Caleb Rogers

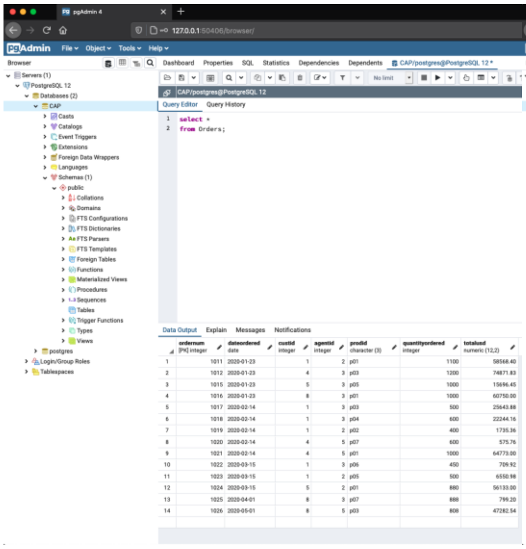
Professor Alan Labouseur

September 8th, 2020

Database Systems

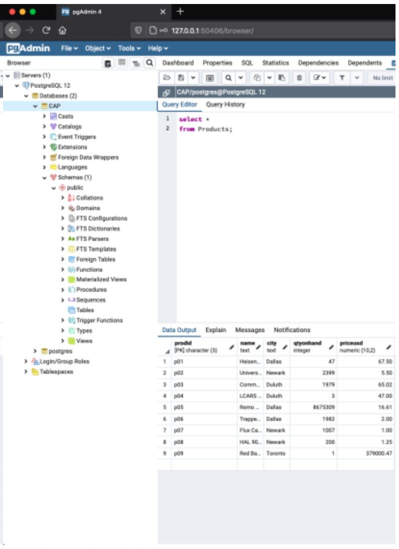
Lab 2

1) Queries' Screenshots:



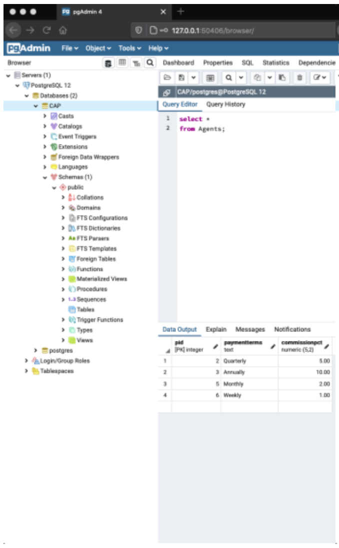
```
select *
from Orders;
```

	ordernum [PK] integer	dateordered date	custid integer	agentid integer	prodid character (3)	quantityordered integer	totalusd numeric (12,2)
1	1011	2020-01-23	1	2	p01	1100	58568.40
2	1012	2020-01-23	4	3	p03	1200	74871.83
3	1015	2020-01-23	5	3	p05	1000	15696.45
4	1016	2020-01-23	8	3	p01	1000	60750.00
5	1017	2020-02-14	1	3	p03	500	25643.88
6	1018	2020-02-14	1	3	p04	600	22244.16
7	1019	2020-02-14	1	2	p02	400	1735.36
8	1020	2020-02-14	4	5	p07	600	575.76
9	1021	2020-02-14	4	5	p01	1000	64773.00
10	1022	2020-03-15	1	3	p06	450	709.92
11	1023	2020-03-15	1	2	p05	500	6550.98
12	1024	2020-03-15	5	2	p01	880	56133.00
13	1025	2020-04-01	8	3	p07	888	799.20
14	1026	2020-05-01	8	5	p03	808	47282.54



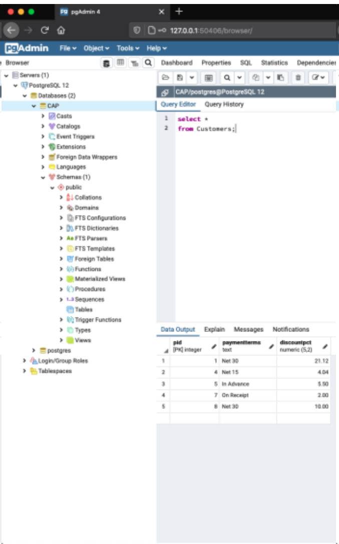
```
select *
from Products;
```

	prodid [PK] character (3)	name text	city text	qtyonhand integer	priceusd numeric (10,2)
1	p01	Heisen...	Dallas	47	67.50
2	p02	Univers...	Newark	2399	5.50
3	p03	Comm...	Duluth	1979	65.02
4	p04	LCARS ...	Duluth	3	47.00
5	p05	Remo ...	Dallas	8675309	16.61
6	p06	Trappe...	Dallas	1982	2.00
7	p07	Flux Ca...	Newark	1007	1.00
8	p08	HAL 90...	Newark	200	1.25
9	p09	Red Ba...	Toronto	1	379000.47



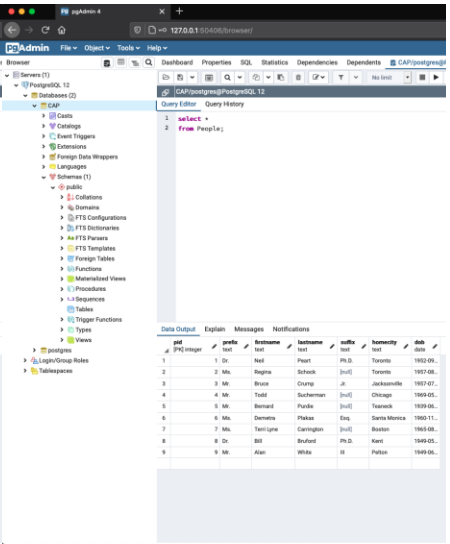
```
select *
from Agents;
```

	pid [PK] integer	paymentterms text	commissionpct numeric (5,2)
1	2	Quarterly	5.00
2	3	Annually	10.00
3	5	Monthly	2.00
4	6	Weekly	1.00



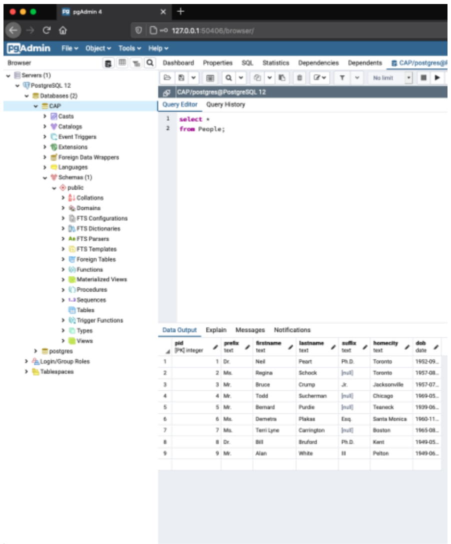
```
select *
from Customers;
```

	pid [PK] integer	paymentterms text	discountpct numeric (5,2)
1	1	Net 30	21.12
2	4	Net 15	4.04
3	5	In Advance	5.50
4	7	On Receipt	2.00
5	8	Net 30	10.00



```
select *
from People;
```

	pid [PK] integer	prefix text	firstname text	lastname text	suffix text	homecity text	dob date
1	1	Dr.	Neil	Peart	Ph.D.	Toronto	1952-09...
2	2	Ms.	Regina	Schock	[null]	Toronto	1957-08...
3	3	Mr.	Bruce	Crump	Jr.	Jacksonville	1957-07...
4	4	Mr.	Todd	Sucherman	[null]	Chicago	1969-05...
5	5	Mr.	Bernard	Purdie	[null]	Teaneck	1939-06...
6	6	Ms.	Demetra	Plakas	Esq.	Santa Monica	1960-11...
7	7	Ms.	Terri Lyne	Carrington	[null]	Boston	1965-08...
8	8	Dr.	Bill	Bruford	Ph.D.	Kent	1949-05...
9	9	Mr.	Alan	White	III	Pelton	1949-06...



```
select *
from People;
```

	pid [PK] integer	prefix text	firstname text	lastname text	suffix text	homecity text	dob date
1	1	Dr.	Neil	Peart	Ph.D.	Toronto	1952-09...
2	2	Ms.	Regina	Schock	[null]	Toronto	1957-08...
3	3	Mr.	Bruce	Crump	Jr.	Jacksonville	1957-07...
4	4	Mr.	Todd	Sucherman	[null]	Chicago	1969-05...
5	5	Mr.	Bernard	Purdie	[null]	Teaneck	1939-06...
6	6	Ms.	Demetra	Plakas	Esq.	Santa Monica	1960-11...
7	7	Ms.	Terri Lyne	Carrington	[null]	Boston	1965-08...
8	8	Dr.	Bill	Bruford	Ph.D.	Kent	1949-05...
9	9	Mr.	Alan	White	III	Pelton	1949-06...

2) *Explain the distinctions among the terms primary key, candidate key, and superkey.*

Primary keys, candidate keys, and superkeys help distinguish different data and helps uniquely identify them, though they are different and fulfill their own roles. Superkeys are the overarching key fields that uniquely identify every row in a table. This is the most basic form compared to its next, candidate keys. Candidate keys are different as they are the minimal super keys. Rather than using long terms like names, candidate keys are simplified, often like a numeric ID. Primary keys differentiate from candidate keys and super keys, as they are the chosen candidate key that uniquely identifies every row in a table. Primary keys are the most unique and identifiable, making them the primary fields used to find a row.

3) *Write a short essay on data types. Select a topic for which you might create a table. Name the table and list its Wields (columns). For each Wield, give its data type and whether or not it is nullable.*

Data types are used to identify attributes in tables. These data types allow programs and databases to be functional, but also serve the purpose of providing context to the data, further providing useful information. Every attribute in a table must require a data type, and there are six main data types. The first are character strings. There are two types of character strings: CHAR (n) and VARCHAR (n), which both allow strings up to n characters. The second data type are BIT (n) strings, which allow strings of bits rather than characters. The third data type is BOOLEAN, which generally states whether an attribute is true or false. The fourth data type is an INT or INTERGER, which are integers that are non-fractional real numbers. The fifth data type is a FLOAT, which are real numbers with a fixed decimal point. Finally, the sixth data type are DATES and TIMES, which are character strings in a specific format. Understanding these data types are very important before creating any tables or databases.

A possible table to be created could be a table of a team's roster of NBA basketball players. Depending on the tables data types, different data would be useful to different approaches of the team. If the data consisted of the financials of the players, then faculty who are involved with the business side of the NBA could easily use the ints, floats, and dates to their advantage when having to deal with salaries and budget management. While a table of different game plans and strategies with certain players are on the court at a time would be a useful way to use character strings, bit strings, and Boolean to help coaches strategize their gamelans. A basic sample table could be used below.

ID	**First Name**	**Last Name**	**Position**	**Accessories**	**Injured**	**Salary (Million)**
1	Kemba	Walker	PG		NO	12
2	Jayson	Tatum	PF	Headband	NO	7.83
3	Jaylen	Brown	SF	Half Sleeve	NO	6.535
4	Enes	Kanter	C	Full Sleeve	YES	17.15
5	Marcus	Smart	SG	Full Sleeve	NO	12.05
6	Daniel	Theis	C	Full Sleeve	NO	5

I would name this table "The Starting Five." There are six players because one of the players are injured. The Injured column has a Boolean data type that determines true or false, whether the player is injured. This Injured column is not nullable, for it is necessary to know whether or not a player is playing in the games. The primary key of this table would be the ID column, which are a set of increasing integers that uniquely identifies each player. This ID column's data type is integers and this column is not nullable, for the player's ID is necessary to properly identify the players. The first and last name columns are character string types that identify the players' names. These columns are nullable, for the players have ID's to identify them. The position is another character string, that helps organize what positions are being covered on the court. The Position column is nullable, for positions are subject to change and can be left empty if necessary. Accessories is a character string that has a nullable data type, for some players do not were accessories, possibly leaving the option blank. Finally, the Salary column has a data type of floats, and is not nullable. This column needs proper inputs, for the NBA has restrictions on how much money is required to pay a player, therefore every player on an active roster like this one, requires a salary.

4) *Explain the following relational "rules" with examples and reasons why they are important.*

a. The "first normal form" rule

b. The "access rows by content only" rule

c. The "all rows must be unique" rule

(a) The first normal rule is the condition that every component of every tuple is an atomic value. This definition essentially means that every element in a table needs to be broken down as much as possible. An example would be if there was a column that had the skills of a person, having multiple skills in the same intersection on a table would violate the "first normal rule." The solution to this error could be to have multiple skills columns so that every skill had its own intersection in the table. This is important to keep data simple and easy to process.

(b) The "access rows by content only" rule disallows markers to indicate rows. For example, referencing a row by searching "the fifth row down" violates this rule because tables in relational databases are due to modify. To solve the issues of this rule, markers to indicate specific elements from the row should be used. For example, an ID used for a person, because no matter which row the searched person is on, the ID remains a reliable source to find the correct row.

(c) The "all rows must be unique" rule requires that every row has an element of data that uniquely identifies that row. For example, a row that only has a person's first name, last name, age, and hometown would violate this rule because other rows could have replicate data like a same first or last name, or same age, or same hometown. The lack of a superkey could make searching for a row difficult if there is replicate data. This is why a reliable superkey like an ID would uniquely identify a row.