Exercise Sheet: Normalization

Data, Data Storage, Data Collection

Exercise 1 – Pizzerias

A food-delivery website lists the menus of local pizzerias in a single table, where each row records one product sold by one pizzeria:

Pizzeria(Name, Address, Product, Price).

The website follows two rules: each pizzeria has a single address, and each pizzeria sets its own price for each product. Here is an extract of the table:

Name	Address	Product	Price
Fratelli	12 Main St	Margherita	9
Fratelli	12 Main St	Calzone	12
Fratelli	12 Main St	Tiramisu	6
Da Rosa	3 Park Ave	Margherita	10
Da Rosa	3 Park Ave	Diavola	11
Bella Napoli	8 Harbor Rd	Margherita	8

(**Q1**) Using rows of the table, give an example of redundancy and an example of each anomaly: insertion, update and deletion.

(**Q2**) What are the functional dependencies implied by the two rules? Show that the only candidate key is $\{\text{Name}, \text{Product}\}$.

(**Q3**) Determine the highest normal form of the table and justify it with the definitions. Then decompose the table into tables in BCNF, check that the split is lossless, and name the foreign key.

Exercise 2 – University Course Registration

A university stores its course registrations in a single table. The registrar follows these rules: student IDs are unique, but two students may share a name; each course is taught by a single instructor, but an instructor may teach several courses.

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

A – First Normal Form (1NF)

(**Q1**) Why is this table not in 1NF? Transform it into 1NF.

B – Second Normal Form (2NF)

(**Q2**) List the functional dependencies implied by the registrar's rules, and find the candidate key(s) of the 1NF table.

(**Q3**) Is the 1NF table in 2NF? If not, explain why and decompose it into 2NF.

C – Third Normal Form (3NF)

The university now records the department of each instructor in the table of courses and instructors, and adds a Statistics course. Each instructor belongs to a single department.

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

(**Q4**) Is this table in 3NF? If not, decompose it into 3NF. Are the resulting tables in BCNF?

The university also runs a tutoring program. Each student has a single tutor per course, and each tutor teaches a single course:

Student	Course	Tutor
Alice	Math	Tom
Bob	Math	Lea
Carol	Math	Tom
Alice	Physics	Nina
Carol	Physics	Nina

(**Q5**) Find the candidate keys of this table. Is it in 3NF? In BCNF?

(**Q6**) Split this table along the dependency that violates BCNF. What do you gain and what do you lose? Which design do you keep?

D – Fourth Normal Form (4NF)

The university also records student activities. Each student can join several clubs and play several sports, and the clubs a student joins have nothing to do with the sports they play:

StudentID	Clubs	Sports
101	Chess, Debate	Swimming, Tennis
102	Robotics, Chess	Tennis, Soccer
103	Debate, Music	Swimming, Soccer

(**Q7**) Transform this table into 1NF. Which rows must it contain for each student? Is it in BCNF?

(**Q8**) Is this 1NF table in 4NF? If not, decompose it into 4NF.

E – Fifth Normal Form (5NF)

The university finally records final-year projects. Each row states that a student works on a topic with a supervisor. No non-trivial functional or multivalued dependency holds, so the only candidate key is the set of all three attributes, and the table is in 4NF:

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

(**Q9**) Project the table onto (Student, Supervisor), (Supervisor, Topic) and (Student, Topic), and join the three projections back. Do you get the original table?

(**Q10**) The university adopts the following rule: *if a student works with a supervisor, that supervisor supervises a topic, and the student works on that topic, then the student works on that topic with that supervisor.* How can we turn the table into 5NF?

Exercise 3 – A Partner's Undocumented Extract

A partner company sends us an extract of one of its tables for a data integration project. The column names were anonymized, so we only know that the schema is $R(A, B, C, D)$, and we receive no documentation and no rules of the domain, only these five rows:

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

Recall that $X \to Y$ holds in a table if any two rows that agree on X also agree on Y. We look for the dependencies that hold on these five rows.

(**Q1**) Which single attributes determine another attribute? Which pairs of attributes determine a third one?

(**Q2**) Deduce the candidate key(s) of R.

(**Q3**) A sixth row arrives: $(a1, b1, c6, d2)$. Which dependencies of (**Q1**) still hold? What does this say about dependencies found on a sample?