

---

# MSDS 420

Atef Bader, PhD

# Agenda

---

1. VPN to Connect to Northwestern Network
2. DSCC – PostgreSQL Server
3. Assignment #3 Walkthrough and Deliverable

# Helpdesk Support Info

---

1. **Northwestern IT Helpdesk:**

- For **VPN** help call this number: 847-491-4357

2. **SPS IT Helpdesk:**

- For **DSCC** help call this number: 312-503-3333

# VPN - Northwestern

File Edit View History Bookmarks Tools Help

Syllabus: 2018SU\_MSDS\_420-D X

https://canvas.northwestern.edu/courses/76687/pages/syllabus?module\_item\_id=... 110% Search

Most Visited Getting Started

## Course Technology

This course will involve a number of different types of interactions. These interactions will involve navigating through the course and become familiar with the course syllabus, structure, and content.

## Network Connection to DSCC

The DSCC serves as a research and training facility for graduate students in the Master of Science in Data Science. The DSCC is a research and training facility for graduate students in the Master of Science in Data Science. They are for student and faculty use only and remain available for student and faculty use only. Each account is tied with its Northwestern University NetID and password. Do not share user account NetIDs and passwords with others. Each account is tied with its Northwestern University NetID and password.

## Technical Support for DSCC

For DSCC technical support related to server connection/login issues, email [sps-it@northwestern.edu](mailto:sps-it@northwestern.edu).

Tutorial: [Working with the School of Professional Studies Data Science Computing Cluster \(DSCC\) \(PDF\)](#)

## Using VPN to connect Northwestern Network

If you are a remote user (not on Northwestern University campus/network), you need to connect to the university network through VPN in order to connect to DSCC. Please follow the [instructions from this link](#) for NU VPN.

## Canvas

The [Canvas Student Center](#) includes information on communicating in Canvas, navigating a Canvas course, grades, additional help, and more. The [Canvas at Northwestern](#) website provides information of getting to know Canvas at Northwestern and getting Canvas support. The [Canvas Student Guide](#) provides tutorials on all the features of Canvas. For additional Canvas help and support, you can always click the Help icon in the lower left corner to begin a live chat with Canvas support or contact the Canvas Support Hotline.

- Bookmark this URL for VPN Documentations
- You need to connect to Northwestern Network over VPN for this homework

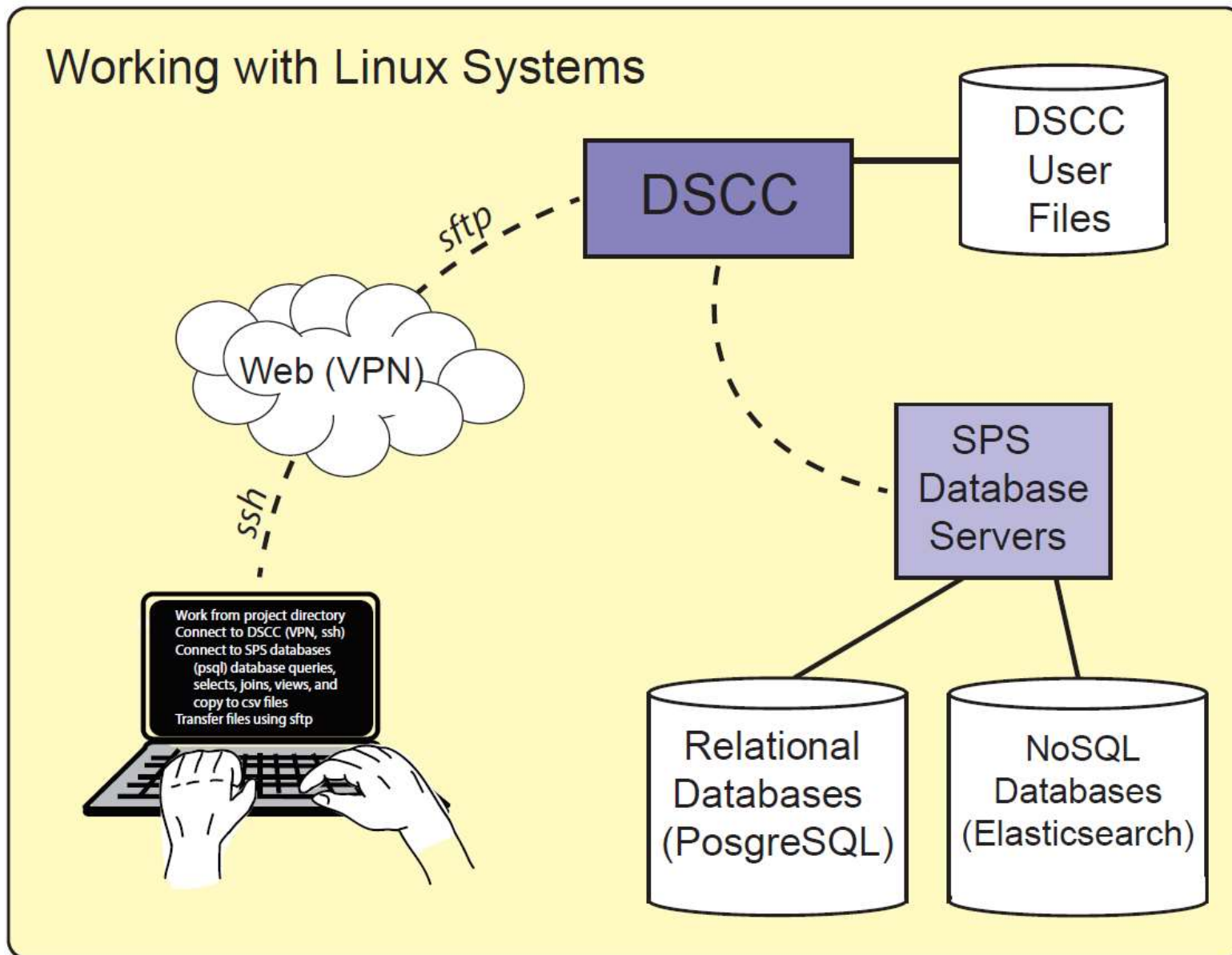
6d You are currently logged into Student View

Resetting the test student will clear all history for this student, allowing you to view the course as a brand new student.

Reset Student Leave Student View

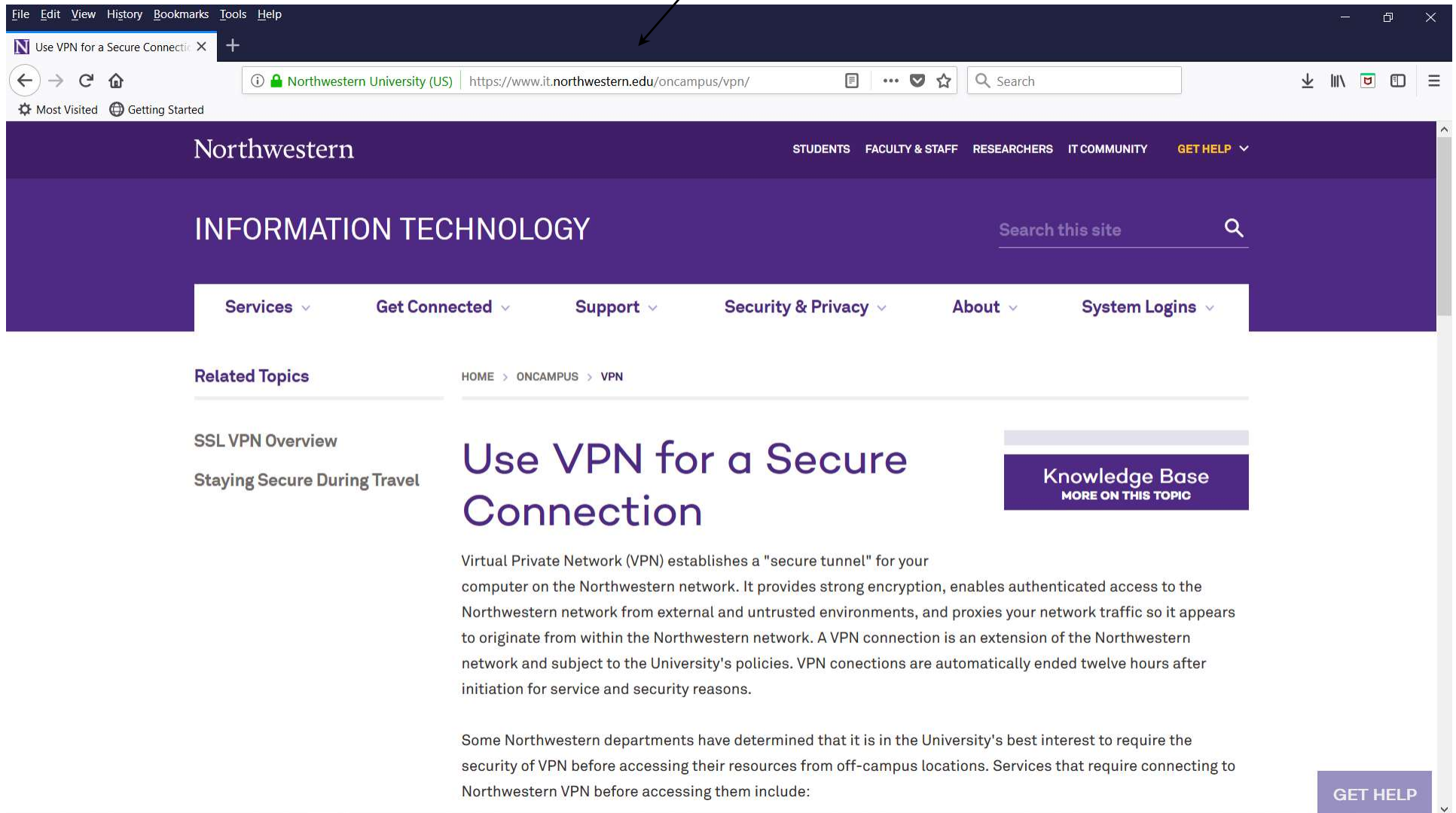
4

## Network elements/systems you need to know for Assignment #3?



# VPN - Northwestern

Bookmark the URL for VPN



File Edit View History Bookmarks Tools Help

Use VPN for a Secure Connection X +

Northwestern University (US) | <https://www.it.northwestern.edu/oncampus/vpn/>

Search

Most Visited Getting Started

Northwestern

STUDENTS FACULTY & STAFF RESEARCHERS IT COMMUNITY GET HELP

INFORMATION TECHNOLOGY

Search this site

Services Get Connected Support Security & Privacy About System Logins

Related Topics

HOME > ONCAMPUS > VPN

SSL VPN Overview

Staying Secure During Travel

## Use VPN for a Secure Connection

Knowledge Base  
MORE ON THIS TOPIC

Virtual Private Network (VPN) establishes a "secure tunnel" for your computer on the Northwestern network. It provides strong encryption, enables authenticated access to the Northwestern network from external and untrusted environments, and proxies your network traffic so it appears to originate from within the Northwestern network. A VPN connection is an extension of the Northwestern network and subject to the University's policies. VPN connections are automatically ended twelve hours after initiation for service and security reasons.

Some Northwestern departments have determined that it is in the University's best interest to require the security of VPN before accessing their resources from off-campus locations. Services that require connecting to Northwestern VPN before accessing them include:

GET HELP

# VPN - Northwestern

The screenshot shows a web browser window displaying the Northwestern University VPN page. The browser's address bar shows the URL <https://www.it.northwestern.edu/oncampus/vpn/>. The page title is "Linux". The main heading is "SSL VPN (Secure Sockets Layer Virtual Private Network)". Below the heading, the text reads: "SSL VPN allows access to administrative systems, critical system administrators. SSL VPN access can be granted to and other external collaborators. Please visit the SSL VPN". A callout box with a black border and a purple background points to the text "administrative systems, critical system administrators". The callout box contains the text "Call or Email Helpdesk for VPN support". Below the main text, it says "Last Updated: 5 March 2018". The footer of the page is purple and contains the Northwestern University logo, contact information for Information Technology, social media links, and a "GET HELP" button.

File Edit View History Bookmarks Tools Help

Use VPN for a Secure Connection X +

Northwestern University (US) <https://www.it.northwestern.edu/oncampus/vpn/> Search

Most Visited Getting Started

Linux

## SSL VPN (Secure Sockets Layer Virtual Private Network)

SSL VPN allows access to administrative systems, critical system administrators. SSL VPN access can be granted to and other external collaborators. Please visit the SSL VPN

Last Updated: 5 March 2018

Call or Email Helpdesk for VPN support

**Northwestern University**

© 2016 Northwestern University  
[Contact Northwestern University](#)  
[Careers](#)  
[Disclaimer](#)  
[Campus Emergency Information](#)  
[University Policies](#)

**Information Technology**  
1800 Sherman Ave  
Evanston, IL 60208  
847-491-4357 (1-HELP)  
[consultant@northwestern.edu](mailto:consultant@northwestern.edu)  
[Share feedback about this page](#)

**Social Media**  
[Twitter](#) [Facebook](#) [YouTube](#)  
[RSS](#)  
[View All Social Media](#)

**Students**  
[Faculty & Staff](#)  
[Researchers](#)  
[IT Community](#)  
[Get Help](#)

**GET HELP**

# DSCC - Helpdesk

---

## □ Network Connection to DSCC

- After you establish VPN connection to Northwestern Network, use your NetID and Password to connect to DSCC server
- The DSCC serves as a research and training facility for graduate students in the Master of Science in Data Science program. User accounts are not associated with individual courses or instructors. They are for student and faculty use only and remain available as long as users maintain valid Northwestern NetIDs. Users should not share user account **NetIDs** and passwords with others. Each account is tied with its Northwestern University network identity.

## □ Technical Support for DSCC:

- For DSCC technical support related to server connection/login issues, email or call : **sps-it@northwestern.edu** and phone number **312-503-3333**



# PostgreSQL Documentations

The screenshot shows a web browser window displaying a Canvas LMS course page. The browser's address bar shows the URL: [https://canvas.northwestern.edu/courses/76687/pages/syllabus?module\\_item\\_id-](https://canvas.northwestern.edu/courses/76687/pages/syllabus?module_item_id-). The page content includes a sidebar with navigation links (Account, Dashboard, Courses, Calendar, Inbox, Help) and a main content area with the following sections:

- Programming Skills Objectives**
  - Use a high-level programming language for data extraction, preparation and exploration.
  - Compare and contrast SQL database and NoSQL database systems and demonstrate the ability to interact with SQL and NOSQL databases.
  - Construct indexes for information retrieval.
  - Store and retrieve geospatial data and plot data on interactive maps.
- Prerequisites**

MSDS 402-DL: Introduction to Data Science is a prerequisite for this class.
- Textbooks & Resources**
  - Required Textbooks
  - Reference Books (Recommended)
  - Online Developer Resources, Tool Manuals, and User Guides

The 'Online Developer Resources, Tool Manuals, and User Guides' section contains the following links:

- [Postgres reference manual, and API documentation](#)
- [ElasticSearch reference manual, and API documentation](#)
- [Python general troubleshooting Q&A for code debugging](#)

A purple box with the text "Bookmark the URL for PostgreSQL Documentations" is overlaid on the page, with an arrow pointing to the first link in the resources section.

**Course Reserves**

Some readings will be available through the Course Reserves in the left navigation menu. The Syllabus and weekly roadmap will note which readings are to be accessed through Course Reserves. For assistance with Course Reserves, e-mail [e-reserve@northwestern.edu](mailto:e-reserve@northwestern.edu). To ask a librarian for

**Footer:** You are currently logged into Student View. Resetting the test student will clear all history for this student, allowing you to view the course as a brand new student. [Reset Student] [Leave Student View]

# Assignment #3

File Edit View History Bookmarks Tools Help

Home Chicago\_Crimes

localhost:8888/notebooks/Chicago\_Crimes.ipynb 110% Search

Most Visited Getting Started

jupyter Chicago\_Crimes Last Checkpoint: 04/18/2018 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Markdown

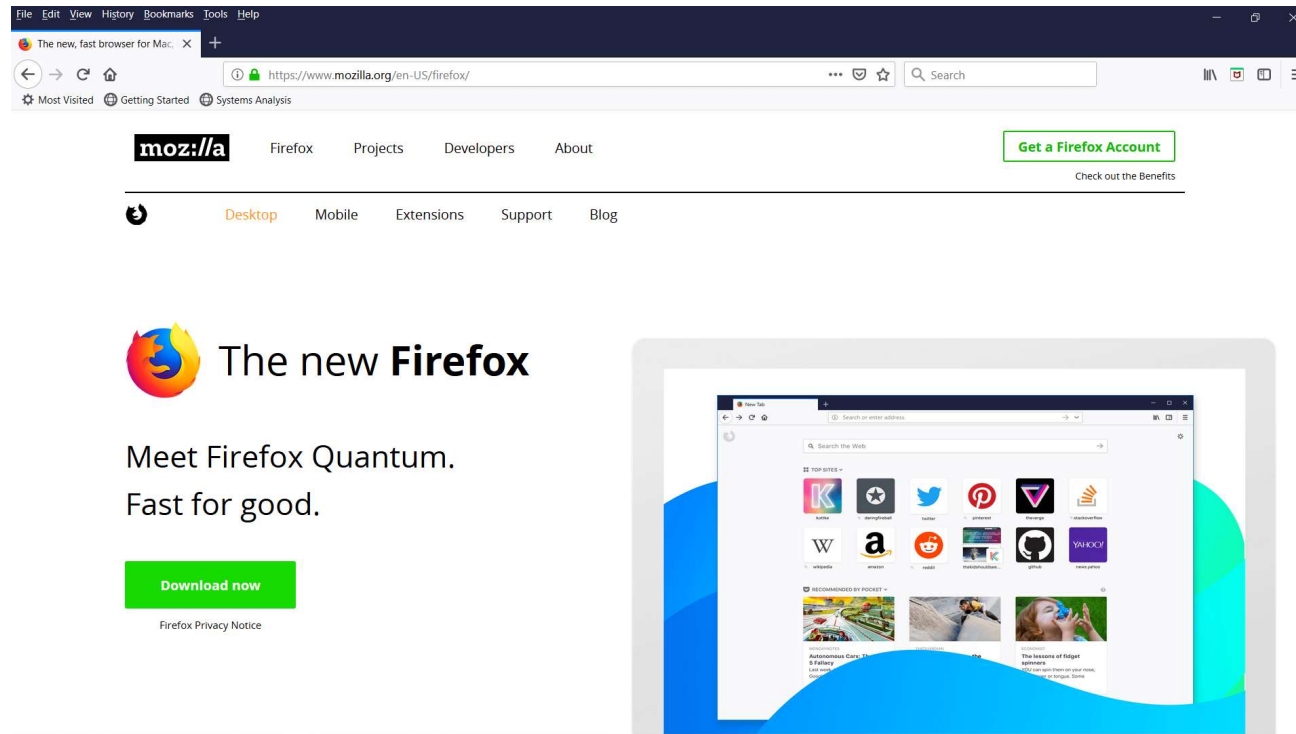
Trump on Chicago: 'A disaster' 'Out of control,' 'Not good!' ... 'It's a great city'

10 times Trump has talked about Chicago violence

Chicago Tribune

# Assignment #3 - Walkthrough

- Use Firefox browser for this assignment; other browsers might not work as expected for interactive Folium maps



# Assignment #3 - Walkthrough

---

- You need to be aware of the following:
  1. You need to update the script to connect to DSCC using your NetID and password
  2. This is a big assignment and will require at least 10 quality hours in order to be intimate with its technical content and provide answers for its deliverables
  3. Query #5 and #6 will take roughly 35 minutes to complete; be patient you have a big dataset with geospatial queries that are very demanding for CPU

# Assignment #3 - Walkthrough

File Edit View History Bookmarks Tools Help

Home Chicago\_Crimes

localhost:8888/notebooks/Chicago\_Crimes.ipynb 110% Search

Most Visited Getting Started

Jupyter

File

Trusted Python 3

Logout

Type your NetID and password

```
from folium import plugins
from folium.plugins import MarkerCluster
import psycopg2
import csv
import pandas as pd
import json
from area import area

from psycopg2.extensions import ISOLATION_LEVEL_AUTOCOMMIT
```

**IMPORTANT NOTES:**

- Use this version of the assignment if you don't want to install PostgreSQL server on your laptop/computer locally to experiment with database and tables creation
- In this version of the assignment you will be connecting to DSCC PostgreSQL server that has the database and tables already created on the server.
- You need to use your NetID and password for login and password to connect PostgreSQL server hosted on DSCC
- For the psycopg2.connect statements listed below, you must provide your NetID and password in order to connect to PostgreSQL server hosted on DSCC

In [2]:

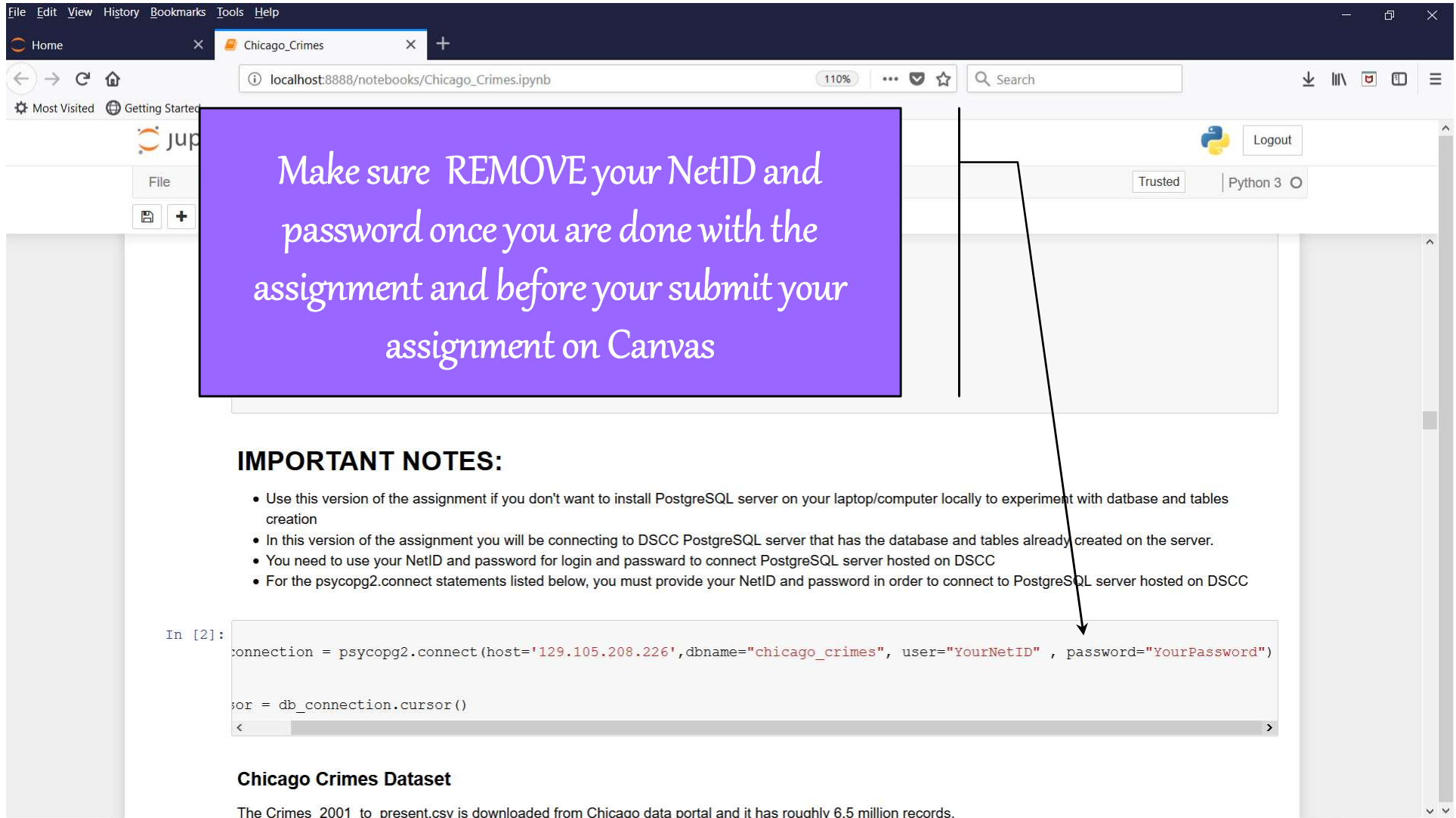
```
connection = psycopg2.connect(host='129.105.208.226', dbname="chicago_crimes", user="YourNetID", password="YourPassword")

cursor = db_connection.cursor()
```

**Chicago Crimes Dataset**

The Crimes\_2001\_to\_present.csv is downloaded from Chicago data portal and it has roughly 6.5 million records.

# Assignment #3 - Walkthrough



The screenshot shows a Jupyter Notebook titled 'Chicago\_Crimes' running on 'localhost:8888/notebooks/Chicago\_Crimes.ipynb'. A purple callout box with white text is overlaid on the notebook, stating: 'Make sure REMOVE your NetID and password once you are done with the assignment and before your submit your assignment on Canvas'. An arrow points from this box to a code cell. The code cell contains the following Python code:

```
In [2]: connection = psycopg2.connect(host='129.105.208.226', dbname="chicago_crimes", user="YourNetID", password="YourPassword")
        cursor = db_connection.cursor()
```

Below the code cell, the text 'Chicago Crimes Dataset' is visible, followed by a line of text: 'The Crimes\_2001\_to\_present.csv is downloaded from Chicago data portal and it has roughly 6.5 million records.'

# Assignment #3 - Deliverables

---

- Submit the 2 documents listed below on Canvas:
  1. Submit on Canvas your ipynb script for Assignment #3 that has the source code and output
  2. Submit on Canvas the PDF document for your ipynb script for Assignment #3 that has the source code and output as a single PDF document