

정보시스템의 데이터 이관에 관한 연구

신현정¹

¹신한대학교 교수

A Study on Data Migration of Legacy Information System

HyunJeong Shin¹

¹Professor, Shinhan University

¹Corresponding author: hjshin@shinhan.ac.kr

Received June 5, 2021; Accepted June 17, 2021

ABSTRACT

데이터 이관(data migration)이란 한 컴퓨터 시스템에서 구축된 데이터베이스를 다른 시스템으로 이전하는 것을 말한다. 데이터 이관은 대부분의 기관에서 비즈니스 활동의 근간을 이루는 정보시스템을 구축할 때 수행해야 하는 중요한 프로세스로서 전용 도구 및 표준 방법론이 거의 없는 상황이다. 이번 연구의 시사점은 데이터 이관 프로젝트를 전략적 관점으로 인식하여 데이터 품질을 개선할 기회로 삼아야 한다는 점이다. 데이터 이관 과정을 통해 가치가 낮은 데이터가 제거되면서 좀 더 빠른 정보 서비스가 가능해지며, 궁극적으로는 개선된 비즈니스 결과를 생성할 수 있기 때문이다. 제시한 이관 단계의 체크 리스트를 작성하여 데이터 이관을 수행할 경우 새로운 플랫폼에 사용자 요구사항을 잘 반영한 정보시스템으로 변화가 일어날 것으로 기대된다. 본 연구에서는 레거시 시스템의 업무기술키, ERD와 Fat 파일만을 고려하였으나 향후 각 단계별 체크 리스트가 정의된다면 데이터 이관의 표준 방법론으로 발전할 수 있을 것이다.

Data migration refers to moving a database built on one computer system to another. Data migration is an important process that most organizations must perform when building information systems that form the basis of business activities, and there are few dedicated tools and standard methodologies. The implication of this study is that the data migration project should be recognized from a strategic perspective and used as an opportunity to improve data quality. This is because, as low-value data is removed through the data migration process, faster information services are possible, and ultimately, improved business results can be generated. When data transfer is performed by creating a checklist of the proposed transfer stage, it is expected that the new platform will change to an information system that reflects user requirements well. In this study, only the work description, ERD, and Fat files of the legacy system were considered, but if a checklist for each step is defined in the future, it can be developed as a standard methodology for data migration.

Keywords: Data migration, Data modeling, Database, Data transfer, Information system

1. 서론

데이터 이관은 대부분의 기관에서 비즈니스 활동의 근간을 이루는 정보시스템을 구축할 때 수행해야 하는 중요한 프로세스이다. 전용 도구 및 표준 방법론이 거의 없는 상황이다. 정보시스템 구축의 다양성과 데이터 이관 프로세스의 복잡성으로 인해 표준 사례를 정의하는 것은 어려운 일이므로 여러 문헌에서 제시한 모범사례를 통해 데이터 이관 계획수립시 참고할 수 있는 시사점을 얻었다. 또한 새로운 플랫폼의 변화로 데이터 이관이 많이 발생할 수 있는 시기가 도래될 것으로 기대되며 본연구에서 제



시한 데이터 이관의 각 단계는 정보시스템의 데이터 이관에 활용 될 것으로 기대된다.

이번 연구에서 얻은 중요한 시사점은 데이터 이관 프로젝트를 전략적 관점으로 인식해야 하고 데이터 품질을 개선 할 수 있는 기회로 삼아야 한다는 점이다. 데이터 이관 과정을 통해 가치가 낮은 데이터를 제거하면 좀 더 빠른 정보서비스가 가능해지며 궁극적으로는 개선된 비즈니스 결과를 생성 할 수 있기 때문이다.

ETL(Extract Transform Load)단계의 체크리스트를 작성하여 데이터 이관을 수행 한다면 새로운 플랫폼에 사용자 요구사항을 잘 반영된 정보시스템으로 변화가 될 것으로 기대된다. 본 연구에서는 레거시 시스템의 업무기술서, ERD만 와 Fat파일만을 고려하였다.

논문의 구성은 2장에서 데이터 이관과 데이터 통합에 대한 정의를 살펴보고 정보시스템의 플랫폼의 변화와 데이터베이스 설계과정을 기술한다. 3장에서 레거시 시스템의 데이터 이관 단계 정의하고 레거시 시스템의 사용자 업무분석서를 ERD로 변환하여 검증하는 것을 제시한다. 향후 연구과제로 정보시스템의 변화에 따른 ETL 각 단계별로 체크리스트가 정의 된다면 데이터 이관의 표준화 방법론을 연구 분야로 남겨 두고자 한다.

2. 관련연구

2.1 데이터 이관의 정의

오늘날 정보시스템은 대부분 기관에서 비즈니스 활동의 근간을 이루는 중요한 인프라이다. 정보시스템은 각자의 생명주기에 따라 일정 기간이 지나면 구식 정보시스템이 되므로 신규 정보시스템 구축이 필요성이 생기며 반드시 수행해야 할 중요한 일로 대두되고 있는 실정이다. 이러한 신규 정보시스템의 일환으로 데이터 이관이 수행되게 되고 이때 정보시스템에 따라 어떤 가이드라인이 필요한 실정이다.

유럽의 시장조사 기관인 Bloor Research가 2011년에 수행한 조사보고서에 따르면 데이터 이관 프로젝트의 84%가 예산초과 또는 사업기간 초과 또는 이 두가지 사항에 모두 해당된다. 동일한 기관이 2011년 봄에 실시한 조사결과에서는 많이 개선되어 데이터 이관 프로젝트의 62% 정도가 예산범위 내에 정시에 완료된 것으로 나타났으며 아직 미비한 실정이다^{1,2}.

국내의 경우 최근 수년간 우리가 경험한 공공기관의 거의 모든 데이터 이관 프로젝트가 시스템통합 업체에 의해 수동 코딩 방식(hand code approach)이나 복제 방식(copying approach)으로 수행되었다. 이러한 경우는 데이터 이관을 신규 정보시스템 구축의 일부분으로 수행되는 단순한 데이터 이동(data moving) 프로젝트라는 인식과 일회성 작업이라는 특징 때문에 데이터 이관의 중요성을 간과하여 노력 및 비용 투입이 미흡한 것으로 추정된다.

임베디드 소프트웨어에서 임베디드 시스템의 목적과 기능이 발휘될 수 있도록 구현된 소프트웨어로 운영체제와 특정 목적에 관련된 응용 프로그램으로 나눌 수 있다. 현재까지 나온 임베디드 운영체제들은 대부분 실시간 운영체제지만, WindowsCE, Embedded Linux, RT-Linux, Embedded Java와 같이 실시간적인 요소를 그다지 충족하지 않는 운영체제들이 존재한다.

선행연구 문헌²에서 데이터 이관의 개념을 명확하게 설명하기 위해 주요한 특징을 사용하였다. 이들이 데이터 이관의 개념을 이해하는데 중요한 요소이므로 이를 반영하여 데이터 이관을 다음과 같이 정의할 수 있다. 데이터 이관은 기술적 또는 비즈니스적인 요구에 따라 기존 환경에서 올바른 품질의 데이터를 선정, 추출, 변환 및 재구성 하여 동일한 데이터 용법의 신규 환경으로 영구적으로 적재하는 단방향 일회성 프로세스이다.

- 올바른 품질의 데이터 구성

데이터 이관은 소스 데이터의 전체를 오류 없이 목표 시스템으로 이동하는 것이 아니라 소스 데이터에서 목표 시스템에 유용하면서 적절한 품질의 데이터만을 선정하여 이동하는 것이다. 그러므로 데이터 이관시 데이터 필터링 및 정제 단계가 필요하다.

- 변환 및 재구성

데이터 이관은 일회성 프로세스이며, 주기적으로 데이터를 이동하는 데이터 통합과는 구별된다. 데이터 이관을 프로세스로 인식해야 하는 이유는 데이터 이관이 단일 기술이 아니고 여러 기술을 사용하고, 이들을 연계해서 여러 단계로 수행해야 하기 때문이다. 그래서 데이터 이관은 단순히 데이터 이동작업이 아닌 프로세스로 관리해야 한다.

대부분 “이관”이란 용어를 “누락 없이 소스 데이터를 이동시킴”이란 의미로 사용하고 있어, 목표 시스템으로 이동된 데이터에 대한 검증 작업은 포함하고 있으나 소스 데이터에 대한 분석 작업, 즉 소스 데이터의 필터링 및 정제 단계에 대한 내용이 미흡하다^{5,6)}.

2.2 데이터 이관과 데이터 통합의 차이점

데이터 이관과 데이터 통합의 차이점을 살펴 봄으로써 데이터 이관의 개념을 좀더 명확하게 규명하고자 한다. 데이터 통합은 여러 곳에 산재된 이질적인 소스 데이터를 모아서 사용자에게 비즈니스 통찰력이 높은 통합 정보로 제공하기 위한 프로세스이며 통합 정보시스템을 구축하는 과정에서 수행된다.

ETL은 1990년대에 데이터 웨어하우스와 함께 발전했다. 전통적인 ETL은 산재되어 있는 여러 소스에서 데이터를 추출하고, 일관된 형식으로 변환하여, 데이터 웨어하우스에 맞춤식(customization)으로 삽입하는 프로세스를 지칭하며 이러한 기능을 지원하는 소프트웨어를 ETL 도구라고 한다. 최근의 확장 ETL 도구들은 데이터 통합을 지원하는 기능을 포함하고 데이터 통합을 위한 주요 도구로 사용되고 있다. 또한 ETL 소프트웨어는 데이터 이관 도구로 많이 사용되고 있다⁴⁾.

데이터 이관과 데이터 통합 모두 한 환경에서 다른 환경으로 데이터를 이동하는 복잡한 프로세스이지만, 다음과 같은 차이점이 있다.

- 신규 환경으로 이동

데이터 통합은 주로 기존 환경내에서 데이터를 이동하는데 반하여, 데이터 이관은 항상 기존의 구식 환경에서 신규 환경으로 데이터를 이동한다.

- 소스 시스템의 폐기

데이터 통합은 데이터 이동 후에도 소스시스템이 유지되고 운영되지만 데이터 이관의 경우는 소스시스템이 가동 정지 또는 폐기된다.

- 일해성 이동

데이터 통합은 소스 시스템에서 목표 시스템으로 데이터가 주기적으로 그리고 반복적으로 이동하지만 데이터 이관은 일회성으로 데이터가 한 방향으로 이동한다.

- 기업 내부 간의 이동

데이터 통합은 일반적으로 데이터 분석을 위해 기업 외부의 데이터를 내부로의 수집을 포함하며, 데이터 이관은 기업 내부의 기존 시스템의 데이터를 기업 내부의 신규 시스템으로 이동한다. 그러므로 데이터 통합 프로세스의 설계 시 이동하는 데이터 규모에 따라 외부 네트워크 대역폭이 중요한 고려사항이 될 수 있다.

- 장기간 축적된 데이터의 이동

데이터 이관에서는 기존 정보시스템의 구축 시부터 축적된 데이터를 신규 정보시스템으로 일회성으로 이동한다. 데이터 통합은 산재되어 있는 기존의 여러 시스템에 축적된 데이터를 기 구축된 통합 정보시스템으로 주기적으로 수집한다. 그러므로 통상적으로 이동하는 데이터의 축적 기간의 데이터 통합에 비해 데이터 이관의 경우에 훨씬 길다. 이로 인해 데이터 통합의 경우보

다 데이터 이관의 경우에 소스 데이터에 대한 데이터 프로파일링(profileing) 작업이 중요하며 이외 여러 가지 문제가 나타날 수 있다. 그래서 데이터 정제와 데이터 변환 작업이 복잡할 수 있다. 그리고 데이터 이관이란 용어를 데이터 통합의 의미로 사용하고 있다.

2.3 정보시스템의 변화

2.3.1 정보시스템의 전략적 역할

많은 조직의 최고 경영진은 정보시스템이 기업 운영에 필요하기는 하지만, 수익이라든가 시장점유, 새로운 모험 사업의 개발 등과 같은 그들의 주요 관심 영역에는 별 관계가 없는 것으로 생각해왔으나 이러한 견해가 변하고 있다. 정보시스템은 기업의 경쟁력에 영향을 미칠 수 있으며 또 실제로 영향을 주고 있다. 기업의 정보시스템들은 기업 경쟁의 도구인 것이다. 그러나 IT를 경쟁적 도구로 이용하려는 기업들이 모두 기대하는 결과를 성취하는 것은 아니다. 그 이유는 기업들이 낡은 방식의 프로세스를 이용하고 있기 때문이다. 과거의 IT를 이용한 경쟁력 우위 쟁취의 추구는 중요한 교훈을 남겼다. 단순히 현재의 작업방식을 자동화 하는 것만으로는 효과가 거의 없다. 조직의 프로세스들은 정보기술을 충분히 이용할 수 있도록 “리엔지니어링” 되어야 한다. 그러므로 IT는 경쟁력의 도구인 동시에 낡은 사업방식의 근본적인 개조를 위한 하나의 촉매로서 전략적 이용도 가능한 것이다. 기존의 사업방식은 종이 중심의 느린 프로세스에서는 적합 했지만 컴퓨터 중심의 빠른 운영에서는 장애가 된다. IT의 이러한 이용은 과거에는 잘 새롭고 혁신적인 컴퓨터의 사용인 것이다. 또한 IT의 정보시스템은 경쟁력 있는 시스템들로 흔히 기업을 공급자 또는 고객들과 전사적으로 연결시켜 주고 있다. 리엔지니어링된 비즈니스 프로세스에서는 주문입력이나 청구서 발송, 대금 지불 등에 조직 간의 데이터 전송을 창의적으로 시도되고 있다. 기업조직들이 글로벌 시장을 목표로 하면서 조직간 “전략적 제휴” 쪽으로의 추세가 강화되어 갈 것이다. 이 모든 것은 지역, 국가, 세계적 단위의 전자 업무가 가능하도록 해주는 조직간 시스템의 개발을 필요로 한다.

2.3.2 지능형 시스템 도래

“지능형 시스템”이란 말은 4차산업 혁명시대에 대 유행하고 있는 용어이다. AI는 인간 행위의 어떤 측면, 예를 들면 사고를 판단든지 또는 의사소통을 하는 것을 모방하거나, 또는 인간의 감각을 흉내 내보려는 일련의 기술을 만든다. AI 기술에는 전문가 시스템, 신경망, 퍼지논리, 기계번역, 음성인식, 자연어처리 등이 포함되며 조금씩 현실화 되기 시작되었으며 4차산업 혁명이 일어나면서 산업 여러분야에서 현실화 되었다. 특히 지식-중심 시스템이라고도 불리는 전문가 시스템은 이미지 처리나 전통적인 정보시스템 같은 여타 기술과 아울러 그것을 이용하고 있는 많은 대기업들에게 주된 컴퓨팅 옵션으로 여겨지게 되었다.

전문가 시스템이란 일종의 분석 또는 문제-해결을 위한 모델로서 거의 언제나 컴퓨터를 통해서 구현되며 “전문가”가 하는 방식으로 어떤 문제를 다룬다. 해결 프로세스는 지식 또는 전문성의 지식베이스를 조회하여 문제의 성격에 기초한 해답을 이끌어 내는 것으로 되어 있다.

전문가 시스템은 “사용자 인터페이스, 추론엔진, 그리고 저장되어 있는 전문성-다시 말하면, 하나의 규칙 세트 지식베이스 또는 전체적인 지식시스템으로 구성되어 있는 하나의 컴퓨터-중심 시스템으로서 그 목적은 어느 특정 영역에 있어서의 문제점에 대해 조언을 하거나 해결책을 제시하는 데 있다. 조언은 해당 문제 영역에 있어서 인간 전문가가 주게 될 조언에 필연적이다.” 고 정의한 바가 있다. 전문가 시스템은 어떤 문제를 해결해 주어야 하고, 그 문제를 어떻게 해결했는지 어느 정도 설명할 수 있어야 하고 유사한 문제를 해결할 수 있는 신뢰할 만한 수단을 제공해 주어야 한다.

전문가 시스템은 1956년 카네기 멜런 대학(Carnegie-Mellon University)의 앨런 뉴얼(Allen Newel)과 허버트 사이몬(Herbert Simon)이 랜드 코퍼레이션(rand Corporation)의 제이씨 쇼(J.C Shaw)와 더불어 개발한 Logic Theorist였다. 또 다른

초기 전문가 시스템으로는 DENDRAL이 있으며 그것은 하나의 질량 분광계가 만들어낸 데이터를 해석하여 분자의 구조를 파악하는 프로그램이다⁷⁾.

전문가 시스템 구성 요소의 하나인 사용자 인터페이스는 전문가 시스템과 외부 세상과의 사이에 있는 접촉면으로 컴퓨터 어플리케이션일 수도 있고 또는 사람일 수도 있다. 추론기관(inference engine)은 지식베이스를 검색해서 문제를 해결하는데 이용되는 추론 방법을 담고 있는 소프트웨어의 해당 부분을 말한다. 지식베이스(knowledge base)는 어떤 특정 애플리케이션에 관련되어 있는 사실들이나 데이터를 저장하고 있는 데이터베이스를 말하며 일종의 정보시스템이 된다. 추론기관은 이정보를 이용해서 문제를 풀어간다⁸⁾.

2.4 정보시스템 데이터 모델링

데이터 모델링이란 기업의 정보구조를 체계적으로 나타내는 방법으로 기업의 데이터를 사용자 관점에서 인식, 분석하여 이를 표준화된 심벌을 이용하여 표현하는 기법을 말한다. 데이터 모델링은 논리적 모델링과 물리적 데이터 모델링으로 구분할 수 있으며 논리적 데이터 모델링은 데이터베이스 구축의 분석단계에 해당하고, 물리적 데이터 모델링은 정보시스템 구축의 설계단계에 해당된다³⁾.

2.4.1 개념적 데이터 모델링

개념적 데이터 모델링은 논리적 데이터 모델링의 첫 번째 단계로서 조직 전체적인 관점에서 데이터 골격을 만드는 것으로 개체(entity), 속성(attribute), 관계(relationship)등을 정의하고 서로 관련을 맺는 작업을 수행한다. 따라서 개념적 데이터 모델링은 논리적 데이터 모델링의 기초 자료가 된다. 개체별 식별자 정의, 데이터 모형 상세화, 데이터 모형 통합 순으로 이루어지며 ERD(Entity Relationship Diagram)를 작성하게 된다.

데이터 모델링은 정보전략계획(ISP:Information Strategy Planning)의 주제영역(subject area)을 중심으로 체계를 추가하며 기존 개체와 관계를 구체화하고 업무규칙을 추가하면 해당 업무영역에 대하여 필요한 데이터, 데이터 간의 연관성, 데이터의 구체적인 내용, 데이터의 정확성을 위한 업무규칙 등이 무엇인지 알 수 있다. 즉 데이터 모델링은 비즈니스에 대한 종합적인 이해를 바탕으로 기업 내에 존재하는 데이터에 대하여 프로세스와는 독립적으로 인식하여 이를 알기 쉽고 체계적으로 표현한 기업의 정보요구라고 할 수 있다. 데이터 모델링 과정은 Fig. 1과 같다.

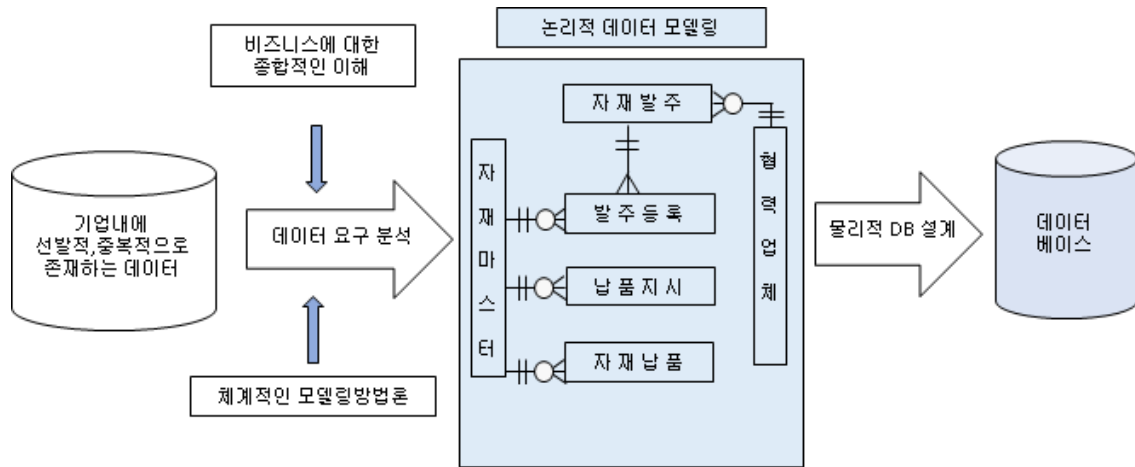


Fig. 1. Data modeling

2.4.2 논리적 데이터 모델링

기업 내에 자발적, 중복적으로 존재하는 데이터를 분석가가 관찰하고 비즈니스에 대한 종합적인 이해와 체계적인 모델링 방법으로 사용자의 데이터 요구사항을 분석하여 논리적 데이터를 모델링한다. 그리고 개체와 개체와 개체를 합치고 통합하기도 하면서 물리적 데이터베이스를 설계하게 되는데 이때 정규화된 산출물이 비정규화되기도 한다. 논리적 데이터 모델링은 아래와 같은 적정성 판단 기준을 지키도록 해야 한다.

- 구조적 타당성(structural validity) : 데이터를 정의하고 구성하는 방법의 일관성
- 단순성(simplicity) : 사용자의 이해 용이
- 비중복성(non-redundancy) : 필요한 데이터는 한 군데에 한 번만 존재
- 공유성(shareability) : 어떠한 적용업무나 기술에 특화되지 않은 다수에 의해 사용가능
- 무결성(integrity) : 비즈니스에서 데이터를 사용하고 관리하는 방식의 일관성
- 확장성(extensibility) : 현재 상태에서 최소한의 노력으로 새로운 요구를 수용할 수 있는 능력

논리적 데이터 모델링의 핵심성공요소(CSF:critical success factor)는 다음과 같다.

- 해당 비즈니스에 대한 종합적인 지식을 가진 현업 담당자와 공동으로 작업
- 논리적 데이터 모델링 전 단계에 걸쳐 체계적인 방법론의 사용
- 프로세스와는 독립적 개념으로 데이터 중심으로 접근
- 정규화(normalization) 기법채택

2.4.3 물리적 데이터 모델링

물리적 데이터베이스 설계는 논리적 데이터베이스 모델링 산출물을 내부모델, 즉 물리적 데이터베이스 구조로 전환하는 과정이다. 물리적 파일과 데이터베이스 설계는 다음과 같은 SDLC 단계를 통해 수집되는 정보들을 이용한다.

- 논리적 설계 단계에서 작성된 물리적 데이터베이스 구조(ERD와 논리적 데이터모델 설계)
- 사용자 요구정의 단계에서 수집된 사용자 정보처리요구(자료사전등에 명시된 사용빈도, 응답시간, 보안사항, 백업, 회복, 유지 등)
- 구현할 DBMS, 운영체제, 네트워크 등의 아키텍처 환경
- 조직환경(사용자 수, 근무지 개수 등)

따라서 물리적 데이터베이스 설계 단계에서는 다음과 같은 작업이 수행된다.

- 데이터 사용량 분석 및 분배전략 수립 : 데이터 사용량 분석표, 데이터 구조도 작성, 데이터 분산 분석표
- 논리적 데이터 구조를 물리적 데이터 구조로 변환
- 역정규화
- 물리적 설계를 위한 SQL문 작성 : 뷰, 인덱스, 트리거 설계
- 테이블 기술서 작성 : 선정된 DBMS를 고려한 테이블 기술서 작성

3. 데이터 이관

3.1 데이터 이관 프로세스

정보시스템은 기업 내 산재되어 있는 여러 종류의 데이터 소스들과 시스템, 응용프로그램 등으로부터 데이터를 모으고, 이 데이터들이 재사용 가능하고 가치 있는 정보, 즉 ‘지능적이며 데이터 가치가 있는 정보’가 되도록 만들어주는 연속적이고 체계적인 과정이라 할 수 있다. 이런 과정에 반드시 필요한 부분이 데이터 이관에 사용되는 기법인 ETL 방법론이다. ETL은 데이터를 소스 시스템에서 추출(Extract) 하여 변환(Transform), 정제(Cleansing)후 타겟 시스템에 적재(Load)하는 작업까지의 전 과정을 의미한다. 다음 Fig. 2는 Legacy System에 산재되어 있는 데이터들과 사용자 요구사항들을 중심으로 새로운 시스템으로 이관하는 프로세스다.

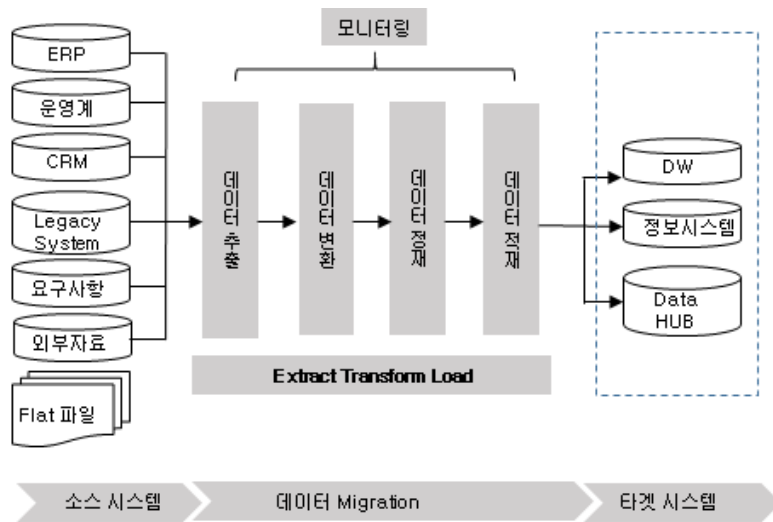


Fig. 2. ETL (Extract Transform Load)

본 연구에서는 최초 레가시 시스템이 개발될 때 작성된 업무 분석서를 고려하여 ERD를 작성하고 업무분석서 체크 리스트를 작성하며 이관되는 데이터의 누락 사항을 체크 할 것을 제안한다. 자금관리업무 분석서와 체크 리스트는 Table 1과 Table 2와 같으며 업무 분석서를 근거로 작성한 자금관리 ERD는 Fig. 3과 같다.

Table 1. Fund management needs analysis

<p>자금관리 업무 분석서</p> <p>1 (자금 계획수립)</p> <p>1-1 차입금 담당자는 차입처에서 차입과목별로 차입한도를 승인받는다.</p> <p>1-2 차입금 담당자는 차입처에서 차입금을 차입한다.</p> <p>1-3 자금 계획을 수립할 때에는 차입금 상환액을 반영한다.</p> <p>1-4 리스 담당자는 발주회사와 리스계약 체결 후에 리스 원장을 기록한다.</p> <p>1-5 자금 계획에는 리스 상환액을 반영한다.</p> <p>1-6 유가증권 담당자는 유가증권을 구매한다.</p> <p>1-7 년자금 계획에는 유가증권 입금액을 반영한다.</p> <p>1-8 예금 담당자는 은행에서 예금 구좌를 개설한다.</p> <p>1-9 자금 계획에는 예금 및 적금 불입액을 반영한다.</p>

Table 2. Fund management check list

No	업무기술	사상	Remark
1-1	차입금 담당자는 차입처에서 차입과목별로 차입한도를 승인받는다.	Y	승인 1-1/1-2
1-2	차입금 담당자는 차입처에서 차입금을 차입한다.	Y	차입 1-2
1-3	자금 계획을 수립할 때에는 차입금 상환액을 반영한다.	Y	차입금상환 1-3
1-4	리스 담당자는 발주회사와 리스계약 체결 후에 리스 원장을 기록한다.	Y	기록 1-4
1-5	자금 계획에는 리스 상환액을 반영한다.	Y	상환 1-5
1-6	유가증권 담당자는 유가증권을 구매한다.	Y	구매 1-6
1-7	년자금 계획에는 유가증권 입금액을 반영한다.	Y	유가증권금액 1-7
1-8	년자금 계획에는 유가증권 입금액을 반영한다.	Y	개설 1-8
1-9	자금 계획에는 예금 및 적금 불입액을 반영한다.	Y	예적금불입 1-9
1-10	모바일 서비스	N	새로운 요구사항

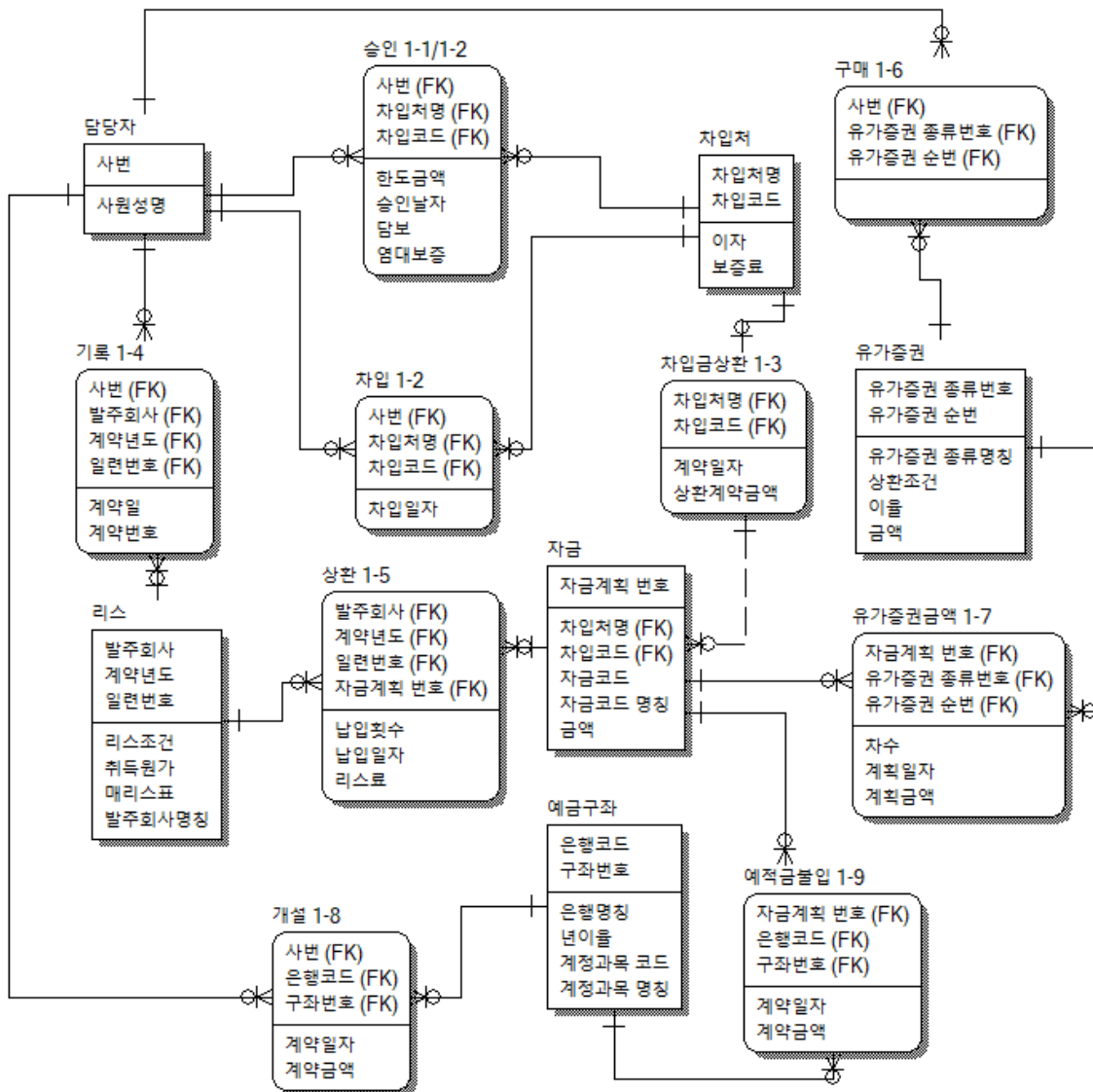


Fig. 3. Fund management ERD

3.2 데이터 이관 단계별 검증

레가시 시스템은 작성된 업무분석서와 이를 근거로 작성된 ERD는 새로운 시스템에서 업무분석서로 이관 되어야 할 자료이며 새로운 요구사항이 추가될 수 있다. 따라서 업무분석서와 데이터베이스 설계 과정에서 작성된 ERD로 토대로 데이터 이관과 데이터 정제 단계를 확인할 수 있다. 자금관리 시스템의 ERD와 업무분석서가 새로운 환경의 목표시스템으로 사상(mapping) 되었을 때 올바른 데이터 이관이 완료 되었다고 판단할 수 있다.

4. 결론

데이터 이관은 대부분의 기관에서 비즈니스 활동의 근간을 이루는 정보시스템을 구축할 때 수행해야 하는 중요한 프로세스이며 전용 도구 및 표준 방법론이 거의 없는 상황이다. 정보시스템 구축의 다양성과 데이터 이관 프로세스의 복잡성으로 인해 표준 사례를 정의하는 것은 어려운 일이다. 선행 연구에서 제시한 모범사례를 통해 데이터 이관 계획수립시 참고할 수 있는 중요한 시사점을 얻었다. 또한 새로운 플랫폼의 변화로 데이터 이관이 많이 발생할 수 있는 시기가 도래 될 것으로 기대 하며 본 연구에서 제시한 데이터 이관의 단계는 데이터 이관을 염두한 정보시스템에서 활용될 것으로 기대된다.

이번 연구에서 얻은 중요한 시사점은 데이터 이관 프로젝트를 전략적 관점으로 인식해야 하고 데이터 품질을 개선 할 수 있는 기회로 삼아야 한다는 점이다. 데이터 이관 과정을 통해 가치가 낮은 데이터를 제거하면 개선된 비즈니스 결과를 얻을 수 있다. 따라서 제시한 ETL단계의 체크리스트를 작성하여 데이터 이관을 수행 한다면 새로운 플랫폼에 사용자 요구사항을 잘 반영된 정보시스템으로 변화가 일어날 것으로 기대된다.

본 연구에서는 자금관리 시스템의 업무분석서, ERD 와 Fat파일만을 고려하였지만 향후 각 단계별 업무분석서와 체크리스트가 정의 된다면 데이터이관의 표준 방법론으로 발전 될 것이다.

Acknowledgement

※ “본 논문은 2020년도 신한대학교 학술연구비 지원으로 연구되었음.”

References

1. 박상현, “ETL 데이터 이관의 성능 향상에 관한 연구”, 송실대학교 정보과학대학원, June 2015.
2. 구흥서, “실용적 관점에 기반한 데이터 이관에 관한 연구”, 융복합지식학회 논문지, 제7권 제1호, pp.109-114, 2019.
3. 구흥서, 신현정, “오라클을 활용한 데이터베이스 이론과 실습”, 이한출판사, 2003.
4. P. Howard, and C. Potter, “Data Migration in the Global 2000-Research”, Forecasts and Survey Results, White Paper.
5. P. Howard, and C. Potter, “Data Migration”, White Paper, Bloor Research, May 2011.
6. H.-S. Park, and H.-C. Kim, “A Data Migration Method for Developing GIS-based Fisheries Resources Information Systems”, Journal of Digital Convergence, Vol. 11, No. 6, pp. 143-149, June 2013.
7. M. Rouse, Data Migration, Blog in Research CIO, posted 2016.10.14.
8. P. Howard, Data Migration, Blog in Research, last update 23rd January, 2018.