

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

Dashboard Properties SQL Statistics De

Rockbuster/postgres@PostgreSQL 15

Query Query History

```
1 SELECT film_id, title
2 FROM film
```

Data Output Messages Notifications

	film_id [PK] integer	title character varying (255)
1	133	Chamber Italian
2	384	Grosse Wonderful
3	8	Airport Pollock
4	98	Bright Encounters
5	1	Academy Dinosaur
6	2	Ace Goldfinger
7	3	Adaptation Holes
8	4	Affair Prejudice
9	5	African Egg
10	6	Agent Truman
11	7	Airplane Sierra
12	9	Alabama Devil
13	10	Aladdin Calendar
14	11	Alamo Videotape
15	12	Alaska Phantom
16	213	Date Speed
17	13	Ali Forever
18	14	Alice Fantasia
19	15	Alien Center
20	16	Alley Evolution

Dashboard Properties SQL Statistics Dependenci

Rockbuster/postgres@PostgreSQL 15

Query Query History

```
1 EXPLAIN
2 SELECT film_id, title
3 FROM film
```

Data Output Messages Notifications

QUERY PLAN

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

Query Query History

```
1 SELECT *
2 FROM film
```

Data Output Messages Notifications

	film_id [PK] integer	title character varying (255)	description text
1	133	Chamber Italian	A Fateful Ri
2	384	Grosse Wonderful	A Epic Dran
3	8	Airport Pollock	A Epic Tale
4	98	Bright Encounters	A Fateful Yi
5	1	Academy Dinosaur	A Epic Dran
6	2	Ace Goldfinger	A Astoundir
7	3	Adaptation Holes	A Astoundir
8	4	Affair Prejudice	A Fanciful L
9	5	African Egg	A Fast-Pace
10	6	Agent Truman	A Intrepid P
11	7	Airplane Sierra	A Touching
12	9	Alabama Devil	A Thoughtfi
13	10	Aladdin Calendar	A Action-Pa
14	11	Alamo Videotape	A Boring Ep
15	12	Alaska Phantom	A Fanciful S
16	213	Date Speed	A Touching
17	13	Ali Forever	A Action-Pa
18	14	Alice Fantasia	A Emotiona
19	15	Alien Center	A Brilliant D

Dashboard Properties SQL Statistics Dependencies

Rockbuster/postgres@PostgreSQL 15

Query Query History

```
1 EXPLAIN
2 SELECT *
3 FROM film
```

Data Output Messages Notifications

QUERY PLAN

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

- 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?

Data shows that the cost for both queries is 64, but we see a difference on width, the first one resulting on 19 and the second one resulting on 384. This could be optimized by adding the LIMIT query.

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.

Dashboard Properties SQL Statistics Dependencies Depen

Rockbuster/postgres@PostgreSQL 15

Query Query History

```

1 SELECT title,
2   release_year,
3   rental_rate
4 FROM film
5 ORDER BY title ASC,
6   release_year,
7   rental_rate DESC

```

Data Output Messages Notifications

	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
5	African Egg	2006	2.99
6	Agent Truman	2006	2.99
7	Airplane Sierra	2006	4.99
8	Airport Pollock	2006	4.99
9	Alabama Devil	2006	2.99
10	Aladdin Calendar	2006	4.99
11	Alamo Videotape	2006	0.99
12	Alaska Phantom	2006	0.99
13	Ali Forever	2006	4.99
14	Alice Fantasia	2006	0.99
15	Alien Center	2006	2.99
16	Alley Evolution	2006	2.99
17	Alone Trip	2006	0.99
18	Alter Victory	2006	0.99
Total rows: 1000 of 1004 Query complete 00:00:00.163			

Dashboard Properties SQL Statistics Dependencies

Rockbuster/postgres@PostgreSQL 15

Query Query History

```

1 EXPLAIN
2 SELECT title,
3   release_year,
4   rental_rate
5 FROM film
6 ORDER BY title ASC,
7   release_year,
8   rental_rate DESC

```

Data Output Messages Notifications

QUERY PLAN

text

1	Sort (cost=113.83..116.33 rows=1000 width=25)
2	Sort Key: title, release_year, rental_rate DESC
3	-> Seq Scan on film (cost=0.00..64.00 rows=1000 width=25)

- 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”:

DONE

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?

Dashboard Properties SQL Statistics Dependencies Dependents Processes

Rockbuster/postgres@PostgreSQL 15

Query Query History

```

1 SELECT rating,
2 AVG(rental_duration) AS average_rental_duration,
3 MIN(rental_duration) AS Min_rental_duration,
4 MAX(rental_duration) AS Max_rental_duration
5 FROM Film
6 GROUP BY rating
7 ORDER BY AVG(rental_duration)

```

Data Output Messages Notifications

rating	mpaa_rating	average_rental_duration	min_rental_duration	max_rental_duration
		numeric	smallint	smallint
1	R	4.774358974358974	3	7
2	G	4.7967032967032965	3	7
3	PG-13	5.053811659192825	3	7
4	PG	5.082474226804123	3	7
5	NC-17	5.142857142857143	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?

The correct procedure to follow during a data migration process would be applying the method of Extraction, Transformation and Loading, better known as ETL. During extraction, this process involves collecting data from multiple sources. Followed by transform, which is when data that was extracted is converted into another format. Lastly, the transformed data is 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?

If data isn't loaded correctly, it won't be consistent and hard to follow. This would make it difficult to analyze and retrieve information.

5. Save your "Answers 3.4" document as a pdf (with screenshots) and your CSV files as a single .xlsx Excel file and upload it here for your tutor to review.