

1. 데이터베이스 시스템의 그릇된 인식

데이터베이스의 보편화

정보화 사회로 일컬어지는 현대 사회에서 사용되고 있는 컴퓨터 관련 용어 중 '데이터베이스'라는 단어는 아마도 가장 보편적으로 사용되는 기술적인 단어 중 하나일 것으로 생각된다. 과학 계산용으로 개발된 컴퓨터가 비즈니스에 사용되기 시작한 60년대에는 느린 진전을 보이던 일상업무 전산화가 데이터베이스 시스템 기술의 출현과 발전으로 인해 빠른 속도로 진척되었는데 오늘날 데이터베이스라는 용어가 보편화된 이유

데이터베이스 시스템 기술 및 시장 동향

최정현

광운대학교 경영정보학과 교수

중 하나가 특별히 경영의 전산화일 것이다.

경영의 많은 부분이 전산화 되면서 경영 활동에서 발생하는 자료를 저장하고 관리하고 경영에 필요한 정보들을 가공해서 만들어 내는데 있어서 데이터베이스 시스템 기술은 우수한 방법과 도구를 제공해 주었고, 경영 정보의 이용에 제반 단계를 컴퓨터가 신속하고 정확하게 처리해 주는 경영의 혁신을 이루었다.

반면에 경영 자료들은 데이터베이스 시스템에 적합한 형태로 전문적 가공과정을 통해 전통적 모습에서 변형되어야 한다는 부가적인 노력이 필요하게 되었다.

잘못된 지식

그러나 '데이터베이스'라는 용어의 보편성에도 불구하고 데이터베이스를 이해하는 데에 있어

전문가들조차도 데이터베이스의 기술적 특징을 무시하고 흔히 컴퓨터에 입력된 많은 양의 자료라는 애매한 의미로 데이터베이스를 잘못 이해하므로서 데이터베이스와 관련된 모든 기술을 포함하는 데이터베이스 시스템의 중요성을 인식하지 못하고 또한 데이터베이스 시스템의 구축에도 신중한 계획이 없이 많은 오류를 범하는 것이 현실이다. 데이터베이스를 구축하고 응용 프로그램을 작성하는 전문적인 인력들도 데이터베이스 설계 기술을 도외시한 채 데이터베이스 시스템을 구축하므로서 4GL(4th Generation Language)을 위한 화일 시스템 수준 정도로 데이터베이스 시스템을 그릇 인식하는 것이 또한 현실이다. 대표적인 예를 들면 매출전표에 기술된 항목들의 레코드를 컴퓨터에 입력시킨 화일을 데이터베이스 관리시스템을 사용해서 저장하고 이를 Powerbuilder 등과 같은 4GL을 사용해서 처리하는 정도를 가지고 데이터베이스 시스템이라 부르는 경우 등이다.

데이터베이스 시스템 기술을 이해하는 것은 먼저 데이터베이스의 기술적 특징을 정확히 이해하므로서 가능하므로 본 글에서는 데이터베이스의 기술적 특징을 먼저 논의한 후에 데이터베이스 시스템의 역할과 필요한 기술들의 중요 이슈(Issue)들을 논해보겠다. 또 해당 기술적 이슈들의 동향이 어떠하며 그에 따른 시장동향을 관계형 데이터베이스 시스템을 중심으로 논하도록 해보겠다.

2. 데이터베이스 시스템의 필요한 기술

데이터베이스 시스템이란?

데이터베이스 시스템이란 데이터베이스를 설계하고 구축하고 관리하고 가공하고 사용하고 유지하는 등의 데이터베이스와 관련된 모든 기술과 도구 그리고 사용자를 총망라하는 포괄적 용어이

다. 그러나 협의로는 데이터베이스를 저장하고 관리하는 컴퓨터와 관련된 기술 즉 데이터베이스 관리 시스템의 기술만을 지칭하기도 한다.

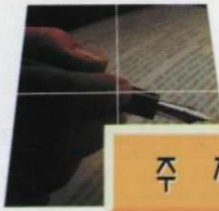
데이터베이스의 출현동기

데이터베이스 시스템의 기술적 의미를 살펴보기 전에 데이터베이스 시스템의 주체가 되는 데이터베이스가 출현했던 배경을 먼저 살펴보자. 60년대 미 국방성을 중심으로 컴퓨터를 일상업무에 사용하려고 시도하면서, 코볼(COBOL)이라는 언어를 만들었고 일상업무의 서류/장부의 자료를 저장하기 위한 효율적 화일 구조를 활발하게 연구했다.

이 노력은 오직 자료를 저장하고 저장된 자료를 읽어서 코볼과 같은 프로그램 언어를 통해 1차원적 가공(보고서 작성 등)만을 허락하는 화일 시스템을 만들었다. 이 화일 시스템은 자료 가공 때마다 하나의 복잡한 응용 프로그램의 작성이 필요하고 응용 프로그램의 복잡성을 줄이기 위해 자료를 변형시킬 필요가 있었는데 변형된 자료는 자연히 해당 응용 프로그램에 종속되기 때문에 동일한 자료가 분산되어 중복되는 현상을 초래했고 이는 정보의 무결성과 일관성을 위협하므로서 정보의 신뢰성이 떨어뜨렸다. 따라서 경영에 필요한 모든 정보를 얻을 수 있다는 실현 불가능한 잠재성만 있을 뿐 현실적으로는 특별한 형태의 자료 가공은 대부분의 경우 특별한 이유로 불가능했다. 이와 같은 문제는 자료의 양과 종류가 늘어나면서 더욱 심각해졌다.

데이터베이스의 기술적 특성

데이터베이스는 이런 문제를 근본적으로 해결하는 열쇠가 되었는데, 데이터베이스는 독립된 여러 자료의 군(群)들이 항목들 사이에 상호 연관성이 있을 경우에 그들의 연관성을 컴퓨터가 현실적으로 인식하고 연관성을 이용해서 자료를



주 제 칼 럼

가공해서 새로운 정보를 만드는 것이 가능하도록 자료 군(群)들을 조직하고 정보의 무결성과 일관성을 유지할 수 있도록 재구성된 서로 연관된 자료의 군(群)들의 집합이다.

이 정의에 대해 본 글의 목적상 이론적으로 보다는 개략적으로 데이터베이스의 특성을 요약해 보면, 첫째, 데이터베이스는 여러 자료의 군(群)들 사이에 공통된 항목을 통해 자료의 군(群)들의 연관성이 표현되어야 하고, 그 연관성에 대한 자유로운 처리가 가능해야 하며, 새로운 모든 정보를 끌어낼 수 있도록 하는 기술이 적용될 수 있어야 한다. 둘째, 데이터베이스는 자료가 최신의 정보를 유지할 수 있도록 자료의 변화가 유기적으로 이루어지되 자료의 일관성과 무결성이 항상 유지할 수 있도록 구성되어 있어야 한다.

데이터베이스 시스템의 기술적 이슈

위에 언급한 첫번째 이슈와 관련된 기술은 데이터베이스의 데이터 모델링이라고 하는데 데이터 모델링은 데이터베이스 설계와 밀접한 관련이 있는 기술이기도 하고 또 한편으로는 데이터베이스의 자료들을 구성하고 저장하는 개념적 방법과 컴퓨터의 저장매체 내부에 저장되는 구조를 결정하는 역할도 담당한다. 데이터 모델링 기술을 따라서 자료를 개념적으로 구성하고 저장매체 내부에 저장하는 것을 자료의 데이터베이스화라고 말한다. 데이터 모델은 크게 망형 모델(Network Model), 계층적 모델(Hierarchical Model), 관계형 모델(Relational Model), 세 가지로 구분할 수 있다.

데이터베이스화된 자료를 사용해서 필요한 모든 정보를 가공해 내고 또 유지하고 관리하기 위해서 데이터베이스 관리 시스템이라는 관리 소프트웨어 즉 DBMS가 데이터베이스 시스템을 운용할 때에 중요한 도구로서 필요하다. DBMS는 데이터베이스 시스템의 사용자들에게 데이터베이

스 시스템의 편리성과 신뢰성과 중요성을 실감하게 하는 위치에 있다.

DBMS와 관련된 기술적 이슈가 데이터베이스 시스템 기술의 핵심일 것이다. DBMS의 기술적 이슈는 첫째, 자료를 가공하는 명령어인 질의어의 처리 기술 둘째, 자료의 유기적인 변화에 일관성과 무결성을 갖도록 하는 보장하는 트랜잭션 처리 기술 셋째, 자료의 안전을 보장하기 위한 데이터베이스 백업(Backup) 기술, 마지막으로 응용 프로그램과의 인터페이스를 위한 개발툴 및 애플리케이션 기술 등으로 크게 구분할 수 있다.

좀더 광의로 데이터베이스 시스템을 이해한다면 업무의 데이터베이스화 하기 위해서 요구되는 업무분석 기술도 데이터베이스 시스템이 다루어야 하는 기술로서 분류할 수 있을 것이다. 이는 주로 엔티티-관계 모델(Entity-Relationship Model)로 대표되는 기술인 데, 이를 근거로 많은 변형된 기법들이 제안되었다.

3. 데이터베이스 시스템의 기술동향

데이터 모델

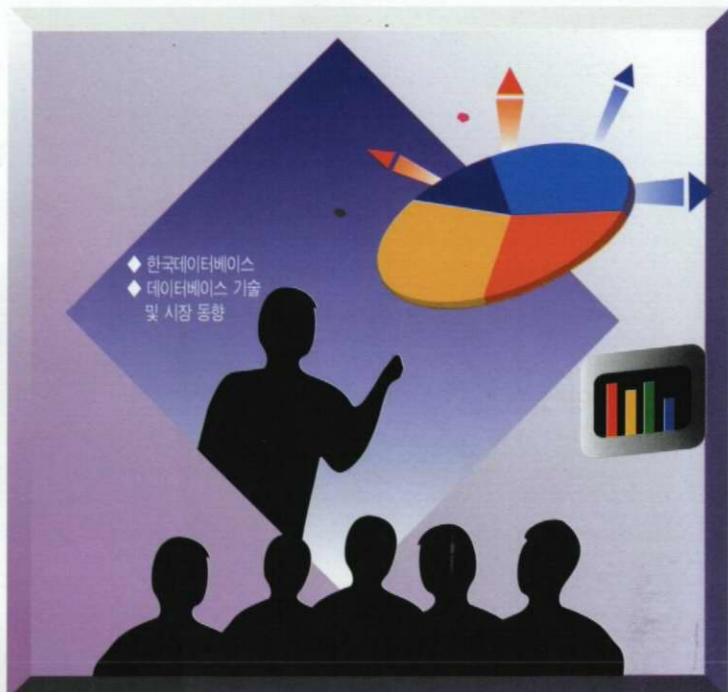
먼저 데이터 모델링 기술을 살펴보자. 필자가 대학 졸업 후 현업에 근무하던 '84년 당시에는 IMS/VS라는 IBM 메인프레임용 계층적 데이터 모델에 근거한 DBMS를 사용했고 당시 IBM 메인프레임을 사용하는 대다수의 한국 기업들이 가장 많이 사용하는 데이터 모델이었다. 당시에는 관계형 데이터 모델은 연구 시제품 또는 초기 상용화 단계 정도에 불과 했고 많은 전문가들도 관계형 데이터 모델에 대한 회의적 반응도 만만치 않았다. 그러나 많은 연구를 거듭한 결과로 '96년 지금은 관계형 데이터베이스 시스템이외에는 거의 상업화된 DBMS가 없는 것이 현실이다.

그 이유는 이러하다. 자료를 데이터베이스화

하는 과정에서 이미 언급한 것처럼 자료들의 연관성 표현해야 하는데 계층적 데이터 모델은 연관성 표현이 포인터 연결 방식으로 이해되므로 소수의 자료 군들 사이의 관계를 표현할 때에는 문제가 없으나 다중의 관계를 많은 자료 군들 사이에 설정하는 경우 예측할 수 없는 연결 포인터의 루핑(Looping), 자료 레코드의 중복 등의 문제가 발생하므로 근본적인 설계변경이 이루어져야 하고 이런 구조적 불안정성은 계층적 데이터 모델에서는 해결할 수 없다. 이를 위해서 망 데이터 모델이 일시 제안되었으나 계층적 데이터 모델의 문제를 해결하기 위해서 구축되는 망 데이터 모델은 설계자의 고려요소, 경험 등의 비 정량적 요소에 따라 너무 다양하고 그 효율성이 천차만별인 결과를 나타내므로 안정적 모델이라고 할 수 없다.

이런 이유 때문에 관계형 데이터 모델이 관련 DBMS의 효율성에도 불구하고 오늘날 가장 많이 사용되는 데이터 모델이 되었다. 관계형 데이터 모델은 업무분석 기술인 Entity-Relationship Model과 거의 1:1의 밀접한 관계가 있으므로 분석과 설계와 구축이 하나로 통일된다는 장점이 있다. 관계형 데이터베이스 설계과정이 정규화라는 이론적 연구로 뒷받침되기 때문에 언제나 동일한 결과를 얻는 장점이 있다. 또 자료의 내부적 표현구조가 테이블이라는 방식으로 나타내므로 자료의 연관성과 자료 항목이 동일하게 처리되는 장점이 있고 화일 시스템의 화일 구조와도 테이블이 그대로 일치되는 장점이 있다.

데이터 모델은 개념적 설계를 변화 없이 그대로 내부적 저장구조에 반영하는 것이 바람직하므로 최근 개념적 설계의 발달로 복잡한 자료 군들의 개념적 관계들, 즉 전문화, 일반화, 카테고리,



하부구조 등등의 관계를 저장구조와 개념적 구조에 반영하기 위해서 최근에 객체지향 데이터 모델이 등장해서 자료 군들 사이의 복잡한 내포관계를 표현하기에 적합한 방법들을 제공하고 있다. 그러나 객체지향 데이터 모델을 그대로 반영하고 처리해 주는 객체지향 DBMS가 연구 단계에 있으므로 아직은 상용화된 제품을 보기가 쉽지 않다.

DBMS 처리기술

DBMS는 어떤 데이터 모델의 개념을 반영하는가에 따라서 망, 계층, 관계형, 객체지향 DBMS 등으로 구분할 수 있겠으나 본 글의 목적상 관계형 DBMS의 기술만을 논하겠다. 상용화 초기에는 관계형 DBMS는 자료의 처리속도가 가장 큰 문제였고 지금도 상용제품의 선정조건이기도 하다. 관계형 DBMS의 처리 속도를 좌우하는 가장 큰 요인은 질의어 처리과정에서 연관 관계를 처리하는 조인(Join) 연산자의 자료를 읽고 쓰는 I/O 시간의 지연이 크다는 것이다. 이를 개



주 제 칼 럼

선하기 위한 기술이 질의어 최적화(Query Optimization) 기술이다.

또 다른 이유는 질의어를 해석해서 필요한 절차적 언어로 변환시키는 질의 해석기가 처리 시간을 소비한다. 그러나 조인 연산의 I/O 시간으로 인한 지연이 지대하므로 상대적으로 질의 해석기의 개선 보다는 질의 최적화가 관계형 DBMS의 성공을 좌우하는 열쇠이다. 질의어는 표준화가 되어서 SQL(System Query Language)/DS라는 표준안이 제안되어 있고 이를 토대로 질의 해석기를 구성하므로 질의어 설계 기술은 이미 정착되어 있다.

현재는 질의를 해석하고 기본적인 관계형 연산을 수행하고 자료를 저장/관리하는 기본적 I/O 시스템을 갖춘 핵심 DBMS 커널을 데이터베이스 엔진이라는 형태로 제안해서 나머지 DBMS의 필요 기술들을 모듈화해서 관계형 DBMS의 외형을 다운사이징 환경에 적응하도록 하는 것이 현재의 추세이다. 주로 부가 모듈들은 트랜잭션 처리, 시스템 백업, 질의어 최적화 등등의 형태로 부분적 모듈 판매방식과 연계해서 관계형 DBMS의 외형을 슬림화 했다. 관계형 DBMS의 전 기능을 전부 사용하려면 고성능의 대형 메인 프레임의 성능을 가지고도 부족한 것이 현실이다.

DBMS 환경기술

DBMS의 운용환경은 과거 10년 사이에 엄청난 변화를 겪었다. 주로 종래에 생각하던 DBMS 운용환경은 중앙 집중식으로 데이터베이스의 모든 자료가 한 사이트에 저장/관리되고 온라인 단말기를 통해서 중앙의 데이터베이스를 접근해서 사용하는 환경이었다.

그러나 통신기술의 발달로 컴퓨터의 분산배치가 이루어지고 자료의 교환이 통신기술에 의해 빠른 속도로 이루어질 수 있으므로 분산 데이터베이스 시스템의 기술이 연구되기 시작했고 현

재 많은 연구로 데이터베이스의 단편들을 빈번히 사용하는 사이트를 중심으로 분산시키면서도 중앙 집중식의 효과를 달성하는 분산 데이터베이스 시스템이 달성되고 있다. 그러나 데이터베이스의 관리적 목적 때문에 많은 호응을 얻지 못하고 있으며, 고속 통신망에 기반을 둔 클라이언트/서버 기술의 발달로 인해 분산 데이터베이스 시스템의 데이터베이스의 분산으로 인한 효과가 미지수이고, 반면에 커다란 프로세싱 오버헤드로 인해 향후 분산 데이터베이스 시스템은 새로운 방향으로 전개되어야 할 것이다. 현재는 3-Tier에 기반을 둔 클라이언트-서버 환경이 DBMS의 주도적 환경기술이 될 가능성이 크다.

4. 관계형 데이터베이스 시스템의 시장동향

데이터베이스 시스템 시장은 PC용 데이터베이스를 제외하고 관계형 데이터베이스 시스템이 주종인 데, 월간 잡지 "컴퓨터월드"의 조사에 의하면 '95년도 상반기에 약 340억 규모로 '94년 대비 약 80%의 성장을 보이는 폭발적 성장을 보이고 있다. 관계형 데이터베이스 시스템의 전체 시장의 80% 이상을 오라클, 인포믹스, 사이베이스의 3사 제품으로 국한된다. 이같은 추세는 앞으로도 계속될 전망이다. 최근 데이터웨어하우징(Data Warehousing)의 바람을 타고 의사결정을 위한 자료분석 및 수집하는 기능이 데이터베이스 시스템과 연계해서 자동으로 이루어지도록 한 제품의 출시로 더욱 가속될 것이다.

향후 데이터베이스 시스템 시장은 4GL과 연계한 애플리케이션 및 개발 툴을 중심으로 확대될 전망이다. 각 기업마다 충분한 데이터가 수집된 오늘날 데이터 가공에 관한 다양한 개념적 기술들이 제안되므로서 정보가공을 위한 애플리케이션이 데이터베이스 시스템의 판매와 직결될 것으로 전망된다. **DC**