

DSC 650

Inman, Gracie

Week 8

5/524

- New Collection in Solr Web Interface

```
gracieinman@big-data-week1-instance:~/dsc650-infra-main/bellevue-bigdata$ cd solr
gracieinman@big-data-week1-instance:~/dsc650-infra-main/bellevue-bigdata/solr$ docker compose up -d
[*] Running 1/2
  Network solr_default Created
  Container solr-solr-1 Started
gracieinman@big-data-week1-instance:~/dsc650-infra-main/bellevue-bigdata/solr$ docker exec -it solr-solr-1 bash
Error response from daemon: No such container: solr-solr-1
gracieinman@big-data-week1-instance:~/dsc650-infra-main/bellevue-bigdata/solr$ docker exec -it solr-solr-1 bash
solr@1668d7cc42f7:/opt/solr-9.5.0$ /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@1668d7cc42f7:/opt/solr-9.5.0$
```

The screenshot displays the Solr Admin web interface for the 'mynewcollection' core. The interface is divided into a sidebar on the left and a main content area. The sidebar contains navigation links for various Solr components. The main content area shows the 'Overview' page for the 'mynewcollection' core, which includes a 'Statistics' section with details like 'Last Modified', 'Num Docs', 'Max Doc', 'Deleted Docs', 'Version', 'Segment Count', and 'Current' status. Below this is a 'Replication (Leader)' table showing the leader's version, generation, and size. A 'Healthcheck' section at the bottom indicates that the ping request handler is not configured with a healthcheck file. The browser's address bar shows the URL 'localhost:8983/solr/#/mynewcollection/core-overview'.

- Successful data ingestion

```
solr@1668d7cc42f7:/opt/solr-9.5.0$ /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-9.5.0.jar -Dautoyes -Dc=mynewcollection -Ddata=files org.apache.solr.cli.SimplePostTool /tmp/products.json
SimplePostTool version 9.5.0
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,ptx,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: 0:00:00.549
solr@1668d7cc42f7:/opt/solr-9.5.0$ █
```

- Queries

```
Time spent: 0:00:00.00v
solr@1668d7cc42f7:/opt/solr-9.5.0$ curl "http://localhost:8983/solr/mynewcollection/select?q=*:*"
{
  "responseHeader":{
    "status":0,
    "QTime":30,
    "params":{"q":"*:*"}
  },
  "response":{
    "numFound":2,
    "start":0,
    "numFoundExact":true,
    "docs":[{
      "id":"1",
      "name":["Product A"],
      "category":["Electronics"],
      "price":100,
      "_version_":1798232010840279040
    },{
      "id":"2",
      "name":["Product B"],
      "category":["Books"],
      "price":20,
      "_version_":1798232010162962432
    }]
  }
}
solr@1668d7cc42f7:/opt/solr-9.5.0$ curl "http://localhost:8983/solr/mynewcollection/select?q=category:Electronics"
{
  "responseHeader":{
    "status":0,
    "QTime":20,
    "params":{"q":"category:Electronics"}
  },
  "response":{
    "numFound":1,
    "start":0,
    "numFoundExact":true,
    "docs":[{
      "id":"1",
      "name":["Product A"],
      "category":["Electronics"],
      "price":100,
      "_version_":1798232010840279040
    }]
  }
}
solr@1668d7cc42f7:/opt/solr-9.5.0$ █
```

- Count number of products

The screenshot displays the Solr Admin web interface in a browser. The left sidebar contains navigation links: Dashboard, Logging, Security, Core Admin, Java Properties, Thread Dump, mynewcollection (selected), Overview, Analysis, Documents, Parameters, Files, Ping, Plugins / Stats, Replication, Schema, and Segments info.

The main content area is divided into two panels. The left panel, titled 'Request-Handler (q)', shows the configuration for the `/select` handler. Key settings include: `q` (set to `*:*`), `q.op` (set to `OR`), `fq` (empty), `sort` (empty), `start` (0), `rows` (10), `df` (empty), `parameters` (empty), `wt` (set to `json`), `indent` (checked), `debugQuery` (unchecked), `facet` (checked), `facet.field` (set to `category`), `facet.prefix` (empty), `facet.contains` (empty), `facet.contains.ignoreCase` (unchecked), `facet.limit` (empty), `facet.matches` (empty), `facet.sort` (empty), and `facet.mincount` (empty).

The right panel displays the JSON response for the query `http://localhost:8983/solr/mynewcollection/select?facet.field=category&facet=true&indent=true&q.op=OR&q=*:*&useParams=`. The response is a JSON object with the following structure:

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 17,
    "params": {
      "q": "*:*",
      "facet.field": "category",
      "indent": "true",
      "q.op": "OR",
      "facet": "true",
      "useParams": ""
    }
  },
  "response": {
    "numFound": 2,
    "start": 0,
    "numFoundExact": true,
    "docs": [
      {
        "id": "1",
        "name": "Product A",
        "category": "Electronics",
        "price": 100,
        "_version_": 1790232018048279040
      },
      {
        "id": "2",
        "name": "Product B",
        "category": "Books",
        "price": 20,
        "_version_": 1790232018162962432
      }
    ]
  },
  "facet_counts": {
    "facet_queries": {
      "facet_fields": {
        "category": [
          "Books", 1,
          "Electronics", 1
        ]
      },
      "facet_ranges": {
      },
      "facet_intervals": {
      },
      "facet_histograms": {
      }
    }
  }
}
```

- Select for Product A

The screenshot shows the Solr Admin interface in a web browser. The browser's address bar displays the URL: `localhost:8983/solr/#/mynewcollection/query?q=%3D%22name%3A%22Product%22%26useParams=true&fq.op=OR&fq.op%3A%3A%22name%3A%22Product%22%26useParams=true`. The Solr Admin page has a left sidebar with navigation links: Dashboard, Logging, Security, Core Admin, Java Properties, Thread Dump, mynewcollection (selected), Overview, Analysis, Documents, Parameters, Files, Ping, Plugins / Stats, Queries (selected), Replication, Schema, and Segments info.

The main content area is divided into two panels. The left panel, titled 'Request-Handler (q)', contains the following fields:

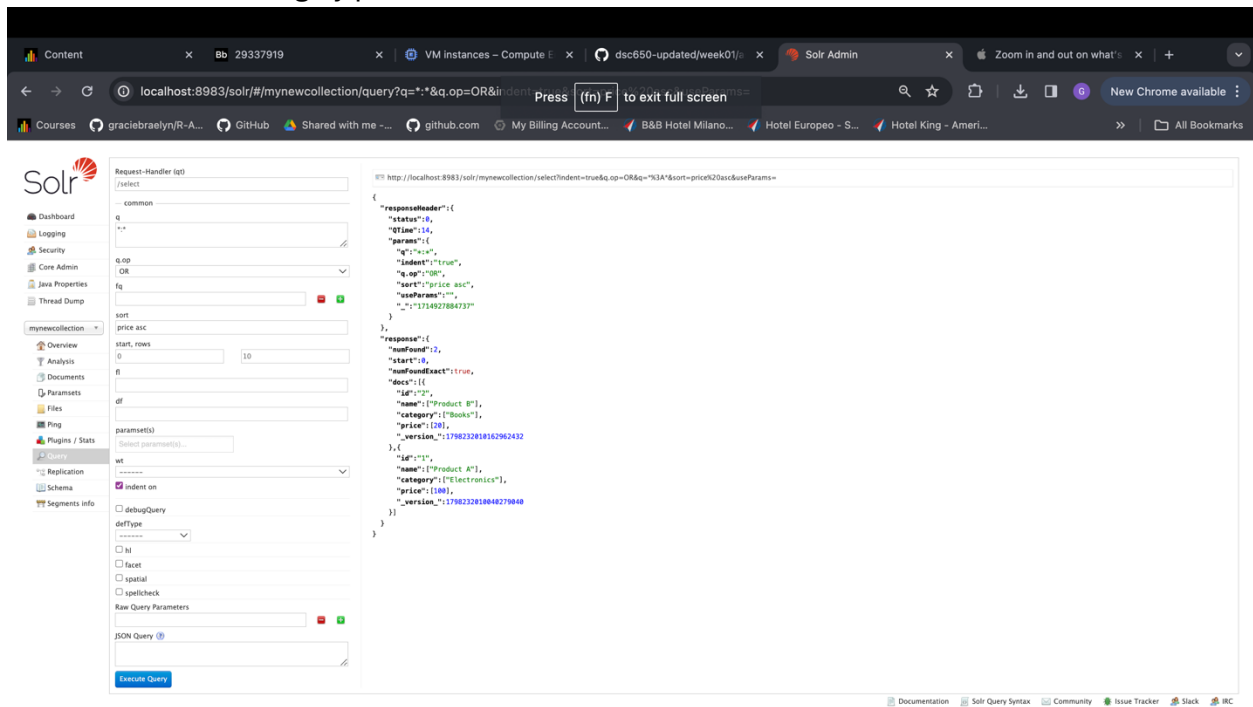
- `/select`
- `common`
- `q`: `name="Product A"`
- `q.op`: `OR`
- `fq`
- `sort`
- `start, rows`: `0, 10`
- `df`
- `parameters()`: `(select parameters())`
- `wt`: `json`
- `indent on`: ☒
- `debugQuery`: ☐
- `defType`: `lucene`
- `hl`: ☐
- `facet`: ☐
- `spatial`: ☐
- `spellcheck`: ☐
- `Raw Query Parameters`
- `JSON Query`
- `Execute Query` button

The right panel displays the JSON response from the query:

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 1,
    "params": {
      "q": "name='Product A'",
      "indent": "true",
      "q.op": "OR",
      "useParams": "true",
      "wt": "json"
    }
  },
  "response": {
    "numFound": 0,
    "start": 0,
    "numFoundExact": true,
    "docs": []
  }
}
```

At the bottom of the page, there are links to Documentation, Solr Query Syntax, Community, Issue Tracker, Slack, and IRC.

- Sort ascending by price



For the on-my-own activity, I obtained a dataset from Kaggle. I loaded the data directly into the Solr web interface using the documents tab. The following images are of the queries.

Reference:

Tatman, R. (2018, April 6). *Iris dataset (JSON version)*. Kaggle.
<https://www.kaggle.com/datasets/ratatman/iris-dataset-json-version?resource=download>

- Query all data

The screenshot shows the Solr Admin web interface in a browser. The address bar displays the URL: `localhost:8983/solr/#/iris_collection/query?q=*:*&q.op=OR&indent=true&useParams=`. The left sidebar contains navigation links for Dashboard, Logging, Security, Core Admin, Java Properties, Thread Dump, and the selected **iris_collection**. Under **iris_collection**, there are links for Overview, Analysis, Documents, Parameters, Files, Ping (live), Plugins / Stats, Query (selected), Replication, Schema, and Segments info.

The main content area is divided into two panels. The left panel, titled "Request-Handler (q)", contains query configuration options: `/select`, `common`, `q` (set to `*:*`), `q.op` (set to `OR`), `fq` (empty), `sort` (empty), `start, rows` (set to `0, 10`), `df` (empty), `params` (empty), `wt` (set to `json`), `indent on` (checked), `debugQuery` (unchecked), `defType` (set to `lucene`), `hl` (unchecked), `facet` (unchecked), `spatial` (unchecked), `spellcheck` (unchecked), `Raw Query Parameters` (empty), and `JSON Query` (checked). A blue "Execute Query" button is at the bottom.

The right panel displays the JSON response of the query. The URL in the browser's address bar is updated to include the query parameters: `http://localhost:8983/solr/iris_collection/select?indent=true&q.op=OR&q=*:*&useParams=`. The JSON response is as follows:

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 8,
    "params": {
      "q": "*:*",
      "q.op": "OR",
      "useParams": true,
      "indent": "true",
      "wt": "json",
      "df": "0",
      "fq": "",
      "sort": "",
      "start": "0",
      "rows": "10",
      "defType": "lucene",
      "hl": "false",
      "facet": "false",
      "spatial": "false",
      "spellcheck": "false",
      "rawQueryParameters": ""
    }
  },
  "response": {
    "numFound": 600,
    "start": 0,
    "numFoundExact": true,
    "docs": [
      {
        "sepalLength": 5.1,
        "sepalWidth": 3.5,
        "petalLength": 1.4,
        "petalWidth": 0.2,
        "species": "setosa",
        "id": "a9f74600-1b11-467c-b9f2-baa72864139f",
        "_version_": 1798236243716218868
      },
      {
        "sepalLength": 4.9,
        "sepalWidth": 3.0,
        "petalLength": 1.4,
        "petalWidth": 0.2,
        "species": "setosa",
        "id": "c4dc6eb-a26a-485d-adb6-46671d9f7228",
        "_version_": 1798236243713558720
      },
      {
        "sepalLength": 4.7,
        "sepalWidth": 3.2,
        "petalLength": 1.3,
        "petalWidth": 0.2,
        "species": "setosa",
        "id": "c4dc6eb-a26a-485d-adb6-46671d9f7228",
        "_version_": 1798236243713558720
      },
      {
        "sepalLength": 4.6,
        "sepalWidth": 3.1,
        "petalLength": 1.3,
        "petalWidth": 0.2,
        "species": "setosa",
        "id": "d5f57845-6cea-4739-a38d-c2d3e8aeb05f",
        "_version_": 1798236243713558720
      },
      {
        "sepalLength": 5.0,
        "sepalWidth": 3.6,
        "petalLength": 1.4,
        "petalWidth": 0.2,
        "species": "setosa",
        "id": "d5f57845-6cea-4739-a38d-c2d3e8aeb05f",
        "_version_": 1798236243713558720
      },
      {
        "sepalLength": 5.4,
        "sepalWidth": 4.4,
        "petalLength": 1.5,
        "petalWidth": 0.2,
        "species": "setosa",
        "id": "d5f57845-6cea-4739-a38d-c2d3e8aeb05f",
        "_version_": 1798236243713558720
      },
      {
        "sepalLength": 5.0,
        "sepalWidth": 3.5,
        "petalLength": 1.6,
        "petalWidth": 0.4,
        "species": "setosa",
        "id": "d5f57845-6cea-4739-a38d-c2d3e8aeb05f",
        "_version_": 1798236243713558720
      },
      {
        "sepalLength": 5.4,
        "sepalWidth": 4.4,
        "petalLength": 1.5,
        "petalWidth": 0.2,
        "species": "setosa",
        "id": "d5f57845-6cea-4739-a38d-c2d3e8aeb05f",
        "_version_": 1798236243713558720
      }
    ]
  }
}
```

- Sort descending by sepal length

The screenshot displays the Solr Admin web application. On the left sidebar, the "Query" option is selected under the "Tools" section. The main area shows a REST client interface with a URL bar containing the endpoint: `localhost:8983/solr/#/iris_collection/query?q=*:*&q.op=OR&indent=true&pg=OR&sort=-sepal.length desc¶ms=`. Below the URL bar, there are input fields for various query parameters like common, q, *, *, g.op, OR, fq, sort, sepal.length desc, start, rows, 0, 10, offset, df, paramters, and indent on. A "Execute Query" button is visible at the bottom left. On the right side, the JSON response of the query is displayed, showing a list of iris species with their corresponding sepal lengths.

- Select for iris' with a petal width between 2.1 and 2.5

The screenshot shows the Solr Admin web interface in a browser. The address bar displays the URL: `localhost:8983/solr/#/iris_collection/query?q=petalWidth:2.1 TO 2.5`. The left sidebar contains navigation links for Dashboard, Logging, Security, Core Admin, Java Properties, Thread Dump, and the selected **iris_collection**. Under **iris_collection**, there are links for Overview, Analysis, Documents, Parameters, Files, Ping (time), Plugins / Stats, Replication, Schema, and Segments info.

The main content area is divided into two panels. The left panel, titled "Request-Handler (q)", contains the following fields:

- `/select` (Request-Handler)
- `common` (Language)
- `q` (Query): `petalWidth (2.1 TO 2.5)`
- `q.op` (Operator): `OR`
- `fq` (Filter Query): (empty)
- `sort` (Sort): (empty)
- `start, rows` (Pagination): `0` to `10`
- `fl` (Fields List): (empty)
- `df` (Default Fields): (empty)
- `params(s)` (Parameters): (empty)
- `wt` (Wrapper): `indent`
- `indent on` (Indent): ☒
- `defType` (Default Type): `lucene`
- `hl` (Highlighting): ☐
- `facet` (Faceting): ☐
- `spatial` (Spatial): ☐
- `spellcheck` (Spellcheck): ☐
- `Raw Query Parameters` (Raw Query Parameters): (empty)
- `JSON Query` (JSON Query): (empty)

The right panel displays the JSON response of the query. The response is a JSON object with the following structure:

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 2,
    "params": {
      "q": "petalWidth (2.1 TO 2.5)",
      "indent": "true",
      "q.op": "OR",
      "useParams": "true",
      "_: "174027804737"
    }
  },
  "response": {
    "numFound": 12,
    "start": 0,
    "numFoundExact": true,
    "docs": [
      {
        "sepalLength": 16.3,
        "sepalWidth": 1.3,
        "petalLength": 16.4,
        "petalWidth": 12.5,
        "species": "virginica",
        "id": "15a3a1b1-628d-464e-8872-b3d8497bf6b3",
        "_version_": 17802362437985585
      },
      {
        "sepalLength": 17.1,
        "sepalWidth": 1.8,
        "petalLength": 15.8,
        "petalWidth": 12.1,
        "species": "virginica",
        "id": "14323ba-6377-491e-b352-b6ad21f6c82f",
        "_version_": 1780236243791788168
      },
      {
        "sepalLength": 16.5,
        "sepalWidth": 1.8,
        "petalLength": 15.8,
        "petalWidth": 12.1,
        "species": "virginica",
        "id": "188e5d37-ab3c-4354-9d4c-eead0fca8d1",
        "_version_": 1780236243792756736
      },
      {
        "sepalLength": 17.6,
        "sepalWidth": 1.8,
        "petalLength": 16.4,
        "petalWidth": 12.1,
        "species": "virginica",
        "id": "1eb251e0-386b-465e-b615-8f380b0ca07",
        "_version_": 1780236243792756737
      },
      {
        "sepalLength": 17.2,
        "sepalWidth": 1.8,
        "petalLength": 16.1,
        "petalWidth": 12.1,
        "species": "virginica",
        "id": "0882a86-a2ef-4c82-aef2-912357ba381b",
        "_version_": 1780236243794653889
      }
    ]
  }
}
```











- Select for petalWidth greater than 2.2, sepallength greater than 7, and sepalWidth greater than 3

The screenshot displays the Solr Admin web interface. On the left, the 'iris_collection' is selected, and the 'Query' tab is active. The query is defined as: `petalWidth:[2.2 TO *] AND sepallength:[7 TO *] AND sepalWidth:[3 TO *]`. The 'Execute Query' button is visible. The main panel shows the JSON response of the query, which is a list of documents. Each document contains fields like `sepalLength`, `sepalWidth`, `petalLength`, `petalWidth`, and `species`. The response is structured as follows:

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 18,
    "params": {
      "q": "petalWidth:[2.2 TO *] AND sepallength:[7 TO *] AND sepalWidth:[3 TO *]",
      "indent": "true",
      "q.op": "AND",
      "useParams": "true",
      "_:": "1714927884737"
    }
  },
  "response": [
    {
      "numFound": 12,
      "start": 0,
      "numFoundExact": true,
      "docs": [
        {
          "sepalLength": 17.2,
          "sepalWidth": 3.8,
          "petalLength": 6.7,
          "petalWidth": 2.2,
          "species": "virginica",
          "id": "c0852ade-2bf4-4c3d-ae72-912357ba381b",
          "_version_": 1798236243794853889
        },
        {
          "sepalLength": 17.7,
          "sepalWidth": 3.8,
          "petalLength": 6.7,
          "petalWidth": 2.2,
          "species": "virginica",
          "id": "61f9acfc-2a57-4985-982c-8886a6d22882",
          "_version_": 1798236243800996768
        },
        {
          "sepalLength": 17.7,
          "sepalWidth": 3.8,
          "petalLength": 6.7,
          "petalWidth": 2.2,
          "species": "virginica",
          "id": "5081a6eb-7b3c-4bd3-9f4a-e38796866982",
          "_version_": 1798236243811631184
        },
        {
          "sepalLength": 17.2,
          "sepalWidth": 3.8,
          "petalLength": 6.7,
          "petalWidth": 2.2,
          "species": "virginica",
          "id": "1d79a7f9-61b4-440b-ab72-4b4b4547786f",
          "_version_": 1798236245966454786
        },
        {
          "sepalLength": 17.7,
          "sepalWidth": 3.8,
          "petalLength": 6.7,
          "petalWidth": 2.2,
          "species": "virginica",
          "id": "51101693-508e-48cb-a99c-61bd9f4e1762",
          "_version_": 1798236245966454786
        }
      ]
    }
  ]
}
```

- Select for virginica species.

The screenshot displays the Solr Admin web interface in a browser. The address bar shows the URL: `localhost:8983/solr/#/iris_collection/query?q=petalWidth:%5B2.2%20TO%20*%5D%20AND%20sepalLength:%5B7%20TO%20*%5...`. The interface is split into two main panels. The left panel, titled 'Solr', contains a configuration form for a query. The right panel displays the query results in a table format. The table has columns for 'id', 'sepalLength', 'sepalWidth', 'petalLength', 'petalWidth', 'species', and 'image'. The results are filtered to show only 'virginica' species. The table contains 10 rows of data. The 'species' column is highlighted in blue. The 'image' column contains small thumbnail images of flowers. The 'id' column contains numerical values. The 'sepalLength', 'sepalWidth', 'petalLength', and 'petalWidth' columns contain numerical values. The 'species' column is highlighted in blue. The 'image' column contains small thumbnail images of flowers. The 'id' column contains numerical values. The 'sepalLength', 'sepalWidth', 'petalLength', and 'petalWidth' columns contain numerical values.

id	sepalLength	sepalWidth	petalLength	petalWidth	species	image
1	5.1	3.5	1.4	0.2	virginica	
2	4.9	3.0	1.4	0.2	virginica	
3	5.0	3.6	1.4	0.2	virginica	
4	5.4	3.7	1.5	0.2	virginica	
5	4.5	2.3	1.3	0.1	virginica	
6	5.2	3.4	1.6	0.4	virginica	
7	4.7	2.8	1.3	0.2	virginica	
8	5.1	3.5	1.4	0.2	virginica	
9	4.9	3.0	1.4	0.2	virginica	
10	5.0	3.6	1.4	0.2	virginica	