

데이터웨어하우스의 개발생산성 향상을 위한 측정지표의 설계 및 분석

Design and Analysis of Metrics for Enhancing Productivity of Datawarehouse

박종모* 조경산**
Jongmo Park Kyungsan Cho

요약

다양한 분석 정보를 추출하여 저장하는 데이터웨어하우스는 기업의 마케팅 및 의사결정 지원을 위해 사용된다. 그러나 데이터웨어하우스는 분산되어 저장된 대규모의 데이터를 통합하기 때문에 처리 시간과 비용이 증가하며 오류 발생의 위험이 높다. 따라서 개발생산성의 향상 및 비용을 절감하기 위해 본 연구에서는 생산성, 프로세스 품질, 데이터 품질의 영역에서 프로세스 측정지표를 설계하여 제안한다. 또한 측정 지표를 이용하여 프로세스를 평가하고, 평가결과를 분석하여 개발생산성의 향상과 프로세스가 개선됨을 보인다.

Abstract

A datawarehouse which extracts and saves the massive analysis data is used for marketing and decision support of business. However, the datawarehouse has the problem of increasing the process time and cost as well as has a high risk of process errors because it integrates vast amount of data from distributed environments. Thus, we propose a metrics for measurement in the area of productivity, process quality and data quality. Also through the evaluation using the proposed metrics, we show that our proposal provides productivity enhancement and process improvement.

☞ keywords: Datawarehouse, Process Improvement, Productivity Enhancement, Measurement and Analysis,
데이터웨어하우스, 프로세스 개선, 개발생산성 향상, 측정 및 분석

1. 서론

정보시스템의 기술혁신에 따라 소프트웨어 개발 프로세스는 규모가 거대해지고 처리되는 데이터의 크기도 증가하는 반면에 짧은 시간 내에 개발해야 하는 등으로 환경이 변하고 있다. 이에 따라 소프트웨어 개발만을 목표로 하지 않고, 개발 생산성을 향상시키고 지속적으로 프로세스를 개선해야 할 필요성이 제시되었다. 소프트웨어 개발 프로세스를 정의하고 평가 및 개선을 위한 모델

로 SEI(Software Engineering Institute)에서는 CMM I(Capability Maturity Model Integration)를 제시하였다[1]. 프로세스 모델을 통해 소프트웨어를 개발하는 조직의 생산성과 품질의 향상을 위한 업무의 틀을 제공받을 수 있으며, 프로세스를 측정하는 기준의 수립이 가능하다. 또한 프로세스 측정은 프로세스의 수행 과정에서 계획과의 편차를 발견하는 기초가 되며, 개선을 위한 기회를 제공하고 감시 및 제어 활동의 기준이 된다[2]. 이는 네 과정으로 수행되는데 첫째, 프로세스의 성과를 측정하기 위해 데이터를 수집하고 둘째, 각 프로세스의 성과를 분석한다. 셋째, 프로세스의 안정성과 분석결과를 해석하고 마지막으로 미래의 비용 및 성과를 예상하기 위해 데이터를 사용하고

* 정 회 원: 단국대학교 정보컴퓨터학부
jmpark@dankook.ac.kr

** 정 회 원: 단국대학교 정보컴퓨터학부
kscho@dankook.ac.kr(교신저자)

[2007/06/14투고 - 2007/06/20심사 - 2007/07/10심사완료]

유지한다.

본 논문의 대상이 되는 데이터웨어하우스는 다양한 종류의 업무 시스템으로부터 대규모의 데이터를 처리하기 때문에 많은 시간과 비용이 소요되며, 잘못 처리된 프로세스와 데이터가 발생하였을 경우 이를 재처리하기 위한 위험과 재작업 비용이 증가한다. 원천 데이터의 소스인 업무 시스템에서 데이터의 품질이 제공되지 않을 경우 분석 시스템인 데이터웨어하우스에 심각한 영향을 미치게 된다[3]. 이러한 문제를 해결하기 위해 본 연구에서는 생산성, 프로세스 품질, 데이터 품질에 관한 측정 지표를 설계하여 제안한다. 실제 프로젝트 사례를 통해 제시된 지표에 의한 개발생산성을 평가하며, 평가결과를 분석하여 프로세스의 개선에 활용한다. 즉, 데이터웨어하우스를 대상으로 프로세스 측정지표의 적용을 통해 계획과의 편차를 분석하여 부족한 부분을 보완함으로써 개발생산성을 향상시킨다.

본 논문의 구성은 다음과 같다. 2장에서는 데이터웨어하우스와 소프트웨어 프로세스 개선 및 평가 모델을 정의하고, 프로세스 측정에 관련된 연구를 제시한다. 3장에서는 데이터웨어하우스의 개발생산성 향상을 위한 측정 지표를 제안하고 4장에서 제안된 측정 지표를 사용하여 프로세스 평가를 수행한 후에 결과를 분석하고 프로세스 개선에 활용한다. 마지막으로 5장에서 결론 및 향후 연구로 본 논문을 마무리한다.

2. 관련연구

데이터웨어하우스는 운영업무 지원시스템인 OLTP(OnLine Transaction Process) 시스템에서 생성된 데이터로부터 다양한 분석 정보를 추출하여 의사결정 지원시스템인 OLAP(OnLine Analysis Process)에 사용하기 위한 데이터의 저장고이다[4, 5]. 데이터웨어하우스는 사용자들의 다양한 분석을 효율적으로 지원하기 위하여 분석의 기준이 되는 차원들과 분석할 내용인 측정값으로 구성된

다차원 형태의 DFM(Dimension Fact Model)으로 데이터를 표현한다. 데이터웨어하우스를 이용하면 사용자가 대화식으로 정보를 분석할 수 있는 의사결정 지원시스템인 OLAP 질의에 빠른 응답을 할 수 있다. 데이터웨어하우스의 설계를 위해 이용되는 스타 스키마는 사실 테이블을 중심으로 차원 테이블들이 별 모양의 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 운영 데이터베이스로부터 수집된 정보를 분석하고, 사용자의 요구 사항을 명세화한 단계를 거친 후, 개념적 설계를 통해 사실과 차원 구조의 생성으로 제시하였다. 개념적 설계 결과에 대한 작업량과 유효성을 판단한 후 테이블 구조로 실체화하기 위한 논리적 설계 단계를 거치고, 물리적 설계를 수행한다. 이러한 과정으로 설계되는 데이터웨어하우스는 대규모의 데이터를 처리하기 때문에 지속적인 프로세스 품질 개선과 개발생산성의 향상이 필요하다.

프로세스를 표준화하고 평가 및 개선하기 위한 모델인 CMMI는 능력성숙도(Capability Maturity) 모델로서 발전되어 왔다. CMMI가 소프트웨어와 관련된 프로세스 개선 활동만을 지원하던 취약점을 보완하여 CMMI는 소프트웨어, 시스템, 하드웨어, 서비스와 관련된 모든 프로세스 개선활동을 지원한다[8]. 또한 조직의 성숙도 및 공정 능력의 평가와 향상을 위한 활동의 우선순위 결정 및 지원 프레임워크로 구성된다[1]. CMMI는 소프트웨어 프로세스를 평가하는 국제 표준 모델이며 SPICE(Software Process Improvement and Capability dEntermination)로 알려져 있는 ISO 15504를 수용한다. ISO 15504는 소프트웨어 프로세스를 고객-공급자, 조직화, 관리, 지원, 엔지니어링의 5개의 영역으로 나누어 각 프로세스에 대한 계획, 관리, 제어, 개선을 수행하는 프레임워크를 제공하며, 프로세스 수립의 기반이나 프로세스 적용 수준을

평가하는 기준으로 활용한다[9]. [10]에서는 소프트웨어 프로세스를 측정하기 위한 프로세스 측정 지표 모델과 역할 및 주요한 활동을 제안하여 조직의 프로세스 성숙도를 향상시키고, 효과적인 통제와 효율성 및 측정의 편리성이 증가함을 보였다.

측정지표의 첫 번째 대상은 프로세스 품질과 생산성이다. [11]은 CMMI의 프로세스 영역에서 MA(Measurement & Analysis)를 기반으로 측정 프로그램을 개발하고, 개발된 측정 프로그램을 통해 CMMI의 각 레벨 단위로 측정을 수행하였다. [12]에서는 <표 1>과 같이 CMMI의 측정 프로세스들에 대하여 프로세스 개선에 대한 객관적인 증거를 제시하기 위해 측정을 위한 정량적인 지표들을 정의하였다. 이 지표들을 통해 프로세스 능력 수준이 결정되고 지속적인 측정값 수집을 통한 조직의 프로세스 개선 예측이 가능하며, 자원의 효과적인 분배와 문제발생에 대한 조기 대응으로 프로젝트 성공확률을 증대시킬 수 있었다. 그러나 데이터웨어하우스는 고객의 분석 정보를 추출하고 로드하는 매핑 프로그램이 대부분의 시간을 차지하기 때문에 데이터가 잘못 처리되었을 경우 재작업 시간이 많이 소요된다. 따라서 매핑 설계 부분을 집중하여 관리하기 위한 측정 지표의 보완이 필요하다.

<표 1> 프로세스의 측정지표

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

측정지표의 두 번째 대상은 데이터이다. 데이터의 품질을 측정하기 위한 지표는 아래의 <표 2>와 같이 완전성, 일관성, 최신성, 정확성의 4가지 기준에 따라 제시되었다[13]. 완전성은 정해진 데이터 구조 안에서 정보가 누락됨 없이 포함되어 있는지를 판단하며, 일관성은 데이터가 상충되지 않고 일관된 상태를 이루고 있는지를 관리한다. 최신성은 최근의 정보를 제공하여 지속적인 갱신이 이루어지는가의 관점이며, 정확성은 데이터의 값과 표현이 실제 값과 동일한지를 측정하는 기준이다. 그러나 데이터웨어하우스는 원천 소스의 데이터에 따라 분석이 되는 정보의 정확성과 신뢰성이 보장되기 때문에 데이터 품질의 향상을 위해서는 원천 소스의 데이터 품질이 매우 중요하다. 따라서 원천 소스의 데이터 품질을 확보하기 위한 측정지표를 중점으로 관리해야 한다.

<표 2> 데이터 품질의 측정지표

측정 기준	데이터 값	데이터 구조	데이터 흐름
완전성	<ul style="list-style-type: none"> •데이터 크기 •데이터 범위 •데이터 값 누락 	<ul style="list-style-type: none"> •중요 속성 누락 •필수 속성 설계 	<ul style="list-style-type: none"> •데이터 생성 가공 시 누락
일관성	<ul style="list-style-type: none"> •데이터 속성과 값의 일치 •데이터 제약조건 일치 •데이터 정의와 레코드 일치 •동일 데이터의 상호일관성 	<ul style="list-style-type: none"> •데이터 표준의 적절성 •도메인 정의의 적절성 •코드 정의의 적절성 	<ul style="list-style-type: none"> •데이터 생성 가공 시 데이터 적용
최신성	<ul style="list-style-type: none"> •최신 데이터 제공 	-	<ul style="list-style-type: none"> •데이터 갱신 주기
정확성	<ul style="list-style-type: none"> •데이터 오타자 •실제 사실과의 일치 •레코드 중복 	<ul style="list-style-type: none"> •참조 무결성 •속성 중복 및 유일성 보장 	<ul style="list-style-type: none"> •원천 데이터의 신뢰성 •데이터 생성 가공 시 오류 및 중복

3. 데이터웨어하우스에서 개발생산성 측정을 위한 지표 제안

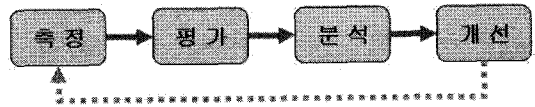
본 장에서는 데이터웨어하우스에서 개발생산성

의 향상을 측정하기 위한 지표를 설계하여 제안한다. 측정지표로 프로세스의 목적달성이나 활동의 진행 정도를 측정하여 원래 계획과의 편차를 분석하고 프로세스의 개선을 수행할 수 있다. 본 논문에서는 데이터웨어하우스에서의 개발생산성을 생산성, 프로세스 품질, 데이터 품질로 정의하였다. 프로세스 개선을 위한 <표 3>의 측정항목은 앞 장의 <표 1>과 <표 2>에서 제시된 프로세스와 데이터 품질 측정에 사용된 항목을 데이터웨어하우스의 특성에 적합하도록 수정한 것이다. 즉, 데이터웨어하우스의 대규모 데이터의 프로세스 처리를 측정하기 위한 생산성 항목, 요구사항 변경 및 오류 등의 처리를 측정하는 프로세스 품질 항목, 원천 소스의 데이터로 인해 발생하는 데이터 자체의 품질을 분석하기 위한 데이터 품질 항목으로 구분하였다.

<표 3> 데이터웨어하우스의 측정항목

구분	측정항목	설명
생산성	S	소프트웨어의 크기
	E	프로젝트에 투입된 공수(Man/Month)
	D	D1 프로젝트 계획일수 D2 프로젝트 지연일수
프로세스 품질	R	R1 개발 기간 동안 제거된 결함개수
		R2 고객에게 인도 후 제거된 결함개수
		R 제거된 결함 수 ($R = R1 + R2$)
	A	A1 요구사항 변경건수
		A2 요구사항 구현건수
		A 요구사항 수
	F	F1 개발 기간 동안 발견된 결함개수
F2 고객에게 인도 후 발견된 결함개수		
F 프로젝트 전체 결함개수 ($F = F1 + F2$)		
데이터 품질	데이터	데이터 자체의 품질
	데이터관리	데이터 관리 정책
	데이터구조	데이터 간의 구조

본 연구에서 측정 프로세스는 (그림 1)에 따라 진행되는데 먼저 프로세스의 특성을 측정항목에 따라 구분한 측정지표로 측정한 후에 개발생산성을 평가한다. 다음으로 평가결과를 분석하고, 프로세스를 개선하기 위한 토대로 활용한다.



(그림 1) 측정 프로세스

<표 3>의 측정항목을 이용하여 프로세스의 개발생산성을 평가하기 위한 측정지표를 <표 4>와 같이 제안한다. 생산성은 투입공수-지연율의 새로운 지표를 제안하였고, 프로세스 품질은 데이터웨어하우스에 적합하도록 수정하였다. 전체 프로세스에서 매핑이 상대적으로 많은 시간이 소요되므로 오류로 인해 발생하는 제작업의 비율을 줄이도록 측정지표를 보완하였다. 데이터 품질의 영역에서는 <표 2>의 데이터의 품질측정 지표 중에 중요한 요소인 데이터 값과 데이터 값의 정확성을 위해 필요한 데이터 구조를 지표로 선택하였다. 왜냐하면 원천 소스의 데이터 품질이 나쁠 경우 데이터웨어하우스의 품질도 보장 할 수 없기 때문이다.

<표 4> 프로세스의 측정지표

구분	측정대상	측정지표
생산성	투입공수	투입공수율 = E / S
	지연일수	계획공수 지연율 = $D2 / D1$ 투입공수 지연율 = $(D2/D1) \times E$
프로세스 품질	결함	결함 주입률 = F / S 결함 제거율 = R / F 개발기간 동안 결함 제거율 = $R1 / F1$ 인도 후 결함 제거율 = $R2 / F2$
		위험
	데이터	데이터 값 및 항목의 누락 데이터 정확성에 대한 신뢰도 데이터 표현형식의 적절성 데이터의 유일성, 제약조건(유효범위) 데이터의 최신성
데이터 품질	데이터 관리	데이터 표준 관리(용어사전, 코드표준) 데이터 관리 정책(데이터 권한 및 보안)
	데이터 구조	참조무결성을 위한 테이블 간의 관계 데이터 구조 변경에 따른 데이터모델 관리 시스템간의 데이터 중복성

생산성은 투입공수와 지연일수로 측정한다. 투입공수에서 소프트웨어의 크기는 데이터웨어하우스의 경우 일반적인 프로그램 규모 산정에 사용되는 LOC(Line Of Code)가 아니라 축적된 데이터베이스의 크기이다. 왜냐하면 데이터웨어하우스가 분석을 위한 데이터의 저장소로서 사용되기 때문에 데이터베이스의 크기가 중요 관점이다. 생산성과 밀접한 관련이 있는 계획공수 지연율은 프로세스 관리에 초점이 맞추어 있으며, 프로세스 지연에 따른 계획일수와 실제일수의 차이를 분석한다. 지연일수가 발생하였을 때 동일한 일정의 지연이라 할지라도 비용이 더 투입되는 곳을 집중 관리하기 위해 본 연구에서는 계획공정 지연일수와 비용의 크기인 투입 공수를 함께 고려하였다. 즉, 지연된 일수에 투입공수를 비용으로 환산한 지연된 비용의 가치인 [M/M]를 곱하여 투입공수-지연율을 측정 기준으로 제안한다.

프로세스 품질영역에서 결함은 결함 제거 측면에서 결함이 발생할 여지가 있는 결함 주입율과 결함의 제거비율로 측정하며, 개발 기간 동안의 결함과 인도 후 결함으로 구분한다. 위험은 재작업 비율과 요구사항의 변경 및 구현율로 측정한다.

데이터 품질영역은 다음의 세 가지로 구성된다. 데이터 자체가 가지는 품질인 데이터 영역과 데이터 표준 및 관리정책인 데이터 관리 영역 그리고 데이터와 데이터의 연결에 관련된 데이터 구조의 세 영역으로 측정지표를 제시하고 분석한다.

4. 제안된 측정지표에 의한 프로세스 개발생산성의 평가 및 분석

본 장에서는 3장에서 제안된 측정지표를 활용하여 생산성, 프로세서 품질, 데이터 품질의 세 가지 영역에서 개발생산성을 평가하기 위해 실제 S사의 데이터웨어하우스 구축 사례를 통해 평가를 수행하였다. 다양한 채널을 통해 온라인 및 오프라인으로 제품을 판매하는 S사는 고객의 자료를 분석하고 통합하여 고객의 특성에 맞는 마케팅을 지원하기 위한 시스템을 구축하였다[3]. 이를 위한 분석 업무용의 데이터웨어하우스 개발 프로세스는 (그림 2)와 같다.

그림 2는 데이터웨어하우스의 개발 프로세스를 보여주는 흐름도이다. 주요 단계는 '요구정의', '설계', '구현'으로 나뉘며, 각 단계에는 세부적인 활동들이 포함되어 있다. '요구정의' 단계에는 '요구사항이해', '요구사항정의', '기술구조정의'가 포함된다. '설계' 단계에는 'DW 모델설계', '매핑설계', '사용자화면설계', '테스트설계'가 포함된다. '구현' 단계에는 '출원', '배포'가 포함된다. 또한 '테스트계획수립'과 '구현계획수립'은 '구현' 단계로 연결된다. '배포' 이후에는 '시스템사용교육'과 '시범운영'이 이루어진다. '출원' 단계에서는 '임플릿명세작성', '데이터구축 단위, 통합테스트실시'가 수행된다.

(그림 2) 데이터웨어하우스의 개발 프로세스

S사의 데이터웨어하우스 개발에서 제시된 문제점은 사용자수 및 신규업무의 증가에 따른 대규모 데이터 처리에 필요한 시간의 증가, 매핑 로직의 복잡성 및 다양성에 따른 통합의 지연이다. 이로 인해 (그림 3)과 같이 주요한 일배치작업이 주간 근무 시간에까지 영향을 미치게 되었다. 따라서 데이터웨어하우스의 개발 프로세스에서 개발생산성의 측정은 대규모의 데이터 처리와 매핑 처리에 중점을 두어야 한다.

일배치작업	근무 시간대																							
	21	22	23	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
작업영역																								
매출집계																								
사용자 식별																								
조직데이터 생성																								
제품데이터 생성																								
환산데이터 입수																								
데이터 정제																								
통합																								
서비스 데이터 전송																								
백업																								

(그림 3) 일배치 작업

4.1 생산성의 평가 및 분석

프로세스의 개발생산성을 평가하기 위한 첫 번째 영역인 생산성은 <표 4>에서 프로젝트에 사용된 투입공수와 지연일수로 측정되며, 지연율은 계획일수 대비 지연일수로 측정된다. 개발 기간의

투입공수와 프로젝트 관리를 통해 변화된 투입공수를 평가하면 <표 5>와 같다. S사의 개발 프로세스는 (그림 2)의 과정과 같이 진행되는데 전체 작업은 총 6개월(24주)이 소요되었으며, 투입공수는 계획단계의 예측보다 점진적 개발 단계에서 프로젝트 관리와 통제를 통해 1 M/M가 감소되었다.

<표 5> 투입공수의 변화

담당업무	A	A+1	A+2	A+3	A+4	A+5	합계
요건정의	0.5	1					1.5
설계		1.5	1.5	0.5			3.5
점진적 개발		1.5	2	2.5 ⇒ 2	3 ⇒ 2.5		9 ⇒ 8
구현				1	1	2	4
합계	0.5	4	3.5	4 ⇒ 3.5	4 ⇒ 3.5	2	18 ⇒ 17

[A: 시작일, A+i: 1개월, 측정단위: M/M]

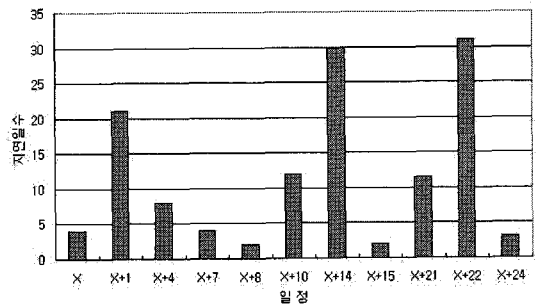
<표 6>은 작업공정에 대해 작업이 지연된 일수와 단계 및 세부작업으로 (그림 4)와 같이 분석되었다. 프로젝트 기간을 24 개의 주로 구분하였으며, 1,133일의 전체 작업 일수(각 업무 담당자의 작업 일수의 합)에 대해 <표 6>의 지연일수 합은 128.5일이다. 계획공정 지연율에 따른 계획 대비 실제 프로젝트 일정은 11.34%의 지연이 발생하였다. 즉, 데이터웨어하우스 개발에서 문제로 지적된 매핑 부분과 분산되어 있는 고객 데이터의 통합 및 정제의 과정이 미흡하였기 때문에 발생한 데이터 추출 및 로드의 지연일수가 많았으며, 이를 관리하기 위해 본 논문에서는 투입공수-지연율의 측정을 제안한다.

<표 6> 작업지연에 따른 작업공정

일정	지연일	[M/M]	단계	세부작업공정
X	4	0.5	단계준비	프로세스별 담당업무 분장
X+1	21	0.5	요구사항이해	요구사항수집 (인터뷰)
X+4	8	0.5	요구사항이해	현행시스템분석
X+7	4	4	요구사항정의	데이터요구사항정의

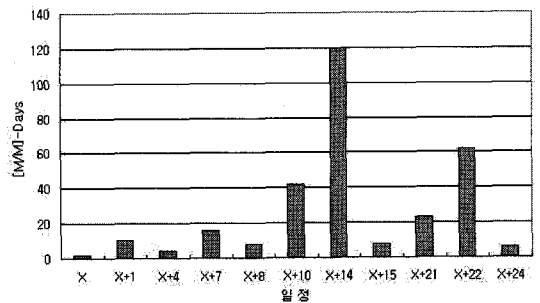
X+8	2	4	기술구조정의	개발환경구축
X+10	12	3.5	DW모델설계	DW모델설계
X+14	30	4	매핑설계	매핑프로그램 설계
X+15	2	4	테스트계획수립	테스트 시나리오 작성
X+21	11.5	2	템플릿명세작성	비즈니스 템플릿
X+22	31	2	데이터 구축	데이터 추출 및 로드
X+24	3	2	통합테스트	통합테스트

[X: 시작주, X+i: i주, 지연일단위: 일(day)]



(그림 4) 일정에 따른 지연일수

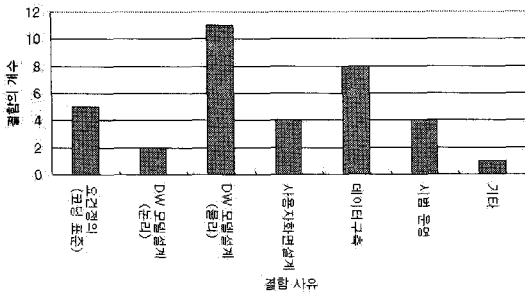
(그림 4)의 지연일수에 제안된 투입공수-지연율의 측정치를 사용하여 지연된 공정의 가치를 적용하면 (그림 5)와 같이 14 주가 가장 심각하였다. 단순히 지연 일수만을 비교하였을 경우에는 14 주와 22 주가 비슷하지만 <표 6>의 [M/M]에 해당하는 공정의 가치와 비교를 하였을 경우에는 작업의 우선 순위를 14 주에 두어야 한다. 즉, 프로젝트 예측 및 관리가 더 필요한 부분에 자원을 투입함으로써 계획대비 공정지연율을 개선시킬 수 있다.



(그림 5) 투입공수-지연율의 작업지연

4.2 프로세스 품질의 평가 및 분석

프로세스의 개발생산성을 평가하기 위한 두 번째 영역인 프로세스 품질은 요건정의, 설계, 점진적 개발, 구현의 프로세스 단계별로 평가한다. 요건정의 단계에서의 코딩 표준과 설계 단계에서의 논리와 물리적 설계 및 사용자 화면 설계를 평가하고, 점진적 개발단계에서의 데이터 구축, 마지막으로 구현단계의 시범운영을 평가한다. 주요한 프로세스에 대한 평가 결과는 (그림 6)과 같이 데이터웨어하우스의 물리적 설계와 데이터 구축에서 많은 결함이 발견되었다. 이를 관리하기 위해 본 논문에서는 결함과 위험에 대한 측정 지표를 제시하고, 측정치에 대한 평가과 분석을 통해 프로세스 품질을 높이고자 한다.



(그림 6) 프로세스 결함분석

발견된 프로세스 결함에 따라 관련된 측정항목 값은 <표 7>과 같고, 이를 이용한 결함과 위험의 평가 결과는 <표 8>과 같다.

<표 7> 측정항목

전체 크기(S) = 60 (결함의 대상 건수)					
전체 결함(D)	19	개발 기간 동안 결함(D1)	10	인도 후 결함(D2)	9
제거된 결함(F)	8	개발 기간 동안 제거(F1)	5	인도 후 제거(F2)	3

<표 8> 결함과 위험의 평가 결과

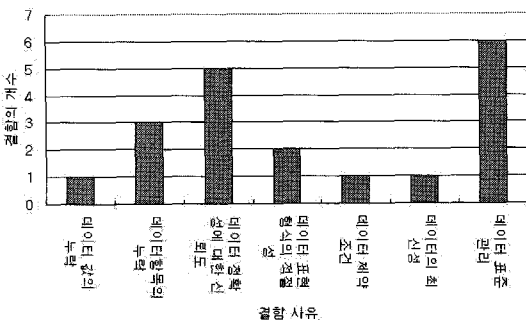
분류	측정 지표	개선전 평가	개선후 평가
결함	결함 주입율(R/S)	(19/60)*100 = 32%	(11/60)*100 = 18%
	결함 제거율(R/F)	(6/19)*100 = 32%	(8/19)*100 = 42%
	개발기간의 결함제거율(R1/F1)	(4/10)*100 = 40%	(5/10)*100 = 50%
	인도 후 결함제거율(R2/F2)	(2/9)*100 = 22%	(3/9)*100 = 33%
위험	재작업 비율(F1 - R1)/S	(10/60)*100 = 17%	(5/60)*100 = 8%
	요구사항 변경율(A1/A)	(37/83)*100 = 45%	
	요구사항 구현율(A2/A)		(21/83)*100 = 25%

<표 8>의 평가결과는 프로젝트가 진행 중일 때 나타난 결함을 평가하는 개선전 평가와 다양한 품질 통제 활동과 프로세스 측정을 통해 결함이 발생할 여지가 있는 데이터와 위험 요소를 제거한 개선후 평가로 구분된다. 이를 분석하면 결함 주입률에 따라 결함 제거율이 결정되는데 개선후에는 결함의 주입율이 줄어들고, 제거율은 향상되었다. 또한 개발 기간 동안과 인도 후의 결함 제거율로 구분하여 분석하면 결함을 제거해 나가는 과정을 통해 제거비율이 향상되었다.

결함의 제거 과정을 통해 개발 기간 동안 발견된 결함에서 제거된 결함을 평가하면 재작업 비율이 향상됨을 볼 수 있다. 요구사항 변경율과 구현율을 통해 요구사항의 변경에 대한 구현 비율을 제공하며 요구사항의 변경율이 많을수록 프로젝트의 위험은 증가한다. 요구사항의 변경율은 높은 편이었으나 프로세스 관리와 결함을 제거해 나가는 과정을 통해 실제 구현해야 할 요구사항 구현율은 상대적으로 낮출 수 있다. 즉, 진행과정의 원인을 분석함으로써 결함의 적시 발견으로 결함에 대한 대처가 가능해지고 프로젝트 품질을 높일 수 있다. 그 결과로 결함을 제거하는데 필요한 공수가 감소되어 생산성을 증가시킬 수 있다.

4.3 데이터 품질의 평가 및 분석

프로세서의 개발생산성을 평가하기 위한 세 번째 영역으로 데이터 품질을 평가한다. <표 4>에서 제시된 측정지표를 적용하여 S사의 고객 정보에 대한 데이터 품질의 결함을 분석하면 (그림 7)과 같다. 데이터의 표준관리와 정확성에 대한 신뢰도의 결함이 가장 심각하였다. 이를 분석하면 대상 시스템에서 데이터 자체는 관리하지만, 표준화된 업무 규정에 의한 데이터 관리와 데이터의 구조 및 전사적 연계가 부족하기 때문이다. 시스템의 구축과정에서 데이터의 값 및 항목의 누락, 전화 번호 등의 기록시에 발생하는 데이터 표현형식의 적절성, 불확실한 데이터의 제약 조건 및 매핑 문제 등의 결함으로 분석되었다. 또한 반복적 시스템 구축으로 인해 가장 최근의 데이터 현황이 파악되지 않아 과거 데이터 현황을 기준으로 분석 데이터를 생성해야 하는 경우도 발생하였다. 따라서 분석된 결과를 토대로 데이터 품질을 향상시키기 위해 데이터에 대한 정확한 파악 작업을 수행하였고, 데이터에 대한 정제 로직의 추가로 결함을 제거하여 유효한 데이터의 건수를 증가시켰다.



(그림 7) 데이터 품질의 결함 분석

데이터 구조 분석 측면에서의 주요한 문제점은 테이블에 대한 참조가 정확하지 않아 외래 키의 불일치가 발생하는 경우가 빈번하고, 데이터의 구

조 변경에 따른 데이터 모델이 관리되지 않으며, 전사적 연계를 위한 데이터 통합 관리가 미약하여 발생한 시스템간 데이터 중복이었다. 이는 다음의 <표 9>와 같이 참조무결성, 데이터의 구조, 데이터 중복성으로 분석되었다.

<표 9> 데이터 구조 분석

측정지표	분석	원인
참조 무결성	판매 채널	지사-지점-대리점의 관리구조가 명확하지 않음
데이터의 구조	제품 데이터의 구조	품목-모델군-모델의 데이터구조가 불확실
데이터 중복성	판매 데이터의 중복	고객-대리점-제품의 데이터 중복

데이터 구조의 분석에서 나타난 문제점은 전체 프로세스와 데이터 관리 및 정제의 과정을 통해 해결이 가능하다. 판매 채널과 제품의 경우 지사, 지점, 대리점의 정보를 분석관점에서 재정의하여 데이터를 정제한다. 정확한 분류 정보와 비정확한 분류정보를 확정하여 각 분류 단계마다 정확하지 않은 정보는 기타 분류에 포함시키거나 하향구조를 만족하지 않은 데이터를 제외시킨다. 판매 데이터의 구조의 경우는 제품과 대리점의 정의되지 않은 코드가 존재할 경우 코드를 재분류한 후에 다시 입력하거나 제외시킨다.

5. 결론

최근의 기업환경에서는 데이터 분석을 통한 마케팅이 기업의 경쟁력에서 중요한 역할을 담당한다. 마케팅 및 기업의 의사결정 지원을 위하여 업무 시스템으로부터 생성된 데이터에서 다양한 분석 정보를 추출하여 저장하는 정보의 저장고로 데이터웨어하우스가 사용된다. 그러나 데이터웨어하우스는 다양한 종류의 업무 시스템으로부터 대규모의 데이터를 처리하기 때문에 많은 시간과

비용이 소요되며, 오류 발생의 위험과 재작업의 비율이 높다. 또한 원천 소스에서 발생하는 데이터의 품질문제로 인해 분석의 신뢰도가 떨어지게 된다. 이와 같이 다양화된 개발 환경에서 개발생산성의 향상 및 비용의 절감을 해결하기 위해 본 연구에서는 프로세스 측정지표를 설계하여 제안하고, 지표를 활용한 측정을 수행한 후에 측정 결과의 분석과 평가를 통해 프로세스 개선이 가능하도록 하였다. 이는 프로세스 측정이 프로세스의 수행 과정에서 계획과의 편차를 발견하는 기초가 되며 개발생산성의 향상과 프로세스 관리를 통한 개선의 기회를 제공하기 때문이다.

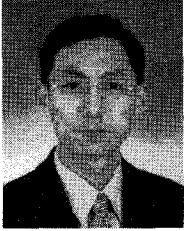
데이터웨어하우스의 개발생산성의 향상을 위한 측정지표는 생산성, 프로세스 품질, 데이터 품질의 영역에서 제안되었다. 생산성 영역에서는 투입공수에 따른 생산성 측정과 지연의 크기를 고려한 투입공수-지연율을 제시하여 동일한 일정의 지연이라 할지라도 비용이 추가되는 곳을 집중 관리해야 함을 보였다. 또한 프로세스 품질 영역에서는 프로세스 결함과 위험의 분석을 위하여 측정 항목과 측정 결과를 제시하여 결함의 제거와 위험 관리를 통한 프로세스 개선이 가능함을 설명하였다. 마지막으로 데이터, 데이터 관리, 데이터 구조 측면에서 측정지표를 제시하였으며, 제시된 데이터 품질 측정지표에 의한 평가를 통해 데이터 품질이 향상될 수 있음을 보였다. 데이터 결함의 원인을 파악하여 데이터 관리 및 데이터에 대한 정제 과정의 추가를 통하여 결함을 제거하였으며, 전사적 연계를 위한 데이터 구조를 재정의하였다. 본 연구는 데이터 품질에 대한 측정지표는 제시하였으나, 정량적인 측정 기준이 미흡하다. 연구의 완성도를 높이기 위한 향후 연구로 데이터 품질에 대한 정량적인 측정지표의 연구가 계속 될 것이다.

참 고 문 헌

[1] Mary C., Meke K. and Sandy S., CMMI:

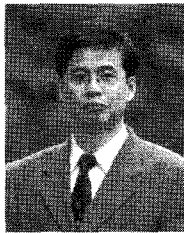
- Guideline for Process Integration and Product Improvement, Addison-Wesley, 2003
- [2] KliWon S. and SunMyung H., "Analyzing Relationships to the Quality Level between CMM and Six Sigma," Proc. of SERA03, pp.34~50, 2003
- [3] 박종모, 조경산, "CMMI의 형상관리를 적용한 데이터웨어하우스 개발 프로세스의 개선," 한국정보처리학회, 제13권 4호, pp.625~632, 2006
- [4] Immon W., Building the Data Warehouse, 3rd Ed., John Wiley & Sons Inc., 2002
- [5] Kimball R., Reeves L., Ross M. and Thornthwaite W., The Data Warehouse LifeCycle Toolkit, John Wiley & Sons Inc., 1998
- [6] Tryfoan N., Busborg F. and Chritiansen J., "starER : A conceptual Model for Data Warehouse Design," Proc. of DOLAP, pp.3~8, 1999
- [7] Golfarelli M. and Rizzi S., "A Methodological Framework for Data Warehouse Design," Proc. of DOLAP, pp.3~9, 1998
- [8] CMU, SEI, <http://www.sei.cmu.edu/cmimi>, 2002
- [9] ISO/IEC 15504-1:2004, Information technology - Process assessment - Part1: Concepts and vocabulary, 2004
- [10] Ruzhi X., Yunjiao X. and Peiyao N., "Research on CMMI-based Software Process Metrics," Proc. of IMSCCS, pp.391~397, 2006
- [11] William F., Robert P. and Anita C., Practical Software Measurement: Measuring for Process Management and Improvement, SEI, 1997
- [12] 황선명, 엄희균, "소프트웨어 프로세스 측정을 위한 척도 설계 및 활용," 한국정보처리학회, 제12권 7호, pp.937~946, 2005
- [13] 김찬수, "데이터의 구조적 품질관리 성숙도 모델 개발," 경희대학교, 박사학위논문, 2004

◎ 저 자 소개 ◎



박 종 모(Jongmo Park)

1995년 단국대학교 전자계산학과(학사)
1997년 단국대학교 전산통계학과(석사)
2007년 단국대학교 정보컴퓨터학과(박사)
1997년~2000년 한국정보시스템 ERP개발부
2000년~2003년 위세아이텍 CRM사업부
2003년~2006년 이지엠텍 전산실
2007년~현재 단국대학교 정보컴퓨터학부 강의전담교수
관심분야 : 소프트웨어공학, 데이터베이스, 컴퓨터시스템
e-mail : jmpark@dankook.ac.kr



조 경 산(Kyungsan Cho)

1979년 서울대학교 전자공학과(학사)
1981년 한국과학원 전기전자공학과(공학석사)
1988년 텍사스 대학원(오스틴) 전기전산공학과(Ph.D.)
1988년~1990년 삼성전자 컴퓨터부분 책임연구원
1990년~현재 단국대학교 정보컴퓨터학부 교수
관심분야 : 네트워크시스템 및 이동통신보안, 컴퓨터시스템
e-mail : kscho@dankook.ac.kr