#### **1. Introduction**

This document provides a high-level design for the Data Engineering Portfolio, including three primary projects: ETL, Web Scraping and Data Extraction, and SQLite Database Operations.

#### **2. System Architecture**

The system architecture consists of three main components:

* **ETL (Extract, Transform, Load)**:
  + Extracts data from various file formats (CSV, JSON, XML).
  + Transforms data (e.g., unit conversion).
  + Loads data into a CSV file.
  + Logs each phase of the ETL process.
* **Web Scraping and Data Extraction**:
  + Scrapes data from a web page.
  + Parses HTML to extract specific data points.
  + Loads extracted data into a CSV file and an SQLite database.
* **SQLite Database Operations**:
  + Manages database operations (creation, insertion, querying).
  + Reads data from CSV files.
  + Loads data into SQLite tables.
  + Performs SQL queries and operations on the database.

**3. Data Flow Diagram**

Data Sources (CSV, JSON, XML) --> ETL Script --> Transformed CSV File

Web Page --> Web Scraping Script --> CSV File, SQLite Database

CSV Files --> SQLite Operations Script --> SQLite Database

#### **4. Technology Stack**

* **Programming Language**: Python
* **Libraries**: Pandas, Requests, BeautifulSoup, sqlite3
* **Database**: SQLite
* **File Formats**: CSV, JSON, XML
* **Logging**: Custom logging to a text file