

Name: Krishna Somani
Roll No. : 2021058

Report - Assignment 1

1. About RDBMS Schema

1. Database Structure: There are Six main tables: Departments, Instructors, Students, Courses, Enrollments, and CourseInstructor.

2. Departments Table: Holds unique department_id and department_name.

3. Instructors Table: Stores instructor details and links to Departments via department_id with cascading deletes.

4. Students Table: Records student info linked to Departments through department_id.

5. Courses Table: Represents courses with foreign keys to ensure valid associations with departments and instructors.

6. Enrollments Table:

Captures student-course relationships, including enrollment_date and grade, with foreign key links.

7. CourseInstructor Table:

Manages many-to-many relationships between courses and instructors with foreign key enforcement.

Full schema is attached as a code file.

2. About Migration and MongoDB Schema

Data Structures Created During Migration

1. Students Collection:

- Each document includes fields such as:
 - `_id`: Unique identifier for the student.
 - `first_name` and `last_name`: Names of the student.
 - `email`: Contact information.
 - `birth_date`: Date of birth.
 - `enrollment_year`: Year the student enrolled.
 - `department`: An embedded document with department details, including:
 - `department_id`: Identifier for the department.
 - `department_name`: Name of the department.
 - `enrollments`: An array of embedded documents representing courses, each with fields like `course_id`, `course_name`, `credits`, and `grade`.

2. Courses Collection:

- Each document comprises:
 - `_id`: Unique identifier for the course.
 - `course_name`: Title of the course.
 - `course_code`: Code for the course.
 - `credits`: Credit hours associated with the course.
 - `is_core`: Boolean indicating if it is a core course.
 - `instructor`: An embedded document containing:
 - `instructor_id`: Identifier for the instructor.
 - `instructor_name`: Name of the instructor.
 - `department`: An embedded document with:
 - `department_id`: Identifier for the department.
 - `department_name`: Name of the department.

3. Instructors Collection:

- Each document consists of:
 - `_id`: Unique identifier for the instructor.
 - `first_name` and `last_name`: Instructor's names.
 - `email`: Contact information.
 - `hire_date`: Date of hiring.
 - `department`: An embedded document containing:
 - `department_id`: Identifier for the department.
 - `department_name`: Name of the department.

More on data migration:

Steps involved:

1. Data Extraction: Data was retrieved from PostgreSQL using SQL queries, including students, enrollments, courses, and instructors, with necessary joins to gather related information.
2. Data Cleaning: Dates were converted to Python *datetime* objects. Null values in critical fields were addressed to prevent inconsistencies.
3. Data Transformation: Enrollments were embedded within student documents as arrays to maintain relationships. Nested structures were created, allowing for instructors and departments to be included in course documents.
4. Data Loading: The cleaned and transformed data was loaded into MongoDB collections (students, courses, and instructors) using *insert_many*.

Overall code is also present in the zip file.