Week 9 Assignment: Exploring Apache NiFi

Objective:

Familiarize yourself with the capabilities of Apache NiFi by designing and implementing dataflows.

1. Environment Initialization

- Change into the nifi directory with
  `cd 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

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

Deliverable: Screenshot confirming successful access to the NiFi UI.

2. Creating a Processor Group

Exercise 1: Drag and drop the "Processor Group" icon onto the canvas. Name this processor group "My First NiFi Flow".

Deliverable: Screenshot of the NiFi canvas showing the Processor Group.

Exercise 2: Create a Parameter Context for your new processor group. Define a parameter (e.g., `File_Size`) that will determine the size of the files generated in the next step.

Deliverable: Screenshot of the defined parameter within the Parameter Context.

3. Designing a Simple Flow

Exercise 3: Enter the "My First NiFi Flow" processor group by double-clicking it.

- Add the `GenerateFlowFile` processor to the canvas.
- Configure its properties:
  - Set `File Size` using the parameter (`File_Size`) you defined earlier.
  - Adjust the `Scheduling` tab to run the processor every 5 seconds.
- Next, add the `LogAttribute` processor to the canvas.
- Change the `Bulletin Level` in the `Settings` to `Info`.
- Change the `Log Payload` property in `Properties` to `true`.

- Connect `GenerateFlowFile` to `LogAttribute`. The relationship should be "success".
- Start both processors and observe the flow of files.

Deliverable: Screenshot of the simple flow (`GenerateFlowFile` to `LogAttribute`).

4. Setting Up Solr Collection

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

```
cd solr
docker-compose up -d
```

Exercise 4: Create a topic named `nifi-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` collection.

5. NiFi Advanced Flow with Solr

Exercise 5: Back in NiFi, on the NiFi canvas:

- Import the provided JSON file. This will load a pre-built dataflow onto your canvas.
- This flow will:
  - Generate log data.
  - Filter the logs with SQL.
  - Convert logs from the syslog format to JSON.
  - Publish the processed data to Solr.
- Start the Flow

Deliverable: Screenshot of the advanced NiFi flow processing and sending data to Solr.

6. Querying Solr Data

Exercise 6: Access the Solr Web Interface for querying:

Query 'syslog' in Solr Web Interface

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 the Solr query results.