CT3532 Database Systems

Colm O'Riordan

Recap - ER modelling

- ER modelling concepts
- Guidelines to map to a relational model

Recap - Normalisation

Normalisation can be used to:

- develop a normalised relational schema given the universal relation
- verify correctness of relational schema developed from conceptual design.

Decompose relation such that it satisfies successively restrictive normal forms

Desirable properties of a relational schema

- Clear semantics of a relation
- Reducing the number of redundant values in tuples
- Reducing the number of null values in tuples
- Disallowing the possibility of generating spurious tuples.

Clear semantics

The semantics of a relation refers to how the attributes grouped together in a relation are to be interpreted.

If ER modelling is done carefully and the mapping undertaken correctly, it is likely the semantics of the resulting relation will be clear.

One should try to design a relation so that it is easy to explain its meaning.

Reducing redundancy

The presence of redundancy leads to:

1. waste of storage space
2. potential for anomalies (deletion, update, insertion). One try to design relations so that no anomalies may occur. If an anomaly can occur, it should be noted.
3. Normalisation will remove many of the potential anomalies.

Reducing the number of null values

Having nulls is often necessary; however having nulls can create problems.

- waste of space
- different interpretations:
  - attribute does not apply to this tuple
  - attribute value is unknown
  - attribute value is known but absent.
- difficulty with aggregate functions
- different meanings with respect to different join operations.

Disallowing generation of spurious tuples

If a relation R is decomposed into R_1 and R_2 and connected via a primary key - foreign key pair, then performing an equi-join between R_1 and R_2 on the involved keys should not produce tuples that were not in the original relation R.

Desirable properties - more formally

Typically we have a relation, R, and a set of functional dependencies, F, defined over R.

We wish to create a decomposition:

$$D = \{R_1, R_2, \ldots, R_n\}$$

We wish to guarantee certain properties of this decomposition.

Desirable properties - more formally

We require that all attributes in the original R be maintained in the decomposition. i.e.,

$$R = R_1 \cup R_2 \cup \ldots \cup R_n$$

Normalisation - Recap

- A relation R is said to be in first normal form if there are no repeating fields.
- A relation R is said to be in 2NF if it is in 1NF and if every non-prime attribute is fully functionally dependent on the key.
- A relation is said to be in third normal form (3NF) if it is in 2NF and if no non-prime attribute is transitively dependent on the key.

Boyce-Codd Normal Form

A relation is said to be in Boyce-Codd Normal form (BCNF) if the relation is in *3NF* and *if every determinant is a candidate key*.

Sample data

StudentNo	Major	Advisor
123	I.T.	Smith
123	Econ	Murphy
444	Biol.	O' Reilly
617	I.T.	Jones
829	I.T.	Smith

Constraints:

- A student may have more than one major
- For each major a student has only one advisor
- Each major can have several advisors
- Each advisor advises one major
- Each advisor can advise several students

Functional Dependencies

- {StudentNo, Major} $\rightarrow$ {Advisor}
- {Advisor} $\rightarrow$ Major

Update anomaly may exist

If student 444 changes major, we lose information that O' Reilly supervises Biology

Decompose tables so as to satisfy BCNF

TAKES: StudentNo, Advisor
ADVISES: Advisor, Major

General Rule

Consider relation R with functional dependencies F.
If $X \rightarrow Y$ violates BCNF, decompose R into

- $\{R - Y\}$
- $\{XY\}$.

Let $R = \{A, B, C, D, E, F, G, H\}$
The functional dependencies defined over R are:

$$A \rightarrow D$$
$$B \rightarrow E$$
$$E \rightarrow F$$
$$F \rightarrow G$$
$$F \rightarrow H$$
$$\{A, B\} \rightarrow C$$
$$C \rightarrow A$$

Decompose R such that BCNF is satisfied.