

Part 2: (ETL Concept Design)

In designing the conceptual diagram (below) for the migration of data and metadata from Climate Watch (or some other source) to the CCDR database hosted on the World Bank, the process involves a systematic flow of information across various components. The source system, Climate Watch (CW) Data, serves as the repository for historical emissions data and metadata. Data can be fetched from this source either systematically through API calls or via manual downloads from the website. I have preferred the former method to fetch source data due to its efficiency and reliability.

Leveraging tools like **Google Cloud Storage**, the extracted data and metadata is then stored in a staging area rather than being loaded directly into the data warehouse. This approach ensures that if any issues arise during the ETL process, users can safely revert the data to its original state. Data extraction sets the stage for the subsequent transformation phase. Within this transformative stage, the data and metadata undergo a series of predefined rules and functions, such as cleaning, auditing, and formatting, facilitated by tools like **Google Cloud Functions**. Extracted metadata is further enhanced with additional contextual information from World Bank's metadata. The purpose of this step is to harmonize the extracted data into a single, unified format and store it in **Google BigQuery** for further processing. In the final step of ETL process, with the help of Google BigQuery, the transformed data is loaded into a single or multiple data warehouses facilitating the need for analysis, reporting, searchability and visualization of the data.

It is important to note that extracting/transforming/loading the metadata, is typically a one-time job as it doesn't change frequently. However, the frequency of running this ETL pipeline for raw data can be determined based on the update frequency of CW Data (for example annually) and the specific needs and reporting requirements of the World Bank.

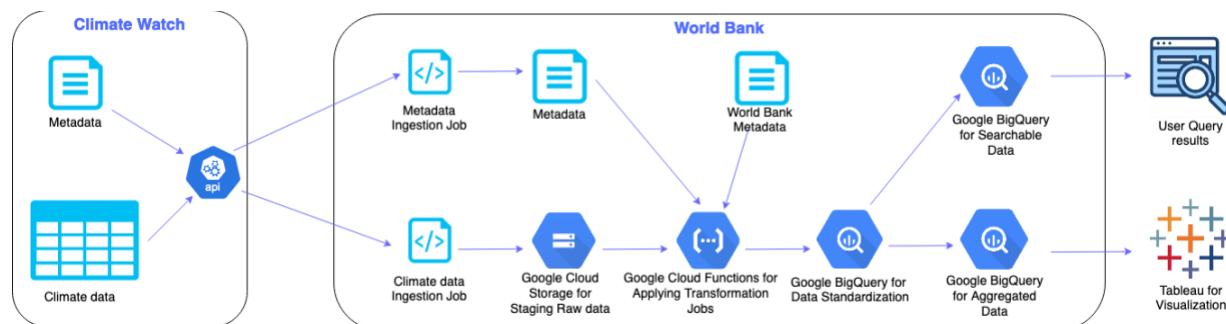


Figure 1: ETL concept diagram depicting data and metadata migration from Climate Watch to World Bank