

오브젝트 스토리지 기반의 데이터 레이크 및 사례

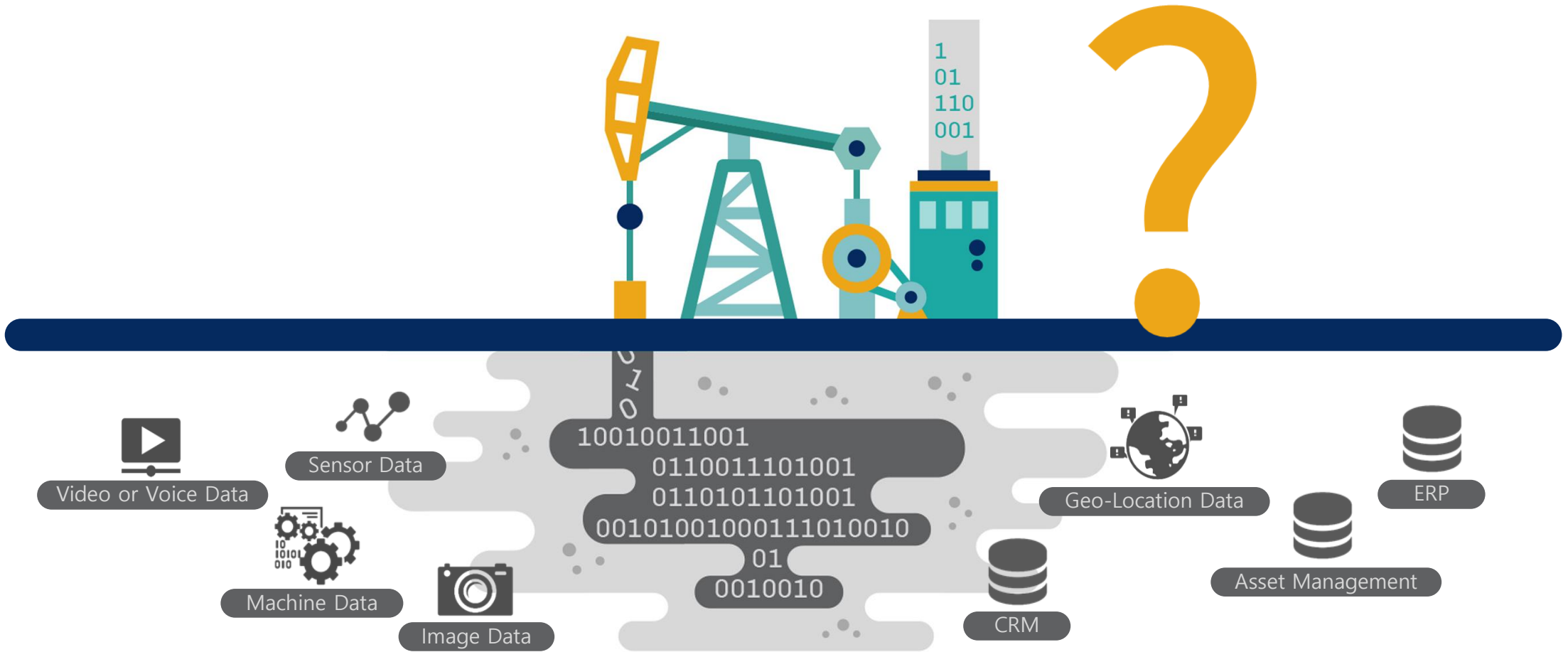
효성인포메이션시스템 권동수 전문위원

효성인포메이션시스템



데이터는 21세기 원유?

데이터는 21세기 원유인 대체품? 아니면 그 이상의 가치가 존재?



데이터 레이크 프로젝트의 함정

8 기업당 평균 8개의
데이터레이크가 존재함
— Forrester

분산된 데이터,
서로 다른
접근 방법

자동화 불가,
IT 운영자
수작업에
지나치게 의존

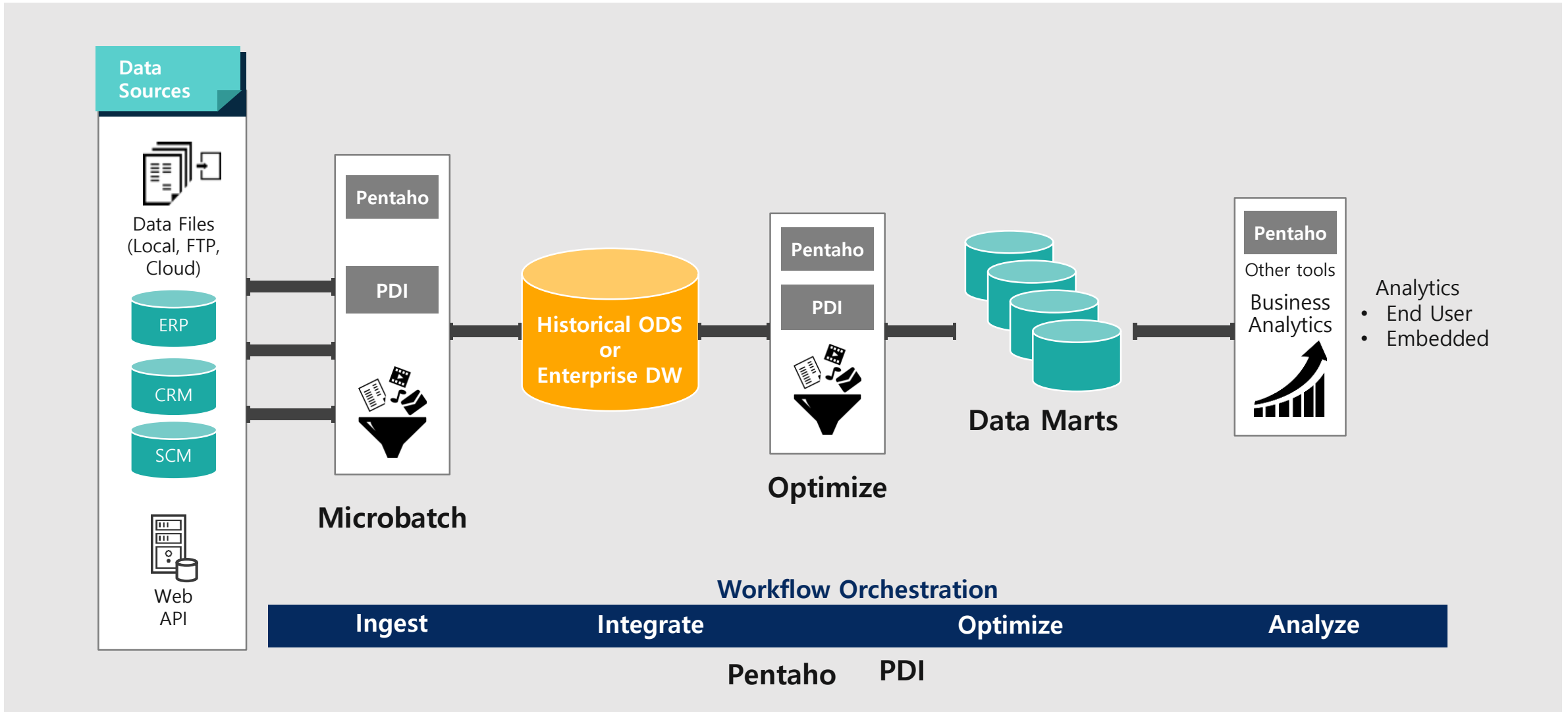
1%의 데이터만이 분석됨
— McKinsey

용도를
모르거나
찾기 어려운
데이터

표준 데이터
운영관리방안
부재

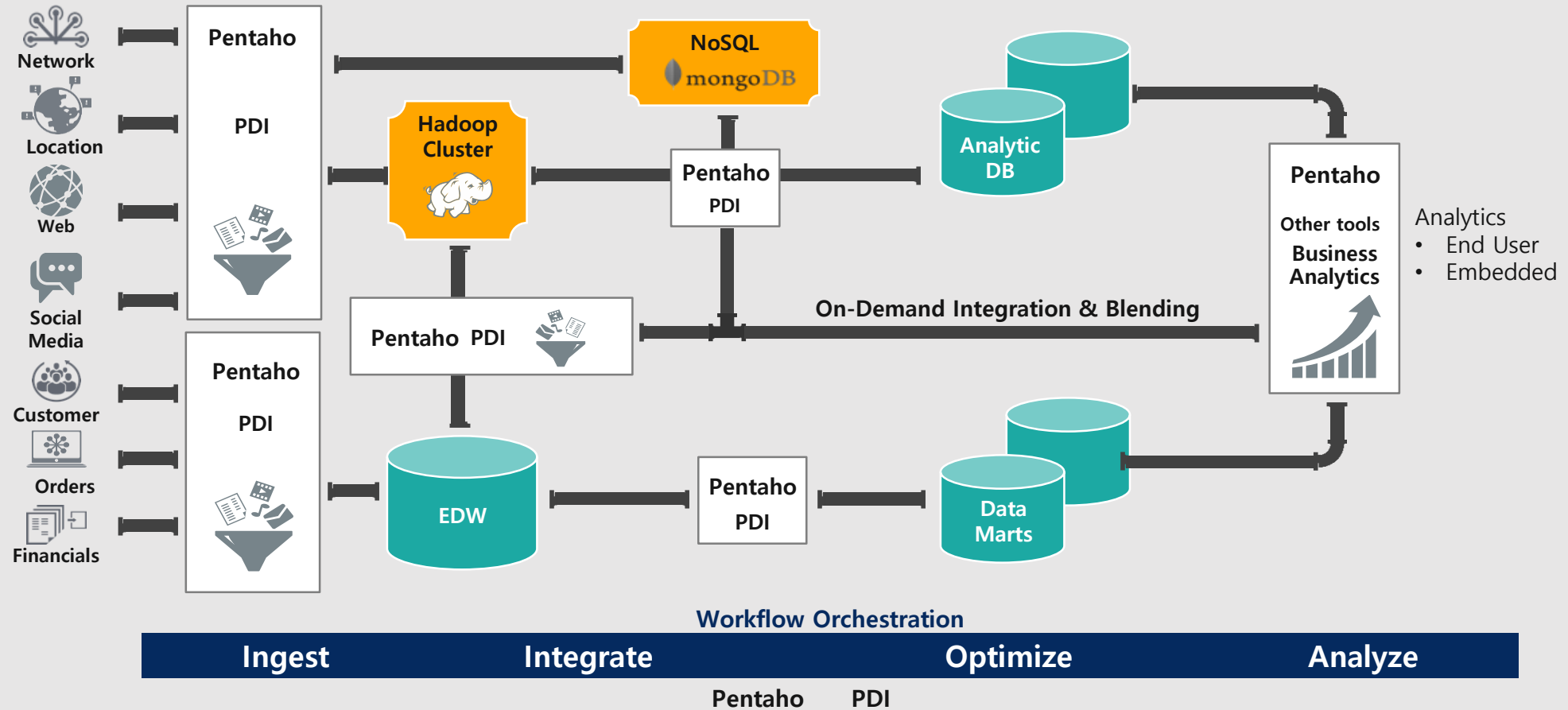
60%의 빅데이터
프로젝트가 실패함
— Gartner

정통적인 DW 아키텍처



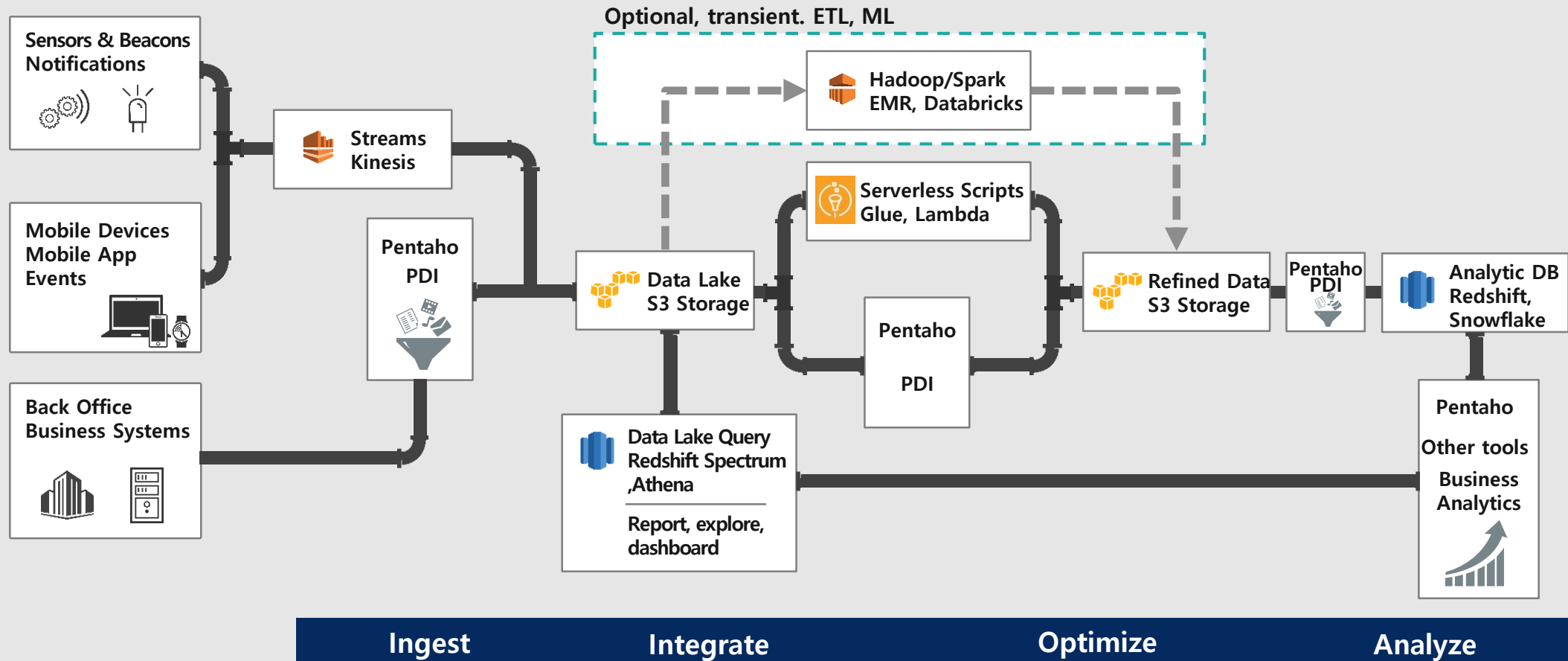
빅데이터 분석 아키텍처

Integrating Hadoop, No SQL and Analytic Databases



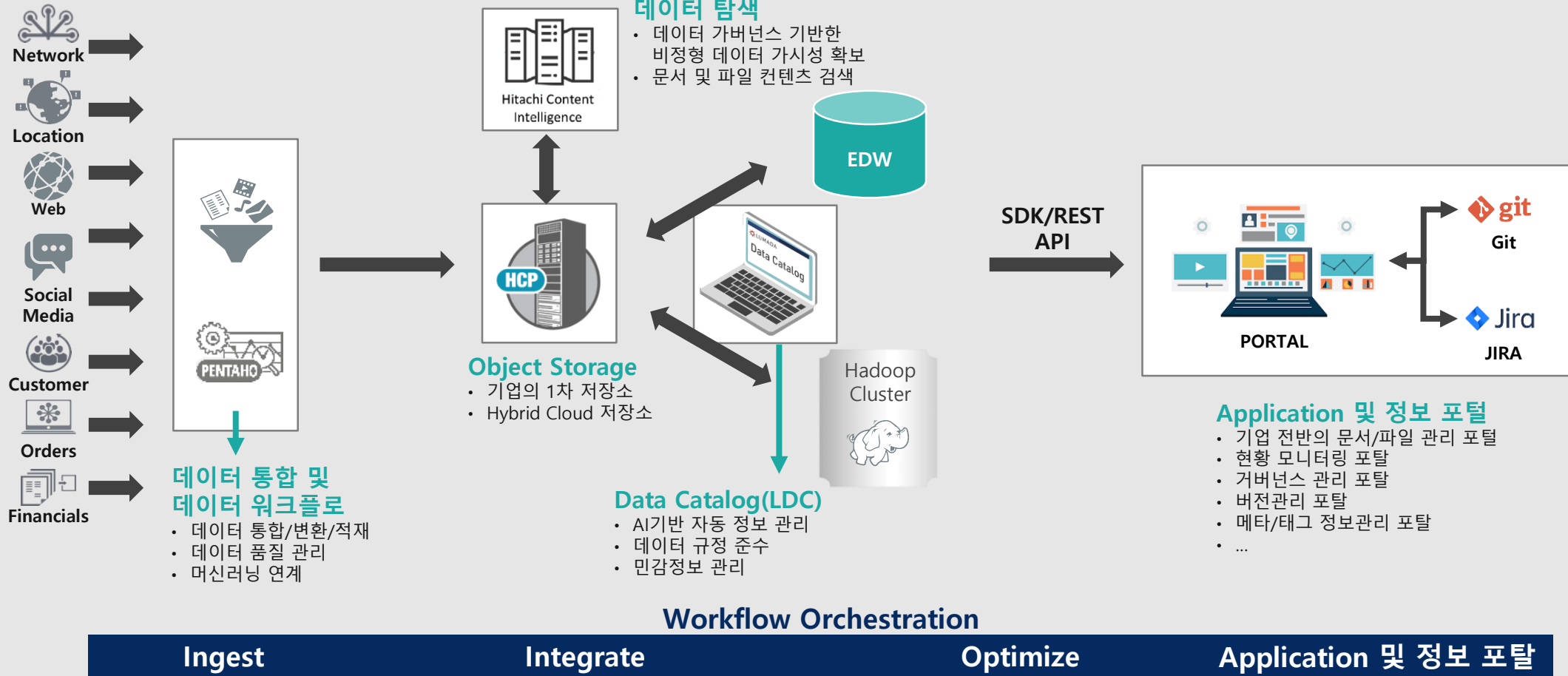
클라우드 분석 아키텍처

AWS Cloud Example



데이터 레이크 - 빅데이터 아키텍처

Object Storage 기반 빅데이터 아키텍처



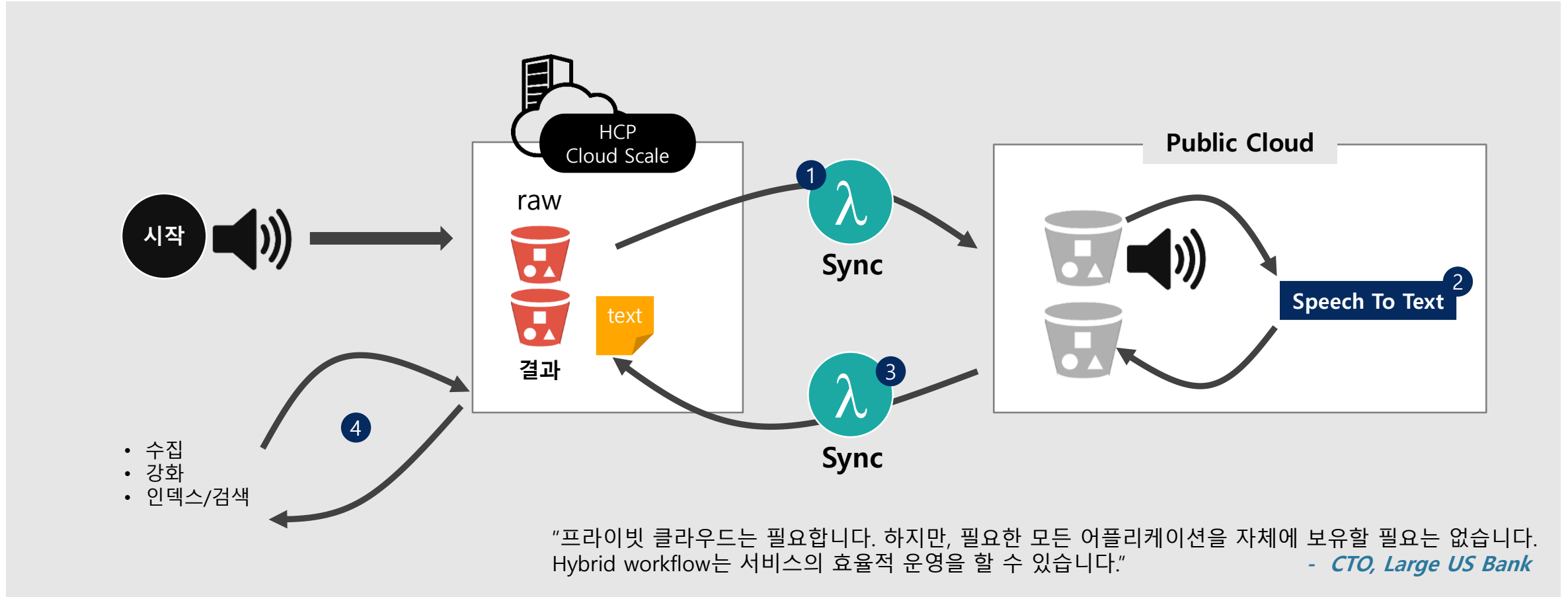
사례



Customers Wins and Successes

US Bank 사례 – Hybrid Cloud 환경 적용

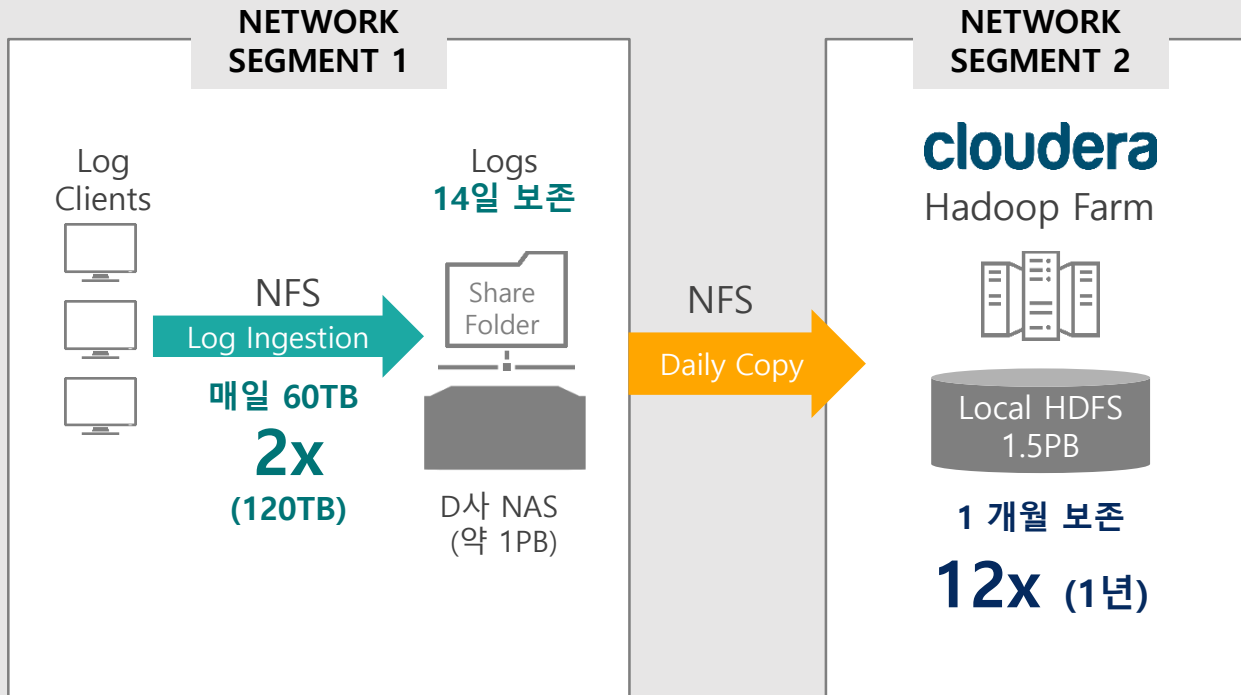
- Hybrid workflow 는 데이터 저장/전송 비용 절감과 데이터 보안 확보가 가능합니다.



Customers Wins and Successes

싱가폴 정부사례 – Hadoop 환경 적용

AS-IS



Growth Requirements



24x

Challenges



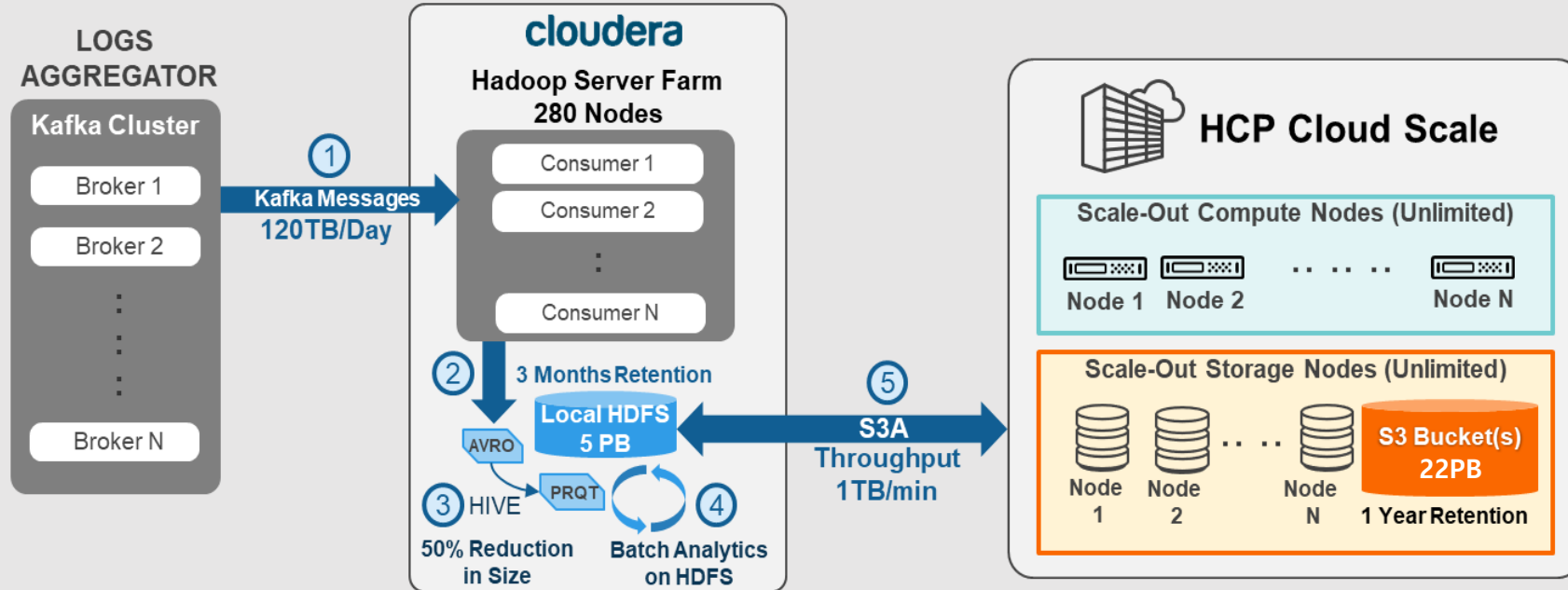
✓ NAS와 HDFS 동시 증설 비용

✓ 노드 증설에 따른 불안정한 NFS 마운트

Customers Wins and Successes

싱가폴 정부사례 – Hadoop 환경 적용

TO-BE



✓ 기존

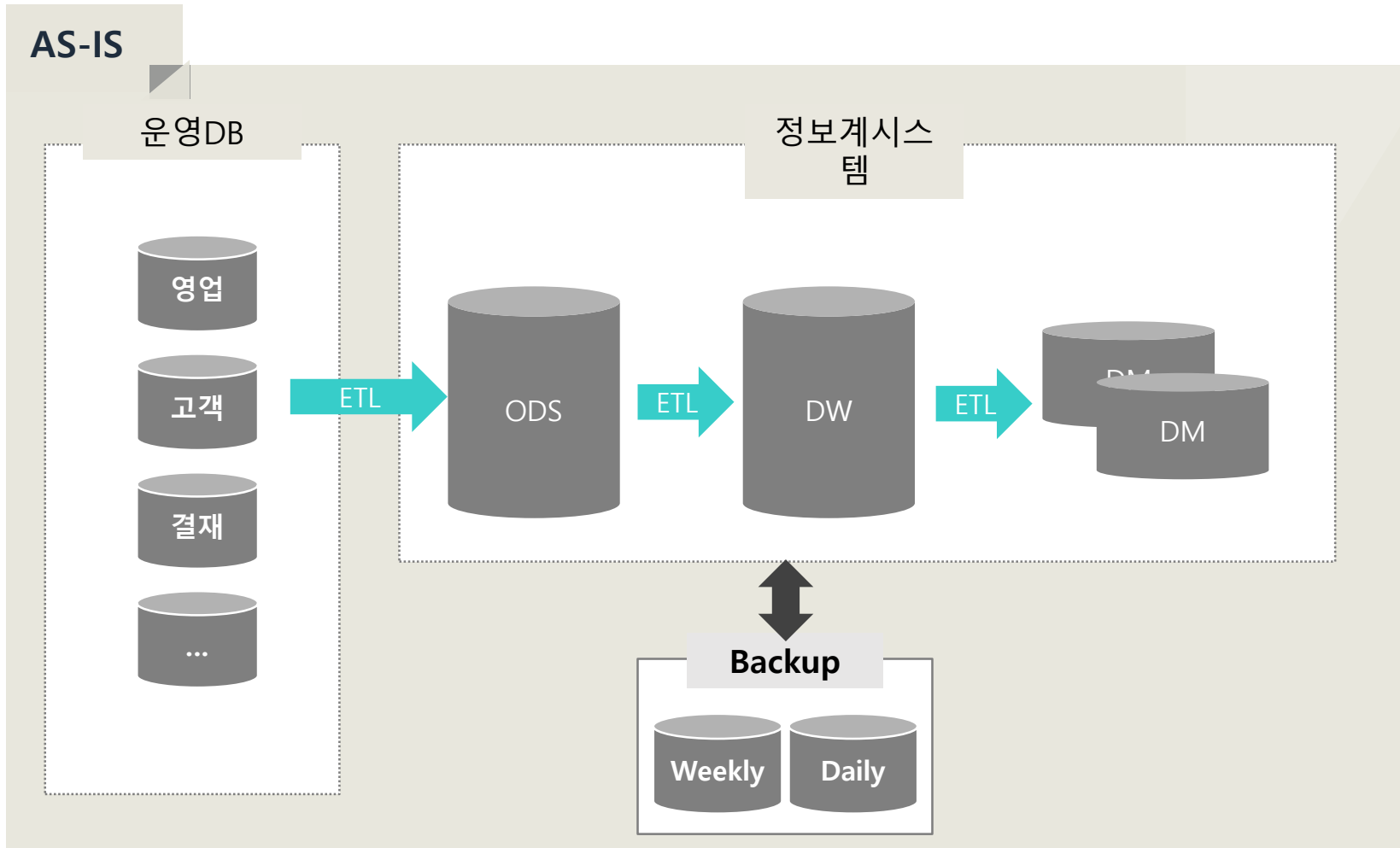
- 로그파일 분석을 위해, 파일을 NFS에 수집하고 분석을 위해 다시 Hadoop에 전송하였음
- 로그 양이 증가하고, 상관분석을 위한 데이터 보관 주기가 늘어남에 따라 NAS 적재 성능 이슈 발생
- NAS 및 Hadoop 스토리지 증가에 대한 비용 및 작업 이슈 발생

✓ 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

Customers Wins and Successes

통신사 사례 - DW 환경 적용



Challenges



✓ DWA 증설 비용

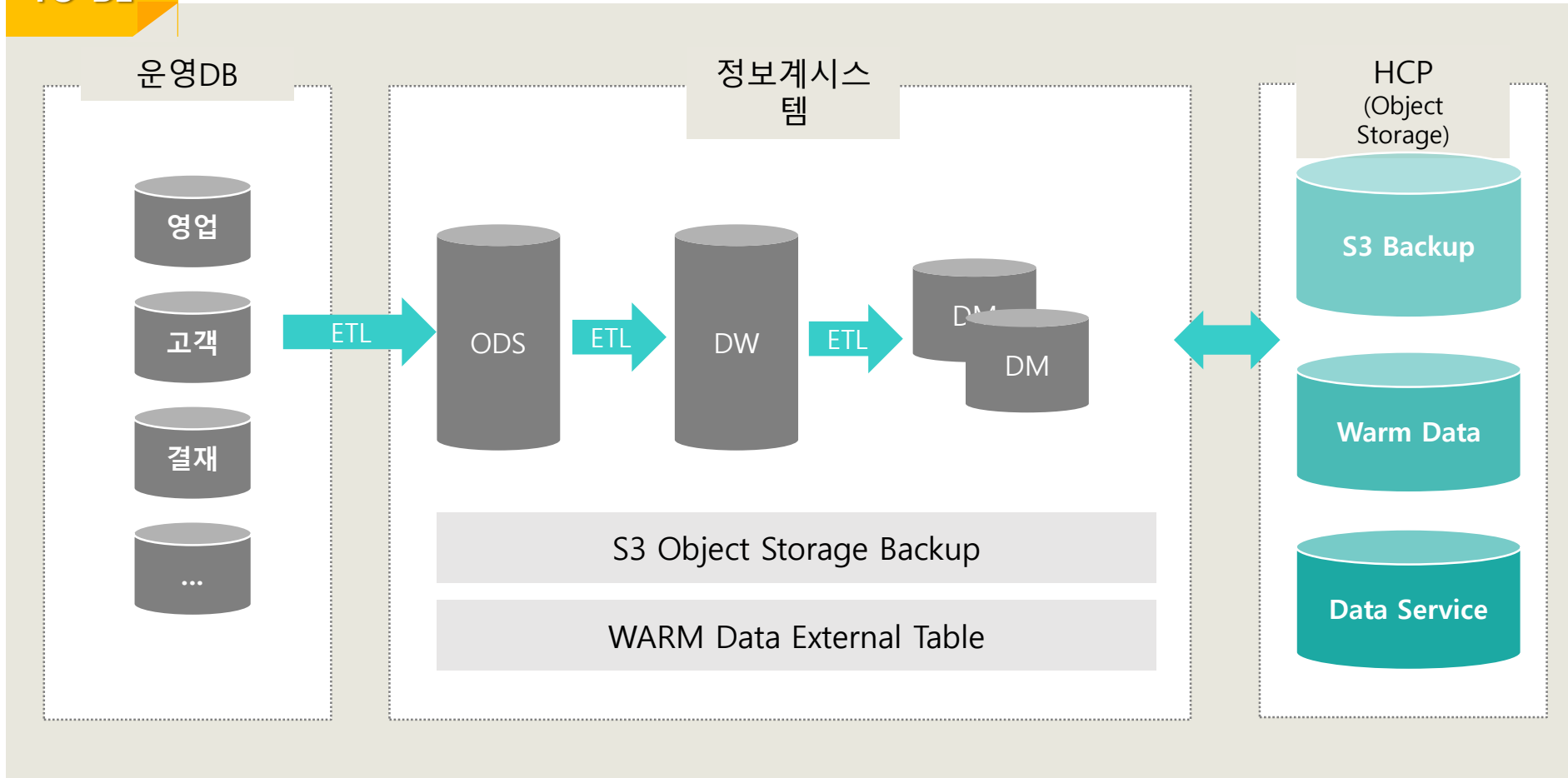
✓ 차세대 시스템 구축

✓ 데이터 서비스 효율성 개선

Customers Wins and Successes

통신사 사례 - DW 환경 적용

TO-BE



✓ 솔루션

- WORM 데이터 오브젝트 스토리지 전환으로 DW 시스템의 운영 리소스 확보
- 오브젝트 스토리지에서 데이터 서비스 제공하여 DW의 Data Mart 운영 부담 낮춤
- 오브젝트 스토리지 백업 전환으로 기존 백업 정책, 성능은 유지하면서 비용 절감

데이터 레이크, 왜 효성인포메이션시스템?



Thank you

