TmaxSoft
Better technology, better tomorrow

데이터 통합의 미래,
실시간 데이터 통합 가상화 솔루션

DataHub 3.0

2015. 03.

© 2015 TmaxSoft Co., Ltd. All Rights Reserved.

Contents

Ⅰ BI 플랫폼의 변화

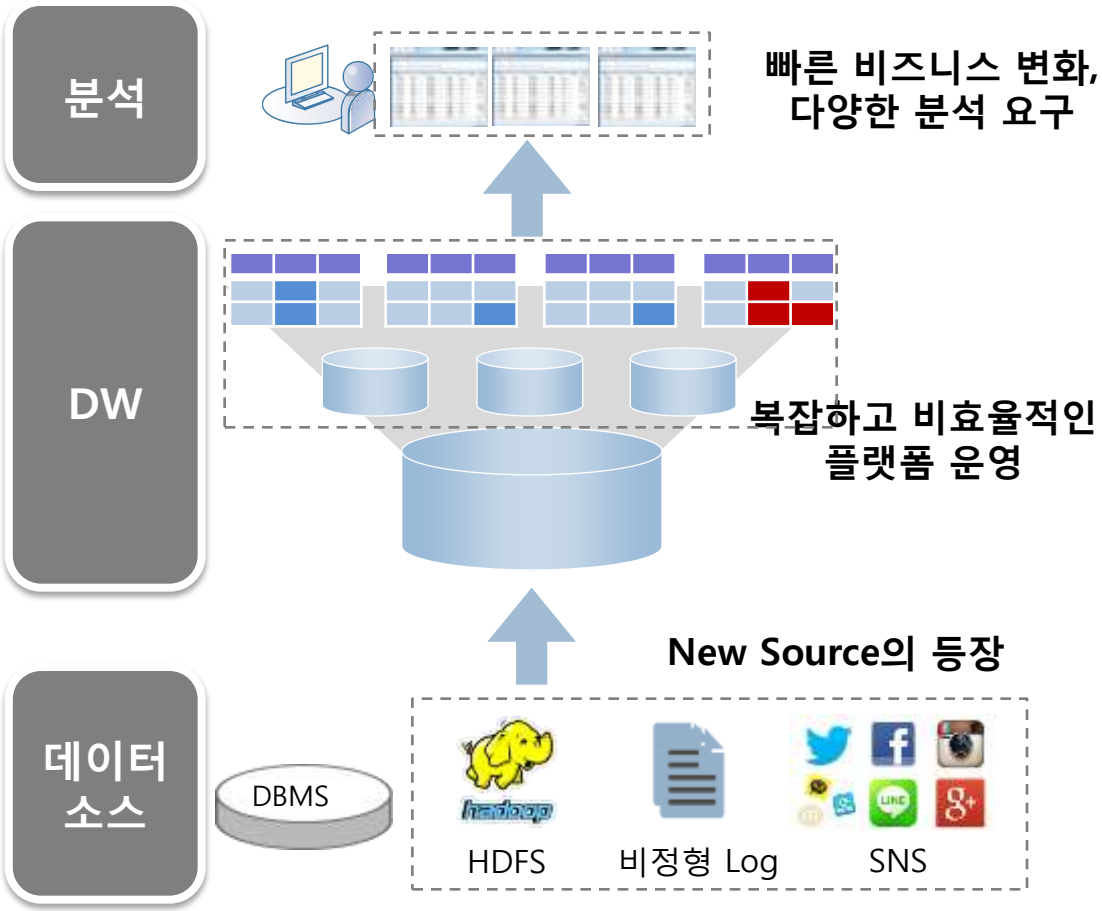
Ⅱ 데이터 가상화 플랫폼 : DataHub

Ⅲ DataHub의 특장점 및 사례

BI 비즈니스 환경의 변화

복잡한 비즈니스 환경, IT인프라의 발전으로 보다 많은 데이터로부터 보다 빠른 의사결정이 요구되는 상황임

BI 운영 현황



요구되는 것

1 On-Time Business 의사 결정

- 새로운 형태의 분석 Report 발생(1~2개월)
- 다차원적, 복잡성 높은 분석 요구

2 Operational BI 강화

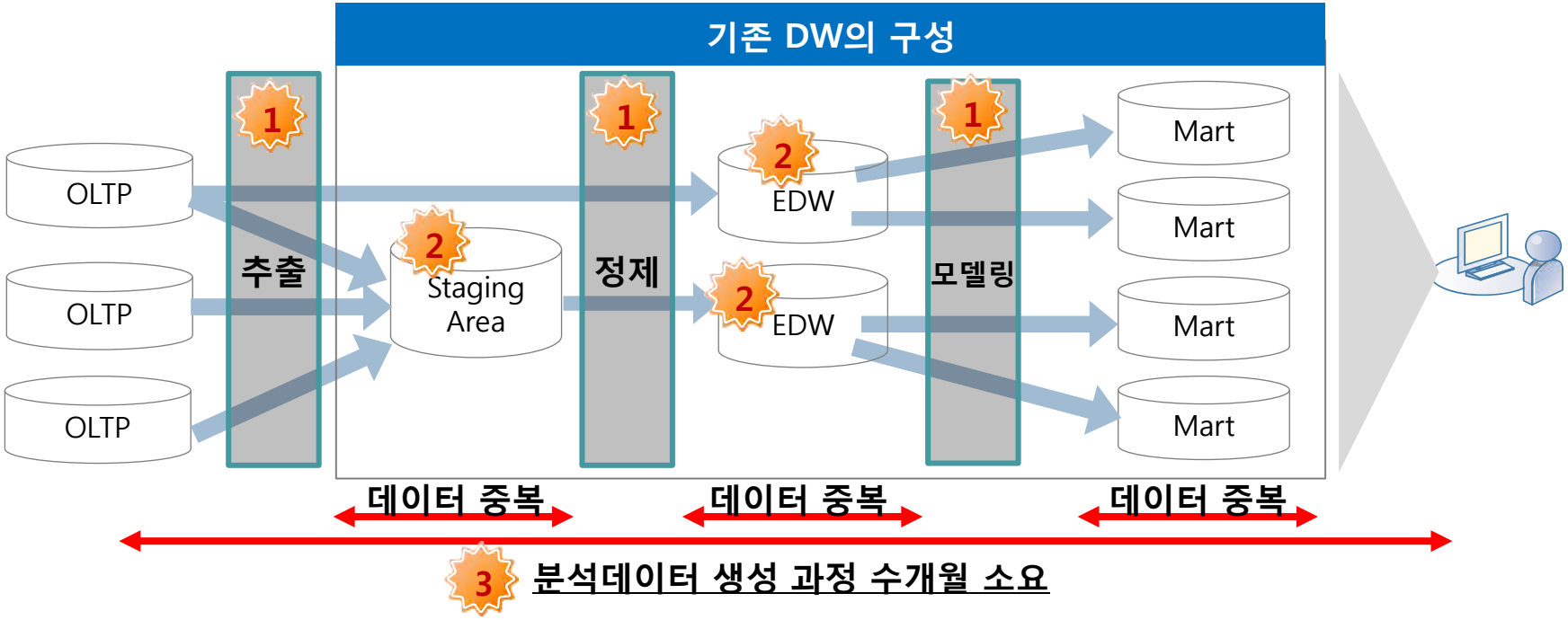
- 실시간 의사결정 횟수 증가, 질적 향상 필요
- 잦은 데이터 변형과 메타관리 필요

3 새로운 형태의 Data 포용

- 대량의 반정형/비정형 데이터 발생
- DW에 Data Source 추가 작업 지속적 발생(1~2개월에 한번)

기존 DW 인프라의 한계

복잡한 데이터 생성 과정으로 인해 시간적, 공간적 비효율 발생 및 On-Time 분석이 어려워짐



기존 DW의 문제.....

1. 데이터 추출, 변환의 별도 작업 필요
2. 데이터 중복 저장 비효율 및 Quality저하
3. 최종 분석까지 Long Term Process

현업의 요구, 다양한 데이터의 유연하고 빠른 분석

다양한 데이터 소스로부터 의미 있는 분석을 적시에 이끌어내기 위한 방안을 고심

필요한 Data를 즉시 찾아서 활용하는 방안

- 어떤 시스템이 있고 어떤 데이터가 존재하는지 쉽게 인지
- 현업, IT담당자, 분석가 제각기 이해할 수 있는 용어 활용

이기종 시스템 One-Step 통합 분석 방안

- 오라클, DB2, MS-SQL, Greenplum, 난립하는 DB의 통합
- 별도의 데이터 추출, 과다한 엑셀 작업 해소

다양한 분석, Ad-hoc 분석 방안

- 현업으로부터의 다채로운 분석 요구의 해소
- 최종분석 도출 중간 중간 즉흥적 분석필요

비정형 빅데이터의 활용 방안

- 방대한 로그데이터를 SQL처럼 분석
- 정형, 비정형 데이터 연계 분석

현업의 요구

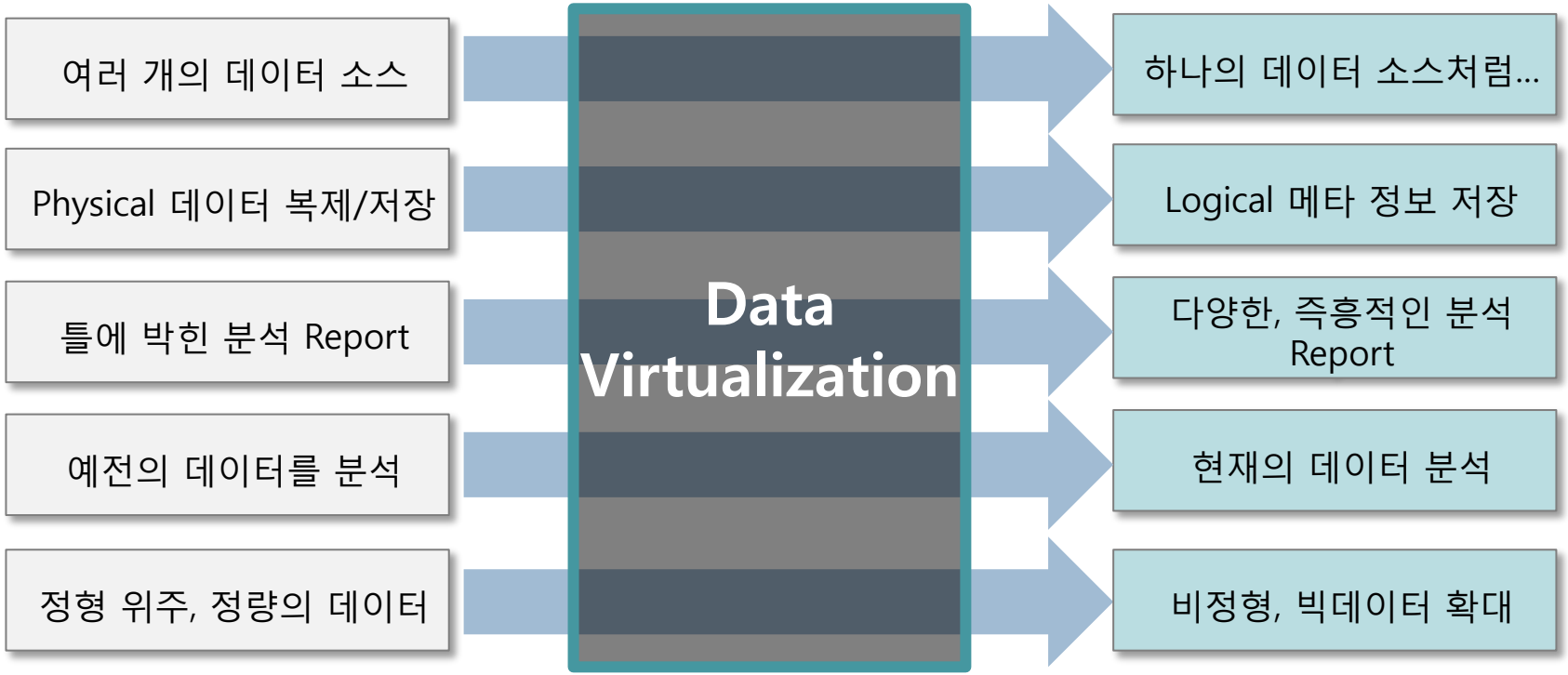
다양한 데이터를 Single DB 처럼...

유연한 모델링을 통하여...

적시에 분석을 하고 싶다

궁극적 대안, Data Virtualization

가상화를 통한 데이터 통합으로 물리적, 시간적 한계를 뛰어넘는 데이터 관리 및 유연하고 다양한 분석이 가능함



데이터 가상화란?

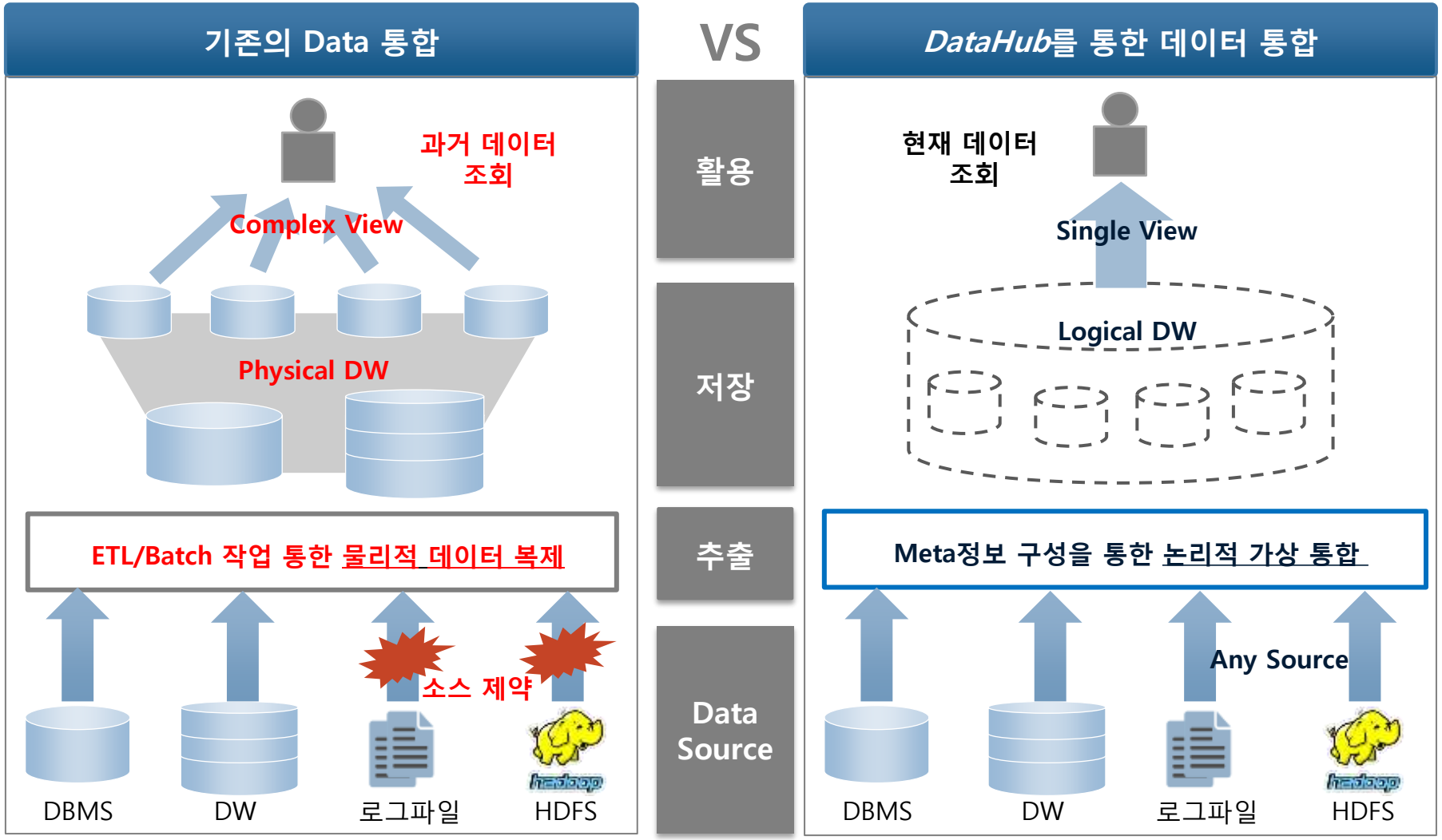
"데이터의 물리적 위치, 형태, 기술적 요건과 상관없이 데이터를 조회, 조작할 수 있는 애플리케이션." – *Wikipedia*

Or more Simply

"여러 데이터 소스를 하나의 SQL기반 DBMS 처럼 다루도록 하는 솔루션"

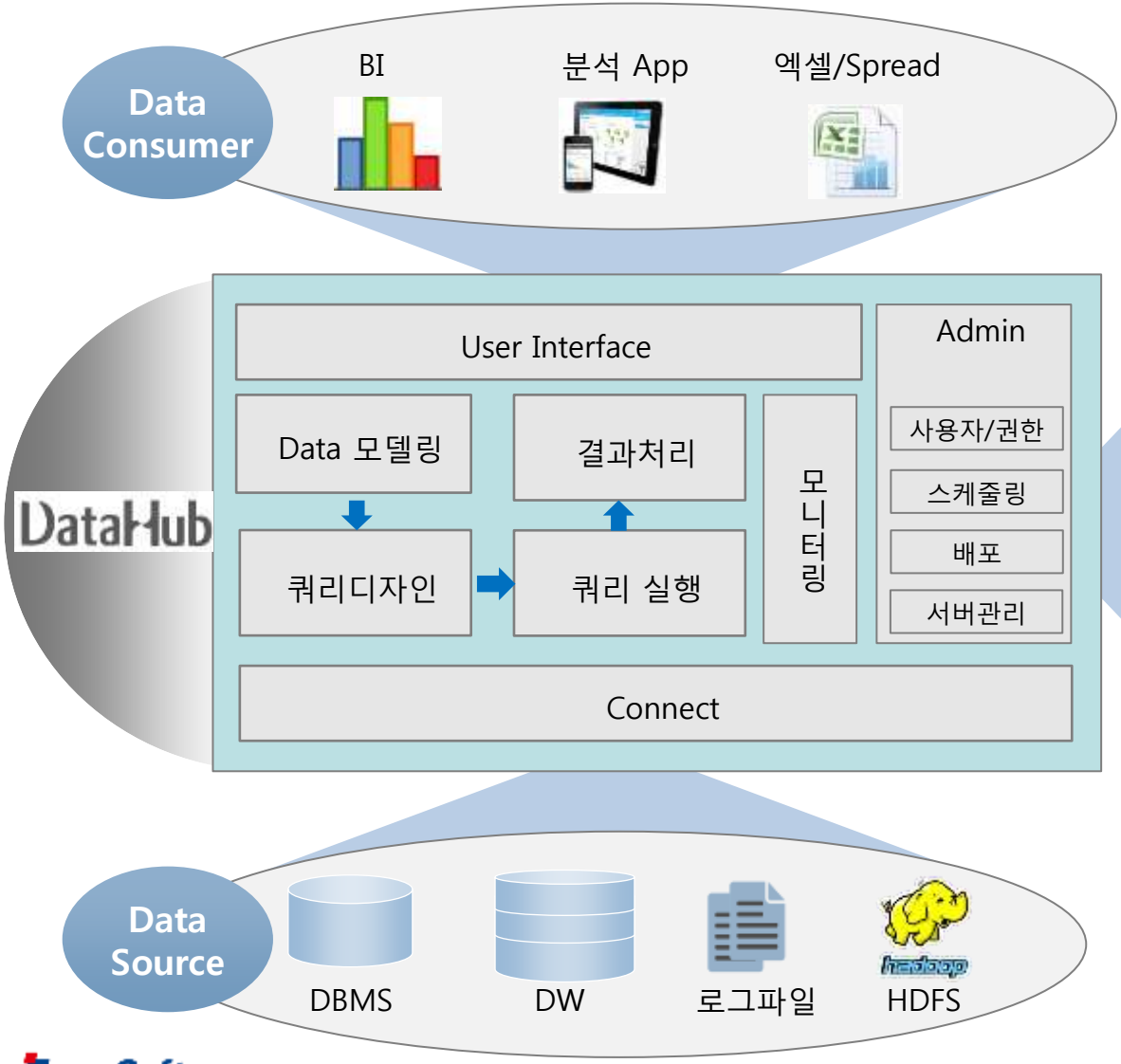
DataHub를 통한 데이터 가상화 구현

DataHub는 다양한 데이터소스를 가상으로 통합하여 모델링 및 분석 처리를 하는 솔루션



데이터 가상화의 기반, DataHub 개요

DataHub는 다양한 데이터소스를 가상으로 통합하여 모델링 및 분석 처리를 하는 솔루션



핵심 Feature

데이터 가상 통합

- Any Data Connection
- 메타정보 활용한 모델링, Single View

유연한 데이터 분석 설계

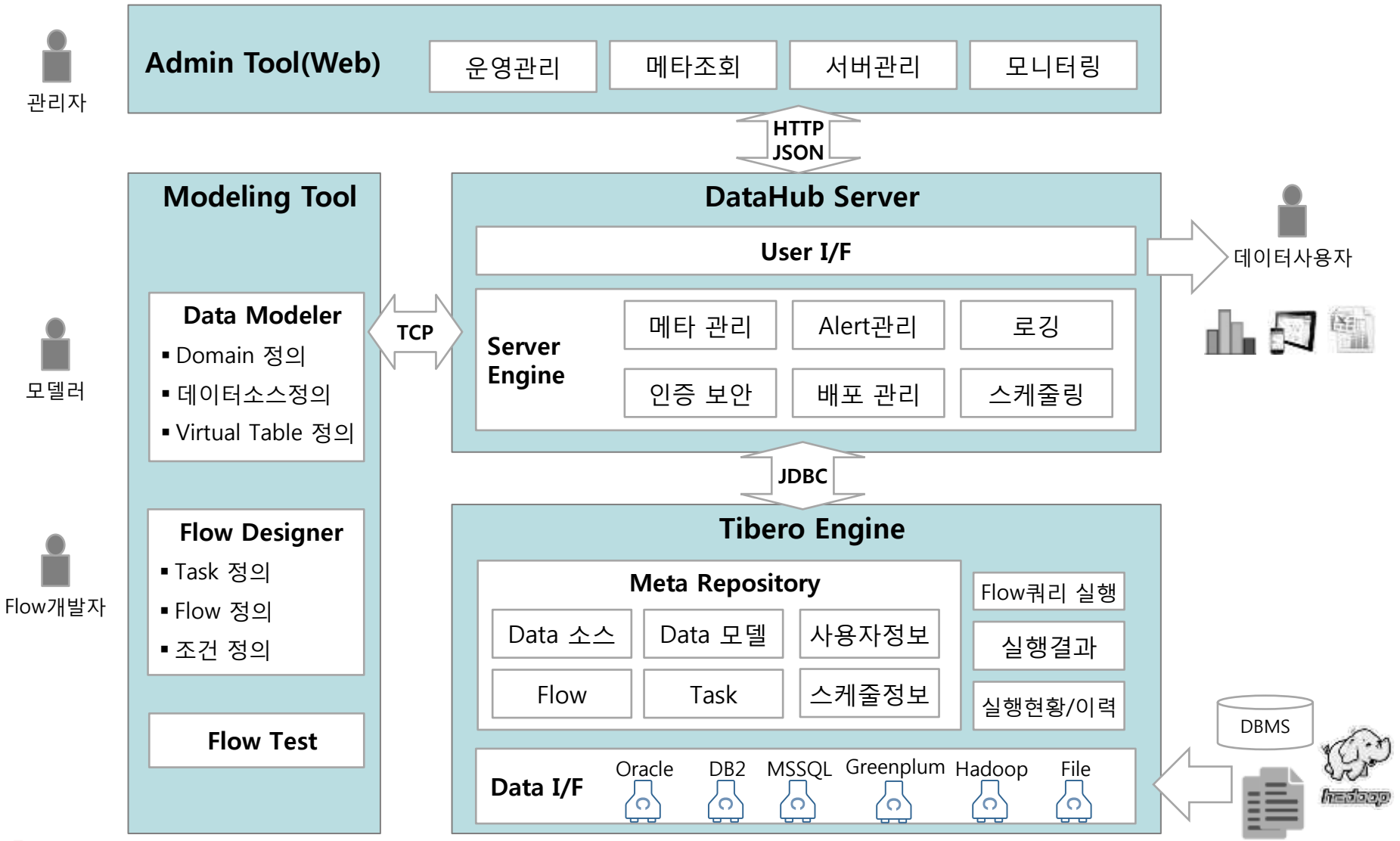
- UI 활용한 Flow 기반 쿼리 설계
- 정규화 과정을 통해 기술용어를 비즈니스 용어로 변환

고성능 쿼리실행 및 결과처리

- RDBMS엔진 내장, 쿼리 최적화
- 분석결과 App, BI 연동 활용

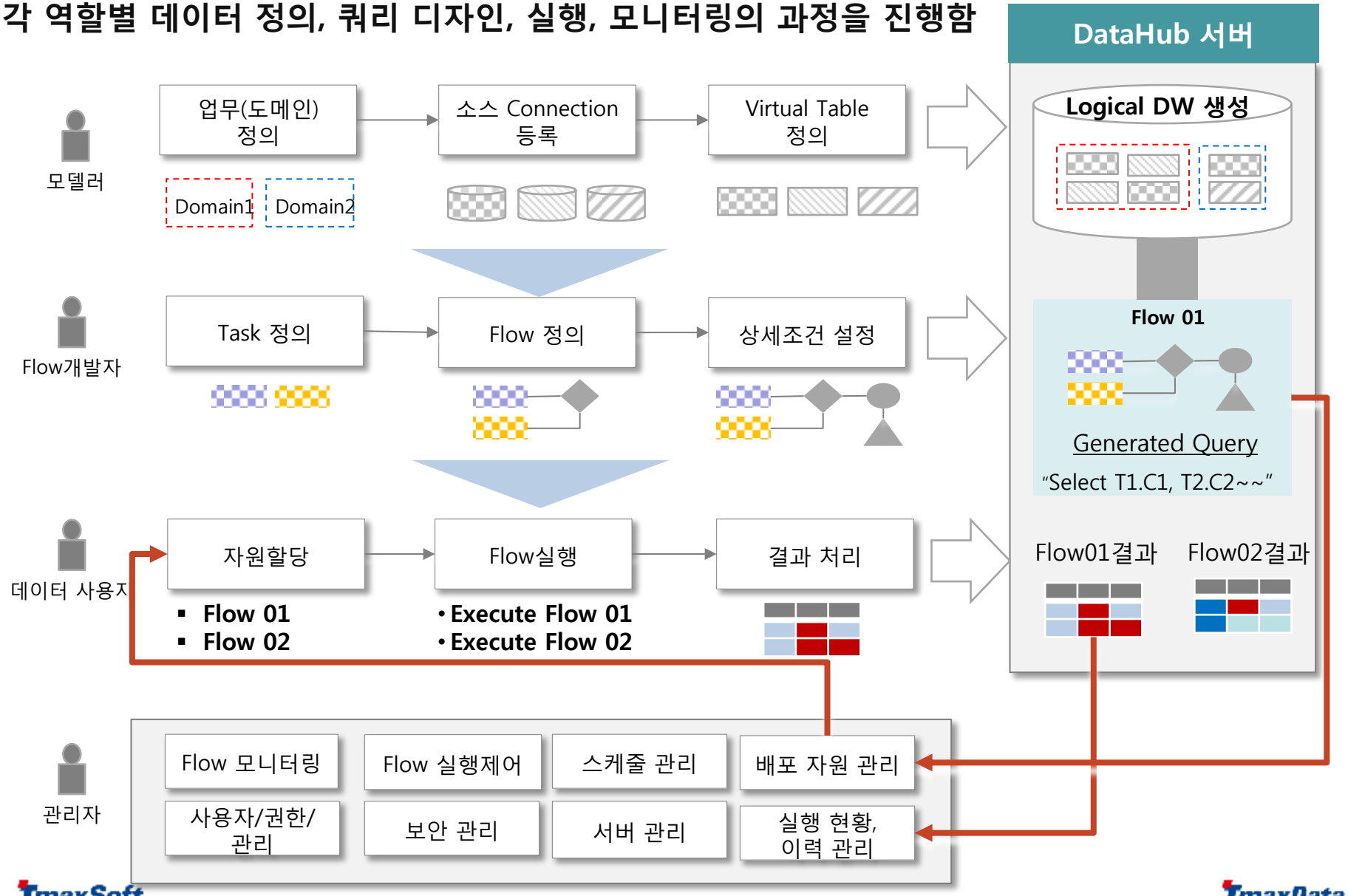
DataHub 아키텍처

Tibero 엔진 기반의 데이터 처리, 관리도구 및 개발도구로 구성됨



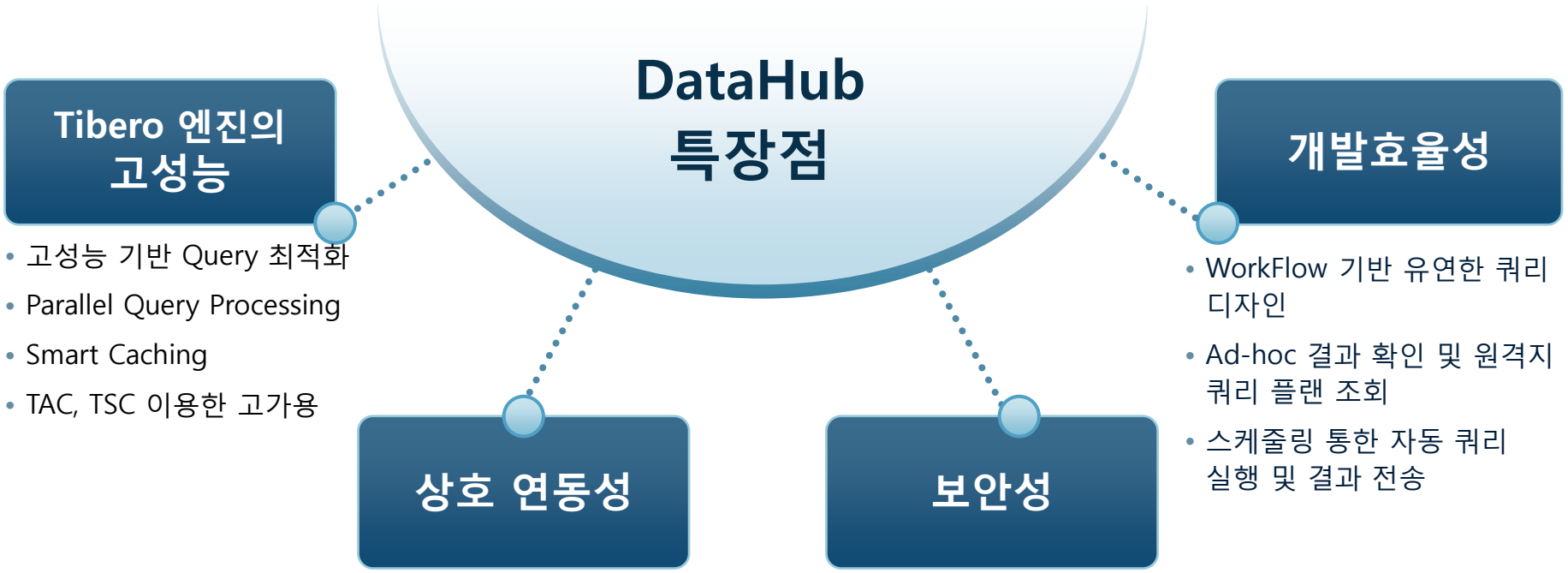
DataHub 사용

각 역할별 데이터 정의, 쿼리 디자인, 실행, 모니터링의 과정을 진행함



DataHub의 우수한 기능

내장된 Tibero 엔진을 이용한 고성능, 다양한 소스접근 및 인터페이스 지원, 데이터 보안성 및 개발 효율성이 향상이 기대됨



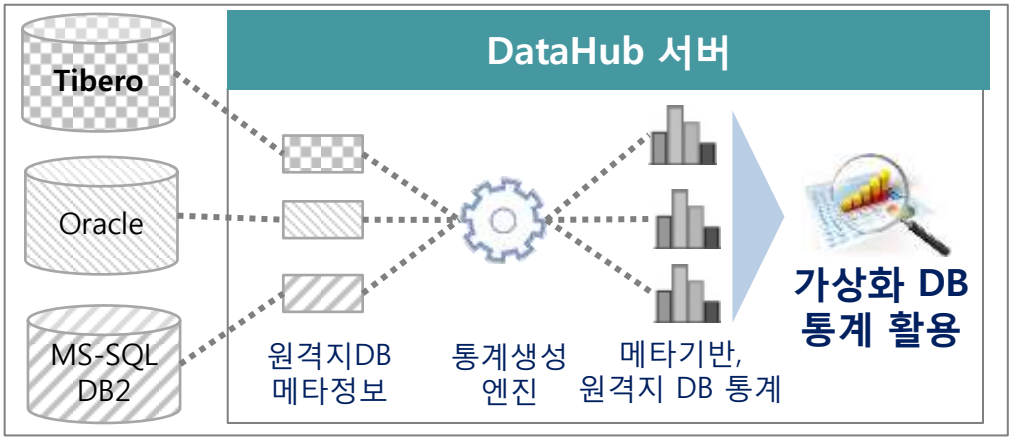
고성능 구현 – Query 옵티마이저

Tibero RDBMS 옵티마이저에 DataHub를 위한 추가된 기능을 통해 원격 DB접속시 최적의 Query Plan 생성 가능

Tibero 옵티마이저

- CBO기반, 최적의 SQL 실행방법
- Query 수행 비용 산정
- 경험기반 통계정보 활용

원격 DB 통계 정보 생성

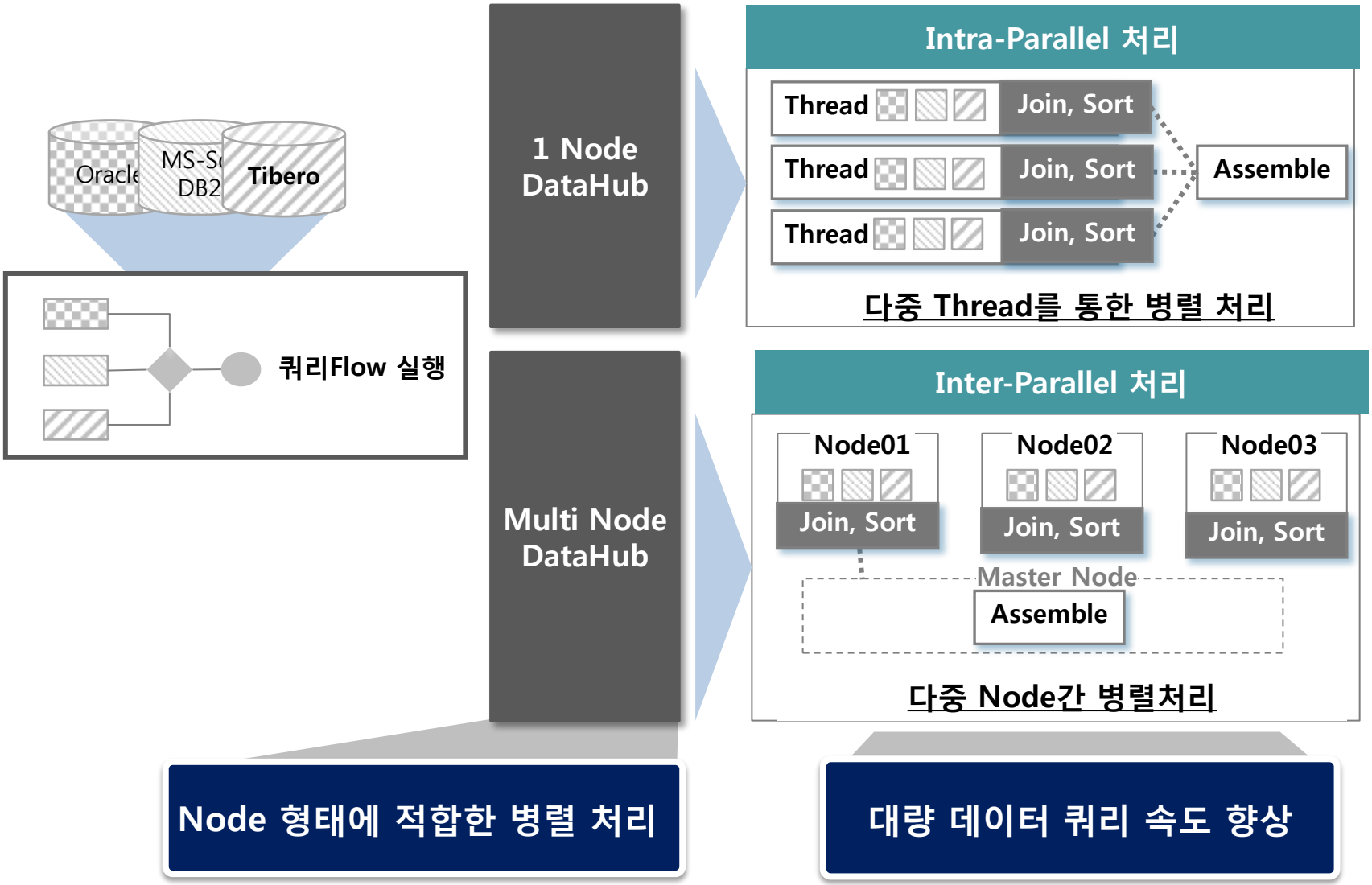


Tibero RDBMS 엔진 옵티마이저 기능 + **가상화 환경에 특화된 최적화 방안**

Logical DW 쿼리 고성능 구현

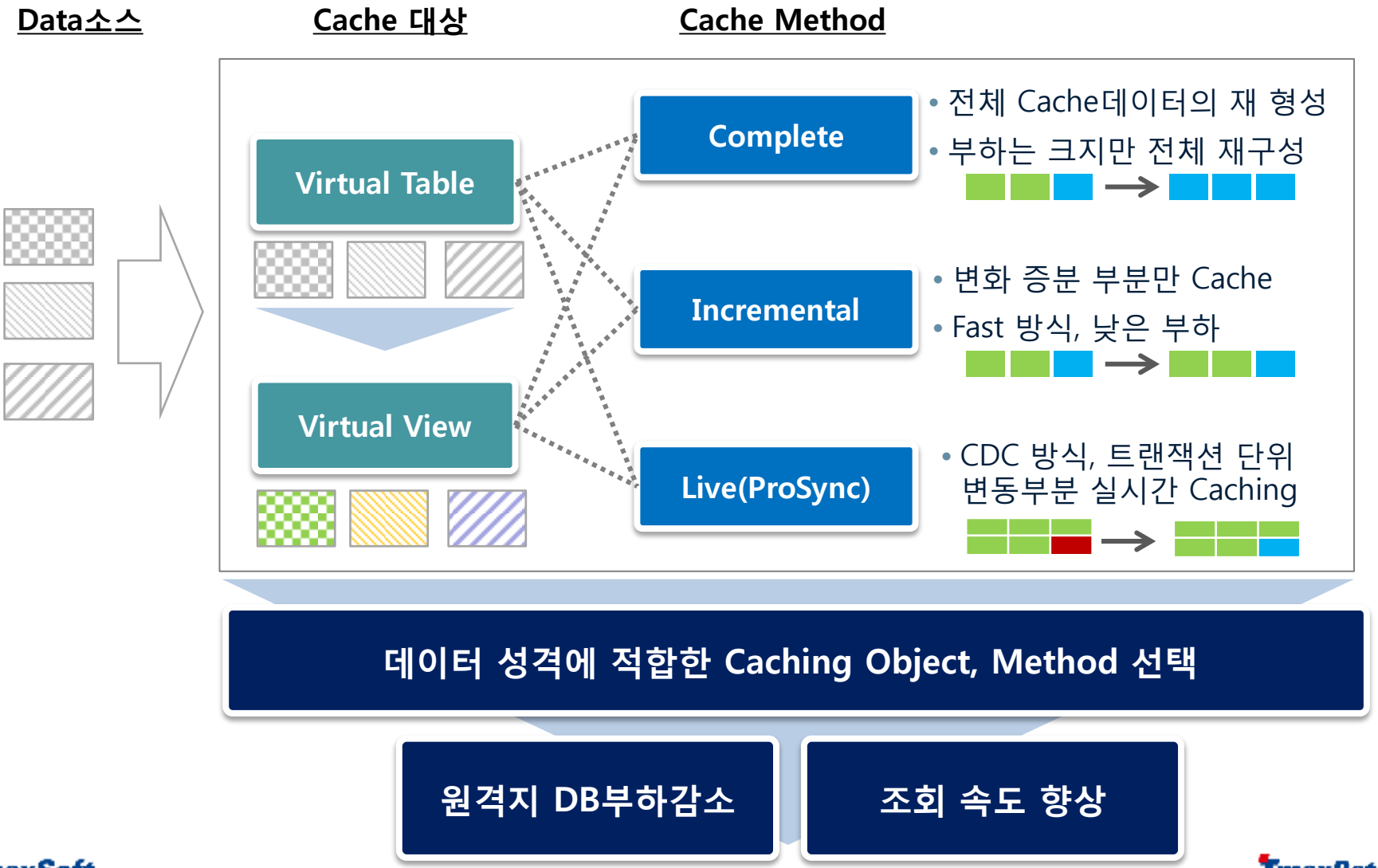
고성능 구현 – Parallel 처리

Intra Parallel 및 Inter Parallel 처리로 대량 데이터 처리시 최적의 성능 보장



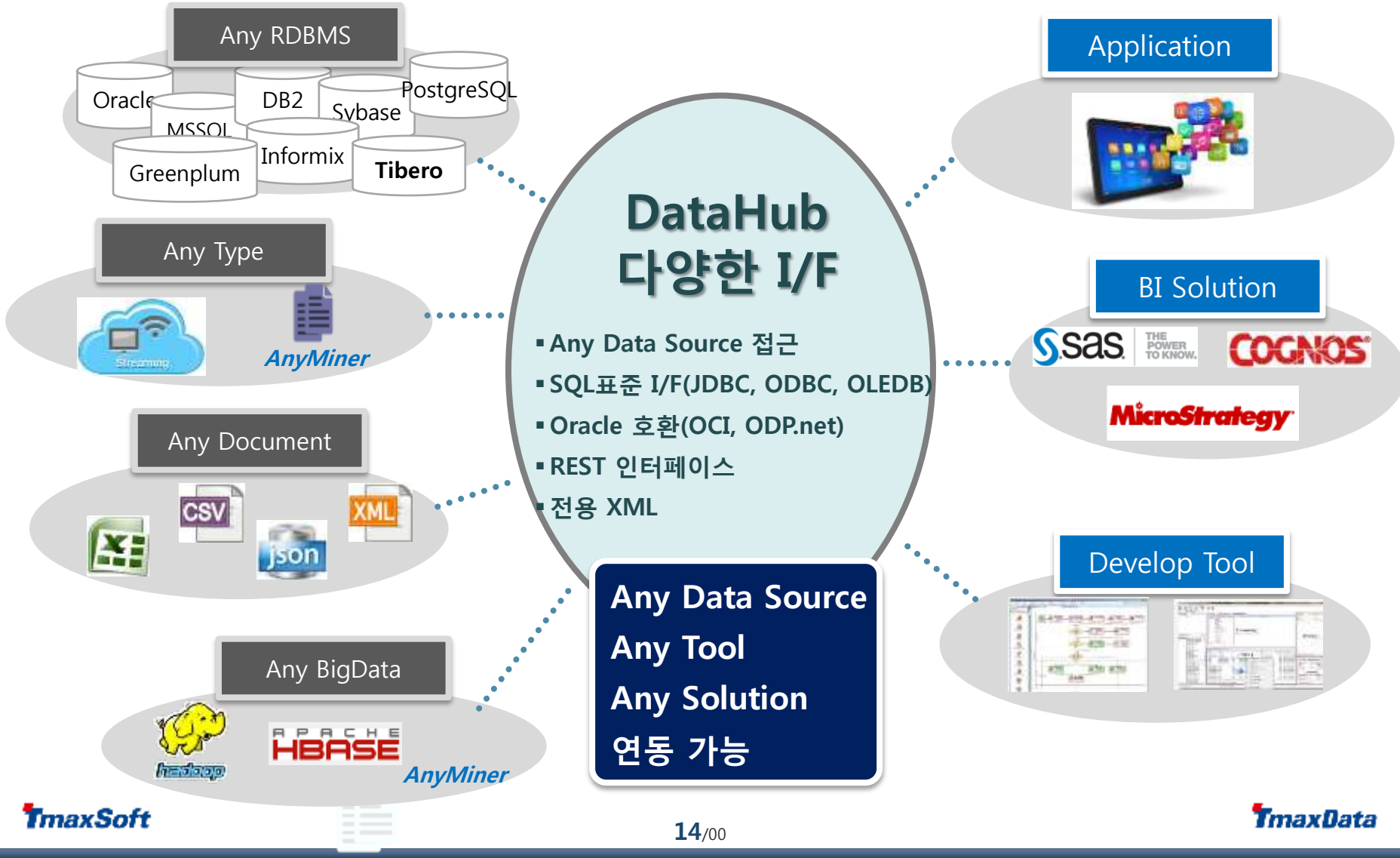
고성능 구현 – Smart Cache

DataHub 내에 존재하는 Virtual Object에 대한 Cache 및 용도별 다양한 방법의 Cache 가능



완벽한 상호 연동성

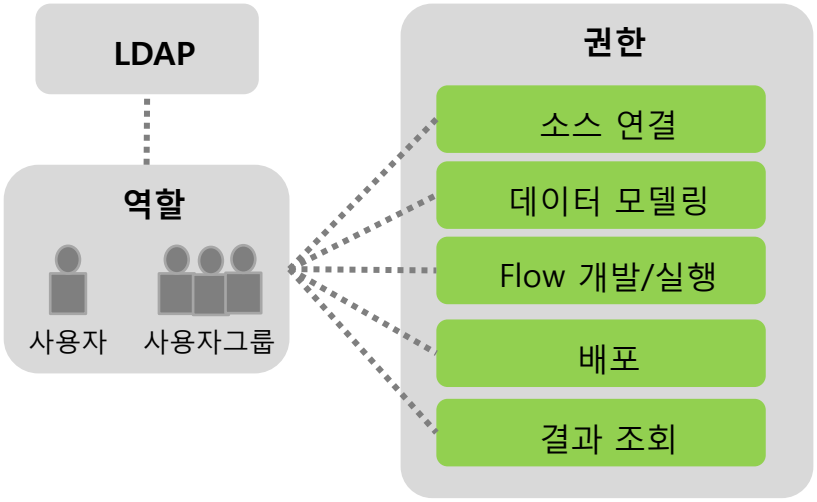
다양한 형태의 소스 데이터와 다양한 계층의 데이터 사용자를 연결하는 진정한 데이터 중개자 Hub 역할



보안성

사용자 권한/인증 관리 및 Access Control, Masking 기능을 통해 원격지 DB 활용 시 발생할 수 있는 보안 문제를 해결

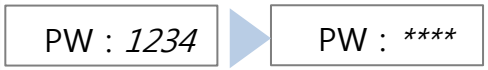
사용자/권한 관리



Advanced 보안

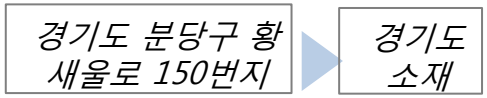
보안 솔루션 연동
- 3rd Party 암호화, 접근제어 솔루션 연동

Data Masking



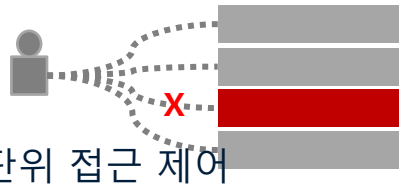
- 주요 보안 데이터의 Masking 처리

Data 익명화



- 주요 정보를 식별 불가능 하게 처리

Row단위 접근제어



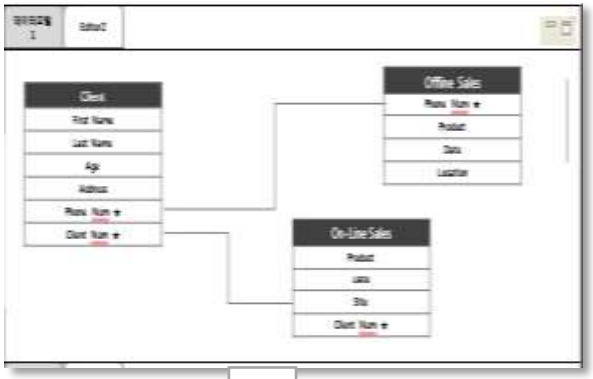
- 사용자/역할별 Row 단위 접근 제어

원격지 Data 보안 강화 기능 제공

개발 효율성

유연한 모델링 및 개발 툴, Ad-hoc 쿼리 작성 및 플랜 확인기능을 통해 개발 생산성을 높임

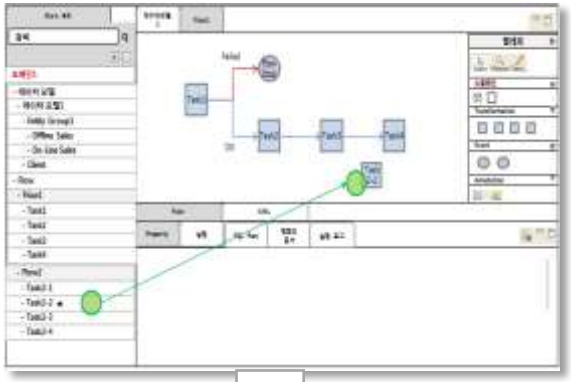
Data Modeler



직관적 Data 모델링

- **Auto Discovery →** 데이터 소스에서 제공하는 정보 기반 모델링
- **관계형 모델링→** Virtual Table간 Relation 정보 확인

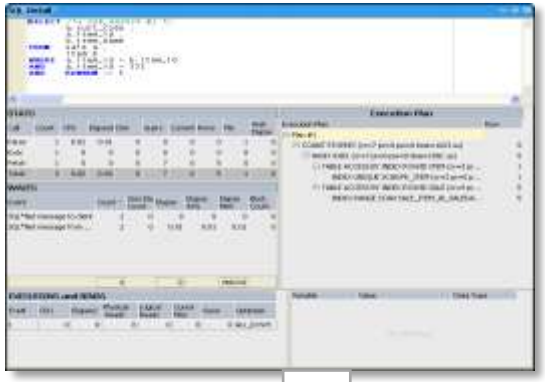
Flow Designer



유연한 Flow 개발

- **Drag & Drop 방식, Task 재사용**
- **다양한 Task 제공** → Query, Join, Filter, Sort, FileWrite, FileRead,등
- **조건형 Flow설계** → 선행Task성공 여부에 따른 분기, 후행Task 실행 순서 지정

View / Monitoring

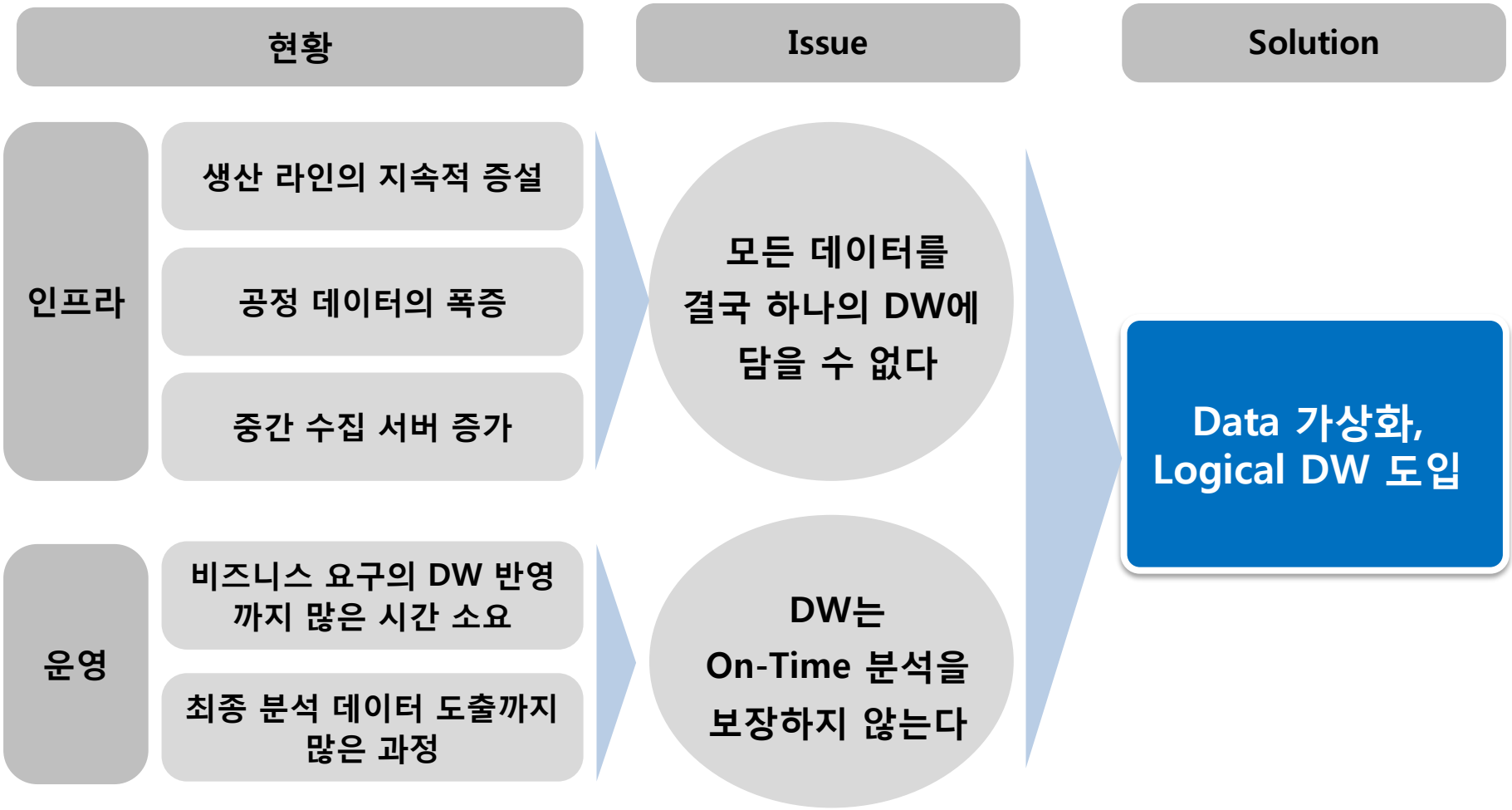


개발 가시성 확보

- **Ad-hoc** 쿼리구문 실행, 결과확인
- Flow의 실행 Plan 확인
- Flow Task 중간 결과 조회 가능
- 실행 Log View 제공

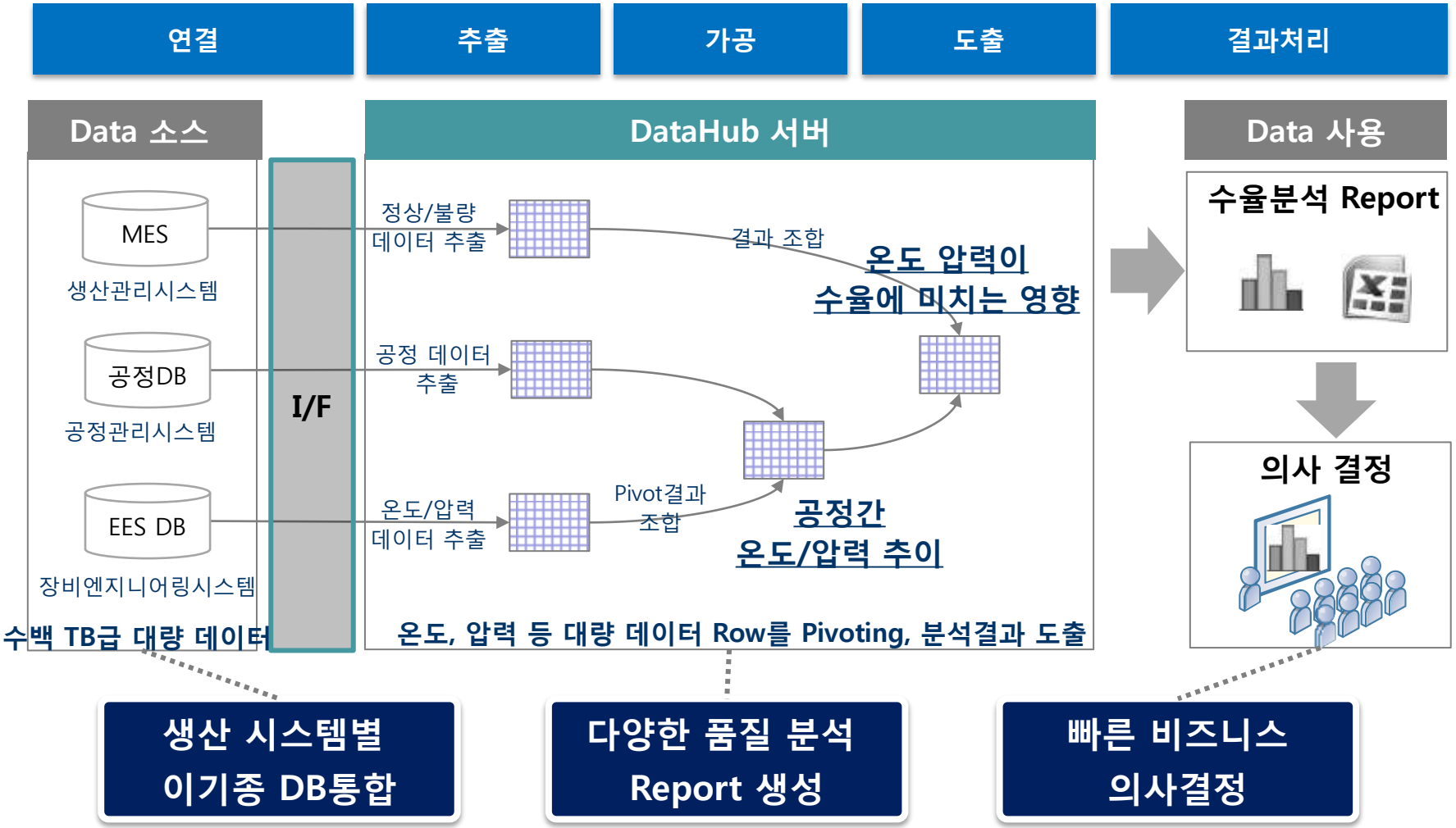
Use Case – 제조업 생산 수율 분석(1/2)

반도체 공정의 폭발적 데이터 증가에 따라 기존 DW의 공간적, 시간적 문제점을 해결하여 유연한 분석과 비즈니스 의사결정 적시성 확보를 위해 DataHub 도입



Use Case – 제조업 생산 수율 분석(2/2)

라인에서 발생하는 실적, 품질 정보를 취합, 수율 분석 결과를 주기적으로 생성 보고하여 공정 및 품질운영 의사 결정에 활용

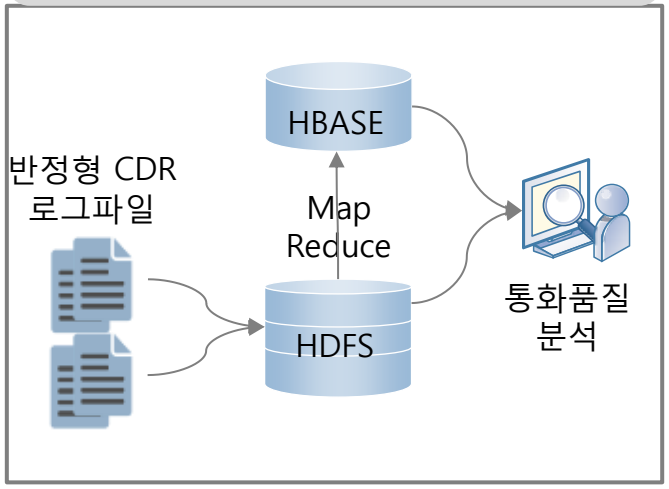


Use Case – 통신사 통화 품질 분석(1/2)

대량 비정형 CDR데이터 분석을 위해 Hadoop 기반 빅데이터 플랫폼을 도입하였으나 RDBMS 정형 데이터와의 실시간 통합분석이 필요한 상황임

현황

- 통화 기록 파일 증가
- 통화 품질 분석의 요구
- Hadoop 인프라 구축



Issue

네트워크 현황과
상관관계 분석 요구 발생

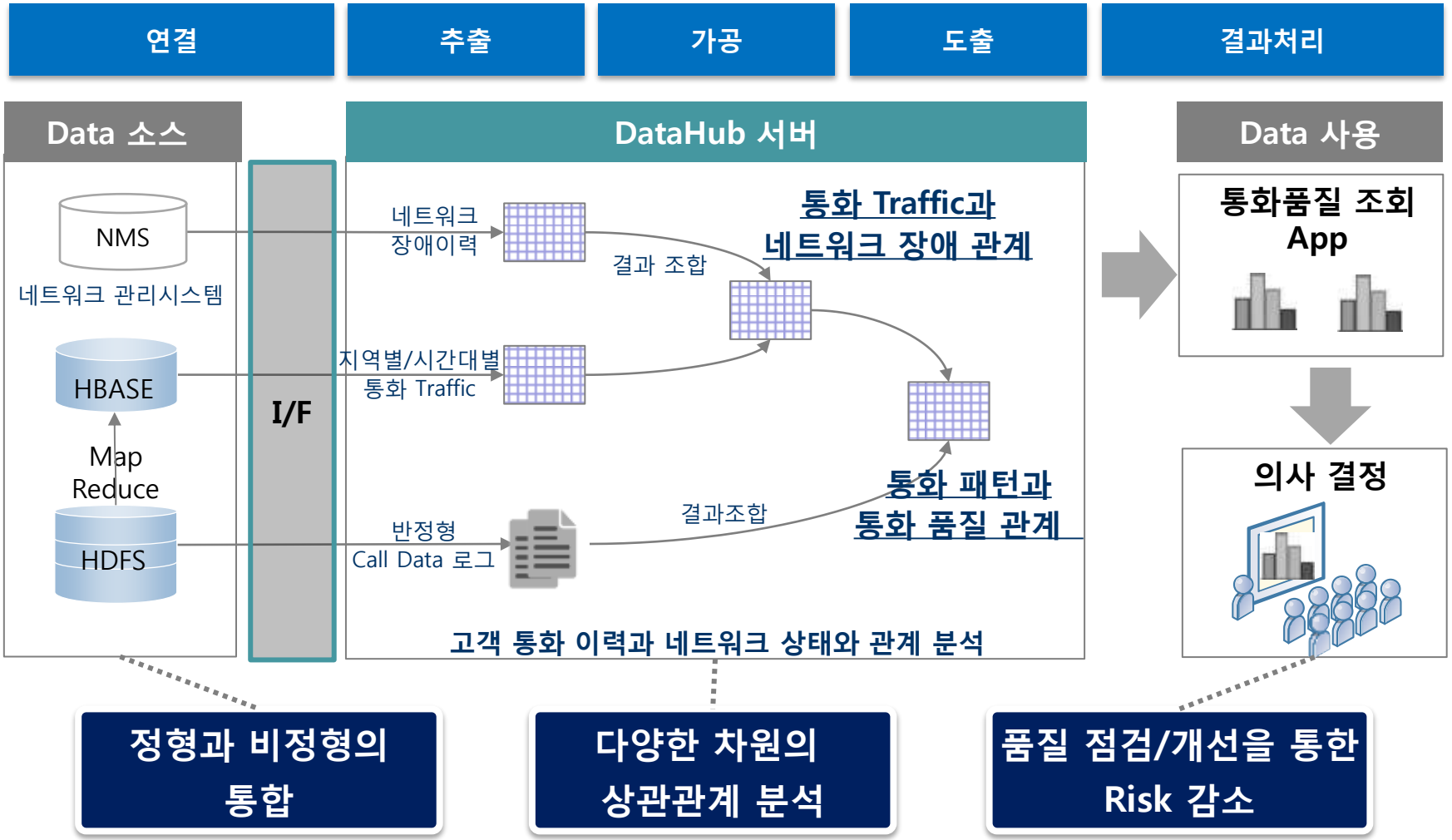
비정형 만으로는
완벽한 분석이
어렵다

Solution

이기종 정형/비정형
통합분석 플랫폼
구축

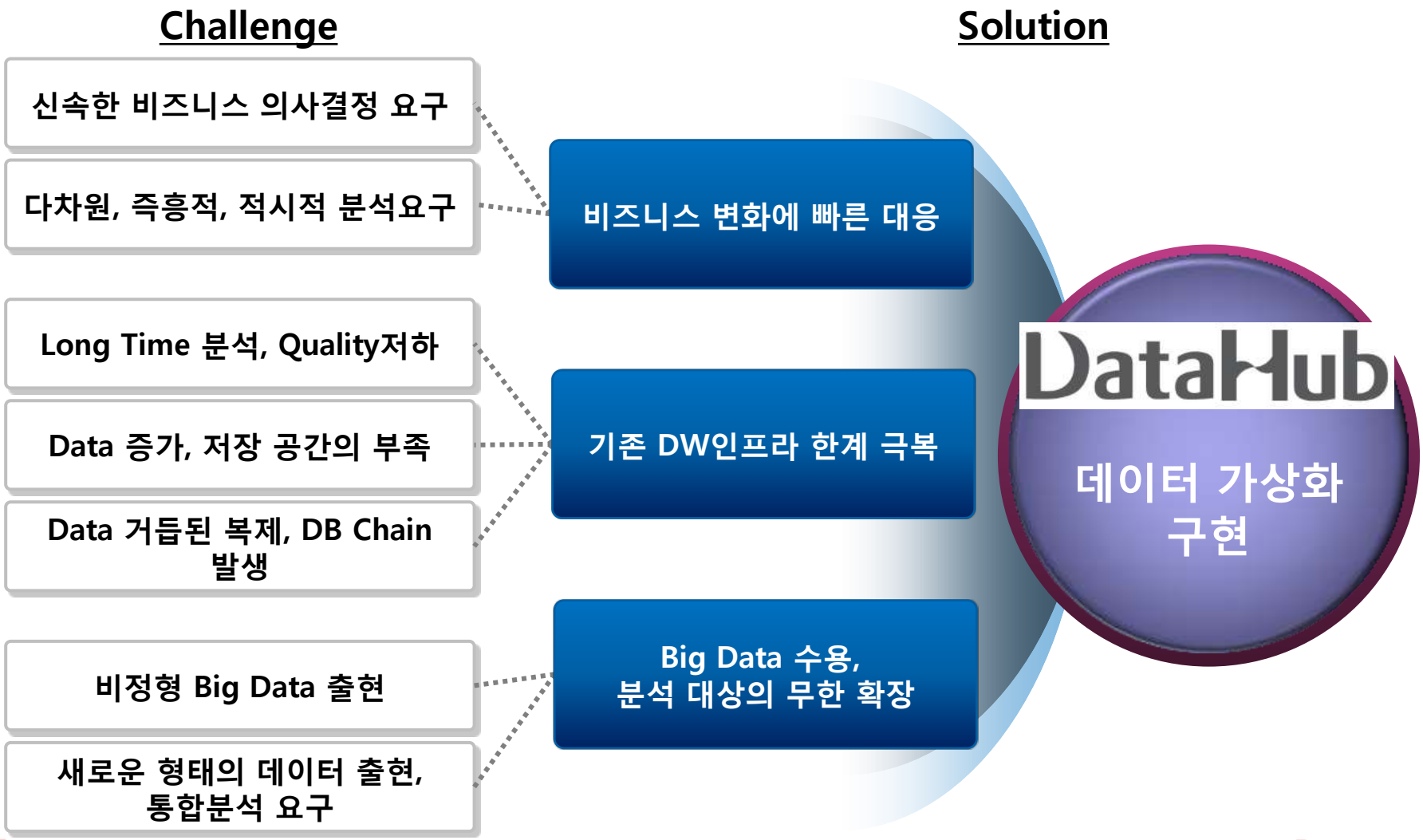
Use Case – 통신사 통화 품질 분석(2/2)

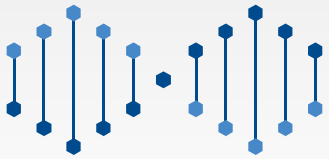
기 구축된 Hadoop 인프라가 보유한 대량의 Call Data 비정형 로그와 NMS 정형 데이터를 연계하여 통화 품질을 분석



맺음말 - BI 시장의 Challenge, DataHub를 통한 극복

Tmax의 DataHub는 기존 DW인프라를 극복한 데이터 가상화 기반의 통합 분석 플랫폼으로서 유연하고 신속한 분석과 On-Time 비즈니스 의사결정을 지원함





Thank you!