

□ 기술개설 □

데이터 웨어하우스 시장 현황 및 향후 전망

삼성SDS 정철흠

1. 서론

우리나라에서도 몇 군데에서 데이터 웨어하우스 구축이 완료되었고 여러 곳에서 추진되고 있다. 삼성전자 마케팅, 쌍방울 등에서 구축이 완료되었고 현대, LG, 삼성생명, 한국통신, 이동통신, 산업은행 등 많은 조직에서 추진되고 있다. 국내의 데이터 웨어하우스 시장은 이제 검토기를 지나 도입기에 접어들었으며 다양한 도구(데이터 추출, OLAP도구, DBMS 등) 벤더와 컨설팅 및 SI(System Integrator)업체(삼성 SDS, SHL, LG 등)가 시장을 주도하고 있다. 기업에서도 경쟁력 강화를 위한 도구로 활용하려 하고 있고 나아가 최신 정보 기술과의 연계를 통한 전사적 정보 시스템으로의 발전까지 고려하고 있다.

본격적인 도입에 앞서 시행착오를 줄일 수 있도록 지금까지의 사례들을 분석하여 우리나라의 실정이 맞는 데이터 웨어하우스 구축방안을 제시하려고 한다. 그리고 향후를 대비할 수 있도록 발전 방향을 전망한다.

2. 구축사례 고찰

삼성 SDS가 직접 추진한 1) SDS Pilot Project, 2) 보험사 Pilot Project, 3) 전자회사 마케팅 Project를 고찰 대상으로 삼았다.

1) 국내에서 처음 시도된 데이터 웨어하우스 프로젝트이었으며 보험사 데이터를 활용하였고 Informix와 공동으로 수행하였다. 이미 여러 경로를 통하여 발표되었으므로 구체적인 설명은 생략한다[3][4].

2) Informix가 ROLAP도구인 메타큐브를 사용하고(그림 1), Oracle이 MOLAP도구인 Express를 사용한(그림 2) ROLA-

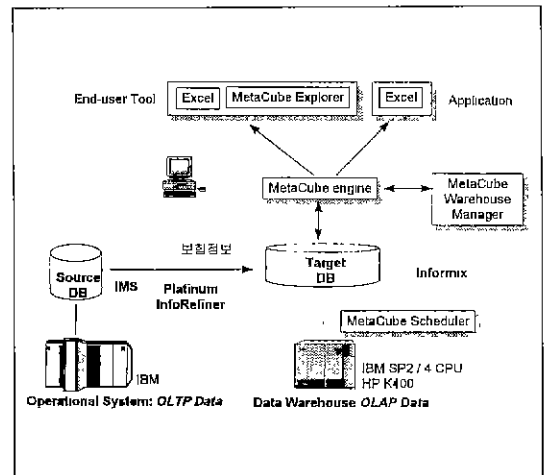


그림 1 Informix의 구축 예

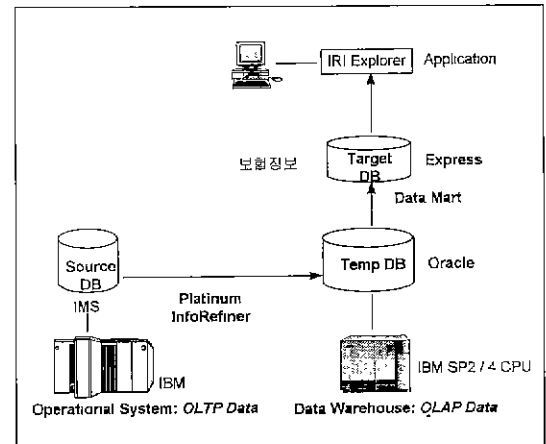


그림 2 Oracle의 구축 예

P대 MOLAP의 벤치마킹이었다. 웨어하우스 DB의 크기는 50GB정도였고 주제는 손익 정보 등이었다. 결과는 세부적으로 밝힐 수 없지만 ROLAP과 MOALP의 장단점들이 정확히 파악된 사례였다.

- 3) Oracle과 DSS Agent를 사용하여 웹 기반으로 구축하였다. 데이터 추출은 EDA/SQL을 사용하였고 매출 정보 등 다양한 주제를 다루었다. DSS Agent는 NT 기반으로 동작하고 웨어하우스 DB용 시스템은 IBM이고 AIX기반이다.

이들 사례에서 몇 가지 주목할 만한 결과가 나타났다. 이 중에는 이미 알려진 사항들도 있지만 주요 주제별로 정리하였다. 평가가 필요한 부분은 나름대로 합리적인 기준을 정하고 다양한 관점에서 평가하였다.

2.1 다차원 데이터 모델링

다차원 분석을 위한 데이터 모델링은 속도 향상을 위해서 반 정규화, 요약 테이블 등의 방법을 사용한다. 일반적인 모델링 방법으로는 스타 스키마와 스노우플레이크 스키마가 있다. 이들을 어떤 방식으로 활용하는가에 대해 살펴본다.

첫째, 스타 스키마나 스노우플레이크 스키마는 서로 어떤 성능 차이가 있을까? 무조건 스노우플레이크 스키마가 스타 스키마보다 성능이 우수하다고 판단하는 것은 금물이다[1]. 어느 스키마를 사용할 것인가는 DBMS의 질의 최적화 기법과 레코드 건수, H/W성능 등을 고려하여 정해야 한다. 그리고 스타 스키마나 스노우플레이크 스키마를 사용한다고 해서 통일해서 사용할 필요는 없다. 일부는 스타 스키마를 사용하고, 나머지는 스노우플레이크 스키마를 사용하는 혼합 형태를 사용할 수도 있다. 그러나 이러한 경우에는 틀마다 일관성이 없기 때문에 혼동될 우려가 있으므로 주의하여야 한다.

둘째, 요약 정도에 따라 어느 정도의 성능 차이가 있을까? 요약 테이블을 많이 만들수록 속도는 빨라진다[2]. 그러나 요약을 많이 할수록 속도는 빠르지만 가공 비용이 비싸지고 데이터 량이 늘어나게 되므로 튜닝 과정에서

적절한 요약 수준을 정해야 한다.

2.2 사례를 통해 본 구축시 유의사항

모델링, 추진 단계, 도구, 활용 및 데이터 관리방법 등에 있어서 기존의 MIS와는 다른 접근 방법이 필요하다. 데이터 웨어하우스의 특성상 반복적인 작업이 필요하고 사용자가 주도하는 형태로 추진하는 것이 효과적이다. 그리고 다음과 같은 현실적인 문제도 고려해야 한다.

첫째, 빈번한 조직 변동이다. 대부분의 데이터 웨어하우스는 조직 또는 부서를 기본적인 차원(dimension)으로 사용하고 있기 때문에 차원의 변화는 웨어하우스 DB자체의 재 구성이 필요함을 의미한다. 따라서 데이터 관리가 어렵고 사용자에게 일관된 활용 방법을 제공하기 어렵다. 장기적인 조직 운영과 함께 기술적으로도 시제열 데이터 처리와 같은 적극적인 해결책이 필요하다. DBMS벤더들의 적극적인 기술개발이 요구된다.

둘째, 대부분의 경영자는 실시간 정보까지 요구한다. 따라서 ODS(Operational Data Store)까지 고려해야 한다. 일반적인 데이터 웨어하우스 방법론으로 접근해서는 우리 사용자들의 욕구를 완전히 충족시키기 어렵다. 다양한 방법론을 가지고 처음부터 정확한 실상을 알리면서 사용자 스스로 선택해 가는 접근 방법이 필요하다.

셋째, 조직 전체의 통합된 뷰를 원한다. 따라서 방대한 주제(subject)를 다루어야 하고 한꺼번에 해결하는 것보다 단계적으로 해결해 가려는 노력이 요구된다. 물론 데이터 웨어하우스로만으로 요구를 모두 해결할 수 없기 때문에 다른 기술과의 연동이 필요하다. 아래 그림 3에서 제시한 향후의 정보관리 체계 모형이 한 가지 방안이 될 수 있다.

3. 발전 전망

데이터 웨어하우스는 이제 유용성을 논할 단계는 지났다. 얼마나 효율적으로 구축하고 활용하는가가 과제이다. 초기의 OLAP중심의 활용에서 데이터 마이닝(data mining)분야로 급격히 확대될 것이며 EIS(Executive Informa-

lion System)와 DSS(Decision Support System)의 근간으로 자리잡게 될 것이다. 세부적인 기술 관점에서 향후의 발전 전망에 대해 살펴본다.

3.1 OLAP

이제 ROLAP(Relational OLAP), MOLAP(Multidimensional OLAP)간의 유용성에 대한 논쟁은 서로가 장단점이 있다는 쪽으로 의견이 모아졌고 양 쪽 진영이 서로 보완적인 관계로 발전하고 있다. 그러나 여전히 각 제품군별 벤더들은 자사 솔루션을 차별화하기 위하여 노력하고 있다.

DBMS벤더는 나름대로 ROLAP, MOLAP, 나아가 DOLAP(Database OLAP)까지 포함하는 통합된 솔루션을 추구하고 있다. DOLAP은 OLAP 엔진의 기능이 DBMS엔진에 내장되는 형태이며 DB로부터 레코드들을 가져온 후 OLAP엔진이 큐브 형태로 변환하여 사용자에게 전달하던 기존 방식을 개선하여 DB에서 곧바로 큐브 형태로 전달하는 기술이다. Informix는 DataBlade기술로, Oracle은 Data Cartridge기술을 기반으로 추진하고 있다.

Microstrategy등 독자적인 OLAP기술 벤더는 멀티 DBMS, 웹 연동 솔루션 개발, 특화 DBMS(예 : Red Brick)와의 연대 등 나름대로의 생존 전략을 추구하고 있다.

3.2 DBMS

DBMS벤더들은 대용량의 웨어하우스 DB를 효과적으로 관리하기 위하여 다양한 인덱싱 방법, 병렬처리와 같은 성능향상 방안을 강구해 가고 있다. 나아가 웹 기술의 활성화에 힘입어 앞서 말한 프로그인 기술(DataBlade, Data Cartridge)들을 기반으로 멀티미디어 DB로의 발전을 염두에 둔 다양한 데이터 타입 지원과 표준 및 오픈 아키텍처를 지향하고 있다.

3.3 웹

데이터 웨어하우스는 향후 OLAP보다는 요약·통계 정보에 대한 데이터 액세스 측면의 활용이 많아질 것이다. 웹 기술의 발전에 따라 데이터 액세스도 웹을 기반하게 될 것이고 웹

과 데이터 웨어하우스의 연동 기술이 계속 발전하고 있다. 웹 기반의 OLAP도구, 웹 출판(web publishing), 보고서 생성기 등의 발전이 이루어지고 있다.

3.4 그룹웨어와의 연계

그룹웨어 제품들은 기존의 메일 백본 방식에서 DB기반 및 웹 기반으로 발전해 가고 있다. 이와 병행하여 데이터 웨어하우스의 요약·통계 정보를 그룹웨어 기반으로 사용자들에게 전달하고 있다. Notes는 데이터 웨어하우스 정보를 Notes서버에 올려(pumping) 사용자들에게 제공하는 기술을 선보이고 있다.

3.5 에이전트

데이터 웨어하우스의 유용한 정보를 보다 능동적으로 사용자에게 전달하기 위하여 에이전트 기능과의 통합이 시도될 것이다. 예를 들면 특정 조건을 정해두고 이를 만족시키는 이벤트가 발생하면 알러트 또는 메일을 자동으로 보낼 수 있다.

3.6 EIS(Enterprise Information System)

향후의 정보기술 분야는 콘텐츠를 중요시할 것이다. 데이터 웨어하우스도 조직 내의 기간 MIS데이터 통합이라는 측면에서 다른 분야로부터 유입되는 다양한 비정형 정보들과의 통합을 이루어 새로운 정보공유 체계로 발전할 것이다.

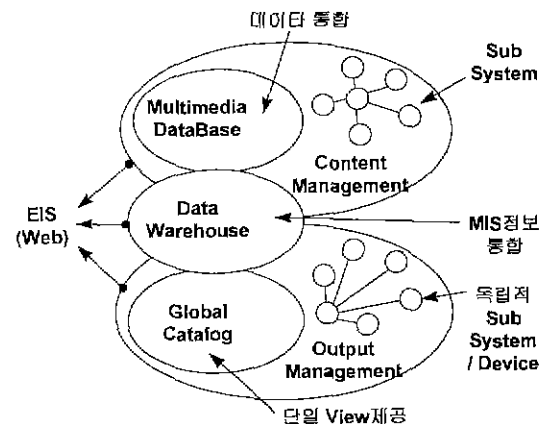


그림 3 향후의 정보관리 체계 모형

MIS데이터는 데이터 웨어하우스로 통합하고 비정형 정보는 멀티미디어 데이터베이스로 통합해 갈 것이다. 아울러 완전한 데이터 통합이 어려운 경우 산출물 관리(Output Management)기술에 의해 연합될 것으로 전망된다. 특히 단일 뷰(single view), 단일 접속(single login)이 요구될 것이고 언제 어디서든 액세스가 가능하고 충분한 정보제공과 쉬운 인터페이스 제공을 위하여 웹을 기반으로 통합될 것이다. 이를 위해 삼성SDS는 데이터 웨어하우스 솔루션과 함께 DIARS(Digital Information Archiving and Retrieval System)라는 정보 공유 플랫폼을 개발하여 EIS(Enterprise Information System)사업을 추진하고 있다.

4. 결 론

국내 시장 동향, 삼성SDS가 추진해온 데이터 웨어하우스 사업의 개략적인 개요와 추진 과정에서 표출된 문제점들, 구축 시 고려 사항들을 살펴 보았다. 그리고 향후의 데이터 웨어하우스 기술 및 시장 발전 방향에 대한 인식을 SI업체 관점에서 사업추진 전략과 결부시켜 설명하였다.

우리나라에서의 데이터 웨어하우스는 요약·통계 정보의 제공 기능이라는 측면이 활성화의 가장 큰 동기로 자리잡을 것으로 전망된다. 우리나라에서도 지금까지 수많은 EIS/DSS가 구축되어 왔고 대부분 실패로 판명되었다. 그러나 경영자들은 재차 요구할 것이고 실무자는 다시 실패하지 않기 위하여 데이터 웨어하우스의 도입을 검토하고 있다. 이러한 배경에서 EIS/DSS의 밑 그림으로써 데이터 웨어하우스의 적용이 활성화될 것이며 데이터베이스 마케팅, 데이터 마이닝, OLAP 등 다양한 활용 분야가 혼합되어 발전될 것이다.

이제 데이터 웨어하우스의 특성이나 활용분야, 활용 방법에 대해서는 공감대가 형성되었다. 벤더는 나름대로 좀더 우리 실정에 맞는 솔루션을 제시해야 할 것이고 사용자는 데이터 웨어하우스의 특성을 잘 파악하여 전략적인 도구로 활용하려는 자세가 필요하다. 특히 모든 것을 해결해 주지 않는다는 점을 인식하여 용

의주도하게 벤더와 기술자와의 협력을 주도해야 한다.

그리고 데이터 웨어하우스를 단순한 요약/통계/OLAP/EIS/DSS를 위한 수단으로 보지 않고 조직의 경쟁력을 강화하기 위한 EIS(Enterprise Information System)의 한 요소로 인식할 필요가 있다. 데이터 웨어하우스 기술이 널리 활용되기도 전에 웹, 그룹웨어, 정보공유 시스템 등이 활성화되고 있기 때문에 향후의 추세에 맞으면서도 단계적인 접근이 요구된다.

참고문헌

- [1] 정철흠, "데이터 웨어하우스 기술적 구성 요소," '96국제학술대회, 정보처리학회, Jun. 1996.
- [2] 정철흠, "데이터 웨어하우스 기술 및 동향," 산학협동프로그램, 경희대학교, Aug. 1996, 1997.
- [3] 정철흠 외, "데이터 웨어하우스 튜토리얼," 삼성SDS, Aug. 1996.
- [4] 정철흠 외, "데이터 웨어하우스 자료집 I, II, III," 삼성SDS, May 1996.
- [5] 정철흠, 관계형 DBMS기반의 데이터 웨어하우스 구축에 관한 연구, 삼성 SDS, Oct. 1996.
- [6] W.H.Inmon. Building the Data Warehouse, A Wiley-QED publication, 1992.
- [7] Vidette Poe, "Building a Data Warehouse," Prentice Hall PTR, 1995.
- [8] Ralph Kimball, "The Data Warehouse Toolkit," John Wiley & Sons, Inc. 1996.
- [9] Patricia Seybold Group, "Open Information System," Vol.9, No8, Aug. 1994.
- [10] Gartner Group, "Housing the Data Warehouse," SDM, June. 1995.
- [11] MicroStrategy, "Decision Support Viewpoint," MicroStrategy white paper, 1994.
- [12] MicroStrategy, "The Case for Relational OLAP," MicroStrategy white paper, 1994.
- [13] Tandem, "Parallel Processing for Deci-

- sion Support," Tandem white paper, 1994.
- [14] Business Objects, "Data Warehousing," Business Objects, 1995.
 - [15] SAS, "A Checklist for Success," SAS Institute Inc. 1996.
 - [16] DBMS, "Data Warehouse Architect," 1995 10~12, 1996 1~6월호.
 - [17] DATAMATION, "Data Warehouse with an OLAP View," 1996 4월호.
 - [18] 삼성SDS, "Technology Trends," 1996 4월호.



정 철 hym

1981 중앙대학교 컴퓨터공학과 졸업
 1983 중앙대학교 대학원 컴퓨터공학과 졸업
 1983~1988 생산기술연구원 근무
 1989~1994 삼성종합기술원 근무
 1994~현재 삼성 SDS 정보기술연구소 근무
 주요경력 SAITDB개발, CODA 개발 참여, ReportMaster 개발, DIARS 개

발 등
 관심분야 : DBMS, Data Warehouse, Information Retrieval, Digital Library, 멀티미디어 DB

● ISAAC '97 학술대회 논문모집 ●

Eighth Annual International Symposium on Algorithms and Computation

- 응모분야 : 알고리즘과 계산이론에 관련된 모든 분야
- 일 자 : 1997. 12. 17~19
- 장 소 : 싱가포르
- 제출마감 : 1997. 5. 30
- 제 출 처 : Hon Wai Leong (Email : leonghw@lscs.nus.sg)
 Department of Information Systems and Computer Science
 National University of Singapore, Singapore 119260
- 문 의 처 : Hon Wai Leong