D326 - Advanced Data Management
Ivy Le
010681661

Part A:

Summarize one real-world written business report that can be created from the DVD Dataset

Mary and John, the owners of CineFlix Rentals, are facing a challenging situation due to the increasing popularity of streaming services. As more people opt for digital streaming platforms, the demand for DVDs has declined significantly.

Due to this changing market, they are planning to implement a loyalty program to incentivize customers to continue renting from them. To launch their program, ***Mary and John need to identify the top 10 customers who have rented the most DVDs.*** By pinpointing these loyal and frequent customers, they can design a targeted loyalty program that acknowledges and appreciates their continued patronage, while also attracting new customers to experience the benefits of renting from CineFlix.

---

1) Identify the specific fields that will be included in the detailed table and summary table:

Detailed Table

Column Name
customer_id
first_name
last_name
email
rental_id

Summary Table

Column Name
customer_name
email
total_rental

---

2) Describe the types of data fields used for the report:

Column Name	Description
customer_id	INT - Unique identifier to one customer
first_name	VARCHAR(45) - Customer's First Name
last_name	VARCHAR(45) - Customer's Last Name
email	VARCHAR(90) - Customer's Email Address
rental_id	INT - Unique identifier to one rental
customer_name	VARCHAR(90) - Combination of First and Last name
email	VARCHAR(90) - Customer's Email Address
total_rental	INT - Total amount of rentals for one customer

---

3) Identify at least two specific tables from the given dataset that will provide the data necessary for the detailed table section and the summary table section of the report.

The specific tables I will be using from the dvdrental sample database are: Rental and Customer

---

4) Identify at least one field in the detailed table section that will require a custom transformation with a user-defined function and explain why it should be transformed

The ‘first_name’ and ‘last_name’ fields from the detailed table will be required to undergo a transformation where they will be combined and populated into the ‘customer_name’ field of the summary table. The format used will be Last_name, First_Name. This will promote scalability for Mary and John, by ensuring accurate and readily accessible customer names as their customer data set grows.

---

5) Explain the different business uses of the detailed table section and the summary table section of the report.

The detailed table contains comprehensive customer information, including first name, last name, email address, rental history, and more. This offers CineFlix Rentals valuable insight to make informed decisions and grants opportunities to enhance their services.

In contrast, the summary table provides a high-level view of customer information by displaying the customer's full name, their email address, and the total amount of DVDs they have rented. CineFlix Rentals can utilize this table in efforts to facilitate the creation of a customer loyalty program. It is also crucial in marketing strategies by offering insights into customer’s behaviors, it can create effective marketing campaigns for their loyalty program.

---

6) Explain how frequently your report should be refreshed to remain relevant to stakeholders.

The ideal time frame for updated reports should be initially more frequent during the launching phase, slowly it can be transitioned to a monthly schedule as the program is fully implemented and when everyone becomes accustomed to it. This allows them to respond to changing dynamics, meet customer needs, and achieve program goals.

PART C:

Provide original SQL code in a text format that creates the detailed and summary tables to hold your report table sections.

--1. Create Tables (Detailed and Summary)

```
CREATE TABLE detailed_table (
        customer_id INT,
        first_name VARCHAR(45),
        last_name VARCHAR(45),
        email VARCHAR(90),
        rental_id INT
);
```

```
CREATE TABLE summary_table (
        customer_name VARCHAR(90),
        email VARCHAR(90),
        total_rental INT
);
```

PART D:

Provide an original SQL query in a text format that will extract the raw data needed for the detailed section of your report from the source database.

```
INSERT INTO detailed_table (
        customer_id,
        first_name,
        last_name,
        email,
        rental_id
)

SELECT c.customer_id, c.first_name, c.last_name, c.email, r.rental_id
FROM rental AS r
INNER JOIN customer AS c
ON r.customer_id = c.customer_id
;
```

Provide original code for function(s) in text format that perform the transformation(s) you identified in part A4.

```
CREATE FUNCTION update_summary()
RETURNS TRIGGER
LANGUAGE PLPGSQL
AS $$
BEGIN
DELETE FROM summary_table;
INSERT INTO summary_table(
        SELECT
                concat_ws(', last_name,first_name) AS customer_name,
                email, COUNT(rental_id) AS total_rental
        FROM detailed_table
        GROUP BY first_name, last_name, email
        ORDER BY total_rental DESC
        LIMIT 10
);
RETURN NEW;
END;
$$
```

Provide original SQL code in a text format that creates a trigger on the detailed table of the report that will continually update the summary table as data is added to the detailed table.

```
CREATE TRIGGER update_summary_tr
AFTER INSERT ON detailed_table
FOR EACH STATEMENT
EXECUTE PROCEDURE update_summary();

CREATE PROCEDURE update_tables()
LANGUAGE PLPGSQL
AS $$
BEGIN

DELETE FROM detailed_table;
INSERT INTO detailed_table(
        customer_id,
        first_name,
        last_name,
        email,
        rental_id
)
SELECT
        c.customer_id, c.first_name, c.last_name, c.email, r.rental_id
FROM rental AS r
INNER JOIN customer AS c
ON r.customer_id = c.customer_id;
END;
$$
```

Provide an original stored procedure in a text format that can be used to refresh the data in both the detailed table and summary table. The procedure should clear the contents of the detailed table and summary table and perform the raw data extraction from part D.

```
CALL update_tables();
SELECT *
FROM detailed_table;
SELECT *
FROM summary_table;
```

1. Identify a relevant job scheduling tool that can be used to automate the stored procedure.

“pg_cron” is a notable scheduling tool that can automate the stored procedure.

H. Acknowledge all utilized sources, including any sources of third-party code, using in-text citations and references. If no sources are used, clearly declare that no sources were used to support your submission.

I did not utilize any third-party code or sources to support my submission.