Tecnológico de Costa Rica

Escuela de Ingeniería en Computación

IC4302 Bases de Datos II

Grupo 20

Proyecto 2: Replicación y análisis de datos
DVD Rental

Profesor:

Alberto Shum Chan

Estudiantes:

Andrés Felipe Arias Corrales
2015028947

Diego Esteban Castro Chaves
200419896

Fabián José Fernández Fernández
2022144383

Fecha de Entrega: 18 de Octubre del 2024

II semestre, 2024

Contents

1 Introduction

This project involves the development of a DVD rental system using PostgreSQL and Business Intelligence (BI) techniques. The system includes data replication, dimensional modeling, and data visualization. In this document, we will describe the process followed for setting up the environment, implementing the database schema, and creating the required data visualizations using Tableau.

2 Objectives

The main objectives of this project are:

- Implement a replication environment separating the transactional and multidimensional models.

- Design and implement a dimensional model for a data mart.

- Set up a PostgreSQL replica and load data into the dimensional model for OLAP queries.

- Create a dashboard using Tableau to visualize key metrics such as the number of rentals and total revenue.

3 Database Design

The transactional database model includes the following tables: *actor*, *film*, *film_actor*, *category*, *film_category*, *store*, *inventory*, *rental*, *payment*, *staff*, *customer*, *address*, *city*, and *country*. Each table was created following the normalization rules and includes appropriate primary and foreign keys.

3.1 ER Diagram

Below is the Entity-Relationship Diagram for the database:

Figure 1: Entity-Relationship Diagram of the DVD Rental Database

3.2 Table Definitions

The following section includes a description of each table along with the attributes and constraints:

Column Name	Data Type	Description
actor_id	SERIAL	Primary key for actors
first_name	VARCHAR(50)	Actor's first name
last_name	VARCHAR(50)	Actor's last name
last_update	TIMESTAMP	Last update timestamp

Table 1: Actor Table Definition

4 Security Configuration

We have implemented the following security measures:

- **Roles**: Two roles were created: *EMP* for employees with limited access and *ADMIN* for administrators with extended privileges.

- **Users**: Three users were created: *video*, *empleado1*, and *administrador1*, each assigned their respective roles.

- All stored procedures are configured to execute with the privileges of the *video* user, ensuring restricted access.

5 Dimensional Model

The dimensional model is designed following the star schema and includes the following fact table and dimensions:

5.1 Fact Table: Rentals

The fact table includes measures such as:

- **Number of Rentals**
- **Total Revenue from Rentals**

5.2 Dimensions

- **Film Dimension**: Hierarchy of category, film, and actors.
- **Location Dimension**: Country and city hierarchy.
- **Date Dimension**: Year, month, and day.
- **Store Dimension**: Store information.

6 Replication Process

We set up a replica of the PostgreSQL database to offload OLAP queries to the slave instance. The following steps were followed to configure replication:

1. Installed and configured PostgreSQL replication settings.
2. Verified data synchronization between the master and slave instances.
3. Documented the steps and tested the configuration.

7 Data Visualization with Tableau

Using Tableau, we created several visualizations to analyze key metrics. The following visualizations were included in the dashboard:

- Rentals and revenue per store, per month.
- Total rentals per film category.

- Top 10 actors by rental revenue.

- A map visualization showing revenue by city.

8 Conclusion

This project allowed us to implement a complete data solution that integrates a transactional system, replication, and business intelligence. We successfully implemented a dimensional model and created meaningful visualizations to assist in decision-making processes.