University of Chittagong

Department of Computer Science & Engineering

Database Systems Lab

Name of the assignment:

Task 3: Refinement, Normalization, and SQL-DDL

CSE 414

Medical Database Analysis

Submitted By:
Debashish Chakraborty
ID: 23701034

Submitted To:
Dr. Rudra Pratap Deb Nath
Associate Professor

July 6, 2025

Contents

Introduction

This report analyzes the medical database schema created in previous tasks, focusing on functional dependencies, normalization theory, and the application of normal forms.

Current Database Schema

The current normalized schema consists of the following relations:

1. **Drug**(drug_name, drug_category, product_name, company_name)
2. **SideEffect**(name)
3. **Disease**(disease_name, disease_category)
4. **ClinicalTrial**(clinical_trial_title, clinical_trial_start_date, clinical_trial_completion_date, clinical_trial_participants, clinical_trial_status, clinical_trial_address, clinical_trial_institution, clinical_trial_address_1, clinical_trial_main_researcher)
5. **ClinicalTrialCondition**(name)
6. **Drug_SideEffect**(drug_name, side_effect_name)
7. **Drug_Interaction**(drug_name, interaction_name)
8. **Drug_Disease**(drug_name, disease_name)
9. **Drug_ClinicalTrial**(drug_name, clinical_trial_title)
10. **ClinicalTrial_Condition**(clinical_trial_title, condition_name)

Functional Dependencies Analysis

Identifying Functional Dependencies

For each relation, we identify the functional dependencies:

3.1.1 Drug Relation

Relation: Drug(drug_name, drug_category, product_name, company_name)

Functional Dependencies:

1. drug_name $\rightarrow$ drug_category
2. product_name $\rightarrow$ company_name
3. drug_name $\rightarrow$ product_name (assuming one product per drug)
4. drug_name $\rightarrow$ company_name (transitive through product_name)

Candidate Keys: {drug_name}

3.1.2 SideEffect Relation

Relation: SideEffect(name)

Functional Dependencies:

1. No non-trivial functional dependencies (single attribute relation)

Candidate Keys: {name}

3.1.3 Interaction Relation

Relation: Interaction(name)

Functional Dependencies:

1. No non-trivial functional dependencies (single attribute relation)

Candidate Keys: {name}

3.1.4 Disease Relation

Relation: Disease(disease_name, disease_category)

Functional Dependencies:

1. FD1: disease_name → disease_category

Candidate Keys: {disease_name}

3.1.5 ClinicalTrial Relation

Relation: ClinicalTrial(clinical_trial_title, clinical_trial_start_date, clinical_trial_completion_date, clinical_trial_participants, clinical_trial_status, clinical_trial_address, clinical_trial_institution, clinical_trial_condition, clinical_trial_main_researcher)

Functional Dependencies:

1. clinical_trial_title → clinical_trial_start_date
2. clinical_trial_title → clinical_trial_completion_date
3. clinical_trial_title → clinical_trial_participants
4. clinical_trial_title → clinical_trial_status
5. clinical_trial_title → clinical_trial_address
6. clinical_trial_title → clinical_trial_institution
7. clinical_trial_title → clinical_trial_comdition
8. clinical_trial_title → clinical_trial_main_researcher

Candidate Keys: {clinical_trial_title}

3.1.6 Drug_SideEffect Relation

Relation: Drug_SideEffect(drug_name, side_effect_name)

Functional Dependencies:

1. FD1: {drug_name, side_effect_name} → {drug_name, side_effect_name} (trivial)
2. No non-trivial functional dependencies (all attributes form the composite primary key)

Candidate Keys: {drug_name, side_effect_name}

3.1.7 Drug_Interaction Relation

Relation: Drug_Interaction(drug_name, interaction_name)

Functional Dependencies:

1. FD1: {drug_name, interaction_name} → {drug_name, interaction_name} (trivial)
2. No non-trivial functional dependencies (all attributes form the composite primary key)

Candidate Keys: {drug_name, interaction_name}

3.1.8 Drug_Disease Relation

Relation: Drug_Disease(drug_name, disease_name)

Functional Dependencies:

1. FD1: {drug_name, disease_name} → {drug_name, disease_name} (trivial)
2. No non-trivial functional dependencies (all attributes form the composite primary key)

Candidate Keys: {drug_name, disease_name}

3.1.9 Drug_ClinicalTrial Relation

Relation: Drug_ClinicalTrial(drug_name, clinical_trial_title)
Functional Dependencies:

1. FD1: {drug_name, clinical_trial_title} → {drug_name, clinical_trial_title} (trivial)
2. No non-trivial functional dependencies (all attributes form the composite primary key)

Candidate Keys: {drug_name, clinical_trial_title}

Lossless Decomposition Analysis

Let R be the original relation and $\{R_1, R_2, \ldots, R_n\}$ be our decomposition.

Theorem: The decomposition is lossless if and only if:

$$R = R_1 \bowtie R_2 \bowtie \cdots \bowtie R_n$$

Let us define our decomposed relations:

$$R_1 = \text{Drug}(drug\_name, drug\_category, product\_name, company\_name) \quad (1)$$
$$R_2 = \text{SideEffect}(name) \quad (2)$$
$$R_3 = \text{Disease}(disease\_name, disease\_category) \quad (3)$$
$$R_4 = \text{ClinicalTrial}(\text{clinical trial attributes}) \quad (4)$$
$$R_5 = \text{Drug\_SideEffect}(drug\_name, side\_effect\_name) \quad (5)$$
$$R_6 = \text{Drug\_Disease}(drug\_name, disease\_name) \quad (6)$$
$$R_7 = \text{Drug\_ClinicalTrial}(drug\_name, clinical\_trial\_title) \quad (7)$$
$$R_8 = \text{ClinicalTrial\_Condition}(clinical\_trial\_title, condition\_name) \quad (8)$$

The total relation is:

$$R' = R_1 \bowtie R_2 \bowtie R_3 \bowtie R_4 \bowtie R_5 \bowtie R_6 \bowtie R_7 \bowtie R_8$$

Union Verification:

$$\bigcup_{i=1}^{8} \text{attributes}(R_i) = \text{attributes}(R)$$

This confirms that all original attributes are preserved in the decomposition. So, it is a lossless decomposition.

Dependency Preservation Analysis

The decomposition preserves dependencies if and only if:

$$F^+ \equiv \left(\bigcup_{i=1}^{n} F_i \right)^+$$

For each relation R_i, we compute the attribute closure and verify functional dependencies:

Drug Relation:

$$F_1 = \{\text{drug\_name} \rightarrow \text{drug\_category, product\_name, company\_name}\} \quad (9)$$
$$\{drug\_name\}^+ = \{drug\_name, drug\_category, product\_name, company\_name\} \quad (10)$$

SideEffect Relation:

$$F_2 = \{\} \text{ (no non-trivial FDs)} \quad (11)$$
$$\{name\}^+ = \{name\} \quad (12)$$

Disease Relation:

$$F_3 = \{\text{disease\_name} \rightarrow \text{disease\_category}\} \tag{13}$$
$$\{disease\_name\}^+ = \{disease\_name, disease\_category\} \tag{14}$$

ClinicalTrial Relation:

$$F_4 = \{\text{clinical\_trial\_title} \rightarrow \text{all other clinical trial attributes}\} \tag{15}$$
$$\{clinical\_trial\_title\}^+ = \{\text{all clinical trial attributes}\} \tag{16}$$

Junction Relations: For each junction relation R_j:

$$F_j = \{\text{composite key} \rightarrow \text{all attributes}\} \tag{17}$$
$$\{\text{composite key}\}^+ = \{\text{all attributes in } R_j\} \tag{18}$$

Dependency Preservation Check:

Original Functional Dependencies:

$$F_{original} = \{\text{drug\_name} \rightarrow \text{drug\_category, product\_name, company\_name}, \tag{19}$$
$$\text{product\_name} \rightarrow \text{company\_name}, \tag{20}$$
$$\text{disease\_name} \rightarrow \text{disease\_category}, \tag{21}$$
$$\text{clinical\_trial\_title} \rightarrow \text{all clinical trial attributes}\} \tag{22}$$

Union of Decomposed Dependencies:

$$F_{decomposed} = F_1 \cup F_2 \cup F_3 \cup F_4 \cup F_5 \cup F_6 \cup F_7 \cup F_8$$

We need to show that, $F^+_{original} = F^+_{decomposed}$.

1. drug_name $\rightarrow$ drug_category: Preserved in F_1
2. product_name $\rightarrow$ company_name: Preserved in F_1
3. disease_name $\rightarrow$ disease_category: Preserved in F_3
4. clinical_trial_title $\rightarrow$ clinical trial attributes: Preserved in F_4
5. All junction table dependencies: Preserved in F_5, F_6, F_7, F_8

Since every functional dependency from $F_{original}$ can be derived from $F_{decomposed}$, and vice versa, the decomposition is dependency-preserving.

Normal Form Analysis

First Normal Form (1NF)

Definition

A relation is in 1NF if:

1. All attributes contain atomic values.
2. No repeating groups exist.

All relations are in 1NF:

1. **Drug**: All attributes (drug_name, drug_category, product_name, company_name) are atomic
2. **SideEffect**: Single atomic attribute (name)
3. **Interaction**: Single atomic attribute (name)
4. **Disease**: Both attributes (disease_name, disease_category) are atomic
5. **ClinicalTrial**: All attributes are atomic values
6. **ClinicalTrialCondition**: Single atomic attribute (name)
7. **Junction Relations**: All contain atomic foreign key values

The original CSV data violated 1NF with repeating groups (side_effect_1, side_effect_2, etc.). This was resolved by creating separate junction tables.

Second Normal Form (2NF)

A relation is in 2NF if:

1. It is in 1NF
2. No non-prime attribute is partially dependent on any candidate key. Meaning no proper subset of candidate key attributes can determine a non-prime attribute.

All relations are in 2NF:

1. **Drug**: Primary key is {drug_name} (single attribute), so no partial dependencies possible
2. **SideEffect, Interaction, ClinicalTrialCondition**: Single attribute relations, automatically in 2NF
3. **Disease**: Primary key is {disease_name} (single attribute), so no partial dependencies possible
4. **ClinicalTrial**: Primary key is {clinical_trial_title} (single attribute), so no partial dependencies possible
5. **Junction Relations**: All non-prime attributes are part of the primary key, so no partial dependencies

Third Normal Form (3NF)

A relation is in 3NF if:

1. It is in 2NF
2. No non-prime attribute is transitively dependent on any non-prime attribute. except non-prime attribute can be transitively dependent on any prime attribute.

In the original relation schema, the Drug relation violates 3NF due to transitive dependency:

1. drug_name → product_name
2. product_name → company_name
3. Therefore: drug_name → company_name (transitive)

To achieve 3NF, we need to decompose the Drug relation:

1. **Drug_New**(drug_name, drug_category, product_name)
2. **Product**(product_name, company_name)

Other Relations in 3NF:

1. **SideEffect, Interaction, ClinicalTrialCondition**: No transitive dependencies
2. **Disease**: No transitive dependencies
3. **ClinicalTrial**: No transitive dependencies (all attributes directly dependent on trial title)
4. **Junction Relations**: No non-prime attributes to create transitive dependencies

Boyce-Codd Normal Form (BCNF)

A relation is in BCNF if:

1. It is in 3NF
2. For every functional dependency X → Y, X is a superkey
3. For X → Y, X is a subset of Y(Trivial)
4. No transitive dependency is allowed.

After 3NF Decomposition:

1. drug_name is a superkey → **BCNF**
2. product_name is a superkey → **BCNF**
3. SideEffect, Interaction, ClinicalTrialCondition: Trivially in BCNF.
4. disease_name is a superkey → **BCNF**
5. clinical_trial_title a superkey → **BCNF**
6. No functional dependencies beyond key constraints → **BCNF**

Revised SQL DDL for 3NF/BCNF Schema

Product Table(New)

```sql
CREATE TABLE Product (
    product_name VARCHAR(255) PRIMARY KEY,
    company_name VARCHAR(255) NOT NULL
);
```

Drug Table

```sql
CREATE TABLE Drug (
    drug_name VARCHAR(255) PRIMARY KEY,
    drug_category VARCHAR(100),
    product_name VARCHAR(255),
    FOREIGN KEY (product_name) REFERENCES Product(product_name)
);
```

SideEffect Table

```sql
CREATE TABLE SideEffect (
    name VARCHAR(255) PRIMARY KEY
);
```

Interaction Table

```sql
CREATE TABLE Interaction (
    name VARCHAR(255) PRIMARY KEY
);
```

Disease Table

```sql
CREATE TABLE Disease (
    disease_name VARCHAR(255) PRIMARY KEY,
    disease_category VARCHAR(100)
);
```

ClinicalTrial Table

```sql
CREATE TABLE ClinicalTrial (
    clinical_trial_title VARCHAR(500) PRIMARY KEY,
    clinical_trial_start_date TEXT,
    clinical_trial_completion_date TEXT,
    clinical_trial_participants NUMERIC,
```

```
    clinical_trial_status VARCHAR(50),
    clinical_trial_address VARCHAR(500),
    clinical_trial_institution VARCHAR(255),
    clinical_trial_address_1 VARCHAR(500),
    clinical_trial_conditions VARCHAR(255),
    clinical_trial_main_researcher VARCHAR(255)
);
```

Drug_SideEffect Table

```
CREATE TABLE Drug_SideEffect (
    drug_name VARCHAR(255) REFERENCES Drug(drug_name),
    side_effect_name VARCHAR(255) REFERENCES SideEffect(name),
    PRIMARY KEY (drug_name, side_effect_name)
);
```

Drug_Interaction Table

```
CREATE TABLE Drug_Interaction (
    drug_name VARCHAR(255) REFERENCES Drug(drug_name),
    interaction_name VARCHAR(255) REFERENCES Interaction(name),
    PRIMARY KEY (drug_name, interaction_name)
);
```

Drug_Disease Table

```
CREATE TABLE Drug_Disease (
    drug_name VARCHAR(255) REFERENCES Drug(drug_name),
    disease_name VARCHAR(255) REFERENCES Disease(disease_name),
    PRIMARY KEY (drug_name, disease_name)
);
```

Drug_ClinicalTrial Table

```
CREATE TABLE Drug_ClinicalTrial (
    drug_name VARCHAR(255) REFERENCES Drug(drug_name) ,
    clinical_trial_title VARCHAR(500) REFERENCES
       ClinicalTrial(clinical_trial_title) ,
    PRIMARY KEY (drug_name, clinical_trial_title)
);
```

```

ALTER TABLE Product ADD CONSTRAINT unique_product_name UNIQUE
   (product_name);

ALTER TABLE Drug ADD CONSTRAINT fk_drug_product
```

```
    FOREIGN KEY (product_name) REFERENCES Product(product_name);

ALTER TABLE Drug ALTER COLUMN drug_category SET NOT NULL;
ALTER TABLE Disease ALTER COLUMN disease_category SET NOT NULL;
```

Revised ER Diagram

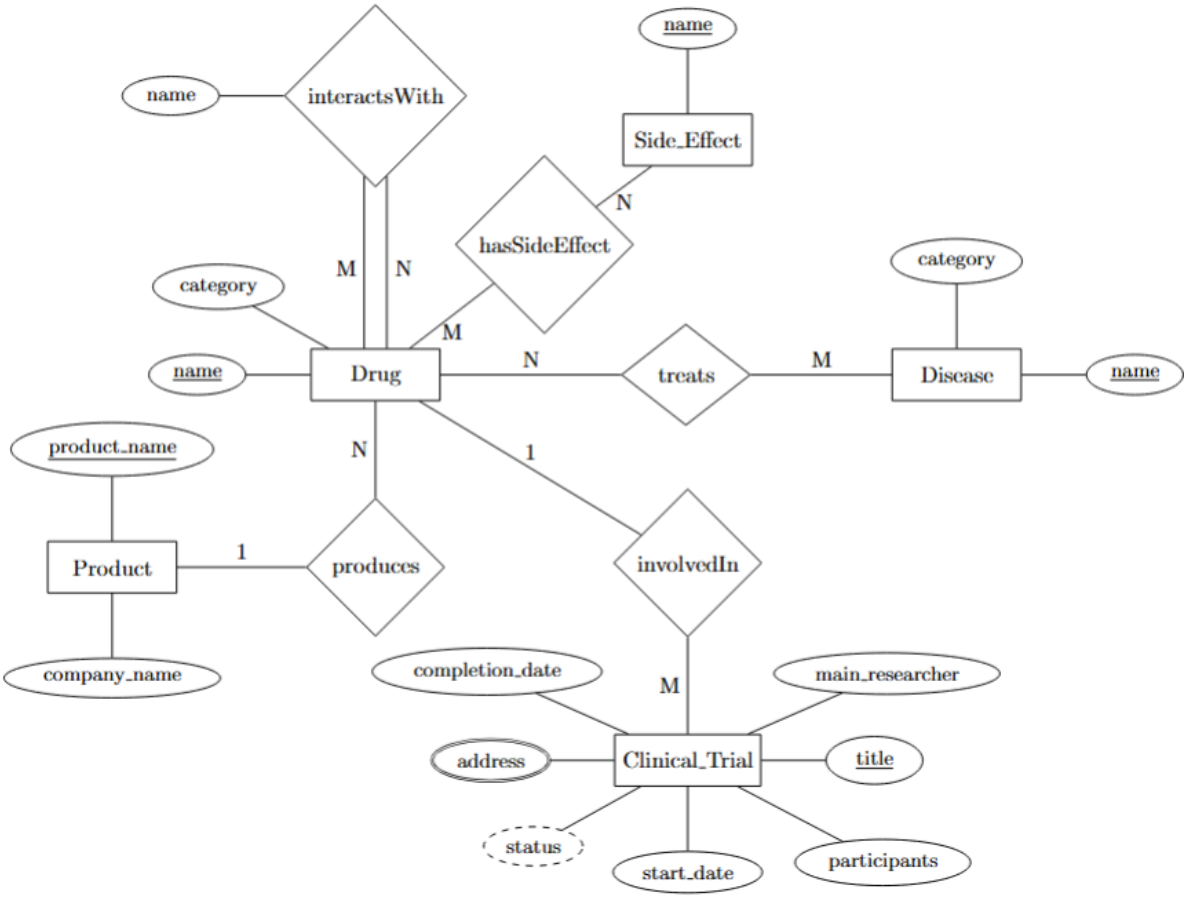

Figure 1: Revised ER Diagram

Comparison with Original Design

1. Removed all Transitive Dependencies. Separated Product entity from Drug.
2. All relations now satisfy BCNF requirements.
3. Explicit foreign key constraints.
4. Reduced Redundancy.