

4차 산업혁명시대의 기본

데이터 품질

산업혁명 4.0 : 데이터의 연결화

What is industry 4.0

Water and steam power ushered in the era of mechanized production.

1차 산업혁명
기계화

The invention of microprocessors brought Automation into plants on a large scale.

3차 산업혁명
자동화(S/W)

2차 산업혁명
대량화

Era of division of labor and mass production,
Epitomized most famously by Henry Ford.

Ubiquitous sensors and big data analytics
connected together into cyber-physical systems.

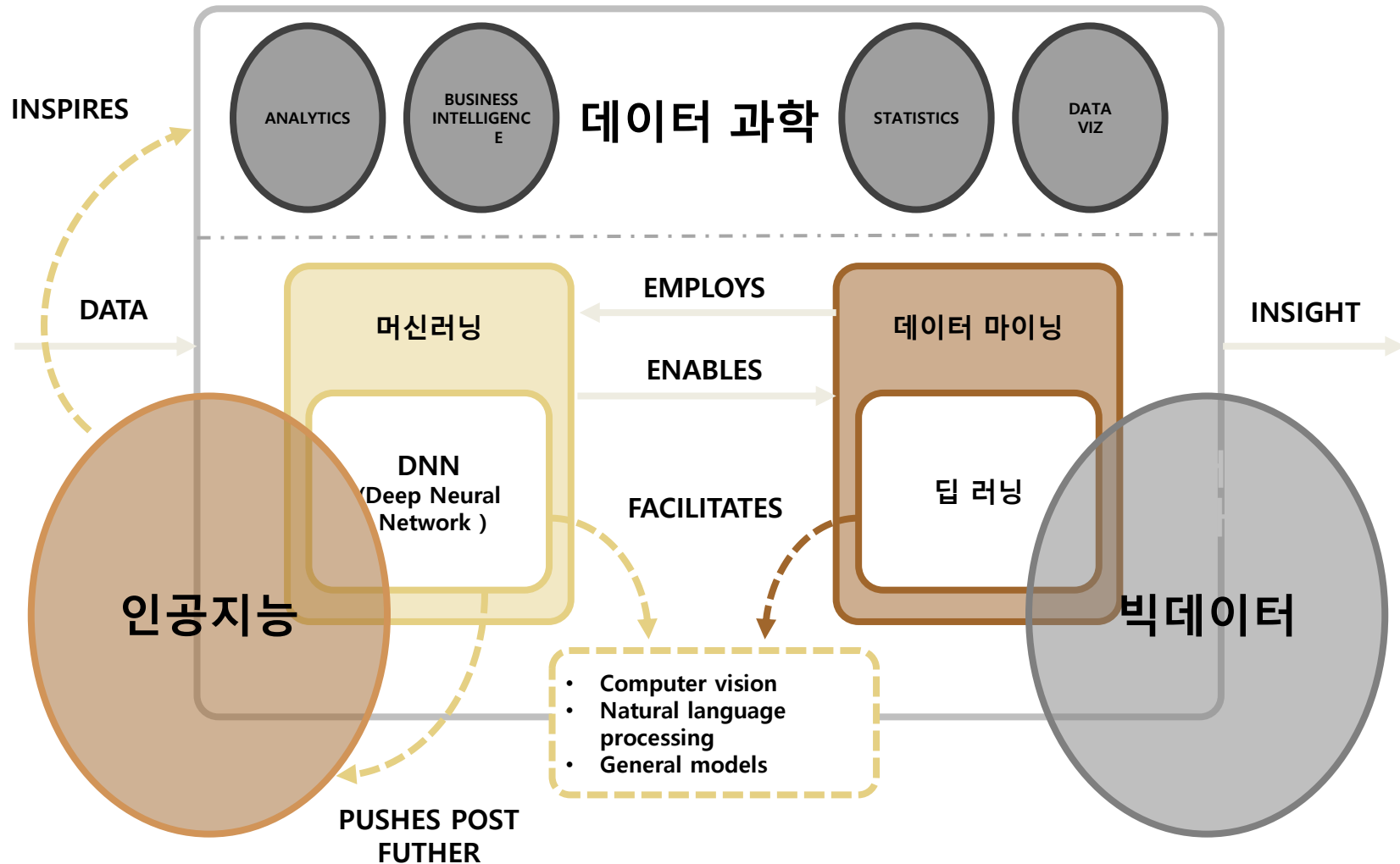
4차 산업혁명
데이터의 연결
화

데이터 혁명

전통 기업의 데이터 혁명	농업혁명	제조혁명	의료혁명
		  	 
	물류혁명	유통혁명	금융혁명
	 	 	 

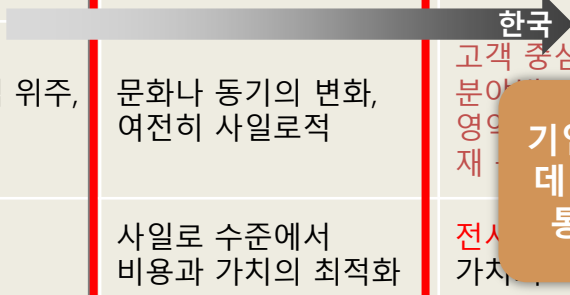
스타트업 기업의 데이터 혁명	DNA 데이터 분석	이상 감지, 데이터 분석	데이터 가공/유통
			
	O2O	개인자산관리	트위터 뉴스
			

Data Science vs. AI vs. Big Data(IOT)



데이터 활용 5단계

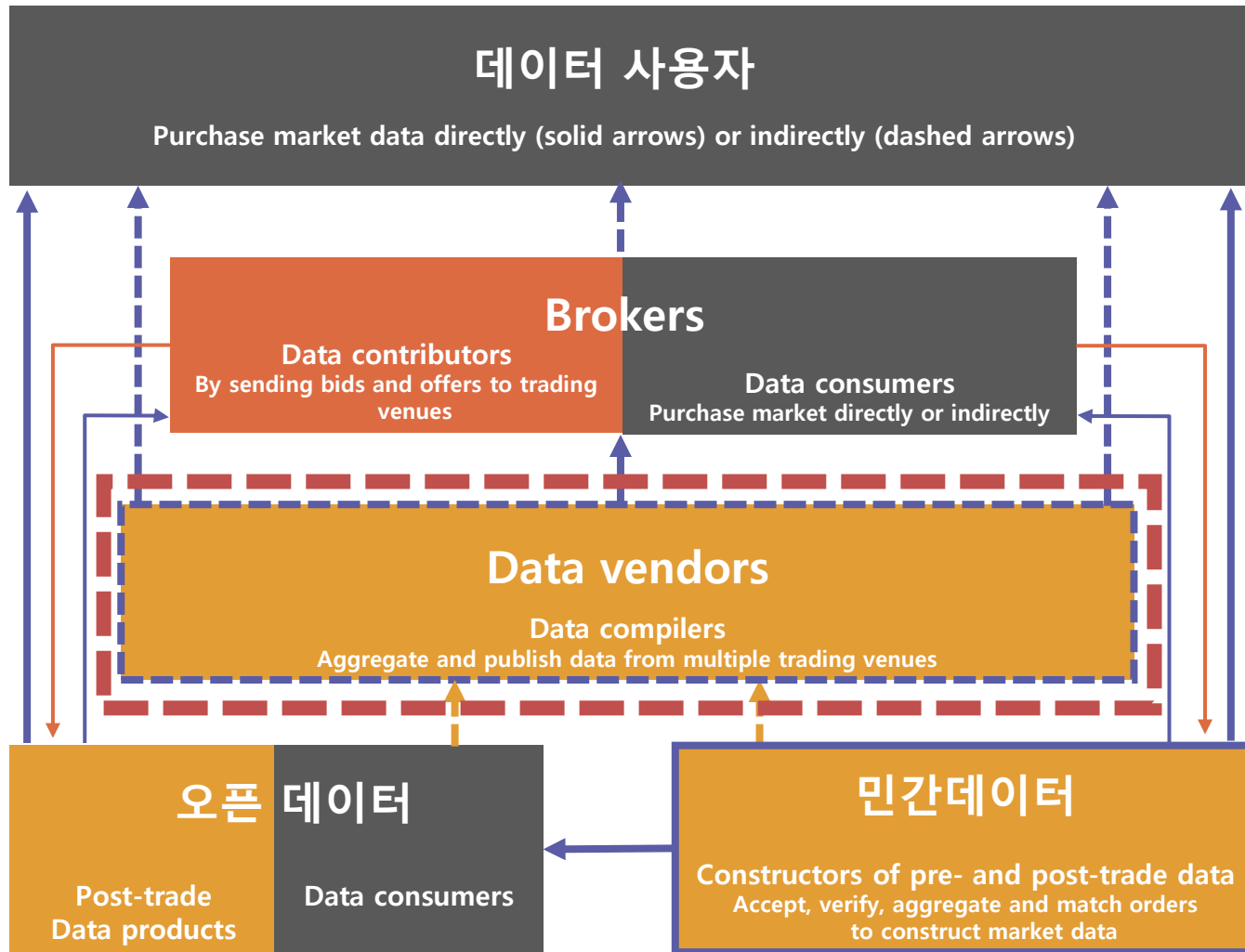
	1단계	2단계	3단계	4단계	5단계
비전 (Vision)	없음	생산 성과 비전 시도	각 영역별 채널의 효율성	전사 내부 통합	연결된 가치 부여 및 인식
전략 (Strategy)	없음	고립적 프로젝트, 아래로부터 시도	좀더 협력적 생각, 여전히 사일로 존재	전사적 통합 CRM 프로그램	서로의 이익을 위해 협력하는 가치 인식
고객 경험 (Customer Experience)	개념 없음	개념 없음	이해하는 사일로 수준에서 집중	각 영역별 연결된 비즈니스로 이해하고 집중	좀더 넓은 영역 이해, 협력
협력 (Collaboration)	내부 영역별로 집중, 사일로(Silo)구조	초기 편협적 고객 위주, 사일로적	문화나 동기의 변화, 여전히 사일로적	고객 중심, 분야별 영역별 데이터 재통합	고객 목표와 기업 목표의 통합
프로세스 (Process)	내부 영역별로 집중, 사일로적	초기 자동화시기, 사일로적	사일로 수준에서 비용과 가치의 최적화	전사적 데이터 통합 비용과 가치	초기 자동화 시기, 실제 프로세스
정보 공유 (Information)	기본적 정보의 산재	팀 기준, 산재, 최소의 인사이트	사일로 수준 정보 공유, 인사이트 발전 과정	전사 관여 정보 공유 및 인사이트	기업을 넘어선 인사이트와 정보 공유, 외부 데이터 활용
기술 (Technology)	몹시 산재되고 미미한 기술	산재 존재, 한정된 역할과 집중	사일로 내에서 높은 수준의 역할	전사 통합 수준의 높은 수준의 역할	기업을 넘어선 높은 수준의 역할
집중 영역 (Metrics)	적은 내부적 집중 영역	산재되고 한정된 집중 영역, 운영 내부적 집중	사일로 내에서 효율적, 고객 집중 부족	전사적 고객 집중/균형 있는 구조	목표 공유, 균형 있는 구조, 잘 연결, 정리된



미국 2000년 초

빅데이터(OT, 외부 데이터, 소셜)

데이터 경제 시스템



빅데이터 활용의 대부분의 문제 → 통합과 가공의 문제



1. 데이터 통합, 관계, 원인 또는 분산 등의 복잡성
2. 빅데이터 기술 수행하기 위한 전문가의 부족
3. 활용 과정의 고비용
4. 빅데이터 기술의 디자인(기획)과 적용을 위한 긴 시간
5. 분석 전문가의 부족
6. 방대한 데이터에서 활용 가능한 데이터의 선택
7. 낮은 데이터 품질
8. 기존의 분석 시스템과의 통합의 어려움
9. 충분하지 않은 컴퓨팅 자원(분석처리)

데이터 품질 관리의 필요성 (1/2)

360° 통합 데이터 관리의 필요성

데이터 급증 – IDG / Mckinsey Global Institute (MGI) / Veritas

- 비즈니스 디지털화 추세, 2005년 이후 전세계적 대역폭 양이 45배 증가, 향후 5년간 현재 수준에서 9배 더 증가
- 포춘 500대 기업 – 매년 39% 데이터 증가
- 스토리지 용량 요구사항의 증가 속도가 기업에서 개별 파일을 생성하는 속도보다 9% 빠름

기업 데이터 관리의 문제점 – IDG / Mckinsey Global Institute (MGI) / Veritas

- 일반적인 기업의 데이터 41%는 3년간 한번도 액세스 된 적이 없으며, 12%는 7년간 수정되지 않은 상태
- 수십 테라에서 수십 페타까지의 불필요한 데이터에 대한 관리 비용
- 기업의 전체 데이터 중 **5%만이 중요한 데이터로 확인**되며 이러한 데이터 조차 가용성/유용성/효율성을 확보하지 못함

기업 경쟁력 확보를 위한 데이터 거버넌스 확보 시급 – Gartner / Infosys / IDG

- 전세계 데이터의 크기는 2020년까지 40 제타바이트를 초과
- 구조화된 데이터는 매년 40%씩 증가, 구조화 및 비정형 데이터 포함 매년 80%씩 증가
- Machine에서 생성된 데이터는 2020년까지 매년 15배씩 증가
- **데이터 분석, 활용, 가공을 위한 핵심은 데이터 거버넌스 체계 수립 (MDM, 메타 데이터, 보안, 품질, 통합, 흐름)**

데이터 품질 관리의 필요성 (2/2)

데이터 기반 비즈니스 분석.활용의 증가

클라우드/오픈소스 데이터 저장소 활용 급증 - IDC

- 2020년 세계 클라우드 시장 규모가 1950억달러(219조원) 전망
- 2017년 전체 퍼블릭 클라우드 시장에서 서비스형(SaaS)과 플랫폼형(PaaS)이 83.7%를 차지했고 인프라형(IaaS)가 16.3%를 기록
- 한국의 경우 2020년까지 연평균 16.3%의 성장세
- 클라우드 서비스 제공자에 의한 오픈소스 DBMS의 자연스러운 사용 증가

데이터 분석 환경의 변화 - IDC / Gartner

- 정형·비정형데이터나 기업 내·외부에 있는 데이터의 수집·관리, 이를 저장하기 위한 DW(Data Warehouse)와 BDP의 연계 확대
- 데이터 분석의 방향이 기존의 Sample Data, 또는 특정 중요 시스템 위주의 분석 환경에서 전사의 모든 데이터를 분석 환경으로 변화.이에 ROI를 극대화할 수 있는 Hybrid 형태의 데이터 분석 환경의 도입이 보편화되고 있는 실정과 동시에 데이터 정합성 문제 대두.
- 인공지능(AI) 및 머신러닝을 위한 알고리즘, 비즈니스를 지원하기 위한 BI(Business Intelligence)와 빅데이터 분석 요구 증가
- 멀티채널 데이터 흐름에 의한 데이터 정합성 및 빅데이터 거버넌스 관리 요구

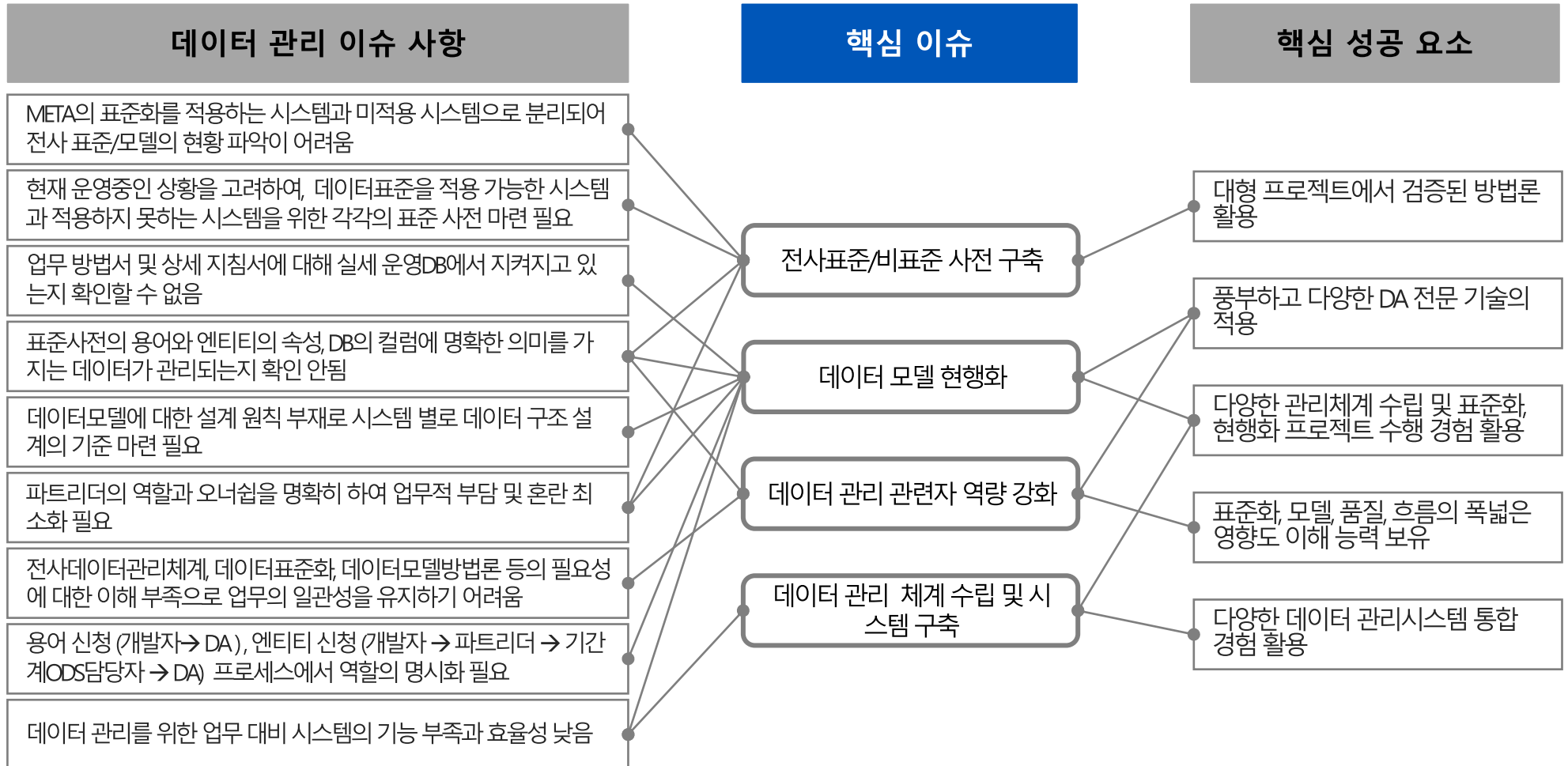
데이터 기반 디지털 트랜스포메이션 - Forrester / IDC

- 모바일, 클라우드, IoT, 인공지능(AI), 로봇 등 지난 10년간 디지털기술의 놀라운 혁신으로 자동화, 지능화가 가속되면서 기업경영, 고객관리, 비즈니스 모델, 운영 프로세스 등에 대해 기존 방식과 다른 새로운 접근방식과 시도
- 옴니채널 (Omni-Channel) 마케팅-세일즈 방식과 온디맨드(On Demand) 서비스의 증가
- 디지털 트랜스포메이션의 기반은 데이터의 정확성.실시간성.유효성 요구

현행 데이터 관리 체계 이슈 사례

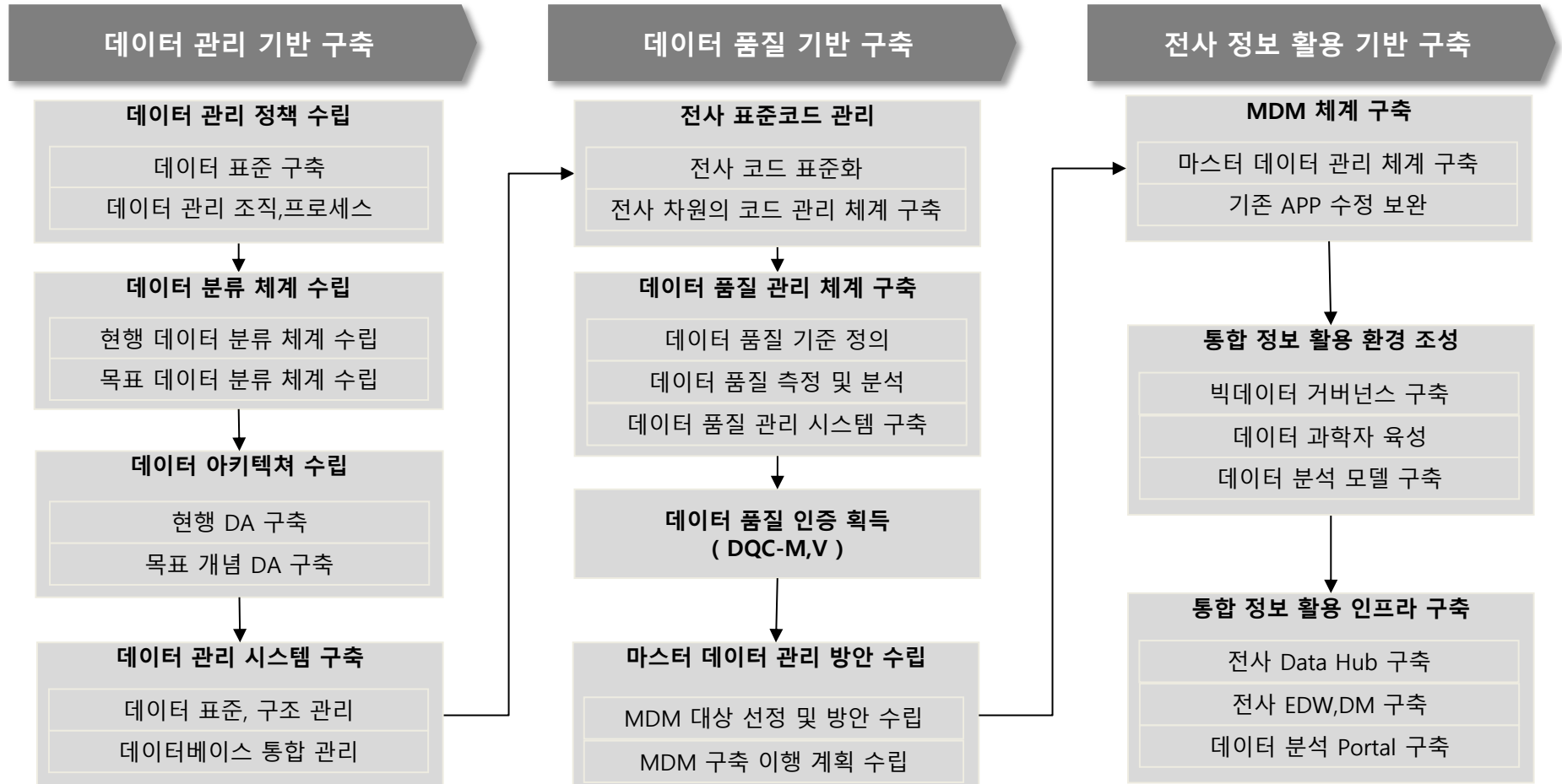
장기적으로 관리할 수 있는 데이터 관리체계 수립을 위해 명확한 분석과 수행 방법을 요구하고 있습니다.

향후 데이터 관리 업무의 효율성을 높이면서 지속적으로 일관되게 관리될 수 있는 주요 이슈에 대해 종합적인 분석과 해결 방안이 필요합니다.



데이터 품질관리 체계 구축을 위한 Roadmap

데이터 품질 관리 체계 구축을 통해 데이터 관리 역량 확대 및 의사결정 역량을 극대화할 수 있습니다. 데이터 품질 관리 체계 구축은 단 기간에 추진하기 보다는 중, 장기적인 Roadmap을 가지고 일관성 있게 추진하는 것이 성공의 필수 요소라고 할 수 있습니다.



데이터 품질 관리 체계

데이터 아키텍처 프레임워크 기반 데이터 관리체계 정의

» 전사 차원의 데이터 아키텍처 프레임워크를 기반으로 현행 데이터 관리체계를 진단하고 목표 데이터 관리체계를 정의합니다.

☑ 전략, 정책, 프로세스, 콘텐츠, 조직, 기반 시스템으로 구성된 데이터 아키텍처 프레임워크

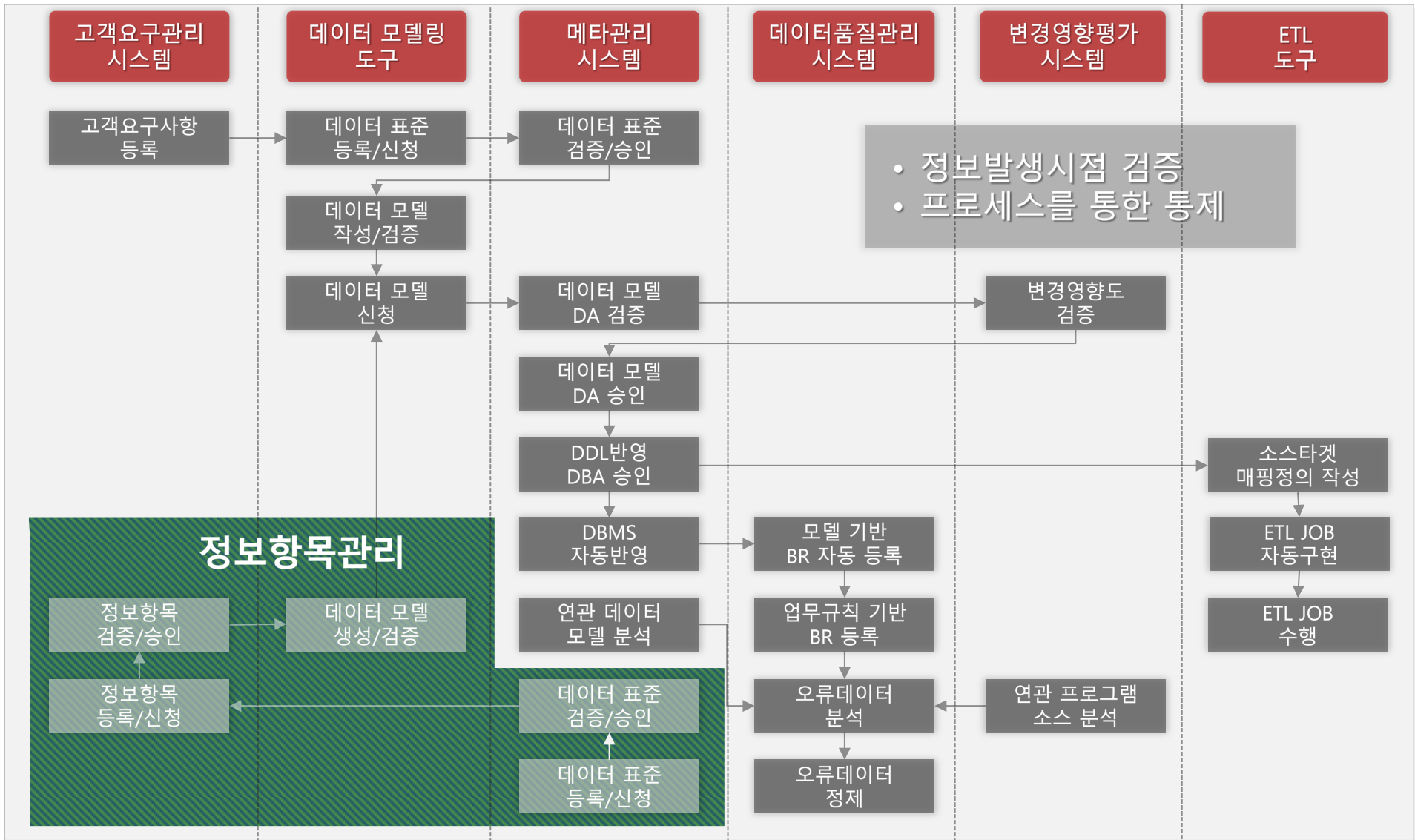


데이터 품질 관리 기능

통합된 관점으로 단계별 프로젝트 기간을 일관되게 유지

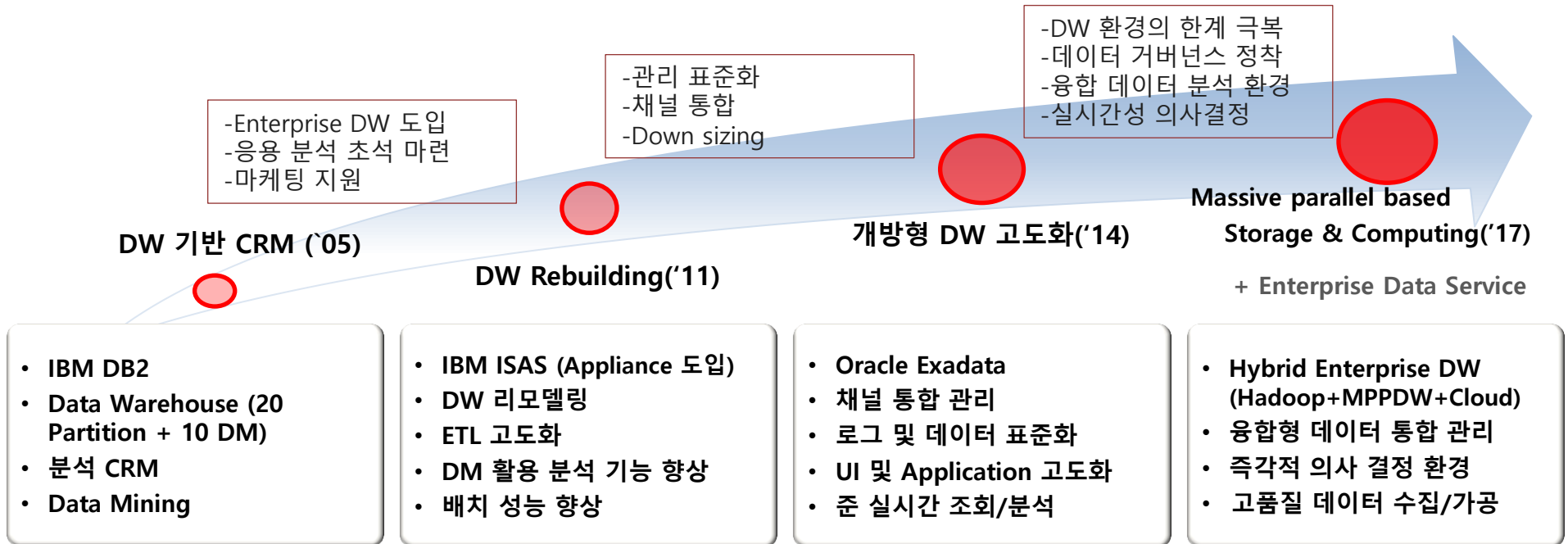
데이터표준	단어 /용어	도메인 코드	명명규칙	항목	준수도 충실도	프로세스	이력	검증 검색 통계 분석 보고서	추적성 영향도 시각화
데이터모델	개괄모델 (분류체계)	개념모델	논리모델	물리모델	프로세스	모델뷰어	이력		
DB카탈로그	TABLE COLUMN	INDEX	VIEW	Const-rains	STO-RAGE	DB PROGRAM	이력		
데이터품질	DQI	CTQ	프로파일링	BR관리	구조품질	프로세스	이력		
AP 표준	UI	Application	표준전문	Interface	검증규칙	프로세스	이력		
응용영향도	기본정보	호출/ 사용 관계	언어별 특성	CRUD Matrix	FUNCTION POINT	변경영향	이력		
데이터흐름	SOURCE	TARGET	MAP-PING	TRANS-FORMATION	JOB	변경영향	이력		
통제/관리	Portal, 사용자/권한, 프로그램 관리, Job Scheduler, Monitoring, 연계								

데이터 관리 시스템 적용 사례



정보 분석 활용 환경 진화 필요성

변화하는 비즈니스 환경에 대응하기 위해 그동안 귀사의 데이터 기반 분석 처리 환경의 진화 방향에 맞추어 통합된 차세대 데이터 서비스 플랫폼과 그에 따르는 거버넌스 체계의 구축 필요성이 증가하고 있습니다.



○ 차세대 BI 기술로서 지향할 4가지 요소에 부합하는 데이터 저장/분석 환경 요구의 수용

1. 기존 인프라와의 융합 (상호 운용성, 확장성)
2. 운영 및 유지보수의 간편성(안정성, 가용성)
3. 비용 (효율성, 재사용성)
4. 의사결정을 위한 쿼리에 대한 응답(성능, 적응성)



데이터 분석/활용 환경의 변화

- 데이터 분석의 방향이 기존의 Sample Data, 또는 특정 중요 시스템 위주의 분석 환경에서 전사의 모든 데이터를 분석 환경으로 변화하고 있다. 이에 투자 대비 효과(ROI)를 극대화할 수 있는 Hybrid 형태의 데이터 분석 환경의 도입이 보편화되고 있는 실정이다.

As-is

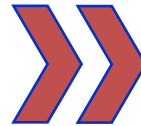
Client-Server & Web



ERP, CRM 등 주요 시스템 Only

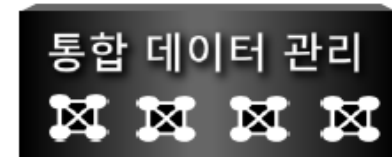
Scale-up / High-end Server & Storage

Relational Database



To-be

Hybrid - Datawarehouse



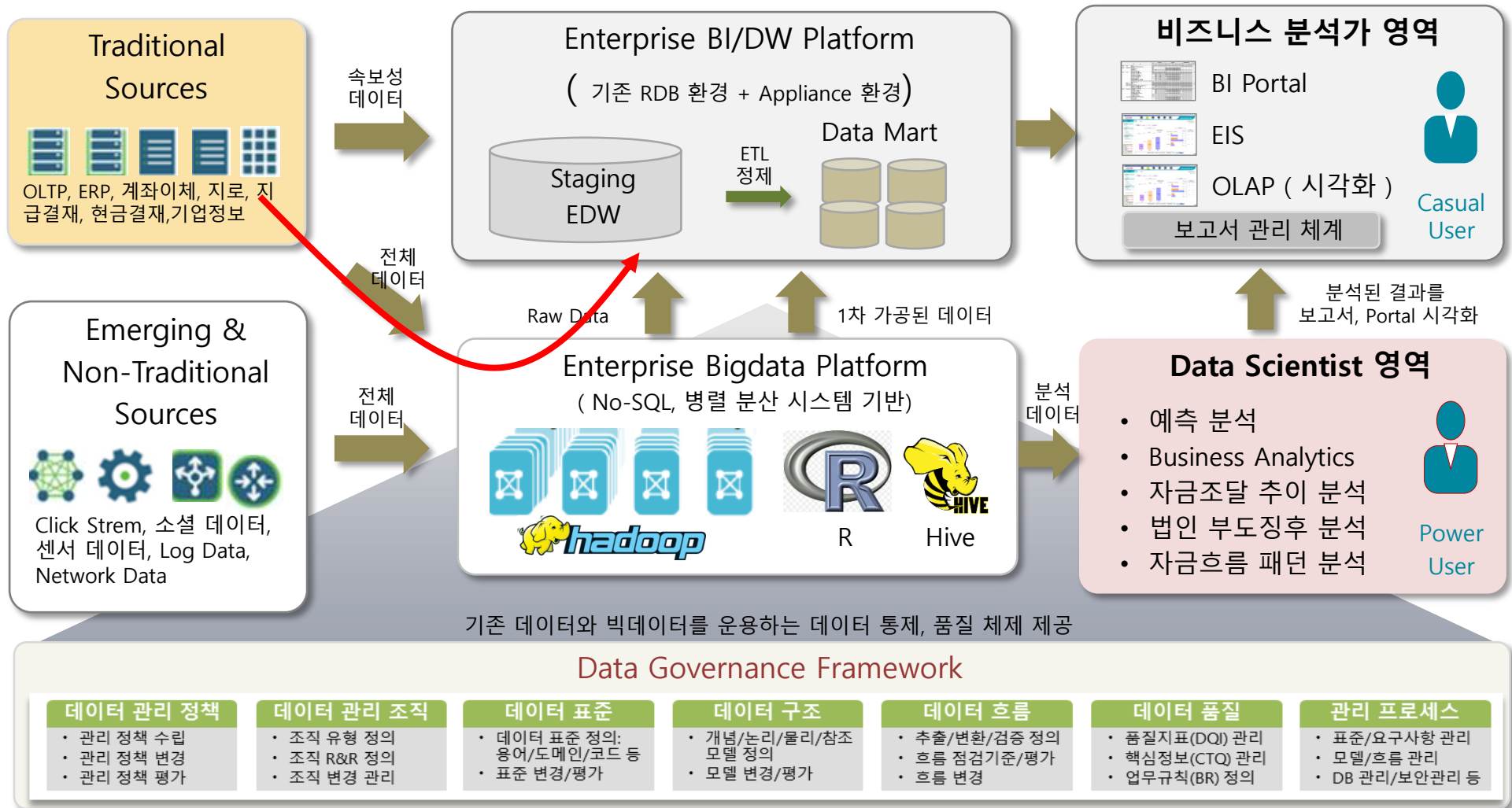
All System, All Data

Scale-out / x86 Server

기존 RDB + Big Data Platform

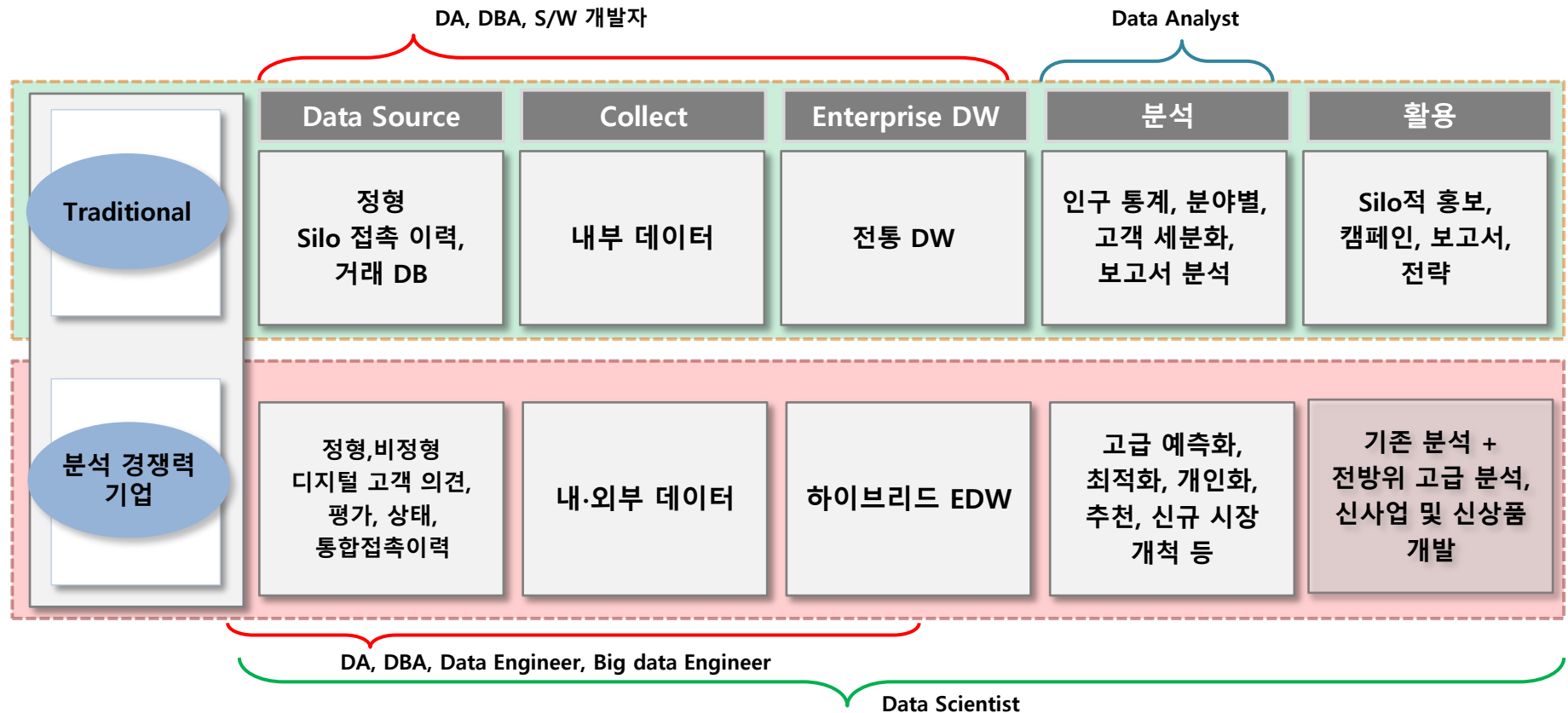
정보 활용/분석 환경 - Hybrid

•데이터 분석의 방향이 기존의 Sample Data, 또는 특정 중요 시스템 위주의 분석 환경에서 전사의 모든 데이터를 분석 환경으로 변화하고 있다. 이에 투자 대비 효과(ROI)를 극대화할 수 있는 Hybrid 형태의 데이터 분석 환경의 도입이 보편화되고 있는 실정이다.



전사 관점의 최적 의사결정을 위한 가공 융합, 분석

유기적인 플랫폼 활용을 위해서 운영 관점만이 아닌 데이터 활용 관점에서의 데이터 흐름의 각 구간별 영역을 전담하는 담당자가 필수적입니다. 특히 분석 경쟁력을 가지고 있는 기업들의 특징은 데이터 과학자, 데이터 엔지니어, 데이터 분석가를 구별하여 이들의 전문성을 극대화하고 있습니다.



[업무관련 전문가 그룹 구성]

- **Data Scientist** : 데이터 활용을 고려하여 Data 항목부터 모델 구성에 이르는 데이터 전반에 대한 코디네이터 역할
- **Data Engineer** : 데이터 수집부터 실제 Analyst 들이 활용할 수 있는 데이터 구성 까지 데이터 운영의 전반을 담당
- **Data Analyst** : 구축 된 분석 모델을 기반으로 주기적인 리포트 및 분석 수행을 담당

감사합니다.