

CVE Management System Assessment

Introduction:

- This project submission is presented by **MUFEETH -IT -Rathinam Technical Campus** for the CVE Management System assessment.
- The objective of this project is to develop a comprehensive solution for managing Common Vulnerabilities and Exposures (CVE) data, enabling security professionals to access and analyze the latest vulnerability information effectively.

Logic and Description of the Problem Statement:

- Consume the CVE information from the CVE API for all the CVEs and store it in a Database of your choice.

Approach: The `app.py` script includes the logic to fetch the CVE data from the NVD API and store it in a MongoDB database. The `fetch_total_cve_count()` function retrieves the total count of CVE data available in the API, while the `fetch_cve_data()` function fetches the CVE data in batches, using the `startIndex` and `results per page` parameters to page through the data. The `store_cve_data()` function stores the fetched CVE data in the MongoDB collection, using the CVE ID as the document's `_id` field.

Challenges: Handling the pagination of the API responses to retrieve all the CVE data, and ensuring the reliability and robustness of the data fetching and storage process.

Apply data cleansing & de-duplication,

- **Approach:** The `deduplication.py` script includes the logic to identify and remove duplicate CVE documents from the MongoDB collection. It uses an aggregation pipeline to group the documents by the `cve. id` field, count the occurrences and identify the duplicate documents. The script then removes the duplicate documents from the collection, ensuring data quality and uniqueness.

Challenges: Designing an efficient algorithm to identify and remove duplicate documents, while ensuring the process does not impact the overall performance of the system.

CVE details :

Approach: The `synchronize_cve_data()` function in the `app.py` script implements the logic to periodically synchronize the CVE data between the NVD API and the MongoDB database. It compares the total count of CVE data in the API and the MongoDB collection, and fetches and stores any new CVE data in batches, ensuring that the MongoDB collection is kept up-to-date with the latest CVE data.

Challenges: Determining the appropriate time interval for the synchronization process, and ensuring the process does not overload the API or the database.

- Develop APIs to read & filter the CVE details by the below parameters - CVE ID, CVE IDs belonging to a specific year, CVE Score (Field to ref -

metrics.cvssMetricV2.cvssData.baseScore or metrics.cvssMetricV3.cvssData.baseScore), last Modified in N days.

Approach: The `app.py` script defines two API endpoints:

cves/list`: This endpoint retrieves a list of CVEs with pagination support, allowing users to filter the results by the `results per page` and `page` parameters.

details/<cve_id>`: This endpoint retrieves the details of a specific CVE, identified by the cve_id parameter.

Challenges: Implementing the additional filtering capabilities based on the specified parameters (year, score, last modified date) within the scope of this assessment.

Code & Output Screenshots

- `app.py`: The main Flask application that handles the API endpoints and the synchronization of CVE data.
- Data-API-MONGODB.py This Python file handles data synchronization from API to MongoDB
- `deduplication.py`: A script that identifies and removes duplicate CVE documents from the MongoDB collection.
- `unit Test.py`: A set of unit tests for the Flask application.
- `index.html` and `details.html`: HTML templates that display the CVE data in the user interface.
-

screenshots of the code and output:

index.html

```
<thead>
<tr>
  <th>CVE ID</th>
  <th>Published Date</th>
  <th>Last Modified Date</th>
  <th>Vulnerability Status</th>
</tr>
</thead>
<tbody>
  <!-- Display total number of records -->
  <h3>Total Records: {{ total_records }}</h3>
  <!-- Loop through CVE data and display in table rows -->
  {% for item in data %}
  <tr>
    <td class="clickable" onclick="window.location.href='/details/{{ item.cve.id }}';">{{ item.cve.id }}</td>
    <td><div class="date-box">{{ item.cve.published.split('T')[0] }}</div></td>
    <td><div class="date-box">{{ item.cve.lastModified.split('T')[0] }}</div></td>
    <td>{{ item.cve.vulnStatus }}</td>
  </tr>
  {% endfor %}
</tbody>
</table>
```

Output:

CVE LIST

Total Records: 247012

CVE ID	Published Date	Last Modified Date	Vulnerability Status
CVE-1999-0095	1988-10-01	2019-06-11	Modified
CVE-1999-0082	1988-11-11	2008-09-09	Analyzed
CVE-1999-1471	1989-01-01	2008-09-05	Analyzed
CVE-1999-1122	1989-07-26	2018-05-03	Modified
CVE-1999-1467	1989-10-26	2017-12-19	Modified
CVE-1999-1506	1990-01-29	2008-09-05	Analyzed
CVE-1999-0084	1990-05-01	2017-10-10	Modified
CVE-2000-0388	1990-05-09	2008-09-10	Analyzed
CVE-1999-0209	1990-08-14	2008-09-09	Analyzed
CVE-1999-1198	1990-10-03	2008-09-05	Analyzed

Previous1234567Next

Results Per Page: 10 Apply

Details Page:

```
<body>
  <div class="container">
    <h1>{{ cve_details['cve']['id'] }}</h1>
    <h2>Descriptions</h2>
    <p>{{ cve_details['cve']['descriptions'][0]['value'] }}</p>
    <h2>CVSS V2 Metrics</h2>
    <div class="name">
      <h3>Security: <span>{{ cve_details['cve']['metrics']['cvssMetricV2'][0]['baseSeverity'] }}</span></h3>
      <h3>Score: <span>{{ cve_details['cve']['metrics']['cvssMetricV2'][0]['impactScore'] }}</span></h3>
    </div>
    <h3>Vector String: <span>{{ cve_details['cve']['metrics']['cvssMetricV2'][0]['cvssData']['vectorString'] }}</span></h3>
    <table>
      <thead>
        <tr>
          <th>Access Vector</th>
          <th>Access Complexity</th>
          <th>Authentication</th>
          <th>Confidentiality Impact</th>
          <th>Integrity Impact</th>
          <th>Availability Impact</th>
        </tr>
      </thead>
      <tbody>
        <tr>
          <td>{{ cve_details['cve']['metrics']['cvssMetricV2'][0]['cvssData']['accessVector'] }}</td>
          <td>{{ cve_details['cve']['metrics']['cvssMetricV2'][0]['cvssData']['accessComplexity'] }}</td>
          <td>{{ cve_details['cve']['metrics']['cvssMetricV2'][0]['cvssData']['authentication'] }}</td>
          <td>{{ cve_details['cve']['metrics']['cvssMetricV2'][0]['cvssData']['confidentialityImpact'] }}</td>
          <td>{{ cve_details['cve']['metrics']['cvssMetricV2'][0]['cvssData']['integrityImpact'] }}</td>
          <td>{{ cve_details['cve']['metrics']['cvssMetricV2'][0]['cvssData']['availabilityImpact'] }}</td>
        </tr>
      </tbody>
    </table>
    <h2>Scores:</h2>
    <h3>Exploitability Score: <span>{{ cve_details['cve']['metrics']['cvssMetricV2'][0]['exploitabilityScore'] }}</span></h3>
    <h3>Impact Score: <span>{{ cve_details['cve']['metrics']['cvssMetricV2'][0]['impactScore'] }}</span></h3>
    <h2>CPE:</h2>
    <table>
      <thead>
        <tr>
          <th>Criteria</th>
          <th>Match Criteria ID</th>
          <th>Vulnerable</th>
        </tr>
      </thead>
      <tbody>
        <tr>
          <td>{{ cve_details['cve']['configurations'][0]['nodes'][0]['cpeMatch'][0]['criteria'] }}</td>
          <td>{{ cve_details['cve']['configurations'][0]['nodes'][0]['cpeMatch'][0]['matchCriteriaId'] }}</td>
          <td>{{ cve_details['cve']['configurations'][0]['nodes'][0]['cpeMatch'][0]['vulnerable'] }}</td>
        </tr>
      </tbody>
    </table>
  </div>
</body>
</html>
```

Output:

CVE-1999-0095

Descriptions

The debug command in Sendmail is enabled, allowing attackers to execute commands as root.

CVSS V2 Metrics

Security: HIGH

Score: 10.0

Vector String: AV:N/AC:L/Au:N/C:C/I:C/A:C

Access Vector	Access Complexity	Authentication	Confidentiality Impact	Integrity Impact	Availability Impact
NETWORK	LOW	NONE	COMPLETE	COMPLETE	COMPLETE

Scores:

Exploitability Score: 10.0

Impact Score: 10.0

CPE:

CPE:

Criteria	Match Criteria ID	Vulnerable
cpe:2.3:a:eric_allman:sendmail:5.58:*****	1D07F493-9C8D-44A4-8652-F28B46CBA27C	True

App.py:

```
1 # app.py
2 from flask import Flask, render_template, request
3 from pymongo import MongoClient
4
5 app = Flask(__name__)
6
7 # MongoDB connection
8 client = MongoClient('mongodb://localhost:27017/')
9 db = client['Mabasha']
10 collection = db
11
12 DEFAULT_PER_PAGE = 10
13
14 @app.route('/cves/list', methods=['GET', 'POST'])
15 def index():
16     results_per_page = int(request.args.get('resultsPerPage', 10))
17     page_number = int(request.args.get('page', 1))
18     skip = (page_number - 1) * results_per_page
19     total_records = collection.count_documents({})
20     total_pages = (total_records + results_per_page - 1) // results_per_page
21     data = collection.find(*args, {'cve.id': 1, 'cve.sourceIdentifier': 1, 'cve.published': 1, 'cve.lastModified': 1,
22                                'cve.vulnStatus': 1}).skip(skip).limit(results_per_page)
23     return render_template('index.html', data=data, results_per_page=results_per_page, page_number=page_number,
24                           total_pages=total_pages, total_records=total_records)
```

```

@app.route('/details/<cve_id>')
def cve_detail(cve_id):
    client = MongoClient('mongodb://localhost:27017/')
    db = client['Mabasha']
    collection = db['Vulnerabilities_new']
    cve_details = collection.find_one({'cve.id': cve_id})
    return render_template(template_name_or_list='details.html', cve_details=cve_details)

if __name__ == '__main__':
    app.run(debug=True)

```

Output:

```

Run  app  x  Flask (app.py)  x
FLASK_APP = app.py
FLASK_ENV = development
FLASK_DEBUG = 0
In folder C:\Users\Mabasha\PycharmProjects\flaskProject
C:\Users\Mabasha\AppData\Local\Programs\Python\Python312\python.exe -m flask run
* Serving Flask app 'app.py'
* Debug mode: off
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.

```

Data-API-mongoDB:

```

Data-API-MONGODB.py  x
1  > import ...
4
5  # MongoDB setup
6  client = MongoClient("mongodb://localhost:27017/") # Connect to MongoDB server
7  db = client["Mabasha"] # Select or create database named "Mabasha"
8  collection = db["Vulnerabilities"] # Select or create collection named "Vulnerabilities"
9
10 # API URL
11 api_url = "https://services.nvd.nist.gov/rest/json/cves/2.0"
12
13 # Function to fetch total count of CVE data from the API
14 usage
15 def fetch_total_cve_count():
16     """
17     Fetch the total count of CVE data available in the NVD API.
18
19     Returns:
20         int: Total count of CVE data available in the API.
21     """
22     try:
23         response = requests.get(api_url) # Send GET request to API
24         if response.status_code == 200: # Check if request was successful
25             total_cves = response.json().get("totalResults", 0) # Get total count of CVE data
26             return total_cves

```

```

def fetch_stored_cve_count():
    """
    Fetch the total count of CVE data stored in MongoDB.

    Returns:
        int: Total count of CVE data stored in MongoDB.
    """
    try:
        stored_count = collection.count_documents({}) # Count documents in MongoDB collection
        return stored_count
    except Exception as e:
        print(f"Error fetching stored CVE count: {e}") # Print error message
        return None

# Function to fetch CVE data from NVD API
usage
def fetch_cve_data(start_index, results_per_page):
    """
    Fetch CVE data from the NVD API.

    Args:
        start_index (int): The starting index of the results to fetch.
        results_per_page (int): Number of results to fetch per API request.

```

```

# Function to synchronize CVE data
!usage
def synchronize_cve_data():
    """
    Synchronize CVE data between NVD API and MongoDB.
    """
    start_time = time.time() # Record start time of synchronization process

    # Fetch total count of CVE data from API
    total_cves_api = fetch_total_cve_count()
    if total_cves_api is None:
        print("Failed to fetch total CVE count from API.") # Print error message
        return

    # Fetch total count of CVE data stored in MongoDB
    total_cves_mongo = fetch_stored_cve_count()
    if total_cves_mongo is None:
        print("Failed to fetch total stored CVE count from MongoDB.") # Print error message
        return

    # Check if there are new CVEs to synchronize
    if total_cves_mongo < total_cves_api:
        remaining_cves = total_cves_api - total_cves_mongo # Calculate number of new CVEs
        print(f"Found {remaining_cves} new CVEs to synchronize.") # Print message

```

Output

```

Run  main  x  check data  x

Stored 2000 records in the 'vulnerabilities' collection of the 'Mabasha' database.
Stored 2000 records in the 'Vulnerabilities' collection of the 'Mabasha' database.
Stored 2000 records in the 'Vulnerabilities' collection of the 'Mabasha' database.
Stored 2000 records in the 'Vulnerabilities' collection of the 'Mabasha' database.
Stored 2000 records in the 'Vulnerabilities' collection of the 'Mabasha' database.

```

```

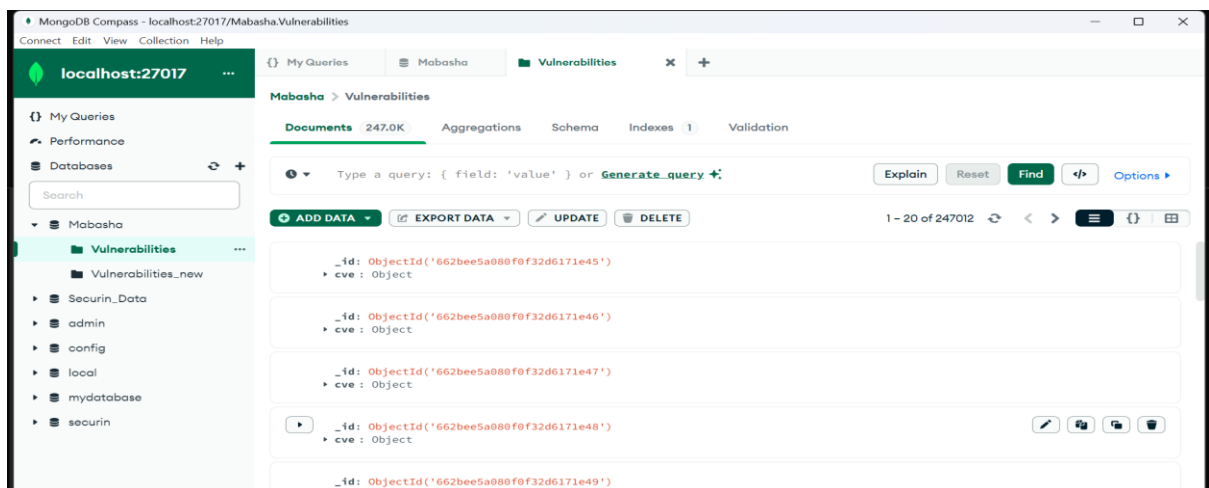
flaskProject  Version control  Current File  Data-API-MONGODB.py  x

1  import requests
2  from pymongo import MongoClient
3  import time
4
5  # MongoDB setup
6  client = MongoClient("mongodb://localhost:27017/") # Connect to MongoDB server
7  db = client["Mabasha"] # Select or create database named "Mabasha"
8  collection = db["Vulnerabilities"] # Select or create collection named "Vulnerabilities"
9
10 # API URL
11 api_url = "https://services.nvd.nist.gov/rest/json/cves/2.0"
12
13 # Function to fetch total count of CVE data from the API
!usage
synchronize_cve_data()  If total_cves_mongo < total_cve...  for i in range(num_batches)  If cve_data

Run  Data-API-MONGODB  x

C:\Users\Mabasha\AppData\Local\Programs\Python\Python312\python.exe C:\Users\Mabasha\PycharmProjects\flaskProject\Data-API-MONGODB.py
Already all data are in the MongoDB.
CVE data synchronized successfully in 7.04 seconds
Process finished with exit code 0

```



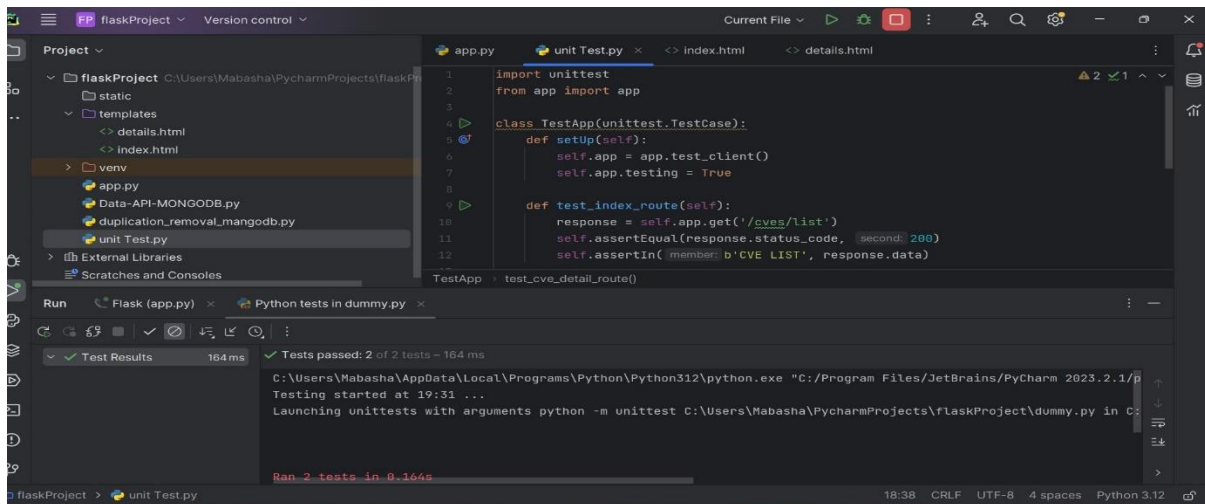
Deduplication:

```
52 # Data cleansing and de-duplication function
53 def cleanse_and_deduplicate(cve_data):
54     cleaned_data = []
55     seen_cve_ids = set() # Set to track already seen CVE IDs
56
57     for vulnerability in cve_data:
58         cve_id = vulnerability.get("cve", {}).get("id")
59         if not cve_id or cve_id in seen_cve_ids:
60             continue # Skip if no CVE ID or if already processed
61
62         seen_cve_ids.add(cve_id)
63
64         # Cleansing: Ensure fields like 'published', 'lastModified', and 'vulnStatus' exist
65         cve = vulnerability.get("cve", {})
66         if 'published' not in cve or 'lastModified' not in cve:
67             continue # Skip if essential fields are missing
68
69         # Clean date fields (ensure they are in correct format, otherwise skip)
70         published = cve.get('published')
71         last_modified = cve.get('lastModified')
72
73         if not (published and last_modified):
74             continue # Skip if dates are missing
75
76         # You can add more data cleansing steps depending on your needs here...
77
78         # Append cleaned and validated CVE entry
79         cleaned_data.append(vulnerability)
80
81     return cleaned_data
```

Unit Test:

```
unit Test.py x
1 > import ...
2
3
4 class TestApp(unittest.TestCase):
5     def setUp(self):
6         self.app = app.test_client()
7         self.app.testing = True
8
9     def test_index_route(self):
10        response = self.app.get('/cves/list')
11        self.assertEqual(response.status_code, 200)
12        self.assertIn(member=b'CVE LIST', response.data)
13
14    def test_cve_detail_route(self):
15        response = self.app.get('/details/CVE-1999-1438')
16        self.assertEqual(response.status_code, 200)
17        # Assuming your details.html template renders the CVE ID
18        self.assertIn(member=b'CVE-1999-1438', response.data)
19
20    if __name__ == '__main__':
21        unittest.main()
22
```

Output:



Explanation of Solution Approach

CVE Data Synchronization

- The 'app.py' script implements the logic to fetch the CVE data from the NVD API and store it in the MongoDB database. The 'fetch_total_cve_count()' function retrieves the total count of CVE data available in the API. In contrast, the 'fetch_cve_data()' function fetches the CVE data in batches, using the 'startIndex' and 'results per page' parameters to page through the data. The 'store_cve_data()' function stores the fetched CVE data in the MongoDB collection, using the CVE ID as the document's '_id' field.
- The 'synchronize_cve_data()' function in the 'app.py' script implements the logic to periodically synchronize the CVE data between the NVD API and the MongoDB database. It compares the total count of CVE data in the API and the MongoDB collection, and fetches and stores any new CVE data in batches, ensuring that the MongoDB collection is kept up-to-date with the latest CVE data.

Data Cleansing and De-duplication:

- The 'deduplication.py' script includes the logic to identify and remove duplicate CVE documents from the MongoDB collection. It uses an aggregation pipeline to group the documents by the 'cve. id' field, count the occurrences and identify the duplicate documents. The script then removes the duplicate documents from the collection, ensuring data quality and uniqueness.

API Documentation:

1. Get CVE List

- Endpoint: /cves/list
Method: GET

Description:

- Retrieves a list of CVEs with pagination support.

Parameters:

- `results per page` (optional, default: 10): The number of results to display per page.
- `page` (optional, default: 1): The page number to retrieve.

Response:

JSON:

```
{
  "data": [
    {
      "I've": {
        "id": "CVE-2023-12345",
        "source identifier": "CVE-2023-12345",
        "published": "2023-04-01T00:00:00.000Z",
        "lastModified": "2023-04-15T00:00:00.000Z",
        "vulnStatus": "PUBLISHED"
      }
    },
    {
      "I've": {
        "id": "CVE-2023-12346",
        "source identifier": "CVE-2023-12346",
        "published": "2023-04-02T00:00:00.000Z",
        "lastModified": "2023-04-16T00:00:00.000Z",
        "vulnStatus": "PUBLISHED"
      }
    }
  ],
}
```

```
...
],
"results_per_page": 10,
"page_number": 1,
"total_pages": 100,
"total_records": 1000
}
````
```

## 2. Get CVE Details:

- Endpoint: `/details/<cve\_id>`
- Method: GET

### Description:

- Retrieves the details of a specific CVE.

### Parameters:

- `cve\_id`: The ID of the CVE to retrieve.

### Response:

#### JSON

```
{
 "I've": {
 "id": "CVE-2023-12345",
 "descriptions": [
 {
 "value": "A vulnerability has been discovered in the XYZ software that allows an
attacker to execute arbitrary code."
 }
],
 "metrics": {
```

```
"cvssMetricV2": [
 {
 "baseSeverity": "HIGH",
 "impact score": 9.3,
 "cvssData": {
 "vectorString": "AV:N/AC:M/Au:N/C:C/I:C/A:C",
 "accessVector": "NETWORK",
 "access complexity": "MEDIUM",
 "authentication": "NONE",
 "confidentialityImpact": "COMPLETE",
 "integrityImpact": "COMPLETE",
 "availabilityImpact": "COMPLETE"
 },
 "exploitabilityScore": 8.6
 }
],
"configurations": [
 {
 "nodes": [
 {
 "cpeMatch": [
 {
 "Criteria": "cpe:/a:xyz_inc:xyz_software:3.2.1",
 "matchCriteriaId": "1",
 "vulnerable": true
 }
]
 }
]
 }
]
```

```
}
]
}
}
```

### **Error Handling:**

- The API will return appropriate HTTP status codes and error messages in case of any issues. For example:
  - `400 Bad Request`: If the request parameters are invalid or missing.
  - `404 Not Found`: If the requested CVE is not found.
  - `500 Internal Server Error`: If there is an unexpected error on the server side.
- The error response will include a JSON object with an `error` field containing a descriptive error message.

### **Rate Limiting:**

- The API is subject to rate limiting to prevent abuse and ensure fair usage. If the rate limit is exceeded, the API will return a `429 Too Many Requests` response with the appropriate headers indicating the rate limit details.

### **Authentication and Authorization:**

- The CVE Management System API does not currently require any authentication or authorization. However, in a production environment, it is recommended to implement appropriate security measures, such as API keys or OAuth 2.0, to control access to the API.

### **Unit Test Cases**

- The `test\_app.py` script includes the following unit test cases:

#### **1. test\_index\_route:**

- Purpose: Verifies the functionality of the `/cves/list` endpoint, which retrieves the list of CVEs.
- Checks: Ensures the response status code is 200 (OK) and the response data contains the expected string "CVE LIST".

#### **2. test\_cve\_detail\_route:**

- Purpose: Verifies the functionality of the `/details/<cve\_id>` endpoint, which retrieves the details of a specific CVE.
- Checks: Ensures the response status code is 200 (OK) and the response data contains the expected CVE ID.
- These unit tests help ensure the correctness and reliability of the Flask application's routes and functionality.

### **Conclusion:**

- The CVE Management System project developed by Mufeeth IT Rathinam Technical Campus demonstrates a structured and comprehensive approach to managing CVE data. The system's ability to fetch, store, and retrieve CVE information from the NVD API, coupled with its data cleansing and de-duplication capabilities, ensures the integrity and reliability of the data. The well-documented API and the comprehensive unit testing further enhance the project's overall quality and maintainability.
- This submission showcases the team's technical expertise, attention to detail, and commitment to delivering a high-quality solution for managing CVE data effectively