

Tutorial 4: Logical Database Design

February 2018

`StudentInfo(S, N, M, A, C, T, I, L, G)`

with the following FDs:

1. $S \rightarrow N$
2. $C \rightarrow T, I$
3. $I \rightarrow L$
4. $S, C, M \rightarrow G$
5. $S, M \rightarrow A$
6. $A \rightarrow M$

(a) Give an instance of `StudentInfo` (i.e., a relation) that illustrates these three anomalies: insertion, deletion, and update.

- **Insertion Anomaly:** We know an instructor `I` and his/her Instructor-Location `L`, but we cannot store this piece of information into `StudentInfo` without storing other attributes like `S`, `M`, `A`, `C`, etc. together with `I`, `L`.
- **Deletion Anomaly:** If we happen to delete all tuples that have a given advisor `A`, then we also lose the information that the advisor `A` is advising for the major `M`.
- **Update Anomaly:** We would like to update an instructor's Instructor-Location `L` since he/she just changed an office. However, if we only update one tuple that has `I` and its corresponding `L`, an inconsistency is created because for the same `I` there are more than one values of `L`, unless all tuples with the same `I` are updated with the new `L`.

(b) Consider the decomposition of the relation `StudentInfo` into: `SI1(S, N, M, A, C)` and `SI2(C, T, I, L, G)`. Is this a lossy or lossless-join? Justify your answer.

Answer:

It is a lossy decomposition.

The attribute shared between SI1 and SI2 is C.

According to FD (2) and FD (3), $(C)^+ = C\ T\ I\ L$

$(C)^+ \neq S\ N\ M\ A\ C$, so C cannot be a key for SI1.

$(C)^+ \neq C\ T\ I\ L\ G$, so C cannot uniquely determine a tuple in SI2, i.e. C is not a key for SI2. So it is possible for SI2 to have 2 tuples that share the same C value but have different G values.

Therefore, when we join SI1 and SI2, a tuple in SI1 may have more than 1 tuples in SI2 to choose from, and vice versa. In this case, the join introduces additional tuples that do not exist in the original relation.

Therefore, the decomposition is lossy with respect to the given set of FDs.

(c) Repeat (b) for the decomposition: SI3(S, M, A, C) and SI4(S, N, M, C, T, I, L, G).

Answer:

It is a lossless-join decomposition.

The attributes shared between SI3 and SI4 are S, M, C.

According to FDs, $(S\ M\ C)^+ = S\ N\ M\ A\ C\ T\ I\ L\ G$.

So (S M C) can uniquely determine tuples in SI3 and SI4, i.e., (S M C) is the key for both SI3 and SI4. Therefore, when we join SI3 and SI4, there is only one tuple in SI3 and one tuple in SI4 can join.

Therefore, the decomposition is a lossless-join decomposition with respect to the given set of FDs.