

# CMMI의 형상관리를 적용한 데이터웨어하우스 개발 프로세스의 개선

박 종 모<sup>†</sup> · 조 경 산<sup>\*\*</sup>

## 요 약

운영 서버의 데이터에서 다양한 분석 정보를 추출하여 저장하는 의사결정지원시스템인 데이터웨어하우스는 데이터의 품질과 대용량의 데이터를 처리하기 위한 처리 시간이 매우 중요하다. 따라서 데이터의 품질 안정화와 생산성 향상을 위해 개발 프로세스를 표준화하고 개선할 필요가 있다. 본 연구에서는 소프트웨어 프로세스 향상모델인 CMMI의 형상관리를 적용하여 개선된 데이터웨어하우스 개발 프로세스를 제안한다. 또한 개선된 개발 프로세스를 평가하기 위해 프로세스 평가 척도를 제시하고, 기존의 개발 프로세스와 비교하여 본 연구의 제안이 처리시간 감소에 따른 비용의 절감과 생산성의 향상을 지원하고, 품질의 향상과 재작업비용을 개선시켰음을 보인다.

**키워드** : 데이터웨어하우스, CMMI, 형상관리, 소프트웨어 프로세스 개선

## Improvement of Datawarehouse Development Process by Applying the Configuration Management of CMMI

Jongmo Park<sup>†</sup> · Kyungsan Cho<sup>\*\*</sup>

### ABSTRACT

A Datawarehouse, which extracts and saves the massive analysis data from the operating servers, is a decision support tool in which data quality and processing time are very important. Thus, it is necessary to standardize and improve datawarehouse development process in order to stabilize data quality and improve the productivity. We propose a novel improved process for datawarehouse development by applying the configuration management of CMMI (Capability Maturity Model Integration) which has become a major force in software development process improvement. In addition, we specify some matrices for evaluating datawarehouse development process. Through the comparison analysis with other existing processes, we show that our proposal is more efficient in cost and productivity as well as improves data quality and reusability.

**Key Words** : Datawarehouse, CMMI, Configuration Management, Software Process Improvement

### 1. 서 론

최근의 소프트웨어 개발 프로세스는 정보시스템의 기술혁신에 따라 대상 범위가 확대되고 규모가 거대해지는 반면에, 짧은 시간 내에 정보시스템을 개발해야 하는 등으로 수행환경이 크게 변화하고 있다. 이에 따라 단순히 소프트웨어 개발의 종료만을 목표로 하지 않고, 개발 프로세스를 표준화하고 실패확률을 낮추며 품질 향상과 비용절감을 해야 할 필요성이 제시되었다. 프로세스란 어떤 일을 하기 위해 정의된 순서가 있는 활동으로 사람이 도구를 사용하여 주어진 절차를 수행하는 작업을 의미한다[1]. 프로세스에 초점을 맞

추면 소프트웨어 제품, 소프트웨어 개발 및 유지보수, 작업 공정 등의 성공적 관리가 가능해지고, 지속적으로 프로세스를 개선함으로써 소프트웨어 제품의 품질과 생산성 및 개발 기간과 예측성 등을 향상시킬 수 있다.

프로세스의 개선을 위해서는 프로세스에 대한 측정이 필요하다. 측정은 프로세스의 수행 과정에서 계획과의 편차를 발견하는 기초가 되며 프로세스 개선을 위한 기회를 제공한다. 프로세스 측정의 과정 및 목적은 다음과 같다[2]. 첫째, 프로세스의 성과를 측정하기 위해 데이터를 수집한다. 둘째, 각 프로세스의 성과를 분석한다. 셋째, 프로세스의 안정성과 분석결과를 해석한다. 넷째, 미래의 비용 및 성과를 예상하기 위해 데이터를 사용하고 유지한다.

프로세스의 개선을 위한 모델로는 품질관리에 중점을 둔 6시그마와 소프트웨어 프로세스 향상에 중점을 둔 SPICE(Software

<sup>†</sup> 정 회 원 : 단국대학교 대학원 정보컴퓨터학과 박사과정

<sup>\*\*</sup> 종신회원 : 단국대학교 정보컴퓨터학부 교수

논문접수 : 2006년 4월 20일, 심사완료 : 2006년 6월 1일

Process Improvement and Capability dEntermination), 능력 성숙도 모델인 CMMI(Capability Maturity Model Integration) 등이 제시되었다. 프로세스 개선 모델은 소프트웨어를 개발하는 조직의 생산성과 품질의 향상을 위한 업무의 형식을 제공하며, 소프트웨어 개발능력의 측정과 소프트웨어를 평가하는 기준을 제공한다. 소프트웨어 프로세스 평가 및 개선을 위한 모델로 SEI(Software Engineering Institute)에서 제시한 CMMI는 소프트웨어에 대한 성숙도를 평가하는 CMM(Capability Maturity Model)을 확장하여 소프트웨어뿐만 아니라 시스템의 설계 및 운영 등 시스템 통합 전반에 대한 성숙도를 평가한다[3].

본 연구에서는 데이터웨어하우스를 개발하는 프로세스에 CMMI의 형상관리(Configuration Management)를 적용하여 데이터웨어하우스 개발 프로세스의 개선안을 제시하고자 한다. 데이터웨어하우스는 운영 시스템으로부터 생성된 데이터에서 다양한 분석 정보를 추출하여 저장하는 정보의 저장고이며, 의사결정 지원시스템인 OLAP(On-Line Analysis Process)에 사용된다[4, 5]. 데이터웨어하우스는 의사결정에 직접적인 영향을 미치기 때문에 데이터의 품질이 매우 중요하다. 부정확하게 가공된 데이터는 잘못된 의사결정을 가져올 수 있으므로 데이터의 품질이 낮을 경우 데이터를 정제하는 과정이 반드시 필요하다. 또한 대규모의 데이터를 처리하기 때문에 잘못 생성된 데이터를 다시 처리하기 위해서는 많은 시간과 비용이 소요되어 생산성과도 밀접하게 관련된다.

그러나 기존의 데이터웨어하우스 개발 프로세스는 상황에 따른 프로세스와 경험에 의존하였으며, 데이터의 품질 안정화를 위한 표준화되거나 개선된 프로세스는 제시되지 않았다. 프로젝트 규모가 커지고 다양화되면서 지속적으로 품질과 생산성을 향상시키기 위해서는 프로세스 개선 모델의 적용이 필요하다. 프로세스의 개선을 위해 본 논문에서는 실제 구축된 데이터웨어하우스를 기반으로 데이터의 품질을 안정화하고 데이터 처리에 필요한 시간과 비용을 절감하기 위하여 형상관리를 적용한 개선된 프로세스를 제안한다. 또한, 제안된 프로세스를 위한 측정치를 제시하여 측정치에 대한 분석 결과를 도출한다.

본 논문의 구성은 다음과 같다. 2장에서는 데이터웨어하우스와 CMMI의 성숙도 레벨에 대한 관련 연구를 제시하고, 3장에서는 A사에서 적용한 기존의 대표적인 데이터웨어하우스 개발 프로세스를 설명한다. 4장에서는 데이터의 품질 및 생산성 향상을 위해 기존의 데이터웨어하우스 개발 프로세스에 CMMI의 형상관리를 적용한 개선방안을 제안하며, 제안된 프로세스의 성능이 기존의 프로세스와 비교하여 향상되었음을 보인다. 5장의 결론으로 본 논문을 마무리한다.

## 2. 관련연구

본 장에서는 본 연구에서 개선할 개발 프로세스의 대상인 데이터웨어하우스와 개선에 적용할 CMMI에 관한 기존 연구를 분석 제시한다.

### 2.1 데이터웨어하우스

데이터웨어하우스는 사용자들에게 다양한 분석정보를 편리하게 제공하기 위하여 분석의 기준이 되는 차원들과 분석할 내용에 해당하는 측정값들로 구성된 다차원 형태의 데이터구조를 가진다[4, 5]. 데이터웨어하우스를 이용하면 사용자가 대화식으로 정보를 분석할 수 있는 의사결정 지원 시스템인 OLAP질의에 빠른 응답을 할 수 있다.

데이터웨어하우스의 설계를 위해 이용되는 스타 스키마는 사실 테이블을 중심으로 차원 테이블들이 마치 별 구조의 구조를 갖는다[6]. 스타 스키마는 다차원 데이터를 관계 데이터베이스의 테이블을 이용하여 저장하면서, 비교적 단순한 형태로 표현하기 때문에 사용자들이 쉽게 이용할 수 있다는 장점이 있다. [7]에서는 데이터웨어하우스의 설계 방법론을 다음과 같은 단계로 제시하였다. 첫째, 운영 데이터베이스로부터 수집된 정보를 분석하고 사용자의 요구 사항을 명세한 단계를 거친 후 개념적 설계를 통해 사실과 차원 구조를 생성한다. 둘째, 개념적 설계 결과에 대한 작업량과 유효성을 판단한 후 테이블 구조로 실체화하기 위한 논리적 설계 단계를 거치고, 마지막으로 물리적 설계를 수행한다.

데이터웨어하우스 응용 분야로 온라인 기업의 고객 관리 및 마케팅에 중요한 정보로 활용되는 웹 로그 분석을 추가한 데이터웨어하우스가 생성되었다. [8]에서는 성공적인 데이터웨어하우스의 개발을 위하여 고객 세분화와 고객 데이터의 정비, 서비스 기획이 정비가 필요하다고 정의했으며, 이를 통해 웹 로그에 대한 고객 분석 업무가 향상된다고 분석하였다.

하지만 데이터웨어하우스의 개발은 상황과 경험에 의존하여 진행되어 왔으며, 표준화된 프로세스가 존재하지 않았다. 따라서 지속적인 품질과 생산성을 향상 시키기 위해 프로세스의 표준화와 개선 모델의 적용이 필요하다.

### 2.2 CMMI

CMMI는 프로세스를 평가하고 개선하기 위한 능력 성숙도 모델로서 발전되어 왔다. CMMI는 SW-CMM이 소프트웨어와 관련된 프로세스 개선 활동만을 지원하던 취약점을 보완하여 소프트웨어, 시스템, 하드웨어, 서비스와 관련된 모든 프로세스 개선활동을 지원한다. 또한 검증된 실무 활동을 반영하여 조직의 성숙도 및 공정 능력 평가, 공정 향상을 위한 활동의 우선순위 결정, 실제 공정 향상을 위한 구현활동을 지원하는 프레임워크로 구성된다.

CMMI는 성숙도 평가를 수행하기 위해서 단계표현(Staged Representation)과 연속표현(Continuous Representation)의 2가지 표현방법을 제공한다[9]. 단계표현이란 조직 차원의 성숙도 수준을 평가하는 것으로, 각 성숙도 단계별로 달성해야 할 핵심적인 프로세스 영역(PA: Process Area)이 정해져 있고, 프로세스 영역을 모두 달성해야만 더 높은 수준의 성숙도로 진전할 수 있도록 적용하는 방법이다. 연속표현이란 단일 프로세스 영역별로 성숙도 수준을 평가하는 것으로 모든 프로세스 영역별로 별도의 레벨이 있고, 이들을 특별한 그룹으로 묶지 않고 개별적으로 측정하여 개선할 수 있도록

〈표 1〉 CMMI 표현방법의 비교

구분	연속표현 (Continuous Representation)	단계표현 (Staged Representation)
개선범위	사업목표에 맞도록 필요 부분 개선	개선단계에 입각해 순차적 개선
비교범위	프로세스 영역을 중심으로 비교	성숙도 레벨을 중심으로 비교
효과	필요한 부분만 개선하여 적은 비용을 투입하여 큰 효과	개선단계에 대한 등급을 재공하여 조직 전체의 개선 및 조직 간의 비교가 가능
구현전략	EIA/IA-731(SE-CMM)을 사용한 조직의 경우 접근 및 전환이 용이	여러 조직에서 SW-CMM의 적용/평가를 통한 경험 지식과 노하우 등을 이용 가능

〈표 2〉 프로세스 영역의 연속표현

구분	프로세스 영역
프로세스 관리영역	1. OPF (Organizational Process Focus)
	2. OPD (Organizational Process Definition)
	3. OT (Organizational Training)
	4. OPP (Organizational Process Performance)
	5. OID (Organizational Innovation & Deployment)
프로젝트 관리영역	1. PP (Project Planning)
	2. PMC (Project Monitoring & Control)
	3. SAM (Supplier Agreement Management)
	4. IPM (Integrated Project Management)
	5. RSKM (Risk Management)
	6. IT (Integrated Team)
	7. ISM (Integrated Supplier Management)
	8. QPM (Quantitative Project Management)
엔지니어링영역	1. RD (Requirements Development)
	2. REQM (Requirements Management)
	3. TS (Technical Solution)
	4. PI (Product Integration)
	5. VEL (Verification)
	6. VAL (Validation)
지원영역	1. CM (Configuration Management)
	2. PPQA (Process & Product Quality Assurance)
	3. MA (Measurement & Analysis)
	4. OEI (Organizational Environment for Integration)
	5. DAR (Decision Analysis & Resolution)
	6. CAR (Casual Analysis & Resolution)

적용하는 방법이다. 두 가지 표현방법은 <표 1>과 같이 분석된다.

연속표현은 프로세스 영역을 4개의 주요한 범주로 나누어 관리하고 있으며, 각 프로세스들의 관계는 <표 2>와 같이 제안되었다[3]. 프로세스 관리영역은 프로젝트가 안정적으로 진행되도록 조직이 표준 프로세스를 보유하고, 개별 프로젝트는 필요에 따라 표준 프로세스를 조정하여 사용한다.

〈표 3〉 프로세스 측정의 척도

구분	범주	척도명	산출 공식	
일정	프로세스	계획공정 준수율	(실행 공정수/계획 공정수)*100	
		공정 진도율	(실제 진도/계획 진도)*100	
비용	M/M	계획 투입공수 준수율	(실제 투입공수/계획 투입공수)*100	
		비용 예산 준수율	(집행예산/계획예산)*100	
생산성	생산성	공정별 생산성	분석: 요구사항 수/투입공수 설계: 설계항목 수/투입공수	
품질	품질	요구사항	요구사항 변경율	(변경된 요구사항 수/최초 요구사항 수)*100
		위험발생 비율	발견대비 결함 제거율	(실현된 위험 수/파악된 위험 수)*100 (제거된 결함 수/발견된 결함 수)*100

〈표 4〉 개선 사례

회사	비용	일정	생산성	품질
Boeing, Australia	결함 수정: 33% ↓	50% ↓	NA	NA
Lockheed Martin MDS	결함 수정: 15% ↓ SW개발: 20% ↓	NA	30% ↑	NA
Lockheed Martin SI		NA	70% ↑	50%의 결함 ↓
Raytheon North Texas	하드웨어: 25억 ↓	일정예측: 20% ↑	NA	NA

↑:향상(개선) ↓:단축(절감) \* NA: Not Available

프로젝트 관리영역은 프로젝트 계획을 수립하고 유지하는 활동 및 계획 대비 진척을 모니터링하고 통제하는 활동을 포함한다. 엔지니어링 영역은 제품 및 서비스의 개발 및 유지보수 활동을 포함한다. 지원 영역은 조직 전반의 지원활동을 포괄하고, 조직의 의사소통의 명확성과 효율성을 지원한다.

[10]에서는 CMMI의 프로세스 영역 중에서 MA(Measurement & Analysis)를 기반으로 측정 프로그램을 수립하고, 개발된 측정 프로그램을 통해 CMMI의 각 레벨 별로 측정 프로그램을 적용하였다. [11]에서는 <표 3>과 같이 CMMI의 핵심 프로세스들에 대하여 정량적인 측정을 위한 척도들을 정의하였다. 이 척도를 통해 프로세스 능력 수준이 결정되고, 지속적인 측정값 수집을 통한 조직의 프로세스 개선 예측이 가능하므로 자원의 효과적인 분배와 문제발생에 대한 조기 대응이 가능하게 되어 프로젝트 성공확률을 증대시킬 수 있다.

프로세스 개선 시에 요구사항 관리나 품질문제와 같이 작업환경에 문제가 생겼을 때 형상관리를 통해 담당자가 문제 해결을 위해 관여하게 된다. 즉, 개발전반에 걸쳐 형상항목을 식별하고 정의해서 문제의 발생시 관련 형상항목이 변경되어가는 것에 대해 기록하고 리포팅하고 완전성과 무결성을 보장해 개발관리 활동에 대한 연구가 진행되었다[12].

프로젝트 관리 측면에서는 CMMI와 PMBOOK(Project Management Body of Knowledge)을 통합하려는 연구가 있었으며[13], 품질관리 측면에서는 ISO 9001:2000과 CMMI 모델에 대한 통합이 제시되었다[14]. 또한 SEPG 컨퍼런스에

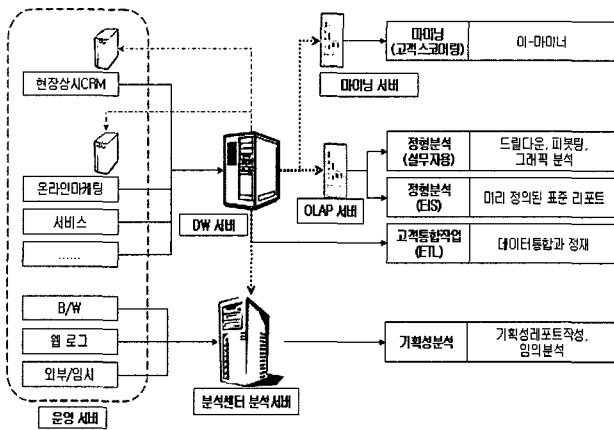
서 <표 4>와 같이 CMMI를 적용하여 프로세스를 개선한 대표적인 사례들이 보고되고 있으며, 개선 효과는 결함 수정 및 개발비용의 절감과, 개발 기간의 단축, 일정 예측 준수로 인한 경쟁력이 확보, 제품의 품질 개선으로 인한 실패 비용이 절감 등으로 나타나고 있다.

### 3. 데이터웨어하우스 개발 프로세스

본 장에서는 기존의 대표적인 데이터웨어하우스 개발 프로세스를 소개하고, 이를 적용한 A사의 데이터웨어하우스의 개발 프로세스를 보인다. A사는 다양한 채널을 통해 온라인 및 오프라인으로 상품과 제품을 판매하는 회사이다. 이에 따라 각종 채널에서 수집된 회원의 통합이 필요하고, 각 채널 별로 제품의 모델과 결제정보의 분석이 필요하다.

#### 3.1 개발 환경과 구현 시스템

(그림 1)은 기존의 데이터웨어하우스 개발 프로세스를 적용한 A사의 시스템 구조도이다. 운영 서버에서 분석에 필요한 데이터를 추출하여 DW(DataWarehouse) 서버와 분석센터 분석서버에 데이터를 적재한다. DW 서버의 데이터를 이용하여 마이닝 서버에서 고객 스코어링을 위한 기초자료를 생성하고, OLAP 서버에서 실무자용과 경영자용을 위한 정형분석을 수행하며, 분석센터 분석서버에서 기획성 분석을 위해 데이터를 가공한다. 구현 시스템의 중점 사항은 운영 서버에서 고객통합작업을 위해 데이터를 추출하여 적재하는 데이터웨어하우스 개발 프로세스이다.



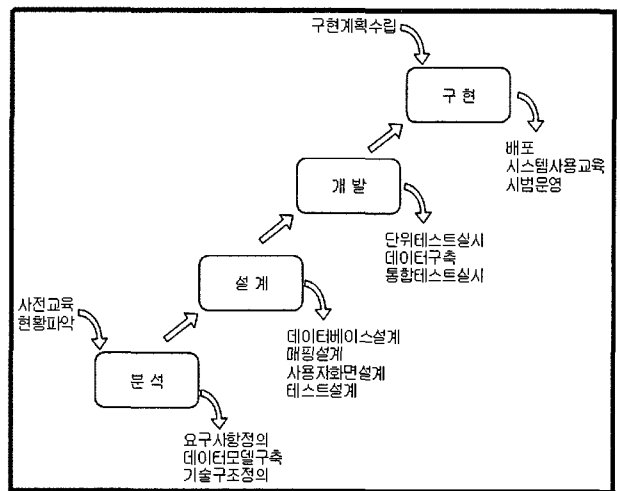
(그림 1) 시스템 구조도

#### 3.2 개발 프로세스

A사의 데이터웨어하우스 개발 프로세스의 목적은 효과적이며 신속 정확한 의사결정 지원시스템을 개발하고, 고객 데이터의 추출 및 변환, 적재와 더불어 고객 데이터 정제 및 통합을 수행하는 것이다. 이를 위한 분석 보고서의 내용은 고객세분화와 연관분석, 상품별 고객분석 업무의 지원이다. 개발 프로세스의 범위는 (그림 2)와 같다.

고객분석 업무수행	고객분석 정형화 지표정의	고객분석 인프라구축
<ul style="list-style-type: none"> <li>○ 정기 고객분석업무                             <ul style="list-style-type: none"> <li>. 기존 모델링 성능향상</li> </ul> </li> <li>○ 기획성 분석업무                             <ul style="list-style-type: none"> <li>. 고객DB 품질분석</li> <li>. 온/오프라인 고객정보 분석</li> <li>. 상관/제품 분석</li> <li>. 마케팅 성과분석</li> <li>. 온라인 판매분석</li> </ul> </li> <li>○ 임의/수시 분석업무                             <ul style="list-style-type: none"> <li>. 비정기적 리포트 작성</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>○ 고객 분석장표 정의                             <ul style="list-style-type: none"> <li>. 업무 지원용 고객 분석장표 정의</li> <li>. 고객기존정보, 고객구매패턴, 이벤트, KPI 등 분석장표 정의</li> <li>. 분석 시스템 포지션 정의 (구축목적, 기대효과, 발전방안)</li> <li>. 주요 분석장표별 활용 방안 작성</li> <li>. 데이터 품질 개선 방안 작성</li> </ul> </li> <li>○ 논리적 데이터 모델링                             <ul style="list-style-type: none"> <li>. 고객 중심의 통합 데이터 모델링</li> </ul> </li> <li>○ 고객 데이터 분석                             <ul style="list-style-type: none"> <li>. 고객분류 개선안 도출</li> <li>. 가구단위 정의 기준 도출</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>○ 고객 데이터웨어하우스 구축                             <ul style="list-style-type: none"> <li>. 고객 데이터 관리 구축</li> <li>. 고객통합 이력개발</li> <li>. 고객정보 공유체계 구축</li> </ul> </li> <li>○ 고객분석 화면 개발                             <ul style="list-style-type: none"> <li>. 일반 사용자용 OLAP 화면 구현</li> <li>. 관리자용 EIS 구축</li> <li>. 통합 타게팅 설계</li> <li>. 분석가용 Ad Hoc-Query 등 사용자별 권한 및 정보보안 강화</li> </ul> </li> </ul>

(그림 2) 개발 프로세스 범위



(그림 3) 데이터웨어하우스 개발 프로세스

A사의 개발 프로세스는 분석, 설계, 개발, 구현의 4 단계를 거치며 각 단계에서의 처리 내용은 (그림 3)과 같다.

분석 단계의 요구사항정의서에서는 사용자와 개발자 사이의 합의에 의해 포함되어야 할 기본조건과 시스템화해야 할 범위를 정의한다. 설계 단계에서는 Erwin을 사용하여 다차원 분석이 가능하도록 사실 테이블과 차원 테이블로 구성된 데이터웨어하우스를 설계한다. 설계된 모델을 기반으로 운영 서버에서 데이터웨어하우스로 데이터를 추출하고 정제하고 적재한다. 개발 단계에서 테스트 후 결함보고서를 작성하며, 요구사항정의서에서 정의된 허용범위 내의 결함이 있을 경우만 통합 테스트를 진행한다. 구현 단계에서는 전체적인 일정을 고려하여 사용자들에게 배포한다.

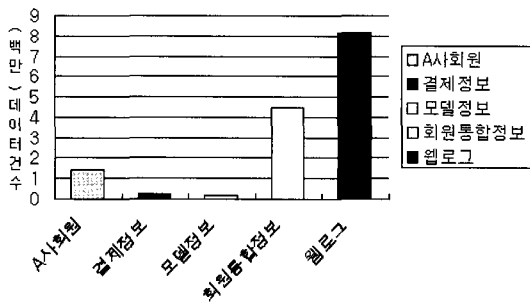
데이터웨어하우스 개발 단계에서 처리하는 데이터의 양은 매우 방대하다. 따라서 한 번 잘못 생성된 데이터를 다시 처리하기 위해서는 많은 시간과 비용이 소요된다. 한 번에 데이터를 올바르게 처리하기 위해서는 데이터의 품질이 보장되어야 한다. 그러나 제시된 데이터웨어하우스 개발 프로세스에는 데이터의 품질 안정화를 위한 최적화된 프로세스가 존재하지 않았다. 따라서 체계적이고 표준화된 프로세스의 필요성이 증대되었고, 지속적으로 품질과 생산성을 향상시키기 위해 프로세스 개선 모델을 적용할 필요가 있다.

#### 4. 데이터웨어하우스 개발 프로세스의 개선 제안 분석

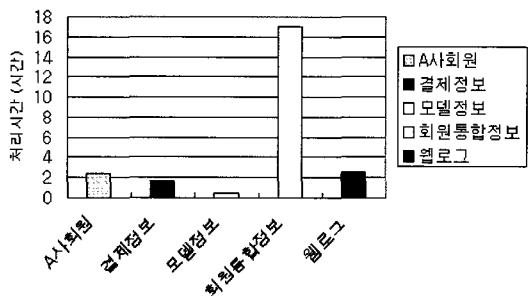
##### 4.1 개발 프로세스의 개선 제안

A사의 경우에 데이터웨어하우스 개발 과정에서 (그림 4)와 같이 방대한 양의 데이터가 적재되며, 해당 데이터를 처리하는 시간은 (그림 5)와 같다. (그림 4)와 같이 회원정보와 결제정보, 그리고 각 제품의 모델정보는 적은 데이터가 적재된다. 그러나 여러 채널에서 발생한 회원을 통합하여 처리하는 통합정보와 웹로그에서 발생하는 정보는 상대적으로 많은 데이터가 발생한다. 데이터 적재량과 비교하여 데이터 처리시간에서 웹로그의 경우는 단순 로그만을 처리하기 때문에 처리시간이 적게 소요되나, 회원통합의 경우는 회원정보를 취합하고 오류를 검출하며, 각 채널의 회원에 대해 다른 규칙으로 발생된 데이터의 정확성을 맞추어야 하기 때문에 비교적 긴 시간이 소요된다. 따라서 데이터를 정확한 방식으로 처리하지 못할 경우 데이터를 재처리하기 위해서는 막대한 비용과 시간이 소요된다. 특히 고객 분석을 통해 의사결정을 할 경우 지연된 데이터 처리 작업으로 인해 적시에 의사결정을 하지 못하는 경우도 발생한다.

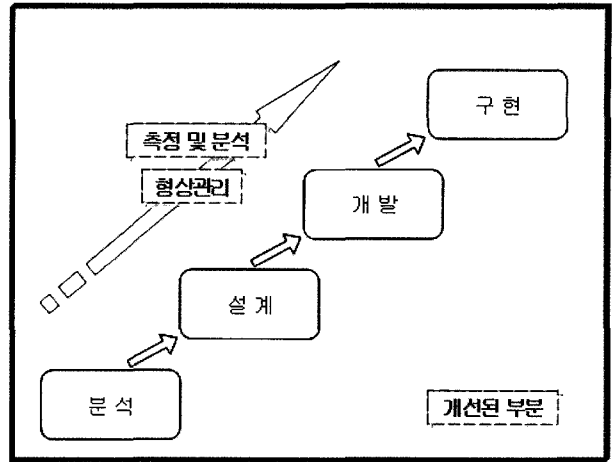
CMMI를 이용하여 (그림 3)에서 제시된 프로세스를 분석하면 데이터 적재에 따른 시간을 절약하고 품질을 향상시키기 위해 형상관리의 적용이 필요하고, 이에 대한 측정 및 분석이 요구된다. 즉 형상관리를 통해 데이터의 품질을 보증하고 생산성을 향상시키며 동시에 프로세스 개발 과정을 통제할 수 있다. 또한 관리자는 프로세스의 진행사항 및 문제점을 측정할 수 있고, 의사결정의 기반이 되는 품질 목표를 가진다. 이로 인해 품질 목표 달성이 점차 정확해지고



(그림 4) A사의 데이터 적재량



(그림 5) A사의 데이터 처리 시간



(그림 6) 개선된 데이터웨어하우스 프로세스

프로세스의 변동폭은 줄어든다. 이러한 관점에서 본 연구가 제안하는 개선된 데이터웨어하우스의 개발 프로세스는 (그림 6)과 같으며, 개발 프로세스의 전 단계에 걸쳐 형상관리를 수행하고 성능 향상을 위해 제시된 측정치를 통해 지속적인 측정 및 분석이 이루어진다.

본 연구에서 제시하는 형상관리가 필요한 근거는 개발과 구현 단계에서 개발된 데이터웨어하우스가 끊임없이 변화하기 때문이다. 즉, 사용자의 요구사항의 변화, 사용자의 관점의 변화 또는 프로세스 주변 환경이 변화이다. 변화된 내용에 대해 형상관리를 추가하면 프로세스 개발과정에 대한 추적이 가능하고, 개발이 완료된 후에도 지속적인 유지보수가 용이하다는 장점이 있다. 형상관리를 위한 측정은 설계와 개발과정중에 발생한 문제들을 명확하게 설명할 수 있고, 측정이 적절하게 수행되면 프로세스 개선을 통해 프로세스의 품질 특성들을 정량화할 수 있으며 프로세스의 안정성과 역량을 평가할 수 있다.

(그림 6)에 보이는 기존의 데이터웨어하우스 개발 프로세스에 형상관리와 측정 및 분석을 적용하여 개선된 프로세스는 다음과 같이 분석, 설계, 개발 단계에 적용된다. 첫째, 분석단계에서 적용되는 요구사항관리는 요구사항에 대한 결함을 발견하고, 요구사항의 변경을 관리하며 해결되는 시점을 추적하여 필요한 사람에게 상태 변경 정보를 통보한다. 둘째, 설계 단계의 데이터웨어하우스 및 매핑 설계에서 적용되는 변경관리를 통해 설계상의 문제점 및 변경요청을 추적하고, 문제점을 해결하기 위한 사용자와 개발자간의 원활한 의사소통을 하며 정의된 요구사항명세서에 기반하여 설계가 진행되는지를 관리한다. 또한 개발 과정 중에 발생한 작업에 대한 우선 순위 할당을 진행한다. 셋째, 개발 단계에서 적용되는 버전관리는 프로세스 자산을 통제하고, 변경의 위험을 줄이며 산출물에 대한 히스토리 정보를 생성하고 관리한다. 특히 데이터웨어하우스와 같이 신속한 의사결정지원이 필요한 시스템에서는 수정된 정보에 의해 최신의 데이터 갱신이 이루어져야 하기 때문에 배포관리를 통해 수정된 배포 형태를 신속히 전달한다.

<표 5> 투입 공수의 변화

담당업무		투입기간(기존 프로세스)							개선프로세스
주업무	부업무	A	A+1	A+2	A+3	A+4	A+5	합 계	합 계
일정 및 조직 관리	분석	0.5	1	1	1	1	0.5	5	6
ETL	설계	0.5	1	1	0.5			3	2
ETL	개발	0.5	1	1	1	1	1	5.5	4
DW	설계	0.5	1	1	1	1		4.5	3.6
DW	개발		1	1	1	1	1	5	4
배포	구현				1	1	1	3	2.4
합 계		2	5	5	5.5	5	3.5	26	22

\* ETL (Extract, Translate, Loading), DW (Datawarehouse)

<표 6> 측정 척도 결과

구분	척도	기존 프로세스	개선 프로세스
처리시간	투입공수(M/M)	26	22
	비용(M/M 기준)	100%	84%
	생산성(투입 M/M의 비율)	$(100/26)*10 = 38.4\%$	$(100/22)*10 = 45.4\%$
품질	결함의 제거비율(제거된 결함수/발견된 결함수)	$(9/36)*100 = 25\%$	$(20/36)*100 = 55\%$
	제작업비율(발견된 결함수 - 제거된 결함수)	$(27/100) = 27\%$	$(16/100) = 16\%$
	데이터 품질(유효건수/전체 데이터건수)	59.51%	76.59%

4.2 제안된 개발 프로세스의 성능 향상 분석

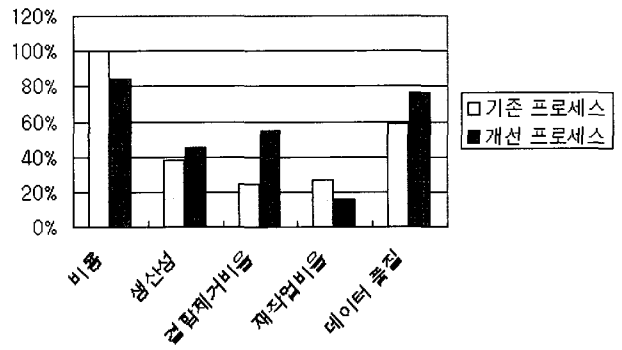
본 연구에서 제안하는 프로세스 성과 측정은 프로세스의 효과와 프로세스 수행 결과에 따른 품질수준 등의 성과를 측정할 수 있는 데이터 항목을 고려하여 제시한다. 즉, <표 3>에서 설명된 척도를 기초로 하여 기존의 프로세스와 개선된 프로세스를 비교하여 측정 결과를 <표 5>, <표 6> 및 <그림 7>에서 제시한다.

4.2.1 처리시간 감소에 따른 변화

측정은 투입 공수(Man/Month)를 비교하고, 투입공수에 따른 비율을 비용으로 책정하여 비용과 생산성을 계산한다. <표 5>에 의하면 투입공수는 일정 및 조직관리에서 기존 프로세스 대비 개선 프로세스가 증가함을 보이는데, 이는 개선 프로세스에서 형상관리를 위해 추가된 투입공수 때문이다. 즉 일정 및 조직관리의 주 업무에 형상관리 업무를 추가하여 관리한다. 그러나 전체 프로젝트의 투입공수는 감소되며, 이는 프로세스 전 과정에서 형상관리가 적용되어 오류로 인해 발생하는 제작업시간이 줄어들었기 때문이다. 결국 <표 5>에서 제시된 것과 같이 투입공수가 줄어드는 만큼 비용절감과 개발 생산성 향상으로 나타난다. 또한 데이터웨어하우스의 데이터 처리 시간은 <그림 5>와 같은 시간은 필요로 하는데 형상관리를 적용한 정확한 명세서 작성을 통해 처리시간을 감축한다.

4.2.2 품질 향상에 따른 변화

품질 향상이 가능한 이유는 형상관리를 통해 결함이 발생할 여지가 있는 데이터가 제거되었기 때문이다. 형상관리를 통해 위험 요소를 사전에 배제하였기 때문에 개선 프로세스



(그림 7) 프로세스 성능 비교

가 <표 6>과 같이 향상되었다. 발견된 결함 수와 제거된 결함 수를 측정하여 결함의 제거비율과 결함이 제거됨으로 감소하게 되는 제작업비율을 계산한다. 또한 데이터의 품질은 기존과 개선 프로세스에서 전체 회원의 데이터건수는 동일하다는 가정하에 유효한 데이터의 건수를 비율로 계산한다.

(그림 7)에서 기존 프로세스와 개선된 프로세스를 비교하였다. 처리시간의 비교는 투입공수와 비용, 생산성으로 구분할 수 있다. 비용의 경우에는 실제 프로젝트 수행 시 투입공수 자체가 비용으로 환산되기 때문에 기존 프로세스를 100%로 가정했을 때 개선된 프로세스의 투입공수가 감소하여 발생한 개선 효과이다. 생산성의 경우에도 같은 일을 수행하는데 투입공수가 줄어들었기 때문에 나타난 향상비율이다. 품질의 비교는 결함과 데이터 자체의 품질로 구분할 수 있다. 결함제거비율과 제작업비율에서는 기존 프로세스와 개선 프로세스의 발견된 결함 수는 동일하다고 가정하였다. 먼저 프로세스 수행 중에 결함을 미리 제거하였기에 결함제

거비용은 향상되었고, 프로세스가 완료된 후에 발생하는 결함에 대해서도 이를 수정하기 위한 재작업비용은 감소하였다. 데이터 품질의 경우 동일한 전체 회원의 데이터 건수에 대해 개선된 프로세스에서는 미리 데이터의 정제가 이루어졌기 때문에 유효한 회원의 데이터 건수가 증가하여 데이터의 품질이 향상되었다.

## 5. 결 론

다차원 분석을 이용한 OLAP 시스템은 사용자가 데이터를 여러 관점에서 비교 분석하여 필요한 데이터를 업무 분야별로 구조화함으로써 경영자에게 전략적인 방향을 제시할 수 있다. OLAP 시스템에서 사용하기 위해 운영 서버에서 생성된 데이터에서 다양한 분석 정보를 추출하여 저장하는 정보의 저장고로 데이터웨어하우스가 사용된다. 이러한 데이터웨어하우스는 의사결정에 직접적인 영향을 미치기 때문에 데이터의 품질이 매우 중요하다. 즉, 부정확하게 가공된 데이터는 잘못된 의사결정을 가져올 수 있기 때문에 데이터의 품질이 낮은 경우에는 데이터의 정제가 반드시 필요하다. 또한 대규모의 데이터를 처리하므로 잘못 생성된 데이터를 다시 처리하기 위해서는 많은 시간과 비용이 소요되어 생산성과도 밀접하게 관련된다.

그러나 기존의 데이터웨어하우스 개발 프로세스들은 상황과 경험에 의지하여 진행되어 왔으며, 데이터의 품질 안정화를 위한 표준화되고 체계적인 프로세스는 제시되지 않았다. 프로젝트 규모가 커지고 다양화되면서 지속적인 품질과 생산성을 향상시키기 위해 개선된 프로세스의 필요성이 증대되었다. 본 연구의 제안은 A사에서 개발한 데이터웨어하우스 개발 프로세스에 CMMI의 형상관리를 적용하여 제기된 필요성을 충족하는 개선된 데이터웨어하우스 개발 프로세스를 제안하였다.

개선된 데이터웨어하우스 개발 프로세스는 데이터의 품질을 안정화하고 데이터를 처리하는데 필요한 시간과 비용을 절감하기 위하여 기존의 데이터웨어하우스 개발 프로세스에 형상관리를 추가하였으며, 추가된 프로세스를 검증하기 위한 측정치를 제시하고 평가하였다. 제안 프로세스에서는 사용자 요구사항의 변화와 사용자 관점의 변화를 점검할 수 있으며 변화된 내용에 대한 프로세스 결과의 추적이 가능하며 프로세스가 개발 완료된 후에도 지속적인 유지보수를 지원한다. 또한 제시된 측정치를 기존의 프로세스와 비교 분석하여 제안된 프로세스가 처리시간 측면에서 투입공수의 감소와 이에 따른 비용의 절감과 생산성 향상이 있음을 보였다. 그리고 품질측면에서 형상관리를 통해 잘못된 데이터의 결함이 제거되고, 이에 따른 효과로 결함의 제거비용이 높아지고 재작업비용을 줄일 수 있으며, 데이터의 품질이 향상되었음을 제시하였다.

본 연구는 향후 데이터웨어하우스를 이용한 OLAP 개발

프로세스 측정에 필요한 척도 설계를 제시하여 OLAP 개발의 생산성을 향상시키고, 데이터의 품질을 안정화하는 작업으로 확장하여 진행할 예정이다.

## 참 고 문 헌

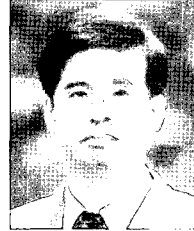
- [1] Watt Humphrey, "Managing the Software Process," Addison-Wesley, 1989.
- [2] William Florac, Robert Park and Anita Caletan, "Practical Software Measurement: Measuring for Process Management and Improvement," CMU/SEI-97HB-003, Software Engineering Institute, 1997.
- [3] Mary Chrissis, Meke Konarad and Sandy Shrum, "CMMI: Guideline for Process Integration and Product Improvement," Addison-Wesley, 2003.
- [4] Immon and H. William, "Building the Data Warehouse," 3rd Ed. NewYork: John Wiley and Sons Inc., 2002.
- [5] R. Kimball, L. Reeves, M. Ross and W. Thornthwaite, "The Data Warehouse LifeCycle Toolkit," New York : John Wiley and Sons Inc., 1998.
- [6] N. Tryfoan, F. Busborg and J. Chritiansen, "starER : A conceptual Model for Data Warehouse Design," Proc. of DOLAP, 1999.
- [7] M. Golfarelli and S. Rizzi, "A Methodological Framework for Data Warehouse Design," Proc. of DOLAP, 1998.
- [8] 박종모, 조경산, "데이터웨어하우스의 성공적인 개발 전략", 한국경영정보학회, 2004년 추계학술대회, pp.766-772, 2004.
- [9] CMU, SEI. <http://www.sei.cmu.edu/cmmi>, 2002.
- [10] 유영무, 한혁수, "CMMI 성숙도 레벨 별 측정 프로그램에 관한 연구", 한국정보처리학회, 제12권 제1호, pp.91-102, 2005.
- [11] 황선명, 엄희균, "소프트웨어 프로세스 측정을 위한 척도 설계 및 활용", 한국정보처리학회, 제12권 제7호, pp.937-946, 2005.
- [12] Anne Mette and Jonassen Hass, 'Configuration Management Principles and Practices', Addison Wesley, 2002.
- [13] 김정환, 김홍재, 박용범, "CMMI와 PMBOOK의 비교 분석을 통한 전략적 프로젝트 관리", 한국정보처리학회, 제12권 제4호, pp.601-608, 2005.
- [14] Chanwoo Yoo, Junho Yoon and Byungjeong Lee, "An Integrated Model of ISO 9001 : 2000 and CMMI for ISO Registered Organizations," Proc. of APSEC, 2004.



### 박종모

e-mail : tinytree@nate.com  
1995년 단국대학교 전자계산학과(학사)  
1997년 단국대학교 전산통계학과(석사)  
1997년~2000년 (주)한국정보시스템 선임  
2000년~2003년 (주)위세아이텍 책임  
2003년~현재 단국대학교 대학원  
정보컴퓨터학과 박사과정

2003년~현재 (주)이지애크 경영지원실  
관심분야 : ERP, CRM, 소프트웨어공학, 데이터베이스,  
시스템설계방법론, 컴퓨터시스템



### 조경산

e-mail : kscho@dankook.ac.kr  
1979년 서울대학교 전자공학과(학사)  
1981년 한국과학원 전기 및 전자공학과  
(공학석사)  
1988년 텍사스 대학원(오스틴) 전기전산  
공학과(Ph.D.)

1988년~1990년 삼성전자 컴퓨터부분 책임연구원  
1990년~현재 단국대학교 정보컴퓨터학부 교수  
관심분야: 컴퓨터 시스템, 컴퓨터 네트워크, 성능분석