U11 Normalization

Jaume Barceló

CIFP Francesc de Borja Moll

Database Management

Contents

Introduction

- If we derive the relational model from an Entity-Relationship Model, most of the times it will be correct.
- There are a number of normalization rules that help us to improve the design of our database.
- We will see, with examples, those that are easier and more useful: NF1, NF2, NF3, BCNF

Functional Dependency

- Given a relation R, a set of attributes X in R is said to functionally determine another set of attributes Y, also in R, (written $X \xrightarrow{FD} Y$) if, and only if, each X value in R is associated with precisely one Y value in R; R is then said to satisfy the functional dependency $X \xrightarrow{FD} Y$.
- customer(customer_id [PK], name, address, phone)
- stock(spare[PK], warehouse[PK], quantity)

Fully Functional Dependency

- Given a relation R, a set of attributes X in R is said to fully functionally determine another set of attributes Y, also in R, (written $X \xrightarrow{FFD} Y$) if, and only if,
- Y depends functionally on X and
- Y does not functionally depend on any proper subset of X.
- customer(customer_id [PK], name, address, phone)
- stock(spare[PK], warehouse[PK], quantity)

First Normal Form

- The domain of each attribute contains only atomic (indivisible) values (the value of each attribute contains only a single value from that domain).
- In other words, we have to get rid of composite atributes and multi-valued attributes.

2NF and 3NF Based on Primary Keys

- When defining the 2NF and the 3NF, it is possible to do so based on primary keys or in a more general form considering all candidate keys.
- For simplicity, we will consider definitions based on primary keys.

Second Normal Form

- It is 1NF
- Every nonprime attribute A in R is fully functionally dependent on the primary key of R.
- purchase(supplier[PK], article[PK], date[PK], quantity, city_supplier)

Second Normal Form (II)

- supplier(supplier[PK], city)
- purchase(supplier[PK][FK], article[PK], date[PK], quantity)

Third Normal Form

- It is 2NF and
- No nonprime attribute of R is transitively dependent on the primary key.
- employee(employee_id[PK], department_id, department_location)

Third Normal Form II

- employee(employee_id[PK], department_id[FK])
- department(department_id[PK], department_location)

Boyce-Codd Normal Form

- It is in 3NF
- For each candidate key it is verified that each attribute that does not belong to it is fully functionally dependent on it.
- school(student_id[PK], course_id[PK], professor_id)
- student_id, course_id $\xrightarrow{FFD}$ professor_id
- professor_id $\rightarrow$ course_id

Boyce-Codd Normal Form (II)

- It is not BCNF because there is a candidate key (student_id, professor_id) that results in the third columm (course_id) not being fully functional dependant on the candidate key, as the course depends only on the professor.

Boyce-Codd Normal Form (III)

- school_professor(professor_id[PK], course_id)
- school_student(student_id[PK], professor_id[PK][FK])