Relational Database Design

Update Anomalies in Relational Databases

There are two interconnected problems which lead to a bad database design:

- Update anomalies.
- Redundancy problems.

Example 1.

Let $F_1 = \{\text{E} \rightarrow \text{D}, \text{D} \rightarrow \text{M}, \text{M} \rightarrow \text{D}\}$.

- E stands for ENAME,
- D stands for DNAME, and
- M stands for MNAME,

$\star$ E is the only key for EMP_1 with respect to F_1.

A relation r_1 over EMP_1:

$ENAME$	$DNAME$	$MNAME$
$Mark$	$Computing$	$Peter$
$Angela$	$Computing$	$Peter$
$Graham$	$Computing$	$Peter$
$Paul$	$Math$	$Donald$
$George$	$Math$	$Donald$

Problems with EMP_1 and F_1:

1. Due to entity integrity we cannot insert a tuple with a null ENAME;
   such a *problem* is called an **insertion anomaly**.
2. For the same reason as (1) we cannot delete all the employees and keep just the department information;
   such a *problem* is called a **deletion anomaly**.

3. E.g. modifying "Peter" to "Philip" in the second tuple, does not violate any FD resulting from a key but D $\rightarrow$ M would be violated (D is not a key for EMP_1 with respect to F_1);
   such a *problem* is called a **modification anomaly**.

4. E.g., same as (3) but this time we modify "Computing" in the first tuple to "Math".

$\star$ In (3) and (4) it is **not** sufficient to check that r_1 satisfies the FDs resulting from the keys of R with respect to F.

$\Rightarrow$ Ideally, we would like **all** the FDs of a relation schema to be inferred from **key dependencies**, i.e. FDs of the form

$$\mathrm{K} \rightarrow \mathrm{schema(R)},$$

where K is a key for R with respect to F.

5. There is **redundancy** in r_1, i.e. for every employee in a given department MNAME is repeated.

$\star$ "Peter" appears three times for "Computing" and "Donald" twice for "Math".

Example 2.

Let $F_2 = \{\text{E} \rightarrow \text{S}\}$.

1. E stands for ENAME,
2. S stands for SAL, and
3. C stands for CNAME.

$\star$ EC is the only key for EMP_2 with respect to F_2.

A relation r_2 over EMP_2:

$ENAME$	$CNAME$	SAL
$Jack$	$Jill$	25
$Jack$	$Jake$	25
$Jack$	$John$	25
$Donald$	Dan	30
$Donald$	$David$	30

Problems with EMP_2 and F_2:

1. An insertion anomaly is present, since we cannot insert an employee without any children.
2. A deletion anomaly is present, since if there is a mistake and "Donald" does not have any children, we cannot record this fact, with deleting the two tuples for "Donald".
3. A modification anomaly occurs if we try and modify the salary of "Jack" to be 27 instead of 25, since **no** FD resulting from a key will be violated, **but** E $\rightarrow$ S would be violated.
4. A redundancy problem is present, since the salary of every employee is repeated for every child.

Example 3.

Let $F_3 = \{\text{SC} \rightarrow \text{P}, \text{P} \rightarrow \text{C}\}$.

1. S stands for Street,
2. C stands for City, and
3. P stands for Postcode.

$\star$ SC and PS are the two keys for R w.r.t.with respect to F; assume PS is the primary key.

A relation r_3 over ADDRESS:

Street	*City*	*Postcode*
Hampstead Way	*London*	*NW*11
Falloden Way	*London*	*NW*11
Oakley Gardens	*London*	*N*8
Gower St	*London*	*WC*1*E*
Amhurst Rd	*London*	*E*8

Problems with ADDRESS and F_3:

1. An insertion anomaly is present, since we cannot insert an address which has not been assigned a postcode.
2. A deletion anomaly is present, since if there is a mistake and a postcode is erroneous, we cannot remove this error without deleting the full address.
3. A modification anomaly occurs if we try and modify the city of the 1st tuple to be "Bristol" instead of "London", since **no** FD resulting from a key will be violated, **but** $\mathrm{P} \rightarrow \mathrm{C}$ would be violated.
4. A redundancy problem is present, since the postcode of every city is repeated for each street.

Formalising Redundancy Problems.

Let R be a relation schema and F be a set of FDs over R.

Definition. The FD resulting from a key K for R with respect to F is K $\rightarrow$ schema(R). Such an FD is called a **key dependency**.

Definition. R has a **redundancy problem** if

1. there exists a relation r over R that satisfies F, and
2. there exists a FD X $\rightarrow$ A in F and two distinct tuples in r that have equal XA values.

• It can be shown that redundancy problems, give rise to update anomalies and vice versa.

$\star$ Verify that the schemas of Examples, 1, 2 and 3 have redundancy problems.