Riley Rutan

RedID: 131091165

GEOG 582 GIS Programming with Python

November 18, 2024

**Final Project Proposal: Python Pipeline to Update 4DVD Database**

**Introduction:** 4DVD is a unique software developed at the Climate Informatics Lab, San Diego State University, for the instant delivery of big climate data to classrooms and households around the world in a convenient and visual way. It is a web application that visualizes historical climate data that is stored in a relational database on an SDSU server. Unfortunately, there are currently gaps in the data that it has stored to visualize, and updating the database is no simple task.

**Problem Statement:** The legacy software that was designed to update the 4dvd database with new data is no longer usable, as it was built with open-source packages that have since changed. I will design a Python package that will allow a user to collect big climate data from NOAA’s cloud storage, process and convert the netCDF data into the appropriate CSV files, and insert this new data into the SQL database with INSERT statements. I want to create code that can be used in the future so students can easily maintain this technology. I will make it a priority to avoid downloading large datasets to the user’s local system, and try and keep files temporarily in cloud storage.

**Data Sources:**

* NOAA ERSST v5 <https://psl.noaa.gov/data/gridded/data.noaa.ersst.v5.html>
* NOAA Gridded Precipitation https://psl.noaa.gov/data/gridded/data.prec.html

**Python Modules:** Numpy, pandas, netcdf4 (for working with gridded data), dask (for lazily loading large datasets), gdown (for google drive connection)

**References: https://4dvd.sdsu.edu/**