Voracious Vikings

Team 3:
Elizabeth City State University

[Team Song: Apples and bananas]

HackHPC@ADMI22
3/31-4/3

Mentors:

Cole Mcknight

Robert Barrett

Hackers:

Tony Guy

Cameron Pendergrass

Marquis Brown-Elahi

Daniel Lamb

Main Objective

To use HPC to analyze the Nhanes dataset and generate graphs that show how population health has changed over time for various races.

NHANES 2017–2018 Examination Data

- NHANES 2017-2018 Examination Variable List
- Exam Procedure Manuals
- 2017-2018 Examination Data Overview
- SAS Universal Viewer

Data File Name	Doc File	Data File	Date Published
Audiometry	AUX_J Doc	AUX_J Data [XPT - 2.2 MB]	November 2021
Audiometry - Acoustic Reflex	AUXAR_J Doc	AUXAR_J Data [XPT - 88.5 MB]	November 2021
Audiometry - Tympanometry	AUXTYM_J Doc	AUXTYM_J Data [XPT - 22 MB]	November 2021
Audiometry - Wideband Reflectance	AUXWBR_J Doc	AUXWBR_J Data [XPT - 14 MB]	November 2021
Blood Pressure	BPX_J Doc	BPX_J Data [XPT - 1.4 MB]	February 2020
Blood Pressure - Oscillometric Measurements	BPXO_J Doc	BPXO_J Data [XPT - 678.1 KB]	April 2021
Body Measures	BMX_J Doc	BMX_J Data [XPT - 1.4 MB]	February 2020
Dual-Energy X-ray Absorptiometry - Android/Gynoid Measurements	DXXAG_J Doc	DXXAG_J Data [XPT - 922.9 KB]	October 2021

What is NHANES?

The National Health and Nutrition Examination Survey (NHANES) is a research program under the National Center for Health Statistics (NCHS) that assesses and tracks the health and nutritional status of adults and children in the United States. Interviews, physical examinations, and laboratory tests are all included in these surveys.

Tools used to accomplish our goal

- Learning and using CloudyCluster to analyze the Nhanes dataset by creating multiple job scripts to download data and generate figures.
  - Used Open-on-Demand to access and develop on our cluster

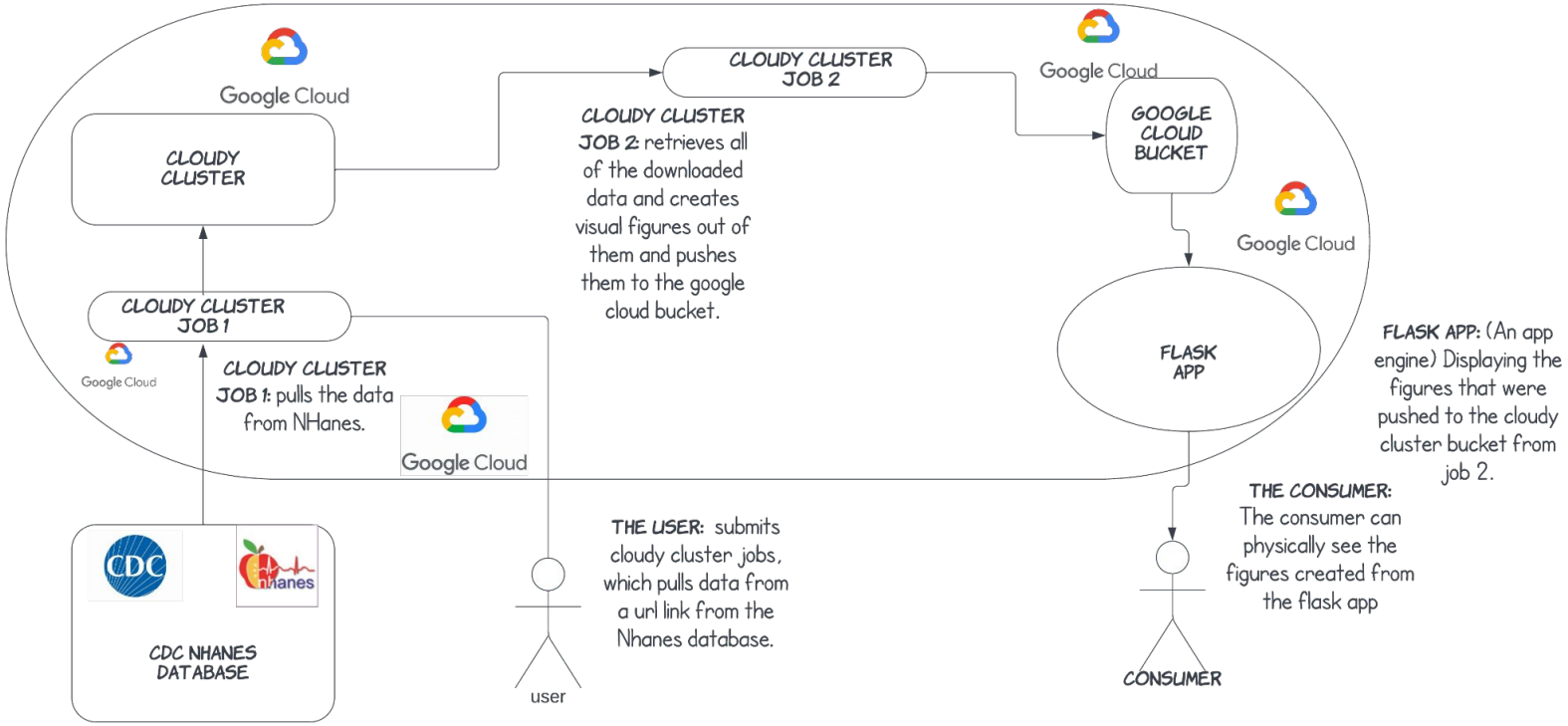
- Learning and using Google Colab and Jupyter Notebook as a development environment.
  - We used Python for the data ingestion, manipulation, and visualization
  - We used Pandas dataframe to organize the data
  - We used Seaborn to create our graphs and then create the figures

- Learning and using Google App Engine to make and host our Flask app

- Learning and using Flask to host our website content
  - HTML/CSS was used to build our frontend which host our figures

- Using github to keep all of our files organized and in a central location

Workflow diagram



Live demo

- CloudyCluster creating the job scripts - Marquis
- Google Colab and Jupyter Notebook creating figures - Cameron
- Flask app/Google Build - Daniel
- Web page review - Cameron

Problems and Solutions

- Flask app not reading the HTML properly
  - When trying to implement a button pop up feature
  - We believe the issue was AJAX not working with Google Engine

- CloudyCluster was not able to properly display our figures when running the python script.
  - We were not able to find a solution to cloudy cluster so we decided to keep the script inside of Google Colab and mount the images into our Google Cloud Storage Bucket. All of the data generation was handled with CloudyCluster

Next Steps

- Making the webpage be interactive
- Utilizing more NHanes Datasets
  - Related to Social History: Drug & Alcohol use
- Comparing health by income groups

SGCI Science Gateways Community Institute

Thank you to all of the mentors and a huge thank you to Robert Barrett and Cole McKnight

Thank you to all of the support organizations and sponsors!

Thank you to ADMI for hosting this amazing opportunity!

CloudyCluster™

Google Cloud

Omnibond™
ORANGEFS · CLOUDYCLUSTER