NORMALIZATION

**Definition**: - Normalization is the process of organizing the data in the database. Normalization is used to minimize the redundancy from a relation or set of relations. It is also used to eliminate undesirable characteristics like Insertion, **Update, and Deletion Anomalies**. Normalization divides the larger table into smaller and links them using relationships. The normal form is used to **reduce redundancy** from the database table.link - https://www.javatpoint.com/dbms-normalization

## Type of normalization

## 1ST normal from 1nf

* in 1nf each column of the table should contain atomic values which means it should not contain multiple values entries like x,y
* do not intermix different type of values in any column
* each column should have a unique value bcoz same name can lead to confusion at the time of the data retrieval

|  |  |  |
| --- | --- | --- |
| Roll no. | Name | subject |
| 1 | a | Math,hindi |
| 2 | b | english |
| 3 | c | science |

Link-https://www.youtube.com/watch?v=mUtAPbb1ECM&list=PLLGlmW7jT-nTr1ory9o2MgsOmmx2w8FB3&index=2

## 2nd normal form 2nf

* 2nf says that the table should be in 1nf and it should not have any partial dependency
* The partial dependency here means the proper subset of the candidate key should give a non-prime attribute.

|  |  |  |  |
| --- | --- | --- | --- |
| id | Student roll no. | Subject id | marks |
| 1 | 11 | pp | 54 |
| 1 | 11 | ii | 32 |
| 3 | 33 | pp | 21 |

Here in this example we can see that there is a uniquely identified values given in the column but if we want to want get marks for the subject if pp, then it will show a proper out bcoz student roll no. is different for every subject id. Thus,it is showing partial dependency which it should not have in the table.

Link- https://www.youtube.com/watch?v=R7UblSu4744&list=PLLGlmW7jT-nTr1ory9o2MgsOmmx2w8FB3&index=3

## 3rd normal form 3nf

* For the 3rd normal form the table should be in 2nd normal form and it should not have any transitive dependency
* Transitive dependency means it occurs when some non key attribute determines some other attribute

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Id | Roll no. | Subject id | Marks | grade |
| 1 | 11 | pp | 54 | A |
| 1 | 11 | ii | 32 | B |
| 3 | 33 | pp | 21 | C |

Here in this example we have created a new column name grade which I have distributed grades according to marks as +50=A , below 50=b and below 30=c

So as we can see it is not following any transitive dependency

Link-https://www.youtube.com/watch?v=aAx\_JoEDXQA&list=PLLGlmW7jT-nTr1ory9o2MgsOmmx2w8FB3&index=4

## Boyce code normal form BCNF

* To perform BCNF the table should be in 3rd normal form and for any dependency if A is indicating B then A should be a super key
* Boyce Codd Normal Form is also known as 3.5 NF. It is the superior version of 3NF and was developed by Raymond F. Boyce and Edgar F.

|  |  |  |
| --- | --- | --- |
| Student id | Subject | professor |
| 101 | Java | A |
| 101 | C++ | B |
| 102 | Java | C |
| 103 | C# | D |

As we see in the above table it does not show any partial dependencies and transitive dependencies becoz each values are differently identified. We can find out any relation by just seeing the professor column becoz all the professors are allocated differently to each subject.

Link - https://www.youtube.com/watch?v=NNjUhvvwOrk&list=PLLGlmW7jT-nTr1ory9o2MgsOmmx2w8FB3&index=5

## 4th normal form 4nf

* 4th normal form must satisfy two conditions that the table should be in BCNF and it should not have any multi valued dependency
* If a single value of A derives more than more than 1 value in other column then it can cause multi value dependency

|  |  |  |
| --- | --- | --- |
| Student id | Cource | hobby |
| 1 | Science | Cricket |
| 1 | Math | Tennis |
| 2 | C++ | Hockey |
| 2 | python | soccer |

Here we can that multiples values has been allocated to student id so it is creating fuss and accuring multi valued dependency. To avoid this we need to create a separate table which identifies each cource and hobby for a particular student ID

Link- https://www.youtube.com/watch?v=OTCuykFHBeA&list=PLLGlmW7jT-nTr1ory9o2MgsOmmx2w8FB3&index=6

## 5th normal form 5nf

* 5th normal form derives that the table should be in 4nf and it should not have any join dependency
* It is also known as PJNF (project join normal form)

|  |  |  |
| --- | --- | --- |
| Supplier | Product | customer |
| A | 72x | Ford |
| A | Gear | GM |
| B | E switch | Ford |
| B | OBD1 | mercedes |

This describes that if the table is decomposed into smaller tables and that leads to some loss of information or some additional information get created then we should not decompose the table bcoz that will lead to some incorrect information

Link - https://www.youtube.com/watch?v=mbj3HSK28Kk&list=PLLGlmW7jT-nTr1ory9o2MgsOmmx2w8FB3&index=7