2022

Automating Data Integration

D191 Performance Assessment

Lydia Strough

1. *Business Question*: Total rentals for each store organized by film category.
   1. *Describe the data used for the report*:   
        
      The report will be composed of two tables, a “detailed\_report” and a “summary\_report” table. The detailed\_report table will be a compiled list of all store rentals. The summary\_report table will be an aggregated version of the detailed table – total rentals per store, grouped by store id and film category.
   2. *Identify two or more specific tables from the given dataset that will provide the data necessary for the detailed and the summary sections of the report*:   
        
      Five tables from the dvdrental database will be used to create the detailed\_report and summary\_report table – the rental, inventory, film, film\_category, and category tables.
   3. *Identify the specific fields that will be included in the detailed and the summary sections of the repor*t:  
        
       The fields included in each table are as follows:

detailed\_report table fields: rental\_id INT, film\_category\_name varchar(45), store\_id INT, rental\_date DATE

summary\_report table fields: store\_id INT, film\_category\_name varchar(45), num\_rentals INT

* 1. *Identify one field in the detailed section that will require a custom transformation and explain why it should be transformed*:   
       
     store\_id INT will require a custom transformation. With a custom transformation, this will organize the film rental categories by store\_id, making it simple for company staff to pinpoint data about specific stores.
  2. *Explain the different business uses of the detailed and the summary sections of the report*:   
       
     The detailed\_report table can be used to identify whether there is a correlation between the popularity of specific genres of films and the time of the year. The summary\_report table can be used to discern the most popular genre of film rented throughout the course of the year. For both tables, once this information is identified, it can be used by the business to adjust company inventory to suit customer demand, and in turn, increase company profit margins as desired.
  3. *Explain how frequently your report should be refreshed to remain relevant to stakeholders*:

In a real-world situation, this report/process would ideally be refreshed on a monthly basis (at the end of the month) in order to gain a variety of data pertinent to customer desires over time.

1. *Write a SQL code that creates the tables to hold your report sections*:

Background pattern

Description automatically generated with low confidence

1. *Write a SQL query that will extract the raw data needed for the detailed section of your report from the source database and verify the data’s accuracy*:  
     
   Graphical user interface, text, application, email

   Description automatically generated
2. *Write code for function(s) that perform the transformation(s) you identified in A4*:  
   Background pattern

   Description automatically generated with low confidence
3. *Write a SQL code 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*:   
     
   
4. *Create a stored procedure that can be used to refresh the data in both your detailed and summary tables*:  
     
   Graphical user interface, text, application, email

   Description automatically generated
   1. *Explain how the stored procedure can be run on a schedule to ensure data freshness*:   
        
      The refresh\_reports() stored procedure can be run on a monthly basis (at the end of every month), creating a fresh detailed\_report table and summary\_report table for the next month. An external tool, such as the pgAgent application, can be used as a scheduling tool for this procedure.
5. *Provide a Panopto video recording that includes a demonstration of the functionality of the code used for the analysis and a summary of the programming environment*:

Panopto video recording link:

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=0c256ac7-687b-402b-8b60-ae4c01820b8e>

Summary of programming environment:

* My Operating system is MacOS.
* Relational DBMS used is PostgreSQL
* GUI used is pgAdmin4
* Database used is dvdrental

1. *Record the web sources you used to acquire data or segments of third-party code to support the application if applicable*:

No sources or third-party code were used in the making of this project.

1. *Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased or summarized*:

No sources or third-party code were used in the making of this project.