

EA를 위한 데이터 아키텍처 구축 모델

김석수*, 이화식**

A Model of implementation Data Architecture for Enterprise Architecture

Seok-Soo Kim*, Hwa-Sik Lee**

요약

데이터는 IT의 핵심요소이다. 다른 아키텍처는 선진 기술과 기법을 참조하고 도입하여 적용이 가능하지만 데이터 아키텍처는 고유한 것이어서 우리 스스로 구축을 하여야 한다. 데이터는 기술의 변화와 진화에 민감하지 않은 영역으로 처음 구축 시 잘 만들면 건물의 철골 구조물과 같이 건설한 정보 시스템을 구축하는데 좋은 방향을 제시할 수 있다. 잘 구축된 데이터 아키텍처는 엔터프라이즈 아키텍처 구축을 용이하게 하고, 구축 후 관리 및 운영을 효과적으로 할 수 있게 한다. 본 논문은 엔터프라이즈 아키텍처를 위한 데이터 아키텍처 구축 모델을 제시한다.

▶ keyword : 데이터 아키텍처, 엔터프라이즈 아키텍처, 데이터 모델

Abstract

Data is the kernel of IT. We can apply to other architectures are refer to technology and advanced method. Data architecture is our's native that must be implement ourself. It will suggest to the best method to implement for a sound information system that likes to framework of building because data is not sensitive field of evolution and change of technology. Well implemented data architecture can be effectively management and operation. Also this is easily to implement of enterprise architecture. This paper propose a model of implementation data architecture for enterprise architecture.

▶ keyword : data architecture, enterprise architecture, data model

• 제1저자 : 김석수 • 교신저자 : 이화식
• 투고일 : 2011. 02. 10, 심사일 : 2011. 03. 17, 게재확정일 : 2011. 08. 26.
* 경원대학교 컴퓨터공학과(Dept. of Computer Science, Kyungwon University)
** (주) 엔코어 대표이사(en-core)
※ 본 연구는 경원대학교 2009년 교내연구비 지원으로 이루어졌음.

I. 서론

데이터는 IT의 핵심요소이다. 다른 아키텍처는 선진 기술과 기법을 참조하고 도입하여 적용이 가능하지만 데이터는 고유한 것이어서 우리 스스로 구축을 하여야 한다. 데이터는 기술의 변화와 진화에 민감하지 않은 영역으로 처음 구축 시 잘 만들면 건물의 철골 구조물과 같이 건실한 정보 시스템을 구축하는데 좋은 방향을 제시할 수 있다. 데이터가 건물의 철골 구조물 같은 역할을 할 수 있도록 하는 것이 데이터 아키텍처(이하: DA)이다. 따라서 DA는 IT의 정보화를 위한 핵심영역이다. 오늘날 정보시스템의 양적 팽창은 정보공유, 비용 등에서 정보시스템의 통합이라는 새로운 이슈를 발생시키고 있다. 정보시스템의 통합은 이제까지 진행되어 온 분산 모델의 진화를 의미하며 그 내용은 인프라, 운영, 조직의 측면으로 볼 수 있다.

최근에 비즈니스 환경은 세계적인 불황과 금융위기에서 조직이 생존하기 위하여 시장 변화에 신속히 대응하는 적응력을 요구하고 있다. 이에 따라 전사적인 차원에서 보유하고 있는 모든 데이터의 관리에 대한 정책, 지침, 표준, 전략 및 방향을 수립하고[1] 데이터를 관리할 수 있는 조직 및 프로세스를 구축하는 데이터 거버넌스 체계 도입의 필요성과 중요성이 높아지고 있다. 데이터 거버넌스가 성공적으로 정착되고 원칙이 준수되기 위해서는 데이터 아키텍처가 구축되어야 한다[2]. 데이터 아키텍처는 EA의 중요한 하부구조로 데이터 측면에서 기업 시스템을 처음부터 끝까지 조망하여 시스템의 데이터를 체계적, 구조적으로 관리하고 설계하는 전 과정을 말한다. 데이터 아키텍처가 중요한 이유는 데이터가 시스템의 근본을 구성하므로 애플리케이션이나 다른 기술적인 요소보다도 먼저 정의되어야 하기 때문이다[3].

많은 프로그래머들은 기업이 각자 고유한 업무적 특성을 가지고 있기 때문에 동일한 업종이라 하더라도 같은 데이터 구조를 가질 수 없다고 믿고 있다. 특히 프로세스를 설계하고 개발하는 사람들은 자신이 사용할 데이터를 자의적으로 설계해 왔기 때문에 데이터 구조는 더욱 체계적일 수가 없었다. 현재 데이터 부문에서 정보 시스템의 실상을 살펴보면 프로그램 개발자 중심의 분석 및 설계가 비즈니스 프로세스 중심으로 이루어짐으로 많은 문제점이 노출되고 있다[4]. Enterprise Architecture(이하: EA)는 조직이 보유하고 활용하고 있는 정보기술을 체계적으로 분석하고 나아가 정보기술의 발전 방향을 제시하여 조직의 전략적 목표를 달성할 수

있는 정보기술의 청사진을 제시한다. 하지만 국내 연구는 사례 연구나 이론적 연구가 부족하다[5]. 따라서 이러한 문제를 개선이나 해결하기 위해서는 전체 과정을 통제, 관리할 수 있는 모델이 필요하다[6].

본 논문에서는 도구를 사용하여 DA를 구축하는 모델을 제안하여 EA를 위한 DA 구축의 자동화 방안을 제시하고자 한다.

II. 이론적 배경 및 고찰

1. DA 프레임워크

DA는 기업의 모든 비즈니스를 데이터 측면에서 처음부터 끝까지 조망해 보려는 것이다. DA는 데이터에 관한 모든 계층을 총망라해서 객관적이고 구체적인 접근 방법을 명시한 체계적인 방법론이다. 업무 프로세스를 고려하지 않고 순수한 데이터만을 체계적으로 분석하는 것은 많은 문제들을 해결할 수 있게 한다. 그림 1은 데이터 아키텍처 프레임워크를 보여준다.

2. DA의 구성요소

정보 시스템에서 데이터는 반드시 존재 하는 것이다. 비즈니스의 결과물이 데이터이다. 데이터 아키텍처는 다음과 같은 내용으로 구성된다[8].

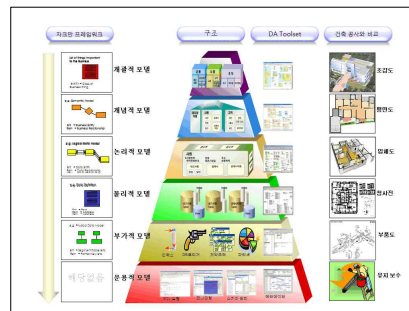


그림 1. 데이터 아키텍처 프레임워크
Fig. 1. Data Architecture Framework

1) 개괄적(contextual) 모델

개괄적 모델은 전사적 차원에서 관리되어야 한다. 이 단계에서 정의되는 내용은 데이터 영역과 데이터 클래스로 구

성된다. 이 내용은 건축물의 조감도에 해당하는 것으로 데이터 아키텍처의 최상위 집합을 정의한다. 데이터 영역이나 데이터 클래스가 상세화 되면 엔티티가 되는 것처럼 상세화가 진행되면 속성이 될 속성의 상위 개념인 아티클(article)을 정의한다. 아티클은 속성들을 하나 이상 보유한 일종의 속성그룹이다. 현행 시스템을 파악하기 위해서 행하는 리버스 모델링 시에는 기존 엔티티 들을 추상화하여 구성한다.

2) 개념적(conceptual) 모델

개념적 모델은 데이터 모델의 골격에 해당한다. 이것은 주요 키 엔티티와 핵심 행위 엔티티들로 구성된다. 이 엔티티는 행위(업무)를 관리하는 엔티티의 의미상의 주어이다. 여기에서는 세부적인 상세 부분집합을 정의하여 명확하게 한 후, 데이터 영역과 데이터 클래스의 엔티티들을 구체적으로 연결(alignment)한다. 개념적 모델에서는 데이터 모델링에서 사용하는 엔티티, 서브타입, 릴레이션십, 속성들을 표현한다.

3) 논리적(logical) 모델

논리적 모델은 데이터 모델링이 최종적으로 완료된 상태이다. 여기에서는 모든 논리적(의미적)인 데이터 객체들을 도출한다. 필요할 때에는 보다 다양한 측면에서 집합을 정의하여 다차원 서브타입으로 구성한다. 논리적 모델은 최종 식별자 형태가 확정된 데이터 모델로 정규화 및 현실적 요소를 감안한 반정규화를 실시한 모델이다. 또한 데이터 모델 상세화를 거쳐서 통합화를 완료한 데이터 모델이고, 비용과 보존에 대한 전략적인 이력관리 방안까지 확정된 모델이다. 현행 시스템을 파악하기 위해서 행하는 리버스 모델링에서는 물리적인 모델을 논리화시켜서 생성한다.

4) 물리적(physical) 모델

물리적 모델은 논리적 모델을 특정 데이터베이스로 설계함으로써 생성된 데이터를 저장할 수 있는 물리적인 스키마이다. 물리적인 개체가 되기 위해서는 반드시 어떤 형태로든 논리적 모델에 정의되어 있어야 한다. 물리적 모델은 논리적 모델과 독립적으로 구성된다. 즉, 하나의 논리 모델에서 N개의 물리 모델이 만들어질 수 있다. 이곳에 데이터 칼럼, 각종 제약 요건을 정의하고, 인덱스, 뷰, 파티션을 정의한다. 또, 체계적인 칼럼 관리를 위해 도메인을 관리한다. 물리적 모델은 메타 데이터 관리와 연계된다[9].

3. DA 성공요인

데이터 아키텍처를 구축하는데 핵심 성공 요인은 다음과

같다. 첫째, 리버스를 통한 전체 단계의 데이터 현행화 구축이다. 문서화가 되어 있는 자료와 기존의 ERD, 자료사전(Data Dictionary), 메타 리파지토리에서 데이터를 추출하여 리파지토리에 저장한다. 둘째, 논리화를 통한 진정한 논리모델을 보유하는 것이다. X-RAY는 단지 뼈의 형태를 보여줄 수 있지만 MRI는 세밀한 부분을 보여줄 수 있듯이 기존의 ERD는 2차원적인 표현으로 평면도의 역할만 할 수 있었다. 이것을 입체도의 역할을 할 수 있는 3차원의 ERD로 작성하여 현재의 물리적인 구조를 의미 중심의 논리적 형태로 재구성한다. 셋째, 개괄모델, 개념모델, 논리모델, 물리모델의 각 단계 간 연결을 확실히 유지하는 것이다. 개괄적 모델의 데이터 영역(클래스)과 개념적/논리적 모델과의 연결, 개괄적 모델의 인터랙션과 개념/논리적 모델과의 연결, 논리적 데이터 모델과 물리적 모델의 테이블 간의 연결, 논리적 모델의 가상의 엔티티와 뷰 간의 연결, 아티클, 애트리뷰트와 오브젝트의 타입간의 연결, 물리적 모델에 관련된 세부 부가적 정의 단계의 연결, 물리적 단계와 운용적 단계 간의 연결들이 정립되어야 한다. 넷째, 메타 데이터 리파지토리에 모든 정보가 저장되어 공유되어야 한다. 차세대 시스템을 구축하는 경우 데이터 거버넌스 유지를 위해서 메타데이터 관리를 도입, 운영한다. 이를 통해 메타데이터에서 관리하는 자원을 연계, 활용하여 데이터 품질관리를 통합적으로 운영할 수 있다. 다섯째, 데이터 모델에 모든 데이터 규칙이 포함되도록 상세화한다. 상세화 단계에서는 정규화 작업과 M:M 릴레이션십의 해소를 통해 최대한 세부적인 엔티티들을 도출하고, 속성들을 그 속에 채우게 된다. 엔티티의 통합에는 계층구조로 되어 있는 엔티티들을 통합하는 수직적 통합과 유사한 내용을 가진 주변 엔티티들을 하나로 모으는 수평적 통합이 있다. 여섯째, 데이터 거버넌스를 통하여 통제가 이루어져야 한다. 데이터 거버넌스가 성공적으로 정착되고 원칙이 준수되기 위해서는 데이터 아키텍처가 구축되어야 한다.

III. 데이터 아키텍처 구축 모델

본 장은 (주)엔코아의 DA# 도구[11]를 사용하여 DA를 구축하는 과정을 프로젝트 사례를 통해 연구하였다. 표준 관리 방안, 구조 관리 방안, 이동 관리 방안, 모델 관리 방안, 품질 관리 방안을 제시하기 위한 기반을 마련하기 위해 시스템 구축과 운용 단계에서의 데이터 관리 영역에 한정하여 연구를 진행하였다.

1. 데이터 아키텍처 구축 모델

시스템 구축과 운용 단계에서의 데이터 관리 영역은 리버스 데이터 모델링, 데이터 표준화, 현행 데이터 아키텍처 구축, 전사 데이터 표준과 데이터 참조 모델을 활용한 목표 데이터 아키텍처 구축, 비즈니스 규칙과 품질 체크 모듈 생성, 오류 데이터 추출과 데이터 클린징을 통한 데이터 이행의 영역으로 구성된다. 그림 2는 시스템 구축과 운용 단계에서의 데이터 아키텍처 구축 모델을 보여준다.

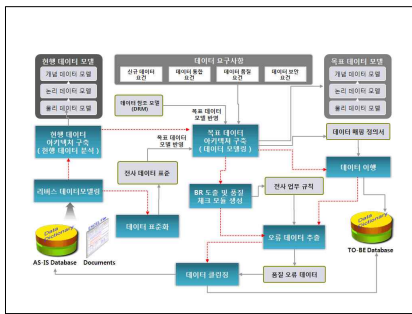


그림 2 데이터 아키텍처 구축 모델
Fig. 2 Model for implement of Data Architecture

2. 리버스(Reverse) 데이터 모델링

리버스의 필요성은 As-Is 시스템을 빠르고, 정확하게 저비용으로 분석하기 위한 것이다. 현재의 시스템은 인력의 이직과 신규 인력의 추가 등으로 업무의 요구사항을 신속하고 정확하게 수용하기가 어렵다. 또한 신 시스템의 추가로 점점 더 복잡한 시스템으로 확장되고 있다. 이러한 상황에서 현재 시스템 분석을 위해 사용자로부터 정보를 얻기 위한 시도로 리버스가 요구된다. 데이터 모델 리버스에서 수행해야 하는 주요 내용은 다음과 같다. 기존 ERD 수집, 시스템 운용정보 분석, 각종 문서 수집, 표준화 자료 구축, Data Dictionary 분석, 관련된 내용을 병합하여 리파지토리에 저장, 형태소 분석, 칼럼 자동 번역 수행 등으로 자동 생성된 1차적인 현행 물리모델을 생성한다. 그리고, 정보 수집을 통하여 획득한 데이터 모델의 관련 정보를 리파지토리에 저장하여 취합하고, 표준화 자료를 이용하여 칼럼 번역 후 자동으로 현행 물리모델을 생성한다. 그림 3은 리버스 모델에 자동관계 찾기 작업을 실시한 초기화면으로 무수히 많은 관계가 연결되어 있다. 이것은 대부분의 속성이 관계로 표현되어 있는 것을 보여준다.

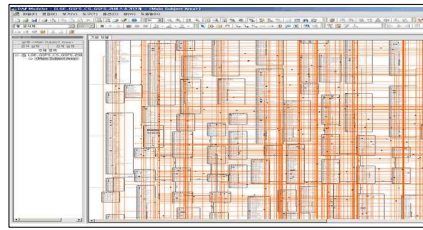


그림 3 리버스 후 생성된 현행 물리모델
Fig. 3. As-Is Physical Model after Reverse

3. 데이터 표준화

데이터가 기업의 전략적 의사결정을 위한 핵심 요소이기 때문에 데이터 통합, 데이터 품질을 위해서는 전사적인 데이터 표준화 과정이 필요하다. 그 내용은 첫째, 데이터 표준화, 규격화를 위한 기본 방침이 설정되어야 한다. 둘째, 전사적인 정보공유를 위해 유지되어야할 공통 데이터 요소를 도출한다. 셋째, 전사적인 데이터 요소 등록 및 관리 체계를 구축한다. 넷째, 정보 시스템 개발 및 유지 보수 시 승인된 데이터 요소를 활용함으로써 시스템 개발의 효율성 및 데이터 고유성을 향상시킨다.

전사적인 데이터 표준화 활동이 수행되면 현업 사용자는 정확한 데이터를 사용할 수 있고, 올바른 의사결정을 내릴 수 있다. 이는 기업의 경쟁력 확보에 많은 영향을 미친다. 데이터 표준화의 기대 효과는 다음과 같다. 첫째, 명칭의 통일로 인한 명확한 의사소통이 이루어진다. 둘째, 필요한 데이터의 소재 파악에 소요되는 시간 및 노력이 감소된다. 셋째, 일관된 데이터 형식 및 규칙의 적용으로 인한 데이터 품질이 향상된다. 넷째, 정보시스템 간 데이터 인터페이스 시 데이터 변환이나 정제 비용이 감소된다. 그림 4는 도구를 사용한 데이터 표준 관리를 보여준다.

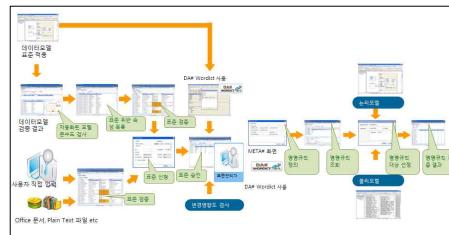


그림 4 도구를 사용한 데이터 표준 관리
Fig. 4. Data standard management using tools

4. 현행 데이터 아키텍처

데이터 모델 현행화에서 수행해야 하는 주요 내용은 다음과 같다. 엔티티 자동 배치, 기존식별자 확인 및 보완, 동일

식별자 정리, 자동 관계 찾기 수행, 필요 시 수동 관계 찾기, 관계선 정리 및 이상 확인, 엔티티 위치 및 크기 조정, 필요 시 주제영역 조정 등이다. 데이터 모델 상세화에서 수행해야 하는 주요 내용은 다음과 같다. 엔티티 명칭 보완, 의미상 주어 식별, 상세한 서브타입 표현, 관계의 명확화, 속성의 분류/그룹핑, 속성 오너십 정의, 엔티티의 구체적 정의, 중요하지 않은 속성 감추기(캡슐화), 파 생성성의 원천 값을 찾아 연결하고, 속성의 특수성을 정의한다. 데이터 모델 논리화에서 수행해야 하는 주요 내용은 다음과 같다. 상세화 모델을 활용하여 논리적 의미 파악, 논리적 집합의 통합, 통합 집합에 공통속성과 공통관계 도출, 다 계층, 다차원의 논리적 서브타입 지정, 논리적 속성의 도출, 논리적 관계 추가, 엔티티 내부의 상세한 관계를 표현한다.

데이터 모델 개념화에서 수행해야 하는 주요 내용은 다음과 같다. 논리모델로부터 도출하고, 각 오브젝트의 레벨은 사전에 정의한다. 적절한 레벨만 선택하여 자동으로 개념모델 생성하며, 논리모델과 동일 리파지토리를 사용하지만 형태는 자유롭게 지정한다. 필요 시 여러 계층의 개념모델이 생성 가능하다. 논리모델과 자동적인 연결을 유지한다. 그림 5는 논리 모델과 동일한 리파지토리를 사용하여 현행 논리화의 완성 이후 키 엔티티, 메인 엔티티, 주요 관계, 핵심 속성 등의 핵심만으로 이루어진 현행 개념 모델이다.

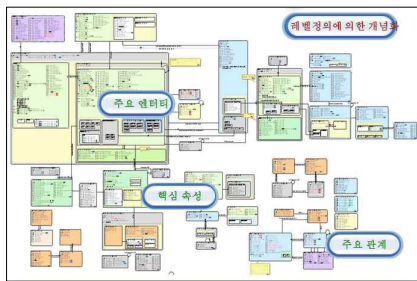


그림 5. 현행 개념 모델
Fig. 5. As-Is Conceptual Model

5. 목표 데이터 아키텍처

목표 개괄 데이터 모델에서 수행해야 하는 주요 내용은 다음과 같다. 전자적 입장에서 데이터 주제영역 도출, 최상위 주제영역 도출, 하위 상세 영역 정의, 개념/논리 데이터모델과 연결 유지, 선진 모델의 참조 등이다. 이것은 엔터프라이즈 데이터 모델을 참조하여 구축한다. 엔터프라이즈 데이터 모델은 “비즈니스에 대한 종합적인 이해를 바탕으로 기업 내에 존재하는 데이터에 대하여 프로세스와는 독립적으로 인식하여 이를 알기 쉽고 체계적으로 표현한 기업의 조감도적 정보요구”로 정의될 수 있으며 주제영역도 작성과 실체관

계도 작성의 2단계로 모델링 작업이 진행된다.

엔터프라이즈 데이터 모델의 주제영역은 전사관계자 (Involved Party) 영역, 상품/서비스(Product or Service)영역, 계정(Account)영역, 약정(Agreement) 영역, 가격정책(Price Policy, Price Plan) 영역, 청구/입금(Charge & Payment)영역, 미납관리(Collection) 영역, 통합자원(Resource) 영역, 오더관리(Order Management) 영역, 인벤토리(Inventory) 영역, 업무처리(Business Action) 영역으로 구성된다. 그림 6은 엔터프라이즈 데이터 모델의 주제영역도이다.

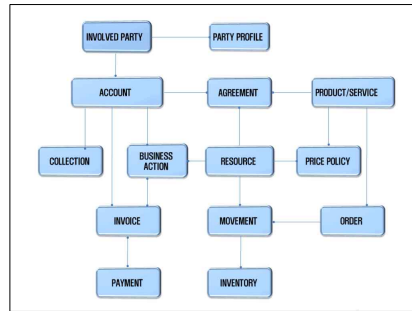


그림 6. 엔터프라이즈 데이터 모델의 주제영역도
Fig. 6. Subject area diagram of Enterprise Data Model

개념 데이터 모델은 데이터 주제영역에 근거하여 개념(상위) 수준에서 핵심 정보 실체와 이들 간의 관계로 체계화하여 표준화된 언어로 모형화한 것이다. 개념 데이터 모델의 목적은 향후 논리모델로 세분화하기 위한 골격을 수립하여 전체 관점에서 데이터 구조를 볼 수 있는 입체적 도면을 관리하고, 프로젝트 기획 및 수행 시 데이터 영향도를 민첩하게 분석하고 구조적 방향성을 제시하는 것이다.

개념 데이터 모델의 역할은 비즈니스 관점과 IT 관점에서의 역할이 있다. 비즈니스 사용자 관점에서는 첫째, 전자적인 비즈니스와 정보를 시각적이고 구조적으로 표현하는 것이다. 둘째, 사용자의 정보요구에 대한 종합적인 이해를 촉진한다. 셋째, 업무를 반영하고 업무 간의 관계를 데이터 관점에서 설정한다. IT 관점에서는 첫째, 설계자 및 사용자 간의 효과적인 의사전달 수단이다. 둘째, 데이터의 저장 및 통제를 위한 청사진의 역할을 한다.

목표 개념 데이터 모델에서 수행해야 하는 주요 내용은 다음과 같다. 현행 개념모델을 기반으로 현재의 문제점을 개선하여 과감한 통합을 한다. 추가 비즈니스 요건을 반영하고, 선진 개념모델 참조하여 혁신적이고 미래지향적인 전자 개념모델을 완성한다.

논리 데이터 모델은 비즈니스에서 관리해야 하는 대상을 논리적 구조로 완전하게 표현한 데이터 모델로 설명 능력을

갖춘 의미 모델(Semantic Model)로서 기업의 모습을 가시화시키는 역할을 수행하고, 전사 데이터 관리의 기준 역할을 수행한다. 논리 데이터 모델의 특징은 다음과 같다. 첫째, 개념 데이터 모델과 함께 데이터 요건에 대하여 IT 실무자와 비즈니스 관계자 간의 의사소통 및 의사 결정을 지원한다. 둘째, 구현 전 단계의 완성된 데이터 모델을 확보하기 위한 것으로 논리 데이터 모델을 통해 목표 비즈니스의 수행 목적과 역할에 필요한 모든 데이터 요소를 파악하고 관리한다. 셋째, 논리 모델을 구성하는 모든 요소(엔티티, 속성, 관계)는 정규화 하여 표현한다. 넷째, 모든 논리적인 데이터 객체들을 도출하고, 최종 식별자가 확정된 모델이다. 다섯째, 정규화 및 필요한 반정규화를 완료한 모델이다. 여섯째, 데이터 모델의 상세화와 통합화를 완료한 모델이다. 일곱째, 구현 단계에서 사용되는 특정 DBMS 및 성능 혹은 기타 제약사항과는 독립적으로 모델링 된다. 여덟째, 논리 데이터 모델은 프로세스 모델의 각 프로세스가 논리적으로 완전히 구성될 수 있도록 형성되어 있는지 상호 검증하는 것이 필요하다. 목표 논리 데이터 모델에서 수행해야 하는 주요 내용은 다음과 같다. 현행논리모델의 문제점과 유연성 개선하고, 목표시스템의 추가요건과 개선방안을 종합한다. 그림 7은 매우 상세한 논리 데이터 모델을 보여주는 목표 논리 데이터 모델이다.

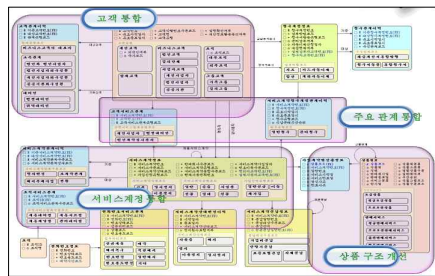


그림 7. 목표 논리 모델
Fig. 7. To-Be Logical Model

물리 데이터 모델은 최적의 정보 시스템으로 실제적인 구현을 위해 사용자가 프로그램에 직접 접근하여 사용할 수 있는 물리적 모습으로의 표현 형태를 나타낸다. 물리 데이터 모델의 특징은 다음과 같다. 첫째, 실제 구현을 목표로 모델을 설계한다. 둘째, H/W, DBMS 및 애플리케이션의 특성을 반영한다. 셋째, 필요에 따라 분할, 병합 또는 반정규화 등의 사항이 반영되어 표현된다. 넷째, 직접 데이터를 저장하고 액세스하는 단위의 테이블 또는 파일 형태로 표현한다. 다섯째, 테이블 명, 레코드 형태, 파일구조 및 칼럼의 데이터 타입 및 길이, Primary Key, Foreign Key 등이 구체적으로 표현된다.

여섯째, 물리 모델은 논리 모델을 근거로 생성된다.

목표 물리 데이터 모델에서 수행해야 하는 주요 내용은 다음과 같다. 논리모델을 매핑해서 자동 물리모델 생성, 표준화 자료를 이용한 자동 칼럼명 생성, DBMS 기술을 바탕으로 효율적인 물리모델 설계 및 검증, 다양한 제약조건 추가, 자동 스키마 생성, 물리모델과 스키마 간의 연결 유지 등이다.

6. BR 도출 및 품질체크 모듈 생성

이 모듈은 모델정보에서 업무 규칙(BR)후보를 자동으로 추출하는 기능으로 데이터 품질 관리체계를 먼저 정립하고 정립된 체계에 맞춰 시스템을 개선하여 구현한다. 데이터 품질 관리활동을 각 단계별로 구현하여 시스템에 따르기만 해도 자동으로 데이터품질 관리체계를 구축하고, 오류데이터 개선활동으로 데이터품질 지수의 개선현황을 분석할 수 있다.

데이터 품질 기본정보를 우선적으로 정의하여 관리하고, 각 기본 정보별로 관련정보를 조회할 수 있는 뷰를 제공한다. 진단대상 테이블의 기초 정보만으로 데이터 프로파일링을 실행한다. 또한 통합된 모델정보를 참조하여 기본적인 대상 업무규칙(참조 무결성, 도메인 정보 유효성) 후보를 자동으로 도출하고, 업무규칙을 정의하고 실행하여 품질지수를 측정한다. 그리고 정의된 업무규칙의 실행주기 및 SQL을 확인하고 검증하며, 프로파일링 결과를 각 분석현황으로 제공한다. 오류데이터를 보여주고 원인분석을 시행하고 결과를 정의할 수 있으며, 원인분석 결과를 대상으로 개선활동을 기록하고 정의하여 품질개선 활동을 주시할 수 있다. 그림 8은 데이터 품질관리 시스템이다. 데이터 품질관리시스템은 웹 기반의 커뮤니케이션 포털 형태로 지표 및 대상관리, 데이터 프로파일링, 다양한 기준의 품질평가, 결과분석 및 데이터 수집 및 측정의 자동화 기능 등을 포함하며 이를 활용하여 데이터 품질관리체계에 기반 해서 데이터 품질관리 활동을 한다. 기준관리에서는 데이터모델 정보를 통한 생성으로 평가 대상관리, DQI(Data Quality Index) 지표 관리, CTQ(Critical To Quality) 정보 관리, 가중치와 목표수준 관리를 한다. 데이터 프로파일링에서는 칼럼분석, 패턴분석, 날짜유형분석, 관계분석, 코드분석을 한다. 구조적 데이터 품질 단계에서는 코드 도메인 검증, 선분이력 데이터 중복 검증, 제약조건 검증, 테이블 구조품질 검증, 상관정합성을 검증한다. 측정 단계에서는 업무규칙 등록, 설정 주기별 자동측정, 스케줄러를 통한 작업 관리, 측정 결과 집계 및 저장, 작업 이력 관리, 상세 오류 내역 관리를 한다.



그림 8. 데이터 품질관리 시스템
Fig. 8. Data quality management system

영향도 분석에서는 데이터표준, 데이터모델 연계, 어플리케이션 영향도 분석, 성능 관점 데이터 연계를 분석한다. 분석 및 평가에서는 기간별/오너별 등의 분석, 현황 및 이력 조회, 추이 분석, 오류 데이터 추적 및 저장 등을 분석, 평가한다.

7. 데이터 이행

지금까지는 모델링 능력에 문제가 있었기 때문에 현행과 목표 시스템의 데이터 구조가 유사했고, 이로 인해 단순한 데이터 매핑이 가능했던 것이다. 이런 형태에서라면 집합의 개체 간의 매핑은 무시하고, 단지 속성 값을 새롭게 변환하는 정도만 처리해도 충분할 수 있다.

현행 데이터 아키텍처를 수립함으로써 현재의 매우 구체적인 데이터 정보가 리파지토리에 보관되어 있고, 목표 데이터 아키텍처를 수립하였다면 그것 또한 리파지토리에 관리되어 있으므로 이를 이용하여 명확한 매핑룰을 수립하는 것이 가장 이상적으로 데이터를 이행하는 방법이다. 데이터 이행 규칙이란 결국 데이터 집합 간의 변환, 이동을 의미하는 것이므로 소스 집합을 알고, 결과 집합을 안다면 그리 어렵지 않게 데이터를 이행을 할 수 있다. 데이터를 정제하는 일 또한 데이터 구조를 정확히 이해하는 것부터 출발되어야 한다.

데이터를 이행하는 애플리케이션에 수많은 예외 처리를 일일이 지정하려고 애쓰지 말고, 규칙을 벗어나는 데이터를 정제하여 규칙을 준수하도록 교정하고, 애플리케이션에서는 단지 규칙을 준수한 집합만을 처리하도록 하는 것이 가장 바람직한 처리 방법이다. 데이터 이행은 수많은 예외 경우를 처리하는 것이 아니라 데이터와 데이터 간의 매핑에 불과하다. 이를 위해서는 현행과 목표 시스템의 데이터 아키텍처를 명확하게 수립하는 것 말고는 절대 더 좋은 대안이 없다. 측정된 오류데이터에 대해 원인분석을 시행하고 결과를 기록하여 원인분석결과에 대한 반영을 실행하고, 그 결과를 전반적인 데이터 품질 개선 활동을 관리하도록 지원한다. 그림 9는 오류 데이터 조회 화면이다.

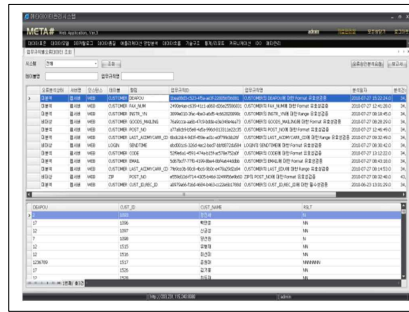


그림 9. 오류 데이터 조회
Fig. 9. Retrieve of error data

그림 10은 데이터 이행 시 GAP 반영과 매핑 설계 화면을 보여준다.

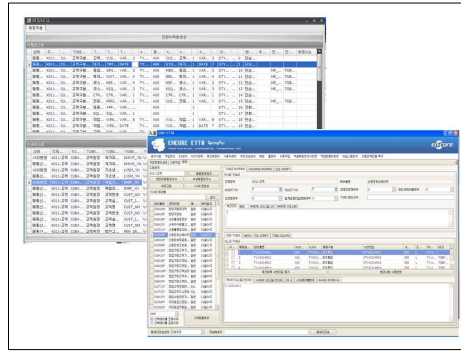


그림 10. 데이터 이행 GAP 반영 및 매핑 설계
Fig. 10. Data migration GAP and mapping design

IV. 구축 모델의 효과와 한계점

D사의 DA 구축 모델에서 표준 관리 방안, 구조 관리 방안, 이동 관리 방안, 모델 관리 방안, 품질 관리 방안 등에서의 효과와 한계점은 다음과 같다.

1. 표준 관리 방안

데이터 모델링 도구에서 직접 데이터 표준을 적용 및 표준 준수 여부 검사를 하고, 존재하지 않는 데이터 표준에 대해서 표준 신청 및 데이터 모델 승인/신청을 데이터 모델링 툴인 DA# MODELER에서 직접 하게 된다. 전사 표준단어 변경에 따른 표준용어 및 모델 오브젝트 중 해당 단어를 사용하는 실제 및 속성의 통계정보를 조회할 수 있으며 특정 실제와 속성 목록을 조회할 수 있도록 하여 변경 영향을 추적할 수 있다.

2. 구조 관리 방안

기업 내의 모든 DBMS의 오브젝트를 통합 관리하여 형상 관리, 저장 공간 사용량 관리, 변화추이 관리 등의 작업을 통합된 환경에서 수행할 수 있다. DBMS 정보 수집 시점에 이전 정보와의 차이를 자동으로 발견하고 선분관리 방식으로 이력관리를 한다.

3. 이동 관리 방안

이동관리 정보로서는 매핑정보, 작업정보, DFD, DBMS접 분석 정보를 통해 현행 이동관련 흐름 정보를 관리하고 조회할 수 있도록 이동관리 기능을 수행한다. 소스(원천) 시스템으로 부터 타겟(목표) 시스템으로의 데이터 흐름 관리를 통해 원천 변경 시 영향을 받는 부분에 대한 분석 및 데이터 사용도를 분석할 수 있다. 소스/타겟 정보, 둘 간의 매핑정보, 변환정보, 검증정보, 데이터 흐름 작업 정보 관리를 한다.

4. 모델 관리 방안

전사 데이터 아키텍처와 관리 프로세스를 통합하고 변화에 대한 영향도를 쉽게 파악할 수 있는 통합 리파지토리 기반의 데이터 품질관리 체계를 갖춘 시스템은 완료된 데이터 모델의 준수도, 활용도, 충실도, 영향도 측면에서 검증과정을 거쳐 리파지토리에 적재 및 관리를 한다. 데이터모델 정보 수집 시점에 이전 모델과의 차이를 자동으로 발견하고 선분관리 방식으로 변경관리 한다. 또한 특정 기간 또는 일자 별로 변경 통계 및 변경내역 정보를 조회할 수 있다. 물리모델과 실제 Database Catalog와의 상시적인 GAP 분석을 통한 데이터 모델, 물리 데이터베이스와의 일치화를 이룩할 수 있도록 한다. 이러한 작업은 주기적인 작업으로 등록하여 항상적인 GAP 분석을 통한 일치성을 확보한다.

5. 품질 관리 방안

데이터 품질관리시스템은 웹 기반의 커뮤니케이션 포털 형태로 지표 및 대상관리, 데이터 프로파일링, 다양한 기준의 품질평가, 결과분석 및 데이터 수집 및 측정의 자동화 기능 등을 포함하며 이를 활용하여 데이터 품질관리체계에 기반한 데이터 품질관리를 한다.

데이터 품질 기본정보인 DQI, CTQ, 진단대상, DQ영역 관리 기능을 지원하여 기초정보를 정의할 수 있다.

구현 모델의 한계점은 시스템 구축과 운용 단계에서의 데이터 관리 영역에 한정하여 엔터프라이즈 아키텍처의 전체

를 관리하고, 통제할 수 있는 부분에는 부족함이 있다. 앞으로 데이터 아키텍처와 어플리케이션 아키텍처의 정렬(alignment)을 위한 연구가 필요하고, 데이터 아키텍처를 중심으로 엔터프라이즈 아키텍처가 구축되는 모델에 대한 연구가 필요하다.

표 1은 D사의 DA 구축 전과 구축 후의 개선된 내용을 비교한 것이다.

표 1. DA 구축 전과 구축 후의 개선된 내용 비교
Table 1. Comparison between before and after implement of DA

항목	DA#사용전	DA#사용후	비 고
표준화기간	장기간소요	2개월	완료
리버스 모델링	없음	차세대 모델링구현	구현
모델 간 얼라이언트	불가능	가능	완료
DA 구현	DBA 수작업	자동화	완료
DA 검증	없음	자동화	완료
데이터 이행 설계	엑셀로 매핑정의서 작성	ETL#으로 매핑정의서 작성	완료
데이터 이행 구현	수작업 프로그램	자동소스 생성	구현
데이터 이행 검증	수작업 프로그램	자동 소스 생성	구현
Entity 의 수	5200	4600	12% 감소
통합된 Entity 의 수		10%	통합됨
부서 간 의사소통	막힘	원활	향상됨

V. 결 론

잘 구축한 데이터 아키텍처를 통해서 얻을 수 있는 기대 효과는 다음과 같다. 데이터 품질에서는 첫째, 부정확하거나 불일치하는 데이터가 없어지고, 시스템별 중복 데이터가 없어서 정확성이 높아진다. 둘째, 데이터 중심의 모델이 고려되어서 프로세스의 설계 시 접근성이 좋아진다. 셋째, 차세대 시스템 구축 시 시스템 통합의 확장성이 가능해진다. 넷째, 데이터 간의 표준화로 인해서 일관성이 확립된다. 다섯째, 데이터 중심의 프로그램 개발에 의한 테이블의 감소로 인하여 운영에 적은 부하가 걸린다. 운영의 적은 부하는 관

리 비용을 감소시키고, 중복 데이터 처리 비용이 감소되어 현 시스템의 관리 비용과 미래 시스템의 투자비용이 감소된다. 여섯째, 테이블의 중복과 조인이 적어짐으로 해서 빠른 데이터 조회가 가능하여 효율성이 높아진다.

본 연구의 기대 효과 및 활용 방안은 다음과 같다. 첫째, 체계적인 데이터 아키텍처 구축방법론을 제시함으로 인해서 비즈니스에 유연한 데이터 아키텍처를 구축할 수가 있고 전사적인 비즈니스를 체계적으로 접근하여 현행시스템과 목표시스템의 손쉬운 식별을 통해서 IT 거버넌스의 중요한 기반으로 활용을 할 수 있다. 둘째, 자동화된 데이터 아키텍처 구현 및 검증방법론을 통해서 신규나 변경사항이 모두 적용되었다는 완전성과 설계한 그대로 구현되었다는 정확성, 빠르게 구현할 수 있는 신속성 및 잘못된 구현을 예방하여 기존의 데이터가 일관성을 가지도록 한다. 이로 인해서 비즈니스 변화에 유연하게 대처할 수 있는 토대를 마련한다. 또한 데이터 아키텍처 구조를 통해서 데이터 값 품질을 관리하는 데이터 프로파일링 기능을 자동으로 구현할 수 있으며, 여기에서 나온 결과를 분석하여 데이터를 정제하는 것은 물론 데이터구조 품질을 높이는 자료로도 활용할 수가 있다.

참고 문헌

- [1] Mi-Young Choi, Jong-Youn Bae, "A Methodology for Developing Data Taxonomy for Data Architecture", Fifth International Joint Conference on INC, IMS and IDC, 2009.
- [2] Won-Young Choi, "IT Governance", Seung Rung Press, 2009.
- [3] Hwa-Sik Lee, "Data Architecture Solution I, Encore, 2004.
- [4] Seok-Soo Kim, "A Efficiency Implement Strategy and Method for Data Architecture", IT Institute of Kyungwon University, Vol 8, pp. 33-40, Dec. 2010.
- [5] Sung-Kun Kim, "A Literature Review Study on Enterprise Architecture", Journal of Information Technology and Architecture, Vol. 7. No. 1, pp. 63-76, Apr. 2010.
- [6] Sangwon Lee, "Automated Enterprise Data Model by Formulating Requirements", Journal of Information technology Application & management, pp. 263-283, Dec. 2009.
- [7] Zachman John, "A Framework for Information System Architecture", IBM System Journal V. 26 No.3, pp. 276-292, 1987.
- [8] Korea Database Agency, "The Guide for Data Architecture Professional", 2010
- [9] Gil-Sang Jang, "A Study on the Scheme of Implementing Meta-data Based Applications for Enterprises", Korea Society of Computer Information, Vol. 14 No. 9, pp. 135-145, Sept. 2009.
- [10] The Zachman Institute for Framework Advancement, <http://www.zifa.com>
- [11] <http://www.en-core.com>

저 자 소개



김 석 수

1982 : 숭실대학교 전자계산학과
공학사

1987 : 숭실대학교 전자계산학과
공학석사

1998 : 숭실대학교 전자계산학과
공학박사

1989 : 경원대학교 컴퓨터공학과
교수

관심분야 : DB, EA, 정보공학

E-mail : sskim@kyungwon.ac.kr



이 화 식

1982 : 경북대학교 수학과 학사

1996 : (주) 엔코아 대표이사

관심분야 : 데이터 아키텍처, DB

E-mail : hslee@encore.com