

Query Query History

```

1  EXPLAIN
2  SELECT * FROM film
3

```

Data Output Messages Notifications

QUERY PLAN
text

1 Seq Scan on film (cost=0.00..98.00 rows=1000 width=384)

Query Query History

```

1  EXPLAIN
2  SELECT film_id,title
3  FROM film
4

```

Data Output Messages Notifications

QUERY PLAN
text

1 Seq Scan on film (cost=0.00..98.00 rows=1000 width=19)

We understand that the cost of query is the time taken by the query to hit the database and return the result; the cost estimate for both queries is (cost=0.00..98.00 rows=1000) means that Postgres expects that it will “cost” 98.00 of an arbitrary unit of computation to find these values. It can be assumed that they might have the same number of rows.

The command can be refined by putting a LIMIT for example from the top 10 films.

Query Query History

```

1  EXPLAIN
2  SELECT film_id,
3  title
4  FROM film
5  ORDER BY title ASC
6  LIMIT 10
7

```

Data Output Messages Notifications

QUERY PLAN
text

1 Limit (cost=0.28..1.63 rows=10 width=19)

2 -> Index Scan using idx_title on film (cost=0.28..135.33 rows=1000 width=...

2.

Query

Query History

Scratch Pad

×

1

SELECT *

2

FROM film

3

ORDER BY title ASC,

4

release_year DESC,

5

rental_rate DESC

6

Data Output

Messages

Notifications

≡

📄

▼

📋

▼

🗑️

🗄️

⬇️

📈

| | film_id [PK] integer | title character varying (255) | description text |
|---|-------------------------|----------------------------------|---|
| 1 | 1 | Academy Dinosaur | A Epic Drama of a Feminist And a Mad Scientist who must Battle a Teacher in The Canadian Rockies |
| 2 | 2 | Ace Goldfinger | A Astounding Epistle of a Database Administrator And a Explorer who must Find a Car in Ancient China |
| 3 | 3 | Adaptation Holes | A Astounding Reflection of a Lumberjack And a Car who must Sink a Lumberjack in A Baloon Factory |
| 4 | 4 | Affair Prejudice | A Fanciful Documentary of a Frisbee And a Lumberjack who must Chase a Monkey in A Shark Tank |
| 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 |
| 6 | 6 | Agent Truman | A Intrepid Panorama of a Robot And a Boy who must Escape a Sumo Wrestler in Ancient China |
| 7 | 7 | Airplane Sierra | A Touching Saga of a Hunter And a Butler who must Discover a Butler in A Jet Boat |
| 8 | 8 | Airport Pollock | A Epic Tale of a Moose And a Girl who must Confront a Monkey in Ancient India |

Total rows: 1000 of 1000

Query complete 00:00:00.225

Ln 5, Col 17

3.

Query

Query History

1

SELECT rating,

2

AVG (rental_rate)

3

FROM film

4

GROUP BY rating

Data Output

Messages

Notifications

≡

📄

▼

📋

▼

🗑️

🗄️

⬇️

📈

| | 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 |

Query

Query History

1

SELECT rating,

2

MIN(rental_duration),

3

MAX(rental_duration)

4

FROM film

5

GROUP BY rating

Data Output

Messages

Notifications

≡

📄

▼

📋

▼

🗑️

🗄️

⬇️

📈

| | 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 |
| 5 | G | 3 | 7 |

4. 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 data engineer(s) will be responsible of completing the ETL process.

- **Extract:** The first step involves collecting the data from multiple data sources. In team with the data engineer(s), put processes in place to ensure data quality.
- **Transform:** During this step, the extracted data is converted into another format. For instance if with the App it is possible to get an overview of where the users click and how far they scroll, or how often users perform actions that matter to the company all that collected data will need to be transformed in order to be paired with the data we currently have.
- **Load:** At this point the transformed data is inserted or loaded into the desired database.

*It may also be necessary to work with the application vendor to take any additional steps that are specific to that software to ensure things go properly and the application functions properly post-migration.

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

The main reason for making a data migration process is turning data which was previously not available into a new source of revenue; it would seem that, since the data comes from different locations, formats and structures, if this data is not ready to be uploaded, a proper analysis cannot be done, the data is not ready for decision-making.

Also very important, the data must be checked to establish there are no errors before any analysis can be done.