Data Processing Calculator

User manual

Abstract

The Data Processing Calculator is a system that calculates measurable differences between ELT and ETL data pipelines. Specifically, this data processing calculator measures the performance of the PARADISE ELT data framework compared to the AWS Glue framework. The software includes a highly interactive user interface and displays historical data allowing for direct comparisons between both pipelines. This software includes utilities for uploading data, querying data, and displaying the resulting data in data visualization bar charts.

This document contains the manual of how to use the Data Processing Calculator. The rest of the document is ordered as follows: Chapter 1 describes that overall program and system requirements. Chapter 2 contains explanations on how to upload and query data. Chapter 3 includes instructions for displaying the data visualization charts from the resulting query.

This document describes functionality of the Data Processing Calculator v. 1.2, dated 2 May 2022.

Table of Contents

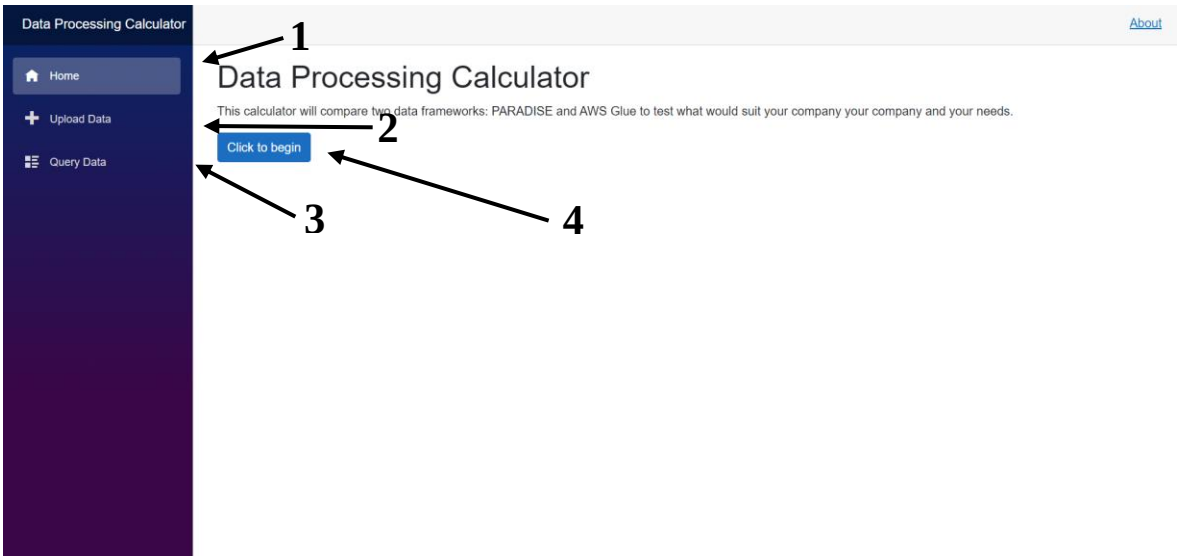
1 System overview 1
1.1 Program interface 1
1.2 System requirements 1
2 Data Management 3
2.1 Data types 3
2.1.1 File format 3
2.1.2 Json file structure 3
2.2 Start Up 4
2.2.1 Main Menu 4
3 Upload Data 5
3.1 Choosing JSON Files 5
3.2 Uploading JSON Files 6
3.3 Upload Data 7
4 Query Data 8
4.1 Query Page 8
4.2 Query Year Input 9
4.3 Successful Query 10
4.4 Unsuccessful Query 11

1 SYSTEM OVERVIEW

1.1 PROGRAM INTERFACE

The Data Processing Calculator includes an interactive user interface as displayed in Figure 1.1. All system functionality is accessible through this main interface.

Figure 1.1 Program interface



1. Home Tab
2. Upload Data Tab
3. Query Data Tab
4. Click to Begin Button

1.2 SYSTEM REQUIREMENTS

The Data Processing Calculator requires users to have a Windows10 operating system, as the application runs on a .NET Framework. In addition, the user must also have access to MongoDB and Amazon Web Services. User will be required to access and login to Amazon Web Services using the credentials listed below. To gain access to the Data Processing Calculator, the user must first clone the following repository, hosted on GitHub, to their computer: https://github.com/rpatel121998/DataCalculator1.git. Following the repository clone, the user must have the listed NuGet Packages installed before running the Data Processing Calculator program.

Amazon Web Service Credentials

https://454515613784.signin.aws.amazon.com/console

Account ID: 454515613784

IAM user name: DataCalculatorAccess

Password: Data-calculator

Required NuGet Packages:

- AWSSDK.S3
- Microsoft.EntityFrameworkCore.Tools
- Radzen.Blazor
- Syncfusion.Blazor
- Syncfucsion.Blazor.SplitButtons

2 DATA MANAGEMENT

2.1 DATA TYPES

The Data Processing Calculator accepts data as a JSON files. JSON files have to conform to specifications of acceptable file format and acceptable file structure.

2.1.1 FILE FORMAT

Acceptable upload file formats are JSON (.json).

2.1.2 JSON FILE STRUCTURE

JSON files define only two data structures: objects and array. A JSON object is an entity in JSON which is enclosed in curly brackets. It is written in the unordered set of name-value pairs in which the name should be followed by ":" (colon), and the name-value pairs need to separated using "," (comma). The first pair describes the "Id" and the second pair describes the "CWE". These pairs will be used to populate the data visualization charts following the query of the data.

Figure 2.1 displays a simplified example of JSON file structure. The Data Processing Calculator data parser will interpret the JSON files as follows:

1. CWE Year - CWE Category Number
2. Top CWE for that Year

Figure 2.1. Example of sample file structure


```
 "cve_items": [
   {
     "id": "CVE-2012-0001",
     "cwe": "CWE-94"
   },
   {
     "id": "CVE-2012-0002",
     "cwe": "CWE-264"
   },
   {
     "id": "CVE-2012-0003",
     "cwe": "CWE-399"
   },
   {
     "id": "CVE-2012-0004",
     "cwe": "CWE-79"
   },
   {
     "id": "CVE-2012-0006",
     "cwe": "CWE-94"
   },

   {
     "id": "CVE-2012-0013",
     "cwe": "CWE-94"
   }
 ]
}
```

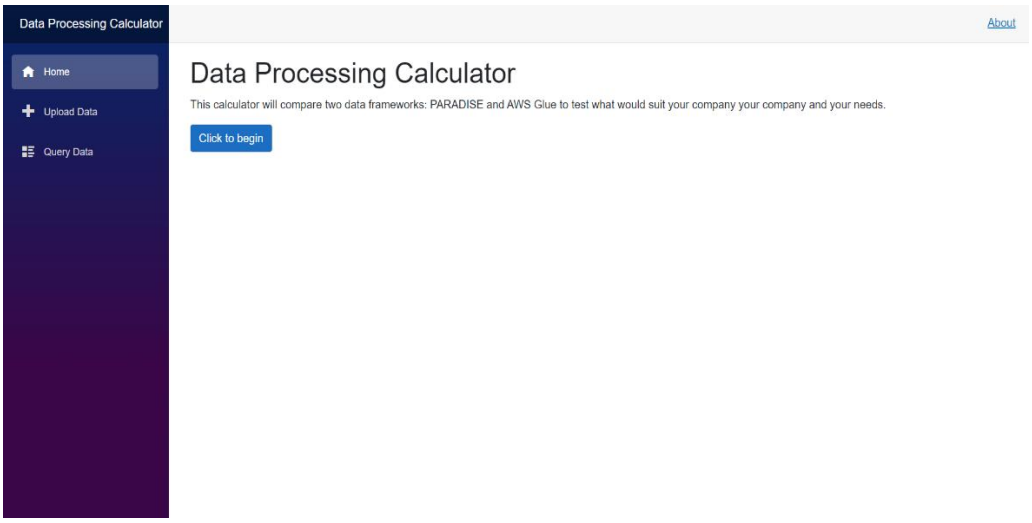
2.2 START UP

2.2.1 MAIN MENU

Upon running the program, the system will open to the main menu page.

Running the program will generate the window illustrated in Figure 2.2.

Figure 2.2. Main Menu



The main menu page provides the user with a small description of what the Data Processing Calculator will do and displays the “Click to Begin” button. This button will be used to navigate to the “Upload Data” screen once the user is ready to begin.

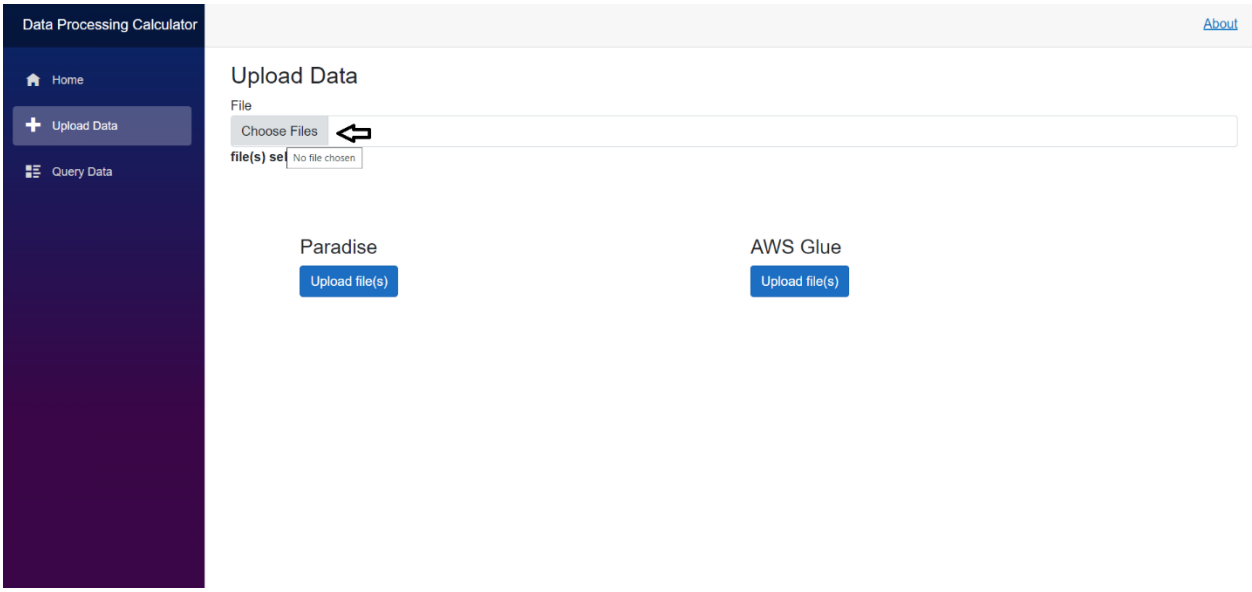
3 UPLOAD DATA

3.1 CHOOSING JSON FILES

To begin the comparison for both the PARADISE and AWS Glue framework, it requires that some data files first be selected to upload onto both frameworks. The Data Processing Calculator gives the user a “Choose Files” button that will allow the user to select and upload multiple JSON files from their computer.

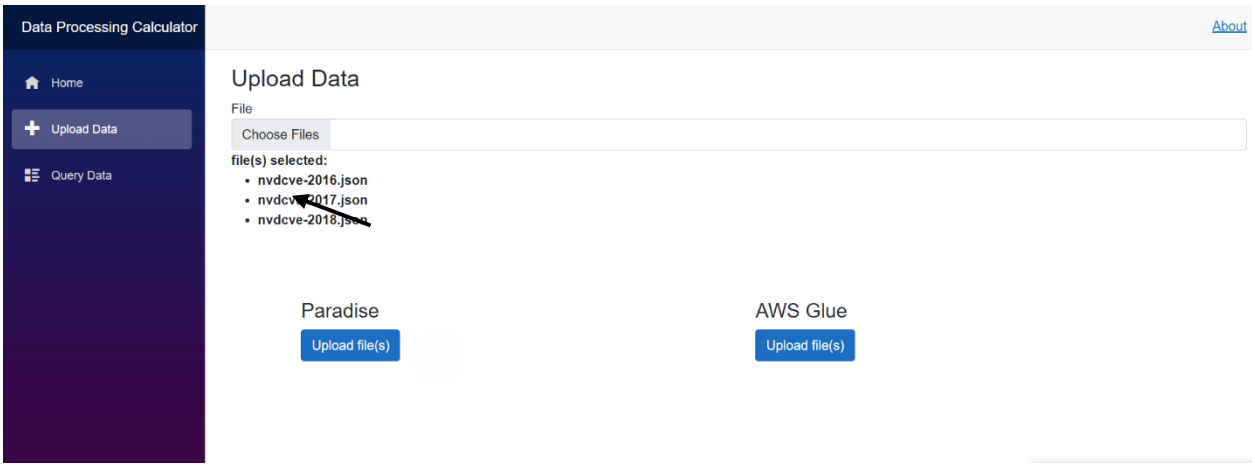
Clicking on the Click to Begin button will generate the window illustrated in Figure 3.1

Figure 3.1. Upload Data page displays, showing choose files button



After selecting one or more files to upload, The Data Processing Calculator gives the user a list of all files that were selected. You can find these files listed below the “file(s) selected” section provided under the “Choose Files” button. The listed files will appear as shown in Figure 3.2.

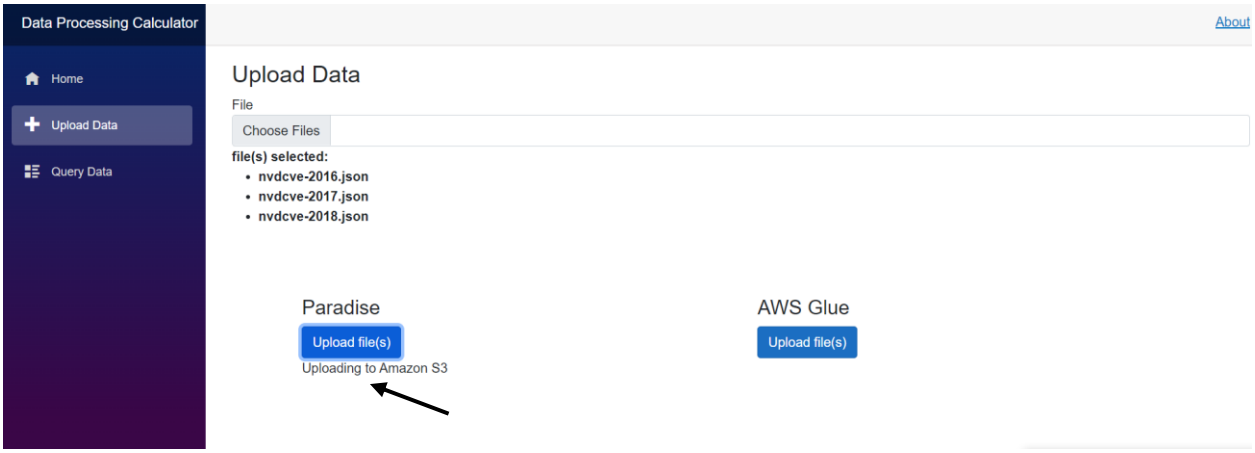
Figure 3.2. Files selected list



3.2 UPLOADING JSON FILES

The Data Processing Calculator gives the user an “Upload file(s)” button for both PARADISE and AWS Glue. These buttons will allow the user to upload JSON files from their computer onto each framework. Once the user clicks the upload button, the system will start a timer and the files will begin uploading to the respective framework. A message will display under each button confirming that the upload has begun as shown in Figure 3.3

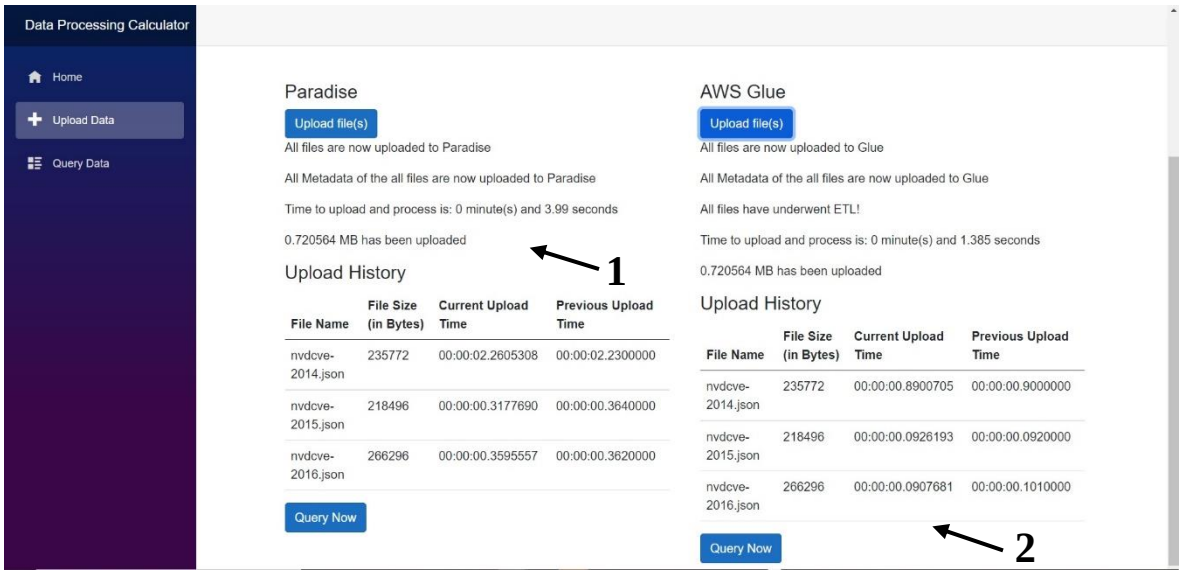
Figure 3.3. Uploading confirmation



3.3 UPLOAD DATA

Upon the completion of the file upload, the timer will stop, and the Data Processing Calculator will display a message confirming that all files have been uploaded. Following the confirmation of the upload, the following metadata will be displayed: upload time (minutes and seconds) and file size of the upload (Megabytes). These values will be recorded and stored upon the completion of each upload. These stored values will be used in the displaying of the historical data (upload history) as shown in Figure 3.4.

Figure 3.4. Upload Data



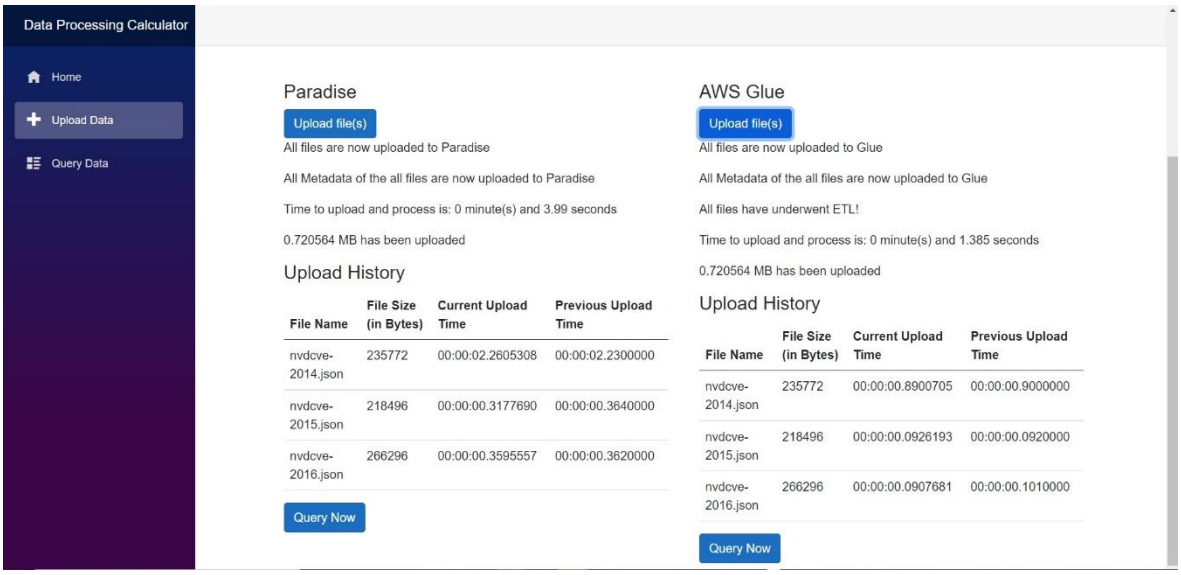
1. Upload Data
2. Upload History

4 QUERY DATA

4.1 QUERY PAGE

Upon the completion of the file upload, the user will click the “Query Now” button as shown in Figure 4.1

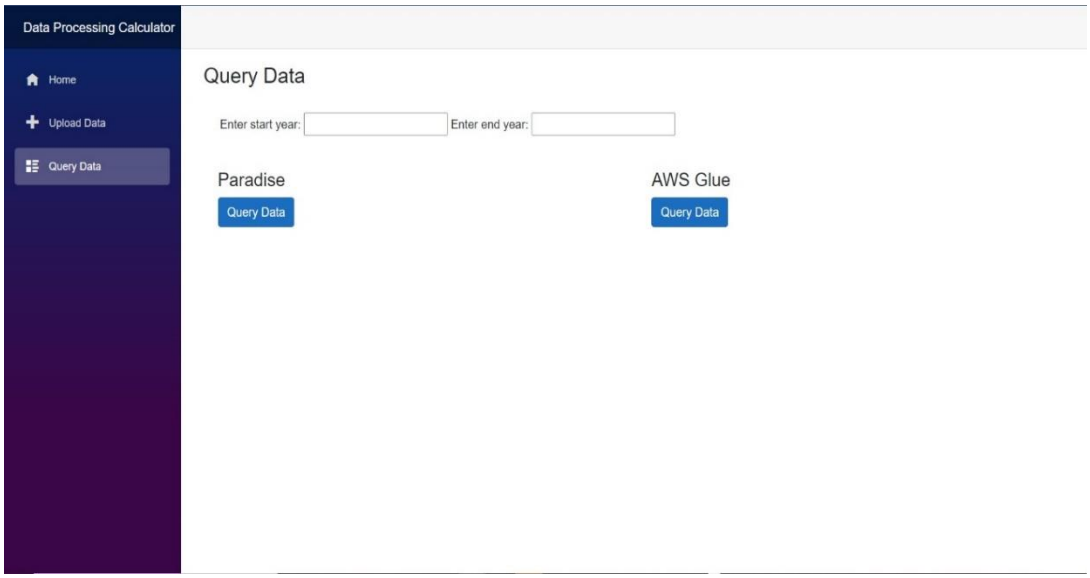
Figure 4.1. Query Now button



After clicking the “Query Now” button, the user will be redirected to the “Query Data” page. This page contains two text fields that will allow the user to enter the range of years they would like to query (all elicited from the ingest data). Directly below these text fields, there will be two “Query Data” buttons, one for the Paradise data pipeline and one for the AWS Glue pipeline.

Clicking on the Query Now button will generate the window illustrated in Figure 4.2.

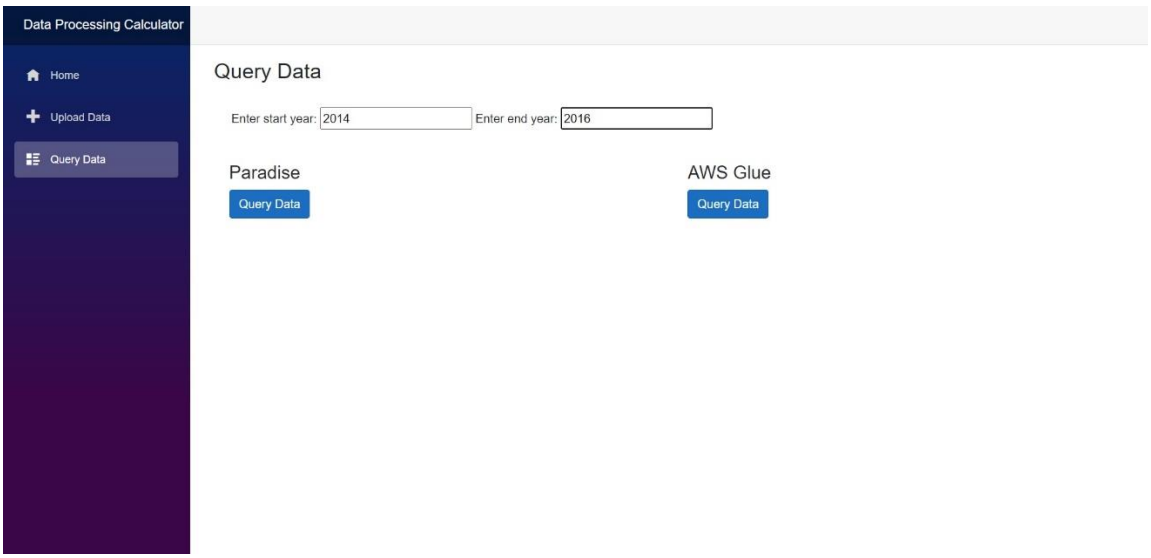
Figure 4.2. Query Data Page



4.2 QUERY YEAR INPUT

On the Query Page, the user must now enter a start year and end year for which they would like to query. After entering both a start year and end year, the user will click the "Query Data" button for either PARADISE or AWS Glue. This is illustrated in figure 4.3.

Figure 4.3. Start and End Year Input



4.3 SUCCESSFUL QUERY

Once the “Query Data” button is selected for either data pipeline, the system will start a timer and the query will be initiated. The results of the query will be displayed as bar graphs, with each graph representing the frequency of the number of occurrences of the top five vulnerabilities between the years input by the user (Figure 4.4). Immediately following the graphs, the user will be able to see the amount time it took to query along with the query history times (Figure 4.5).

Figure 4.4. Bar Charts Generated for Each Year

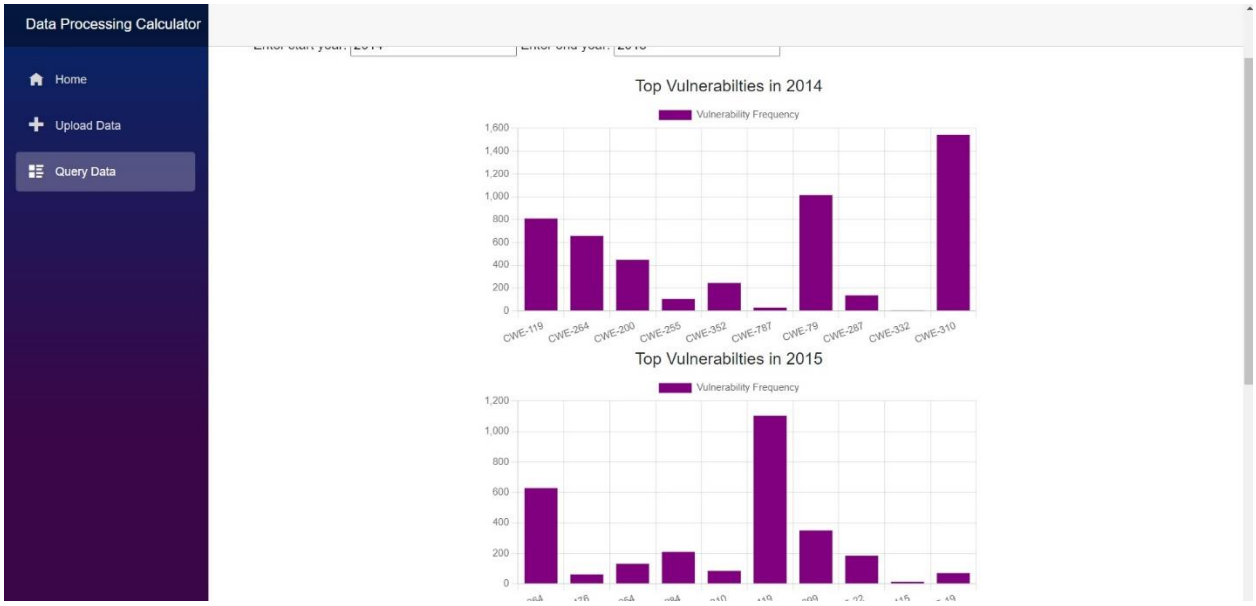
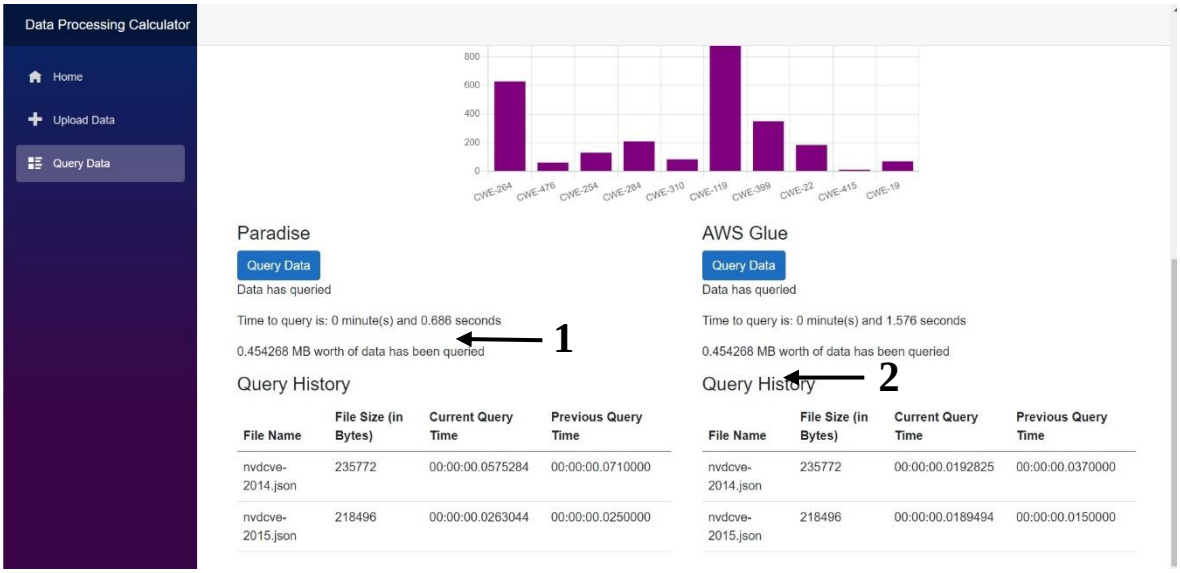


Figure 4.5. Query Time and Query History Displayed



1. Query Time
2. Query History

4.4 UNSUCCESSFUL QUERY

Once the “Query Data” button is selected for either data pipeline, the system will display a message that reads “Invalid year input. Please try again” (Figure 4.6). The user will then need to enter a valid start and end year and re-query the data.

Figure 4.6. Unsuccessful Query Error