

Theory I: Database Foundations

07.2012

20: Formal Design

- We want to distinguish good from bad database design.
- What kind of additional information do we need?
- Can we transform a bad into a good design?
- By which cost?

20.1 Motivation

Relations and anomalies

Stadt

<u>SNr</u>	SName	LCode	LFläche
7	Freiburg	D	357
9	Berlin	D	357
40	Moscow	RU	17075
43	St.Petersburg	RU	17075

Kontinent

<u>KName</u>	<u>LCode</u>	KFläche	Prozent
Europe	D	3234	100
Europe	RU	3234	20
Asia	RU	44400	80

Having removed anomalies

Stadt'

<u>SNr</u>	SName	LCode
7	Freiburg	D
9	Berlin	D
40	Moscow	RU
43	St.Petersburg	RU

Land'

<u>LCode</u>	LFläche
D	357
RU	17075

Lage'

<u>LCode</u>	<u>KName</u>	Prozent
D	Europe	100
RU	Europe	20
RU	Asia	80

Kontinent'

<u>KName</u>	KFläche
Europe	3234
Asia	44400

20.2 Functional Dependencies

20.2.1 Definition

- Let a relation schema be given by its format V and let $X, Y \subseteq V$.
- Let $r \in \text{Rel}(V)$. r fulfills a *functional dependency (FD)* $X \rightarrow Y$, if for all $\mu, \nu \in r$:

$$\mu[X] = \nu[X] \Rightarrow \mu[Y] = \nu[Y].$$

- Let \mathcal{F} a set of functional dependencies over V and $X, Y \subseteq V$. The set of all relations $r \in \text{Rel}(V)$, which fulfill all FD's in \mathcal{F} , is called $\text{Sat}(V, \mathcal{F})$.

20.2.2 Membership-Test

- The FD $X \rightarrow Y$, $\mathcal{F} \models X \rightarrow Y$ is implied by \mathcal{F} , if for each relation r , whenever $r \in \text{Sat}(V, \mathcal{F})$ then r fulfills $X \rightarrow Y$.
- The set $\mathcal{F}^+ = \{X \rightarrow Y \mid \mathcal{F} \models X \rightarrow Y\}$ is called *closure* of \mathcal{F} .
- $X \rightarrow Y \in \mathcal{F}^+$ is called *Membership-Test*.

key

Let $V = \{A_1, \dots, A_n\}$. $X \subseteq V$ is called *key* of V (bzgl. \mathcal{F}), if

- $X \rightarrow A_1 \dots A_n \in \mathcal{F}^+$,
- $Y \subset X \Rightarrow Y \rightarrow A_1 \dots A_n \notin \mathcal{F}^+$.

Armstrong-Axiome

Let $r \in \text{Sat}(V, \mathcal{F})$.

(A1) Reflexivity: If $Y \subseteq X \subseteq V$, then r fulfills FD $X \rightarrow Y$.

(A2) Augmentation: If $X \rightarrow Y \in \mathcal{F}, Z \subseteq V$, then r fulfills FD $XZ \rightarrow YZ$.

(A3) Transitivity: If $X \rightarrow Y, Y \rightarrow Z \in \mathcal{F}$, then r fulfills FD $X \rightarrow Z$.

(A1): *trivial FD's*.

Soundness and Completeness

- Every FD derivable by the Armstrong axioms is an element of the closure (soundness).
- Every FD in \mathcal{F}^+ is derivable by the Armstrong axioms (completeness)
 - To show completeness: If $X \rightarrow Y$ not derivable by (A1)–(A3), then $X \rightarrow Y \notin \mathcal{F}^+$, i.e. $\exists r, r$ fulfills \mathcal{F} , however does not $X \rightarrow Y$.

Membership-Test Variante 1:

Starting from \mathcal{F} apply (A1)–(A3) until $X \rightarrow Y$ is derived, or \mathcal{F}^+ is derived and $X \rightarrow Y \notin \mathcal{F}^+$.

Complexity?

more axioms

Let $r \in \text{Sat}(V, \mathcal{F})$. Let $X, Y, Z, W \subseteq V$ und $A \in V$.

- (A4) Union: If $X \rightarrow Y, X \rightarrow Z \in \mathcal{F}$, r fulfills FD $X \rightarrow YZ$.
- (A5) Pseudotransitivity: If $X \rightarrow Y, WY \rightarrow Z \in \mathcal{F}$, r fulfills FD $XW \rightarrow Z$.
- (A6) Decomposition: If $X \rightarrow Y \in \mathcal{F}, Z \subseteq Y$, r fulfills FD $X \rightarrow Z$.
- (A7) Reflexivity: If $X \subseteq V$, r fulfills FD $X \rightarrow X$.
- (A8) Accumulation: If $X \rightarrow YZ, Z \rightarrow AW \in \mathcal{F}$, r fulfills $X \rightarrow YZA$.

Axiom systems $\{(A1), (A2), (A3)\}$ and $\{(A6), (A7), (A8)\}$ are equivalent.

Membership-Test Variante 2:

- (Attribut-)closure X^+ of X (w.r.t. \mathcal{F}):

$$X^+ = \{A \mid A \in V \text{ and } X \rightarrow A \text{ is derivable by (A1) - (A3)}\}.$$

- First compute X^+ by (A6) - (A8) and afterwards test whether $Y \subseteq X^+$.

XPlus-Algorithm

```

XPlus( $X, Y, \mathcal{F}$ ) boolean {
  result :=  $X$ ;
  WHILE (changes to result) DO
    FOR each  $X' \rightarrow Y' \in \mathcal{F}$  DO
      IF ( $X' \subseteq$  result) THEN result := result  $\cup$   $Y'$ ;
    end.
  IF ( $Y \subseteq$  result) RETURN true ELSE false;
}

```

Example XPlus-Algorithm

Let $V = \{A, B, C, D, E, F, G, H, I\}$ and
 $\mathcal{F} = \{AB \rightarrow E, BE \rightarrow I, E \rightarrow G, GI \rightarrow H\}$.

$AB \rightarrow GH \in \mathcal{F}^+$?

Axiom	Anwendung	result
(A7)	$AB \rightarrow AB$	$\{A, B\}$
...

Using XPlus-Algorithm we can, given V, \mathcal{F} , compute a key.

How?

20.2.3 Minimal Cover

Equivalence

- Let \mathcal{F}, \mathcal{G} sets of FD's.
- \mathcal{F}, \mathcal{G} are called *equivalent*, $\mathcal{F} \equiv \mathcal{G}$, if $\mathcal{F}^+ = \mathcal{G}^+$.

Left and right reduction

- A set \mathcal{F} of FD's is called *left-reduced*, if the following condition is fulfilled.

If $X \rightarrow Y \in \mathcal{F}$, $Z \subset X$, then $\mathcal{F}' = (\mathcal{F} \setminus \{X \rightarrow Y\}) \cup \{Z \rightarrow Y\}$ not equivalent \mathcal{F} .

left-reduction: replace $X \rightarrow Y$ in \mathcal{F} by $Z \rightarrow Y$.

- It is called *right-reduced*, if $X \rightarrow Y \in \mathcal{F}$, $Z \subset Y$, then

$\mathcal{F}' = (\mathcal{F} \setminus \{X \rightarrow Y\}) \cup \{X \rightarrow Z\}$ not equivalent \mathcal{F} .

right-reduction: replace $X \rightarrow Y$ in \mathcal{F} by $X \rightarrow Z$.

looking for possible reductions

- Let $X \rightarrow Y$ be a FD in \mathcal{F} and let $Z \rightarrow Y$, where $Z \subset X$.
We perform a left-reduction, if $XPlus(Z, Y, \mathcal{F})$ is true.
- Let $X \rightarrow Y$ a FD in \mathcal{F} and let $X \rightarrow Z$, where $Z \subset Y$.
We perform a right-reduction, if $XPlus(X, Y, \mathcal{F}')$ is true.

Theorem

Let \mathcal{F} be a set of FD's and \mathcal{F}' be derived from \mathcal{F} by left-, resp. right-reduction.
 $\mathcal{F} \equiv \mathcal{F}'$.

Example

- $\mathcal{F}_1 = \{A \rightarrow B, B \rightarrow A, B \rightarrow C, A \rightarrow C, C \rightarrow A\}$.

right-reduction?

- $\mathcal{F}_2 = \{AB \rightarrow C, A \rightarrow B, B \rightarrow A\}$.

left-reduction?

minimal cover

\mathcal{F}^{min} is a *minimal cover* of \mathcal{F} , if it is derived from \mathcal{F} by the following steps:

- Perform all possible left-reductions.
- Perform all possible right-reductions.
- Delete all trivial FD's of the form $X \rightarrow \emptyset$.
- Compute the union of all FD's $X \rightarrow Y_1, \dots, X \rightarrow Y_n$ to derive $X \rightarrow Y_1 \dots Y_n$.

- A Minimal cover can be computed in polynomial time.

How?

- \mathcal{F}^{min} is not unique, in general.

Why?

20.3 Decomposition

20.3.1 Lossless

Let $\rho = \{X_1, \dots, X_k\}$ a *decomposition* of V , \mathcal{F} a set of FD's.

- Let $r \in \text{Sat}(V, \mathcal{F})$ and let $r_i = \pi[X_i]r$, $1 \leq i \leq k$.

ρ is called *lossless*, if for any $r \in \text{Sat}(V, \mathcal{F})$ there holds:

$$r = \pi[X_1]r \bowtie \dots \bowtie \pi[X_k]r.$$

Example

- $V = \{A, B, C\}$ and $\mathcal{F} = \{A \rightarrow B, A \rightarrow C\}$.
- $r \in \text{Sat}(V, \mathcal{F})$:

$$r = \begin{array}{ccc} A & B & C \\ \hline a_1 & b_1 & c_1 \\ a_2 & b_1 & c_2 \end{array}$$

- $\rho_1 = \{AB, BC\}$ and $\rho_2 = \{AB, AC\}$.
- $r \quad \pi[AB]r \bowtie \pi[BC]r,$
- $r \quad \pi[AB]r \bowtie \pi[AC]r.$

Theorem

Let a format V and set \mathcal{F} of FD's. Let $\rho = (X_1, X_2)$ be a decomposition of V .
 ρ is lossless, iff

$$(X_1 \cap X_2) \rightarrow (X_1 \setminus X_2) \in \mathcal{F}^+, \text{ oder } (X_1 \cap X_2) \rightarrow (X_2 \setminus X_1) \in \mathcal{F}^+.$$

20.3.2 Dependency Preserving

Example

$V = \{A, B, C, D\}, \rho = \{AB, BC\}.$

- $\mathcal{F} = \{A \rightarrow B, B \rightarrow C, C \rightarrow A\}.$

Is ρ dependency preserving w.r.t. \mathcal{F} ?

- Consider $\mathcal{F}' = \{A \rightarrow B, B \rightarrow C, C \rightarrow B, B \rightarrow A\}.$

Is ρ dependency preserving w.r.t. \mathcal{F}' ?

Definition

- Let $R = (V, \mathcal{F})$ and $Z \subseteq V$.
- Define the *projection* of \mathcal{F} on Z

$$\pi[Z]\mathcal{F} = \{X \rightarrow Y \in \mathcal{F}^+ \mid XY \subseteq Z\}.$$

- A decomposition $\rho = \{X_1, \dots, X_k\}$ of V is called *dependency preserving* w.r.t. \mathcal{F} , if

$$\bigcup_{i=1}^k \pi[X_i]\mathcal{F} \equiv \mathcal{F}.$$

There exist lossless decompositions which are not dependency preserving!

- $R = (V, \mathcal{F})$, where $V = \{\text{Stadt, Adresse, PLZ}\}$,
- $\mathcal{F} = \{\text{Stadt Adresse} \rightarrow \text{PLZ}, \text{PLZ} \rightarrow \text{Stadt}\}$.
- $\rho = \{X_1, X_2\}$: $X_1 = \{\text{Adresse, PLZ}\}$ und $X_2 = \{\text{Stadt, PLZ}\}$.
- ρ is lossless, as $(X_1 \cap X_2) \rightarrow (X_2 \setminus X_1) \in \mathcal{F}$.
- ρ is not dependency preserving.

What are the keys!

20.4 Normalform

Let $R = (V, \mathcal{F})$. We are looking for a decomposition $\rho = (X_1, \dots, X_k)$ of R with the following properties:

- each $R_i = (X_i, \pi[X_i]\mathcal{F})$, $1 \leq i \leq k$ is in normalform,
- ρ is lossless and, if possible, dependency preserving.
- k minimal.

Terminology

- Let X key of R and $X \subseteq Y \subseteq V$, then Y *Superkey* of R .
- If $A \in X$ for any key X of R , then A *Keyattribute (KA)* of R ;
- if $A \notin X$ for every key X , then A *Non-Keyattribute (NKA)*.

3rd Normalform

Schema $R = (V, \mathcal{F})$ is in *3rd Normalform* (3NF), if for any NKA $A \in V$ there holds:

If $X \rightarrow A \in \mathcal{F}$, $A \notin X$, then X Superkey.

3NF?

Stadt			
<u>SNr</u>	SName	LCode	LFläche
7	Freiburg	D	357
9	Berlin	D	357
40	Moscow	RU	17075
43	St.Petersburg	RU	17075

Kontinent			
<u>KName</u>	<u>LCode</u>	KFläche	Prozent
Europe	D	3234	100
Europe	RU	3234	20
Asia	RU	44400	80

3NF?

Stadt'		
<u>SNr</u>	SName	LCode
7	Freiburg	D
9	Berlin	D
40	Moscow	RU
43	St.Petersburg	RU

Land'	
<u>LCode</u>	LFläche
D	357
RU	17075

Lage'		
<u>LCode</u>	<u>KName</u>	Prozent
D	Europe	100
RU	Europe	20
RU	Asia	80

Kontinent'	
<u>KName</u>	KFläche
Europe	3234
Asia	44400

Boyce-Codd-Normalform

Schema $R = (V, \mathcal{F})$ is in *Boyce-Codd-Normalform* (BCNF), if the following holds. If $X \rightarrow A \in \mathcal{F}$, $A \notin X$, then X superkey.

BCNF implies 3NF.

- Consider $R = (V, \mathcal{F})$, where $V = \{ \text{Stadt, Adresse, PLZ} \}$, and $\mathcal{F} = \{ \text{Stadt Adresse} \rightarrow \text{PLZ}, \text{PLZ} \rightarrow \text{Stadt} \}$.
- R is in 3NF, however not in BCNF.
- Let $\rho = \{ \text{Adresse PLZ}, \text{Stadt PLZ} \}$ a decomposition, then ρ is in BCNF, lossless and not dependency preserving.

20.5 Normalization Algorithm

BCNF-Analysis: lossless and not dependency-preserving

Let $R = (V, \mathcal{F})$ a schema.

- 1 Let $X \subset V$, $A \in V$ and $X \rightarrow A \in \mathcal{F}$ a FD, which violates BCNF. Let $V' = V \setminus \{A\}$.

Decompose R in

$$R_1 = (V', \pi[V']\mathcal{F}), \quad R_2 = (XA, \pi[XA]\mathcal{F}).$$

- 2 Test for BCNF w.r.t. R_1 and R_2 and proceed recursively.

3NF-Analysis: lossless and dependency-preserving

Let $R = (V, \mathcal{F})$ a schema and let $\rho = (X_1, \dots, X_k)$ a decomposition of V , such that the Schemata $R_1 = (X_1, \pi[X_1]\mathcal{F}), \dots, R_k = (X_k, \pi[X_k]\mathcal{F})$ in BCNF.

- 1 Let \mathcal{F}^{min} a minimal cover of \mathcal{F} .
- 2 Identify the set $\mathcal{F}' \subseteq \mathcal{F}^{min}$ of those FD's, which are not dependency preserving.
- 3 For any such FA, $X \rightarrow A$ extend ρ by XA , resp. schema $(XA, \pi[XA]\mathcal{F})$.