Project Title: Hotel Chain Database Design

Project Overview:

In this project, you will design, implement, and optimize a relational database for a hotel chain management system. This database will manage multiple properties and critical assets, including rooms, guests, bookings, staff, and services. The system will be designed to handle operations across all hotel properties in the chain.

You will apply the knowledge gained from their Database module in the Data Engineering Bootcamp, focusing on building a schema for complex relationships, optimizing query performance, and ensuring data security.

Core Assets (Entities) To Consider

1. **Hotels**: Information about each hotel property in the chain, including location, rating, and contact information.
2. **Rooms**: Details of individual rooms across all properties (room type, availability, price, etc.).
3. **Guests**: Customer profiles, preferences, and stay history.
4. **Bookings**: Information about guest reservations, including dates, rooms, and billing.
5. **Staff**: Employee records, schedules, and role assignments.
6. **Services**: Additional hotel services offered, such as restaurants, spa, and gym, and their associated costs.

Project Objectives:

• Design a scalable relational database schema for a multi-property hotel chain. •

Develop and run complex queries for business analysis.

• Apply performance optimisation techniques (indexing, normalization, query optimization).

• Implement security measures like user roles and permissions to protect sensitive data.

• Handle operational challenges for a chain-wide system, such as data consistency and scalability.

Project Phases:
Phase 1: Requirements Gathering and Data Mapping

- **Task**: Identify the specific data fields for each core entity (Hotels, Rooms, Guests, Bookings, Staff, Services).

Deliverable: A document outlining attributes for each entity. For example:

- **Hotels**: Hotel ID, Name, Location, Rating, Contact Info.
- **Rooms**: Room ID, Hotel ID (FK), Room Type, Price, Availability, Capacity.
- **Guests**: Guest ID, Name, Email, Phone, Preferences.
- **Bookings**: Booking ID, Guest ID (FK), Room ID (FK), Check-in Date, Check-out Date, Total Cost.
- **Staff**: Staff ID, Name, Hotel ID (FK), Role, Shift Schedule, Salary.
- **Services**: Service ID, Hotel ID (FK), Service Type (e.g., Spa, Restaurant), Cost.

Phase 2: Database Schema Design

- **Task**: Design an ER (Entity-Relationship) diagram representing the relationships between entities, ensuring scalability to manage multiple properties.

Deliverable: An ER diagram and corresponding SQL code for creating tables and relationships (including primary/foreign keys, constraints). Consider normalization (3NF or higher) to reduce redundancy and improve database efficiency.

Example Relationships:

- **One-to-Many**: One hotel has many rooms; one guest can have multiple bookings.
- **Many-to-Many**: A booking may include multiple services (e.g., room service, spa treatments), and one service can be associated with multiple bookings.

Phase 3: Data Insertion

- **Task**: Populate the database with realistic sample data that covers multiple hotel properties. Ensure data diversity across properties (e.g., different room types, guest preferences, and services). You can use the random data generator “Mockaroo” to help you.

Deliverable: SQL scripts to insert sample data into each table, covering at least:

- 5 different hotels.
- 100 rooms distributed across the hotels.
- 200 guest records.
- 300 bookings (with various services utilised).
- 50 staff.
- 5 services

Phase 4: Querying and Business Analysis

- **Task**: Develop SQL queries to retrieve critical business insights for hotel chain management.

Deliverable: A report containing both basic and advanced queries, including:

Basic Queries:

0. List all available rooms at a specific hotel.
1. Retrieve guest details and preferences for all guests who stayed in the last month.
2. Calculate total revenue generated by each hotel property in a specific time period.

Advanced Queries:

3. Determine which room types generate the most revenue across the chain. 4. Identify guests who have stayed in more than two hotels within the chain. 5. Calculate the average occupancy rate for each hotel over the past six months. **Business Analysis**:

- Perform analysis on the most requested services and their revenue contribution.
- Identify peak booking periods and recommend staff scheduling optimizations.

Phase 5: Database Optimization

- **Task**: Implement optimisation techniques to improve performance for large-scale data retrieval, given the multi-property structure.
  - Create indexes on frequently queried columns (e.g., room availability, booking dates).
  - Refactor slow queries by analysing their execution plans.

- Use appropriate joins and aggregation strategies to handle complex reporting queries.

Deliverable:

- A performance report comparing query execution times before and after optimisation.
- SQL scripts for creating indexes and optimised queries.

Phase 6: Security Implementation

- **Task**: Implement role-based security and enforce data protection standards. Use techniques to prevent SQL injection and secure sensitive guest and staff data.
  - Define user roles (e.g., Admin, Hotel Manager, Receptionist) with appropriate access control.
  - Encrypt sensitive fields like guest contact information.
  - Ensure prepared statements and parameterised queries to avoid SQL injection vulnerabilities.

Deliverable:

- SQL scripts for user roles, permissions, and security measures.
- A summary of the security protocols implemented and how they mitigate risks.

GOOD LUCK!