Database Management Systems

Vijaya Saradhi

IIT Guwahati

Thu, 30th Jan 2020

Normalization - 04

An example table

Proj-ID	Proj-Name	Proj-Mgr-ID	Emp-ID	Emp-Name	Emp-Dpt	Emp-Hrly-Rate	Total-Hrs
100	E-commerce	789487453	123423479 980808980 234809000 542298973	Heydary Jones Alexander Johnson	MIS TechSupport TechSupport TechDoc	65 45 35 30	10 6 6 12
110	Distance-Ed	820972445	432329700 689231199 712093093	Mantle Richardson Howard	MIS TechSupport TechDoc	50 35 30	5 12 8
120	Cyber	980212343	834920043 380802233 553208932 123423479	Lopez Harrison Olivier Heydary	Engineering TechSupport TechDoc MIS	80 35 30 65	4 11 12 07
130	Nitts	550227043	340783453	Shaw	MIS	65	07

Fig. 5-2. The PROJECT table.

Normalization - 05

Issues with the table

- Some attributes have multiple values
- Example: Emp-ID, Emp-Name, Emp-Dpt, Emp-Hrly-Rate & Total-Hrs
- Such a relation is known to be in unnormalized form
- To represent this table as relation and it is necessary to normalize the table.
- Need to convert the table into First Normal Form

1NF - 01

Definition

A relation r(R) is in first normal form if and only if every entry of the relation has at most a single value

1NF - 01

Definition

A relation r(R) is in first normal form if and only all its attribtues are based upon a simple domain

1NF - 01

Definition

Flatten the table by (i) removing repeating groups, (ii) filling the missing entries with copies of corresponding nonrepeating attributes

1NF - 02

An example table

Proj-ID	Proj-Name	Proj-Mgr-ID	Emp-ID	Emp-Name	Emp-Dpt	Emp-Hrly-Rate	Total-Hrs
100	E-commerce	789487453	123423479	Heydary	MIS	65	10
100	E-commerce	789487453	980808980	Jones	TechSupport	45	6
100	E-commerce	789487453	234809000	Alexander	TechSupport	35	6
100	E-commerce	789487453	542298973	Johnson	TechDoc	30	12
110	Distance-Ed	820972445	432329700	Mantle	MIS	50	5
110	Distance-Ed	820972445	689231199	Richardson	TechSupport	35	12
110	Distance-Ed	820972445	712093093	Howard	TechDoc	30	8
120	Cyber	980212343	834920043	Lopez	Engineering	80	4
120	Cyber	980212343	380802233	Harrison	TechSupport	35	11
120	Cyber	980212343	553208932	Olivier	TechDoc	30	12
120	Cyber	789487453	123423479	Heydary	MIS	65	10
130	Nitts	550227043	340783453	Shaw	Cabling	40	27

1NF - 03

Not a relation

- The above flattened table is not a relation
- It has no (primay) key
- It has redundancy

Decomposing relation

Process

- Given a relation R($A_1, A_2, \cdots, A_n$)
- We may decompose R into two relations S and T
- S($B_1, B_2, \cdots, B_m$)
- T($C_1, C_2, \cdots, C_k$)
- Such that $\{A_1, A_2, \cdots, A_n\} = \{B_1, B_2, \cdots, B_m\} \cup \{C_1, C_2, \cdots, C_k\}$
- Tuples in R with attribute values ($B_1, B_2, \cdots, B_m$) are placed in S
- Tuples in R with attribute values ($C_1, C_2, \cdots, C_k$) are placed in T
- Duplicates are not included

Example - decomposing relation

R(title, year, length, filmType, studioName, starName)

title	year	length	filmType	studioName	starName
Star wars	1977	124	color	Fox	Carrie Fisher
Star wars	1977	124	color	Fox	Mark Hamill
Star wars	1977	124	color	Fox	Harrison Ford
Mighty Ducks	1991	104	color	Disney	Emilio Estevez
Wayne's world	1992	95	color	Paramount	Data Carvey
Wayne's world	1992	95	color	Paramount	Mike Meyers

S(title, year, length, fileType, studioName)

title	year	length	filmType	studioName
Star wars	1977	124	color	Fox
Mighty Ducks	1991	104	color	Disney
Wayne's world	1992	95	color	Paramount

Example - decomposing relation

R(title, year, length, filmType, studioName, starName)

title	year	length	filmType	studioName	starName
Star wars	1977	124	color	Fox	Carrie Fisher
Star wars	1977	124	color	Fox	Mark Hamill
Star wars	1977	124	color	Fox	Harrison Ford
Mighty Ducks	1991	104	color	Disney	Emilio Estevez
Wayne's world	1992	95	color	Paramount	Data Carvey
Wayne's world	1992	95	color	Paramount	Mike Meyers

T(title, year, starName)

title	year	starName
Star wars	1977	Carrie Fisher
Star wars	1977	Mark Hamill
Star wars	1977	Harrison Ford
Mighty Ducks	1991	Emilio Estevez
Wayne's world	1992	Data Carvey
Wayne's world	1992	Mike Meyers

1NF - 04

Convert table into - 1 NF

- Decompose the original table into two tables
- Such that each table is a relation with a (primary) key

1NF - 05

An example table

Proj-ID	Proj-Name	Proj-Mgr-ID	Emp-ID	Emp-Name	Emp-Dpt	Emp-Hrly-Rate	Total-Hrs
100	E-commerce	789487453	123423479	Heydary	MIS	65	10
100	E-commerce	789487453	980808980	Jones	TechSupport	45	6
100	E-commerce	789487453	234809000	Alexander	TechSupport	35	6
100	E-commerce	789487453	542298973	Johnson	TechDoc	30	12
110	Distance-Ed	820972445	432329700	Mantle	MIS	50	5
110	Distance-Ed	820972445	689231199	Richardson	TechSupport	35	12
110	Distance-Ed	820972445	712093093	Howard	TechDoc	30	8
120	Cyber	980212343	834920043	Lopez	Engineering	80	4
120	Cyber	980212343	380802233	Harrison	TechSupport	35	11
120	Cyber	980212343	553208932	Olivier	TechDoc	30	12
120	Cyber	789487453	123423479	Heydary	MIS	65	10
130	Nitts	550227043	340783453	Shaw	Cabling	40	27

Decomposed table in 1NF

PROJECT

Proj-ID	Proj-Name	Proj-Mgr-ID
100	E-commerce	789487453
110	Distance-Ed	820972445
120	Cyber	980212343
130	Nitts	550227043

1NF - 06

An example table

Proj-ID	Proj-Name	Proj-Mgr-ID	Emp-ID	Emp-Name	Emp-Dpt	Emp-Hrly-Rate	Total-Hrs
100	E-commerce	789487453	123423479	Heydary	MIS	65	10
100	E-commerce	789487453	980808980	Jones	TechSupport	45	6
100	E-commerce	789487453	234809000	Alexander	TechSupport	35	6
100	E-commerce	789487453	542298973	Johnson	TechDoc	30	12
110	Distance-Ed	820972445	432329700	Mantle	MIS	50	5
110	Distance-Ed	820972445	689231199	Richardson	TechSupport	35	12
110	Distance-Ed	820972445	712093093	Howard	TechDoc	30	8
120	Cyber	980212343	834920043	Lopez	Engineering	80	4
120	Cyber	980212343	380802233	Harrison	TechSupport	35	11
120	Cyber	980212343	553208932	Olivier	TechDoc	30	12
120	Cyber	789487453	123423479	Heydary	MIS	65	10
130	Nitts	550227043	340783453	Shaw	Cabling	40	27

Decomposed table in 1NF

PROJECT-EMPLOYEE

Proj-ID	Emp-ID	Emp-Name	Emp-Dpt	Emp-Hrly-Rate	Total-Hrs
100	123423479	Heydary	MIS	65	10
100	980808980	Jones	TechSupport	45	6
100	234809000	Alexander	TechSupport	45	6
100	542298973	Johnson	TechDoc	30	12
110	432329700	Mantle	MIS	65	5
110	689231199	Richardson	TechSupport	45	12
110	712093093	Howard	TechDoc	30	8
120	834920043	Lopez	Engineering	80	4
120	380802233	Harrison	TechSupport	45	11
120	553208932	Olivier	TechDoc	30	12
120	123423479	Heydary	MIS	65	10
130	340783453	Shaw	Cabling	40	27

1NF - 07

Anomalies in 1NF Tables

- Insertion into `project-employee` table
  - The FD `EMP-ID` $\rightarrow$ `EMP-DPT`
  - For a new employee, when department is not assigned yet, inserting this into `project-employee` result in issues
- Deletion from `project-employee` table
  - When deleting employee, we loose employee-project information
  - We also loose the employee department information
- Update Anomaly We may miss updating one redundant row

Partial Dependencies

Discussion

- Let X, Y be set of attributes such that X $\subset$ R and Y $\subset$ R
- We also are given the FD: X $\rightarrow$ Y
- We say Y is fully dependent on X if and only if there no proper subset W of X such that W $\rightarrow$ Y
- We say Y is partially dependent on X if there a proper subset W of X such that W $\rightarrow$ Y
- Exmaple: $A_1 A_2 \cdots A_m \rightarrow B_1 B_2 \cdots B_n$
- Let $AcA_d \cdots A_k \subset A_1 A_2 \cdots A_m$ such that
- $AcA_d \cdots A_k \rightarrow B_1 B_2 \cdots B_n$

2NF - No partial dependency

Definition

A relation r(R) is in 2NF if and only if

- r(R) is already in 1NF
- No non-prime attribute is partially dependent on any key non-prime attribute

2NF - No partial dependency

Definition Continued

- Let X be key; Y be (set of) non-prime attribute(s)
- And let $W \subset X$; $W \not\rightarrow Y$
- That is no proper subset of the key fully determines

2NF - No partial dependency

Definition

A relation r(R) is in 2NF if and only if

- r(R) is already in 1NF
- Each non-prime attribute is fully dependent on every key

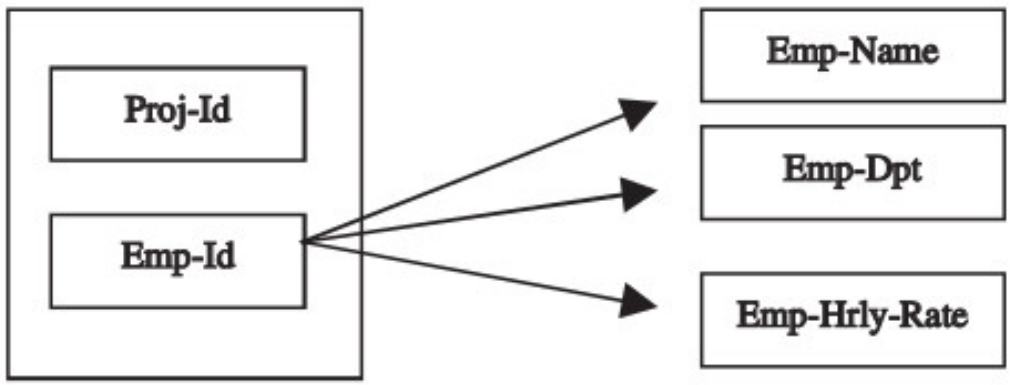
Non-prime attributes

Discussion

- To find non-prime attribute of R we need to identify all prime attributes of R
- That is we need to identify all possible keys of R
- non-prime attributes(R) = all attributes(R) - prime attribtues(R)

2NF - 01

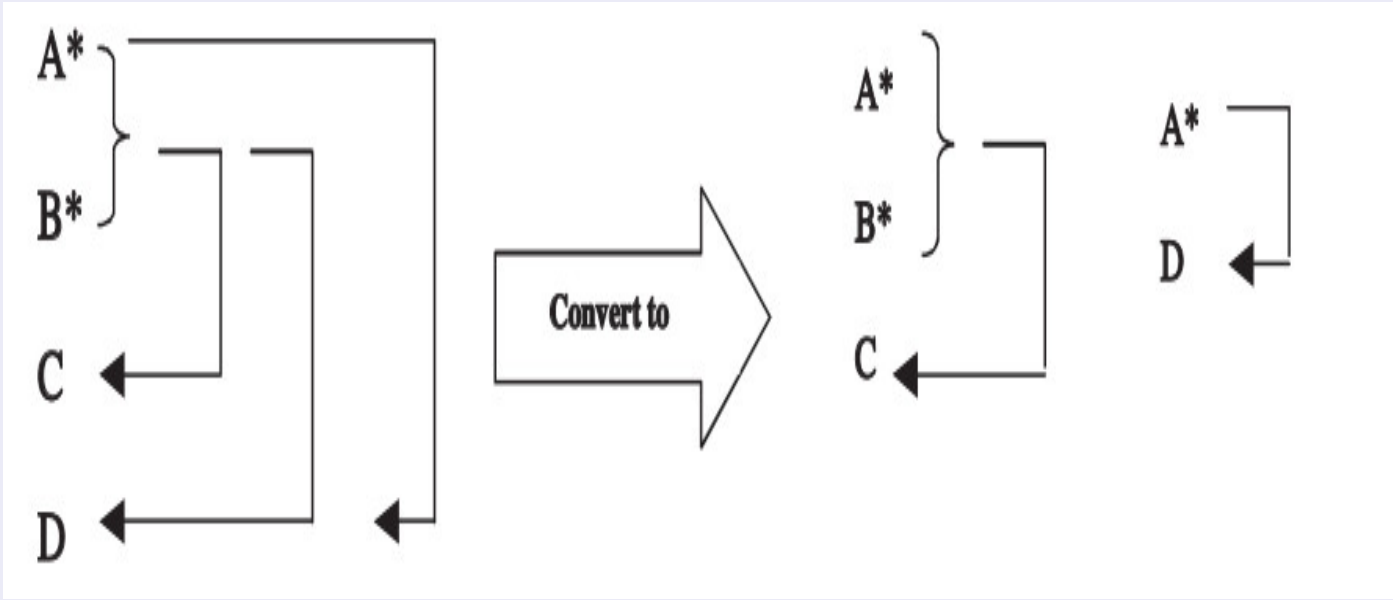
An example table



- Prime attribute(s) X = {Proj-Id, Emp-Id}
- no-prime attribute(s) Y = {Emp-Name, Emp-Dpt, Emp-Hrly-Rate}
- W = {Emp-Id} $\subset$ X
- W $\rightarrow$ Y; Emp-Id $\rightarrow$ {Emp-Name, Emp-Dpt, Emp-Hrly-Rate}
- non-prime attribute Y partially depend on X

2NF - 02

Decomposing Relation into 2NF



- Prime attribute(s) X = {A*, B*}
- no-prime attribute(s) Y = {C, D}
- X → Y
- W = {A*} ⊂ X
- W → Y; A* → {D}
- non-prime attribute Y partially depend on X

2NF - 03

Decomposing Relation into 2NF

- Original table `PROJECT-EMPLOYEE(Proj-ID, Emp-ID, Emp-Name, Emp-Dpt, Emp-Hrly-Rate)` is split into two tables given the above partial dependency FD
- Place partial dependency FD into a new table
- `HOURS-ASSIGNED(Proj-ID, Emp-ID, Total-Hours)`
- `EMPLOYEE(Emp-ID, Emp-Name, Emp-Dpt, Emp-Hrly-Rate)`

2NF - 04

Project-Employee Table split into 2 - Hours-Assigned

Proj-ID	Emp-ID	Total-Hrs
100	123423479	10
100	980808980	6
100	234809000	6
100	542298973	12
110	432329700	5
110	689231199	12
110	712093093	8
120	834920043	4
120	380802233	11
120	553208932	12
120	123423479	10
130	340783453	27

2NF - 04

Project-Employee Table split into 2 - Employee

Emp-ID	Emp-Name	Emp-Dpt	Emp-Hrly-Rate
123423479	Heydary	MIS	65
980808980	Jones	TechSupport	45
234809000	Alexander	TechSupport	45
542298973	Johnson	TechDoc	30
432329700	Mantle	MIS	65
689231199	Richardson	TechSupport	45
712093093	Howard	TechDoc	30
834920043	Leopez	Engineering	80
380802233	Harrison	TechSupport	45
553208932	Olivier	TechDoc	30
123423479	Heydary	MIS	65
340783453	Shaw	Cabling	40

2NF - 05

Data Anomalies in 2NF

- Relations in 2NF are still subject to data anomalies
- Consider the FD: `Emp-Dpt` $\rightarrow$ `Emp-Hrly-Rate`
- This fact was not considered perviously
- Insertion Anomaly Rate to be charged to be set before hand to employees of new department
- Deletion Anomaly The only employee in the department left there by loosing the information about that department charges
- Update Anomalies Employees of the same department working on different projects data need to be updated

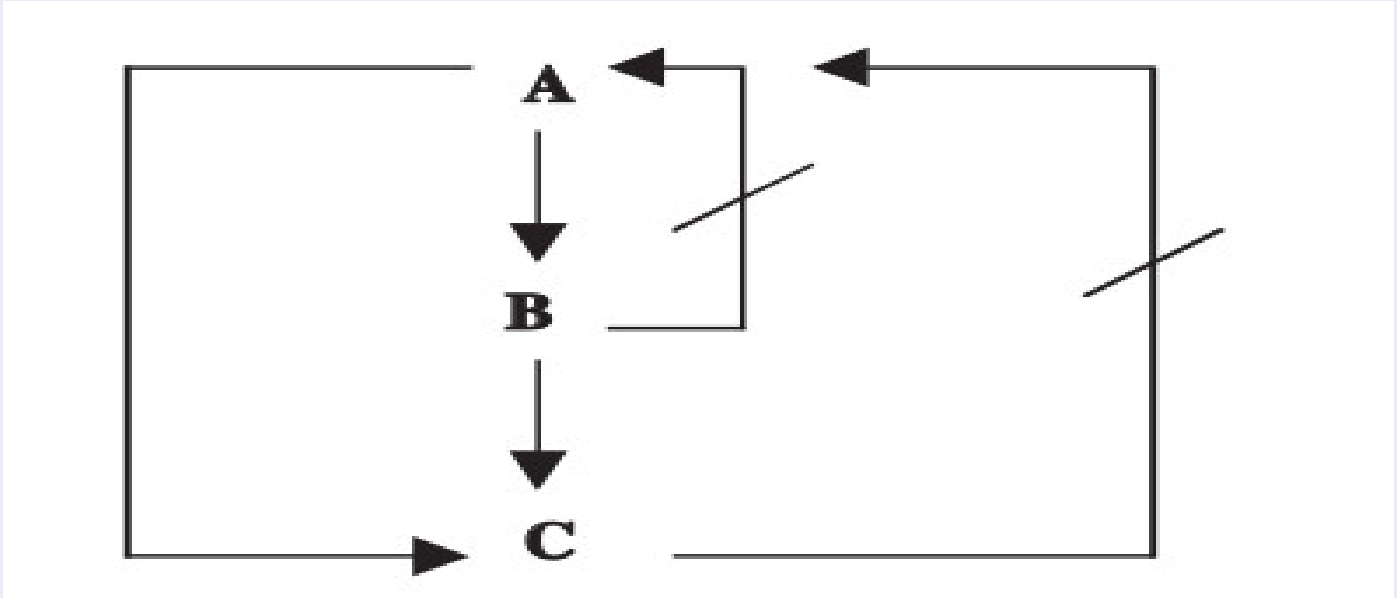
Transitive Dependencies - 01

Discussion

- Let A, B, C be set of attribute of r(R)
- Assume FDs $\{ A \rightarrow B, B \rightarrow C, A \rightarrow C, B \not\rightarrow A, C \not\rightarrow A \}$
- Then C is transitively dependent on A

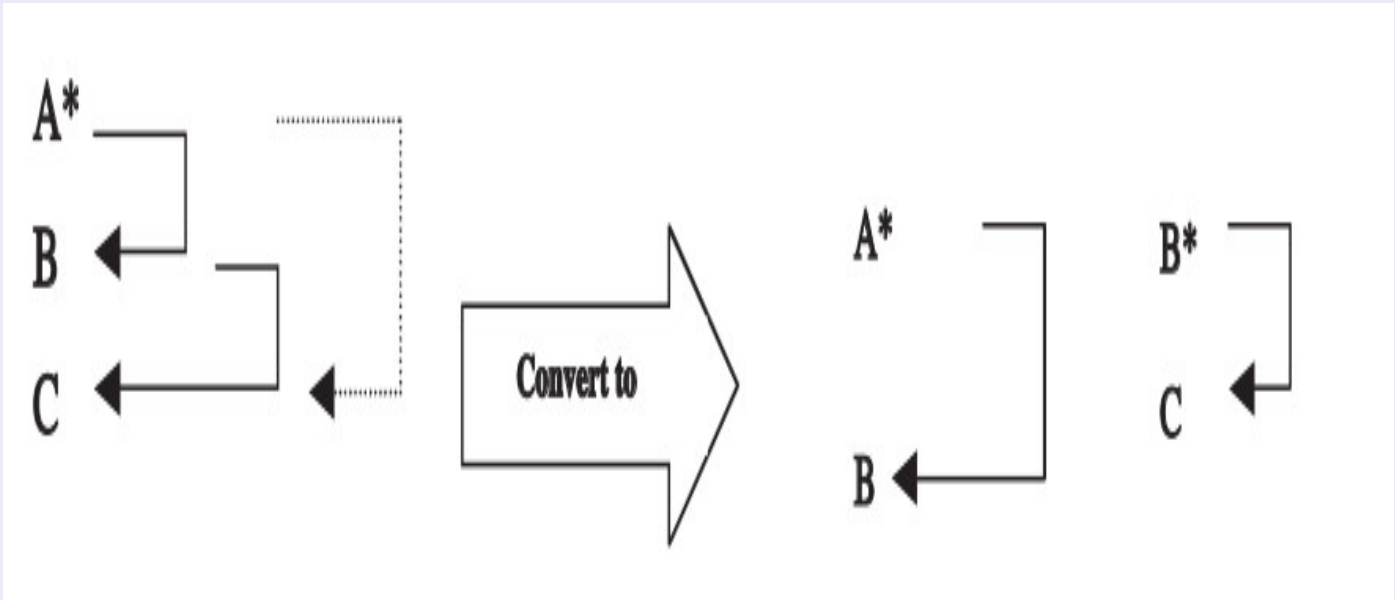
Transitive Dependencies - 02

Transitive Dependency - Diagram



3NF - 01

Decomposing Relation into 3NF



3NF - 02

Decompose Empoyee into two tables to get to 3NF

- EMPLOYEE(Emp-ID, Emp-Name, Emp-Dpt)
- CHARGES(Emp-Dpt, Emp-Hrly-Rate)

3NF - 03

Employee Table split into 2 - Employee

Emp-ID	Emp-Name	Emp-Dpt
123423479	Heydary	MIS
980808980	Jones	TechSupport
234809000	Alexander	TechSupport
542298973	Johnson	TechDoc
432329700	Mantle	MIS
689231199	Richardson	TechSupport
712093093	Howard	TechDoc
834920043	Leopez	Engineering
380802233	Harrison	TechSupport
553208932	Olivier	TechDoc
123423479	Heydary	MIS
340783453	Shaw	Cabling

3NF - 04

Employee Table split into 2 - Charges

Emp-Dpt	Emp-Hrly-Rate
MIS	65
TechSupport	45
TechDoc	30
MIS	65
Engineering	80
Cabling	40

Given table to 1NF, 2NF and 3NF

Given table

PROJECT(Proj-ID, Proj-Name, Proj-Mgr-ID, Emp-ID, Emp-Name, Emp-Dpt, Emp-Hrly-Rate, Total-Hrs)

Given table to 1NF, 2NF and 3NF

Convert to 1NF

PROJECT(Proj-ID, Proj-Name, Proj-Mgr-ID, Emp-ID, Emp-Name, Emp-Dpt, Emp-Hrly-Rate, Total-Hrs) is split into two tables

- PROJECT(Proj-ID, Proj-Name, Proj-Mgr-ID)
- PROJECT-EMPLOYEE(Proj-ID, Emp-ID, Emp-Name, Emp-Dpt, Emp-Hrly-Rate, Total-Hrs)

Given table to 1NF, 2NF and 3NF

Convert to 2NF

PROJECT-EMPLOYEE(Proj-ID, Emp-ID, Emp-Name, Emp-Dpt, Emp-Hrly-Rate, Total-Hrs) is split into two tables due to the partial dependency FD: Emp-ID $\rightarrow$ {Emp-Name, Emp-Dpt, Emp-Hrly-Rate, Total-Hrs}

- PROJECT(Proj-ID, Proj-Name, Proj-Mgr-ID)
- HOURS-ASSIGNED(Proj-ID, Emp-ID, Total-Hours)
- EMPLOYEE(Emp-ID, Emp-Name, Emp-Dpt, Emp-Hrly-Rate)

Given table to 1NF, 2NF and 3NF

Convert to 3NF

EMPLOYEE(Emp-ID, Emp-Name, Emp-Dpt, Emp-Hrly-Rate) is split into two tables due to the transitive dependency FD Emp-Dpt $\rightarrow$ Emp-Hrly-Rate

- PROJECT(Proj-ID, Proj-Name, Proj-Mgr-ID)
- HOURS-ASSIGNED(Proj-ID, Emp-ID, Total-Hours)
- EMPLOYEE(Emp-ID, Emp-Name, Emp-Dpt)
- CHARGES(Emp-Dpt, Emp-Hrly-Rate)

Hospital Form Example - 01

Post Surgery Form

Hospital Number:	H17	Hospital Name:	St Vincent's	Operation Number:	48		
Hospital Category:	P			Contact at Hospital:	Fred Fleming		
Operation Name:	Heart Transplant			Operation Code:	7A	Procedure Group:	Transplant
Surgeon Number:	S15	Surgeon Specialty:	Cardiology	Total Drug Cost:	$75.50		

Drug Code	Full Name of Drug	Manufacturer	Method of Admin.	Cost of Dose ($)	Number of Doses
MAX 150mg	Maxicillin	ABC Pharmaceuticals	ORAL	$3.50	15
MIN 500mg	Minicillin	Silver Bullet Drug Co.	IV	$1.00	20
MIN 250mg	Minicillin	Silver Bullet Drug Co.	ORAL	$0.30	10

Hospital Form Example - 02

Observations

- One form is filled for each operation
- Each hospital is given a unique number
- All hospitals are prefixed by "H"
- Operations numbers are assigned sequentially
- Hospital categories {Teaching, Public, Private}
- Operation Code - standard international code is adopted
- Surgeon name - recorded in terms of surgeon number
- Total drug cost
- Drug Code, name, method of administration

Hospital Form Example - 03

Initial Data Model

```
OPERATIONS(Hospital Number, Operation Number,
  Hospital Name, Hospital Category, Contact Person,
  Operation Name, Operation Code, Procedure Group,
  Surgeon Number, Surgeon Speciality, Total Drug Cost,
  Drug Code 1, Drug Name 1, Manufacturer 1, Method of Administration   1, Dose Cost 1, Number of Doses 1,
  Drug Code 2, Drug Name 2, Manufacturer 2, Method of Administration   2, Dose Cost 2, Number of Doses 2,
  Drug Code 3, Drug Name 3, Manufacturer 3, Method of Administration   3, Dose Cost 3, Number of Doses 3,
  Drug Code 4, Drug Name 4, Manufacturer 4, Method of Administration   4, Dose Cost 4, Number of Doses 4
)
```

Hospital Form Example - 04

One Fact Per Column

- `Drug Code` Multivalued - holds short form of name of drug and dosage size
- `Dosage Size` holds two facts numeric size and units
- Three facts are recorded into one column of `Durg Code`
- `Hospital Category`: Multivalued - Private, Public, teaching or non-teaching
- 1NF: tells us not to include multivalued attributes
- These attribtues to be split from the OPERATION table

Hospital Form Example - 04

Hidden Data

- Say hospital files the operations forms as per a sequence
- This information that hospital files the forms in sequence is not explicitly recorded in the form
- It must be included in the OPERATION table
- If the forms are color coded: red for EMERGENCY operations, blue for elective operations etc. These facts must be recorded in the OPERATION table

Hospital Form Example - 04

Derivable Data, Primary Key

- Total Cost a deriable column. Its presence must be argued
- Primary key: (Hospital Number, Operation Number)
- Split the repeating columns/multivalued columns from OPERATION table

Hospital Form Example - 05

Removing Repeating Groups

- Put the data in table form by identifying and eliminating repeating groups
- Remove each separate set of repeating group into a new table
- In the new table, each of occurance of the group becomes a row
- Include key of the original table into each new table
- Name each of the new table
- Identify the primary key for each of the new table

Hospital Form Example - 06

Repeating Groups in OPERATION table

- `Drug Short Name`
- `Drug Name`
- `Manufacturer`
- `Size of Dose`
- `Unit of Measure`
- `Method of Administration`
- `Dose Cost`
- `Number of Doses`
- Place these into a new table named DRUG ADMINISTRATION
- With key (Hospital Name, Operation Number, Drug Short Name, Size of Dose, Unit of Measure, Method of Administration)
- Rest are non-repeating attributes are retained in the original table

Hospital Form Example - 07

Convert to 1NF

- OPERATIONS(Hospital Number, Operation Number, Hospital Name, Hospital Category, Teaching Status, Contact Person, Operation Name, Operation Code, Procedure Group, Surgeon Number, Surgeon Speciality, Total Drug Cost)
- HOSPITAL(Hospital Number, Operation Number, Drug Name, Size of Dose, Unit of Measure, Method of Administ
Dose Cost, Number of Doses)

Hospital Form Example - 08

Identify & Eliminate Redundancy

- Hospital Number $\rightarrow$ { Hospital Name, Hospital Type, Teaching Status, Contact Person }
- Remove this FD from the OPERATION table resulting in three tables
- OPERATIONS(Hospital Number, Operation Number, Hospital Name, Hospital Category, Contact Person, Operation Name, Operation Code, Procedure Group, Surgeon Number, Surgeon Speciality, Total Drug Cost)
- HOSPITAL(Hospital Number, Hospital Name, Hospital Category, Contact Person)
- DRUG ADMINISTRATION(Hospital Number, Operation Number, Drug Sho Dose Cost, Number of Doses)

Hospital Form Example - 09

Identify FDs

- Identify FDs in a given relation and place them in separate table
- Remove the above FD columns from the original table
- FD 2: Surgeon: (Hospital Number, Operation Number) → Surgeon Name
- FD 3: Operation Type: Operatoin Code → (Operation Name, Procedure Group)
- FD 4: Standard Drug Dosage (Drug Short Name, Size of Dose, Unit of Measure, Method of Administration) → Standard Dose Cost
- FD 5: Drug: Drug Short Name → (Drug Name, Manufacturer)

Hospital Form Example - 10

Summary - Original Table

```
OPERATIONS(Hospital Number, Operation Number,
  Hospital Name, Hospital Category, Contact Person,
  Operation Name, Operation Code, Procedure Group,
  Surgeon Number, Surgeon Speciality, Total Drug Cost,
  Drug Code 1, Drug Name 1, Manufacturer 1, Method of Administration   1, Dose Cost 1, Number of Doses 1,
  Drug Code 2, Drug Name 2, Manufacturer 2, Method of Administration   2, Dose Cost 2, Number of Doses 2,
  Drug Code 3, Drug Name 3, Manufacturer 3, Method of Administration   3, Dose Cost 3, Number of Doses 3,
  Drug Code 4, Drug Name 4, Manufacturer 4, Method of Administration   4, Dose Cost 4, Number of Doses 4
)
```

Tables after eliminating redundancy

- OPERATION(Hospital Number, Operation Number, Operation Code, Surgeon Name)
- SURGEON((Hospital Number, Operation Number), Surgeon Name)
- OPERATION TYPE(Operatoin Code, (Operation Name, Procedure Group))
- STANDARD DRUG DOSAGE((Drug Short Name, Size of Dose, Unit of Measure, Method of Administration) , Standard Dose Cost)
- DRUG(Drug Short Name , (Drug Name, Manufacturer))
- HOSPITAL(Hospital Number, Hospital Name, Hospital Type, Teaching Status, Contact Person)
- DRUG ADMINISTRA-
  TION(Hospital Number, Operation Number, Drug Name, Drug Short Name, Size of Dose, Unit of Measure, Metho
  Dose Cost, Number of Doses)