Tehtävä 5: Normaalimuodot

Tietokantojen teoriassa normalisointi (normalization) on prosessi, jonka tavoitteena on parantaa tietokannan rakennetta. Tähän liittyy erilaisia normaalimuotoja (normal forms), jotka ohjaavat taulujen suunnittelua.

Lue William Kentin artikkeli A Simple Guide to Five Normal Forms in Relational Database Theory (linkki) ja vastaa sen jälkeen seuraaviin tehtäviin. Voit mielellään etsiä tietoa myös muista lähteistä.

Tehtävä 1

Mitä hyötyä normaalimuodoista on? Miten ne liittyvät kurssimateriaalin luvun 6 periaatteisiin?

The normalisation forms are made with a purpose of preventing abnormal data updates and data inconsistencies. They perform the role of restructuring database, mostly relational database, in order to remove data redundancies and enhance data integrity.

One of the main responsibility of data normalisation is to ensure that data dependencies are working probably enforced by certain data constraints without conflicting with other columns and tables.
With respect to performance efficiency, it tends to assume that all non-key fields will be updated regularly. So it is likely to penalise retrieval, since data which may retrieve from one record in an unnormalised design or from several records in the normalized form. There is no requirement to normalise all records when desired implementation are taken into account."

Relation to chapter 6:
-A optimal goal in design is that the desired database is comfortable to use in SQL. The structure of the database should be such that we can easily retrieve and modify data with SQL commands.
- Useful Database design principles often lead to workable solutions. The result should always be that database is beneficial to serve the desired purpose and not that the principles are randomly generated without one's thinking and supervision.

Tehtävä 2

Kent määrittelee 1. normaalimuodon näin: "Under first normal form, all occurrences of a record type must contain the same number of fields."

Mitä tarkoittavat "record type" ja "fields"?

A Record type means a complex data system, which allows the programmer to manage data types within the desired column structure. Record type groups one or more columns to establish a new data type for management. A Record type can accept the data information.

A SQL field is a calculated field, which contains a mount of data type. Fields in database management system perform its calculation at the data source. They are passed through to the database and is calculated at the database.

Tehtävä 3

Kent mainitsee 2. normaalimuodosta: "It is only relevant when the key is composite, i.e., consists of several fields."

Anna esimerkki tilanteesta, jossa ehto "key is composite" ei pidä paikkaansa, ja perustele asia.

Composite key is a combination of two or more columns that is used to retrieve the unique information from the table . The combination of columns guarantees uniqueness, though individually uniqueness is not guaranteed. Hence, they are combined to uniquely identify records in a table. If a relation has a composite key, all non-key attributes must depend on all components of the key.

If a table has duplicates or null values and these values prevent the columns from identifying a row, those columns cannot be composite keys and return error. In other words, the combination of two columns is not a composite key when its data is not unique. For example, if columns ('student ID','student task','status') in a table have several rows containing the same values ('1','task1','pass'), which means that student#1 has succeeded to pass task#1 several times in the records, then those columns cannot be composite keys because they cannot identify a unique row in the database.

Tehtävä 4

Mitkä seuraavista ovat funktionaalisia riippuvuuksia?
elokuvan nimi → elokuvan julkaisuvuosi
elokuvan julkaisuvuosi → elokuvan nimi
opiskelijanumero → opiskelijan nimi
opiskelijan nimi → opiskelijanumero
käyttäjän sähköpostiosoite → käyttäjän nimi
käyttäjän nimi → käyttäjän sähköpostiosoite
Perustele joka kohdasta lyhyesti, miksi se on tai ei ole funktionaalinen riippuvuus.

Case 1-3-6 are functional dependencies because:
- Case 1: A certain movie name has only one unique release year and movie release year is a part of the movie name's detail information, which means release year functionally depends on move name

- Case 3: A student number could return a unique student name, but a student name may return many student numbers --> student name functional depends on student number

- Case 6: The user email address could return a unique username, and user email address is a part of a user name's detail information, which means that user email address functionally depends on user name.

Case 2-4-5 are not functional dependency because case 2-4-5 cannot return desired results when their data may return many data results (for example, case 2 returns many movie names in the same release year, case 4 returns many student numbers from the same student name, and case 5 returns many user names using the same user name) and hardly define the end results based on the available data resource.

Tehtävä 5

Kent mainitsee 4. normaalimuodosta: "The main problem with violating fourth normal form is that it leads to uncertainties in the maintenance policies."

Mitä tarkoittaa tässä "maintenance policy"?

The fourth normal form is used to minimise the update problems in the database, for example the situation that database manager has to update multiple records manually without ability to use sql commands because

the same data information may keep repeating in several records and is hard to control.Or the problem may happen during the insertion of a new cell value or deletion of a cell value in a table.

The maintenance policies could control the maintenance of two independent multi-valued facts in a table by using several policies as below:

- A disjoint format, which leads to change the format to more blank fields of record type
- A random mix which is used to minimise the number of records, with repetitions or with null values; or proceed with an unrestricted mix
- A "cross-product" form, which creates a record for every possibilities of pairing all available cell values

Tehtävä 6

Kent mainitsee 5. normaalimuodosta: "Fifth normal form does not differ from fourth normal form unless there exists a symmetric constraint --"
Mihin sana "symmetric" viittaa tässä?

A table is shifted to the fifth normal form for normalisation, when it cannot be broken into any smaller tables with different keys.

A symmetric constraint caused by the fifth normal form does differ from the fourth normal form. A symmetric constraint is used to generate a certain regulation for a combination of all defined attributes in a table in 4NF and assess, which combination is valid and which is not. As a results, certain redundancies could be eliminated and the database becomes cleaned.