**Required Viewing**

Before starting this assignment, you must watch three videos:

1. **Week 9 Fundamentals Lecture Video**
   * Explains NiFi architecture, the ETL process, and an IoT example.
   * Provides the background needed to understand what you are learning.
   * Link: <https://youtu.be/YiCo70SV2O8>
2. **Week 9 Assignment Walkthrough Video**
   * Shows step by step how to complete the tasks, including commands and expected outputs.
   * It is not enough to just run commands. You must verify that your commands executed correctly. Incorrect or incomplete results will lose points.

Watching all videos is mandatory. The fundamentals video explains what you are learning and why it matters, the interface walkthrough shows how to navigate the NiFi UI, and the walkthrough video demonstrates 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 show your ability to communicate technical work effectively.

**Week 9 Assignment – Objectives and Points**

* **Objective 1 – Conceptual Foundations: 8 pts**
* **Objective 2 – Environment Initialization (NiFi UI Access): 6 pts**
* **Objective 3 – Creating a Processor Group and Parameter Context: 10 pts**
* **Objective 4 – Designing a Simple Flow (GenerateFlowFile → LogAttribute): 12 pts**
* **Objective 5 – Setting Up Solr Collection: 12 pts**
* **Objective 6 – NiFi Advanced Flow with Solr Integration: 16 pts**
* **Objective 7 – Querying Solr Data: 24 pts**

**Total: 88 points**

# Week 9 Assignment: Exploring Apache NiFi

In this assignment, you will explore **Apache NiFi**, a powerful tool for automating and managing data flows between systems. NiFi is widely used for moving and transforming data across different platforms, offering real-time control over data flow, routing, transformation, and system integration. You will design and implement dataflows, working with processors, parameter contexts, and integrating with **Apache Solr** for data indexing and querying.

By the end of this assignment, you will:

* Understand how to create and configure processor groups in NiFi.
* Gain experience in designing simple and advanced data flows using NiFi processors.
* Learn how to integrate NiFi with Solr for processing and indexing log data.
* Use Solr’s interface to query and analyze processed 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:  
**Link:** <https://youtu.be/YiCo70SV2O8>

**Deliverable:** Write a 3–4 paragraph summary of the fundamentals video. Explain the key concepts in your own words, why they are important, and how they connect to NiFi and Solr in this assignment.

# Objective 2 - Environment Initialization (6 points)

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

* Change into the nifi directory with

cd dsc650-infra/bellevue-bigdata/nifi

* Start NiFi using the command:

/bin/bash nifi-\*/bin/nifi.sh start

* Access the NiFi User Interface using the instructions in the Week 1 assignment.

[Access NiFi UI](https://localhost:8443/nifi)

**Note:** Use the username and password obtained in Week 1 to access the interface.

**Deliverable:** Screenshot confirming successful access to the NiFi UI. Include 1–2 sentences explaining what you did to start NiFi and how you know the UI is working.

# Objective 3 - Creating a Processor Group and Context (10 points)

#### **1. Creating a Processor Group**

In this section, you will create your first **Processor Group** in NiFi, which serves as a logical container for grouping processors that will work together to form a data flow.

***Please follow the video tutorial which covers creating a process group and parameter context.***

**Deliverable:** Screenshot of the defined parameter within the Parameter Context. Add 1–2 sentences explaining what parameter you created and why it matters in managing your flow.

# Objective 4 - Designing a Simple Flow (12 points)

#### **1. Designing a Simple Flow**

In this step, you will design a simple data flow using NiFi processors to generate files and log their attributes. This introduces you to how processors can be linked and configured to automate data processing.

***Please follow the video tutorial which covers creating designing your first flow.***

**Deliverable:** Screenshot of your simple flow (GenerateFlowFile connected to LogAttribute). Include 1–2 sentences explaining what the flow does and what the output confirms.

# Objective 5 - Setting Up Solr Collection (12 points)

#### **1. Setting Up Solr Collection**

In this section, you’ll set up a Solr collection to store logs. This will integrate your NiFi flow with Solr, allowing you to push data into a searchable index.

* Navigate to the Solr directory and initiate the Solr Docker container, as instructed in the Solr assignment:

cd solr  
docker-compose up -d

Create a Solr collection named syslog.

* Enter the Solr docker 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

* Create the Solr Collection

/opt/solr/bin/solr create -c syslog

**Deliverable:** Screenshot confirming the successful creation of the syslog Solr collection. Add 1–2 sentences explaining what this collection will be used for in the assignment.

# Objective 6 - NiFi Advanced Flow with Solr Integration (16 points)

#### **1. NiFi Advanced Flow with Solr**

Now, you’ll implement a more advanced data flow that integrates NiFi with Solr for log data processing and storage. You’ll import a pre-built flow to generate, filter, and index logs in Solr.

***Please follow the video tutorial which covers importing and starting your flow.***

**Deliverable:** Include a screenshot of the advanced NiFi flow running and a screenshot of data visible in Solr, with 1–2 sentences explaining what the results confirm.

# Objective 7 – Querying Solr Data (24 points)

#### **1. Querying Solr Data**

Finally, you will query the processed data stored in the Solr syslog collection. Use the Solr Web Interface to craft and execute queries, exploring different query parameters and filters.

Access the Solr Web Interface for querying:

[Query ‘syslog’ in Solr Web Interface](http://localhost:8983/solr/#/syslog/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.

**Deliverable:** Screenshot of Solr query results from the syslog collection. Add 1–2 sentences explaining the query you ran and what the results show about your indexed data.

## Shutting Down

#### **1. NiFI Shutdown**

* Change into the nifi directory with

cd dsc650-infra/bellevue-bigdata/nifi

* Stop NiFi using the command:

/bin/bash nifi-\*/bin/nifi.sh stop

#### **2. Solr Shutdown**

* Change (cd) into the solr directory
* Stop Solr

docker-compose down