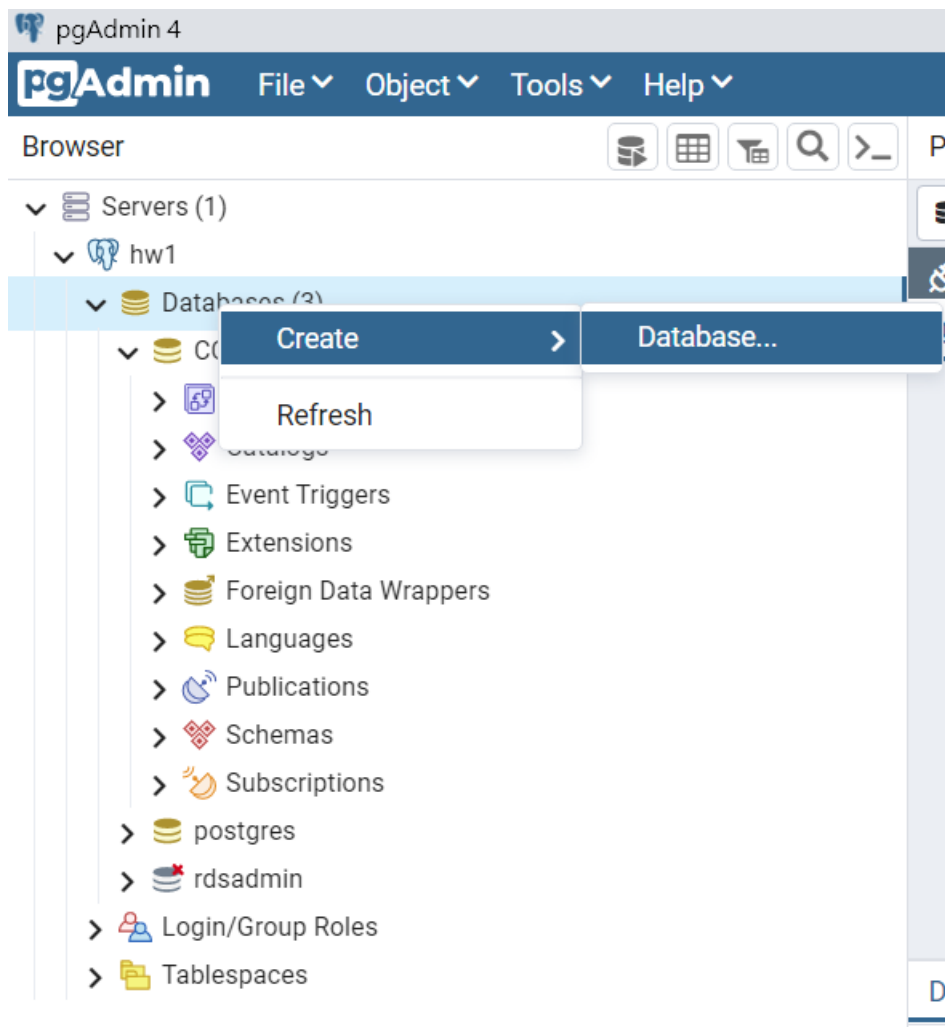


# HW1 Report

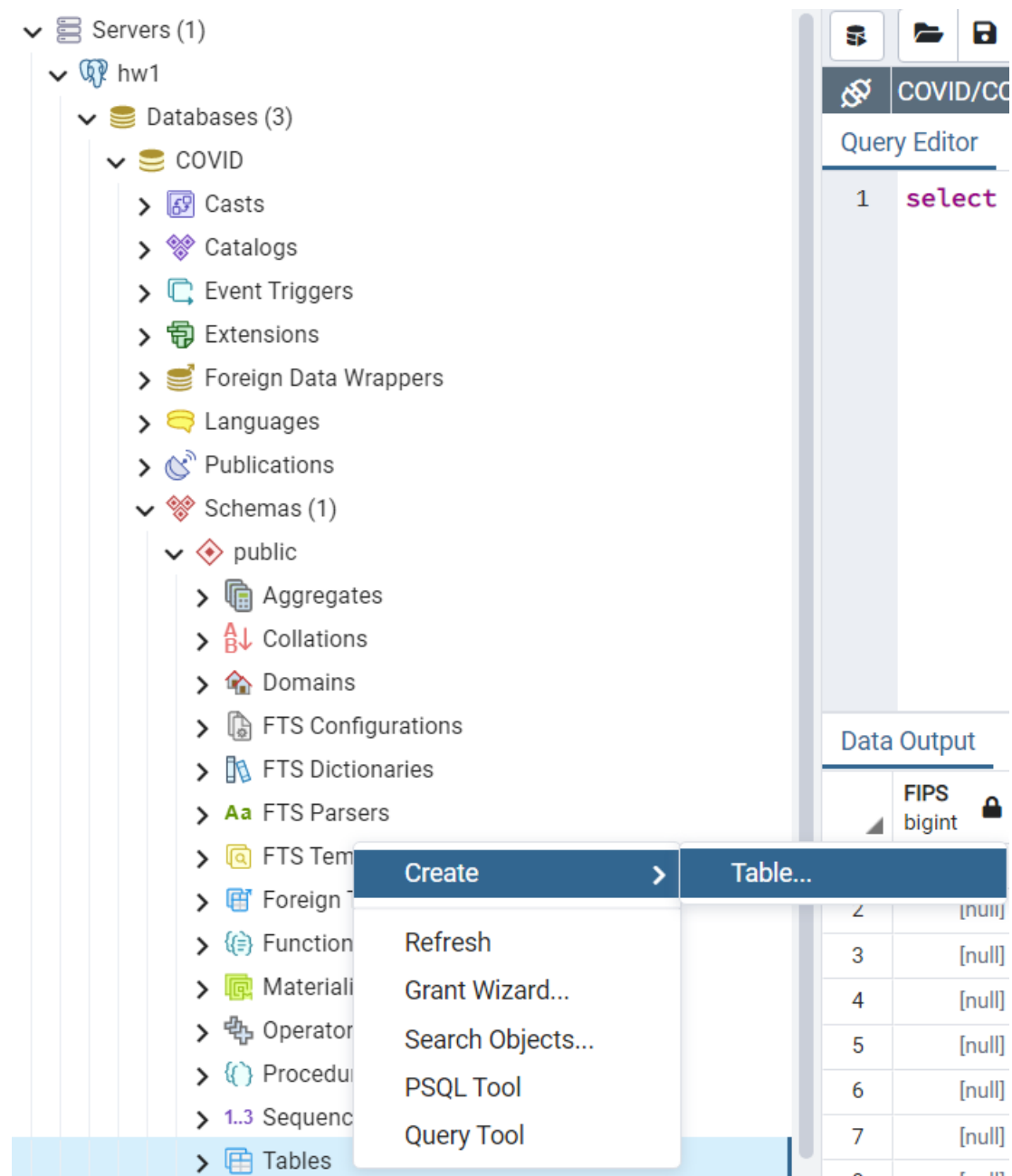
照著老師給的步驟設定好並開啟 Amazon 的雲端服務後，用 pgAdmin4 和雲端服務進行連線存取。連線完成後，執行以下步驟。

1. create database: 對 Servers 底下的 Database 欄按右鍵，選擇

Create Database，並填寫 database 名稱。此作業為 COVID。



2. create table: 點開 COVID 並選到 Schemas，對底下的 Tables 欄按  
右鍵選擇 Create Table。



3. setting attributes: 將 csv 檔的 attributes 填到 Columns 欄位中。

Name	Data type	Length/Precision	Scale	Not NULL?	Primary key?	Default
FIPS	bigint			<input type="checkbox"/>	<input type="checkbox"/>	
Admin2	character varying	100		<input type="checkbox"/>	<input type="checkbox"/>	
Province_State	character varying	100		<input type="checkbox"/>	<input type="checkbox"/>	
Country_Region	character varying	100		<input type="checkbox"/>	<input type="checkbox"/>	
Last_Update	character varying	50		<input type="checkbox"/>	<input type="checkbox"/>	
Lat	double precision			<input type="checkbox"/>	<input type="checkbox"/>	
Long_	double precision			<input type="checkbox"/>	<input type="checkbox"/>	
Confirmed	bigint			<input type="checkbox"/>	<input type="checkbox"/>	
Deaths	bigint			<input type="checkbox"/>	<input type="checkbox"/>	
Recovered	bigint			<input type="checkbox"/>	<input type="checkbox"/>	
Active	bigint			<input type="checkbox"/>	<input type="checkbox"/>	
Combined_Key	character varying	100		<input type="checkbox"/>	<input type="checkbox"/>	
Incident_Rate	double precision			<input type="checkbox"/>	<input type="checkbox"/>	
Case_Fatality_Ratio	double precision			<input type="checkbox"/>	<input type="checkbox"/>	

總共有 14 個 attributes，其中整數數字用 bigint、浮點數用 double precision、剩下的就都用 character varying。至於長度指定 100 是因為 Combined\_Key 這個 attribute 有少數的 row 擁有較長的字串。另外，在這個作業中，因為資料是從外部匯入的，不會做修改，因此 Not Null 和 Primary key 都不做設定也不會影響 query 的結果。將 3 月 22 和 3 月 23 兩個 table 都進行一樣的設定。

4. query process: 按下下圖中的 Query Tool 按鍵，可以開啟 Query Editor 介面，就可以在上面寫 SQL。



Q1. extract the total case number (Confirmed) in Kansas, US on 2022-03-24 (data update date)

Query Editor   Query History

```
1 select sum("Confirmed") from "323"
2 where "Province_State"='Kansas' and "Country_Region"='US'
3 and "Last_Update" like '2022-03-24%'
```

Data Output   Explain   Messages   Notifications

	sum numeric
1	770314

第一題我們用 `sum("Confirmed")` 並選擇 `Province_State = 'Kansas'` 和 `Country_Region = 'US'`，還有 `Last_Update` 為 2022-03-24 開頭的即可。另外要注意的是 table 和 attributes 最好都加上雙引號比較安全。還有匯入 csv 時斜線(/)會被自動轉為減號(-)。這題得到的 result 是 770314。

Q2. extract the country names with more than 1,000,000 total COVID cases on 2022-03-24 (data update date)

Query Editor

Query History

```
1 select "Country_Region" from(  
2     select "Country_Region", sum("Confirmed") as s from "323"  
3     where "Last_Update" like '2022-03-24%'  
4     group by "Country_Region"  
5 ) as tmp  
6 where tmp.s > 1000000
```

Data Output

Explain

Messages

Notifications

	Country_Region character varying (100)
1	Indonesia
2	Bangladesh
3	Sweden
4	Jordan
5	Ireland
6	Singapore
7	Portugal
8	Colombia
9	Cuba
--	-

這題使用 group by 將相同的 country 整合起來，並用 sum 將他們的 Confirm 數分別加總，並在外層的 query 選擇總和超過 1000000 的，印出他們的 country name。這題的 result 較長，我將它放在底下附錄。

Q3. extract the country names and newly diagnosed case number of countries with a newly diagnosed case number > 100,000 on 2022-03-24 (compare with 2022-03-23)

Query Editor		Query History
<pre>1 select "C1" as "Country_Region", tmp2."S2" - tmp1."S1" as "newly diagnosed" from ( 2     select "Country_Region" as "C1", sum("Confirmed") as "S1" from "322" 3     where "Last_Update" like '2022-03-23%' 4     group by "Country_Region" 5 ) as tmp1, ( 6     select "Country_Region" as "C2", sum("Confirmed") as "S2" from "323" 7     where "Last_Update" like '2022-03-24%' 8     group by "Country_Region" 9 ) as tmp2 10 where tmp1."C1" = tmp2."C2" and tmp2."S2" - tmp1."S1" &gt; 100000</pre>		
Data Output		Explain Messages Notifications
	Country_Region character varying (100)	newly diagnosed numeric
1	France	154118
2	Korea, South	395589
3	US	1450756
4	United Kingdom	103078
5	Germany	158385
6	Vietnam	140837

這題和用上題一樣的方式選擇 3 月 23 和 24 的兩個 subquery，接著選出 Country name 一樣的 row 並將兩個 sum 做相減。若相減結果超過 100000，則就是我們要選的 country。得到的 result 如上圖。

## Appendix: Q2 result

Country_Region
Indonesia
Bangladesh
Sweden
Jordan
Ireland
Singapore
Portugal
Colombia
Cuba
France
Slovakia
Israel
Malaysia
Philippines
Turkey
China
Netherlands
Australia
Serbia
Spain
Georgia
Belgium
Thailand
Korea, South
Italy
US
United Kingdom
Germany
Canada
Argentina
Greece
India
Iran

Chile
Vietnam
South Africa
Peru
Kazakhstan
Japan
Denmark
Iraq
Switzerland
Hungary
Russia
Norway
Pakistan
Romania
Brazil
Austria
Lithuania
Bulgaria
Croatia
Tunisia
Morocco
Mexico
Poland
Lebanon
Czechia