# 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

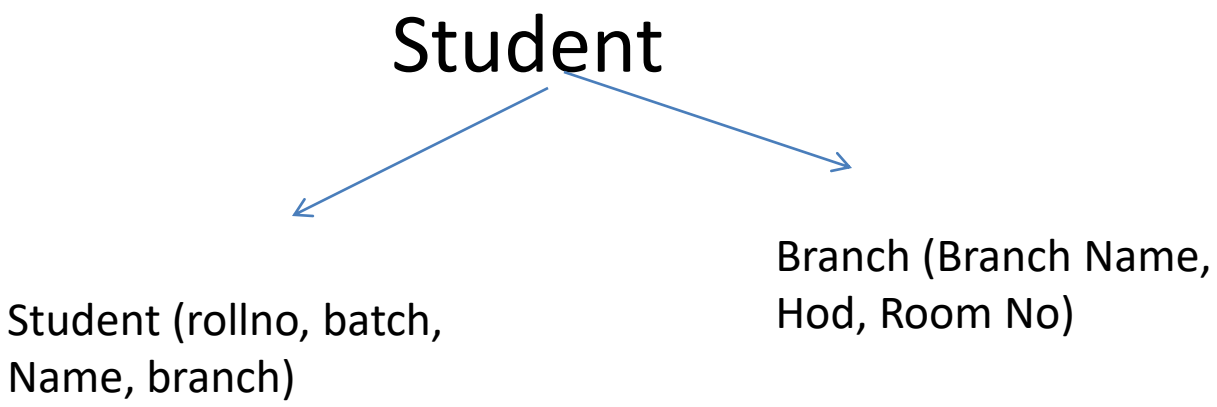
| Roll No | Batch | Name | Branch | Hod | Hod_room No |
|---|---|---|---|---|---|
| 20 | A | ASHLEY | CSE | Manilal | 202 |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

# 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 |

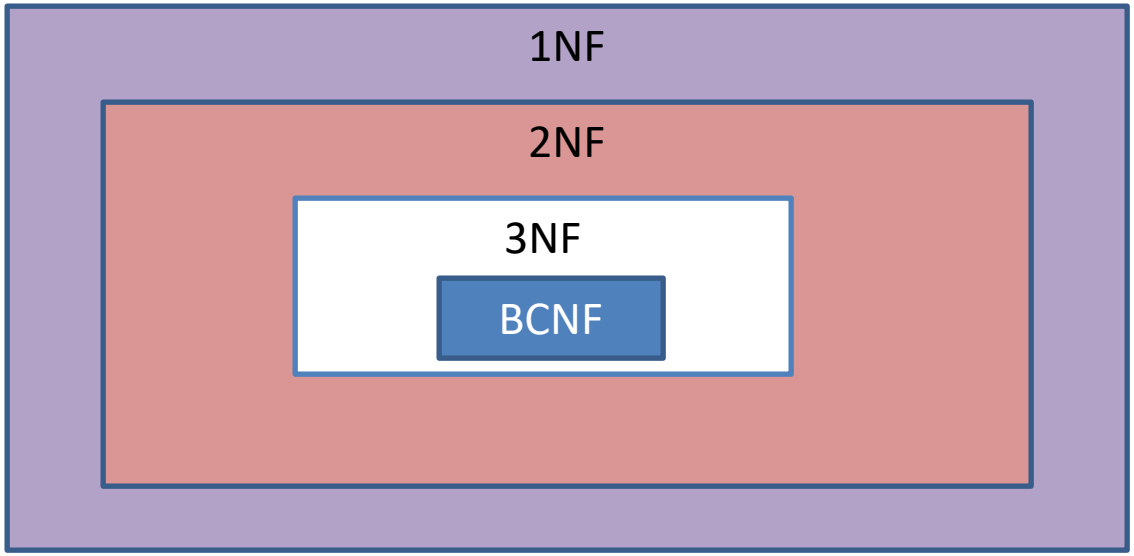
# Branch Table After Normalization

| 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

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

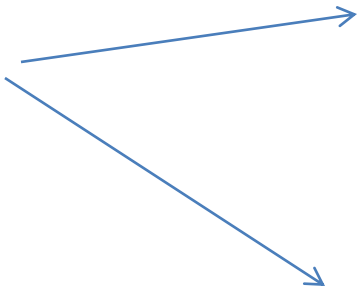
- Prime attributes={A,B}
- Non Prime attributes={C,D}

- 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
- 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

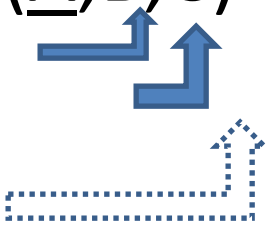
1.R1(A,D)

2.R2(C,E)

3.R3(A,B,C)

# Third Normal Form

- The relation should be in 2NF
- No non prime attribute is transitively dependent on key
- R(A,B,C) WITH FDs A→B,B→C



- A----→C can be inferred transitively

Conversion to 3NF

1.R1(A,B)

2.R2(B,C)

- R(X,Y,Z,W) with FDS X→Y, Y→W. Given that X is the candidate key. Find the highest normal form. Find the next highest NF

  Since all attributes are atomic , R is in 1NF

  Since all there is no partial dependency , R is in 2NF

X-----→W can be infered transitively,the relation is not in 3NF

- Highest Normal form is 2NF

Conversion to 3NF

R1(X,Y,Z)

R2(Y,W)

- Consider the relation R(A,B,C,D,E) and the set F={AB→CE,E→AB,C→D}.What is the highest normal form of this relation?
- AB→C
- AB→E
- E→AB
- C→D

# F={AB→CE,E→AB,C→D}. AB & E are Candidate keys

{AB}+= {A,B,C,D,E}

{E}+= {A,B,C,D,E}

Both {AB} and {E} are candidate keys

- AB→CE is full dependency
- E→AB is full dependency
- C→ D is full dependency , Since both {C, D} are nonprime attribute

Is there any transitive dependency????????

- Yes AB→C, C→D
- AB----→ D is a trasitive dependency . So the relation is not in 3NF
- Highest Normal form is 2NF

# F={AB→CE,E→AB,C→D}. AB & E are Candidate keys

- Decomposition to 3NF

R1(A,B,C,E) with {AB} and {E} are Candidate keys

R2(C,D}

# Example

- Consider the relation R(X,Y,Z,W) AND a set F={Y→W,W→Y,XY→Z} What are the candidate keys of the relation ? What is the highest normal form of the relation
- KEYS-={XY, XW}
- {xw}+={xwyz}
- Superkeys={xy, xw,xyw,xyz,xwz}
- Y→W
- W→Y

# Y→W,W→Y,XY→Z

- Two candidate keys {XY} and {WX}

Prime attributes={X,Y,W}

Non Prime attributes={Z}

All dependencies are full

Y→W and W→Y are dependencies between prime attributes

XY→Z is Full Dependency

There is no transitive dependency.

So the highest normal form is 3NF

# BCNF(Boyce Codd Normal Form)

- The Relation R is in 1NF (3NF)
- If X→A , X must be a superkey or X→A is a trivial dependency.

- AB→A is reflexive
- AB→B is reflexive
- AB→AB is reflexive

Consider the previous example

R is not in BCNF W→Y and Y→W

# Example 1

- EMP_Proj(SSN,Pno,Hours,Ename,Pname,Ploc)

With FDs SSN,Pno→Hours

SSN→Ename

Pno→Pname,Ploc. Find the higest normal form? Find a BCNF decomposition of the above relation

SSN,Pno→Hours SSN→Ename

Pno→Pname,Ploc

- Key ={SSN,Pno}
- Prime attributes={Ssn,Pno}
- Non Prime attributes={Hours,Ename,Pname,Ploc}
- SSN→ Ename is partial dependency
- Pno→{Pname, Ploc } is partial dependency
- So highest normal form is 1NF

SSN,Pno→Hours SSN→Ename

Pno→Pname,Ploc

- 2 NF decomposition
- R1(SSN,Pno,Hours)
- R2(SSN, Ename)
- R3(Pno,Pname,Ploc)

- What about 3NF????????

SSN,Pno→Hours SSN→Ename

Pno→Pname,Ploc

- 2 NF decomposition
- R1(SSN,Pno,Hours)
- R2(SSN, Ename)
- R3(Pno,Pname,Ploc)
- Since no relation contain transitive dependency , R1 ,R2 and R3 are in 3NF
- What about BCNF????????

SSN,Pno→Hours SSN→Ename

Pno→Pname,Ploc

- 3 NF decomposition
- R1(SSN,Pno,Hours)
- R2(SSN, Ename)
- R3(Pno,Pname,Ploc)

left hand side of all FDs are candidate keys,

All relations are in BCNF

# Example 2

- EmpDep(Ename,SSn,Bdate,address,Dno,dname,Dmgr_ssN) Here address is atomic. Given FDs are
- SSN→{Ename,Bdate,address,Dno}
- Dno→{Dname,Dmgr_SSN}

Find the highest normal form? Find a BCNF decomposition

SSN→{Ename,Bdate,address,Dno}
Dno→{Dname,Dmgr_SSN}

- Prime attribute={SSN}
- Non-Prime attributes={Ename,Bdate,address,Dno,dname,Dmgr_ssN}
- Since the there is no partial dependency,
- Relation is in 2NF
- (Every single attribute candidate key is in 2NF by default)
- What about 3NF??????????

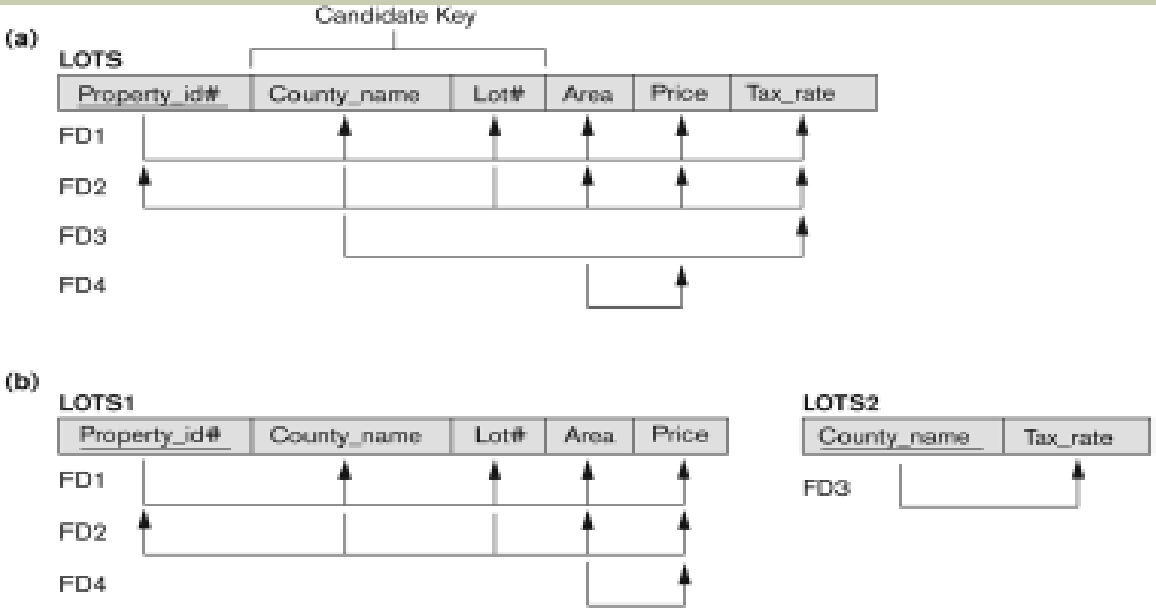
SSN→{Ename,Bdate,address,Dno}
Dno→{Dname,Dmgr_SSN}

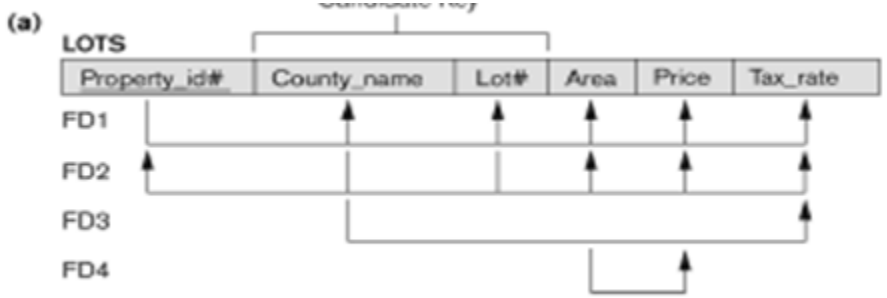
- SSN is transitively dependent on Dname and Dmgr_SSN
- So the Relation is not in 3NF
- 3NF Decomposition
- R1{SSN,Ename,Bdate,address,Dno}
  R2{Dno,Dname,Dmgr_SSN}
- Is this relation is in BCNF???????

SSN→{Ename,Bdate,address,Dno}

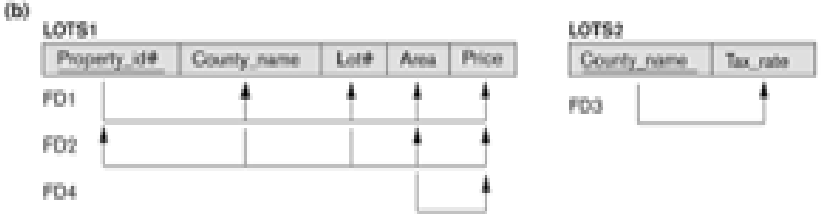
Dno→{Dname,Dmgr_SSN}

- Relation is in BCNF


(a)
Candidate Key
LOTS
Property_id#
County_name
Lot#
Area
Price
Tax_rate
FD1
FD2
FD3
FD4
(b)
LOTS1
Property_id#
County_name
Lot#
Area
Price
FD1
FD2
FD4
LOTS2
County_name
Tax_rate
FD3


(a)
LOTS
Property_id#
County_name
Lot#
Area
Price
Tax_rate
FD1
FD2
FD3
FD4

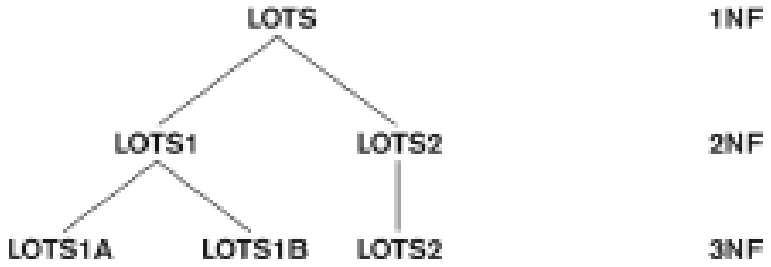
(b)


LOTS1
Property_id#
County_name
Lot#
Area
Price
FD1
FD2
FD4
LOTS2
County_name
Tax_rate
FD3

(c)


LOTS1A
Property_id#
County_name
Lot#
Area
FD1
FD2
LOTS1B
Area
Price
FD4

(d)


LOTS
LOTS1
LOTS2
LOTS1A
LOTS1B
LOTS2
1NF
2NF
3NF