Week: 4

Normalization:

Database normalization is a technique for designing relational database tables to minimize duplication of information and, in doing so , to safeguard the database against certain types of logical or structural problems namely data anomalies.

The normalization forms are:

1. **First Normal Form**: 1NF requires that the values in each column of a table are atomic. By atomic we mean that there are no sets of values within a column.
2. **Second Normal Form**: where the 1NF deals with atomicity of data, the 2NF deals with relationships between composite key columns and non-key columns. To achieve 2NF the tables should be in 1NF. The 2NF any non-key columns must depend on the entire primary key. In case of a composite primary key, this means that non-key column can't depend on only part of the composite key.
3. **Third Normal Form**: 3NF requires that all columns depend directly on the primary key. Tables violate the third normal form when one column depends an another column, which in turn depends on the primary key(transitive dependency). One way to identify transitive dependency is to look at your tables and see if any columns would require updating if another column in the table was updated. If such a column exists, it probably violates 3NF.

Let's normalize our entities:

Normalization of Passenger entity:

In the passenger entity there exists a passenger with two phone numbers, but atomic values should be there. So we normalize the relation as follows.

Passenger:

Pnrno	pname	age	sex	ticketno	address	phno	Catno
1001	Subbu	31	M	1111	5-4,srpt	9492506282, 9848845985	Cap5112
1002	Achaith	22	M	2222	6-8,hyd	9949060540	Cap6900
1003	Padma	25	F	3333	h/7,vij	9704054050	Cap5772
1004	Ravi	23	M	4444	8-9,hyd	9704613151	Cap6132
1005	Satyam	42	F	5555	9-11,hyd	9848354941	Cap6732

Pnrno	pname	age	sex	ticketno	address	phno	Catno
1001	Subbu	31	M	1111	5-4,srpt	9492506282	Cap5112

1001	Subbu	31	M	1111	5-4,srpt	9848845985	Cap5112
1002	Achaith	22	M	2222	6-8,hyd	9949060540	Cap6900
1003	Padma	25	F	3333	h/7,vij	9704054050	Cap5772
1004	Ravi	23	M	4444	8-9,hyd	9704613151	Cap6132
1005	Satyam	42	F	5555	9-11,hyd	9848354941	Cap6732

The above relation is now in 1NF and the relation is 2NF as there are no partial functional dependencies and the relation is also in 3NF as there are no transitive dependencies.

Normalization of Bus entity:

Bus:

Busno	**serviceno**	**source**	**destination**	**deptime**	**retime**	**bustype**	**Noofseats**
Ap555	3889	Srpt	Hyd	9:00:00	19:15:00	Ac	36
Ap501	3891	Srpt	Hyd	10:00:15	20:15:00	Ac	36
Ap444	3601	Hyd	Srpt	9:00:00	19:30:00	Nonac	52
Ap891	3555	Hyd	Srpt	9:30:00	20:30:00	Nonac	52
Ap8830	3239	Hyd	Vij	9:00:00	22:30:00	Metro	45

In this relation the values in each column are atomic so it is already in 1NF.
In the Bus entity **Busno+serviceno** is the primary key.
There exists following partial dependencies.

Busno ----> Bustype,Noofseats
Serviceno---->Source,Dest

So the relation will be in 2NF as follows.

Busno	**serviceno**	**deptime**	**retime**
Ap555	3889	9:00:00	19:15:00
Ap501	3891	10:00:15	20:15:00
Ap444	3601	9:00:00	19:30:00
Ap891	3555	9:30:00	20:30:00
Ap8830	3239	9:00:00	22:30:00

Busno	**bustype**	**Noofseats**
Ap555	Ac	36
Ap501	Ac	36
Ap444	Nonac	52
Ap891	Nonac	52
Ap8830	Metro	45

serviceno	**source**	**destination**
3889	Srpt	Hyd
3891	Srpt	Hyd
3601	Hyd	Srpt
3555	Hyd	Srpt
3239	Hyd	Vij

The above relation is 2NF. And all columns directly depend on primary key. So there is no transitive dependency and the relation is 3NF.

Normalization of Ticket entity:

Ticketno	Joudate	Joutime	Source	Destination	Seatno	Amount	Catcard
1111	2010-8-5	9:00:00	Srpt	Hyd	5	96	No
2222	2010-8-5	10:00:15	Srpt	Hyd	10	88	Yes
3333	2010-8-15	9:00:00	Hyd	Srpt	15	88	Yes
4444	2010-8-18	9:30:00	Hyd	Srpt	20	96	No
5555	2010-8-6	9:00:00	Hyd	Vij	18	172	Yes

In this relation the values in each column are atomic so it is already in 1NF.
In the above relation there are no partial functional dependencies so the relation is in 2NF.
The ticket entity might face the following transitive dependency

Ticketno-------> catcard
Catcard--------->amount

So the relation is in 3NF.

Ticketno	Joudate	Joutime	Source	Destination	Seatno	Catcard
1111	2010-8-5	9:00:00	Srpt	Hyd	5	No
2222	2010-8-5	10:00:15	Srpt	Hyd	10	Yes
3333	2010-8-15	9:00:00	Hyd	Srpt	15	Yes
4444	2010-8-18	9:30:00	Hyd	Srpt	20	No
5555	2010-8-6	9:00:00	Hyd	Vij	18	Yes

Put the catcard and amount attributes in a separate table. Then the relation should be in 3NF.

Catcard	Amount
No	96
Yes	88
Yes	88
No	96
Yes	172

The above relation is 3NF as we have eliminated the transitive dependencies.

Finally all the tables are normalized and free from data redundancy, partial functional dependencies and transitive dependencies.