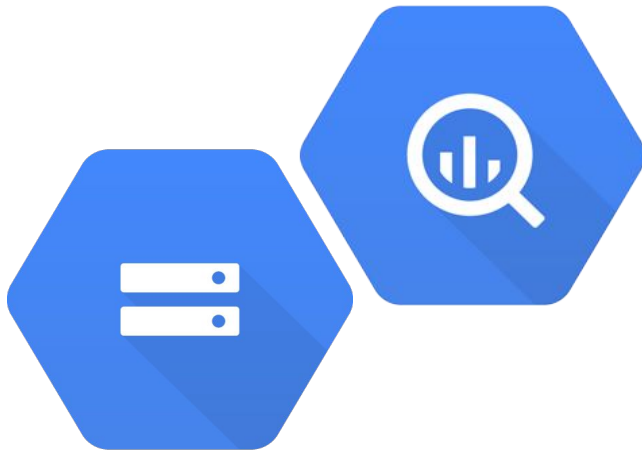




Google Cloud Public Datasets:

What they are, and how to use them

Google Cloud



Helping you tell your story



Google Cloud

Cloud Public Datasets Lead



Shane Glass

- ✓ Previously managed NOAA's Big Data Project
- ✓ Data visualization enthusiast
- ✓ Public datasets user

Maximize the **accessibility** and **usability** of high-value, high-demand public data assets through Google Cloud Platform

Mission Statement

Public Datasets Program

Current catalog >130 datasets

Onboarded and maintained by Googler(s)
with data provider input/guidance

g.co/cloud/marketplace-datasets

Data providers:

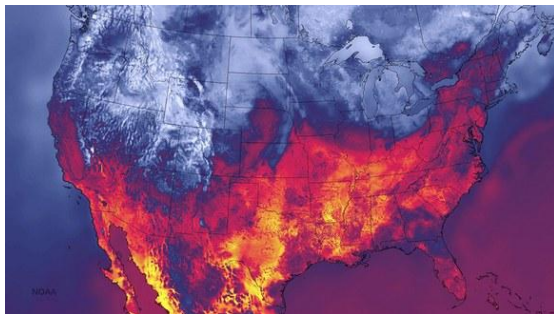


Unstructured datasets in Google Cloud Storage



Sentinel-2

NEW



High-Resolution Rapid
Refresh Model (HRRR)

NEW



National Water Model



The "Data Science" method

You need to **discover** the dataset and where to access it.

You need to **access** your data in a bulk, machine-readable format.

You need to **write** a program to parse the data into your analysis tool.

You need to **load** the data into a database

You need to **update, maintain, and secure** your data and database.

You need to **manage access** to data.

You need to **update** the data regularly.

You need to **link** public data with private data.

You need to **analyze** it.

You need to **visualize and communicate** your results.

What if someone else did this...

Discover the dataset and where to access it.

Access data in a bulk, machine-readable format.

Write a program to parse the data into your analysis tool.

Load the data into a database

Manage, maintain, and secure a database (BigQuery).

Manage access to data.

Update the data regularly.

Help you **link** public data with private data.

You need to analyze your data.

You need to visualize and communicate your results.

...and you do this:

You need to discover the dataset and where to access it.

You need to access your data in a bulk, machine-readable format.

You need to load the data into a database by writing a program so it can be analyzed, or write a program to analyze it.

You need to manage, update, maintain, and secure your database.

You need to manage access to data.

You need to update the data regularly.

You need to link public data with private data. Repeat this whole process.

You **analyze** the data.

You **communicate, visualize, and implement** the results and insights.

To the demos!



Thank you!

@shanecglass

Demo code and queries from the session

BigQuery Public Datasets Demos:

- [Weather: Urban areas with the most hurricanes](#)
- [Weather: Cities with the most hurricanes](#)
- [Transportation: NYC taxi tips by NYC zones](#)
- [Transportation: NYC subway riders by borough](#)
- [Transportation: NYC subway stops per borough](#)
- [Tornadoes: Where tornadoes crossed roads in Oklahoma between 2007 and 2017?](#)

Google Cloud Storage Public Datasets Demos:

- [Analyzing the 2018 Kīlauea volcanic eruption using GOES-16](#)