



---

# MET CS 689 DESIGNING AND IMPLEMENTING A DATA WAREHOUSE

---

Assignment 1A: Install Database and Python tools



MARCH 17, 2023  
VINCENT NGUYEN  
BOSTON UNIVERSITY

# Overview of the Assignment:

This first assignment installs some of the tools that will be needed for the first weeks of this class.

## Part 1 – Install a database system.

Take a screenshot showing that you have a working database system on your machine by getting the version information.

- Below is screenshot about PostgreSQL database server.

The screenshot displays the pgAdmin 4 web interface. The left sidebar shows a tree view of the database structure, including Servers, Databases, and Schemas. The main pane shows a SQL query window with a query that joins customer and order data, calculates row numbers, and ranks customers by total order amount. Below the query, the 'Data output' tab shows the results of the query as a table with 8 rows and 8 columns.

customer_id	order_id	order_total	order_date	rownumberforeachcustomer	rankcustomer	dencerankcustomer	
numeric (10)	[PK] numeric (10)	numeric (12,2)	date	bigint	bigint	bigint	
1	2	5	1584.00	2003-12-18	4	1	1
2	1	2	1000.00	2003-12-17	2	2	2
3	1	1	506.00	2003-12-18	3	3	3
4	4	6	100.00	2003-12-17	7	4	4
5	1	8	100.00	2003-12-19	1	4	4
6	5	7	40.00	2003-12-18	8	6	5
7	3	3	15.00	2003-12-18	6	7	6
8	3	4	15.00	2003-12-17	5	7	6

Total rows: 8 of 8    Query complete 00:00:00.129    Ln 114, Col 1

## Part 2 – Restore US National Statistics Database

*Alter and run the following SQL commands and take a screenshot showing the results of each:*

*Look to add two additional columns to the code below (both select statements), these two columns will repeat for each record:*

- **MyName**
- **CurrentDateTime**

**SQL Commands to alter and run:**

1. ***SELECT education\_codes.\* FROM education\_codes;***
2. ***SELECT COUNT(\*) as CountOfRecords FROM household\_income***

- Below are screenshots for SQL commands above with MyName, CurrentDataAndTime without time zone:

The screenshot shows the pgAdmin 4 interface. The left sidebar displays the database structure, including Schemas (1) and Tables (6). The main pane shows a SQL query being executed:

```
1 SELECT education_codes.*, 'Vincent Nguyen' as MyName, NOW():timestamp as CurrentDateTime FROM education_codes;
```

The results are displayed in a table with the following columns: code [PK] integer, education\_level\_achieved character varying (50), myname text, and currentdatetime timestamp without time zone. The data shows five records for different education levels, all with the name 'Vincent Nguyen' and the same timestamp.

code [PK] integer	education_level_achieved character varying (50)	myname text	currentdatetime timestamp without time zone
1	no diploma	Vincent Nguyen	2023-03-16 23:32:45.89102
2	high school	Vincent Nguyen	2023-03-16 23:32:45.89102
3	bachelor	Vincent Nguyen	2023-03-16 23:32:45.89102
4	master	Vincent Nguyen	2023-03-16 23:32:45.89102
5	doctorate	Vincent Nguyen	2023-03-16 23:32:45.89102

pgAdmin 4

File Object Tools Help

us\_national\_statistics/postgres@PostgreSQL 14\*

us\_national\_statistics/postgres@PostgreSQL 14

Query Query History

```
1 SELECT COUNT(*) as CountOfRecords, 'Vincent Nguyen' as MyName, NOW()::timestamp as CurrentDateAndTime FROM household_income;
```

Data output Messages Notifications

	countofrecords bigint	myname text	currentdateandtime timestamp without time zone
1	32526	Vincent Nguyen	2023-03-16 23:31:37.445435

- Below are screenshots for SQL commands above with MyName, CurrentDateAndTime with time zone:

The screenshot shows the pgAdmin 4 interface. The left sidebar displays the database structure, including Schemas (1) and public. The main pane shows a SQL query in the Query editor:

```
1 SELECT education_codes.*, 'Vincent Nguyen' as MyName, NOW() as CurrentDateAndTime FROM education_codes;
```

The Query History tab shows the same query. The Data output tab displays the results in a table:

	code [PK] integer	education_level_achieved character varying (50)	myname text	currentdateandtime timestamp with time zone
1	1	no diploma	Vincent Nguyen	2023-03-16 23:24:06.773557-02:30
2	2	high school	Vincent Nguyen	2023-03-16 23:24:06.773557-02:30
3	3	bachelor	Vincent Nguyen	2023-03-16 23:24:06.773557-02:30
4	4	master	Vincent Nguyen	2023-03-16 23:24:06.773557-02:30
5	5	doctorate	Vincent Nguyen	2023-03-16 23:24:06.773557-02:30

The screenshot shows the pgAdmin 4 interface. The left sidebar displays the database structure, including Schemas (1) and public. The main pane shows a SQL query in the Query editor:

```
1 SELECT COUNT(*) as CountOfRecords, 'Vincent Nguyen' as MyName, NOW() as CurrentDateAndTime FROM household_income;
```

The Query History tab shows the same query. The Data output tab displays the results in a table:

	countofrecords bigint	myname text	currentdateandtime timestamp with time zone
1	32526	Vincent Nguyen	2023-03-16 23:25:44.292141-02:30

## Part 3 – Install Anaconda / Jupyter notebook.

Take a screen shot of the opened file.

```
In [ ]: import random

In [ ]: random.randrange(50,500)

In [ ]: str(random.randrange(50,500))

In [ ]: print ("Hello " + str(random.randrange(50,500)))

In [ ]: int((1 + 2) / 3)

In [ ]: int((2 + 2) / 3)

In [ ]: int((3 + 2) / 3)

In [ ]: for i in range(12):
        print (str(i + 1) + " -> " + str(int((i + 3) / 3)) )

In [ ]: from datetime import date
        from dateutil.relativedelta import relativedelta
        days_back = 1
        days_total = 2
        startDt = date.today() + relativedelta(days=+(-1 * days_back))
        for dateOff in range(days_total):
            dimDate = startDt + relativedelta(days=+dateOff)
            dateYear = dimDate.year
            fiscalPer = str (dimDate.year) + 'm' + str(dimDate.month)
            calQuarter = int((dimDate.month + 2) / 3)
            for hr in range(24):
                primaryKey = 'k' + str(dateOff * 24 + hr)
                print (primaryKey, dimDate, dateYear, fiscalPer, calQuarter, hr)
        print ((dateOff + 1) * 24)

In [ ]: run LoadTitanic.py

In [ ]: list(titanic_ppl.columns.values)

In [ ]: titanic_ppl.count()

In [ ]: len(titanic_ppl)

In [ ]: run dimPeople.py

In [ ]: run LoadFromDatabase.py

In [ ]: print (is_it_a_state('NH'))
```

## Grading Criteria

Use the **Ask the Teaching Team Discussion Forum** if you have any questions regarding the how to approach this assignment.

Save your assignment as ***lastnameFirstname\_assign1\_A.docx*** and submit it in the *Assignments* section of the course.

For help uploading files please refer to the *Technical Support* page in the syllabus.

Criterion	A	B	C	D	F	Letter Grade
Correctness and Completeness of Results (70%)	All steps' results are entirely complete and correct	About ¾ of the steps' results are correct and complete	About half of the steps' results are correct and complete	About ¼ of the steps' results are correct and complete	Virtually none of the step's results are correct and complete.	
Constitution of SQL/Python and Explanations (30%)	Excellent use and integration of appropriate SQL/Python constructs and supporting explanations	Good use and integration of appropriate SQL/Python constructs and supporting explanations	Mediocre use and integration of appropriate SQL/Python constructs and supporting explanations	Substandard use and integration of appropriate SQL/Python constructs and supporting explanations	Virtually all SQL/Python constructs and supporting explanations are unsuitable or improperly integrated	
					Assignment Grade:	