The relational data model

1 The relational data model

- DDL is too low level and not easily understandable by most users
- We need a data model: A collection of intuitive concepts describing data, their relationships and constraints
- Relational data model
  - Relations between data are stored in tables
  - Based on the concept of mathematical relations
  - The most widely used data model (for structured data)
- Intuitively in our data, every entity combines various attributes together
- The schema of a relation - the description of a particular collection of data in the model
- Let $A_1, A_2, ..., A_n$ be a set of attributes that can be related
- The $R(A_1, A_2, ..., A_n)$ is the scheme of the relation R
- In a relation schema the ordering of the attributes does not matter
- A database has many entities, each with its own attributes
- This information is decomposed into smaller pieces where every relation stores only one piece of the information
- However there is data duplication where two customers have the same account, and null values are needed where data is not complete

2 Relation Model Terminology

- **Relation** - A table
- **Attribute** - A named column of a relation. Every attribute has a unique name
- **Domain** - The set of allowable values of an attribute
- **Tuple** - A row of a relation - every tuple has a concrete value for every attribute (not left empty, use NULL if no data)
- **Cell** - The intersection of a row and a column
- **Degree** - The number of attributes (every row stores as many values as the degree of the relation)
- **Cardinality** - The number of tuples
- **Normalized** - Appropriately structured (every cell has exactly one value, no repetitions of two identical rows)
- **Relational Database** - A collection of normalised relations

3 Instances of branch and staff relations

- NULL Value
  - a special case of a cell entry
  - It represents an attribute value that is either currently unknown or not applicable
  - Not the same as 0
  - May or may not belong to the domain of the attribute

4 Properties of relations

- The relation name is distinct from all other relation names in the relational schema
- Each attribute within a relation has a distinct name
- Values of an attribute are all from the same domain
- Each cell of relation contains exactly one atomic value
- Each tuple is distinct among the tuples of the relation
- The ordering of the attributes has no significance
- The ordering of tuples has no significance

5 Keys

- How do we uniquely identify a tuple in a normalized table?
  - attribute names are unique within a table
  - but two tuples may share attribute values
- Every table must have some attributes, such as:
  - Their value uniquely determines a tuple of the table
  - These attribute are the primary key of the table
- **Candidate key:** a minimal set of attributes whose values uniquely identify the tuples
- **Primary key:** The candidate key selected to identify rows uniquely with the table
- **Alternate key:** Those candidate keys not selected as primary key
- **Simple key:** The key consists of only one attribute
- **Composite key:** The key consists of only one several attributes

6 Integrity constraints

- So far we have seen **domain constraints** for the attributes
- Entity integrity - every attribute of a **primary key** can **not** be **NULL**
- Purpose of entity integrity
  - guarantees that each entity has a unique identifier
  - ensures that foreign key values can reference primary key values

7 Integrity constraints

- Referential integrity
  - a foreign key either matches the primary key it refers to or it is null
- Purpose of referential integrity
  - any reference between tables is valid (or it has not been set yet)
  - Prevents deleting a row in a table B, if the primary key of B has a matching foreign key in another table A

8 Summary: Characteristics of a relational table

- A relation is represented by a two dimensional table
- Each row (tuple) signifies a entity occurence
- No two rows can be identical (each row of the table is unique)
- Each column represents an attribute and has a distinct name
- The intersection of a row and column has a single value (atomic)
- All values in a column must be of the same type
- One (or more) attributes uniquely identify each row (primary key)
- Two tables can be dependent (the primary key is the foreign key of another table)
- The ordering of rows and columns does not matter

9 Views

- So far all relations we have seen
  - Base relations
  - Its tuples are **physically stored** in the database
- A different type of relation: a view
  - a virtual relation
  - it does not exist physically in the database
- The content of a view
  - is derived from one (or more) base relations
  - is computer upon request by a user, at the time of request
  - changes when the underlying base relations change
- Main use
  - show customised information to every user
  - computer dynamic quantities

10 Alternatives to the relational data model

- Network data model
  - records appear as nodes
  - relationships appear as edges
- Hierarchical data model
  - Special case of the network data model, where the graph is a tree graph
  - its structure mirrors parent child relationship
  - limitations of the model e.g.
    - deleting a parent
    - adding a record without a parent