

## 3.4 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?

### ANSWER

- a. Select \*

1	EXPLAIN
2	SELECT *
3	FROM film;

QUERY PLAN	
	text
1	Seq Scan on film (cost=0.00..64.00 rows=1000 width=384)

- b. Select film\_id, title

EXPLAIN
SELECT film_id, title
FROM film;

QUERY PLAN	
	text
1	Seq Scan on film (cost=0.00..64.00 rows=1000 width=19)

In both cases the cost of returning is 0 and the cost is 64. Although the width of each query is different which makes the query faster when you limit the columns to what you actually need. Besides, if you think in creating a longer script with more functions, the best would be to optimize the script from the beginning to make it clear, short and faster. In general being more accurate in the information you need makes the query faster and cheaper.

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. To do this, click the button “Save results to file”:

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

	title character varying (255) 🔒	release_year integer 🔒	rental_rate numeric (4,2) 🔒
1	Academy Dinosaur	2006	0.99
2	Ace Goldfinger	2006	4.99
3	Adaptation Holes	2006	2.99
4	Affair Prejudice	2006	2.99

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?

```
SELECT rating, round( AVG (rental_rate),2)
FROM film
GROUP BY rating;
```

	rating mpaa_rating 🔒	round numeric 🔒
1	PG	3.05
2	R	2.94
3	NC-17	2.97
4	PG-13	3.03
5	R	2.89

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

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

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

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?

According to the process for extracting data (ETL):

Android app > Data converted into data warehouse format > Data warehouse

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

- Data type is not determined equal, at the end the data analyzed is not compatible with the data warehouse structure.
- Make the process longer, since the data will be already processed it could take longer the process of extracting and loading.