

□정보산업동향□

데이터 웨어하우스를 활용한 데이터 마이닝 기술

김 영 남[†]

◆ 목 차 ◆

- | | |
|-------------------|--------------|
| 1 서 론 | 4 활용 분야 |
| 2 용어 정의 | 5 결론 및 향후 과제 |
| 3 데이터 웨어하우스 구성 요소 | |

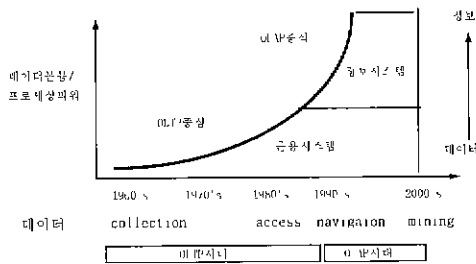
1. 서 론

기업에서의 정보시스템 변화는 크게 3가지로 구분할 수 있다. 첫째로 사용자 요구의 변화이다. 과거 1960,70년대의 메인프레임 시대에서는 컴퓨터 도입에 막대한 비용이 소요되어 비교적 규모가 큰 조직에서 단말기를 네트워크로 연결하여 한정된 소수의 사용자가 전산 부서에서 제공하는 입출력물을 활용하는 전산 환경이었기 때문에 업무의 전산화 주도권은 사용자가 아닌 전산요원이 가졌다. 그러나 1980년대 PC의 보급이 대중화되면서 사용자가 스프레드시트 같은 소프트웨어를 이용하여 직접 데이터를 취급, 처리할 수 있을 뿐만 아니라 데이터를 용이하게 분석할 수 있게 되었다. 즉 전산화의 주역이 전산요원이 아닌 사용자로 변화하게 되었다. 둘째로 정보기술,

특히 반도체 기술의 급격한 발전으로 인하여 컴퓨터 가격이 크게 하락되어 컴퓨터 도입이 대중화되고 정보기술이 공급업체의 독점적인 기술에서 오픈 시스템이라는 개방형 기술로 변화하여 컴퓨터간의 연동이 용이하게 되었으며 또한 네트워크 기술의 혁신적인 발달로 인하여 조직내·외부의 유연한 정보통신 기반을 구축할 수 있게 되었다. 셋째로 기업의 환경 변화이다. 경쟁우위를 유지 또는 선점하기 위해서 정보가 인력, 자본, 물질보다 더 중요한 요소로 인식됨에 따라 단순한 데이터 처리인 운영 시스템에서 타 기업과는 다른 차별적인 정보 분석 업무로 (그림 1)처럼 변화하고 하고 있다. 즉 정보시스템을 활용한 보다 과학적인 의사결정 정보의 필요성이 증가되었다. 그러나 이러한 정보시스템 환경에서 현재 조직 내부의 정보처리는 다음과 같은 현실적인 문제점을 내포하고 있다. 첫째로 분산된 정보자원 중에서 실제로 필요로 하는 데이터가 어느 장소에 저장되고 있는지를 모르고 있고 또한 데이터

[†] 정회원 : 시스템공학연구소 선임연구원

이름, 포맷 등의 데이터 처리 규칙이 상이하여 의사결정에 필요한 정보를 적시에 제공하지 못하고 있으며 둘째로 정보시스템 업무의 60 - 70%가 단순 통계 출력보고서를 작성하는 업무임에도 불구하고 적시에 사용자에게 제공되지 못하는 출력보고서 적체 현상이 나타나고 있다. 셋째로 데이터 분석 업무를 기존의 운영계 시스템과 공존하여 수행하는데 상호간의 업무 성격이 <표 1>과 같이 배타적이어서 데이터 처리에 효율적이지 못하며 넷째로 현행 이질적 전산환경에서의 분산처리 기술은 지역적으로 분산되어 있는 데이터의 무결성 문제를 완전하게 해결하지 못하고 있고 또한 데이터 보안도 취약하여 현실적으로 실용적이고 단순한 데이터 관리 기술이 요구되고 있다. 이러한 제반 문제점을 해결하기 위한 새로운 개념 및 정보기술이 데이터 웨어하우스 및 데이터 마이닝 기술이다.



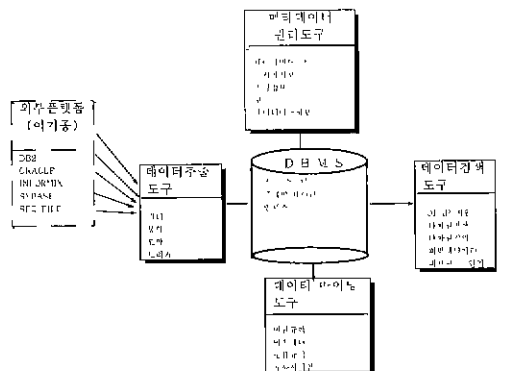
(그림 1) 정보시스템 발전 배경

<표 1> OLTP 와 OLAP 비교

OLTP (Online Transaction Processing)	OLAP (Online Analysis Processing)
정확성/처리 속도 강조	분산성/통합성 강조
현재자료	과거 기간별 축적자료
계속적인 동적 수정	정적
프로세스 지향	주제별 분석 지향
운영 업무	DSS/EIS
한 시점에 적은 데이터 접근	한 시점에 많은 데이터 접근
상세 자료	요약상세 자료
데이터 구조 복잡	데이터 구조 단순
구조적 정형화 검색	AD HOC 질의 검색
내상전체를 관리	주제별 관리
정규화	필요시 비정규화
정적 구조	유연, 확장성 구조
SQL	SQL + 데이터분석(OLAP 도구 이용)

2. 용어 정의

데이터 웨어하우스의 용어 어원은 1980년대 중반 IBM의 인포메이션 웨어하우스에서 출발하여 현재는 데이터 웨어하우스라는 용어로 많이 쓰이고 있고 1992년 인문은 데이터 웨어하우스는 기업의 의사결정 과정을 지원하기 위한 주제 중심적이고 통합적이고 시간성을 가지는 비휘발성 자료의 집합이다 라고 정의하고 있다.[1] 주제 중심적이라는 의미는 기업의 업무별 구분없이 분산, 산재되어 있는 데이터를 영업별, 고객별 등의 업무별, 기능별로 구분하여 데이터를 효율적으로 관리한다는 뜻이고 통합적이라는 의미는 서로 다른 하드웨어, 운영체제에 저장되어 있는 각각의 데이터의 이질적인 문제점을 극복하기 위하여 데이터를 표준화시켜 통합 관리한다는 의미이다.



(그림 2) 데이터 웨어하우스 구성 요소

데이터 마이닝은 대량의 실제 데이터로부터 묵시적이고 미리 알 수는 없지만 잠재적으로 쓸모있는 정보를 추출하는 작업으로 정의한다.[3] 특히 단순 의사결정 정보와의 차별성을 위하여 데이터 마이닝은 대규모 과거 자료의 패턴, 추세를 분석하기 위한 방안으로 사용자가 가설을 제공하고 통계학적으로 결과를 검증하는 기존의 통계학적 개념이 아닌 시스템 내부에서 연관규칙, 연속패턴,

분류규칙, 유사 시계열 등의 알고리즘을 이용하여 사용자의 가설 제공없이 데이터간의 의미를 자동적으로 분석하여 유용한 정보를 추출한다.

3. 데이터 웨어하우스 구성 요소

데이터 웨어하우스 구성 요소로는 (그림 2)와 같이 데이터 추출 도구, 메타 데이터 관리 도구, 데이터 검색 도구, 데이터 마이닝 도구로 구분할 수 있다.

3.1 데이터 추출 도구

한개 이상의 운영계 시스템에서 정확하고 일관성 있게 데이터를 추출하고 목표 시스템에 데이터를 저장하여 주는 도구로서 동일 DBMS 또는 이질적인 DBMS에 저장된 데이터의 형식, 정의 등의 불일치성을 해결하고 데이터를 표준화시키는 방법은 <표 2>와 같이 관련 운영계 데이터의 통합 기능을 수행한다.

3.2 메타 데이터 관리 도구

메타 데이터는 운영계 시스템 구조로부터 데이터 웨어하우스를 독립시키는 정보 사전으로서 데이터 웨어하우스에서는 다음과 같은 정보를 가지고 있다.

- (1)데이터 추출에 관한 정보
 - . 변환 알고리즘(치환, 계산)
 - . 운영시스템에서 데이터 웨어하우스로 매핑
 - . 원시 데이터 식별, 어트리뷰트간 매핑, 전환, 인코딩, 키 변화 등의 정보를 관리
 - . 데이터 추출 이력 정보
- (2)데이터 웨어하우스 관리에 관한 정보
 - . 주제별 정보
 - . 데이터의 보관 주기 및 폐기
 - . 데이터/테이블 간의 관계 정보
 - . 로드 이력 정보
 - . 보안
- (3)데이터 검색에 관한 정보

<표 2> 데이터 추출도구 기능

기능	세 부 사 항
. 변환 - 인코딩 - 이름 변환 - 속성추정	남자 M, 여자 F 또는 남자 1, 여자 0으로 표기한 것을 남자 M, 여자 F로 통일 시키는 기능 KOREA, KOR, 대한민국, 한국으로 표기한 것을 KOREA로 통일 시키는 기능 센티 미터, 인치, 야드 단위를 센티 미터로 통일시키는 기능
. 필터링	DBMS내의 여러 데이터 중 필요한 데이터만 추출하여 데이터를 저장시키는 기능
. 통합	서로 다른 하드웨어, DBMS내의 데이터를 하나의 목표시스템으로 합병, 통합시키는 기능
. 요약	신속한 다차원 검색을 위하여 분기별, 지역별 등으로 데이터를 요약 저장시키는 기능
. 코드 생성	COBOL, C 등의 언어와 JCL, 유닉스 스크립트를 자동적으로 생성하는 기능
. 메타 데이터 기능	개발자, 사용자, 관리자를 위하여 원시 DBMS와 목표 DBMS 사이에 데이터가 어떻게 매핑되었는가를 관리하는 메타 데이터관리 기능
. CASE와의 통합	사용자가 CASE의 리퍼지토리에 있는 데이터 저장소를 자동적으로 추출하여 데이터 추출도구의 메타데이터로 이동하는 기능
. 변경정보 관리	운영계 시스템에서 추출된 모든 데이터의 변경정보를 관리하는 기능
. 양방향 추출	이질적인 DBMS간 데이터를 양방향으로 추출, 변환, 전송하는 기능
. 기타	광범위한 하드웨어, 운영체제 지원 및 GUI 기능

- . 검색 패턴
- . 데이터 웨어하우스 구조
- . 어트리뷰트 변경 내역
- . 드릴다운 검색 정보

이러한 메타 데이터 관리 기능을 이기종 플랫폼간에 호환을 갖기 위한 표준화 작업은 1995년 7월 메타그룹의 일부가 주도하여 메타데이터 협의회가 발족되었으며, Arbor, Business Object, Cognos, EII, Platinum, TI 등의 회사가 참가하여 데이터 웨어하우스 용도의 메타 데이터의 표준화가 진행중이다.

3.3 데이터 웨어하우스의 DBMS 엔진

데이터 웨어하우스 시스템에서 실제 데이터를 저장하는 DBMS는 기존의 관계형 DBMS 외에 다차원 DBMS도 사용되고 있다. 데이터 웨어하우스 성능을 높이기 위하여 병렬 연산, 병렬 유틸리티, 병렬 I/O 등의 병렬처리, 질의 최적, 멀티 조인, 대규모 데이터의 백업 등의 기능이 우수한 DBMS를 사용하여야 하며 데이터 웨어하우스 용도의 DBMS를 선택할 때 고려해야 할 사항으로서는 대용량의 데이터를 추출 저장하는 시간, 보다 더 많은 사용자들을 지원하기 위한 시스템 확장성, 관리자가 프로세싱 자원의 효율성을 크게 하고 성능을 증가 시키기 위하여 요약, 인덱스, 조인등을 변화시킬수 있도록 엔진 질의의 형태,

빈도 등의 모니터링 기능 등이 있다. 기존의 관계형 DBMS와 다차원 DBMS 사이에는 <표 3>과 같은 차이가 있으므로 데이터의 규모, 사용자 인원 등에 따라 적합한 DBMS를 선택하여야 한다.

3.4 데이터 검색 도구

데이터 웨어하우스에서의 데이터 검색 도구로서는 기존의 OLTP와 대조되는 OLAP 개념이 도입되고 있다. 전자는 트랜잭션의 처리에 관계된 복구, 동시 작업 처리에 중점을 두는 반면 후자는 의사결정과 관계된 분석 자료를 얻기 위한 기능에 중점을 두는 처리 과정이다. OLAP은 데이터 웨어하우스의 정보검색 부분의 중심 기능이며 이들 기능을 구현한 도구가 OLAP 도구이다. OLAP 도구는 관계형 DBMS를 기반으로 하여 다중 차원의 뷰를 제공하는 ROLAP 도구와 다차원 DBMS를 기반으로 하는 MOLAP 도구로 나누어지고 있다.[2]

3.4.1 OLAP 도구

ROLAP제품들은 관계형 DBMS를 이용하여 다차원 분석을 수행하기 때문에 질의 응답성 능은 물리적 데이터 모델과 강하게 연관된다. 따라서 응답 성능을 향상 시키기 위해 다양한 최적화 기법이 사용된다. 비정규화, 요약, 분할과 같은 기법을 포함하여 사용자 질의에 최적의 성능을 발휘할 수 있도록 스타 스키마(Star Schema) 모델이나

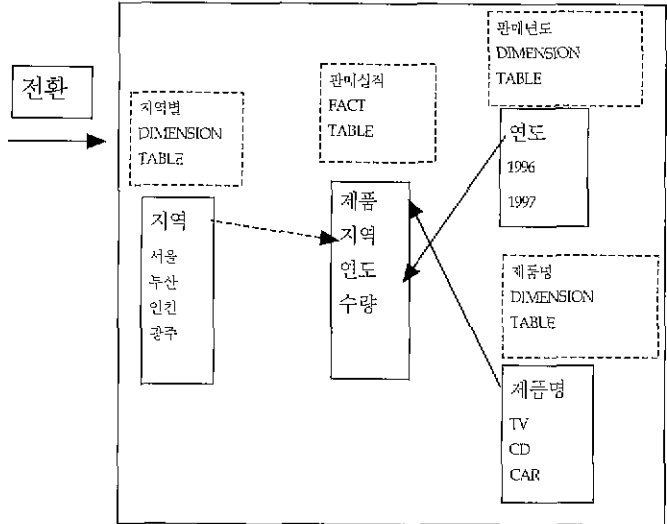
<표 3> 관계형 DBMS와 다차원 DBMS 비교

구 분	관계형 DBMS (Relational DBMS)	다차원 DBMS (Multidimensional DBMS)
원시 데이터 공간	대규모 데이터(20GB 이상)	소규모 데이터(20GB 이하)
인터페이스 상세도	적음(기본량의 2-3배)	많음(기본량의 5-10배)
유연성	표준 인터페이스 제공	표준 인터페이스 부재
공인된 기술자	세부기능 부족	세부기능 충분
제3 공급업체 DBMS 제품	일부 변동시 쉽게 대처	일부 변동시 재구축
	수십만명	수만명
	수천개	수십개
	Oracle, Informix, Sybase	Essbase, Express

관계형 DBMS

제품	지역	연도	수량
TV	서울	1996	200
TV	서울	1997	100
CD	서울	1996	50
CD	부산	1996	100
CD	인천	1997	20
CAR	광주	1996	100

STAR SCHEMA



(그림 3) MOLAP의 구조 및 스타 스키마

스노우 플레이크 스키마(Snowflake Schema) 모델과 같은 데이터 모델이 사용된다.[4] 스타 스키마는 비정규화의 특수한 형태로 1980년대 초반 메타 포어사에 의해 개발되었다. (그림 3)은 ROLAP의 구조 및 스타 스키마의 예를 보여 주고 있다.

데이터 웨어하우스 관리자는 DBMS에 미리 등록된 차원 테이블과 스타 스키마를 이용할 때 필요한 속성 테이블 간의 관계를 맺어 주고, 각 차원에 대한 다단계 요약 순서를 지정하여 준다. 또 사실 테이블은 이렇게 설정된 차원 테이블과 역시 미리 등록된 사실 테이블의 차원 키들을 연결하여 차원 테이블이 사용자에게 다차원 뷰를 갖도록 지정한다. 이러한 OLAP 도구로서는 인포믹스사의 메타 큐브 제품, 마이크로 스트릭사의 DSS 슈트 등이 있다.

3.4.2 MOLAP 도구

기존의 관계형 DBMS와는 달리 MOLAP 도구는 다차원 DBMS에서 OLAP기능을 제공해주는 도구이다. 다차원 DBMS는 다차원 정보가 저장될

때의 물리적인 구조가 다차원 배열 구조이며 각 차원과 차원 항목들은 다차원 배열 내의 값을 구분하기 위한 인덱스로서의 역할을 수행한다. 다차원 DBMS의 가장 큰 특징은 다차원 DBMS 내의 항목들 사이의 다양한 관계가 내장된다는 점이다. 즉 관계형 DBMS가 단순히 데이터만을 저장하고 있는 반면, 다차원 DBMS는 데이터와 이들 데이터간의 상호 관계 즉 업무 구조를 함께 저장하고 있다. 이에 따라 사용자 뷰를 제공하기 위해서 별도의 개발 노력이 필요없이 어플리케이션이 제공된다는 의미에서 “다차원 DBMS = 데이터베이스 + 어플리케이션”으로 풀이 될 수도 있다. 다차원 DBMS는 항목들 사이의 다양한 관계를 쉽게 설정하고 비즈니스 로직을 정의할 수 있도록 다차원 연산과 다차원 질의 검색을 편리하기 위한 다차원 질의 기능을 제공한다. MOLAP도구로서는 IRI/Oracle사의 Express, Arbor Software사의 Essbase, Pilot Software사의 Lightship 등이 있다. OLAP도구들은 관계형 OLAP 및 다차원 OLAP 등으로 분류

되어 개발되어 졌으나 최근에는 다차원 OLAP도구가 관계형 OLAP의 특성을 수용하는 형태의 혼합형(Hybrid) OLAP 도구 개발이 진행중이다. 또한 객체 지향형 데이터베이스를 기반으로 다양한 표현 방법이 시도되고 있다. 실시간 데이터 웨어하우스와 운영계 데이터 웨어하우스를 혼합하고, 인터넷 및 인트라넷과 접목되어 실질적인 OLAP 환경을 구축하기 위한 연구가 활발히 진행중이다.

3.5 데이터 마이닝 도구

데이터 마이닝은 기계학습, 통계학, 인공지능 등 많은 다른 연구 결과로부터 발전한 기술로서 대량 데이터 정보처리 기술, 영상 시각 기술을 추가적으로 응용하여 데이터 웨어하우스와 함께 급속히 발전하는 최신 데이터 분석 기술이다. 데이터 마이닝은 대규모 데이터 중에서 숨겨진 관계와 패턴을 발견하는 것으로 데이터의 가치 있는 지식을 표현한다. 기본적으로 데이터 마이닝은 이전에 알려지지 않은 데이터의 규칙, 패턴을 발견하는 사용자 친화적인 소프트웨어 기술이다. 사용자 친화적인 의미는 기존의 데이터 분석 방법인 전통적인 통계 기법 보다 데이터 분석이 훨씬 쉽다는 뜻이다. 기존의 통계학으로의 데이터 분석은 사전에 가설, 또는 모델을 설정하고 전문적인 데이터 분석가가 반복적인 시뮬레이션을 통해 가설을 검증한 반면 데이터 마이닝은 이러한 모든 작업을 전적으로 대규모 데이터를 처리하는 데이터 마이닝 소프트웨어가 처리한다. 데이터 마이닝 기술은 분류규칙, 연관규칙, 연속패턴, 유사 시계열 등을 중심으로 연구중이다. 분류규칙은 클래스 레벨이 이미 알려진 데이터인 트레이닝 데이터의 집합을 분석하고 이러한 데이터의 성질을 기초로 하여 데이터베이스를 클래스로 분류한 뒤 이 클래스에 영향을 줄 규칙의 집합을 공식화한다. 연관규칙은 시험된 레코드중에서 반영된 데이터들

사이의 유사성을 인식하는 것이다. 이러한 유사성은 규칙으로 표현된다. 이러한 발견 등에 대한 퍼센트는 연관성의 신뢰요인으로 사용된다. 대표적 응용으로 시장 바구니 분석은 연관규칙을 제품의 유사성을 인식하기 위한 POS(Point Of Sales) 트랜잭션 데이터에 적용한 것이다. 연속패턴은 연관규칙의 변형으로서 시간적인 관계를 가지고 있다. 예를 들어 배낭을 구입한 고객은 다음에 텐트를 구입하는 경향이 있는 경우, 배낭의 구입과 텐트의 구입에는 연속성이 존재한다는 의미를 부여한다. 유사 시계열의 탐사는 주식, 물가, 판매량, 과학적 실험 데이터 등이 시간축으로 관측값 등이 나열된 시계열의 형태를 갖게되는 경우 서로 유사한 시계열 데이터들을 발견하는 것이다. 예를 들면 특정 물품의 판매량 추이와 비슷한 판매량 추이를 갖는 다른 물품을 발견할 수 있다면 두 물품의 판매 전략에 유사성을 부여할 수 있다. 데이터 마이닝 기술은 관계형 데이터베이스에서 특정 레코드 또는 레코드와 유사한 레코드 등을 추출 해내는 케이스 베이스 리즈닝(case based reasoning) 분야, 상이한 예측에 대한 정보를 그래픽 디스플레이를 이용하여 쉽고 빠르게 추출해 내는 데이터 시각화(data visualization) 분야, 현 상태와 가장 근접한 결과를 추출하기 위한 퍼지 어널리시스(fuzzy analysis) 분야, 변수들 사이에 존재하는 상관관계를 인식하기 위해 디자인된 지식 발견(knowledge discovery) 분야 및 데이터들 간의 상관관계를 발견하고 예측하기 위해 훈련 되어 질수 있는 뉴럴 네트워크 분야 등을 중심으로 활발하게 연구중에 있다. 지금까지의 외국의 대학 및 연구소에서의 데이터 마이닝 관련 제품 및 연구 동향은 다음과 같다.

- Simon Fraser 대학의 DBMiner 데이터 마이닝 제품
대용량 관계형 데이터베이스 시스템에서 다제

총 지식을 추출하기 위한 유닉스 및 윈도우 환경의 GUI 사용자 인터페이스를 갖춘 데이터 마이닝 시스템으로 특정 작업과 관련된 데이터들을 수집하여 서로 다른 관점에서 다수의 개념 단계를 생성하는 특성 도출기, 목표 클래스와 계층 클래스 등의 차이점을 분석 요약하기 위한 차별 규칙 등을 발견하는 차별 규칙 생성기, 다수의 개념 단계에서 데이터베이스와 관련된 데이터들과의 연관 규칙을 발견하는 연관규칙 생성기, 트레이닝 데이터를 분석하고 데이터의 특성에 따라 모델을 구축하는 클래스 생성기, 수집되지 못한 데이터의 가치나 특정 객체의 특성에 따른 가치를 분배하기 위한 예측을 하는 예측기, 사용자 요구에 따른 메타 규칙을 생성하기 위한 메타 규칙 안내자, 데이터베이스내의 작업과 관련된 편차 패턴을 평가하기 위한 편차 패턴 평가기 등의 기능을 가지고 있다.

- IBM의 DB2 Parallel Edition 데이터 마이닝 제품
 데이터를 정보로 변환하여 의사결정을 위한 처리과정으로 데이터의 타입을 결정하는 선택 과정, 선택된 데이터를 한가지 타입 또는 새로 도출될 특성으로 변환하는 변환과정, 변환된 데이터를 가지고 추출하는 추출과정 및 추출된 데이터를 가지고 기업 경영의 목적을 분석하고 의사결정자에게 최선의 정보를 제공하는 해석 과정 등이 사용된다. 또한 데이터 마이닝을 지원하는 기능은 추론 기법, 연관 규칙 또는 연속 발견 기술에 의한 분석, 시각화, 클러스터링 기술에 의한 데이터베이스 세그멘테이션, 통계 기술을 이용한 편차 발견 등이 지원되고 있다.

4. 활용분야

데이터 웨어하우스에 기반한 데이터 마이닝 기술은 특정 분야에만 적용될 수 있는 단말 기술이

아니고 취급하는 데이터의 분량이 수백만 바이트 이상 대규모인 분야, 대량으로 축적된 업무 데이터의 특성이 업무 현장 자체의 특성을 대변하는 분야, 업무 현장의 행태가 복잡하여 단순한 산술 모형으로 업무를 모델링 하기 어려운 분야 등 다양한 분야에서 활용될 수 있는 핵심 기반 기술이다. 이러한 특성을 가지고 있는 분야를 예시하면 <표 4>와 같다.

5. 결론 및 향후 과제

데이터 웨어하우스 관련 기술의 표준화 동향은 비록 데이터 웨어하우스, OLAP 협의체 등의 제품 업체가 협의체를 구성하여 표준화를 지향하고 있지만 업체별 대립 관계가 잠재하고 있어 단기일 내에 표준화를 달성하기가 어려울 것이라고 예상하고 있다. 1995년 10월 가트너 발표에 의하면 향후 5년 동안 DBMS 관련 미들웨어 제품들은 산업 표준을 따르지 않을 것이고 SQL3는 적어도 2000년 까지는 표준으로 자리 잡지 못할 것으로 전망하고 있다. 향후 데이터 웨어하우스 분야에서 향후 발전할 기술 분야는 다음과 같다.

<표 4> 데이터 웨어하우스 및 데이터 마이닝 활용 분야

분 야	용 도
통 신	특정 고객의 차별적 선정 특수 상품의 수익성 분석 및 개발 유무선 통신량 분석 및 요금 정책
금 융	특수 상품의 수익성 분석 및 개발 위험 및 신용평가(기업 부도 예측 분석) 고객 정보관리
유 통	매출분석, 광고 효과분석 수익성 분석, 시장분석 고객의 구매 패턴 분석
연구소	실험 결과 예측

- 대용량DB(수십 테라바이트) 처리가 가능한 데이터 병렬 처리 기술
- 관계형 DBMS와 다차원 DBMS 간의 동적인 연동 기술
- 다양한 GUI(애니메이션, 그래픽) 기술
- 인터넷/인트라넷 간의 연동 기술
- 멀티미디어 데이터와의 연동 기술
- 더욱 진보된 데이터 마이닝 알고리즘 개발
- 유닉스 오픈 시스템을 기반으로 하는 MPP 활용 기술

데이터 웨어하우스는 새로운 분야인 만큼 신기술이 어느 분야에서 얼마 만큼의 영향을 미칠지에 대해서는 현 상태에서 예측하기 힘들다. 데이터 웨어하우스의 본질에 대해서 관계형 데이터베이스의 연장선상에서 해석하는 측과 다차원 데이터베이스의 연장선상에서 해석하는 측이 대립 상태를 보이고 있다. 그러나 대부분의 전문가들은 다차원 데이터베이스와 관계형 데이터베이스가 궁극적으로 서로의 장점을 극대화하기 위해 통합화를 지향할 것으로 전망하고 있다. 또한 데이터 웨어하우스는 단순히 기업 내 데이터의 효율적 분석 차원을 넘어 인터넷 등 외부 데이터 소스와 연계되는 한편 분석 정보의 질적 수준의 향상에 기술적 관심이 집중될 것이다. 특히 표현 방법도 단순한 정지영상을 넘어서 동영상 기능 등 멀티미디어화된 다양한 방법이 도입될 것이다. 따라서

데이터 웨어하우스가 각종 분산된 기업용 정보 기술과 결합되어 기업 전략 정보를 제공하는 핵심 인프라가 될 것으로 전망되므로 국내에서도 이와 관련된 연구 개발이 필요로 하고 있다.

참고문헌

- [1] Rob Mattison, 'Data Warehouse', McGraw-Hill, 1996
- [2] Vdette Poe, 'Building Data Warehouse for Decision Support', Prentice Hall PTR, 1995
- [3] W. H. Inmon, 'Buildinging the Data Warehouse', A Wiley-Qed publication, 1992
- [4] W. H. Inmon & R. D. Hackathorn, 'Using the Data Warehouse', A Wiley-Qed publication, 1994

김 영 남



1982년 한양대학교 공과대학 (공학사)
 1994년 한국 외국어대학교 경영대학원(경영학 석사)
 1982년-현재 시스템공학연구소 선임 연구원

1991년 정보처리기술사
 1995년 공인 전산 감사사(CISA)
 관심분야 : DBMS, 정보시스템 분석 및 설계, 전산 감리