

엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스 개발 프로세스

최 성 만[†] · 유 철 중^{††} · 장 옥 배^{†††}

요 약

기존의 연구비 관리업무는 예산계획, 예산편성 및 예산정산 부분으로 관리됨으로써 여러 가지 문제점이 발생되었다. 이러한 문제점을 해결하고 연구비의 안정적인 확보와 효율적인 운영 및 투명한 집행을 위하여 연구비의 통합관리가 절실하게 요구되고있는 실정이다. 최근의 이러한 추세를 반영하여 기존의 시스템 통합업체(Inmon, IBM)의 데이터 웨어하우스 개발 프로세스에 대해서 연구한 결과 Inmon의 데이터 웨어하우스 개발 프로세스에서는 고전적인 개발주기 기법으로 단계적이며 순차적인 접근방법을 이용한다. 이로 인해 개발단계에서 중복되어지는 부분이 나타나게 되어 각 단계의 진행과정에서 그 이전단계로 피드백이 요청되는 심각한 문제를 유발하게 된다. 또한, IBM의 데이터 웨어하우스 개발 프로세스에서는 개발 프로세스가 수행되는 동안 기능과 데이터가 분리되어 어떤 기능이 데이터를 참조하고 수정하는지 알기 어려운 문제점이 발생되었다. 따라서, 본 논문에서는 이러한 문제점들을 해결하고자 계획 및 분석단계, 설계단계, 구현 및 시험단계에서 UML을 적용한 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스 개발 프로세스를 제안하였다. 본 논문에서는 기존의 예산계획 DB, 예산편성 DB, 예산정산 DB의 데이터를 이용하여 사용자가 원하는 정보를 찾아주는 역할을 정보검색 에이전트에서 수행한다. 또한, 정보통합 에이전트에서는 정보검색 에이전트에서 수집한 데이터를 추출, 전송, 가공, 로딩하여 통합 데이터베이스에 저장한다. 결과적으로, 정보통합 에이전트에서는 다수의 정보소스를 사용자가 하나하나 접근하여 검사하는 노력을 줄여주고 사용자에게 불필요하다고 판단되는 데이터를 걸러주는 역할을 수행한다. 이러한 결과로 사용자의 요구사항을 최대한 반영하여 연구비 관리정책의 수립에 필요한 다양한 형태의 의사결정 지원정보를 제공할 수 있도록 하였다. 최종 사용자에게는 원하는 분석정보를 신속하게 접근하여 단편적인 관점보다는 종합적인 관점에서 다양한 분석자료를 제공받을 수 있도록 하였다. 또한, 3개의 시스템을 하나로 통합한 결과 데이터의 공유, 시스템 통합, 운영비용 절감, 의사결정 지원환경을 단순화시키는 효과를 제공하였다.

Integrated Management Data Warehouse Development Process of Research Expenses in Enterprise Environment

Seong-Man Choi[†] · Cheol-Jung Yoo^{††} · Ok-Bae Chang^{†††}

ABSTRACT

The existing management job of research expenses has been divided into three parts : budget planning, budget draw-up, and exact settlement of budget. However, it caused some problems. Under this current circumstance it is required to obtain research expenses steadily, to operate efficiently and to use them clearly to solve such problems. As a result of a study on data warehouse development process of existing system integration company (Inmon, IBM) to reflect current trend described above, data warehouse development process of Inmon uses systematic and gradual access as a classical development cycle method. It causes overlap and feedback to the previous step in the process of each step is requested. And another problem that it is difficult to tell what function refers and corrects data because functions and data are separated during performing development process at data warehouse development process of IBM is caused. Integrated management data warehouse development process of research expenses in the enterprise environment which applies UML at planning and analysis step, design step and implement and test step is suggested in this paper. Information retrieval agent uses existing budget plan DB, budget draw-up DB and budget settlement DB to find out information that a user wants to know. Information retrieval agent collects and saves information at integration database and information integration agent extracts, transports, transforms and loads the data. Information integration agent reduces a user's efforts to access to a number of information sources and check each of them. It also screens out data that a user may not need. As a result, integrated management data warehouse development process of research expenses in the enterprise environment reflects a user's requirements as much as possible and provides various types of information to make a decision which is needed to establish the policy of research expense management. It helps an end user approach his/her desired analysis information quickly and get various data from the comprehensive viewpoint rather than the fragmentary viewpoint. Furthermore, as it integrated three systems into one, it is possible to share data, to integrate the system, to reduce operating expenses and to simplify supporting environment for the decision making.

키워드 : UML, 엔터프라이즈 환경(Enterprise Environment), 연구비 통합관리(Research Expenses Integrated Management), 데이터 웨어하우스(Data Warehouse), 정보통합 에이전트(Information Integration Agent)

† 준 회 원 : 전북대학교 대학원 컴퓨터통계정보학과
†† 종신회원 : 전북대학교 컴퓨터학과 교수

††† 정 회 원 : 전북대학교 전자정보공학부 교수
논문접수 : 2003년 9월 2일, 심사완료 : 2003년 10월 31일

1. 서론

최근 네트워크와 클라이언트/서버 환경 및 소프트웨어 기술의 혁신적인 발전으로 인해 기업들간의 정보경쟁이 점점 가속화되고 있다. 이러한 시장변화로 인해 기업의 경영에 있어 데이터 처리에 보다 정확성과 신속성이 요구되어 데이터 웨어하우스(Data Warehouse)가 등장하게 되었다[1, 2]. W. H. Inmon에 의해 처음으로 제창된 데이터 웨어하우스는 데이터를 저장하고 필요시 가져오는 곳을 의미한다. 그 본질은 기업이 보유하고 있는 여러 가지 형태의 데이터를 차원 모델링(Dimensional Modeling)을 통해 정보화 또는 지식화하여 기업활동의 효율성을 증가시키기 위해 의사결정을 지원하는 정보체계라 할 수 있다[2, 3].

본 논문을 연구하게된 배경은 기존의 연구비 관리업무가 예산계획, 예산편성 및 예산정산 부분으로 관리됨으로써 사용자의 요구사항을 실시간으로 반영하지 못하며 또한, 연구비 예산정책과 관련된 의사결정 과정을 제대로 지원하지 못하였다. 이러한 문제점을 해결하고자 기존의 연구비 관리업무의 재사용성을 위한 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스 개발 프로세스를 제안하였다. 재사용성을 위해 연구비 관리 시스템의 이론적인 체계를 바탕으로 하여 데이터 웨어하우스의 주제 중심적이면서 통합적인 특징을 기반으로 삼았다. 본 논문의 연구내용은 비 정형화된 데이터들을 사용자가 원하는 정보로 찾아주는 정보검색 에이전트를 이용한다. 또한, 정보통합 에이전트에서는 정보검색 에이전트에서 수집한 데이터를 추출, 데이터 전송, 데이터 가공, 데이터 로딩을 할 수 있도록 하였다[4].

엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스를 개발한 결과로는 서로 다른 목적으로 운영되던 시스템을 통합하여 시스템의 총 운영비용을 줄이고 의사결정 지원환경을 단순화하여 복잡성을 감소시켰다. 또한, 의사결정에 필요한 데이터를 모두 데이터 웨어하우스에 저장하기 때문에 모든 사용자들로 하여금 시스템간의 이질성을 최소화하고 연구비 집행업무의 투명성을 부여하여 상호간의 유기적인 정보교환과 조직의 계획수립 및 분석적 업무를 효과적으로 지원할 수 있도록 하였다.

본 논문의 구성은 2장에서는 관련 연구로서 시스템 통합업체의 데이터 웨어하우스 개발 프로세스에 대해서 알아본다. 3장에서는 재사용성을 위한 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스 개발 프로세스에 대해서 설명한다. 4장에서는 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스의 실행과정 및 평가를 보여준다. 마지막으로 5장에서는 결론 및 향후 연구과제를 제시한다.

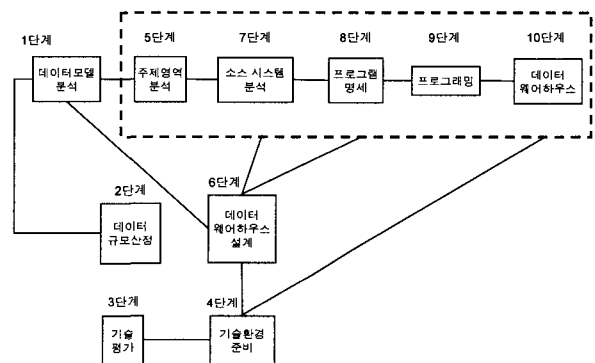
2. 관련 연구

본 장에서는 데이터 웨어하우스 개발 경험이 있는 시스

템 통합업체인 Inmon, IBM의 개발 프로세스 특징과 문제점에 대해서 알아본다[1, 5-7].

2.1 Inmon의 데이터 웨어하우스 개발 프로세스

(그림 1)에서 보는 개발 프로세스는 Inmon의 데이터 웨어하우스 개발 프로세스이다. 이 개발 프로세스에서 데이터 모델 분석에서는 주제영역 정의, 데이터 모델의 명확한 범위 정의, 과생 데이터에서 원시 데이터의 분리를 포함한다[5]. 데이터규모 산정에서는 DSS(Decision Support System : 의사결정 지원 시스템) 환경을 위한 데이터량을 추정하는 단계이다[8]. 즉, 데이터 웨어하우스에 얼마나 많은 데이터를 포함해야 하는지를 조사하기 위한 간단한 프로젝트이다. 기술 평가에서는 데이터 웨어하우스를 관리하기 위한 기술상의 요구사항으로 운영환경을 처리하고 관리한다. 기술 평가에서는 대량의 데이터를 관리할 수 있으며 데이터 모델에 따라 데이터를 조직화할 수 있는 능력을 가져야 한다. 기술환경 준비에서는 아키텍처 형태들을 어떻게 수용해야할 것인지를 기술적으로 정의하는 부분이다. 주제영역 분석에서는 함축적이면서도 의미가 있고 구현가능한 정도의 작은 규모로 선정한다. 데이터 웨어하우스 설계에서는 데이터 모델을 근거로 하여 데이터 웨어하우스를 설계한다. 소스 시스템 분석에서는 기존의 시스템 환경에서 해당 주제영역에 대한 소스 데이터를 정의하여 분석한다. 프로그램 명세서 작성에서는 운영환경과 DSS 환경간의 인터페이스가 개략적으로 가시화되면 이에 대한 프로그램 명세를 작성한다. 프로그래밍에서는 개발, 코드, 컴파일, 단위 테스트와 다양한 형태의 스트레스 테스트를 통한 프로그램 활동들을 포함한다. 구축에서는 지금까지 개발된 DSS 프로그램을 실행한다.

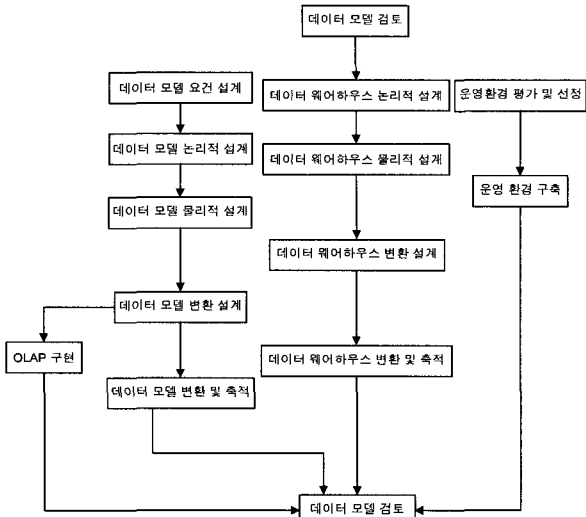


(그림 1) Inmon의 데이터 웨어하우스 개발 프로세스

(그림 1)의 개발 프로세스는 고전적인 개발주기 기법으로 단계적이며 순차적인 접근 방법을 이용한다. 실제로 (그림 1)의 개발 프로세스를 이용하여 시스템을 개발하다 보면 개발 단계에서 중복되어지는 부분이 나타나게 되어 각 단계의 진행과정에서 그 이전단계로 피드백(feedback)이 요청되는 심각한 문제를 유발하게 된다.

2.2 IBM의 데이터 웨어하우스 개발 프로세스

(그림 2)에서 보는 개발 프로세스는 기술 인프라와 전체 동적 확장 솔루션을 설계하였으며 표준 피드 및 인터페이스를 통해 기업의 트랜잭션 정보를 통합하는 방식이다. 이를 통해 에이전트들은 고객들에게 보다 빠르고 효과적인 조인을 제공할 수 있으며 고객들과 보다 밀접한 관계를 유지할 수 있게 되었다[5,6].



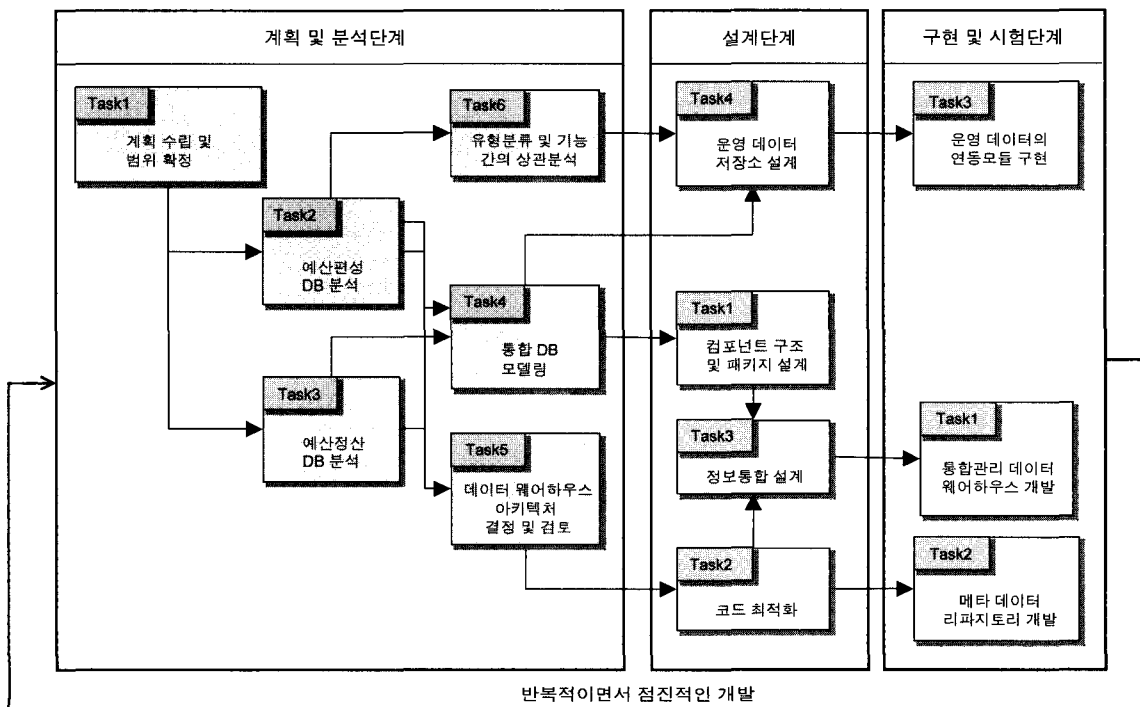
(그림 2) IBM의 데이터 웨어하우스 개발 프로세스

(그림 2)의 개발 프로세스는 인포믹스를 통해 우리가 언제 어디서든지 고객 정보를 액세스할 수 있다. 이로 인해

고객 서비스 측면에서 경쟁우위를 확보할 수 있도록 했을 뿐만 아니라 새로운 비즈니스 시스템이 기존의 보고서 작성시간을 크게 줄여 전체 효율성을 향상시킨다. 그러나, 수행되는 기능과 데이터가 분리되어 어떤 기능이 데이터를 참조하고 수정하는지 알기 어려워 유지보수하는데 많은 어려움이 나타난다.

3. 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스 개발 프로세스

본 논문에서 제안한 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스 개발 프로세스는 객체지향 분석 및 개발을 위한 UML을 적용하였다[9]. UML을 계획 및 분석, 설계, 구현에 적용하면 심각한 위험요소(기술, 요구사항, 목적, 유용성 등)를 초기에 완화시킨다. 또한, 초기의 피드백, 사용자 의견반영, 적용으로 실제 요구사항에 더욱 적합한 시스템을 구축할 수 있도록 해준다. 결과적으로 UML의 가장 큰 특징인 한번의 개발주기로 완성되기보다는 여러 개발주기를 통해 반복적이면서 점진적으로 개발하여 개발 프로세스 자체를 개선시키는 효과가 있다. 이렇게 함으로써 (그림 1)에서 제시된 개발주기의 단계적이며 순차적인 접근 방법으로 인한 문제점을 해결하고 (그림 2)에서 제시된 수행되는 기능과 데이터가 분리되어 유지보수하는데 많은 어려움이 발생하는 문제점을 해결한다. (그림 3)은 UML의 특징들이 적용된 개발프로세스로 반복적이면서 점진적으로 새로운 기능들을 추가하여 컴포넌트의 개발이 이루어지게



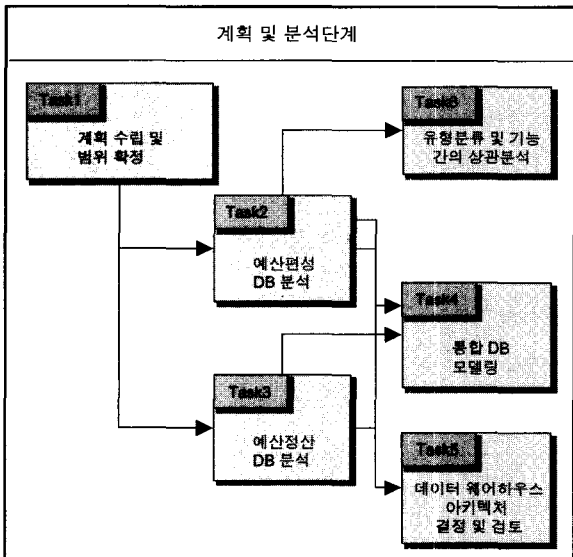
(그림 3) 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스 개발 프로세스

된다[10].

(그림 3)은 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스 개발 프로세스를 보여주고 있다. 연구비 통합관리 개발하기 위한 프로세스는 크게 계획 및 분석단계, 설계단계, 구현 및 시험단계로 이루어지며 각 단계별로 업무들이 정의되어 있다.

3.1 계획 및 분석단계

계획 및 분석단계에서는 어떠한 기술적인 사항이나 구현 사항으로부터도 자유로워야 하며 문제 도메인에 대한 필요한 지식을 얻고 해결하고자 하는 문제를 정의하는 단계이다[11, 12]. 계획 및 분석단계는 (그림 4)에서 보는 바와 같이 6개의 업무로 이루어져 있으며 각각의 업무들에 대한 정의 및 업무간 작업흐름을 보여주고 있다.



(그림 4) 계획 및 분석단계

3.1.1 계획수립 및 범위확정

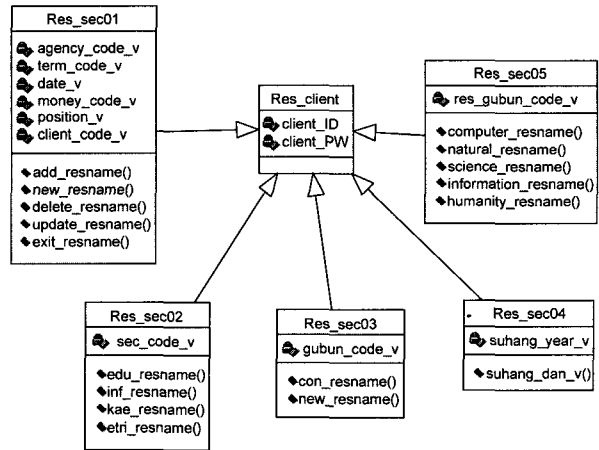
계획수립 및 범위확정 업무에서는 프로젝트 진행계획 및 개발범위를 확정하는 작업을 수행한다. 즉, 해결하고자 하는 문제가 무엇이며 시스템이 무엇을 해야하는지 등을 확정한다. 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스 개발범위로는 연구비 예산편성 DB와 예산정산 DB에 저장되어 있는 데이터만을 대상으로 하며 기존의 연구비 관리 정보체계를 활용한 추세분석 및 예측업무에 반드시 필요한 예산편성, 예산집행, 예산정산 데이터와 외부 데이터로 구축범위를 한정하였다. 즉, 연구비관리 정보분석으로 요구사항 분석(예산편성 클래스 다이어그램, 예산정산 클래스 다이어그램), 통합 데이터 클래스 다이어그램, 데이터 웨어하우스 아키텍처를 결정하기 위한 단계별 통합 클래스 다이어그램, 예산편성 통합 클래스 다이어그램, 예산정산 통합 클래스 다이어그램, 정보통합 설계, 통합관리 데

이터 웨어하우스 개발로 한정하였다.

3.1.2 예산편성 DB분석 및 예산정산 DB분석

예산편성 DB분석 및 예산정산 DB분석 업무에서는 요구사항을 명확하게 이해한 상태에서 데이터 요소들을 식별하고 요약 및 세분화 수준을 결정하며 데이터의 유효성이나 무결성을 점검한다. 즉, 기존의 예산편성 및 예산정산에 대한 요구사항을 분석하여 시스템에 존재하는 기본적인 엔티티에 대한 개념모델이나 중요한 용어에 대한 카탈로그를 작성한다[13]. 예산편성 DB분석 및 예산정산 DB분석의 결과로 (그림 5)와 (그림 6)의 클래스 다이어그램을 추출하였다.

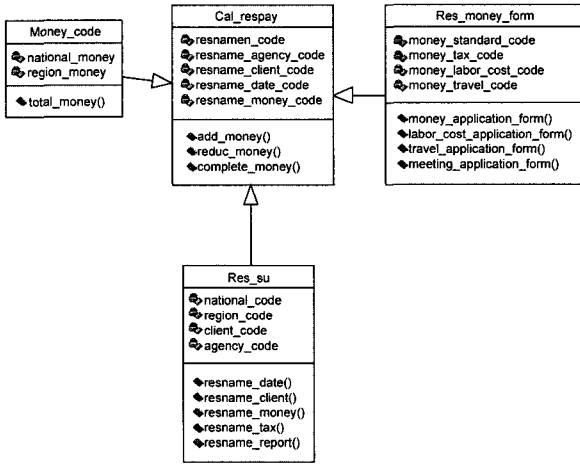
(그림 5)는 예산편성 클래스 다이어그램으로 이 클래스 다이어그램에는 6개의 클래스인 Res_client, Res_sec01, Res_sec02, Res_sec03, Res_sec04, Res_sec05 클래스로 구성된다. Res_Client는 연구과제를 수행하게 될 책임자의 ID와 PW를 속성으로 갖는다. Res_sec01에서는 연구과제의 발주기관, 수행기간, 연구비의 정보를 통해 연구과제의 종합적인 정보를 알 수 있다. Res_sec02에서는 연구과제의 발주기관을 구분하는 정보를 가지며 Res_sec03에서는 연구과제가 신규인지 아니면 계속과제인지를 구분한다. Res_sec04에서는 기존의 연구과제를 수행한 년도와 일자의 정보를 가지며 Res_sec05에서는 연구과제가 어느 분야에 속하는지를 알려준다.



(그림 5) 예산편성 클래스 다이어그램

(그림 6)은 예산정산 클래스 다이어그램으로 이 클래스 다이어그램에는 4개의 클래스인 Cal_respay, Money_code, Res_su, Res_money_form 클래스로 구성된다. Cal_respay에서는 연구비를 계산하기 위해서 연구과제 코드, 연구과제 발주기관 코드, 연구수행자 코드, 연구과제 수행일자 코드, 연구비 코드 속성을 갖는다. Money_code에서는 연구비의 지급방식에 따라 국비, 지방비, 연구비 총액의 정보를 가지며 Res_su에서는 국가기관 코드, 지방기관 코드, 연구수행자

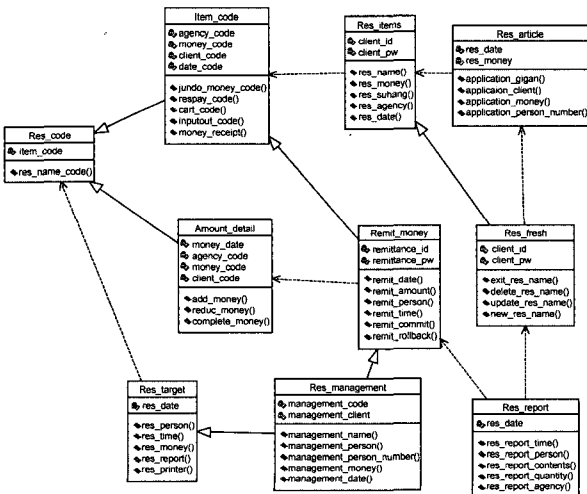
코드, 발주기관 코드의 정보를 가지면서 연구과제가 어느 기관에서 발주한 과제인지를 확인할 수 있다. Res_money_form에서는 연구비 집행기준 코드, 연구비 세금 코드, 연구비 인건비 코드, 연구비 여비 코드의 정보를 가지고 있으면서 연구비를 집행할 때 참고하도록 하였다.



(그림 6) 예산정산 클래스 다이어그램

3.1.3 통합 DB 모델링

통합 DB 모델링 업무에서는 일반적으로 기업 모델링으로 불리는 단계로 예산관련 데이터를 취합하여 하나의 데이터 모델로 만들어가는 과정이다. 외부 데이터에 대해서도 별도의 모델링 과정을 거쳐 데이터를 추가로 입력한다. 즉, 예산편성 DB분석 및 예산정산 DB분석 결과를 통합하여 통합 DB 모델링을 작성한다. (그림 7)은 예산편성 클래스 다이어그램 및 예산정산 클래스 다이어그램을 통합한 결과를 보여주고 있다. 통합 데이터 클래스 다이어그램은 Res_code, Item_code, Res_Items, Res_article, Amount_detail, Remit_money, Res_fresh, Res_target, Res_management, Res_report



(그림 7) 통합 데이터 클래스 다이어그램

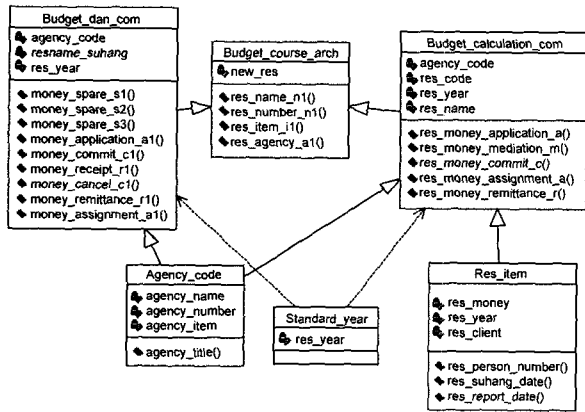
클래스로 이루어져 있으며 예산편성 클래스 다이어그램 및 예산정산 클래스 다이어그램 외에 연구과제를 수행하는데 있어 필요한 내용을 추가하였다. Res_code에서는 연구항목에 대한 코드와 연구과제의 제목에 대한 코드를 갖도록 하였으며, Item_code에서는 연구과제의 기관, 금액, 책임자, 연구기관의 정보를 가지도록 하였다. 또한, Res_items와 Res_article, Res_target, Res_management, Res_report, Res_fresh에서는 연구과제에 대한 책임자 ID, PW 및 연구과제에 대한 연구기간, 금액, 신규여부, 지속된 과제, 연구비 신청정보, 연구비 관리정보, 연구비 지출결과에 대한 정보를 각각 가지고 있다. Amount_detail과 Remit_money에서는 연구비 지출에 대한 정산정보를 가지고 있다.

3.1.4 데이터 웨어하우스 아키텍처 결정 및 검토

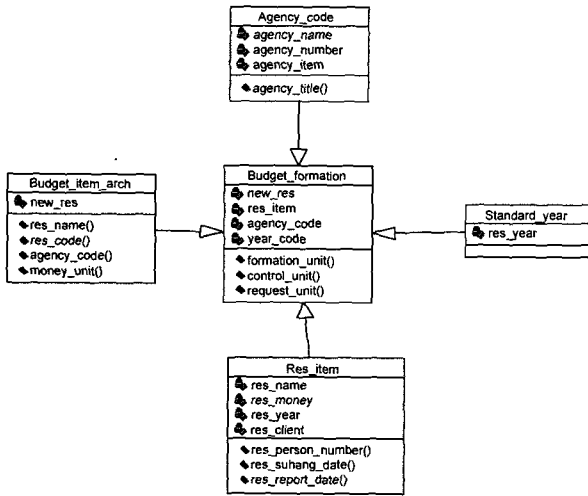
데이터 웨어하우스 아키텍처 결정 및 검토 업무에서는 예산편성 DB 분석 및 예산정산 DB 분석 결과를 토대로 개발하고자 하는 데이터 웨어하우스의 아키텍처를 결정하는 업무를 수행한다. 데이터 웨어하우스 아키텍처를 결정하기 위해서 단계별 통합 클래스 다이어그램, 예산편성 통합 클래스 다이어그램, 예산정산 통합 클래스 다이어그램을 산출하였다. 3개의 통합 클래스 다이어그램은 사용자의 다양한 요구사항을 만족시키는 복잡한 질의를 신속하게 처리할 수 있도록 설계하였다. (그림 8)은 단계별 통합 클래스 다이어그램으로 Budget_dan_com, Budget_course_arch, Budget_calculation_com, Agency_code, Standard_year, Res_item 클래스로 구성된다. Budget_dan_com에서는 기존의 클래스에서 추출된 연구비 집행기관 코드, 연구과제 수행정보, 연구과제 년도의 정보를 가지고 있다. Budget_course_arch에서는 연구과제에 대한 신규여부에 대한 정보를 확인할 수 있다. Budget_calculation_com에서는 기존의 클래스에서 추출된 연구기관 코드, 연구과제 코드, 연구과제 년도, 연구과제 제목의 정보를 가지고 있다. Agency_code에서는 연구과제 기관명, 기관 고유번호, 기관소속에 대한 정보를 가지고 있다. Standard_year에서는 연구과제의 해당년도에 대한 정보를 가지고 있다. Res_item에서는 연구과제에 대한 제목, 연구금액, 연구년도, 연구책임자에 대한 정보를 가지고 있다. 결과적으로, 단계별 클래스 다이어그램은 의사결정 지원시스템에서 분석하고자 하는 주제를 중심으로 하여 예산편성 DB 및 예산정산 DB를 비교 분석한 결과를 통합하여 의사결정에 필요한 데이터만을 추출할 수 있도록 하였다.

(그림 9)는 예산편성 통합 클래스 다이어그램으로 Budget_formation, Agency_code, Budget_item_arch, Res_item, Standard_year 클래스로 구성된다. Budget_formation에서는 Agency_code, Budget_item_arch, Standard_year, Res_item의 클래스를 통합한 정보를 가지고 있다. Budget_item_arch에서는 신규과제에 대한 과제명, 과제코드, 기관코드, 연구금액에 대한 정보를 가지고 있다. Agency_code에서는 연구

기관명, 연구기관번호, 연구기관소속에 대한 정보를 가지고 있다. Standard_year에서는 연구과제의 수행년도에 대한 정보를 가지고 있으며, Res_item에서는 연구과제 수행인원에 대한 정보를 가지고 있다. 결과적으로, 예산편성 통합 클래스 다이어그램에서는 연구비를 기준으로 하여 편성된 금액과 최종 확정된 금액을 년도별로 분석할 수 있게 설계하였다.



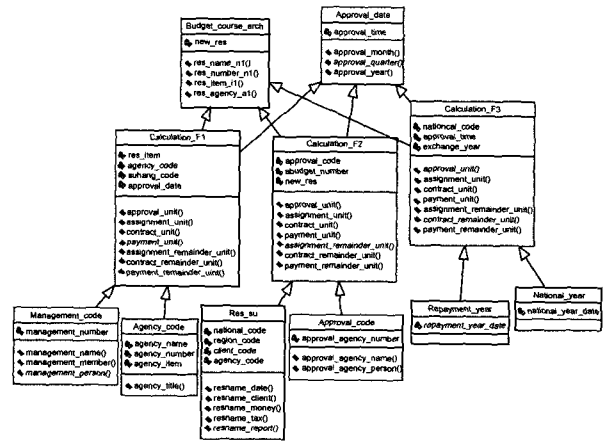
(그림 8) 단계별 통합 클래스 다이어그램



(그림 9) 예산편성 통합 클래스 다이어그램

(그림 10)은 예산정산 통합 클래스 다이어그램으로 Budget_course_arch, Approval_date, Calculation_F1, Calculation_F2, Calculation_F3, Management_code, Agency_code, Res_su, Approval_code, Repayment_year, National_year 클래스로 구성된다. Budget_course_arch에서는 연구과제의 신규 및 지속여부에 대한 정보를 가지고 있다. Approval_date에서는 연구비 신청 및 지출에 대한 승인정보를 가지고 있다. Calculation_F1, Calculation_F2, Calculation_F3에서는 연구과제에 대한 연구비 신청 및 승인, 지불, 세금, 배정에 대한 정보를 가지고 있다. Management_code, Agency_code에서는 연구비 관리기관 번호, 관리기관명, 관리기관 부서, 관리기관 담당자에 대한 정보를 가지고 있다. Res_su, Ap-

proval_code에서는 연구비의 지출기관에 대한 정보와 연구비 지출승인에 대한 정보를 가지고 있다. Repayment_year, National_year에서는 당해연도의 환수금액에 대한 정보를 가지고 있다. 결과적으로, 예산정산 통합 클래스 다이어그램에서는 실제 신청한 연구비에 대한 집행 및 정산결과를 보여줄 수 있도록 설계하였다. 또한, 예산편성, 예산승인, 관리기관, 승인코드 등을 참고하여 데이터를 추출할 수 있도록 하였다.

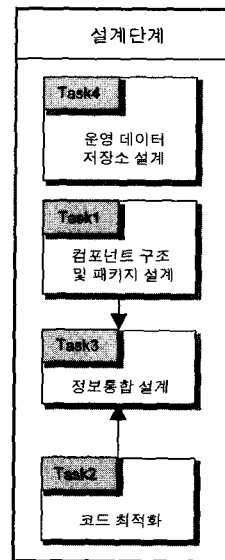


(그림 10) 예산정산 통합클래스 다이어그램

3.1.5 유형분류 및 기능간의 상관분석

유형분류 및 기능간의 상관분석 업무에서는 예산관리 정보체계에 대한 추이분석과 예측분석을 위하여 주제를 설정하고 이와 관련된 데이터를 정의하여 목적과 성격에 적합한 분석업무를 수행한다.

3.2 설계단계



(그림 11) 설계단계

설계단계는 4개의 업무들로 구성되어 있으며, 4개의 업무들이 (그림 11)에서 보는 바와 같이 반복적으로 수행된다. 설계단계는 계획 및 분석결과에 대한 기술적인 확장 및 적용 단계이다. 또한, 기술적인 모든 세부사항과 구현환경에 대한 제약조건을 고려해야 한다[14, 15].

3.2.1 컴포넌트 구조 및 패키지 설계

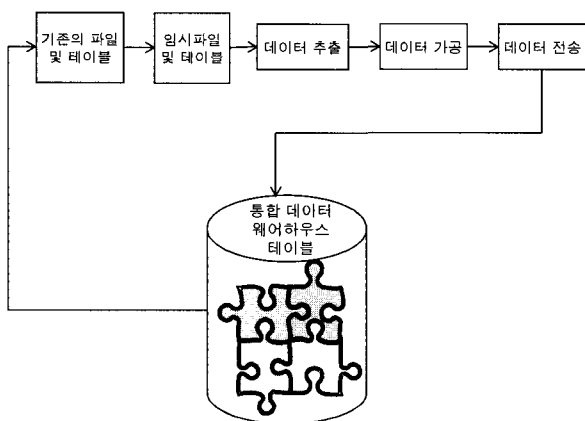
컴포넌트 구조 및 패키지 설계 업무에서는 이전 업무에서 정의한 통합 DB 모델링을 기반으로 예산편성 DB분석과 예산정산 DB분석에서 식별된 컴포넌트들에 대한 컴포넌트 인터페이스를 명확히 기술하고 각각의 컴포넌트를 설계한다. 또한, 패키지 설계업무에서는 각각에 해당하는 전개파일(deployment descriptor)의 생성과 모듈들을 하나로 묶어 주는 역할을 한다.

3.2.2 코드 최적화

코드 최적화 업무에서는 최종 실행 프로그램이 보다 적은 기억장소를 사용하여 보다 빠르게 작업을 처리할 수 있도록 주어진 환경에서 최상의 명령어 코드를 사용하여 작업을 수행한다.

3.2.3 정보통합 설계

정보통합 설계 업무에서는 정보통합 에이전트를 통해 데이터의 추출, 데이터 전송, 데이터 가공, 데이터 로딩시키는 과정의 업무를 수행한다. 본 논문에서는 일정한 시기에 운영 시스템에서 데이터를 추출하여 데이터 웨어하우스로 보내는 방식을 선택하였다. 또한, 데이터 정제작업은 데이터 웨어하우스의 사실 테이블로 가는 중간단계 저장소인 ODS (Operational Data Store)를 이용하였다. ODS의 역할은 소스 데이터를 가지고 있기 때문에 데이터의 가공 및 정제작업을 쉽게 해준다. 이러한 결과로 사용자의 요구사항 변경에 신속하게 대응할 수 있을 뿐만 아니라 사실 테이블에 문제가 발생하였을 경우 신속하게 복구할 수 있도록 편리성을 제공해준다.



(그림 12) 정보통합 설계과정

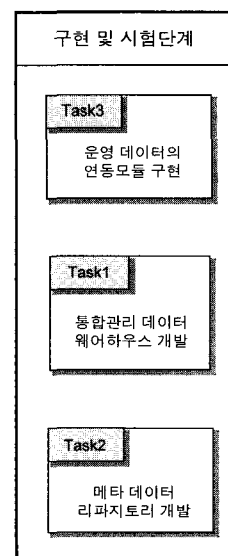
(그림 12)는 정보통합 설계과정으로서 기존의 파일에서 데이터를 데이터 웨어하우스 서버의 임시 테이블에 로드하여 논리적인 데이터 검증을 실시한다. 에러가 제거되면 임시 테이블의 데이터를 통합 데이터 웨어하우스의 테이블에 가공 및 정제하여 로드한다. 이러한 작업을 마치면 오라클 익스프레스의 MDM에 데이터를 로드하기 위하여 임시요소 테이블을 가공 및 생성한다. 마지막으로 임시요소 테이블에서 오라클 익스프레스 MDM으로 데이터를 로드한다. 이러한 결과로 사용자의 요구사항이나 사실 테이블의 변화에 신속하게 대처하도록 하였다. 또한, 통합 데이터 웨어하우스에서부터 사실 테이블을 구성하기 위해 데이터를 가공하여 의사결정에 필요한 분석자료들을 질의할 수 있도록 하였다.

3.2.4 운영 데이터 저장소 설계

운영 데이터 저장소 설계 업무에서는 유형분류 및 기능간의 상관분석 결과와 통합 DB 모델링을 통해 운영 데이터 저장소에 대한 아키텍처를 설계한다.

3.3 구현 및 시험단계

구현 및 시험단계는 3개의 업무들로 구성되어 있으며, 3개의 업무들이 (그림 13)에서와 같이 반복적으로 수행된다. 구현 및 시험단계에서는 정보통합 에이전트에서 데이터를 추출, 전송, 가공, 로딩하는 정보통합 작업을 수행하여 저장고에 사용자가 직접 접근하여 다차원적 질의를 통해 대화식으로 정보를 분석하여 구현 및 시험을 수행한다[15, 16]. 구현단계에서는 설계에 대한 최종적인 결정을 내리고 다이어그램과 명세서를 프로그래밍 언어의 구문으로 전환하는 실질적인 코딩을 수행한다. 이러한 구현단계가 종료되면 시험단계가 이루어진다. 시험단계의 목적은 코드에 존재하는 에



(그림 13) 구현 및 시험단계

를 찾는 것이다. 시험단계는 목적에 따라 단위시험 및 통합시험 단계로 수행한다. 단위시험 단계에서는 설계단계에서 설계지침을 기반으로 클래스들의 인터페이스, 자료구조, 경계조건, 제어흐름 및 오류처리 등에 대한 다양한 시험을 수행한다. 단위시험을 수행하고 나면 통합시험을 진행한다. 통합시험에서는 설계단계에서 발생한 오류를 발견하며 이러한 오류를 수정한 후 개발된 컴포넌트가 요구사항을 만족하는지를 확인한다.

3.3.1 통합관리 데이터 웨어하우스 개발

통합관리 데이터 웨어하우스 개발 업무에서는 운영 데이터베이스로부터 데이터를 추출하여 가공한 다음에 데이터 웨어하우스의 사실 테이블로 가는 중간단계 저장소인 ODS에 저장한다. 이렇게 함으로써 사용자의 요구사항이나 사실 테이블의 변화에 신속하게 대처하도록 구성하였다. 또한, 통합 데이터베이스(IDB)로부터 사실 테이블을 구성하기 위해 데이터를 가공하여 통합 데이터베이스에 로딩하여 의사결정에 필요한 분석자료들을 질의할 수 있도록 하였다.

연구비 통합관리 데이터 웨어하우스 구축에 요구되는 개발업무 분야는 크게 두 분야로 나누어진다. 연구비 관리 정보체계에 대한 다양한 측면의 분석과 기존의 연구비 관리 정보체계를 활용한 추세분석 및 예측으로 이루어진다. 시스템 개발업무는 데이터 웨어하우스의 OLAP(On-Line Ana-

lytic Processing) 기술을 이용하고 기존의 연구비 관리 정보체계를 활용한 추세분석 및 예측업무는 특정 예산항목에 대한 증감추이 등으로 분석한다[6, 7]. 또한, 미래의 예측치를 추정할 수 있는 예측모형을 개발함으로써 정확한 연구비 예산편성 지침을 작성할 수 있도록 하였다.

(그림 14)는 본 논문에서 제안한 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스의 모델이다. 정보검색 에이전트를 통해 기존의 데이터인 예산계획 DB, 예산편성 DB, 예산정산 DB의 데이터를 이용한다. 정보검색 에이전트에서는 사용자가 원하는 정보를 찾아주는 역할을 수행하며 검색엔진을 이용한다. 검색엔진은 검색로봇(search robot), 인덱스(index), 질의서버(query server)로 구성된다. 정보통합 에이전트에서는 데이터 추출, 데이터 전송, 데이터 가공, 데이터 로딩을 통합 데이터베이스의 ODS(Operational Data Store)에 저장하게 된다. 다음으로 업무성격에 알맞은 구조의 데이터 마트에 데이터를 추출 및 저장하고 OLAP 기술과 데이터 마이닝 기술을 이용하여 분석업무와 추세분석 및 예측업무를 수행하게 한다. 또한, 정보통합 에이전트에서는 다수의 정보소스를 사용자가 하나하나 접근하여 검사하는 노력을 줄여주고 각 정보 사이트에서 사용자에게 불필요하다고 판단되는 것을 걸러주는 역할을 수행한다.

3.3.2 메타 데이터 리파지토리 개발

메타 데이터 리파지토리 개발 업무에서는 대용량의 데이터가 저장되어 있는 리파지토리에 사용자가 직접 접근하여 다차원적 질의를 통해 대화식으로 정보를 분석하고 탐색하는 메타 데이터 리파지토리를 개발한다.

3.3.3 운영 데이터의 연동모듈 구현

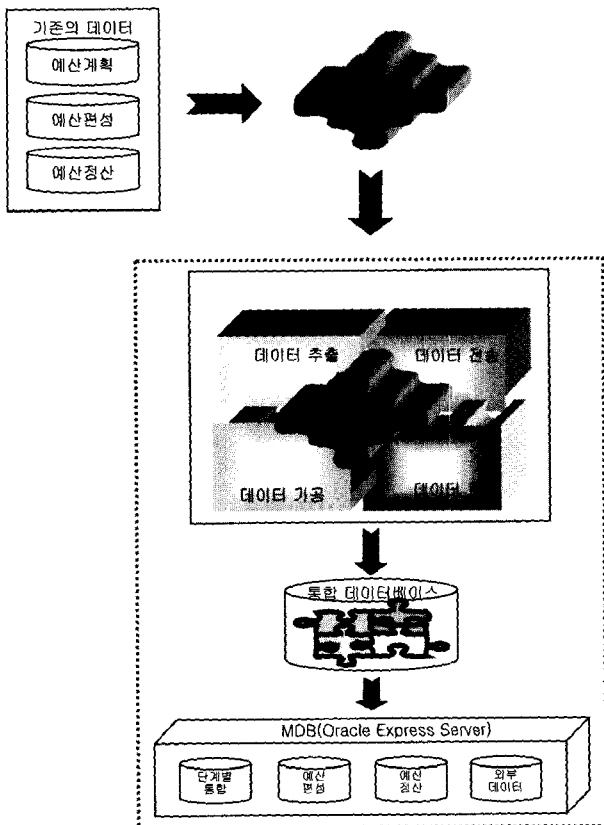
운영 데이터의 연동모듈 구현 업무에서는 목적과 성격에 적합한 복수의 데이터 마이닝 기법을 선택하여 연동모듈을 구현한다.

4. 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스

본 장에서는 UML로 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스 실행과정 및 평가에 대해서 알아본다.

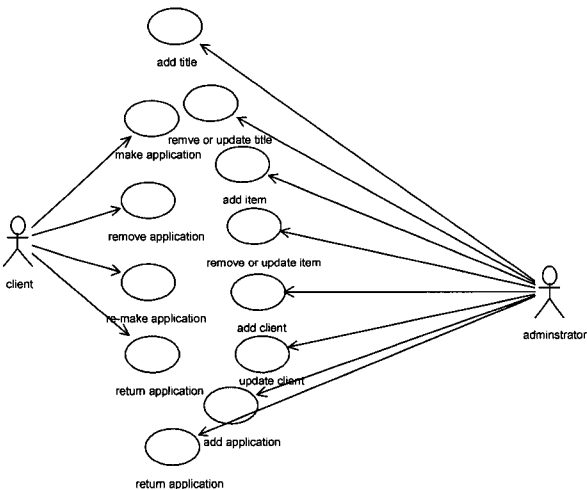
4.1 UML로 표현한 연구비 통합관리 데이터 웨어하우스 실행과정

연구비 통합관리 데이터 웨어하우스에 대한 요구사항 명세들을 수집하여 공통된 기능들을 추출하면 (그림 15)와 같은 유스케이스 다이어그램을 산출할 수 있다. 연구비 통합관리 데이터 웨어하우스의 유스케이스 다이어그램은 시스템, 서브시스템 또는 클래스의 행위를 모델링하는데 중요한



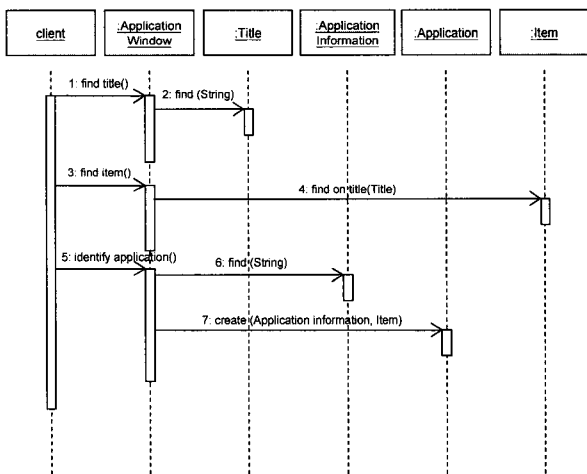
(그림 14) 연구비 통합관리 데이터 웨어하우스 모델설계

역할을 수행한다. (그림 15)에서의 유스케이스 다이어그램은 시스템의 정적인 유스케이스 뷰를 모델링하는데 이용된다[10, 16]. 클라이언트와 관리자간의 공통 기능들은 사용자 번호, 비밀번호, 인증, 과제등록, 과제검색, 연구비 신청, 연구비 신청취소, 과제입력, 결제, 결제확인, 결제취소, 연구비 입력, 연구비 입금, 입금확인 등이 있다.



(그림 15) 연구비 통합관리 데이터 웨어하우스의 유스케이스 다이어그램

유스케이스 다이어그램을 모델링한 결과를 가지고 시간 순서에 의한 제어흐름을 모델링하기 위하여 (그림 16)과 같은 시퀀스 다이어그램을 생성한다[10, 16].

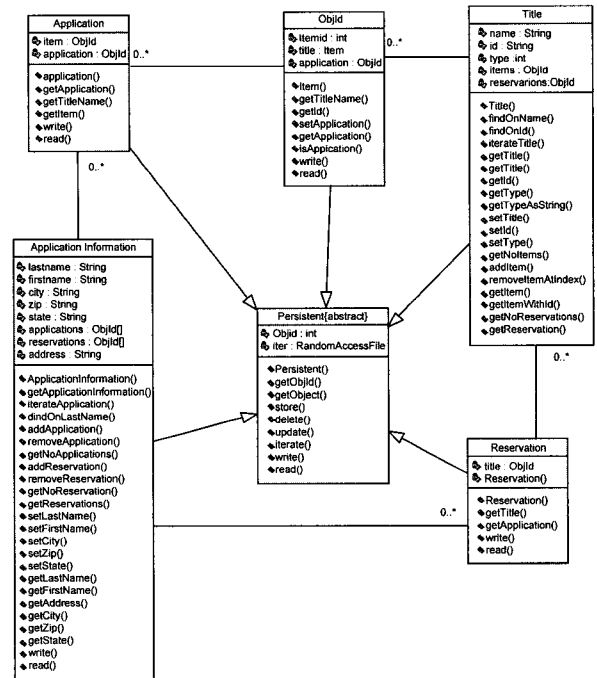


(그림 16) 연구비 통합관리 데이터 웨어하우스의 시퀀스 다이어그램

(그림 16)은 연구비 통합관리 데이터 웨어하우스의 시퀀스 다이어그램으로 유스케이스 시나리오에 의한 동적 행위의 표현방법을 메시지의 시간순서로 강조한 다이어그램이다. 이러한 시퀀스 다이어그램은 일련의 객체와 이들 객체 간에 송수신되는 메시지를 보여준다. 연구비 통합관리 데이

터 웨어하우스의 시퀀스 다이어그램을 이용함으로써 사용자는 시스템의 동적인 뷰를 쉽게 설명할 수 있으며 간단하게 반복과 분기를 표현하는데 매우 효과적이다.

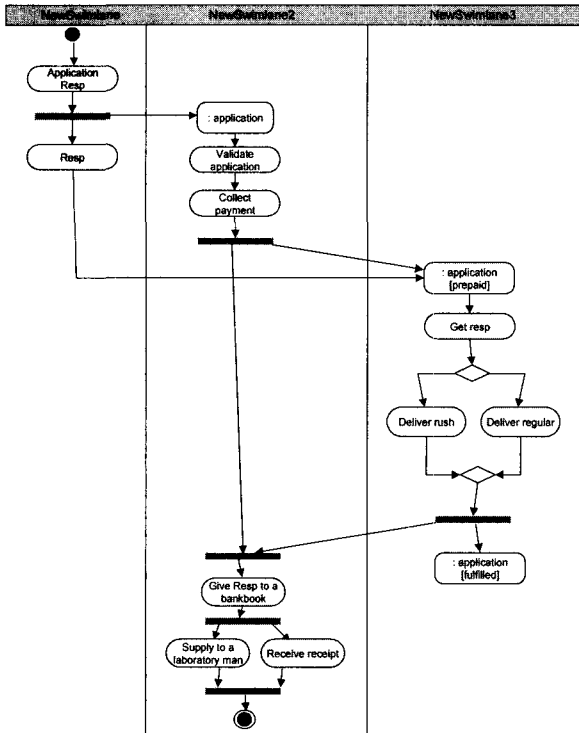
(그림 17)은 연구비 통합관리 데이터 웨어하우스의 클래스 다이어그램이다. 통합관리 데이터 웨어하우스는 Application, Item, Title, Application Information, Persistent(abstract), Reservation 클래스로 구성된다. Application에서는 연구과제에 대한 정보를 입력 및 출력할 수 있도록 해준다. Item에서는 연구과제의 ID와 과제명에 대한 정보를 가지고 있다. Title에서는 연구과제명을 입력할 수 있는 기반정보를 가지고 있다. Application Information에서는 수행기관에 대한 정보를 가지고 있으며, Reservation에서는 연구비 신청에 대한 지출승인, 보류, 취소 등에 대한 정보를 가지고 있다. Persistent(abstract)에서는 각 클래스에서 가지는 고유정보를 통합한 정보를 가지고 있다. 따라서, 연구비 통합관리 데이터 웨어하우스의 클래스 다이어그램에서는 일련의 클래스, 인터페이스 및 협동과 그들간의 종속, 일반화, 연관 관계를 보여준다. 또한, 객체지향 시스템의 모델링에서 가장 많이 작성하는 다이어그램으로 사용자에게 정적인 설계 뷰를 설계하는데 유용하다. 이러한 클래스 다이어그램을 통해 컴포넌트 다이어그램과 배치 다이어그램을 산출할 수 있다.



(그림 17) 연구비 통합관리 데이터 웨어하우스의 클래스 다이어그램

(그림 18)은 연구비 통합관리 데이터 웨어하우스의 액티비티 다이어그램으로 활동상태와 행동상태, 변환, 객체를 포함하며 다른 다이어그램과 마찬가지로 노트와 제약조건

을 갖는다. (그림 18)에서의 액티비티 다이어그램은 시스템의 동적인 사항을 모델링하는데 이용하며 비즈니스 프로세스 및 업무흐름을 가시화 해준다.



(그림 18) 연구비 통합관리 데이터 웨어하우스의 액티비티 다이어그램

4.2 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스 평가

본 논문에서 제안한 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스는 기존의 업무환경이 계획, 편성, 정산부분이 별개로 관리됨으로 인해 발생하는 문제점을 해결하고 연구비의 효율적인 운영과 기존의 데이터를 통한 향후 의사결정을 지원하기 위한 시스템이다. 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스 개발 프로세스와 기존의 개발 프로세스와의 차이점은 Inmon의 데이터 웨어하우스 개발 프로세스의 문제점인 개발 단계에서 중복되어지는 부분이 각 단계의 진행과정에서 그 이전단계로 피드백이 요청되는 문제를 해결하였고, IBM의 데이터 웨어하우스 개발 프로세스의 문제점인 수행되는 기능과 데이터가 분리되어 어떤 기능이 데이터를 참조하고 수정하는지 알기 어려워 유지보수하는데 많이 시간이 소요되는 문제점을 하나의 통합데이터베이스로 관리하여 이러한 문제점을 해결하였다. 또한, 기존의 시스템에서 각각 분리된 데이터베이스 시스템을 하나로 통합한 데이터 웨어하우스로 개발한 이점은 다음과 같다.

- 데이터의 공유 : 엔터프라이즈 환경의 연구비 통합관리 데

이터 웨어하우스는 의사결정에 필요한 데이터를 모두 데이터 웨어하우스에 저장하기 때문에 모든 사용자들은 데이터를 공유할 수 있다.

- 시스템 통합 : 기존의 정보 시스템들은 서로 다른 목적으로 인해 각각 구축되었기 때문에 통합하기가 매우 어려웠다. 그러나 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스는 이러한 시스템을 통합하는 문제를 해결하였다.
- 운영비용 절감 : 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스는 과거에 비효율적으로 운영되던 시스템들을 최적화함으로써 시스템의 총 운영비용을 줄여주는 효과가 있다.
- 의사결정 지원환경의 단순화 : 기존의 의사결정 지원환경이 각각의 애플리케이션마다 종속적이어서 관리하기가 복잡하였다. 그러나 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스를 통해 단일 의사결정 지원환경을 제공하기 때문에 이러한 복잡성을 감소시켰다.

기존의 개발 프로세스와 본 논문에서 제안한 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스와의 비교항목들을 통해 비교한 차이점을 <표 1>에서 보여주고 있다.

<표 1> 개발 프로세스들간의 비교

프로세스 비교 항목	Inmon의 데이터 웨어하우스 개발 프로세스	IBM의 데이터 웨어하우스 개발 프로세스	엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스 개발 프로세스
구축 내용	구축	구축 + 운영	구축 + 운영
구축 범위	주제 영역	주제 영역	주제영역 + 전사적 (enterprise)
구축 위험	상	하	하
모델링 관점	데이터관점 + 조직관점	데이터 관점	데이터관점 + 프로세스 관점
도구 사용	종속적	종속적	독립적
피드백	무	무	유

5. 결론 및 향후 연구과제

사용자의 요구사항을 최대한 반영하고 연구비 예산정책과 관련된 의사결정 과정을 최대한 지원할 수 있는 시스템이 절대적으로 필요하게 되었다. 이러한 현실적인 상황을 반영하고자 연구비의 예산편성 및 예산집행의 효율성을 높이고자 엔터프라이즈 환경의 연구비 통합관리 데이터 웨어하우스를 개발하였다.

본 논문의 연구내용은 데이터 웨어하우스 개발 경험이 있는 시스템 통합업체 Inmon, IBM의 데이터 웨어하우스 개

발 프로세스 과정에서 발생하는 문제점을 해결하고자 통합 관리 데이터 웨어하우스 개발 프로세스를 제안하여 분석 및 설계, 구현단계에서 객체지향 분석 및 설계를 지원하는 UML을 이용하였다. UML을 이용한 결과 Inmon의 개발 프로세스에서 발생된 문제점과 IBM의 개발 프로세스에서 발생된 문제점을 해결하였다. 또한, 통합관리 데이터 웨어하우스로 개발하기 위하여 설계단계에서 정보검색 에이전트와 정보통합 에이전트를 이용하였다. 정보검색 에이전트를 통해 기존의 데이터인 예산계획 DB, 예산편성 DB, 예산정산 DB의 데이터를 이용하여 사용자가 원하는 정보를 찾아주는 역할을 수행하도록 하였다. 또한, 정보통합 에이전트에서는 정보검색 에이전트에서 수집한 데이터를 추출, 데이터 전송, 데이터 가공, 데이터 로딩을 통합 데이터베이스의 ODS에 저장한다. 또한, 정보통합 에이전트에서는 다수의 정보소스를 사용자가 하나하나 접근하여 검사하는 노력을 줄여주고 각 정보 사이트에서 사용자에게 불필요하다고 판단되는 것을 걸러주는 역할을 수행한다.

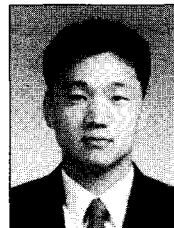
이러한 결과로 연구비 계획단계에서부터 예산편성 및 예산집행, 예산정산까지 독립적으로 관리하고 있는 운영 시스템의 데이터베이스들을 최적화하였다. 또한, 통합관리 데이터 웨어하우스로 개발하여 각 시스템간의 이질성을 최소화하고 연구비 집행업무의 투명성을 부여하여 상호간의 유기적인 정보교환과 조직의 계획수립 및 분석적 업무를 효과적으로 지원할 수 있도록 하였다. 이로 인해 최종 사용자가 원하는 분석정보에 신속하게 접근하여 단편적인 관점보다는 종합적인 관점에서 다양한 분석자료를 제공받을 수 있도록 하였다. 또한, 업무의 생산성 향상과 효율적인 연구비 예산 정책수립에 기여할 뿐만 아니라 관련정보를 사용자들 사이에서 공유할 수 있도록 하였다.

향후 연구과제로는 자기탐지 기능을 통해 통합관리 데이터 웨어하우스 시스템을 검증하여 성능이 떨어지는 곳을 찾아 성능을 개선할 수 있는 인공지능 개념을 추가하여 연구할 필요가 있다.

참 고 문 헌

[1] Eric Sperley, "The Enterprise Data Warehouse : Planning, Building, and Implementation," Prentic Hall PTR, Inc., 1999.
 [2] Nielson Debevoise, Thomas Debevoise, "The Data Warehouse Method," Prentice Hall PTR, Inc., 1999.
 [3] Larissa Terpeluk Moss, Sid Adelman, "Data Warehouse Project Management," Addison-Wesley Longman, Inc., 2000.
 [4] 최중민, "인터넷 정보 추출 에이전트", 정보과학회지, 제18권 제5호, pp.48-53, 2000.
 [5] Barry Devlin, "Data Warehouse : From Architecture to

Implementation," Addison-Wesley Longman, Inc., 1996.
 [6] M. C. McCabe, D. Grossman, "The Role of Tools in Development of a Data Warehouse," in Proceedings of Assessment of Software Tools, pp.139-145, 1996.
 [7] Aluizio Haendchen Filho, Hercules A. Prado, Simao S. Toscani, "Evolving a Legacy Data Warehouse System to an Object-Oriented Architecture," in Proceedings of Computer Science Society, pp.32-40, 2000.
 [8] Surajit Chaudhuri, Umesbwar Dayal, VenKatesh Ganti, "Database technology for decision support systems," in Computer, Vol.34, Issue.12, pp.48-55, 2001.
 [9] Taku Fujii, Yahiko Kambayashi, "An Application of Data Warehouse Technology to the Measurement System for UML based Artifacts," in Proceedings of Conference Computer Software and Applications, pp.459-467, 2001.
 [10] 최성만, 김송주, 유철중, 장옥배, 이정열, "재공학 환경에서 적용성 향상을 위한 디자인 패턴의 UML 표현", 정보과학회 봄 학술발표 논문집, 제30권 제1호, pp.148-150, 2003.
 [11] Hans-Erik Eriksson, Magnus Penker, "UML Toolkit," John Wiley & Sons, Inc., 1998.
 [12] 조은숙, 김수동, 류성열, "UML 기반의 객체지향 프레임워크 모델링 기법", 정보과학회논문지(B), 제26권 제4호, pp. 533-545, 1999.
 [13] Young Jong Yang, Soon Yong Kim, Gui Ja Choi, Eun Sook Cho, Chul Jin Kim, Soo Dong Kim, "A UML-based Object-Oriented Framework Development Methodology," in Proceedings of Conference Software Engineering, pp. 211-218, 1998.
 [14] Joseph D. Schulz, Craig Larman, "Requirements-Based UML," in Proceedings of Technology of Object-Oriented Languages and Systems, pp.307-316, 2001.
 [15] Craig Larman, "APPLYING UML AND PATTERNS : An Introduction to Object-Oriented Analysis and Design and the Unified Process," Prentice Hall PTR, Inc., 2002.
 [16] Rational Software Corp., "Unified Modeling Language (UML) Summary," 1997.



최 성 만

e-mail : sm3099@jnue.ac.kr
 1999년 전주대학교 전자계산학과(이학사)
 2003년 전북대학교 대학원 전산통계학과
 멀티미디어 소프트웨어 전공(이학석사)
 2003년~현재 전북대학교 대학원 컴퓨터
 통계정보학과 박사과정

1997년~1999년 정인대학 사무자동화과 조교
 관심분야 : 소프트웨어공학, CBD, CBSE, UML 기반 모델링,
 객체지향 프로그래밍, 디자인 패턴, 데이터 웨어하우스,
 소프트웨어 개발 프로세스



유철중

e-mail : cjyoo@chonbuk.ac.kr

1982년 전북대학교 전산통계학과(이학사)

1984년 전남대학교 대학원 계산통계학과
(이학석사)

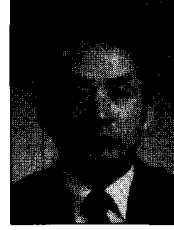
1994년 전북대학교 대학원 전산통계학과
(이학박사)

1982년~1985년 전북대학교 전자계산소 조교

1985년~1996년 전주기전여자대학 전자계산과 전임강사~부교수

1997년~현재 전북대학교 자연과학대학 컴퓨터과학과 전임강사
~부교수

관심분야 : 소프트웨어 개발 프로세스, 소프트웨어 품질, 컴포넌트 소프트웨어, 소프트웨어 매트릭스, 소프트웨어 에이전트, GNSS, GIS, 교육공학, 인지과학



장옥배

e-mail : okjang@chonbuk.ac.kr

1966년 고려대학교 수학과(이학사)

1973년 고려대학교 수학과(이학석사)

1974년~1980년 조지아 주립대 박사과정
수료

1988년 산타바바라대 대학원(Ph. D.)

1980년~현재 전북대학교 공과대학 전자정보공학부 교수

관심분야 : 소프트웨어공학, 전산교육, 수치해석, 인공지능