Data Modelling and Normalization

Mullins chapter 3

Bjarte Kileng

HVL

January 15, 2018

Entities, occurences and attributes

- Data modelling and UML (or E/R diagrams) have been studied in earlier courses (e.g. DAT104).
- Data model – Abstraction of real world things.
- The *entities* are the «objects» that are stored in the database, e.g. *student*.
- Occurence is an instance of an entity, e.g. the student *Anne Annesen* is an occurence of *student*.
- Attributes are the characteristics of an entity. *Name*, *birth date* and *phone number* are three attributes of a occurence of *student*.

Attributes

An attribute does one of three things:

1. Identifies an entity:
   - *Candidate key*.
   - Immutable.
2. Relates entities:
   - *Foreign key*.
   - Refers to the primary key of an occurrence of another entity.
3. Describes an occurrence of an entity.

Primary key

- One key chosen from the candidate keys.
- Used to identify an entity occurrence.

Conceptual data model

- High level, business oriented view.
- Focus on the most important entities, attributes and relationships.
- Can contain many-to-many relationships.
- Cardinality, optionality and data types can be skipped.

Logical data model

- Fully normalised entities.
- All attributes are defined.
- All candidate keys, primary keys and foreign keys are defined.
- No many-to-many relationships.

Physical data model

- The logical model must be transformed into a physical implementation in a DBMS.
- Details in chapter 4.

Normalisation

- There are many correct models of the world.
- Not all equivalent models are equally good when it comes to:
  - reading data,
  - manipulating data.

Normalisation

Identify the one best place where each fact belongs.

Normalisation

Design approach that minimises data redundancy and optimises data structures.

First normal form

Domain

Domain of an attribute is the universe of values of the attribute.

1NF

A row is in first normal form if and only if all underlying domains contain atomic values only.

Atomic value

Wheter or not a value is atomic depends on the use of the value.

An unnormalised entity

StudentID	StudentName	MajorID	StudentMajor	CourseNum	CourseName	CourseCompDate
12	Olsen, Ole	INF	Informatics	TOD062 TOD072	Programmering Databaser 1	2006-11-23 2008-11-25
14	Annesen, Anne	INF	Informatics	TOD072 FOA031 FOA052	Databaser 1 Fysikk Kjemi og miljø	2008-11-25 2008-12-3 2009-5-27
17	Gretesen, Grete	EL	Elkraft	TOE152 HOE076	Elektriske anlegg Hovedprosjekt	2007-5-22 2009-6-16

Table: Unnormalised Student data

- How does this entity break with the first normal form?
  - Repeating groups (the courses).
  - Attribute *StudentName* is not atomic.
    - But this depends on our use of the data.

Entities in 1NF

StudentID	LastName	FirstName	MajorID	StudentMajor
12	Olsen	Ole	INF	Informatics
14	Annesen	Anne	INF	Informatics
17	Gretesen	Grete	EL	Elkraft

Table: Entity *Student* in 1NF

StudentID	CourseNum	CourseName	CourseCompDate
12	TOD062	Programmering	2006-11-23
12	TOD072	Databaser 1	2008-11-25
14	TOD072	Databaser 1	2008-11-25
14	FOA031	Fysikk	2008-12-3
14	FOA052	Kjemi og miljø	2009-5-27
17	TOE152	Elektriske anlegg	2007-5-22
17	HOE076	Hovedprosjekt	2009-6-16

Table: Entity *Course* in 1NF

Second Normal Form

2NF

A row is in second normal form if and only if it is in first normal form and every non-key attribute (i.e. not part of any candidate key) is fully dependent on a candidate key (or on another non-key attribute).

- Can you see any problems with entity *Course* ?

StudentID	CourseNum	CourseName	CourseCompDate
12	TOD062	Programmering	2006-11-23
12	TOD072	Databaser 1	2008-11-25
14	TOD072	Databaser 1	2008-11-25
14	FOA031	Fysikk	2008-12-3
14	FOA052	Kjemi og miljø	2009-5-27
17	TOE152	Elektriske anlegg	2007-5-22
17	HOE076	Hovedprosjekt	2009-6-16

  - *CourseName* depends on *CourseNum* but not on *StudentID*.
- Solution?
  - Move the attribute with the part of the primary key on which it depends to a new table.

Entities in 2NF

StudentID	LastName	FirstName	MajorID	StudentMajor
12	Olsen	Ole	INF	Informatics
14	Annesen	Anne	INF	Informatics
17	Gretesen	Grete	EL	Elkraft

Table: Entity *Student* (unchanged from the 1NF form)

StudentID	CourseNum	CourseCompDate
12	TOD062	2006-11-23
12	TOD072	2008-11-25
14	TOD072	2008-11-25
14	FOA031	2008-12-3
14	FOA052	2009-5-27
17	TOE152	2007-5-22
17	HOE076	2009-6-16

Table: Entity *Enrolment* in 2NF

CourseNum	CourseName	Credits
TOD062	Programmering	10
TOD072	Databaser 1	5
FOA031	Fysikk	10
FOA052	Kjemi og miljø	10
TOE152	Elektriske anlegg	10
HOE076	Hovedprosjekt	15

Table: Entity *Course* in 2NF

Null values and normalisation

- Attribute value **null** may indicate either that a value is unknown or that the attribute is "not applicable" for this occurence of the entity.
- Value **null**, meaning "not applicable" may indicate a normalisation problem.
- How can an attribute with value "not applicable" depend fully on a candidate key?

Third normal form

3NF

A row is in third normal form if and only if it is in 2NF and every non-key attribute is nontransitively dependent (i.e. directly dependent) on the primary key (PK).

- Do you see any problems with the 2NF Student entity?

StudentID	LastName	FirstName	MajorID	StudentMajor
12	Olsen	Ole	INF	Informatics
14	Annesen	Anne	INF	Informatics
17	Gretesen	Grete	EL	Elkraft

  - *StudentMajor* depends on *StudentID* transitively through *MajorID*.
  - *StudentMajor* is not a key.
- Solution?
  - Move attributes that do not depend on the PK to a new table, together with the non-PK attribute on which they depend.

Entities in 3NF

StudentID	LastName	FirstName	MajorID
12	Olsen	Ole	INF
14	Annesen	Anne	INF
17	Gretesen	Grete	EL

Table: Entity *Student* in 3NF

MajorID	StudentMajor
INF	Informatics
EL	Elkraft

Table: Entity *Major* in 3NF

StudentID	CourseNum	CourseCompDate
12	TOD062	2006-11-23
12	TOD072	2008-11-25
14	TOD072	2008-11-25
14	FOA031	2008-12-3
14	FOA052	2009-5-27
17	TOE152	2007-5-22
17	HOE076	2009-6-16

Table: Entity *Enrolment* (unchanged)

CourseNum	CourseName	Credits
TOD062	Programmering	10
TOD072	Databaser 1	5
FOA031	Fysikk	10
FOA052	Kjemi og miljø	10
TOE152	Elektriske anlegg	10
HOE076	Hovedprosjekt	15

Table: Entity *Course* (unchanged)

More details

- Additional normal forms exist, but normalisation beyond 3NF seldom occur in practise.
- Normalisation is done when moving from the conceptual level to the logical level.
- Only 1NF is required for a relational database.
- 3NF make it easier to manage and maintain data integrity.
- Due to performance issues, the physical model may deviate from 3NF.