**Vespa**

**VESPA**
Shweta Ajay Shinde
Masters in Data Analytics, San Jose State University
Data 226: Data warehousing
Instructor: Keeyong Han
06th November 2024

**Vespa**

## Introduction to VESPA

1. **Download, Process and Generate an input file to VESPA (2pt)**
   - o **Should have "doc_id", "title" and "text". Refer to this scriptLinks to an external site. as a reference**

```
PS F:\Shweta MSDA Sem1\Data 226 Datawarehousing\sjsu-data226-main\sjsu-data226-main\week11> python process_tmdb_csv_2_jsonl.py
                        put                                          fields
0    id:hybrid-search:doc::19995   {'doc_id': 19995, 'title': 'Avatar', 'text': '...
1     id:hybrid-search:doc::285   {'doc_id': 285, 'title': 'Pirates of the Carib...
2   id:hybrid-search:doc::206647  {'doc_id': 206647, 'title': 'Spectre', 'text':...
3    id:hybrid-search:doc::49026  {'doc_id': 49026, 'title': 'The Dark Knight Ri...
4    id:hybrid-search:doc::49529  {'doc_id': 49529, 'title': 'John Carter', 'tex...
```
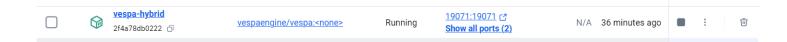
| | | | | |
|---|---|---|---|---|
| 📁 app | 11/5/2024 4:02 PM | File folder | | |
| 📄 clean_tmdb.jsonl | 11/6/2024 6:24 PM | JSONL File | 2,000 KB | |
| 📄 process_tmdb_csv_2_jsonl.py | 11/5/2024 4:01 PM | Python Source File | 4 KB | |
| 📄 pyvesap_search.py | 11/6/2024 6:37 PM | Python Source File | 3 KB | |
| 📄 README.md | 11/5/2024 4:01 PM | Markdown Source ... | 1 KB | |
| 📄 tmdb_5000_movies.csv | 11/5/2024 3:24 PM | Microsoft Excel Co... | 5,566 KB | |

2. **Run VESPA as a Docker Container (1pt)**

```
PS F:\Shweta MSDA Sem1\Data 226 Datawarehousing\sjsu-data226-main\sjsu-data226-main\week11> docker run -d --name vespa-hybrid --hostname vespa-container --p
ublish 8082:8080 --publish 19071:19071 vespaengine/vespa
Unable to find image 'vespaengine/vespa:latest' locally
latest: Pulling from vespaengine/vespa
9e8596c68b38: Download complete
2609da11fd88: Download complete
8405cd9efa2e: Download complete
5e4c4aad9d46: Download complete
44f111e695cf: Download complete
Digest: sha256:6e4090bfa309788ed3b695276d796870aeb9f60e8f76d076f1264c63d598a98b
Status: Downloaded newer image for vespaengine/vespa:latest
2f4a78db0222d60f10c27bf3877f4996eed251193fad5ba3d13c277557769c1c
```

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ☐ | 📦 | **vespa-hybrid**<br>2f4a78db0222 📋 | vespaengine/vespa:\<none\> | Running | 19071:19071 ↗<br>Show all ports (2) | N/A  36 minutes ago | ■ | ⋮  🗑 |

**Vespa**

3. **Configure VESPA and Ingest the file into VESPA (2pt).**

```
Successfully installed vespacli 0.391.25
PS F:\Shweta MSDA Sem1\Data 226 Datawarehousing\sjsu-data226-main\sjsu-data226-main\week11> vespa config set target local
PS F:\Shweta MSDA Sem1\Data 226 Datawarehousing\sjsu-data226-main\sjsu-data226-main\week11> vespa status deploy --wait 300
Waiting up to 5m0s for deploy API...
Deploy API at http://127.0.0.1:19071 is ready
PS F:\Shweta MSDA Sem1\Data 226 Datawarehousing\sjsu-data226-main\sjsu-data226-main\week11> vi pyvesap_search.py
PS F:\Shweta MSDA Sem1\Data 226 Datawarehousing\sjsu-data226-main\sjsu-data226-main\week11> vespa deploy --wait 300 app
Waiting up to 5m0s for deploy API...
Uploading application package... done

Success: Deployed 'app' with session ID 2
Waiting up to 5m0s for deployment to converge...
Waiting up to 5m0s for cluster discovery...
Waiting up to 5m0s for container default...
```

4. **Run search against VESPA (3pt)**
   o **Use [this scriptLinks to an external site.](#) as a reference (keyword search, semantic search and recommendation search)**

   **Output of all searches:-**

```
PS F:\Shweta MSDA Sem1\Data 226 Datawarehousing\sjsu-data226-main\sjsu-data226-main\week11> python pyvesap_search.py
   doc_id                               title
0     767       Harry Potter and the Half-Blood Prince
1     671  Harry Potter and the Philosopher's Stone
2     674         Harry Potter and the Goblet of Fire
3     673  Harry Potter and the Prisoner of Azkaban
4  13967                               Miss Potter
   doc_id                               title
0     767       Harry Potter and the Half-Blood Prince
1     675  Harry Potter and the Order of the Phoenix
2     672     Harry Potter and the Chamber of Secrets
3     674         Harry Potter and the Goblet of Fire
4  168705                               BloodRayne
   doc_id                               title                                       text
0     767       Harry Potter and the Half-Blood Prince  As Harry begins his sixth year at Hogwarts, he...
1     675  Harry Potter and the Order of the Phoenix  Returning for his fifth year of study at Hogwa...
2     672     Harry Potter and the Chamber of Secrets  Ignoring threats to his life, Harry returns to...
3     671  Harry Potter and the Philosopher's Stone  Harry Potter has lived under the stairs at his...
4     674         Harry Potter and the Goblet of Fire  Harry starts his fourth year at Hogwarts, comp...
PS F:\Shweta MSDA Sem1\Data 226 Datawarehousing\sjsu-data226-main\sjsu-data226-main\week11>
```

   **Keyword Search:-**

```
PS F:\Shweta MSDA Sem1\Data 226 Datawarehousing\sjsu-data226-main\sjsu-data226-main\week11> python pyvesap_search.py
   doc_id                               title
0     767       Harry Potter and the Half-Blood Prince
1     671  Harry Potter and the Philosopher's Stone
2     674         Harry Potter and the Goblet of Fire
3     673  Harry Potter and the Prisoner of Azkaban
4  13967                               Miss Potter
```

   **Semantic Search:-**

```
   doc_id                               title
0     767       Harry Potter and the Half-Blood Prince
1     675  Harry Potter and the Order of the Phoenix
2     672     Harry Potter and the Chamber of Secrets
3     674         Harry Potter and the Goblet of Fire
4  168705                               BloodRayne
```

**Vespa**

**Recommandation Search:-**

```
  doc_id                                title                                              text
0   767         Harry Potter and the Half-Blood Prince  As Harry begins his sixth year at Hogwarts, he...
1   675  Harry Potter and the Order of the Phoenix  Returning for his fifth year of study at Hogwa...
2   672      Harry Potter and the Chamber of Secrets  Ignoring threats to his life, Harry returns to...
3   671    Harry Potter and the Philosopher's Stone  Harry Potter has lived under the stairs at his...
4   674             Harry Potter and the Goblet of Fire  Harry starts his fourth year at Hogwarts, comp...
PS F:\Shweta MSDA Sem1\Data 226 Datawarehousing\sjsu-data226-main\sjsu-data226-main\week11>
```

Git Link:

https://github.com/ShindeShwetaK/DW_Homework9/tree/main/week11