IS5102
Database Management Systems
Lecture 5: Relational Model

Alexander Konovalov
alexander.konovalov@st-andrews.ac.uk

(with thanks to Susmit Sarkar)

2021

Before

- Data models
- Entity–Relationship data models
- E-R diagrams and graphical notation

This week

- Lower level (logical) data models
- Relational data models and translations
- Formal query analysis

- A relational database consists of
  - a collection of tables each with a unique name
    - similar to a spreadsheet document containing a number of worksheets
- Each row in a table represents a relationship among a set of values
  - similar to mathematical notion of a **relation**
    - from which the model gets its name
- Each column in a table represents a distinct kind of value

- Table with a number of rows and columns

- Terms
  - **Relation** is the same as table
  - **Tuple** is the same as row or record
  - **Attribute** is the same as column or field

Restrictions

- No two rows with same values in all column positions
- All the values within a column in a relation have the same type
  - simple types only
  - no complex values such as sets or other rows

- Attributes (fields or columns)
  - a field is used to store an individual item of data
- Key Attributes
  - **Candidate key** – candidate key is an attribute or combination of attributes that uniquely identifies rows in a table
  - **Primary key** – there can be more than one candidate from which the primary key is chosen
  - **Foreign Key** – an attribute or combination of attributes that match attribute(s) in another table.
- Record (or tuple)
  - a record is a group of related attributes
  - identifiable by its primary key (or any candidate key)

Data Types

- Data Types:
  - the fields are set to accept a particular data type
  - helps check for the wrong type of data being entered
  - stores data as efficiently as possible
  - sorts data correctly

- Examples:
  - Integer number
  - Floating point number
  - String of given maximum length
  - Date and time
  - ...

- Proposed by E. F. Codd in 1970
- A means of storing information in tables called relations
  - Each table has multiple columns each with a unique name
- Use of data structures to access information quickly
  - E.g. indexes can help perform operations quickly
- High level means of expressing queries
  - Powerful means of expressing queries using relational algebra
  - Expressions can be optimised for faster evaluation

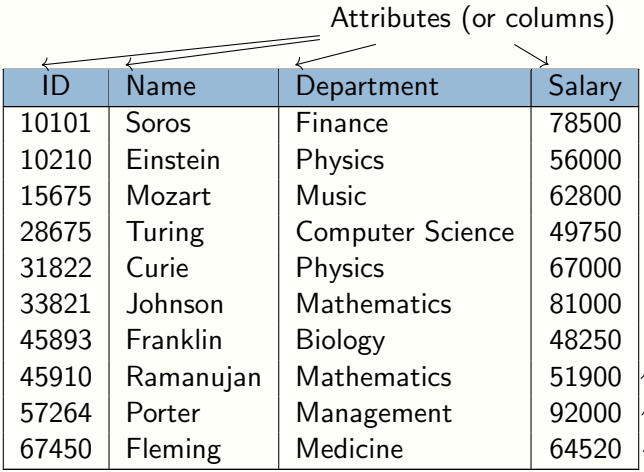
The Relational Model is based on the mathematical concept of a relation

- **Relation** – table
  - represented by a table with columns and rows
- **Tuple** – row
  - order of tuples (rows) is not important
- **Attributes** – named column headers
  - All values in a column have the same type
  - Attributes can have only simple types

More Terminology

- **Domain** of an attribute
  - Set of permitted values for that attribute
- **Degree** of a relation
  - Number of attributes it contains
- **Cardinality** of a relation
  - Number of tuples it contains
  - Changes as tuples are added or deleted

Example of a Relation (instructor)



ID	Name	Department	Salary
10101	Soros	Finance	78500
10210	Einstein	Physics	56000
15675	Mozart	Music	62800
28675	Turing	Computer Science	49750
31822	Curie	Physics	67000
33821	Johnson	Mathematics	81000
45893	Franklin	Biology	48250
45910	Ramanujan	Mathematics	51900
57264	Porter	Management	92000
67450	Fleming	Medicine	64520

Attribute Types

- The set of allowed values for each attribute is called the domain of the attribute
- Attribute values are (normally) required to be atomic; that is, indivisible
- The special value **null** is a member of every domain
- The null value causes complications in the definition of many operations

Attribute	Meaning	Domain Definition
ID	The set of all possible instructor IDs	character: size 5, range 00000 – 99999
name	The set of all possible names	character : size 50
dept_name	The set of all possible department names within the university	character: size 30
salary	Possible values of the instructor salaries	currency: 6 digits, range 10000 - 150000

- A_1, A_2, ..., A_n are attributes
- $R = (A_1, A_2, \ldots, A_n)$ is a relation schema

  Example:

  `instructor = (ID, name, dept_name, salary)`

- Formally, given sets D_1, D_2, ..., D_n
  a relation r is a subset of $D_1 \times D_2 \times \ldots \times D_n$

  Thus, a relation is a set of n-tuples $(a_1, a_2, \ldots, a_n)$
  where each $a_i \in D_i$

- The current values (relation instance) of a relation are specified by a table
- An element t of r is a **tuple**, represented by a row in a table

Order of tuples is irrelevant (tuples may be stored in an arbitrary order)

Example: instructor relation with unordered tuples

ID	Name	Department	Salary
10101	Soros	Finance	78500
31822	Curie	Physics	67000
45910	Ramanujan	Mathematics	51900
15675	Mozart	Music	62800
33821	Johnson	Mathematics	81000
67450	Fleming	Medicine	64520
10210	Einstein	Physics	56000
45893	Franklin	Biology	48250
57264	Porter	Management	92000
28675	Turing	Computer Science	49750

- A database often consists of multiple relations
- Information about an enterprise is broken up into parts:
  - instructor
  - student
  - advisor
- Bad design:

```
univ (instructor_id, name, dept_name, salary, student_id,...)
```

  Results in
  - repetition of information (e.g., two students have the same instructor)
  - the need for null values (e.g., represent an student with no advisor)
- Normalization theory (later) deals with how to design "good" relational schemas

Reading

- Chapter 7, Database Design
- Chapter 2, Database System Concepts
- Chapter 4 & 5.1, Database Systems