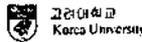


# 효율적인 전자상거래를 위한 데이터 웨어하우스의 활용

정인정



고려대학교  
Korea University

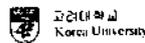


한국전산원  
National Computerization Agency

## 전자상거래(Electronic Commerce)(1)

■ 정의 : 상업행위(거래)에 관한 제반 업무들을 컴퓨터 네트워크를 통해 수행하는 제반 상거래 행위

- ◆ 전자금융
- ◆ 전자문서교환 (HTML, SGML, CGI, SSL, Java, EDI, SET, JEPI, ...)
- ◆ 전자인증 (Certified Authority)
- ◆ 의사결정



고려대학교  
Korea University



한국전산원  
National Computerization Agency

## 전자상거래(Electronic Commerce)(2)

### ◆ 전자상거래 프로토콜

- IKP (Internet Keyed Payment) : IBM
- SEPP (Secured Electronic Payment Protocol) :  
개방형 표준을 제공하는 최초의 다자간 프로토콜
- SET (Secure Electronic Transaction) :
  - 인터넷과 같은 네트워크 상에서 안전한 신용카드 거래를 위한 기술사양
  - RSA 데이터 보안회사의 암호화 기술에 기초
  - 상호운용성(interoperability) 보장

## EC의 특징

- 분산 컴퓨팅 및 네트워크 환경에서의 상업행위
- 인터넷 :

국부적이며 실제적인 상업시장  
(local and physical marketplace)



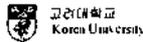
광역적이며 가상적인 상업시장  
(global and virtual marketplace)

- 전자 쇼핑을 이용한 상품의 판매 및 구매에 대한 새로운 전략 방법
- 시간 및 공간적 제약의 극복 → 신속하며 능률적인 상거래

## OLTP vs. OLAP

### - 종래의 OLTP (1) -

- 운영 데이터베이스 (operational DB)
  - ◆ DB의 내용 : 현재상태 중심으로 구축
  - ◆ 업무처리 중심으로 구성 → 트랜잭션 처리중심
  - ◆ 트랜잭션이 발생할 때 마다 갱신 → 최신의 스냅샷만을 보관
- 데이터 모델 : 개체 - 관계 (ER : Entity - Relationship) 모델
  - ◆ 개체간의 데이터의 의존성 (dependency)을 표현
  - ◆ 트랜잭션 처리에 적합
  - ◆ 데이터를 여러 개의 개체(entity)로 나누고 각 개체를 하나의 테이블로 정의 → 데이터가 여러 개체에 산재 → 다수의 Join 명령문 필요
  - ◆ 이해하기 복잡하며 사용자 관점에 따른 질의어 처리가 어려움
  - ◆ 스키마가 복잡 + 질의에 대한 다수의 경로 (path) 존재



고려대학교  
Korea University

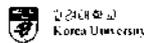


한국직업능력  
National Competency Agency

## OLTP vs. OLAP

### - 종래의 OLTP (2) -

- 업무 application마다 데이터가 산재
  - ◆ 데이터를 횡적으로 통합하여 실시간 분석이 어려움
- 요약되지 않는 정보 : 데이터의 양이 증가  
(※ 거미집(spider web) 부작용 및 자연적 진화 구조 (naturally evolving architecture))
  - ◆ 많은 비용의 증가
  - ◆ 데이터의 신뢰성 부족
- 정리되지 않은 데이터를 현장의 최일선에 적시에 공급이 어려움
- 과거의 데이터를 바탕으로 현재의 문제를 유추하는 것이 어려움
  - ◆ 데이터에서 정보로의 변환이 어려움
- 과거에서 현재까지의 누적된 이질적인 데이터 관리가 어려움



고려대학교  
Korea University

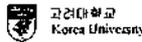


한국직업능력  
National Competency Agency

## OLTP vs. OLAP

### - OLAP(Online Analytical Processing) -

- 1993년 Codd의 정의 : 동적자료에서 필요한 정보를 생성, 조작, 활성화, 종합하는데 필요한 역동적인 자료분석
- 1995년 Pendse 및 Creeth의 정의 : 공유되는 다차원 정보에 대한 신속한 분석
- OLTP의 상대적 개념 :
  - ◆ OLTP는 업무처리 중심으로 구성되어 각 순간마다의 조직(시스템)의 현재 상황을 알려 줌
  - ◆ OLAP은 최종 사용자가 조직(시스템)의 전반적인 상황을 이해하고 최적의 의사결정을 하도록 지원함
  - ◆ 그 조직(시스템)이 나아갈 방향을 제시
- OLAP 서버 (데이터마트) :
  - ◆ OLAP을 사용자 환경으로 구축해 놓은 것
  - ◆ DWH 환경에서 다차원 정보를 제공하는 분석용 데이터마트



고려대학교  
Korea University

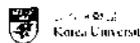


한국전산원  
National Computerization Agency

## 데이터 웨어하우스(Data Warehouse)

### - 정의 -

- DSS 및 EIS를 지원하는 다차원 DB
  - 1) 주제중심적 (subject-oriented)
  - 2) 통합적 (integrated)
  - 3) 시간변이적 (time variant)
  - 4) 비휘발적 (non-volatile)
- 현재상태의 데이터 + 과거에서 지금까지 누적된 통합 데이터의 저장 및 분석 → DSS 및 EIS에 필요한 정보의 추출 및 제공



고려대학교  
Korea University



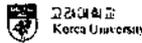
한국전산원  
National Computerization Agency

## 데이터 웨어하우스(Data Warehouse)

### - 개념(1) -

#### ■ 특징 :

- 1) 아키텍처 환경의 핵심 : 두 가지 종류의 데이터
  - 원시데이터 (primitive data) :
    - 트랜잭션 업무처리에 대한 데이터
    - 갱신가능
  - 파생데이터(derived data):
    - DSS + EIS에 필요한 요약된 데이터
    - 과거에서 현재까지 누적된 데이터
- 2) DSS 처리를 위한 기반 : 주제 중심적인 구조 + 데이터와 시간을 함께 저장
- 3) 하나의 통합된 데이터 원천
- 4) 데이터 액세스가 용이한 IT의 중심



고려대학교  
Korea University

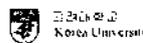


한국전산원  
National Computerization Agency

## 데이터 웨어하우스(Data Warehouse)

### - 개념(2) -

- 경쟁력 우위를 확보할 수 있는 전략을 갖춘 정보기술의 총체
- 의사결정 지원시스템(DSS) 및 임원정보시스템(EIS)이 가능한 새로운 정보시스템
- 현재 상태만의 데이터 뿐만 아니라 과거의 데이터도 유지
  - ◆ 누적된 통합 데이터를 분석
  - ◆ 필요한 정보의 공급
- OLTP 개념에서 OLAP 개념의 주제 중심적인 구조 + 데이터와 함께 시간을 저장 → 분석하고자 하는 데이터의 통합된 과거 데이터의 플랫폼을 제공하여 정보처리
- 1990년대 후반에 등장한 정보기술(IT)의 중심 핵
- 과거의 데이터와 내외부 데이터를 주제별로 통합
  - ◆ 별도의 프로그램 없이 즉시(on-line) 여러 각도에서 분석이 가능



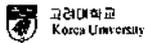
고려대학교  
Korea University



한국전산원  
National Computerization Agency

## EC에서의 데이터 웨어하우스의 필요성

- 이익의 극대화를 위한 전략
- 유연성(적응성) 및 인내성
- 기동성(maneuverability)
- 효과적인 의사결정 및 동향파악과 문제점 해결(전략적 사고)
  - ◆ 시장 경향분석
  - ◆ 주요 상업지표 측정 및 추적
  - ◆ drill-down 분석
  - ◆ 문제점 파악 및 해결제시
  - ◆ 누적된 통합 데이터의 분석 및 정책결정
  - ◆ 주제중심적 데이터 구조 및 정보추출



고려대학교  
Korea University



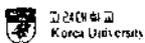
한국전산원  
National Computerization Agency

## Motivating Example

	Store		Product		Date			Sales
	Loc	Item	Manu	y	m	d		
1	New York	Mac	Apple	1996	May	15	7000	
2	Washington	Mac	Apple	1996	May	15	3800	
3	Los Angeles	Mac	Apple	1996	May	15	15500	
4	New York	PC	IBM	1996	May	15	2700	
5	Washington	PC	IBM	1996	May	15	3900	
6	Los Angeles	PC	IBM	1996	May	15	6530	

(a) For each member store, find out the year-to-date total sales amounts for each day of this year(i.e., the daily cumulative sales amounts over this year).

(b) Find the year-to-date total sales amounts, in each region(cast, west, etc.), of each product whose last year's nation-wide total sales ranked among top five.



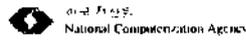
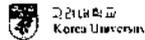
고려대학교  
Korea University



한국전산원  
National Computerization Agency

## 데이터 웨어하우스의 뷰 형상화 (View Materialization)

- 분산된 데이터 소스들로부터 정보를 수집하여 단일 데이터베이스로 결합 → client로부터의 질의어 처리
- 외부 데이터 소스의 기초데이터에 관한 뷰 형상화
  - ◆ 대량의 배치파일의 주기적인 변경
- 2VNL 알고리즘 : 다층적 동시성 제어
- 효과적인 질의어 처리 및 의사결정
  - ◆ 다층적 질의 최적화(MQO : Multiple Query Optimization)



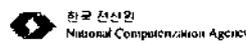
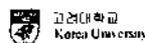
## 관리 트랜잭션의 보기

	city	state	product line	date	total sales
insert :	San Jose	CA	golf equip	10/16/96	11,000
insert :	Novato	CA	rollerblades	10/13/96	6,000
update :	San Jose	CA	golf equip	10/14/96	10,000 → 10,200
delete :	Berkeley	CA	racquetball	10/14/96	12,000

Example maintenance transaction

Tuple VN	operation	city	state	production line	date	total sales	pre_total_sales
5	update	San Jose	CA	golf equip	10/14/96	10,200	10,000
4	insert	San Jose	CA	golf equip	10/15/96	1,500	null
5	delete	Berkeley	CA	racquetball	10/14/96	12,000	12,000
5	insert	Novato	CA	roller blades	10/13/96	6,000	null
5	insert	San Jose	CA	golf equip	10/16/96	11,000	null

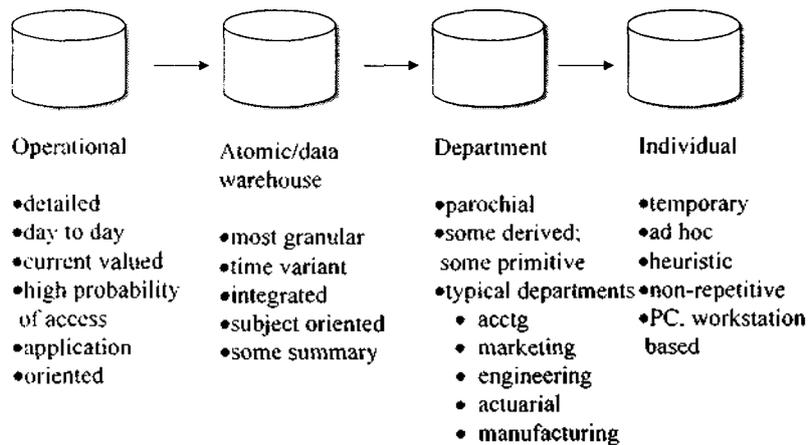
Result of Daily Sales after maintenance transaction



## 데이터 웨어하우스의 아키텍처 환경

- 운영수준(operational level)
- 데이터 웨어하우스 수준(data warehouse level)
- 부서수준(departmental level)
- 개인수준(individual level)

## Levels of the Architecture



## 다차원 데이터 모델링(1) (Multidimensional Data Modeling)

- 관계형 데이터 모델링(RDM)
  - ◆ OLTP 목적의 운영 데이터베이스(operational database) 지원
- 다층차원 데이터 모델링(MRDM)
  - ◆ 보고 및 분석(reporting and analysis) 지원
- RDM과 MRDM의 비교
  - ◆ 트랜잭션 vs. 시간개념
  - ◆ 국부적 일치성(local consistency) vs. 광역적 일치성(global consistency)
  - ◆ audit trail vs. big picture
  - ◆ 명시적 관계 vs. 함축적 관계

## 다차원 데이터 모델링(2) (Multidimensional Data Modeling)

- 다중 데이터 뷰(multiple data view)
  - ◆ 보기 : sales by category, brand, season, store, etc
- 효과적인 의사지원 시스템
- 종류 :
  - ◆ client - side
  - ◆ server - based
  - ◆ spreadsheets
- roll-up(개략적인 정보) / roll-down(상세한 정보)

## 다차원 데이터 모델링 도구

- Based on the notion of arrays
- an organizational principle for arranging and storing related data
- views and analysis from multiple perspectives
- summarization and arrangements : cross-tabular views



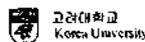
고려대학교  
Korea University



한국전산원  
National Computerization Agency

## 다차원 데이터베이스

- 복잡한 의사결정에 관한 질의어 처리를 위한 데이터 저장소
- 데이터의 유연한 액세스 및 다양한 방법에 의한 데이터 분류
- 효과적인 의사지원 시스템
- 대용량의 자료를 저장, 분석
- 다중차원으로 배열될 수 있도록 array에 저장
- 형상화된 뷰들의 선택문제
  - ◆ 다중차원 격자구조(multidimensional lattice) :
    - 문제의 해답공간
    - 뷰 사이의 종속관계 표현
- 데이터 큐브(data cube)
  - ◆ graphical representation of multiple aggregate views
  - ◆ derived from a fact table with different sets of attributes
  - ◆ intuitive way to navigate various levels of summary information in database

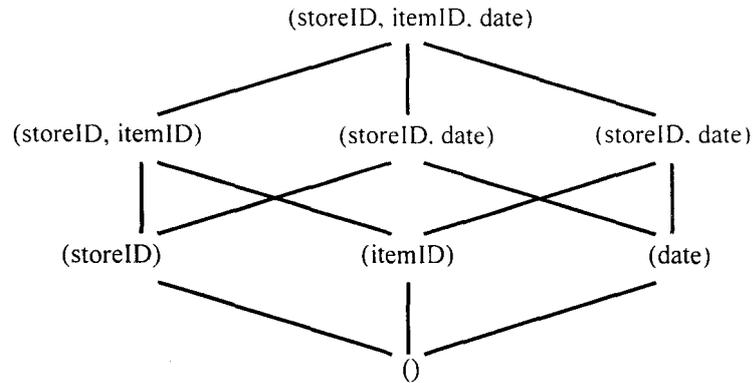


고려대학교  
Korea University



한국전산원  
National Computerization Agency

## Data Cube Lattice



## Star Join

- 데이터 웨어하우스에서 entity에 대량의 데이터가 존재할 때 필요한 설계구조
- 사실 테이블(fact table)
  - ◆ star join에서 중심이 되는 entity
  - ◆ 숫자 데이터 + foreign key
- 차원 테이블(dimension table)
  - ◆ star join에서 사실 테이블 주변의 테이블들
  - ◆ 문자 데이터

