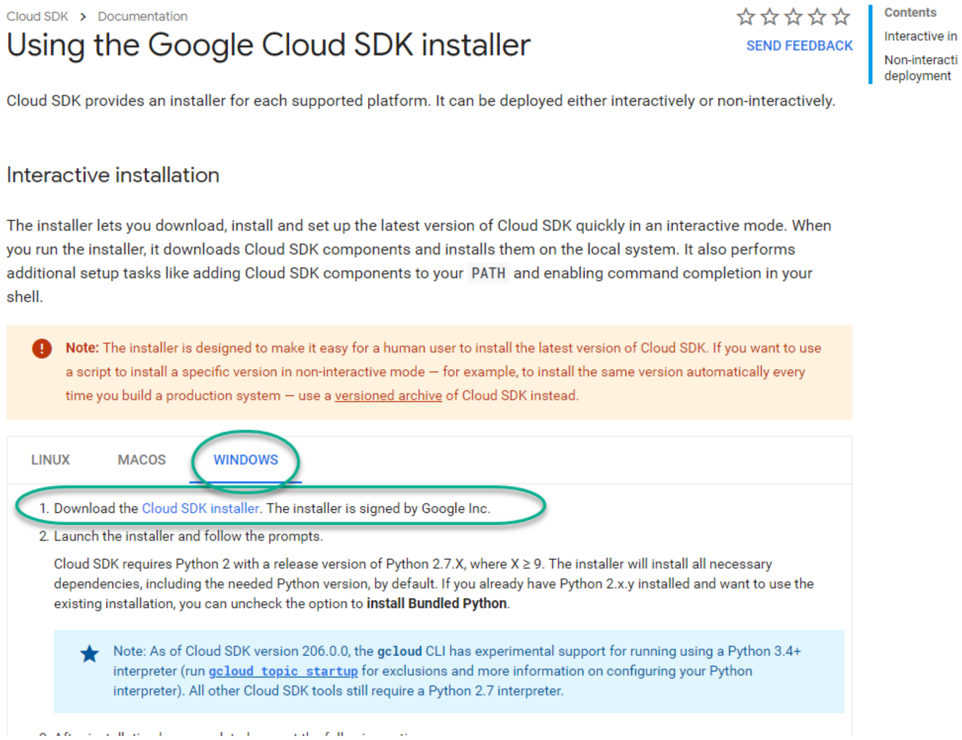
Connecting iPython Notebook to query GCP

- Ensure that you are logged into GCP with your Macy’s account credentials
  - https://console.cloud.google.com
  - More login info here: https://confluence.federated.fds/pages/viewpage.action?spaceKey=DaaS&title=Using+Google+Web+GUI+to+Query+DaaS+Data

- Download the Google Cloud SDK
  - https://cloud.google.com/sdk/docs/downloads-interactive



Using the Google Cloud SDK installer

Cloud SDK provides an installer for each supported platform. It can be deployed either interactively or non-interactively.

Interactive installation

The installer lets you download, install and set up the latest version of Cloud SDK quickly in an interactive mode. When you run the installer, it downloads Cloud SDK components and installs them on the local system. It also performs additional setup tasks like adding Cloud SDK components to your `PATH` and enabling command completion in your shell.

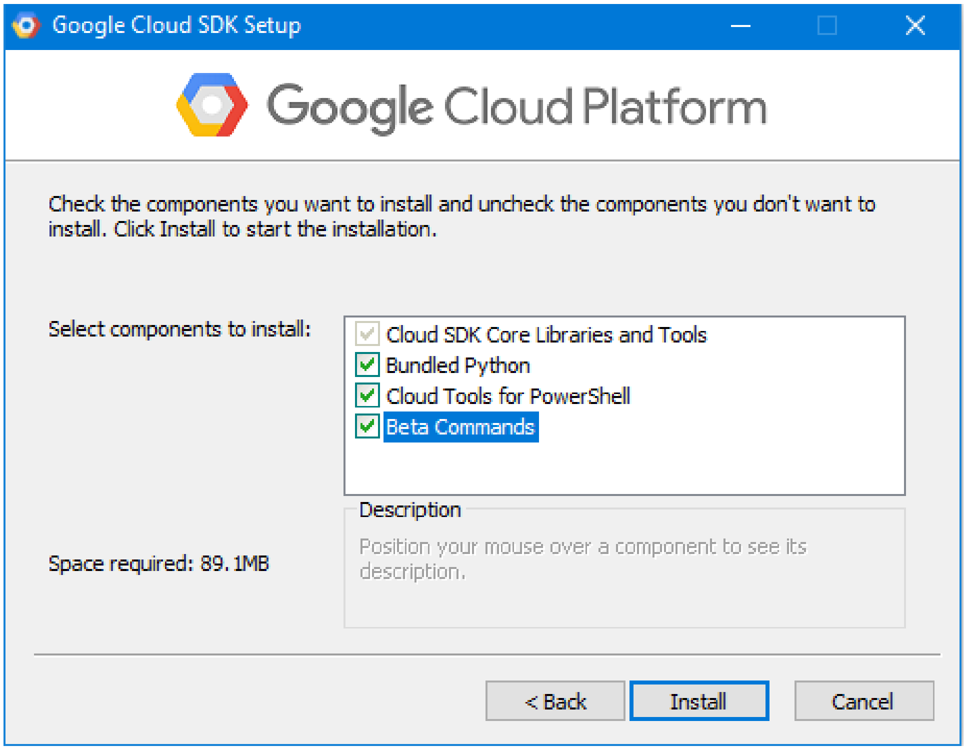
Note: The installer is designed to make it easy for a human user to install the latest version of Cloud SDK. If you want to use a script to install a specific version in non-interactive mode — for example, to install the same version automatically every time you build a production system — use a versioned archive of Cloud SDK instead.

1. Download the Cloud SDK installer. The installer is signed by Google Inc.
2. Launch the installer and follow the prompts.

   Cloud SDK requires Python 2 with a release version of Python 2.7.X, where X ≥ 9. The installer will install all necessary dependencies, including the needed Python version, by default. If you already have Python 2.x.y installed and want to use the existing installation, you can uncheck the option to **install Bundled Python**.

   Note: As of Cloud SDK version 206.0.0, the **`gcloud`** CLI has experimental support for running using a Python 3.4+ interpreter (run `gcloud topic startup` for exclusions and more information on configuring your Python interpreter). All other Cloud SDK tools still require a Python 2.7 interpreter.

  - After the file has downloaded, run the file to install on your computer
  - Install for Single User
  - Select destination folder to install; defualt select should be fine.
  - Select ALL components to install; click install



Google Cloud SDK Setup

Google Cloud Platform

Check the components you want to install and uncheck the components you don't want to install. Click Install to start the installation.

Select components to install:

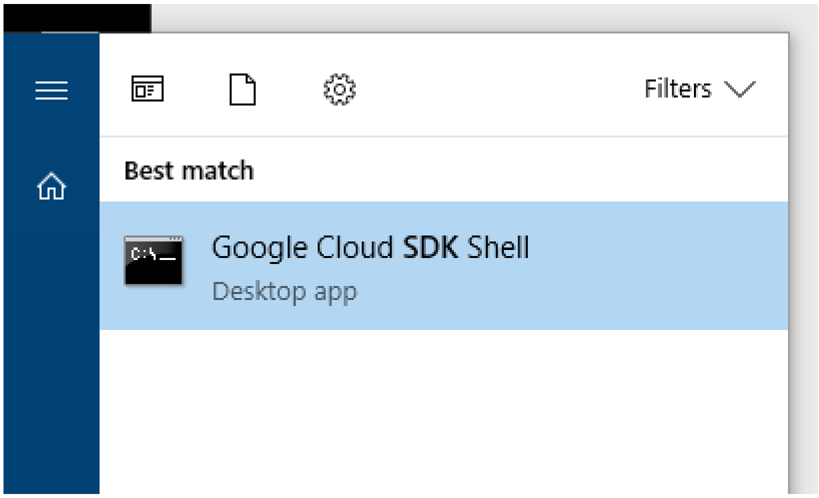
☑ Cloud SDK Core Libraries and Tools
☑ Bundled Python
☑ Cloud Tools for PowerShell
☑ Beta Commands

Description

Position your mouse over a component to see its description.

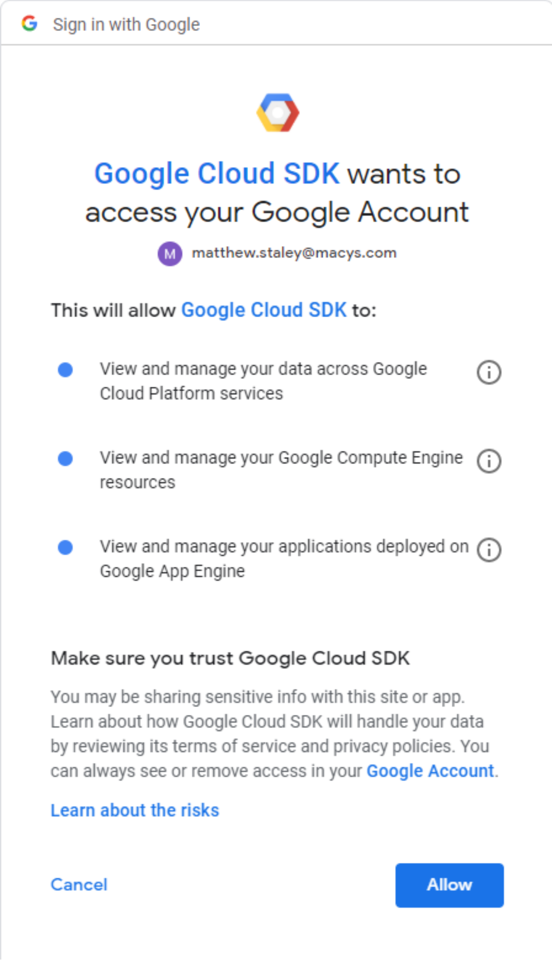
Space required: 89.1MB

- Open the SDK terminal shell
  - Pin to your taskbar for quick access in the future



- You will be prompted through a quick setup.
  - Enter in the selection for **macys-product-queries** as your default library when prompted.
  - You will be prompted to authenticate the SDK; make sure you authenticate to your Macy’s account. You may need to copy and paste the URL from this page into an

incognito/private web browser if you are defaulted to a personal account.



Sign in with Google

Google Cloud SDK wants to access your Google Account

matthew.staley@macys.com

This will allow Google Cloud SDK to:

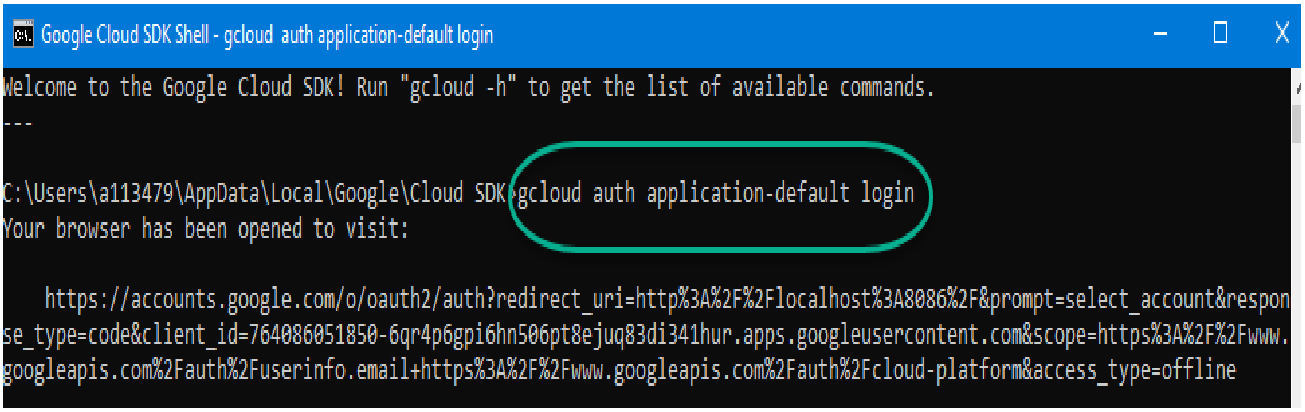
- View and manage your data across Google Cloud Platform services
- View and manage your Google Compute Engine resources
- View and manage your applications deployed on Google App Engine

Make sure you trust Google Cloud SDK

You may be sharing sensitive info with this site or app. Learn about how Google Cloud SDK will handle your data by reviewing its terms of service and privacy policies. You can always see or remove access in your Google Account.

Once the SDK is installed and authenticated, you need to authorize the Google Auth Library. This allows Google to "know" that your computer is authenticated when you are logged in to your Macy's account.

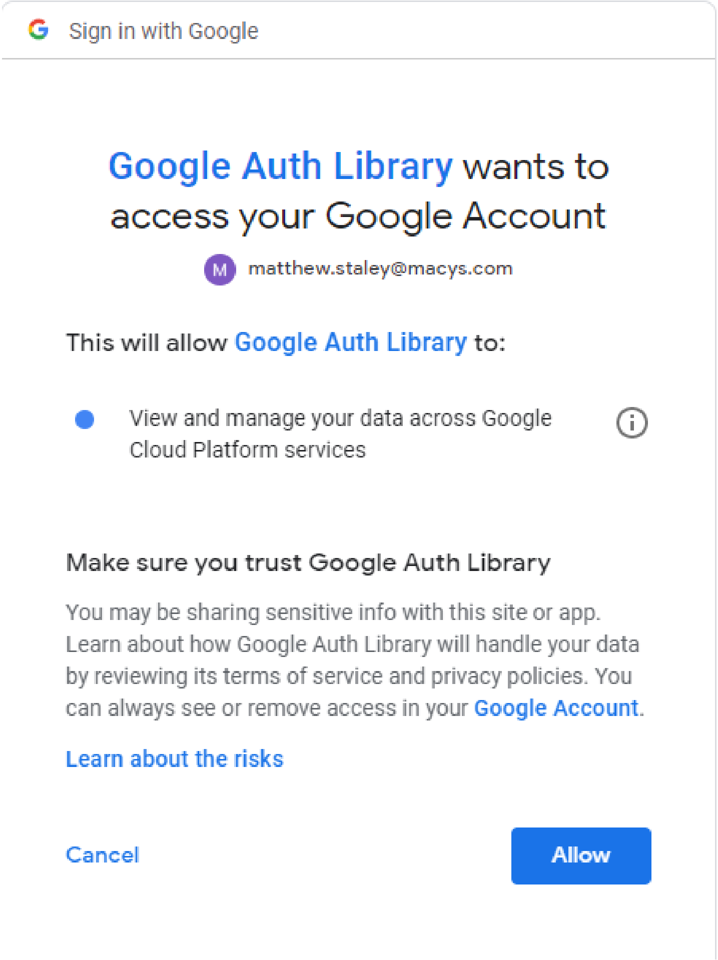
- From SDK terminal run this command:
  - gcloud auth application-default login



Google Cloud SDK Shell - gcloud auth application-default login

```
Welcome to the Google Cloud SDK! Run "gcloud -h" to get the list of available commands.
---

C:\Users\a113479\AppData\Local\Google\Cloud SDK>gcloud auth application-default login
Your browser has been opened to visit:

    https://accounts.google.com/o/oauth2/auth?redirect_uri=http%3A%2F%2Flocalhost%3A8086%2F&prompt=select_account&respon
se_type=code&client_id=764086051850-6qr4p6gpi6hn506pt8ejuq83di341hur.apps.googleusercontent.com&scope=https%3A%2F%2Fwww.
googleapis.com%2Fauth%2Fuserinfo.email+https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fcloud-platform&access_type=offline
```

        - You will be taken to this web url to allow access. If you Macy's account is not showing up, you can copy and paste the URL into a private/incognito web browser.



Sign in with Google

Google Auth Library wants to access your Google Account

M matthew.staley@macys.com

This will allow Google Auth Library to:

- View and manage your data across Google Cloud Platform services

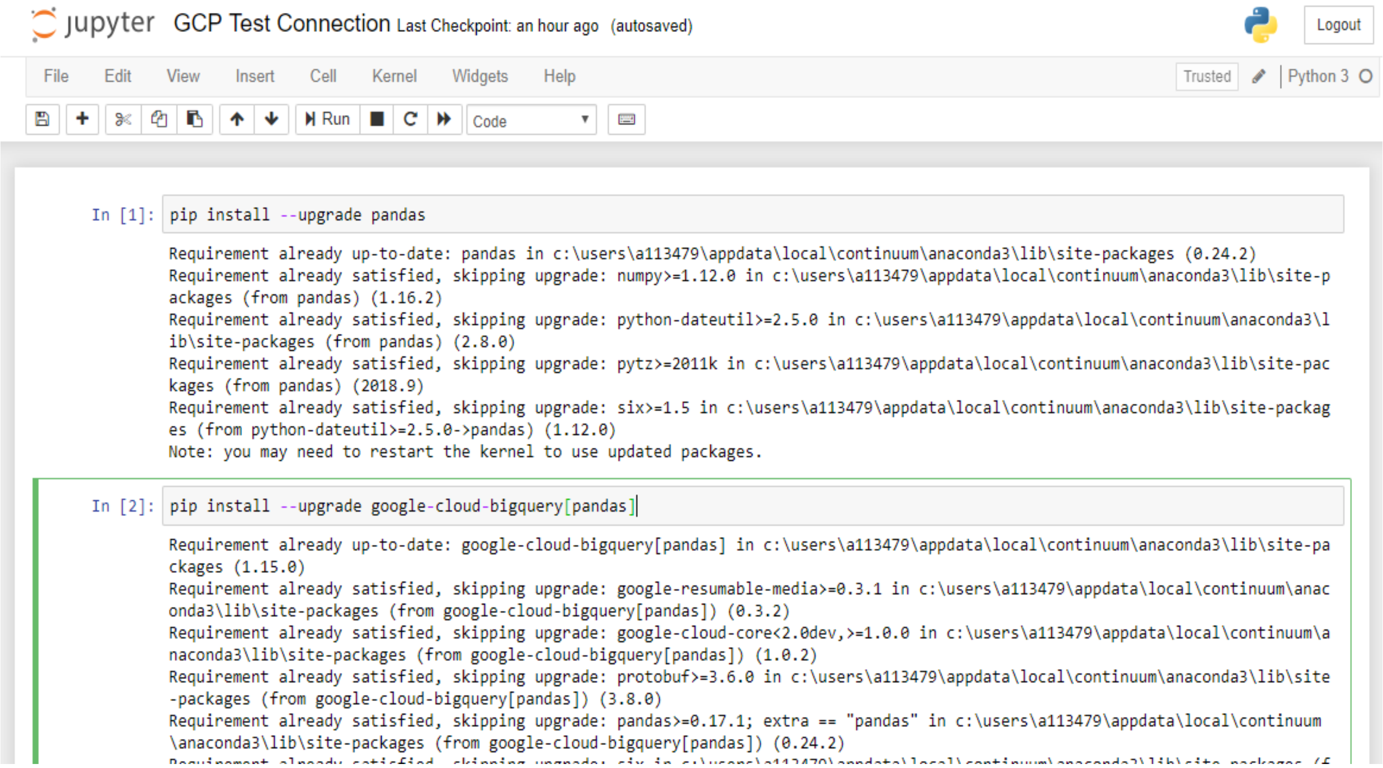
Make sure you trust Google Auth Library

You may be sharing sensitive info with this site or app. Learn about how Google Auth Library will handle your data by reviewing its terms of service and privacy policies. You can always see or remove access in your Google Account.

Learn about the risks

Cancel Allow

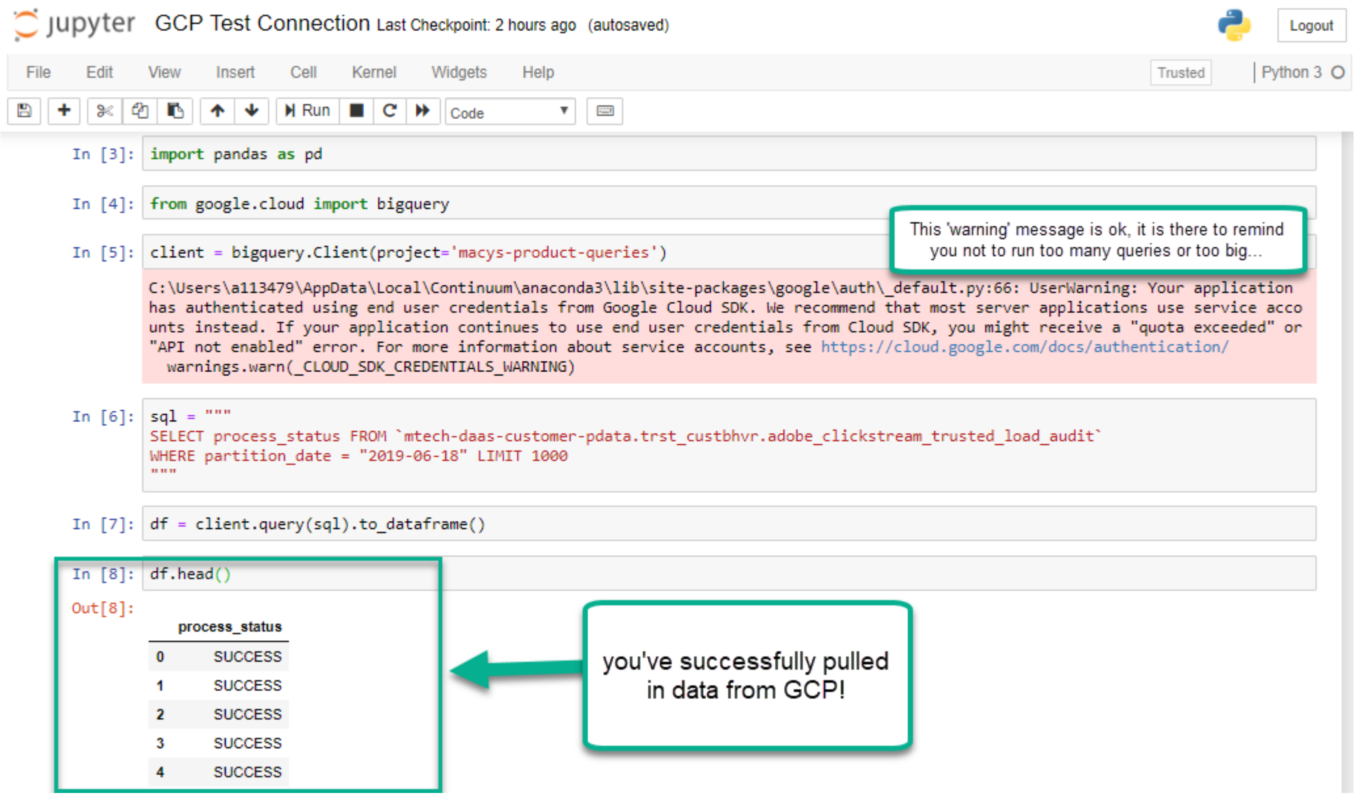
- Open Anaconda Navigator and Launch a new instance of iPython
    - Download Anaconda here: https://www.anaconda.com/distribution/#download-section
    - Run these lines separately to install updates
        - pip install --upgrade pandas
        - pip install --upgrade google-cloud-bigquery[pandas]
    - Restart the kernel and comment out or delete those lines of code, you only need to run them once, or whenever prompted to update
        - (hit the refresh button between the stop and fast forward button in iPython to restart kernal)



```
In [1]: pip install --upgrade pandas

        Requirement already up-to-date: pandas in c:\users\a113479\appdata\local\continuum\anaconda3\lib\site-packages (0.24.2)
        Requirement already satisfied, skipping upgrade: numpy>=1.12.0 in c:\users\a113479\appdata\local\continuum\anaconda3\lib\site-p
        ackages (from pandas) (1.16.2)
        Requirement already satisfied, skipping upgrade: python-dateutil>=2.5.0 in c:\users\a113479\appdata\local\continuum\anaconda3\l
        ib\site-packages (from pandas) (2.8.0)
        Requirement already satisfied, skipping upgrade: pytz>=2011k in c:\users\a113479\appdata\local\continuum\anaconda3\lib\site-pac
        kages (from pandas) (2018.9)
        Requirement already satisfied, skipping upgrade: six>=1.5 in c:\users\a113479\appdata\local\continuum\anaconda3\lib\site-packag
        es (from python-dateutil>=2.5.0->pandas) (1.12.0)
        Note: you may need to restart the kernel to use updated packages.

In [2]: pip install --upgrade google-cloud-bigquery[pandas]

        Requirement already up-to-date: google-cloud-bigquery[pandas] in c:\users\a113479\appdata\local\continuum\anaconda3\lib\site-pa
        ckages (1.15.0)
        Requirement already satisfied, skipping upgrade: google-resumable-media>=0.3.1 in c:\users\a113479\appdata\local\continuum\anac
        onda3\lib\site-packages (from google-cloud-bigquery[pandas]) (0.3.2)
        Requirement already satisfied, skipping upgrade: google-cloud-core<2.0dev,>=1.0.0 in c:\users\a113479\appdata\local\continuum\a
        naconda3\lib\site-packages (from google-cloud-bigquery[pandas]) (1.0.2)
        Requirement already satisfied, skipping upgrade: protobuf>=3.6.0 in c:\users\a113479\appdata\local\continuum\anaconda3\lib\site
        -packages (from google-cloud-bigquery[pandas]) (3.8.0)
        Requirement already satisfied, skipping upgrade: pandas>=0.17.1; extra == "pandas" in c:\users\a113479\appdata\local\continuum
        \anaconda3\lib\site-packages (from google-cloud-bigquery[pandas]) (0.24.2)
        Requirement already satisfied, skipping upgrade: six in c:\users\a113479\appdata\local\continuum\anaconda3\lib\site-packages (f
```

- Paste and run these lines of code one at a time to help with de-bugging

```
import pandas as pd

from google.cloud import bigquery

client = bigquery.Client(project='macys-product-queries')

sql = ""”SELECT process_status FROM `mtech-daas-customer-pdata.trst_custbhvr.adobe_clickstream_trusted_load_audit` WHERE partition_date = "2019-06-18" LIMIT 1000"""

df = client.query(sql).to_dataframe()

df.head()
```



```
In [3]: import pandas as pd

In [4]: from google.cloud import bigquery

In [5]: client = bigquery.Client(project='macys-product-queries')

C:\Users\a113479\AppData\Local\Continuum\anaconda3\lib\site-packages\google\auth\_default.py:66: UserWarning: Your application
has authenticated using end user credentials from Google Cloud SDK. We recommend that most server applications use service acco
unts instead. If your application continues to use end user credentials from Cloud SDK, you might receive a "quota exceeded" or
"API not enabled" error. For more information about service accounts, see https://cloud.google.com/docs/authentication/
  warnings.warn(_CLOUD_SDK_CREDENTIALS_WARNING)

In [6]: sql = """
        SELECT process_status FROM `mtech-daas-customer-pdata.trst_custbhvr.adobe_clickstream_trusted_load_audit`
        WHERE partition_date = "2019-06-18" LIMIT 1000
        """

In [7]: df = client.query(sql).to_dataframe()

In [8]: df.head()

Out[8]:
     process_status
  0  SUCCESS
  1  SUCCESS
  2  SUCCESS
  3  SUCCESS
  4  SUCCESS
```

This 'warning' message is ok, it is there to remind you not to run too many queries or too big...

you've successfully pulled in data from GCP!

More Resources

- **Pinned location to work on query**
  - https://console.cloud.google.com/bigquery?p=mtech-daas-customer-pdata&project=macys-product-queries&organizationId=397937668872
- **More documentation for connecting to GCP with Google/R**
  - https://googleapis.github.io/google-cloud-python/latest/bigquery/usage/pandas.html
  - https://cloud.google.com/blog/products/gcp/google-cloud-platform-for-data-scientists-using-r-with-google-bigquery
- **SAS may use ODBC "Simba Drivers", which aren't great but work**
  - https://communities.sas.com/t5/SAS-Data-Management/Simba-Drivers-for-Google-s-BigQuery/td-p/520989