ICS 434
Advanced Database Systems

Dr. Abdallah Al-Sukairi
sukairi@kfupm.edu.sa

Second Semester 2003 - 2004 (032)

King Fahd University of Petroleum & Minerals
Information & Computer Science Department

Outline

- The Relational Data Model: Version2
- Advanced Data Modeling
- Databases on the Web
- Client-Server Architecture
- Client-Server Databases & Tools
- The System Catalog
- Query Processing and Optimization
- Transaction Processing
- Concurrency Control
- Recovery
- Administration & Security
- Distributed Databases
- Database Replication
- Object-Oriented Databases
- Data Warehousing and Data Mining
- Other Emerging Database Technologies

History of DBMS Hardware

- 1970s
  - Mainframes & Minicomputers
- 1980s
  - Personal Computers & Workstations (desktop machines)
- Mid 1980s
  - LAN to share resources
- 1990s and beyond
  - LAN + PCs → client-server computing
  - LAN + WAN → distributed computing
  - Internet

History of DBMS Software

- 1970s
  - Hierarchical & Network Data Models
- 1980s
  - Relational model
  - Mainframe relational DBMSs (DB2, Oracle, ...)
- Late 1980s
  - Database servers (Oracle, Sybase, SQL Server)
- 1990s and beyond
  - Database serves + Front-end tools → client-server databases
  - Distributed & Replicated Databases
  - Data Warehouses
  - Web/Database Integration
  - Object-Oriented Databases

1. The Relational Model

The Relational Model - Version 1

- Data models, Schema, and Instances
- Rests on a solid mathematical foundation
- Simple and elegant
- Flexible enough to accepts changes
- The basic relational model consists of:
  - Structural Part (relations, domains, etc.)
  - Integrity Part (entity integrity, referential integrity, etc.)
  - Manipulative Part (select, join, etc.)

Relation

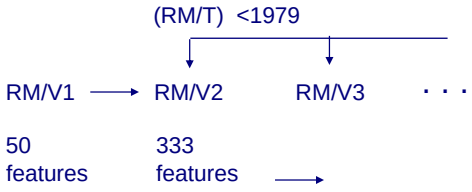
- Every relation is a set
- Every relation can be perceived as a table
- When a relation R is conceived as a table (special table R-table), it has the following properties:
  - Each row represents a tuple of R
  - The ordering of rows is immaterial (no nextness)
  - No duplicate rows
- Why duplicate rows are not allowed ?

Normalization

- A theory to measure how ‘good’ a relation schema is
- First Normal Form (1NF)
  - Repeating groups
- Second Normal Form (2NF)
  - A column’s dependency on only part of a composite key
- Third Normal Form (3NF)
  - A nonkey column representing a fact about another nonkey column

History of the Relational Model

- The model was originated by E. F. Codd who published 30 papers on the model from 1969 through 1988
- All his papers and more are now collected in a book (1990)



Version 2 of the Relational Model

- Goals of RM/V2
  - Simplify interaction with data
  - Increase the productivity of programmers
  - Support more powerful tools for the DBA
- No implementation details
- RM/V2 features are divider among 18 classes:
  - Structural
    Basic operators
    Naming
    Qualifiers
    Extended data types
    Advanced operators
    Elementary commands
    Indicators

Domain Concept

- Full support for many of the features in the relational model depends on the full support of the domain concept
- Domain (extended data type) CREATE DOMAIN
  - name + basic data type + range of values permitted + whether the less than comparator (<) is meaningfully applicable to its values
- Primary Key: minimal, unique, and not null
- Foreign Key

. . . Domain Concept

- Referential Integrity
  - Let D be a domain from which one or more primary keys draw their values. Let K be a foreign key, which draws its values from domain D. Every unmarked value which occurs in K must also exist in the database as a value of the primary key on domain D of some base relation

When Is a DBMS Relational ?

- 'A DBMS product is fully relational in the 1990s if it fully supports each and every one of the features of RM/V2 defined in this book'
  [Codd, 1990]
- Codd's Twelve Rules
  - Published in 1985, now included in RM/V2
  - A system is practically relational if it implements a significant fraction of these rules
- Rule 0
  'For any system that is advertised as, or claimed to be, a relational database management system, that system must be able to manage database entirely through its relational capabilities'

Codd's Twelve Rules

- R1: Information Rule
- R2: Guaranteed Access
- R3: Missing Information
- R4: Active Catalog
- R5: Comprehensive Data Sublanguage
- R6: View Updatability
- R7: High level insert, update, and delete
- R8: Physical Data Independence
- R9: Logical Data Independence
- R10: Integrity Independence
- R11: Distribution Independence
- R12: Non-Subversion

Missing Information

- Treating missing information in a systematic and general way
- Missing db-value: ("null" in RM/V1)
  - missing-but-applicable (unknown) (A-mark)
  - missing-but-inapplicable (unknowable) (I-mark)
- A mark is neither a value nor a variable
- In column C, each row contains a value for C or an A-mark or an I-mark
- A-mark — db-value — I-mark
  - with authorization
  - with authorization

. . . Missing Information

- A composite foreign key can have some (or all) components with A-marks. Unmarked components should adhere to the referential integrity constraint
- If I-marks are placed in the top class, A-marks in the second class, and all db-values in the third class, the combination (arithmetic or otherwise) of any two items is an item of whichever class is the higher of the two operands

. . . Missing Information

- Equality of marks
  - symbolic (used in ORDER BY of SQL)
    A-mark = A-mark
  - semantic
    A-mark may = A-mark
    Salary > 1000 ? where Salary has A-mark
    the answer should be MAYBE
- (Birthdate > 66-1-1) or (Salary < 10,000)

Three-Valued Logic

- The truth tables of the three-valued logic of RM/V1
  - t = true f = false m = maybe

not	
t	f
m	m
f	t

or	t	m	f
t	t	t	t
m	t	m	m
f	t	m	f

and	t	m	f
t	t	m	f
m	m	m	f
f	f	f	f

- Tautologies cause problems:
  - (Salary < 1000) or (Salary = 1000) or (Salary > 1000)

Four-Valued Logic

- The truth tables of the four-valued logic of RM/V2
  - t = true f = false a = maybe-and-applicable
  - i = maybe-and-inapplicable

not	
t	f
a	a
i	i
f	t

or	t	a	i	f
t	t	t	t	t
a	t	a	a	a
i	t	a	i	f
f	t	a	f	f

Qualifiers

- MAYBE_A
- MAYBE_I
- MAYBE

EMP ...	Salary	Bonus
1	10,000	- I
2	- A	2,000
3	12,000	1,800
4	- A	- I
5	15,000	- A
6	- I	2,500

- Q ← EMP [(Salary > 11,000) and (Bonus < 4,000)]
- Q = 3 Q MAYBE_A = 2, 5 Q MAYBE_I = 4, 6
- Q union (Q MAYBE) = ?

Criticisms

- Default-value approach
  - Many problems
  - Q1: Find the suppliers in Jubail
  - Q2: Find the suppliers NOT in Jubail
- The rules of functional dependencies do not apply to marks
- RM/V2 does not handle the "half-missing" case