Normalization

Database normalization is the process of structuring a relational database in order to improve data integrity and reduce data redundancy.

Data integrity : By creating rules for data and enforcing this rules

Reduce data redundancy : by splitting tables into smaller logical unit and creatin relationship between them .

Rules : Normal Forms

1NF: First Normal Form

2NF: Second Normal Form

3NF: Third Normal Form

Below three are not asked in exam:

BCNF : Boyce-Codd Normal Form

4NF: Fourth Noral Form

5NF: Fifth Normal Form

1NF

* No duplicate rows
* No multivalued attribute
* Column entries must be of same data type
* No repeting attributes

multivalued attribute makes it difficult to search

repeating attribute

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Student ID | Name | DOB | Course 1 | Course 2 |
| 1 | Suhail, sayyad | July | Chem. | Math |
| 2 | Arya,sayyad | Jan | Phy | Chem. |
| 1 | Suhail, sayyad | July | Phy | Chem. |
| 2 | Arya, Sayyad | Jan | Chem. | Math |

Ie: this also has repeating entries

Then

Removed multivalued entry

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Student ID | First Name | Last Name | DOB | Course 1 | Course 2 |
| 1 | Suhail | Sayyad | July | Chem. | Math |
| 2 | Arya | Sayyad | Jan | Phy | Chem. |
| 1 | Suhail | Sayyad | July | Phy | Chem. |
| 2 | Arya | Sayyad | Jan | Chem. | Math |

Repeating Attributes

Now, will create three tables

|  |  |  |  |
| --- | --- | --- | --- |
| Student ID | First Name | Last Name | DOB |
| 1 | Suhail | Sayyad | July |
| 2 | Arya | Sayyad | Jan |

|  |  |
| --- | --- |
| Course ID | Corse Name |
| 1 | Chem |
| 2 | Math |
| 3 | Phy |
|  |  |

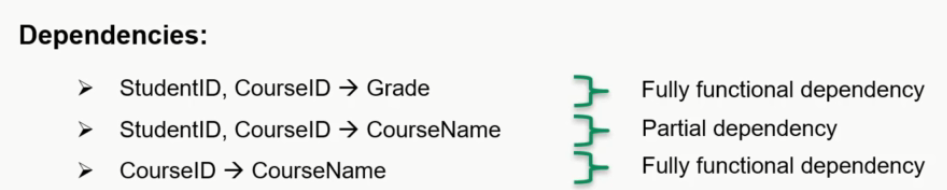
|  |  |
| --- | --- |
| Student ID | Course ID |
| 1 | 1 |
| 1 | 2 |
| 1 | 3 |
| 2 | 1 |
| 2 | 2 |
| 2 | 3 |

2NF

* The relation must be in the first Normal form
* No partial dependency (every non-primary-key attribute should be fully functionally dependent on the **primary key**)

|  |  |  |  |
| --- | --- | --- | --- |
| Student ID | Course ID | Grade | Cource Name |
| 1 | 1 | 90 | Chem. |
| 1 | 2 | 54 | Phy |
| 1 | 3 | 67 | math |
| 2 | 1 | 43 | chem |
| 2 | 2 | 87 | Phy |
| 2 | 3 | 56 | math |
|  |  |  |  |

Composite key



Here we were ABLE TO GET COURCE NAME DIRECTLY FROM COURCE ID

|  |  |  |
| --- | --- | --- |
| Student ID | Course ID | Grade |
| 1 | 1 | 90 |
| 1 | 2 | 54 |
| 1 | 3 | 67 |
| 2 | 1 | 43 |
| 2 | 2 | 87 |
| 2 | 3 | 56 |

|  |  |  |  |
| --- | --- | --- | --- |
| Student ID | Course ID | Grade | Cource Name |
| 1 | 1 | 90 | Chem. |
| 1 | 2 | 54 | Phy |
| 1 | 3 | 67 | math |
| 2 | 1 | 43 | chem |
| 2 | 2 | 87 | Phy |
| 2 | 3 | 56 | math |

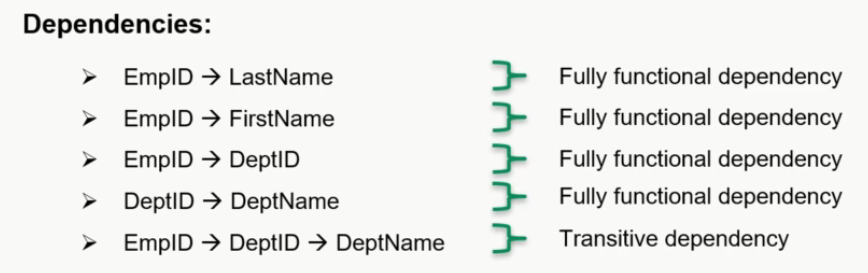
|  |  |
| --- | --- |
| Course ID | Corse Name |
| 1 | Chem |
| 2 | Math |
| 3 | Phy |

3NF

* The relation must be in the second normal form
* No transitive dependency (Non-primary-key attribute should **not be** transitively dependent on primary key )

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| EMP ID | FIRST NAME | Last name | Dep ID | DEPT NAME |
| 1 | Suhail | sayyad | 1 | Computer |
| 2 | Arya | Sayyad | 2 | HR |
| 3 | Shanu | Sayyad | 2 | HR |
| 4 | shabana | sayyad | 1 | Computer |

Here we can observe that Dept name is dependent on Dep ID and Dep ID is dependent on EMP ID



3nf ->

|  |  |  |  |
| --- | --- | --- | --- |
| EMP ID | FIRST NAME | Last name | Dep ID |
| 1 | Suhail | sayyad | 1 |
| 2 | Arya | Sayyad | 2 |
| 3 | Shanu | Sayyad | 2 |
| 4 | shabana | sayyad | 1 |

|  |  |
| --- | --- |
| Dep ID | DEPT NAME |
| 1 | Computer |
| 2 | HR |