

데이터 웨어하우징과 기업정보시스템

광운대학교 조재희

1. 서 론

이제까지 기업의 전산자원은 거래처리시스템 (TPS) 개발과 이 시스템으로부터 수집되는 현황정보를 조직 내에 공유하기 위한 경영정보시스템 (MIS) 및 임원정보시스템 (EIS) 개발에 집중투자 되어왔다. 그러나 의사결정을 내리려면 보다 분석적이고 다양한 정보가 제공되어야 하는데, 현재의 기업정보시스템은 이러한 요구에 부응하지 못하고 있는 실정이다. 이런 맥락에서 볼 때, 구미에서 활발히 보급되고 있는 데이터웨어하우징 솔루션과 이를 기반으로 대두되기 시작한 실시간 업무분석(OLAP: On-Line Analytical Processing; '3.2 데이터마트와 실시간 업무분석' 참조)에 관한 연구는 기업의 경영진과 정보담당 부서원들에게 시사하는 바 크다[3, 5, 6, 10].

1995년에 데이터베이스 시장이 'OLTP 시장과 데이터웨어하우징 시장으로 양분되고 있다'는 Kimball의 주장을 인터넷에서 접했을 당시, 데이터웨어하우징 솔루션이 국내에 미처 상륙하기 전이었기 때문에 그 말의 의미를 체감할 수 없었다[14]. 현재 국내외 데이터웨어하우징 발전동향과 흐름을 조망해 보면, 그가 말한 추세가 여실히 드러나고 있음을 알 수 있다. 데이터웨어하우스 구축 프로젝트는 인터넷을 이용한 기업통신망 및 전자상거래 구축 프로젝트와 더불어 국내기업들이 의욕적으로 계획 또는 추진 중인 핵심 프로젝트이다. 작년과 올해 상반기의 '문제 이해기'와 '솔루션 이해기'를 거쳐 올해 하반기가 되면 '솔루션 적용기'로 접어들게 되어 많은 수의 프로젝트가 진행될 전망

이다. 이에 대비하여 국내 SI업체 및 IT 컨설팅 회사들이 구미의 데이터웨어하우징 및 OLAP용 소프트웨어 벤더들과 계약을 맺고 적극적으로 마케팅이나 기술지원 체제를 정비하고 있다.

주지하다시피 OLTP 시스템은 판매자와 고객간의 거래가 순조롭게 진행되도록 신속하게 거래데이터를 처리하는 것이 목적이다. 항상 데이터의 현재성이 강조되며, 따라서 갱신된 데이터가 반드시 실시간에 반영되어야만 하는, 즉 데이터의 휘발성이 매우 높은 데이터의 집합이다. 반면 데이터웨어하우스는 경영상에 발생하는 여러 의사결정 상황을 지원하기 위하여 데이터를 일정주기로 계속 축적하는 저장고로서, 추세분석이나 기타 비교분석을 가능케 한다. 그러나 예를 들어, 고객별로, 상품별로, 지역별로, 판매경로별로 데이터를 일정한 주기로 축적하다 보면 DB의 크기가 상당히 커지게 되는 단점이 있다.

한편, 은행의 입출금 시스템이나 편의점/백화점의 POS 시스템과 같은 OLTP 시스템이 없으면 비즈니스 운영조차 할 수 없게 되지만, 데이터웨어하우스라는 의사결정용 DB를 구축하지 않았다고 해서 영업을 마비된다거나 생산라인이 멈추게 되지는 않는다. 그러나 효과적인 마케팅 활동, 경쟁력 있는 제품 및 서비스 개발, 적시에 정확한 의사결정 등을 수행하려면 이러한 정보기반이 필요하다. 이상을 요약하면 표 1과 같다.

본 논문에서는 첫째, 데이터웨어하우징 솔루션의 목표, 정의 및 출현배경은 무엇이며, 둘째, 기업전산 및 정보분석 환경이 어떻게 변화

표 1 OLTP와 데이터 웨어하우스 비교

	OLTP	데이터 웨어하우스
목적	데이터 수집 (getting the data in)	데이터 분석 (getting the data out)
적용업무	거래처리	의사결정 지원
내용	업무필 데이터 집합	주제밀 데이터 집합
시간성	현재	과거/현재
상세정도	상세	요약/상세
취발성	강함(자주 갱신)	없음(스냅샷 보관)
가치	비즈니스 운영	비즈니스 선도

하는지, 셋째, 어떠한 아키텍처와 요소기술이 사용되는지, 마지막으로 기업경영에 미치는 효과는 무엇인지에 관하여 논의해 본다.

2. 데이터웨어하우징과 기업 정보분석 환경 리엔지니어링

‘데이터웨어하우징’이란 용어는 매우 포괄적인 용어로서, 데이터웨어하우스나 데이터마트와 같은 의사결정지원용 데이터저장고를 구축 또는 사용하는 일련의 과정을 일컫는다. Orr는 데이터웨어하우징을 “질 높고 통합된 전사데이터를 비전문적 최종사용자와 전문적인 사용자가 공히 쉽게 접근할 수 있는 제반시설을 구축하는 과정”이라고 정의하며, Kelly는 “기업 정보인프라와 아키텍처를 구축하는 과정”이라고 정의한다[13, 17]. 필자는 “OLAP을 구현하기 위한 정보분석환경 리엔지니어링 과정”이라고 정의하고자 한다.

일반적으로 우리가 데이터저장고란 의미로 사용하는 데이터웨어하우스는 Inmon과 Hackathorn의 정의가 가장 최초이자 널리 알려져 있는데, 그들은 “기업의사결정 과정을 지원하기 위한 주제 중심적이고, 통합적이며, 시간성을 가지는 비휘발성 자료의 집합”이라고 정의하고[15], ‘Data Warehouse World’라는 컨퍼런스를 주최하는 메타그룹에서는 “의사결정 지원용으로 특별히 설계된 주제 중심적인 정보저장고”라고 정의하며, Kelly는 그의 저서에서 “기업 내의 의사결정지원 어플리케이션들을 위한 정보기반을 제공하는 하나의 통합된 데이터저장공간”이라고 정의한다[12]. 이상의 정의들

이 공통적으로 포함하고 있는 단어는 ‘의사결정’이며, 데이터웨어하우징의 목표가 의사결정 지원이라는 사실을 선명하게 나타내고 있다.

2.1 기업정보시스템의 문제점

기업경영환경은 매우 다차원적이다. 즉, 고객, 시장, 제품, 생산공장, 부품, 공급자, 광고매체, 판매경로, 생산기술, 그리고 정보기술 등과 같은 여러 경영요소 중 한가지 측면에 국한되어 문제가 발생하는 것이 아니라 복수의 요소들이 복합적으로 작용하여 의사결정을 한층 어렵게 만든다. 그러나 현재의 기업정보시스템은 각 문제상황에 관련 있는 측면들에 관한 다양하고 심층적인 정보를 제공해 주지 못하는 데, 그 주된 원인은 정보인프라의 취약성이라고 할 수 있다.

이러한 기업정보기반, 즉 ‘의사결정용 데이터베이스’가 부재한 상황에서 진행되어 온 국내 기업의 정보시스템 포트폴리오의 문제점은 그림 1의 오른쪽 그림과 같이 기본수 형태를 나타내고 있다는 점이다. 효과적인 DSS를 통해 정보분석활동이 활발히 이루어져 왔다면 EIS에 제공될 정보가 현재처럼 과다해지지 않았을 것이다. 즉, EIS를 통해 경영진에게 제공하는 정보들이 정보분석이라는 일종의 여과과정을 거치지 않았기 때문에 임원이 보는 정보의 질이 MIS를 통해 관리자에게 제공되는 정보의 질과 차이가 없는 기현상이 발생하게 되었다.

정보기반의 부재로 인한 두번째 문제점은 최종사용자 컴퓨팅(EUC)이라는 명목 아래 전사적인 데이터 통합 노력 없이 필요할 때마다 한시적으로 또는 일회성으로 분석활동이 이루어

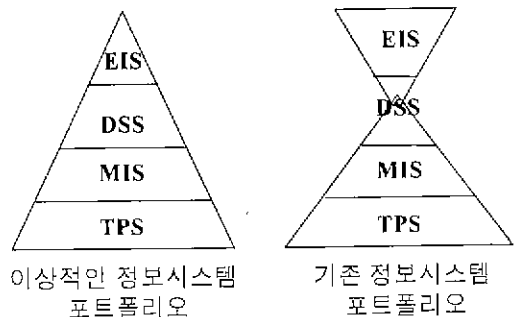


그림 1 기업정보시스템 포트폴리오 문제점

져 왔다는 점이다. 사용자의 요청에 의해 OLTP 데이터가 데스크탑DB나 서버DB에 내려졌고, 이에 따라 데이터의 불일치, 전산실의 지원능력 한계, 부서간의 컴퓨팅 기술 격차심화 등 여러 가지 문제점들이 노정 되었다.

2.2 데이터웨어하우징 출현배경

데이터웨어하우징 솔루션은 위에 언급한 문제점들과 기업 데이터 자산의 급격한 증가로 인한 문제점을 해결하려고 출현했다고 볼 수 있다. 1970년대만 하더라도 미국 500대 기업의 평균 데이터 량이 8 기가바이트에 불과 했으나, 1980년대에는 350 기가바이트, 1990년대에는 28 테라바이트로 급증하고 있으며, 2000년대에는 400 테라바이트로 증가될 전망이다. 이렇게 날로 급증하는 다량의 데이터를 효과적으로 분석하여 정보화하기 위해서는 기존의 분석 방식, 즉 필요한 데이터를 전산실에 요청하여 받거나, 정보시스템에 디스플레이 되는 데이터를 하드 카피하여 엑셀 등과 같은 사용자 틀에 재 입력하여 데이터를 분석하는 방식은 지양되어야 한다. 이러한 정보분석환경에서는 경쟁력 있는 기업이 탄생 될 수 없다.

2000년대가 되어 데이터 량이 더욱 방대해지면, 도저히 사용자에게 의해 일일이 데이터가 분석되어질 수 없다. 이때가 되면 데이터마이닝 기술을 도입하여 컴퓨터에 의한 지식발견(데이터 안에 내재되어 있는 패턴이나 추세, 또는 데이터간의 상관관계를 찾는 방식)으로 분석방식의 일대 전환이 예상된다. 따라서 그림 2와

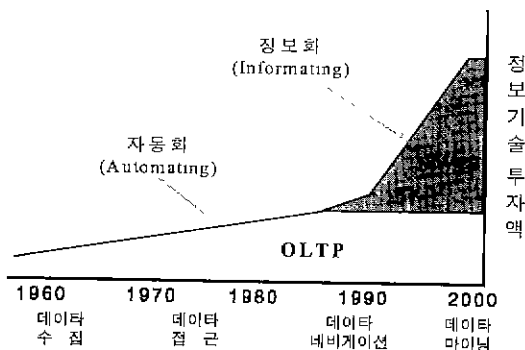


그림 2 정보기술 투자 추세변화

같이 시간이 흐를수록 정보기술에 대한 투자 중에서 OLTP에 대한 투자는 일정 수준에 머물고, 데이터웨어하우징 분야(네비게이션과 마이닝)에 대한 투자는 기하급수적으로 증가하는 정보기술 투자패턴의 변화가 예상된다[12].

성숙된 전산 및 정보분석 환경도 데이터웨어하우징 출현을 가능하게 한 또 다른 배경으로서 작용했다고 볼 수 있다. 기업들이 클라이언트/서버 환경을 구축함으로써 다량의 데이터 공유 및 분배 능력이 이전의 중앙집중식 호스트 환경보다 용이해 졌고, EUC용 툴들의 분석 기능은 물론 데이터 관리능력이 향상되었다. 따라서 최종사용자는 이제 어느 정도의 정보욕구는 전산실의 지원 없이 직접 해결할 수 있게 되었고, 이러한 경험에서 생긴 일종의 자신감 덕분에 다량의 데이터에 최종사용자의 접근을 허용하고 스스로 분석하게 하는 정보기반을 제공하기에 이르렀다.

이상은 기술적인 배경이라고 할 수 있는데, 이 외에도 경영 전략적인 배경도 강조되고 있다. 즉, 상품중심의 기업으로부터 고객중심의 기업으로 변신을 시도하려는 기업들에게 '고객 DB' (어떤 구입자가, 언제, 어디서, 무슨 상품을, 어떤 이유로, 어떻게 구입했는지에 관한 정보가 담긴 DB)라는 주제의 데이터웨어하우스는 필수적인 전략무기로 인식되고 있으며, 현재 진행 중인 거의 모든 데이터웨어하우징 프로젝트의 공통적인 핵심주제는 '고객'이라는 사실이 이를 반증하고 있다. 고객 또는 소비자에 관한 정보(demographic, behavioral, psychographic 정보)는 날로 경쟁이 치열해지는 금융회사, 자동차회사, 통신회사, 유통회사 등에게 매우 중요한 정보가 아닐 수 없다. 이러한 소비자 정보가 POS 데이터와 연결되어 분석되면 구매패턴의 변화(예를 들어, 성별, 연령별, 소득별, 직업별 등)가 파악될 수 있으며, 이 정보를 이용하여 기존 고객을 보호함과 동시에 새로운 고객을 확보한다는 전략적인 목적을 배경에 깔고 있다.

2.3 정보분석환경 리엔지니어링의 필요성

오늘날의 경영환경은 과거보다 더욱 경쟁적이고 역동적이다. 세계시장에서 굴지의 기업들

과 경쟁을 벌이고 있는 국내 기업들은 하루가 다르게 증폭되는 변화의 물결을 체험하게 된다. 이렇게 급박한 시대일수록 기업의 사활은 그들의 정보시스템의 상황적응능력과 정보를 분석/취합하는 능력에 따라 좌우되며, 따라서 조직 구성원은 점차 데이터 분석에 과거보다 많은 시간을 할애하게 된다. 이전에는 고도로 숙련된 요원만이 경영전략 데이터를 분석하였고, 그렇게 획득한 정보를 의사결정에 반영하거나 조직 내에 전달하는 역할을 하였다. 90년대 후반부터는 기업의 거의 모든 직원이 데이터 분석업무를 수행할 것이다. 분석 프로세싱은 계속적으로 조직에 퍼짐으로써 경영자에게 적시에 전략적인 방향을 제시해 줄 것이다. 따라서, 기업이 경쟁에서 우위를 차지하는 능력은 중국적으로는 기업이 가지고 있는 분석 프로세싱 능력의 질과 효율성, 그리고 얼마만큼 조직 내에서 활성화되어 있는가에 달려 있다고 해도 과언이 아니다.

기업에서 데이터를 분석할 때는 거래 처리할 때와는 달리 한꺼번에 다량의 데이터에 접근해야 하고, 업무요소들간의 다양한 관계를 규명해야 하며, 분석을 위한 질의문도 복잡하게 표현된다. 뿐만 아니라 조직에서 필요로 하는 정보는 부서마다 정도의 차이는 있겠으나 매우 역동적이어서 그때그때마다 해당 자료를 공급하기가 어렵다. 의사결정을 하기 위해서는 당면한 문제의 전체적인 윤곽이나 출처를 파악하여야 하며 그러기 위해서는 보고서나 그래프의 축을 손쉽게 거의 무작위로 바꿔볼 수 있어야 한다[2]. 기존의 보고체계나 정보시스템은 이런 요구에 부응하지 못하고 있는 실정인데, 이러한 문제점을 극복하기 위해 의사결정용 DB를 구축하고 이를 기반으로 OLAP을 실현하는 체제로의 전환, 즉 데이터웨어하우징을 통한 정보분석환경 리엔지니어링이 필요하다.

3. 데이터웨어하우징 솔루션과 의사결정지원

3.1 데이터웨어하우징 프레임워크 구축 가이드라인

한마디로 데이터웨어하우징은 전산실이 보유

하고 있는 OLTP 데이터를 조직의 여러 최종 사용자들이 이해하기 쉽고 사용하기 편리한 형태로 변환하는 과정이라고 볼 수 있다. 최종 사용자가 분석업무를 원활히 수행하기 위해서는 OLTP 시스템과 분리되어 작업해야 된다. 물론 원시데이터는 OLTP 데이터베이스에서 가지고 오지만 거래처리 업무에 방해가 되지 않도록 앞서 언급한 데이터웨어하우스라고 불리는 분석용 DB를 별도로 구축해야 한다는 이론이다. 그림 3은 가장 널리 인용되는 데이터웨어하우징 프레임워크인데, 분석업무를 효과적으로 지원하기 위해 여러 가지 층으로 이루어진다. 참고로 '데이터마트 층'의 원통형 모양은 관계형 DB를 의미하며, 직육면체 모양은 다차원DB를 표시한다.

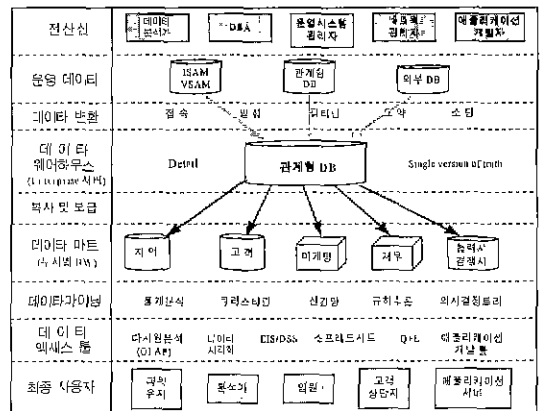


그림 3 데이터웨어하우징 프레임워크

우리가 일반적으로 지칭하는 데이터웨어하우스는 이 프레임워크를 구성하는 여러 층 중의 하나인 'Enterprise 서버 층'이다. 주로 고객정보를 기반으로 하여 데이터베이스 마케팅에 활용하기 위해 구축되는 데이터웨어하우스는 일정한 주기로 계속 과거 데이터를 축적하기 때문에 대용량 데이터베이스가 될 수밖에 없는데, 최초의 데이터웨어하우스로서 인구에 회자되고 있는 MCI사의 고객 데이터웨어하우스의 현재 크기는 4.3 테라바이트나 된다고 한다[1]. 이렇게 대용량DB가 되면 분석 프로세싱을 수행하는데 있어서 응답속도, 분석의 깊이와 폭 등에 제한을 받게 된다.

이러한 단점을 보완하기 위해서 데이터마트

라는 층을 별도로 구축한다. 데이터웨어하우스와 데이터마트에 관한 비교 설명은 '3.2 데이터마트와 실시간 업무분석'에서 자세히 다루고 있다. 데이터웨어하우징 프로젝트의 초창기에는 '데이터 액세스 툴'에 의한 정보전달이 우선 순위이지만, 미국 기업과 같이 데이터웨어하우스가 이미 사용되고 있는 곳에서는 '데이터마케팅' 층이 새로운 정보분석 방법론으로서 대두되고 있다.

그림 4는 필자의 데이터웨어하우징 구축 경험과 이제까지 수집한 각 벤더들의 자료를 기초로 작성한 다섯 가지 구축 방법이다. 복수의 방법을 혼합하여 구축하는 것이 보편적인데, 예를 들어, 1번 방법과 2번 방법(데이터웨어하우스와 데이터마트 공히 관계형DB로 구축) 또는 1번 방법과 3번 방법(데이터웨어하우스는 관계형DB로, 데이터마트는 다차원DB로 구축)이 혼합된 구축 시나리오가 있을 수 있다.

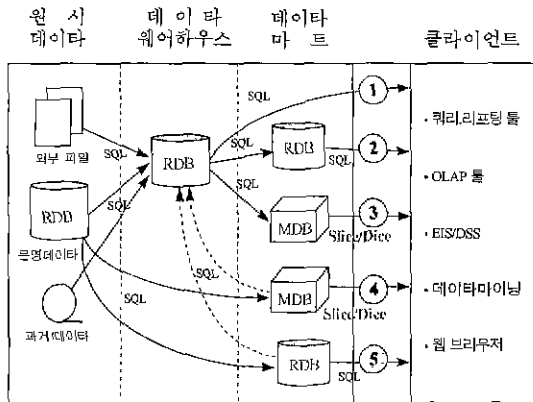


그림 4 데이터웨어하우징 구축 가이드라인

기업에서는 업무의 복잡도, 전산환경, 사용자의 컴퓨팅 스킬 등과 같은 자사의 여건에 따라 여러 시나리오 중에서 취사선택하여 구축하게 된다[15]. 예를 들어, 분석 업무량이 많고, 전산자원도 풍부할 경우는 데이터웨어하우스와 데이터마트를 동시에 구축하여 역할을 분담시키고, 중소기업이나 단위 부서와 같이 데이터량이 많지 않거나 자원이 부족할 경우에는 데이터마트 층만 구축하는 것도 솔루션이 될 수 있다. 일반적으로 데이터웨어하우징 구축 초창기에는 데이터웨어하우스나 데이터마트 둘 중

에 하나를 먼저 구축하여 정보기반을 갖추는 방법도 있다. 최근 들어 4번과 5번 방법처럼 데이터마트를 먼저 구축하는 기업도 생기는데, 단기간 안에 정보기반의 가치를 먼저 가시화해 보려는 발상에서 비롯된다.

아마도 데이터웨어하우징 관련 요소기술 선정과정 중에서 가장 고심의 대상이 되는 부분이 타겟DB 선정 문제일 것이다. '데이터웨어하우스 층'의 타겟DB에는 논란이 일지 않으나, '데이터마트 층'의 타겟DB 선정과정에는 다차원DB라는 다소 생소한 형태의 데이터베이스와 이 DB가 제공하는 응답속도와 분석기능의 우수성 때문에 프로젝트 담당자들은 고심하게 된다. 데이터웨어하우징 프로젝트 담당자는 타겟DB를 선정할 때, 관계형DB와 다차원DB 간에 존재하는 현격한 기능 및 역할의 차이를 이해하고, 이 두 종류의 DB를 배타적이지 아니라, 상호보완적으로 보급해야 한다. 타겟DB 선정 기준을 몇 가지 소개하면 표 2와 같은데, 다른 기준들은 논란의 소지가 있지만, 2번 DB크기 (capacity)에 관한 기준과 4번 응답속도(performance)에 관한 기준은 일반적으로 수긍되고 있다.

표 2 타겟DB 선정기준

	관계형DB	다차원DB
1. 데이터의 휘발성이 높을 경우	○	
2. 데이터베이스의 크기가 클 경우	○	
3. 한번에 접근하는 데이터 량이 많은 경우		○
4. 사용자가 빠른 응답속도를 요구할 경우		○
5. 분석의 복잡도가 높을 경우		○
6. 사용자의 업무스킬 수준이 높을 경우		○

3.2 데이터마트와 실시간 업무분석 (OLAP)

데이터마트는 의사결정지원과 관련하여 주목할 만한 솔루션인데, 전사적인 통합데이터 서버인 데이터웨어하우스와는 달리 특정주제 혹은 부서 단위 데이터저장고라고 할 수 있으며, 수주일에서 수개월 내로 구축할 수 있다는 장점이 있다. 크기는 약 100 기가바이트 미만으

로서 데이터웨어하우스보다 크기는 작지만 더욱 복잡한 분석업무를 수행한다는 차별성이 강조된다. 또한 데이터마트는 데이터웨어하우스와는 달리 읽기뿐만 아니라 쓰기도 가능한 환경을 제공한다. 표 3은 데이터웨어하우스와 데이터마트의 차이점을 요약한 것이다.

표 3 데이터웨어하우스와 데이터마트 비교

	데이터웨어하우스	데이터마트
목표	잠재적인 모든 유형의 질의에 대처	특화된 분석지원
특성	데이터 저장고	모델링 도구, 연산 엔진
질의유형	읽기	읽기/쓰기
응답속도	질의 유형에 따라 가변적	일관성, 신속성
내용	과거, 현재	과거, 현재, 미래
자료구조	비정규화, 평면적	다차원적, 계층적
데이터량	초대량, 매우 상세	대량, 상세
구축기간	수개월~수년	수주일~수개월

그림 5는 데이터웨어하우스와 데이터마트의 역할을 비교한 것이다. OLTP 시스템으로 거래 데이터를 주기적으로 계속 데이터웨어하우스에 저장한 후, 일상적이고 반복적인 질의나 보고서 작성은 여기서 이루어진다. 예산을 계산한 다거나 예측과 같은 복잡한 분석은 사용자의 입력과 그에 따른 결과를 여러 차례 계산해 볼 필요가 있게 되는데, 이러한 업무는 부서별로 데이터마트를 구축하여 별도로 실시하도록 한다.

OLAP은 “최종사용자가 대규모 기업 데이터에 직접 접근해 능동적으로 정보를 분석하는 과정”이라고 정의할 수 있다. OLAP 서버는

“다차원 데이터 구조를 운영하기 위해 디자인된 고성능의 다수 사용자용 데이터 분석 엔진”이라고 정의할 수 있으며, 다차원DB를 기반으로 개발되어야 한다는 주장이 일반적인 주장이다 [6, 8, 9, 16]. OLAP 서버에 대한 정의는 연구자나 소프트웨어 하우스마다 이견이 있지만, 다차원DB가 내장된 OLAP 서버를 이용하여 경영의사결정에 필요한 정확하고 신속한 정보를 제공하고, 현재 우수측순차등 재생하는 기업의 EUC 체제를 체계화하는 정보기반으로서 자리매김하려는 움직임이 지배적이다.

OLAP에 관한 연구는 이론적이기보다는 실무 지향적인 연구이다. 이 연구는 1980년대 초에 소개되기 시작한 DSS의 실패를 거울 삼아 “새로운 전산/정보기술로 정보를 분석하여 기업 내에 배분하려는 전산화 노력”, 또는 “분석 프로세스를 조직 내에 활성화시키기 위한 노력”이라고 해석 할 수 있다[3]. 25년전, 12가지 관계형DB 원칙을 발표했던 Codd 박사가 이번에는 12가지 OLAP 법칙을 발표했는데, 그는 관계형DB가 유연성 있는 데이터 분석에는 적합치 않다고 인정하며, 최종사용자들의 주업무인 분석 프로세스를 위해서 필요한 기능을 선정하였다[6]. 무엇보다도 다차원 뷰를 제공하여야 하며, 클라이언트/서버 환경에서 작동되어야 하며, 직관적인 데이터 조작 기능, 복수 사용자의 동시접속 등을 갖추어야 한다는 것이다. Codd 박사의 이 연구발표는 특정 소프트웨어 회사의 연구비를 받고 수행되었기 때문에 발표논문 내용에 관해 많은 논란을 일으키기도 하였지만, OLAP 기술 도입에 관심을 가지고 있는 기업은 이 기능들을 OLAP 툴 선정시 제품평가 가이드라인으로 사용하고 있다.

4. 결론 : 기업경쟁우위 확보

아직 실증적이거나 계량적인 데이터웨어하우스, 데이터마트, OLAP의 효과 분석은 발표된 바 없으나, EIS/DSS와 EUC를 기업 내에 활성화시키기 위한 정보분석 기반으로 보는 시각과 데이터 자산의 전략적 이용 가능성에 대한 견해에는 이견이 없다. 데이터웨어하우스 프로젝트에 의해 정보분석환경을 재 구축한 기업은

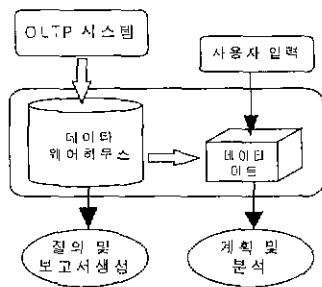


그림 5 데이터웨어하우스와 데이터마트 역할비교

다음과 같은 이점을 얻게 된다. 첫째, 쉽게 다양한 각도로 업무를 분석할 수 있기 때문에 기업의 관리자는 항상 선명한 경영 상황을 볼 수 있다. 둘째, 내장되어 있는 분석 기능과 리포팅 기능을 이용하여 최종사용자 혼자 처리할 수 있는 업무의 량이 증대될 수 있다. 셋째, 데이터를 새로운 관점에서 보게 됨으로써 새로운 아이디어나 대안을 마련할 수 있게 된다. 넷째, OLAP이 제공하는 다차원분석을 통해 사용자는 종합표나 비즈니스 차트를 통해 데이터의 여러 관점과 계층을 향해함으로써 중요한 추세, 또는 문제점의 출처를 발견할 수 있도록 해준다. 다섯째, 부서 내에서 필요한 데이터를 업무별로 구조화함으로써 관리항목의 범위를 명확히 하고, 보고서 작성을 최소화할 수 있다 [3]. 이상과 같이 데이터웨어하우징이 최종사용자 부서와 전산 부서에 각각 미치는 영향을 정리해 보면 표 4와 같다.

표 4 데이터웨어하우징의 부서별 효과

최종사용자 분석	전산 부서
전산 부서에 대한 의존도 감소	OLTP 시스템 작업량 감소
사용 편리 및 다양한 분석 수행	SQL 프로그래밍 작업량 감소
원하는 정보에 신속히 접근	애플리케이션 개발 부담 감소
워크그룹 생산성 증대	정기 리포팅 출력물 생산량 감소
업무 프로세스 개선	유지보수 비용 절감
환경변화에 신속히 대응	기투자된 시스템의 가치 상승

표 5는 유럽의 데이터웨어하우징 사용자를 대상으로 설문 조사한 결과로서, 기업들이 어떠한 목적을 가지고 데이터웨어하우징 기술을 적용하는지를 알려 주는 유용한 자료이다. 응답율은 각 항목에 대하여 해당 순위를 매긴 응답자 수를 전체 응답자 수로 나눈 비율이다.

데이터웨어하우징 솔루션과 관련된 주요이슈 중의 하나는 인터넷 기술과의 통합 움직임이다. 인터넷만큼 저렴한 비용으로 전세계 어느 곳으로나 통신 및 정보 공유를 가능하게 해주는 통신망은 없기 때문에, 기업들은 데이터웨어하우스에 저장되어 있는 지식을 원격지에 있는 해외지사, 공장, 고객, 또는 협력사에게 학습시킬 수단으로서 웹기술을 이용하려고 한다. 특히 고객이나 협력사에게 데이터웨어하우스

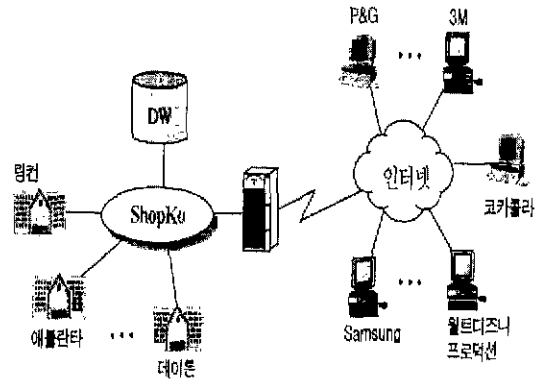


그림 6 ShopKo사의 전략적 데이터웨어하우스 활용 사례

표 5 데이터웨어하우징 현재 효과와 기대효과

순위	현재 효과	응답율	기대 효과	응답율
1	사용자에게 직접 데이터 제공	52%	기업 경쟁력 향상	42%
2	새로운 시장기회 발견	37%	하나의 일관된 데이터 제공	34%
3	고객기반에 관한 이해	33%	사용자에게 직접 데이터 제공	28%
4	기업 경쟁력 향상	33%	데이터 자산의 효율적 이용	23%
5	하나의 일관된 데이터 제공	33%	새로운 시장기회 발견	22%
6	양질의 정보 제공	32%	의사결정정보에 관한 전산실 업무감소	20%
7	데이터 자산의 효율적 이용	24%	양질의 정보 제공	17%
8	의사결정정보에 관한 전산실 업무감소	19%	비즈니스 프로세스 리엔지니어링	16%
9	조직내의 의사소통 장벽 제거	16%	과거의 낡은 정보시스템 리엔지니어링	8%
10	전산환경 질서확립	16%	전산환경 질서확립	7%

자료 : Data Warehouse Report, 1995년 겨울호

정보를 제공하는 서비스는 그림 6의 ShopKo 사 사례처럼 매우 전략적인 의미를 지니게 된다.

ShopKo사는 미국의 유통회사(retailer)로서 전국적으로 130개 매장이 있으며, 수천여 협력사로부터 상품을 구매하여 소비자에게 판매하고 있다. 이들이 다루는 상품 품목 수는 200,000가지이며, 일 평균 거래 데이터는 수만 건에 달하는데, 이렇게 방대한 분량의 POS 데이터를 인터넷을 통해 협력사에 공개하려고 시스템을 구축 중이다. 저장용량이 400 기가바이트로 예상되는 이 데이터웨어하우스에는 2년간의 POS 데이터(일별, 매장별, 상품별 매출 데이터)가 담길 예정인데, 이렇게 되면 제조회사인 협력사들은 그 동안 가장 아쉬웠던 자사 제품들에 관한 판매정보를 다이내믹하게 제공받게 된다[11]. ShopKo사는 다른 유통회사가 제공하지 못하는 이러한 정보서비스를 협력사에게 제공함으로써 경쟁사들보다 유리한 조건으로 상품을 구매하겠다는 전략적인 목적을 갖고 있다.

이제 방대한 기업데이터와 복잡하게 느껴지던 업무구조/프로세스를 데이터웨어하우스 또는 데이터마트리는 정보기반으로 체계화하고, 이를 기반으로 OLAP을 실현하는 것은 더 이상 지연할 수 없는 필수 불가결한 리엔지니어링 과제이다. 왜냐하면 정보기반이 갖춰진 기업은 첫째, 고객 행동양식에 대한 이해가 증진되어 더욱 경쟁력 있는 상품과 서비스를 생산하게 되고, 둘째, 비즈니스 프로세스의 개선 또는 재창출을 통해 경영 효율성이 향상되며, 셋째, 최종사용자가 기업 데이터에 직접 접근하여 유연성 있게 분석업무를 수행하므로 업무생산성이 제고되기 때문이다. 한마디로 데이터웨어하우스에 의한 정보분석환경 리엔지니어링 프로젝트는 기업의 경쟁우위 획득과 확보를 위한 원천이 될 것이며, 체계적인 이론이나 프레임워크, 포괄적인 아키텍처, 우수한 클라이언트/서버 환경의 틀에 의해 뒷받침되고 있기 때문에 기업정보기반의 표준으로써 자리매김할 가능성이 높다고 전망한다.

참고문헌

- [1] 인포믹스, “데이터웨어하우스 구축사례 : 미국MCI 고객정보시스템”, DW 길타잡이, 1996.
- [2] 조재희, “최종사용자를 위한 다차원 데이터베이스”, 한국경영정보학회 EIS/DSS 연구분과위원회 학술대회 논문집, 1994년 12월, pp. 115-125.
- [3] 조재희, “기업의 정보기반 구축을 위한 OLAP 서버”. 한국경영정보학회 춘계학술대회 논문집, 1995년 6월, pp. 563-577.
- [4] 조재희, “튜토리얼 : 데이터웨어하우징 : 기술, 응용, 그리고 미래”, 한국경영정보학회 추계학술대회 논문집, 1996년 12월, pp. 29-40.
- [5] 조재희, 박성진, 데이터웨어하우징과 OLAP, 대청정보시스템, 1996년 8월.
- [6] Codd, E. F., S. B. Codd, and C. T. Salley, ‘Providing OLAP to User-Analysts : An IT Mandate,’ White Paper, Codd & Date Inc. 1993.
- [7] Crandall, Richard L., “Multi-Dimensionality in a Decision Support System,” White Paper, Comshare, 1983.
- [8] Dorrian, Jim, “OLAP-The Multi-dimensional Approach to Data Analysis,” *Mini-Micro Systems*, April 1994, pp. 10-11.
- [9] Dresner, Howard, “Multidimensionality : Ready or Not, Here it Comes,” OIS Research Note, Gartner Group, June 3, 1993.
- [10] Finkelstein, Richard, “Understanding the Need for On-Line Analytical Servers,” White Paper, Comshare, 1994.
- [11] Foley, John and Bruce Caldwell, “Dangerous Data,” *InformationWeek*, Sept. 30, 1996.
- [12] Kelly, Sean, *Data Warehousing : The Route to Mass Customisation*, John Wiley & Sons, New York, 1994.
- [13] _____, “The Second Wave is

Volatile! Operational Data Warehousing," *DCP's Data Warehouse World Conference Proceedings*, August 1996, pp. D1-1~D1-11.

- [14] Kimball, Ralph, "The Database Market Splits," *DBMS Magazine*, September 1995, pp. 12, 17.
- [15] Inmon, W. H., and R. D. Hackathorn, *Using the Data Warehouse*, John Wiley and Sons, New York, 1992.
- [16] Light, Matt, "On-Line Analytic Processing : Getting the Most From the Data Warehouse," *Inside Gartner Group This Week*, January 25, 1995, pp. 8-11.
- [17] Orr, Ken, "Data Warehousing : Phase 2," *DCP's Data Warehouse World Conference Proceedings*, August 1996, pp. C31-1~C31-50.



조 재 희

1993~1994 LG-EDS 기술연구소 컨설턴트
 1994~현재 광운대학교 경영정보학과 조교수
 1997~현재 경영정보학과 학과장
 관심분야 : 메이타웨어하우징(다차원DB, 데이터마트, OLAP, 웹과 데이터웨어하우스 통합), 데이터마이닝, 의사결정지원시스템.

● WAAC '97 학술대회 논문모집 ●

JAPAN-KOREA Joint Workshop '97 on Algorithms and Computation

- 응모분야 : 알고리즘과 계산이론에 관련된 모든 분야
- 일 자 : 1997. 7. 25~26
- 장 소 : 일본 후쿠오카
- 제출마감 : 1997. 5. 15
- 제출처 : Takao Assno (Email : waac 97@ise.chuo-u.ac.jp)
 Dept. of Information and System Engineering
 Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112, Japan
 Tel : +81-3-3817-1686 Fax : +81-3-3817-1681
- 주 최 : 일본정보처리학회(IPSJ) SIGAL, 한국정보과학회 컴퓨터이론연구회
- 문 의 처 : 좌경룡 (Email : kyohwa@jupiter, kaist.ac.kr)
 Tel : 042-869-3513