3.1 - Intro to Relational Database

Directions:

Step 1

If you haven't done so already, install PostgreSQL and load the Rockbuster database using the instructions in the Exercise. Then download your Achievement 3 project brief (PDF) to get an idea of what each Exercise will cover.

Step 2

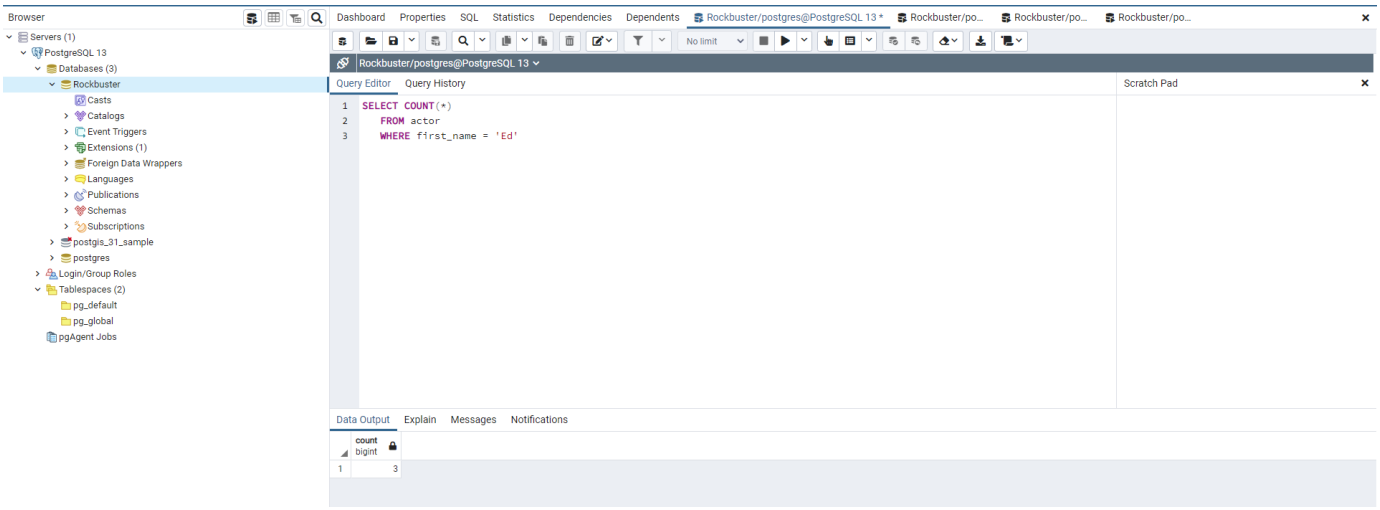
Compare and contrast spreadsheets and databases by following the steps below:

- Download the Rockbuster "actor.csv" file and open it in Excel.
- Drawing on what you've learned in previous Achievements, use the appropriate functions in Excel to count all the actors whose first name is "Ed." Write down the result in a text document.
    - 3
- Launch pgAdmin 4, open the Query Tool, copy-paste the SQL statement below into the Query Editor, and execute it. If you're unsure how to do this, reread the "Bonus Content: Walkthrough of PostgreSQL pgAdmin 4" section of the Exercise. The **Query Tool** and **Execute/Refresh** buttons may look different in your pgAdmin 4 console (i.e., not a lightning bolt) depending on which version you're using.

```
SELECT COUNT(*)
FROM actor
WHERE first_name = 'Ed'
```

TIP
This statement will count all the instances of an actor with a first name "Ed" in the table actor.

- Copy the result that tells you the number of times the first name "Ed" appears in the "actor" table from the **Data Output** window into your text document from step 2b. Check that your answer matches your answer from step 2a. Was it easier to use Excel or the SQL statement and database to count the number of "Eds"? Provide an explanation for your answer in the same text document.



- Was it easier to use Excel or the SQL statement and database to count the number of "Eds"?
    - If for no other reason, my familiarity Excel makes it far easier to navigate in that program than in pgAdmin4. Further, I can actually see everything. I think the two programs will likely have to be used side by side in any serious investigations. But I imagine that future questions and commands will highlight the limitation of programs like Excel and need for work with pgAdmin4.
    - **Also, as a side note, I had to redirect the PostgreSQL Binary Path to the 'bin folder' with the pg_restore.exe program in it for the program to work – this took an hour or two. Maybe mention this for future students**
        - How to Link PosgreSQL Binary Path

Step 3

To answer the next set of questions, you'll be pasting the queries provided into the Query Editor in pgAdmin 4. Note down your answers in your running text document.

- Execute the following query and list the names of the columns in the payment table.

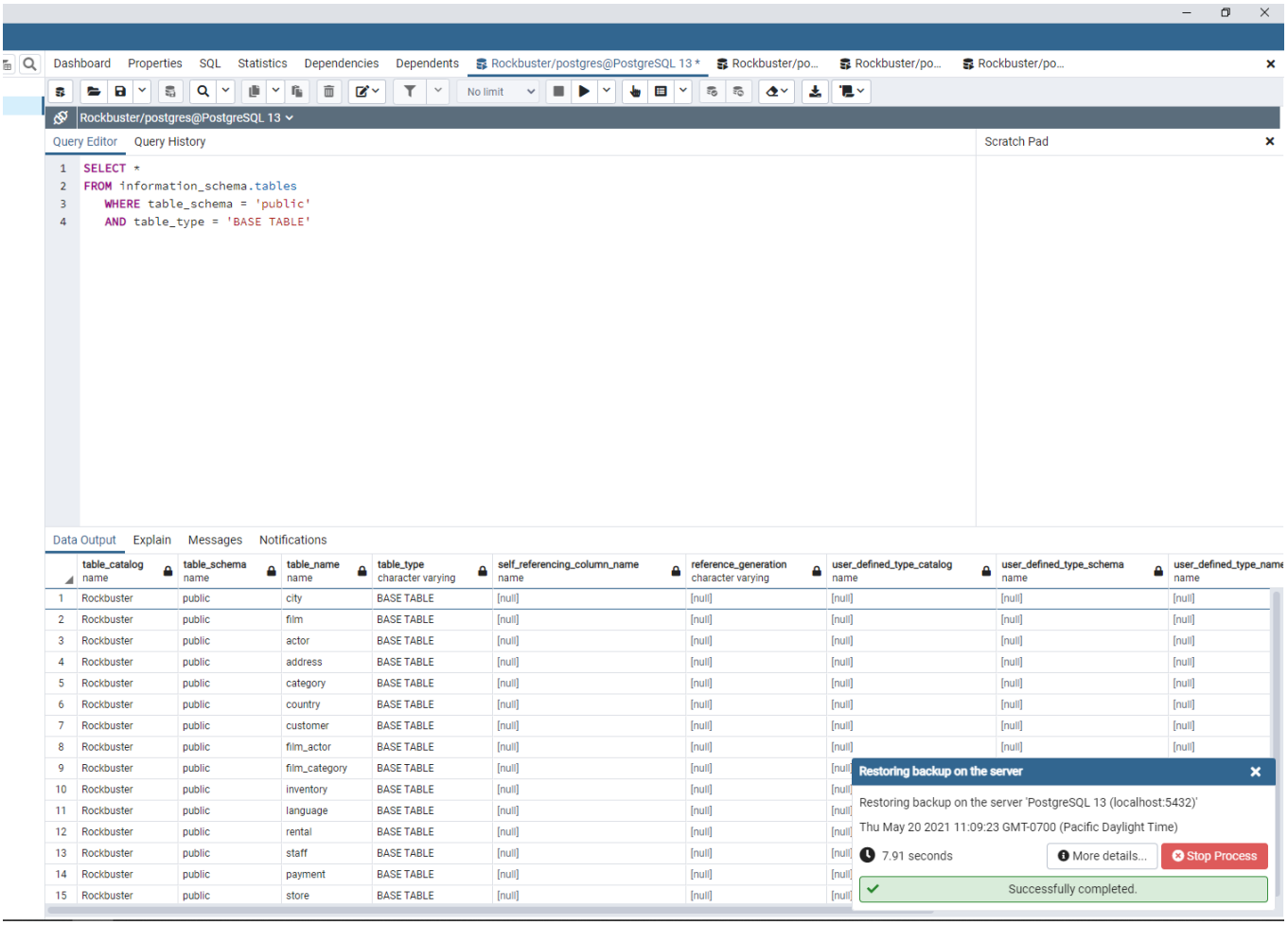
```
SELECT * FROM payment LIMIT 10;
```

- Under the "table_name" column, what are the names of the tables that are available in the Rockbuster database? (List all names.)
- 


```
SELECT * FROM information_schema.tables

WHERE table_schema = 'public'

AND table_type = 'BASE TABLE'
```



- Within the pgAdmin 4 console, can you think of another way to list all the table names in the database instead of the SQL statement above?
- Analyze the rental duration distribution. How many days are most films rented for?
    - 6 days

```
SELECT rental_duration AS "rented for (in days)", COUNT(*) AS "number of films"

FROM film

GROUP BY 1

ORDER BY 2
```

Query Editor | Query History

```
1  SELECT rental_duration
2  AS "rented for (in days)",
3  COUNT(*) AS "number of films"
4    FROM film
5    GROUP BY 1
6    ORDER BY 2
```

Data Output | Explain | Messages | Notifications

	rented for (in days) smallint	number of films bigint
1	7	191
2	5	191
3	4	203
4	3	203
5	6	212

Step 4

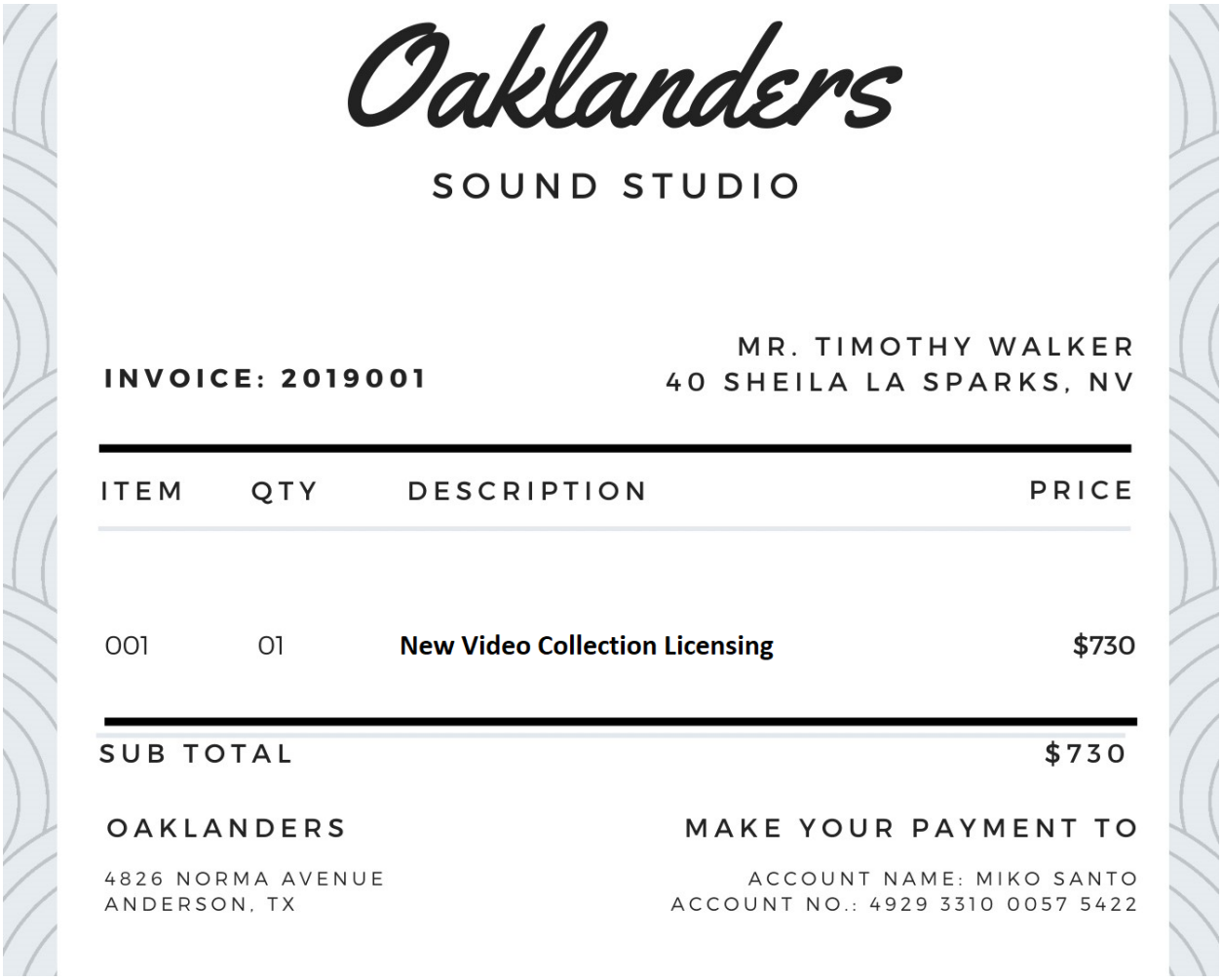
Think about who in Rockbuster Stealth might want to use an OLAP or OLTP system for their data needs; for example, the sales department, which is interested in sales trends, would likely use an OLAP system. Describe at least 2 situations for each type of system.

1 (OLTP): to track the number of payments done in a specific period of time, considering possible seasonality (i.e., during cold months people use to watch more films and look for online entertainment)

2 (OLAP): existing data could be used to analyze upcoming trends and define which “film genre/film category” might be more popular than another in a specific year.

Step 5

Rockbuster Stealth has received an invoice for the licenses for its new video collection.



Oaklanders

SOUND STUDIO

INVOICE: 2019001

MR. TIMOTHY WALKER
40 SHEILA LA SPARKS, NV

ITEM	QTY	DESCRIPTION	PRICE
001	01	**New Video Collection Licensing**	$730

SUB TOTAL $730

OAKLANDERS
4826 NORMA AVENUE
ANDERSON, TX

MAKE YOUR PAYMENT TO
ACCOUNT NAME: MIKO SANTO
ACCOUNT NO.: 4929 3310 0057 5422

Take a moment to familiarize yourself with data in the invoice, then note down the answers to the questions below.

- Does the invoice contain structured or unstructured data? Write an explanation for your answer.
    - The data presented are structured. These are easily searchable data that is quantitative. We have a receipt here detailing the purchase of 1 license for the price ($730).
- Organize and store the information on the invoice in a database. Step one will be to create a table in the text document you’ve started (you can insert a table if you’re using MS Word or Google Docs, for example). Make sure your table contains columns with the appropriate labels, as well as the values from the invoice in each column. You're focusing, here, on a high-level structuring of your data.

Invoice #	Item	Quantity	Description	Billing address	Billing Address provider	Subtotal	Total	Bank account
201901	001	1	New Video collection licensing	Mr. Timothy Licensing 40 Sheila La Sparks, NV	Oaklanders 4826 Nora Avenue Anderon TX	$730	$730	4929 3310 0057 5422

Step 6

Save the text document containing your answers as a PDF and upload it here for your tutor to review. Don’t hesitate to contact your tutor or mentor if you have any questions!