**Database Querying in SQL**

1. **Refining Your Query:** You need to get some data from the “film” table and decide to use the query SELECT \* FROM film.
   * You realize that only the “film\_id” and “title” columns are needed. Write a new query that selects only those 2 columns.
   * Compare the cost of the original query and the revised query, and write a few sentences explaining the comparison. Can you suggest any ways to optimize this query?

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

From the data output message the cost of the original query and the revised query are the same (cost=0.00..64.00). Although they have different runtimes as shown in each of their output. The best way to optimise and safe cost is to create a script.

1. **Ordering the Data:**

* In the pgAdmin Query Tool, run a query that selects every film from the “film” table, with the movies sorted by title from A to Z, then by most recent release year, and then by highest to lowest rental rate.

Graphical user interface, text, application, Word

Description automatically generated

* Extract the data output of your query into a csv file for the film collection department to analyze in Excel. To do this, click the button “Save results to file”:

1. **Grouping Data:** The strategy department has asked you the questions below. Write a SQL query to retrieve the correct answers, then extract your results as a csv file.
   * What is the average rental rate for each rating category?

Graphical user interface, text, application, Word

Description automatically generated

* + What are the minimum and maximum rental durations for each rating category?

Graphical user interface, application, table

Description automatically generated

1. **Database Migration:** Your team has decided to use an external tool to collect data on user behavior in the new Rockbuster Android app. Data collected from this new source will need to be loaded into the data warehouse before you can analyze it.
   * **Can you outline the procedure for migrating the data and who will be responsible for it?**

**1. Extract:** This the first step and it involves collection of data from various data sources .

**2. Transform:** In this step, the extracted data is converted into another format. This could mean calculating ages from dates of birth or combining multiple data points like area codes and telephone numbers to get a contact number, for example.

**3. Load:** In this step, the transformed data is inserted or loaded into the new database.

* + **What problems do you foresee if you start analyzing the data before it’s been loaded into the data warehouse?**

We can run the risk of working with unclean and incompletely formatted data which in turn can lead to an inaccurate analysis.

**Bonus Task**

You’ve not yet covered custom sorting; however, let’s imagine you’ve found the two resources below that explain it. Read each one, then try to write a query to answer the following question: What are the minimum and the maximum replacement costs for each rating category ordered by rating as follows: G, PG, PG-13, R, NC-17?

Graphical user interface, text, application

Description automatically generated