kt cloudware

Hadoop EcoSystem을 활용한
Hybrid DW 구축 사례

2013-05-02

KT cloudware / NexR
Project Manager 정구범
klaus.jung@{kt|nexr}.com

NEXR
TOWARD OPEN PLATFORM

KT의 대용량 데이터 처리 이슈

□ 적재 Data의 폭발적인 증가

→ LTE 등 초고속 무선 Data 통신 : 트래픽이 예상보다 빨리/많이 증가

→ 비통신(컨텐츠/플랫폼/Bio/금융 등) 영역의 대용량 Data 처리수요 증가

□ 비용절감

→ 주력매출의 정체 : 비싼 장비와 SW에 계속 투자해서는 이득이 없음

→ 내재화를 통한 자체 역량 확보 및 구축/확대/유지보수 비용 절감

□ 보다 정밀해지는 분석처리

→ 정밀한 분석 : 분석에 필요한 Data의 크기와 종류의 증가

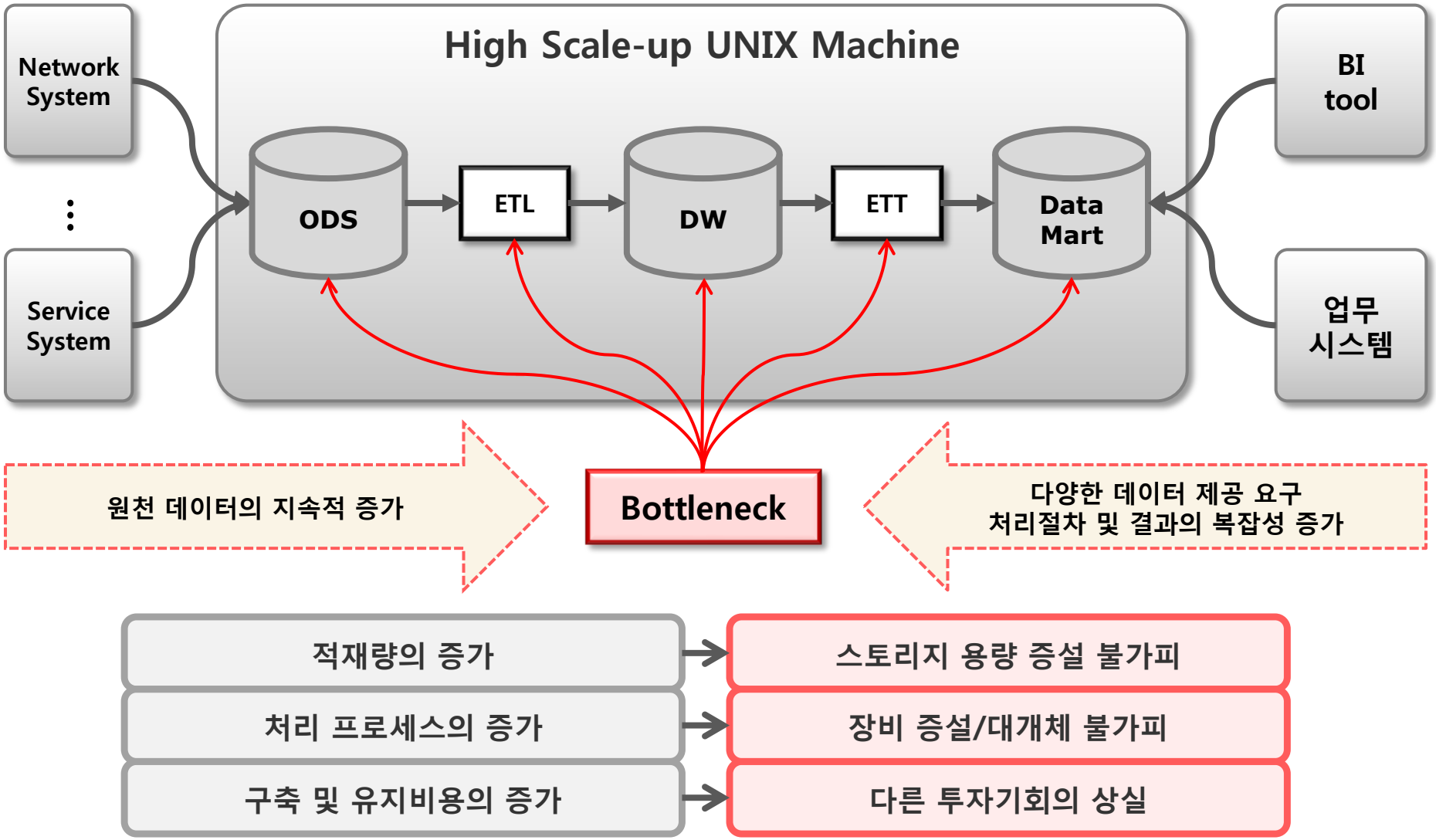
→ 교차분석 대상 및 관계요건 복잡성의 대폭 증가

□ 처리결과의 적시성 확보

→ Data의 크기와 복잡성이 증가해도 처리 결과는 항상 일정한 시간에 제공

→ 처리시간의 여유 부족 = 처리 중 사소한 문제라도 발생되면 적시 제공불가

상용DW 기반의 기존 분석 시스템



상용DW 대체의 전제 조건

❑ 경제성 확보

- 데이터와 성능 증가목표에 대응하여 **단계별 적기 투자**가 가능한 **비용 합리성** 확보
- **Commodity Hardware**에서 운용 가능한 **비용 효율성** 확보

❑ 동일한 처리결과 확보

- 기존의 **SQL을 최대한 활용**하여 적용할 수 있는 **호환성** 확보
- 기존의 업무처리 결과와 **동일한 처리결과** 확보

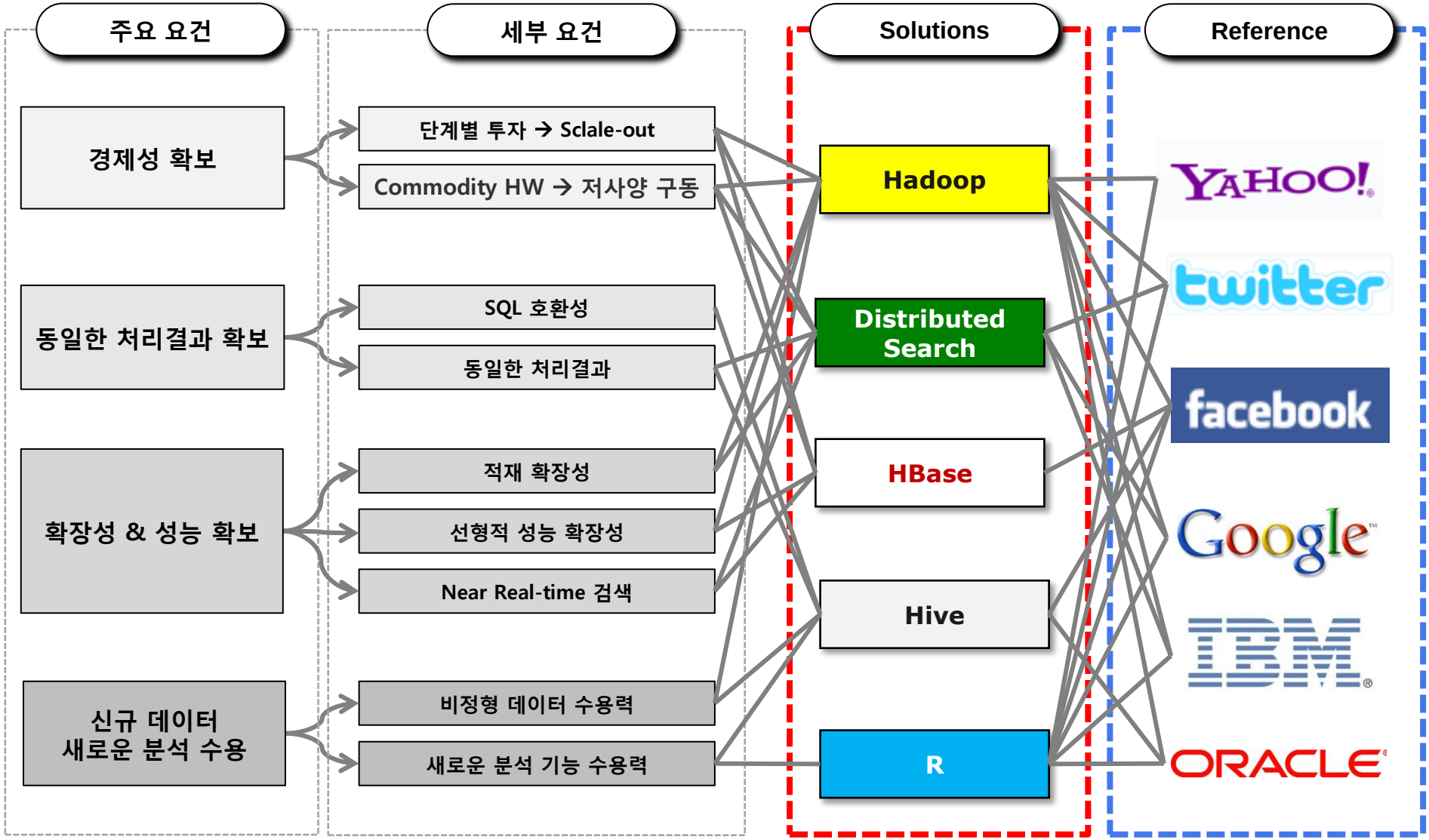
❑ 확장성과 Real-time 성능 확보

- 지속적으로 **증가하는 데이터의 수용**이 가능한 **적재 확장성** 확보
- 장비 추가에 따른 **처리성능**의 **선형적 확장성** 확보
- 원하는 조건의 데이터를 **Real-time** 이내에 확인 가능한 **성능** 확보

❑ 새로운 데이터/분석 기법에 대한 수용력 확보

- 기존 시스템 통합으로 데이터 포맷 등의 추가적용이 가능한 **비정형 데이터 수용력** 확보
- 새로운 분석 기법/알고리즘의 추가 적용이 용이한 **분석기능 수용력** 확보

요구사항 분석

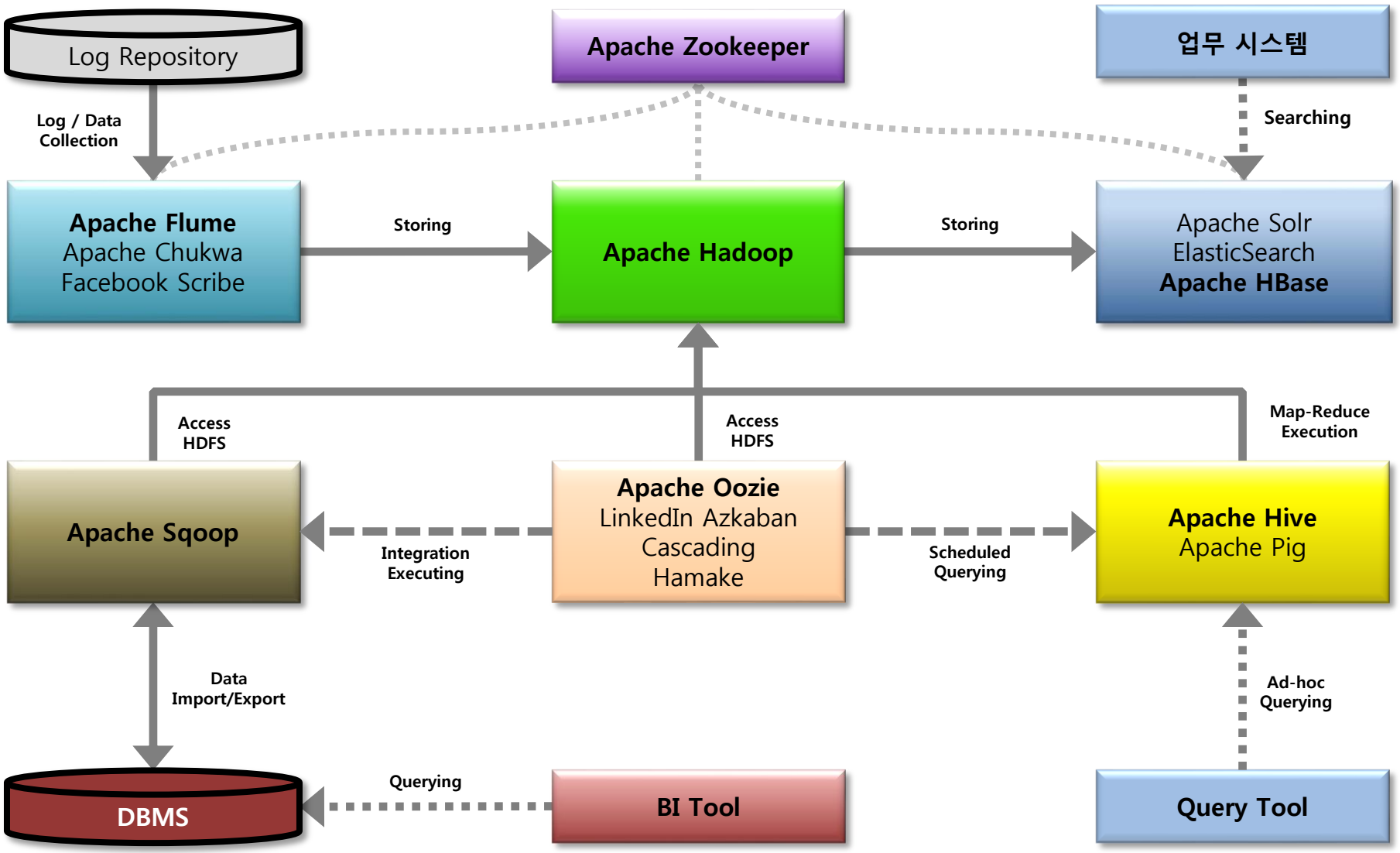

주요 요건
세부 요건
Solutions
Reference
경제성 확보
동일한 처리결과 확보
확장성 & 성능 확보
신규 데이터
새로운 분석 수용
단계별 투자 → Sclale-out
Commodity HW → 저사양 구동
SQL 호환성
동일한 처리결과
적재 확장성
선형적 성능 확장성
Near Real-time 검색
비정형 데이터 수용력
새로운 분석 기능 수용력
Hadoop
Distributed Search
HBase
Hive
R
YAHOO!
twitter
facebook
Google
IBM
ORACLE

Hadoop EcoSystem

- 구글은 방대한 데이터의 수집/적재/처리를 보다 저렴한 비용으로 현실화
- 범용부품 기반의 맞춤/조립식 저비용 하드웨어 기반, 모든 문제를 소프트웨어로 해결
- Hadoop = 구글의 비공개 기술을 모방, 오픈소스에 기반한 집단지성의 결정체

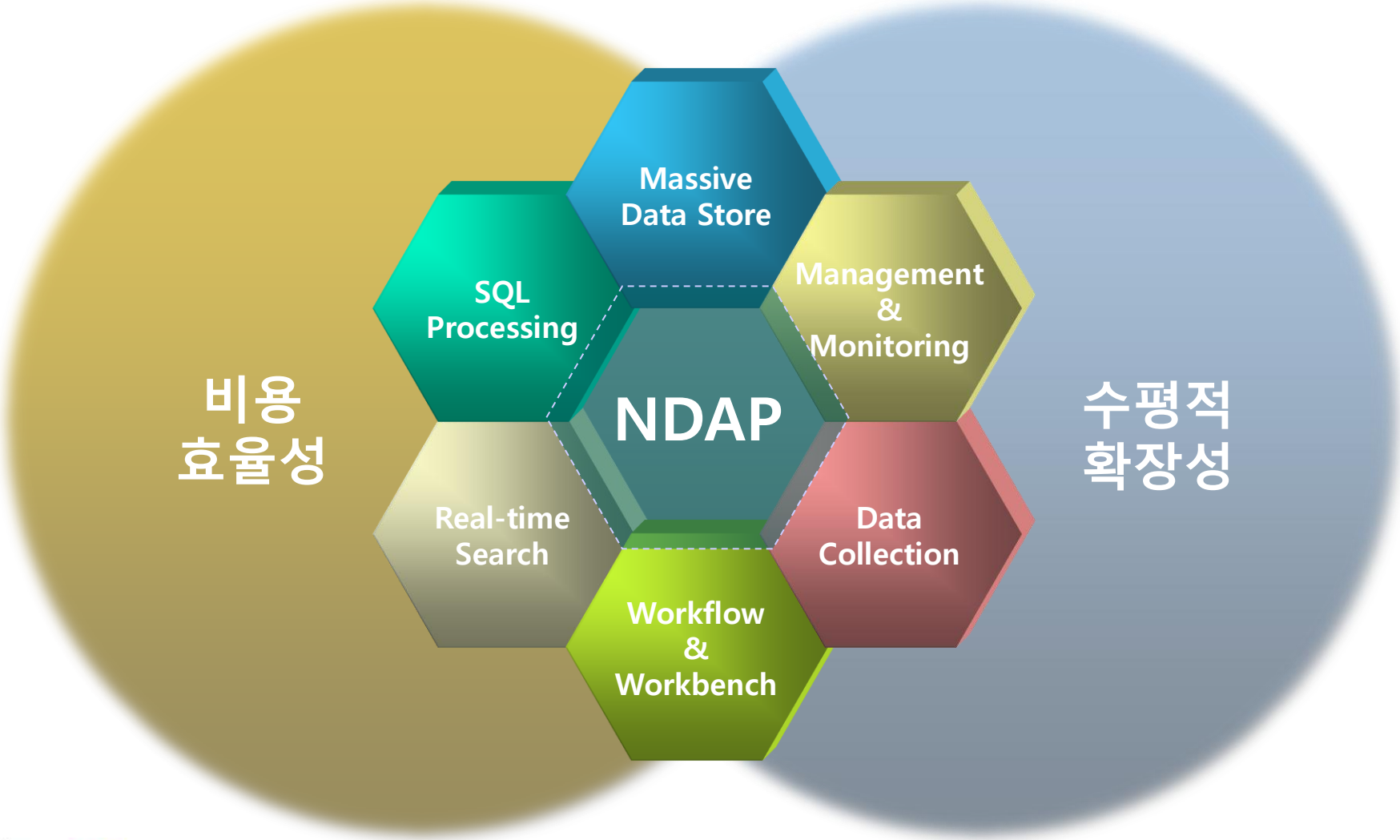
기술/서비스	Google Eco-System	Hadoop Eco-System
분산 파일 시스템	Google File System	Apache Hadoop (HDFS)
분산 데이터 처리	Map-Reduce	Apache Hadoop (MapReduce)
분산 락킹, 이벤트 통지	Chubby	Apache Zookeeper
분산 데이터 저장 (Row-level ACID)	BigTable	Apache HBase
SQL기반 데이터 처리	BigQuery (외부 서비스 제공. 최대 70TB)	Apache Hive
분산 데이터베이스 (Full ACID)	Spanner	-

Hadoop EcoSystem based Realization Architecture

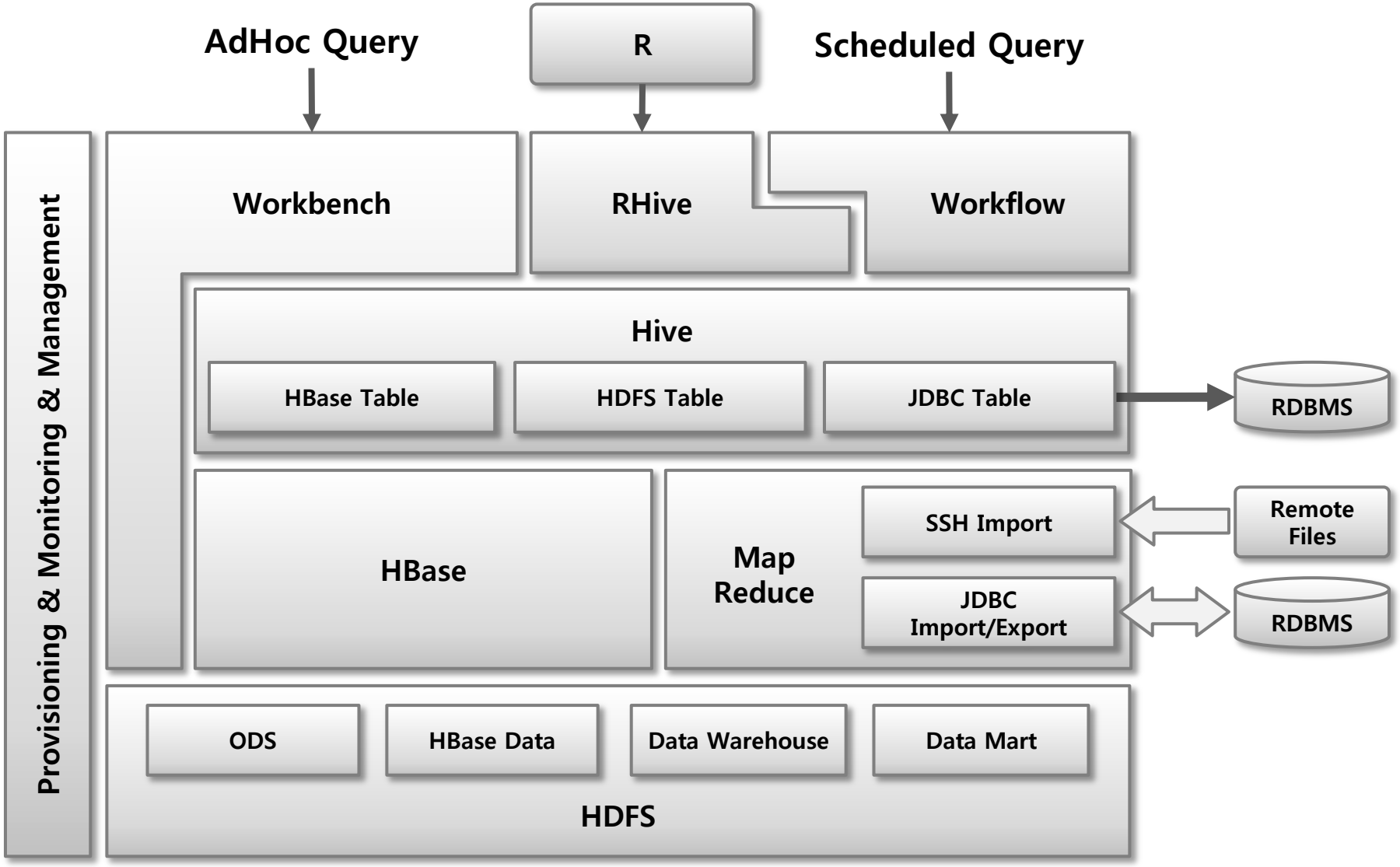


NDAP (NexR Data Analytics Platform)

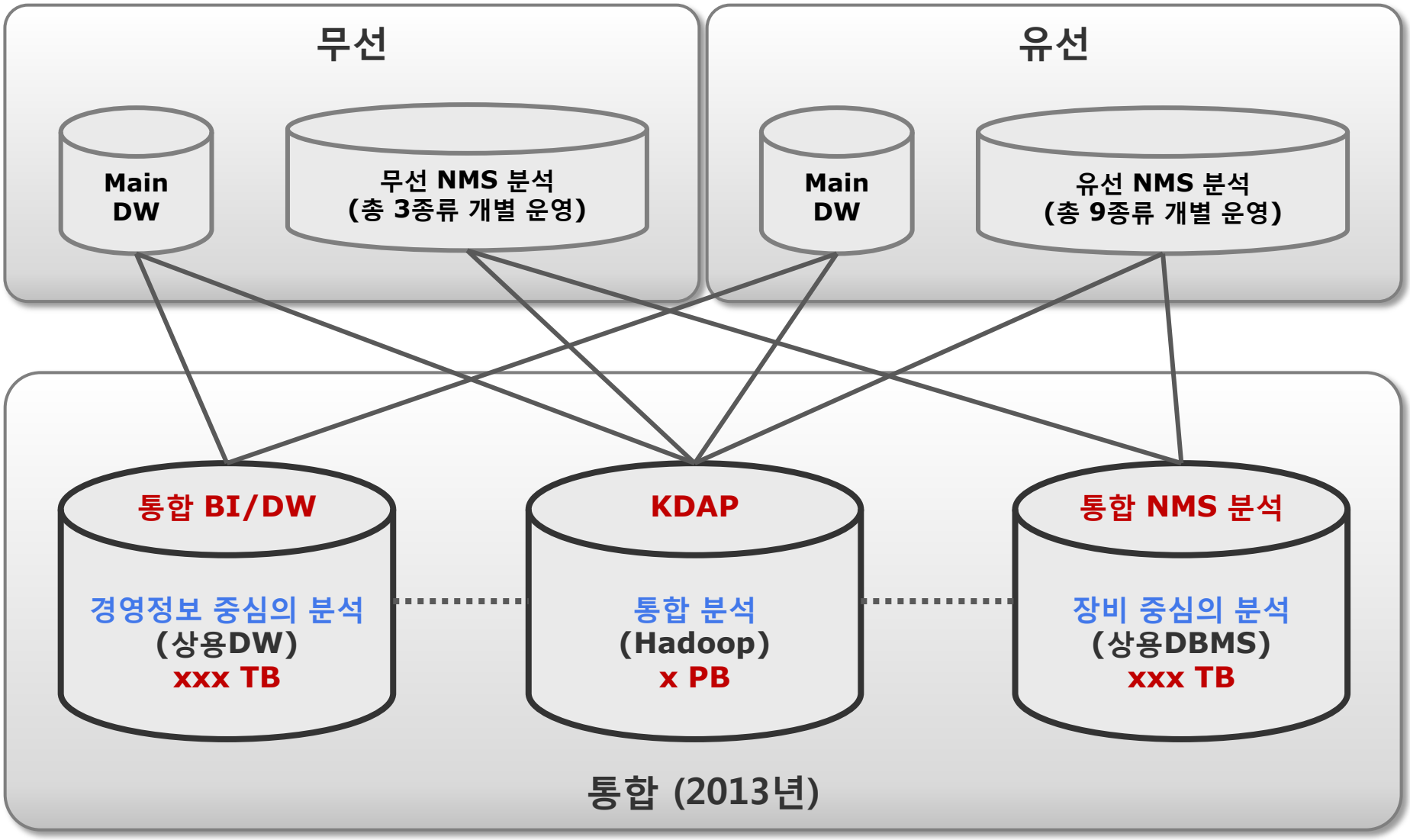
Big Data 처리를 위한 All-in-One 솔루션
: Big Data의 수집/저장/분석/검색/변환 등 모든 처리를 하나의 솔루션 기반으로 수행



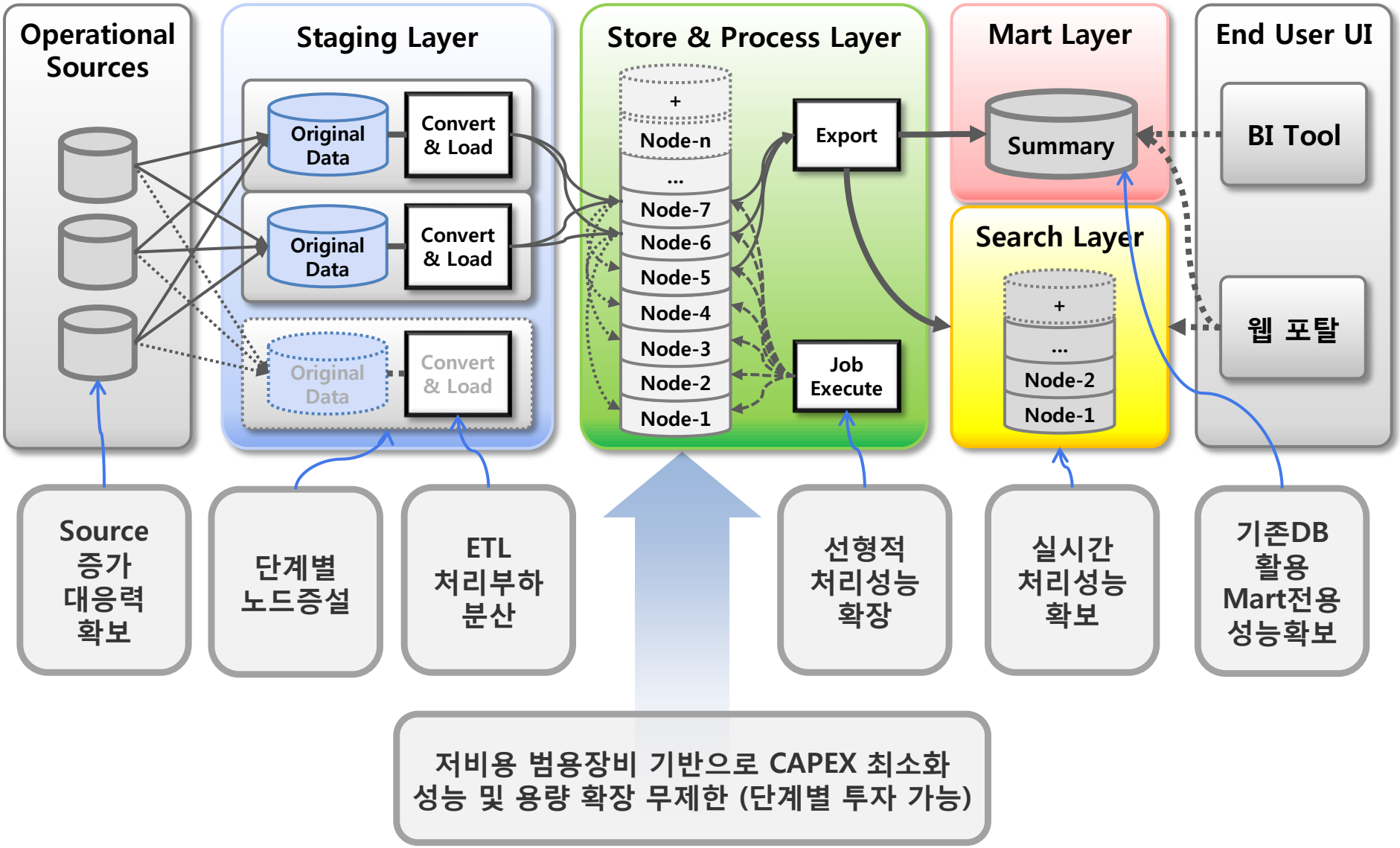
NDAP Software Stack



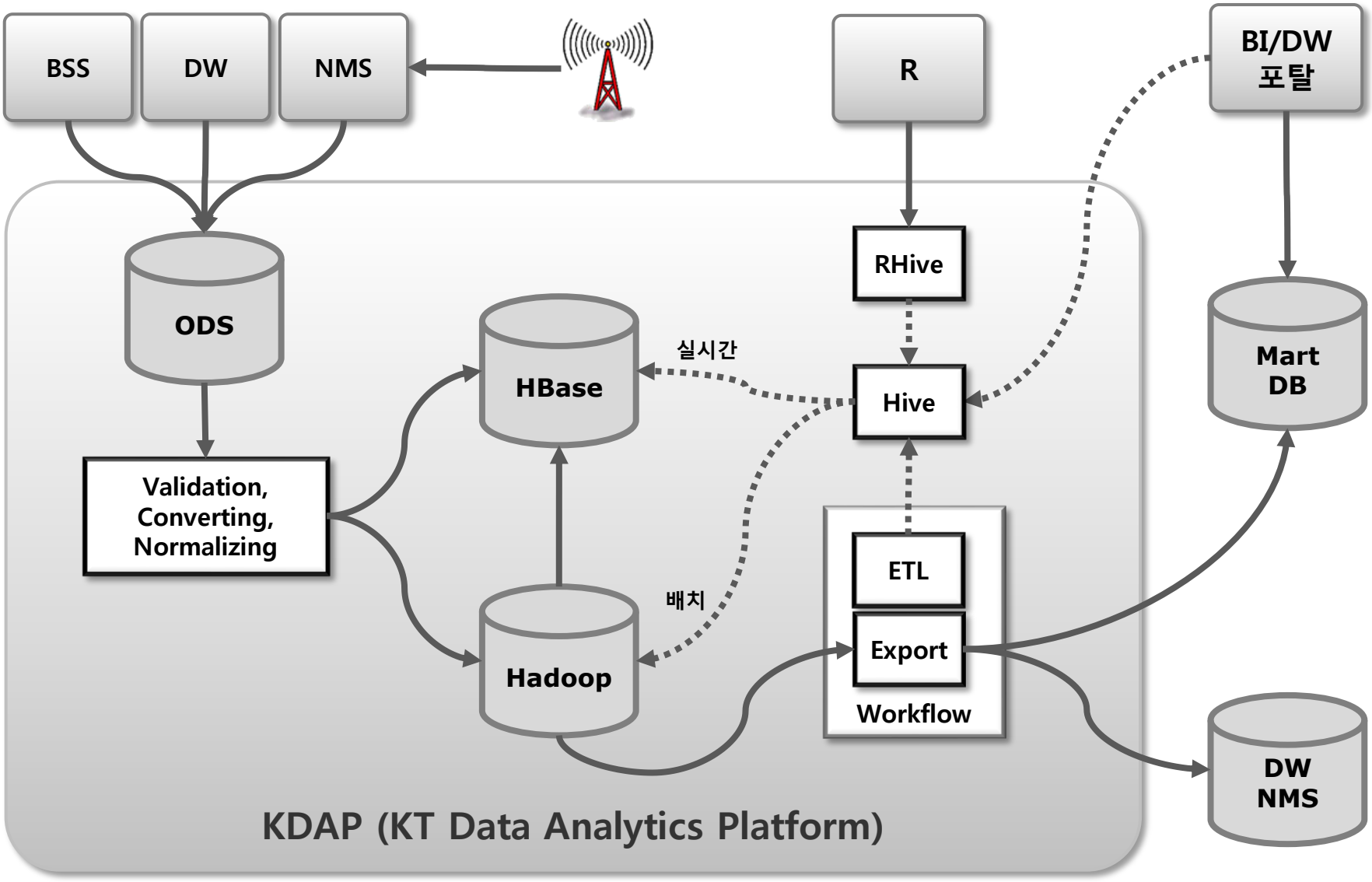
KT의 주요 분석 시스템 구축 현황



Platform Architecture



KDAP Architecture



Data 수용 및 확대 적용 계획

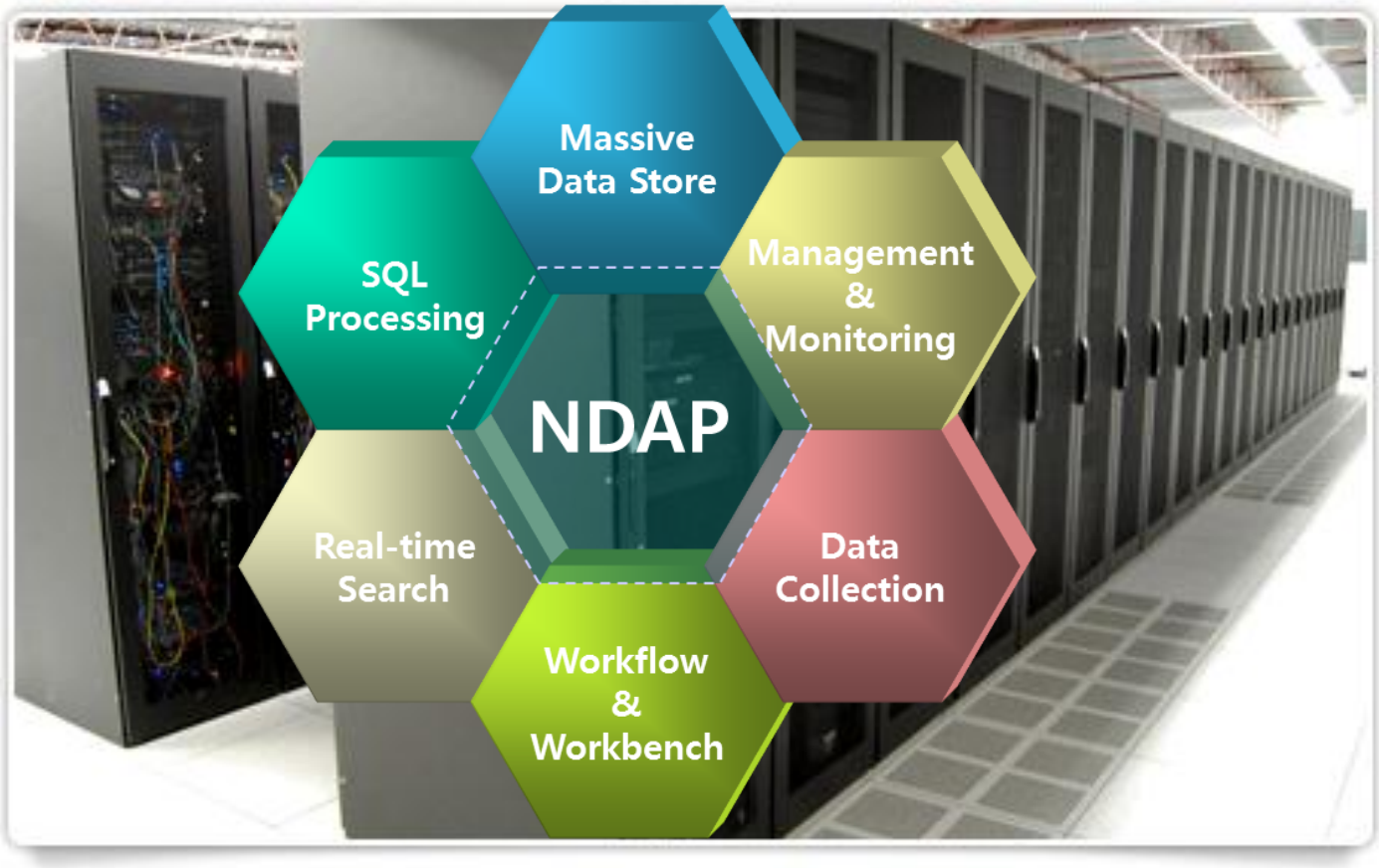
KT Platforms

- 3G Voice/data
- LTE data/VoLTE
- SMS / MMS
- olleh Wibro
- olleh Wifi
- Packet Data Log
- Wifi Call / VoIP
- Wired Internet
- olleh TV

2013년 통합

2014년 확대

KT Data Analytics Platform



2011년 부터 NDAP을 기반 SW로 사용하여 구축 수행
2013년 100 TB 규모 → 2014년 1 PB 이상 확대

Thank you

contact@ktcloudware.com