

مقدمه

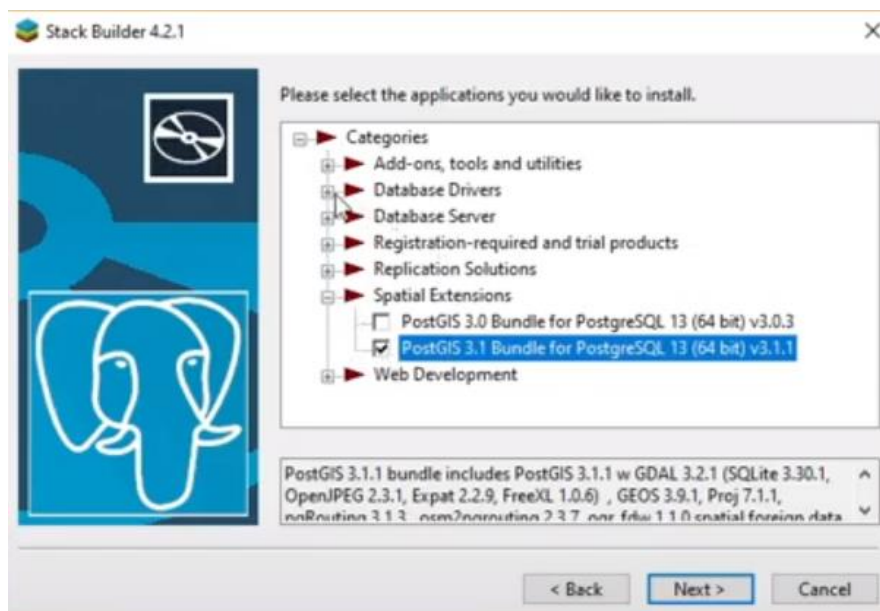
در این تمرین ما قصد داریم تا با databaseها کار کنیم و با تعدادی از ابزارهای معروف که به منظور ارتباط با databaseها استفاده می‌شود، آشنا شویم.

در این تمرین از PostgreSQL به عنوان یکی از RDBMSهای (دیتابیس‌های رابطه‌ای) رایج و حرفه‌ای استفاده می‌کنیم، و قصد بررسی و کار با آن و تعدادی از ابزارهای مرتبطش را داریم.

ابزارهایی که در این تمرین به بررسی آنها می‌پردازیم عبارت‌اند از: pgAdmin، psycopg2 و همین‌طور سعی می‌کنیم تا با ابزار QGIS کار کنیم و یک آشنایی اولیه با آن داشته باشیم.

ابتدا شما باید PostgreSQL را بر روی سیستم خود نصب کنید و به منظور دانلود آن می‌توانید که از [این لینک](#) اقدام کنید. همچنین به منظور آشنایی با نحوه نصب PostgreSQL می‌توانید از [این ویدیو](#) استفاده کنید.

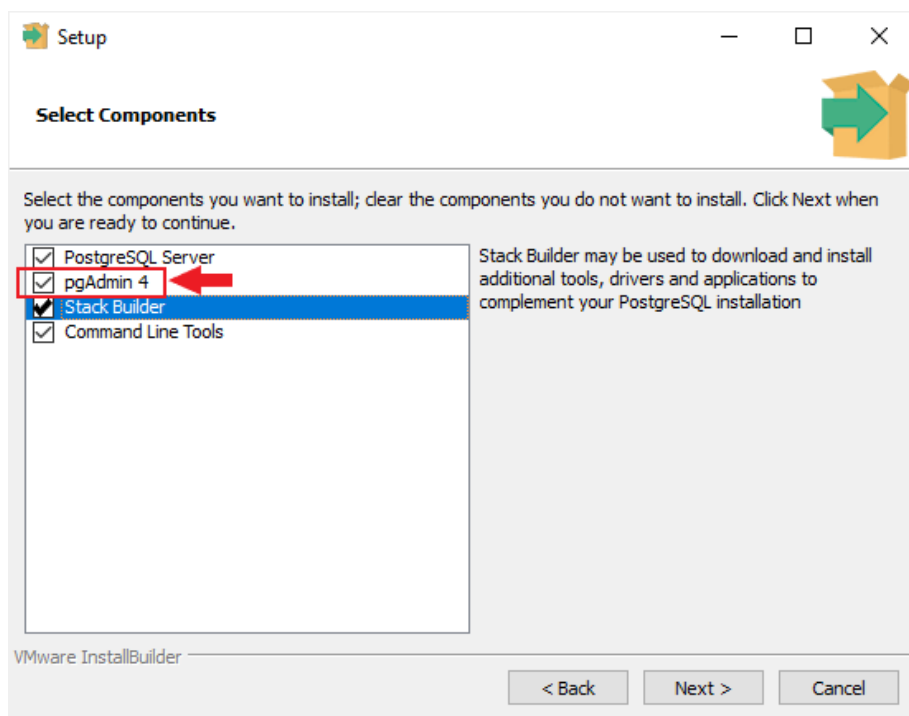
هنگام نصب PostgreSQL دقت کنید که مطابق تصویر زیر تیک مربوط به نصب ماژول PostGIS را نیز بزنید تا بتوانید در قسمت مربوط با این ماژول نیز کار کنید.



شکل ۱: نصب ماژول PostGIS (در سیستم عامل windows)

به هنگام نصب PostgreSQL نرم افزار pgAdmin نیز نصب می شود (باید component آن را هنگام نصب انتخاب کنید). اما اگر بخواهید به صورت جداگانه آخرین نسخه pgAdmin را نیز نصب کنید می توانید از طریق [این لینک](#) آن را دانلود کرده و نصب کنید.

به منظور نصب شدن pgAdmin هنگام نصب نرم افزار تیک زیر را بزنید.



شکل ۲: نصب کامپوننت های مختلف هنگام نصب PostgreSQL

همچنین به منظور نصب نرم افزار QGIS از طریق [این لینک](#) می توانید اقدام کنید.

بخش اول – آشنایی با SQL

در این بخش از تمرین با دستورات پایه SQL آشنا میشویم. اگر قصد کار تخصصی در حوزه نرم افزار و داده را دارید، آشنایی با SQL به صورت تخصصی جزو ملزومات اصلی این کار خواهد بود و در این تمرین هم برآنیم که تا حد امکان به این هدف برسیم.



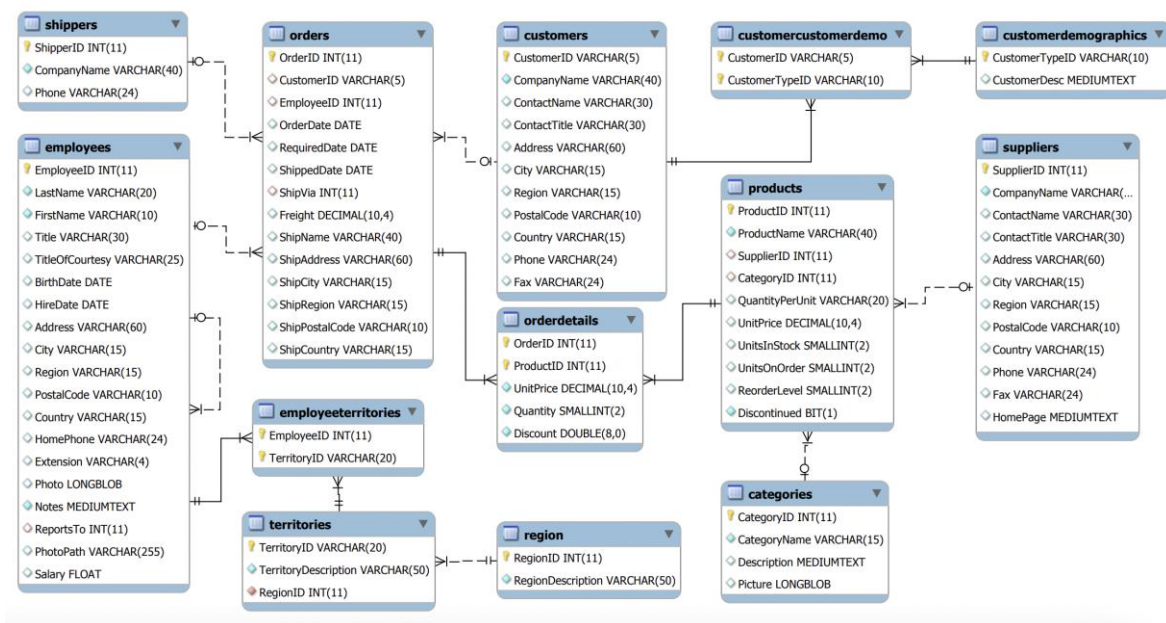
پایگاه داده‌ای که برای این تمرین در نظر گرفته شده است، database معروف آموزشی میکروسافت با نام Northwind^۱ است که برای اهداف آموزشی ما، بسیار مناسب است و نسخه Postgres آن را در این تمرین استفاده خواهیم کرد.

این دیتابیس که یک فروشگاه بزرگ را مدل‌سازی می‌کند، از جداول اصلی زیر تشکیل شده است:

- **Customer**: مشتریان فروشگاه
- **Order**: سفارشات هر مشتری – اطلاعات کلی
- **OrderDetails**: جزییات هر سفارش شامل آیتم‌های خریداری شده
- **Products**: محصولات فروشگاه
- **Suppliers**: تامین کنندگان کالا که محصولات از آن‌ها تهیه می‌شوند.
- **Categories**: گروه‌بندی کالاها
- **Shippers**: شرکت‌های حمل و نقلی که وظیفه ارسال کالا به دست مشتری را بر عهده دارند.
- **Employee**: کارمندان فروشگاه که هر سفارش، توسط یک کارمند ثبت یا تایید می‌شود.
- **Territory & Region**: مناطق و نواحی‌ای که این فروشگاه در آن‌ها شعبه دارد. هر منطقه شامل چندین ناحیه است و هر کارمند می‌تواند به سفارشات چندین ناحیه رسیدگی کند.
- **CustomerDemographics**: برای رتبه‌بندی مشتریان استفاده می‌شود که در این تمرین با آن سر و کار نخواهیم داشت.

در ادامه می‌توانید نمودار ERD این database را مشاهده کنید:

^۱ <https://github.com/Microsoft/sql-server-samples/tree/master/samples/databases/northwind-pubs>



شکل ۳: نمودار ERD پایگاه داده Northwind

در ابتدا باید جداول این دیتاست به همراه دیتای موجود در آنها را وارد database ساخته شده در PostgreSQL کنید.

برای شما دو فایل به نامهای northwind.sql و northwind_data.sql در داخل پوشه Northwind_SQL فراهم شده است که فایل اول شامل دستورات مربوط به ایجاد جداول و فایل دوم شامل دستورات مربوط به وارد کردن دیتاهای اولیه به داخل database است. شما باید محتوای این دو فایل را وارد یک database که در PostgreSQL می‌سازید کنید. (ابتدا باید با فایل اول جداول را ساخته و سپس با فایل دوم دیتا را وارد کنید).

در این قسمت می‌توانید از ویدیوی آموزشی شماره ۴۸ به نام Database Operations in pgAdmin که برای شما فراهم شده است استفاده کنید.

به منظور اجرای queryها در این بخش از تمرین می‌توانید از دو روش مختلف استفاده کنید. روش اول این است که ماژول مربوط به اجرای دستورات SQL در Jupyter Notebook را نصب کنید و سپس اقدام به اجرای دستورات SQL در Jupyter Notebook کنید. روش دوم این است که مطابق روش گفته شده در ویدیو شماره ۴۸، queryهای خود را در قسمت مربوطه در pgAdmin اجرا کنید. هر دو روش یکسان هست و فقط باید در گزارش کد مربوطه به همراه قسمتی از خروجی را قرار دهید.



A: ابتدا query ها و کدهای نوشته شده در ویدیو SQL Hands-on داخل سامانه را خودتان در قسمت‌های مربوطه در فایل نوتبوک SQL_HandsOn.ipunb که در پوشه HandsOn_Notebooks قرار دارد، بنویسید. (یا می‌توانید در PgAdmin آنها را نوشته و خروجی به همراه دستور نوشته شده را در گزارش خود بیاورید).

B: query های زیر را در یک نوتبوک پایتون یا pgAdmin نوشته و خروجی مربوطه به همراه query نوشته شده را در گزارش خود بیاورید.

۱. تعداد کارمندان در هر منطقه (region) را بدست آورید.
۲. به ازای هر منطقه (region)، چه نواحی‌ای (territories) تعریف شده است؟ فقط نام منطقه و نام ناحیه در خروجی باید ظاهر شود.
۳. نام و نام‌خانوادگی تمام کارمندان ناحیه Chicago (حرف اول آن بزرگ است) را به دست آورید.
۴. می‌خواهیم سفارش جدیدی برای محصولاتی که تعداد آنها (UnitsInStock) از حد مجاز (ReorderLevel) کمتر شده است، بدهیم. این محصولات را بیابید و آنها بر اساس میزان موجودی (به صورت صعودی) نمایش دهید.
۵. کدام محصول بیشترین سفارش را در یک ماه اخیر داشته است؟
۶. مشتریانی را بیابید که شماره فکس آنها در سیستم موجود نیست.
۷. می‌خواهیم براساس میزان عملکرد کارکنان، به آنها پاداش بدهیم. سه کارمندی را بیابید که بیشترین تعداد سفارش را تایید یا ثبت کرده‌اند.
۸. از کدام دسته محصولات، به آلمان چیزی ارسال نشده است؟
۹. می‌خواهیم کارکنان Chicago و Orlando را به New York منتقل کنیم. دستور آپدیت متناظر را بنویسید. دقت کنید که تنها نام این شهرها را می‌دانیم و در دستور آپدیت هم از این سه نام استفاده کنید. (امتیازی)
۱۰. به دلیل بالا رفتن بهای شیر ما باید محصولاتی را که Category آنها Dairy Products هست را به میزان ۱۵٪ افزایش قیمت دهیم. (این تغییر را در جدول مربوطه اعمال کنید)
۱۱. همچنین تامین کننده‌ی مالک شرکت Mayumi's اقدام به افزایش ۱۰٪ ای محصولات خود کرده است. این تغییر را نیز در جدول مربوطه اعمال کنید. همچنین این تامین کننده به تعداد ۱۵ عدد از محصولات خود را پس از این افزایش قیمت به انبار ما ارسال کرده است. (موجودی مربوطه را نیز به روز کنید). (امتیازی)

بخش دوم – کار با psycopg2

در این قسمت ما قصد داریم تا با کتابخانه psycopg2 پایتون که به منظور تعامل با PostgreSQL ایجاد شده است کار کنیم. ابتدا مطابق راهنمایی‌های موجود در فایل `psycopg2_HandsOn.ipynb` موجود در پوشه `HandsOn_Notebooks` اقدام به نصب کتابخانه‌های موجود و همین‌طور آشنایی با روال کارکرد این کتابخانه کنید. همچنین به منظور آشنایی شما با این ابزار ویدیوی شماره ۵۰ رفرنس درسی را حتما مشاهده کنید.

A: اقدام به بازنویسی موارد جای خالی در فایل Hands-on مطابق با فیلم شماره ۵۰ با عنوان Psycopg2 HandsOn کنید.

B: در این قسمت از شما خواسته می‌شود که با یک دیتاست دیگر کار کنید و اقدام به طراحی یک پایگاه داده به منظور ذخیره داده‌های این جدول در PostgreSQL کنید.

۱. از شما به منظور ارزیابی جهت استخدام در مرکز ملی آمار ایران خواسته می‌شود که یک دیتابیس به منظور نظم‌دهی و همگام‌سازی اطلاعات مرگ و میر کشور طراحی کنید.

a. بدین منظور یک مجموعه دیتاست در اختیار شما قرار داده شده است که مربوط به آمار مرگ و میر در ایران به صورت هفتگی از سال ۱۳۹۵ تا ۱۳۹۹ می‌باشد. این مجموعه فایل‌ها را می‌توانید داخل پوشه Datasets بیابید.

b. این اطلاعات به صورت استانی دسته‌بندی شده است و همچنین اطلاعات سنی و جنسیت افراد را نیز شامل می‌شود.

c. شما باید از این دیتاست و با توجه به مفاهیمی که تاکنون از پایگاه‌داده فراگرفته‌اید اقدام به طراحی یک پایگاه‌داده به منظور ذخیره این اطلاعات کنید به صورتی که این پایگاه داده این قابلیت را داشته باشد تا توسط سازمان آمار بنابر اطلاعاتی که به صورت هفتگی از استان‌های مختلف دریافت می‌کند، آن را به روز کند و مرگ و میرهای جدید را در آن ثبت نماید.

d. به منظور طراحی این دیتابیس می‌توانید از سایت [DBDesigner](http://dbdesigner.org) استفاده کنید. این وبسایت به شما این امکان را می‌دهد که نمودار ER طراحی کنید و سپس خروجی SQL مطابق هر RDBMS ای که با آن کار می‌کنید بگیرید. (این وبسایت قبلا کاملا

رایگان بود اما مدتی است برای طراحی دیتابیس‌های بزرگ نیاز به پرداخت هزینه دارد اما برای پروژه شما میتوان از همان نسخه رایگان استفاده کرد.

e. به منظور آشنایی بیشتر با نمودارهای ER پیشنهاد می‌کنم که [این پلی‌لیست](#) یوتیوب را مشاهده کنید.

f. پس از طراحی نمودار ER و ساخت و طراحی پایگاه‌داده اقدام به وارد کردن داده‌ها به داخل پایگاه داده کنید. (طراحی ER اجباری نیست و صرفاً برای راحت‌تر شدن طراحی جداول است و می‌توانید که مستقیماً کد جداول را بنویسید).

g. توجه: هدف ما در این درس با توجه به این که دوره تخصصی پایگاه‌داده نیست، این نیست که شما یک پایگاه‌داده را به صورت اصولی طراحی کنید و فرم‌های نرمال مختلف و سایر ویژگی‌های یک سیستم پایگاه داده را به صورت دقیق و کامل پیاده کنید به همین دلیل در زمینه طراحی پایگاه داده سخت‌گیری زیادی نخواهد شد.

۲. پس از ساخت دیتابیس و آماده کردن آن اقدام به نوشتن queryهای زیر با استفاده از Psycopg2 کنید و سپس خروجی مربوطه را به صورت خواسته شده نمایش دهید.

a. مجموع مرگ و میر کل کشور در سال‌های مختلف را بدست آورید و با استفاده از bar-chart آن را نمایش دهید.

b. نمودار قسمت قبل را مجدد به صورت stack bar-chart و به تفکیک زن و مرد رسم کنید.

c. در سال ۱۳۹۹، ۱۰ عدد از استان‌های دارای بیشترین مرگ و میر در سن ۰ سال (هنگام تولد) را بدست آورید و نتایج را بر روی bar-chart نمایش دهید.

d. در سال ۱۳۹۹، ۱۰ استان دارای بیشترین مرگ و میر مردان و زنان را به صورت جداگانه بدست آورید و نتایج را نمایش دهید.

e. سوال قسمت c را برای سال ۱۳۹۵ نیز انجام دهید.

f. داده‌های موجود در فایل Iran_population.csv که اطلاعات جمعیتی استان‌های مختلف ایران را نشان می‌دهد را در table مربوط به هر استان در پایگاه داده جایگذاری کنید و سپس نرخ مرگ‌ومیر (نسبت تعداد مرگ به جمعیت هر استان) در سال‌های مختلف را با استفاده از این فیلد جدید اضافه شده بدست آورید و در نمودار نمایش دهید. (*امتیازی)

۳. از نتایج و داده‌های رسم شده چه نتایجی را دریافت می‌کنید؟ تحلیل کنید.

۴. تحقیق کنید که مشکل استفاده زیاد و متوالی از `cursor.connect(**params)` چیست؟ یکی از راه‌های رفع این مشکل را بیان کنید.

بخش سوم – کار با QGIS (امتیازی*)

در این سوال هدف یک آشنایی اولیه با نرم‌افزار QGIS و همچنین اتصال یک پایگاه‌داده PostgreSQL به این نرم‌افزار از طریق ماژول PostGIS است.

QGIS یک نرم‌افزار سیستم اطلاعات جغرافیایی متن‌باز و cross-platform می‌باشد که به ما امکان مشاهده، تحلیل و ویرایش داده‌های جغرافیایی را می‌دهد.

به منظور آشنایی با این نرم‌افزار و همچنین قابلیت‌های مختلف آن در امور جغرافیایی و نقشه‌برداری و... پیشنهاد می‌کنم تا [این پلی‌لیست](#) که مربوط به یک آموزش نسبتاً کامل از این نرم‌افزار توسط خانم دکتر کفاش استاد دانشگاه تبریز است را در آپارات مشاهده کنید.

A: سه عدد فایل SQL به نام‌های `Iran_States.sql`، `Iran_County.sql` و نیز `Iran_Country.sql` به همراه پروژه شما آپلود شده است. شما باید با استفاده از این فایل‌ها مشابه سوال اول یک پایگاه داده بسازید و دستورات مورد نیاز را برای وارد کردن جدول‌های مربوط به مختصات‌های جغرافیایی استان‌ها، شهرستان‌ها و کل کشور ایران وارد کنید تا جداول مربوطه ساخته شوند. سپس در نرم‌افزار QGIS از طریق ماژول PostGIS اقدام به ایجاد یک ارتباط جدید با پایگاه‌داده ساخته شده در PostgreSQL کنید و به این طریق داده‌ها را داخل نرم‌افزار QGIS بارگزاری کنید. پس از انجام این موارد می‌توانید با استفاده از داده‌های بارگزاری شده لایه‌های جدید ایجاد کنید و این موارد را بر روی نقشه مشاهده کنید.

برای مشاهده نقشه‌ها پیشنهاد می‌کنم که لایه OpenStreetMap را نیز در نرم‌افزار QGIS فعال کنید.

توجه: بعد از ساختن پایگاه‌داده باید حتماً با استفاده از دستور `CREATE EXTENSION postgis` اقدام به نصب ماژول مربوطه در pgAdmin کنید تا بتوانید با نرم‌افزار QGIS تعامل داشته باشید.

B: پیدا کردن موقعیت: در این قسمت به شما یک موقعیت داده می‌شود و شما باید با نوشتن query مورد نظر محل (مانند اسم استان، شهرستان و...) را بدست آورید.



توجه: سیستم CRS (Project Coordinate Reference System) استفاده شده در داده‌هایی که در اختیار شما قرار گرفته شده است، EPSG:4326 می‌باشد.

a. مشخص کنید که موقعیت (35.7242345 51.3867718) کدام استان و شهرستان را مشخص می‌کند.

b. اکنون به همان شیوه قبل فایل Iran_inland_waters.sql را نیز وارد پایگاه داده خود کنید. این فایل شامل داده‌های مربوط به آب‌های درون سرزمینی ایران است. این پایگاه داده را نیز در QGIS مشابه قبل بارگذاری کنید و آن را مشاهده کنید.

c. دو مختصات (31.0624 53.6166) و (31.2736 61.2768) مشخص کننده چه دریاچه‌هایی هستند و همچنین در کدام استان و شهرستان قرار دارند؟



نکات تحویل

- مهلت ارسال این تمرین تا پایان روز جمعه ۵ آذر ماه خواهد بود.
- انجام این تمرین به صورت یک نفره می باشد.
- لطفا هر گونه فرض در حل سوالات را در گزارش خود ذکر کنید.
- به منظور تحویل این تمرین می توانید که گزارش ها را در فایل Notebook مربوطه بنویسید یا این که به صورت pdf گزارش خود را ارسال کنید.
- توجه کنید که خروجی نمودارها و تمامی فرض ها و نیز نتایجی که بدست می آورید را باید در گزارش خود بیاورید.
- لطفا گزارش ، فایل کدها و سایر ضmann مورد نیاز را با فرمت زیر در سامانه مدیریت دروس بارگذاری نمائید.

HW4_[Lastname]_[StudentNumber].zip

- در صورت وجود سوال و یا ابهام می توانید از طریق رایانامه زیر با دستیار آموزشی در ارتباط باشید:
p.zilouchian@ut.ac.ir