

Second Challenge for the Hash Agile Internship

Task of the second challenge:

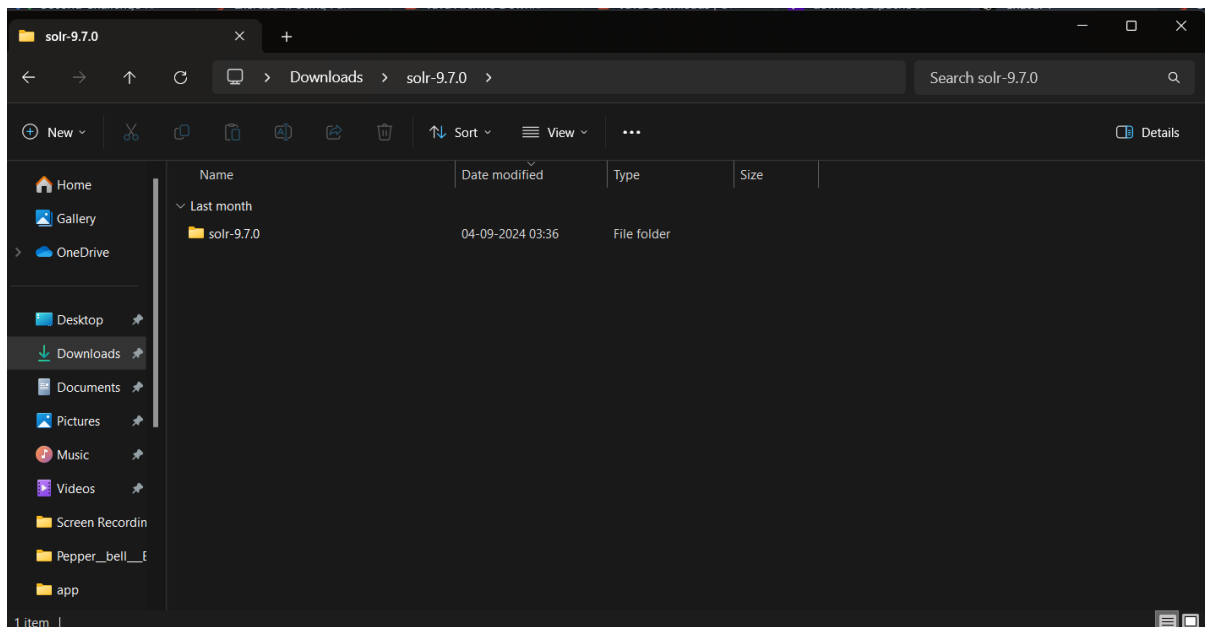
1. Read about Apache Solr: [Solr Tutorial](#).
2. Install Solr on your local machine.
3. Create a collection in Solr.
4. Index the Employee data
from <https://www.kaggle.com/datasets/williamlucas0/employee-sample-data>

Install Apache Solr on your local machine.

Step 1: Download Apache solr

1. Visit the Apache Solr website: Go to [Solr Downloads](#).
2. Download the latest version: Select the latest stable version (e.g., 9.0.0) and download the .tgz or .zip archive depending on your system.

Step 2: Extract the Downloaded file.



After the Extraction of the .tgz File. Want to navigate to the solr directory and run.

Windows : Open the command prompt, navigate to the solr directory

```
C:\Windows\system32\cmd.e: X + v
Microsoft Windows [Version 10.0.22631.4169]
(c) Microsoft Corporation. All rights reserved.

C:\Users\DHARANISH E>cd Downloads

C:\Users\DHARANISH E\Downloads>cd solr-9.7.0

C:\Users\DHARANISH E\Downloads\solr-9.7.0>cd solr-9.7.0

C:\Users\DHARANISH E\Downloads\solr-9.7.0\solr-9.7.0>bin\solr.cmd start
Java 17 detected. Enabled workaround for SOLR-16463
Java HotSpot(TM) 64-Bit Server VM warning: JVM cannot use large page memory because it does not have enough privilege to lock pages in memory.
WARNING: A command line option has enabled the Security Manager
WARNING: The Security Manager is deprecated and will be removed in a future release
Failed to parse command-line arguments due to: Unrecognized option: --max-wait-secs

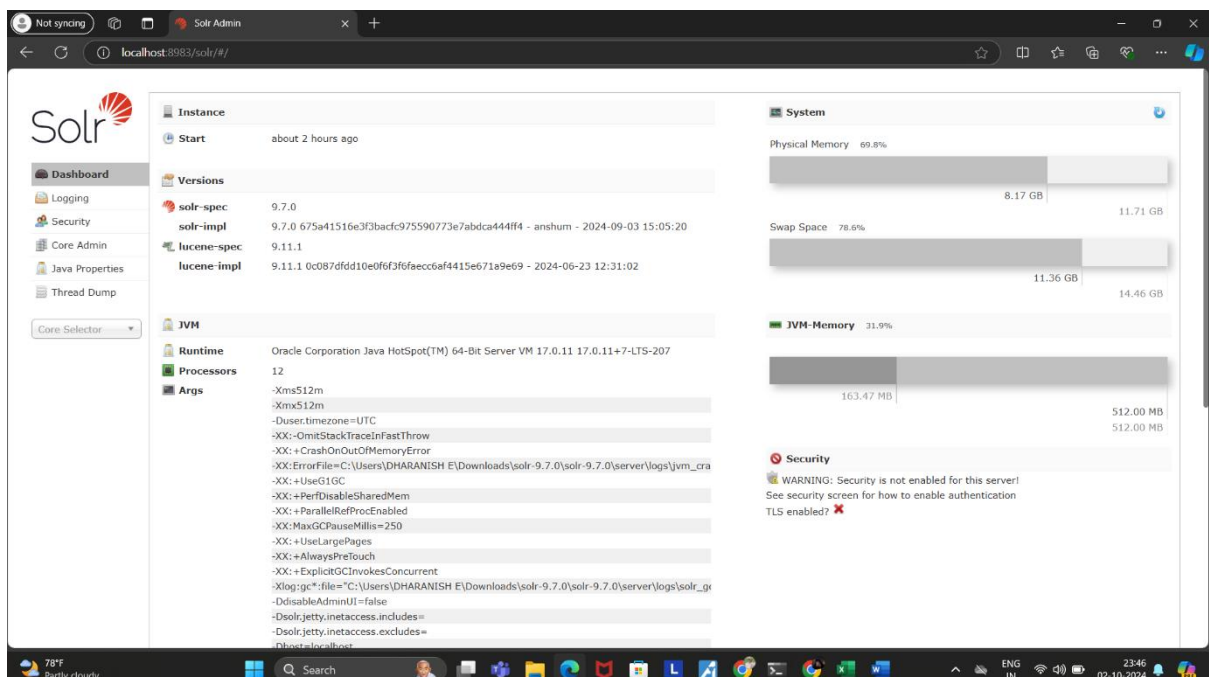
usage: bin/solr status [-maxWaitSecs <SECS>] [--solr-url <URL>]

List of options:
  -maxWaitSecs <SECS>  Wait up to the specified number of seconds to see Solr running.
  --solr-url <URL>     Address of the Solr Web application, defaults to: http://localhost:8983.

Please see the Reference Guide for more tools documentation:
https://solr.apache.org/guide/solr/latest/deployment-guide/solr-control-script-reference.html
```

Step 3: Verify the Installation:

Open your browser and go to <http://localhost:8983/solr>. The Solr Admin UI should appear, confirming that Solr is running.



Step 4: Create a collection

To create a collection where data will be stored:

```
C:\Windows\system32\cmd.e: X + v

C:\Users\DHARANISH E\Downloads\solr-9.7.0\solr-9.7.0>

C:\Users\DHARANISH E\Downloads\solr-9.7.0\solr-9.7.0>bin\solr.cmd create -c employee_data
WARNING: Using _default configset with data driven schema functionality. NOT RECOMMENDED for production use.
  To turn off: bin\solr config -c employee_data -p 8983 -action set-user-property -property update.autoCreateFields -value false
Option '-solrUrl','--solrUrl': Deprecatcd for removal since 9.7: Use --solr-url instead
WARNING: URLs provided to this tool needn't include Solr's context-root (e.g. "/solr"). Such URLs are deprecated and support for them will be removed in a future release. Correcting from [http://localhost:8983/solr] to [http://localhost:8983/].
Option '-confdir','--confdir': Deprecatcd for removal since 9.7: Use --conf-dir instead
Option '-confdir','--confdir': Deprecatcd for removal since 9.7: Use --conf-dir instead
Option '-confdir','--confdir': Deprecatcd for removal since 9.7: Use --conf-dir instead
Option '-confdir','--confdir': Deprecatcd for removal since 9.7: Use --conf-dir instead
Option '-confname','--confname': Deprecatcd for removal since 9.7: Use --conf-name instead

Created new core 'employee_data'
```

Step 5: Indexing and the Query Data

Once Solr is running, you can upload and index your data using the Solr Admin UI or via curl commands.

```
C:\Windows\system32\cmd.e: X + v

Created new core 'employee_data'

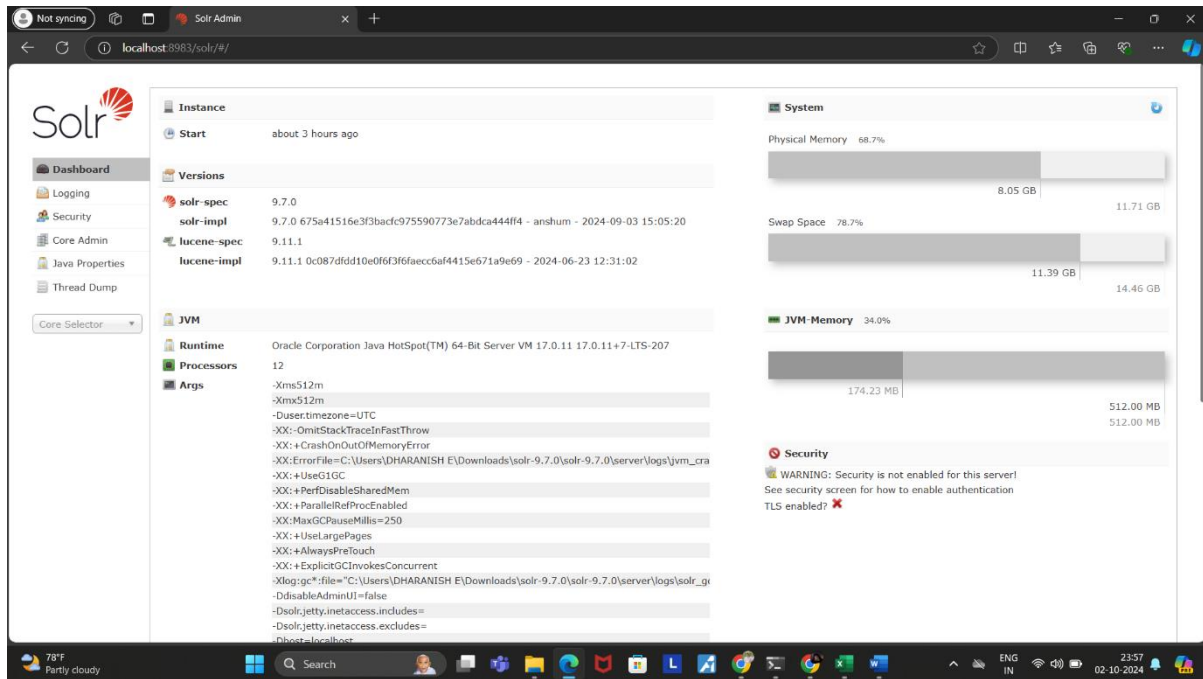
C:\Users\DHARANISH E\Downloads\solr-9.7.0\solr-9.7.0>bin\solr.cmd post -c employee employee_data.csv
Posting files to [base] url http://localhost:8983/solr/employee/update...
Entering auto mode. File endings considered are xml,json,jsonl,csv,pdf,doc,docx,ppt,pptx,xls,xlsx,odt,odp,ods,ott,otp,ots,rtf,htm,html,txt,log
PostTool: WARNING: No files or directories matching employee_data.csv
0 files indexed.
COMMITting Solr index changes to http://localhost:8983/solr/employee/update...
WARNING: URLs provided to this tool needn't include Solr's context-root (e.g. "/solr"). Such URLs are deprecated and support for them will be removed in a future release. Correcting from [http://localhost:8983/solr] to [http://localhost:8983/].

ERROR: Error from server at http://localhost:8983/solr/employee/update: Expected mime type in [application/vnd.apache.solr.javabin, application/octet-stream] but got text/html. <p>
  Searching for Solr?<br/>
  You must type the correct path.<br/>
  Solr will respond.
</p>
```

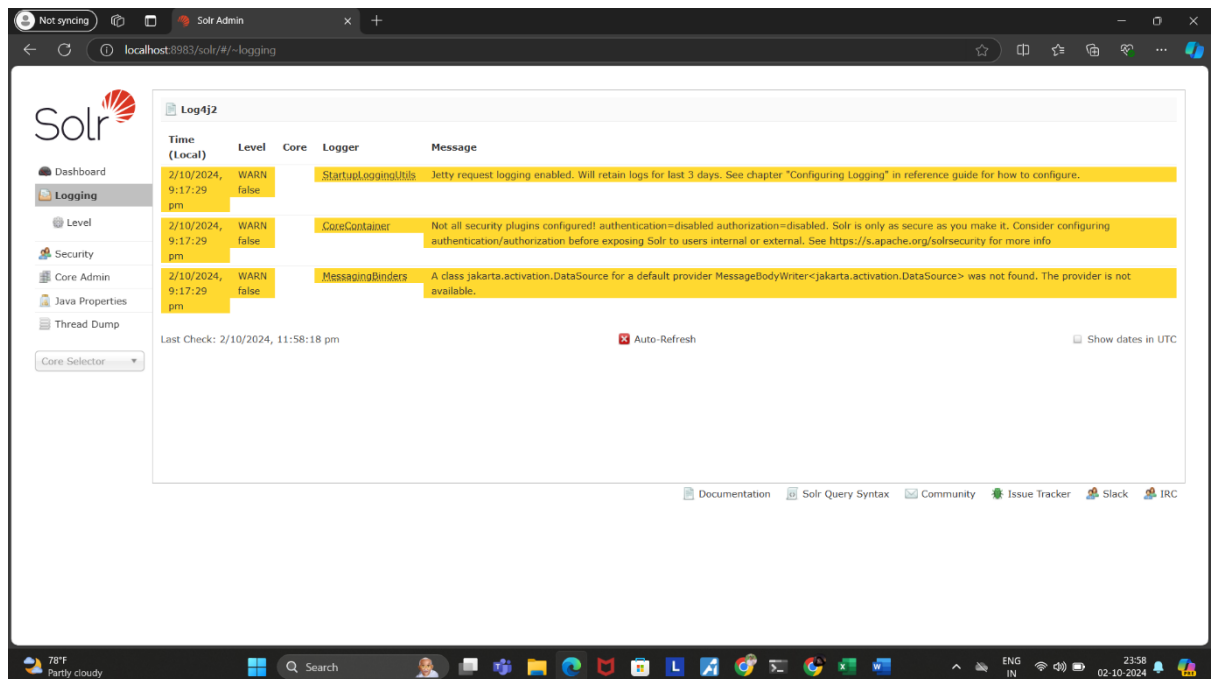
These are the step to Install, Extract, Create the table and Indexing the table.

Interface of the Solr:

1. Dashboard:

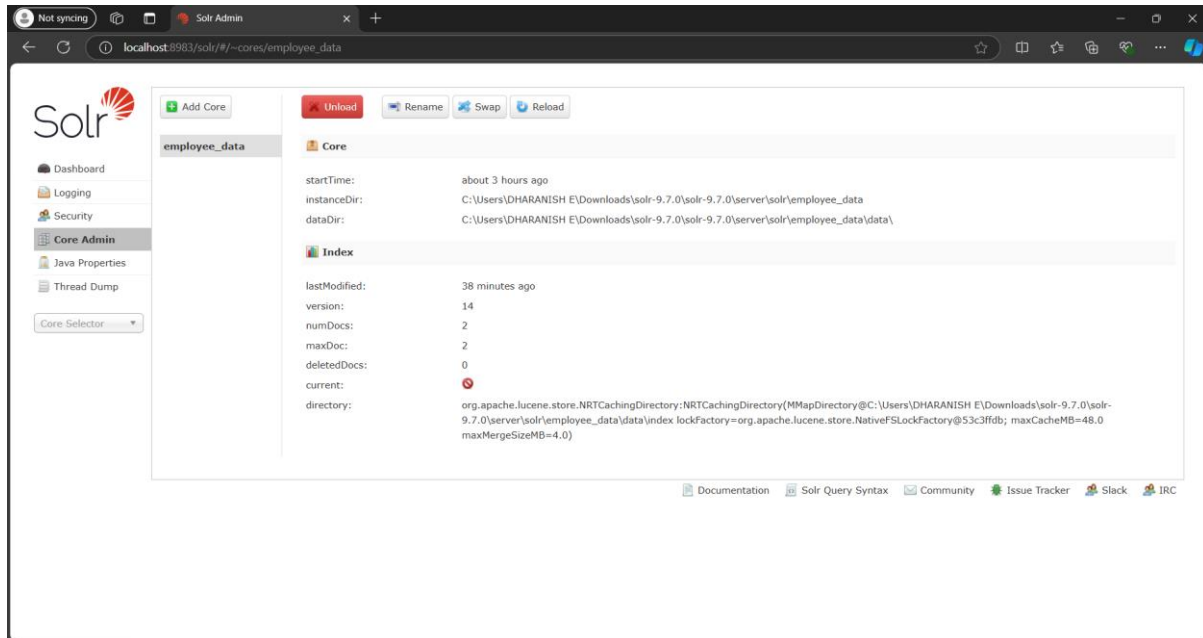


2. Logging:



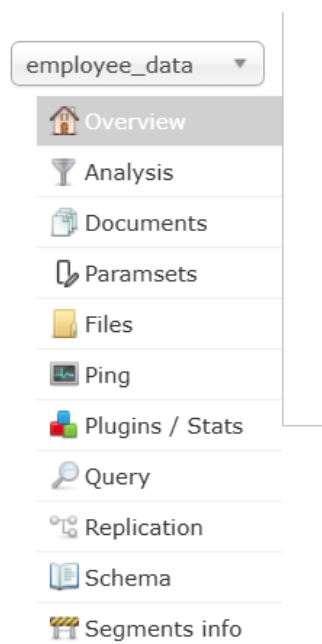
3. Admin UI:

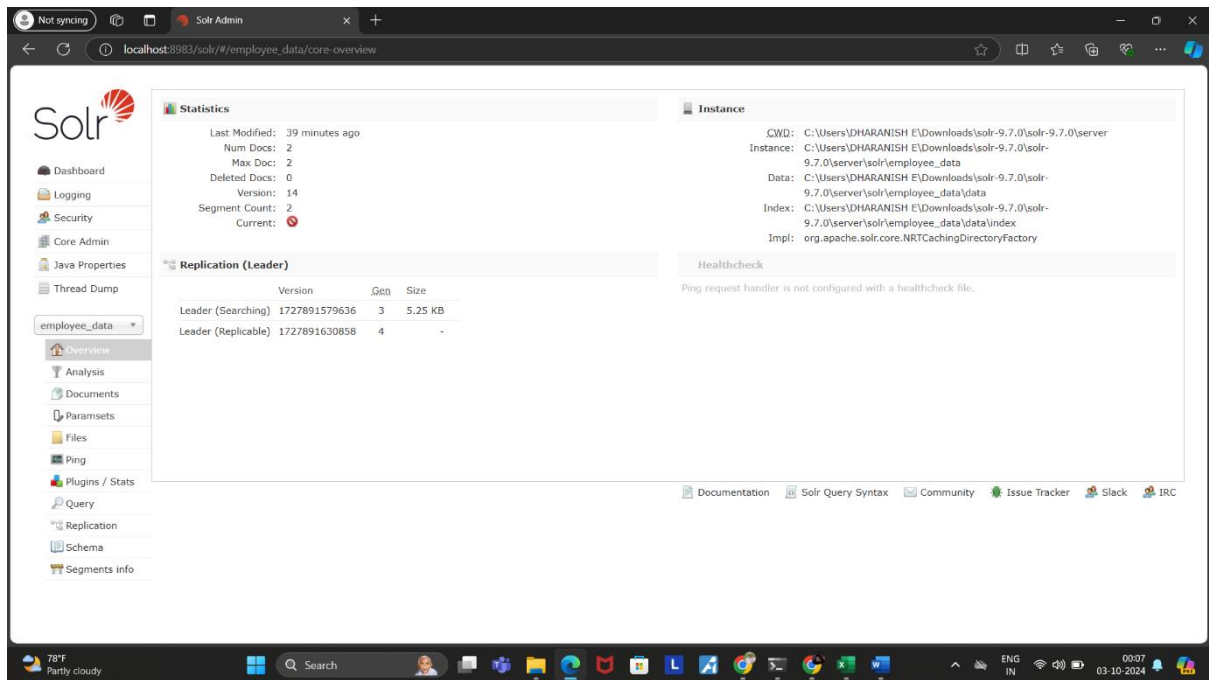
In This phase we can see the index of the data.



Core Selector in the Selecting the collection of the Created table.

Here, we can see the overall of the collection of the data, the indexing of the data. From there We can see the analysis , Plugins , Document and the Paramsets, etc.





To Stop the Solr:

ERROR: Must specify the port when trying to stop Solr, or use --all to stop all running host.

```
C:\Users\DHARANISH E\Downloads\solr-9.7.0\solr-9.7.0>bin\solr.cmd stop --all
Stopping Solr process 13020 running on port 8983
Waiting up to 180 seconds for process 13020 to exit
C:\Users\DHARANISH E\Downloads\solr-9.7.0\solr-9.7.0>
```

1. Stop Solr on port 8983: