## ! STOP!

**Required Viewing**

Before starting this assignment, you must watch two videos:

1. **Week 8 Fundamentals Lecture Video –** Explains Solr architecture and the inverted index, giving you the background needed to understand what you are learning.  
   Link: <https://youtu.be/dM-0545k9as>
2. **Week 8 Assignment Walkthrough Video** – Shows you step by step how to complete the tasks, including each command and the expected output.

It is not enough to just run commands without checking results. You must verify that your commands execute correctly. If they do not, you will lose points.

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

## Week 8 Assignment: Diving into Apache Solr

# Conceptual Foundations

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 that demonstrates your understanding of the material presented in the video. Your writeup should explain the main ideas in your own words, highlight why these concepts are important, and connect them to the technologies you will be working with in the assignment.

### Objective: Exploring Apache Solr for Data Indexing and Querying

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.

#### **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:** Screenshot showing ‘mynewcollection’ in the Solr Web Interface.

#### **3. 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:** Screenshot showing successful data ingestion messages in the terminal.

#### **4. 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’.

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:** Screenshots of the JSON response for both queries.

#### **5. 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.

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

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

#### **6. 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:** Screenshots of different queries on your Solr collection and the query results from either the command line or the Solr Web Interface.

## 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.