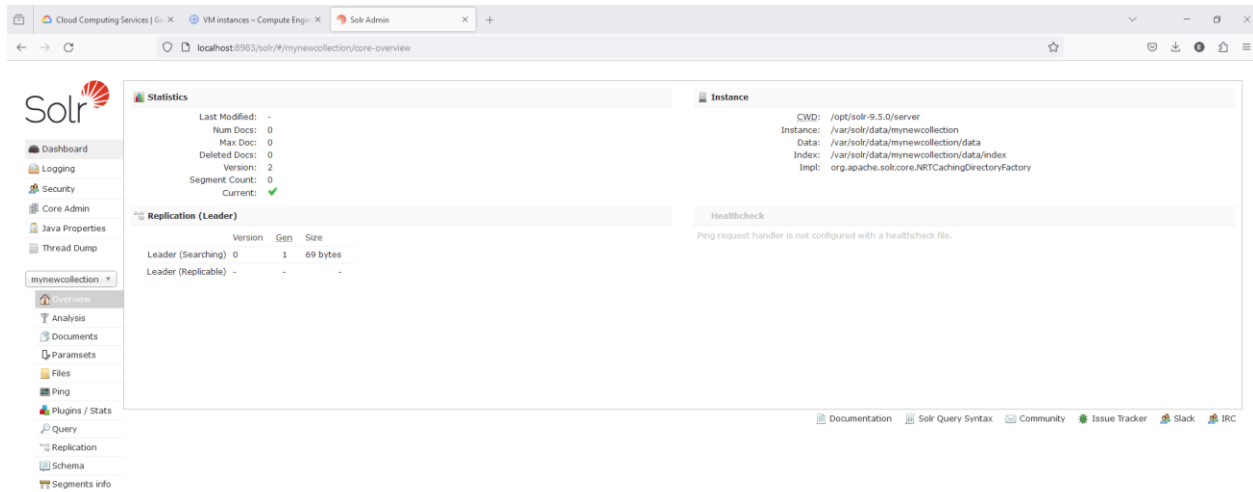


## DSC 650 – Big Data

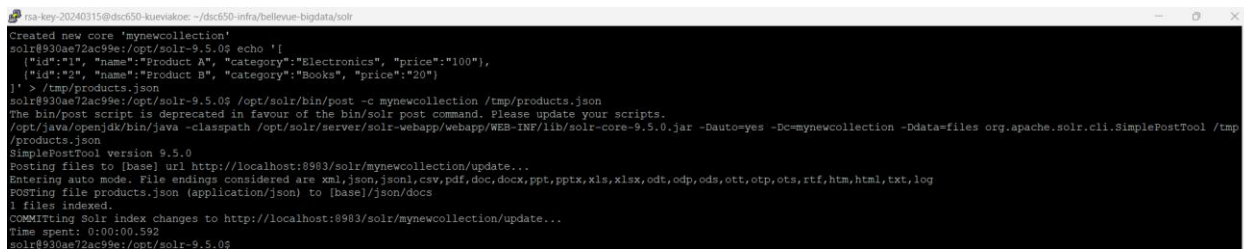
### Week 8 Assignment

Eyram Kueviakoe

#### Screenshot showing 'mynewcollection' in the Solr Web Interface



#### Screenshot showing successful data ingestion messages in the terminal.



#### Screenshots of the JSON response for both queries.

## Query 1:

```
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/solr
solr@930ae72ac99e:/opt/solr-9.5.0$ 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_":1797897163528732672
    },{
      "id":"2",
      "name":["Product B"],
      "category":["Books"],
      "price":[20],
      "_version_":1797897163611570176
    }]
  }
}
solr@930ae72ac99e:/opt/solr-9.5.0$
```

## Query 2:

```
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/solr
solr@930ae72ac99e:/opt/solr-9.5.0$ curl "http://localhost:8983/solr/mynewcollection/select?q=category:Electronics"
{
  "responseHeader":{
    "status":0,
    "QTime":0,
    "params":{
      "q":"category:Electronics"
    }
  },
  "response":{
    "numFound":1,
    "start":0,
    "numFoundExact":true,
    "docs":[{
      "id":"1",
      "name":["Product A"],
      "category":["Electronics"],
      "price":[100],
      "_version_":1797897163528732672
    }]
  }
}
solr@930ae72ac99e:/opt/solr-9.5.0$
```

Screenshots of different queries and their results from the Solr Web Interface.

## Number of products in each category

The screenshot shows the Solr Admin interface with the 'Query' tab selected. The 'Request-Handler (qt)' section shows the following configuration:

- q: /select
- q.op: OR
- fq: (empty)
- sort: (empty)
- start, rows: 0, 10
- fl: (empty)
- df: (empty)
- paramset(s): (empty)
- wt: (empty)
- indent on: ☒
- debugQuery: ☐
- defType: (empty)
- facet: ☒
- facet.query: (empty)
- facet.field: category
- facet.prefix: (empty)

The URL bar shows: `localhost:8983/solr/#/mynewcollection/query?q=*:*&q.op=OR&indent=true&facet=true&facet.field=category&facet.mincount=1&facet.sort=count&useParams=`

The response is a JSON object:

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 13,
    "params": {
      "q": "*:*",
      "facet.field": "category",
      "indent": "true",
      "q.op": "OR",
      "facet.mincount": "1",
      "facet.sort": "count",
      "useParams": ""
    }
  },
  "response": {
    "numFound": 2,
    "start": 0,
    "numFoundExact": true,
    "docs": [
      {
        "id": "1",
        "name": "Product A",
        "category": ["Electronics"],
        "price": [100],
        "_version_": 1797897163528732672
      },
      {
        "id": "2",
        "name": "Product B",
        "category": ["Books"],
        "price": [20],
        "_version_": 1797897163611570176
      }
    ],
    "facet_counts": {
      "facet_queries": { },
      "facet_fields": {
        "category": ["books", 1, "electronics", 1]
      },
      "facet_ranges": { },
      "facet_intervals": { },
      "facet_heatmaps": { }
    }
  }
}
```

## Products in the 'Electronics' category.

The screenshot shows the Solr Admin interface with the 'Query' tab selected. The 'Request-Handler (qt)' section shows the following configuration:

- q: /select
- q: category:Electronics
- q.op: OR
- fq: (empty)
- sort: (empty)
- start, rows: 0, 10
- fl: (empty)
- df: (empty)
- paramset(s): (empty)
- wt: (empty)
- indent on: ☒
- debugQuery: ☐
- defType: (empty)
- facet: ☐
- facet.query: (empty)
- facet.field: (empty)
- facet.prefix: (empty)

The URL bar shows: `localhost:8983/solr/#/mynewcollection/query?q=category:Electronics&q.op=OR&indent=true&useParams=`

The response is a JSON object:

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 1,
    "params": {
      "q": "category:Electronics",
      "indent": "true",
      "q.op": "OR",
      "useParams": ""
    }
  },
  "response": {
    "numFound": 1,
    "start": 0,
    "numFoundExact": true,
    "docs": [
      {
        "id": "1",
        "name": "Product A",
        "category": ["Electronics"],
        "price": [100],
        "_version_": 1797897163528732672
      }
    ]
  }
}
```

## Products in the 'Books' category.

The screenshot shows the Solr Admin web interface. The left sidebar contains navigation links: Dashboard, Logging, Security, Core Admin, Java Properties, Thread Dump, mynewcollection (selected), Overview, Analysis, Documents, Paramsets, Files, Ping, Plugins / Stats, Query (selected), Replication, Schema, and Segments info. The main content area is divided into two panels. The left panel, titled 'Request-Handler (qt)', contains the following fields:   
 - Request-Handler: /select   
 - q: category:Books   
 - q.op: OR   
 - fq:   
 - sort:   
 - start, rows: 0 to 10   
 - fl:   
 - df:   
 - paramset(s): Select paramset(s) ...   
 - wt:   
 - indent on: ☒   
 - debugQuery: ☐   
 - defType:   
 The right panel displays the JSON response for the query: 

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 0,
    "params": {
      "q": "category:Books",
      "indent": "true",
      "q.op": "OR",
      "useParams": "",
      "_": "1714608760713"
    }
  },
  "response": {
    "numFound": 1,
    "start": 0,
    "numFoundExact": true,
    "docs": [
      {
        "id": "2",
        "name": ["Product B"],
        "category": ["Books"],
        "price": [20],
        "_version_": 1797897163611570176
      }
    ]
  }
}
```

## Prices greater than or equal to 50

The screenshot shows the Solr Admin web interface with the same sidebar as the previous image. The main content area is divided into two panels. The left panel, titled 'Request-Handler (qt)', contains the following fields:   
 - Request-Handler: /select   
 - q: price:[50 TO \*]   
 - q.op: OR   
 - fq:   
 - sort:   
 - start, rows: 0 to 10   
 - fl:   
 - df:   
 - paramset(s): Select paramset(s) ...   
 - wt:   
 - indent on: ☒   
 The right panel displays the JSON response for the query: 

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 0,
    "params": {
      "q": "price:[50 TO *]",
      "indent": "true",
      "q.op": "OR",
      "useParams": "",
      "_": "1714608760713"
    }
  },
  "response": {
    "numFound": 1,
    "start": 0,
    "numFoundExact": true,
    "docs": [
      {
        "id": "1",
        "name": ["Product A"],
        "category": ["Electronics"],
        "price": [100],
        "_version_": 1797897163528732672
      }
    ]
  }
}
```

## Create a new collection in Solr

```
/opt/solr/bin/solr create -c week8_sales_collection
```

```
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/solr
solr@930ae72ac99e:/opt/solr-9.5.0$ /opt/solr/bin/solr create -c week8_sales_collection
WARNING: Using default configset with data driven schema functionality. NOT RECOMMENDED for production use.
        To turn off: bin/solr config -c week8_sales_collection -p 8983 -action set-user-property -property update.autoCreateFields -value false
Created new core 'week8_sales_collection'
solr@930ae72ac99e:/opt/solr-9.5.0$
```

## Let's create a json file and call it sales.json

```
rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/solr
solr@930ae72ac99e:/opt/solr-9.5.0$ echo '[
{
  "sale_id": 1,
  "product": "Product 1",
  "category": "Books",
  "quantity": 5,
  "unit_price": 11.35,
  "total_price": 56.75
},
{
  "sale_id": 2,
  "product": "Product 2",
  "category": "Electronics",
  "quantity": 1,
  "unit_price": 16.33,
  "total_price": 16.33
},
{
  "sale_id": 3,
  "product": "Product 3",
  "category": "Toys",
  "quantity": 4,
  "unit_price": 59.45,
  "total_price": 237.8
},
{
  "sale_id": 4,
  "product": "Product 4",
  "category": "Clothing",
  "quantity": 8,
  "unit_price": 81.1,
  "total_price": 648.8
},
{
  "sale_id": 5,
  "product": "Product 5",
  "category": "Toys",
  "quantity": 8,
  "unit_price": 85.35,
  "total_price": 682.8
},
{
  "sale_id": 6,
  "product": "Product 6",
  "category": "Home & Kitchen",
  "quantity": 7,
  "unit_price": 18.29,
  "total_price": 128.03
},
{
  "sale_id": 7,
  "product": "Product 7",
  "category": "Toys",
  "quantity": 6,
  "unit_price": 99.28,
  "total_price": 595.68
},
{
  "sale_id": 8,
  "product": "Product 8",
  "category": "Toys",
  "quantity": 7,
  "unit_price": 73.46,
```

## Let's add the data to the collection

```
nsa-key-20240315@dc650-kueviakoe: ~/dc650-infra/bellevue-bigdata/solr
solr@930ae72ac99e:/opt/solr-9.5.0$ /opt/solr/bin/post -c week8_sales_collection /tmp/sales.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 -Dauto=yes -Dc=week8_sales_collection -Ddata=files org.apache.solr.cli.SimplePostTool /tmp/sales.json
SimplePostTool version 9.5.0
Posting files to [base] url http://localhost:8983/solr/week8_sales_collection/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 sales.json (application/json) to [base]/json/docs
3 files indexed.
COMMITting Solr index changes to http://localhost:8983/solr/week8_sales_collection/update...
Time spent: 0:00:00.250
solr@930ae72ac99e:/opt/solr-9.5.0$
```

## Queries:

All documents from the collection

Sales where category = clothing

The screenshot shows the Solr Admin web interface. On the left is a sidebar with navigation links: Dashboard, Logging, Security, Core Admin, Java Properties, Thread Dump, and a dropdown for 'week8\_sales\_...'. The 'Query' link is selected. The main panel is titled 'Request-Handler (qt)' and shows the following configuration:

- Request-Handler (qt): /select
- q: category:Clothing
- q.op: OR
- fq: (empty)
- sort: (empty)
- start, rows: 0, 10
- fl: (empty)
- df: (empty)
- paramset(s): Select paramset(s)...
- wt: json
- indent on: ☒
- debugQuery: ☐
- defType: (empty)
- hl: ☐
- facet: ☐
- spatial: ☐
- spellcheck: ☐

The URL bar shows: `localhost:8983/solr/#/week8_sales_collection/query?q=category:Clothing&q.op=OR&indent=true&useParams=`

The response is displayed in a JSON format:

```
{
  "responseHeader": {
    "status": 0,
    "QTime": 1,
    "params": {
      "q": "category:Clothing",
      "indent": "true",
      "q.op": "OR",
      "useParams": "",
      "_": "1714612568393"
    }
  },
  "response": {
    "numFound": 2,
    "start": 0,
    "numFoundExact": true,
    "docs": [
      {
        "sale_id": [4],
        "product": ["Product 4"],
        "category": ["Clothing"],
        "quantity": [8],
        "unit_price": [81.1],
        "total_price": [648.8],
        "id": "87e981c7-3c5c-4bf3-bec2-f52f99620c3f",
        "_version_": 1797901104420225024
      },
      {
        "sale_id": [10],
        "product": ["Product 10"],
        "category": ["Clothing"],
        "quantity": [6],
        "unit_price": [97.72],
        "total_price": [586.32],
        "id": "47232a19-4c7e-4d55-ac77-cf765f8517f6",
        "_version_": 1797901104426516400
      }
    ]
  }
}
```

## Number of sales by category

The screenshot shows the Solr Admin interface. The browser address bar displays the URL: `localhost:8983/solr/#/week8_sales_collection/query?q=*:*&q.op=OR&indent=true&facet=true&facet.field=category&facet.sort=count&useParams=`. The interface includes a sidebar with navigation options: Dashboard, Logging, Security, Core Admin, Java Properties, Thread Dump, week8\_sales\_..., Overview, Analysis, Documents, Paramsets, Files, Ping, Plugins / Stats, Query, Replication, Schema, and Segments info. The main area shows a JSON query input field with an "Execute Query" button. The output is a JSON response showing sale details and facet counts for the category 'kitchen'.

```
{
  "sale_id": [7],
  "product": ["Product 7"],
  "quantity": [4],
  "unit_price": [19.68],
  "total_price": [78.72],
  "id": "3800e43d-ae50-491a-8c15-1bd58c269b67",
  "_version_": 1797900106132881408
}, {
  "sale_id": [8],
  "product": ["Product 8"],
  "quantity": [1],
  "unit_price": [56.2],
  "total_price": [56.2],
  "id": "33dd331d-a286-4c8a-905a-88e277cc8400",
  "_version_": 1797900106133929984
}, {
  "sale_id": [9],
  "product": ["Product 9"],
  "quantity": [4],
  "unit_price": [38.6],
  "total_price": [154.4],
  "id": "1bd1dc8f-7ea4-44d6-b880-27ac0806e2ae",
  "_version_": 1797900106133929985
}, {
  "sale_id": [10],
  "product": ["Product 10"],
  "quantity": [8],
  "unit_price": [18.3],
  "total_price": [146.4],
  "id": "7b71859d-a362-4eb0-9833-bd9f934ac11b",
  "_version_": 1797900106134978560
}]
}, {
  "facet_counts": {
    "facet_queries": { },
    "facet_fields": {
      "category": ["toys", 4, "clothing", 2, "home", 2, "kitchen", 2, "books", 1, "electronics", 1]
    },
    "facet_ranges": { },
    "facet_intervals": { },
    "facet_heatmaps": { }
  }
}
```

Category = kitchen

`curl "http://localhost:8983/solr/week8_sales_collection/select?q=category:Kitchen"`

```

rsa-key-20240315@dsc650-kueviakoe: ~/dsc650-infra/bellevue-bigdata/solr
solr@930ae72ac99e:/opt/solr-9.5.0$ curl "http://localhost:8983/solr/week8_sales_collection/select?q=category:Kitchen"
{
  "responseHeader":{
    "status":0,
    "QTime":1,
    "params":{
      "q":"category:Kitchen"
    }
  },
  "response":{
    "numFound":2,
    "start":0,
    "numFoundExact":true,
    "docs":[
      {
        "sale_id":[6],
        "product":["Product 6"],
        "category":["Home & Kitchen"],
        "quantity":[7],
        "unit_price":[18.29],
        "total_price":[128.03],
        "id":"38e550ba-e3aa-4b59-998f-ale944ael0a6",
        "_version_":179790110442322176
      },{
        "sale_id":[9],
        "product":["Product 9"],
        "category":["Home & Kitchen"],
        "quantity":[5],
        "unit_price":[45.69],
        "total_price":[228.45],
        "id":"fla3d5b9-fc95-4858-a430-343b1dlbc049",
        "_version_":1797901104425467904
      }
    ]
  }
}
solr@930ae72ac99e:/opt/solr-9.5.0$

```

Total sales price is higher or equal to 650

Solr Admin

localhost:8983/solr/#/week8\_sales\_collection/query?q=total\_price:[650 TO \*]&q.op=OR&indent=true&useParams=

Solr

Dashboard
Logging
Security
Core Admin
Java Properties
Thread Dump
week8\_sales\_
Overview
Analysis
Documents
Params
Files
Ping
Plugins / Stats
Query
Replication
Schema
Segments info

Request-Handler (qt)

/select

common

q

total\_price:[650 TO \*]

q.op

OR

fq

sort

start, rows

0 10

fl

df

paramset(s)

Select paramset(s)

wt

-----

indent on

debugQuery

defType

-----

hl

facet

spatial

spellcheck

Raw Query Parameters

JSON Query

http://localhost:8983/solr/week8\_sales\_collection/select?indent=true&q.op=OR&q=total\_price%3A%5B650%20TO%20%2A%5D&useParams=

```

{
  "responseHeader":{
    "status":0,
    "QTime":1,
    "params":{
      "q":"total_price:[650 TO *]",
      "indent":"true",
      "q.op":"OR",
      "useParams":"","
      "_":1714612568393"
    }
  },
  "response":{
    "numFound":2,
    "start":0,
    "numFoundExact":true,
    "docs":[
      {
        "sale_id":[4],
        "product":["Product 4"],
        "quantity":[10],
        "unit_price":[96.73],
        "total_price":[967.3],
        "id":"918b9ad4-f4d4-4435-b640-447ef99f3363",
        "_version_":1797900106129735680
      },{
        "sale_id":[5],
        "product":["Product 5"],
        "category":["Toys"],
        "quantity":[8],
        "unit_price":[85.35],
        "total_price":[682.8],
        "id":"4cd4219-22cd-4e9f-b2f9-ec192a7c2543",
        "_version_":1797901104421273600
      }
    ]
  }
}

```

8