In ETL, data can be a simple delimited text file, other types of files such as JSON, Excel, or XML, or databases such as MySQL or MongoDB. NiFi gives you the ability to extract data from these different sources and prepare it to be loaded into another system. NiFi is designed to handle large volumes of data efficiently.

In this mini-lesson, you will learn how to use NiFi to extract data from an Excel file and prepare it to be loaded into another system for processing.

For this mini-lesson, you will be using Docker *containers*. First, create a Docker network called NifiNetwork by running the command below in a Terminal window:

| docker network create NifiNetwork |
| --- |

Next, create the Docker *container* running NiFi and associate it to the NifiNetwork that you created:

| docker run --name nificontainer -p 8080:8080 --network NifiNetwork -d apache/nifi:1.13.2 |
| --- |

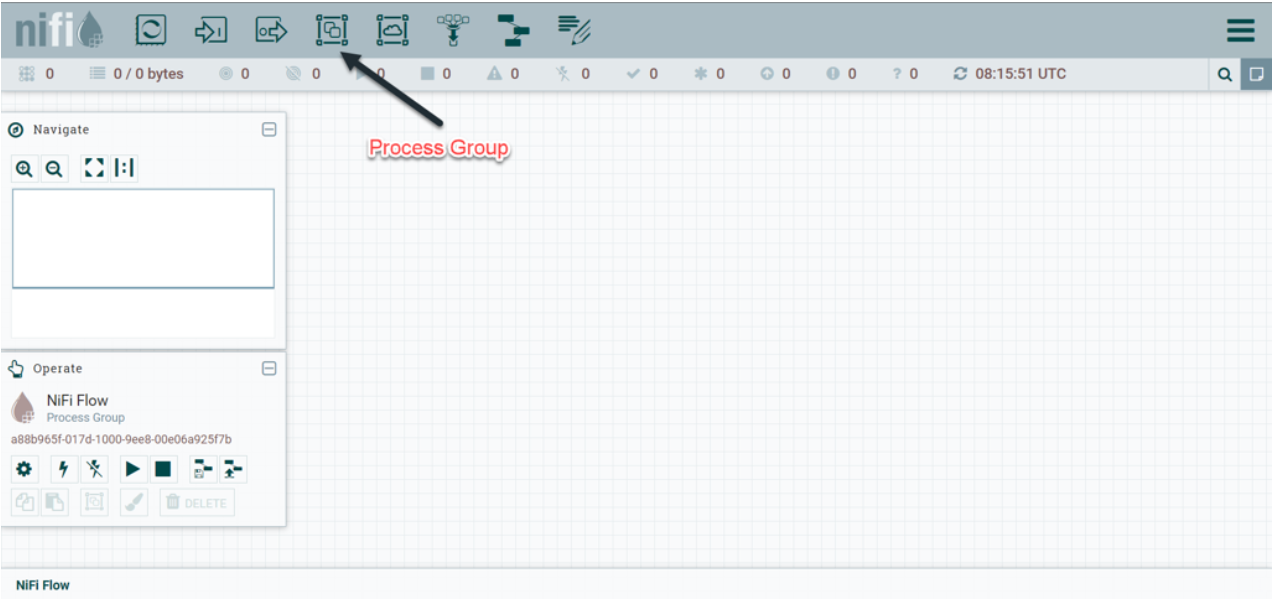
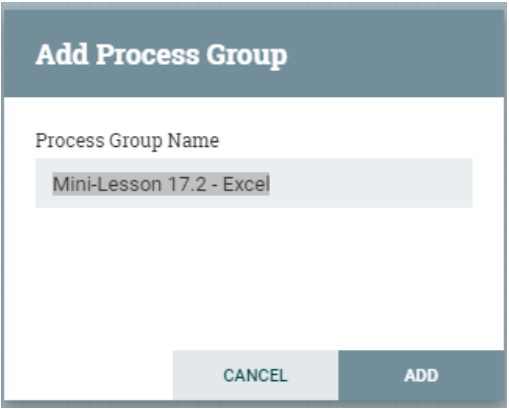
Note that setting up a NiFi Docker *container* may take several minutes depending on the speed of your network connection.

To prepare the Docker NiFi *server* to extract and load the data from an Excel file, follow the steps below:

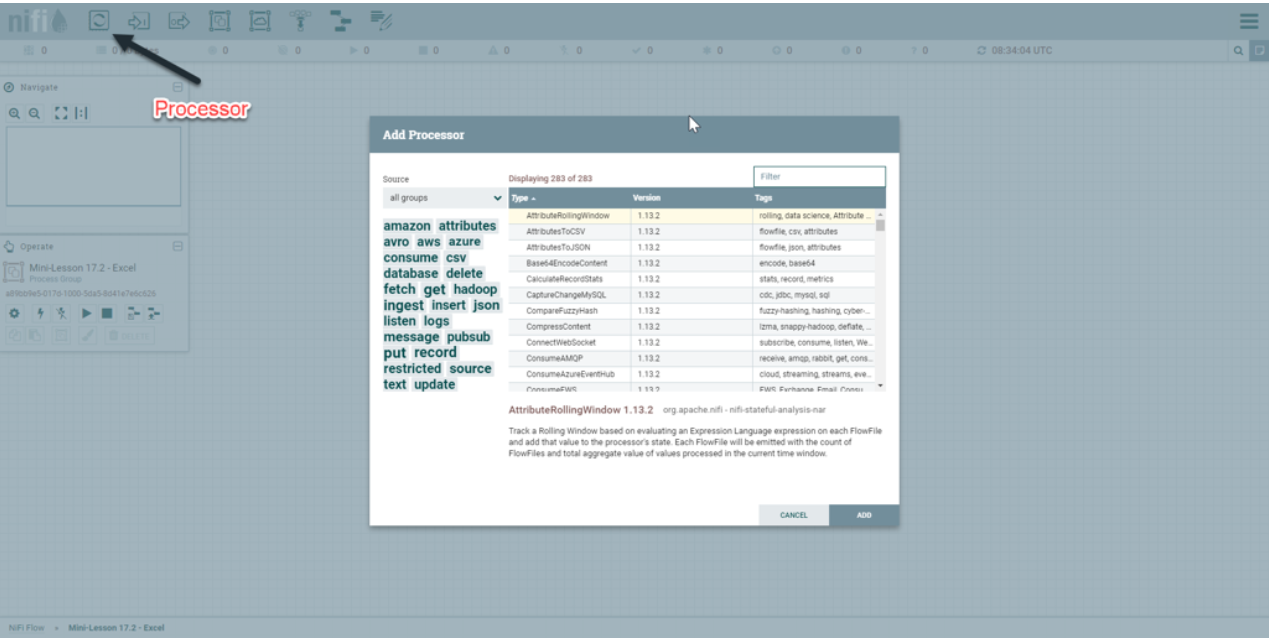
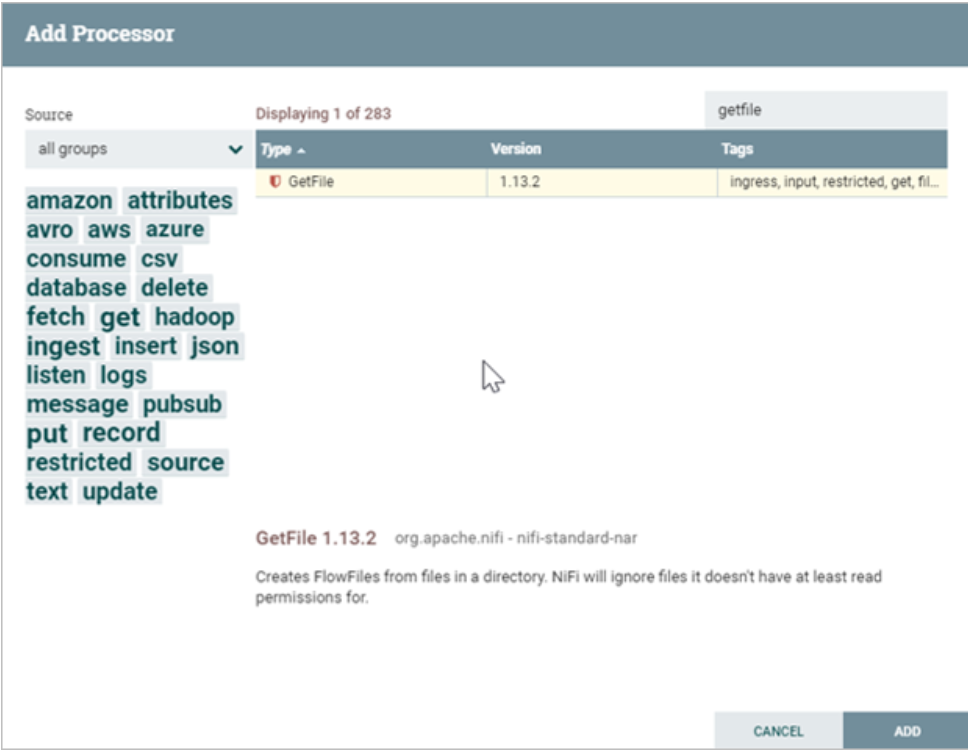
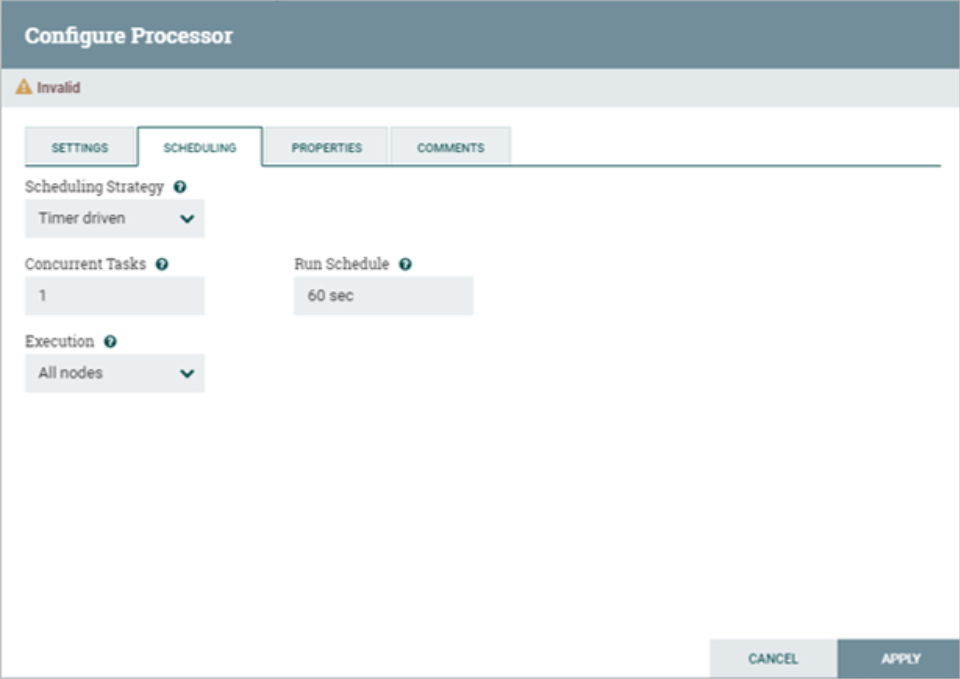
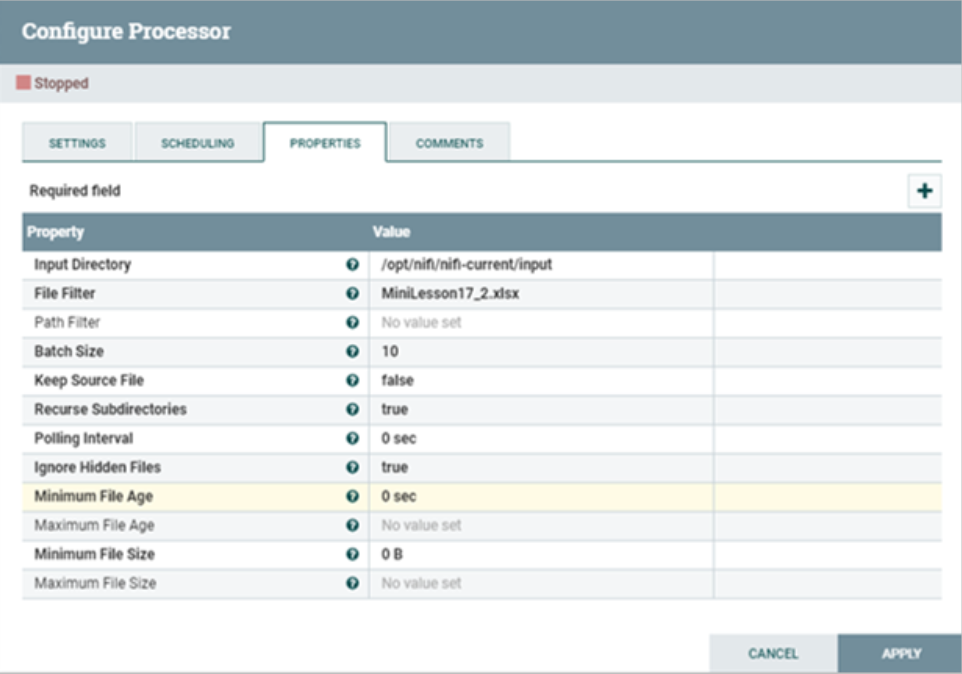
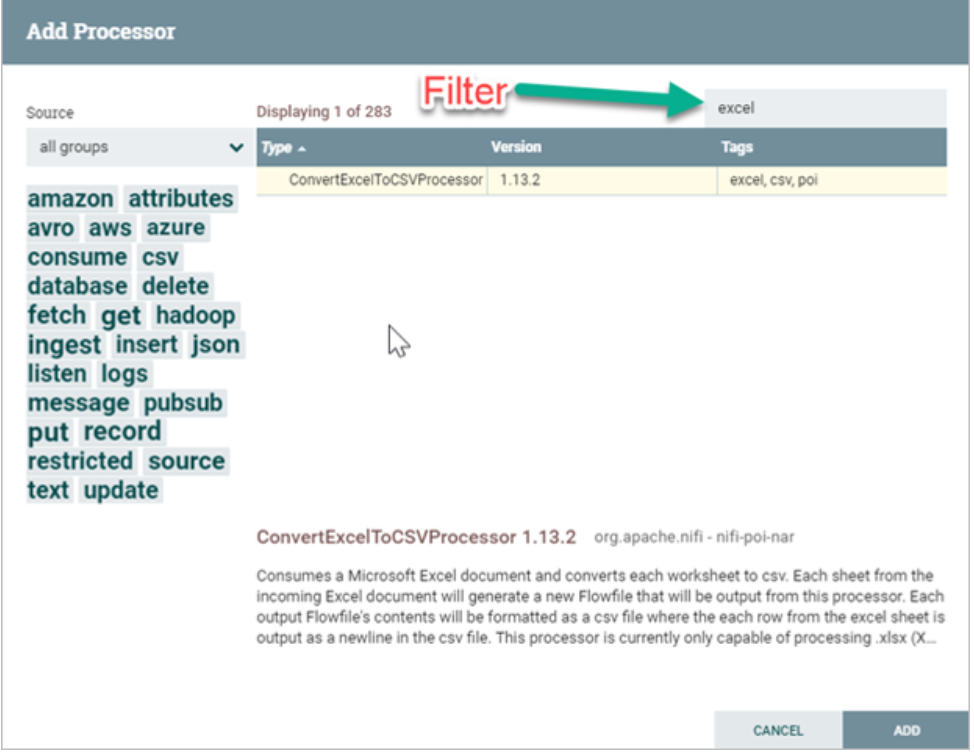
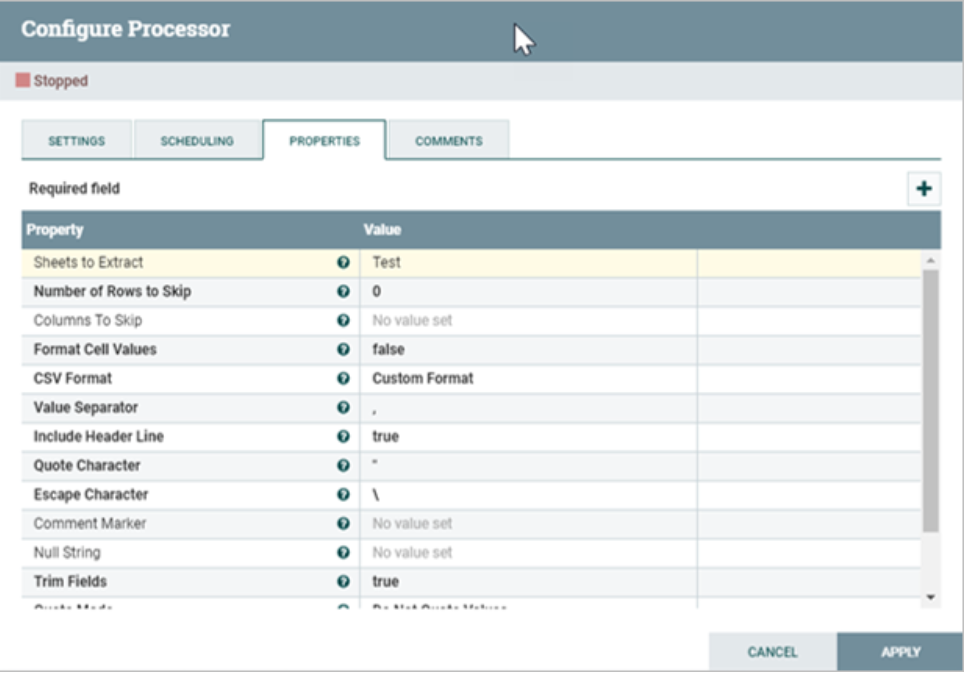
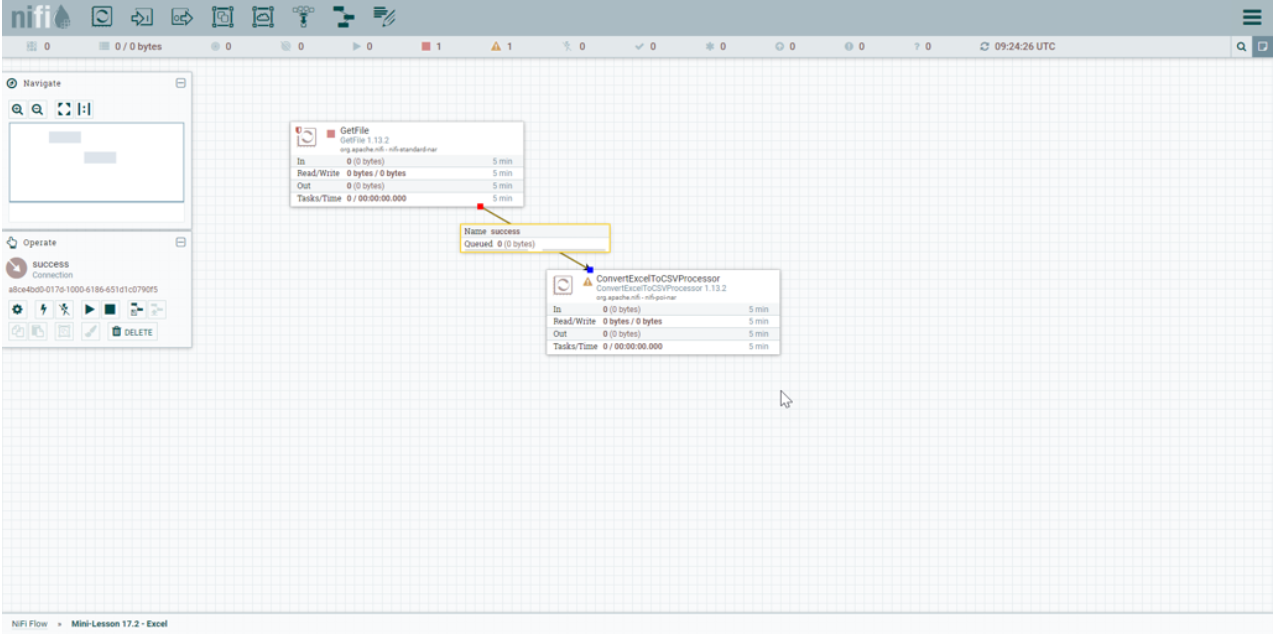
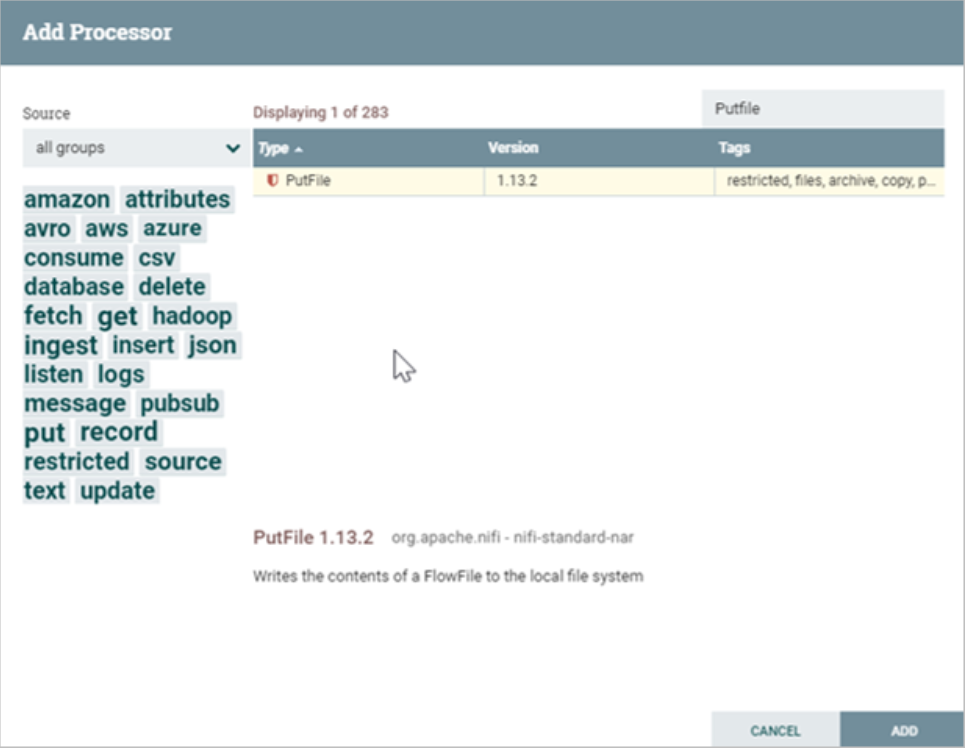
1. Open the NiFi CLI via the Docker GUI where it resides under "Container/Apps". By hovering over the *container* with your pointer, you will reveal a clickable icon to open the CLI Terminal. Select the icon to open the NiFi CLI.
2. In the bash command window, ensure you are in the /opt/nifi/nifi-current/ folder and create two folders within it that are titled input and output. The input folder will store the Excel files once they arrive in the system. The output folder will store the files after they have been processed in NiFi.
3. Use the Docker copy command below to copy the [MiniLesson17 2.xlsx](https://classroom.emeritus.org/courses/10605/files/3007359/download) file from your local machine into the input folder inside your NiFi Docker *container*.  
   docker cp Activity17-2.xlsx nificontainer:/opt/nifi/nifi-current/input

In your browser, navigate to http://localhost:8080/nifi/. If you don't see the NiFi app in your browser, be sure that you have opened the CLI via the Docker app GUI as instructed in Step 1. Then, you can open the browser from the same location in the GUI by selecting the "Open in Browser" icon.

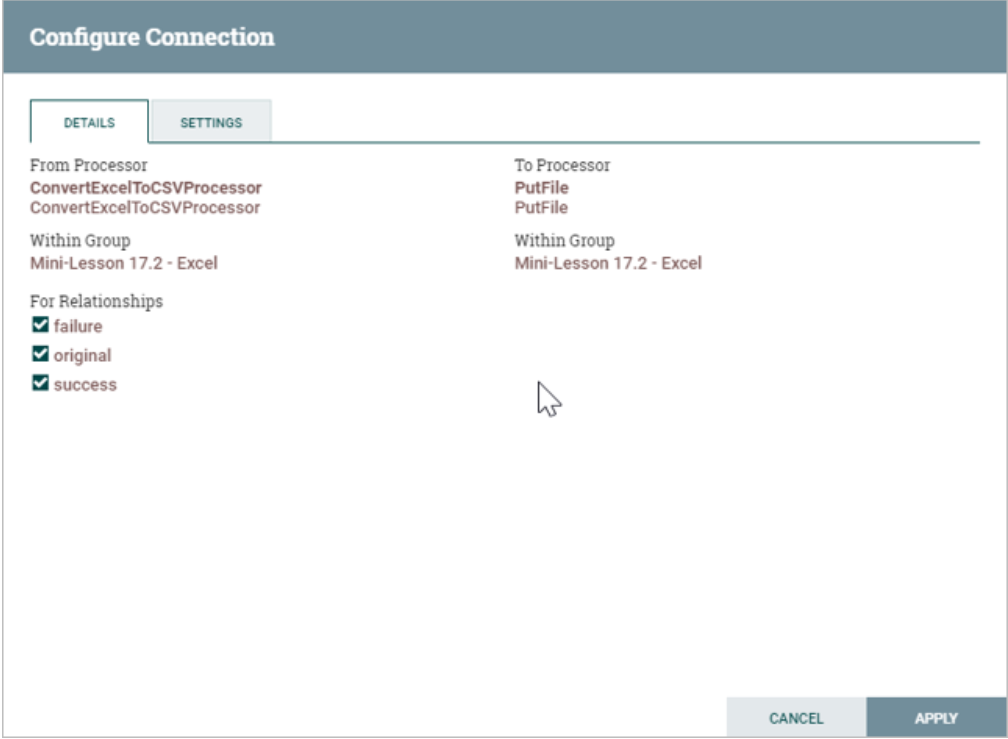
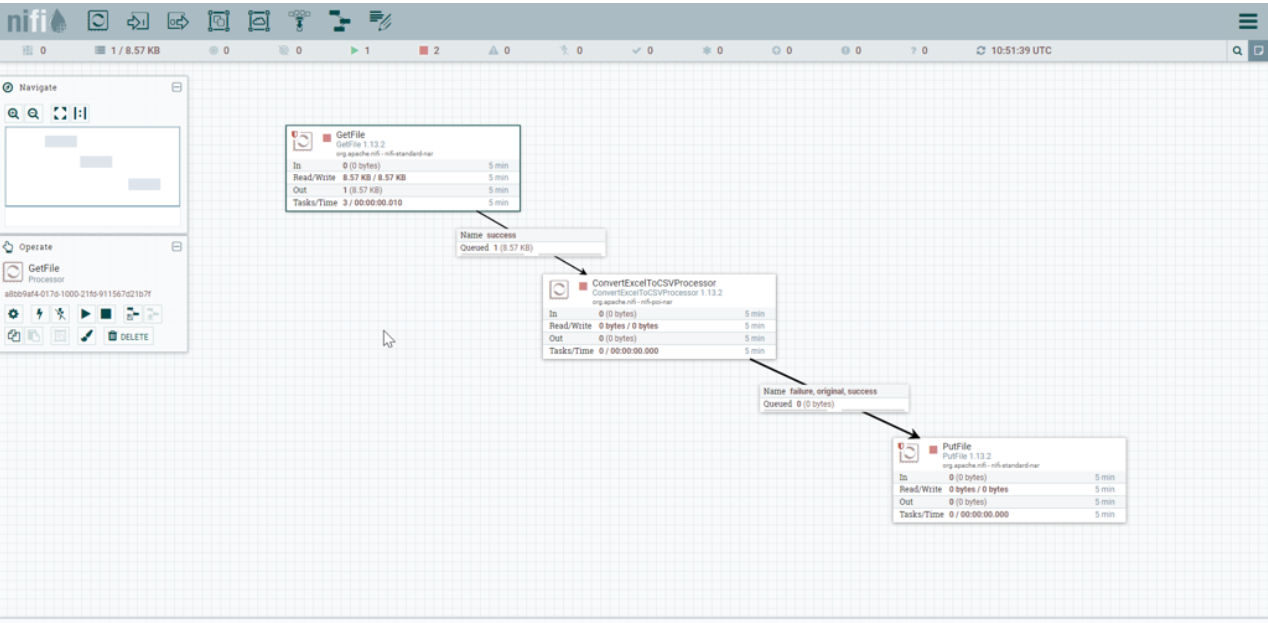
To start, you will need to create a process group, which is used to logically group *flows* and *processors* together. In NiFi, all workflows live together. Therefore, you can use a process group to help you organize your workflows.

Drag the process group icon onto the NiFi canvas and name it Mini-Lesson 17.2 - Excel. Select “Add” to add the process group:

Next, you will need to add *processors* to your workflow.

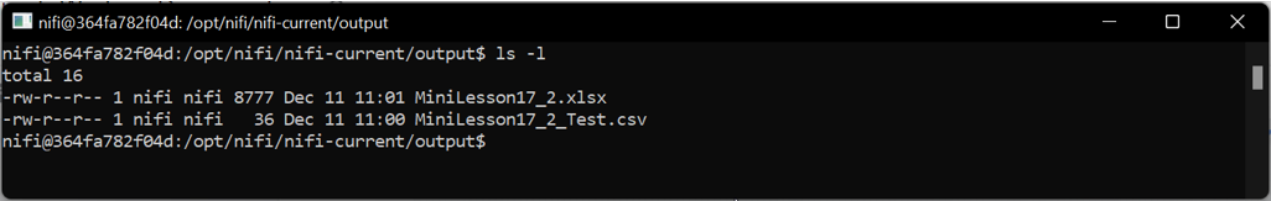
Double-click the Mini-Lesson 17.2 - Excel process group to get a blank canvas to start your workflow. Drag the *Processor* icon onto your blank NiFi canvas:Add the GetFile *processor* onto the canvas. This *processor* will be responsible for getting the Excel file from the input folder.Right-click on the GetFile *processor* and select Configure. In the SCHEDULING tab, set the Run Schedule field equal to 60 seconds.In the PROPERTIES tab, set the Input Directory field equal to /opt/nifi/nifi-current/input, and set the File Filter field equal to the Excel file name to be processed. For this example, the Excel file name is: Activity17\_2.xlsx.Add another *processor*, this time using the ConvertExcelToCSVProcessor *processor* type.Right-click on the ConvertExcelToCSVProcessor *processor* and select Configure. In the SCHEDULING tab, set the Run Schedule field equal to 60 seconds. In the PROPERTIES tab, set the Sheets to Extract field equal to Test.Connect the two *processors* by dragging the arrow from the GetFile *processor* to the ConvertExcelToCSVProcessor *processor:*a . Add another processor, PutFile, to your canvas. This processor will write the processed file inside the /opt/nifi/nifi-current/output folder in your bash command window.In the SCHEDULING tab, set the Run Schedule field equal to 60 seconds. In the PROPERTIES tab, set the Directory field equal to /opt/nifi/nifi-current/output.

Create a *connector* from the ConvertExcelToCSVProcessor *processor* to the PutFile *processor*. In the DETAILS tab, select failure, original, and success for the relationships.

Now you are ready to run your *flow.* Your NiFi canvas should contain three *processors* with two *connectors*, as shown below:For any given file run, if there are still files in *connector* queues (>0) be sure to clear the queue by right clicking on the background and select 'Empty all queues".

Right-click on the GetFile *processor* and select Start. Alternatively, select the *processor*, and then select the triangle icon that means Start in the left panel that will appear for the *processor*. The GetFile *processor* should start. It should read the file and remove it from the /opt/nifi/nifi-current/input *directory*.

Right-click on the ConvertExcelToCSVProcessor *processor* and select Start. Do the same for the PutFile *processor.* In your bash window, navigate to the /opt/nifi/nifi-current/output *directory*. Listing the contents of the output folder should display the CSV files created by the PutFile *processor*:

Notice that the PutFile *processor* created two files: a copy of the original file that you were reading and a CSV file named MiniLesson17\_2\_Test.csv. The name of this file is the original file name plus the sheet name set in the ConvertExcelToCSVProcessor *processor*.

In this mini-lesson, you learned how to read an Excel file and convert it to CSV format for further processing. This is the extract portion of the ETL process. The file is then transformed from Excel format to CSV format and copied to another folder for loading.