

SEN-209 Designing and Implementing Databases – Take-Home Exam

Duration: **12** hours (8:00 AM → 8:00 PM)

Expected Completion Time: ~**6** hours

Resources Allowed:

1. Notes, Internet, GPT/AI tools (but answers must be your own work).
2. **dvdrental** database can be downloaded from <https://neon.tech/unify?a=c0afe242-89e1-4987-81f7-488e6d865f0b&n=postgresltutorial/dvdrental.zip>

Deliverables: Submit a single **PDF** document with all answers. Validation must be accompanied by screenshot.

Question 1 – Advanced ER Modeling & Mapping (20 points)

A national research institute manages **Projects, Researchers, Departments, Grants, and Publications**.

- Each **Project** has an ID, title, start date, budget, and is led by exactly one researcher.
- Each **Researcher** has an ID, name, email, position, and belongs to exactly one department.
- Each **Department** has an ID, name, head researcher (a researcher who manages it), and location.
- A researcher can work on multiple projects, and a project can involve multiple researchers. Each researcher's role (Lead, Collaborator, Consultant) and weekly hours on the project must be recorded.
- Each **Grant** funds one or more projects. A project can be funded by multiple grants. Each grant has an ID, funding agency, total amount, and funding period.
- Each **Publication** is linked to exactly one project, but may have multiple co-authors (researchers). Each publication has a DOI, title, year, and type (journal, conference, report).

Tasks:

- a. Draw an **ER diagram** with all entities, attributes, PKs, relationships, and cardinalities.
- b. Convert it into **relational schemas** (tables, attributes, keys, constraints).
- c. Explain **how your schema enforces referential integrity**.

Question 2 – Normalization to 3NF & Implementation (20 points)

The research institute initially kept all its project data in a **single unstructured table**:

Research Project
ProjectID
ProjectTitle
ResearcherID
ResearcherName
ResearcherEmail
DepartmentName
DepartmentHead
GrantID
GrantAgency

GrantAmount
PublicationDOI
PublicationTitle
PubYear
ResearcherRole
HoursPerWeek

Notes:

- Each project can involve multiple researchers.
- A researcher belongs to exactly one department.
- A department is managed by exactly one researcher (the department head).
- A project can be funded by multiple grants.
- A project can result in multiple publications.
- Each researcher's **role** and **weekly hours** are tracked for each project.

Tasks

a. Functional Dependencies (FDs)

- Identify at least **6 functional dependencies** from the table (you may assume reasonable ones if not explicitly stated).

b. 1NF → 2NF → 3NF Normalization

- Show step-by-step decomposition:
 - **1NF:** Remove repeating groups (separate multivalued attributes).
 - **2NF:** Remove partial dependencies (no non-key attribute depends on part of a composite key).
 - **3NF:** Remove transitive dependencies (non-key attributes depend only on candidate keys).

c. Final Normalized Schema (3NF)

- List all relations (tables) in **3NF**, including:
 - Primary keys
 - Foreign keys
 - Attributes

d. SQL Implementation

- Write **CREATE TABLE** statements for your final 3NF schema.
- Each table must have at least:
 - PRIMARY KEY
 - FOREIGN KEY (where appropriate)
 - At least one additional constraint (NOT NULL, UNIQUE, or CHECK).

Question 3 – Reverse Query Writing (20 points)

You are given **result descriptions** from the **dvdrental** database. For each, you must **write the SQL query** that produces it. There may be multiple plausible answers, but your query must be correct and efficient.

a. Result 1 – Nested Aggregation

- Output: A table with two columns: Customer ID, Rental Count.
- But it only includes customers whose rental count is **above the overall average rental count** across all customers.

b. Result 2 – Multi-Join Revenue Analysis

- Output: A table with Film Title, Category Name, and Total Revenue.
- Shows only the **top-grossing film in each category**.
- Hint: Requires combining film, category, payment, and rental.

c. Result 3 – Correlated Subquery

- Output: A list of films (Film ID, Title) that have a **rental rate higher than the average rental rate** of all films in the same rating category.
- Example: Compare a film's rental rate to the average for other films rated "R" or "PG".

d. Result 4 – Complex Join with Time Constraint

- Output: A table with Staff Name, Month, Total Payments.
- Only for months in **2006** where the staff member processed **more than \$3,000** in payments.
- Hint: Requires grouping by staff + month, and filtering with HAVING.

Question 4 – Complex Analytical Queries (20 points)

a. Managerial Question 1:

- "Which **5 customers** spent the most in total payments, and what percentage of overall revenue did each contribute?"

b. Managerial Question 2:

- "Which **3 categories** of films generated the highest revenue, and what is the average rental duration of films in those categories?"

c. Managerial Question 3:

- "Across all stores, which **city** had the highest number of unique customers in 2005, and how many rentals did they make on average?"

For each question:

- Write the SQL query.
- Show output.

- Provide a short (2–3 sentence) explanation of your approach.

Question 5 – Database Design Critique & Cloud Migration (20 points)

The dvdrental database was designed for a local DVD rental shop but is now moving to the cloud for global use.

Tasks:

- a. Identify two major limitations of the current schema (e.g., scalability, redundancy, lack of constraints, performance).
- b. Propose concrete schema redesigns to address them (with example SQL if needed).
- c. Justify how your improvements enhance data integrity, efficiency, or cloud scalability.