Information Systems 2

1. Modelling Information Systems I: Databases

Lars Schmidt-Thieme

Information Systems and Machine Learning Lab (ISMLL)
Institute for Business Economics and Information Systems
& Institute for Computer Science
University of Hildesheim
http://www.isml.uni-hildesheim.de

1. What is a database?

2. Entity Relationship Models

3. The Relational Model

4. Basics of SQL
Why Databases?

Benefits of databases:
- Data can be shared.
- Redundancy can be reduced.
- Inconsistency can be avoided (to some extent).
- Transaction support can be provided.
- Integrity can be maintained.

[Dat04]

Schema and State

- **Schema** (intension):
  describes the structure of a database with so-called *schema constructs*.
  The schema often is represented graphically.

- **State** (extension, contents):
  describes the contents of a database at a given point in time (*snapshot*).
  The state can be described by a set of instances of schema constructs.
Example:
A small company keeps a list of its customers in an OO Calc table. It contains a column each for name, address, phone number and email address. By now, there are just two rows:

<table>
<thead>
<tr>
<th>name</th>
<th>address</th>
<th>phone</th>
<th>email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anna Müller</td>
<td>Schuhstraße 3, 31139 Hildesheim</td>
<td>05121 / 123456</td>
<td><a href="mailto:mueller@example.com">mueller@example.com</a></td>
</tr>
<tr>
<td>Bert Meier</td>
<td>Hauptstraße 11, 30300 Hannover</td>
<td>050 / 12480</td>
<td><a href="mailto:meier@beispiel.de">meier@beispiel.de</a></td>
</tr>
</tbody>
</table>

Here,
- “name”, “address” etc. are schema constructs.
- “Anna Müller”, “Schuhstraße 3” etc. describe the state.

Data Model

A data model provides a formal description of how data may be structured and accessed.
It covers:
- **data structures**: that may be used to define the schema of a database.
- **integrity rules**: placed on the data structures to enforce integrity constraints.
- **data manipulation operators**: which allow to query and change the data.
Information Systems 2 / 1. What is a database?

Levels of Data Models

Usually one distinguishes 3 different levels of data models:

**Conceptual models**
(also logical models; high-level models): describe data in terms close to the concepts of users, e.g.:
- **Object model**: describes data as objects that are instances of classes, which have properties and methods; classes are organized in an inheritance hierarchy.
- **Entity Relationship model** (P. Chen 1976): describes data as entities with attributes and relations.

**Representation models**
(also Implementation Models): describe data in terms that are close to implementations, e.g.:
- **Relational data model** (Edgar Codd, 1969): describes data as tables (relations).
- **Network data model** (Charles Bachman, 1969): describes data as a network of records (example: LDAP).
- **Hierarchical data model** (mainframe era): describes data as a tree of records.

**Physical models**
(also low-level models; storage models; internal models): describe storage of data in detail.

Sometimes the logical level is split in
- **external models** (also user logical models): describes data from the perspective of different users.
- **conceptual models** (also community logical models): describes data from the perspective of the community of all users.

Accordingly, one distinguishes three different levels of schemata:

- **Internal schema**: describes physical structures in which the data is stored.

- **Conceptual schema**: describes the structure of the whole database for users.

- **External schema** (also user view): describes the structure of a a part of the database for a specific user or user group.

Data is stored accordingly to the internal schema.

There are mappings between internal schema and higher schemata.
1. What is a database?

2. Entity Relationship Models

3. The Relational Model

4. Basics of SQL
Properties of Attributes

simple vs. complex:
simple attributes cannot be decomposed into parts, complex attributes are composed of other attributes.

single-valued vs. multi-valued:
for single-valued attributes each entity has at most one value, for multi-valued attributes an entity may have several values.

stored vs. derived:
stored attributes are stored explicitly in the database, derived attributes can be computed from other information.

value domain:
the values of an attribute come from a fixed set, e.g., integers, real numbers, strings of a maximal length etc.

null value:
the special value null marks missing values or attributes that do not apply for a given entity.

Entities

Things of the same type, i.e., things that can be described by the same attributes and relations, are modeled as instances of an entity.

key attributes: an attribute whichs value occurs among all instances of an entity at most once, i.e., allows to identify an instance, is called a key.

Most entities have exactly one key. But they may also have none or several keys.

regular vs. weak:
an entity with at least one key is called regular. An entity without key is called weak.
Relations

Relations are used to model relationships between entities. The entities involved are called **participating entities**.

**Arity**: the number of participating entities of a relation is called its arity.

**Role names**: the positions at which entities can enter a relation are called roles. E.g., Auftragserteiler, Auftrag, Auftragsbearbeiter.

Role names are especially important if the same entity can participate in several roles in a relationship.

**Cardinality constraint**: restricts how often an instance of an entity may participate in a relation.

- **1:1**: an instance is allowed to participate at most with one other instance.
- **1:n**: an instance of role 1 may participate at most with n other instances in role 2, but an instance of role 2 may participate at most with one other instances in role 1.
- **n:m**: an instance of role 1 may participate at most with n other instances in role 2, an instance of role 2 may participate at most with m other instances in role 1.
**Total Participation:**
An entity participates totally in a relation, if each instance must be related to some other instances by that relation. (lower bound cardinality restriction $\geq 1$).

**Partial Participation:**
An entity participates partially in a relation, otherwise.

---

**Attributes:**
Relations may have attributes, too.

Each instance of the relation, i.e., each tuple of instances of entities between which the relation holds, has a value for each of the attributes of the relation.
Weak entity: an entity without a key attribute.

Instances of weak entities are identified indirectly by means of a relation. An instance of a weak entity is identified as the instance that

- is related to a given other entity (identifying entity; owner entity)
- with respect to a given relation (identifying relation) and
- has a given value for a given attribute (partial key).

Weak entities always participate completely in their identifying relation.

Weak entities can be converted to regular entities by introducing a key.
1. What is a database?

2. Entity Relationship Models

3. The Relational Model

4. Basics of SQL
Basic Concepts

The Relational model organizes data in tables.

<table>
<thead>
<tr>
<th>Relational model</th>
<th>common sense table</th>
</tr>
</thead>
<tbody>
<tr>
<td>attribute</td>
<td>column</td>
</tr>
<tr>
<td>attribute domain</td>
<td>value domain of a column</td>
</tr>
<tr>
<td>tuple</td>
<td>row</td>
</tr>
<tr>
<td>relation</td>
<td>table</td>
</tr>
<tr>
<td>null value</td>
<td>cells without entry (missing values, unappropriate attributes)</td>
</tr>
<tr>
<td>key</td>
<td>set of columns which values uniquely identify a row</td>
</tr>
<tr>
<td>primary key</td>
<td>key usually used for identifying rows</td>
</tr>
</tbody>
</table>

### KUNDE

<table>
<thead>
<tr>
<th>Kundennr</th>
<th>Name</th>
<th>Geburtstag</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Frank Müller</td>
<td>20.11.1980</td>
</tr>
<tr>
<td>2</td>
<td>Fred Schmidt</td>
<td>6.6.1972</td>
</tr>
<tr>
<td>3</td>
<td>Heribert Mayer</td>
<td>11.1.1954</td>
</tr>
<tr>
<td>4</td>
<td>Frank Müller</td>
<td>3.7.1978</td>
</tr>
</tbody>
</table>

### BESTELLUNG

<table>
<thead>
<tr>
<th>Rechnungsnr</th>
<th>Rechnungsdatum</th>
<th>Rechnungsbetrag</th>
<th>Kundennr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1099</td>
<td>12.2.2000</td>
<td>2099,-</td>
<td>2</td>
</tr>
<tr>
<td>1100</td>
<td>12.2.2000</td>
<td>589,-</td>
<td>1</td>
</tr>
<tr>
<td>1101</td>
<td>13.2.2000</td>
<td>4490,-</td>
<td>3</td>
</tr>
<tr>
<td>1102</td>
<td>15.2.2000</td>
<td>3349,-</td>
<td>2</td>
</tr>
<tr>
<td>1103</td>
<td>18.2.2000</td>
<td>10500,-</td>
<td>5</td>
</tr>
</tbody>
</table>

### Foreign Key:

An attribute (or set of attributes) that contains the key value of another relation.

The value domain of the foreign key must be the same as the value domain of the key of the referenced relation.

For each tuple of the referencing relation the value of the foreign keys must occur among the values of the key attribute of the referenced relation or be null (referential integrity).
Regular entity with relation (entity relation) with
- simple attributes → attribute
- complex attributes → one attribute for each component (neglect structure)
- key → primary key (select one)
- keys → secondary keys (all other)

Multi-valued attributes are mapped to an own relation that contains
- the foreign key of the entity relation and
- the value of the attribute.

Together they define the primary key of the new relation.

Binary relation 1:n → add a foreign key
- referencing the entity relationship on the 1-side
- to the entity relationship on the n-side.

All attributes of the relation are added to the entity relationship on the n-side.

variant:
Binary relation n:m \rightarrow \text{relation (relationship relation) with}
- foreign key to the entity relation on the n-side and
- foreign key to the entity relation on the m-side
- (jointly defining the primary key)

All attributes of the relation are mapped to attributes of the relationship relation.

Weak entity \rightarrow \text{relation (entity relation) with}
- foreign key to the identifying entity relation
- (jointly with the partial key defining the primary key)

All attributes are mapped as for regular entities.
### Mapping an ER Model to a Relational Model / Summary

<table>
<thead>
<tr>
<th>ER model</th>
<th>Relational model</th>
</tr>
</thead>
<tbody>
<tr>
<td>entity</td>
<td>relation (entity relation)</td>
</tr>
<tr>
<td>relation, binary, 1:1</td>
<td>add foreign key to one of the participating entity relations</td>
</tr>
<tr>
<td>relation, binary, 1:n</td>
<td>add foreign key to the participating entity relation on the n-side</td>
</tr>
<tr>
<td>relation, binary, n:m</td>
<td>relation (relationship relation) with 2 foreign keys</td>
</tr>
<tr>
<td>relation, n-ary</td>
<td>relation (relationship relation) with n foreign keys</td>
</tr>
<tr>
<td>attribute, simple</td>
<td>add attribute to relation</td>
</tr>
<tr>
<td>attribute, complex</td>
<td>add attributes to relation, one for each component</td>
</tr>
<tr>
<td>attribute, multi-valued</td>
<td>relation with foreign key</td>
</tr>
<tr>
<td>key</td>
<td>primary or secondary key</td>
</tr>
</tbody>
</table>

### Mapping an ER Model to a Relational Model / Example

```
KUNDE
  Kundennr | Name

BESTELLUNG
  Rechnungsnr | Rechnungsdatum | Rechnungsbetrag | Kundennr

PRODUKT
  Bezeichnung | Stückkosten

POSTEN
  Rechnungsnr | Produktbezeichnung | Stückzahl

AKTION
  Bezeichnung | Aktionskosten | Aktionsdatum

AKTIONPOSTEN
  Kundennr | Aktionsbezeichnung
```
1. What is a database?

2. Entity Relationship Models

3. The Relational Model

4. Basics of SQL

**History**

SQL (Standard Query Language) is the industrial standard for

- the definition of relation schemata
  (schema definition language, DDL) and

- the manipulation of the contents of a relational database
  (data manipulation language, DML) and

SQL is an ISO/ANSI standard:

- based on SEQUEL (Structured English Query Language) by Donald D. Chamberlin and Raymond F. Boyce (IBM) in the early 1970s

- 1986 standardized by ISO/ANSI (SQL/1, SQL-86)

- 1992 update to SQL/2, SQL-92; 1999 update to SQL/3, SQL:1999


SQL is supported by nearly all relational database management systems:

- Many simpler DBMS (such as mysql) do not implement parts of the standard.

- Most DBMS provide (mutually incompatible) non-standard extensions.
SQL defines a text format for database queries.

Every SQL query returns
- an error code or
- a result set.

Creating a Database

DBMS can manage several databases at the same time.

One can create a new database via:

```
create database ⟨database-name⟩;
use ⟨database-name⟩;
```

All following operations will be applied to this database.

One can destroy a database via:

```
drop database ⟨database-name⟩;
```
Creating a Table

One can create a table via

```
create table ⟨table-name⟩ ( ⟨table-spec⟩ ) ;
```

where ⟨table-spec⟩ is a comma-separated list of table specifications, i.e., of

- columns:
  
  ```
  ⟨column-name⟩ ⟨column-type⟩ [not null] [primary key]
  [references ⟨table-name⟩ [((⟨column-commaalist⟩))]]
  ```

- primary keys:
  
  ```
  primary key (⟨column-commaalist⟩)
  ```

- foreign keys:
  
  ```
  foreign key (⟨column-commaalist⟩)
  references ⟨table-name⟩ [((⟨column-commaalist⟩))]
  ```

One can destroy a table via:

```
drop table ⟨table-name⟩;
```
Creating a Table / Example

```
create table kunde (kundennr integer not null primary key, name varchar(30)) ;
create table bestellung (rechnungsnr integer not null primary key, rechnungsdatum date, rechnungsbetrag integer, kundennr integer references kunde);
create table produkt (bezeichnung varchar(30) not null primary key, stueckkosten integer);
create table posten (rechnungsnr integer not null references bestellung, produktbezeichnung varchar(30) not null references produkt (bezeichnung), stueckzahl integer, primary key (rechnungsnr, produktbezeichnung) ) ;
```

Modifying the Contents of a Table / Insert

Insert rows in a table:

```
insert into ⟨table-name⟩ [column-commalist] table-expr ;
where ⟨table-expr⟩ in the simplest case is as
values (⟨scalar-expr-commalist⟩)
```

Example:

```
insert into kunde values (1, "Frank Mueller");
insert into produkt values ("Spark II", 400);
insert into bestellung values (2001, "2000-06-31", 2000, 1);
insert into posten values (2001, "Spark II", 1);
```
Delete rows in a table:

\[
\text{delete from } \langle \text{table-name} \rangle \ [\text{where } \langle \text{cond-expr} \rangle];
\]

where \( \langle \text{cond-expr} \rangle \) in the simplest case is as

\[
\langle \text{column-name} \rangle \ (\text{comparision-operator}) \langle \text{value} \rangle
\]

or a combination of such expressions with the boolean operators “and,” “or” and “not”.

Example:

\[
\begin{align*}
\text{delete from kunde where kundennr = 2000;} \\
\text{delete from bestellung where rechnungsdatum < "2000-07-01";}
\end{align*}
\]

Update rows in a table:

\[
\text{update } \langle \text{table-name} \rangle \ \text{set } \langle \text{column-name} \rangle = \langle \text{scalar-expr} \rangle \\
[\text{where } \langle \text{cond-expr} \rangle];
\]

Example:

\[
\text{update produkt set stueckkosten = 1.2 * stueckkosten ;}
\]
SQL Queries

```sql
select ⟨select-item-commalist⟩
from ⟨table-ref-commalist⟩
[where ⟨cond-expr⟩]
[group by ⟨column-ref-commalist⟩]
[having ⟨cond-expr⟩]

where ⟨select-item⟩ is as
⟨scalar-expr⟩ [as ⟨column-name⟩] | [range-variable .] *

and ⟨table-ref⟩ is as
⟨table-name⟩ [AS ⟨range-variable⟩ [(column-commalist)]]
| ⟨table-expr⟩ [AS ⟨range-variable⟩ [(column-commalist)]]
| ⟨join-table-expr⟩
```

Examples:

```sql
select * from bestellung;
```

<table>
<thead>
<tr>
<th>rechnungsnr</th>
<th>rechnungsdatum</th>
<th>rechnungsbetrag</th>
<th>kundennr</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>2000-06-31</td>
<td>2000</td>
<td>1</td>
</tr>
<tr>
<td>2002</td>
<td>2000-07-01</td>
<td>6000</td>
<td>3</td>
</tr>
<tr>
<td>2003</td>
<td>2000-07-04</td>
<td>1600</td>
<td>1</td>
</tr>
</tbody>
</table>

```sql
select kundennr as knr, rechnungsbetrag from bestellung;
```

<table>
<thead>
<tr>
<th>knr</th>
<th>rechnungsbetrag</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2000</td>
</tr>
<tr>
<td>3</td>
<td>6000</td>
</tr>
<tr>
<td>1</td>
<td>1600</td>
</tr>
</tbody>
</table>
```
More complex queries combine several tables.

The **join operator** (represented as comma or by “join”) builds the cartesian product of two tables.

Usually, one is not interested in all combinations of the rows of two tables, but just the ones that are joined by a foreign key. This can be accomplished by:

- filtering by a “where” clause,
- a “left join” or “right join” operator with “on” clause or
- a “natural left join” or “natural right join” operator (join on all attributes with the same name).

```sql
select name, rechnungsbetrag from kunde, bestellung
   where kunde.kundennr = bestellung.kundennr;
```

```sql
select name, rechnungsbetrag from bestellung left join kunde
   on kunde.kundennr = bestellung.kundennr;
```

```sql
select name, rechnungsbetrag from bestellung natural left join kunde;
```

<table>
<thead>
<tr>
<th>name</th>
<th>rechnungsbetrag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frank Mueller</td>
<td>2000</td>
</tr>
<tr>
<td>Heribert Mayer</td>
<td>6000</td>
</tr>
<tr>
<td>Frank Mueller</td>
<td>1600</td>
</tr>
</tbody>
</table>
SQL Queries / Aggregation

One can aggregate the values of a column groupwise by

- defining groups of rows by a “group by" clause and
- use an aggregation function such as
  sum, count, max, min, avg
in the “select-expr”.

Example:

```sql
select name,sum(rechnungsbetrag)
from bestellung natural left join kunde
group by name;
```

<table>
<thead>
<tr>
<th>name</th>
<th>sum(rechnungsbetrag)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frank Mueller</td>
<td>3600</td>
</tr>
<tr>
<td>Heribert Mayer</td>
<td>6000</td>
</tr>
</tbody>
</table>

With the “having" clause one can filter those aggregated rows that meet some specified criteria.

Example:

```sql
select name,sum(rechnungsbetrag)
from bestellung natural left join kunde
group by name
having count(*) > 1;
```

<table>
<thead>
<tr>
<th>name</th>
<th>sum(rechnungsbetrag)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frank Mueller</td>
<td>3600</td>
</tr>
</tbody>
</table>
With the “order by” clause one can sort the rows of the result set.

One has to provide

- the name of the columns to sort by and
- “asc” for ascending or “desc” for descending sorting.

Example:

```
select kundennr, rechnungsbetrag
from bestellung
order by rechnungsbetrag desc;
```

<table>
<thead>
<tr>
<th>kundennr</th>
<th>rechnungsbetrag</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>6000</td>
</tr>
<tr>
<td>1</td>
<td>2000</td>
</tr>
<tr>
<td>1</td>
<td>1600</td>
</tr>
</tbody>
</table>

One can use “select” expressions in “where” clauses.

Example:

```
select name
from posten natural left join bestellung natural left join kunde
where produktbezeichnung in
    (select produktbezeichnung
     from posten natural left join bestellung natural left join kunde
     where name="Frank Mueller");
```

<table>
<thead>
<tr>
<th>name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heribert Mayer</td>
</tr>
</tbody>
</table>
Further SQL Concepts

- transactions
- views
- access rights
- trigger
- cursor
- stored procedures

Summary

- Databases can be modeled on different levels, especially the
  - conceptual level (e.g., entity relationship models) and
  - representation level (e.g., relational models).
- There is a map mapping ER models to relational models.
- SQL is a language for manipulating and querying data in relational databases.
- We will see three further modelling techniques in this course:
  - tagged data (XML; chapter 2).
  - ontologies (Knowledge Management; chapter 5).
  - data warehouses (Business Intelligence; chapter 6).
- You will learn all the details about databases in the course
  Databases (Computer Science 3) next term.
References