A diagram of data storage

Description automatically generated

The diagram illustrates the workflow for migrating data from CLIMATEWATCH to the World Bank's DataBank CCDR database, a process which includes three stages: staging, standardization, and serving. Initially, raw data and metadata are extracted from CLIMATEWATCH using Azure Data Factory and the provided API calls. This raw data is then converted into the Delta Lake format for processing upon arrival at Data Lake Storage. It's then ingested into the Bronze layer via Apache Spark APIs in Azure Databricks.

Next, a series of tasks are performed to filter, clean, and transform the data into the Silver layer. This ensures the data aligns with the CCDR’s schema and is conducted through Azure Databricks. Finally, the Gold layer contains the aggregated and enriched data, prepared and ready for analytics and reporting. This assumes that the data in the CCDR is stored in the Azure Data Lake.

During the transformation, the Country column in CLIMATEWATCH is mapped to the Country Name column in the CCDR data table. Likewise, year columns in CLIMATEWATCH align with the year columns in CCDR data table. Data Source, Sector, Gas, and unit columns are mapped to the indicator columns in CCDR data table. The metadata from CLIMATEWATCH is also migrated to CCDR and transformed into a structure similar to the CCDR metadata table.