What is normalization

* It is a process or technique to arrange data in efficient way
* To reduce the redundancy
* maintain data integrity
* to avoid insertion,updation and deletion anamoly.

in banking application if you design following table

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Acid | Custid | Cname | Address | Type | Balance | Data of opening |
| 1 | 100 | Kishori | Aundh | Saving | 22222 | 22 nov 1997 |
| 2 | 100 | Kishori | Baner | Saving | 44444 | 20 nov2000 |
| 3 | 100 | Kishori | Baner | Demat | 55555 |  |
| 4 | 200 | Rajan | Baner | Saving | 55555 |  |
|  | 300 | Revati | Baner |  |  |  |

What is primary key

acid is primary key

1. and hence if any customer does enquiry and do not open a/c we cannot insert/ stroe information of the customer for future use. because acid cannot be null

this is insertion anamoly

1. when we try to modify address of Kishori. it is possible that it is changed only at one place

it is not correct information because of redundancy

This Is updation anamoly

1. If Rajan closes account then while deleting account customer details will also get lost

this is deletion anamoly.

Account

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Acid | Custid | Type | Balance | Data of opening |
| 1 | 100 | Saving | 22222 | 22 nov 1997 |
| 2 | 100 | Saving | 44444 | 20 nov2000 |
| 3 | 100 | Demat | 55555 |  |
| 4 | 200 | Saving | 55555 |  |
|  | 300 |  |  |  |

Customer

|  |  |  |
| --- | --- | --- |
| Custid | Cname | Address |
| 100 | Kishori | Aundh |
| 200 | Rajan | Baner |
| 300 | Revati | Baner |

Normalization:

types of normaliztion

1. 1NF---a table should contain one value in one row and one column

every column should have atomic value

|  |  |  |  |
| --- | --- | --- | --- |
| Custid | Cname | Address | Mobile |
| 100 | Kishori | Aundh | 9822022222 |
| 200 | Rajan | Baner | 982341345,983453333 |
| 300 | Revati | Baner | 98220111,982203333 |

mobile column contains more than one value in one row and one column

hence the table is not in 1 NF

|  |  |  |
| --- | --- | --- |
| Custid | Cname | Address |
| 100 | Kishori | Aundh |
| 200 | Rajan | Baner |
| 300 | Revati | Baner |

|  |  |
| --- | --- |
| Custid | Mobile |
| 100 | 9822022222 |
| 200 | 982341345 |
| 200 | 983453333 |
| 300 | 982203333 |
| 300 | 98220111 |
|  |  |

flight booking system

flight -- not in 1 NF

|  |  |  |
| --- | --- | --- |
| Flightno | Filghname | Seatno |
| 1234 | Luftaza | 1,2,3,4,5,,,,,150 |
|  |  |  |

flight

|  |  |
| --- | --- |
| Flightno | Filghname |
| 1234 | Luftaza |
| 1111 | Indigo |

fligt-seatbooking

|  |  |  |  |
| --- | --- | --- | --- |
| Filgtnum | Seatno | cname | date |
| 1234 | 1 | Kishori | 22 nov |
| 1234 | 2 | Rajan | 3 oct |
|  |  |  |  |

1. 2NF
2. the table should be in 1NF
3. all non key columns are fully functional dependent on the complete primary key

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Empid | Ename | Mobile | Projectid | Pname | Duration |
| 123 | Ashu | 111 | 100 | Inssurance | 6 |
| 124 | Rajas | 2222 | 200 | Loan | 10 |
| 125 | Deven | 3333 | 100 | Inssurance | 8 |
| 123 | Ashu | 111 | 200 | Loan | 24 |

primary key : empid + projectid

prime attribute--- empid, projected

non prime attribute---ename,mobile,pname,duration

empid -🡪 ename, mobile

projectid-🡪pname

empid+projected-🡪 duration

There is a partial dependency hence it is not in 2NF

Employee

|  |  |  |
| --- | --- | --- |
| Empid | Ename | Mobile |
| 123 | Ashu | 111 |
| 124 | Rajas | 2222 |
| 125 | Deven | 3333 |
| 123 | Ashu | 111 |

project

|  |  |
| --- | --- |
| Projectid | Pname |
| 100 | Inssurance |
| 200 | Loan |
| 100 | Inssurance |
| 200 | Loan |

employee-project

|  |  |  |
| --- | --- | --- |
| Empid | Projectid | Duration |
| 123 | 100 | 6 |
| 124 | 200 | 10 |
| 125 | 100 | 8 |
| 123 | 200 | 24 |

Appoinment system

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Patientid | *Pname* | *Drid* | Dname | Date of appointment | Time |
| 100 |  | 1 |  | 22 nov | 10.00am |
| 100 |  | 1 |  | 22 nov | 12.00pm |

patientid-🡪pname

drid-🡪dname

patientid\_drid+date+time--🡪

|  |  |
| --- | --- |
| Patientid | *Pname* |
| 100 |  |
| 100 |  |
|  |  |

|  |  |
| --- | --- |
| *Drid* | Dname |
| 1 |  |
| 1 |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Patientid | *Drid* | Date of appointment | Time |
| 100 | 1 | 22 nov | 10.00am |
| 100 | 1 | 22 nov | 12.00pm |

All these 3 table are in 2NF

3NF

1. a relation should be in 2NF
2. there should not be any transitive relation ship

a->b->c

a relation is in 3NF, then it holds one of the following condition for every transitive dependeancy

1. a should be super key
2. b is prime attribute, each element of b is part of primary key

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Empid | Ename | State | City | zipcode |
| 1 | Rajat | Maharashtra | Mumbai | 23456 |
| 2 | Rashmi | Us | Boston | 45566 |
| 3 | Sunil | MP | Bhopal | 3434 |
| 4 | Ashu | Maharashtra | Mumbai | 2345611 |

empid-🡪ename

city -🡪 state

empid-🡪city -🡪state

if you have transitive dependency then it is not in 3 NF

city

|  |  |
| --- | --- |
| State | City |
| Maharashtra | Mumbai |
| Us | Boston |
| MP | Bhopal |
| Maharashtra | Pune |

employee

|  |  |  |  |
| --- | --- | --- | --- |
| Empid | Ename | City | zipcode |
| 1 | Rajat | Mumbai | 23456 |
| 2 | Rashmi | Boston | 45566 |
| 3 | Sunil | Bhopal | 3434 |
| 4 | Ashu | Mumbai | 23456 |

these tables are in 3NF

4 NF

a table is in BCNF

1. the table should be in 3NF
2. a table is in BCNF if every function dependency x->y , x is a super key

emp---dept

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Empid | Country | Dname | Dtype | deptno |
| 1234 | India | Design | UX | 100 |
| 1234 | India | Testing | UX | 120 |
| 2222 | Japan | Purchase | Product | 130 |
| 2222 | Japan | Sales | Product | 140 |
| 3333 | Japan | sales | Product | 140 |

In this table there is no column which can be primary key.

hence find functional dependencies and break the table accordingly.

Emp-id--🡪 country

dname--🡪dtype, deptno

this table is not in BCNF because empid, dname

this is multivalued dependency

|  |  |
| --- | --- |
| Empid | Country |
| 1234 | India |
| 2222 | Japan |
| 3333 | Japan |

|  |  |  |
| --- | --- | --- |
| Dname | Dtype | deptno |
| Design | UX | 100 |
| Testing | UX | 120 |
| Purchase | Product | 130 |
| Sales | Product | 140 |

|  |  |
| --- | --- |
| Empid | Dname |
| 1234 | Design |
| 1234 | Testing |
| 2222 | Purchase |
| 2222 | Sales |
| 3333 | sales |

Normalize the given table

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Proj  Code | Proj  Type | Proj  Desc | Empno | Ename | Grade | Sal scale | Proj  Join Date Time | Time allocated |  |
| 001 | APP | LNG | 46 | JONES | A1 | 5 | 12/1/1998 | 24 |  |
| 001 | APP | LNG | 92 | SMITH | A2 | 4 | 2/1/1999 | 24 |  |
| 001 | APP | LNG | 96 | BLACK | B1 | 9 | 2/1/1999 | 18 |  |
| 004 | MAI | SHO | 72 | JACK | A2 | 4 | 2/4/1999 | 6 |  |
| 004 | MAI | SHO | 92 | SMITH | A2 | 4 | 5/5/1999 | 6 |  |

**example 2**

* **Orderno**
* **Orderdate**
* **Itemno**
* **Qty**
* **Price**
* **Cname**
* **Custno**
* **Email**
* **Orderamt**
* **Salespersonno**
* **Salespersonname**
* **Locationid ----------location from where item dispatched**
* **Location name**

one order will be placed by one customer, but one customer can place many order

one order items may dispatch from different location

one order will be managed by one salesperson