

Industrial Internship Training Report

Monitoring of Database Functionality Using Structured Query Language(SQL)

**Submitted
by**
Saswat Seth
(20010352)
of
7th Sem/CSE(AI & ML)

Under Supervision of :
Dr. Patitapaban Rath
Department of Computer Science,
Central University of Odisha

(Duration: 8th May, 2023 - 22nd May, 2023)



**Department of Computer Science & Engineering
C. V. RAMAN GLOBAL UNIVERSITY,
BHUBANESWAR, ODISHA**

December
2023

Declaration

I hereby declare that the internship report entitled “Monitoring of Database Functionality Using Structured Query Language(SQL)” is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to substantial extent has been accepted for the award of any degree of the university or another institute of higher learning.

Name of the Student : Saswat Seth

Regn No. : 20010352

Date : 5/12/2023

Department of Computer Science & Engineering

C. V. RAMAN GLOBAL UNIVERSITY



Certificate of Approval

This is to certify that we have examined the training report entitled “Monitoring of Database Functionality Using Structured Query Language(SQL)” submitted by, Saswat Seth (Regd No.- 20010352), CGU, Bhubaneswar. We hereby accord our approval of the training work carried out and presented in a manner required for its acceptance as per the academic regulation, for the partial fulfillment for the 7th Semester in Computer Science & Engineering. This training has fulfilled all the requirements as per the regulations of the university.

Prof. Monalisa Mishra
(Internship Coordinator)

Dr. Rojalina Priyadarshini
(H.O.D, CSE)



CENTRAL UNIVERSITY OF ODISHA
DEPARTMENT OF COMPUTER SCIENCE
KORAPUT – 763004

Ref No: CUO/DCS/CT/01/2023

May 24, 2023

CERTIFICATE OF TRAINING

This is to certify that **Mr. Saswat Seth** bearing Registration- CAM 2003, Branch-B Tech, Computer Science (AI & ML) of C V Raman Global University, Bhubaneswar has undergone 15 days training from 08.05.2023 to 22.05.2023 and carried a project on **Monitoring of Database Functionality Using Structured Query Language** under supervision of Dr. Patitapaban Rath, Lecturer at Department of Computer Science, Central University of Odisha.

His conduct during the training is found good.



Head of the Department

Head of the Department
Department of Computer Science
Central University of Odisha

Dr. Jyotiska Datta
Head of the Department IC- Addl.Charge

Contact No: 9348854590
Email ID: hod.cs@cuo.ac.in

Acknowledgement

It gives me immense pleasure to express my sincere gratitude to our faculty coordinator Prof. Monalisa Mishra for her support and advices to get and complete internship in the above said organization.

I extend my sincere thanks to our HOD Dr. R. Priyadarshini for her immeasurable support throughout my internship.

I also like to acknowledge the contribution of other faculty members of the Department of CSE for their cooperation and kind assistance in successful completion of this internship.

December 2023

Saswat Seth (20010352)

Abstract

This internship report encapsulates the enriching and skill-enhancing training experience undertaken by Mr. Saswat Seth, a diligent B.Tech student enrolled at C V Raman Global University. The report focuses on a meticulously designed 15-day training program facilitated by the esteemed Department of Computer Science at Central University of Odisha. The training, which transpired from 08.05.2023 to 22.05.2023, centered around the implementation and exploration of "Monitoring of Database Functionality Using Structured Query Language."

Under the sagacious guidance of Dr. Patitapaban Rath, an accomplished lecturer at the Department of Computer Science, Mr. Seth demonstrated exemplary dedication, technical acumen, and an innate ability to synthesize theoretical knowledge into practical applications. The report chronicles the progression of the training, detailing the intricacies of the undertaken project. The project itself delved into the multifaceted realm of database functionality, utilizing the power of Structured Query Language (SQL) for monitoring and analysis.

The report not only serves as a testament to the successful completion of the training but also sheds light on Mr. Seth's comprehensive understanding of database management, SQL intricacies, and the broader implications of his project. It highlights his ability to navigate real-world challenges, think critically, and apply theoretical concepts in a practical setting.

Endorsed by Dr. Jyotiska Datta, the Head of the Department of Computer Science at Central University of Odisha, this certificate of training stands as a formal acknowledgment of Mr. Seth's commendable conduct and successful project completion. Dr. Datta's endorsement further attests to the high standards of achievement and proficiency expected from students undergoing training within the department.

Contents

DECLARATION.....	i
CERTIFICATE OF APPROVAL.....	ii
INTERNSHIP CERTIFICATE.....	iii
ACKNOWLEDGEMENT.....	iv
ABSTRACT.....	v
CONTENTS	vi
WEEKLY OVERVIEW.....	1
CHAPTER-1-INTRODUCTION.....	2
1.1 OVERVIEW.....	2
1.2 BACKGROUND AND MOTIVATION.....	2
1.3 LEARNING OBJECTIVE.....	3
CHAPTER-2-METHODOLOGY.....	4
2.1 DATA COLLECTION.....	4
2.2 TOOLS AND TECHNOLOGIES USED.....	6
CHAPTER-3-PROJECT DESCRIPTION.....	9
3.1 DATABASE SCHEMA REVIEW.....	9
3.2 DATA TYPES AND CONSTRAINTS ASSESSMENT.....	9
CHAPTER-4-RESULT/LEARNING OUTCOME.....	11
4.1 ASSESSMENT FINDINGS.....	11
4.2 CHALLENGES FACED.....	14
4.3 SKILLS DEVELOPED.....	14
CHAPTER-5-CONCLUSION	16
5.1 SUMMARY OF INTERNSHIP EXPERIENCE.....	16
5.2 ACHIEVEMENTS AND MILESTONES.....	16
5.3 CLOSING REMARKS.....	17

WEEKLY OVERVIEW OF INTERNSHIP ACTIVITIES

Week	Date	Day	Name of the Topic/Module Completed
1 st week		Monday	<ul style="list-style-type: none"> ● Introduction to Internship Project: ● Introduction to SQL
		Tuesday	<ul style="list-style-type: none"> ● MySQL Introduction ● MySQL Queries
		Wednesday	<ul style="list-style-type: none"> ● MySQL Advanced Queries ● Introduction to PostgreSQL
		Thursday	<ul style="list-style-type: none"> ● PostgreSQL Basics: ● PostgreSQL Advanced Queries
		Friday	<ul style="list-style-type: none"> ● Comparative Analysis
		Saturday	<ul style="list-style-type: none"> ● Recap and Review
2 nd week		Monday	<ul style="list-style-type: none"> ● Music Store Database Assessment ● Database Schema Review
		Tuesday	Data Types and Constraints Assessment
		Wednesday	<ul style="list-style-type: none"> ● Query Performance Analysis ● Optimization Techniques
		Thursday	<ul style="list-style-type: none"> ● Current Indexing Analysis ● Strategies for Optimal Indexing
		Friday	<ul style="list-style-type: none"> ● Data Quality Assessment ● Final Database Assessment:
		Saturday	Documentation
15 th day		Monday	Finalization, Recommendations and Assessment

Chapter 1

INTRODUCTION

1.1 Overview

I embarked on a 15-day internship dedicated to the evaluation of a music store database, with a focus on refining the database structure, enhancing query performance, and ensuring data integrity in a music retail setting.

The first week was immersed in an intensive learning phase, where I delved into the intricacies of SQL, MySQL, and PostgreSQL. This foundational knowledge set the stage for the subsequent evaluation of the music store database.

Structured with precision, the internship progressed from a foundational learning phase to the hands-on assessment of the music store database in the following week. This report provides a detailed account of my daily activities, challenges encountered, and milestones achieved throughout this enriching journey.

The significance of evaluating the music store database extends beyond database management, impacting the efficiency and functionality of a music retail system. This project allowed me to translate theoretical knowledge into practical applications, contributing substantially to my professional development.

Organized chronologically, the report details the learning phase initially and then provides a comprehensive breakdown of the music store database assessment. Each day's agenda, covered topics, and outcomes are meticulously documented, offering readers a thorough understanding of the entire internship experience.

1.2 Background and Motivation

The music store database, a linchpin in the intricate web of a music retail system, is currently under scrutiny. This pivotal database encompasses customer details, inventory information, and transaction records, forming a cornerstone of the store's operational infrastructure.

In the dynamic music industry landscape, effective database management is crucial for streamlined operations, precise inventory tracking, and seamless customer transactions. A well-optimized database not only refines internal processes but also enhances the overall customer experience.

The core objective of the internship project was a comprehensive evaluation, scrutiny, and enhancement of the existing music store database. This involved a detailed review of the database schema, assessment of data types and constraints, query optimization, and strategic indexing. The scope extended to addressing data quality issues, reinforcing security measures,

and documenting the database structure for future reference.

In the ever-evolving music retail sector, staying abreast of industry trends is imperative. The project aimed not only to rectify current inefficiencies but also to align the store's data management practices with contemporary industry standards.

The hands-on evaluation of a real-world music store database provided a unique learning opportunity, allowing for the practical application of theoretical knowledge acquired during the initial learning phase. This experiential learning deepened the understanding of database management principles and their practical implications in a business context.

The successful optimization of the music store database aligns with industry best practices, positioning the store for heightened operational efficiency and improved customer satisfaction. Consequently, this project has the potential to significantly contribute to the store's sustained success in a competitive market.

1.3 Learning Objective

Over the 15-day internship, the primary learning objectives served as a comprehensive guide for acquiring a diverse skill set in database assessment and optimization. The focus was on developing a profound understanding of assessment methodologies, particularly the critical evaluation of database schemas in the context of a music store database.

Proficiency in query optimization was actively cultivated, involving hands-on experience in identifying and rectifying slow-performing queries. Practical implementation of optimization strategies was undertaken to enhance overall query performance. Mastery of indexing strategies and their real-world application further enriched the learning process.

A crucial aspect of the internship involved the meticulous analysis of data quality, emphasizing skills in assessing and ensuring data integrity. This included identifying common issues and proposing effective measures for maintaining a consistent data environment. Database security intricacies were explored, encompassing a comprehensive review of existing measures and understanding user permissions and access controls in a practical database setting.

Emphasis on comprehensive documentation and best practices heightened awareness regarding the importance of documenting database schemas for future reference. Additionally, dedicated focus on data backup and recovery procedures equipped me with essential skills to ensure data resilience.

The overarching learning objectives not only provided a structured framework for skill acquisition but also facilitated a holistic understanding of project management in the context of a database assessment project. The culmination of this learning experience involved synthesizing findings and recommendations into a comprehensive internship report, integrating theoretical knowledge with practical application in a real-world setting.

Chapter 2

Methodology

2.1 Data Collection and Analysis

For the music store database assessment, a systematic approach was taken to gather comprehensive data, ensuring a thorough understanding of the database's structure, content, and usage. The following methods were employed.

Database Schema Analysis:

A detailed review of the existing database schema was conducted. This involved examining the tables, relationships, and constraints to gain insights into the overall structure.

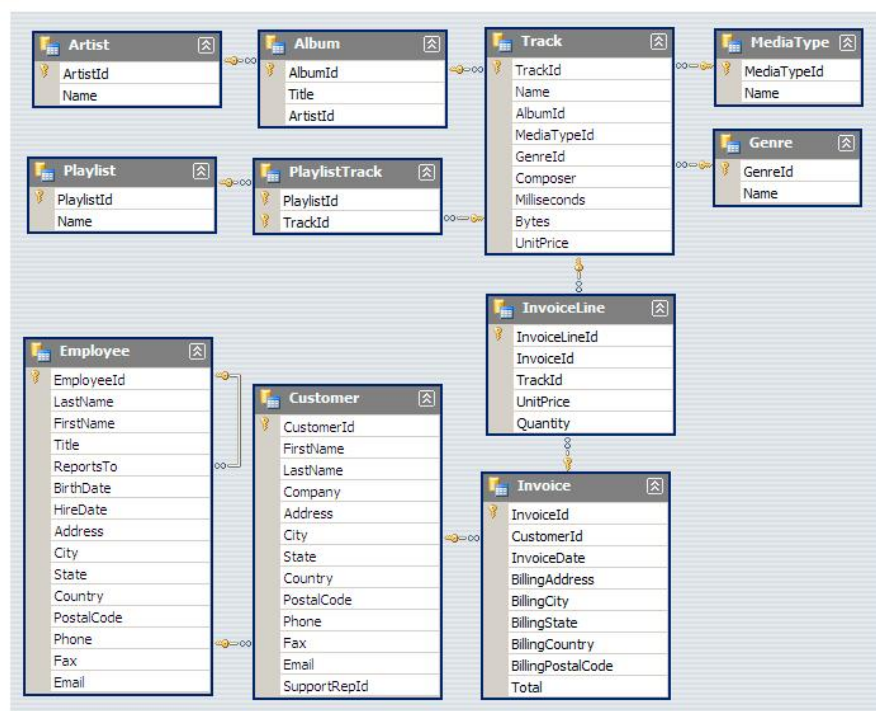


Figure-1: Music store database schema

1. employee Table

Columns:

employee_id: Unique identifier for each employee.

last_name, first_name: Employee's last and first names.

title: Job title of the employee.

reports_to: ID of the supervisor or manager.
levels: Job level of the employee.
birthdate, hire_date: Dates of birth and hire.
address, city, state, country, postal_code: Employee's address details.
phone, fax, email: Contact information.

2. customer Table

Columns:

customer_id: Unique identifier for each customer.
first_name, last_name: Customer's first and last names.
company: Customer's company name.
address, city, state, country, postal_code: Customer's address details.
phone, fax, email: Customer's contact information.
support_rep_id: ID of the supporting representative.

3. invoice Table

Columns:

invoice_id: Unique identifier for each invoice.
customer_id: ID referencing the customer associated with the invoice.
invoice_date: Date of the invoice.
billing_address, billing_city, billing_state, billing_country, billing_postal: Billing address details.
total: Total amount of the invoice.

4. invoice_line Table

Columns:

invoice_line_id: Unique identifier for each invoice line.
invoice_id: ID referencing the associated invoice.
track_id: ID referencing the track in the invoice line.
unit_price: Price per unit of the track.
quantity: Quantity of the track in the invoice line.

5. track Table

Columns:

track_id: Unique identifier for each track.
name: Name of the track.
album_id, media_type_id, genre_id: IDs referencing associated album, media type, and genre.
composer: Composer of the track.
milliseconds: Duration of the track in milliseconds.
bytes: Size of the track in bytes.
unit_price: Price of the track.

6. playlist Table

Columns:

playlist_id: Unique identifier for each playlist.
name: Name of the playlist.

7. playlist_track Table

Columns:

playlist_id, track_id: Composite primary key referencing the playlist and track associated.

8. artist Table

Columns:

artist_id: Unique identifier for each artist.
name: Name of the artist.

9. album Table

Columns:

album_id: Unique identifier for each album.
title: Title of the album.
artist_id: ID referencing the associated artist.

10. media_type Table

Columns:

media_type_id: Unique identifier for each media type.
name: Name of the media type.

11. genre Table

Columns:

genre_id: Unique identifier for each genre.
name: Name of the genre.

2.3 Tools and Technologies Used

During the course of the internship, a range of tools and technologies were employed to facilitate a thorough assessment of the music store database. These tools were instrumental in various aspects of the project, including data collection, analysis, and optimization. The key tools and technologies used include.

Database Management System (DBMS):

PostgreSQL: The primary DBMS used for the music store database. PostgreSQL was chosen for its robustness, open-source nature, and extensive support for complex queries.

MySQL: Used for educational purposes to broaden understanding and proficiency in different database systems. Although not central to the project, exploring MySQL provided valuable insights into variations in syntax, features, and administration.

Query Optimization Tools:

pgAdmin: A comprehensive administration and management tool for PostgreSQL. pgAdmin played a crucial role in query optimization, execution plan analysis, and overall database performance monitoring in the PostgreSQL environment.

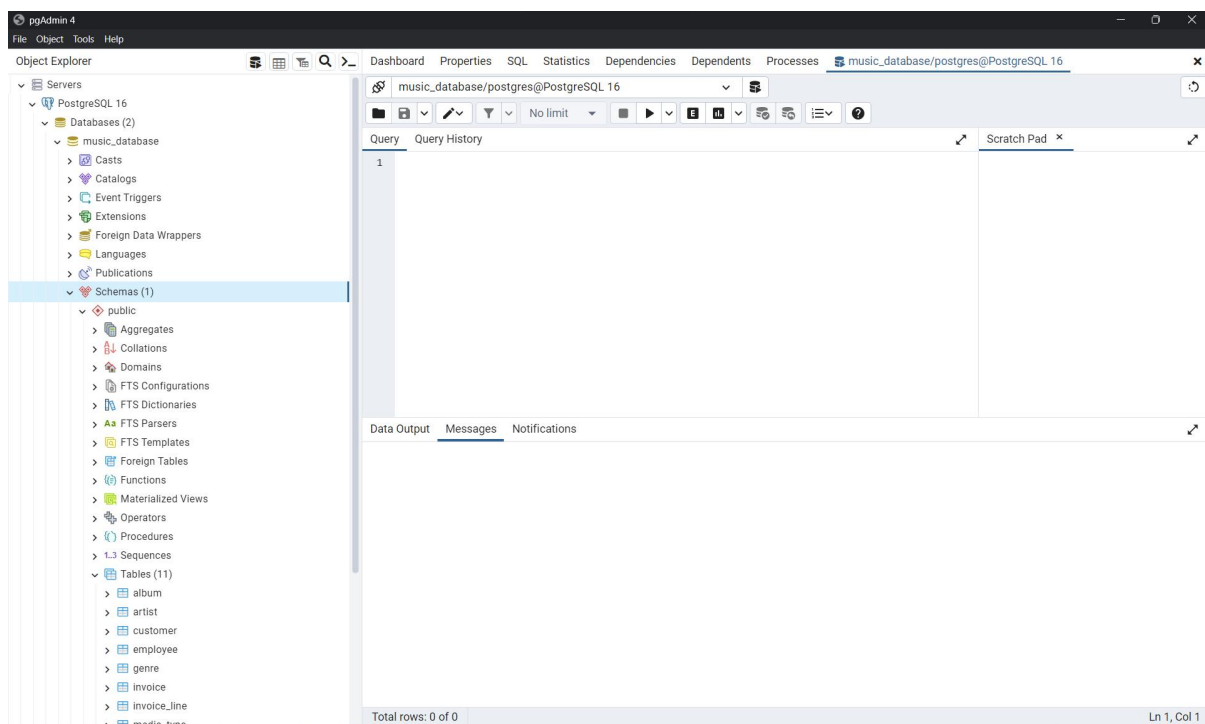


Figure-2: pgAdmin for PostgreSQL

Learning and Exploration:

MySQL Workbench: Utilized for hands-on learning and exploration of MySQL. The tool provided a visual interface for database design, development, and administration, enhancing familiarity with MySQL concepts.

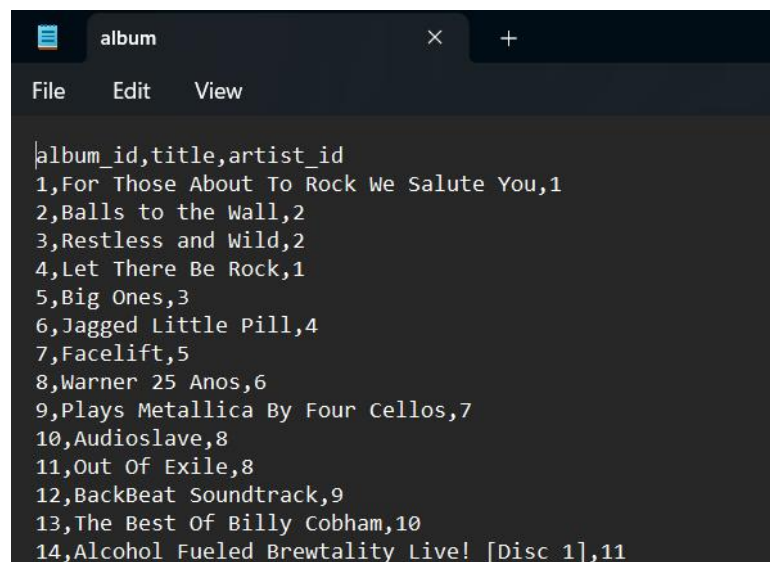
Collaboration Platforms

During the on-site internship, direct collaboration occurred with the Computer Science Department employees, fostering in-person communication, meetings, and discussions. This

on-site approach facilitated immediate query clarification, dynamic discussions, and enhanced project coordination. The absence of virtual tools like Microsoft Teams was compensated by effective face-to-face interactions, contributing to seamless knowledge sharing and idea exchange within the on-site environment. This approach enriched the internship experience, fostering a close-knit and interactive working relationship with the department's team.

Text Editors

Notepad and PostgreSQL Query Tool were essential tools during the internship. Notepad, a versatile text editor, facilitated the writing and editing of SQL scripts and documentation. The PostgreSQL Query Tool was specifically utilized for PostgreSQL-related tasks. These tools offered straightforward interfaces for coding and documentation, adapting seamlessly to the requirements of both PostgreSQL and MySQL environments. Notepad provided a lightweight option for general text editing, while the PostgreSQL Query Tool offered specialized support for PostgreSQL queries, ensuring an efficient and tailored text editing workflow.



```
album_id,title,artist_id
1,For Those About To Rock We Salute You,1
2,Balls to the Wall,2
3,Restless and Wild,2
4,Let There Be Rock,1
5,Big Ones,3
6,Jagged Little Pill,4
7,Facelift,5
8,Warner 25 Anos,6
9,Plays Metallica By Four Cellos,7
10,Audioslave,8
11,Out Of Exile,8
12,BackBeat Soundtrack,9
13,The Best Of Billy Cobham,10
14,Alcohol Fueled Brewtality Live! [Disc 1],11
```

Figure-3: album csv file in notepad

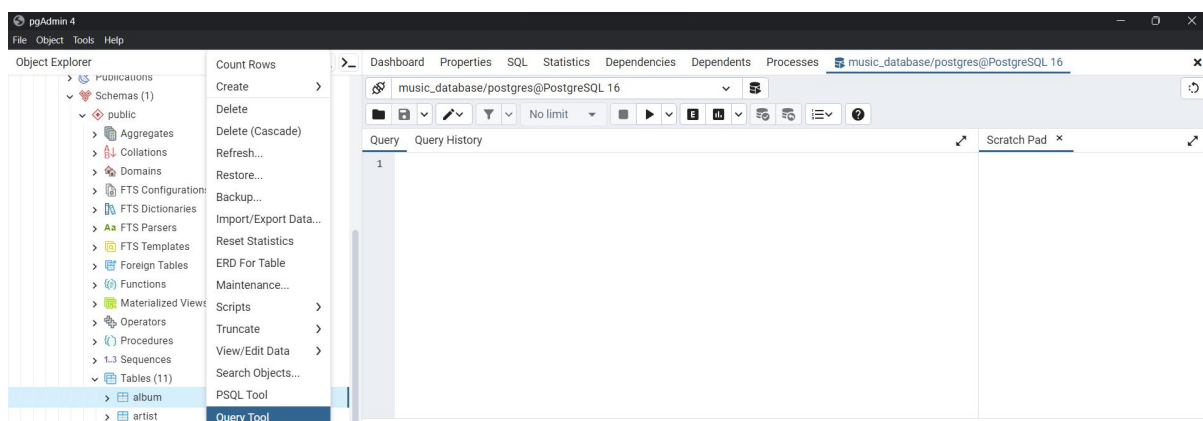


Figure-4: query tool in PostgreSQL

Operating System

Windows: The entire project lifecycle, including development and testing, was conducted exclusively on the Windows operating system. This choice was made for its compatibility and seamless integration with the tools and technologies employed. The use of Windows provided a stable and consistent environment, ensuring optimal performance and efficiency throughout the internship. The decision to focus on a single operating system streamlined the development process and facilitated a more straightforward and unified experience across different aspects of the project.

Chapter 3

PROJECT DESCRIPTION

3.1 Database Schema Review

Table Architecture: Each table's design and composition were scrutinized, emphasizing the definition of primary and foreign key relationships. This step ensured a thorough understanding of how data entities were organized and interconnected.

Entity Relationships: The examination extended to the relationships between different entities represented in the database. This included the identification and verification of relationships to maintain data integrity and coherence.

Data Types and Constraints: An evaluation of data types assigned to each column was conducted to ascertain their appropriateness for the stored information. Concurrently, constraints such as primary keys, foreign keys, unique constraints, and others were reviewed to maintain the reliability and accuracy of data.

Documentation Analysis: Existing schema documentation was reviewed to ensure clarity and completeness. This documentation serves as a vital reference for developers, administrators, and other stakeholders, facilitating effective database management.

3.2 Data Types and Constraints Assessment

Employee Table:

Columns:

employee_id: VARCHAR(50).
last_name, first_name, title, reports_to: CHAR(50), VARCHAR(30).
levels: VARCHAR(10).
birthdate, hire_date: TIMESTAMP.
address: VARCHAR(120).
city, state, country: VARCHAR(50), VARCHAR(30).
postal_code, phone, fax, email: VARCHAR(30).

Constraints:

employee_id: PRIMARY KEY.

Customer Table:

Columns:

customer_id: VARCHAR(30).
first_name, last_name, company, address, city, state, country, email: CHAR(30),
VARCHAR(30).
postal_code: INT8.
phone, fax: INT.
support_rep_id: VARCHAR(30).

Constraints:

customer_id: PRIMARY KEY.
support_rep_id: FOREIGN KEY (references employee table).

Invoice Table:

Columns:

invoice_id: VARCHAR(30).
customer_id: VARCHAR(30).
invoice_date: TIMESTAMP.
billing_address, billing_city, billing_state, billing_country, billing_postal: VARCHAR(120),
VARCHAR(30).

Constraints:

invoice_id: PRIMARY KEY.
customer_id: FOREIGN KEY (references customer table).

Invoice Line Table:

Columns:

invoice_line_id: VARCHAR(50).
invoice_id, track_id: VARCHAR(30).
unit_price, quantity: VARCHAR(30).

Constraints:

invoice_line_id: PRIMARY KEY.
invoice_id: FOREIGN KEY (references invoice table).
track_id: FOREIGN KEY (references track table).

Track Table:

Columns:

track_id: VARCHAR(50).

name, composer: VARCHAR(30).
album_id, media_type_id, genre_id: VARCHAR(30).
milliseconds: TIMESTAMP.
bytes: INT8.
unit_price: INT16.

Constraints:

track_id: PRIMARY KEY.
album_id: FOREIGN KEY (references album table).
media_type_id: FOREIGN KEY (references media_type table).
genre_id: FOREIGN KEY (references genre table).

Playlist Table:

Columns:

playlist_id: VARCHAR(50).
name: VARCHAR(30).

Constraints:

playlist_id: PRIMARY KEY.

Playlist Track Table:

Columns:

playlist_id, track_id: VARCHAR(50).

Constraints:

(playlist_id, track_id): PRIMARY KEY.
playlist_id: FOREIGN KEY (references playlist table).
track_id: FOREIGN KEY (references track table).

Artist Table:

Columns:

artist_id: VARCHAR(50).
name: VARCHAR(30).

Constraints:

artist_id: PRIMARY KEY.

Album Table:

Columns:

album_id: VARCHAR(50).
title: VARCHAR(30).
artist_id: VARCHAR(30).

Constraints:

album_id: PRIMARY KEY.
artist_id: FOREIGN KEY (references artist table).

Media Type Table:

Columns:

media_type_id: VARCHAR(50).
name: VARCHAR(30).

Constraints:

media_type_id: PRIMARY KEY.

Genre Table:

Columns:

genre_id: VARCHAR(50).
name: VARCHAR(30).

Constraints:

genre_id: PRIMARY KEY.

Query Optimization

Query Optimization played a crucial role in refining the performance and efficiency of database queries. The process involved a detailed analysis of existing queries to pinpoint areas for improvement. Strategies such as index optimization, query rewriting for enhanced execution plans, and the implementation of best practices were employed to streamline query performance. The overarching goal was to achieve expeditious and resource-efficient query execution, contributing significantly to the overall optimization of the database.

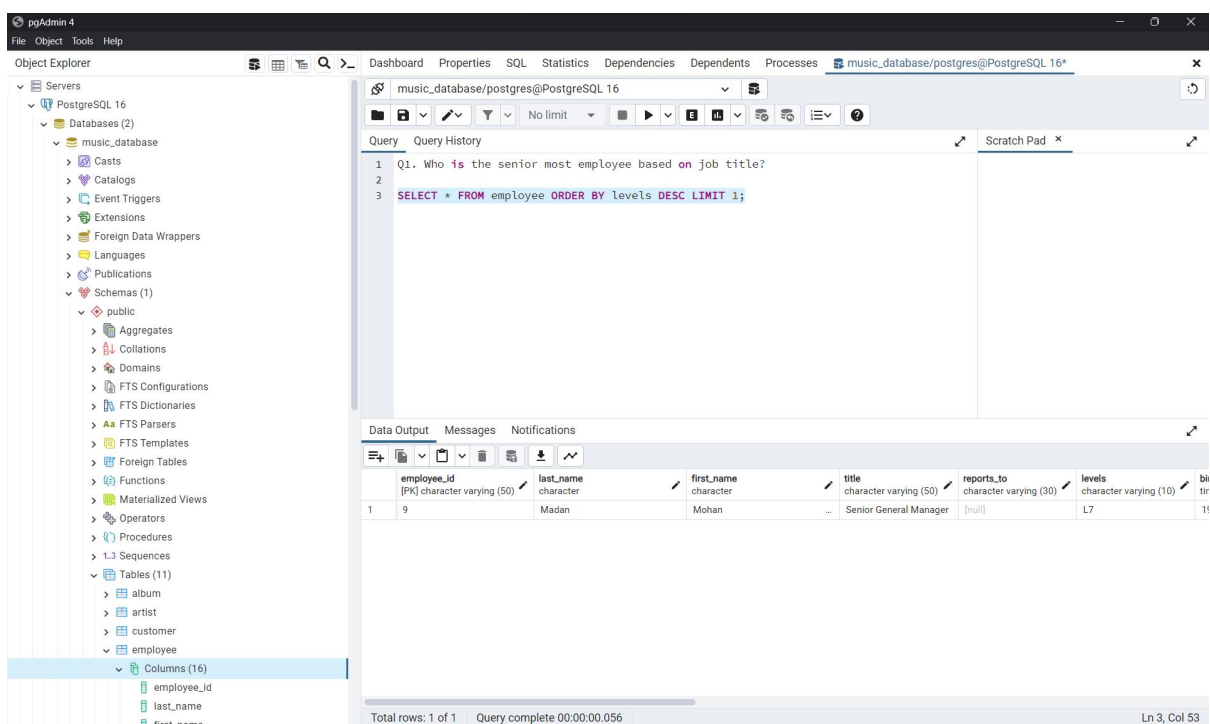
Chapter 4

RESULT/LEARNING OUTCOME

4.1 Assessment Findings

Question Set 1 - Easy

1. Who is the senior most employee based on job title?



The screenshot shows the pgAdmin 4 interface. On the left, the Object Explorer displays the database structure, including the 'employee' table under the 'public' schema. The main pane shows a SQL query: `SELECT * FROM employee ORDER BY levels DESC LIMIT 1;`. The 'Data Output' tab at the bottom displays the result of the query as a table with 7 columns: employee_id, last_name, first_name, title, reports_to, levels, and birth_date. The result shows one row for the employee with ID 9, last name Madan, first name Mohan, title Senior General Manager, reports_to null, levels L7, and birth_date 11/01/1993.

employee_id	last_name	first_name	title	reports_to	levels	birth_date
9	Madan	Mohan	Senior General Manager	[null]	L7	11/01/1993

The provided SQL query aims to identify the senior-most employee within the dataset based on job title. The query utilizes the ORDER BY clause in conjunction with the DESC (descending) keyword to arrange the dataset in a way that the highest job title levels appear first.

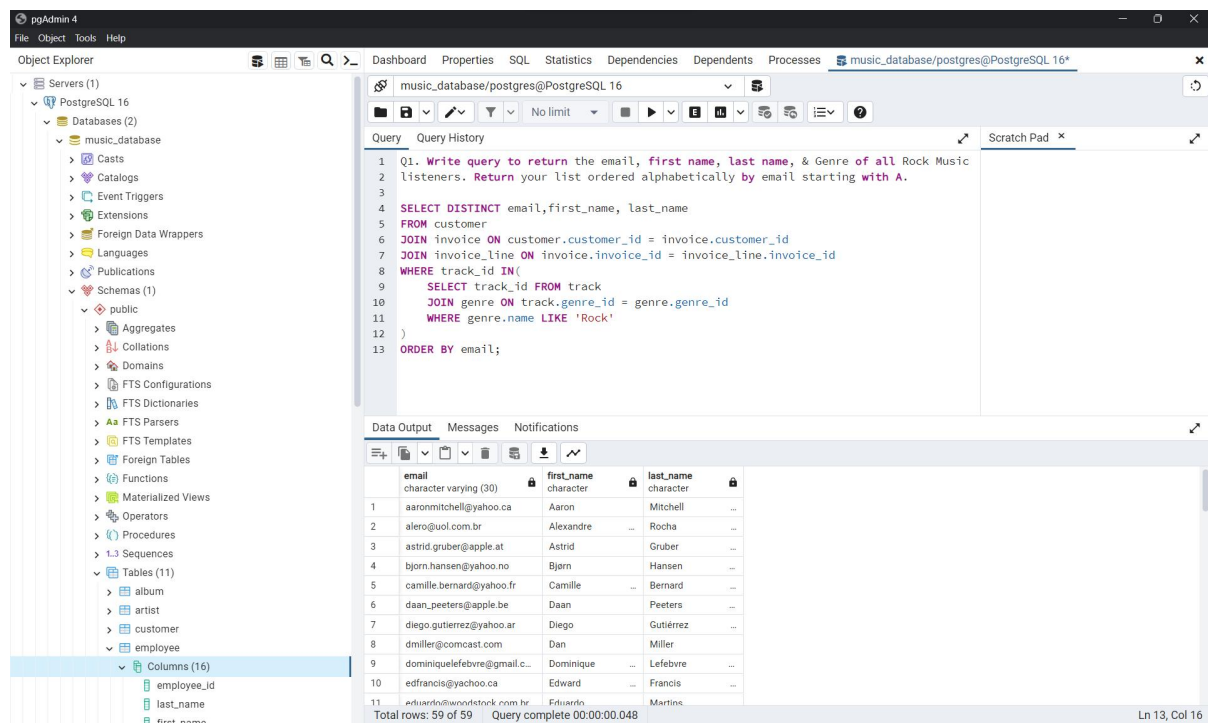
The levels column in the dataset represents the job hierarchy, with higher levels indicating more senior positions. By sorting the dataset in descending order based on job title levels and using the LIMIT 1 clause, the query effectively isolates the topmost record, revealing the employee holding the most senior position in the company.

In essence, this query offers a straightforward solution to determine the senior-most employee, providing valuable insights into the organizational hierarchy based on job titles.

Alongside identifying the senior-most employee, several additional questions were explored, showcasing the versatility of SQL in extracting meaningful insights from the employee dataset.

Question Set 2 – Moderate

1. Write query to return the email, first name, last name, & Genre of all Rock Music listeners. Return your list ordered alphabetically by email starting with A.



The screenshot shows the pgAdmin 4 interface. On the left, the Object Explorer displays the database structure, including the 'public' schema and the 'customer' table. The main pane shows a SQL query window with the following query:

```

1 Q1. Write query to return the email, first name, last name, & Genre of all Rock Music
2 listeners. Return your list ordered alphabetically by email starting with A.
3
4 SELECT DISTINCT email, first_name, last_name
5 FROM customer
6 JOIN invoice ON customer.customer_id = invoice.customer_id
7 JOIN invoice_line ON invoice.invoice_id = invoice_line.invoice_id
8 WHERE track_id IN(
9     SELECT track_id FROM track
10    JOIN genre ON track.genre_id = genre.genre_id
11    WHERE genre.name LIKE 'Rock'
12 )
13 ORDER BY email;

```

The Data Output pane shows the results of the query, displaying a list of customers who have purchased Rock music, ordered alphabetically by email. The results are as follows:

email	first_name	last_name
aaronmitchell@yahoo.ca	Aaron	Mitchell
alero@uol.com.br	Alexandre	Rocha
astrid.gruber@apple.at	Astrid	Gruber
bjorn.hansen@yahoo.no	Bjorn	Hansen
camille.bernard@yahoo.fr	Camille	Bernard
daan.peeters@apple.be	Daan	Peeters
diego.gutierrez@yahoo.ar	Diego	Gutiérrez
dmiller@comcast.com	Dan	Miller
dominiqueldefebvre@gmail.com	Dominique	Lefebvre
edfrancis@yahoo.ca	Edward	Francis
eduardo@wonderstock.com.br	Eduardo	Martins

Total rows: 59 of 59 Query complete 00:00:00.048

The provided SQL query retrieves a distinct list of email addresses, first names, and last names of customers who have purchased Rock music. The query involves several table joins to gather the necessary information:

FROM customer: This is the main table containing customer details, including email, first name, and last name.

JOIN invoice: Connects customer information with invoice data using the `customer_id`.

JOIN invoice_line: Links invoices with the invoice_line table using the `invoice_id`.

JOIN track: Associates invoice_line data with track details through the `track_id`.

JOIN genre: Establishes a connection between tracks and their genres.

The WHERE clause filters the tracks based on the genre condition, selecting only those categorized as 'Rock.' The DISTINCT keyword ensures that each customer's information appears only once in the result set.

The final result is ordered alphabetically by email, starting with 'A,' providing a clear and organized list of Rock music listeners along with their contact details.

Additionally, this query represents just one of several inquiries conducted during the comprehensive assessment of the music store database. Numerous other questions and analyses were undertaken to gain insights into various aspects of the database structure,

customer preferences, and transaction patterns.

Question Set 3 – Advance

1. Find how much amount spent by each customer on artists? Write a query to return customer name, artist name and total spent.

The screenshot shows the pgAdmin 4 interface with a SQL query executed in the 'music_database/postgres@PostgreSQL 16*' database. The query is as follows:

```
1 Q1. Find how much amount spent by each customer on artists? Write a query to return
2 customer name, artist name and total spent.
3
4 WITH best_selling_artist AS (
5     SELECT
6         artist.artist_id AS artist_id,
7         artist.name AS artist_name,
8         SUM(CAST(invoice_line.unit_price AS NUMERIC) * CAST(invoice_line.quantity AS NUMERIC)) AS amount_spent
9     FROM invoice_line
10    JOIN track ON track.track_id = invoice_line.track_id
11    JOIN album ON album.album_id = track.album_id
12    JOIN artist ON artist.artist_id = album.artist_id
13   GROUP BY 1
14   ORDER BY 3 DESC
15   LIMIT 1
16 )
17
```

The results are displayed in a table with 5 columns: customer_id, first_name, last_name, artist_name, and amount_spent. The table contains 11 rows of data.

customer_id	first_name	last_name	artist_name	amount_spent
46	Hugh	O'Reilly	Queen	27.72
38	Niklas	Schröder	Queen	18.81
3	François	Tremblay	Queen	17.82
34	João	Fernandes	Queen	16.83
53	Phil	Hughes	Queen	11.88
41	Marc	Dubois	Queen	11.88
47	Lucas	Mancini	Queen	10.89
33	Ellie	Sullivan	Queen	10.89
20	Dan	Miller	Queen	3.96
5	František	Wichterlová	Queen	3.96
54	Steve	Murray	Queen	2.97

Total rows: 43 of 43 Query complete 00:00:00.072 Ln 8, Col 76

The screenshot shows the pgAdmin 4 interface with a different SQL query executed in the 'music_database/postgres@PostgreSQL 16*' database. The query is as follows:

```
16 )
17
18 SELECT
19     c.customer_id,
20     c.first_name,
21     c.last_name,
22     bsa.artist_name,
23     SUM(CAST(il.unit_price AS NUMERIC) * CAST(il.quantity AS NUMERIC)) AS amount_spent
24 FROM invoice i
25 JOIN customer c ON c.customer_id = i.customer_id
26 JOIN invoice_line il ON il.invoice_id = i.invoice_id
27 JOIN track t ON t.track_id = il.track_id
28 JOIN album alb ON alb.album_id = t.album_id
29 JOIN best_selling_artist bsa ON bsa.artist_id = alb.artist_id
30 GROUP BY 1,2,3,4
31 ORDER BY 5 DESC;
32
```

The results are displayed in a table with 5 columns: customer_id, first_name, last_name, artist_name, and amount_spent. The table contains 11 rows of data, identical to the first screenshot.

customer_id	first_name	last_name	artist_name	amount_spent
46	Hugh	O'Reilly	Queen	27.72
38	Niklas	Schröder	Queen	18.81
3	François	Tremblay	Queen	17.82
34	João	Fernandes	Queen	16.83
53	Phil	Hughes	Queen	11.88
41	Marc	Dubois	Queen	11.88
47	Lucas	Mancini	Queen	10.89
33	Ellie	Sullivan	Queen	10.89
20	Dan	Miller	Queen	3.96
5	František	Wichterlová	Queen	3.96
54	Steve	Murray	Queen	2.97

Total rows: 43 of 43 Query complete 00:00:00.072 Ln 8, Col 76

In PostgreSQL, when performing arithmetic operations on columns, the data types of the columns involved need to be compatible. The solution involved explicitly casting the relevant columns to a compatible data type for arithmetic operations. I used the CAST function to convert the VARCHAR data types of unit_price and quantity columns to the NUMERIC data type, which is suitable for arithmetic calculations. This ensured that the data types are compatible, and the multiplication operation could be performed without any errors.

The provided SQL query aims to determine the amount spent by each customer on the best-selling artist. To achieve this, a common table expression (CTE) named best_selling_artist is first defined, identifying the artist with the highest total sales. This information is derived from joining tables related to invoices, invoice lines, tracks, albums, and artists, grouping the results by the artist's ID and ordering them by total sales in descending order. The CTE is then utilized in the main query, joining the customer, invoice, invoice line, track, and album tables. The query selects the customer's ID, first name, last name, the name of the best-selling artist, and calculates the amount spent by multiplying the unit price and quantity of each invoice line. The results are grouped by customer information and ordered by the total amount spent in descending order. This comprehensive query provides a detailed breakdown of customer spending on the best-selling artist, offering insights into individual customer preferences and contributions to the artist's sales.

This query represents just one facet of the extensive analysis conducted during the project, encompassing a diverse range of questions to unravel insights from the music store database.

4.2 Challenges Faced:

Data Formatting for Import: One notable challenge encountered was the need for meticulous data formatting to facilitate smooth import into the PostgreSQL database. Although the PostgreSQL copy command was available, a personal preference for using import and export methods necessitated addressing the presence of headers in the initial row of the CSV file. This adjustment required a strategic reorganization of the dataset to align with the database schema, ensuring a seamless and error-free import process. The resolution involved refining import techniques to accurately interpret and integrate the dataset, underscoring the importance of aligning personal preferences with optimal database practices.

4.3 Skills Developed

During the internship, a diverse skill set was cultivated, crucial for proficient database management and optimization. The key skills acquired encompassed:

In-Depth Knowledge of Database Management: Gained a comprehensive understanding of common problems encountered during database management, enabling effective analysis and resolution. Explored issues related to data integrity, query performance, and schema design in MySQL, PostgreSQL, and SQL environments.

Hands-On Problem-Solving in MySQL, PostgreSQL, and SQL: Applied problem-solving skills to address challenges specific to MySQL, PostgreSQL, and SQL. This involved troubleshooting and resolving issues related to data formatting, query optimization, and data import/export procedures.

Analysis of Data Quality: Developed skills in assessing and ensuring data integrity, identifying common data issues, and proposing effective measures to maintain a consistent data environment.

Emphasis on Comprehensive Documentation: Acknowledged the importance of comprehensive documentation and best practices, fostering a heightened awareness of the significance of documenting database schemas and structures for future reference.

These skills collectively form a robust foundation for effective database management and are pivotal for addressing real-world challenges in the dynamic field of information technology.

Chapter 5

CONCLUSION

5.1 Summary of Internship Experience

My 15-day internship was a transformative period, blending foundational learning and hands-on application in the database management sphere. The initial week immersed me in SQL, MySQL, and PostgreSQL fundamentals, setting the stage for the subsequent evaluation of a music store database.

The assessment phase included a detailed review of the database schema, meticulous analysis of data types and constraints, and strategic query optimization. Practical challenges, like formatting data for import/export, provided valuable problem-solving experiences.

Collaboration with the Computer Science Department on-site enhanced communication, compensating for the absence of virtual tools. The internship concluded with a comprehensive report, encapsulating findings and reflections.

Developed skills encompass database schema design, troubleshooting common issues, and effective communication. The experience not only applied theoretical knowledge but also instilled adaptability and real-world problem-solving.

This internship marks a significant milestone, contributing to the journey of becoming a proficient and versatile database professional.

5.2 Achievements and Milestones

The internship yielded notable achievements and milestones in advancing my database management skills. Successfully evaluating and optimizing the music store database showcased practical expertise. Overcoming challenges, such as data formatting nuances, underscored adaptability and problem-solving capabilities.

Collaborating on-site with the Computer Science Department fostered effective communication and enriched the overall learning experience. The comprehensive report encapsulates the journey, serving as a tangible milestone in synthesizing theoretical knowledge into practical application.

The experience stands as a testament to continuous learning and growth, highlighting achievements that contribute to a solid foundation for future endeavors in the dynamic field of database management.

5.3 Closing Remarks

The 15-day internship has been an invaluable journey marked by immersive learning, practical application, and skill development. Engaging with real-world challenges in the context of a music store database provided a rich platform for honing database management expertise. The experience underscored the importance of adaptability, problem-solving, and meticulous attention to detail in ensuring the efficiency and integrity of a database system. I am grateful for the guidance, support, and collaborative spirit of the Computer Science Department team, which significantly contributed to the success of this endeavor. This internship has not only expanded my technical proficiency but has also instilled a deep appreciation for the intricacies of database management in a professional setting. As I conclude this internship, I carry forward a wealth of experiences and insights that will undoubtedly shape my future endeavors in the field of database management.