

3.3: SQL for Data Analysts

Step 1:

QUERY:

```
SELECT category_id, name  
FROM category;
```

OUTPUT

category_id	name
1	Action
2	Animation
3	Children
4	Classics
5	Comedy
6	Documentary
7	Drama
8	Family
9	Foreign
10	Games
11	Horror
12	Music
13	New
14	Sci-Fi
15	Sports
16	Travel

Step 2:

Dashboard X Properties X SQL X Statistics X Dependencies X Dependents X Processes X **ROCBUSTER/postgres@PostgreSQL 18*** X

ROCBUSTER/postgres@PostgreSQL 18

No limit ▾ E M I S Q L

Query Query History

```

1 INSERT INTO category(name)VALUES
2 ('Thriller'),
3 ('Crime'),
4 ('Mystery'),
5 ('Romance'),
6 ('War')
7
8
9
10

```

Data Output **Messages** Notifications

INSERT 0 5

Query returned successfully in 49 msec.

Total rows: Query complete 00:00:00.049

Search

ses(3)

KBUSTER

asts

atalogs

rent Triggers

tensions

oreign Data Wrappers

nguages

ublications

chemas(1)

public

Aggregates

Collations

Domains

FTS Configurations

FTS Dictionaries

FTS Parsers

FTS Templates

Foreign Tables

Functions

Materialized Views

Operators

Procedures

Sequences

Tables(15)

actor

address

category

Data Output Messages Notifications

Showing rows: 1 to 21

	name
1	Thriller
2	Sci-Fi
3	Comedy
4	War
5	Family
6	Games
7	Crime
8	Animation
9	Romance
10	Documentary
11	Classics
12	Sports
13	New
14	Mystery
15	Children
16	Music
17	Travel
18	Foreign
19	Horror
20	Drama
21	Action

Write a short paragraph explaining the various constraints that have been applied to the columns. What do these constraints do exactly?

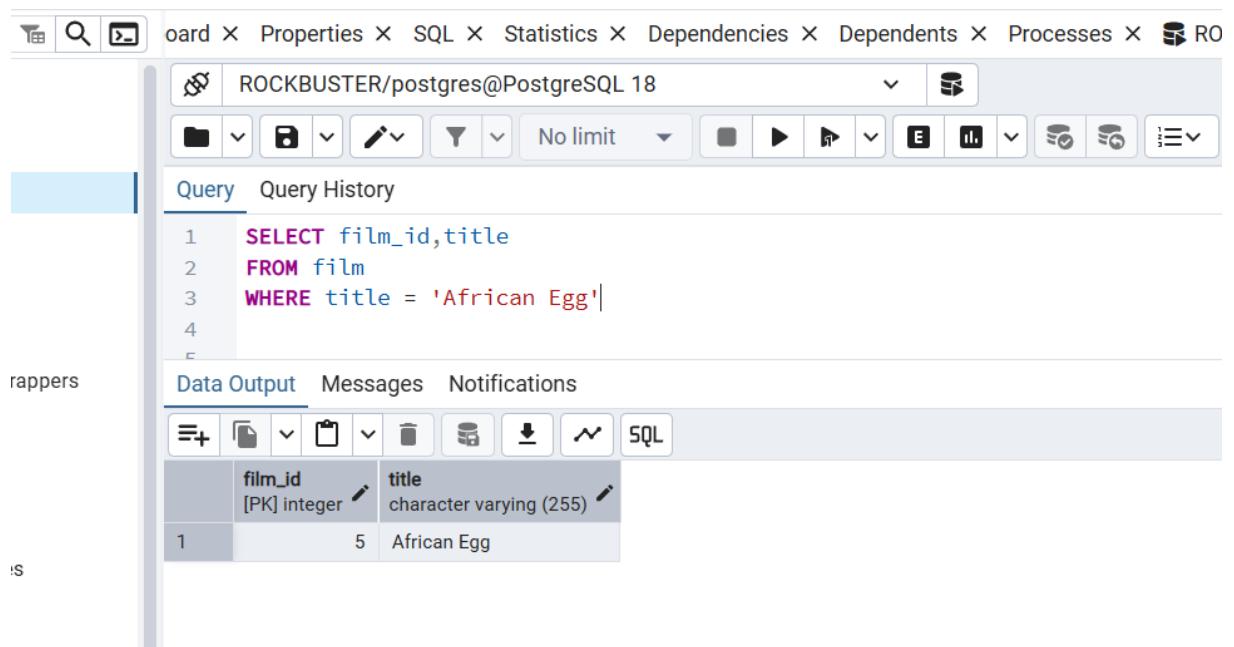
The primary key: the primary key column can't contain any null or duplicate values.

NOT NULL Constraint: This ensures that the column of category_id can't have any empty or missing values

Why are they important? Because it keeps the data organized and in making it easier and faster to query the database.

Step 3:

Write the SELECT statement to find the film_id for the movie *African Egg*.



The screenshot shows the pgAdmin 4 interface. The top bar has tabs for Board, Properties, SQL, Statistics, Dependencies, Dependents, Processes, and a connection dropdown for 'RO'. Below the toolbar are buttons for file operations, search, and connection status. The main area is divided into two sections: 'Query' and 'Query History'. The 'Query' section contains the following SQL code:

```
1  SELECT film_id,title
2  FROM film
3  WHERE title = 'African Egg'
4
```

The 'Data Output' section shows the results of the query:

	film_id	title
1	5	African Egg

Once you have the film_ID and category_ID, write an UPDATE command to change the category in the film_category table

v Window Help

18
:(3)
JSTER
S
logs
t Triggers
visions
gn Data Wrappers
uages
ications
mas(1)
iblic
Aggregates
Collations
Domains
FTS Configurations
FTS Dictionaries
FTS Parsers
FTS Templates

encies X Dependents X Processes X ROCKBUSTER/postgres@PostgreSQL 18 X ROCKBUSTER/postgres@PostgreSQL 18 X ROCKBUSTER/postgres@PostgreSQL 18

Query History

```
1 UPDATE film_category
2 SET category_id = 17
3 WHERE film_id = 5;
```

Data Output Messages Notifications

UPDATE 1

Query returned successfully in 43 msec.

Step 4:

Query History

```
1 SELECT category_id, name
2 FROM category
3 WHERE category_id = 19;
```

Data Output Messages Notifications

category_id [PK] integer name character varying (25)

	category_id	name
1	19	Mystery

The screenshot shows a MySQL Workbench interface. At the top, there's a toolbar with various icons for file operations like Open, Save, and Print. Below the toolbar, the title bar says "Query History". The main area contains a query log:

```
1  DELETE FROM category
2  Where category_id=19 |
3
```

Below the query log, there are tabs for "Data Output", "Messages", and "Notifications". The "Messages" tab is selected, showing the output of the query:

```
DELETE 1

Query returned successfully in 59 msec.
```

Step 5:

Pros of SQL: SQL efficiently handles millions of rows, allows for complex queries and joins, and makes multi-user collaboration possible.

Cons of SQL: Its syntax is more difficult to learn, it does not provide an intuitive graphical interface, and there is no automatic “Undo” function.

Pros of Excel: Excel offers an intuitive interface, quick visualization through charts, pivot tables, and filters, and flexibility since you can mix text, numbers, and formulas in the same worksheet.

Cons of Excel: It has a volume limit that makes it difficult to manage beyond a few hundred thousand rows, carries a risk of errors where a bad formula or a copy-paste mistake can corrupt the entire file, is not designed for relational databases, and becomes slow with complex calculations or large files.