Exercise Sheet: Data Normalization and Denormalization

Data, Data Storage, Data Collection

Exercise 1 - Pizzerias

Consider the following schema, where each row represents one product sold by a pizzeria:

Pizzeria(**Name**, **Address**, **ProductSold**, **Price**).

(**a**) Using examples, identify the anomalies and redundancies caused by this schema.

Exercise 2 - Functional Dependencies in a Sample Relation

Consider the relation r over the schema $R(A, B, C, D, E)$ with the following tuples:

A	**B**	**C**	**D**	**E**
a1	b1	c1	d1	e1
a1	b2	c2	d2	e1
a2	b1	c3	d3	e1
a2	b1	c4	d3	e1
a3	b2	c5	d1	e1

(**a**) Which attributes take only a single value in the dataset? What functional dependencies follow immediately?
(**b**) For each attribute $X \in \{A, B, C, D\}$, examine whether $X \to Y$ holds for any other attribute Y.
(**c**) Check whether any *pairs* of attributes (e.g. AB, AC, AD) determine another attribute. List all FDs of the form $XY \to Z$.
(**d**) From all the above, keep only the *minimal* functional dependencies.

Note that these dependencies hold for this dataset because equal values of the left-hand side always correspond to equal values of the right-hand side. They are not guaranteed to hold in a larger or real-world dataset.

Exercise 3 - University Course Registration

A university maintains a database of course registrations. The current table is as follows:

StudentID	**StudentName**	**Courses**	**Instructor**
101	Alice	Math, Physics	Dr. Smith, Dr. Johnson
102	Bob	Chemistry, Math	Dr. Brown, Dr. Smith
103	Carol	Biology, Chemistry	Dr. Green, Dr. Brown

(1) First Normal Form (1NF)

(**a**) Identify the issues with the current table. Why is it not in 1NF?
(**b**) Transform the table into 1NF, explain your steps.

(2) Second Normal Form (2NF)

(**a**) List all superkeys of the 1NF table.
(**b**) Identify the candidate key(s).
(**c**) List the minimal functional dependencies.
(**d**) Is the table in 2NF? If not, explain why and transform it into 2NF.

(3) Third Normal Form (3NF)

Suppose you add a **Department** column to the table. Each **Instructor** belong to one **Department**.

Course	Instructor	Department
Math	Dr. Smith	Mathematics
Physics	Dr. Johnson	Physics
Chemistry	Dr. Brown	Chemistry
Biology	Dr. Green	Biology

(**a**) Is this table in 3NF? Explain your answer. If not, normalize it to 3NF.

(4) Fourth Normal Form (4NF)

Now, suppose you add a table for student activities, where each student can join multiple clubs and participate in multiple sports:

StudentID	Club	Sport
101	Chess, Debate	Swimming, Tennis
102	Robotics, Chess	Tennis, Soccer
103	Debate, Music	Swimming, Soccer

(**a**) Transform this table into 3NF.
(**b**) Is the resulting table in 4NF? Explain your answer. If not, normalize it to 4NF.

(5) Fifth Normal Form (5NF)

Consider the following table representing student projects, where each project involves a team of students, a supervisor, and a topic:

Student	Supervisor	Topic
Alice	Dr. Smith	AI
Alice	Dr. Smith	Data Science
Alice	Dr. Brown	AI
Bob	Dr. Brown	AI
Carol	Dr. Smith	Data Science

(**a**) Is this table in 5NF? Explain your answer. If not, decompose it into 5NF.

Exercise 4 - Hospital Patient Records

A hospital database is designed as follows:

- Patients(**PatientID**, **Name**, **Address**, **AdmissionDate**)
- Doctors(**DoctorID**, **Name**, **Specialization**)
- Visits(**VisitID**, **PatientID**, **DoctorID**, **VisitDate**, **Diagnosis**)

(1) Normalization

(**a**) Is this schema normalized? If not, identify the issues and propose a normalized schema.

(2) Denormalization

Suppose the hospital imports daily insurance information from an external partner. The provided dataset is a flat file in the form

Insurance(**PatientName**, **Address**, **InsuranceProvider**, **InsurancePlan**, **LastUpdate**)

with no stable patient identifier.

(**a**) Describe a denormalization strategy that allows the hospital database to store the imported data efficiently without reconstructing unique identifiers for every row.

Hospitals must retain the *exact* doctor specialization that was in effect at the time of each visit, even if the doctor later changes specialization.

(**b**) Show how this requirement forces a denormalization of the normalized schema. Provide a modified version of the Visits table that satisfies the rule, and explain why it is not redundant in this context.

The emergency unit performs thousands of queries per hour to retrieve:

PatientName, DoctorName, Specialization, Diagnosis, VisitDate

for the most recent visit of each patient. The normalized schema requires joining three tables.

(**c**) Propose a denormalized structure that speeds up this workload. Explain the potential anomalies or maintenance costs introduced.