**Required Viewing**

Before starting this assignment, you must watch two videos:

1. **Week 8 Fundamentals Lecture Video**
   * Explains Solr architecture and the inverted index.
   * Provides the background needed to understand what you are learning.
   * Link: <https://youtu.be/dM-0545k9as>
2. **Week 8 Assignment Walkthrough Video**
   * Demonstrates step by step how to complete the tasks, including commands and expected outputs.
   * It is not enough to simply run commands. You must verify that your commands execute correctly. Incorrect or incomplete results will lose points.

Watching both videos is mandatory. The **fundamentals video** explains what you are learning and why it matters, while the **walkthrough video** shows how to complete the assignment.

**Submission Guidelines**

* Submit your work as a **single Word or PDF document** (no raw screenshots or multiple files).
* Include the following in your submission:
  + Screenshots of each required step.
  + A short explanation for each screenshot:
    - The command/action you ran.
    - What the output shows.
    - Whether the result matched your expectation.
* Organize your work in the **same order as the assignment guide** so it is easy to follow.
* This is a **master’s level course** – professionalism and clarity are expected. Well-structured submissions demonstrate your ability to communicate technical work effectively.

**Week 8 Assignment – Objectives and Points**

* **Objective 1 – Conceptual Foundations (Solr Overview)**: 8 pts
* **Objective 2 – Creating a Solr Collection**: 10 pts
* **Objective 3 – Adding Data to the Collection**: 10 pts
* **Objective 4 – Querying the Collection**: 12 pts
* **Objective 5 – Advanced Querying in the Solr Web Interface**: 20 pts
* **Objective 6 – Working with Custom JSON Data**: 28 pts

**Total: 88 points**

## Week 8 Assignment: Diving into Apache Solr

In this assignment, you will gain hands-on experience with **Apache Solr**, a powerful search platform used for indexing and querying large datasets. Solr is widely used for full-text search, faceted search, and real-time indexing. Through a series of exercises, you will learn how to create collections, ingest data, and execute queries both from the command line and through Solr’s Web Interface.

By the end of this assignment, you will:

* Understand how to create and manage Solr collections.
* Gain experience in adding structured data to Solr for indexing.
* Learn how to query data using Solr’s query syntax and the Solr Web Interface.
* Explore advanced search capabilities such as faceted search and filters.
* Work with custom JSON datasets to understand how Solr handles and indexes different types of data.

# Objective 1 - Conceptual Foundations (8 points)

Before beginning the assignment, watch the instructor-led fundamentals video, which introduces and explains the key concepts for this week:  
<https://youtu.be/dM-0545k9as>

**Deliverable:** Write a 3–4 paragraph summary in your own words. Your write-up should explain the main ideas from the fundamentals video, highlight why these concepts are important, and connect them to the Solr exercises in this assignment.

# Objective 2 - Creating a Solr Collection (10 points)

#### **1. Environment Initialization**

* Navigate to the Solr directory:

cd dsc650-infra/bellevue-bigdata/solr

* Start the Docker containers:

docker-compose up -d

* Access the Solr container:

docker exec -it solr\_solr\_1 bash

* If you can’t access the Solr container, it could be due to a container name change. In this cause use:

docker exec -it solr-solr-1 bash

#### **2. Creating a Solr Collection**

In this section, you will create a Solr collection named mynewcollection. Solr collections are logical groupings of indexed data, and each collection is capable of being queried independently.

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

/opt/solr/bin/solr create -c mynewcollection

**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.

**Deliverable 1:** Screenshot showing mynewcollection in the Solr Web Interface, plus 1–2 sentences explaining how you know the collection was created successfully.

# Objective 3 - Adding Data to the Collection (10 points)

#### **1. Generating and Adding Data to the Collection**

You will now generate sample data in JSON format and add it to the mynewcollection for indexing. This step allows you to see how Solr ingests structured data into its indexing engine.

For this exercise, create a json file called products.json using this commands:

echo '[

{"id":"1", "name":"Product A", "category":"Electronics", "price":"100"},

{"id":"2", "name":"Product B", "category":"Books", "price":"20"}

]' > /tmp/products.json

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

/opt/solr/bin/post -c mynewcollection /tmp/products.json

**Deliverable 1:** Screenshot of the terminal output after running the post command, plus 1–2 sentences explaining what the ingestion messages confirm about the data load.

# Objective 4 - Querying the Collection (12 points)

#### **1. Querying the Collection**

Now that the data has been ingested, you will perform queries on the mynewcollection using Solr’s command line query interface. This demonstrates how Solr processes and retrieves indexed data.

**Exercise 4:** Use Solr’s query interface to retrieve all documents from ‘mynewcollection’.

**Deliverable 1:** Screenshot of the JSON response showing all documents from the collection, plus 1–2 sentences explaining what the output shows.

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"

**Deliverable 2:** Screenshot of the query filtering by category Electronics, plus 1–2 sentences explaining the result.

# Objective 5 - Advanced Querying in the Solr Web Interface (20 points)

#### **1. Advanced Querying in Solr Web Interface**

In this section, you will access the Solr Web Interface for more advanced querying. The Web Interface provides a user-friendly platform for running and experimenting with different queries, filters, and parameters.

* **Faceted Search**:  
  Run a faceted search to count the number of products in each category. Faceted search is a powerful feature in Solr that allows you to retrieve aggregated data, such as counts of documents grouped by a particular field.
* **Experiment with Filters and Query Parameters**:  
  Use the Web Interface to experiment with additional query parameters, such as sorting, filtering, and pagination.

Access the Solr Web Interface for querying:

[Query ‘mynewcollection’ in Solr Web Interface](http://localhost:8983/solr/#/mynewcollection/query)

Follow the link, and you’ll be presented with a user-friendly interface to craft and execute your queries. Experiment with different parameters and filters.

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

**Deliverable 1:** Screenshot of a faceted search showing counts of products by category, plus 1–2 sentences explaining the output.

**Exercise 7:** Use the Web Interface to experiment further with various query parameters and filters.

**Deliverable 2:** Screenshots of at least two additional queries using filters, sorting, or pagination, each with a short explanation of what the query did and what the results mean.

# Objective 6 - Working with Custom JSON Data (28 points)

#### **1. Custom JSON Data**

To extend your learning, you will now work with custom data. Either generate your own JSON data or download a JSON dataset from the web and add it to a new Solr collection. This allows you to see how Solr indexes and handles different datasets.

* Create a new collection in Solr
* Either generate JSON data or use JSON data from the web and add it to your collection.
* Query your collection.

**Deliverable 1:** Screenshot showing the new collection you created for your custom dataset, plus 1–2 sentences explaining what data you loaded.

**Deliverable 2:** Screenshot of at least two different queries against your custom dataset (either via Solr Web Interface or curl), each with a short explanation of what the query did and what the results show.

**Deliverable 3:** A short paragraph summarizing what you learned by working with your own dataset and how Solr handled it.

## Shutting Down

Ensure all Docker containers are turned off with docker-compose down for each directory. If you’re using google cloud, please shut down your virtual machine to preserve cloud costs.