

AI&BIG DATA를 위한 데이터레이크

효성인포메이션시스템 권동수 전문위원

효성인포메이션시스템



데이터는 21세기 원유?

데이터는 21세기 원유인 대체품? 아니면 그 이상의 가치가 존재?



다양하고 복잡하며 고비용의 데이터 분석 요구사항

각 워크플로 단계에는
고유한 컴퓨팅, 스토리지 및
네트워킹 요구사항 존재

데이터 관리 및 통합 문제
발생으로 데이터 사일로 발생

비용과 시간이 증가하고
근본적인 인사이트 도출에
방해 요소



데이터 레이크 프로젝트의 함정

8 기업당 평균 8개의
데이터레이크가 존재함
— Forrester

분산된 데이터,
서로 다른
접근 방법

1%
의 데이터만이 분석됨
— McKinsey

자동화 불가,
IT 운영자
수작업에
지나치게 의존

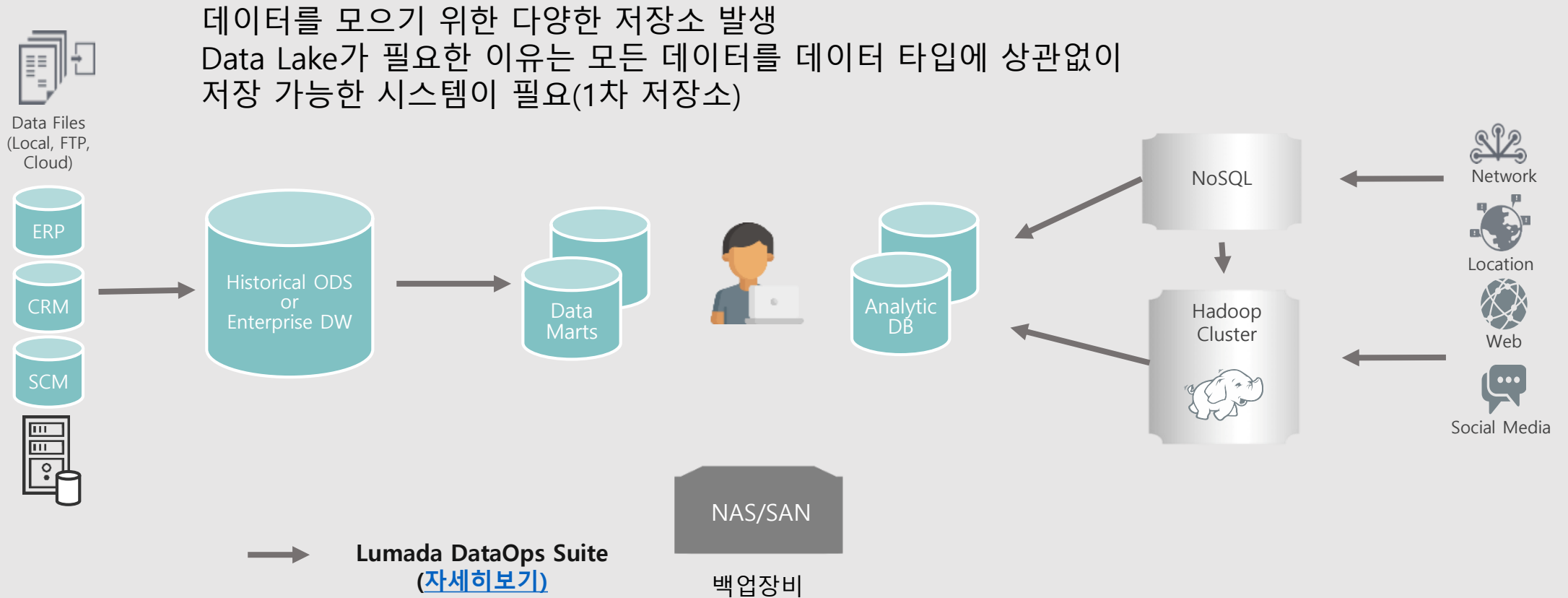
용도를
모르거나
찾기 어려운
데이터

표준 데이터
운영관리방안
부재

60%의 빅데이터
프로젝트가 실패함
— Gartner

데이터 흐름

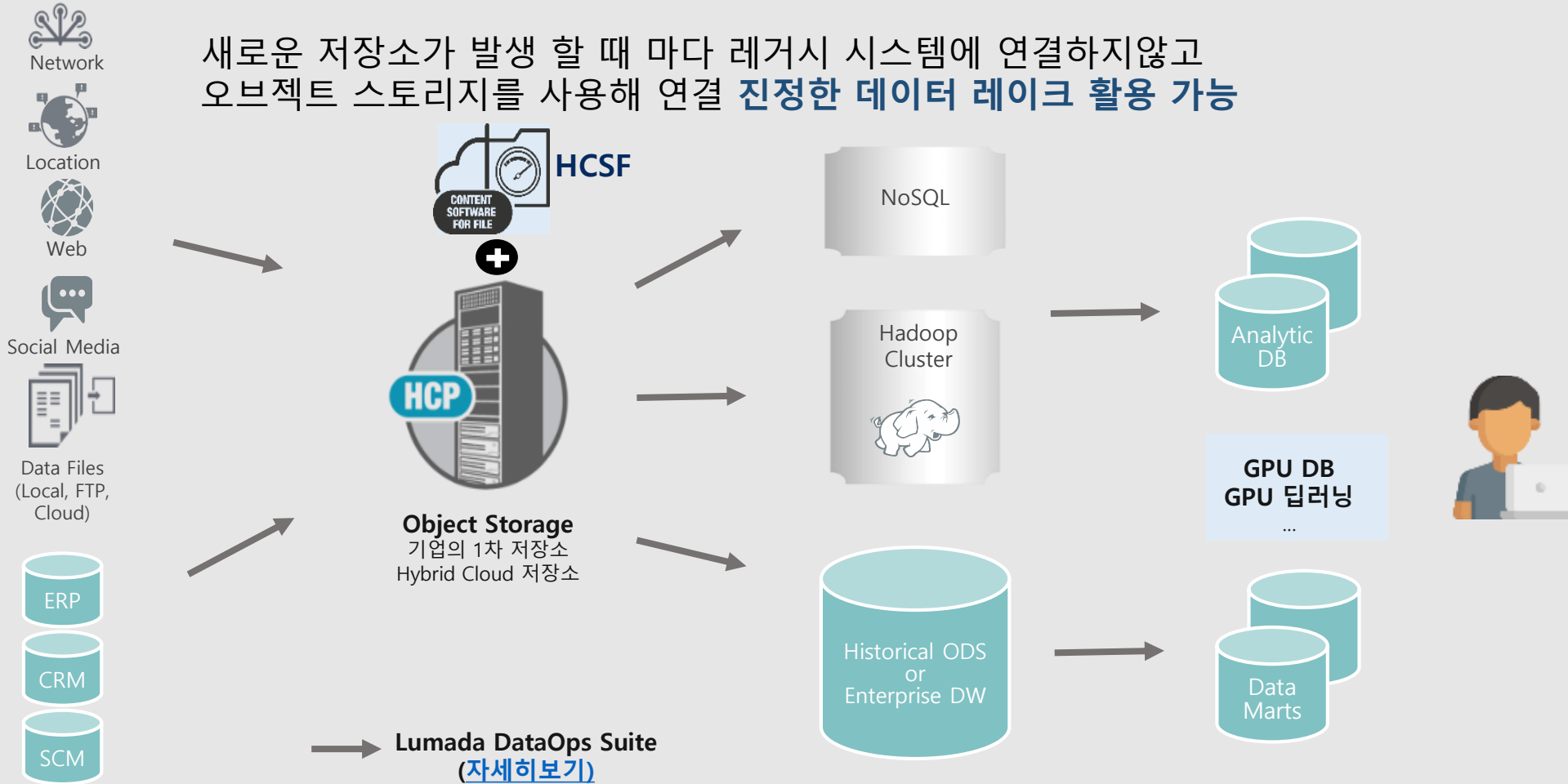
DW 정보계시스템, Big Data Hadoop 시스템??



데이터 흐름

DW 정보계시스템, Big Data Hadoop 시스템, 데이터 레이크

새로운 저장소가 발생 할 때 마다 레거시 시스템에 연결하지 않고
오브젝트 스토리지를 사용해 연결 진정한 데이터 레이크 활용 가능

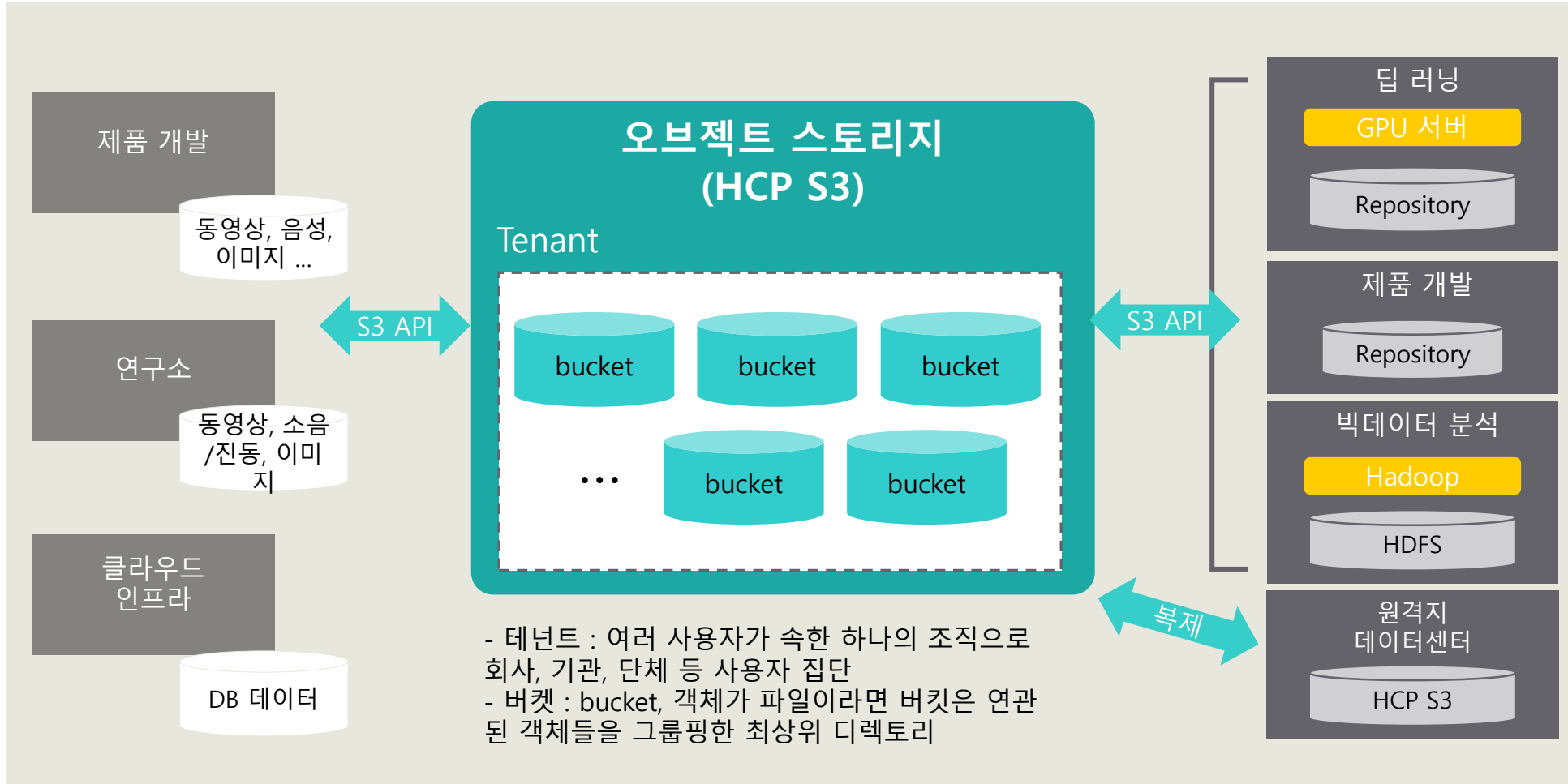


사례



Customers Wins and Successes

제조 사례 - 비정형 데이터 환경 적용

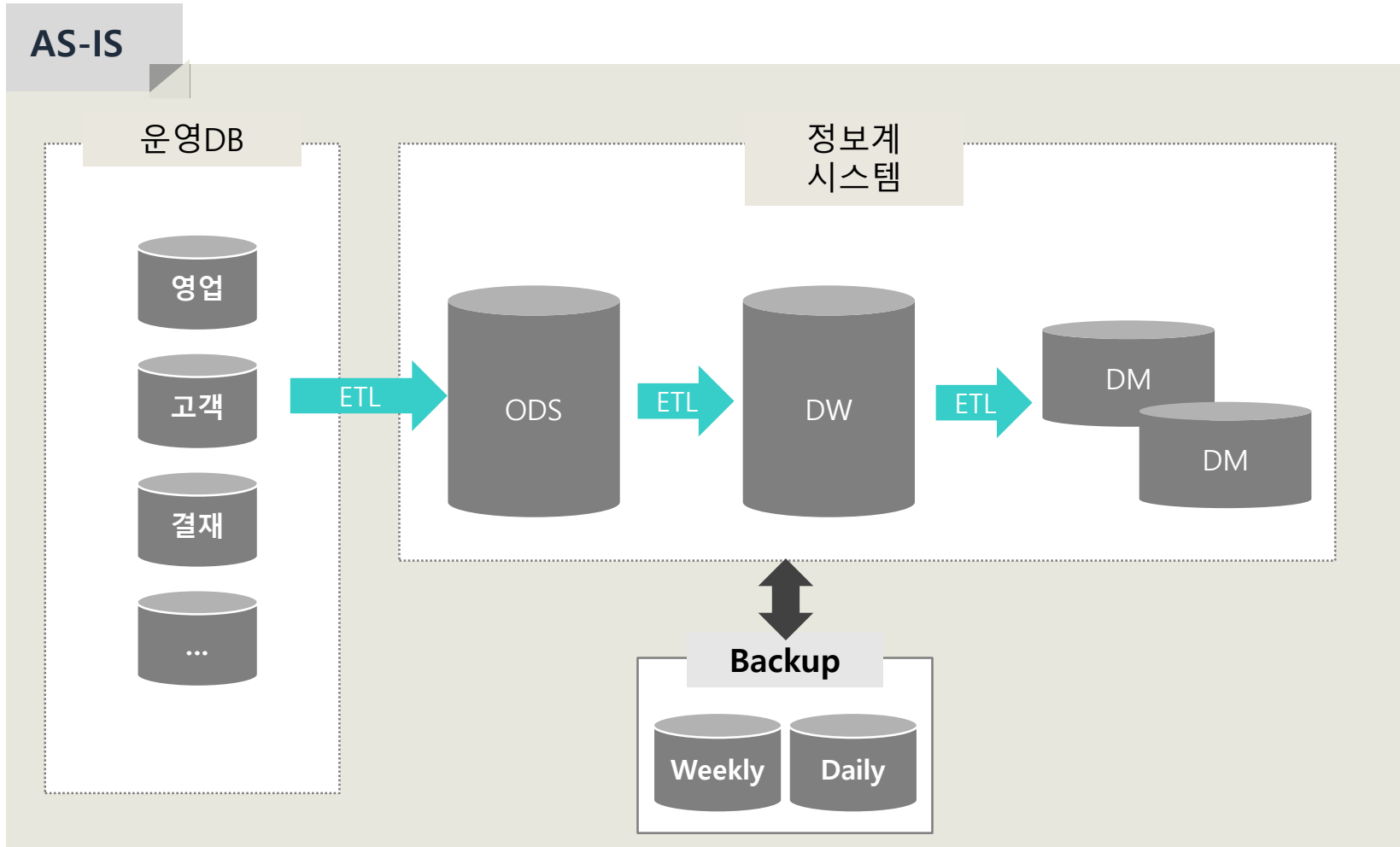


✓ 솔루션

- 대량 비정형 데이터 운영 위한 성능과 확장성 보장
- S3 전환을 위한 HIS 기술 서비스
 - 데이터 이관, S3 API 개발 가이드

Customers Wins and Successes

통신사 사례 - DW 환경 적용



Challenges

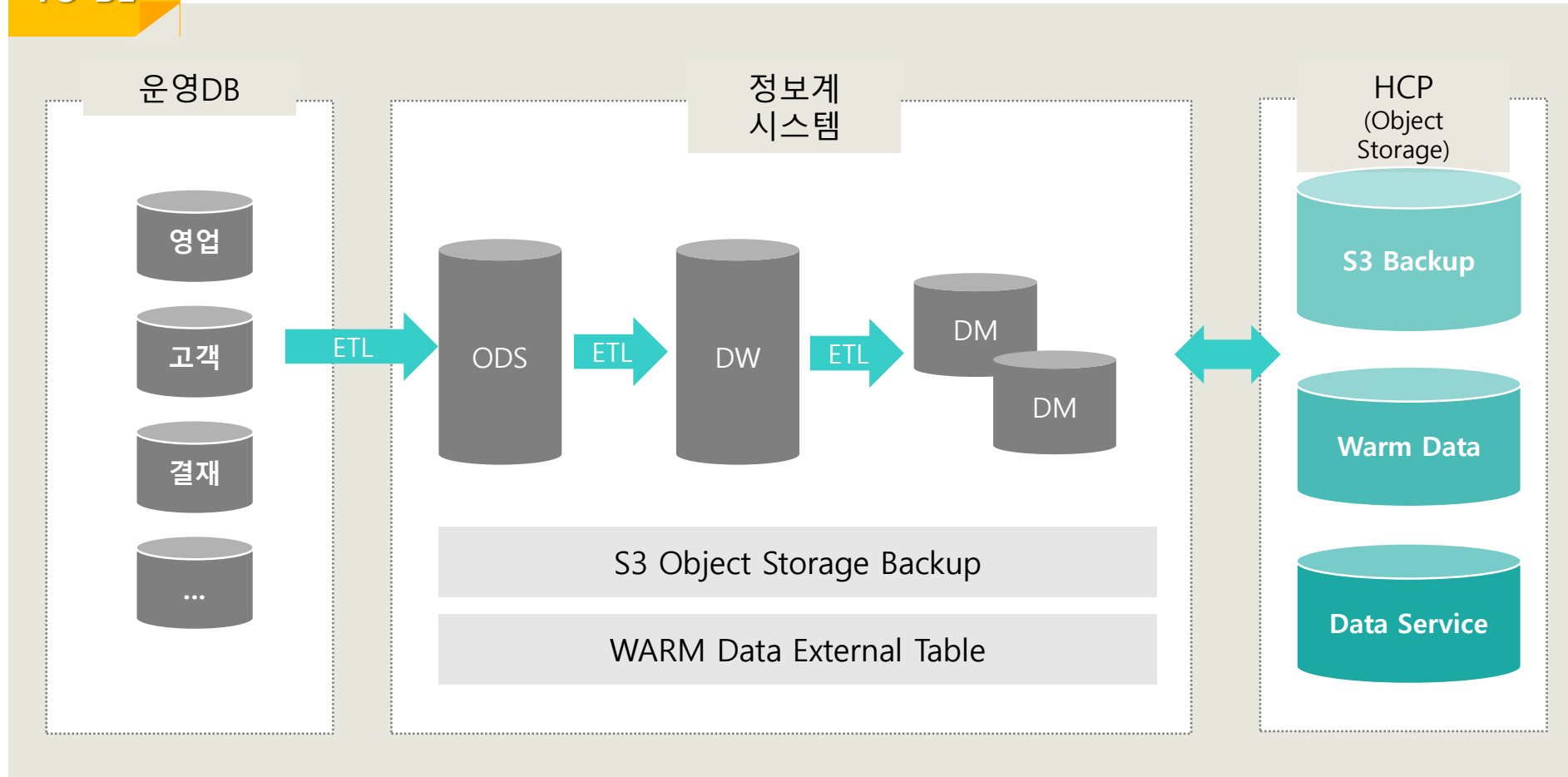


- ✓ DWA 증설 비용
- ✓ 차세대 시스템 구축
- ✓ 데이터 서비스 효율성 개선

Customers Wins and Successes

통신사 사례 - DW 환경 적용

TO-BE



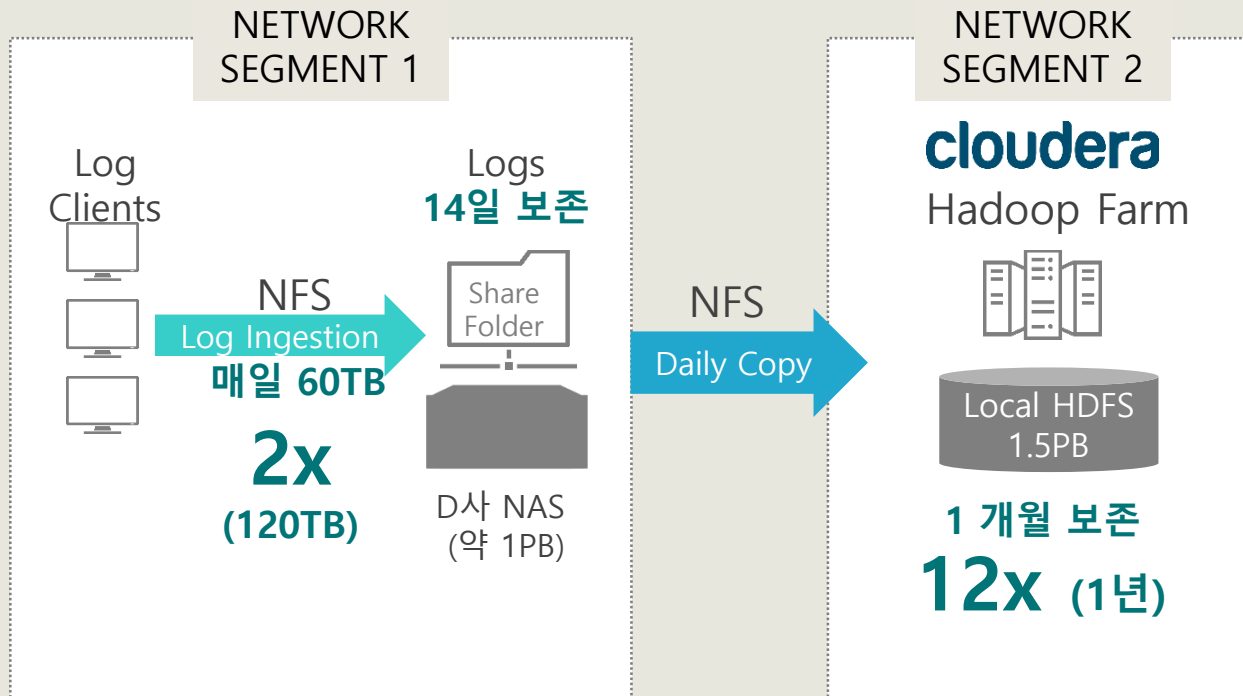
✓ 솔루션

- 고비용 시스템의 비용 절감 효과
- 새로운 데이터 서비스 방향 확보
- S3 도입 위한 HIS 기술 지원
 - 인프라 아키텍트, 데이터 아키텍트, S3 API 개발 가이드

Customers Wins and Successes

해외 정부사례 – Hadoop환경 적용

AS-IS



Growth Requirements



24x

Challenges

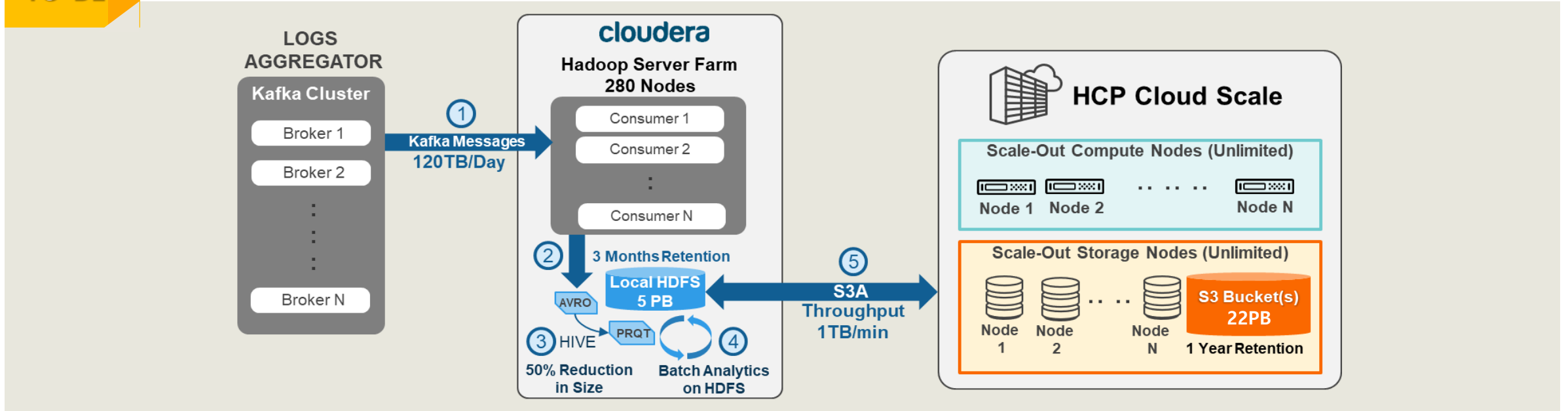


- ✓ NAS와 HDFS 동시 증설 비용
- ✓ 노드 증설에 따른 불안정한 NFS 마운트

Customers Wins and Successes

해외 정부사례 – Hadoop환경 적용

TO-BE



✓ 기존

- 로그파일 분석을 위해, 파일을 NFS에 수집하고 분석을 위해 다시 Hadoop에 전송
- 로그 양이 증가하고, 상관분석을 위한 데이터 보관 주기가 늘어남에 따라 NAS 적재 성능 이슈 발생
- NAS 및 Hadoop 스토리지 증가에 대한 비용 및 작업 이슈 발생

✓ 솔루션

- 실시간 대용량 로그 처리 방식으로 변경 및 파일 포맷 변환
- 데이터 수집 성능 및 안정성을 높이고, Hadoop 저장공간 50% 절감 : 하둡 분석(Hive/Impala) 성능 향상을 위한 Hadoop-Obj. Storage 간 목표 throughput 제시
- 원할한 분석을 위한 적정 Throughput을 찾고, 물리적 compute/storage 노드 구성 향후 개별 노드 확장이 가능한 유연한 환경 제시

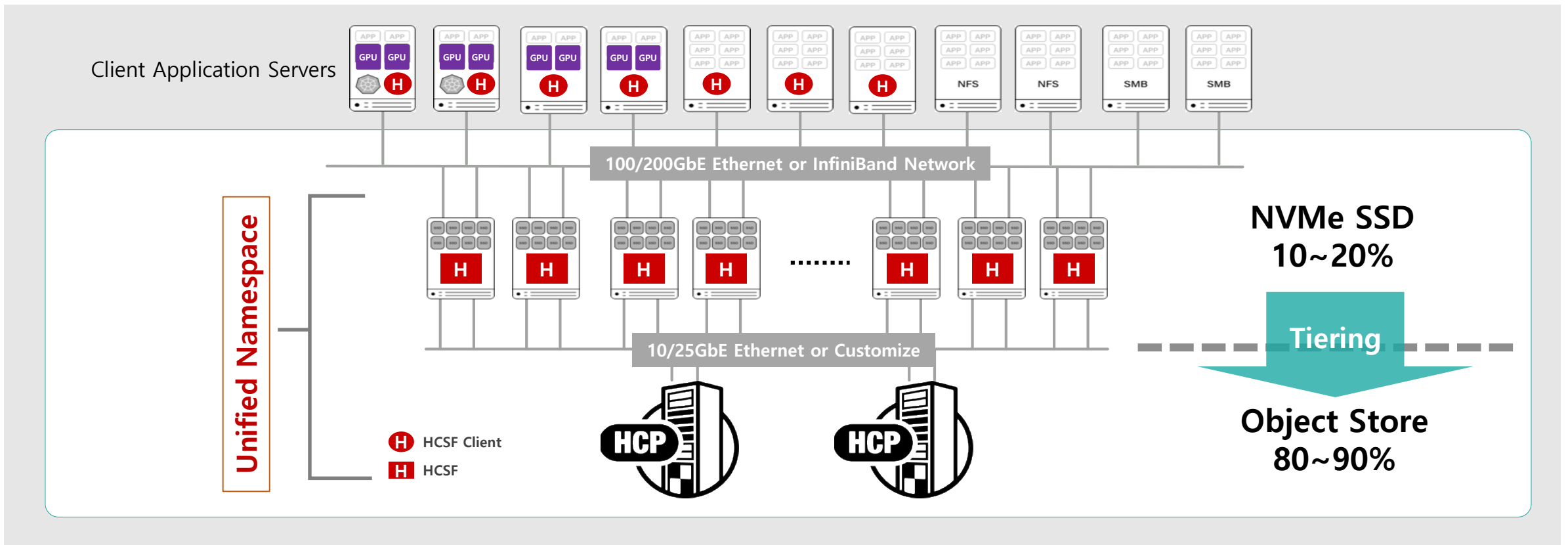
HCSF(초고성능 병렬 파일시스템) 사례



지질 데이터 분석 및 정보 제공 서비스 환경(22PB) - 석유/가스 탐사 및 생산에 정보 제공을 위한 분석 시스템



새로운 분석 워크로드를 위한 고성능 스토리지(40PB)- 실시간 부정거래 방지 시스템 및 고성능 데이터 보호



데이터 레이크, 왜 효성인포메이션시스템?



Thank you

