

Exercise 3.4

May 26, 2022

STEP 1

Refining Your Query: You need to get some data from the “film” table and decide to use the query **SELECT * FROM film.**

EXPLAIN

SELECT *

FROM film

"Seq Scan on film (cost=0.00..64.00 rows=1000 width=388)"

You realize that only the “film_id” and “title” columns are needed. Write a new query that selects only those 2 columns.

EXPLAIN

SELECT film_id, title

FROM film

"Seq Scan on film (cost=0.00..64.00 rows=1000 width=19)"

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?

Looking at the two EXPLANATIONS, we see that the first part is the same, the difference lies in the “width”, * = 388, while film_id & title = 19. To optimize, it’s important to know what data you’re looking for, this will save time.

STEP 2

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.
- Extract the data output of your query into a csv file for the film collection department to analyze in Excel. (You may need to explore how to save your output as a csv file in the Query Tool.)

STEP 3

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

```
SELECT rating, AVG(rental_rate)
```

```
FROM film
```

```
GROUP BY rating
```

```
SELECT rating, MIN(rental_duration), MAX(rental_duration)
```

```
FROM film
```

```
GROUP BY rating
```

STEP 4

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?

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

The procedure for migrating the data to the data warehouse would begin with collecting the data; this step was performed by the Android app. Once the data has been collected, the data will need to be transformed so that the data from the different sources are in the same format, for example ensuring dates are all either month/day/year or day/month/year. The final step is loading the data into the warehouse. These steps are usually completed by a data engineer.

If one were to analyze the data before it's been properly loaded, you will be working with incomplete data, this means that some things will be missing or you will receive error messages, some data may also be duplicated and skew any analysis.