Databases

Lecture 5

Functional Dependencies, Normal Forms

- the notion of *transitive dependency* is required for the third normal form

Definition. An attribute Z is transitively dependent on an attribute X, if $\exists Y$ such that $X \to Y, Y \to Z, Y \to X$ does not hold and Z is not included in $X \cup Y$.

Definition. A relation is in the third normal form (3NF) if it is in the second normal form and no non-prime attribute is transitively dependent on any key in the relation.

- let R be a 2NF relation
- if K is the (sole) key and β an attribute that is transitively dependent on K, then there is an attribute α such that: $K \rightarrow \alpha$ (this dependency always holds) and $\alpha \rightarrow \beta$
- since the relation is in 2NF, β is fully functionally dependent on K, therefore $\alpha \not\subset K$

$\Rightarrow$ a relation that is in 2NF and not in 3NF has a dependency $\alpha \rightarrow \beta$, where α is a non-prime attribute

- decomposition (to eliminate this dependency) - similar to the 2NF decomposition

Example 9. The BSc examination results are stored in the relation:

BSC_EXAM [StudentName, Grade, Supervisor, Department]

- the relation stores the supervisor and the department in which she works
- since the relation contains data about students (i.e., one row per student), *StudentName* can be chosen as the key
- the following functional dependency holds: *{Supervisor}* → *{Department}* ⇒ the relation is not in 3NF
- to eliminate this dependency, the relation is decomposed into the following 2 relations:

RESULTS [StudentName, Grade, Supervisor]

SUPERVISORS [Supervisor, Department]

Example 10. The following relation stores addresses for a group of people:

ADDRESSES [CNP, LastName, FirstName, ZipCode, City, Street, No]

- key: {CNP}
- identified dependency: *{ZipCode}* $\rightarrow$ *{City}* (can you identify another dependency in this relation?)
- since this dependency holds, relation ADDRESSES is not in 3NF, therefore it must be decomposed:

ADDRESSES'[CNP, LastName, FirstName, ZipCode, Street, No]

ZIPCODES[ZipCode, City]

Example 11. The following relation stores the exam session schedule:

EX_SCHEDULE[Date, Hour, Faculty_member, Room, Group]

- the following restrictions are expressed via <u>key definitions</u> and <u>functional dependencies</u>:

    1. a group of students has at most one exam per day

    => *{Date, Group}* is a key

    2. on a certain date and time, a faculty member has an exam with a group

    => *{Faculty_member, Date, Hour}* is a key

    3. on a certain date and time, there is at most one exam in a room

    => *{Room, Date, Hour}* is a key

Example 11. The following relation stores the exam session schedule:

EX_SCHEDULE[Date, Hour, Faculty_member, Room, Group]

- the following restrictions are expressed via <u>key definitions</u> and <u>functional dependencies</u>:

    4. a faculty member doesn't change the room in a day

    => the following dependency holds: *{Faculty_member, Date}* $\rightarrow$ *{Room}*

- all the attributes appear in at least one key, i.e., there are no non-prime attributes
- given the normal forms' definitions specified thus far, the relation is in 3NF
- objective: eliminate the *{Faculty_member, Date}* → *{Room}* functional dependency

Definition. A relation is in the Boyce-Codd (BCNF) normal form if every determinant (for a functional dependency) is a candidate key, i.e., there are no functional dependencies $\alpha \rightarrow \beta$, where α is not a candidate key.

- to eliminate the functional dependency, the original relation must be decomposed into:

EX_SCHEDULE [Date, Faculty_member, Hour, Group],

ROOM_ALLOCATION [Faculty_member, Date, Room]

- these relations don't contain other functional dependencies, i.e., they are in BCNF
- however, the key associated with the 3rd constraint, *{Room, Date, Hour}*, does not exist anymore
- if this constraint is to be kept, it needs to be checked in a different manner (e.g., through the program)

Obs. The following simple properties for functional dependencies hold (they can be easily demonstrated; the first one has already been introduced):

1. If *K* is a key of $R[A_1, A_2, \dots, A_n]$, then $K \rightarrow \beta, \forall\beta \subset \{A_1, A_2, \dots, A_n\}$; such a dependency is always true, hence it will not be eliminated via decompositions.

2. If $\beta \subseteq \alpha$, then $\alpha \rightarrow \beta$ - the *trivial functional dependency* or *reflexivity*.

$$\Pi_\alpha(r_1) = \Pi_\alpha(r_2) \Rightarrow \Pi_\beta\ (r_1) = \Pi_\beta(r_2)$$

$$\beta \subseteq \alpha$$

$$\Rightarrow \alpha \rightarrow \beta$$

Obs. The following simple properties for functional dependencies hold (they can be easily demonstrated; the first one has already been introduced):

3. If $\alpha \to \beta$, then $\gamma \to \beta$, $\forall\gamma$ with $\alpha \subset \gamma$.

$$\Pi_\gamma(r_1) = \Pi_\gamma(r_2) \underset{\alpha \subset \gamma,\ \text{propr.2}}{\Rightarrow} \Pi_\alpha(r_1) = \Pi_\alpha(r_2) \underset{\alpha \to \beta}{\Rightarrow} \Pi_\beta(r_1) = \Pi_\beta(r_2)$$

$$\Rightarrow \gamma \to \beta$$

4. If $\alpha \to \beta$ and $\beta \to \gamma$, then $\alpha \to \gamma$ - *transitivity*.

$$\Pi_\alpha(r_1) = \Pi_\alpha(r_2) \underset{\alpha \to \beta}{\Rightarrow} \Pi_\beta(r_1) = \Pi_\beta(r_2) \underset{\beta \to \gamma.}{\Rightarrow} \Pi_\gamma(r_1) = \Pi_\gamma(r_2)$$

$$\Rightarrow \alpha \to \gamma$$

Obs. The following simple properties for functional dependencies hold (they can be easily demonstrated; the first one has already been introduced):

5. If $\alpha \to \beta$ and $\gamma \subset \{A_1, \dots, A_n\}$, then $\alpha\gamma \to \beta\gamma$, where $\alpha\gamma = \alpha \cup \gamma$.

$$\Pi_{\alpha\gamma}(r_1) = \Pi_{\alpha\gamma}(r_2) \Rightarrow \left\{\begin{array}{l} \Pi_{\alpha}(r_1) = \Pi_{\alpha}(r_2) \Rightarrow \Pi_{\beta}(r_1) = \Pi_{\beta}(r_2) \\ \Pi_{\gamma}(r_1) = \Pi_{\gamma}(r_2) \end{array}\right\} \Rightarrow \Pi_{\beta\gamma}(r_1) = \Pi_{\beta\gamma}(r_2)$$

Example 12: consider the relation

DFM [Department, FacultyMember, MeetingDate],

where *FacultyMember* and *MeetingDate* are repeating attributes

- a record in this relation could look like the ones in the following table:

Computer Science	FCS1 FCS2 … FCSm	DCS1 DCS2 … DCSn
Mathematics	FM1 FM2 … FMp	DM1 DM2 … DMq

- to avoid repeating attributes (such that the relation is at least in 1NF), the data must be stored like in the following table:

Computer Science	FCS1	DCS1
Computer Science	FCS1	DCS2
…	…	…
Computer Science	FCS1	DCSn
Computer Science	FCS2	DCS1
Computer Science	FCS2	DCS2
…	…	…
Mathematics	FM1	DM1
…	…	…
Mathematics	FMp	DMq

Computer Science	FCS1	DCS1
Computer Science	FCS1	DCS2
…	…	…
Computer Science	FCS1	DCSn
Computer Science	FCS2	DCS1
Computer Science	FCS2	DCS2
…	…	…
Mathematics	FM1	DM1
…	…	…
Mathematics	FMp	DMq

- in this table, every faculty member has the same meeting dates
- therefore, when adding / changing / removing rows, additional checks must be carried out

- a simple functional dependency $\alpha \rightarrow \beta$ means, by definition, that every value u of α is associated with a unique value v for β

Definition. multi-valued dependency $\alpha \rightrightarrows \beta$ (α multi-determines β)

- a value u of α is associated with a set of values v for β: $\beta(u) = \{v_1, v_2, \dots, v_n\}$, and this association holds regardless of the values of γ = A − α − β (where A is the set of attributes in the relation, i.e., A = $\alpha \cup \beta \cup \gamma$).

- obs. $\sigma_{\alpha=u}(R)$ determines a relation that contains the tuples of R where $\alpha = u$ (this operator will be detailed in a subsequent lecture)

- let R[A] be a relation, $\alpha \rightrightarrows \beta$ a multi-valued dependency, and A = $\alpha \cup \beta \cup \gamma$, with γ a non-empty set
- the association among the values in $\beta(u)$ for β and the value u of α holds regardless of the values of γ (the context)
- i.e., these associations (between u and an element in $\beta(u)$) exist for any value w in γ:
  - $\forall w \in \Pi_\gamma\left(\sigma_{\alpha=u}(R)\right)$, $\exists r_1, r_2, \ldots, r_n$ such that $\Pi_\alpha(r_i) = u$, $\Pi_\beta(r_i) = v_i$, $\Pi_\gamma(r_i) = w$

- if $\alpha \rightrightarrows \beta$ and the following rows exist:

α	β	γ
u_1	v_1	w_1
u_1	v_2	w_2

then the following rows must exist as well:

α	β	γ
u_1	v_1	w_2
u_1	v_2	w_1

Property. Let R[A] be a relation, A = $\alpha \cup \beta \cup \gamma$. If $\alpha \rightrightarrows \beta$, then $\alpha \rightrightarrows \gamma$.

Justification.

- Let u be a value of α in R.
- Let $\beta(u) = \Pi_\beta\big(\sigma_{\alpha=u}(R)\big)$, $\gamma(u) = \Pi_\gamma\big(\sigma_{\alpha=u}(R)\big)$ (the β and γ values in the tuples where $\alpha = u$).

Since $\alpha \rightrightarrows \beta \Rightarrow$

$\forall w \in \gamma(u), \forall v \in \beta(u), \exists r = (u, v, w)$, or

$\forall v \in \beta(u), \forall w \in \gamma(u), \exists r = (u, v, w)$,

therefore $\alpha \rightrightarrows \gamma$.

- for relation DFM (in the previous example):

{Department} $\rightrightarrows$ *{FacultyMember}*, *{Department}* $\rightrightarrows$ *{MeetingDate}*

Definition. A relation R is in 4NF if, for every multi-valued dependency $\alpha \rightrightarrows \beta$ that holds over R, one of the statements below is true:

- $\beta \subseteq \alpha$ or $\alpha \cup \beta = R$
- α is a superkey.

- if R[α, β, γ] and $\alpha \rightrightarrows \beta$, R is decomposed into the following relations:

$R_1[\alpha, \beta] = \Pi_{\alpha \cup \beta}(R)$

$R_2[\alpha, \gamma] = \Pi_{\alpha \cup \gamma}(R)$

- relation DFM is decomposed into:

  DF [Department, FacultyMember]

  DM [Department, MeetingDate]

References

- [Ta13] ȚÂMBULEA, L., Curs Baze de date, Facultatea de Matematică și Informatică, UBB, 2013-2014
- [Ra00] RAMAKRISHNAN, R., GEHRKE, J., Database Management Systems (2nd Edition), McGraw-Hill, 2000
- [Da03] DATE, C.J., An Introduction to Database Systems (8th Edition), Addison-Wesley, 2003
- [Ga08] GARCIA-MOLINA, H., ULLMAN, J., WIDOM, J., Database Systems: The Complete Book, Prentice Hall Press, 2008
- [Ha96] HANSEN, G., HANSEN, J., Database Management And Design (2nd Edition), Prentice Hall, 1996
- [Ra07] RAMAKRISHNAN, R., GEHRKE, J., Database Management Systems, McGraw-Hill, 2007, http://pages.cs.wisc.edu/~dbbook/openAccess/thirdEdition/slides/slides3ed.html
- [Ul11] ULLMAN, J., WIDOM, J., A First Course in Database Systems, http://infolab.stanford.edu/~ullman/fcdb.html
- [Ta03] ȚÂMBULEA, L., Baze de date, Litografiat Cluj-Napoca 2003