

데이터 웨어하우스 구축을 위한 상향식·하향식 접근 방법에 관한 연구

- 방법론적 고찰을 통한 접근

박주석, 전대성, 박진휘
경희대학교 경영학과

Recently, domestic datawarehouse marketplaces take up the most important part of information technology investments with ERP(Enterprise Resource Planning), EC(Electronic Commerce) construction projects in spite of a business slowdown for IMF environment.

Therefore, in industrial fields as finances, communications, manufactures, distributions, and so on, datawarehouse was constructed or is under construction, and datawarehouse success stories was showed around some enterprises that has constructed previously datawarehouse.

This study is intended to be henceforth an index and direction for an actual proof study by discussing a study of methodology for datawarehouse construction in order to compare and analyze Top-down & Bottom-up approach for datawarehouse that has constructed in actual enterprises.

I. 서론

최근 몇 년에 걸쳐 대기업을 중심으로 대량의 데이터를 포함하고 있는 정보시스템을 구축, 관리할 수 있는 데이터 웨어하우스가 새로운 아키텍처로 각광받고 있다. 국내의 데이터 웨어하우스 시장은 IMF 환경으로 인한 경기 부진에도 불구하고 전사적 자원관리(ERP ; Enterprise Resource Planning), 전자상거래(EC ; Electronic Commerce) 구축 프로젝트 등과 더불어 정보기술(IT ; Information Technology) 투자의 가장 중요한 부분을 차지하고 있다. 따라서 금융, 통신, 유통, 제조 등 각 산업 분야별로 데이터 웨어하우스가 구축 중이거나 구축되었으며, 이미 데이터 웨어하우스를 구축한 일부 업체들을 중심으로 데이터 웨어하우스 성공 사례가 나타나고 있다.

그러므로 데이터 웨어하우스 구축을 위한 상향식·하향식 접근 방법을 비교·분석하여 그 특성을 파악하기 위해 데이터 웨어하우스 구축에 따른 방법론적 고찰을 논함으로써 데이터 웨어하우스 구축을 위한 지침과 방향성을 모색하여야 한다.

본 연구는 7가지의 데이터 웨어하우스 방법론을 여러 가지의 특성 요인별로 비교·분석하여, 실제 기업에서 구축한 데이터 웨어하우스 접근 방법을 분석하고 측정하기 위한 향후 실증 연구의 토대

와 방향을 제시하고자 한다.

II. 상향식·하향식 접근방법

데이터 웨어하우스 도입 추진에 있어 생각할 수 있는 접근 방법은 크게 상향식과 하향식 방식이 있다. 우선, 상향식(Bottom-up) 접근법은 개별 데이터 마트 통합으로 전사적인 데이터 웨어하우스를 구축하는 것이고, 하향식(Top-down) 접근법은 데이터 웨어하우스를 구축한 후 필요하면 데이터 마트를 구축하는 것이다. 이러한 접근 방식을 선택할 때 고려해야 할 요소들로는 구축 소요시간, 초기 비용, 분석의 어려움, 분석의 어려움, 위험 요소 등을 고려해야 한다.

상향식 접근 방법은 데이터 마트를 먼저 구축하기 때문에 실제적으로 비용과 시간을 단축시킬 수 있다는 점과 단계적으로 구축함으로써 각 단계에 대한 구축 성공 여부를 조기에 결정할 수 있다는 장점이 있다. 따라서, 최고 경영층에서 의지를 갖고 지원할 수 있다. 반면에, 데이터 마트는 데이터 웨어하우스 만큼 데이터를 담을 수 없다는 문제점이 발생한다. 또한 수 개의 데이터 마트를 구축하면, 또다시 데이터 중복성, 무결성, 일치성 등에 대한 문제가 야기되며, 향후 데이터 웨어하우스 통합 시 전체 모델링을 선행해야 한다는 어려움이 있다.

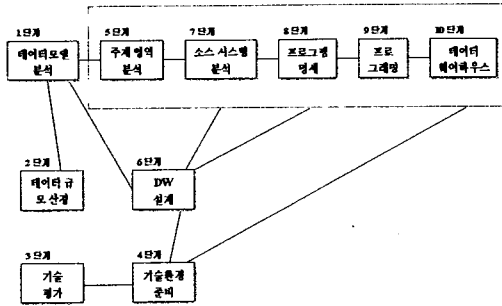
하향식 접근 방법은 장기적으로 보면, 통합에 따른 문제가 없으며, 기업의 모든 데이터를 일관성 있는 형태로 사용자가 이용할 수 있다는 점에서 커다란 가치가 있다. 반면에, 구축 비용과 시간이 많이 들고, 구축 성공에 대한 확신이 부족하다는 단점도 야기된다. 그러므로, 최고 경영층에서 이를 적극적으로 지원하기가 힘들었다.

결국, 데이터 웨어하우스 구축을 고려중인 기업은 사용자의 요구와 전략적 목표를 세심하게 검토한 후, 전사차원의 데이터 웨어하우스를 구축할 것인지 신속하고 특정 기능만 지원하는 데이터 마트를 구축할 지를 결정해야 한다.

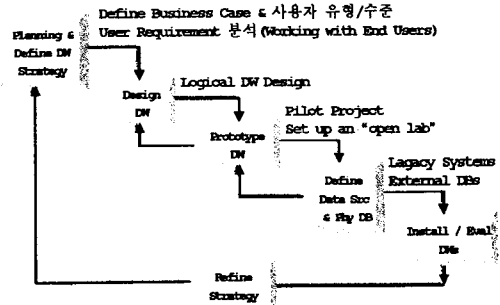
또한 상향식·하향식 접근 방법을 통한 데이터 웨어하우스의 구축이 데이터 발견 기법인 데이터 마이닝을 위한 플랫폼을 제공한다는 잠재적인 특징을 인식하여야 한다.

III. 데이터 웨어하우스 구축방법론 분석 (7가지 방법론을 중심으로)

본 연구에서는 실제 데이터 웨어하우스 구축 경험이 있는 5개 SI 업체의 데이터 웨어하우스 구축 방법론과 인먼(W.H.Inmon) 및 박주석 교수의 데이터 웨어하우스 구축 방법론을 토대로 그 특성 요인별로 비교·분석하였다. 그리고 각각의 방법론들을 간략히 도식화하여 아래에 제시하였다.



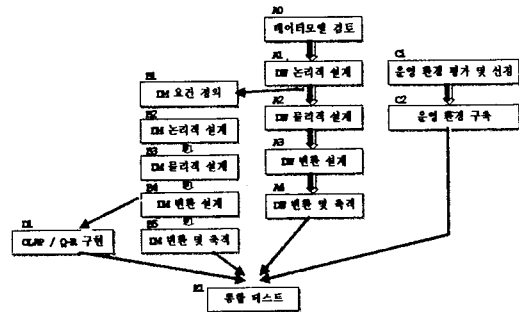
<그림 1> Inmon의 데이터 웨어하우스 구축 방법론



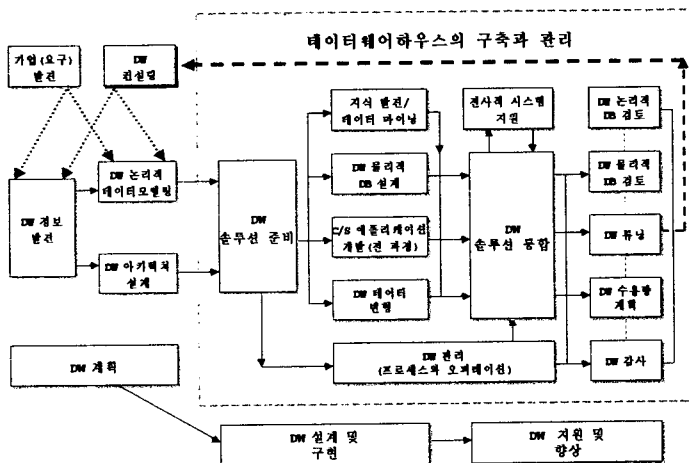
<그림 2> SAS의 데이터 웨어하우스 구축 방법론

- Step 1 사용자 요구조건 (Business Requirement)의 정의
- Step 2 DW 성격의 사전정의 (Pre-definition)
- Step 3 통합환경 구성 (H/W, DBMS, DW 관리도구, Data access tools)
- Step 4 Pilot Project (3개월 - 6개월 정도 소요)
- Step 5 실제 DW구축
 - 5.1 DW Data Model
 - 5.2 Meta data
 - 5.3 Data 통합 프로그램작성
 - 5.4 Operational Service : source file의 추출 및 통합, Data Refresh, Resources 관리
 - 5.5 DW 구축

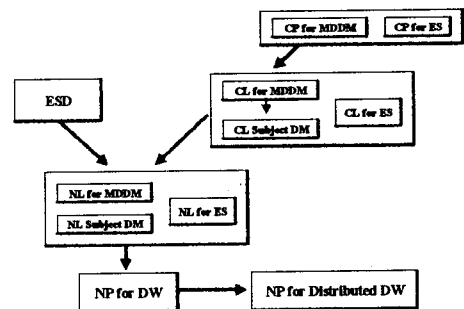
<그림 3> INFORMIX의 데이터 웨어하우스 구축 방법론



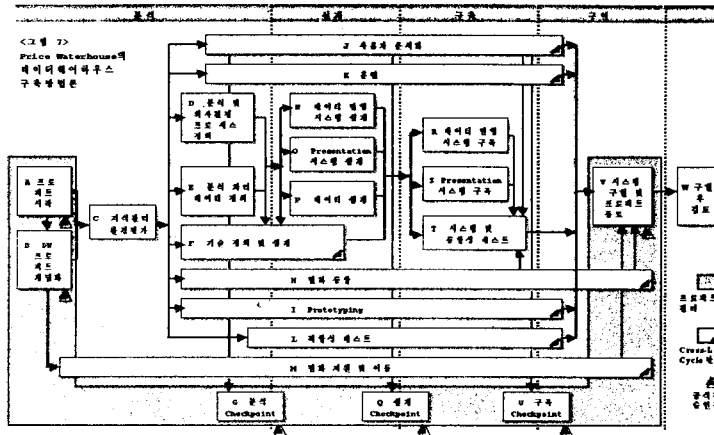
<그림 4> IBM의 데이터 웨어하우스 구축 방법론



<그림 5> NCR의 데이터 웨어하우스 구축 방법론



<그림 6> 데이터베이스 마케팅을 위한 데이터 웨어하우스 구축 방법론 (박주석 교수)



위 7가지 방법론들을 유형별로 그 특성을 요약하면 다음과 같다.

방법론 유형	Inmon	SDS	INFORMIX	IBM	PW	NCR	박주석
구축 내용	구축	구축	구축, 관리	구축, 운영	구축, 관리	구축, 관리	구축
구축 목표	DW	DW	DW	DW + Data Mart	DW	DW + Data Mining	DW
규모(범위)	주제영역	전사적	주제영역	주제영역	전사적	전사적	주제영역 + 전사적
구축 기간	중·장기	중·장기	중·장기	장기	중·장기	장기	중기
구축 비용	상	상	상	상	상	상	중
구축 위험	상	하	하	하	하	하	하
전략 관점	무	유	무	유	유	유	유
모델링 관점	데이터관점 + 프로세스관점 + 조직관점	데이터관점	데이터관점	데이터관점	데이터관점	데이터관점	데이터관점
분석 시점	데이터모델	사용자 요구사항	사용자 요구사항	데이터모델	사용자 요구사항	사용자 요구사항	데이터모델 + 사용자 요구사항
도구 사용	종속적	종속적	종속적	종속적	종속적	종속적	독립적
Pilot	무	유	유	유	유	유	무
Feedback	무	유	무	무	무	유	무
구조적 상·하향식	NA	NA	NA	하향식	NA	NA	NA
방법론적 상·하향식	상향식	하향식	하향식	상향식 + 하향식	하향식	하향식	상향식 + 하향식

<표 1> 7가지 데이터 웨어하우스 방법론의 유형별 특성

위 <표1>에서 보는 바와 같이, 가장 주목할 만한 점은 IBM의 방법론을 제외하고 다른 방법론에서는 구조적 상향식·하향식 접근 방법-데이터 웨어하우스를 먼저 구축할 것이나, 데이터 마트를 먼저 구축할 것이나-이 제시되어 있지 않다는 것이다. 또한, 각각의 특성 요인별로 방법론 상의 차이는 크게 나타나지 않았다. 그러나 특성 요인별로 나타난 각 방법론들의 특징을 요약하면 다음과 같다.

Inmon의 방법론에서 나타난 가장 주요한 특징은 다른 방법론과는 달리 데이터, 프로세스, 조직이라는 모델링 관점이 제시되어 있다. 또한 일반적인 전사적 관점의 데이터 웨어하우스 구축을 위한 방법론이라고 할 수 있다. 그리고 특성 요인별 비교·분석에서 나타난 바와 같이, SDS, INFORMIX, PW의 방법론은 구축 내용, 구축 범위, 전략의 관점, Feedback 등에서 서로 차이가 있을 뿐, 어느 정도 유사한 점이 많은 방법론이라고 말할 수 있다. 단지 INFORMIX와 PW의 방법론은 구축(DW 아키텍처)의 관점과 관리의 관점을 가지고 있다는 점에서 의미가 있다.

IBM의 방법론은 앞에서 언급한 바와 같이, 데이터 웨어하우스와 데이터 마트의 관점이 제시된 하향식 접근 방식의 형태를 가지고 있으며, 또한 방법론적인 면에서는 상향식 및 하향식 접근을 취하고 있다는 점이 다른 방법론과 다른 가장 큰 차이점이다.

NCR의 방법론에서 제시된 가장 큰 특징은 데이터 웨어하우스의 구축과 더불어 데이터 마이닝 기법을 모델화한다는 점이다. 이는 데이터 웨어하우스의 구축을 통해 경쟁력있는 정보를 얻기 위한 전략적 측면까지 고려한 것이라 말할 수 있다.

마지막으로, 박주석 교수의 방법론에서는 주제 영역과 전사적인 측면을 구축 범위로 하고, 데이터 모델과 사용자의 요구사항을 분석 시점으로 한다는 점, 방법론적인 면에서 상향식, 하향식 접근을 취하고 있다는 점이 가장 큰 특징이라고 할 수 있다.

결론적으로 말하면, 위에서 제시된 방법론들은 각각 그 특성 요인별로 유사점과 차이점이 나타나 있으므로, 어떠한 방법론이 데이터 웨어하우스를 구축하기 위한 방법론으로 더 유리하다고 말할 수 없으며, 그 나름대로의 특징을 통해 기업의 전략과 요구 및 환경에 맞는 방법론의 선택이 필요하다고 하겠다.

IV. 연구의 한계 및 결론

지금까지 7개의 방법론을 통해 상호 차이가 있는 특성 요인을 찾아내고 비교·분석하였다. 그러나, 방법론상에서 나타나 있는 상세함의 부족으로 깊은 분석에는 이르지 못했다는 연구의 한계를 가지고 있다. 이는 기업에서 가지고 있는 자신의 방법론의 외부 유출을 꺼려한다는 점과 실제 위 방법론을 통한 데이터 웨어하우스 구축을 직접적으로 경험한 구성원과의 인터뷰나 커뮤니케이션을 통해서 그 나름대로의 관점을 살펴보는 점이 부족했기 때문이다.

그러나 실제로 기업에서 사용하고 있는 방법론을 통해 이를 비교·분석하여 그 특성요인들을 살펴봄으로써, 향후 데이터 웨어하우스 구축을 고려 중에 있는 기업들에게 필요한 지침이나 기준, 방향성 등을 제시하기 위한 향후 실증연구의 토대가 될

것이다. 아울러 데이터 웨어하우스 구축의 수학적 모델의 연구가 이루어진다면 더욱 의미있는 연구가 될 것이다.

참고문헌

- [1] Ronald S.S., "Planning and Architecting a SCALABLE DATA WAREHOUSE", DCI's Data Warehouse World Conference Proceeding, May 1998, pp.65-108.
- [2] George Z., "Data Warehousing - Client/Server for the Rest of the Decade", Prentice Hall, 1997, Vol 2, p.8.
- [3] Paul S.B., "Mining for Profits in the Data Warehouse", Prentice Hall, 1997, Vol 2, p.149.
- [4] Peter B., "March of the Data Marts", DBMS, March 1997.
- [5] Dick M., "The FAQs on Data Warehousing", BANK Marketing, 1997.4.
- [6] Inmon W.H., "The Data Warehouse and Data Mining", Communications of the ACM, 1996, Vol 39, No.11, p.49.
- [7] Charles B.D., "How To Integrate your DATA WAREHOUSE", Datamation, May 1996.
- [8] 조재희, "데이터 웨어하우징을 통한 기업 정보 분석환경 리엔지니어링", SIGDM '98 발표자료집, 1998.6.
- [9] 김관호, "통합 DB 이용해 의사결정 시간·비용 절약", 컴퓨터월드, 1997.12, p.289.
- [10] 이지훈, "모든 길은 데이터 웨어하우징으로", 경영과 컴퓨터, 1997.11.
- [11] 장동인, "데이터 웨어하우스의 최신기술 동향", 경영과 컴퓨터, 1997.10, pp.50-53.
- [12] 이성주, "데이터 웨어하우징 : 정보화 투자는 '비용'이라는 관념부터 깨야한다", 1997.9, <http://cio.seoul.kr>.
- [13] 조재희, "데이터 웨어하우징과 기업정보시스템", 정보과학회지, 1997.5, 제15권 제5호.
- [14] _____, "막상 프로젝트하기에는 아직 이해가 부족", 1997.4, <http://cio.seoul.kr>.
- [15] 박주석, 장인혁, "데이터베이스 마케팅을 위한 데이터 웨어하우스 구축방법론", 경희대학교 대학원, 1997.2. pp.37-44.
- [16] 정철홍, "Data Warehouse", 경희대학교 데이터모델링 및 DB 전문가 양성 교육 자료집, 1997.1.
- [17] 조재희, 박성진, "데이터 웨어하우징과 OLAP", 대청출판사, 1997, pp.91-95.
- [18] _____, "사용목적에 따라 선택한다", 1996.12, <http://cio.seoul.kr>.
- [19] _____, "Data warehousing with INFORMIX", 한국 인포믹스 마케팅부 자료집, 1995.
- [20] _____, "System Management Methodology", Price Waterhouse, pp.31-647.