
Chapter 5

비즈니스 인텔리전스의 기초 : 데이터베이스와 정보관리



Essentials of Management Information Systems

Chapter. 5 비즈니스 인텔리전스의 기초 : 데이터베이스와 정보관리

학습목표

- 관계형 데이터베이스가 데이터를 어떻게 구성하고, 객체지향 데이터베이스와 어떠한 차이가 존재하는가?
- 데이터베이스 관리 시스템의 원리는 무엇인가?
- 기업의 성과와 의사결정력을 향상시키기 위한 데이터베이스의 정보에 접근하기 위한 주요 도구와 기술들은 무엇인가?
- 구조적 데이터 자원의 관리에 있어서 정보정책과 데이터 관리의 역할은 무엇인가?
- 데이터의 품질에 대한 보증이 기업에서 왜 중요한가?

R.R. Donnelley의 데이터관리

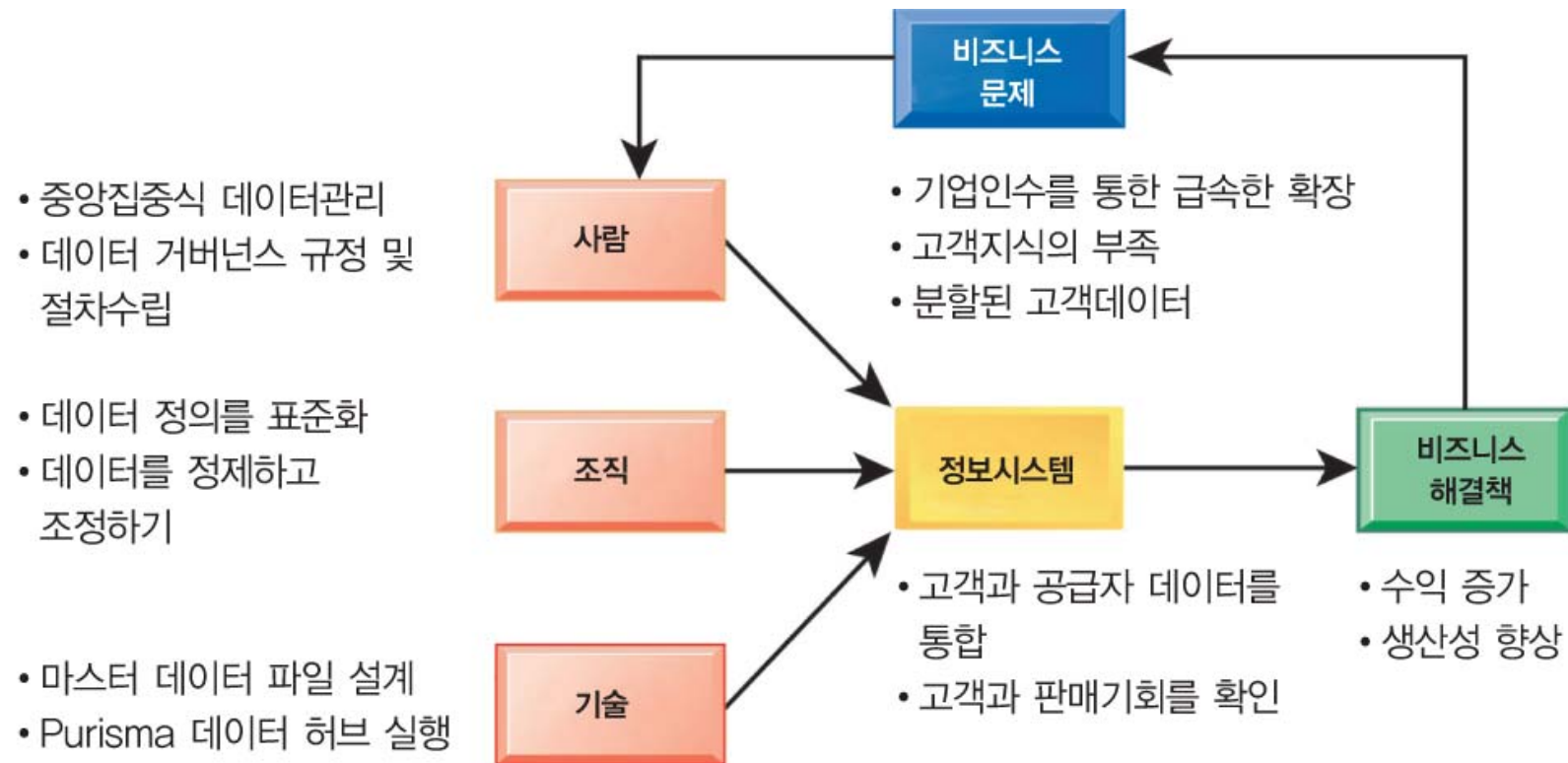
- **Problem:** 정보관리와 관련하여 직면한 도전의 폭발적 증가
- **Solutions:** 불필요한 데이터 중복을 방지하면서 전사적 형태의 데이터 집합체를 만들기 위한 MDM의 활용



R.R. Donnelley의 데이터관리

- **Master data management (MDM)** :R.R. Donnelley와 같은 기업들이 오래되고, 불완전하며, 부정확한 형태의 데이터를 제거하도록 하기 위하여 마스터데이터베이스(MDM)을 사용한다.
- 성공적인 데이터관리에 있어서 정보기술의 역할을 증명하는 사례
- 데이터를 저장하고 구성하기 위한 디지털 기술의 역할을 보여주는 사례

R.R. Donnelley의 데이터관리



데이터 관리를 위한 데이터베이스 접근

- **데이터베이스(Database):**
 - 사람, 장소, 사물에 대한 기록을 담고있는 관계된 파일들의 집합
 - 데이터베이스가 사용되기 이전에 기업들은 종이서류를 보관하기 위해 캐비닛을 사용
- **개체(Entity):**
 - 정보를 저장하고 유지하는 사람, 장소, 사물 등을 나타내는 일반화된 유형(Generalized category)
 - 예) 공급자, 부품
- **속성(Attributes):**
 - 각각의 개체에 대한 특성:
 - 공급자 이름, 주소
 - 부품 설명, 단위가격, 공급자

데이터 관리를 위한 데이터베이스 접근

- **관계형 데이터 베이스(Relational database):**
 - 행과 열로 구성된 이차원의 테이블 형태로 데이터를 구성하는 것
 - 각각의 개체에 대한 하나의 테이블:
 - 예: CUSTOMER, SUPPLIER, PART, SALES
 - 필드(Fields :columns) 속성을 나타내는 데이터의 저장소
 - 열(Rows) 분리된 레코드에 대한 자료 저장소
 - Key field: 각각의 레코드를 식별하기 위한 필드
 - 일차키(Primary key):
 - 각각의 테이블에 하나의 필드
 - 중복이 불가능
 - 특정 열에 있는 모든 정보에 대해 식별자 역할을 함

데이터 관리를 위한 데이터베이스 접근

A Relational Database Table

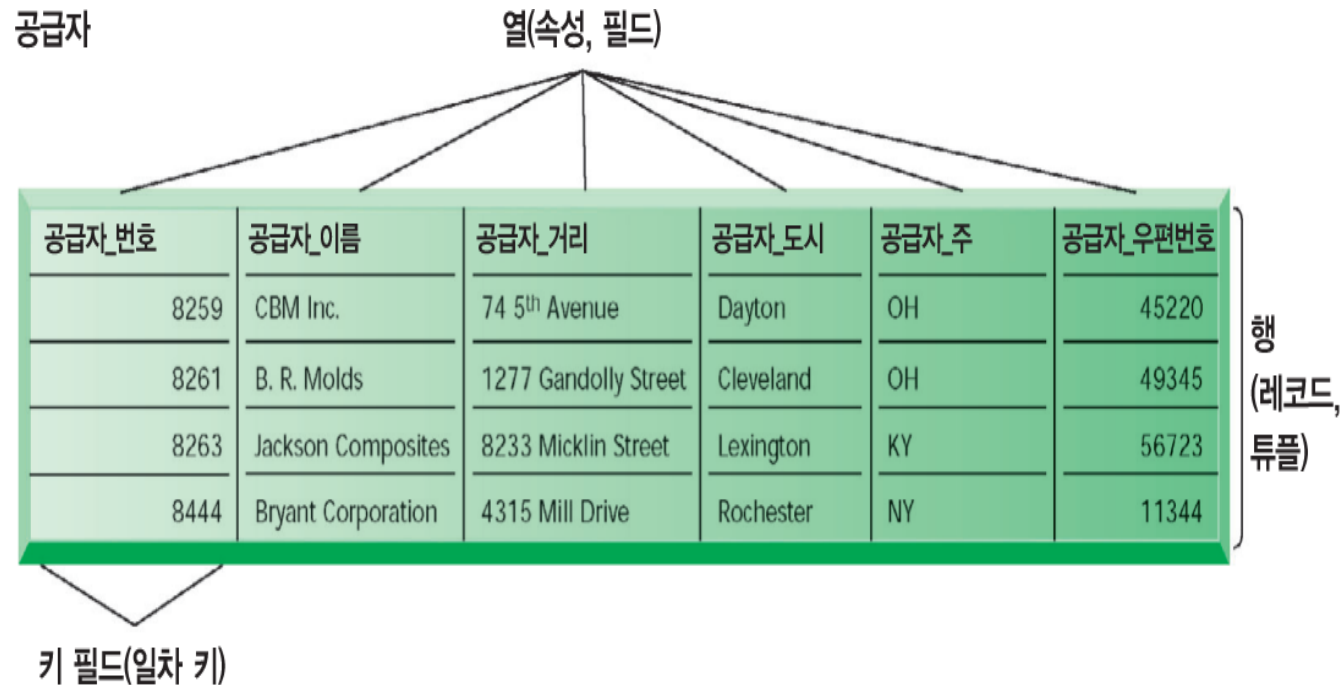


그림 5-1 관계형 데이터베이스 테이블

관계형 데이터베이스는 데이터를 2차원 형태의 테이블로 조직한다. 이 그림에서는 공급자 개체에 대한 테이블을 보여줌으로써 개체와 속성을 표현하는 방법에 대해 설명하고 있다. 여기서 공급자 번호는 키 필드에 해당한다.

데이터 관리를 위한 데이터베이스 접근

부품 테이블

그림 5-2 부품 테이블

부품 개체에 대한 데이터는 다른 테이블로 분류된다. 부품 번호는 일차 키, 공급자 번호는 외래 키로서 사용자가 공급자 테이블로부터 각 부품을 공급하는 공급자와 관련된 정보를 검색할 수 있도록 한다.

부품

부품_번호	부품_이름	단가	공급자_번호
137	Door latch	22.00	8259
145	Side mirror	12.00	8444
150	Door molding	6.00	8263
152	Door lock	31.00	8259
155	Compressor	54.00	8261
178	Door handle	10.00	8259

일차 키

외래 키

데이터 관리를 위한 데이터베이스 접근

- 관계 정의 (Establishing relationships)
 - 개체관계도 : Entity-relationship diagram
 - 관계형 데이터베이스 안의 테이블의 관계를 명확하게 하기위해 사용
 - 관계형 데이터베이스 테이블:
 - 1 대 1 관계
 - 1 대 다 관계
 - 다 대 다 관계
 - 정보를 조합하기 위하여 두 테이블을 연결하는 추가의 테이블을 만들어야 함.

데이터 관리를 위한 데이터베이스 접근

단순 개체 관계도와 주문서(샘플)



주문번호: 3502
주문날짜: 1/15/2010

공급자 번호: 8259
공급자 이름: CBM Inc.
공급자 주소: 74 5th Avenue, Dayton, OH 45220

주문_번호	부품_번호	부품_수량	부품_이름	단가	부품금액
3502	137	10	Door latch	22.00	\$220.00
3502	152	20	Door lock	31.00	620.00
3502	178	5	Door handle	10.00	50.00

총 주문금액: \$890.00

그림 5-3 단순 개체 관계도

이 그림은 공급자와 부품 간의 관계를 보여준다.

그림 5-4 단순 주문서

음영부분은 공급자, 주문, 그리고 주문부품 테이블로부터 가져온 것이다. 데이터베이스에는 부품 금액이 포함되어 있지 않는데 이는 부품 금액이 다른 데이터 테이블로부터 파생될 수 있기 때문이다.

데이터 관리를 위한 데이터베이스 접근

- 정규화 (Normalization)
 - 복잡한 데이터의 집합을 간소화:
 - 중복 데이터 요소를 제거
 - 복잡한 다대다 관계를 최소화
 - 데이터의 안정성과 유연성을 증대
- 참조무결성 규칙(Referential integrity rules)
 - 관련된 테이블간의 관계가 일관성을 유지하도록하기 위해 관계형 데이터베이스를 사용
 - 예) 하나의 테이블이 다른 테이블의 외부 키를 가지고 있을 때, 연결된 테이블에 상응하는 레코드가 없을 때 외부 키를 가진 테이블의 레코드를 추가하지 않는 경우

데이터 관리를 위한 데이터베이스 접근

샘플 레코드에 대한 최종 데이터 설계

부품

부품 번호	부품 이름	단가	공급자 번호
137	Door latch	22.00	8259
145	Side mirror	12.00	8444
150	Door molding	6.00	8263
152	Door lock	31.00	8259
155	Compressor	54.00	8261
178	Door handle	10.00	8259

주문 부품

주문 번호	부품 번호	부품 수량
3502	137	10
3502	152	20
3502	178	5

주문

주문 번호	주문 날짜
3502	1/15/2010
3502	1/15/2010
3502	1/15/2010

공급자

공급자 번호	공급자 이름	공급자 거리	공급자 도시	공급자 주	공급자_우편번호
8259	CBM Inc.	74 5 th Avenue	Dayton	OH	45220
8261	B. R. Molds	1277 Gandolly Street	Cleveland	OH	49345
8263	Jackson Components	8233 Micklin Street	Lexington	KY	56723
8444	Bryant Corporation	4315 Mill Drive	Rochester	NY	11344

그림 5-5 최종 데이터베이스 설계 사례

공급자, 부품, 그리고 주문에 대하여 최종적으로 설계된 데이터베이스는 4개의 테이블을 갖는다. 주문 부품 테이블은 결합 테이블로서 주문과 부품 사이에 존재하는 다대다 관계를 제거하기 위한 테이블이다.

데이터 관리를 위한 데이터베이스 접근

4개의 테이블을 가진 데이터베이스의 개체관계도

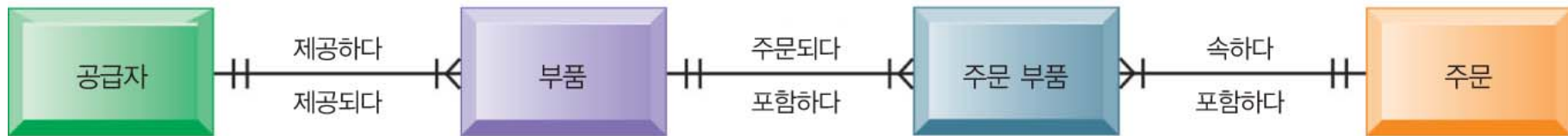


그림 5-6 4개 테이블을 포함하는 데이터베이스에 대한 개체관계도

이 그림은 공급자, 부품, 주문 부품, 그리고 주문 개체들 간의 관계를 보여주고 있다.

데이터베이스 관리 시스템(Database Management Systems)

데이터베이스 관리시스템(DBMS)

- 데이터베이스로 부터 데이터를 만들고, 저장, 구성하여, 접근 가능하게 하는 특성 유형의 소프트웨어
- 데이터의 논리적/물리적 관전을 분리
- 논리적 관점: 최종사용자가 데이터를 어떻게 보는가?
- 물리적 관점: 데이터가 어떻게 정확하게 구조화되고, 조직화되었는가?
- 예) DBMS: Microsoft Access, DB2, Oracle Database, Microsoft SQL Server, MySQL

데이터베이스 관리 시스템(Database Management Systems)

다양한 관점의 인적자원 데이터베이스

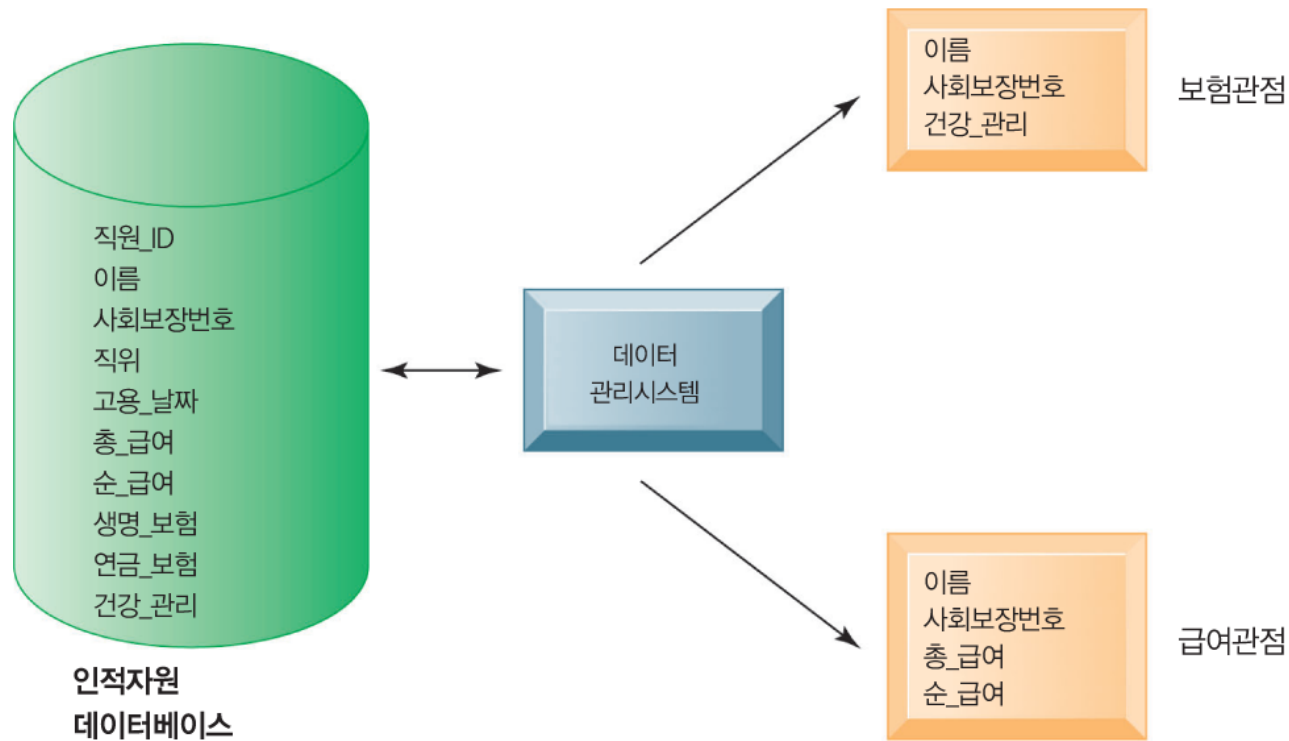


그림 5-7 다양한 관점의 인적자원 데이터베이스

하나의 인적자원 데이터베이스를 통해 사용자가 요구하는 정보에 따라 다양한 관점에서 데이터를 살펴볼 수 있다. 그림에서는 두 가지 관점, 즉 보험 전문가 관점과 기업의 급여를 담당하고 있는 부서 직원의 관점에 대하여 설명하고 있다.

데이터베이스 관리 시스템(Database Management Systems)

관계형 DBMS의 운용

- **선택(Select):**
 - 기술된 영역에 맞는 모든 레코드의 집합을 만들어냄.
- **병합(Join):**
 - 개별 테이블로부터 가용한 것 이상의 정보를 보여주는 관계형 테이블의 결합
- **추출(Project):**
 - 하나의 테이블에 특정 열을 포함하고 있는 부분 집합체를 생성
 - 사용자가 필요로하는 정보만을 담고 있는 테이블을 생성하도록 함

데이터베이스 관리 시스템(Database Management Systems)

관계형 DBMS의 세가지 기본 연산

부품

부품_번호	부품_이름	단가	공급자_번호
137	Door latch	22.00	8259
145	Side mirror	12.00	8444
150	Door molding	6.00	8263
152	Door lock	31.00	8259
155	Compressor	54.00	8261
178	Door handle	10.00	8259

부품번호 = 137 또는 150 Select

공급자

공급자_번호	공급자_이름	공급자_거리	공급자_도시	공급자_주	공급자_우편번호
8259	CBM Inc.	74 5 th Avenue	Dayton	OH	45220
8261	B. R. Molds	1277 Gandolly Street	Cleveland	OH	49345
8263	Jackson Components	8233 Micklin Street	Lexington	KY	56723
8444	Bryant Corporation	4315 Mill Drive	Rochester	NY	11344

공급자 번호에 의한 Join

부품_번호	부품_이름	공급자_번호	공급자_이름
137	Door latch	8259	CBM Inc.
150	Door molding	8263	Jackson Components

선택된 열 Project

그림 5-8 관계형 데이터베이스에서의 세 가지 기본 연산

Select, Project 그리고 Join 연산은 2개의 서로 다른 테이블의 데이터를 조합하여 선택한 속성만을 표시할 수 있도록 한다.

데이터베이스 관리 시스템(Database Management Systems)

DBMS의 기능

- 데이터 정의:
 - 데이터베이스 내용의 구조를 정의
- 데이터 사전(Data dictionary):
 - 데이터 요소와 특성에 대한 정의를 저장하는 자동화 또는 수동적 파일
- 질의(Querying)와 보고(reporting):
 - 데이터 조작어(Data manipulation language)
 - Structured query language (SQL)
 - Microsoft Access query-building tools
 - 리포트 생성

데이터베이스 관리 시스템(Database Management Systems)

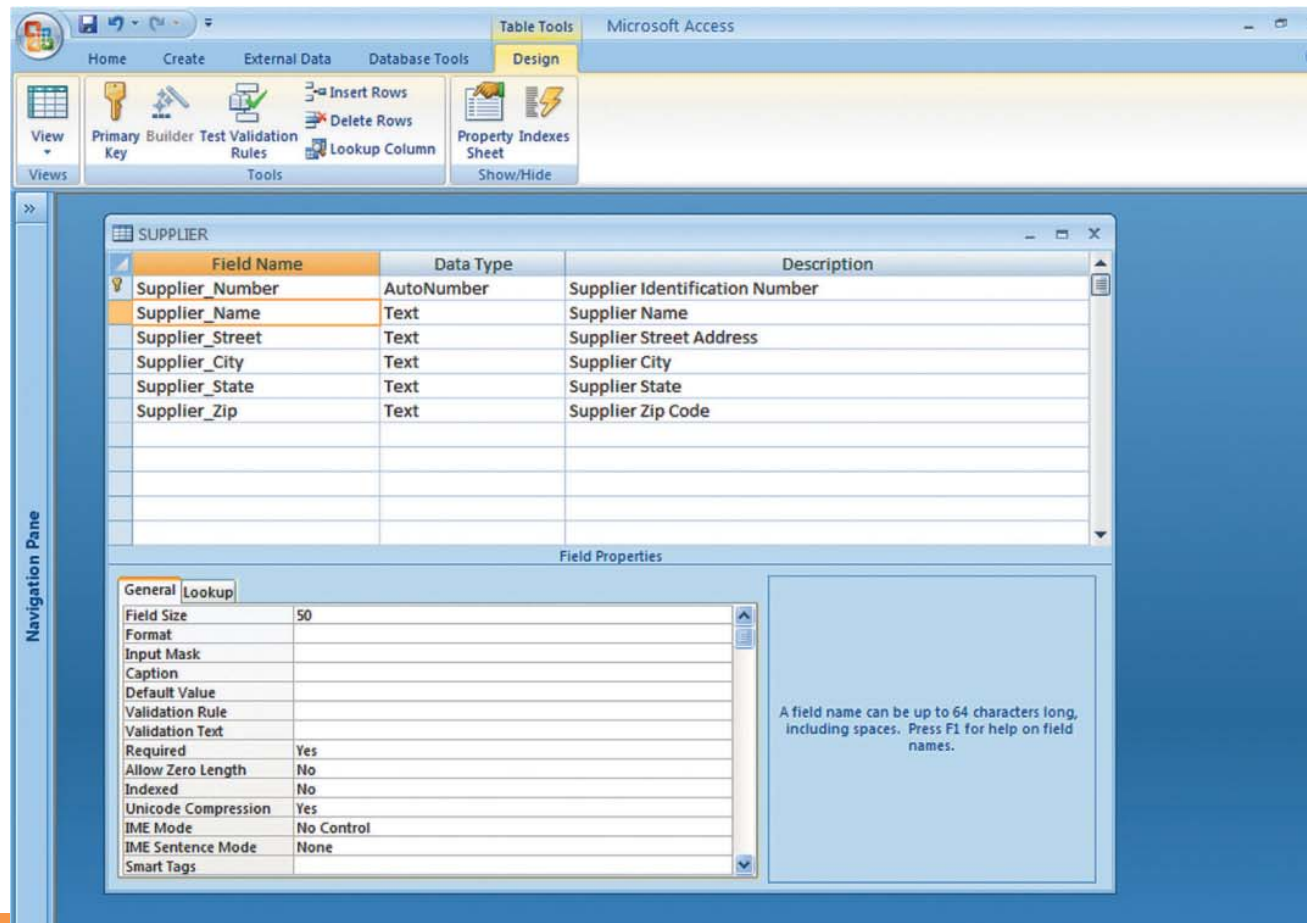


그림 5-9 액세스 데이터 사전 기능

마이크로소프트 액세스는 데이터베이스에 포함된 각 필드의 크기, 유형, 그리고 기타 특징들을 표현하기 위한 기본적인 데이터 사전 기능을 가지고 있다. 그림에서는 공급자 테이블에 대한 정보를 표시하고 있다. 공급자 번호의 왼쪽에 표시된 작은 열쇠 모양은 키 필드(key field)를 나타낸다.

데이터베이스 관리 시스템(Database Management Systems)

Example of an SQL Query

```
SELECT PART.Part_Number, PART.Part_Name, SUPPLIER.Supplier_Number,  
SUPPLIER.Supplier_Name  
FROM PART, SUPPLIER  
WHERE PART.Supplier_Number = SUPPLIER.Supplier_Number AND  
Part_Number = 137 OR Part_Number = 150;
```

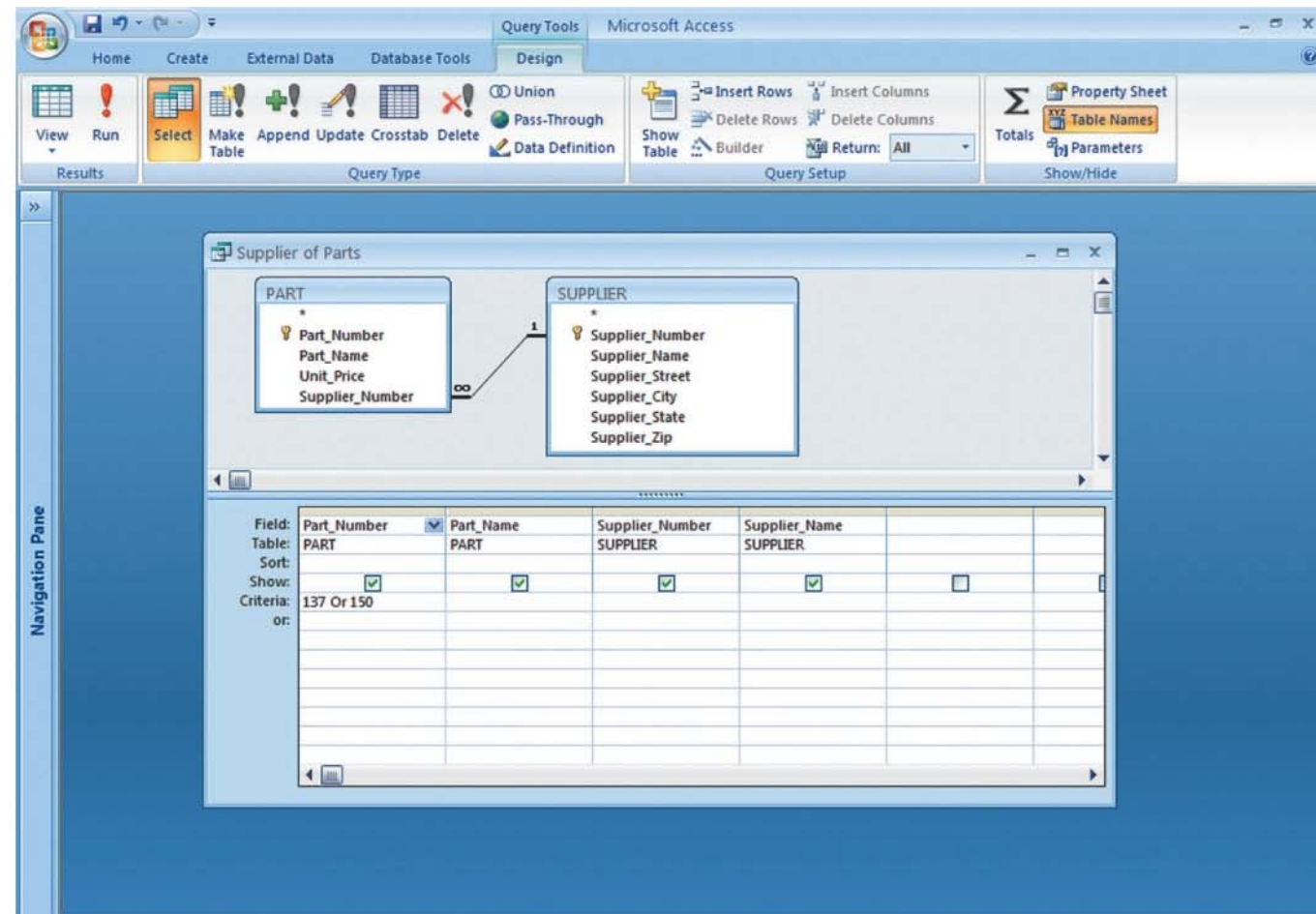
137 또는 150번 부품의 공급자를 선택하도록 하는 SQL 구문

Figure 5-10

데이터베이스 관리 시스템(Database Management Systems)

그림 5-11 액세스 질의

이 그림에서는 그림 5-10의 질의가 액세스 질의 편집기의 질의 구축 도구(query-building tools)를 사용하였을 경우에 어떻게 구축되는지 설명하고 있다. 이것은 테이블과 필드, 그리고 질의에 대한 선택기준을 나타내고 있다.



데이터베이스 관리 시스템(Database Management Systems)

객체지향 DBMS (OODBMS)

- 데이터가 하나의 객체로써 조회되고 공유되어지는 것처럼 작동하게 하도록 데이터와 절차를 저장
- 단순한 구조화된 데이터를 위한 DBMS보다 그래픽, 그림, 비디오 등을 저장하는데 더욱 적합
- 웹응용프로그램에서 멀티미디어 요소와 자바 애플릿을 관리하기 위해 사용됨.
- 관계형 DBMS와 비교할 때 상대적으로 느린 특성을 가짐.
- 객체지향 DBMS(Object-relational DBMS):

객체지향 + DBMS의 결합 형태



비즈니스 성과와 의사결정 향상을 위한 데이터베이스의 활용

- 데이터베이스는 기업을 더욱 효율적으로 운영하고, 관리자와 직원들이 보다 나은 의사결정을 할 수 있도록 정보를 제공한다.
- 방대한 양의 데이터에 접근하여 분석하게 하는 도구:
 - 데이터 웨어하우징 (Data warehousing)
 - 다차원 데이터 분석(Multidimensional data analysis)
 - 데이터 마이닝(Data mining)
 - 데이터베이스로의 웹 인터페이스 활용
(Utilizing Web interfaces to databases)

비즈니스 성과와 의사결정 향상을 위한 데이터베이스의 활용

Data Warehouses

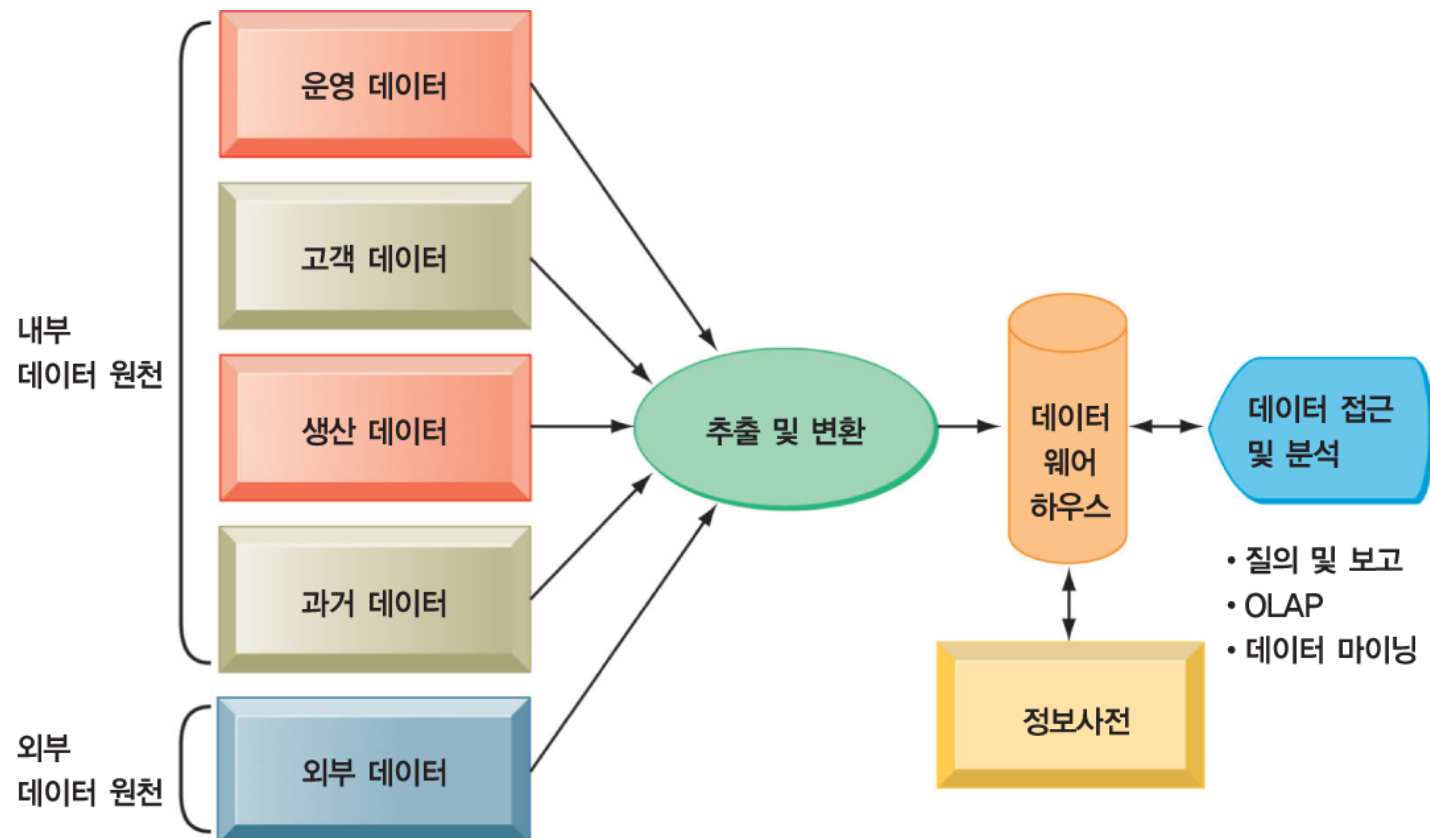
- 데이터 웨어하우스(Data warehouse):
 - 의사결정자가 관심을 가질 수 있는 과거에서 현재까지의 데이터를 저장하고 있는 데이터베이스
 - 다양한 시스템과 운영 및 거래 데이터베이스로부터 데이터를 통합하고 표준화함
 - 데이터는 접근 가능하지만 교체(수정)되지는 않음
- 데이터 마트(Data mart):
 - 특정 사용자를 위해 고도로 초점이 맞춰져 있으며, 독립된 형태의 데이터 웨어하우스의 부분 집합

비즈니스 성과와 의사결정 향상을 위한 데이터베이스의 활용

Components of a Data Warehouse

그림 5-12 데이터 웨어하우스의 구성요소

데이터 웨어하우스는 기업 내부의 다중 운영시스템들로부터 현재 및 과거 데이터를 추출한다. 이러한 데이터들은 외부 원천으로부터 제공되는 데이터와 결합되어 경영 보고 및 분석을 위해 설계된 중앙 데이터베이스로 재구성된다. 정보사전은 사용자들에게 데이터 웨어하우스에서 사용 가능한 데이터들에 관한 정보를 제공한다.



비즈니스 성과와 의사결정 향상을 위한 데이터베이스의 활용

사례연구 (조직의 관점):
미 국세청의 데이터 웨어하우스를 이용한 조세사기 적발

- 사례 연구를 읽고 다음의 질문에 답하시오.:
 - 미 국세청이 그들이 수집한 조세 자료를 분석하기 어려운 이유는 무엇이었나?
 - 미국세청이 데이터 웨어하우스(CDW)를 구축할 때 겪게 된 도전은 무엇인가?
 - 미 국세청의 데이터 웨어하우스는 어떻게 의사결정과 운영을 향상시켰는가?
납세자에게 이로운 점은 무엇인가?
 - 데이터 웨어하우스가 정부의 다른 분야에서 유용할 것으로 판단되는 가? 그
예는? 그 이유는 무엇인가?

비즈니스 성과와 의사결정 향상을 위한 데이터베이스의 활용

비즈니스 인텔리전스, 다차원 데이터분석, 데이터마이닝

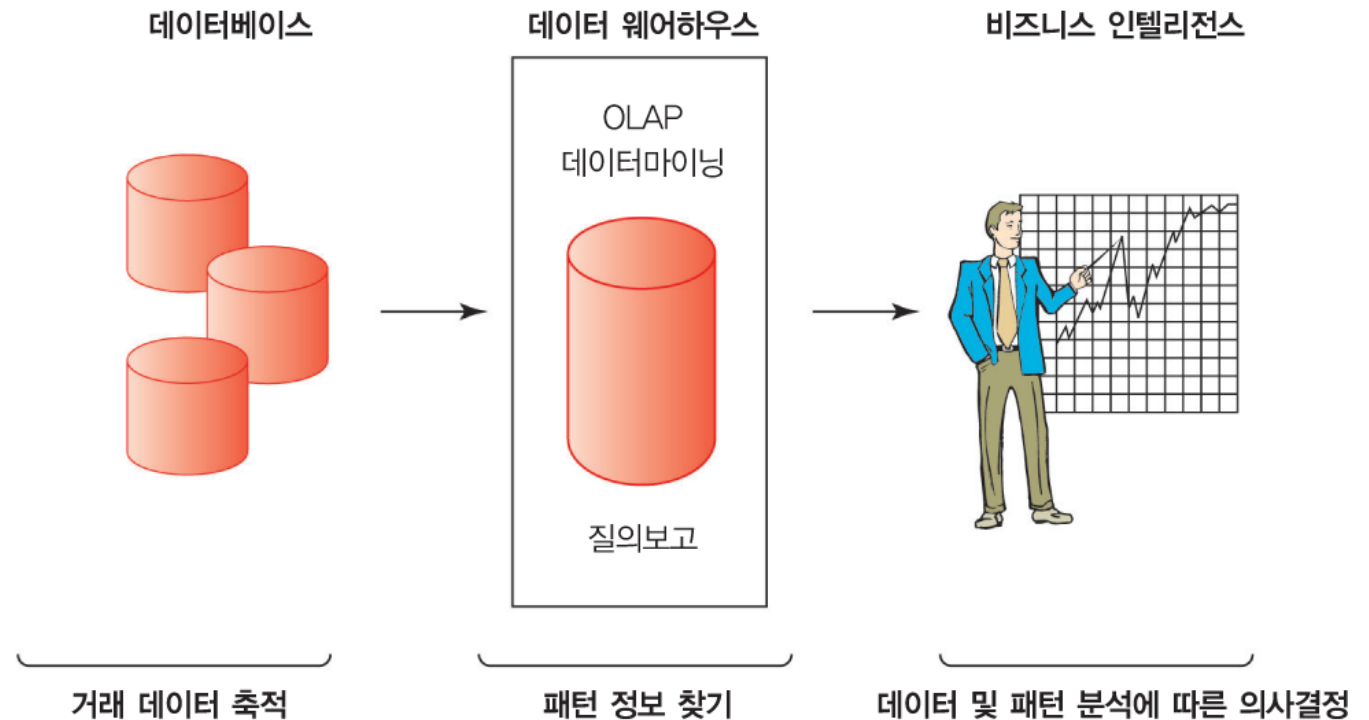
- 비즈니스 인텔리전스 (Business intelligence): 의사결정을 향상시키기 위해 방대한 양의 데이터에 접근하게하고, 통합하고 분석하게 하는 도구
 - 데이터베이스 레포팅과 질의를 위한 소프트웨어 (Software for database reporting and querying)
 - 다차원 분석을 위한 도구 (OLAP :online analytical processing)
 - 데이터 마이닝(Data mining)
- 예)카지노 기업인 Harrah ' s Entertainment는 고객의 데이터를 분석하여 도박 프로파일을 만듦으로써 가장 높은 수익의 고객을 식별하게 한다.

비즈니스 성과와 의사결정 향상을 위한 데이터베이스의 활용

Business Intelligence

그림 5-13 비즈니스 인텔리전스

일련의 분석도구를 이용하여 데이터베이스에 저장된 데이터를 분석함으로써 관리자와 종업원들은 조직성과의 향상을 위해 더 나은 의사결정을 하는데 도움을 줄 수 있는 패턴과 통찰력을 발견할 수 있다.



비즈니스 성과와 의사결정 향상을 위한 데이터베이스의 활용

Online Analytical Processing (OLAP)

- 사용자가 똑같은 데이터를 다차원의 관점을 통하여 다른 방식으로 볼 수 있도록 하게함으로써 다차원 데이터 분석을 지원
 - 정보의 개별적 측면—product, pricing, cost, region, 또는 time period—각기 다른 차원을 나타낸다.
 - 예) 미 동부에서의 월별 매출 차이를 분석
- 사용자가 짧은 시간에 즉각적인 응답을 온라인을 통해 제공받을 수 있도록 한다.

비즈니스 성과와 의사결정 향상을 위한 데이터베이스의 활용

다차원 Data Model

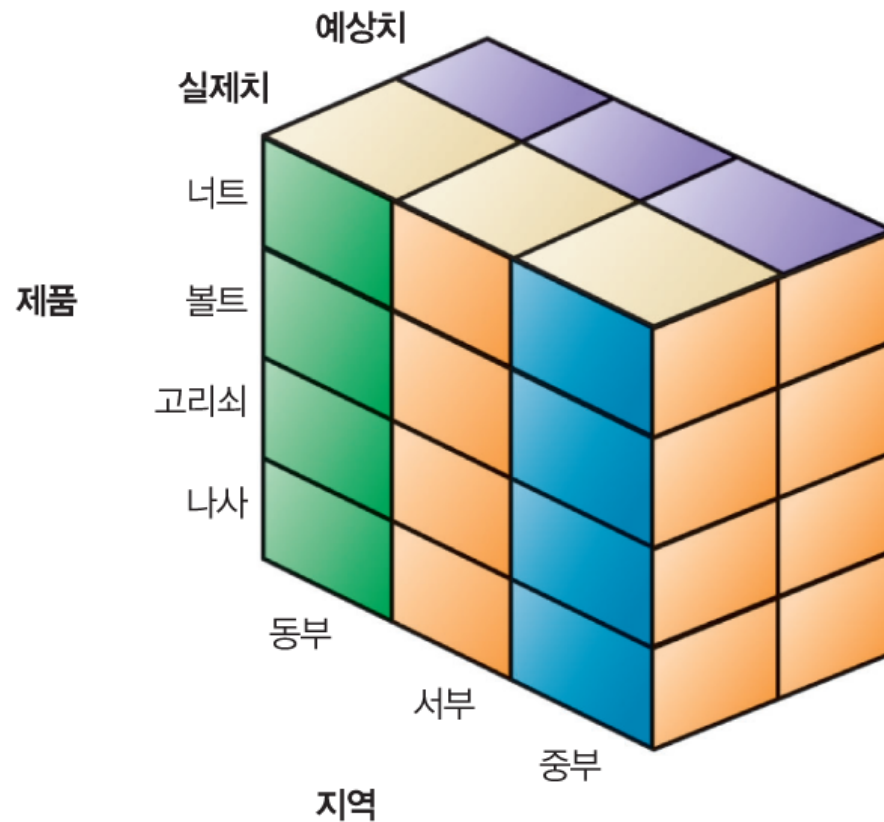


그림 5-14 다차원 데이터 모델

그림에서 보여주고 있는 관점은 제품 대 지역에 대한 것이다. 만일 이 입방체를 90도 회전시키면, 제품 대비 실제 및 예상 판매액에 대한 관점이 나타날 것이다. 그리고 또 다시 90도 회전시키면, 지역 대비 실제 및 예상 판매액을 볼 수 있다. 다른 관점들도 가능하다.

비즈니스 성과와 의사결정 향상을 위한 데이터베이스의 활용

데이터마이닝 (Data Mining)

- 대량의 데이터베이스 속에서 숨겨진 패턴과 관계를 찾아내어, 밝혀진 규칙을 통해 미래의 행위를 예측하게 하는 것
- 데이터 마이닝으로 부터 얻을 수 있는 정보 유형
 - 연관성(Associations): occurrences linked to single event
 - 순차(Sequences): events linked over time
 - 분류(Classifications): patterns describing a group an item belongs to
 - 군집(Clusters): discovering as yet unclassified groupings
 - 예측(Forecasting): uses series of values to forecast future values

비즈니스 성과와 의사결정 향상을 위한 데이터베이스의 활용

데이터 마이닝

- 데이터 마이닝의 활용 : 예) 고객 데이터를 분석하여 수익성 있는 고객을 찾아내어 일대일 마케팅을 실시하게 한다.
- 예측적 분석(Predictive analysis):
 - 데이터마이닝 기법과 과거의 데이터, 미래 상황에 대한 가정들을 사용하여 이벤트의 결과를 예측하는 것 : 예) 한 고객이 특정 제품에 대한 구매 혹은 제안에 응답할 가능성을 예측
- 데이터마이닝 versus 프라이버시 침해 (privacy concerns)
 - 각각의 개인에 대한 상세 데이터를 만들기 위해 사용

- **텍스트 마이닝 (Text Mining)**
 - 비구조화된 텍스트 파일들은 한 조직의 유용한 정보의 80%를 차지한다.
 - 텍스트 마이닝은 기업들이 대량의 비 구조화된 데이터 집합으로부터 주요 요소를 추출하고, 패턴을 발견하고, 요약하게 하는 기능을 제공한다.
- **웹(Web Mining)**
 - 웹으로부터 유용한 패턴과 정보를 발견하고 분석하는 것
 - 예) Content mining, structure mining, usage mining

비즈니스 성과와 의사결정 향상을 위한 데이터베이스의 활용

데이터베이스와 웹

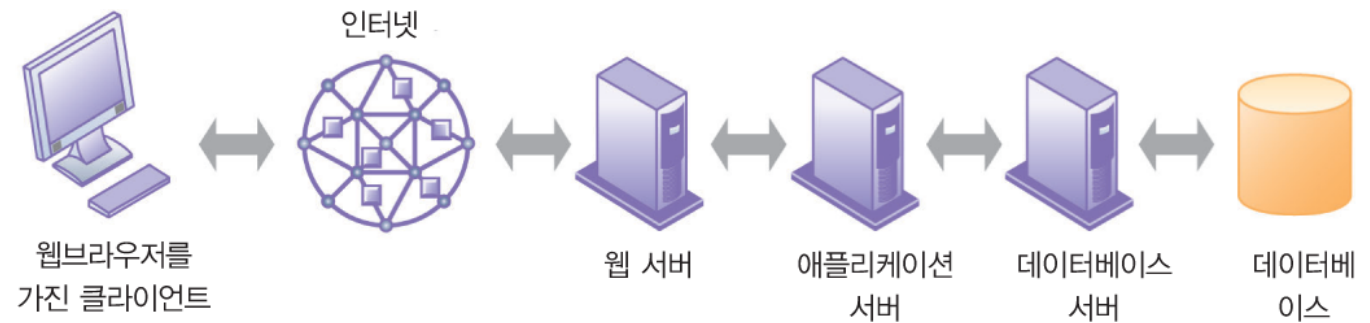
- 기업들은 웹을 이용하여 그들 내부의 데이터베이스로부터 정보를 산출하여 고객과 파트너에게 제공
- Middleware와 그 밖의 소프트웨어가 필요
 - Web server
 - Application servers or CGI
 - Database server
- 웹인터페이스는 사용자에게 사용적 측면의 친밀감 제공, 현존하는 시스템(Leagacy system)을 재설계하거나 재구축하는 비용을 절감

비즈니스 성과와 의사결정 향상을 위한 데이터베이스의 활용

내부데이터와 웹과의 연결

그림 5-15 내부 데이터를
웹에 연결하기

사용자들은 데스크톱 PC
와 웹 브라우저 소프트웨어를
사용함으로써 웹을
통해 조직의 내부 데이터
베이스에 접근할 수 있다.



데이터 자원의 관리

정보정책의 수립

- 정보정책(Information policy)
 - 정보를 조직하고, 관리하고, 저장하여, 공유하게하는 조직의 규칙을 기술
- 데이터 관리(Data administration)
 - 데이터가 자원으로 관리되어질 수 있도록 하는 특정 정책 및 절차
- 데이터베이스 관리(Database administration)
 - 데이터베이스의 구조와 내용을 정의하고 구성하며, 이를 유지하는 책임을 나타내는 데이터베이스 설계 및 관리 집단의 역할을 기술

데이터 자원의 관리

데이터 품질 보장 (Ensuring Data Quality)

- 데이터 품질의 저하: 성공적인 고객 관계 관리의 주요 장애
 - 데이터 품질 문제의 야기:
 - 다양한 시스템으로부터 산출되는 중복되고 일치하지 않는 데이터
 - 데이터 입력 오류
 - 데이터 품질 감사(Data quality audit): 데이터의 정확성과 완전성에 대한 구조화된 조사
 - 데이터 정제(Data cleansing): 부정확, 불완전, 부적절한 형태 및 중복 데이터의 발견 및 수정
- 

데이터 자원의 관리

사례연구 (인간의 관점) 개인 신용평가 기관의 오류-중요한 인적 문제

- 사례연구를 읽고 다음의 물음에 답하시오. :
 - 신용평가 기관의 데이터 품질과 관련된 문제가 미칠 수 있는 비즈니스 충격에 대하여 평가해보시오.
 - 신용평가 기관의 데이터품질 문제에 의해 야기 될 수 있는 윤리적 문제가 있는가?
 - 신용평가 기관의 데이터 품질 문제에 대한 책임이 있는 인간, 조직, 기술의 요인들을 분석하시오.
 - 이러한 문제를 해결할 수 있는 해결책은 무엇인가?