CO_2 Around the World

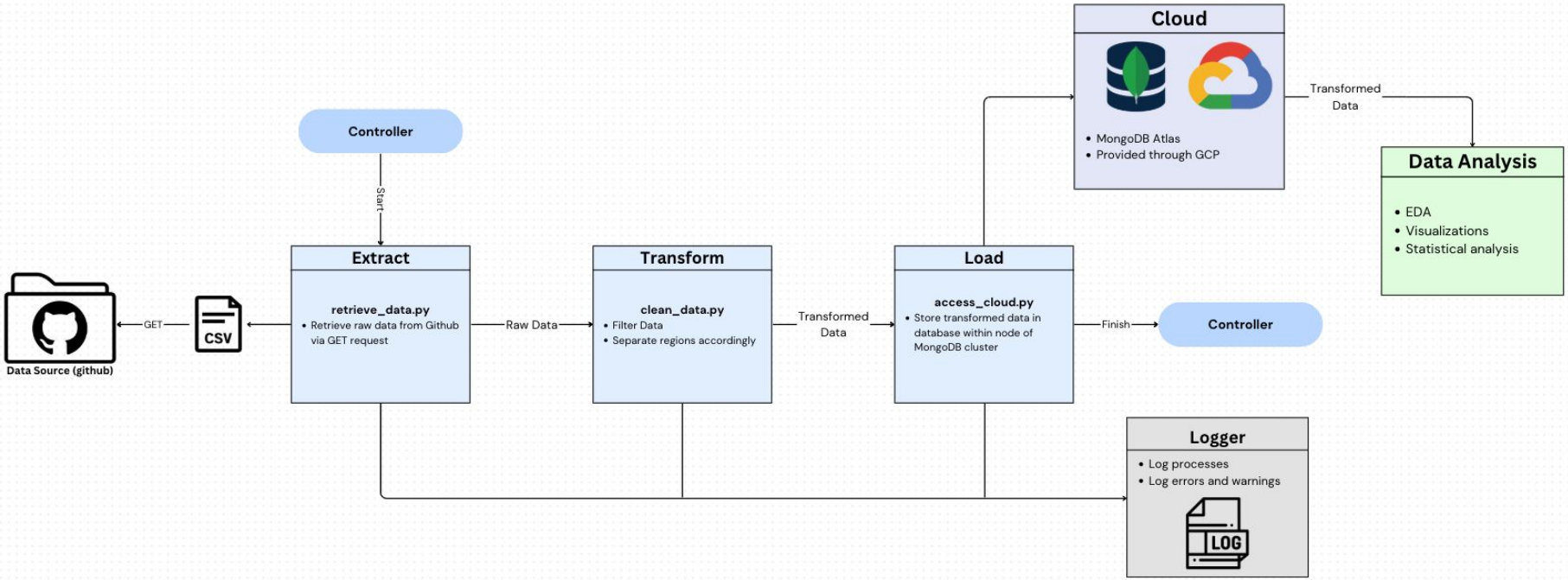
DS 2002 Final Project

Benjamin Yeh, Jaelyn Do, Jonah Lee, Kevin Lam

Data Selection

- Initial struggle to find datasets (size, depth, usability)
- The dataset we decided on is a collection of key metrics maintained by *Our World in Data*
  - built upon a number of datasets that include GDP and global carbon emissions
- There is a dataset for each of the following countries (including territories), continents, and socioeconomic (income levels)
- Our focus was on the relationship between CO2 emissions, GDP, and population

ETL Pipeline



Cloud Storage

- Opted for MongoDB Atlas (NoSQL) over SQL
- Resolved issues with database user permissions and URI configuration to establish a successful connection.
- Implemented environment variables to securely manage database credentials, avoiding exposure in the code repository.

Analysis

- Looking at correlation between CO2 and GDP and CO2 and population
- Strong, positive correlation for both relationships
- Performed bootstrapping to find confidence interval for R squared values
- Narrow intervals with high values