



# 데이터 중심 기업으로 거듭나기, 하둡 기반 빅데이터 성공사례



Reshaping Business with Next Generation Big Data Platform

최성호 부장

2013. 02

kt cloudware <sup>NEXR</sup>

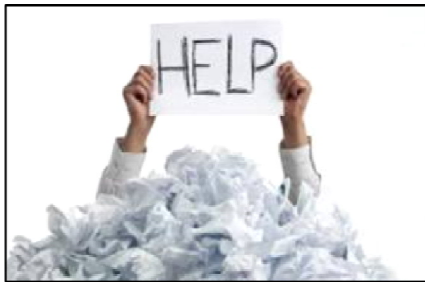
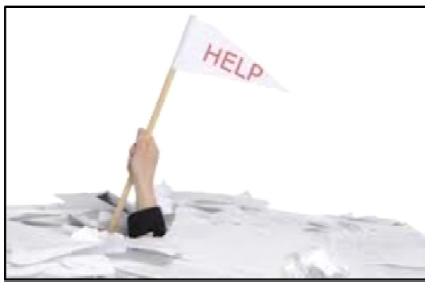
# 목 차

1	빅데이터에 대한 혼란
2	글로벌 빅데이터 구축 사례
3	기업환경에서 빅데이터의 접근방식
4	국내 빅데이터 구축 사례
5	차세대 빅데이터 플랫폼 NDAP 소개
6	맺음말
7	회사소개



# 빅데이터에 대한 혼란

WHAT PART OF  
$$i\hbar \frac{\partial}{\partial t} \Psi(\vec{r}, t) = \left( -\frac{\hbar^2}{2m} \nabla^2 + V(\vec{r}, t) \right) \Psi(\vec{r}, t)$$
  
DON'T YOU UNDERSTAND?



# 빅데이터에 대한 혼란



3V?

*Volume, Velocity, Variety*

*4V? Value?*

*+ Complexity*

# 빅데이터로의 접근

# No.

빅데이터의 정의,  
기준에 대한  
이해로부터의 출발



# YES

빅데이터 기술  
(technology)  
활용 관점에서의 접근

# 글로벌 빅데이터 구축 사례



MICRON TECHNOLOGY, INC.

**10 %**  
제품 생산시간 단축



**50만대 → 1,000대**  
판매 시점 결함 포착



**1초 주기 데이터 수집**  
오작동 징후 포착




**Scale-Out 아키텍처**  
2PB



**10만 명 → 5만 명**  
이탈 고객 수 절감



**빅데이터 서비스**  
신규비즈니스




**79%**      **65%**  
매장방문      쿠폰이용 구매



**200TB**  
유전자 데이터 확보



**150 PB**  
100TB DW



**신상품 개발**  
GPS 위치정보



**1/4 비용절감**  
Transaction 데이터 활용



**5억명**  
빅데이터 분석

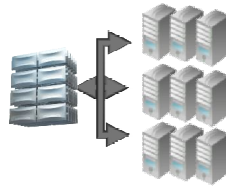
# 글로벌 빅데이터 구축 사례 – Key Takeaways



효율적 IT 인프라 구조  
**운영/도입 비용 절감**



대용량 데이터 고급 분석  
**새로운 Biz. 패턴 파악**



대용량 데이터 저장/처리  
**확장성 / 유연성 제공**

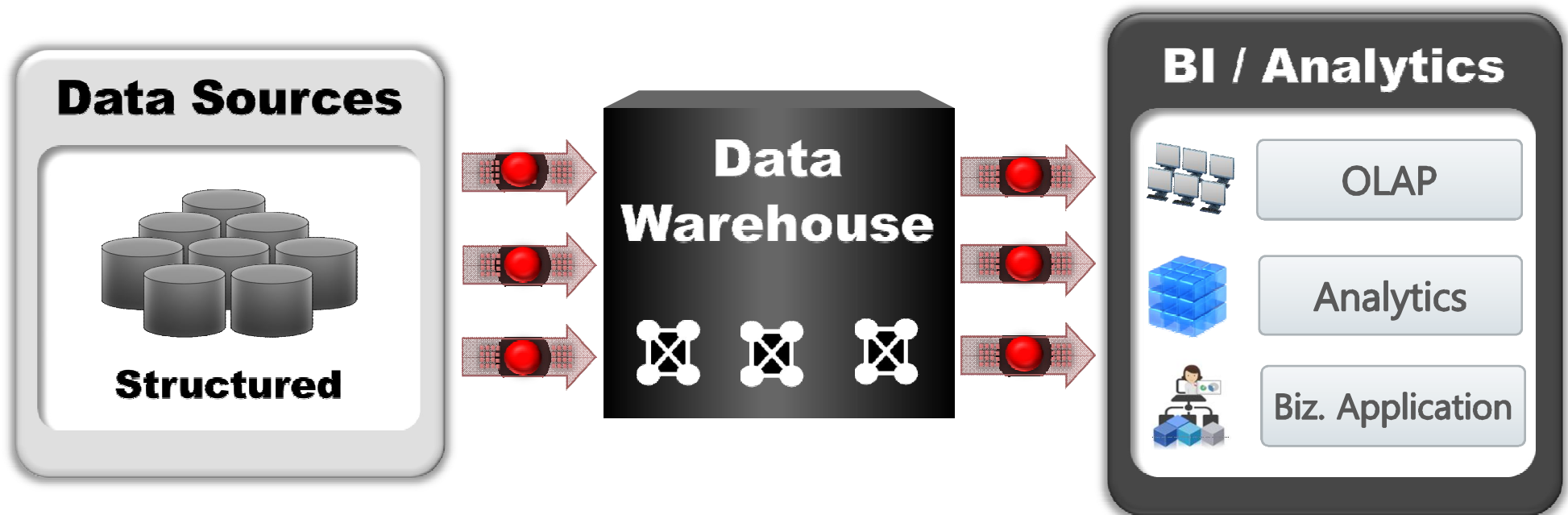


기존 DW 시스템 연계  
**작업 수행속도 개선**

# 전통적 아키텍처의 한계

전통적 중앙집중식 DBMS 기반 아키텍처의 한계

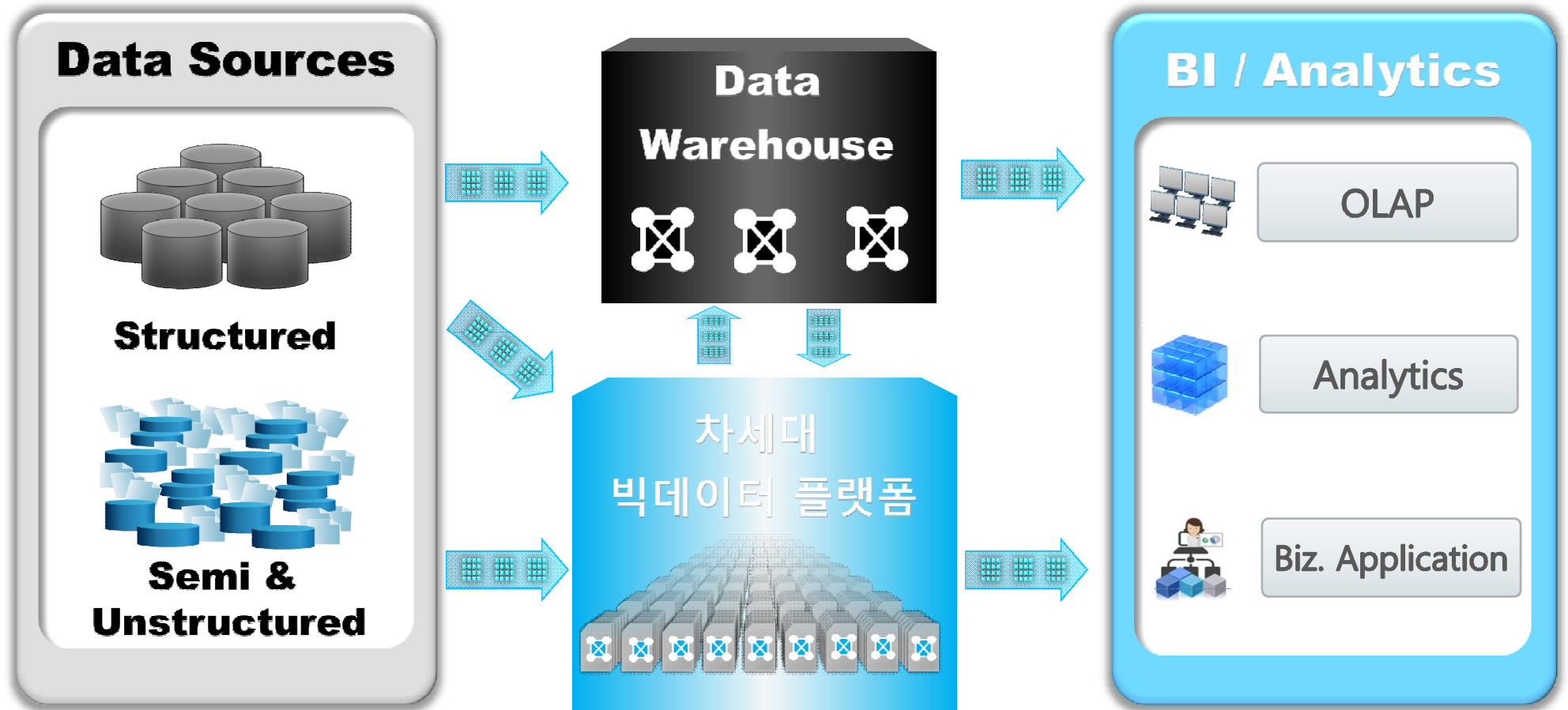
Scale-Up 구조 아키텍처 - 주요 병목구간: ●



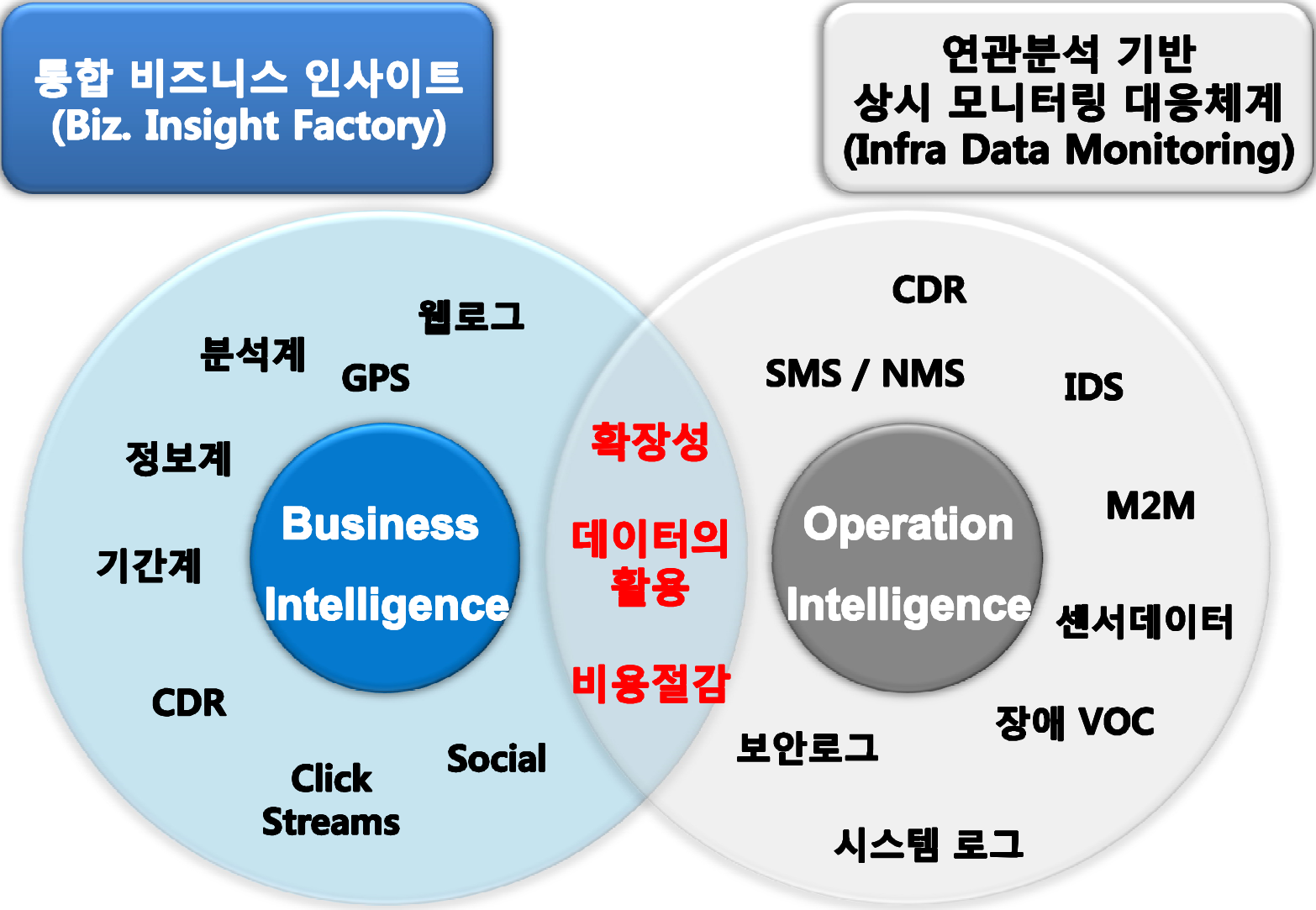


# 기존 아키텍처의 한계를 극복한 차세대 빅데이터 플랫폼

기존 아키텍처의 한계를 극복한 차세대 빅데이터 플랫폼  
x86 서버 기반 Scale-Out 구조 아키텍처

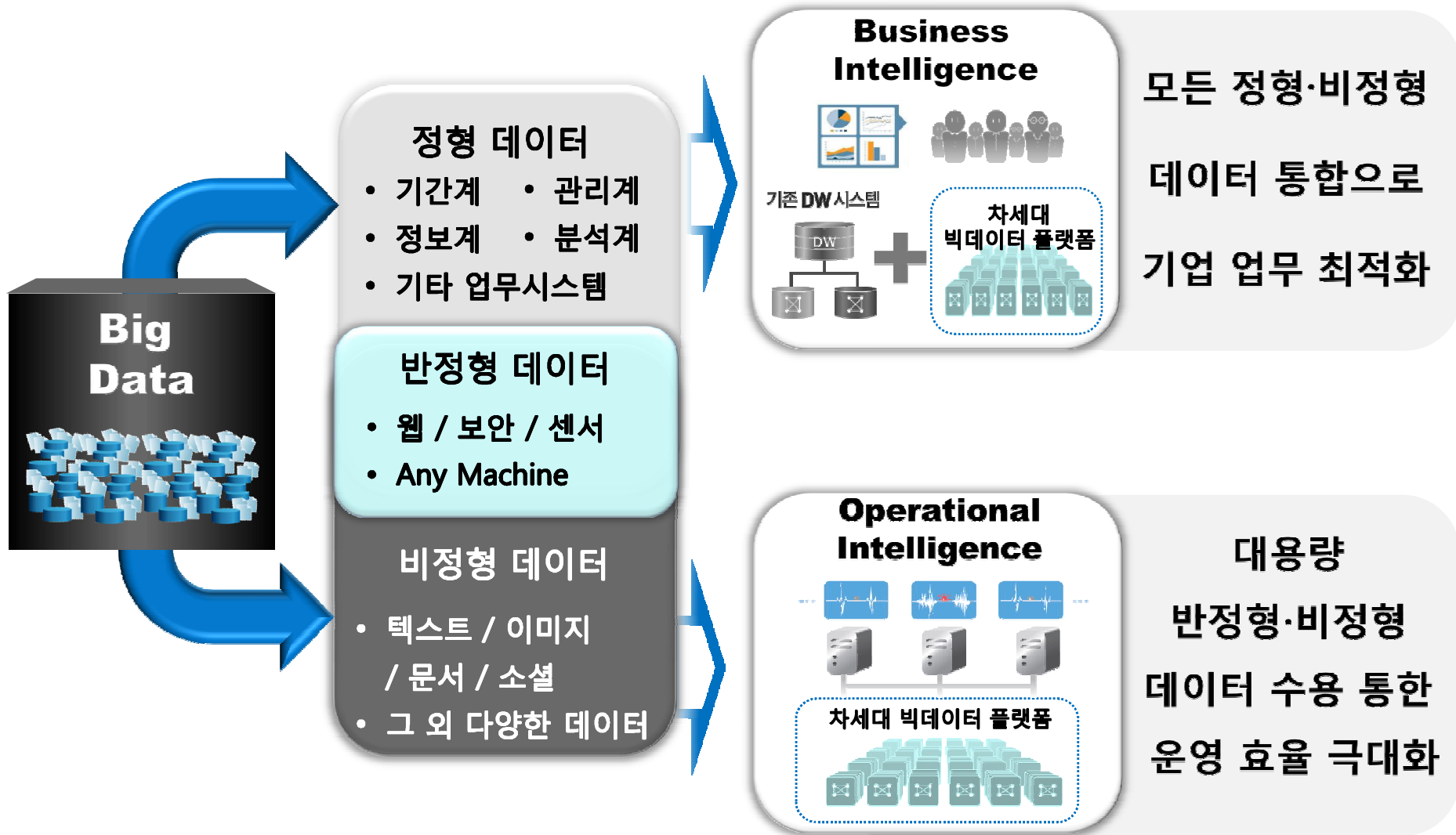


# 빅데이터로의 접근방향



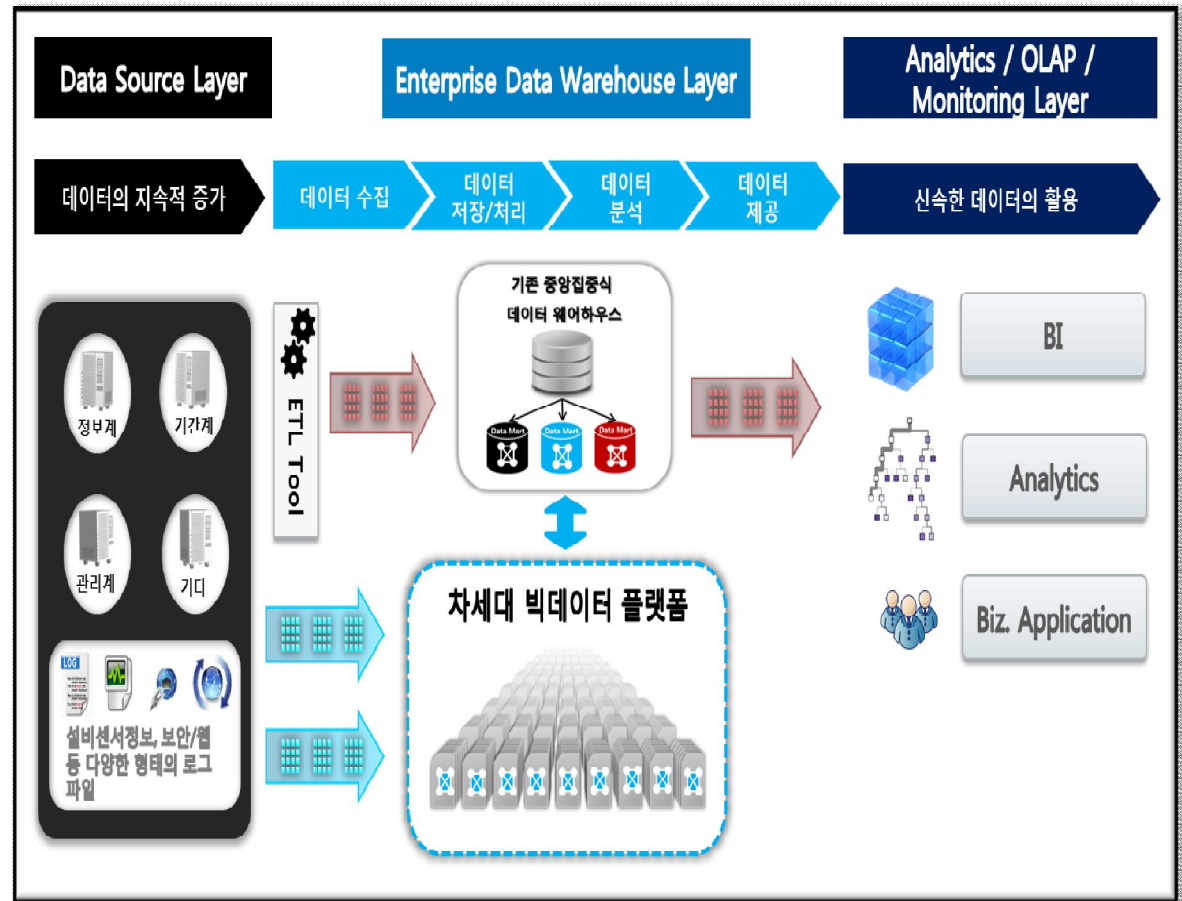
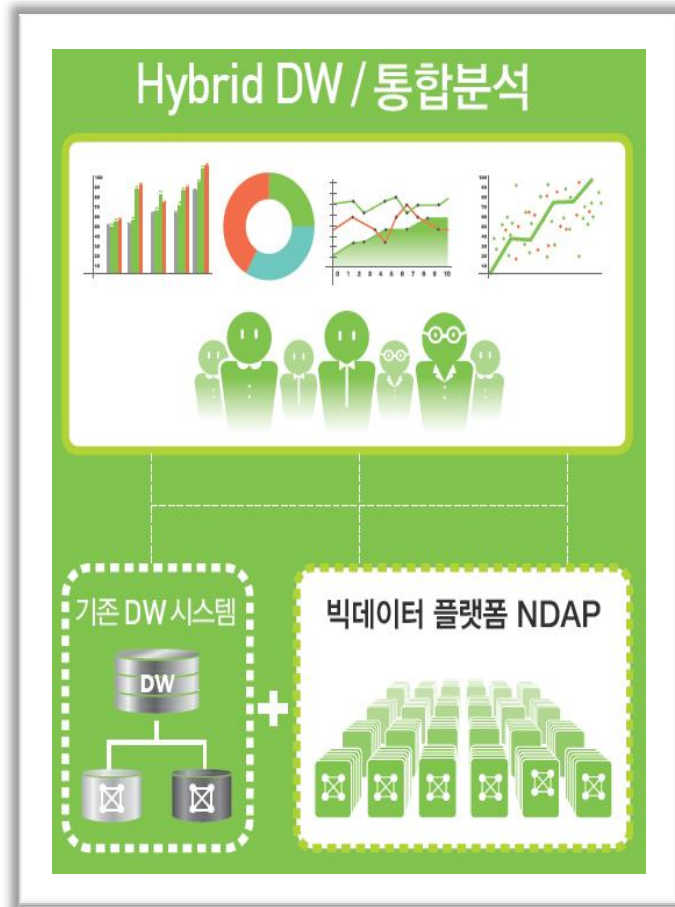


# 빅데이터로의 접근방향



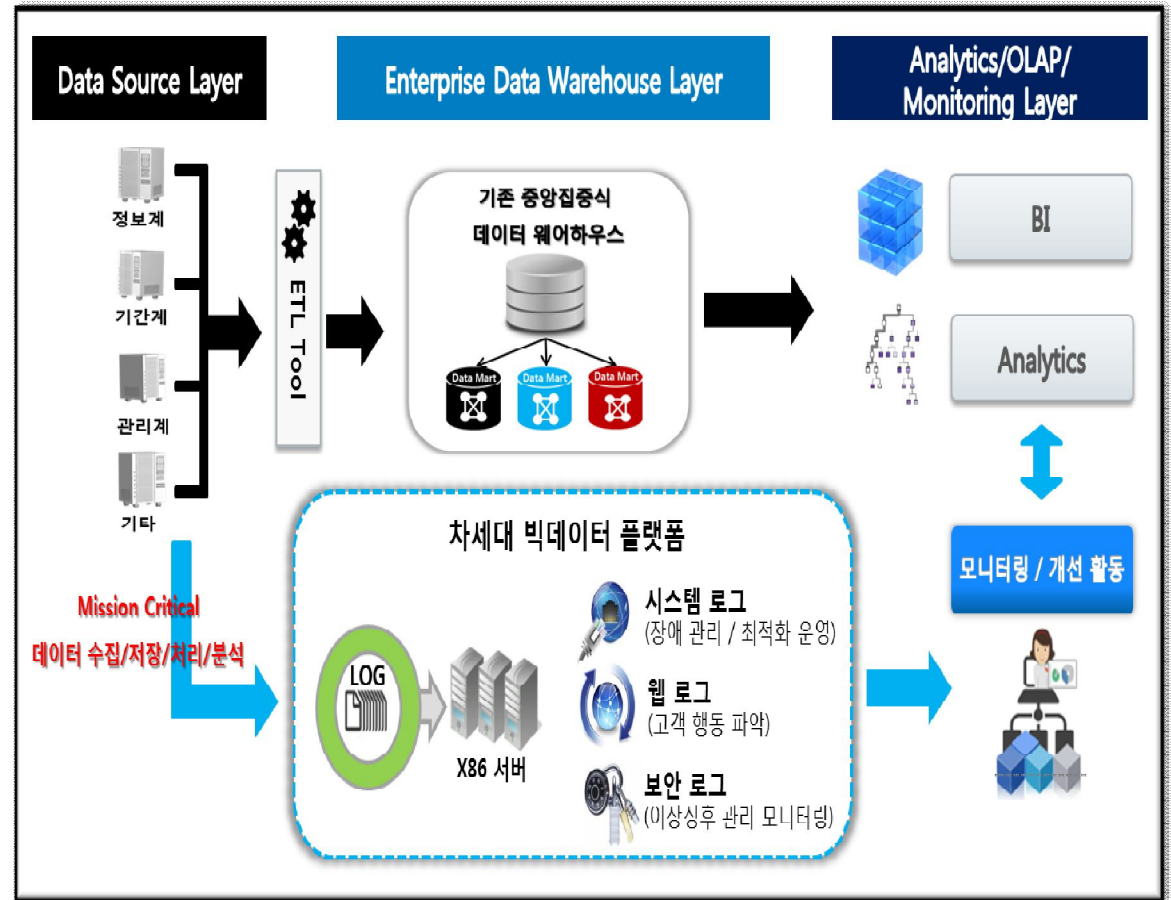
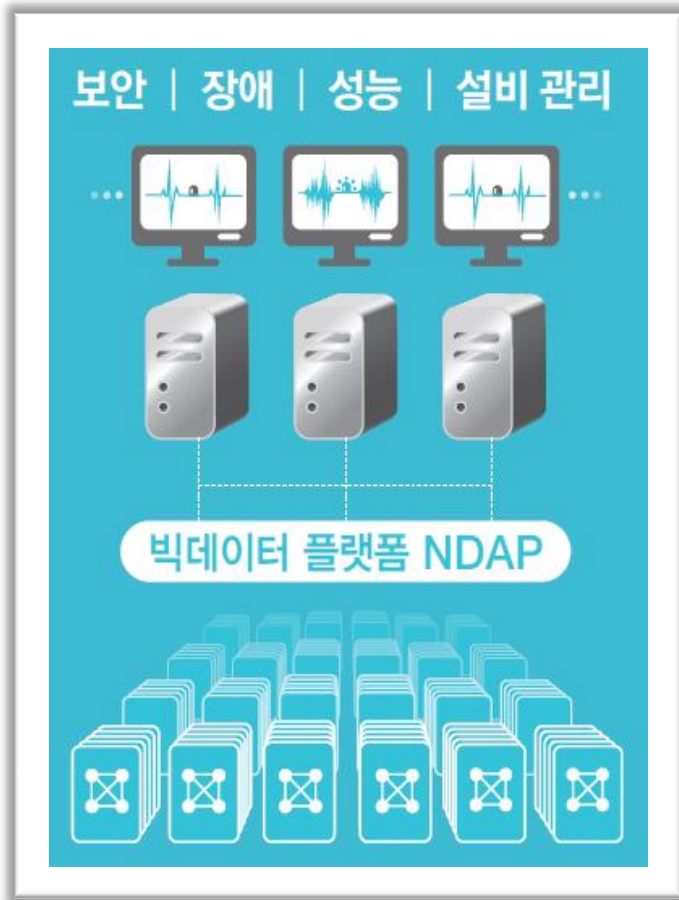
# 빅데이터로의 접근방향: Business Intelligence

모든 정형·비정형 데이터를 수용할 수 있는 효율적인 빅데이터 인프라는 보다 신속하고 고도화된 비즈니스 통찰력을 확보할 수 있게 합니다.



# 빅데이터의 접근방향: Operational Intelligence

기존에 버려지거나 활용되지 못하던 대량의 데이터를 비용 효율적으로 처리·분석함으로써 능동적인 대응 체계 및 가시성을 확보하여 통합 IT 환경을 개선할 수 있습니다.



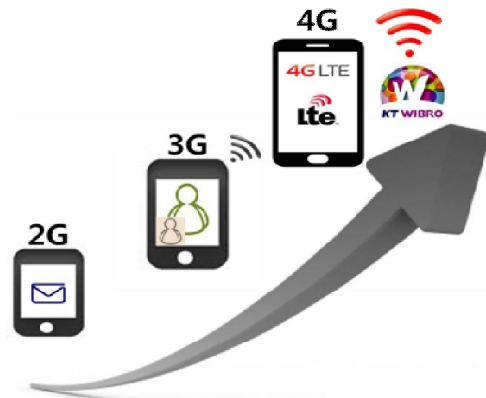
# 국내 구축 사례 #1 : Hybrid DW (가입자 분석 시스템)

## Korea Telecom



- 담당 부서 : BIT 경영인프라혁신담당 BI/DW 구축팀
- 목표 : Call 데이터 급증을 수용, 처리할 수 있도록 비용 효율적인 가입자 분석 시스템 구축

### 현황 및 이슈



**CDR 데이터**  
**증가 속도**  
**예측 불가**

- 기존 RDBMS 기반의 인프라 확장으로 수용, 처리 불가
- 과도한 구매, 유지 보수 비용

### 빅데이터 플랫폼 주요 효과

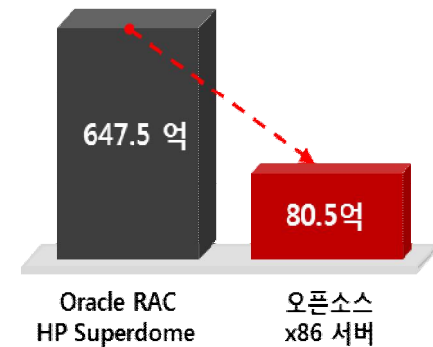
성능 개선과 함께

RDBMS 기반 아키텍처 대비 5년 TCO 기준

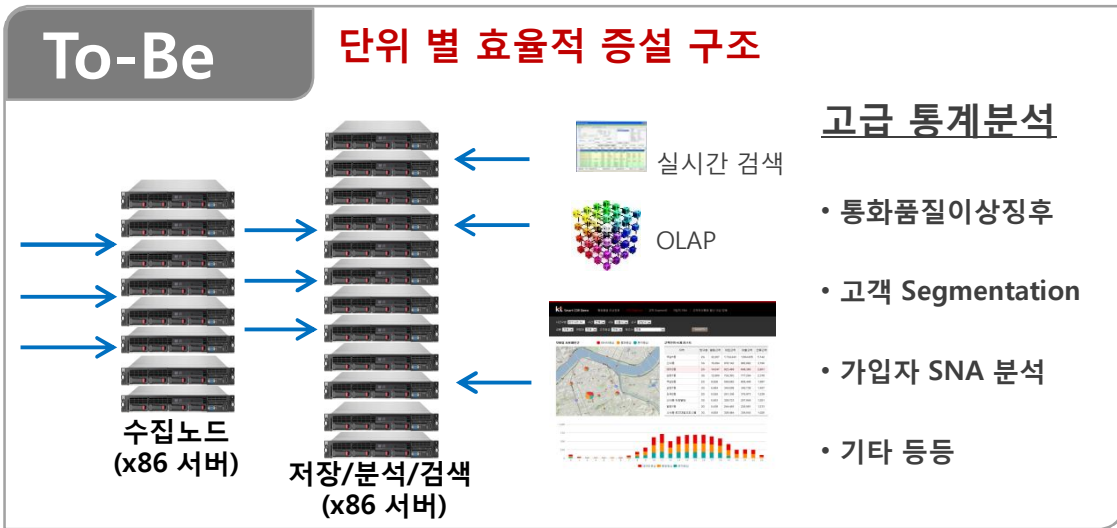
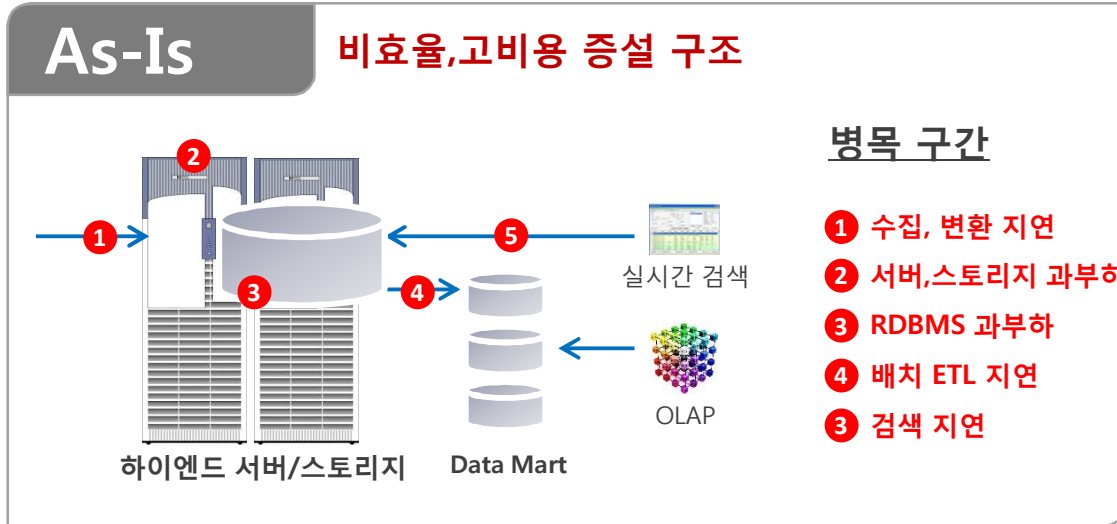
**567억**

**(88%)**

**비용 절감**



# 국내 구축 사례 #1 : Hybrid DW (가입자 분석 시스템)



## 개선 사항

**1.5x**

선형적 배치 성능 개선 (서버 2배 증설)

**6x**

ETL 병목현상 개선

**2.5x**

BI 성능 개선

**1sec 미만**

실시간 검색 평균 처리시간

# 국내 구축 사례 #1 : Hybrid DW (가입자 분석 시스템)

## CDR 기반 빅데이터 로드맵

### 1단계

- 인프라 구축
- 데이터 통합

빅데이터 통합/융합  
**인프라 근간 마련**

데이터 저장, 분석  
**인프라 TCO 절감**

고객 다차원 분석을 통한  
타겟 마케팅, 서비스 개선  
등을 통한  
**고객 만족도 향상**

### 2단계

- 데이터 융합
- 계열사, 이종사업

계열사 간 데이터 융합  
을 통한 서비스 차별화  
를 통한  
**기업 경쟁력 확보**

외부 이종 사업 간  
데이터 융합을  
**새로운 비즈니스  
서비스**

### 3단계

- 공공부문 확대

대국민 공공 서비스  
**국가차원에서의  
데이터 자원의  
부가가치 증대**

# 국내 구축 사례 #2 : IT 인프라 운영 품질 개선

## Korea Telecom

olleh ucloud

- 담당 부서 : 클라우드 추진본부 서비스품질기획팀
- 목표 : 서비스 가용성, 성능 향상을 통한 고객 만족도 향상 및 운영 환경 개선

### 현황 및 이슈



Amazon, ucloud 등 서비스 가용성 이슈

Too many HW, Too many VM -> 관리 이슈

Too Many 로그 종류 -> 수집, 변환 이슈

### 빅데이터 플랫폼 주요 효과

다양하고, 방대한 클라우드 로그 통합  
저장, 분석을 통해 ...

서비스 가용성 30% 개선

고객 만족도 및 uCloud 시스템

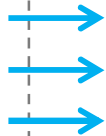
신뢰성, 관리성 향상



# 국내 구축 사례 #2 : IT 인프라 운영 품질 개선

## 시스템 현황

Apache Log,  
Webapp log,  
syslog,  
machine(netflow) log 등  
다양한 소스 및  
형식의 로그 수용



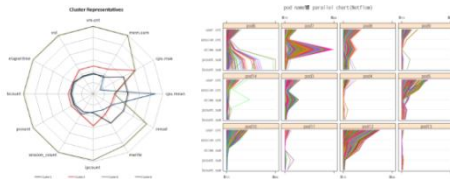
다양한 형태 로그 수용  
성능저하 없이 장기보관 가능  
고속의 병렬 고급 분석



수집노드  
(x86 서버)

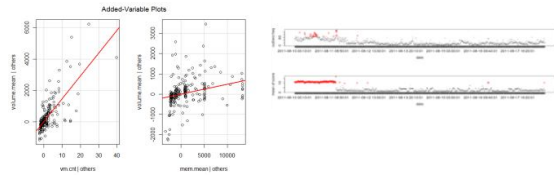


저장/분석/검색  
(x86서버)



### 고급 통계 분석

- 자원 수요 예측
- 사전 장애 탐지 및 선재 대응
- 고객 군집 분석 및 추이 분석  
을 통한 자원 재배치 등



## 개선 사항

서비스 가용성

**30% 향상**

장애 대응 시간

**45% 단축**

증설, 구매 프로세스

**30% 개선**

성능 (분산을 통한)

**35% 개선**



# 국내 구축 사례 #3 : 제조사 신규 서비스 개발

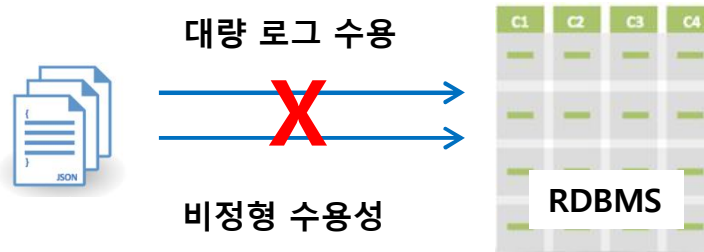
## A 제조사



- 담당 부서 : R&D 연구소
- 목표 : 제품 상의 센서 데이터 분석을 통한 **제품품질개선 및 대고객 서비스 개발**

### 현황 및 이슈

RDBMS로 해당 데이터 수용 부적합



기존 상용 분석 툴 (SAS, SPSS) 고비용  
서비스화 고려 장기적인 관점 접근 필요

### 주요 효과 및 확산 계획

결함 예측 모형 개발 및 데이터 집계를 통한  
**품질 개선, 조기 리콜 비용 절감,  
R&D 프로세스 개선**

빅데이터 통합 분석 플랫폼 로그맵



빅데이터를 활용한 대고객 개인화 서비스,  
품질조기 경보, 부품수급예측 등 계획

# 국내 구축 사례 #4 : 금융사 빅데이터 Pilot 프로젝트

## B 금융사



- 담당 부서 : CIO & 빅데이터 TFT팀
- 목표 : 빅데이터 통한 TCO 개선 검증, 기존 분석 불가능한 대용량 데이터 분석 타당성 검토

### 현황 및 이슈

RDBMS 기반 DW의 처리 성능 한계..  
또 증설?

인프라의 수용 제약으로  
버려지는 로그 데이터

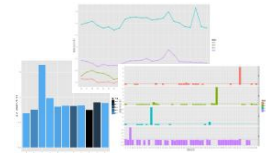


기존 분석 작업들에 정확도  
측면에서의 개선 필요

### Pilot 프로젝트 결과

저비용 (x86 서버 7대) 구조로  
대용량 온라인 거래로그 분석

**성능 5~10배 개선**



다양한 설명 인자 고려한 빅데이터 고급  
분석을 통해 기존 Rule 기반의

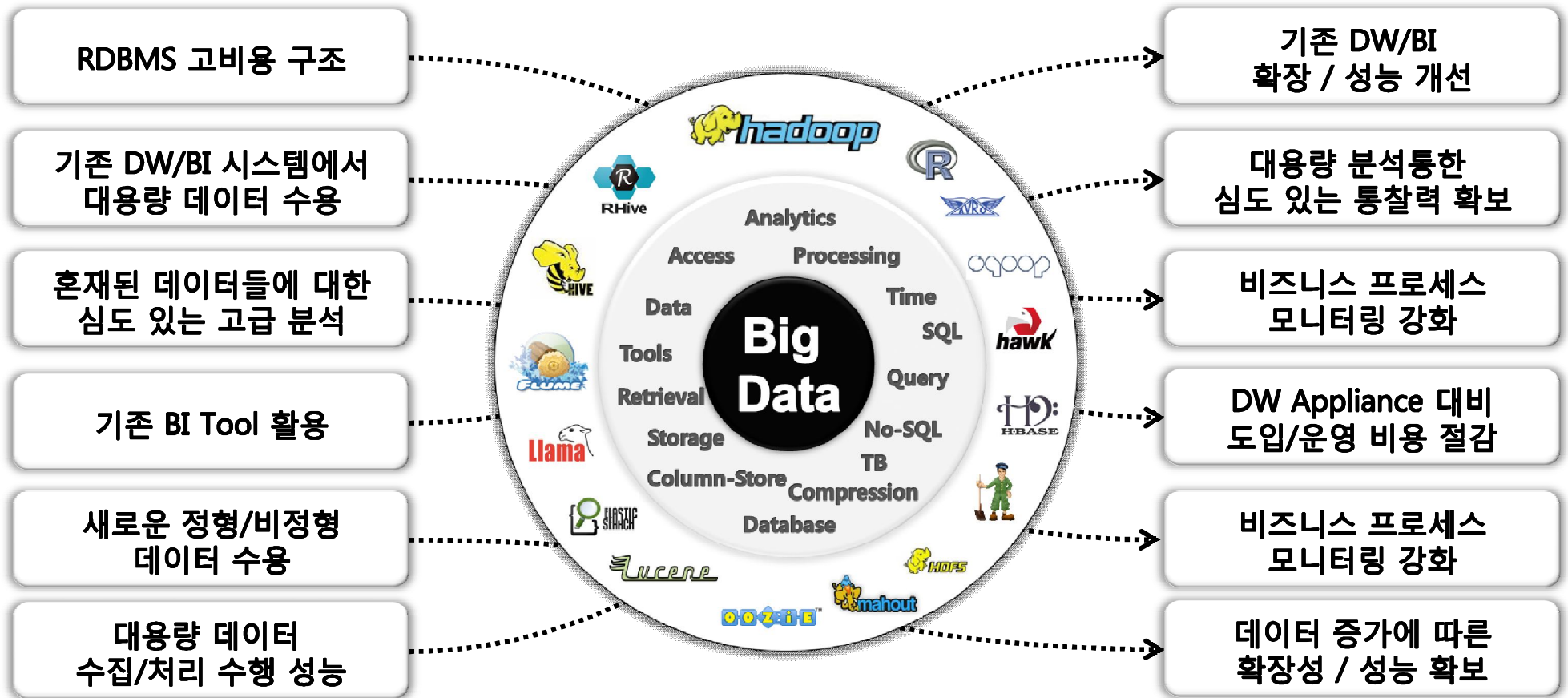
**중복매출처리 프로세스 및  
정확도 74% 개선**

Open Source



# 차세대 빅데이터 플랫폼 구현 조건

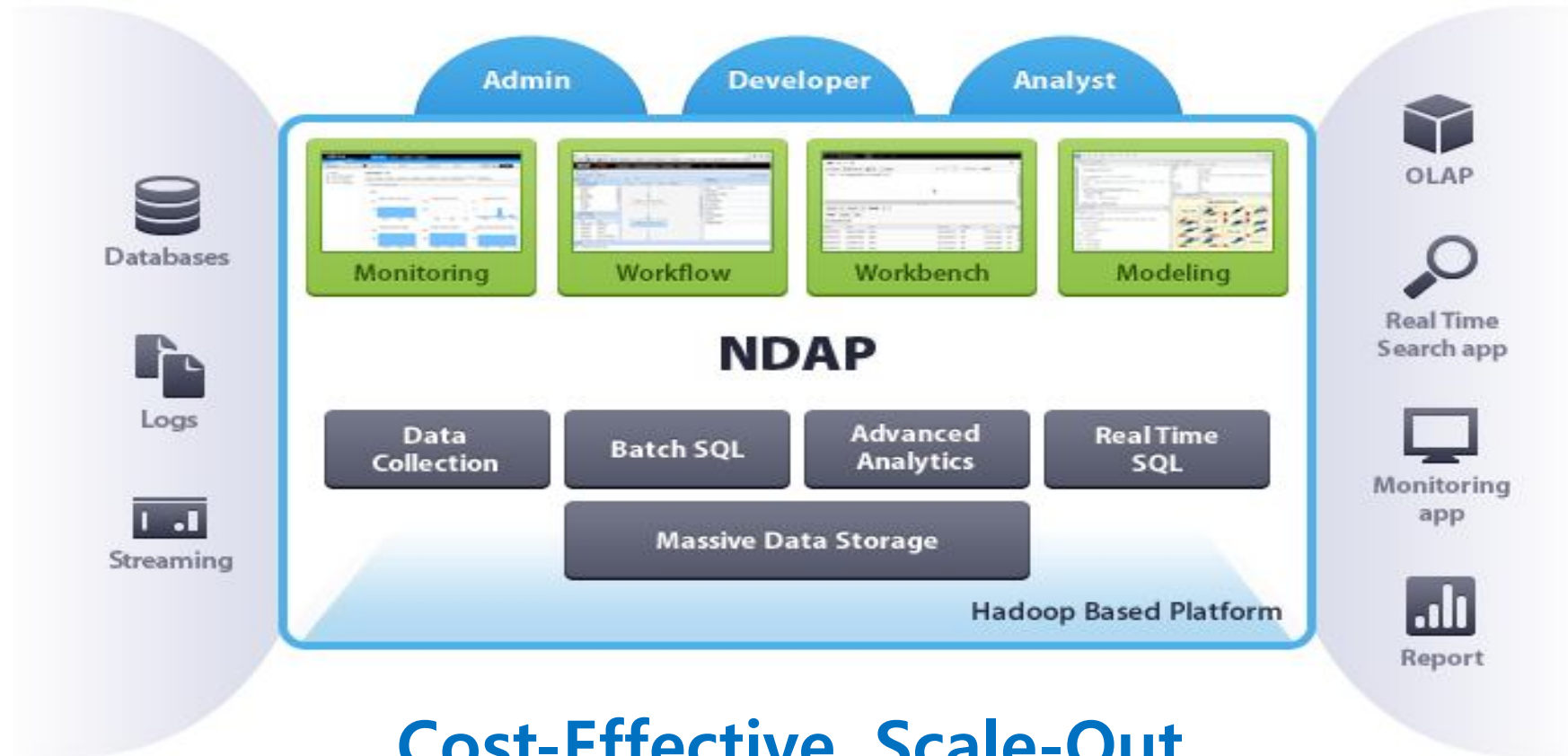
## 기업 환경에 맞는 다양한 빅데이터 기술 검토 및 Skill Set 필요



# NDAP (NexR Data Analytic Platform) 소개

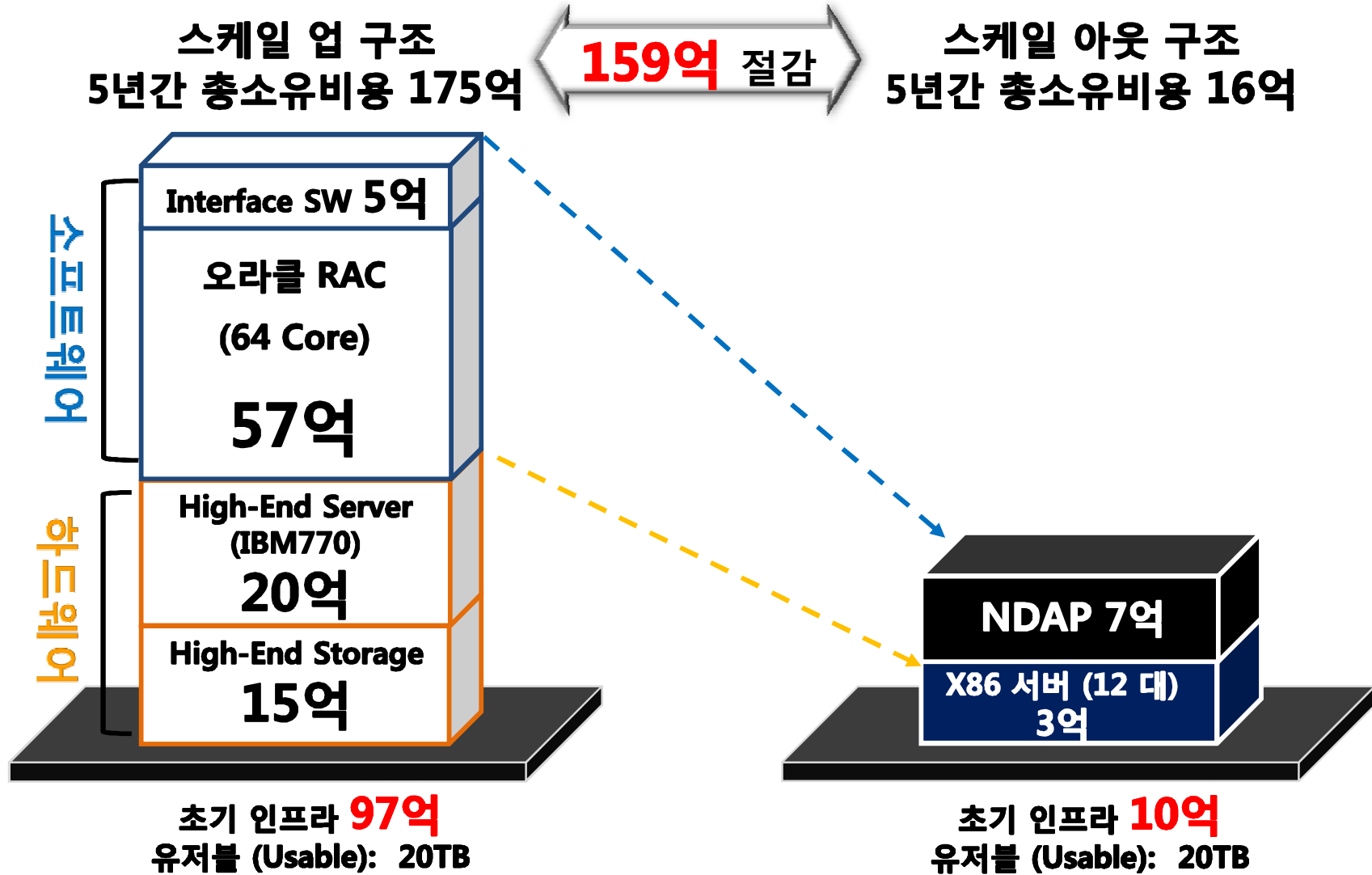
Big Data 처리를 위한 **All-In-One 소프트웨어**

→ Big Data의 수집/저장/분석/검색/변환 등 모든 처리를 하나의 솔루션 기반으로 수행



**Cost-Effective, Scale-Out**

# NDAP 의 차별성 - 초기 투자비용 Big Economic 실현



\* 기준: Usable 20TB / HW&SW List Price \*\* 유지보수율: NDAP 15% / 타사 평균 20%

# Summary

---

1. 빅데이터 기술 적용을 통해 선진기업들은 이미 비즈니스 혜택을 얻고 있음
2. Cost 이슈로 인해 기존 담지 못했던 데이터 저장/수집/처리 가능 함
3. 기존 BI/DW와 공존을 통해 비즈니스 연속성상에서 추가적인 가치 창출이 가능해짐

## 성공적인 빅데이터 프로젝트를 위해서.....

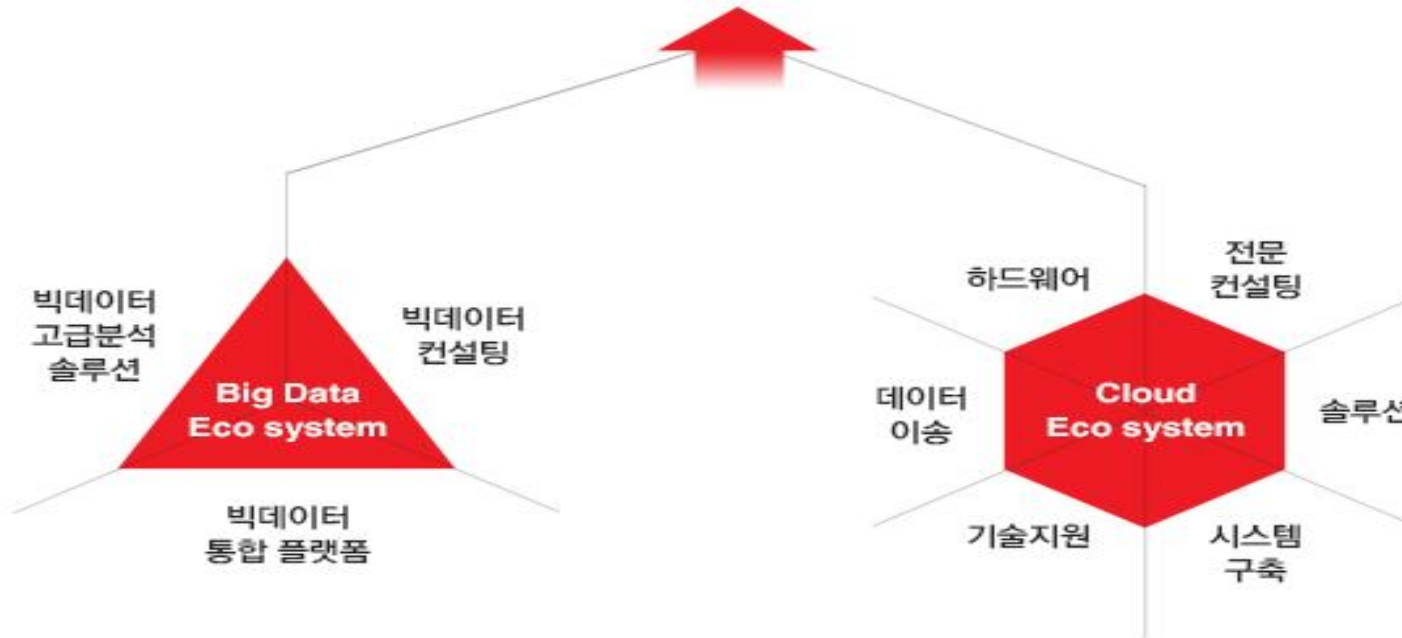
1. 소규모 Pilot 프로젝트로 부터 시작
2. 임원의 Support
3. 현업과 긴밀한 공조
4. Right Partner와의 협업

# 회사소개

## About kt Cloudware

**kt cloudware**

빅데이터 · 클라우드 엔터프라이즈 End-To-End 솔루션 및 서비스 제공



# 회사소개

빅데이터 경험 풍부한  
오픈 소스 **전문가** 보유



100명 이상의 전문 엔지니어 보유

- Korea R User Group
- Korea Hadoop Community
- 국내 최초 아파치 Hadoop Committer

다양한 빅데이터  
프로젝트 **경험** 보유



국내 다양한 Industry에서  
엔터프라이즈 빅데이터 프로젝트  
구축 경험 보유

전문가 · 경험 기반  
End-To-End  
**빅데이터 SW**



차세대 빅데이터 통합 플랫폼 /  
빅데이터 고급 분석 솔루션



# 감사합니다

[contact@ktcloudware.com](mailto:contact@ktcloudware.com)

**kt** **cloudware** **NEXR**