**INTERNSHIP PROBLEM STATEMENT**

**Problem Title:**

Data files to be loaded to structured database and automate email triggers with success or failure messages.

**Problem description:**

Data maintained as files had to be loaded to structured database with a set of rules for data validation based on values in each column.

**Scenario :**

Create sample files (2 or more) files that have dependency on the main file(File1).

1. File1 can be a csv file that has a few columns with different data types and integer columns with sequential/unique values.
2. File2 has multiple columns with different data types and minimum of 10-15 columns with integer data type.
3. File2 integer columns value should fall within the range of integer column value in File1.
4. Design database/tables as needed with necessary constraints.
5. Automate email trigger based on data validations in input file.

**Technologies:**

* Python
* Postgresql

**Expected result:**

Trigger email based on data validation.

**STEPS TO SOLVE THE PROBLEM STATEMENT**

**Understanding the problem statement:**

The main goal is to load the multiple files into a structured database, to perform data validation based on predefined rules and trigger emails based on validation results.

Step 1:

Creating of Sample files

* Create two sample files, especially focusing on File1 as it serves as the main file.
* File2 integer column values should fall within range of integer column values of File1.

Step 2:

Database Design

* Create a table with the help of PostgreSQL to store the data from File1and File2.
* Establish a relationship between File1 and File2 with the help of foreign key constraints.
* Implement validation rules using SQL constraints.

Step 3:

Data Loading

* Write Python scripts to read the data from the files and insert it into PostgreSQL table by using libraries.
* Implement error handling mechanism to address the issues such as duplicate file, data format discrepancies or violation of validation rules.

Step 4:

Data Validation

* Data validation can be done by using SQL queries and by Python scripts.
* SQL checks whether the loaded data meet the predefined rules and Python checks the complex data that cannot be handled easily by SQL alone.

Step 5:

Email Trigger Automation

* We can use Python ‘smtplib’ library to send email programmatically.
* Define conditions under which email should be triggered like successful data loading, validation failure.
* Craft informative email messages that provide details about the outcome of the data loading/validation process.

Step 6:

Testing and Deployment

* Test each process individually whether they all function correctly.
* Deploy the solution in a suitable environment, considering factors like scalability, reliability, and security.