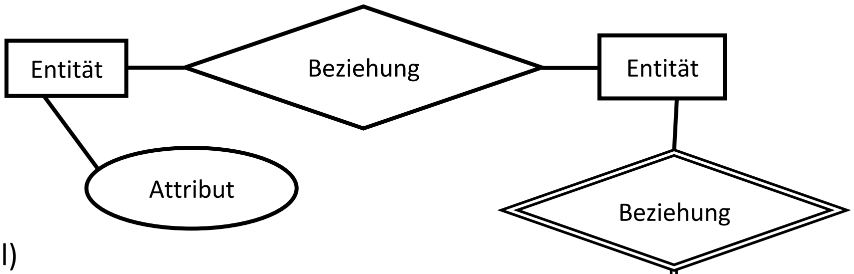
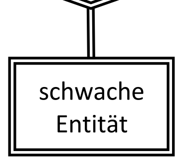
Zusammenfassung Datenbanksysteme

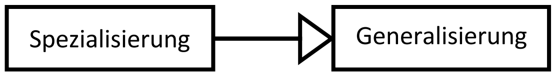
ER-Modell



Bestandteile:

- **Entitäten**
- **Beziehungen**
- **Attribute** (Schlüssel)
- **Kardinalitäten** (Funktionalität vs. min-max-Notation)
- **schwache Entitäten**
- **Generalisierung**





Relationenschema

Relation

Eine **Relation** R ist eine Teilmenge des kartesischen Produkts der Domänen:

$$R \subseteq D_1 \times \ldots \times D_n$$

⇒ Eine Datenbank ist eine Menge von Relationen.

Tupel

Ein **Tupel** t ist ein Element einer Relation R: $t \in R$

Schema

Das **Schema** einer Relation besteht aus dem **Namen** der Relation und der Liste der **Attribute**:

$$Name: (A_1: D_1, \ldots, A_n: D_n)$$

Schlüssel

Ein **Schlüssel** ist eine **minimale Menge** von Attributen, deren Werte ein Tupel **eindeutig** identifizieren. Ein ***Fremdschlüssel*** ist eine Menge von Attributen, welche auf den Schlüssel einer anderen Relation verweist. Schlüssel werden im Relationenschema unterstrichen, *Fremdschlüssel* zusätzlich kursiv geschrieben.

Anomalien

- **Problem-Schlüssel**
- **Update-Anomalie**
- **Einfüge-Anomalie**
- **Lösch-Anomalie**

Relationale Algebra

Basisoperatoren

- σ: **Selektion**
- π: **Projektion**
- $\cup$: **Vereinigung**
- $\cap$: **Durchschnitt**
- $-$: **Mengendifferenz**
- $\times$: **kartesisches Produkt**
- $\div$: **Division**
- ρ: **Umbenennung**

Selektion: $\sigma_F(R)$

- Schema: $att(\sigma_F(R)) = att(R)$

- Ausprägung: $\text{sigma}_{\text{F}}(\text{R}) = \{t \in R \mid t \text{ erfüllt } F\}$

Projektion: $\pi_A(R)$

- Vorbedingung: $A \subseteq \text{att}(R)$
- Schema: $att(\pi_A(R)) = A$
- Ausprägung: $\pi_A = \{t' \mid \exists t \in R: t.A = t'\}$
- $\Rightarrow$ Duplikate werden eliminiert.

Vereinigung: $R \cup S$

- Vorbedingung: $att(R) = att(S)$
- Schema: $att(R \cup S) = att(R) = att(S)$
- Ausprägung: $R \cup S = \{t \mid t \in R \vee t \in S\}$

Mengendifferenz: $R - S$

- Vorbedingung: $att(R) = att(S)$
- Schema: $att(R - S) = att(R) = att(S)$
- Ausprägung: $R - S = \{t \mid t \in R \wedge t \notin S\}$

kartesisches Produkt: $R \times S$

- Schema: $att(R \times S) = att(R) \cup att(S)$
- Ausprägung: $\text{R} \times S = \{t \mid \exists t_R \in R: t.att(R) = t_R \wedge \exists t_S \in S: t.att(S) = T_S\}$
- Ergebnisgröße: $|R \times S| = |R| \cdot |S|$

Division: $R \div S$

- Vorbedingung: $S \subseteq R$
- $R \div S = T \Leftrightarrow R = T \times S$
- Schema: $att(R \div S) = att(S)$
- Ausprägung: $R \div S = \{t \mid \forall s \in S: \exists r \in R: r.att(S) = s \wedge r.(att(R \setminus att(S)) = t\}$

Tupelkalkül

Anfragen **relationalen Tupelkalküls** sind von der Form:

$$\{t \mid P(t)\}$$

Die **Tupelvariable** t ist eine freie Variable der Formel $P(t)$ und durch keinen Quantor gebunden. Ein Tupel t ist im Ergebnis, wenn es die Formel $P(t)$ erfüllt.

Domänenkalkül

Anfragen **relationalen Domänenkalküls** sind von der Form:

$$\{[v_1, v_2, \dots, v_n] \mid P(v_1, v_2, \dots, v_n))\}$$

Die **Domänenvariablen** $v_1, v_2, \dots, v_n$ sind freie Variablen der Formel $P(v_1, v_2, \dots, v_n)$ und durch keinen Quantor gebunden. Ein Tupel $[v_1, v_2, \dots, v_n]$ ist im Ergebnis, wenn es die Formel $P(v_1, v_2, \dots, v_n)$ erfüllt.

funktionale Abhängigkeit

Sei $\mathcal{R}$ ein Relationenschema mit $\alpha \subseteq \text{R}$, $\beta \subseteq \text{R}$: Eine **funktionale Abhängigkeit** ist eine Beziehung $\alpha \rightarrow \beta$. Eine Relation R erfüllt eine funktionale Abhängigkeit $\alpha \rightarrow \beta$ genau dann, wenn gilt:

$$\forall r, t \in R: r.\alpha = t.\alpha \Rightarrow r.\beta = t.\beta$$

Attributhülle

Die Menge γ^+ der Attribute, welche von γ funktional abhängen nennt man die **Hülle** der Attributmenge γ:

$$\gamma^+ = \{A \in R \mid \gamma \rightarrow A\}$$

Hülle von funktionalen Abhängigkeiten
Die Menge F_2 von FDs ist aus der Menge F_1 ableitbar, wenn jede Relation, die alle FDs in F_1 erfüllt, auch alle FDs in F_2 erfüllt. Die Menge aller aus F ableitbaren FDs wird Hülle F^+ von F genannt:

$$F \vdash \{\gamma \rightarrow AttrHülle(F, \gamma)\} = F^+$$
$$F \vdash \{\alpha \rightarrow \beta\} \Leftrightarrow \beta \subseteq AttrHülle(F, \alpha)$$

Armstromg Axiome
- **Reflexivität**: $\beta \subseteq \alpha \Rightarrow (\alpha \rightarrow \beta \wedge \alpha \rightarrow \alpha)$
- **Verstärkung**: $\alpha \rightarrow \beta \Rightarrow (\alpha\gamma \rightarrow \beta\gamma)$
- **Transitivität**: $(\alpha \rightarrow \beta \wedge \beta \rightarrow \gamma) \Rightarrow (\alpha \rightarrow \gamma)$
- **Vereinigung**: $(\alpha \rightarrow \beta \wedge \alpha \rightarrow \gamma) \Rightarrow (\alpha \rightarrow \beta\gamma)$
- **Dekomposition**: $(\alpha \rightarrow \beta\gamma) \Rightarrow (\alpha \rightarrow \beta \wedge \alpha \rightarrow \gamma)$
- **Pseudotransivität**: $(\alpha \rightarrow \beta \wedge \gamma\beta \rightarrow \delta) \Rightarrow (\alpha\gamma \rightarrow \delta)$

Äquivalenz
Zwei Mengen F, G sind **äquivalent**, wenn sie **dieselbe Hülle** besitzen:

$$F \equiv G \Leftarrow F^+ = G^+$$
$$F \equiv G \Leftrightarrow (F \subseteq G^+ \wedge G \subseteq F^+)$$

kanonische Überdeckung
F_C heißt **kanonische Überdeckung** einer Menge F von FDs, wenn folgende Kriterien erfüllt sind:
- $F_C^+ = F^+$
- F_C enthält keine FDs mit überflüssigen Attributen
- Jede linke Seite einer FD in F_C ist einzigartig.

Berechnung:
1. **Zerlegung** durch **Dekomposition** auf der rechten Seite
2. **Kürzung** überflüssiger Attribute
   a. **Linksreduktion**:
      $B \in AttrHülle(F, \alpha - A) \Rightarrow$ ersetze $\alpha \rightarrow B$ durch $(\alpha - A) \rightarrow B$
   b. **Rechtsreduktion**:
      $B \in AttrHülle(F - (\alpha \rightarrow B), \alpha) \Rightarrow$ streiche $\alpha \rightarrow B$
3. **Zusammenfassung** mittels Vereinigungsregel

Schlüssel
$\gamma \subseteq \mathrm{R}$ ist ein Schlüssel, wenn:
- $\gamma \rightarrow \mathcal{R}$: $AttrHülle(F, \gamma) = \mathcal{R}$
- γ ist minimal: $\forall A \in \gamma: (\gamma - \{A\}) \nrightarrow \mathcal{R}$

Zerlegung
Eine Zerlegung eines Relationenschemas $\mathcal{R}$ ist eine Menge von Schemata $\mathcal{R}_1, \dots, \mathcal{R}_\mathrm{n}$, sodass gilt:

$$att(\mathcal{R}_1) \cup \dots \cup att(\mathcal{R}_\mathrm{n}) = \mathrm{att}(\mathcal{R})$$

Verlustlosigkeit
Eine Zerlegung $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_\mathrm{n}$ von $\mathcal{R}$ ist **verlustlos**, falls für jede Ausprägung R von $\mathcal{R}$, welche $F_\mathcal{R}$ erfüllt gilt:

$$R = \pi_{\mathcal{R}_1}(\mathrm{R}) \bowtie \pi_{\mathcal{R}_2}(\mathcal{R}) \bowtie \cdots \bowtie \pi_{\mathcal{R}_\mathrm{n}}(\mathrm{R})$$

Eine Zerlegung von $\mathcal{R}$ in $\mathcal{R}_1$ und $\mathcal{R}_2$ ist **verlustlos**, wenn die Joinattribute in einer der Teilrelationen (Super-)Schlüssel sind.

Abhängigkeitstreue

Sei F eine Menge an FDs über $\mathcal{R}$ und $\mathcal{R}' \subseteq \mathcal{R}$:

$$F[\mathcal{R}'] = \{\alpha \to \beta \in \mathrm{F} \mid \alpha \cup \beta \subseteq \mathcal{R}'\}$$

Eine Zerlegung $\mathcal{R}_1, \dots, \mathcal{R}_\mathrm{n}$ von $\mathcal{R}$ ist **abhängigkeitstreu**, wenn:

$$F \equiv (F^+[\mathcal{R}_1] \cup \dots \cup \mathrm{F}^+[\mathcal{R}_\mathrm{n}])^+$$

Normalformen

1. Normalform

Ein Relationenschema $\mathcal{R}$ ist in 1. Normalform, wenn die Domänen von $\mathcal{R}$ atomar sind.

2. Normalform

Ein Relationenschema $\mathcal{R}$ ist in 2. Normalform, wenn es in 1. Normalform ist und jedes Nichtschlüsselattribut von jedem Schlüssel funktional abhängig ist.

3. Normalform & Boyce-Codd-Normalform

1. $B \in \alpha$
2. α ist Superschlüssel von $\mathcal{R}$
3. $\exists \sigma$: σ ist Schlüssel $\wedge B \in \sigma$

Ein Relationenschema $\mathcal{R}$ mit F ist in dritter Normalform, wenn gilt:

$$\forall (\alpha \to \mathrm{B}) \in F, \alpha \subseteq \mathcal{R}, \mathrm{B} \in \mathcal{R}: 1 \vee 2 \vee 3$$

Ein Relationenschema $\mathcal{R}$ mit F ist in Boyce-Codd-Normalform, wenn gilt:

$$\forall (\alpha \to \mathrm{B}) \in F, \alpha \subseteq \mathcal{R}, \mathrm{B} \in \mathcal{R}: 1 \vee 2$$

Die Normalformen hängen wie folgt zusammen:

$$BCNF \subseteq 3NF \subseteq 2NF \subseteq 1NF$$

Synthesealgorithmus

Der Synthesealgorithmus erzeugt eine verlustlose, abhängigkeitstreue Zerlegung von $\mathcal{R}$ Teilschemata, die jeweils in dritter Normalform liegen:

1. Bestimmung der **kanonischen Überdeckung** F_C
2. für jede FD $\alpha \to \beta \in F_C$:
   a. Erstellung des Relationenschemas $\mathcal{R}_\mathrm{i} \coloneqq \alpha \cup \beta$
   b. Zuordnung von $F_i \coloneqq F_C[\mathcal{R}_\mathrm{i}]$ zu $\mathcal{R}_\alpha$
3. falls keines der in 2. erzeugten Teilschemata einen Schlüssel von $\mathcal{R}$ enthält: Definition von zusätzlichem Schema $\mathcal{R}_\kappa \coloneqq \kappa$ mit $F_\kappa \coloneqq \emptyset$ (beliebiger Schlüssel κ)
4. Elimination von Schemata, die bereits einem anderen Schema enthalten sind

Dekompositionsalgorithmus

Der Dekompositionsalgorithmus erzeugt eine verlustlose, aber nicht zwingend abhängigkeitstreue, Zerlegung von $\mathcal{R}$ in Teilschemata, die jeweils in Boyce-Codd-Normalform liegen:

1. $Z \coloneqq \{(\mathcal{R}, \mathrm{F})\}$
2. Existenz eines Teilschemas $(\mathcal{R}_\mathrm{i}, \mathrm{F}_\mathrm{i}) \in \mathrm{Z}$, dass die BCNF verletzt:
   a. beliebige Auswahl $\alpha \to \beta \in F_i$
   b. $\mathcal{R}_{\mathrm{i}_1} \coloneqq (\alpha \cup \beta)$; $F_{i_1} \coloneqq F_i^+[\mathcal{R}_{\mathrm{i}_1}]$
   c. $\mathcal{R}_{\mathrm{i}_2} \coloneqq (\alpha \cup \beta)$; $F_{i_2} \coloneqq F_i^+[\mathcal{R}_{\mathrm{i}_2}]$
   d. Entfernung von $(\mathcal{R}_\mathrm{i}, \mathrm{F}_\mathrm{i})$ und Ersetzung durch $(\mathcal{R}_{\mathrm{i}_1}, \mathrm{F}_{\mathrm{i}_1})$ und $(\mathcal{R}_{\mathrm{i}_2}, \mathrm{F}_{\mathrm{i}_2})$