CS3043 Database Systems

Laboratory Exercise 5

Database Normalization

Database Normalization is a technique of organizing the data in the database. Normalization is a systematic approach of decomposing tables to eliminate data redundancy(repetition) and undesirable characteristics like Insertion, Update and Deletion Anamolies. It is a multi-step process that puts data into a tabular form, removing duplicated data from the relation tables. Normalization is used for mainly two purposes,

- Eliminating redundant (useless) data.
- Ensuring data dependencies make sense i.e data is logically stored.

1 Issues Without Normalization

If a table is not properly normalized and has data redundancy then it will not only eat up extra memory space but will also make it difficult to handle and update the database, without facing data loss. Insertion, Updation and Deletion Anamolies are very frequent if the database is not normalized. To understand these anomalies let us take an example of a **Student** table.

rollno	name	branch	hod	office_tel
401	Akon	CSE	Mr. X	53337
402	Bkon	CSE	Mr. X	53337
403	Ckon	CSE	Mr. X	53337
404	Dkon	CSE	Mr. X	53337

In the table above, we have data of 4 Computer Sci. students. As we can see, data for the fields branch, hod(Head of Department) and office_tel is repeated for the students who are in the same branch in the college, this is **Data Redundancy**.

1.1 Insertion Anomaly

Suppose for a new admission, until and unless a student opts for a branch, data of the student cannot be inserted, or else we will have to set the branch information as **NULL**. Also, if we have to insert data of 100 students of the same branch, then the branch information will be repeated for all those 100 students.

These scenarios are nothing but **Insertion anomalies**.

1.2 Updation Anomaly

What if Mr X leaves the college? or is no longer the HOD of the computer science department? In that case, all the student records will have to be updated, and if by mistake we miss any record, it will lead to data inconsistency. This is Updation anomaly.

1.3 Deletion Anomaly

In our **Student** table, two different pieces of information are kept together, Student information and Branch information. Hence, at the end of the academic year, if student records are deleted, we will also lose the branch information. This is Deletion anomaly.

2 Normalization Rule

Normalization rules are divided into the following normal forms:

1. First Normal Form
2. Second Normal Form
3. Third Normal Form
4. BCNF
5. Fourth Normal Form

2.1 First Normal Form (1NF)

For a table to be in the First Normal Form, it should follow the following 4 rules:

1. It should only have a single(atomic) valued attributes/columns.
2. Values stored in a column should be of the same domain
3. All the columns in a table should have unique names.
4. And the order in which data is stored does not matter.

Link: https://www.youtube.com/watch?v=mUtAPbb1ECM

2.2 Second Normal Form (2NF)

For a table to be in the Second Normal Form,

1. It should be in the First Normal form.
2. And, it should not have Partial Dependency.

Link: https://www.youtube.com/watch?v=R7UblSu4744

2.3 Third Normal Form (3NF)

A table is said to be in the Third Normal Form when,

1. It is in the Second Normal form.
2. And, it doesn't have Transitive Dependency.

Link: https://www.youtube.com/watch?time_continue=7&v=aAx_JoEDXQA

2.4 Boyce and Codd Normal Form (BCNF)

Boyce and Codd's Normal Form is a higher version of the Third Normal form. This form deals with the certain type of anomaly that is not handled by 3NF. A 3NF table which does

not have multiple overlapping candidate keys is said to be in BCNF. For a table to be in BCNF, the following conditions must be satisfied:

- ❖ R must be in 3rd Normal Form
- ❖ For each functional dependency (X → Y), X should be a super Key.

Link: https://www.youtube.com/watch?v=NNjUhvvwOrk

3 Lab Work

FastDel is a delivery company who is planning on moving from its traditional file-based system to a DBMS. During the process of designing the DBMS, they have met with few optimization problems. As the DB design engineer of the company, you are assigned the task of solving the following problems as a first step to identifying an optimized DB design.

1. FastDel tracks their customer home locations by recording the latitude and longitude in the address column as follows. Convert the following table to **First Normal Form**.

customer_ID	first_name	second_name	contact_no	location
0001	Michael	Phelps	941111111111	33.680565, 73.020199
0002	Roger	Federer	942222222222	33.646104, 72.990074

2. The table proposed to include the details of FastDel's customers and their respective sales representatives are given below. Convert the table into **second normal form**.

customer_id	customer_name	sales_rep_id	sales_rep_name	payment
0009	Tom	0987	Andrew	100000
0013	Jane	0034	Peterson	150000

3. The table maintained to track the delivery services provided by FastDel and the amount charged is given below. Here the package ID determines the Courier ID while the Courier ID determines the Courier Type. Convert the table into **third normal form**.

package_id	courier_id	courier_type	amount
1	3	International	145.00
2	2	Same Day	14.00
3	1	Standard	16.50
4	4	Overnight	80.00
5	1	Standard	20.40

4. Find the set of attributes that are functionally determined by **IK** under F (set of functional dependencies)

$$R = \{H, I, K, L, M, N, O\}$$
$$F = \{H \rightarrow I, \quad K \rightarrow OM, \quad O \rightarrow MN, \quad KO \rightarrow I\}$$

5. Find the canonical cover of the following set of F of functional dependencies.

$$F = \{A \rightarrow BC$$
$$B \rightarrow C$$
$$A \rightarrow B$$
$$AB \rightarrow C\}$$

6. The delivery company stores employee information in the following manner. Due to the shortage of labour, each employee can belong to more than one department. Part of the employee details table used by the company is shown below. Is the below table in **BCNF**? If not, perform suitable operations in order to make the table comply with BCNF

Note: emp_dept - the department to which employee belongs to
dept_no_of_emp - number of employees in the department

emp_id	emp_nationality	emp_dept	dept_type	dept_no_of_emp
002	Sri Lankan	warehousing	D003	6
002	Sri Lankan	logistics	D002	15
003	Canadian	logistics	D002	15
003	Canadian	finance	D001	9