3.4: Database Querying in SQL

Innocent Bayai (23/5/24)

Question 1. Refining the query

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

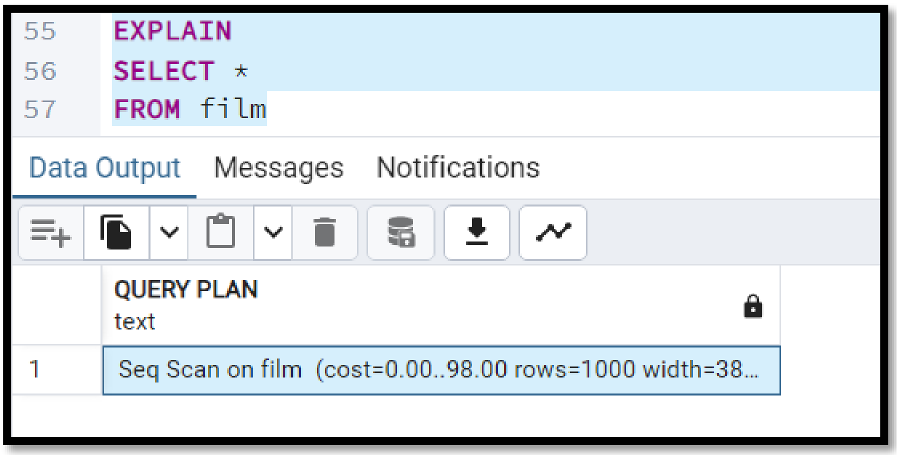
a. **You realize that only the “film_id” and “title” columns are needed. Write a new query that selects only those 2 columns.**

New query:

SELECT film_id, title
FROM film

b. **Compare the cost of the original query and the revised query**

Cost of original query

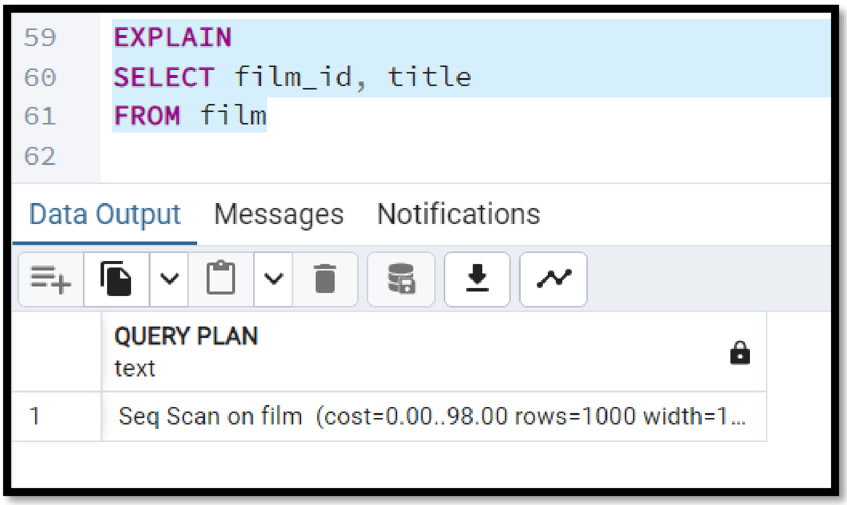


```
EXPLAIN
SELECT *
FROM film
```

	QUERY PLAN text
1	Seq Scan on film (cost=0.00..98.00 rows=1000 width=38...

Full cost explanation: *Seq Scan on film (cost=0.00..98.00 rows=1000 width=384)*

Cost of new query



```
EXPLAIN
SELECT film_id, title
FROM film

```

	QUERY PLAN text
1	Seq Scan on film (cost=0.00..98.00 rows=1000 width=1...

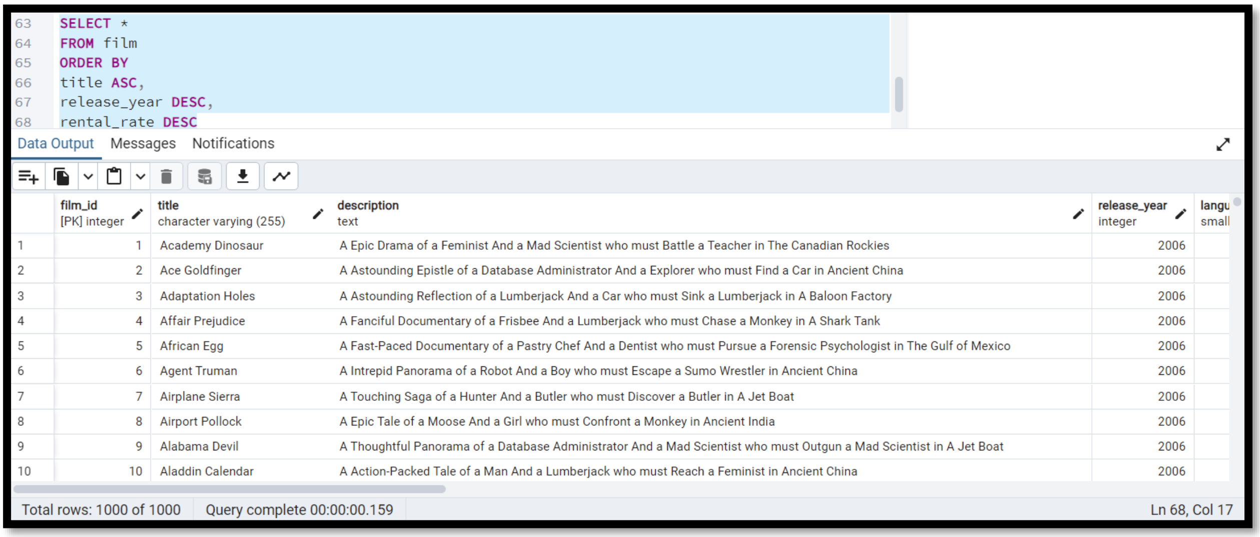
Full cost explanation: *Seq Scan on film (cost=0.00..98.00 rows=1000 width=19)*

Write a few sentences explaining the comparison. Can you suggest any ways to optimize this query?

The cost/time for the two queries is the same, that is time to return the first row is 0.00 and it costs 98.00 to return all the rows. However, the width for the first query is 384 compared to 19 for the second query. This means optimization must limit the number of rows the query is supposed to work with to pick the records required.

Question 2: Ordering the Data

a. 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.



```
SELECT *
FROM film
ORDER BY
title ASC,
release_year DESC,
rental_rate DESC
```

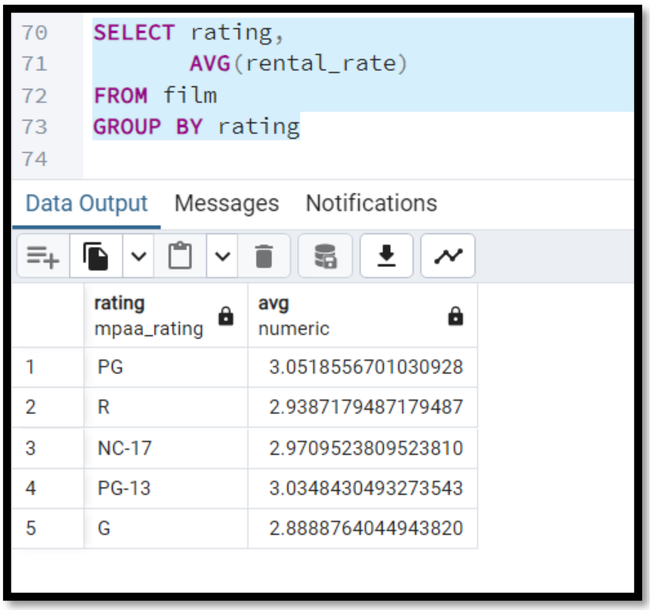
	film_id [PK] integer	title character varying (255)	description text	release_year integer	langu small
1	1	Academy Dinosaur	A Epic Drama of a Feminist And a Mad Scientist who must Battle a Teacher in The Canadian Rockies	2006	
2	2	Ace Goldfinger	A Astounding Epistle of a Database Administrator And a Explorer who must Find a Car in Ancient China	2006	
3	3	Adaptation Holes	A Astounding Reflection of a Lumberjack And a Car who must Sink a Lumberjack in A Baloon Factory	2006	
4	4	Affair Prejudice	A Fanciful Documentary of a Frisbee And a Lumberjack who must Chase a Monkey in A Shark Tank	2006	
5	5	African Egg	A Fast-Paced Documentary of a Pastry Chef And a Dentist who must Pursue a Forensic Psychologist in The Gulf of Mexico	2006	
6	6	Agent Truman	A Intrepid Panorama of a Robot And a Boy who must Escape a Sumo Wrestler in Ancient China	2006	
7	7	Airplane Sierra	A Touching Saga of a Hunter And a Butler who must Discover a Butler in A Jet Boat	2006	
8	8	Airport Pollock	A Epic Tale of a Moose And a Girl who must Confront a Monkey in Ancient India	2006	
9	9	Alabama Devil	A Thoughtful Panorama of a Database Administrator And a Mad Scientist who must Outgun a Mad Scientist in A Jet Boat	2006	
10	10	Aladdin Calendar	A Action-Packed Tale of a Man And a Lumberjack who must Reach a Feminist in Ancient China	2006	

Total rows: 1000 of 1000 Query complete 00:00:00.159 Ln 68, Col 17

Question 3. Grouping Data

Write a SQL query to retrieve the correct answers, then extract your results as a CSV file.

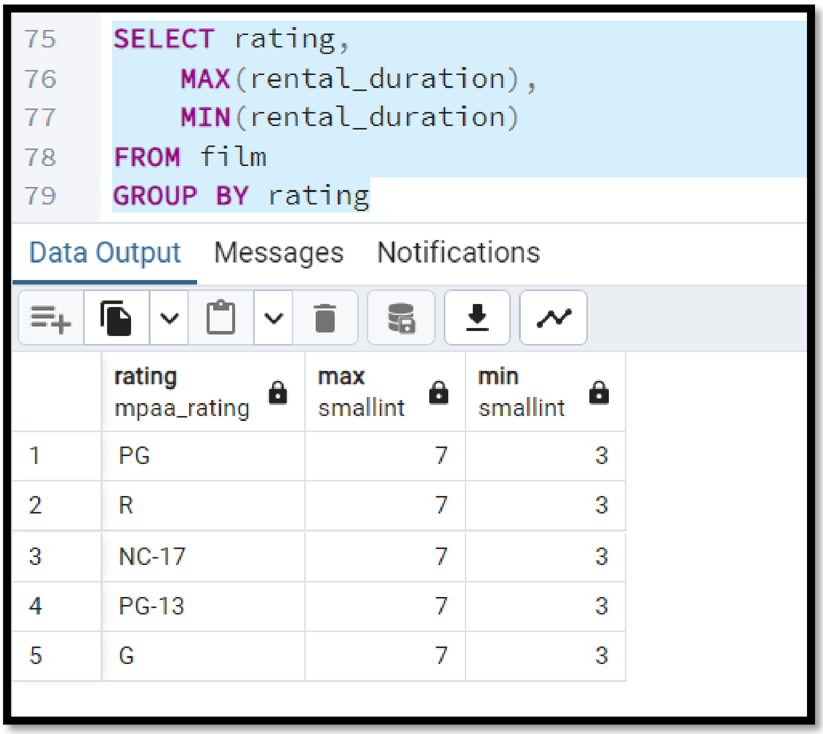
a. What is the average rental rate for each rating category?



```
SELECT rating,
       AVG(rental_rate)
FROM film
GROUP BY rating
```

	rating mpaa_rating	avg numeric
1	PG	3.0518556701030928
2	R	2.9387179487179487
3	NC-17	2.9709523809523810
4	PG-13	3.0348430493273543
5	G	2.8888764044943820

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



```sql
SELECT rating,
    MAX(rental_duration),
    MIN(rental_duration)
FROM film
GROUP BY rating
```

	rating mpaa_rating	max smallint	min smallint
1	PG	7	3
2	R	7	3
3	NC-17	7	3
4	PG-13	7	3
5	G	7	3

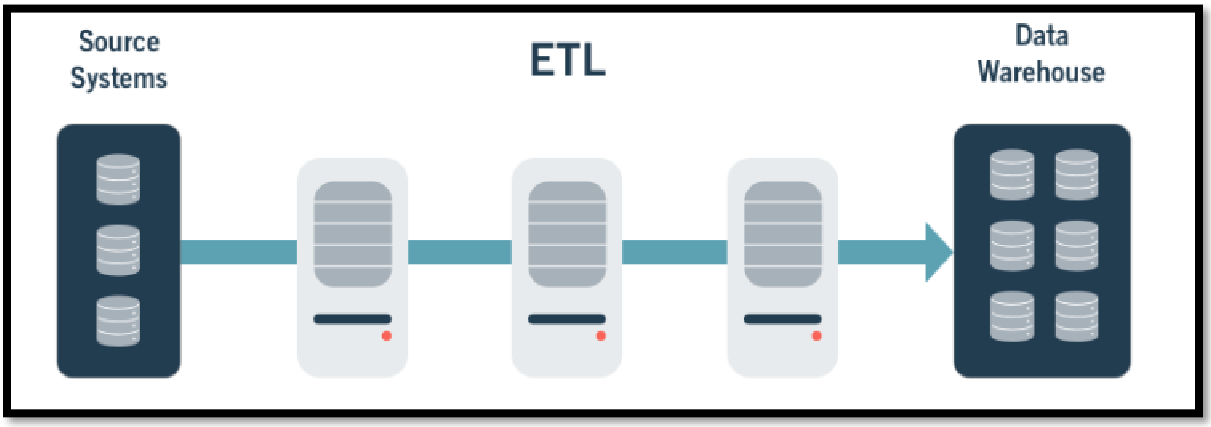
Question 4: Data Migration

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.

a. **Can you outline the procedure for migrating the data and who will be responsible for it?**

- Extraction includes picking the user behavior data from the Rockbuster Android App.
- Transformation involves changing the format of the data and presenting it in a format that gives the right user details required.
- Loading involves inserting of the transformed data into the database.

The summarized visualization of the data migration process is presented hereunder.



Data migration is mostly the pre-occupation of data engineers although a data analyst still needs to understand the process as role might switch along way. Also, a data analyst must know of this process so that they are privy of the process behind the data they are using and can make amendments to the data if the need arises.

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

Analyzing the data before it is loaded implies that the behaviour data won't be connected or linked to the main database where other variables of concern are. The analysis becomes one independent analysis not integrated with the other.