**Introduction**

In this tutorial, we are going to build a simple two processor flow. This dataflow aims to generated fake data to simulate a computer motherboard load. It will generated randomly memory and cpu usage as well the motherboard temperature. Then the Json generated data is stored on the local file system.

For information about building advanced data flows (using parameters, custom properties, controller services, or reporting tasks), see the [*Apache NiFi User Guide*](https://nifi.apache.org/docs/nifi-docs/html/user-guide.html).

**Steps To Build Your Nifi Flow**

**Apache NiFi** is an easy task. The following steps describe all the tasks you need to perform to create your flow.

1. Start by adding a Process Group to the canvas.
2. You can add all the processors and other components to build your data flow as you need.
3. Configure the processors according to the behavior you desire in your data flow.
4. Connect data flow components.
5. Start your data flow.

First of all, we will add a Process Group on the canvas, then we will add two processors and configure each of them. After configuring both the processor, we will connect and run them.

**Build The Nifi Flow**

In Apache NiFi, a user can maintain different data flows in different Process Groups. These groups can be based on different projects or the organizations, which Apache NiFi instance supports.

**Add a Process Group**

The fourth symbol in the menu at the top of the NiFi UI as shown in the above picture is used to add a Process Group in the NiFi canvas. Drag and Drop a new Process Group on the canvas and name it “First Nifi Flow”.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-add-process-group-1.png)

Process groups can be created in hierarchical manner to manage the data flows in better structure, which is easy to understand.

[A screenshot of a computer

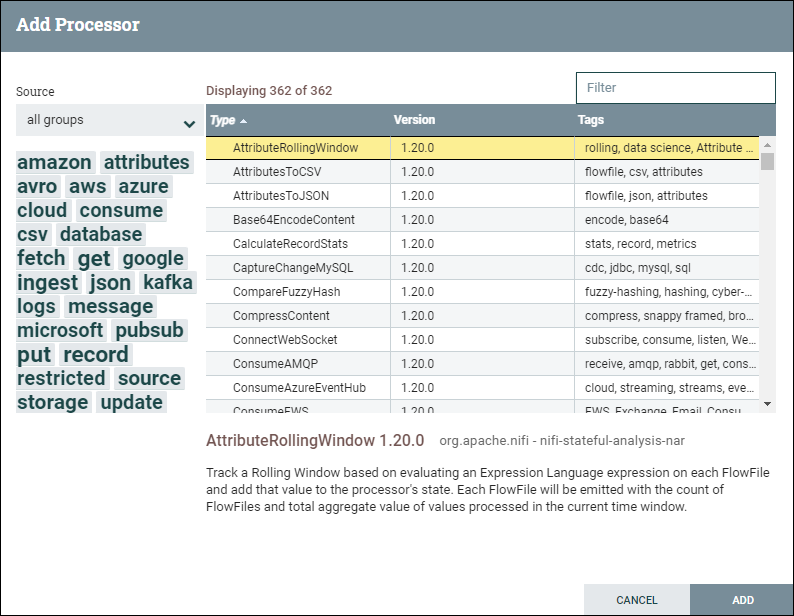
AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-procees-group-added.png)

Double click on the Process Group to enter into it. You can navigate back out and in by clicking on the corresponding level on the breadcrumb bar.

[](http://localhost/wp-content/uploads/2024/04/nifi-breadcombbar.png)

**Add Processors**

**Step 1:** To add a processor on the canvas, go to the component section in the toolbar and drag a processor component. This will open a Add Processor window, where you go through the list of processors.

[](http://localhost/wp-content/uploads/2024/04/Capture-decran-2024-04-23-205206-1.png)

**Step 2:** Find the required processor ***GenerateFlowFile*** or click on its tag under the Tag Cloud to reduce the list of processors by category and functionality that you are looking for a processor.

**Step 3:** Click on the processor that you want to select and add it to the canvas by clicking on the **Add** button. Similarly, drag the processor icon again, type the name of the processor you want and double-click to add it to the canvas.

**Step 4:** If you already know the processor name that you want, you can simply type the name of the processor up here in the filter bar. After that, double-click on the processor to add it to the canvas.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-generateflowfile-processor.png)

By following the above steps, search for the second processor to add on the canvas (***PutFile***).

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-first-flow-processors.png)

**Configure Processors**

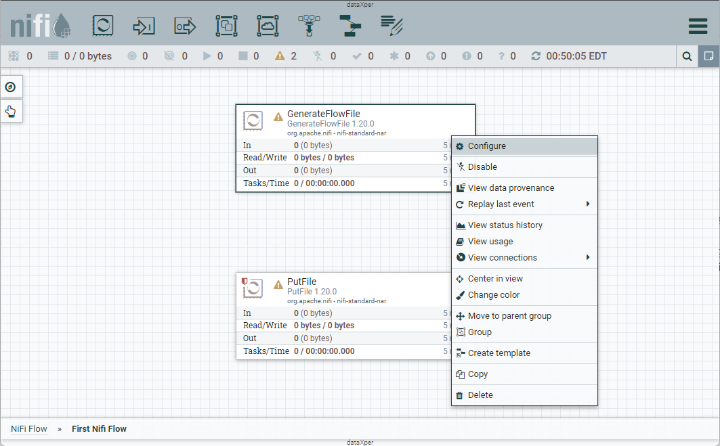
At this level, all the required processors are on the canvas. But you will see that both processors are invalid because they have a warning symbol in the upper left corner of the processor base. This is because you need to configure these processors. Hover the mouse over the warning icon, it will show the minimum requirements need to be configured to make processors valid and able to run.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-processor-warnings.png)

**Note:** To know more about the processor, right-click on it and go to View Usage option. It will show you the documentation for the processor.

To configure the processor, simply right click on the respective processor and go to **Configure**.

[](http://localhost/wp-content/uploads/2024/04/nifi-configuure-processor.png)

A window of Configure Processor with default values will pop up to you. This will bring up a new window containing five tabs, i.e., **Setting**, **Scheduling**, **Properties**, **Relationships** and **Comment**.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-configure-processor-settings.png)

**Configure GenerateFlowFile Processor**

To configure the **GenerateFlowFile** processor, right click on it and select **Configure**. This will open the Settings configuration window for this processor.

**Settings**

From the **Setting** tab you can change the name of the processor in **Name field**, because by default its name is the processor type. Each processor has a unique id number, which is not configurable.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-configure-processor-settings-name1.png)

**Scheduling**

The **Scheduling** tab defines:

* How to run
* How often to run
* How long to run.

Set the Run Schedule to **10 sec** because this processor can produce test files very fast. Leave all other fields as default for now.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-configure-processor-settings-scheduling1.png)

**Properties**

The **Properties** tab is the main tab where you configure the information that the processor needs to run properly, where properties that are not in bold letters are optional.

* From the **Value** column, click on the **Custom Text**row to change the content of this property and enter the following code snippet. This code will generate random values using the [Nifi Expression Language](https://nifi.apache.org/docs/nifi-docs/html/expression-language-guide.html" \t "_blank). The expected output is one row in Json format.

*{"id": "${UUID()}",*

*"memory": ${random():mod(95):plus(10)},*

*"cpu": ${nextInt()}.${random():mod(99):plus(1)},*

*"host": "${ip()}/${hostname(true)}",*

*"temperature": "${random():mod(60):plus(60)}",*

*"systemtime": "${now():format("MM/dd/yyyy HH:mm:ss", "EST")}"}*

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-configure-processor-settings-properties-text.png)

* Enter the code and click on the **OK** button.
* Repeat the same and set the value of the **Mime Type** as :  **application/json**

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-configure-processor-settings-properties-final.png)

**Relationships**

This processor only has success relationship. So, leave the Automatically Terminate Relationship unchecked because we want to continue to the next processor in the flow.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-configure-processor-settings-relationships.png)

**Comments**

In the **Comments** tab, type any comments to describe the processor and the tasks accomplished, like why you configured the processor or the way you did.

**Applying the Settings**

Now, click on the **Apply** button to save all the changes are made and complete the configuration of GenerateFlowFile processor.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-configure-processor-settings-properties-final-apply.png)

**Configure PutFile Processor**

Similar to the **GenerateFlowFile** processor, right-click on the **PutFile** processor and go to **Configure**, where we will change a couple of things that we didn’t in GenerateFlowFile processor. Here you will also get a new window containing five tabs, i.e., Settings, Scheduling, Properties, Relationships and Comments.

* Go to the Setting tab and change the name of the processor : **Save data to disk**.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-configure-putprocessor-name1.png)

**Settings**

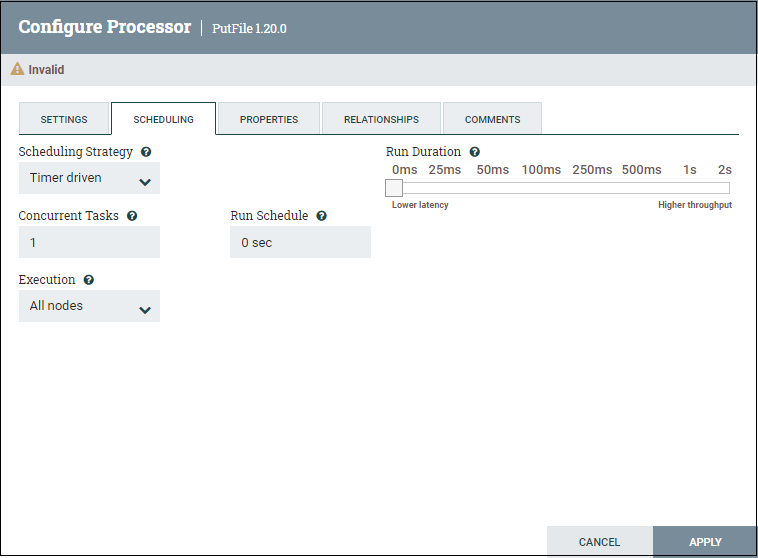
From the **Setting** tab you can change the name of the processor in **Name field**, because by default its name is the processor type. Each processor has a unique id number, which is not configurable.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-configure-processor-settings-name1.png)

**Scheduling**

Nothing to do here for this processor. We will keep all the default settings values for now.

[](http://localhost/wp-content/uploads/2024/04/nifi-configure-putprocessor-scheduling.png)

**Properties**

From the **Properties** tab we will set only the output directory as this is a required property. We will keep all the other properties  default values.

* From the **Value** column, click on the **Directory**row to change the content of this property and enter the output directory.  This is the directory is where Nifi is expected to write the generated data.

**Directory** = **/home/training/tutorials/nifi/firstflow**

A screenshot of a computer

AI-generated content may be incorrect.

**Relationships**

This processor has two relationships : failure and success. You need to Automatically Terminate Relationship because the dataflow ends here and there is no other next processor in the flow.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-configure-putprocessor-relashionships.png)

**Comments**

In the **Comments** tab, type any comments to describe the processor and the tasks accomplished, like why you configured the processor or the way you did.

**Applying the Settings**

Now, click on the **Apply** button to save all the changes are made and complete the configuration of **PutFile** processor.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-configure-putprocessor-properties-apply.png)

After completing the above steps, you can see that both the processors are still invalid. This is because we have not connected them. So, we will now connect both processors and run.

**Connect And Run Processors**

Now our processors are configured and ready to run, you need to connect them together.

**Connecting**

**Step 1:** To connect the processors to each other, hover the mouse over the center of the processor, an aero in a circle will show. Drag the mouse from that circle to another processor until it highlighted in green.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-connecting-processors.png)

**Step 2:** Release the mouse here. A **Create Connection** window will open containing details and setting tab. The Details tab shows what the connection is going from and to. It also shows the list of relationships that will be included in the connection. In our case, the Details tab window shows that a connection is created from the **GenerateFlowFile** processor to the **PutFile** processor using its **success** relationship.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-connectiom-settings2.png)

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-connectiom-settings1.png)

Leave the current settings as default for now and click on the **ADD** button.

**Ready to Run**

Now, you can see that both the processors are valid now as they have a stop symbol in place of warning symbol at the upper left-hand corner

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-dataflow-ready-to-run1.png)

**Start the Dataflow**

To run the dataflow (both processors) , open the left panel and click on the **Start** button present under the Operate section.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-run-dataflow.png)

You can also choose to run each processor individually. Just right click on the processor and select **Start** from the contextual menu.

**Dataflow Statistics**

Now you started the dataflow, you can see that the **GenerateFlowFile** is producing data. This data is consumed by the **PutFile** processor and written to disk in the output directory as expected. On each processor we can see the amount of Read/Write data. This information represents what happened in the processors over the last five minutes.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-dataflow-running-stats1.png)

**View Queued Elements**

To view the queued flowfiles, just right click the connection and select View Queue from the contextual menu. This will open a new window that shows a list of all the flowfiles in the queue.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-queue-flowfiles-list.png)

From here you can choose any row and click on the **eye** icon to view the content of the flowfile.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-queue-flowfiles-content.png)

**Stop the Dataflow**

To stop the dataflow, click on the **Stop** icon in the Operate panel or right click on any empty space on the canvas and select **Stop**from the contextual menu.

[A screenshot of a computer

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-datalow-stop1.png)

**Check the result**

The dataflow purpose is to generate data in Json format and store this data on the local machine in the folder : **/home/training/tutorials/nifi/firstflow**

To check the output, SSH to the sandbox and list the content of the output directory. You should see a list of files outputted by the Nifi dataflow.

***$ ls /home/training/tutorials/nifi/firstflow***

[A screen shot of a computer screen

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/nifi-dataflow-output-fileslist.png)

To view the content of one file, run the following command:

***$ cd /home/training/tutorials/nifi/firstflow/***

***$ cat***

[A computer screen shot of white text

AI-generated content may be incorrect.](http://localhost/wp-content/uploads/2024/04/Capture-decran-2024-04-24-143708.png)

**Note**: The filenames are generated randomly and are supposed to be unique. The output filenames you will get might be different.

**Summary**

In this tutorial we covered all the steps you need to create your first Nifi dataflow. This dataflow aims to generate fake data in Json format and save this data into a directory on the local file system. We also covered how to start and stop the dataflow, how to explore the processors statistics and how to view the flowfiles queued in the connections buffers.

Download Nifi Template (right click and Save Link As…)

[First Nifi Flow](http://localhost/wp-content/uploads/2024/04/First_Nifi_Flow.xml)