

3.4 Answers- Database Querying

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.
 - Both queries have the same cost, but the 2nd one took less time to process and that’s because it selected less columns.

The screenshot shows a database query interface with a query editor and a query plan. The query editor contains the following SQL:

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

The query plan is displayed in the "Data output" tab. It shows a single step: "Seq Scan on film (cost=0.00..64.00 rows=1000 width=...)". The status bar at the bottom indicates "Total rows: 1 of 1" and "Query complete 00:00:00.157".

The screenshot shows a database query interface with a query editor and a query plan. The query editor contains the following SQL:

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

The query plan is displayed in the "Data output" tab. It shows a single step: "Seq Scan on film (cost=0.00..64.00 rows=1000 width=19)". The status bar at the bottom indicates "Total rows: 1 of 1" and "Query complete 00:00:00.151".

2. **Ordering the Data** - Movies sorted by title from A to Z, then by most recent release year, and then by highest to lowest rental rate.

The screenshot shows a database query interface with a query editor and a query plan. The query editor contains the following SQL:

```
1 SELECT title, release_year, rental_rate
2 FROM film
3 ORDER BY title, release_year, rental_rate DESC
4
```

The query plan is displayed in the "Data output" tab. It shows a single step: "Seq Scan on film (cost=0.00..64.00 rows=1000 width=...)". The status bar at the bottom indicates "Total rows: 1000 of 1000" and "Query complete 00:00:00.058".

	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

[filmcd.csv](#)

3. Grouping Data - What is the average rental rate for each rating category? What are the minimum and maximum rental durations for each rating category?

Query		Query History	
1	SELECT	rating,	
2	AVG	(rental_rate) AS avg_rental_rate	
3	FROM	film	
4	GROUP BY	rating	
5			
6			

Data output		Messages	Notifications
rating	avg_rental_rate		
mpaa_rating	numeric		
1	G		
2	R		
3	PG-13		
4	NC-17		
5	PG		

Query		Query History	
1	SELECT	rating,	
2	MIN	(rental_duration) AS min_rental_duration,	
3	MAX	(rental_duration) AS max_rental_duration	
4	FROM	film	
5	GROUP BY	rating	
6			

Data output		Messages	Notifications
rating	min_rental_duration	max_rental_duration	
mpaa_rating	smallint	smallint	
1	G	3	7
2	R	3	7
3	PG-13	3	7
4	NC-17	3	7
5	PG	3	7

Avg_rate.csv

rental duration.csv

4. Database Migration – Collect data on user behavior of Android Apps
- **Extract:** The first step involves collecting the data from the external source
 - **Transform:** During this step, the extracted data is converted into another format. This could mean calculating number of title views, duration, ratings, film_id.
 - **Load:** The transformed data is inserted or loaded into the new database, usually by a data engineer.