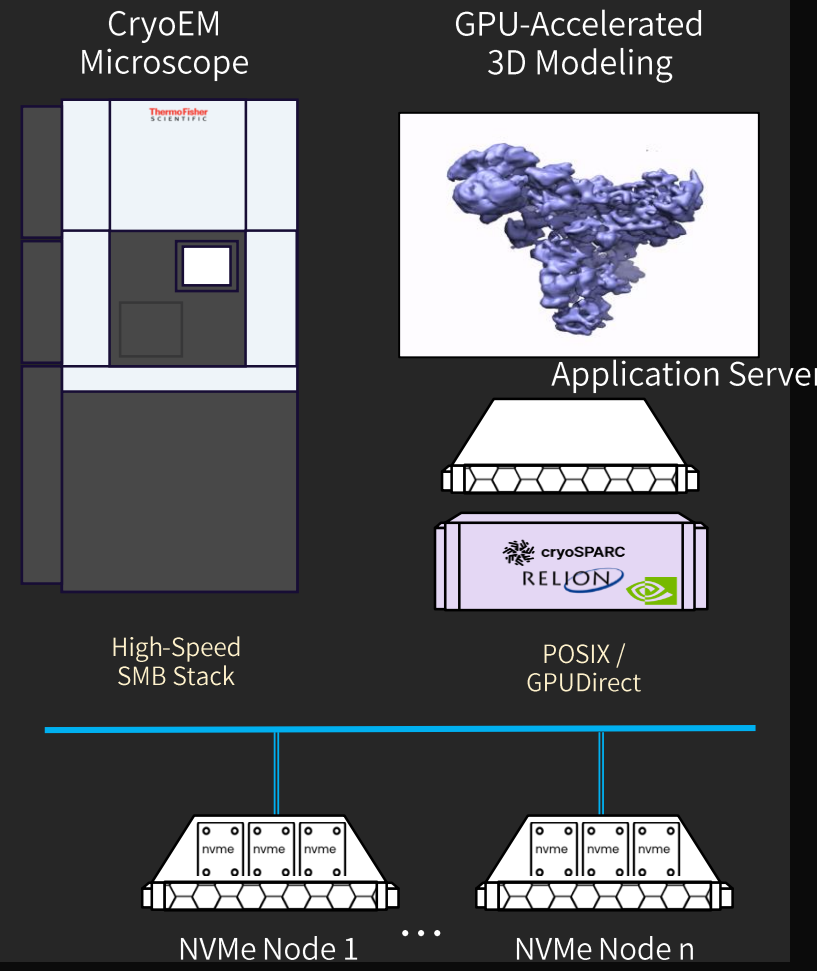
- 생체 입자를 극저온 상태로 유지하면서 전자현미경으로 2D 이미지 생성
- 대량의 2D 생체 이미지를 고해상도 3차원 구조로 모델링

➔

- 입자 손상의 극단의 시간동안 이미지 Write 및 3D 시각화를 위한 대량의 Read IOPs가 필수 요건

고성능 데이터 분석 플랫폼

- 최적의 운영성 확보
  - 4K 블록에 저장된 데이터를 위한 IO 최적화
  - 제로 튜닝으로 운영성 확보
  - Microsecond latency
  - 단일 디렉토리에 수십억 개의 파일로 정교한 모델링 가능
  - 공용 Namespace : 엑사바이트 규모
  - 완전히 분산된 메타데이터 : 병목 최소화



금융 통합 FDS / 분석 데이터 플랫폼

Category : 금융, FDS

필수 고려사항 : 시계열 로그분석, 데이터분석 일원화

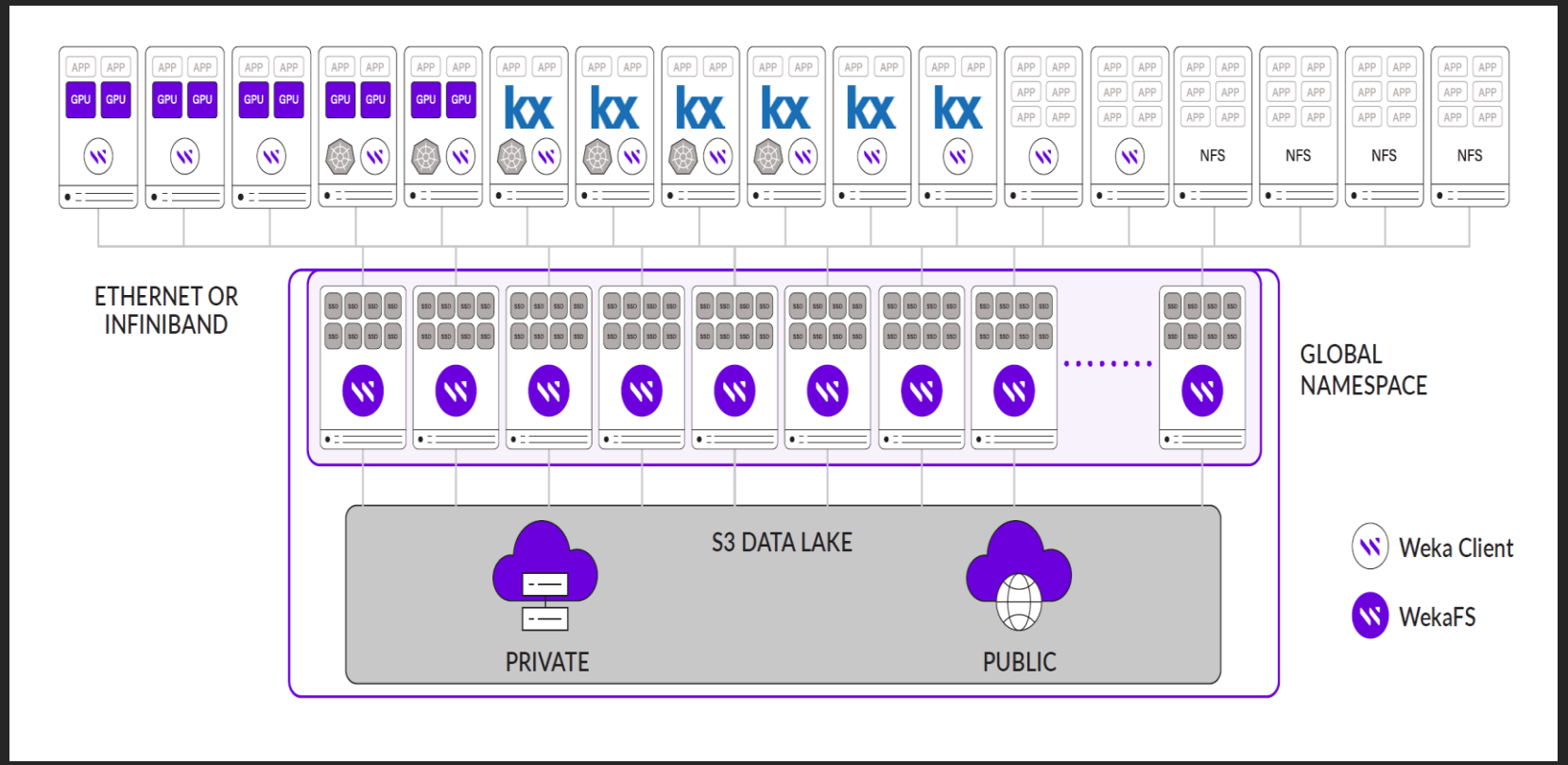
- 실시간 부정거래 방지 시스템(FDS) 및 거래 데이터의 고성능 분석 플랫폼을 구축

고객사 현황 및 요구사항

- 글로벌 txn이 초당 24,000 건까지 증가 (2020년 추산)
- 부정 사용의 지속적 증가 : 기존 FDS와 함께 ML 학습에 의한 시나리오 분석 플랫폼 필요

=======

- 대용량 dataset을 고속으로 저장, 분석, 처리 및 검색에 활용 (수십억개의 record 동시 처리
- 시계열 DB인 kdb+ 활용 기반



퀀트 트레이딩 플랫폼

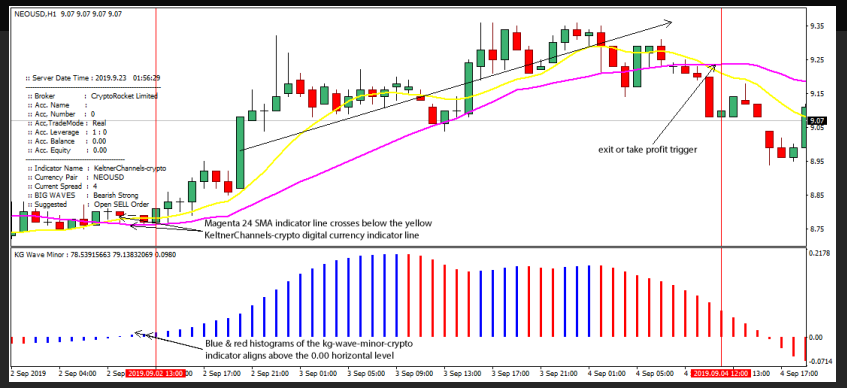
Category : 금융, Machine Learning

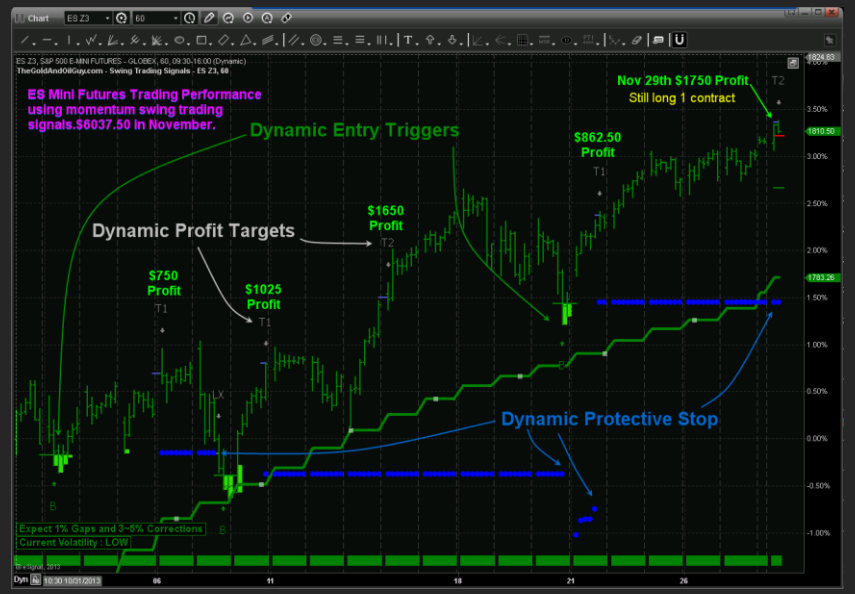
필수 고려 사항 : Zero-Copy , 실시간 강화학습

- AI 모델링에 기반한 자동화 된 증권 거래 시스템

Business Challenge

- 1000개 이상의 서버 자동화된 트레이딩 작업 : SSD 필수 요건
- Legacy NAS를 대체하는 로컬 NVMe 대체
- latency 증가 (학습 데이터 증가 및 강화학습 활용) ➔ 거래 지연시 대규모 손실 예측





고성능 스토리지

- 2018년 9월 실거래 시스템 적용
- 8 node NVMe 병렬 처리 cluster
- 로컬 NVMe 드라이브 대비 3배 우위 성능 제공

구축 효과

- Zero-Copy Architecture : Application 서버에 별도의 데이터 복사 불필요
- NVMe drive 번아웃 현상으로 인해 소요됐던 스토리지 비용 65% 절감 효과

뉴욕 증권 거래 플랫폼

Category : 금융, 실시간 대용량 처리

필수 고려 사항 : 확장성, 저비용 DR

기존 문제점

- NYSE 일일 평균 주식 거래량의 폭증
  (Tick Data : 2019년 70억, 2020년 109억, 2021년 147억)
- 유연하고 비용 효율적인 확장 구조 필수
  - Legacy 시스템, 매우 어려운 확장 구조
  - DR 시스템 구축 시
    과도한 예상 비용 추정
- 기존 데이터 분석 결과,
  IOPs 가 최대 관건

WEKA

- 2019년 구축
- NVMe 188TB / 600TB Object storage
- 3-site DR 구축

구축 효과

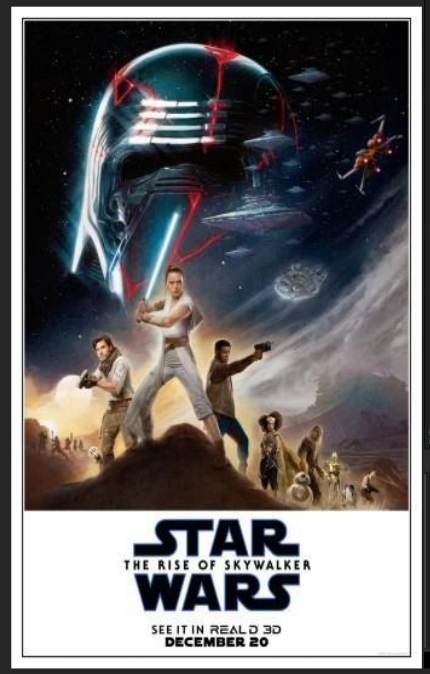
- 성능 향상 : 기존 대비 체결 7배 성능 확인
  (Latency 극소화)
- 운용 비용 1/7로 절감
- 중대한 재해로부터 복구 시스템 마련
- 향후 확장을 위한 Simple Architecture

그래픽 편집, VFX(visual effect) 활용 사례

Category : Media & Entertainment

필수 고려 사항 : 초고속 SMB

- Animation Studio 고객사 / 글로벌 OTT 서비스 사업자



Business Challenge

- Color Correction / Rendering 에 인프라 성능 저하로 인한 작업 효율성 저하
- 그래픽 편집 작업을 위한 이미지 Read 성능 강화 필수 요건

구축 효과

- Phase 1: 500 TB NVMe, Phase 2 진행
- 최적의 Low-Latency 확보로 업무효율성 최적화 평가
- 주요 프로토콜: SMB –W

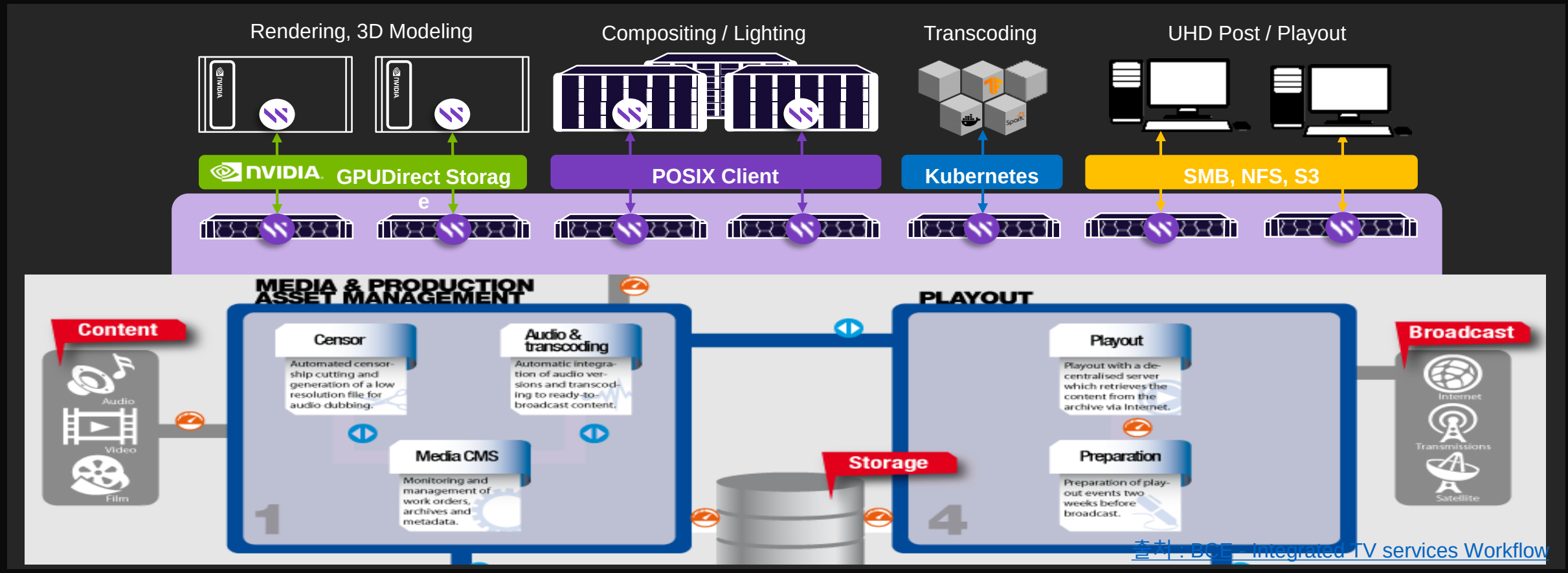
글로벌 OTT 서비스 진행 : 현재 600PB 규모
- 평균 2MB 파일, 3 조 개의 Object에 대한 안정적인 서비스 실시중

권장 활용 사례 : The Unified DataLake (방송/ 미디어)

Category : Media & Entertainment

필수 고려 사항 : The Unified 스토리지

- 컨텐츠의 생산부터 소비까지의 전 워크플로우 관점에서 일원화된 공간에서 작업
  - 로컬 복사본을 만들지 않고도 데이터에 액세스하고 공유
- 완벽한 협업 워크플로우 제공 : 어디에서나 접근이 가능하여 이동성과 생산성이 향상
- 3D 모델링, 시각효과, 트랜스코딩, 합성 및 영상 편집, 및 송출까지 모두 하나의 작업공간에서 수행.



출처 : BCE - Integrated TV services Workflow

감사합니다.