DSC-650-Week-08-Mukherjee

Author : CHitramoy MUkherjee

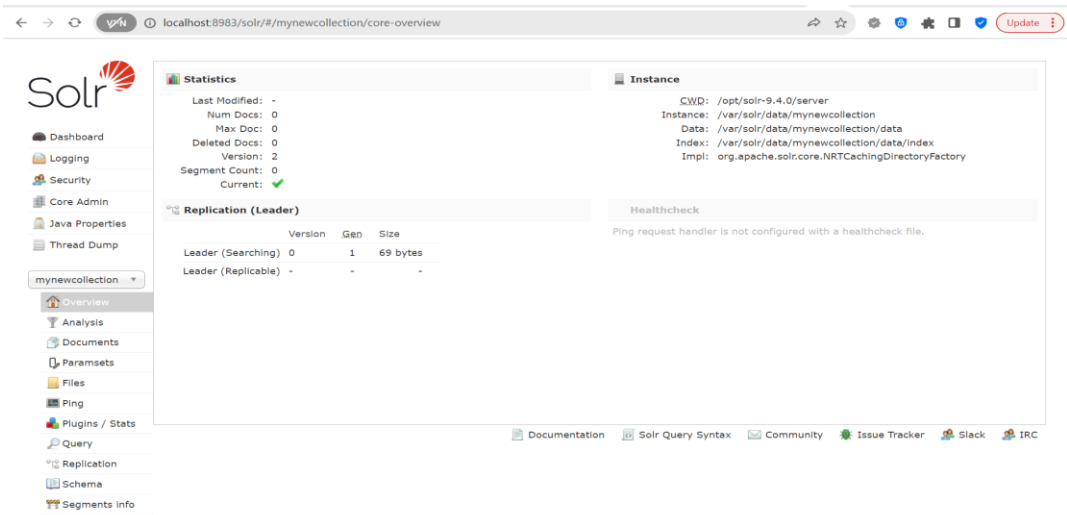
Date : 02/01/2024

Exercise 1: Create a Solr collection named ‘mynewcollection’.

Screenshot showing ‘mynewcollection’ in the Solr Web Interface.

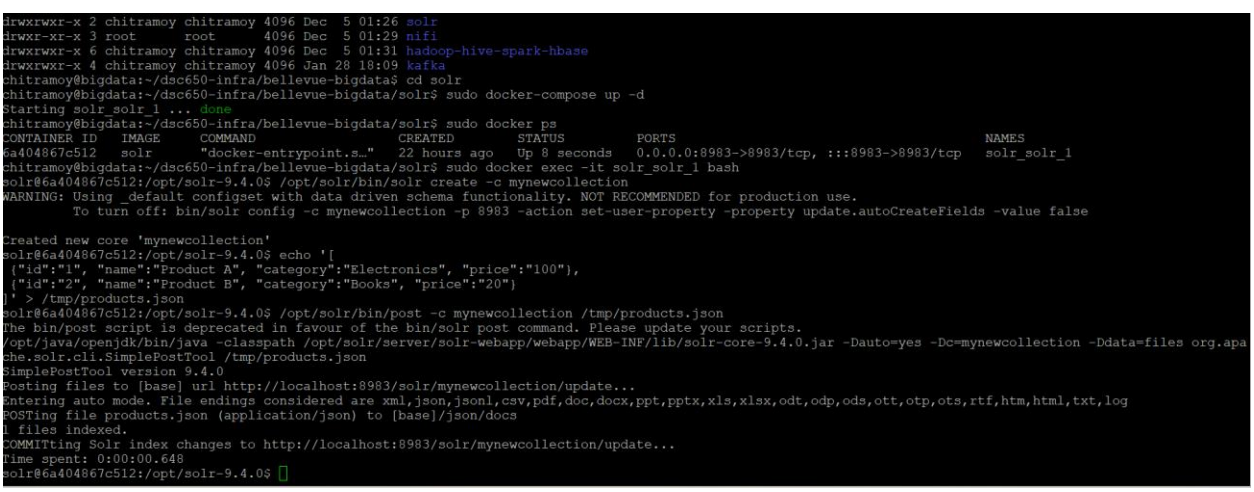
Exercise 2: Verify that ‘mynewcollection’ has been successfully created.

Access Solr’s Web Interface at http://localhost:8983/solr/ and check for ‘mynewcollection’ under the “Core Selector” dropdown.



Exercise 3: Let’s add the generated data to our collection.

Screenshot showing successful data ingestion messages in the terminal.



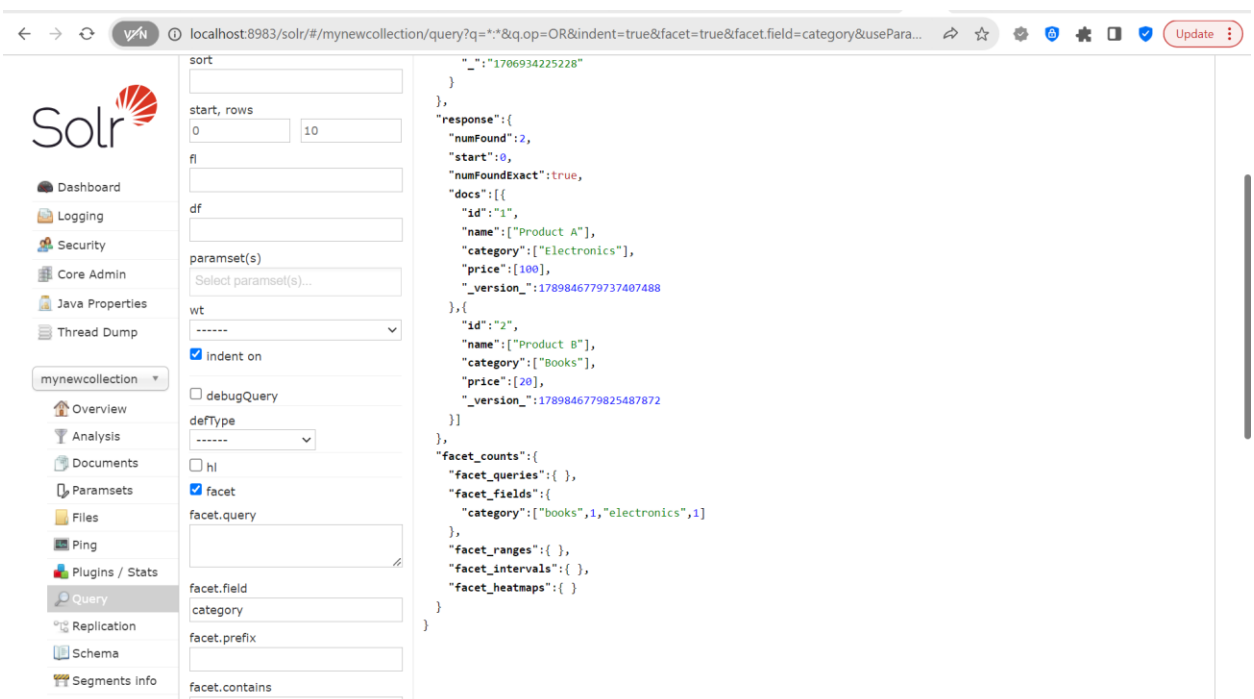
```
drwxrwxr-x 2 chitramoy chitramoy 4096 Dec  5 01:26 solr
drwxr-xr-x 3 root      root      4096 Dec  5 01:29 nifi
drwxrwxr-x 6 chitramoy chitramoy 4096 Dec  5 01:31 hadoop-hive-spark-hbase
drwxrwxr-x 4 chitramoy chitramoy 4096 Jan 28 18:09 kafka
chitramoy@bigdata:~/dsc650-infra/bellevue-bigdata$ cd solr
chitramoy@bigdata:~/dsc650-infra/bellevue-bigdata/solr$ sudo docker-compose up -d
Starting solr_solr_1 ... done
chitramoy@bigdata:~/dsc650-infra/bellevue-bigdata/solr$ sudo docker ps
CONTAINER ID   IMAGE     COMMAND                  CREATED        STATUS         PORTS                                       NAMES
6a404867c512   solr      "docker-entrypoint.s…"   22 hours ago   Up 8 seconds   0.0.0.0:8983->8983/tcp, :::8983->8983/tcp   solr_solr_1
chitramoy@bigdata:~/dsc650-infra/bellevue-bigdata/solr$ sudo docker exec -it solr_solr_1 bash
solr@6a404867c512:/opt/solr-9.4.0$ /opt/solr/bin/solr create -c mynewcollection
WARNING: Using _default configset with data driven schema functionality. NOT RECOMMENDED for production use.
         To turn off: bin/solr config -c mynewcollection -p 8983 -action set-user-property -property update.autoCreateFields -value false

Created new core 'mynewcollection'
solr@6a404867c512:/opt/solr-9.4.0$ echo '[
 {"id":"1", "name":"Product A", "category":"Electronics", "price":"100"},
 {"id":"2", "name":"Product B", "category":"Books", "price":"20"}
]' > /tmp/products.json
solr@6a404867c512:/opt/solr-9.4.0$ /opt/solr/bin/post -c mynewcollection /tmp/products.json
The bin/post script is deprecated in favour of the bin/solr post command. Please update your scripts.
/opt/java/openjdk/bin/java -classpath /opt/solr/server/solr-webapp/webapp/WEB-INF/lib/solr-core-9.4.0.jar -Dauto=yes -Dc=mynewcollection -Ddata=files org.apa
che.solr.cli.SimplePostTool /tmp/products.json
SimplePostTool version 9.4.0
Posting files to [base] url http://localhost:8983/solr/mynewcollection/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
POSTing file products.json (application/json) to [base]/json/docs
1 files indexed.
COMMITting Solr index changes to http://localhost:8983/solr/mynewcollection/update...
Time spent: 0:00:00.648
solr@6a404867c512:/opt/solr-9.4.0$
```

Exercise 4: Use Solr's query interface to retrieve all documents from 'mynewcollection'. curl "http://localhost:8983/solr/mynewcollection/select?q=*:*"

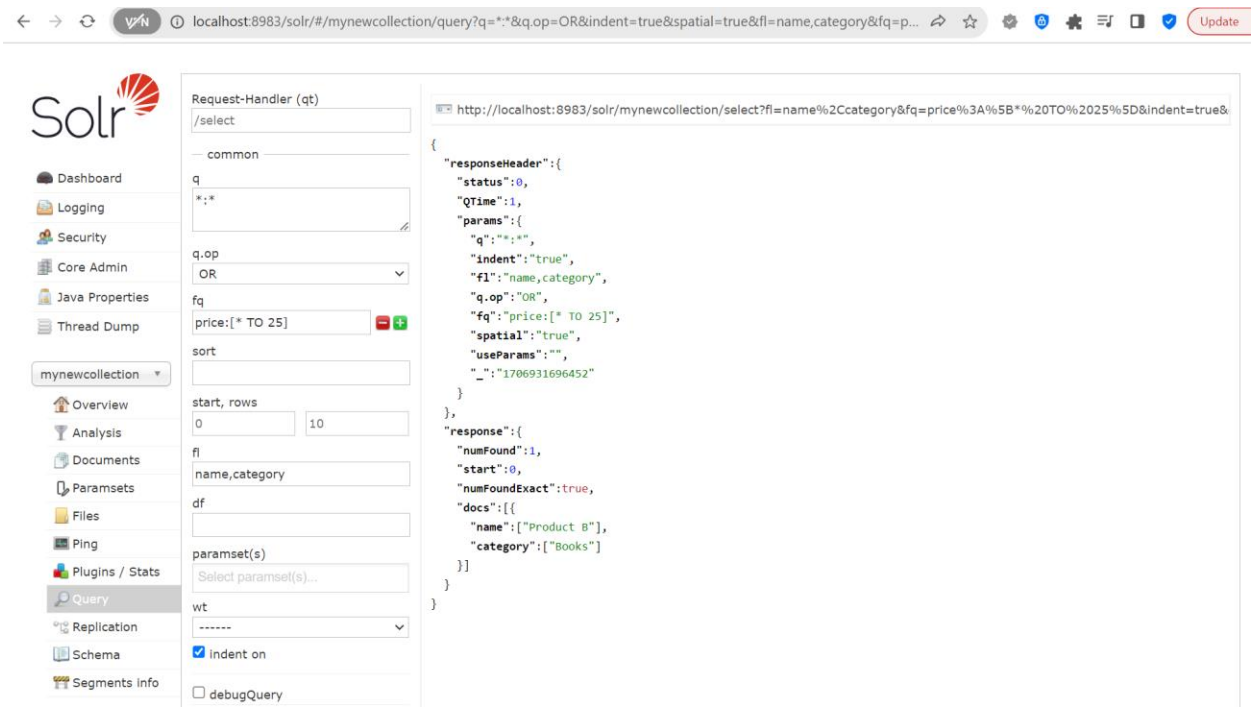
Exercise 5: Query for products in the 'Electronics' category. curl "http://localhost:8983/solr/mynewcollection/select?q=category:Electronics"

```
solr@6a404867c512:/opt/solr-9.4.0$ curl "http://localhost:8983/solr/mynewcollection/select?q=*:*"
{
  "responseHeader":{
    "status":0,
    "QTime":0,
    "params":{
      "q":"*:*"
    }
  },
  "response":{
    "numFound":2,
    "start":0,
    "numFoundExact":true,
    "docs":[{
      "id":"1",
      "name":["Product A"],
      "category":["Electronics"],
      "price":[100],
      "_version_":1789846779737407488
    },{
      "id":"2",
      "name":["Product B"],
      "category":["Books"],
      "price":[20],
      "_version_":1789846779825487872
    }]
  }
}solr@6a404867c512:/opt/solr-9.4.0$curl "http://localhost:8983/solr/mynewcollection/select?q=category:Electronics"
{
  "responseHeader":{
    "status":0,
    "QTime":1,
    "params":{
      "q":"category:Electronics"
    }
  },
  "response":{
    "numFound":1,
    "start":0,
    "numFoundExact":true,
    "docs":[{
      "id":"1",
      "name":["Product A"],
      "category":["Electronics"],
      "price":[100],
      "_version_":1789846779737407488
    }]
  }
}solr@6a404867c512:/opt/solr-9.4.0$
```

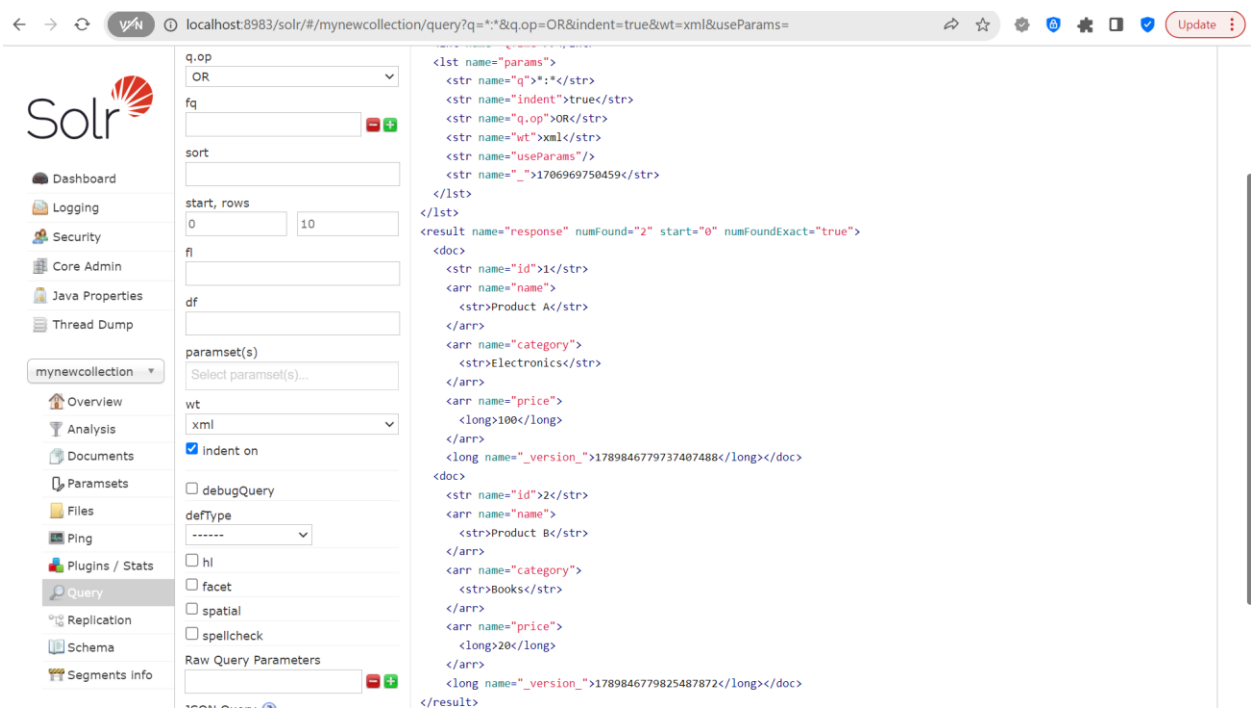
Exercise 6: Try a faceted search to count the number of products in each category.



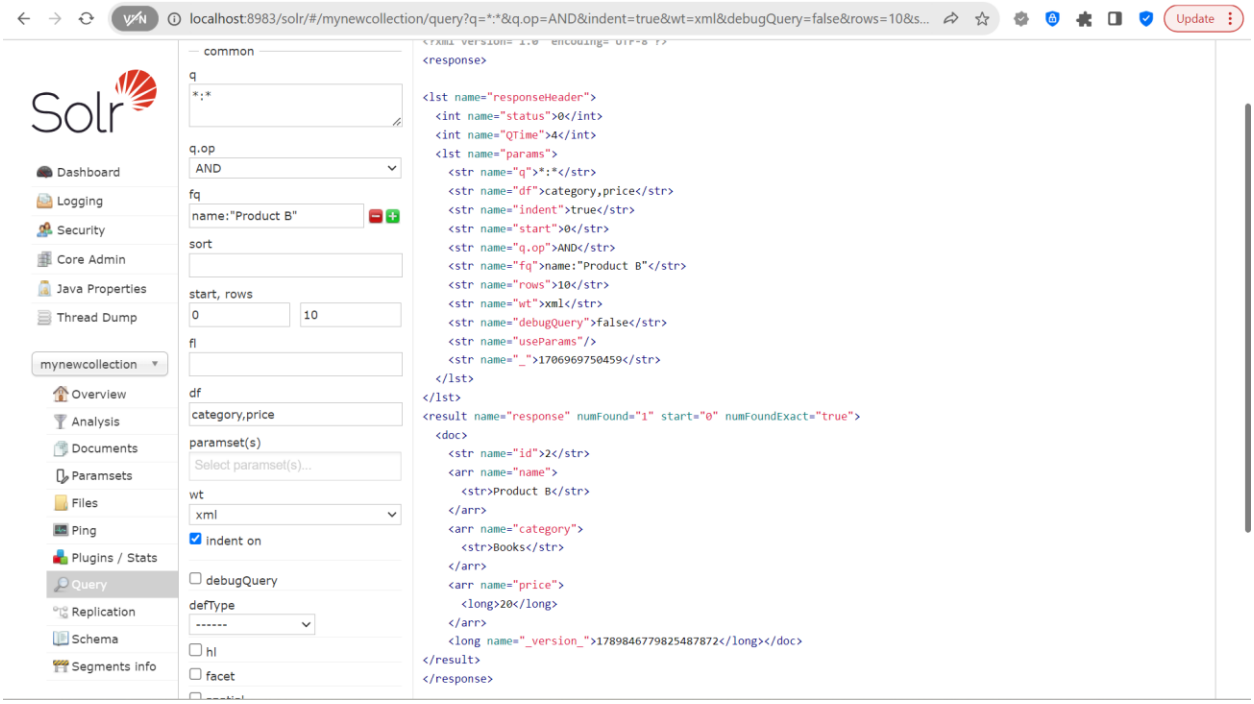
Select name, category where price less than 25.



Display data in xml format.



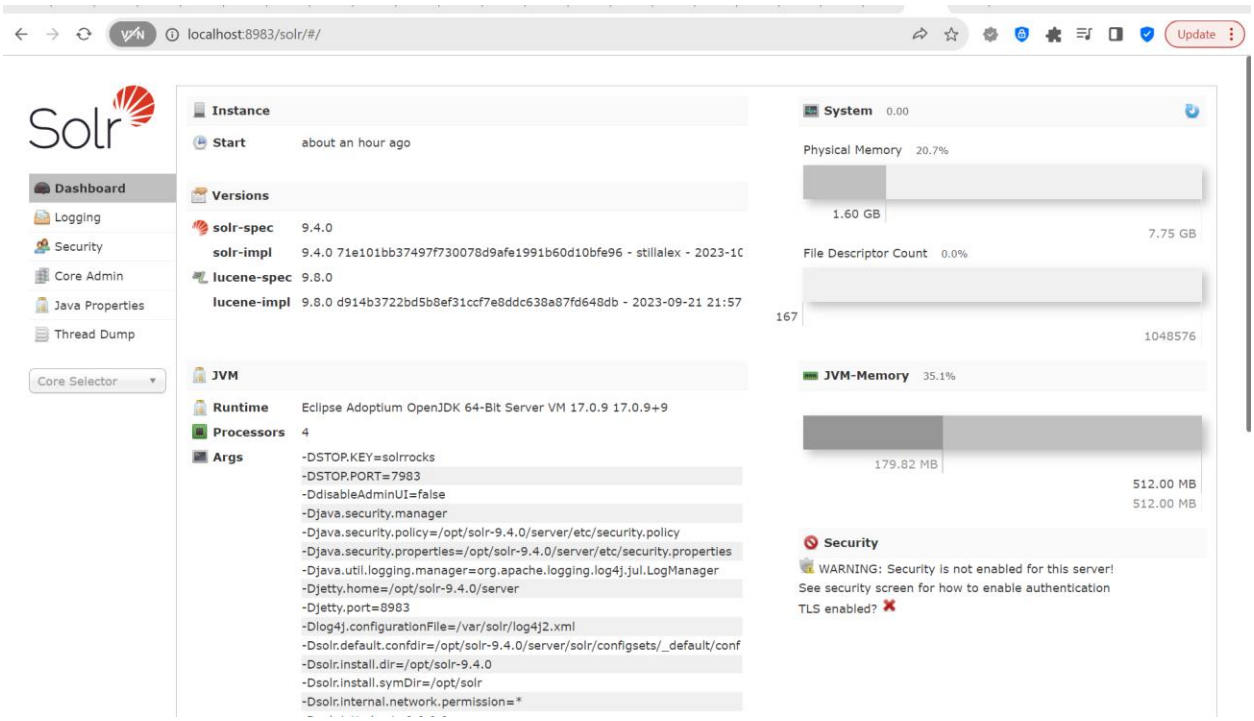
Select category, price for name:"Product B" and xml output.



Create a new collection in Solr mynewcollection_1.

```
}solr@6a404867c512:/opt/solr-9.4.0$/opt/solr/bin/solr create -c mynewcollection_1
WARNING: Using _default configset with data driven schema functionality. NOT RECOMMENDED for production use.
         To turn off: bin/solr config -c mynewcollection_1 -p 8983 -action set-user-property -property update.autoCreateFields -value false

Created new core 'mynewcollection_1'
solr@6a404867c512:/opt/solr-9.4.0$
```

Verify that ‘mynewcollection_1’ has been successfully created



Generate JSON data and add it to mycollection_1.

Structure of result.json file

```
echo '[
{"student_id":"1", "name":"John", "Grade":"A", "Result":"Pass"},
{"student_id":"2", "name":"Ron", "Grade":"A+", "Result":"Pass"},
{"student_id":"3", "name":"Andrew", "Grade":"B", "Result":"Pass"},
{"student_id":"4", "name":"Sophia", "Grade":"B", "Result":"Pass"},
{"student_id":"5", "name":"Jason", "Grade":"F", "Result":"Fail"},
{"student_id":"6", "name":"Mady", "Grade":"A", "Result":"Pass"},
{"student_id":"7", "name":"Grayson", "Grade":"F", "Result":"Fail"}
]' > /tmp/result.json
```

```
Created new core 'mynewcollection_1'
solr@6a404867c512:/opt/solr-9.4.0$ echo '[
 {"student_id":"1", "name":"John", "Grade":"A", "Result":"Pass"},
 {"student_id":"2", "name":"Ron", "Grade":"A+", "Result":"Pass"},
 {"student_id":"3", "name":"Andrew", "Grade":"B", "Result":"Pass"},
 {"student_id":"4", "name":"Sophia", "Grade":"B", "Result":"Pass"},
 {"student_id":"5", "name":"Jason", "Grade":"F", "Result":"Fail"},
 {"student_id":"6", "name":"Mady", "Grade":"A", "Result":"Pass"},
 {"student_id":"7", "name":"Grayson", "Grade":"F", "Result":"Fail"}
]' > /tmp/result.json
solr@6a404867c512:/opt/solr-9.4.0$ /opt/solr/bin/post -c mynewcollection_1 /tmp/result.json
The bin/post script is deprecated in favour of the bin/solr post command. Please update your scripts.
/opt/java/openjdk/bin/java -classpath /opt/solr/server/solr-webapp/webapp/WEB-INF/lib/solr-core-9.4.0.jar -Dauto=yes -Dc=mynewcollection_1 -Ddata=files org.a
pache.solr.cli.SimplePostTool /tmp/result.json
SimplePostTool version 9.4.0
Posting files to [base] url http://localhost:8983/solr/mynewcollection_1/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
POSTing file result.json (application/json) to [base]/json/docs
1 files indexed.
COMMITting Solr index changes to http://localhost:8983/solr/mynewcollection_1/update...
Time spent: 0:00:00.304
solr@6a404867c512:/opt/solr-9.4.0$
```

Query your collection from command line and web interface :

```
chitramoy@bigdata:~/dsc650-infra/bellevue-bigdata/solr$ docker exec -it solr_solr_1 bash
solr@6a404867c512:/opt/solr-9.4.0$ curl "http://localhost:8983/solr/mynewcollection_1/select?q=*:*"
{
  "responseHeader":{
    "status":0,
    "QTime":1,
    "params":{
      "q":"*:*"
    }
  },
  "response":{
    "numFound":7,
    "start":0,
    "numFoundExact":true,
    "docs":[{
      "student_id":[1],
      "name":["John"],
      "Grade":["A"],
      "Result":["Pass"],
      "id":"ca971317-fde8-4a3e-a29e-36c43622453c",
      "_version_":1789849572365828096
    },{
      "student_id":[2],
      "name":["Ron"],
      "Grade":["A+"],
      "Result":["Pass"],
      "id":"bd9a6560-6b18-412b-9651-1208b93aca5b",
      "_version_":1789849572371070976
    },{
      "student_id":[3],
      "name":["Andrew"],
      "Grade":["B"],
      "Result":["Pass"],
      "id":"53185e92-53f6-49d1-a749-eff9f63569b9",
      "_version_":1789849572372119552
    },{
      "student_id":[4],
      "name":["Sophia"],
      "Grade":["B"],
      "Result":["Pass"],
      "id":"2439635f-00ac-4adf-a00d-15aa27d2996b",
      "_version_":1789849572373168128
```

```
      "Grade":["A"],
      "Result":["Pass"],
      "id":"ca971317-fde8-4a3e-a29e-36c43622453c",
      "_version_":1789849572365828096
    },{
      "student_id":[2],
      "name":["Ron"],
      "Grade":["A+"],
      "Result":["Pass"],
      "id":"bd9a6560-6b18-412b-9651-1208b93aca5b",
      "_version_":1789849572371070976
    },{
      "student_id":[3],
      "name":["Andrew"],
      "Grade":["B"],
      "Result":["Pass"],
      "id":"53185e92-53f6-49d1-a749-eff9f63569b9",
      "_version_":1789849572372119552
    },{
      "student_id":[4],
      "name":["Sophia"],
      "Grade":["B"],
      "Result":["Pass"],
      "id":"2439635f-00ac-4adf-a00d-15aa27d2996b",
      "_version_":1789849572373168128
    },{
      "student_id":[5],
      "name":["Jason"],
      "Grade":["F"],
      "Result":["Fail"],
      "id":"226a7293-51ee-44d5-87e7-9a27bf33b4ff",
      "_version_":1789849572374216704
    },{
      "student_id":[6],
      "name":["Mady"],
      "Grade":["A"],
      "Result":["Pass"],
      "id":"05941724-5b8e-490b-86d5-b427525a180b",
      "_version_":1789849572374216705
    },{
      "student_id":[7],
      "name":["Grayson"],
      "Grade":["F"],
      "Result":["Fail"],
      "id":"dbc961c0-0a56-47ac-a470-0ae35d5ce843",
      "_version_":1789849572375265280
    }]
  }
}solr@6a404867c512:/opt/solr-9.4.0$
```

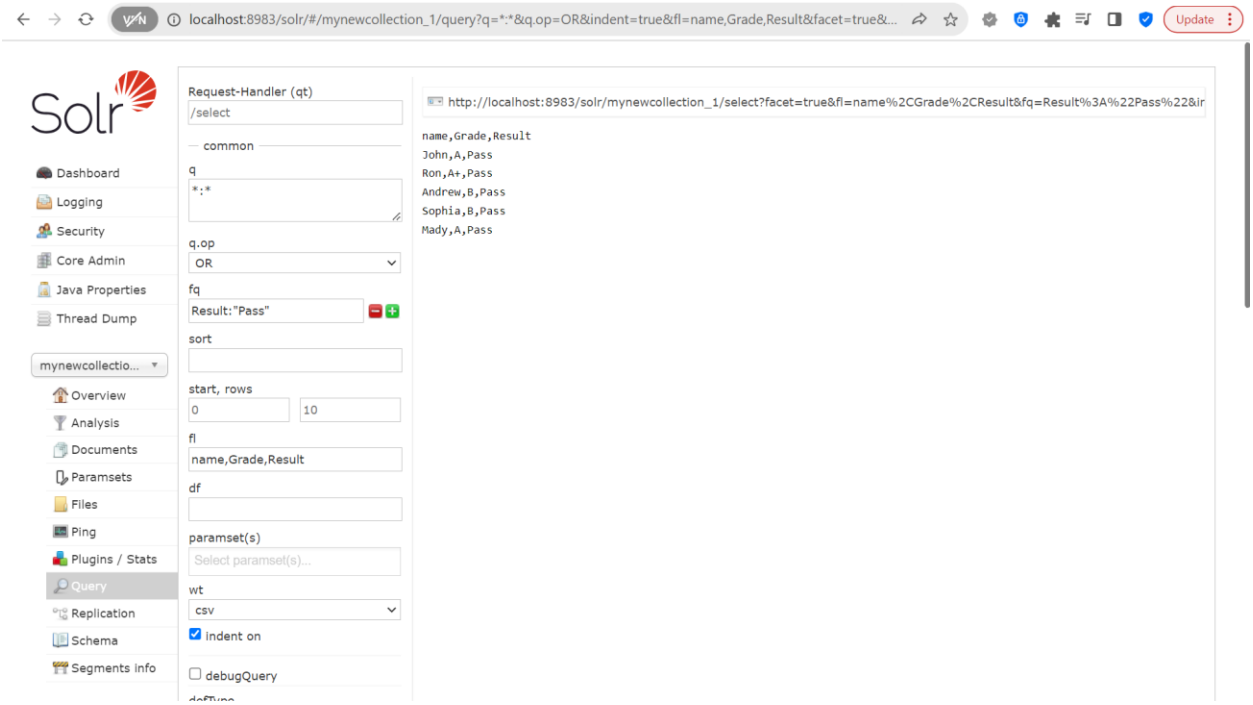

Query your collection from command line and web interface :

Select where name is Ron and select * where Garde = ‘B’.

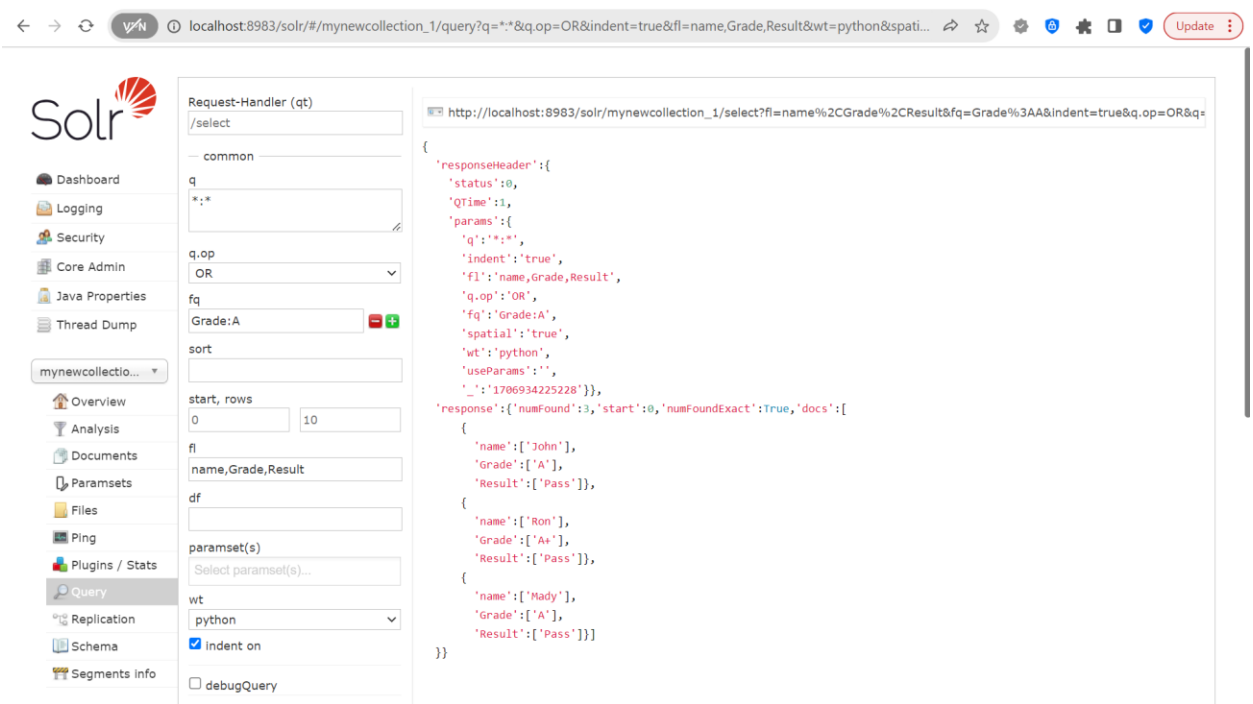

```
solr@6a404867c512:/opt/solr-9.4.0$ curl "http://localhost:8983/solr/mynewcollection_1/select?q=name:Ron"
{
  "responseHeader":{
    "status":0,
    "QTime":1,
    "params":{
      "q":"name:Ron"
    }
  },
  "response":{
    "numFound":1,
    "start":0,
    "numFoundExact":true,
    "docs":[{
      "student_id":[2],
      "name":["Ron"],
      "Grade":["A+"],
      "Result":["Pass"],
      "id":"bd9a6560-6b18-412b-9651-1208b93aca5b",
      "_version_":1789849572371070976
    }]
  }
}solr@6a404867c512:/opt/solr-9.4.0$curl "http://localhost:8983/solr/mynewcollection_1/select?q=Grade:B"
{
  "responseHeader":{
    "status":0,
    "QTime":2,
    "params":{
      "q":"Grade:B"
    }
  },
  "response":{
    "numFound":2,
    "start":0,
    "numFoundExact":true,
    "docs":[{
      "student_id":[3],
      "name":["Andrew"],
      "Grade":["B"],
      "Result":["Pass"],
      "id":"53185e92-53f6-49d1-a749-eff9f63569b9",
      "_version_":1789849572372119552
    },{
      "student_id":[4],
      "name":["Sophia"],
      "Grade":["B"],
      "Result":["Pass"],
      "id":"2439635f-00ac-4adf-a00d-15aa27d2996b",
      "_version_":1789849572373168128
```


Query your collection from web interface:

Select name, Grade, Result from mynewcollection_1 where Result:”Pass” and extract in csv format.



Select name, Grade, Result from mynewcollection_1 where Grade = ‘A’ and extract in python format (df).



Grade wise count from from mynewcollection_1.

localhost:8983/solr/#/mynewcollection_1/query?q=*:*&q.op=AND&indent=true&facet=true&facet.field=Grade&wt=json...

Solr

- Dashboard
- Logging
- Security
- Core Admin
- Java Properties
- Thread Dump

mynewcollectio...

- Overview
- Analysis
- Documents
- Paramsets
- Files
- Ping
- Plugins / Stats
- Query
- Replication
- Schema
- Segments Info

facet.contains

☐ facet.contains.ignoreCase

facet.limit

facet.matches

facet.sort

---

facet.mincount

☐ facet.missing

json.facet

☑ spatial

pt

sfield

d


```
    "Grade":["B"],
    "Result":["Pass"],
    "id":"2439635f-00ac-4adf-a00d-15aa27d2996b",
    "_version_":1789849572373168128
  },{
    "student_id":[5],
    "name":["Jason"],
    "Grade":["F"],
    "Result":["Fail"],
    "id":"226a7293-51ee-44d5-87e7-9a27bf33b4ff",
    "_version_":1789849572374216704
  },{
    "student_id":[6],
    "name":["Mady"],
    "Grade":["A"],
    "Result":["Pass"],
    "id":"05941724-5b8e-490b-86d5-b427525a180b",
    "_version_":1789849572374216705
  },{
    "student_id":[7],
    "name":["Grayson"],
    "Grade":["F"],
    "Result":["Fail"],
    "id":"dbc961c0-0a56-47ac-a470-0ae35d5ce843",
    "_version_":1789849572375265280
  }]
},
"facet_counts":{
  "facet_queries":{ },
  "facet_fields":{
    "Grade":["a",3,"b",2,"f",2]
  },
  "facet_ranges":{ },
  "facet_intervals":{ },
  "facet_heatmaps":{ }
```