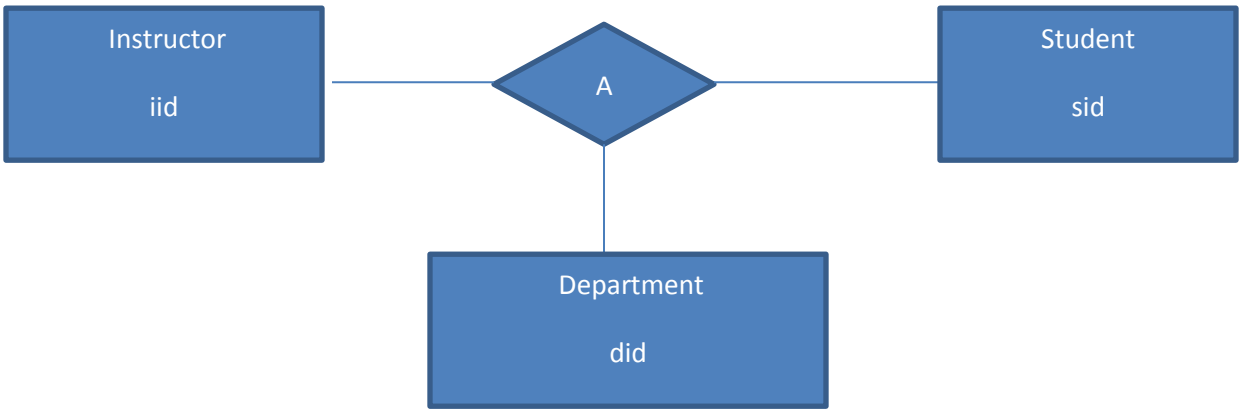
Schema Normalization

30th August 2011

Submitted By:
Saurabh Singla 09010146
Rahul Bhatnagar 09010136

Normalization

Consider the following ER diagram with some FD:



Here FD's are:

iid $\longrightarrow$ did

sid, did $\longrightarrow$ iid

For A schema is

iid, sid, did , with (sid, did) as its key

- A is in 1st NF as all the attributes are atomic
- A is in 2nd NF
- A is in 3rd NF
- A is not in BCNF due to the dependency iid $\longrightarrow$ did

For converting A to BCNF we decompose A into 2 tables (R1 and R2):

R1's schema (sid, iid) with both attributes as key.

R2's schema (iid, did) with iid as primary key.

- R1 and R2 both are in BCNF as there are only two attributes in each of the decomposed table

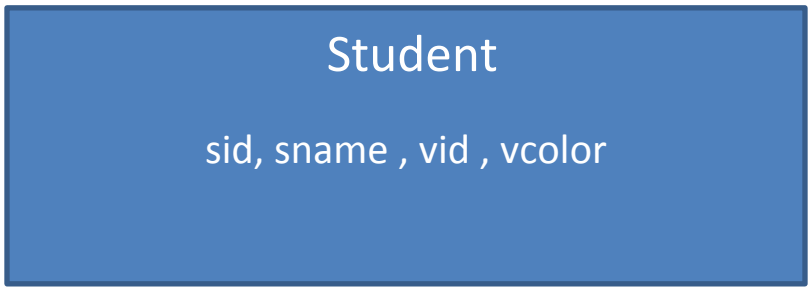
This decomposition is Lossless Decomposition as R1 ⋂ R2 = iid. And iid is the key for R2

But this Decomposition is not Dependency Preserving as the FD sid, did $\longrightarrow$ iid is lost.

Consider a Table Student with attributes sid, sname, vid, vcolor where:

(sid forms the key for the table)

- sid represents student_id
- sname represents student's name
- vid represents vehicle_id , vehicle is owned by students
- vcolor represents color of the vehicle owned by the student



Now consider different situations:

1. Students don't share vehicle and every student has only one.

In this case the FD's are:

1. sid $\longrightarrow$ vid
2. vid $\longrightarrow$ sid

3. sid $\longrightarrow$ sname
4. vid $\longrightarrow$ vcolor

- Table is in 2nd NF as there is only one attribute in candidate key.
- It is not in 3rd NF due to dependency 4 as vcolor is not part of candidate key.
- So it is also not in BCNF as not in 3rd NF

2. Now Let Students share Vehicle.

In this Case all the FD that are in Case 1 hold except vid $\rightarrow$ sid.

- Table is in 2nd NF as there is only one attribute in candidate key.
- It is not in 3rd NF due to dependency 4 as vcolor is not part of candidate key.
- So it is also not in BCNF as not in 3rd NF.

So we decompose Student table into 2 tables Student and Vehicle.

- Student Scheme is (sid, sname, vid) with key sid and vid as foreign key to Vehicle table.
- Vehicle Schema is (vid, vcolor) with key as vcolor.

    - Vehicle is in BCNF as only two attributes.
    - Student is also in BCNF as only dependency for student is either trivial and the others are sid implied and sid is only the key for the table.
    - Decomposition is Lossless as student $\cap$ vehicle = vid and vid is key for vehicle table.
    - It is not dependency preserving decomposition as dependency sid $\rightarrow$ vcolor is lost.

3. Students can have multiple vehicles and vehicles can be shared between students.

Consider the two tables' student and vehicle created in previous case, FD's for vehicle table remains same, but the FD's for student table changes. **Also Candidate key for student table now becomes sid, vid.**

In this case the FD's for student table are:

1. sid, vid → sname
2. sid → sname

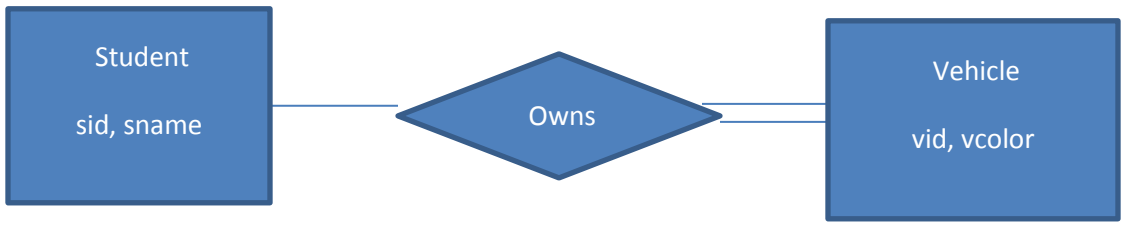
- Vehicle Table is in BCNF as only two attributes.
- Student Table is in 1st NF, But not in 2nd NF and so not in 3rd NF and BCNF. Table not in 2nd NF because of FD:
  - sid → sname as sname is not a part of candidate key for student table and depends on sid, which is proper subset of candidate key for student table.

To attain higher NF's for student table decompose student table further into two tables:

- Student1 with schema (sid, sname) and sid as its key.
- Student2 with schema (sid, vid) and both attributes as key, i.e. vid, sid together form candidate key for student2.

Now Student1, Student2, Vehicle all three tables have only two attributes each which ultimately make them in BCNF.

Originally the ER diagram for above problem must be as follows:



While converting ER diagrams to relational model some information is always lost, as from the above ER diagram it is clear that for vehicle table their must be full participation in the relation owns, but in relation model tables we created there is no such information, but if we do ER modeling Properly and convert ER diagrams to relational tables properly, there will be no need for decomposition's to be done to attain the NF's.