CPE Datacollection and API Development

By: Rathinadevan E M _ SVCE _ CSE

Tech Stack:

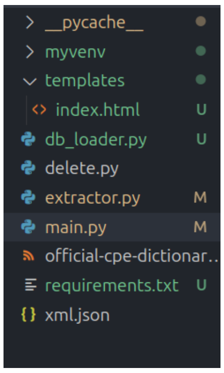
1. Fastapi
2. Jinja templating Engine(for frontend)
3. mongodb
4. uvicorn(for server)
5. xmltodict(conversion of xml to json)

Logical Approach Taken:

1. Conversion of xml to json
2. Creation of mongodb database by running a container with port mapping done
3. Uploading of data to it following a schema
4. Creation of fastapi app and running it through uvicorn
5. Creation of endpoints for db fetching
6. Parsing parameters from endpoints and performing operations

Q1: Converting XML to json :

File used: extractor.py
Why : Fastapi has native support to json and hence easier to use
What it does: creates a file named xml.json which contains all data converted into xml format
What tools used : xmltodict and json package

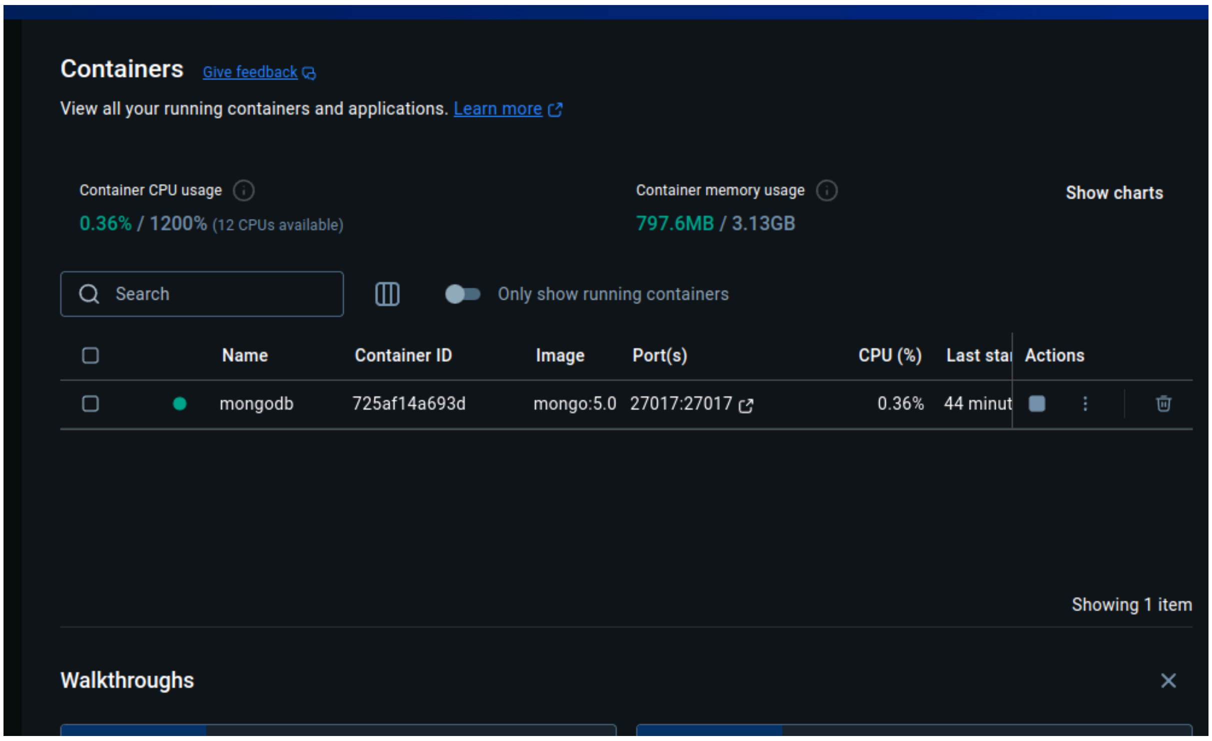


```
> __pycache__
> myvenv
v templates
    index.html          U
  db_loader.py          U
  delete.py
  extractor.py          M
  main.py               M
  official-cpe-dictionar...
  requirements.txt      U
  xml.json
```

Q: Parsing the json data and pushing it into mongodb :

Running : As a container with port mapping done
Access Methods : Connecting through Mongodb Compass
Tools used : json and pymongo , Docker(to run mongodb instance)
Url of running mongodb: mongodb://localhost:27017/
File used to load data : db_loader.py

Images :



Containers Give feedback

View all your running containers and applications. Learn more

Container CPU usage
0.36% / 1200% (12 CPUs available)

Container memory usage
797.6MB / 3.13GB

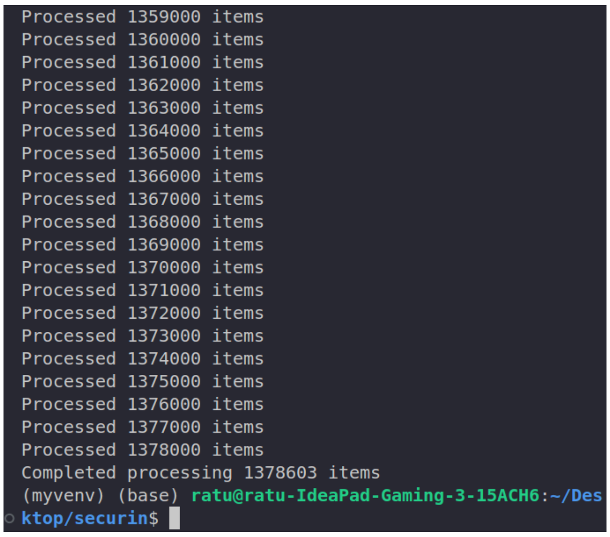
Show charts

Only show running containers

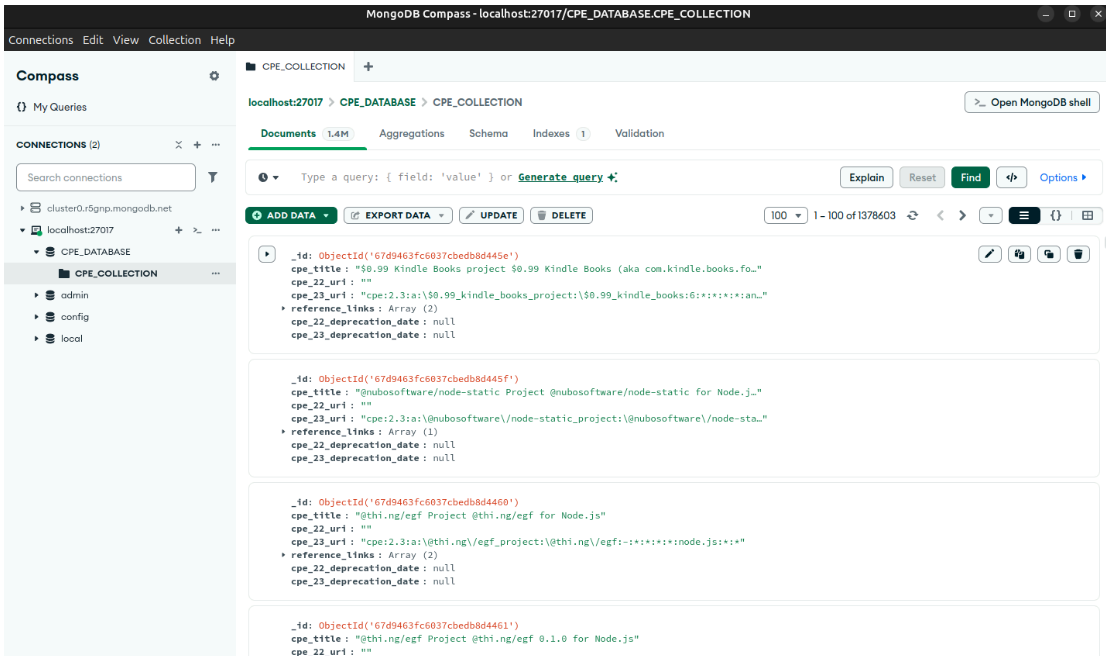
Name	Container ID	Image	Port(s)	CPU (%)	Last sta	Actions
mongodb	725af14a693d	mongo:5.0	27017:27017	0.36%	44 minut	

Showing 1 item

Walkthroughs



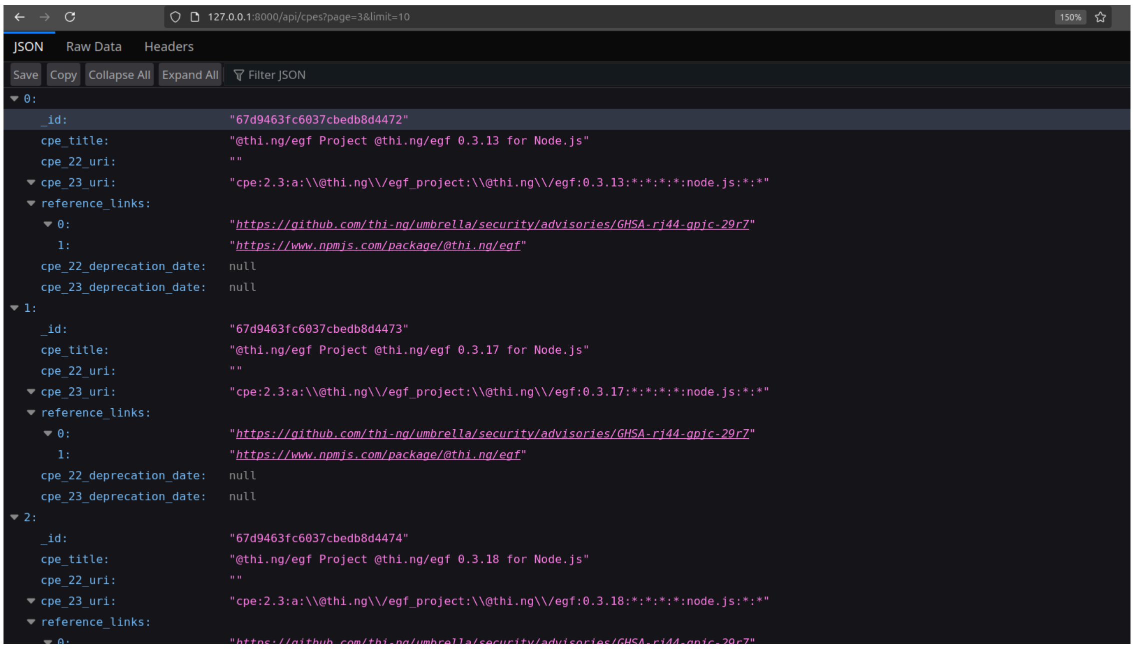
```
Processed 1359000 items
Processed 1360000 items
Processed 1361000 items
Processed 1362000 items
Processed 1363000 items
Processed 1364000 items
Processed 1365000 items
Processed 1366000 items
Processed 1367000 items
Processed 1368000 items
Processed 1369000 items
Processed 1370000 items
Processed 1371000 items
Processed 1372000 items
Processed 1373000 items
Processed 1374000 items
Processed 1375000 items
Processed 1376000 items
Processed 1377000 items
Processed 1378000 items
Completed processing 1378603 items
(myvenv) (base) ratu@ratu-IdeaPad-Gaming-3-15ACH6:~/Des
ktop/securin$
```



```
_id: ObjectId('67d9463fc6037cbedb8d445e')
cpe_title : "$0.99 Kindle Books project $0.99 Kindle Books (aka com.kindle.books.fo…"
cpe_22_uri : ""
cpe_23_uri : "cpe:2.3:a:\$0.99_kindle_books_project:\$0.99_kindle_books:6:*:*:*:*:an…"
reference_links : Array (2)
cpe_22_deprecation_date : null
cpe_23_deprecation_date : null

_id: ObjectId('67d9463fc6037cbedb8d445f')
cpe_title : "@nubosoftware/node-static Project @nubosoftware/node-static for Node.j…"
cpe_22_uri : ""
cpe_23_uri : "cpe:2.3:a:\@nubosoftware\/node-static_project:\@nubosoftware\/node-sta…"
reference_links : Array (1)
cpe_22_deprecation_date : null
cpe_23_deprecation_date : null

_id: ObjectId('67d9463fc6037cbedb8d4460')
cpe_title : "@thi.ng/egf Project @thi.ng/egf for Node.js"
cpe_22_uri : ""
cpe_23_uri : "cpe:2.3:a:\@thi.ng\/egf_project:\@thi.ng\/egf:-:*:*:*:*:node.js:*:*"
reference_links : Array (2)
cpe_22_deprecation_date : null
cpe_23_deprecation_date : null

_id: ObjectId('67d9463fc6037cbedb8d4461')
cpe_title : "@thi.ng/egf Project @thi.ng/egf 0.1.0 for Node.js"
cpe_22_uri : ""
```

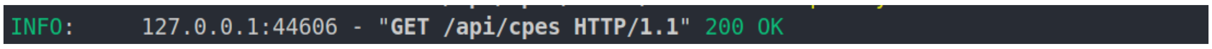
Q 3: Working of endpoint (`/api/cpes`)

Parameters used: page(optional),limit(optional)
Use : Retrieves the data from the server while simultaneously utilizing pagination with parameters such as page and limit
Notes : Default limit is set to 10 and can be changed by the user



```
0:
  _id:                      "67d9463fc6037cbedb8d4472"
  cpe_title:                "@thi.ng/egf Project @thi.ng/egf 0.3.13 for Node.js"
  cpe_22_uri:               ""
  cpe_23_uri:               "cpe:2.3:a:\\@thi.ng\\/egf_project:\\@thi.ng\\/egf:0.3.13:*:*:*:*:node.js:*:*"
  reference_links:
    0:                      "https://github.com/thi-ng/umbrella/security/advisories/GHSA-rj44-gpjc-29r7"
    1:                      "https://www.npmjs.com/package/@thi.ng/egf"
  cpe_22_deprecation_date:  null
  cpe_23_deprecation_date:  null
1:
  _id:                      "67d9463fc6037cbedb8d4473"
  cpe_title:                "@thi.ng/egf Project @thi.ng/egf 0.3.17 for Node.js"
  cpe_22_uri:               ""
  cpe_23_uri:               "cpe:2.3:a:\\@thi.ng\\/egf_project:\\@thi.ng\\/egf:0.3.17:*:*:*:*:node.js:*:*"
  reference_links:
    0:                      "https://github.com/thi-ng/umbrella/security/advisories/GHSA-rj44-gpjc-29r7"
    1:                      "https://www.npmjs.com/package/@thi.ng/egf"
  cpe_22_deprecation_date:  null
  cpe_23_deprecation_date:  null
2:
  _id:                      "67d9463fc6037cbedb8d4474"
  cpe_title:                "@thi.ng/egf Project @thi.ng/egf 0.3.18 for Node.js"
  cpe_22_uri:               ""
  cpe_23_uri:               "cpe:2.3:a:\\@thi.ng\\/egf_project:\\@thi.ng\\/egf:0.3.18:*:*:*:*:node.js:*:*"
  reference_links:
```

Notes : Default limit is set to 10 and can be changed by the user

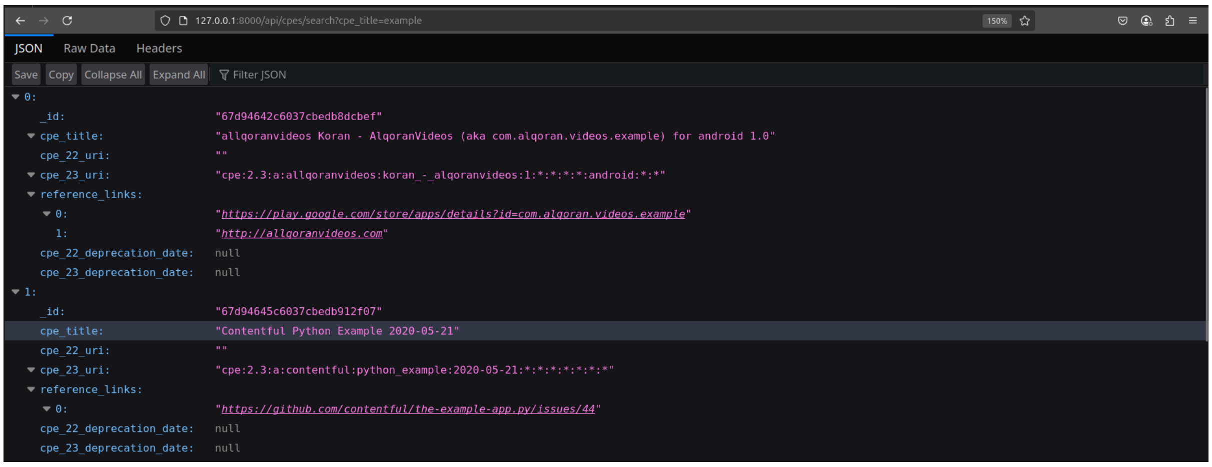


```
INFO:     127.0.0.1:44606 - "GET /api/cpes HTTP/1.1" 200 OK
```

Q 4 : **Working of endpoint** (api/cpes/search):

```
Query Parameters : cpe_title: Optional[str] = None,
      cpe_22_uri: Optional[str] = None,
      cpe_23_uri: Optional[str] = None,
      deprecation_date: Optional[str] = None,
```

> These parameters are optional and the api displays all the data if the option not given
> Partial parameters are also taken care as shown below



127.0.0.1:8000/api/cpes/search?cpe_title=example

```
0:
  _id:                      "67d94642c6037cbedb8dcbef"
  cpe_title:                "allqoranvideos Koran - AlqoranVideos (aka com.alqoran.videos.example) for android 1.0"
  cpe_22_uri:               ""
  cpe_23_uri:               "cpe:2.3:a:allqoranvideos:koran_-_alqoranvideos:1:*:*:*:*:android:*:*"
  reference_links:
    0:                      "https://play.google.com/store/apps/details?id=com.alqoran.videos.example"
    1:                      "http://allqoranvideos.com"
  cpe_22_deprecation_date:  null
  cpe_23_deprecation_date:  null
1:
  _id:                      "67d94645c6037cbedb912f07"
  cpe_title:                "Contentful Python Example 2020-05-21"
  cpe_22_uri:               ""
  cpe_23_uri:               "cpe:2.3:a:contentful:python_example:2020-05-21:*:*:*:*:*:*:*"
  reference_links:
    0:                      "https://github.com/contentful/the-example-app.py/issues/44"
  cpe_22_deprecation_date:  null
  cpe_23_deprecation_date:  null
```

Q 5 : Serving Frontend :

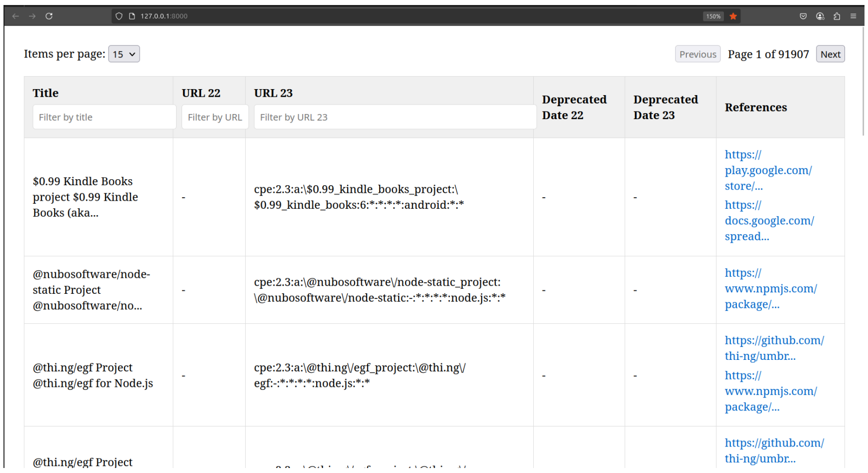
Done using : Jinja Templating Engine

Steps :

1. Index.html template is placed in the templates folder .
2. The templated directory selected as default directory for templates for jinja
3. The user is redirected to the frontend whenever they access the default route or /.
4. The redirection is done by rendering the frontend whenever the user accesses / directory.
5. The Ui was designed to enable user to decide the number of data displayed at a page at a time.
6. Filters can also be done through the UI itself enabling easier user interaction

Images:

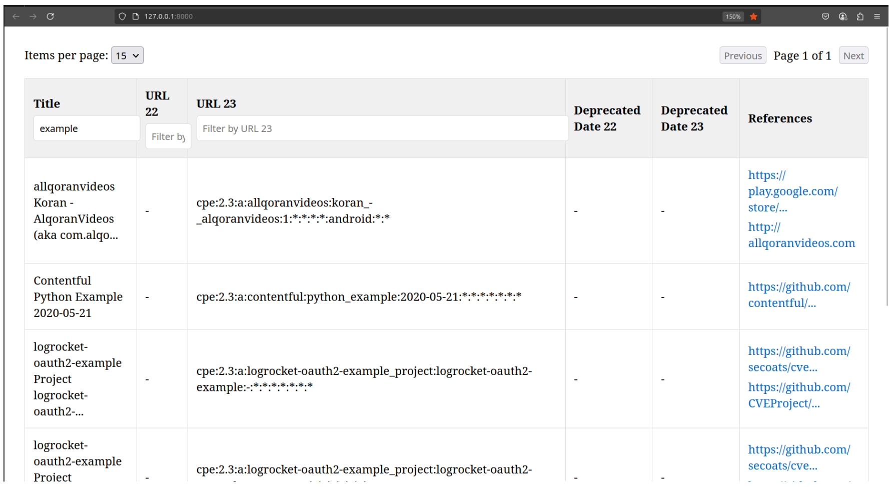
Images showing default rendering along with pagination


127.0.0.1:8000

Items per page: 15 | Previous Page 1 of 91907 Next

Title	URL 22	URL 23	Deprecated Date 22	Deprecated Date 23	References
Filter by title	Filter by URL	Filter by URL 23			
$0.99 Kindle Books project $0.99 Kindle Books (aka...	-	cpe:2.3:a:\$0.99_kindle_books_project:\$0.99_kindle_books:6:*:*:*:*:android:*:*	-	-	https://play.google.com/store/... https://docs.google.com/spread...
@nubosoftware/node-static Project @nubosoftware/no...	-	cpe:2.3:a:\@nubosoftware\/node-static_project:\@nubosoftware\/node-static:-:*:*:*:*:node.js:*:*	-	-	https://www.npmjs.com/package/...
@thi.ng/egf Project @thi.ng/egf for Node.js	-	cpe:2.3:a:\@thi.ng\/egf_project:\@thi.ng\/egf:-:*:*:*:*:node.js:*:*	-	-	https://github.com/thi-ng/umbr... https://www.npmjs.com/package/...
@thi.ng/egf Project					https://github.com/thi-ng/umbr...

Implementing search in the frontend:

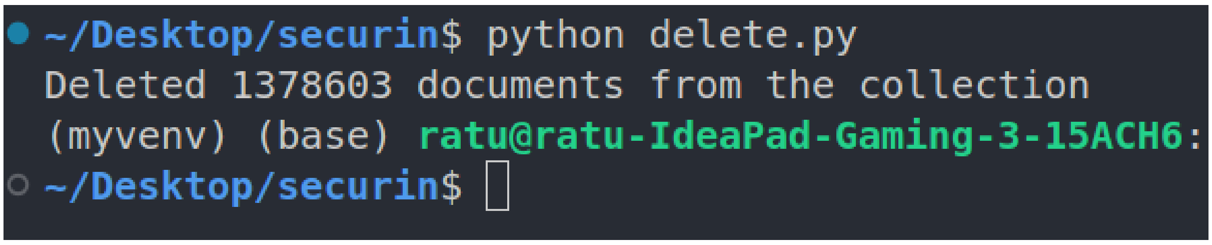

127.0.0.1:8000

Items per page: 15 | Previous Page 1 of 1 Next

Title	URL 22	URL 23	Deprecated Date 22	Deprecated Date 23	References
example	Filter by	Filter by URL 23			
allqoranvideos Koran - AlqoranVideos (aka com.alqo...	-	cpe:2.3:a:allqoranvideos:koran_-_alqoranvideos:1:*:*:*:*:android:*:*	-	-	https://play.google.com/store/... http://allqoranvideos.com
Contentful Python Example 2020-05-21	-	cpe:2.3:a:contentful:python_example:2020-05-21:*:*:*:*:*:*:*	-	-	https://github.com/contentful/...
logrocket-oauth2-example Project logrocket-oauth2-...	-	cpe:2.3:a:logrocket-oauth2-example_project:logrocket-oauth2-example:-:*:*:*:*:*:*:*	-	-	https://github.com/secoats/cve... https://github.com/CVEProject/...
logrocket-oauth2-example Project	-	cpe:2.3:a:logrocket-oauth2-example_project:logrocket-oauth2-	-	-	https://github.com/secoats/cve...

Q 6 : Deletion from database:

Can use the delete.py to delete the mongodb database if needed clearing
File used: delete.py


```
~/Desktop/securin$ python delete.py
Deleted 1378603 documents from the collection
(myvenv) (base) ratu@ratu-IdeaPad-Gaming-3-15ACH6:
~/Desktop/securin$
```

Challenged Encountered:

1. Parsing of the very large xml data
2. Pushing the data into the database , due to its heavy nature the pushing was done in parts of 1000

3. Pushing of data into the database , due to its large size the pushing took much time .
4. Parsing of Date type from the mongodb and actually using it as a query
5. Processing or iterating through all the data in the mongodb as the _id is of ObjectId type , the conversion of object ID into str was needed in order to iterate through all the instances of the data
6. References and titles had to done special iteration methods as they were sometimes dictionary and sometimes list , this iteration was also required while changing the ObjectId into string
7. Using date parameters and querying based on it were tougher as dates was implemented in different format while querying using different format
8. The null values of data such as in cpe_22_deprecation_date,cpe_23_deprecation_date,cpe_22_uri,cpe_22_uri should be taken care due to their nature of having null value sometimes and should be taken care with retrieving such data or iterating through them
9. Mongodb server not running locally , requiring a container with port mapping to be spun in docker to used as mongodb server.
10. Uploading and using the data sometimes crashed the system due to its size , hence the upload had to done as chunks so that it can be used and .