










3.4: Database Querying in SQL










1. Refining Your Query

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

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

➔ This query is much cleaner and only pulls the required information for mailing addresses

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

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

➔ This query may pull in other data also stored in the customer table, such as phone numbers, activity dates, and notes from sales and customer service.

➔

2. Ordering the Data

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

Please find the extracted sheet in the other attachment (3_4_2)

3. Grouping Data

```
1 SELECT rating, AVG (rental_rate), MIN (rental_duration) ASC, MAX (rental_duration) DESC
2 FROM film
3 GROUP BY rating
```

Data Output Messages Notifications				
	rating mpaa_rating	avg numeric	asc smallint	desc smallint
1	R	2.9387179487179487	3	7
2	PG	3.0518556701030928	3	7
3	NC-17	2.970952380952381	3	7
4	PG-13	3.034843049327354	3	7
5	G	2.888876404494382	3	7

Please find the extracted sheet in the attachment (3_4_3)

4. Database Migration

- Can you outline the procedure for migrating the data and who will be responsible for it?

1. Extract: Gathering data from multiple different sources
2. Transform: Extracted data converted into another format.
3. Load: Transformed data is inserted 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 a data analyst analyzes the data before it has been loaded into the data warehouse, the data will not be consistently formatted. Thereby it will be very difficult to retrieve and manipulate the data or to get insights and meaningful conclusions from it.