



메타데이터 기반 통합 데이터품질관리

목차

I. WHY ?

II. WHAT/HOW ?

III. 메타데이터 활용 방안

IV. 메타데이터 기반 통합 데이터품질관리

왜 메타데이터관리와 데이터 품질관리가 중요한지에 대해 다음의 4가지 사안으로 살펴봅니다.

구분	필요성	시사점
<p>1 미국방성(DoD) 표준화 방법론 대동</p>	<ul style="list-style-type: none"> • 군사작전 수행시 육군, 해군, 공군이 동일한 사항을 각기 다른 용어로 불일치하게 사용함에 따라 작전 실패의 원인 • 국가 존폐의 문제, 막대한 비용 소모 	<ul style="list-style-type: none"> • 통합 전시 수행능력 확보 • 유지관리 비용 절감
<p>2 범 정부차원 표준화(참조모델) 추진</p>	<ul style="list-style-type: none"> • 국가 차원의 행정부, 국방부등에서 DB표준화 정책 진행 중 • 최근 도량형 통일화(평→ M², 돈→g(그램)) 	<ul style="list-style-type: none"> • 정부 부처간 원활한 정보 전달 및 의사소통으로 활용 • 재정비 비용 절감
<p>3 일반 기업 차세대 시스템 사업 구축</p>	<ul style="list-style-type: none"> • 운영계와 정보계 용어 불일치로 데이터 정합성 불신 • 유지관리비용 증가 	<ul style="list-style-type: none"> • 의사소통의 일치를 위한 데이터 표준화(품질) 정책, 관리조직, 시스템 필요 • 고품질의 데이터를 활용한 마케팅
<p>4 데이터품질대한 고객의 높은 수준 요구</p>	<ul style="list-style-type: none"> • 고객에게 신뢰성 있는 정보 제공 필요 • 고객만족 → 고객감동 → 고객전율에 대한 신뢰성 있는 요구가 높다 • DW, CRM, Data Mining 시 데이터 품질 	<ul style="list-style-type: none"> • 고품질의 정보 서비스 제공 • 회사 이미지 향상

주) DoD: Department of Defense(미국방성)

빠른 비즈니스 환경 변화와 IT환경의 변화는 급속한 데이터 및 애플리케이션의 변화를 초래하기 때문에 품질 저하를 가져오고 이에 따른 시스템 장애에 대한 위험요소를 내포하고 있습니다. 이러한 문제를 해결하기 위해 **효율적이고 체계적 데이터표준화 방안 및 메타시스템 구축이 필요합니다.**

비즈니스 환경의 변화

1. 산업간 융합 추세
2. 망 개방 요구 및 정부 규제 강화
3. 고객의 선택권 확대
4. 정보 품질 요구 수준의 향상

IT환경의 변화

1. 데이터 입력채널의 다양화
2. 시스템간 데이터 이동 폭증
3. 애플리케이션 규모, 복잡도 증가
4. IT유지보수 비용 증가

우리의 현실

1. 프로세스 중심의 접근
2. 문서적인 IT관리
3. 업무담당자의 단편적인 요구
4. 전문 모델러 부재
5. 데이터 품질관리 체계 부재

빠른 환경 변화 대응 및 데이터의 품질 향상을 위한 메타데이터 관리시스템 필요

메타데이터관리시스템

전사데이터 관리체계 수립

- 데이터 표준화 관리 체계 수립
- 데이터 표준화 업무 수행
- 데이터 품질 관리 체계 수립
- 데이터 품질 관리 업무 수행

메타데이터 관리시스템 구축

- 메타데이터 저장소 구축
- 표준화 및 모델 관리시스템
- 데이터 품질관리 시스템
- 데이터 영향분석 시스템

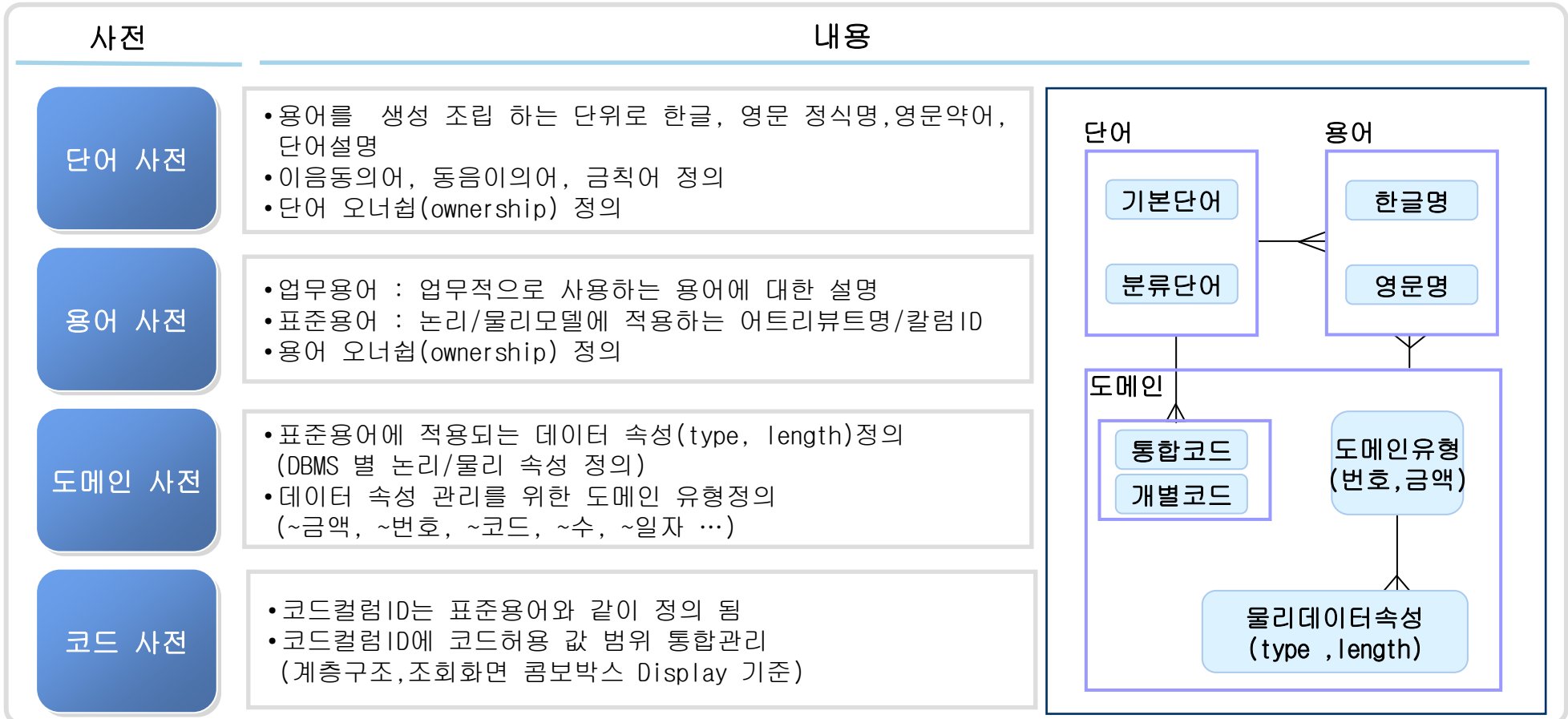
실행조직 R & R

- 현업 : 비즈니스 를 도출 지원
- 개발팀 : 표준화/모델, 개발 패턴
- 운영팀 : 데이터/소스관리, 시스템관리
- DA : 표준화 관리 및 데이터품질관리 주관자

데이터 표준화 지침은 전사차원의 데이터표준화에 대한 정책을 수립하여 회사의 방향을 직원간에 공유하고 추진하고자 하는 지침입니다. 표준에 대한 정의, 데이터 표준화 범위, 데이터 표준 용도, 데이터관리조직 및 주요 표준화 프로세스에 대한 전사적인 데이터 표준 관리 기준을 수립하고 배포하여 관리 합니다.

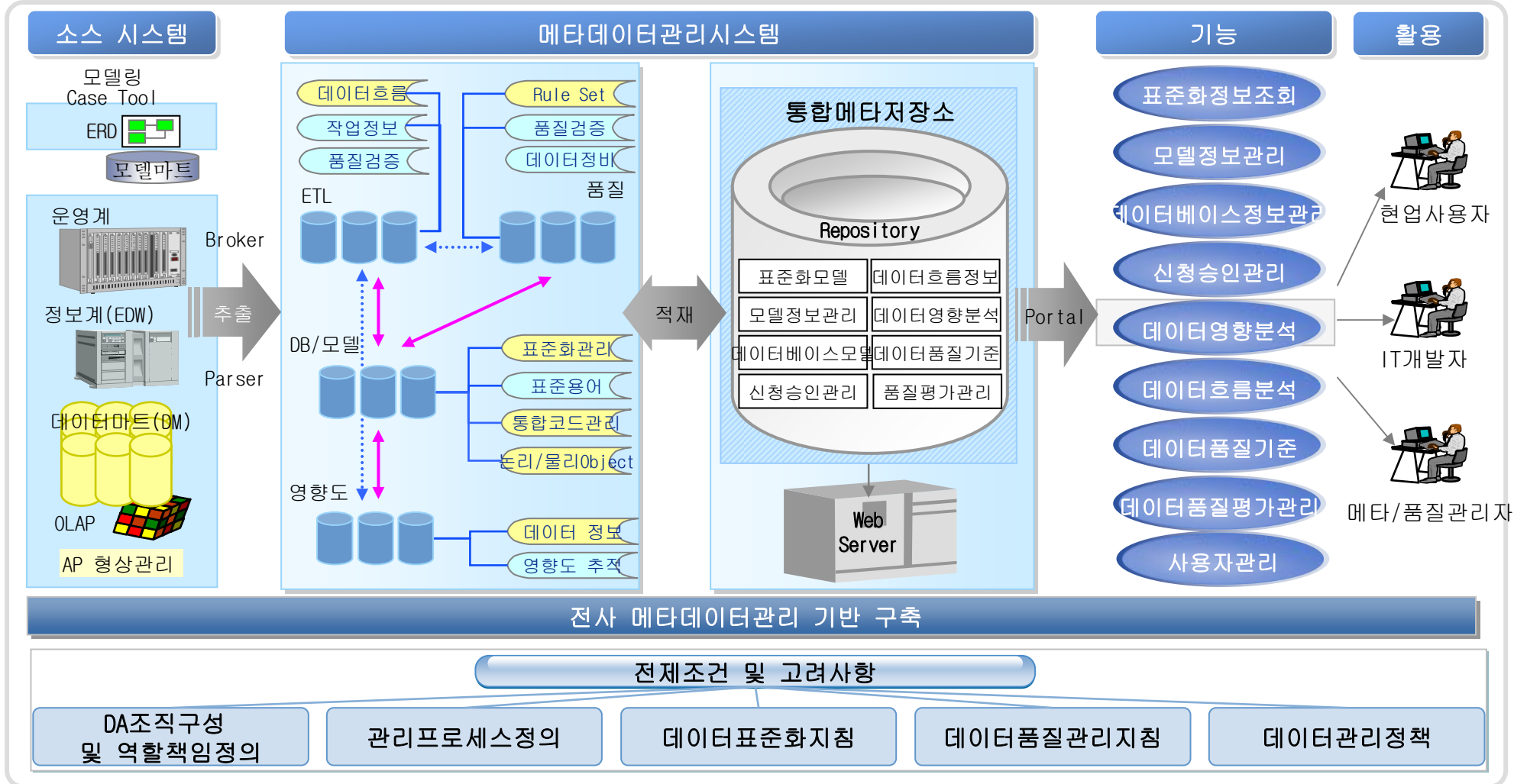
주요 지침	내용
데이터 표준의 목적 정의	<ul style="list-style-type: none"> 프로젝트 개발, 유지보수 대상 오브젝트들에 대하여 명확한 명칭부여 기준을 정함으로써 프로그램의 Readability를 향상시키고 일관성을 유지하여 개발자의 생산성 향상과 인터페이스의 개선 및 시스템의 품질 향상을 꾀하는데 그 목적을 둔다.
데이터 표준 정의	<ul style="list-style-type: none"> 전사에서 데이터를 정의하고 활용하는 데 적용하는 통일된 기준을 의미한다. 전사의 데이터표준은 메타관리시스템에 등록된 기본단어, 업무용어 및 도메인을 말한다.
데이터 표준 범위 정의	<ul style="list-style-type: none"> 명명규칙: 업무용어의 한글 및 영문 이름을 정하는 규칙. 정의규칙: 업무용어에 설명(산출기준)을 붙이는 규칙. 표현규칙: 데이터 타입과 길이를 정하는 규칙. (도메인 유형정의)
데이터 표준 용도	<ul style="list-style-type: none"> 업무적인 데이터의 이름과 속성을 정하는 기준이 된다. 논리모델에서 엔티티명, 속성(어트리뷰트)을 정의하는 기준이 된다. 물리모델에서 테이블, 컬럼을 정의하는 기준이 된다.
데이터 관리 조직 및 주요 표준화 프로세스 정의	<ul style="list-style-type: none"> 데이터 관리를 위한 조직과 조직간(DA, 모델러, IT운영, 현업, DBA)의 R&R을 명확히 한다. 주요 표준화 데이터 생성 및 적용에 대한 운영 프로세스를 수립한다.

기업에서 사용하고 있는 용어는 모든 구성원들이 유일한 의미로 공유되어야 하기 때문에 데이터 사전을 생성 구축되어야 합니다. 데이터 사전 구축 방법은 기본단어를 정의하고 단어들의 조합에 의한 표준용어를 생성하며 표준용어에 대한 오너쉽, 설명정보, 데이터 값 산출 기준 등을 정의하고 시스템화 합니다. 또한 하나의 용어에 하나의 도메인(type, length)를 정의하여 논리에서 물리까지 데이터 일치성을 확보합니다.



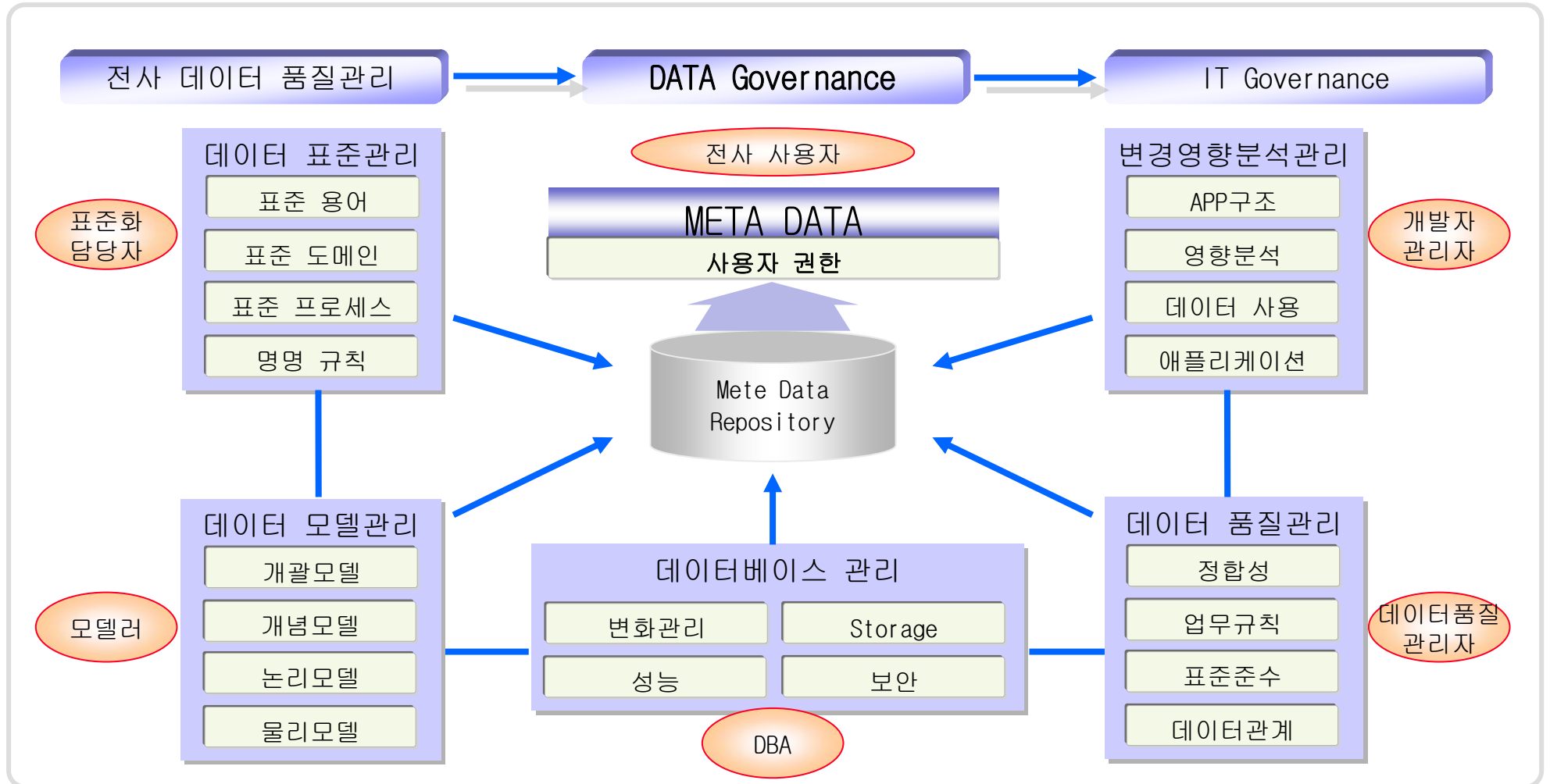
2. 메타데이터관리 2.1 메타데이터 저장소 구축 2.2 데이터 거버넌스(DATA Governance) II.WHAT/HOW

IT정보에 대한 변경에는 관련 업무 또는 응용 프로그램에 오류 발생의 위험이 따르며 또한 잘못된 변경으로 인한 데이터에 심각한 훼손 또는 데이터오류가 발생 할 수 있습니다. 이러한 위험을 해결하기 위해서는 메타데이터 정보를 활용하여 IT자산 정보를 체계적으로 통합 관리하여야 합니다.



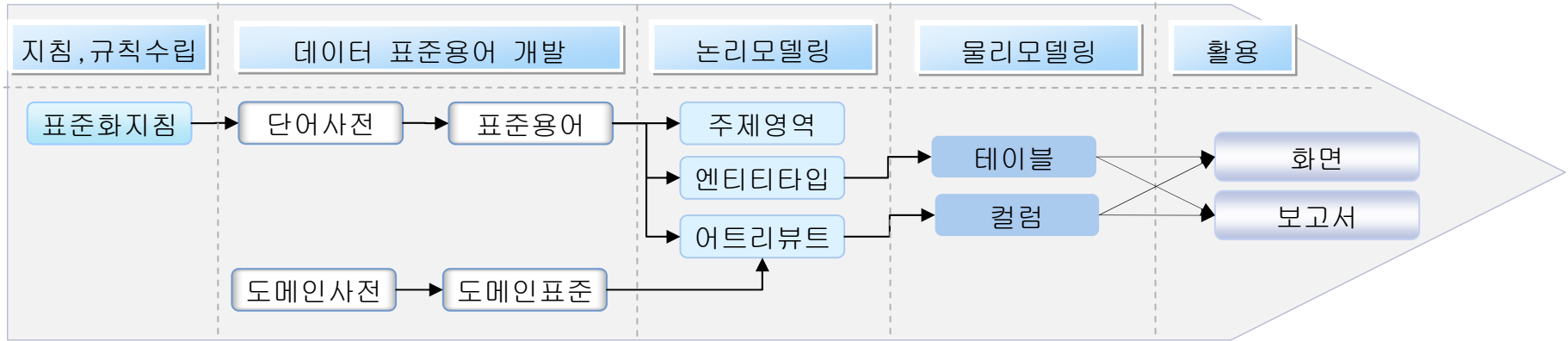
2. 메타데이터관리 2.1 메타데이터 저장소 구축 2.2 데이터 거버넌스(DATA Governance) II.WHAT/HOW

통합 메타 저장소관리를 함으로써 메타 데이터의 정합성관리 와 동일한 데이터를 연관 시스템에 연계 등이 가능해져 기업의 데이터관리에 대한 데이터 거버넌스 체계가 구축 됩니다.



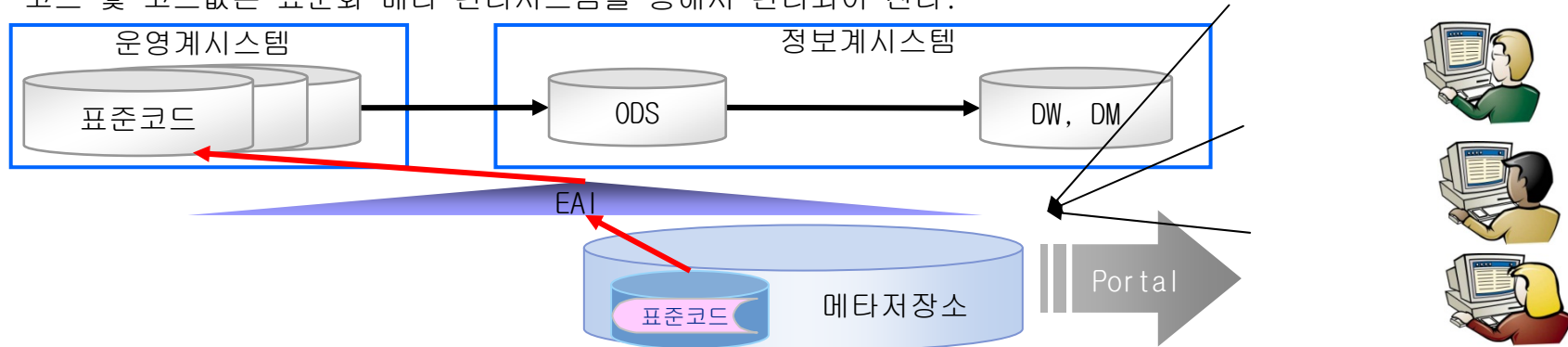
표준화 데이터사전 구축을 기반으로 ER 논리모델을 설계하고 메타정보를 활용하여 물리모델정보(칼럼ID, 칼럼속성)를 자동연동하여 제공하며 데이터베이스생성 DDL Script 정보를 제공하여 표준화와 모델 정합성 동기화를 지원합니다.

표준용어 활용

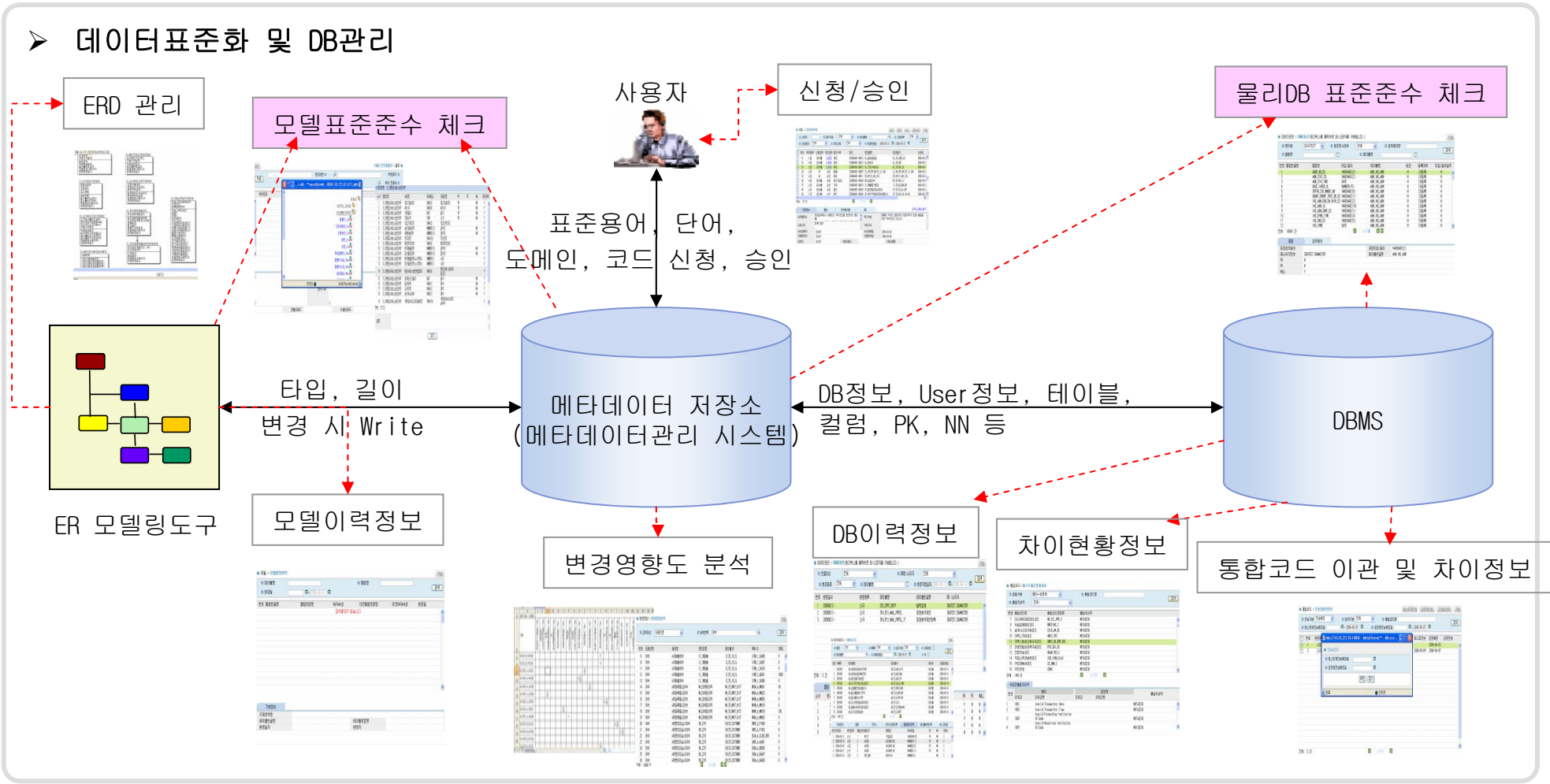


표준코드 활용

통합된 코드는 운영계와 정보계에 EAI를 통해 Real time으로 동기화하여 각 시스템에서 활용한다. 또한 추가, 변경되는 코드 및 코드값은 표준화 메타 관리시스템을 통해서 관리되어 진다.



데이터의 Life Cycle을 관리 함으로써 데이터에 대한 변경이력, 표준준수, 변경영향분석, 전송기능들을 제공합니다.



2. 데이터 정합성 및 시스템 안정성 유지

III. 메타데이터 활용 방안

관련 시스템과 연동하여 데이터베이스 및 어플리케이션의 동기화를 하여 메타/모델/데이터베이스 3위 일체 정합성 확보를 통해 시스템 안정성을 유지합니다.

▶ 데이터 정합성 및 시스템 안정성 유지

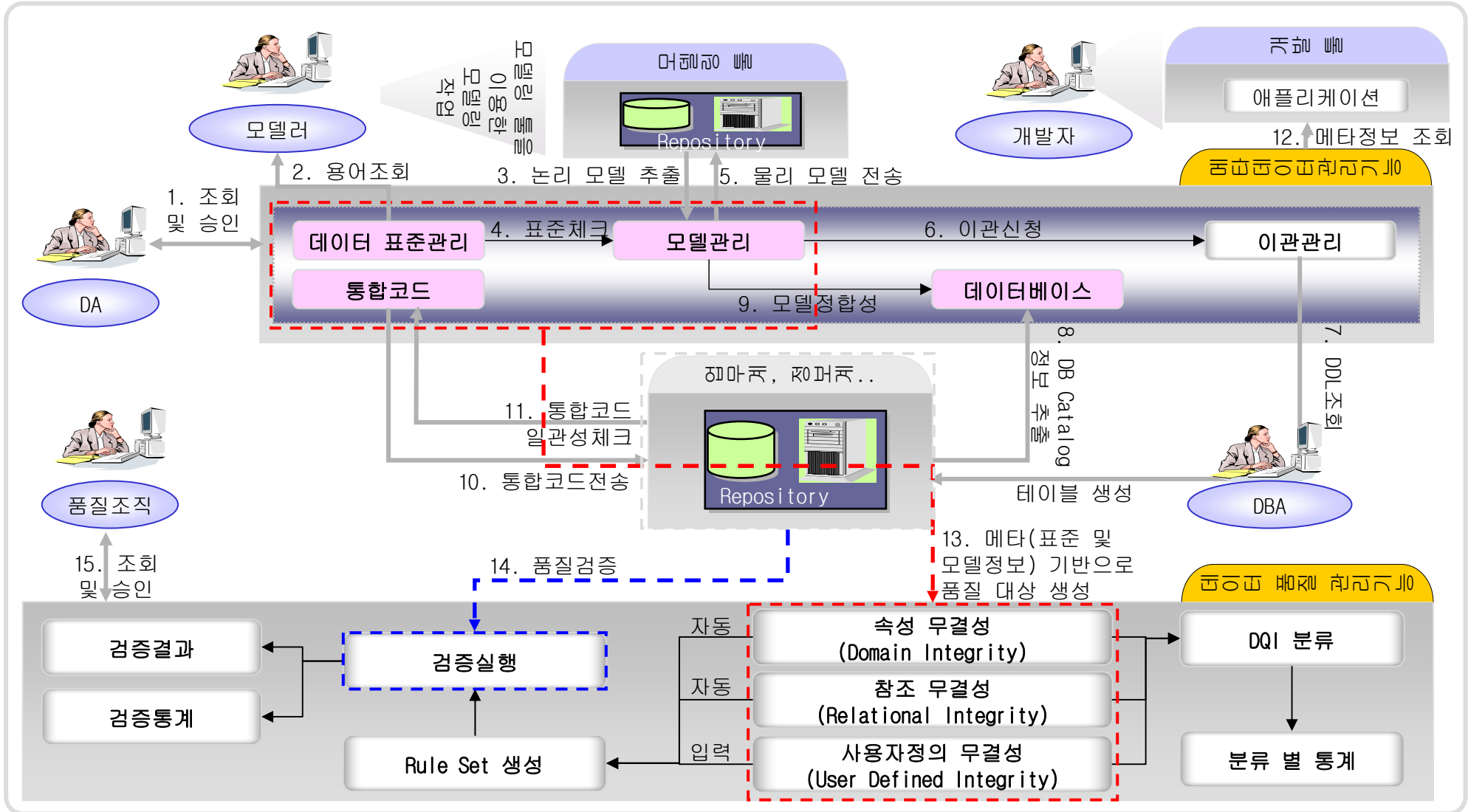
시스템운영 기능 구분	개발계 (A)	테스트계 (B)	운영계 (C)	비고
ITPMS (ITSM, ITAMS)				ITPMS 에서 SR 정의
어플리케이션 개발	2.1 어플리케이션 개발			
형상관리	3.1 프로그램 개발완료 등록			내부 status 관리
모델관리	4.1 논리모델 반영			모델 버전관리
메타관리	5.1. 메타 물리변경 신청(SR 연동) 5.2. DDL SCRIPT 5.3. 어플리케이션 영향분석 6.2 물리반영여부 등록			1. ITPMS 에서 SR을 읽어서 모델신청 시 해당업무와 연계신청 2. 물리 DB 정합성 체크 (개발계, 테스트계, 운영계)
DBA(물리DB생성)	6.1 물리변경			

주) ITPMS : IT Process Management Services

1. 메타데이터 기반의 통합 데이터품질관리

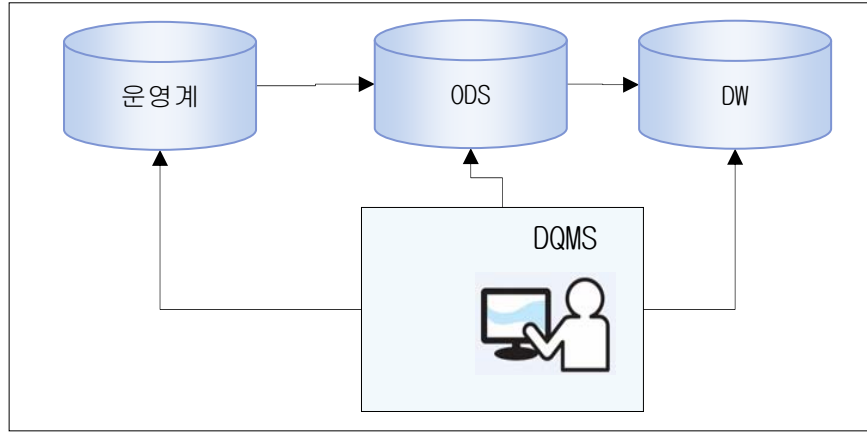
IV. 메타데이터 기반 통합 데이터품질관리

메타데이터 기반의 품질관리 프로세스를 통하여 데이터의 구조적 품질 향상과 데이터 Value의 품질 측정을 통하여 데이터 품질 향상을 기합니다.

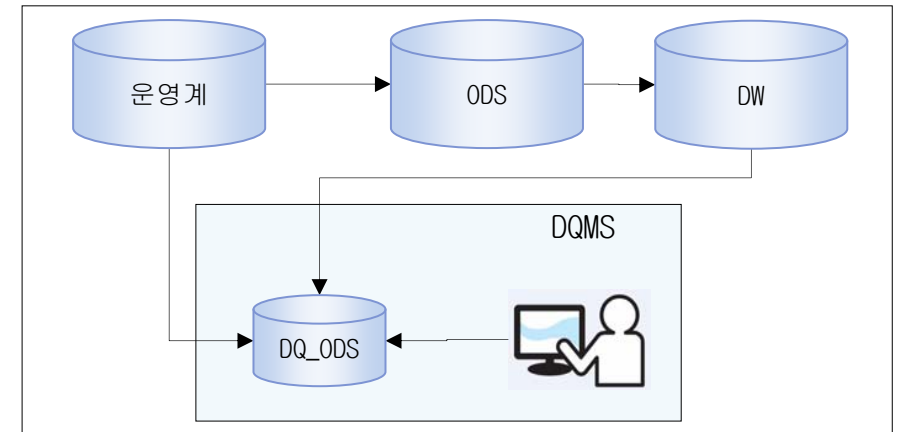


현재 데이터 품질을 검증하기 위해 구축 되는 아키텍처에 대한 사례 예시 입니다.

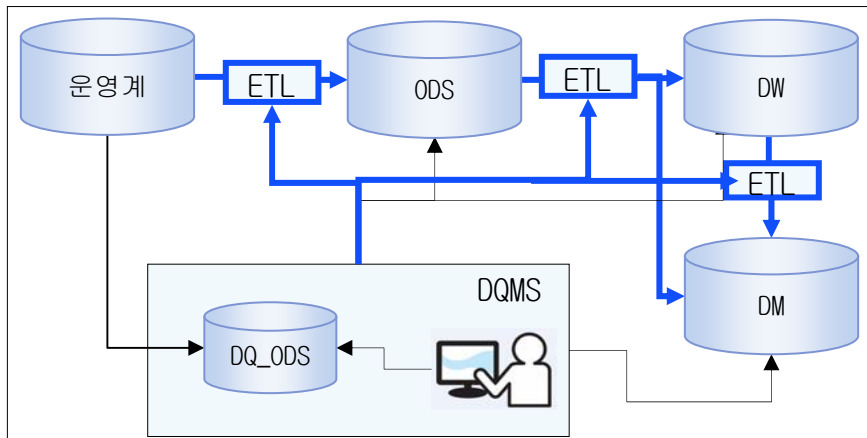
구성도(예1) SDQE



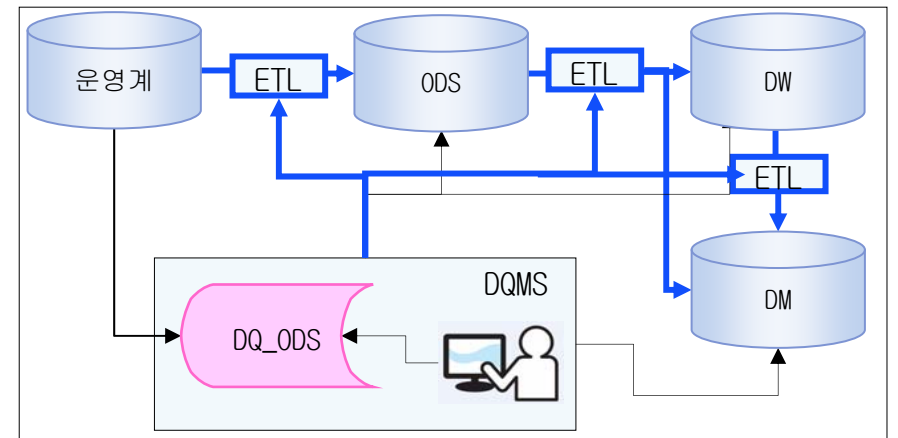
구성도 (예2) SDQE+DQ_ODS



구성도 (예3) : DDQE+SDQE+DQ_ODS



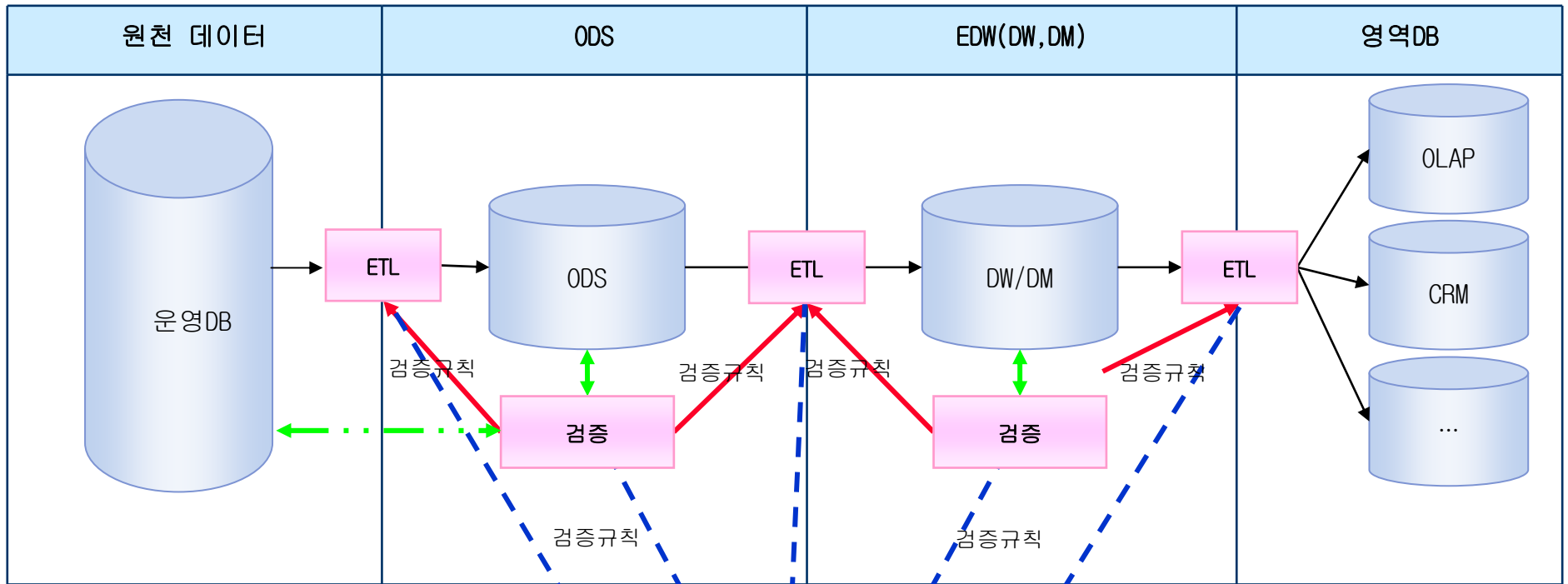
구성도 (예4) : DDQE+SDQE+DQ_ODS(SAM-F)



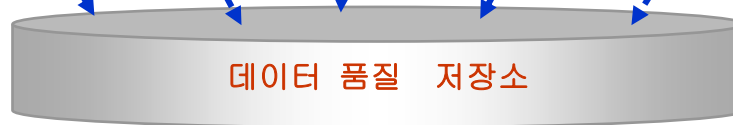
- 주) 1. 사전검증(동적 데이터 무결성 확보) : DDQE (Dynamic Data Quality Engine) : ETL 단계 적용 검증
 2. 사후검증(정적 데이터 무결성 확보) : SDQE (Static Data Quality Engine) : 적재 테이블 대상 검증
 3. 데이터품질관리 검증용 Storage : DQ_ODS , DQ_ODS(SAM-F) : 별도의 품질검증용 DB/파일 구성 검증

데이터 품질을 검증하는 방법 중 품질관리 시스템의 검증규칙을 활용 ETL 단계부터 적용한 사전 검증 프로세스입니다.

➤ 동적 데이터 무결성 검증(DDQE : Dynamic Data Quality Engine)



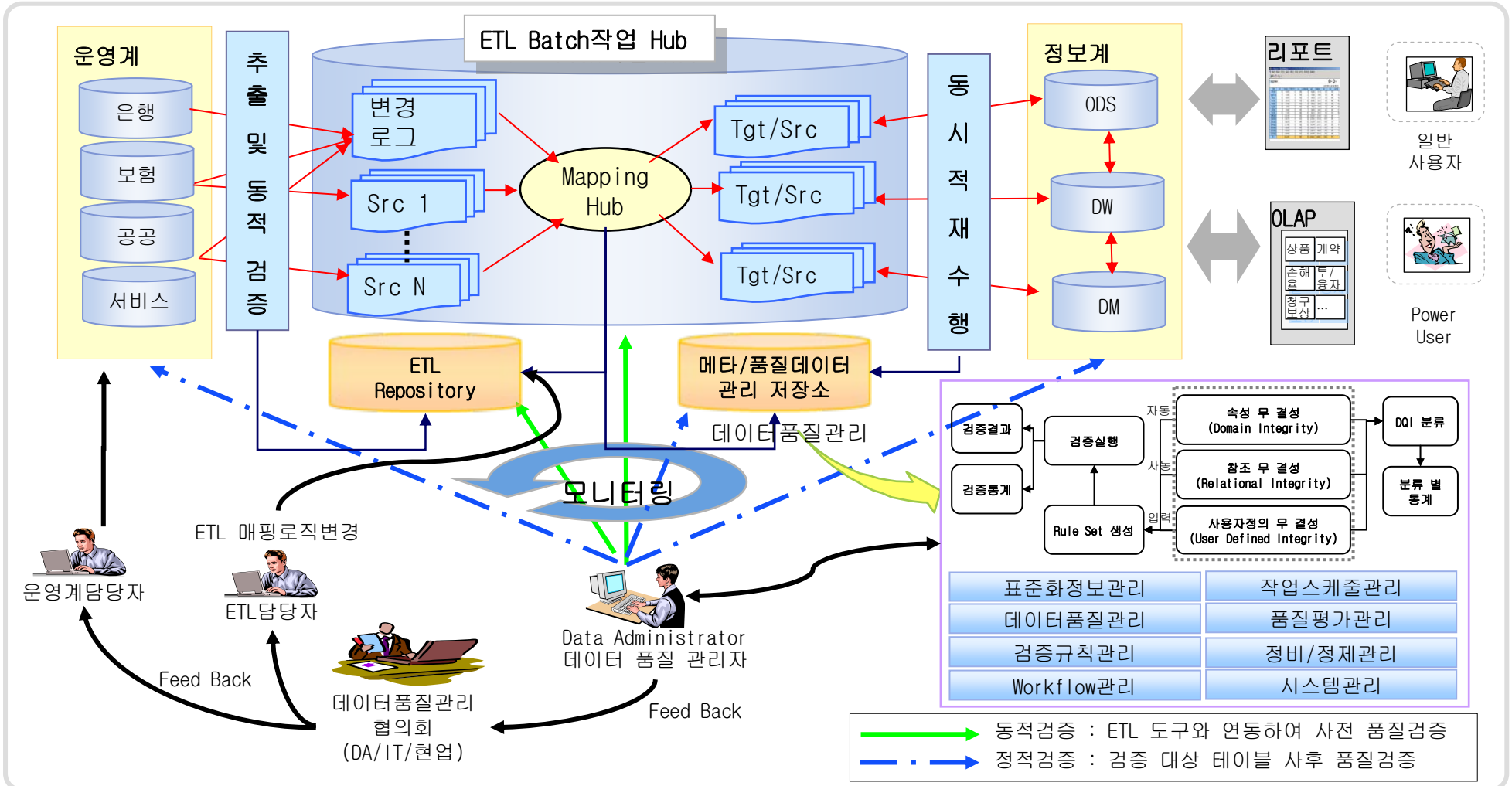
- ETL단계 동적 데이터 품질검증
- - - 목적테이블 대상 정적 데이터 품질검증
- - - 데이터 품질 결과 정보



4.상시 데이터 품질모니터링 체계

IV. 메타데이터 기반 통합 데이터품질관리

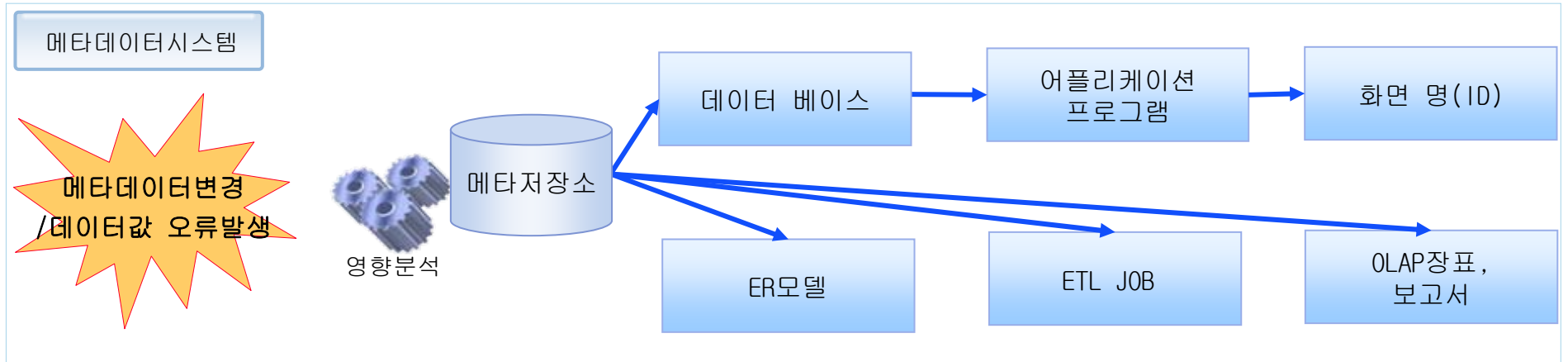
ETL Batch작업 Hub를 통하여 ETL 단계를 간소화 하고 메타데이터 시스템을 기반으로 표준 데이터 관리 및 데이터 품질관리를 사전 동적 검증 및 사후 정적 검증이 가능한 다중 검증 체계를 구축합니다. 이러한 데이터품질검증체계를 통해 데이터 이행 전 품질 모니터링에 의한 재 작업 방지 와 데이터품질검증에 대한 부하분산 기대효과가 있습니다.



5. 영향분석

IV. 메타데이터 기반 통합 데이터품질관리

영향도 분석의 목적을 2가지 측면에서 보면 첫째는 메타데이터의 변경 발생시 영향 받는 모델이나 데이터베이스, ETL JOB 등을 추적하여 변경에 대한 동기화를 하고자 함이고 두 번째는 데이터 값의 오류가 발생시 관련 소스를 추적해서 단순 결과에 재한 데이터클린징이 아닌 원인을 발생시키는 어플리케이션 소스를 추적하여 정비함으로써 데이터 품질을 확보하고자 함입니다.



유형	분석 부문	내용
데이터 영향분석	모델영향분석	영향 받는 엔티티목록(주제영역, 업무구분, 엔티티, 에트리뷰트)
	물리 DB 영향분석	시스템구분, 오너, 업무구분, 테이블ID, 칼럼ID
	ETL JOB 영향분석	ETL JOB 목록, 선 후행 관계, 작업실행결과
	OLAP 장표	장표 위치, 장표명
	REPORT(보고서)	보고서 위치, 보고서 명
어플리케이션 프로그램 영향분석(DBIO, 화면ID)	Shell 프로그램	Shell 프로그램 명, 프로그램위치, 프로그램 구조
	어플리케이션 프로그램영향분석(DBIO, 화면ID)	어플리케이션 call 구조, CRUD 매트릭스, 관련 SQL 소스, 화면 명, 화면ID

5. 영향분석 (예시)

IV. 메타데이터 기반 통합 데이터품질관리

데이터 흐름정보 검색

소스/타겟 테이블목록조회

소스/타겟 조건선택

테이블 영향

모델 영향

OLAP/보고서 영향

ETL JOB 영향

CRUD 매트릭스

업무구분	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
구분	DBC TABLES	DMWACQAMBROJL STDA	DMWACQAMBROJL STLIS	DMWACQSPDTCMD JSTTEMP03	DMWACQSPDTCMD JSTTEMP04	DMWACQSPDTCMD MASTTEMP02	DMWACQSPDTCMD MASTTEMP03	DMWACQSPDTCMD MASTTEMP04	DMWACQSPDTCMD MASTTEMP05	DMWACQSPDTCMD MASTTEMP06	DMWACQSPDTCMD MASTTEMP07	DMWACQSPDTCMD MASTTEMP08	DMWACQSPDTCMD MASTTEMP09	DMWACQSPDTCMD MASTTEMP10	DMWACQSPDTCMD MASTTEMP11	DMWACQSPDTCMD MASTTEMP12	DMWACQSPDTCMD MASTTEMP13	DMWACQSPDTCMD MASTTEMP14
02ftp_core_front																		
lod_tmp_card_siin																		
bdwh_big_daejun_dmts																		
bdwh_dwm_card_cusinfo																		
lod_tmp_card_siin_o_lg																		
bdwh_dwm_cuset_ctr_rst																		
bdwh_dwm_cum_haddr_ddtt																		
bdwh_dwm_ysac_d_rst_ddtt																		
bdwh_dwm_accm_md_rst_ddtt																		
bdwh_dwm_accm_rnw_rst_ddtt																		
bdwh_dwm_card_cusinfo_lg																		

SQL구분 : brkma0040 SQL구분 : 1번째

```
SELECT TableName FROM DBC.Tables WHERE DatabaseName = 'SFB_WORK' AND TableName = 'RKM_ALM_DAY_BS';
```

관련 SQL Script

프로그램간 선 후행 흐름정보 TREE 구조

호출 프로그램
[open all](#) | [close all](#)

- 호출계층
 - isis.mct.basc.creditHistory.dbio.MCTCRD01MTV0013
 - isis.mct.basc.creditHistory.ec.CreditHistoryImp1
 - isis.mct.basc.creditHistory.pc.CreditHistoryProcessImp1
 - isis.mct.basc.creditHistory.service.CreditHistoryService
 - 신용정보종합검색 [MCTBASCO061M0]
 - TreeView [MCTBASCO067M0]
 - 고객신용정보관리 일괄처리 [MCTBASCO063M0]
 - 고객신용정보관리 [MCTBASCO062M0]
 - 공동소유자 등록 [MUWMORT0003P0]

화면명 (화면 ID)

프로그램 구조도(PL/SQL)

SQL Name: PLSQL_PACKAGE

Creation Date: 2007-08-16 12:04:07 PM

Owner: PLSQL

Description: PLSQL_PACKAGE

감사합니다