

Task 3.1 Intro to Relational Databases

By Lee Heng Chuah

Imagine you've been hired as a data analyst at Rockbuster Stealth, a fictional company that provides online video rentals. Your job is look at the data to answer various business questions. As you work through this Achievement and learn more about SQL, the questions you answer and data queries you perform will gradually become more complex.

On your first day of work, your supervisor sends you the data dump for Rockbuster Stealth and asks you to work on the following tasks.

Directions:

Step 1

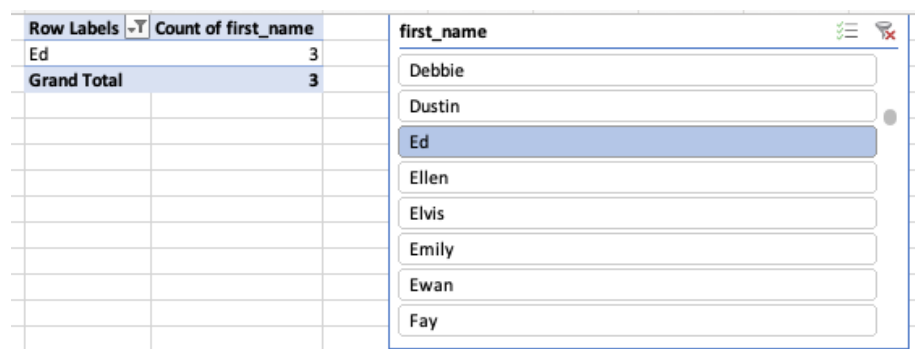
If you haven't done so already, install PostgreSQL and load the Rockbuster database using the instructions in the Exercise. Then download your Achievement 3 project brief (PDF) to get an idea of what each Exercise will cover.

- Completed

Step 2

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

- Download the Rockbuster "actor.csv" file and open it in Excel. - Completed
- Drawing on what you have learned in previous Achievements, use the appropriate functions in Excel to count all the actors whose first name is "Ed". Write down the results in a text document.
 - Using function - `=COUNTIF(B2:B201,"Ed")`, the result showed 3.
 - Using Pivot Table, then drag the "first_name" into Row, Slicer the "first_name" with "Ed", the result showed 3.



Row Labels	Count of first_name
Ed	3
Grand Total	3

first_name
Debbie
Dustin
Ed
Ellen
Elvis
Emily
Ewan
Fay

- Launch pgAdmin 4, open the Query Tool, copy-paste the SQL statement below into the Query Editor, and execute it. If you are unsure how to do this, re-read 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 your are using.

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

The screenshot shows the pgAdmin 4 interface with the following components:

- Top Bar:** Displays the connection name "actor/postgres@PostgreSQL 14".
- Query Editor:** Contains the following SQL query:


```
1 SELECT COUNT(*)
2 FROM PUBLIC."actor"
3 WHERE first_name = 'Ed'
```
- Data Output Window:** Shows the result of the query in a table:

	count bigint
1	3
- Status Bar:** Indicates "Total rows: 1 of 1", "Query complete 00:00:00.044", and "Ln 3, Col 24".

- Using the PgAdmin 4, with the code copied into Query, the result showed 3 as well.

- 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. – Both produced the same result of '3'. I think both Excel and query in SQL equally easy to use. SQL would require creating a table, and then import the csv file into pgAdmin 4, which took longer time. However, it was easy to find the COUNT of 'Ed' once the query was pasted. While Excel can open the csv file, and find the COUNT of 'Ed' by using the function of COUNTIF().

Step 3

To answer the next set of questions, you will 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;

Rockbuster/postgres@PostgreSQL 14

Query Query History Scratch Pad

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

Data output Messages Notifications

	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

- 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'

Dashboard Properties SQL Statistics Dependencies Dependents Rockbuster.tar*

Rockbuster/postgres@PostgreSQL 14

Query Query History Scratch Pad

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

Data output Messages Notifications

	table_catalog name	table_schema name	table_name name	table_type character varying	self_referencing_column_name name	reference_generation character varying	user_defined_type_catalog name	user_defined_type_schema name	user_defined_type_name name
1	Rockbuster	public	actor	BASE TABLE	[null]	[null]	[null]	[null]	[null]
2	Rockbuster	public	store	BASE TABLE	[null]	[null]	[null]	[null]	[null]
3	Rockbuster	public	address	BASE TABLE	[null]	[null]	[null]	[null]	[null]
4	Rockbuster	public	category	BASE TABLE	[null]	[null]	[null]	[null]	[null]
5	Rockbuster	public	city	BASE TABLE	[null]	[null]	[null]	[null]	[null]
6	Rockbuster	public	country	BASE TABLE	[null]	[null]	[null]	[null]	[null]
7	Rockbuster	public	customer	BASE TABLE	[null]	[null]	[null]	[null]	[null]
8	Rockbuster	public	film_actor	BASE TABLE	[null]	[null]	[null]	[null]	[null]
9	Rockbuster	public	film_catego...	BASE TABLE	[null]	[null]	[null]	[null]	[null]
10	Rockbuster	public	inventory	BASE TABLE	[null]	[null]	[null]	[null]	[null]
11	Rockbuster	public	language	BASE TABLE	[null]	[null]	[null]	[null]	[null]
12	Rockbuster	public	rental	BASE TABLE	[null]	[null]	[null]	[null]	[null]
13	Rockbuster	public	staff	BASE TABLE	[null]	[null]	[null]	[null]	[null]
14	Rockbuster	public	payment	BASE TABLE	[null]	[null]	[null]	[null]	[null]
15	Rockbuster	public	film	BASE TABLE	[null]	[null]	[null]	[null]	[null]

Total rows: 15 of 15 Query complete 00:00:00.055 Ln 3, Col 30

table_name: actor, store, address, category, city, country, customer, film_actor, film_category, inventory, language, rental, staff, payment, film

- 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? – [Left hand panel, go to Database → Rockbuster → Schemas → Tables \(15\)](#)
- Analyse the rental duration distribution. How many days are most films rented for? – [Most film 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
```

The screenshot shows the pgAdmin 4 interface. The top navigation bar includes Dashboard, Properties, SQL, Statistics, Dependencies, and Dependents. The current tab is 'Rockbuster.tar*'. The left sidebar shows the database structure: Rockbuster/postgres@PostgreSQL 14. The main query editor shows the following SQL query:

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

The 'Data output' tab is active, displaying the results of the query in a table format:

	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.

- OLAP System – Analyst in the stores might use the customer data to fetch an order from the previous year, location of customers and manipulate that data to look for trends over time and location of customer to rent
- OLTP System – Inventory in the stores might constantly insert, delete and update the record the rents in the system.

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 familiarise 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.
 - Structured data. All information has been labelled with data on it.
- Organise and store the information on the invoice in a database. Step one will be to create a table in the text document you have started (you can insert a table if you are 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 are focusing, here, on a high-level structuring of your data.

customer_database

Last_name	First_name	Address	City	State
Walker	Timothy	40 Sheila La	Sparks	NV

transaction_database

Invoice_id	item	Quantity	Description	Price
2019001	001	01	New Video Collection Licencing	\$730

supplier_database

Supplier_name	Account_name	Account.no	Address	City	State
Oaklanders	MIKO SANTO	4929331000575422	4826 Norma Avenue	Anderson	TX

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. - [completed](#)