GALGOTIAS UNIVERSITY

Program   :B.Sc

Course Code   :BSCS3570

Course Name   :Advances in Databases

Faculty   :Dr Satheesh A

GALGOTIAS
UNIVERSITY

Normalization

What is Normalization?

❑**It was first proposed by Edgar F. Codd as part of his relational model.**

❑**Database design technique**

❑**Reduces data redundancy**

❑**Eliminates undesirable characteristics like Insertion, Update and Deletion Anomalies.**

❑**Normalization rules divides larger tables into smaller tables and links them using relationships.**

Problem Without Normalization

- Without Normalization, it becomes difficult to handle and update the database, without facing data loss.

- Insertion, Updation and Deletion Anamolies are very frequent if Database is not Normalized.

- To understand these anomalies let us take an example of Student table.

S_id	S_Name	S_Address	Subject_opted
401	Adam	Noida	Bio
402	Alex	Panipat	Maths
403	Stuart	Jammu	Maths
404	Adam	Noida	Physics

Problem Without Normalization

- Updation Anamoly : To update address of a student who occurs twice or more than twice in a table, we will have to update *S_Address* column in all the rows, else data will become inconsistent.

Problem Without Normalization

▪ Insertion Anamoly : Suppose for a new admission, we have a Student id(S_id), name and address of a student but if student has not opted for any subjects yet then we have to insert NULL there, leading to Insertion Anamoly.

EmployeeID	SalesPerson	SalesOffice	OfficeNumber	Customer1	Customer2	Customer3
1003	Mary Smith	Chicago	312-555-1212	Ford	GM	
1004	John Hunt	New York	212-555-1212	Dell	HP	Apple
1005	Martin Hap	Chicago	312-555-1212	Boeing		
???	???	Atlanta	312-555-1212			

Problem Without Normalization

- Deletion Anamoly : If (*S_id*) 401 has only one subject and temporarily he drops it, when we delete that row, entire student record will be deleted along with it.

S_id	S_Name	S_Address	Subject_opted
401	Adam	Noida	Bio
402	Alex	Panipat	Maths
403	Stuart	Jammu	Maths
404	Adam	Noida	Physics

Normalization Rule

- Normalization rule are divided into following normal form.

    1.First Normal Form

    2.Second Normal Form

    3.Third Normal Form

    4.BCNF

    5. Fourth Normal Form

    6. 5NF

First Normal Form (1NF)

- There are no repeating or duplicate fields.
- Each cell contains only a single value.
- Each record is unique.
  - Identified by primary key

Example

item	colors	price	tax
T-shirt	red, blue	12.00	0.60
polo	red, yellow	12.00	0.60
T-shirt	red, blue	12.00	0.60
sweatshirt	blue, black	25.00	1.25

Table is not in first normal form because:

- Multiple items in color field

- Duplicate records / no primary key

Example

item	color	price	tax
T-shirt	red	12.00	0.60
T-shirt	blue	12.00	0.60
polo	red	12.00	0.60
polo	yellow	12.00	0.60
sweatshirt	blue	25.00	1.25
sweatshirt	black	25.00	1.25

Table is now in first normal form.

Example 2

Student Table

Student	Age	Subject
Adam	15	Biology, Maths
Alex	14	Maths
Stuart	17	Maths

- In First Normal Form, any row must not have a column in which more than one value is saved, like separated with commas. Rather than that, we must separate such data into multiple rows.

Example 2

Student Table following 1NF will be :

Student	Age	Subject
Adam	15	Biology
Adam	15	Maths
Alex	14	Maths
Stuart	17	Maths

- Using the First Normal Form, data redundancy increases, as there will be many columns with same data in multiple rows but each row as a whole will be unique.

GALGOTIAS
UNIVERSITY

2 NF

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.

Second Normal Form (2NF)

Example: Suppose a school wants to store the data of teachers and the subjects they teach. They create a table that looks like this: Since a teacher can teach more than one subjects, the table can have multiple rows for a same teacher.

Teacher_id	subject	Teacher_age
111	Maths	38
111	Physics	38
222	Biology	38
333	Physics	40
333	Chemistry	40

Candidate Keys:
{teacher_id, subject}

Non prime attribute:
teacher_age

Is it Second Normal Form (2NF)?

StudentID	ProjectID	StudentName	ProjectName
S89	P09	Olivia	Geo Location
S76	P07	Jacob	Cluster Exploration
S56	P03	Ava	IoT Devices
S92	P05	Alexandra	Cloud Deployment

Example

item	color	price	tax
T-shirt	red	12.00	0.60
T-shirt	blue	12.00	0.60
polo	red	12.00	0.60
polo	yellow	12.00	0.60
sweatshirt	blue	25.00	1.25
sweatshirt	black	25.00	1.25

Table is not in second normal form because:

- **price** and **tax** depend on **item**, but not **color**

Example

item	color
T-shirt	red
T-shirt	blue
polo	red
polo	yellow
sweatshirt	blue
sweatshirt	black

item	price	tax
T-shirt	12.00	0.60
polo	12.00	0.60
sweatshirt	25.00	1.25

Tables are now in second normal form.

Example 2

New Student Table following 2NF will be :

Student	Age
Adam	15
Alex	14
Stuart	17

- In Student Table the candidate key will be Student column, because all other column i.e Age is dependent on it.

Example 2

New Subject Table introduced for 2NF will be :

Student	Subject
Adam	Biology
Adam	Maths
Alex	Maths
Stuart	Maths

- In Subject Table the candidate key will be {Student, Subject} column. Now, both the above tables qualifies for Second Normal Form and will never suffer from Update Anomalies.

- Although there are a few complex cases in which table in Second Normal Form suffers Update Anomalies, and to handle those scenarios Third Normal Form is there.

3 NF

3 NF

A table is in third normal form if:

- ❑ **A table is in 2nd normal form.**

- ❑ It contains only columns that are **non-transitively dependent** on the primary key

Transitive

- When something is *transitive*, then a meaning or relationship is the same in the middle as it is across the whole. If it helps think of the prefix *trans* as meaning "across." When something is transitive, then if something applies from the beginning to the end, it also applies from the middle to the end.

- Since ten is greater than five, and five is greater than three, you can infer that ten is greater than three

Dependence

- An object has a dependence on another object when it relies upon it. In the case of databases, when we say that a column has a dependence on another column, we mean that the value can be derived from the other.

- For example, my age is dependent on my birthday.

- Dependence also plays an important role in the definition of the second normal form.

Transitive Dependence

❑ Consider three columns: AuthorNationality, Author, and Book. Column values for AuthorNationality and Author rely on the Book; once the book is known, you can find out the Author or AuthorNationality. But also notice that the AuthorNationality relies upon Author. That is, once you know the Author, you can determine their nationality. In this sense then, the AuthorNationality relies upon Book, via Author. This is a transitive dependence.

❑ This can be generalized as being three columns: A, B and PK. If the value of A relies on PK, and B relies on PK, and A also relies on B, then you can say that A *relies* on PK *though* B. That is A is transitively dependent on PK.

Primary Key (PK)	Column A	Column B	Transitive Dependence?
PersonID	FirstName	Last Name	No, In Western cultures a person's last name is based on their father's LastName, whereas their FirstName is given to them.
PersonID	BodyMassIndex	Is Overweight?	Yes, BMI over 25 is considered overweight.It wouldn't make sense to have the value IsOverweight be true when the BodyMassIndex was < 25.
PersonID	Weight	Sex	No:There is no direct link between the weight of a person and their sex.
VehicleID	Model	Manufacturer	Yes:Manufacturers make specific models. For instance, Ford creates the Fiesta; whereas, Toyota manufacturers the Camry.

Example

item	color
T-shirt	red
T-shirt	blue
polo	red
polo	yellow
sweatshirt	blue
sweatshirt	black

item	price	tax
T-shirt	12.00	0.60
polo	12.00	0.60
sweatshirt	25.00	1.25

Tables are not in third normal form because:

- **tax** depends on **price**, not **item**

Example

item	color
T-shirt	red
T-shirt	blue
polo	red
polo	yellow
sweatshirt	blue
sweatshirt	black

item	price
T-shirt	12.00
polo	12.00
sweatshirt	25.00

price	tax
12.00	0.60
25.00	1.25

Tables are now in third normal form.

Another Example

Name	Assignment 1	Assignment 2
Jeff Smith	Article Summary	Poetry Analysis
Nancy Jones	Article Summary	Reaction Paper
Jane Scott	Article Summary	Poetry Analysis

Table is not in first normal form because:

- Assignment field repeating
- First and last name in one field
- No (guaranteed unique) primary key field

Another Example

Assignment ID	Description
1	Article Summary
2	Poetry Analysis
3	Reaction Paper

Student ID	First Name	Last Name
1	Jeff	Smith
2	Nancy	Jones
3	Jane	Scott

Assignment ID	Student ID
1	1
1	2
1	3
2	1
2	3
3	2

Tables are in third normal form.

Example 3

Student_Detail Table :

Student_id	Student_name	DOB	Street	city	State	Zip

- In this table Student_id is Primary key, but street, city and state depends upon Zip.

- The dependency between zip and other fields is called transitive dependency.

- Hence to apply 3NF, we need to move the street, city and state to new table, with Zip as primary key.

Example 3

New Student_Detail Table :

Student_id	Student_name	DOB	Zip

Address Table :

Zip	Street	city	state

The advantage of removing transtive dependency is,

- Amount of data duplication is reduced.
- Data integrity achieved.

References

http://www.studytonight.com/dbms/database-normalization.php

https://www.essentialsql.com/get-ready-to-learn-sql-database-normalization-explained-in-simple-english/

http://beginnersbook.com/2015/05/normalization-in-dbms/

https://www.sqa.org.uk/e-learning/MDBS01CD/page_03.htm

http://www.1keydata.com/database-normalization/