

# Normalization



# Normalization: What and why?

- A systematic approach to:
  - Restructure tables to eliminate data redundancy (duplicated data)
  - Eliminate Update Anomalies
    - Insertion
    - Deletion
    - Modification
- Bottom up approach
  - Starts from the data
  - Ends with multiple related tables, with a minimum of duplicated data

# Update Anomaly

## Employees' Skills

| Employee ID | Employee Address   | Skill           |
|-------------|--------------------|-----------------|
| 426         | 87 Sycamore Grove  | Typing          |
| 426         | 87 Sycamore Grove  | Shorthand       |
| 519         | 94 Chestnut Street | Public Speaking |
| 519         | 96 Walnut Avenue   | Carpentry       |

→ Employee 519 is shown as having different addresses on different records.

# Insertion Anomaly

## Faculty and Their Courses

| Faculty ID | Faculty Name   | Faculty Hire Date | Course Code |
|------------|----------------|-------------------|-------------|
| 389        | Dr. Giddens    | 10-Feb-1985       | ENG-206     |
| 407        | Dr. Saperstein | 19-Apr-1999       | CMP-101     |
| 407        | Dr. Saperstein | 19-Apr-1999       | CMP-201     |

|     |             |             |   |
|-----|-------------|-------------|---|
| 424 | Dr. Newsome | 29-Mar-2007 | ? |
|-----|-------------|-------------|---|

→ Until the new faculty member, Dr. Newsome, is assigned to teach at least one course, his or her details cannot be recorded.

# Insertion Anomaly

Faculty and Their Courses

| Faculty ID | Faculty Name   | Faculty Hire Date | Course Code |
|------------|----------------|-------------------|-------------|
| 389        | Dr. Giddens    | 10-Feb-1985       | ENG-206     |
| 407        | Dr. Saperstein | 19-Apr-1999       | CMP-101     |
| 407        | Dr. Saperstein | 19-Apr-1999       | CMP-201     |

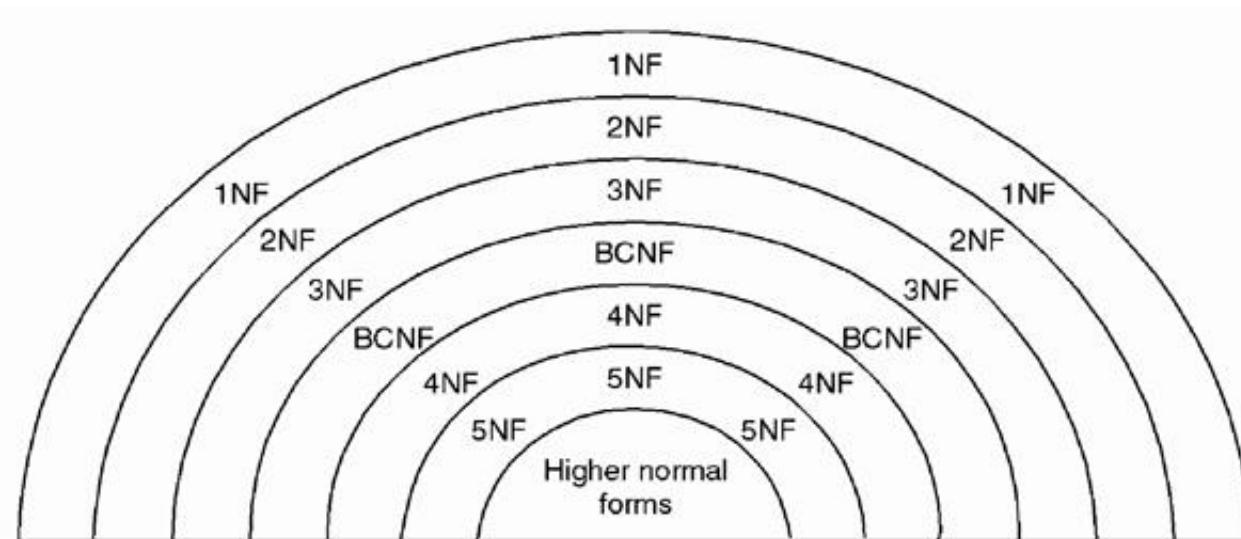
DELETE

→ All information about Dr. Giddens is lost if he or she temporarily ceases to be assigned to any courses.

# Normal Forms

- UNF: Unnormalized Form
- NF1 – First Normal Form
- NF2 – Second Normal Form
- NF3 – Third Normal Form
- BCNF – Boyce-Codd Normal Form
- *NF4 and NF5*

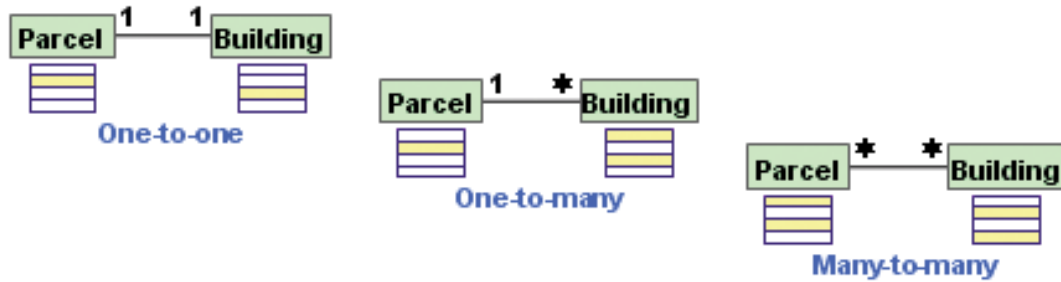
→ Practical use of NF's go to BCNF, or at most NF4



# Guidelines

- Anomalies that cause redundant work to be done during insertion into and modification of a relation, and that may cause accidental loss of information during a deletion from a relation
- Waste of storage space due to NULLs and the difficulty of performing selections, aggregation operations, and joins due to NULL values
- Generation of invalid and spurious data during joins on base relations with matched attributes that may not represent a proper (foreign key, primary key) relationship

# Cardinalities



- 1 - \* , one to many
- \* - 1 , many to one
- \* - \* , many to many
- 1 - 1 , one to one
- 0..1 - 1 , zero/one to many



# UNF: Unnormalized Form

→ Data idea on loan from YouTube

| Customer Name    | Item              | Shipping Address            | Newsletter                  | Supplier  | Supplier Phone | Price |
|------------------|-------------------|-----------------------------|-----------------------------|-----------|----------------|-------|
| Jens Bergholm    | XBOX One          | Mølletofte 20, 7100 Vejle   | Xbox News                   | Microsoft | 12341324       | 1800  |
| Dennis Jørgensen | Playstation 4     | Vinkelvej 167, 7100 Vejle   | Playstation News            | Sony      | 23452345       | 1900  |
| Hans Jensen      | XBOX One, PS Vita | Kongensgade 45, 5000 Odense | Xbox News, Playstation News | Wholesale | 00000000       | 3300  |
| Anne Johansen    | Playstation 4     | Vinkelvej 167, 7100 Vejle   | Playstation News            | Sony      | 23452345       | 1900  |

# NF1 – First Normal Form

- Each column should contain atomic values – entries like “x, y” violate this rule
- A column should contain values that are of the same type
- Each column name must be unique
- Each row must be unique

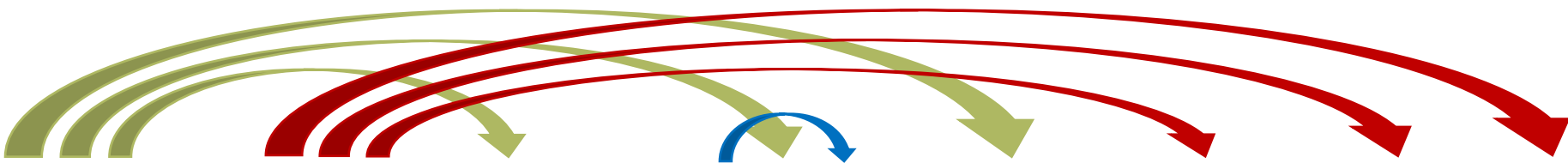
| Customer Name    | Item              | Shipping Address            | Newsletter                  | Supplier  | Supplier Phone | Price |
|------------------|-------------------|-----------------------------|-----------------------------|-----------|----------------|-------|
| Jens Bergholm    | XBOX One          | Mølletofte 20, 7100 Vejle   | Xbox News                   | Microsoft | 12341324       | 1800  |
| Dennis Jørgensen | Playstation 4     | Vinkelvej 167, 7100 Vejle   | Playstation News            | Sony      | 23452345       | 1900  |
| Hans Jensen      | XBOX One, PS Vita | Kongensgade 45, 5000 Odense | Xbox News, Playstation News | Wholesale | 00000000       | 3300  |
| Anne Johansen    | Playstation 4     | Vinkelvej 167, 7100 Vejle   | Playstation News            | Sony      | 23452345       | 1900  |



| ID | Customer Name    | Item          | Shipping Address | Zip  | City   | Newsletter       | Supplier  | Supplier Phone | Price |
|----|------------------|---------------|------------------|------|--------|------------------|-----------|----------------|-------|
| 1  | Jens Bergholm    | XBOX One      | Mølletofte 20    | 7100 | Vejle  | Xbox News        | Microsoft | 12341324       | 1800  |
| 2  | Dennis Jørgensen | Playstation 4 | Vinkelvej 167    | 7100 | Vejle  | Playstation News | Sony      | 23452345       | 1900  |
| 3  | Hans Jensen      | XBOX One      | Kongensgade 45   | 5000 | Odense | Xbox News        | Microsoft | 12341324       | 1800  |
| 3  | Hans Jensen      | PS Vita       | Kongensgade 45   | 5000 | Odense | Playstation News | Sony      | 23452345       | 1500  |
| 4  | Anne Johansen    | Playstation 4 | Vinkelvej 167    | 7100 | Vejle  | Playstation News | Sony      | 23452345       | 1900  |

# Functional Dependencies

- A strong connection between two attributes in a table
  - Denoted as  $A \rightarrow B$
  - A functionally determines B, or B is functionally dependent on A.
  - Customer Name and Item is the **Determinant**.
- Customer Name  $\rightarrow$  {Shipping Address, City, Newsletter}
  - Item  $\rightarrow$  {Supplier, Supplier Phone, Price}
  - Zip  $\rightarrow$  City



| ID | Customer Name    | Item          | Shipping Address | Zip  | City   | Newsletter       | Supplier  | Supplier Phone | Price |
|----|------------------|---------------|------------------|------|--------|------------------|-----------|----------------|-------|
| 1  | Jens Bergholm    | XBOX One      | Mølletofte 20    | 7100 | Vejle  | Xbox News        | Microsoft | 12341324       | 1800  |
| 2  | Dennis Jørgensen | Playstation 4 | Vinkelvej 167    | 7100 | Vejle  | Playstation News | Sony      | 23452345       | 1900  |
| 3  | Hans Jensen      | XBOX One      | Kongensgade 45   | 5000 | Odense | Xbox News        | Microsoft | 12341324       | 1800  |
| 3  | Hans Jensen      | PS Vita       | Kongensgade 45   | 5000 | Odense | Playstation News | Sony      | 23452345       | 1500  |
| 4  | Anne Johansen    | Playstation 4 | Vinkelvej 167    | 7100 | Vejle  | Playstation News | Sony      | 23452345       | 1900  |

# Partial Dependencies

- An Xbox one purchase does not require you to be “Jens Bergholm”
- But there is a dependency, as “Jens Bergholm” purchased this item



| ID | Customer Name    | Item          | Shipping Address | Zip  | City   | Newsletter       | Supplier  | Supplier Phone | Price |
|----|------------------|---------------|------------------|------|--------|------------------|-----------|----------------|-------|
| 1  | Jens Bergholm    | XBOX One      | Mølletoft 20     | 7100 | Vejle  | Xbox News        | Microsoft | 12341324       | 1800  |
| 2  | Dennis Jørgensen | Playstation 4 | Vinkelvej 167    | 7100 | Vejle  | Playstation News | Sony      | 23452345       | 1900  |
| 3  | Hans Jensen      | XBOX One      | Kongensgade 45   | 5000 | Odense | Xbox News        | Microsoft | 12341324       | 1800  |
| 3  | Hans Jensen      | PS Vita       | Kongensgade 45   | 5000 | Odense | Playstation News | Sony      | 23452345       | 1500  |
| 4  | Anne Johansen    | Playstation 4 | Vinkelvej 167    | 7100 | Vejle  | Playstation News | Sony      | 23452345       | 1900  |

# NF2 – Second Normal Form

→ No partial Dependencies

| Customer Table |                  |                  |      |        |                  |
|----------------|------------------|------------------|------|--------|------------------|
| ID             | Customer Name    | Shipping Address | Zip  | City   | Newsletter       |
| 1              | Jens Bergholm    | Mølletoften 20   | 7100 | Vejle  | Xbox News        |
| 2              | Dennis Jørgensen | Vinkelvej 167    | 7100 | Vejle  | Playstation News |
| 3              | Hans Jensen      | Kongensgade 45   | 5000 | Odense | Xbox News        |
| 3              | Hans Jensen      | Kongensgade 45   | 5000 | Odense | Playstation News |
| 4              | Anne Johansen    | Vinkelvej 167    | 7100 | Vejle  | Playstation News |

| Product Table |               |           |                |       |
|---------------|---------------|-----------|----------------|-------|
| ID            | Item          | Supplier  | Supplier Phone | Price |
| 1             | XBOX One      | Microsoft | 12341324       | 1800  |
| 2             | Playstation 4 | Sony      | 23452345       | 1900  |
| 3             | PS Vita       | Sony      | 23452345       | 1500  |

| Order Table |            |
|-------------|------------|
| Customer_Id | Product_Id |
| 1           | 1          |
| 2           | 2          |
| 3           | 1          |
| 3           | 3          |
| 4           | 2          |

# NF3 – Third Normal Form

→ A table is said to be in 3NF, if and only if:

→ That table is in the second normal form (2NF)

→ Every attribute in the table that do not belong to a candidate key should directly depend on every candidate key of that table.

| Customer Table |                  |                  |                  |      |        |
|----------------|------------------|------------------|------------------|------|--------|
| ID             | Customer Name    | Shipping Address | Newsletter       | Zip  | City   |
| 1              | Jens Bergholm    | Mølletofte 20    | Xbox News        | 7100 | Vejle  |
| 2              | Dennis Jørgensen | Vinkelvej 167    | Playstation News | 7100 | Vejle  |
| 3              | Hans Jensen      | Kongensgade 45   | Xbox News        | 5000 | Odense |
| 3              | Hans Jensen      | Kongensgade 45   | Playstation News | 5000 | Odense |
| 4              | Anne Johansen    | Vinkelvej 167    | Playstation News | 7100 | Vejle  |

| Order Table |            |
|-------------|------------|
| Customer_Id | Product_Id |
| 1           | 1          |
| 2           | 2          |
| 3           | 1          |
| 3           | 3          |
| 4           | 2          |

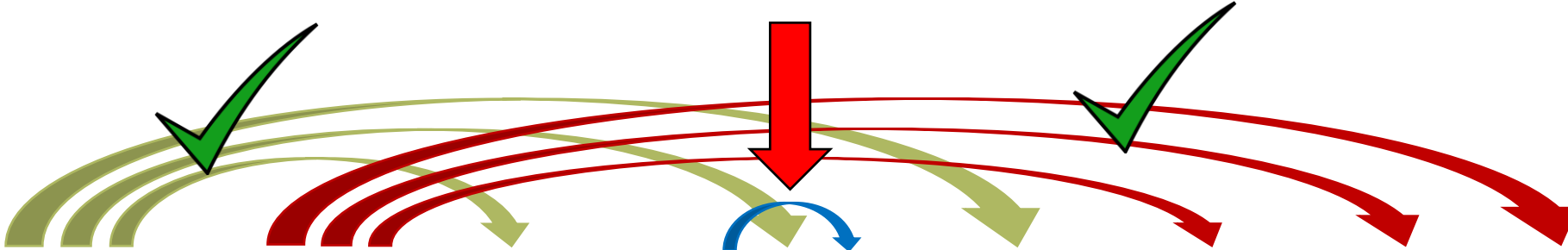
| Product Table |               |       |             |
|---------------|---------------|-------|-------------|
| ID            | Item          | Price | Supplier_ID |
| 1             | XBOX One      | 1800  | 1           |
| 2             | Playstation 4 | 1900  | 2           |
| 3             | PS Vita       | 1500  | 2           |

| Supplier Table |           |          |
|----------------|-----------|----------|
| ID             | Name      | Phone    |
| 1              | Microsoft | 12341234 |
| 2              | Sony      | 23452345 |

# BCNF – Boyce–Codd Normal Form

→ AKA: NF3.5

→ If a relational schema is in BCNF then all redundancy based on functional dependency has been removed, although other types of redundancy may still exist.



| ID | Customer Name    | Item          | Shipping Address | Zip  | City   | Newsletter       | Supplier  | Supplier Phone | Price |
|----|------------------|---------------|------------------|------|--------|------------------|-----------|----------------|-------|
| 1  | Jens Bergholm    | XBOX One      | Mølletoft 20     | 7100 | Vejle  | Xbox News        | Microsoft | 12341324       | 1800  |
| 2  | Dennis Jørgensen | Playstation 4 | Vinkelvej 167    | 7100 | Vejle  | Playstation News | Sony      | 23452345       | 1900  |
| 3  | Hans Jensen      | XBOX One      | Kongensgade 45   | 5000 | Odense | Xbox News        | Microsoft | 12341324       | 1800  |
| 3  | Hans Jensen      | PS Vita       | Kongensgade 45   | 5000 | Odense | Playstation News | Sony      | 23452345       | 1500  |
| 4  | Anne Johansen    | Playstation 4 | Vinkelvej 167    | 7100 | Vejle  | Playstation News | Sony      | 23452345       | 1900  |

# BCNF – Boyce–Codd Normal Form

→ If a relational schema is in BCNF then all redundancy based on functional dependency has been removed, although other types of redundancy may still exist.

| Customer Table |                  |                  |                  |        |
|----------------|------------------|------------------|------------------|--------|
| ID             | Customer Name    | Shipping Address | Newsletter       | Zip_ID |
| 1              | Jens Bergholm    | Mølletofte 20    | Xbox News        | 7100   |
| 2              | Dennis Jørgensen | Vinkelvej 167    | Playstation News | 7100   |
| 3              | Hans Jensen      | Kongensgade 45   | Xbox News        | 5000   |
| 3              | Hans Jensen      | Kongensgade 45   | Playstation News | 5000   |
| 4              | Anne Johansen    | Vinkelvej 167    | Playstation News | 7100   |

| Zip Table |        |
|-----------|--------|
| Zip       | Name   |
| 5000      | Odense |
| 7100      | Vejle  |

| Order Table |            |
|-------------|------------|
| Customer_Id | Product_Id |
| 1           | 1          |
| 2           | 2          |
| 3           | 1          |
| 3           | 3          |
| 4           | 2          |

| Product Table |               |       |             |
|---------------|---------------|-------|-------------|
| ID            | Item          | Price | Supplier_ID |
| 1             | XBOX One      | 1800  | 1           |
| 2             | Playstation 4 | 1900  | 2           |
| 3             | PS Vita       | 1500  | 2           |

| Supplier Table |           |          |
|----------------|-----------|----------|
| ID             | Name      | Phone    |
| 1              | Microsoft | 12341234 |
| 2              | Sony      | 23452345 |



# NF4 – Fourth Normal Form

→ All columns can be determined only by the key in the table and no other column

→ No Multi-valued Dependencies

| Newsletters Tab. |                  |
|------------------|------------------|
| ID               | Newsletter       |
| 1                | Xbox News        |
| 2                | Playstation News |

| Subscriptions Table |             |
|---------------------|-------------|
| Newsletter_ID       | Customer_ID |
| 1                   | 1           |
| 2                   | 2           |
| 1                   | 3           |
| 2                   | 3           |
| 2                   | 4           |

| Customer Table |                  |                  |        |
|----------------|------------------|------------------|--------|
| ID             | Customer Name    | Shipping Address | Zip_ID |
| 1              | Jens Bergholm    | Mølletofte 20    | 7100   |
| 2              | Dennis Jørgensen | Vinkelvej 167    | 7100   |
| 3              | Hans Jensen      | Kongensgade 45   | 5000   |
| 4              | Anne Johansen    | Vinkelvej 167    | 7100   |

| Zip Table |        |
|-----------|--------|
| Zip       | Name   |
| 5000      | Odense |
| 7100      | Vejle  |

| Order Table |            |
|-------------|------------|
| Customer_Id | Product_Id |
| 1           | 1          |
| 2           | 2          |
| 3           | 1          |
| 3           | 3          |
| 4           | 2          |

| Product Table |               |       |             |
|---------------|---------------|-------|-------------|
| ID            | Item          | Price | Supplier_ID |
| 1             | XBOX One      | 1800  | 1           |
| 2             | Playstation 4 | 1900  | 2           |
| 3             | PS Vita       | 1500  | 2           |

| Supplier Table |           |          |
|----------------|-----------|----------|
| ID             | Name      | Phone    |
| 1              | Microsoft | 12341234 |
| 2              | Sony      | 23452345 |



# Recap with using simple rules

- Rows identify single entities
- A tables attributes should be directly related
  - Think object oriented, or about physical entites
- Avoid Data Redundancy
- Preserve Relationships

# Working from NF1

| Entity |                  |                  | Entity<br>+<br>Redundancy |        | Entity<br>+<br>Redundancy |               | Entity<br>+<br>Redundancy |           | Entity<br>+<br>Redundancy |  | #sdu.dk |
|--------|------------------|------------------|---------------------------|--------|---------------------------|---------------|---------------------------|-----------|---------------------------|--|---------|
| ID     | Customer Name    | Shipping Address | Zip                       | City   | Newsletter                | Item          | Price                     | Supplier  | Supplier Phone            |  |         |
| 1      | Jens Bergholm    | Mølletofte 20    | 7100                      | Vejle  | Xbox News                 | XBOX One      | 1800                      | Microsoft | 12341324                  |  |         |
| 2      | Dennis Jørgensen | Vinkelvej 167    | 7100                      | Vejle  | Playstation News          | Playstation 4 | 1900                      | Sony      | 23452345                  |  |         |
| 3      | Hans Jensen      | Kongensgade 45   | 5000                      | Odense | Xbox News                 | XBOX One      | 1800                      | Microsoft | 12341324                  |  |         |
| 3      | Hans Jensen      | Kongensgade 45   | 5000                      | Odense | Playstation News          | PS Vita       | 1500                      | Sony      | 23452345                  |  |         |
| 4      | Anne Johansen    | Vinkelvej 167    | 7100                      | Vejle  | Playstation News          | Playstation 4 | 1900                      | Sony      | 23452345                  |  |         |

# Exercises

| Doctor<br>Number | Name           | Room<br>Address | Phone    | Department Id    | Designation         | Charges per<br>hour | Patient<br>Number | Patient<br>Name | CPR<br>Number | Patient<br>Phone | Room<br>Number | Room<br>Type | Bed<br>Number |
|------------------|----------------|-----------------|----------|------------------|---------------------|---------------------|-------------------|-----------------|---------------|------------------|----------------|--------------|---------------|
| D1               | Dr. Peterson   | U45             | 12341234 | Neurology        | Professor           | 5000                | P1                | Jan             | 190582-1113   | 98769876         | R2             | Normal       | B1            |
| D1               | Dr. Peterson   | U45             | 12341324 | Neurology        | Professor           | 5000                | P5                | Peter           | 300175-2359   | 87658765         | R2             | Normal       | B1            |
| D1               | Dr. Peterson   | U45             | 12341234 | Neurology        | Professor           | 5000                | P7                | Jens            | 041298-1257   | 76547654         | Null           |              | Null          |
| D2               | Dr. Jensen     | U32             | 24352435 | Orthopedic       | Professor           | 5000                | P4                | Ole             | 051165-9863   | 65436543         | R2             | Normal       | B1            |
| D2               | Dr. Jensen     | U32             | 23452345 | Orthopedic       | Professor           | 5000                | P7                | Jens            | 041298-1257   | 76547654         | R4             | Two Bed      | B5            |
| D2               | Dr. Jensen     | U32             | 23452435 | Orthopedic       | Professor           | 5000                | P9                | Anna            | 260792-1050   | 54325432         | R4             | Two Bed      | B7            |
| D4               | Dr. Poetch     | U186            | 34563456 | ENT/Neurology    | Assistant Professor | 3000                | P10               | Dennis          | 150893-1151   | 43214321         | Null           |              | Null          |
| D4               | Dr. Poetch     | U186            | 34563456 | ENT/Neurology    | Assistant Professor | 3000                | P1                | Jan             | 190582-1113   | 98769876         | R5             | Special      | B8            |
| D5               | Dr. Neurenheim | U150            | 45674567 | Skin/Orthepoedic | Assistant Professor | 3000                | P12               | Ahmed           | 010211-7853   | 32103210         | Null           |              | Null          |
| D5               | Dr. Neurenheim | U150            | 45674567 | Skin/Orthepoedic | Assistant Professor | 3000                | P13               | Annika          | 051285-8072   | 21092109         | R6             | Special      | B9            |

→ From the dataset above:

→ Normalize the table to the fourth normal form.

→ Create SQL Script to generate all the tables, and the associated data.

→ Query the database for all patients of Dr. Peterson, and find their name, cpr, what room they were in, and the bed number.