

7강. 데이터웨어하우스의 구축 및 응용

1. 데이터웨어하우스 구축
2. 데이터웨어하우스 응용

1. 데이터웨어하우스 구축

◆ 데이터웨어하우스 구축 수립 단계

→ 1단계 공급(Feed)

- 운영 시스템 및 외부 소스의 적절한 데이터 세트를 규명, 추출, 정화, 수정하고 데이터웨어하우스로 옮기는 모든 작업

→ 2단계 저장(Store)

- 데이터 서버에 데이터를 저장하여 데이터에 대한 신속하고 효율적인 액세스를 제공하며 데이터웨어하우스에 대한 관리를 지속적으로 해나가는 단계

→ 3단계 사용(Use)

- 데이터를 실제 사용하는 것으로 데이터에 대한 질의, 분석, 보고서, 차트, 그래프의 생성과 출판을 포함
- 일반적으로 최종 사용자는 그래픽 프론트-엔드 툴을 사용하여 데이터베이스에 대한 질의를 수행하고 그 결과를 조회

1. 데이터웨어하우스 구축

◆ 데이터웨어하우스 구축 단계

구분	의미	내용
단계 1	실행계획 수립	<ul style="list-style-type: none"> 조직의 전략계획, 이행계획, 기술적 검토, 데이터웨어하우스 구성요소에 대한 평가 대한 보고서 작성
단계 2	데이터모델 검토	<ul style="list-style-type: none"> 데이터베이스 설계와 프로젝트관리 과정을 포함 기존 논리모델에 대한 검토 작업과 현업의 개괄적인 요건 조사를 통해 프로젝트 범위 확정
단계 3	데이터성격 규정	<ul style="list-style-type: none"> 데이터웨어하우스 내에서 데이터를 입력, 보관, 관리하며 최종 사용자가 가장 효율적으로 활용할 수 있도록 데이터의 성격을 규정하는 작업 유지 관리시 중요한 정보로 활용
단계 4	데이터 변환 프로그램 작성	<ul style="list-style-type: none"> 운영환경에 있는 데이터를 데이터웨어하우스 데이터베이스로 변환하기 위해 필요한 정보를 정리한 후, 데이터변환 도구에 의해 작성하고 그 결과를 확인
단계 5	소스데이터 추출	<ul style="list-style-type: none"> 생성된 데이터 변환 프로그램을 이용하여 실제로 데이터웨어하우스를 구축하는 반복적 작업
단계 6	메타데이터 설정	<ul style="list-style-type: none"> 특정 기능별 부분에 대한 자료의 집중과 질의를 위해 이미 구축된 데이터웨어하우스로부터 데이터 복제 기능을 이용하여 필요한 부분과 기능별 데이터마트를 별도 구성
단계 7	데이터 액세스 도구 지원	<ul style="list-style-type: none"> 데이터웨어하우스를 구축한 후에 최종 사용자가 데이터를 가장 완벽하게 활용할 수 있도록 하는 과정 즉, 데이터마이닝의 과정
단계 8	운영환경 점검, 테스트	<ul style="list-style-type: none"> 구축 후 지속적 모니터링을 통하여 문제점을 파악하여 시스템을 개선 사용자에게 데이터웨어하우스 프로토타입 제시

1. 데이터웨어하우스 구축

◆ 데이터웨어하우스 구축시 고려사항

→ 정보시스템 팀과 현업의 조화와 협조로 시작

- 정보시스템 팀은 현업의 요구사항을 충분히 반영해야 하며 현업도 요구사항을 현실적으로 조정할 수 있도록 IT의 능력에 대한 이해가 있어야 함

→ 최고 경영층의 의지

- 데이터웨어하우스의 효과에 대한 강한 믿음 과 중요성을 인식

→ 성급하게 진행해서는 안됨

- 데이터웨어하우스를 2~3년 내에 끝내겠다는 생각보다는 시간적인 여유를 가지고 구축에 대한 충분한 검토와 과정을 거쳐 구축

→ 처음부터 과도한 투자를 해서는 안됨

- 데이터웨어하우스가 병원에 어느 정도 기여할 것인지 정확히 알 수 없는 상황에서 과도한 투자는 병원 경영의 경제적 상황을 악화시킬 수 있음

2. 데이터웨어하우스의 응용

◆ 데이터웨어하우스에 적재된 데이터의 활용방법

→ 통계분석 패키지

▪ SPSS, SAS, R

→ 데이터마트

→ 온라인분석시스템 (On-Line Analytical Processing)

→ 데이터마이닝(Data Mining) 도구

1) 온라인분석시스템

◆ 온라인분석시스템(OLAP : On Line Analytical Processing)

→ 데이터웨어하우스에 체계적으로 쌓여 있는 데이터 속에 담겨 있는 정보를 효율적으로 끌어내어 분석하는데 사용

◆ OLTP와의 차이점

→ OLTP시스템 :

- 은행의 창구 업무나 항공사의 예약 업무, 접수/진료/수납 등
- 원시데이터가 실제로 발생하고 기록되는 '**무엇(What)**'에 초점
- 현재 거래 상태를 정확하게 기록하고 갱신할 수 있도록 하는 것이 목표
- 기업의 운영을 지원

→ OLAP시스템 :

- 사용자가 기업의 전반적인 상황을 이해할 수 있게 하고, 의사결정을 지원
- 수집된 데이터를 의사결정에 활용하는 측면을 담당하여 '**왜(Why)**'에 초점
- 기업의 방향을 설정

1) 온라인분석시스템

◆ OLTP와 OLAP 비교

표 10-8 OLTP와 OLAP 비교

OLTP	OLAP
<ul style="list-style-type: none"> 워크플로우 기반 (<u>업무 프로세스 중심</u>) 	<ul style="list-style-type: none"> 사용자의 분석 수행 기반 (<u>주제 중심</u>)
<ul style="list-style-type: none"> <u>트랜잭션 처리</u> (입력, 조회, 삭제, 수정) 운영자 계층 시스템 	<ul style="list-style-type: none"> <u>보고서, 분석, 계획</u> (조회, 제한적 입력/수정) 분석가 및 의사결정자 계층 시스템
<ul style="list-style-type: none"> 2차원, 정규화 	<ul style="list-style-type: none"> <u>다차원</u>, 계층구조
<ul style="list-style-type: none"> <u>상세 데이터, 중복성 배제</u> 	<ul style="list-style-type: none"> <u>요약정보, 중복성 수용</u>
<ul style="list-style-type: none"> 소량의 데이터 처리 활용 패턴 단순, 고른 시간대 분포 시스템 자원 사용량 예측이 용이 	<ul style="list-style-type: none"> 대량의 데이터 처리 활용 패턴 다양, 시간대 불규칙 분포 <u>시스템 자원 사용량 예측 어려움</u>
<ul style="list-style-type: none"> 구축 후 데이터 축적 중심 전통적 개발 주기 시스템 구축 후 유지보수 단순 	<ul style="list-style-type: none"> 구축 후 데이터 축적 및 <u>스키마 변경</u> 반복 확장 개발 주기 시스템 구축 후 유지보수 전략 필요
<ul style="list-style-type: none"> 사용자 중심 응용프로그램 (4GL) 사용자 맞춤형(Customizing) 용이 정형화된 보고서/변경이 어려움 단순한 화면 조작 	<ul style="list-style-type: none"> <u>전용 도구 (Off-the-Shelf, Out-of-Box)</u> 사용자 맞춤형(Customizing) 제한적 <u>동적인 비정형 분석/변경이 용이</u> EUC(End User Computing) 활성화 필요

1) 온라인분석시스템

◆ 온라인분석시스템의 특성

- 최종사용자(End-user)가 다차원 정보에 직접 접근하여 대화식으로 정보를 분석
- 최종사용자가 기업의 전반적인 상황을 이해하고 이를 통해 의사결정에 이용

표 10-9 온라인분석시스템의 특성

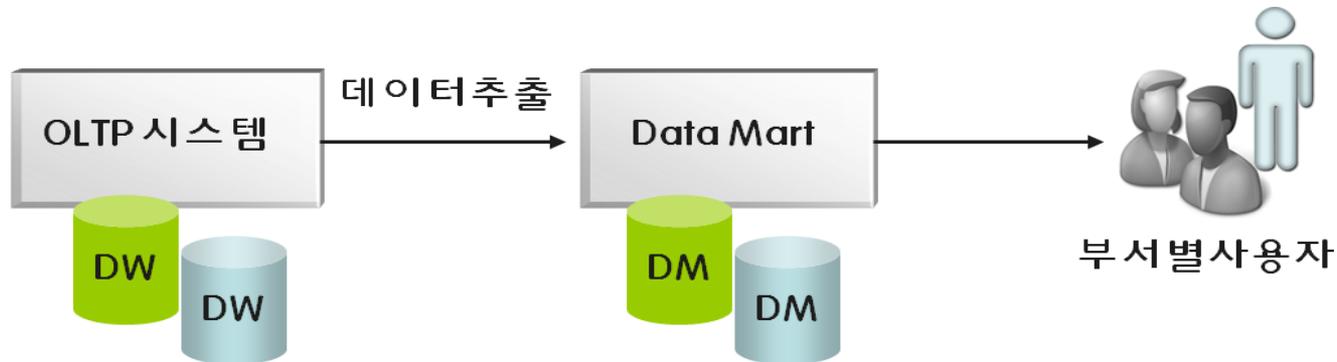
특성	의미
다차원 정보	사용자들이 실질적인 차원에서 정보를 분석하도록 한다.
직접접근	사용자가 전산 부서와 같은 정보 매개자를 거치지 않고 자신이 원하는 정보에 직접 접근한다.
대화식분석	사용자는 시스템과의 상호작용을 통해 정보를 분석하며 원하는 결과를 얻을 때까지 계속해서 분석을 수행한다.
의사결정에 이용	사용자가 기업의 전반적인 상황을 이해할 수 있게 하고 의사결정을 지원한다.

2) 데이터마트

◆ 데이터마트(Data Mart)

- 데이터웨어하우스(DW)와 데이터마트의 역할은 흔히 유통업에 비유
- 운영시스템 : 데이터라는 상품을 생산하는 곳
 - DW : 데이터를 소비자들에게 판매하기 위해 체계적으로 분류해서 저장하고 분배하는 기능을 수행하는 도매상
 - 데이터마트 : 도매상과 소비자 사이에 위치하여 이용자가 일상적으로 필요한 대부분의 물품들을 쉽고 빠르고 간편하게 구매할 수 있는 것

그림 10-18 데이터마트의 개념



2) 데이터마트

◆ 데이터마트(Data Mart)

- DW : 데이터의 중앙 집합체, 전략적이지만 다소 덜 다듬어진 개념
- 데이터마트 : 데이터의 저장소로서 특별히 계획된 목적을 위해 접근의 용이성과 유용성을 강조한 것, 전술적이며 당장의 요구에 부합하는 데에 목표

◆ 데이터마트(Data Mart)의 정의

- 데이터웨어하우스로부터 구축된 데이터 속에서 한 가지 주제 또는 한 부서 중심으로 구축된 소규모, 단일 주제의 웨어하우스
- 예측 가능한 질의에 대해서 매우 빠르게 응답할 수 있도록 데이터를 제공하는 시스템
- 다양한 정보를 사용자의 요구항목에 따라서 체계적으로 분석, 기업의 경영활동을 돕기 위한 시스템
 - 월별 병상이용률, 월별 진료과별 진료실적, 주간 의약품 재고현황
 - 환자의 진료기록, 수입이나 환자의 통계분석, 검사 및 치료결과 분석
 - 약품, 진료,치료재료 분석, 인건비, 회계, 직원 인사정보 분석, 병원 경영지원

3) 데이터마이닝

◆ 데이터마이닝(Data Mining)

- 데이터웨어하우스에 존재하는 대량의 데이터 사이에 묻혀 있는 전략적인 정보를 발견하고 유용한 패턴을 찾아내는 규칙을 추론함으로써 의사결정을 제 공하고 그 효과를 예측
- 주로 통계학, 데이터 분석 및 경영정보학 분야에서 사용되어 왔으며, 컴퓨터 공학, 의료정보학 등의 학문분야를 중심으로 큰 관심을 갖고 있음

표 10-11 데이터마이닝 작업에 이용되는 기법

작업(Task)	기법(Technique)
분류(Classification)	의사결정 트리(Decision Tree) 신경망(Neural Network) 유전자 알고리즘(Genetic Algorithm) 베이지안 알고리즘(Bayesian Algorithm)
측정(Estimation)	신경망(Neural Network)
예측(Prediction)	신경망(Neural Network) 연관규칙(Association Rule) 의사결정 트리(Decision Tree) 순차패턴(Sequential Pattern)
연관(Affinity Grouping)	연관규칙(Association Rule)
군집(Clustering)	클러스터링(Clustering)
묘사(Description)	의사결정 트리(Decision Tree) 연관규칙(Association Rule) 순차패턴(Sequential Pattern)

7강. 데이터웨어하우스의 구축 및 응용 - 요약

◆ 1. 데이터웨어하우스 구축

→ 데이터웨어하우스 구축 수립 단계

- 공급(feed) → 저장(Store) → 사용(Use)

→ 데이터웨어하우스 구축단계

- 실행계획 수립 → 데이터모델 검토 → 데이터성격 규정 → 데이터변환 프로그램
- 소스데이터 추출 → 메타데이터 설정 → 데이터 액세스 도구 지원 → 운영환경 점검,테스트

◆ 2. 데이터웨어하우스 응용

→ 1. 온라인분석시스템(OLAP)

- OLTP시스템 - 기업의 운영을 지원
 - 원시데이터가 실제로 발생하고 기록되는 '**무엇(What)**'에 초점
 - 현재 거래 상태를 정확하게 기록하고 갱신할 수 있도록 하는 것이 목표
- OLAP시스템 - 기업의 방향을 설정
 - 사용자가 기업의 전반적인 상황을 이해할 수 있게 하고, 의사결정을 지원
 - 수집된 데이터를 의사결정에 활용하는 측면을 담당하여 '**왜(Why)**'에 초점
 - 최종사용자(End-user)가 다차원 정보에 직접 접근하여 대화식으로 정보를 분석하여 이를 통해 의사결정에 이용

→ 2. 데이터 마트 (Data Mart)

- 한 가지 주제 또는 한 부서 중심으로 구축된 소규모, 단일 주제의 웨어하우스
- 예측 가능한 질의에 대해서 매우 빠르게 응답할 수 있도록 데이터를 제공

→ 3. 데이터 마이닝 (Data Mining)

- 대량의 데이터 사이에 묻혀 있는 전략적인 정보를 발견하고 유용한 패턴을 찾아내는 규칙을 추론함으로써 의사결정을 제공