

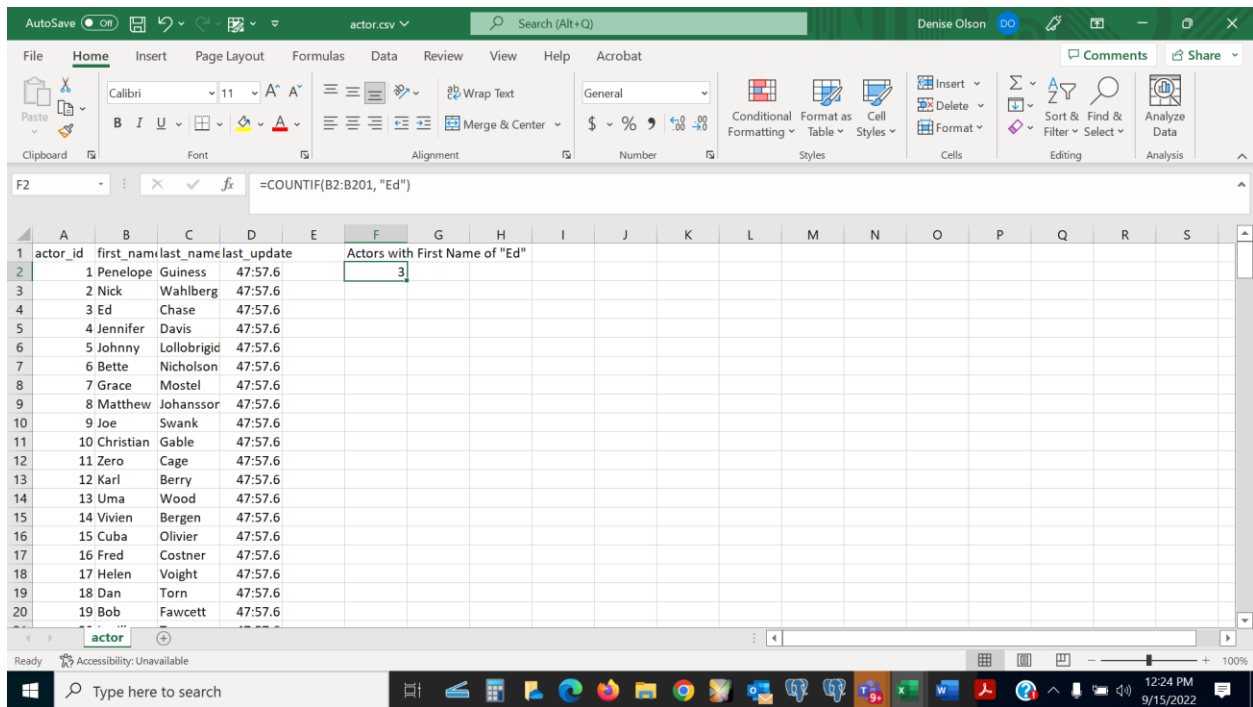
Task 3.1 Introduction to Relational Databases

1. Step 1

- Install PostgreSQL and load Rockbuster database – [completed](#)

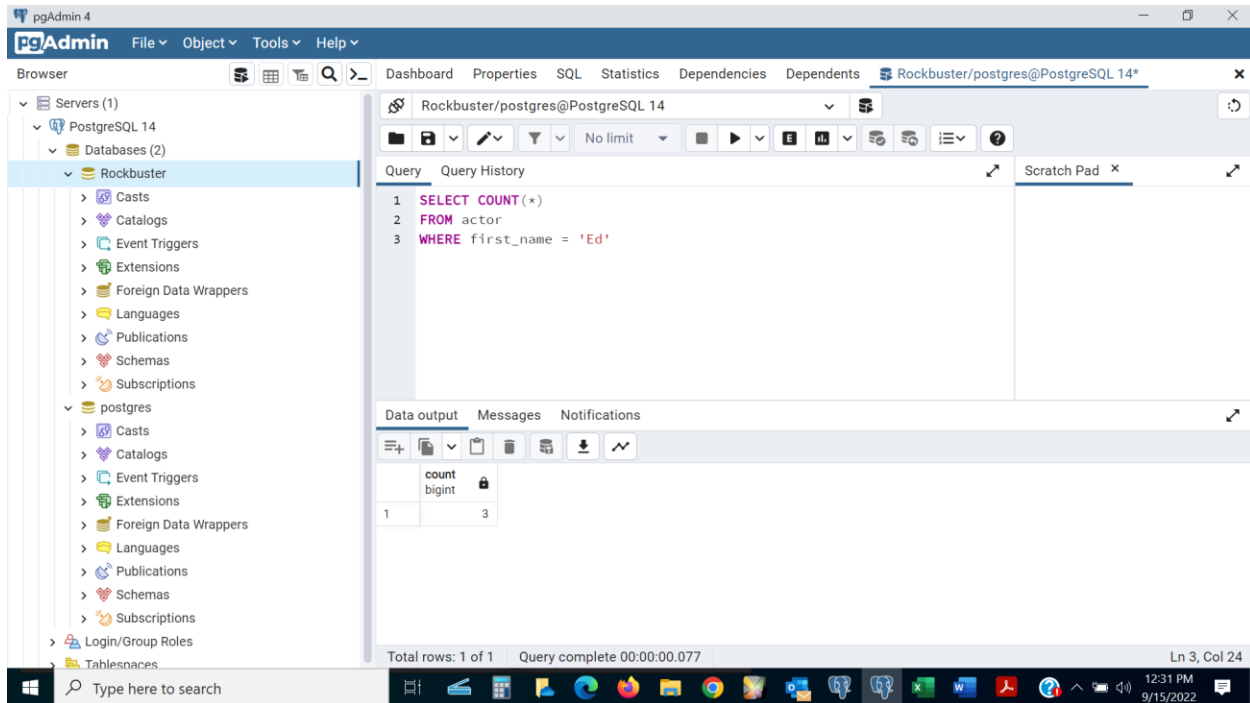
2. Step 2

- Compare and contrast the spreadsheets and databases by following steps below:
 - Download Rockbuster “actor.csv” file, open in Excel
 - Use appropriate functions in Excel to count all the actors whose first name is “Ed”.
Write down the result in a text document. – [See below Excel file](#)



actor_id	first_name	last_name	last_update
1	Penelope	Guinness	47:57.6
2	Nick	Wahlberg	47:57.6
3	Ed	Chase	47:57.6
4	Jennifer	Davis	47:57.6
5	Johnny	Lolobrigid	47:57.6
6	Bette	Nicholson	47:57.6
7	Grace	Mostel	47:57.6
8	Matthew	Johansson	47:57.6
9	Joe	Swank	47:57.6
10	Christian	Gable	47:57.6
11	Zero	Cage	47:57.6
12	Karl	Berry	47:57.6
13	Uma	Wood	47:57.6
14	Vivien	Bergen	47:57.6
15	Cuba	Olivier	47:57.6
16	Fred	Costner	47:57.6
17	Helen	Voight	47:57.6
18	Dan	Torn	47:57.6
19	Bob	Fawcett	47:57.6

- Launch pgAdmin 4, open the query tool, copy-paste the SQL statement below into the query editor and execute it – [See below pgAdmin query results](#)



- 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. *I felt it was easier to use SQL statement to retrieve the count of the number of “Eds” for I am familiar with the SQL language more so than Excel functions.*

3. 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.
 - Columns are: `payment_id`, `customer_id`, `staff_id`, `rental_id`, `amount`, `payment_date`

`SELECT * FROM payment LIMIT 10;`

The screenshot shows the pgAdmin 4 interface. On the left, the 'Servers' tree is expanded to show the 'Rockbuster' database. The 'Query' tab is active, displaying the query: `SELECT * FROM payment LIMIT 10;`. The 'Data output' tab shows the results of the query, which are 10 rows from the 'payment' table. The table has columns: `payment_id` (PK integer), `customer_id` (smallint), `staff_id` (smallint), `rental_id` (integer), `amount` (numeric (5,2)), and `payment_date` (timestamp without time zone).

	payment_id [PK] integer	customer_id smallint	staff_id smallint	rental_id integer	amount numeric (5,2)	payment_date timestamp without time zone
1	17503	341	2	1520	7.99	2007-02-15 22:25:46.996577
2	17504	341	1	1778	1.99	2007-02-16 17:23:14.996577
3	17505	341	1	1849	7.99	2007-02-16 22:41:45.996577
4	17506	341	2	2829	2.99	2007-02-19 19:39:56.996577
5	17507	341	2	3130	7.99	2007-02-20 17:31:48.996577
6	17508	341	1	3382	5.99	2007-02-21 12:33:49.996577
7	17509	342	2	2190	5.99	2007-02-17 23:58:17.996577
8	17510	342	1	2914	5.99	2007-02-20 02:11:44.996577
9	17511	342	1	3081	2.99	2007-02-20 13:57:39.996577
10	17512	343	2	1547	4.99	2007-02-16 00:10:50.996577

Total rows: 10 of 10 Query complete 00:00:01.011 Ln 1, Col 35

- Under the "table_name" column, what are the names of the tables that are available in the Rockbuster database? (List all names.) Available tables are 15 tables: 1.actor, 2.store, 3.address, 4.category, 5.city, 6.county, 7.customer, 8.film_actor, 9. film_category, 10.inventory, 11.language, 12.rental, 13.staff, 14. payment, 15. film.

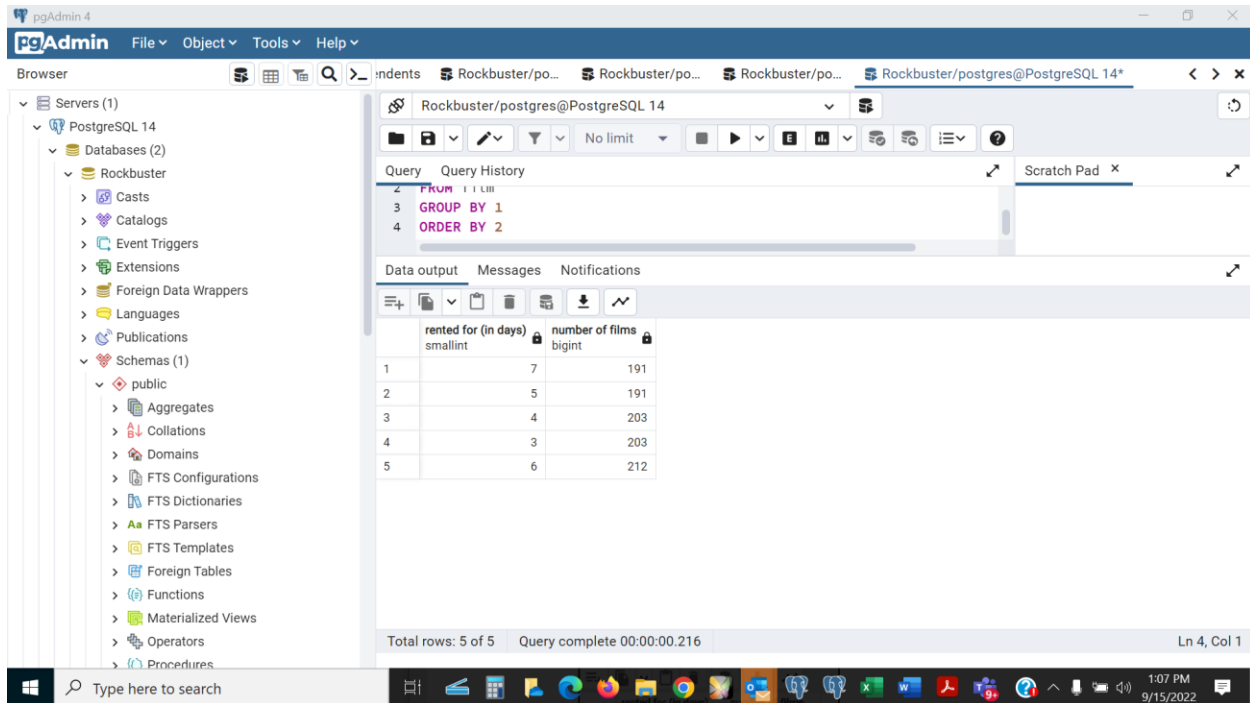
```
SELECT * FROM information_schema.tables
WHERE table_schema = 'public'
AND table_type = 'BASE TABLE'
```

The screenshot shows the pgAdmin 4 interface. On the left, the 'Rockbuster' database is selected under 'PostgreSQL 14'. The main pane displays the query result for the SQL statement: `SELECT * FROM information_schema.tables WHERE table_schema = 'public' AND table_type = 'BASE TABLE'`. The result is a table with 11 rows and 8 columns: `table_catalog`, `table_schema`, `table_name`, `table_type`, `self_referencing_column_name`, `reference_generation`, `user_defined_t`, and `name`. The status bar at the bottom indicates 'Total rows: 15 of 15' and 'Query complete 00:00:00.142'.

	table_catalog	table_schema	table_name	table_type	self_referencing_column_name	reference_generation	user_defined_t	name
1	Rockbuster	public	actor	BASE TABLE	[null]	[null]	[null]	
2	Rockbuster	public	store	BASE TABLE	[null]	[null]	[null]	
3	Rockbuster	public	address	BASE TABLE	[null]	[null]	[null]	
4	Rockbuster	public	category	BASE TABLE	[null]	[null]	[null]	
5	Rockbuster	public	city	BASE TABLE	[null]	[null]	[null]	
6	Rockbuster	public	country	BASE TABLE	[null]	[null]	[null]	
7	Rockbuster	public	customer	BASE TABLE	[null]	[null]	[null]	
8	Rockbuster	public	film_actor	BASE TABLE	[null]	[null]	[null]	
9	Rockbuster	public	film_catego...	BASE TABLE	[null]	[null]	[null]	
10	Rockbuster	public	inventory	BASE TABLE	[null]	[null]	[null]	
11	Rockbuster	public	language	BASE TABLE	[null]	[null]	[null]	

- 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? [Yes, you can select Schemas -> Tables\(15\) from the Rockbuster database.](#)
- Analyze the rental duration distribution. How many days are most films rented for? [Most films are rented for 6 days – screen shot below of query results.](#)

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



4. 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.

- **OLAP (online analytical processing system)** – data warehouse, used for reading operations and look at data for trends over time (Analytical Data Mining)
 - Marketing / eCommerce teams – can use OLAP for market research, sales forecasting, customer analysis, promotions.
- **OLTP (online transaction processing system)** – used for inserting, delete, and update large quantities of transactions. (Transactional system).
 - Financial / Inventory teams – used for transactional items, insert, delete and updating records from customers.

5. Step 5

Rockbuster Stealth has received an invoice for the licenses for its new video collection. 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. **Data for the invoice is structured, it is predefined data and can be organized as rows and columns in a table.**

- 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.

TRANSACTION					
Invoice Number	Item	Quantity	Description	Currency	Price
2019001	001	01	New video Collection Licensing	US Dollar	730

MERCHANT						
Name	Address	City	State	State Abbreviation	Account Name	Account Number
Oaklanders Sound Studio	4826 Norma Avenue	Anderson	Texas	TX	Miko Santo	4929331000575422

CUSTOMER							
Title	First Name	Last Name	Gender	Address	City	State	State Abbreviation
Mr.	Timothy	Walker	Male	40 Shelia La	Sparks	Nevada	NV