

[지디넷코리아 ACC+컨퍼런스]

미래 AI환경을 위한 데이터 레이크하우스 전략

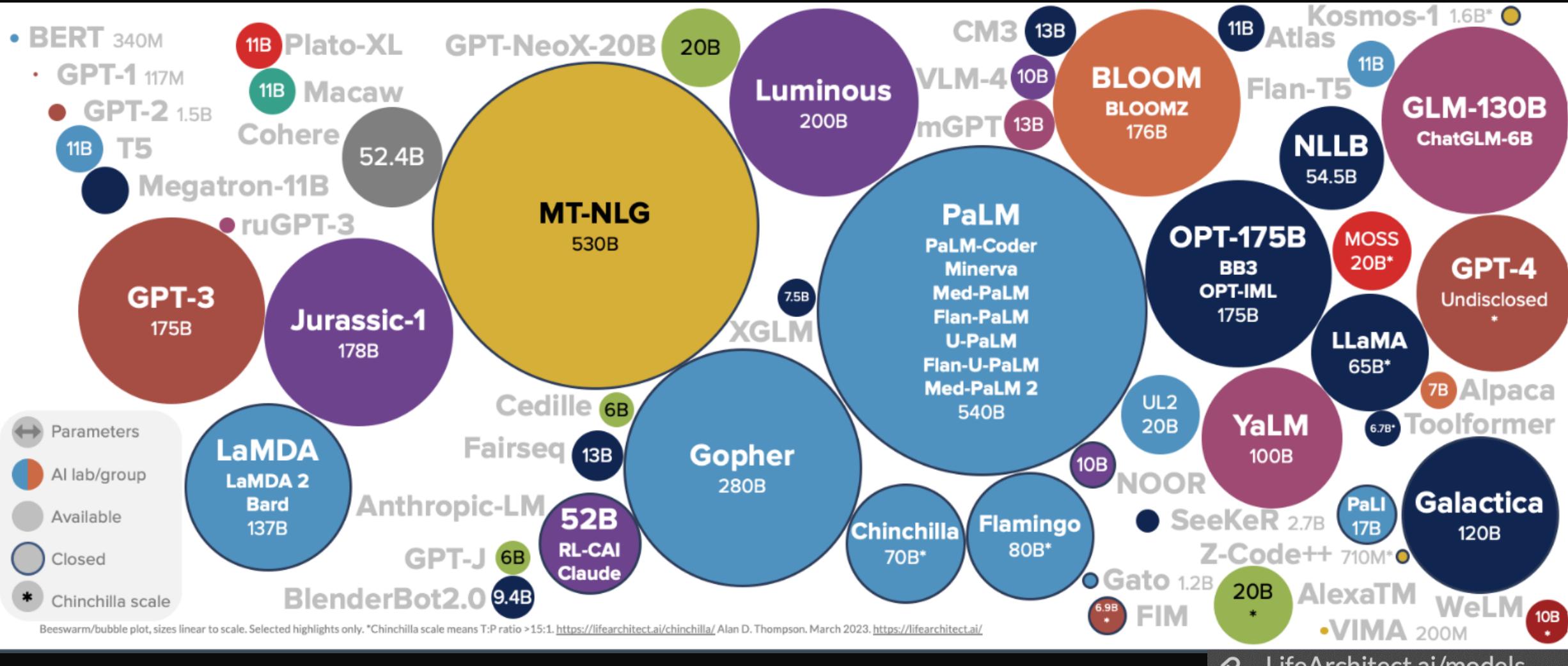
효성인포메이션시스템
데이터사업팀
권동수 전문위원



GPT 모델 크기

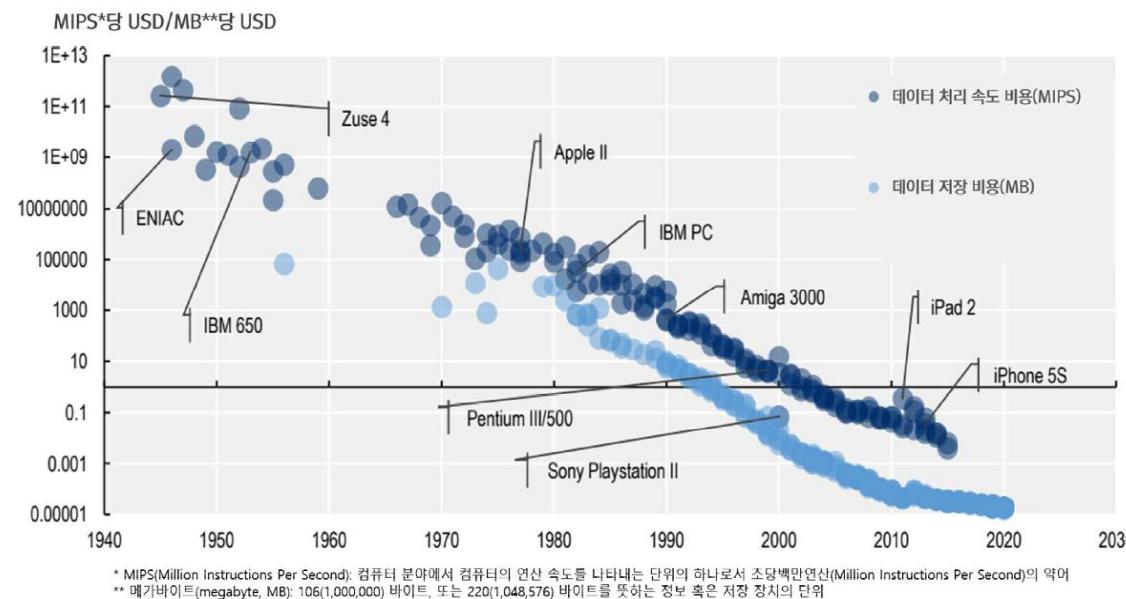
GPT	GPT-2	GPT-3	GPT-4
1억 1700만	15억	1750억	미공개
미공개	40GB	570GB(45TB)	미공개

LANGUAGE MODEL SIZES TO MAR / 2023



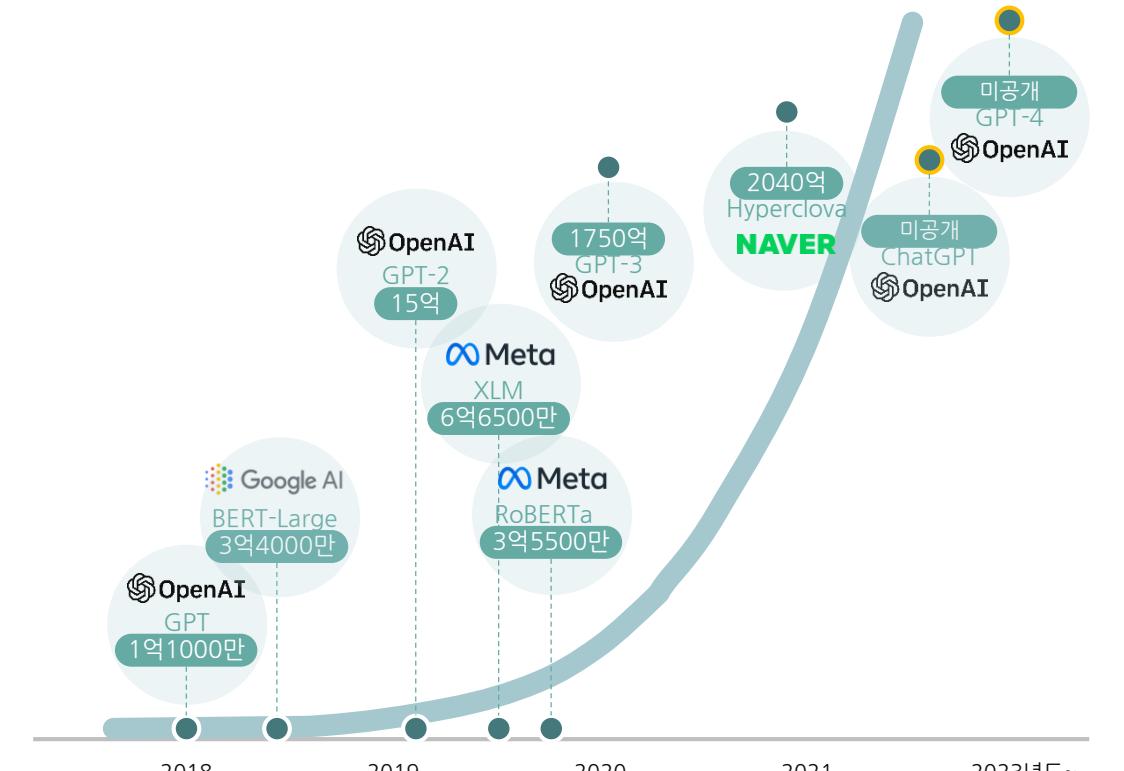
데이터 시장 가치 및 모델 데이터 크기 변화

글로벌 IT 제품별 데이터의 처리 및 저장 비용



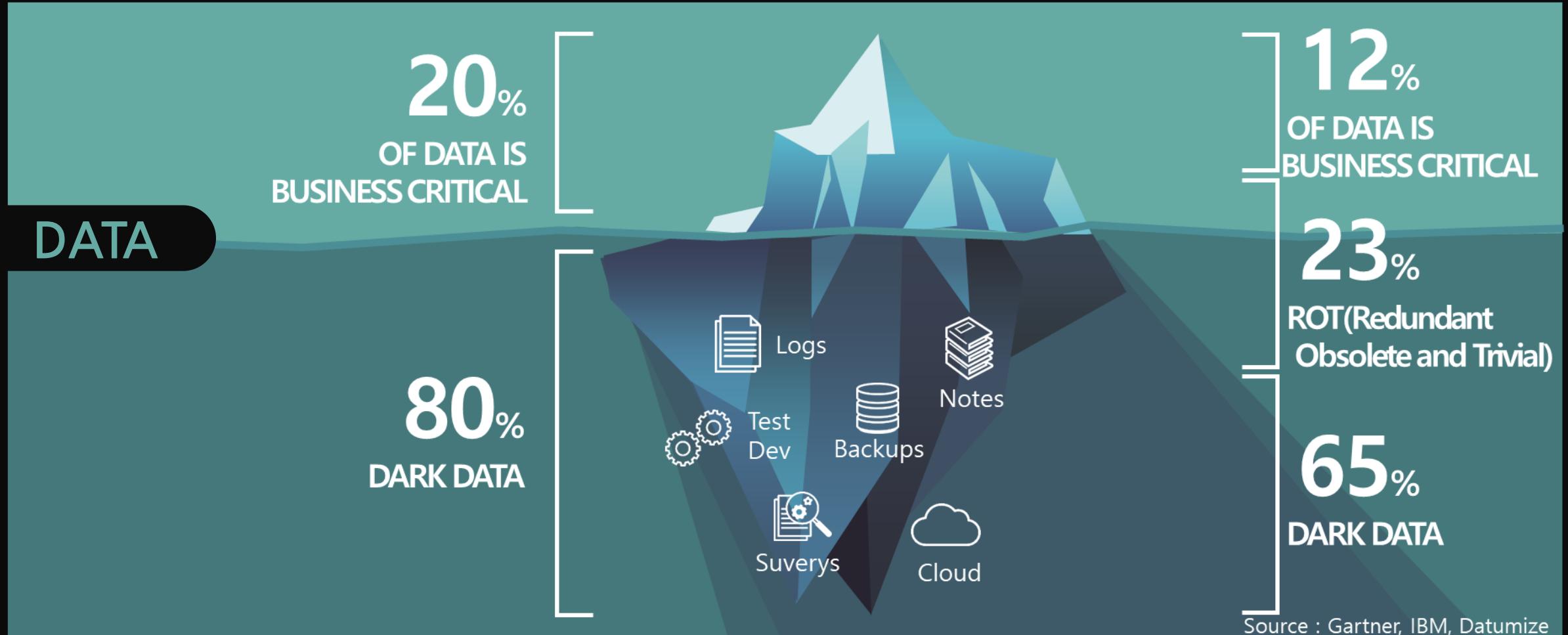
Source : 한국데이터산업진흥원

large language model 연도별 모델 크기



다크 데이터?

다크 데이터, 수집 및 분석 가능한 [도구 부재](#)! 너무 많은 데이터! 불완전한 데이터!

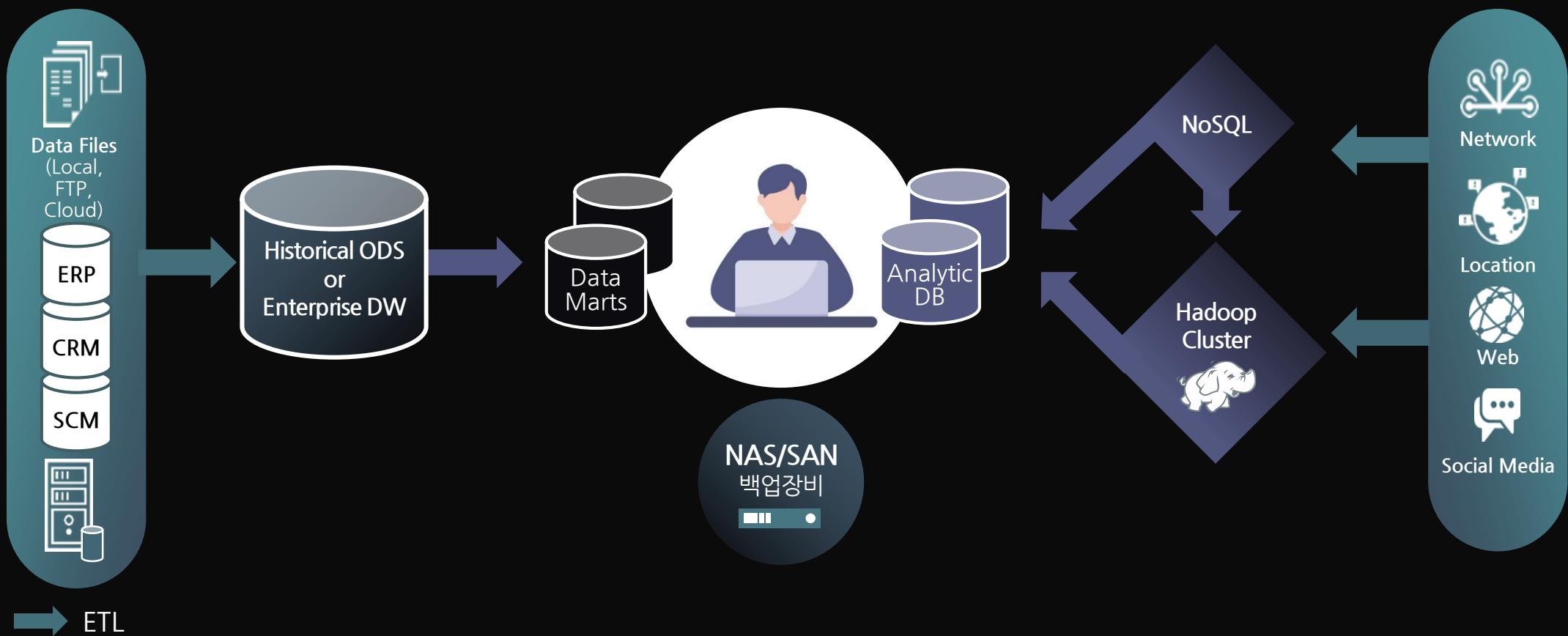


다크 데이터(dark data)는 다양한 컴퓨터 네트워크 운영을 통해 얻는 데이터이지만 의사 결정이나 이해를 위한 수단으로 사용되지는 않는다.

1세대 데이터 레이크

DW 정보계 시스템, 빅데이터 Hadoop 시스템??

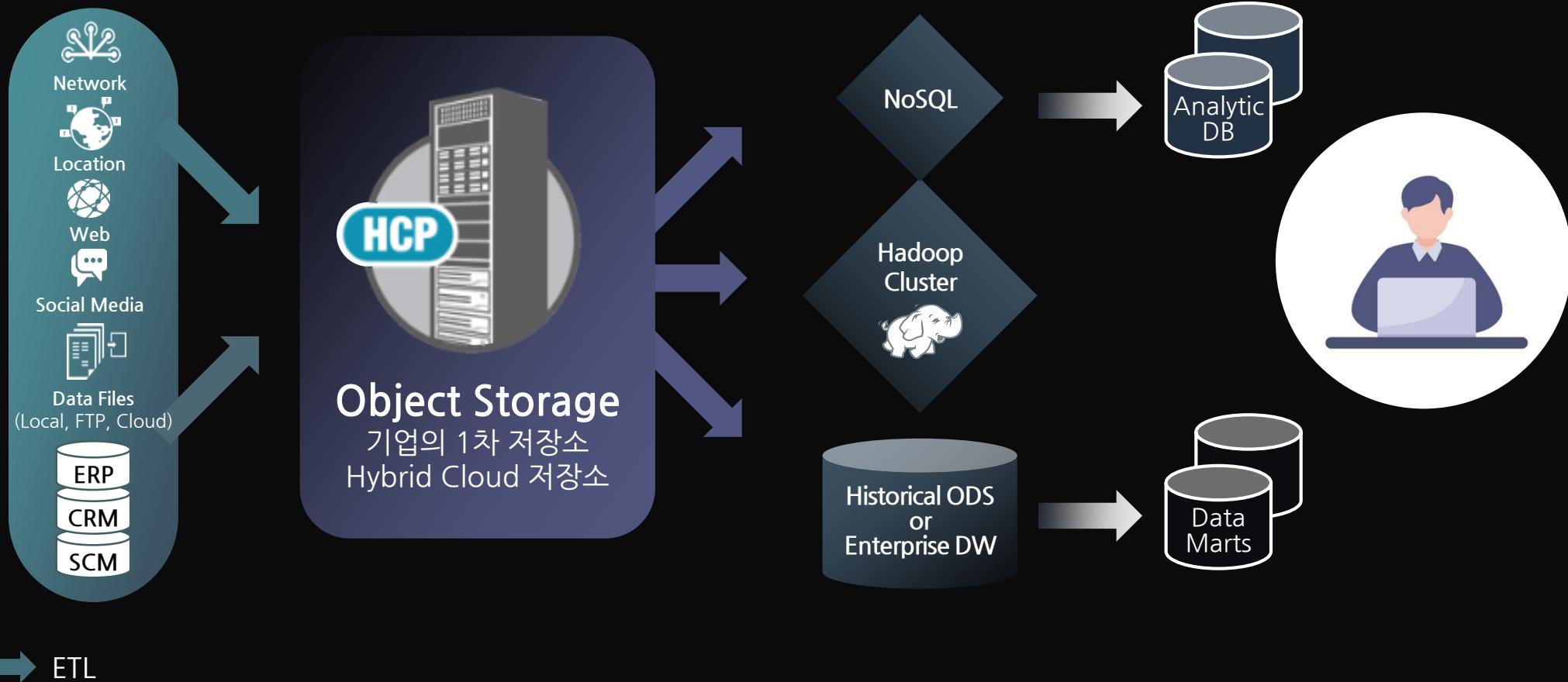
- ✓ 데이터를 모으기 위한 다양한 저장소 발생
데이터 레이크가 필요한 이유는 모든 데이터를 데이터 타입에 상관없이 저장 가능한 시스템이 필요 (1차 저장소)



차세대 데이터 레이크

오브젝트 스토리지 기반 데이터 레이크

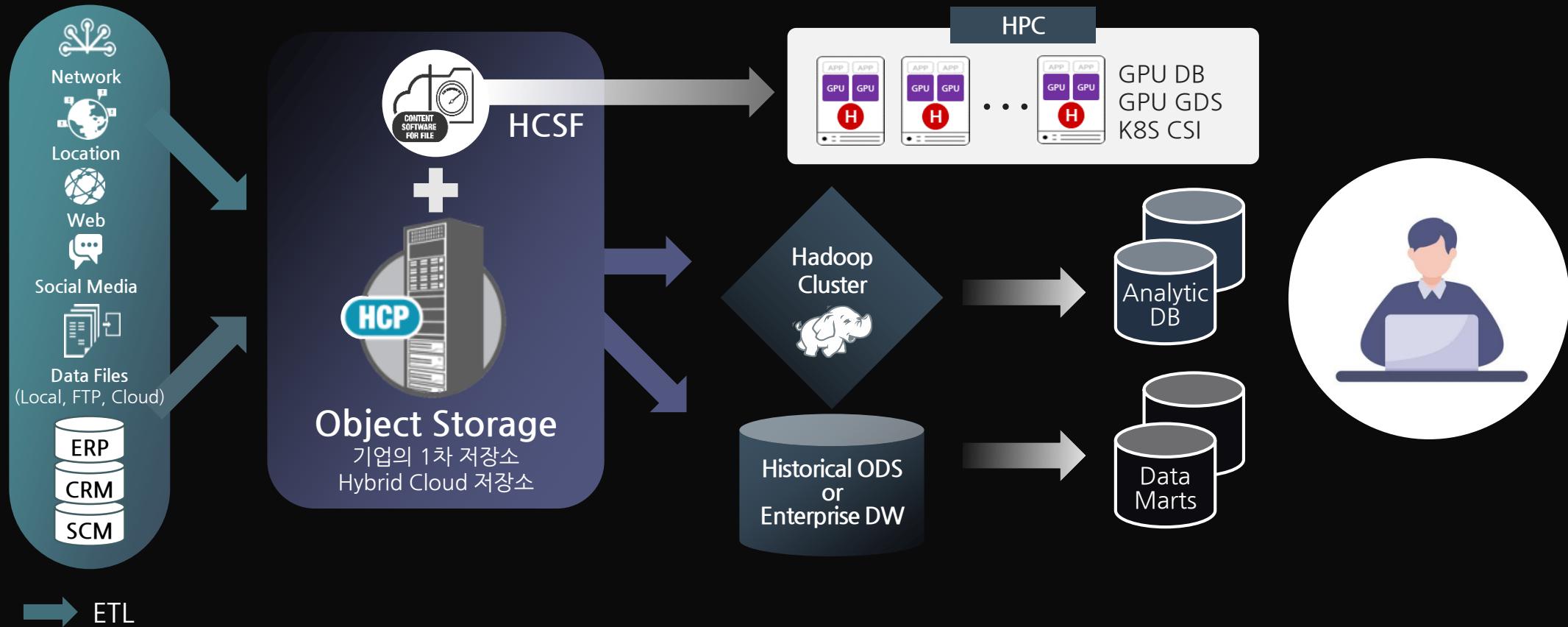
- ✓ 다양한 데이터를 저장할 수 있는 오브젝트 스토리지를 사용해 **진정한 데이터 레이크 구축**



차세대 데이터 레이크

오브젝트 스토리지 + 초고성능 병렬 파일시스템 기반 데이터 레이크

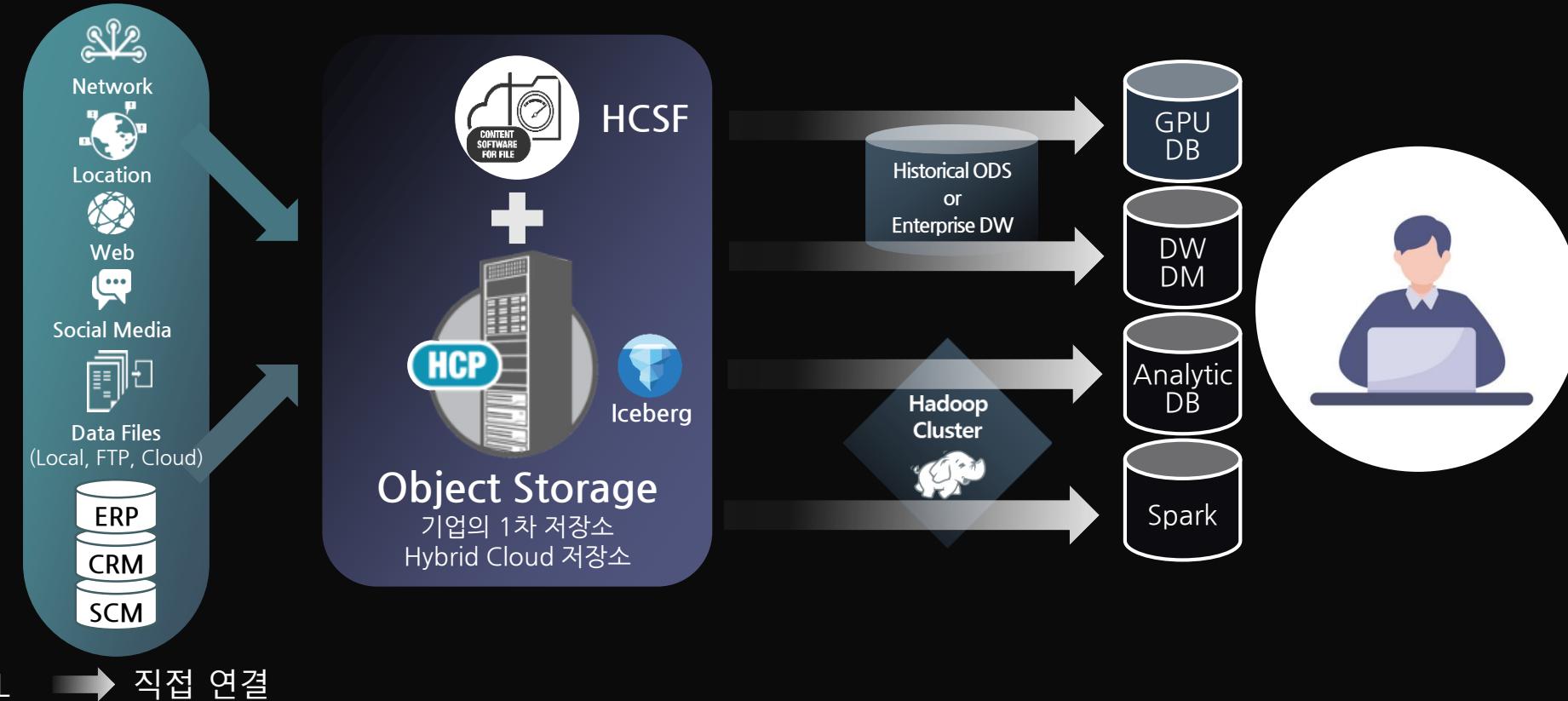
- ✓ 다양한 데이터를 저장할 수 있는 오브젝트 스토리지를 사용해 **진정한 데이터 레이크 구축**
- ✓ **쿠버네티스 환경 지원(CSI 지원)**
- ✓ **GPU 직접 연결을 통한 효율적인 자원 활용(GPUDirect Storage 지원)**



데이터 레이크하우스

데이터 웨어하우스 + 데이터 레이크 = 데이터 레이크하우스

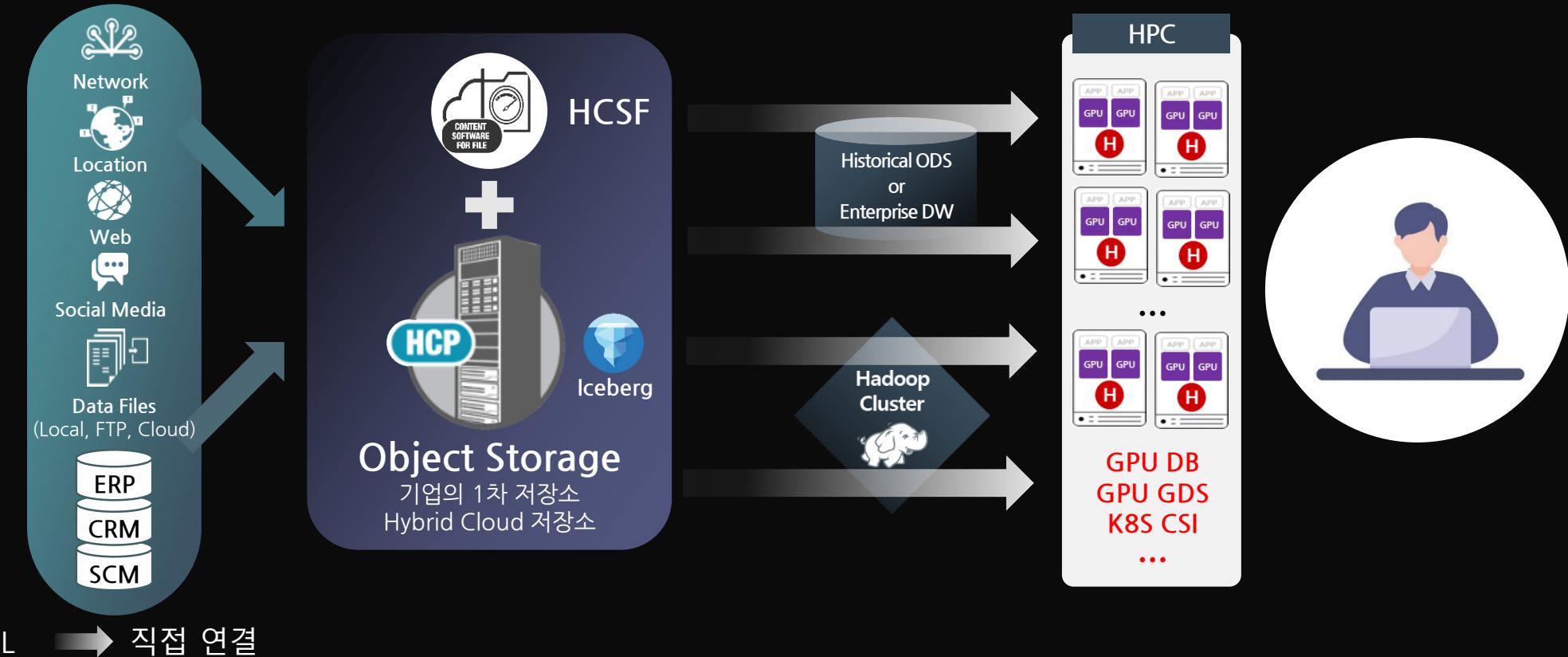
- ✓ 오브젝트 스토리지(HCP)의 개방형 테이블 형식(MetaData)을 제공하는 Apache Iceberg 등장으로 Spark, Trino, PrestoDB, Flink, Hive, Impala 를 통해 테이블 데이터 조회 가능한 **데이터 레이크하우스** 구축
- ✓ 오브젝트 스토리지(HCP)를 데이터 노드로 활용하는 DW솔루션 등장
- ✓ 초고성능 병렬파일시스템(HCSF)를 데이터 노드로 활용하는 GPU DB 솔루션 등장



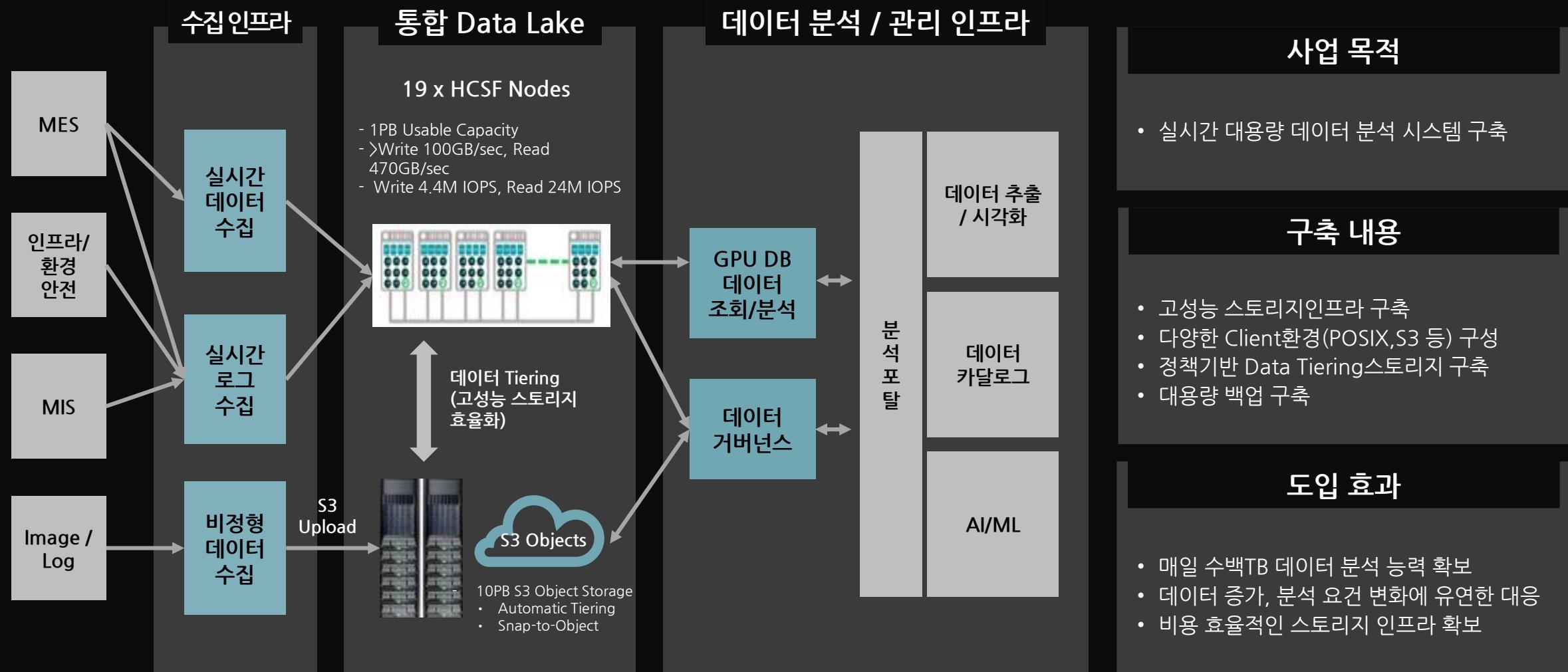
데이터 레이크하우스

데이터 웨어하우스 + 데이터 레이크 = 데이터 레이크하우스

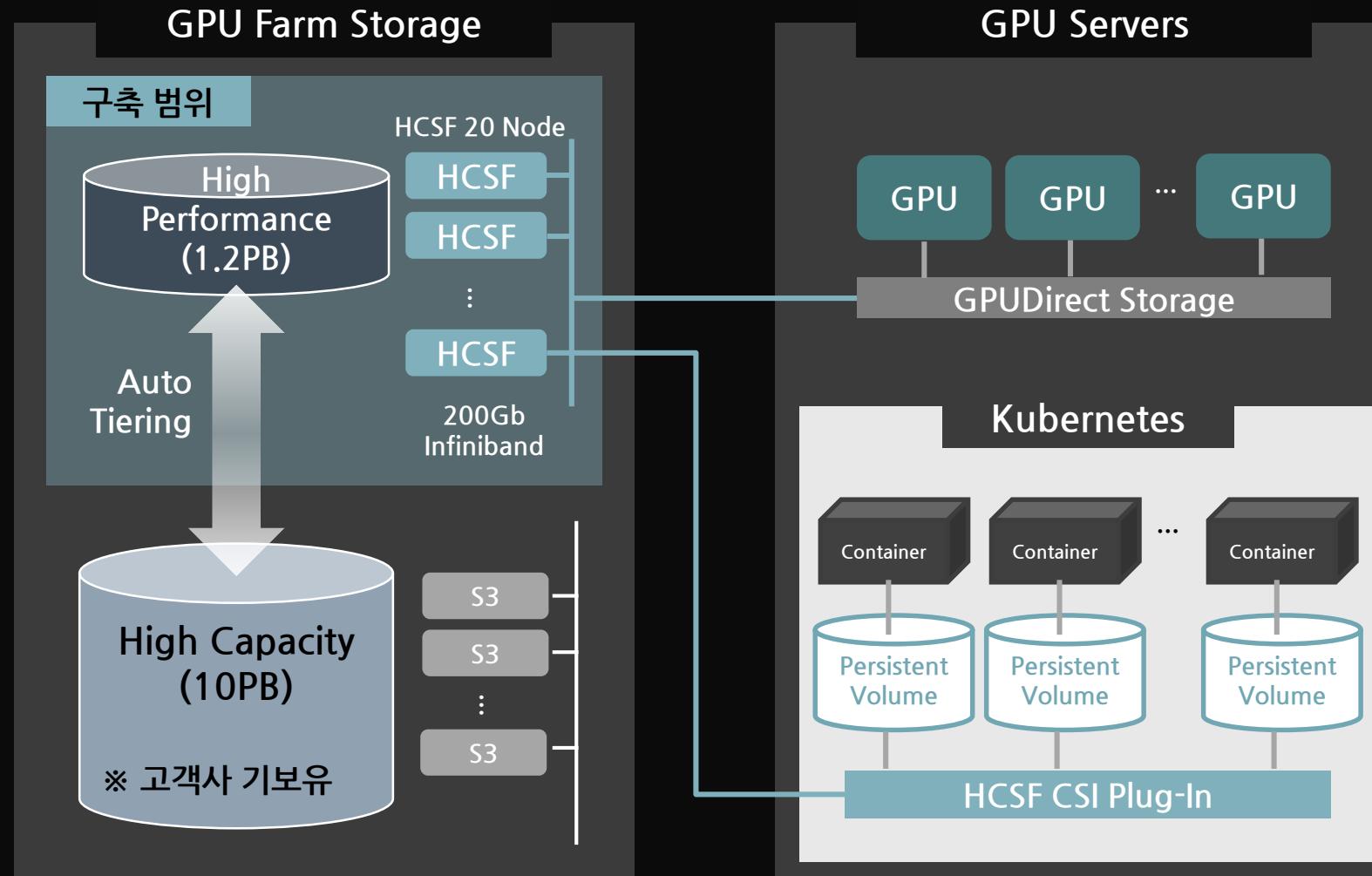
- ✓ 오브젝트 스토리지(HCP)의 개방형 테이블 형식(MetaData)을 제공하는 Apache Iceberg 등장으로 Spark, Trino, PrestoDB, Flink, Hive, Impala 를 통해 테이블 데이터 조회 가능한 **데이터 레이크하우스** 구축
- ✓ 오브젝트 스토리지(HCP)를 데이터 노드로 활용하는 DW솔루션 등장
- ✓ 초고성능 병렬파일시스템(HCSF)를 데이터 노드로 활용하는 GPU DB 솔루션 등장



사례 01. 국내 제조업 데이터 분석 체계 구축 사업



사례 02. 국내 R&D기업



사업 목적

- GPU 서버와 전용 스토리지를 도입하여 자연어 처리 알고리즘 학습 성능 개선

구축 내용

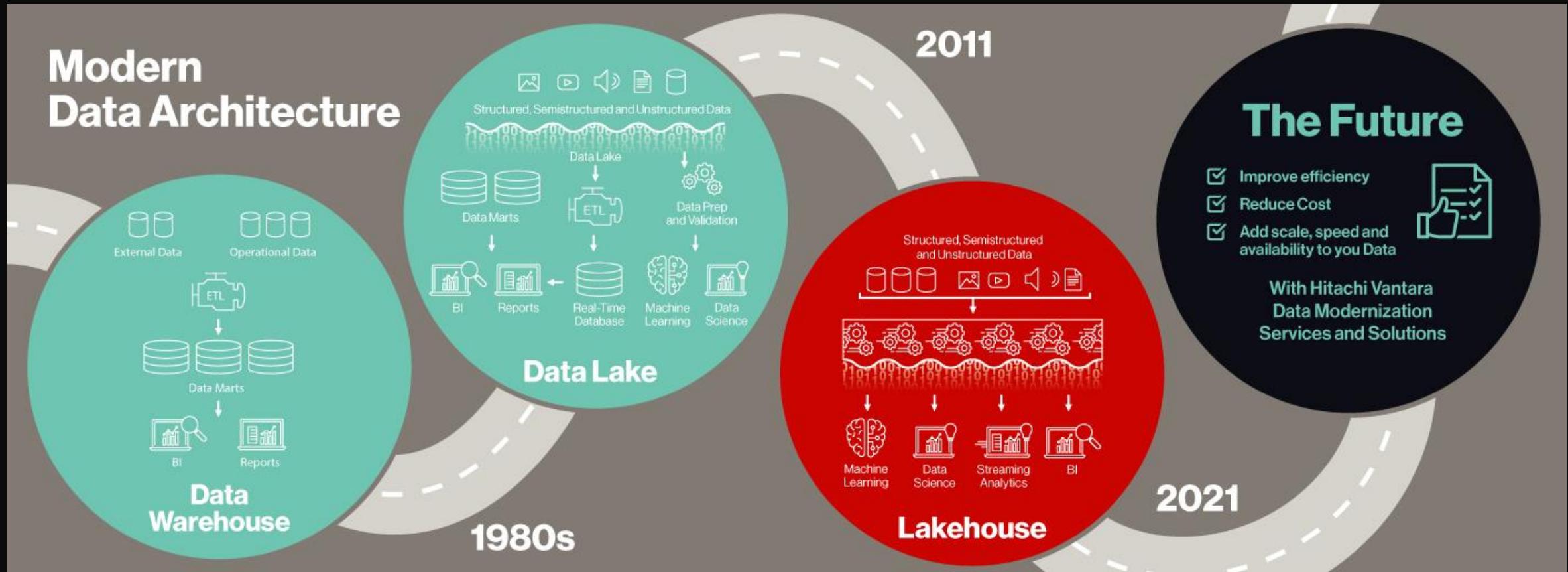
- 고성능(NVMe) 스토리지 구축
- GPUDirect Storage로 응답 성능 최적화
- Container Storage Interface Plug-In 제공으로 볼륨 운영 최적화 / 효율화 지원
- 고객사 S3 스토리지와 Tiering 구성

도입 효과

- 자연어 처리 기계학습 연산 성능 보장
- 서버 인프라 구성 고려한 스토리지 최적화
- 자체 S3 스토리지 활용성 확보

현대적 데이터 아키텍처

데이터 레이크의 다양하고 방대한 데이터를 통해 인사이트 도출
새롭게 도출된 인사이트를 통한 가치 창출



Source : Hitachi Vantara



감사합니다.