Database Revision

Dr Na Yao

Topics covered

- Week 1
  - Introduction
  - Relational model
  - Relational Algebra
  - Entity-relationship(ER) modelling

- Week 2
  - EER model
  - ER to Relational Model mapping
  - Database design
  - SQL

Topics covered

- Week 3
  - Normalization
  - Advanced Normalization
  - Transaction management
  - Distributed DBMS

- Week 4
  - XML
  - Data warehouse
  - Data mining
  - NoSQL

Database Introduction

Concepts

- Data
- Database
- Database management system (DBMS)
- Data model
- Schema vs data
- Database languages:
  - Data Definition Language (DDL)
  - Data Manipulation Language (DML)
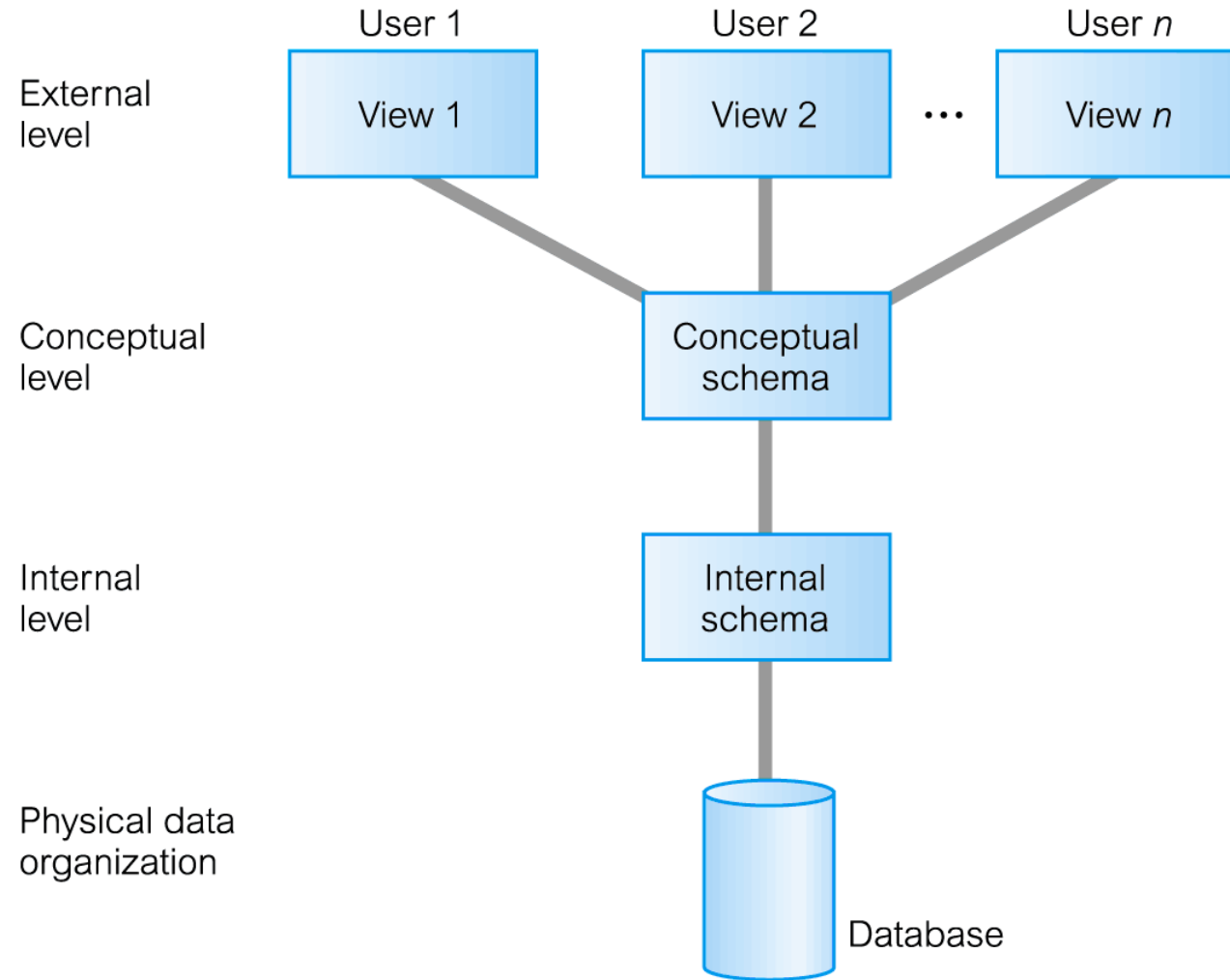
Components of DBMS Environment
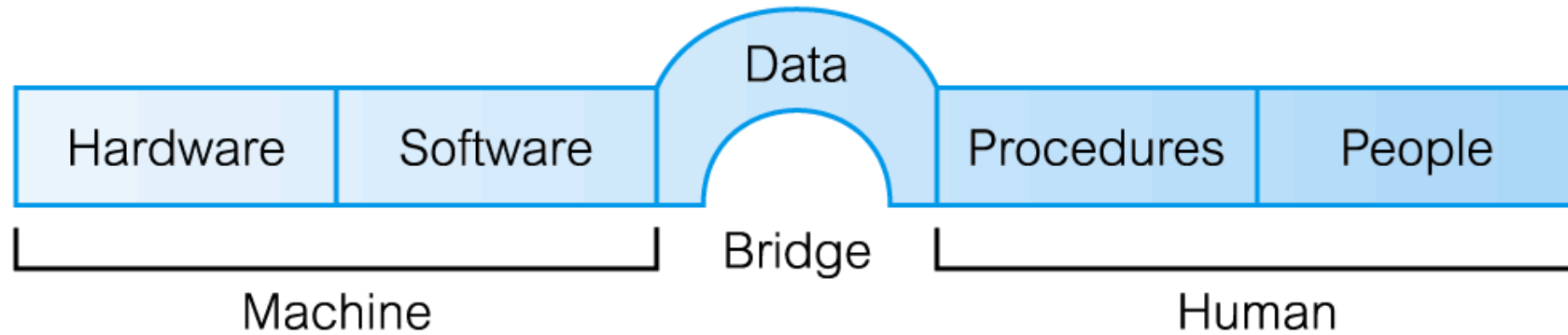
Roles in the Database Environment

- Data Administrator (DA)
- Database Administrator (DBA)
- Database Designers (Logical and Physical)
- Application Developers
- End Users (naive and sophisticated)

Data abstraction

The three-level
ANSI-SPARC
architecture

Relational Model

Concepts

- Relational model:
  - Relation (table)
  - Attribute (column)
  - Tuple (row)
  - Domain
- Keys:
  - Candidate key
  - Primary key
  - Foreign key

Integrity constraints

- Entity Integrity
  - In a base relation, no attribute of a primary key can be null.

- Referential Integrity
  - If foreign key exists in a relation, either foreign key value must match a candidate key value of some tuple in its home relation or foreign key value must be null.

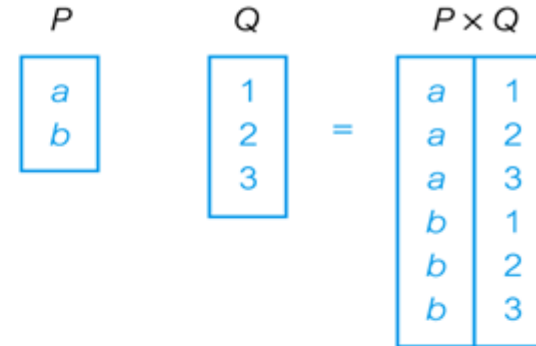- General Constraints

Relational Algebra
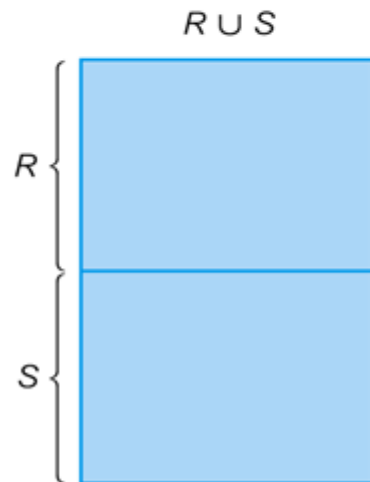
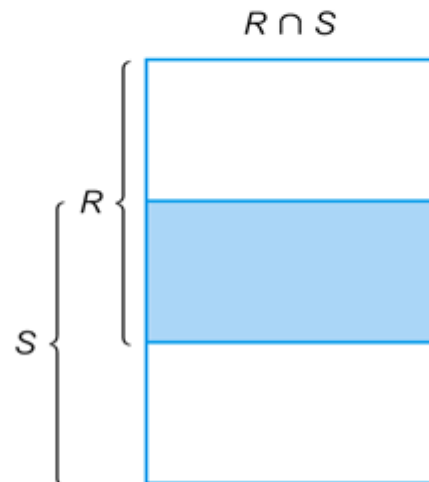# Relational Algebra Operations
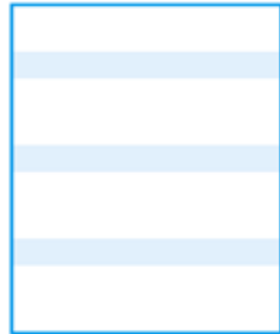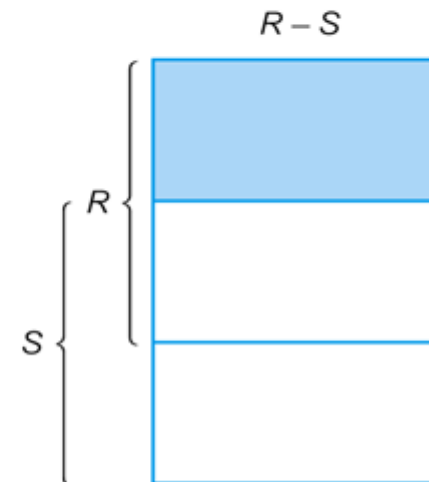

Selection

Projection

Cartesian product

Union

Intersection

Set difference

Relational Algebra Operations

Natural join

Divis on (shaded area)

Example of division

ER & EER

Entity Relationship (ER) Model

- Entity, Relationship, Attributes

- Multiplicity of relationship types:
  - one-to-one (1:1)
  - one-to-many (1:*)
  - many-to-many (*:*)
- Cardinality (maximum number) and participation (minimum number) of a relationship.
- Ternary relationship
- Making Assumptions

Enhanced Entity Relationship (EER) Model

- Specialization/Generalization
  - participation constraints (*mandatory* or *optional)*
  - disjoint constraints (*disjoint* (or) or *nondisjoint (and)*)

Logical Database Design (ER to Relational Model mapping)

Mapping ER model concepts to relations

- Entity to relations

- Binary 1-1, 1-N, N-M relationships

- Complex relationships

- Multi-valued attributes

SQL

- SELECT statement

  SELECT [DISTINCT | ALL]
     {* | [columnExpression [AS newName]] [,...] }
  FROM   TableName [alias] [, ...]
  [WHERE   condition]
  [GROUP BY   columnList]
  [HAVING   group condition]
  [ORDER BY   columnList]

- Aggregates

  COUNT, SUM, AVG, MIN, MAX

- Subqueries

Normalization

Normalization

- Purpose of normalization

- Data redundancy

- Update Anomalies
  - Insertion
  - Deletion
  - Modification

- Normalization is done via decomposition
  - Lossless-join
  - Dependency preservation
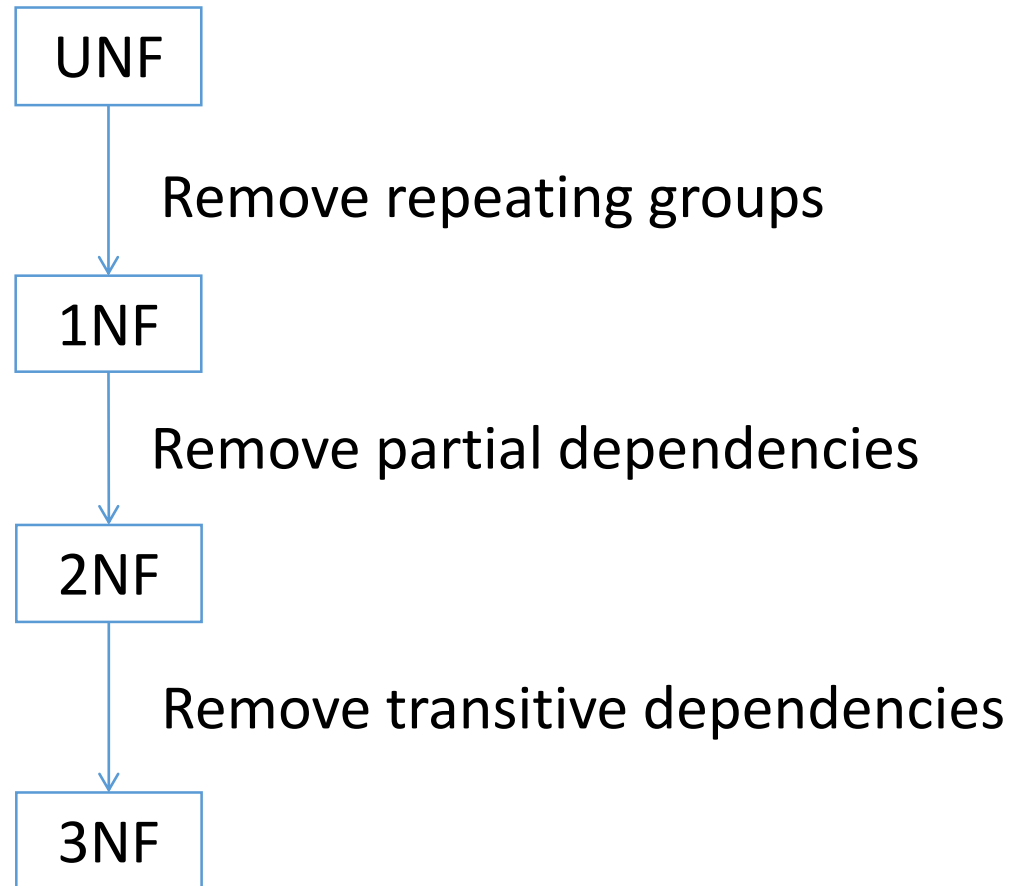
Normalization

- Functional dependencies:

    A and B are attributes of relation R, B is functionally dependent on A (denoted A $\rightarrow$ B), if each value of A in R is associated with exactly one value of B in R.

- Full functional dependencies

- Transitive functional dependencies

Normalization

- UNF, 1NF, 2NF, 3NF, BCNF, 4NF

Normalization

- BCNF: every determinant is a candidate key.

- 4NF: in Boyce-Codd Normal Form and contains no nontrivial multi-valued dependencies.

- Multi-valued Dependency (MVD):
  - for each value of A there is a set of values for B and a set of values for C. However, the set of values for B and C are independent of each other.
  - Trivial MVD: A MVD A –>> B in relation R is defined as being trivial if
    (a) B is a subset of A *or* (b) A $\cup$ B = R.
  - Nontrivial MVD: if neither (a) nor (b) are satisfied

Algorithm for decomposing relations into BCNF

- Relation R with FDs

- Compute keys for R

- Repeat until all relations are in BCNF:
  - Pick any R' with A->B that violates BCNF
  - Decompose R' into R1(A, B) and R2(A, rest)
  - Compute FDs for R1 and R2
  - Compute keys for R1 and R2

Transaction Management

Transactions

- Transaction: Action, or series of actions, carried out by user or application, which reads or updates contents of database.
- ACID properties of a transaction:

  Atomicity, Consistency, Isolation, Durability
- Concurrency control purpose
- Concurrency control problems:
  - Lost update problem.
  - Uncommitted dependency problem.
  - Inconsistent analysis problem.

Transaction management

- Serializability
  - Schedule
  - Serial schedule/nonserial schedule
  - Aim of serializability
- Two-Phase Locking (2PL)
- Deadlock
- Database recovery
  - log file
  - checkpointing

Distributed DBMSs

- Client/Server Architecture
- Distributed database
- Distributed DBMS
- Distributed processing
- Advantages/disadvantages of distributed DBMS

XML

- XML definition and basic concepts
- Relational model vs XML
- Well-formed XML, Valid XML
- DTD, XSD

Data warehouse & Data mining

- Data Warehouse
- Definition: subject-oriented, integrated, time-variant, and non-volatile
- OLTP
- OLAP
- Data mining
- Data warehouse & data mining

NoSQL

- NoSQL concept
- NoSQL application areas

Revision

- Read all lecture notes
- Make your own notes!
- Understand is the key!
- Check – have you met the learning outcomes at the beginning of each lecture notes?
- Practice on **writing**

Exam techniques

- Read the questions!
  - Explain, illustrate, describe, list, use examples…
- Clear handwriting
- Write something even if you don't know the answer…
- Time management
- Sleep, eat and drink before the exam