

Apache NiFi 3

Monitoring a DataFlow

Date of Publish: 2018-12-18



<https://docs.hortonworks.com/>

Contents

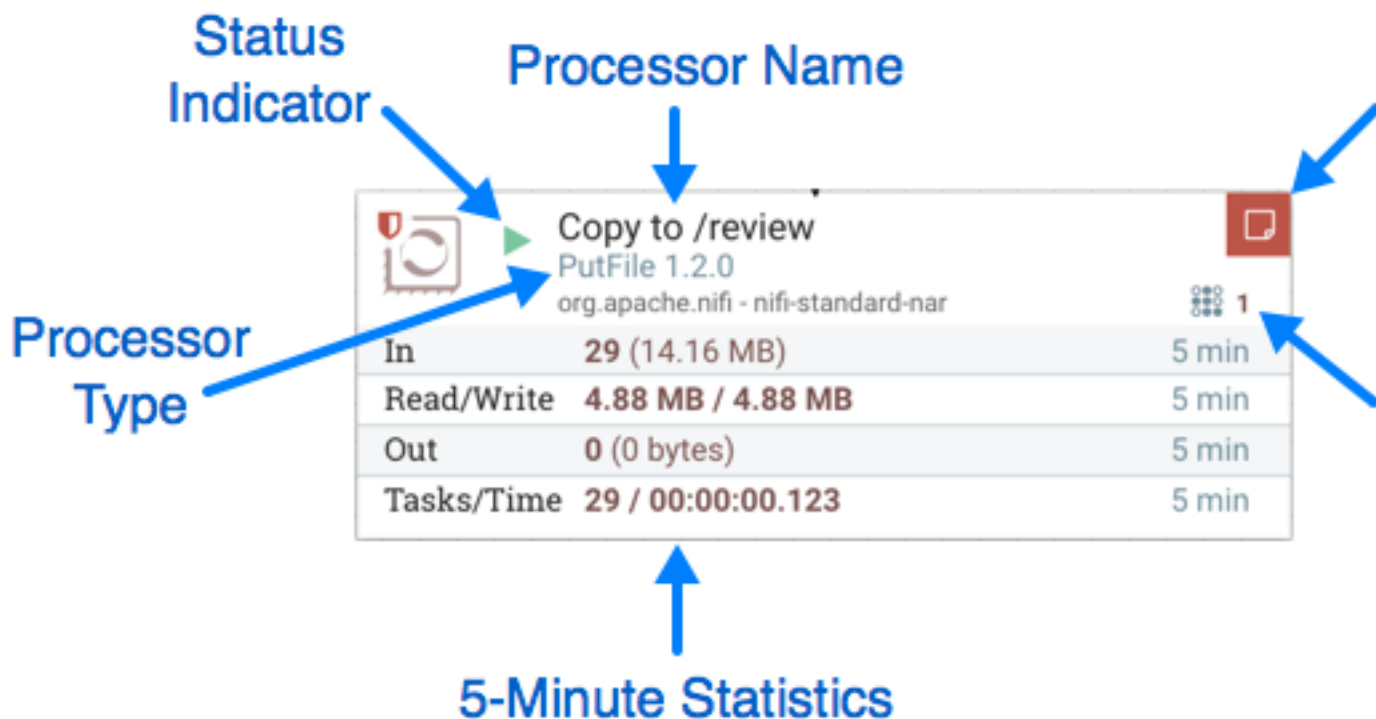
Monitoring of DataFlow.....	3
Anatomy of a Processor.....	3
Anatomy of a Process Group.....	5
Anatomy of a Remote Process Group.....	6
Queue Interaction.....	8
Summary Page.....	8
Historical Statistics of a Component.....	11

Monitoring of DataFlow

NiFi provides a great deal of information about the DataFlow in order to monitor its health and status. The Status bar provides information about the overall system health. Processors, Process Groups, and Remote Process Groups provide fine-grained details about their operations. Connections and Process Groups provide information about the amount of data in their queues. The Summary Page provides information about all of the components on the canvas in a tabular format and also provides System Diagnostics that include disk usage, CPU utilization, and Java Heap and Garbage Collection information. In a clustered environment, this information is available per-node or as aggregates across the entire cluster. We will explore each of these monitoring artifacts below.

Anatomy of a Processor





NiFi provides a significant amount of information about each Processor on the canvas. The following diagram shows the anatomy of a Processor:



The image outlines the following elements:

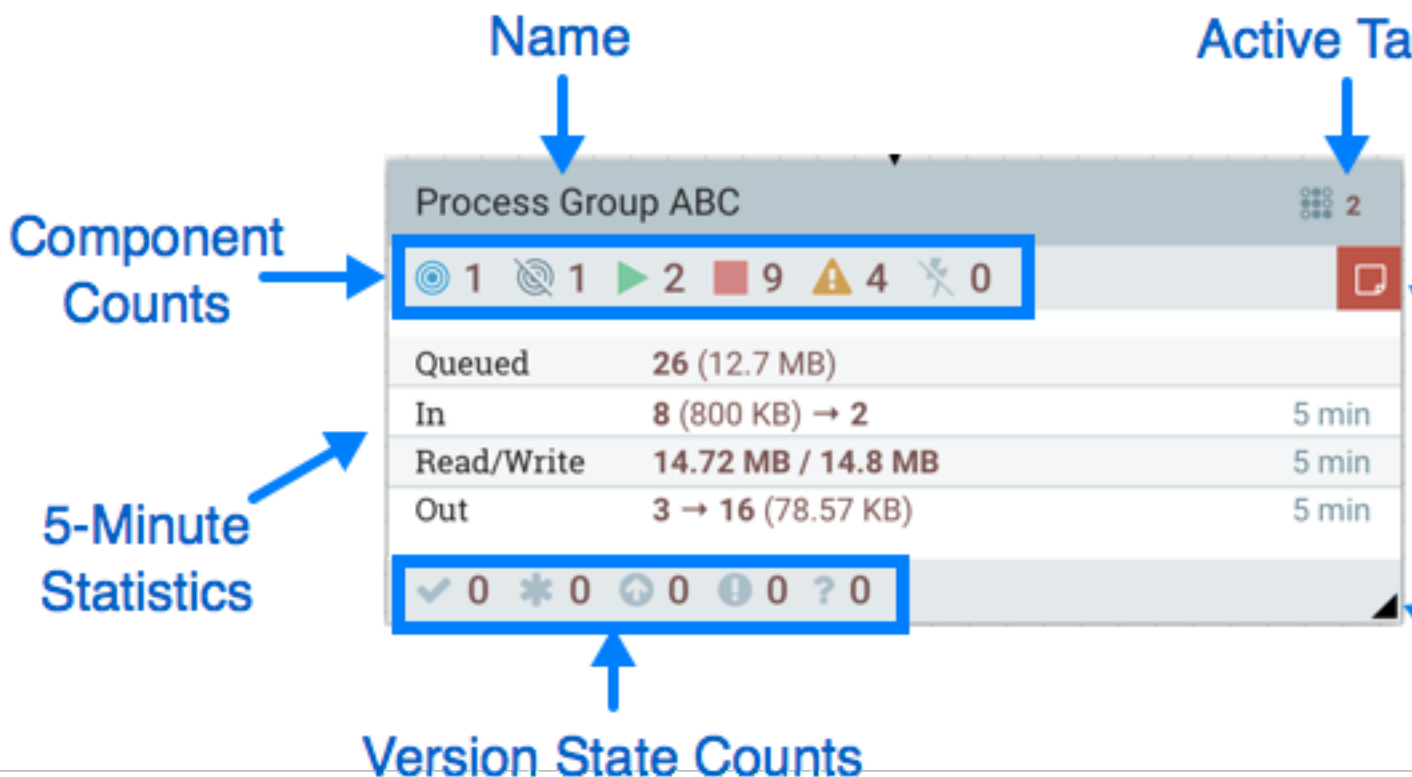
- **Processor Type:** NiFi provides several different types of Processors in order to allow for a wide range of tasks to be performed. Each type of Processor is designed to perform one specific task. The Processor type (PutFile, in this example) describes the task that this Processor performs. In this case, the Processor writes a FlowFile to disk - or "Puts" a FlowFile to a File.
- **Bulletin Indicator:** When a Processor logs that some event has occurred, it generates a Bulletin to notify those who are monitoring NiFi via the User Interface. The DFM is able to configure which bulletins should be displayed in the User Interface by updating the "Bulletin level" field in the "Settings" tab of the Processor configuration dialog. The default value is WARN, which means that only warnings and errors will be displayed in the UI. This icon is not present unless a Bulletin exists for this Processor. When it is present, hovering over the icon with the mouse will provide a tooltip explaining the message provided by the Processor as well as the Bulletin level. If the

instance of NiFi is clustered, it will also show the Node that emitted the Bulletin. Bulletins automatically expire after five minutes.

- Status Indicator: Shows the current Status of the Processor. The following indicators are possible:
 -  Running: The Processor is currently running.
 -  Stopped: The Processor is valid and enabled but is not running.
 -  Invalid: The Processor is enabled but is not currently valid and cannot be started. Hovering over this icon will provide a tooltip indicating why the Processor is not valid.
 -  Disabled: The Processor is not running and cannot be started until it has been enabled. This status does not indicate whether or not the Processor is valid.
- Processor Name: This is the user-defined name of the Processor. By default, the name of the Processor is the same as the Processor Type. In the example, this value is "Copy to /review".
- Active Tasks: The number of tasks that this Processor is currently executing. This number is constrained by the "Concurrent tasks" setting in the "Scheduling" tab of the Processor configuration dialog. Here, we can see that the Processor is currently performing one task. If the NiFi instance is clustered, this value represents the number of tasks that are currently executing across all nodes in the cluster.
- 5-Minute Statistics: The Processor shows several different statistics in tabular form. Each of these statistics represents the amount of work that has been performed in the past five minutes. If the NiFi instance is clustered, these values indicate how much work has been done by all of the Nodes combined in the past five minutes. These metrics are:
 - In: The amount of data that the Processor has pulled from the queues of its incoming Connections. This value is represented as <count> (<size>) where <count> is the number of FlowFiles that have been pulled from the queues and <size> is the total size of those FlowFiles' content. In this example, the Processor has pulled 29 FlowFiles from the input queues, for a total of 14.16 megabytes (MB).
 - Read/Write: The total size of the FlowFile content that the Processor has read from disk and written to disk. This provides valuable information about the I/O performance that this Processor requires. Some Processors may only read the data without writing anything while some will not read the data but will only write data. Others will neither read nor write data, and some Processors will both read and write data. In this example, we see that in the past five minutes, this Processor has read 4.88 MB of the FlowFile content and has written 4.88 MB as well. This is what we would expect, since this Processor simply copies the contents of a FlowFile to disk. Note, however, that this is not the same as the amount of data that it pulled from its input queues. This is because some of the files that it pulled from the input queues already exist in the output directory, and the Processor is configured to route FlowFiles to failure when this occurs. Therefore, for those files which already existed in the output directory, data was neither read nor written to disk.
 - Out: The amount of data that the Processor has transferred to its outbound Connections. This does not include FlowFiles that the Processor removes itself, or FlowFiles that are routed to connections that are auto-terminated. Like the "In" metric above, this value is represented as <count> (<size>) where <count> is the number of FlowFiles that have been transferred to outbound Connections and <size> is the total size of those FlowFiles' content. In this example, all of the Relationships are configured to be auto-terminated, so no FlowFiles are reported as having been transferred Out.
 - Tasks/Time: The number of times that this Processor has been triggered to run in the past 5 minutes, and the amount of time taken to perform those tasks. The format of the time is <hour>:<minute>:<second>. Note that the amount of time taken can exceed five minutes, because many tasks can be executed in parallel. For instance, if the Processor is scheduled to run with 60 Concurrent tasks, and each of those tasks takes one second to complete, it is possible that all 60 tasks will be completed in a single second. However, in this case we will see the Time metric showing that it took 60 seconds, instead of 1 second. This time can be thought of as "System Time," or said another way, this value is 60 seconds because that's the amount of time it would have taken to perform the action if only a single concurrent task were used.

Anatomy of a Process Group







The Process Group provides a mechanism for grouping components together into a logical construct in order to organize the DataFlow in a way that makes it more understandable from a higher level. The following image highlights the different elements that make up the anatomy of a Process Group:



The Process Group consists of the following elements:

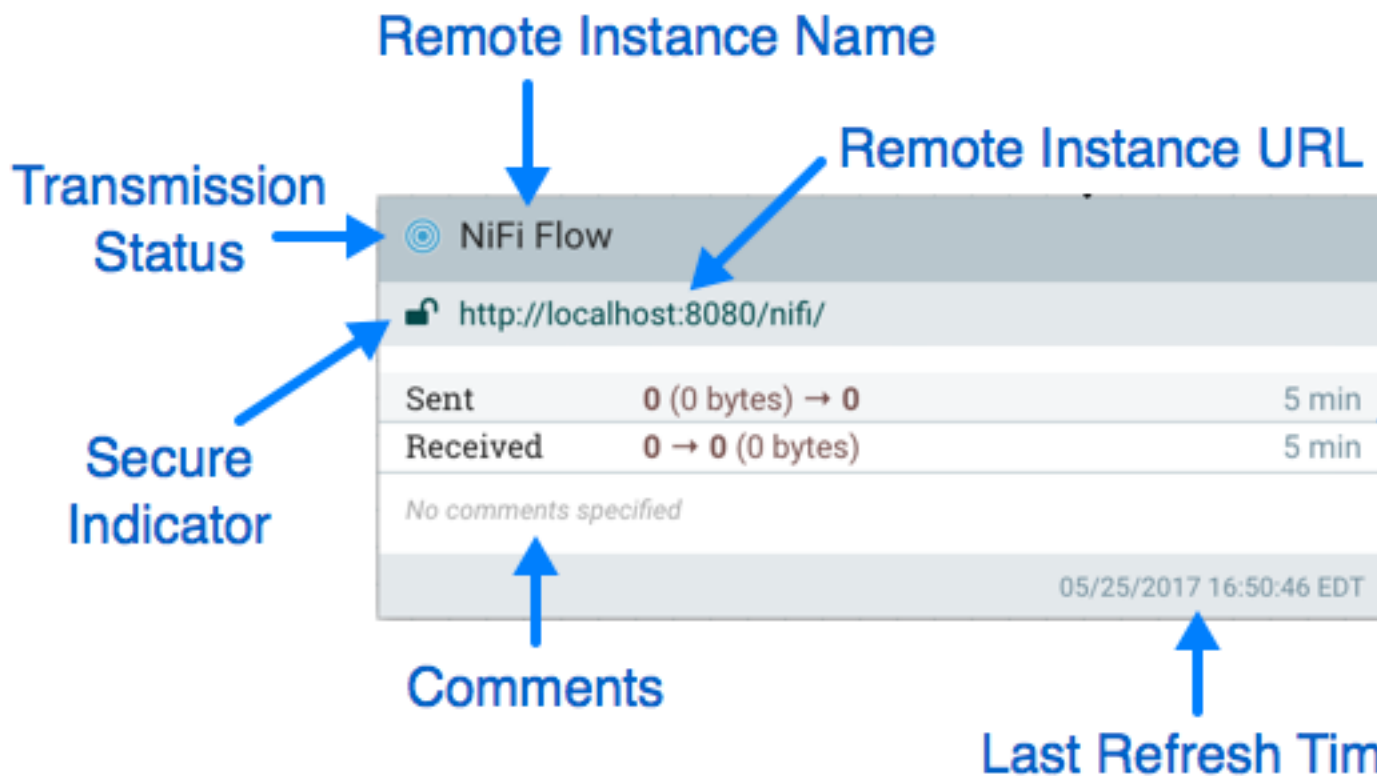
- **Name:** This is the user-defined name of the Process Group. This name is set when the Process Group is added to the canvas. The name can later be changed by right-clicking on the Process Group and clicking the "Configure" menu option. In this example, the name of the Process Group is "Process Group ABC."
- **Bulletin Indicator:** When a child component of a Process Group emits a bulletin, that bulletin is propagated to the component's parent Process Group, as well. When any component has an active Bulletin, this indicator will appear, allowing the user to hover over the icon with the mouse to see the Bulletin.
- **Active Tasks:** The number of tasks that are currently executing by the components within this Process Group. Here, we can see that the Process Group is currently performing two tasks. If the NiFi instance is clustered, this value represents the number of tasks that are currently executing across all nodes in the cluster.
- **Statistics:** Process Groups provide statistics about the amount of data that has been processed by the Process Group in the past 5 minutes as well as the amount of data currently enqueued within the Process Group. The following elements comprise the "Statistics" portion of a Process Group:
 - **Queued:** The number of FlowFiles currently enqueued within the Process Group. This field is represented as <count> (<size>) where <count> is the number of FlowFiles that are currently enqueued in the Process Group

and <size> is the total size of those FlowFiles' content. In this example, the Process Group currently has 26 FlowFiles enqueued with a total size of 12.7 megabytes (MB).





- **In:** The number of FlowFiles that have been transferred into the Process Group through all of its Input Ports over the past 5 minutes. This field is represented as <count> / <size> # <ports> where <count> is the number of FlowFiles that have entered the Process Group in the past 5 minutes, <size> is the total size of those FlowFiles' content and <ports> is the number of Input Ports. In this example, 8 FlowFiles have entered the Process Group with a total size of 800 KB and two Input Ports exist.
- **Read/Write:** The total size of the FlowFile content that the components within the Process Group have read from disk and written to disk. This provides valuable information about the I/O performance that this Process Group requires. In this example, we see that in the past five minutes, components within this Process Group have read 14.72 MB of the FlowFile content and have written 14.8 MB.
- **Out:** The number of FlowFiles that have been transferred out of the Process Group through its Output Ports over the past 5 minutes. This field is represented as <ports> # <count> (<size>) where <ports> is the number of Output Ports, <count> is the number of FlowFiles that have exited the Process Group in the past 5 minutes and <size> is the total size of those FlowFiles' content. In this example, there are three Output Ports, 16 FlowFiles have exited the Process Group and their total size is 78.57 KB.
- **Component Counts:** The Component Counts element provides information about how many components of each type exist within the Process Group. The following provides information about each of these icons and their meanings:
 -  **Transmitting Ports:** The number of Remote Process Group Ports that currently are configured to transmit data to remote instances of NiFi or pull data from remote instances of NiFi.
 -  **Non-Transmitting Ports:** The number of Remote Process Group Ports that are currently connected to components within this Process Group but currently have their transmission disabled.
 -  **Running Components:** The number of Processors, Input Ports, and Output Ports that are currently running within this Process Group.
 -  **Stopped Components:** The number of Processors, Input Ports, and Output Ports that are currently not running but are valid and enabled. These components are ready to be started.
 -  **Invalid Components:** The number of Processors, Input Ports, and Output Ports that are enabled but are currently not in a valid state. This may be due to misconfigured properties or missing Relationships.
 -  **Disabled Components:** The number of Processors, Input Ports, and Output Ports that are currently disabled. These components may or may not be valid. If the Process Group is started, these components will not cause any errors but will not be started.
- **Version State Counts:** The Version State Counts element provides information about how many versioned process groups are within the Process Group.
- **Comments:** When the Process Group is added to the canvas, the user is given the option of specifying Comments in order to provide information about the Process Group. The comments can later be changed by right-clicking on the Process Group and clicking the "Configure" menu option.

Anatomy of a Remote Process Group

When creating a DataFlow, it is often necessary to transfer data from one instance of NiFi to another. In this case, the remote instance of NiFi can be thought of as a Process Group. For this reason, NiFi provides the concept of a Remote Process Group. From the User Interface, the Remote Process Group looks similar to the Process Group. However, rather than showing information about the inner workings and state of a Remote Process Group, such as queue sizes, the information rendered about a Remote Process Group is related to the interaction that occurs between this instance of NiFi and the remote instance.




The image above shows the different elements that make up a Remote Process Group. Here, we provide an explanation of the icons and details about the information provided.

- **Transmission Status:** The Transmission Status indicates whether or not data Transmission between this instance of NiFi and the remote instance is currently enabled. The icon shown will be the Transmission Enabled icon () if any of the Input Ports or Output Ports is currently configured to transmit or the Transmission Disabled icon () if all of the Input Ports and Output Ports that are currently connected are stopped.
- **Remote Instance Name:** This is the name of the NiFi instance that was reported by the remote instance. When the Remote Process Group is first created, before this information has been obtained, the URL of the remote instance will be shown here instead.
- **Remote Instance URL:** This is the URL of the remote instance that the Remote Process Group points to. This URL is entered when the Remote Process Group is added to the canvas and it cannot be changed.
- **Secure Indicator:** This icon indicates whether or not communications with the remote NiFi instance are secure. If communications with the remote instance are secure, this will be indicated by the "Locked" icon (). If the communications are not secure, this will be indicated by the "Unlocked" icon (). If the communications are secure, this instance of NiFi will not be able to communicate with the remote instance until an administrator for the remote instance grants access. Whenever the Remote Process Group is added to the canvas, this will automatically initiate a request to have a user for this instance of NiFi created on the remote instance. This instance will be unable to communicate with the remote instance until an administrator on the remote instance adds the user to the system and adds the "NiFi" role to the user. In the event that communications are not secure, the Remote Process Group is able to receive data from anyone, and the data is not encrypted while it is transferred between instances of NiFi.

- **5-Minute Statistics:** Two statistics are shown for Remote Process Groups: Sent and Received. Both of these are in the format <count> (<size>) where <count> is the number of FlowFiles that have been sent or received in the previous five minutes and <size> is the total size of those FlowFiles' content.
- **Comments:** The Comments that are provided for a Remote Process Group are not comments added by the users of this NiFi but rather the Comments added by the administrators of the remote instance. These comments indicate the purpose of the NiFi instance as a whole.
- **Last Refreshed Time:** The information that is pulled from a remote instance and rendered on the Remote Process Group in the User Interface is periodically refreshed in the background. This element indicates the time at which that refresh last happened, or if the information has not been refreshed for a significant amount of time, the value will change to indicate Remote flow not current. NiFi can be triggered to initiate a refresh of this information by right-clicking on the Remote Process Group and choosing the "Refresh flow" menu item.

Queue Interaction

The FlowFiles enqueued in a Connection can be viewed when necessary. The Queue listing is opened via List queue in a Connection's context menu. The listing will return the top 100 FlowFiles in the active queue according to the configured priority. The listing can be performed even if the source and destination are actively running.

Additionally, details for a Flowfile in the listing can be viewed by clicking on the Details icon () in the left most column. From here, the FlowFile details and attributes are available as well buttons for downloading or viewing the content. Viewing the content is only available if the `nifi.content.viewer.url` has been configured. If the source or destination of the Connection are actively running, there is a chance that the desired FlowFile will no longer be in the active queue.

The FlowFiles enqueued in a Connection can also be deleted when necessary. The removal of the FlowFiles is initiated via Empty queue in the Connection's context menu. This action can also be performed if the source and destination are actively running.

Summary Page


While the NiFi canvas is useful for understanding how the configured DataFlow is laid out, this view is not always optimal when trying to discern the status of the system. In order to help the user understand how the DataFlow is functioning at a higher level, NiFi provides a Summary page. This page is available in the Global Menu in the top-right corner of the User Interface.
























The Summary Page is opened by selecting Summary from the Global Menu. This opens the Summary table dialog:


NiFi Summary

PROCESSORS
INPUT PORTS
OUTPUT PORTS
REMOTE PROCESS GROUPS

Displaying 60 of 60

Filter by name 

	Name 	Type	Run Status	In / Size 5 min
	Base64EncodeCo...	Base64EncodeCo...	 Invalid	0 (0 bytes)
	Capture Network ...	ExecuteProcess	 Running (5)	0 (0 bytes)
	Check if Dataset ...	IdentifyMimeType	 Running	0 (0 bytes)
	Decompress Data...	CompressContent	 Running	0 (0 bytes)
	Detect Interesting...	RouteOnAttribute	 Running (2)	11,741,700 (13 GB)
	Empty Contents	ReplaceText	 Running	0 (0 bytes)
	EvaluateJsonPath	EvaluateJsonPath	 Running	0 (0 bytes)
	EvaluateJsonPath	EvaluateJsonPath	 Running (4)	11,742,510 (13 GB)
	Extract File ID	UpdateAttribute	 Running	0 (0 bytes)
	Extract File ID	UpdateAttribute	 Running	0 (0 bytes)
	ExtractText	ExtractText	 Running	1,498,914 (244.45 MB)























 Last updated: **17:45:24 UTC**

This dialog provides a great deal of information about each of the components on the canvas. Below, we have annotated the different elements within the dialog in order to make the discussion of the dialog easier.


NiFi Summary

PROCESSORS	INPUT PORTS	OUTPUT PORTS	REMOTE PROCESS GROUPS
-------------------	--------------------	---------------------	------------------------------

Displaying 60 of 60

Filter	by name			▼
Name ▲	Type	Run Status	In / Size 5 min	
 Base64EncodeCo...	Base64EncodeCo...	 Invalid	0 (0 bytes)	
 Capture Network ...	ExecuteProcess	 Running (5)	0 (0 bytes)	
 Check if Dataset ...	IdentifyMimeType	 Running	0 (0 bytes)	
 DecompressContent	DecompressContent	 Running	0 (0 bytes)	
 DetectDuplicate	DetectDuplicate	 Running (2)	11,741,700 (13 C)	
 Empty Contents	ReplaceText	 Running	0 (0 bytes)	
 EvaluateJsonPath	EvaluateJsonPath	 Running	0 (0 bytes)	
 EvaluateJsonPath	EvaluateJsonPath	 Running (4)	11,742,510 (13 C)	
 Extract File ID	UpdateAttribute	 Running	0 (0 bytes)	
 Extract File ID	UpdateAttribute	 Running	0 (0 bytes)	
 ExtractText	ExtractText	 Running	1,498,914 (244.4)	

Show Processor Details

 Last updated: 17:45:24 UTC

The Summary page is primarily comprised of a table that provides information about each of the components on the canvas. Above this table is a set of five tabs that can be used to view the different types of components. The information provided in the table is the same information that is provided for each component on the canvas. Each of the columns in the table may be sorted by clicking on the heading of the column.

The Summary page also includes the following elements:

- **Bulletin Indicator:** As in other places throughout the User Interface, when this icon is present, hovering over the icon will provide information about the Bulletin that was generated, including the message, the severity level, the time at which the Bulletin was generated, and (in a clustered environment) the node that generated the Bulletin.

Like all the columns in the Summary table, this column where bulletins are shown may be sorted by clicking on the heading so that all the currently existing bulletins are shown at the top of the list.

- **Details:** Clicking the Details icon will provide the user with the details of the component. This dialog is the same as the dialog provided when the user right-clicks on the component and chooses the "View Configuration" menu item.
- **Go To:** Clicking this button will close the Summary page and take the user directly to the component on the NiFi canvas. This may change the Process Group that the user is currently in. This icon is not available if the Summary page has been opened in a new browser tab or window (by clicking the "Pop Out" button, as described below).
- **Status History:** Clicking the Status History icon will open a new dialog that shows a historical view of the statistics that are rendered for this component.
- **Refresh:** The Refresh button allows the user to refresh the information displayed without closing the dialog and opening it again. The time at which the information was last refreshed is shown just to the right of the Refresh button. The information on the page is not automatically refreshed.
- **Filter:** The Filter element allows users to filter the contents of the Summary table by typing in all or part of some criteria, such as a Processor Type or Processor Name. The types of filters available differ according to the selected tab. For instance, if viewing the Processor tab, the user is able to filter by name or by type. When viewing the Connections tab, the user is able to filter by source, by name, or by destination name. The filter is automatically applied when the contents of the text box are changed. Below the text box is an indicator of how many entries in the table match the filter and how many entries exist in the table.
- **Pop-Out:** When monitoring a flow, it is helpful to be able to open the Summary table in a separate browser tab or window. The Pop-Out button, next to the Close button, will cause the entire Summary dialog to be opened in a new browser tab or window (depending on the configuration of the browser). Once the page is "popped out", the dialog is closed in the original browser tab/window. In the new tab/window, the Pop-Out button and the Go-To button will no longer be available.
- **System Diagnostics:** The System Diagnostics window provides information about how the system is performing with respect to system resource utilization. While this is intended mostly for administrators, it is provided in this view because it does provide a summary of the system. This dialog shows information such as CPU utilization, how full the disks are, and Java-specific metrics, such as memory size and utilization, as well as Garbage Collection information.

Historical Statistics of a Component

While the Summary table and the canvas show numeric statistics pertaining to the performance of a component over the past five minutes, it is often useful to have a view of historical statistics as well. This information is available by right-clicking on a component and choosing the "Status History" menu option or by clicking on the Status History in the Summary page.

The amount of historical information that is stored is configurable in the NiFi properties but defaults to 24 hours. When the Status History dialog is opened, it provides a graph of historical statistics:

Status History

Status History

Id
b0a3b095-5408-38b8-b0dd-7b31babb2e04

Group Id
50380170-a353-47ad-8b01-14f3ac4b4551

Name
Twitter Garden Host

Type
GetTwitter

Start
07/28/2016 19:13:47.262


End
07/28/2016 22:47:51.162

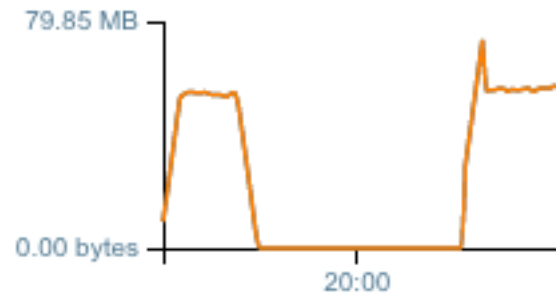
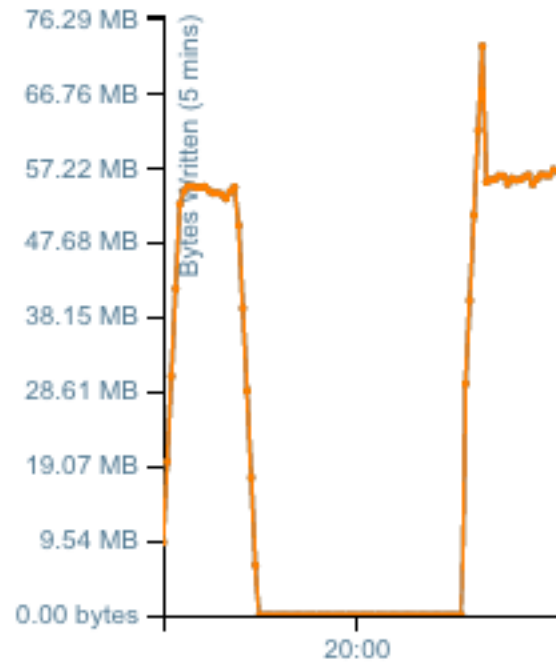
NiFi
Min / Max / Mean
0.00 bytes / 76.78 MB / 35.89 MB

NiFi

Nodes
Min / Max / Mean
0.00 bytes / 76.78 MB / 35.89 MB

nifi-04:8080

 Last updated: 22:48:13 UTC




The left-hand side of the dialog provides information about the component that the stats are for, as well as a textual representation of the statistics being graphed. The following information is provided on the left-hand side:

- Id: The ID of the component for which the stats are being shown.
- Group Id: The ID of the Process Group in which the component resides.
- Name: The Name of the Component for which the stats are being shown.
- Component-Specific Entries: Information is shown for each different type of component. For example, for a Processor, the type of Processor is displayed. For a Connection, the source and destination names and IDs are shown.

- **Start:** The earliest time shown on the graph.
- **End:** The latest time shown on the graph.
- **Min/Max/Mean:** The minimum, maximum, and mean (arithmetic mean, or average) values are shown. These values are based only on the range of time selected, if any time range is selected. If this instance of NiFi is clustered, these values are shown for the cluster as a whole, as well as each individual node. In a clustered environment, each node is shown in a different color. This also serves as the graph's legend, showing the color of each node that is shown in the graph. Hovering the mouse over the Cluster or one of the nodes in the legend will also make the corresponding node bold in the graph.

The right-hand side of the dialog provides a drop-down list of the different types of metrics to render in the graphs below. The top graph is larger so as to provide an easier-to-read rendering of the information. In the bottom-right

corner of this graph is a small handle () that can be dragged to resize the graph. The blank areas of the dialog can also be dragged around to move the entire dialog.

The bottom graph is much shorter and provides the ability to select a time range. Selecting a time range here will cause the top graph to show only the time range selected, but in a more detailed manner. Additionally, this will cause the Min/Max/Mean values on the left-hand side to be recalculated. Once a selection has been created by dragging a rectangle over the graph, double-clicking on the selected portion will cause the selection to fully expand in the vertical direction (i.e., it will select all values in this time range). Clicking on the bottom graph without dragging will remove the selection.