Normalization

Normalization

- It is the process of decomposing a database table into smaller tables so as to minimize data redundancy and data anomalies

Why we need to normalize

- Redundancy
- Anomalies
    1. Insetion Anomaly
    2. deletion Anomaly
    3. Updation Anomaly

Student table

Roll No	Batch	Name	Branch	Hod	Hod_room No
20	A	ASHLEY	CSE	Manilal	202
55	A	Savion	CSE	Manilal	202
31	B	Malavika	CSE	Manilal	202
46	B	Rindish	CSE	Manilal	202

Insertion Anomaly

Roll No	Batch	Name	Branch	Hod	Hod_room No
20	A	ASHLEY	CSE	Manilal	202
55	A	Savion	CSE	Manilal	202
31	B	Malavika	CSE	Manilal	202
46	B	Rindish	CSE	Manilal	202
28	B	Kim	CSE	Manilal	202

Deletion Anomaly

Roll No	Batch	Name	Branch	Hod	Hod_room No
20	A	ASHLEY	CSE	Manilal	202
55	A	Savion	CSE	Manilal	202
31	B	Malavika	CSE	Manilal	202
46	B	Rindish	CSE	Manilal	202

Deletion Anomaly

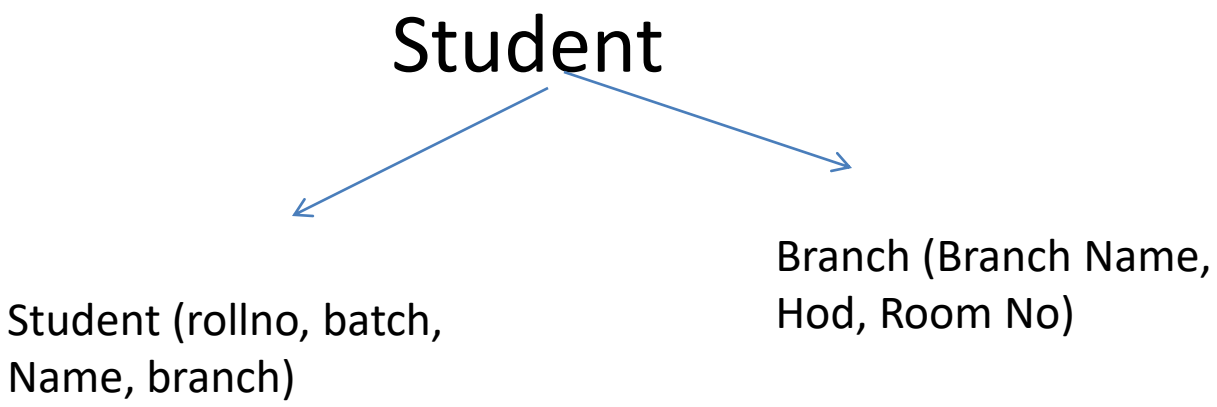
Roll No	Batch	Name	Branch	Hod	Hod_room No
20	A	ASHLEY	CSE	Manilal	202

Updation Anomaly

Roll No	Batch	Name	Branch	Hod	Hod_room No
20	A	ASHLEY	CSE	Manilal	202
55	A	Savion	CSE	Manilal	202
31	B	Malavika	CSE	Manilal	202
46	B	Rindish	CSE	Manilal	202
42	B	Prithvi	CSE	Manilal	202
45	B	Rajath R	CSE	Manilal	202
9	A	Aleena	CSE	Manilal	202
56	A	shalu	CSE	Manilal	202

- It is the process of decomposing a database table into smaller tables so as to minimize data redundancy and inconsistency

How do we Normalize it?????



Student table After Normalization

Roll No	Batch	Name	Branch
20	A	ASHLEY	CSE
55	A	Savion	CSE
31	B	Malavika	CSE
46	B	Rindish	CSE
42	B	Prithvi	CSE
45	B	Rajath R	CSE
9	A	Aleena	CSE
56	A	shalu	CSE

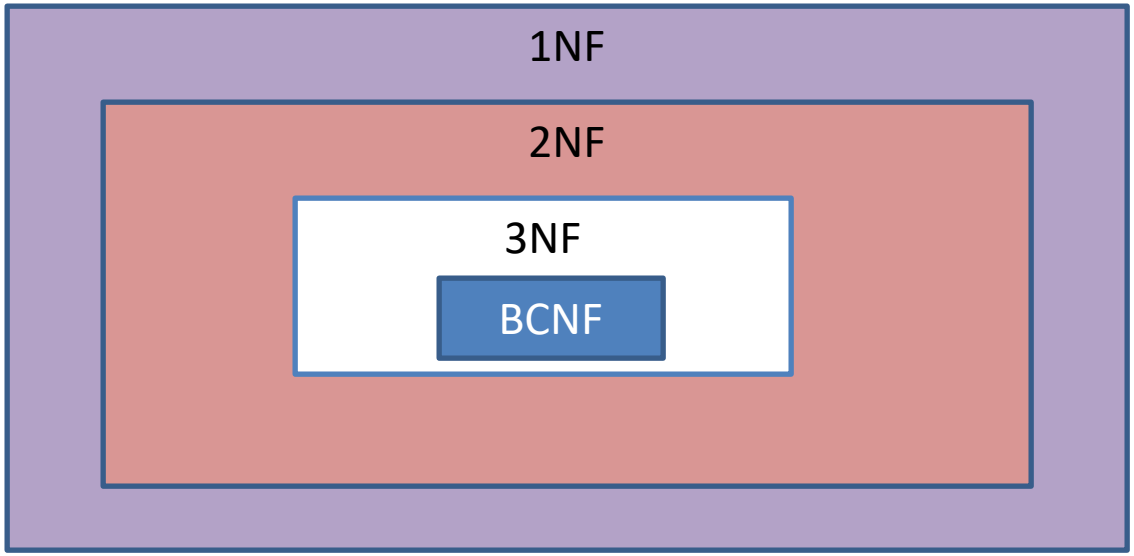
Branch Table After Normalization

Branch	Hod	Room No
CSE	Manilal	202

Normal Forms

- First Normal Form(1NF)
- Second Normal Form(2NF)
- Third Normal Form(3NF)
- Boyce Codd Normal Form(BCNF)

Hierarchy of the Normal forms



First Normal Form(1NF)

- All entries in the table should be atomic

- Atomic means
  - No Multivalued
  - No Composite

  - Two approaches

Dealing with Multivalued Attribute

Roll No	Batch	Name	Phone Number
20	A	ASHLEY	984720,985632
55	A	Savion	99975,986258
31	B	Malavika	89885,72588,9856
46	B	Rindish	958548

Create a separate table for each multi valued attribute

Roll No	Phone Number
20	984720
20	985632
55	99975,
55	986258
31	89885
31	72588
31	9856
46	958548

Dealing with Composite attributes

Roll No	Name	Address
20	ASHLEY	10A, Plalarivattom,Kochi
55	Savion	30S,Changanassery, Kottayam
31	Malavika	23A,Thrikkaakara,Kochi

Method 2: Add separate column for each atomic values

Roll No	Name	Address1	Address2	Address3
20	ASHLEY	10A,	Palarivattom,	Kochi
55	Savion	30S,	Changanassery,	Kottayam
31	Malavika	23A,	Thrikkaakara,	Kochi

Second Normal Form

- The relation should be in I NF
- No nonprime attribute is partially dependent on key

- Normalize R(A,B,C,D) with AB-→C,B->D

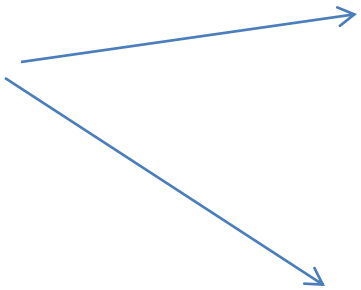
- Prime attributes={A,B}
- Non Prime attributes={C,D}

- Assumption
  - All attributes are atomic

- AB→C
- AB is the full key
- C is the Non Prime attribute
- Non Prime attribute C is dependent on full key
- So full dependency
- B→D
- B is the partial Key(it is a part of key)
- Non Prime attribute D is dependent on partial key
- So partial dependency

- AB→C
- AB is the full key
- C is the Non Prime attribute
- Non Prime attribute C is dependent on full key
- So full dependency
- B→D
- B is the partial Key(it is a part of key)
- Non Prime attribute D is dependent on partial key
- So partial dependency

- R(<u>A</u>,<u>B</u>,C,D)



R1(<u>A</u>,<u>B</u>,C)

R2(<u>B</u>,D)

- Now both R1 and R2 are in second Normal form

- Consider the relation schema of the relation schedule shown below. Find the highest normal form?Transform it to next highest form
- Schedule(Studentid,ClassNo,Student Name,Student Major,Class Time, Room, Instuctor)
- Student id -→student Name
- Student id -→student Major
- Class No→Class time
- Class No →Room
- Class No→Instructor

Here student id, classno combination is the key

- S1(Studentid, Student Name,Student Major,)
- S2(ClassNo,class time, Room, Instuctor)
- S3(Studentid, ClassNo)

We must ensure that all decomposition is reversible, that means when we take natural join of the decomposed relation , orginal relation should be obtained . For this we added the relation S3

- R(A,B,C,D,E) WITH A→D, C→E below.Find the highest normal form?Transform it to next highest form

- R(A,B,C,D,E) WITH A→D, C→E below.Find the highest normal form?Transform it to next highest form
- Highest normal form-1NF
- R1(A,D)
- R2(C,E)
- R3(A,C,B)

- R(A,B,C,D,E) WITH A-->D, B→E .Find the highest normal form?Transform it to next highest form

Highest Normal form -1NF

Conversion to 2NF

R1(A,D,B,E)

R2(A,C)

- R(A,B,C,D,E) WITH A-->B,B→D, A→E.Find the highest normal form?Transform it to next highest form
- {AC}+={A,B,C,D,E} Since {AC}+ contains all the attributes of the relation R , AC is the key
- AC is the Candidate key
- Highest normal form-1NF

Conversion to 2NF

- R1(A,B,D,E)
- R2(A,C)

- R(A,B,C,D,E) WITH A-->D,C→E. Find the highest normal form?Transform it to next highest form
- {ABC}+={A,B,C,D,E}

Highest normal form-1NF

Partial dependency present in both FD

2NF decomposition

1.R1(A,D)

2.R2(C,E)

3.R3(A,B,C)