Relational Database Basics Review

- Overview
- Database approach
- Database system
- Relational model

1

File Processing Approaches

- Based on file systems
  - Data are recorded in various types of files organized in folders (directories)
- File types
  - Sequential data files
  - Name-value pair files
  - Spreadsheets or list files
  - XML files

2

Files Processing Problems

- Difficult to handle complex data
- Low data quality: redundancy and inconsistency
- No central management
- Difficult to maintain and share in multi-user environments

3

Files Processing Problems

- Limited security
- Not scalable: cannot handle large quantity of data efficiently
- Lack of specialized and standardized data management and processing capabilities

4

Database Approach

- Database is a structured and self-describing collection of data
- Structured: structures and rules are consistently and rigorously defined and enforced (integrity)
- Self-describing: the description of data (data definition, or metadata) is contained within the database

5

Database Approach

- Centralized management
- Managed and controlled by specialized programs, called database management systems (DBMS),
- which provides rich data management functionalities

6

Advantages of Databases

- High data quality, integrity, and consistency
- Reduced data redundancy and application maintenance
- Easy access and sharing
- Scalable
- Improved security
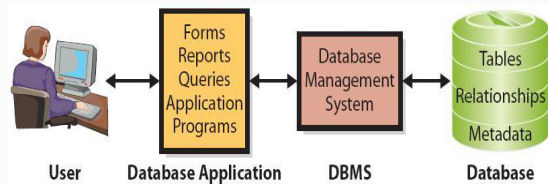- Specialized and productive management tool

7

Major disadvantages

- Increased complexity
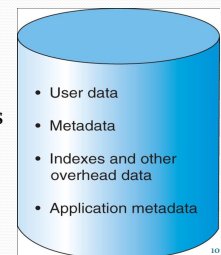- Greater impact of failure

8

Database System

- A database system is a complete information system
- Basic layers of a database system

9

Database

- A database is a storage place for data
- What's in the database?
  - Data (tables)
  - Metadata
  - Other data and structures

- User data
- Metadata
- Indexes and other overhead data
- Application metadata

10

Metadata

- Metadata are data that describe data (data definitions)
- Metadata is always a part of a database.

11

Database Management System

- DBMS serves as a controller (gatekeeper) for databases
- DBMS provides common functionalities and interfaces for managing and controlling database activities, such as
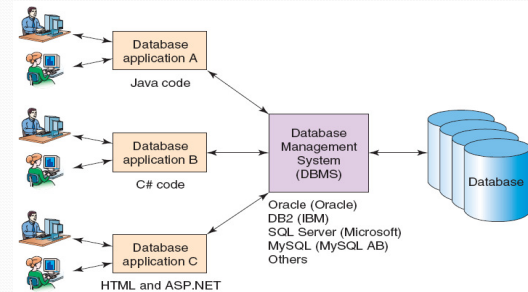- creating and maintaining databases and other structures

12

Database Management System

- reading, updating and deleting data
- data backup and recovery
- controlling concurrency, consistency, and enforcing other rules
- providing security

13

Enterprise Database System

14

Introduction to Relational Databases

- The relational approach was originally proposed in 1970's .
- The first project that proved the practicality of the relational model is System R, developed at IBM's San Jose Research Laboratory in 1976. examples oracle, Sybase, access

15

Introduction to Relational Databases cont:

- The model was proposed as a disciplined way of handling data using the rigour of mathematics, particularly set theory.
- This would enhance the concept of program-data independence
- and improve programmer activities.

16

Introduction to Relational Databases cont:

- The relational model will have only values.
- Even references between data in different sets (relations) are represented by means of values.
- In the hierarchical and network model there are explicit references (pointers), which make them more complicated.

17

Relational Data structure

- The Relational approach is based on elementary mathematical relation theory.
- Its basic construct is a relation. A relation is also called a table.
- The data is organized in tables. The table has columns and rows.

18

3

Relational Model Terminology

- A relation is a table with columns and rows.
  - Only applies to logical structure of the database, not the physical structure.
- Attribute is a named column of a relation.
- Domain is the set of allowable values for one or more attributes.

19

Relational Model Terminology Cont:

- Tuple is a row of a relation.
- Degree is the number of attributes in a relation.
- Cardinality is the number of tuples in a relation.
- Relational Database is a collection of normalized relations with distinct relation names.

20

Instances of Branch and Staff (part) Relations

21

Examples of Attribute Domains

Attribute	Domain Name	Meaning	Domain Definition
branchNo	BranchNumbers	The set of all possible branch numbers	character: size 4, range B001–B999
street	StreetNames	The set of all street names in Britain	character: size 25
city	CityNames	The set of all city names in Britain	character: size 15
postcode	Postcodes	The set of all postcodes in Britain	character: size 8
sex	Sex	The sex of a person	character: size 1, value M or F
DOB	DatesOfBirth	Possible values of staff birth dates	date, range from 1-Jan-20, format dd-mmm-yy
salary	Salaries	Possible values of staff salaries	monetary: 7 digits, range 6000.00–40000.00

22

Alternative Terminology for Relational Model

Table 3.1 Alternative terminology for relational model terms.

Formal terms	Alternative 1	Alternative 2
Relation	Table	File
Tuple	Row	Record
Attribute	Column	Field

23

Properties of Relations

- Relation name is distinct from all other relation names in relational schema.
- Each cell of relation contains exactly one atomic (single) value.
- Each attribute has a distinct name.
- Values of an attribute are all from the same domain.

24

4

Properties of Relations

- Each tuple is distinct; there are no duplicate tuples.
- Order of attributes has no significance.
- Order of tuples has no significance, theoretically.

25

Relational keys

- A key is one or more columns of a relation that is used to uniquely identify a record
- Primary key
- Candidate key
- Alternate key
- Surrogate key
- Composite key
- Foreign key

26

Candidate Key

- The minimum set of column(s) that uniquely identifies a single record (row)
- Each value in this column is unique in this relation

27

Primary Key

- Primary key is a column/attribute that is used to uniquely identify a record
- Is one of the candidate keys chosen to be the identifying key; others become alternate keys
- Each value of this key column uniquely identifies a single record (row)
- There is only ONE primary key for a table

28

Composite Key

- A composite key contains two or more attributes (columns)
- Example:
- "FirstName" + "LastName"
- "FirstName" + "LastName" + "BirthDate"
- "FirstName" + "LastName" + "BirthDate" + "BirthCity"

29

Artificial Primary Key/Surrogate Key

- Sometimes it is difficult to find a natural attribute as a primary key, or it is difficult to use a composite key.
- A column is created arbitrarily and assign each record a unique number/id
- Product Number, Product Id, Movie Id, Actor Id, etc.

30

Artificial Primary Key/Surrogate Key

	AdviserID	AdviserName	Department	Phone	Office
+	1	Johnson	Biology	236-8879	Sci-123
+	2	Wu	Chemistry	236-0091	Sci-260
+	3	Horan	Math	236-0098	AR-45
*	(AutoNumber)				

Record: ◄ ◄ 1 ► ►► ►* of 3

31

PK Selection Guidelines

- Do not use a field whose value is frequently changed as PK
- Look for single-attribute PK first
- If a PK contains more than 3 columns, consider a surrogate key
- Don't be limited to sample data; think beyond and consider possible scenarios and requirements

32

Relationship and Foreign Key (FK)

- Relationship is how tables (relations) are linked
- It is defined by the foreign key (FK) constraint
- A foreign key references a primary key (or any other unique keys) in another table
- This pair of keys are of the same kind (may be of different name)

33

Relationship and FK Example

34

Relationship in SQL Server Database

35