Python to Extract JSON and Create MySQL Table

Author
Kaustubh Oak

Problem Statements

1. Given this JSON as a sample, how will you write a short script/program to process this Json and store into databases

2. Explain selection of coding language and DB of your choice.

3. Given that its large set of data, how will you ensure that no records are missed when writing into the DB

4. In the case of duplicate entries, how will you handle them?

Proposed Approach / Solution

Extract Keys from JSON File

To know the schema required for database

Identify Keys

Examine the JSON structure to determine the relevant keys to extract.

Read JSON File

Use a Python library like json to read and parse the JSON data.

Extract Key-Value Pairs

Iteratively access the keys and their corresponding values from the JSON.

Code Snippet

```
with open("data.json" , "r") as f:
    data = json.load(f)

extract keys from json file.
    keys = set()
    for record in data:
        for ks in record.keys():
            keys.add(ks)
    print(keys)
```

Output

```
{'comment',
 'decryptor',
 'encryptionAlgorithm',
 'extensionPattern',
 'extensions',
 'iocs',
 'microsoftDetectionName',
 'microsoftInfo',
 'name',
 'ransomNoteFilenames',
 'resources',
 'sandbox',
 'screenshots',
 'snort'}
```

MySQL Queries to Create Schema

To setup database for accepting data

Create Database

Use SQL to create a new database to hold the table.

Create Table

Write a CREATE TABLE statement to define the structure of the new table.

```
CREATE TABLE ransomwaredata (
    id INT AUTO_INCREMENT PRIMARY KEY,
    name VARCHAR(255),
    iocs TEXT DEFAULT NULL,
    comment TEXT DEFAULT NULL,
    encryptionAlgorithm TEXT DEFAULT NULL,
    snort TEXT DEFAULT NULL,
    extensions TEXT DEFAULT NULL,
    screenshots TEXT DEFAULT NULL,
    microsoftInfo TEXT DEFAULT NULL,
    ransomNoteFilenames TEXT DEFAULT NULL,
    extensionPattern TEXT DEFAULT NULL,
    decryptor TEXT DEFAULT NULL,
    sandbox TEXT DEFAULT NULL,
    microsoftDetectionName TEXT DEFAULT NULL
);
```

```
CREATE TABLE resource (
    id INT AUTO_INCREMENT PRIMARY KEY,
    name VARCHAR(255),
    resources TEXT
);
```

Insert Data to MySQL Database

To parse json data into sql queries and insert data into database

1 Prepare Data

Parse the JSON data for insertion into the MySQL table.

```
# parsing the key and valuses for sql queries
    for record in data:
        k = []
        v = []
        tresources = record['resources']   # storing recources key temporary
        del(record['resources'])
        tname = record['name']             # storing name key temporary
        del(record['name'])
        for key, value in record.items():
          k.append(key)
          v.append(value)
        v = [ json.dumps(x) for x in v]    # Adding double quotes to Vaule string
        for n in tname:
```

2 Execute Queries

Use SQL INSERT statements to add the data rows to the database.

```
# Create records for each ransomeware name ; execpt resources key
sql = "INSERT INTO ransomwaredata ({},{}) VALUES ({},{})".format('name',','.join(k),  json.dumps(n), ','.join(v))
mycursor.execute(sql)
```

```
# Create records for each resource key
    sql2 = "INSERT INTO resource (name,resources) VALUES ({},{})".format(json.dumps(n),json.dumps(rs))
    mycursor.execute(sql2)
```

3 Verify Insertion

Query the table to ensure the data was successfully added.

114	CryptXXX	https://otx.alienvault.com/browse?q=CryptXXX	Comes with Bedep
115	CryptProjectXXX	https://otx.alienvault.com/browse?q=CryptXXX	Comes with Bedep
116	CryptXXX 2.0	https://otx.alienvault.com/browse?q=CryptXXX+2.0	Locks screen. Ransom note names are an ID.¶Comes wit
117	CryptProjectXXX	https://otx.alienvault.com/browse?q=CryptXXX+2.0	Locks screen. Ransom note names are an ID.¶Comes wit

4 Duplicate Record Handling

This solution suggests to keep track of duplicate ransomware names with additional properties. This helps to keep track of all properties and information versions.

Thank you !