

# Getting Started with PostgreSQL

M. Edward (Ed) Borasky

2018-03-02

## Introduction: PostgreSQL in Context

## Industry trends: NoSQL databases (Redmond and Wilson 2012)

- ▶ Graph databases
- ▶ Key-value stores
- ▶ JSON document stores
- ▶ In-memory databases

## Industry trends: Backend frameworks (Apache/PHP, Ruby on Rails, Django)

- ▶ Work with *any* database (MySQL/MariaDB, SQLite, PostgreSQL)
- ▶ Just use the database for CRUD, application logic is all in PHP / Ruby / Python / JavaScript code!
- ▶ Wait - CRUD?
  - ▶ Create
  - ▶ Read (aka SELECT)
  - ▶ Update
  - ▶ Delete

## If all you want . . .

- ▶ If all you want is an industrial-strength open-source permissive-licensed CRUD engine that
  - ▶ Is fully ACID compliant
  - ▶ Scales to huge installations
  - ▶ Has replication / failover / high availability as standard equipment
- ▶ Yeah, PostgreSQL's got that.

## But if you also want . . .

- ▶ Full-text search
- ▶ Stored procedures in Python, Perl, Ruby, R, Tcl and Lua
- ▶ Foreign data (Text files, GIS data, MySQL/MariaDB, Redis) mapped into your database (foreign data wrappers)
- ▶ Key-value stores (hstore)
- ▶ JSON document stores (jsonb)
- ▶ Yeah, PostgreSQL's got that too!

## Speaking of industrial strength ...

- ▶ Geographic Information Systems (GIS)
- ▶ PostGIS
  - ▶ Read and write GIS data files
  - ▶ Process geometric, geographic and topology GIS data types
  - ▶ Both vector and raster data
  - ▶ Geocoding, reverse geocoding, address standardization
- ▶ pgRouting
  - ▶ Shortest / fastest / lowest cost routes from point A to point B
  - ▶ Traveling salesperson problem
  - ▶ Turn-by-turn directions for cars, bikes and pedestrians!
- ▶ Yeah, I want that CRUD too!

## PostgreSQL on the Desktop - single user



## Windows or Mac

- ▶ Go to EnterpriseDB download site <https://www.enterprisedb.com/downloads/postgres-postgresql-downloads>.
- ▶ Select the latest version (10.3)
- ▶ Select your operating system (Windows or Mac)
- ▶ Installation: install everything but don't run StackBuilder yet

## Linux: use the PGDG repositories

- ▶ RHEL / CentOS / Fedora:  
<https://www.postgresql.org/download/linux/redhat/>
- ▶ Ubuntu: <https://www.postgresql.org/download/linux/ubuntu/>  
(probably works for Linux Mint)
- ▶ Debian: <https://www.postgresql.org/download/linux/debian/>
- ▶ Installation: install PostgreSQL 10.3 and pgAdmin 4 for desktop

## Connecting with pgAdmin

- ▶ Reference: R. Obe and Hsu (2017b), chapter 4

## Exploring the tree

## Creating a database

## Creating tables - Data Definition Language (DDL)

To create a table, you need to do two things:

1. Define the names and data types of every column in the table.
2. Load the data into the table.

We'll be working with a sample file of Oregon highway mileposts.

You'll find it in

[https://github.com/hackoregon/data-science-pet-containers/blob/master/examples/mileposts/Mileposts\\_2014/Mileposts\\_2014.csv](https://github.com/hackoregon/data-science-pet-containers/blob/master/examples/mileposts/Mileposts_2014/Mileposts_2014.csv).

## A cheat code if you're in a hurry

If you don't know the data types, you can always just set them all to `text` and re-cast them to the correct type later! But in this case it's mostly obvious which columns are numeric or timestamps, and we can use `text` for the rest.

For the mileposts, we'll use `double precision` for the latitude and longitude and `date` for `GIS_PRC_DATE` and `text` for the others.

## The finished DDL

```
CREATE TABLE mileposts_2014 (  
    hwyname text, hwynumb text, st_hwy_sfx text,  
    rdwy_id text, mlge_typ text, ovlp_cd text,  
    mp text, mp_desc text, mp_disp text,  
    lrs_key text, lrm_key text,  
    lat double precision, longtd double precision,  
    hrz_col_m text, crd_rf_dtm text, effectv_dt text,  
    gis_prc_dt date);
```



## Hey, Ed, what's with all the lower case names?

PostgreSQL requires special care in coding SQL queries when column names have anything besides lower-case letters, numbers or underscores. You have to enclose them in double-quotes. It's a real hassle, so “snake\_case” rules!

## Reading a CSV file into a table - COPY

```
COPY mileposts_2014 FROM  
'/d/Projects/data-science-pet-containers/examples/mileposts  
WITH CSV HEADER;
```

## Backing up a database

## Restoring a database

## Creating a query

## PostgreSQL on a (Linux) Server

## Two dialects of Linux

- ▶ Ubuntu 16.04.x LTS (most popular)
- ▶ RHEL 7 / CentOS 7

## Docker container

- ▶ Debian stable - similar to Ubuntu
- ▶ Alpine - avoid this unless you want to do a lot of research
- ▶ I have a full PostgreSQL / PostGIS / pgRouting stack in a container at

<https://github.com/hackoregon/data-science-pet-containers>.

- ▶ See <https://github.com/hackoregon/data-science-pet-containers/blob/master/containers/small.yml> for the Docker compose file
- ▶ The documentation's a bit sparse still
- ▶ Native executables will perform better
- ▶ Docker hosting on a desktop / laptop isn't exactly end-user friendly yet



## The PGDG repositories

- ▶ Linux distributions ship with the version of PostgreSQL that was stable when their release process froze features.
- ▶ You'll still get security updates and bug fixes, but no new features.
- ▶ So the PostgreSQL Global Development Group (PGDG) maintains up-to-date binary repositories for all the major Linux distros.
- ▶ These binaries are as well tested and supported as those that ship with the Linux distros.
- ▶ Download page: <https://www.postgresql.org/download/>

## Configuring

- ▶ Initializing the cluster
- ▶ Enabling and starting the service

## Listening address and port

# Authentication

## Fine-grained permissions and roles

# Tablespaces

# PostGIS

## Reading in GIS data (Obe and Hsu 2015, chap. 4)

- ▶ Shapefiles
- ▶ “GDB” databases
- ▶ OpenStreetMap data



## Tagging points with a geometry column

## Geocoding (Obe and Hsu 2015, chap. 8)

- ▶ There's a demo using a Docker container at <https://github.com/hackoregon/data-science-pet-containers/tree/master/examples/geocoding>.
- ▶ This uses a container running Debian Linux with a full PostgreSQL / PostGIS / pgRouting stack.
- ▶ ***I have not tested it on a Windows Docker host, just Linux. It undoubtedly needs some adjustments for the volumes mounted on host filesystems.***

# Spatial joins

## pgRouting - what it can do

## Turn-by-turn directions

- ▶ Reference: (R. Obe and Hsu 2017a)

## References

## References

Obe, Regina O., and Leo S. Hsu. 2015. *PostGIS in Action, Second Edition*. Manning Publications Co. <https://www.manning.com/books/postgis-in-action-second-edition>.

Obe, Regina, and Leo Hsu. 2017a. *PgRouting: A Practical Guide*. Locate Press LLC. <https://locatepress.com/pgrouting>.

———. 2017b. *PostgreSQL: Up and Running, 3rd Edition*. O'Reilly Media, Inc. <http://shop.oreilly.com/product/0636920052715.do>.

Redmond, Eric, and Jim R. Wilson. 2012. *Seven Databases in Seven Weeks*. The Pragmatic Programmers. <https://pragprog.com/book/rwdata/seven-databases-in-seven-weeks>.