

Apache NiFi

Training

Mohammed Nadeem



About Me

6+ Years of experience working on Apache NiFi

Contributor to Apache NiFi project. I have submitted bug fixes, small patches and features to the project

Skills: Java, Spring Framework, Spring Boot, Spring Security, Distributed Systems, Apache NiFi, Kafka, Zookeeper, Big Data, Docker, Kubernetes, Jenkins, Nginx, AWS, Full Stack (Javascript, ReactJS, NodeJS, SQL....), Microservices



Staff Software Engineer
General Electric (GE)

Course RoadMap

Course Roadmap



- DataFlow Changes in an Enterprise
- History of Apache NiFi
- Introduction to Apache NiFi
- Core Features
- Exploring NiFi concepts
- Process Group and Funnel

Course Roadmap



- **NiFi Expression Language**
 - ◆ Understanding EL
 - ◆ Building DataFlow with FlowFile
Attributes as core operation
- **NiFi Structured Processing**
 - ◆ Splitting, Transforming,
Aggregating
 - ◆ Exposing a Web service

Course Roadmap



- Installation and configuration of NiFi Cluster
 - ◆ Architecture
 - ◆ Understanding NiFi configuration properties
 - ◆ Standalone Installation (Secure/Insecure)
 - ◆ 2 node secure cluster integrated with LDAP authentication

Course Roadmap



- **NiFi Registry and Version Control System**
 - ◆ Understanding NiFi Registry
 - ◆ Integrating NiFi with NiFi Registry
- **Integration with Apache Kafka**
 - ◆ NiFi as data producer/consumer

Course Roadmap



- Integration with AWS Services
 - ◆ SNS, SQS, S3, Lambda
- Integration with NoSQL Database
 - ◆ MongoDB & AWS DynamoDB
- Integration with ElasticSearch

Course Roadmap



- **NiFi Monitoring and Logging**
 - ◆ **Inline Monitoring (System Diagnostics)**
 - ◆ **External Monitoring (Grafana/Prometheus)**
 - ◆ **NiFi Logging (Customising nifi logback configuration files)**

Course Roadmap



- Advanced NiFi
 - ◆ Record Based Processing
 - ◆ Building Custom Processor
 - ◆ ExecuteScript Processor as equivalent

How I take this Course

Presentations (Diagrams) - 600+ Slides

Annotations/Animations to simplify concepts

Hands-on building dataflows from simple to complex

Hands-on with Cluster Integration and Operations

Please feel free to ask questions. I will be thankful for your questions

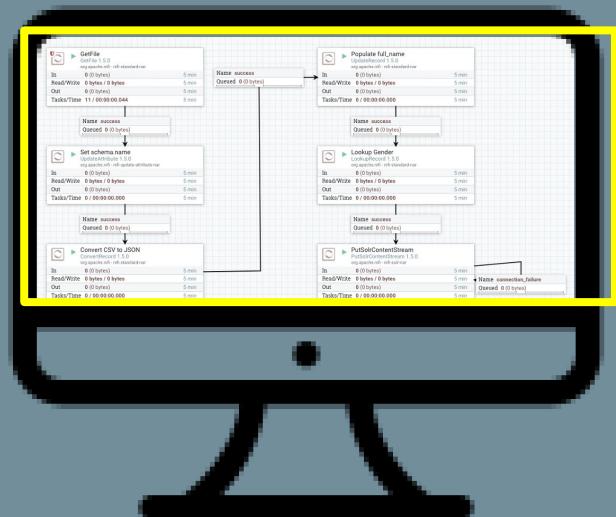
Please use Microsoft Teams chat window for raising questions, your PC screenshot if facing any issues

I will be using Mac as working machine throughout the course

Introduction to Apache NiFi

What is Apache NiFi

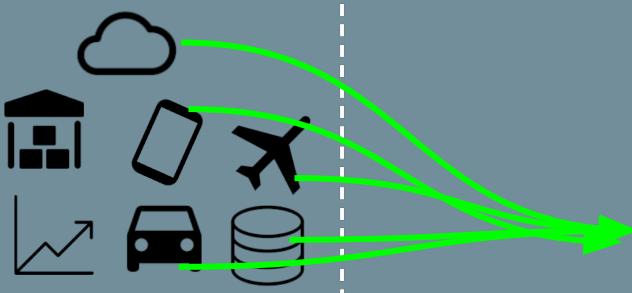
NiFi (short for “Niagara Files” is a powerful enterprise grade dataflow tool that can collect, route, enrich, transform and process data in scalable manner



NiFi is based on the concepts of flow-based programming (FBP). FBP is a programming paradigm that defines applications as networks of “blackbox” processes which exchange data across predefined connections

What is Apache NiFi

Data Producers



Apache NiFi



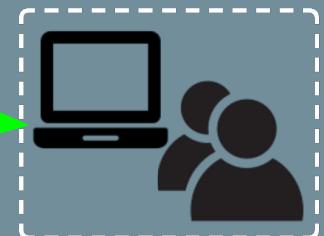
collects data

transforms data

routes data

processes data

- Consumers
 - Users
 - Systems
 - Storage
 -Many More



Short History

Developed by National Security Agency (NSA) for over 8 years



Open sourced in Nov 2014 (Apache). Major contributions were ex-NSA who formed the company named Onyara

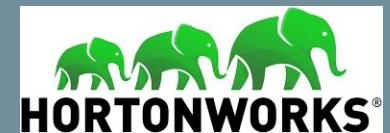


Become Apache Top Level Project in July 2015



In August 2015, Hortonworks acquires Onyara

In September 2015, Hortonworks released HDF 1.0 powered by NiFi.

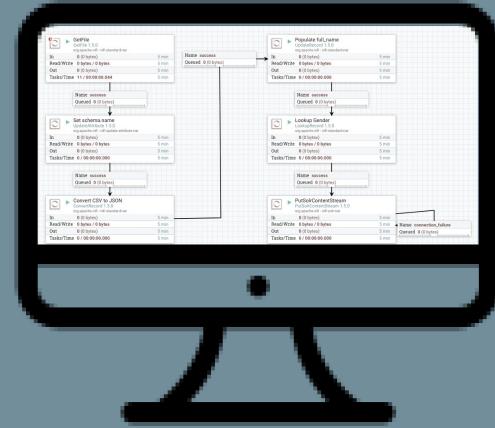


In October 2018, Cloudera acquired Hortonworks. Cloudera actively maintains the Apache NiFi project



Apache NiFi Features

- Web-based user interface
 - Seamless experience between design, control, feedback
- Highly configurable
 - Loss tolerant vs guaranteed delivery
 - Low latency vs high throughput
 - Dynamic prioritization
 - Flow can be modified at runtime
 - Back pressure

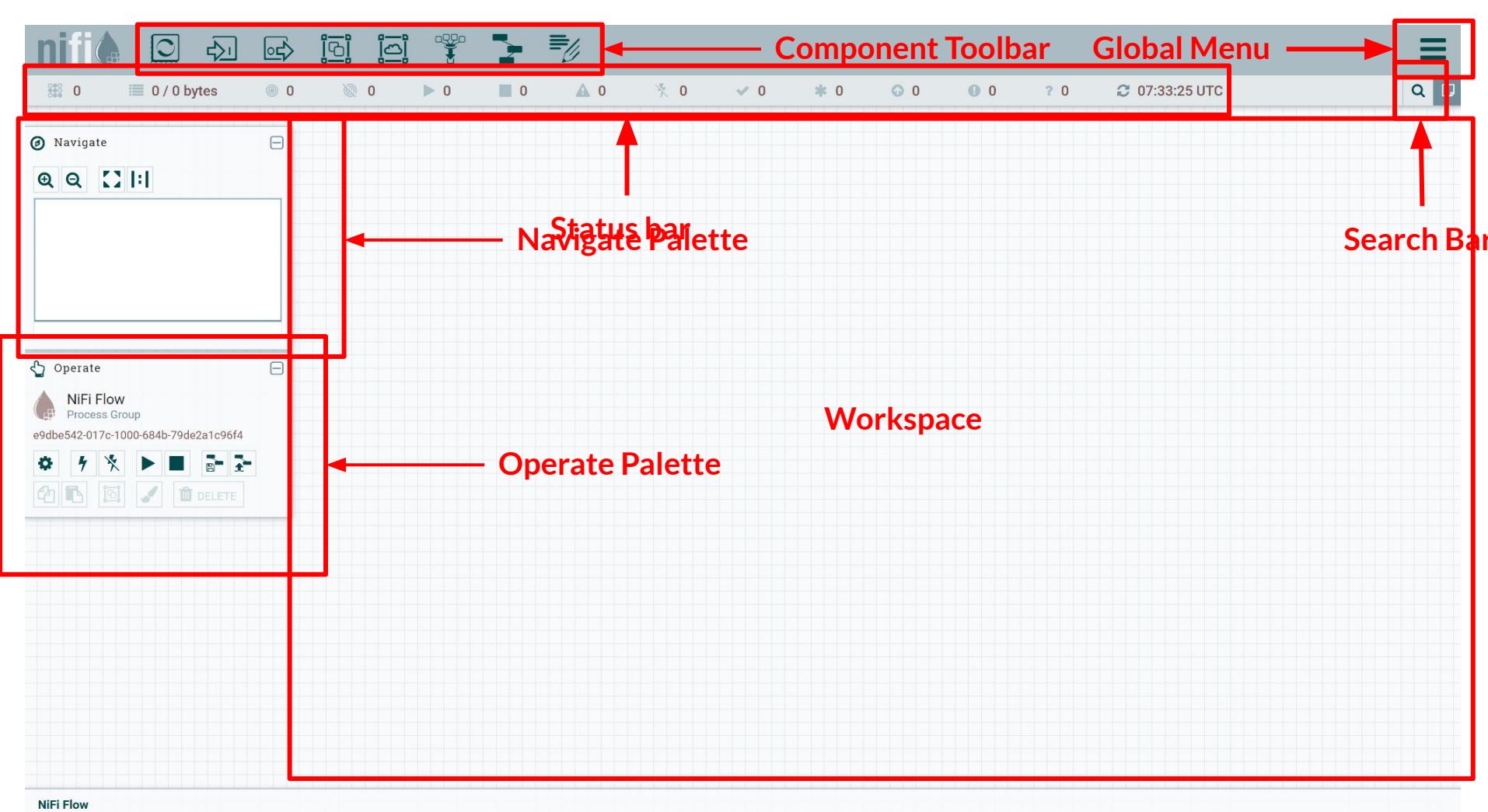


Apache NiFi Features

- Data Provenance
 - Track dataflow from beginning to end
- Secure
 - SSL, SSH, HTTPS, encrypted content etc
 - Multi-tenant authorization and internal authorization / policy management



NiFi User Interface





Component Toolbar



0 / 0 bytes

0

0

0

0

0

0

0

0

0

0

0

0

0

0

0

07:33:25 UTC



Navigate



Operate



e9dbe542-017c-1000-684b-79de2a1c96f4



Processor



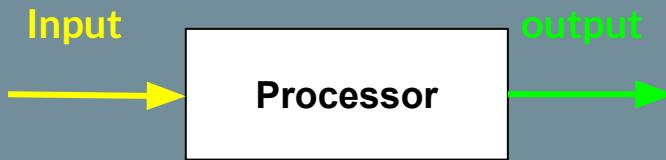
Processor

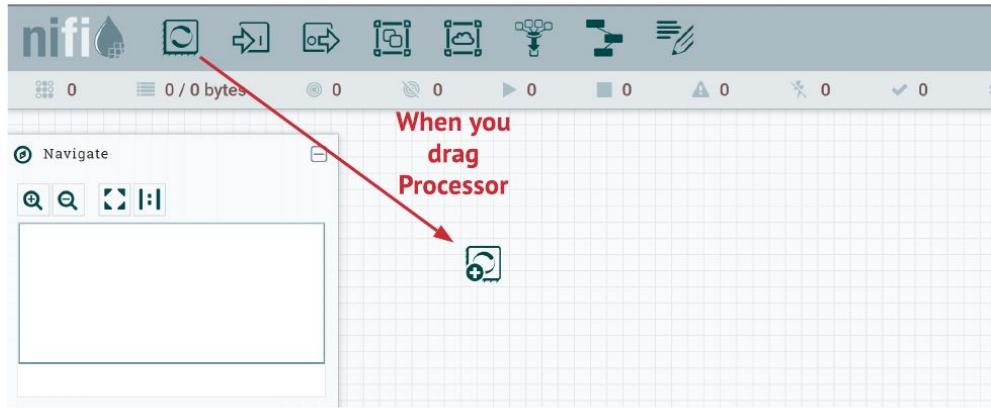
Processor



Black Box,

Performs the work,
doing some combination
of data routing,
transformation or
mediation between
systems





When you
drag
Processor

Add Processor

Source	Type	Version	Tags
all groups	Type ▾	Displaying 288 of 288	Filter
amazon	AttributeRollingWindow	1.14.0	rolling, data science, Attribute ...
attributes	AttributesToCSV	1.14.0	flowfile, csv, attributes
avro	AttributesToJSON	1.14.0	flowfile, json, attributes
aws	Base64EncodeContent	1.14.0	encode, base64
azure	CalculateRecordStats	1.14.0	stats, record, metrics
consume	CaptureChangeMySQL	1.14.0	cdc, jdbc, mysql, sql
csv	CompareFuzzyHash	1.14.0	fuzzy-hashing, hashing, cyber... lzma, snappy-hadoop, deflate, ...
database	CompressContent	1.14.0	lzma, snappy-hadoop, deflate, ...
delete	ConnectWebSocket	1.14.0	subscribe, consume, listen, We...
fetch	ConsumeAMQP	1.14.0	receive, amqp, rabbit, get, cons...
get	ConsumeAzureEventHub	1.14.0	cloud, streaming, streams, eve...
hadoop	ConsumeFWS	1.14.0	FWS, Exchange, Email, Conne...
ingest	ConsumeFile	1.14.0	receive, file, read, write, cons...
insert	ConsumeKafka	1.14.0	cloud, streaming, streams, eve...
json	ConsumeNATS	1.14.0	receive, nats, publish, subscribe
listen	ConsumeRedis	1.14.0	cloud, streaming, streams, eve...
logs	ConsumeRabbitMQ	1.14.0	receive, amqp, rabbit, get, cons...
message	ConsumeSFTP	1.14.0	cloud, streaming, streams, eve...
pubsub	ConsumeSocket	1.14.0	receive, socket, publish, subscribe
put	ConsumeStream	1.14.0	cloud, streaming, streams, eve...
record	ConsumeTwitter	1.14.0	receive, twitter, stream, cons...
restricted	ConsumeWebService	1.14.0	cloud, streaming, streams, eve...
source	ConsumeWebsocket	1.14.0	receive, websocket, subscribe, ...
text	ConsumeZookeeper	1.14.0	cloud, streaming, streams, eve...
update	ConsumeZookeeper	1.14.0	cloud, streaming, streams, eve...

AttributeRollingWindow 1.14.0 org.apache.nifi - nifi-stateful-analysis-nar
Track a Rolling Window based on evaluating an Expression Language expression on each FlowFile and add that value to the processor's state. Each FlowFile will be emitted with the count of FlowFiles and total aggregate value of values processed in the current time window.

Add Processor

Source all groups **amazon attributes**

Displaying 288 of 288

Type Version Tags

Type	Version	Tags
AttributeRollingWindow	1.14.0	rolling, data science, Attribute...
AttributesToCSV	1.14.0	flowfile, csv, attributes
AttributesToJson	1.14.0	flowfile, json, attributes
Base64EncodeContent	1.14.0	encode, base64
CalculateRecordStats	1.14.0	stats, record, metrics
CaptureChangeMySQL	1.14.0	cdc, jdbc, mysql, sql
CompareFuzzyHash	1.14.0	fuzzy-hashing, hashing, cyber...
CompressContent	1.14.0	Izma, snappy-hadoop, deflate, ...
ConnectWebSocket	1.14.0	subscribe, consume, listen, We...
ConsumeAMQP	1.14.0	receive, amqp, rabbit, get, cons...
ConsumeAzureEventHub	1.14.0	cloud, streaming, streams, eve...
ConsumeFWS	1.14.0	FWS Exchange Email Consum...

AttributeRollingWindow 1.14.0 org.apache.nifi - nifi-stateful-analysis-nar

Track a Rolling Window based on evaluating an Expression Language expression on each FlowFile and add that value to the processor's state. Each FlowFile will be emitted with the count of FlowFiles and total aggregate value of values processed in the current time window.

Number of Processors

Search for Processor Filter

Highlighted color denotes Selected processor

Popular Processor Tags

Add Processor

Source Displaying 5 of 288 Type ▾ Version Tags

Source	Type	Version	Tags
all groups	DeleteS3Object	1.14.0	S3, Delete, Archive, Amazon, A...
amazon attributes	FetchS3Object	1.14.0	S3, Fetch, Get, Amazon, AWS
avro aws azure	ListS3	1.14.0	S3, Amazon, list, AWS
consume csv	PutS3Object	1.14.0	S3, Archive, Amazon, AWS, Put
database delete	TagS3Object	1.14.0	S3, Archive, Tag, Amazon, AWS

Processor name

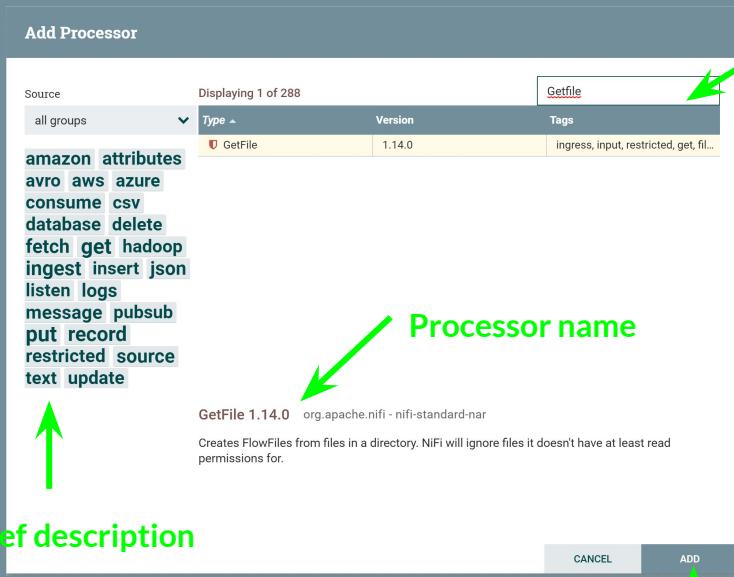
Processors with 'S3' word search

Cancel Add

Processor name: DeleteS3Object 1.14.0 org.apache.nifi - nifi-aws-nar

Brief description: Deletes FlowFiles on an Amazon S3 Bucket. If attempting to delete a file that does not exist, FlowFile is routed to success.

Adding Processor to canvas



GetFile Processor exact search

Processor name

Brief description

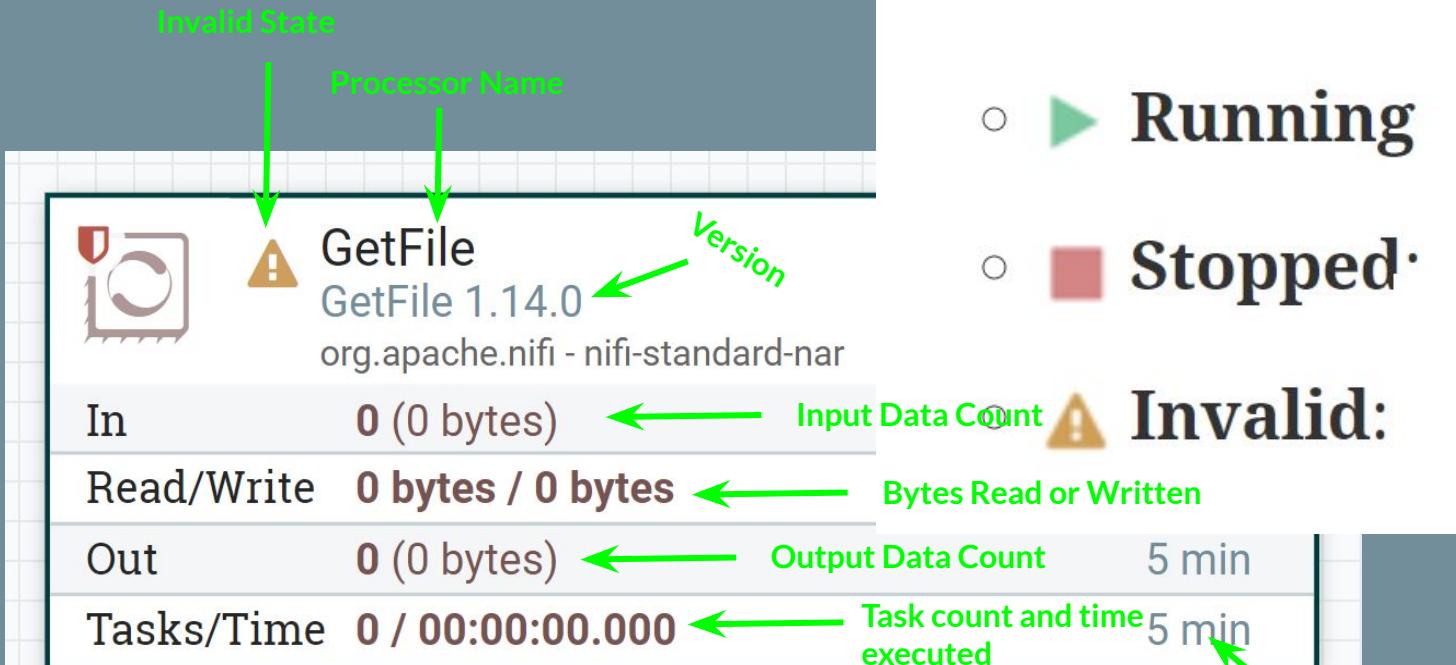
Adding Processor to canvas

- **Status Indicator:**

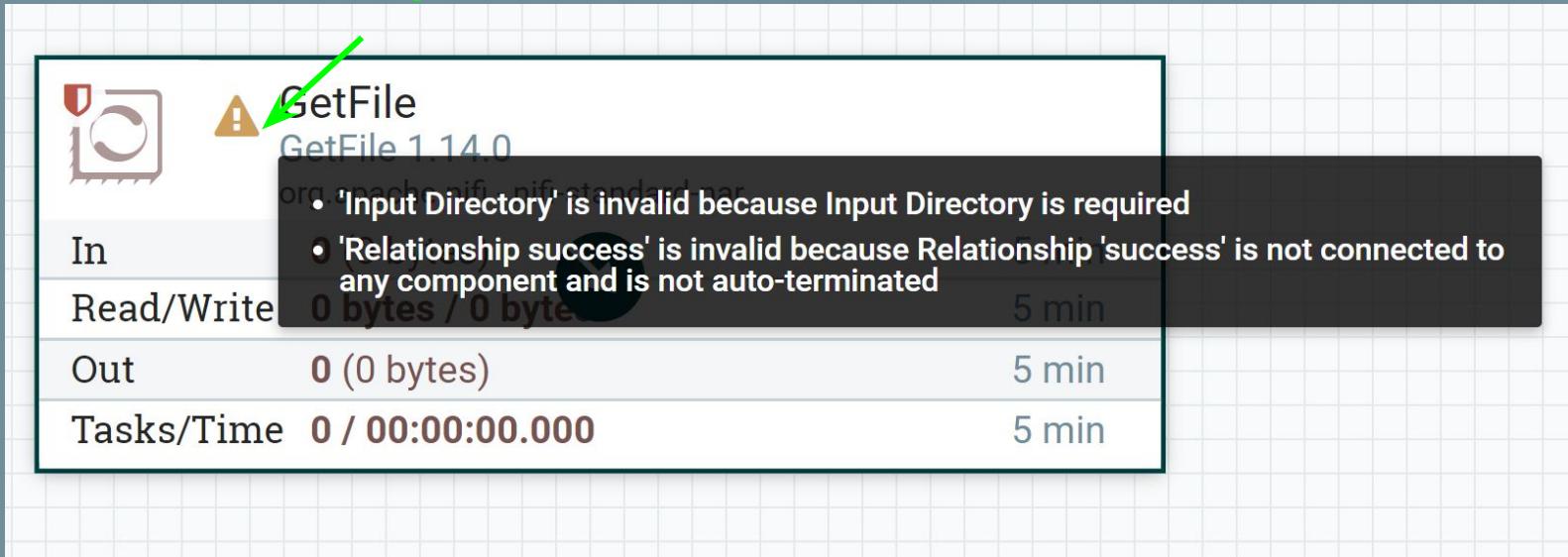
-  **Running**

-  **Stopped**

-  **Invalid:**



Hovering over the status indicator



Right clicking the Processor

 **GetFile**
GetFile 1.14.0
org.apache.nifi - nifi-standard-nar

In 0 (0 bytes)
Read/Write 0 bytes / 0 bytes
Out 0 (0 bytes)
Tasks/Time 0 / 00:00:00.000



-  Configure
-  Disable
-  View data provenance
-  View status history
-  View usage
-  View connections
-  Center in view
-  Change color
-  Group
-  Create template
-  Copy
-  Delete

Configure Processor

Number of threads
for simultaneously
Running the
processor

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

Scheduling Strategy ? Timer driven

Concurrent Tasks ? 1

Run Schedule ? 0 sec

Execution ? All nodes

All nodes ✓

All nodes ?

Primary node ?

CANCEL APPLY

Time based or CRON based

How often to run the processor

Where to run this Processor

Configure Processor

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Bold text properties are required

Property	Description	Value
Input Directory	When Recurse Subdirectories is true, then only subdirectories whose path matches the given regular expression will be scanned	No value set
File Filter	Expression language scope: Not Supported	No value set
Path Filter	Sensitive property: false	[^\\].*
Batch Size		10
Keep Source File		false
Recurse Subdirectories		true
Polling Interval		0 sec
Ignore Hidden Files		true
Minimum File Age		0 sec
Maximum File Age		No value set
Minimum File Size		0 B
Maximum File Size		No value set

Description of property

Light text properties Are not required

Setting value to a property

CANCEL APPLY

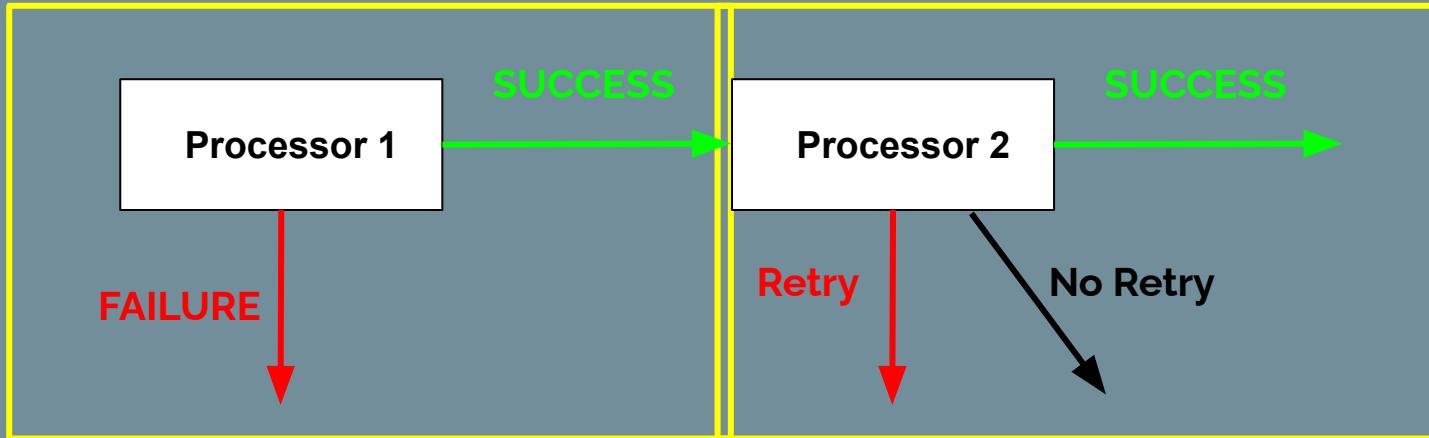
Configure Processor

⚠ Invalid

SETTINGS	SCHEDULING	PROPERTIES	COMMENTS

CANCEL APPLY

Understanding Relationships



2 Relationships

3 Relationships

Lab : Moving Files

- **Before we proceed for the Lab**

Java 8 should be installed

<https://www.oracle.com/in/java/technologies/javase/javase8-archive-downloads.html>

Apache NiFi 1.15.3

Moving Files

get-file



put-file



/home/nadeem/Documents

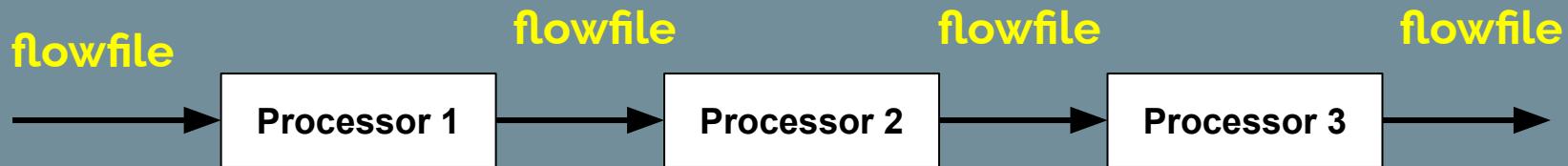
FlowFile

FlowFile

Packet of Information,

Unit of Work that moves between Processors

Made up of **Attributes + Content**



FlowFile

Packet of Information,

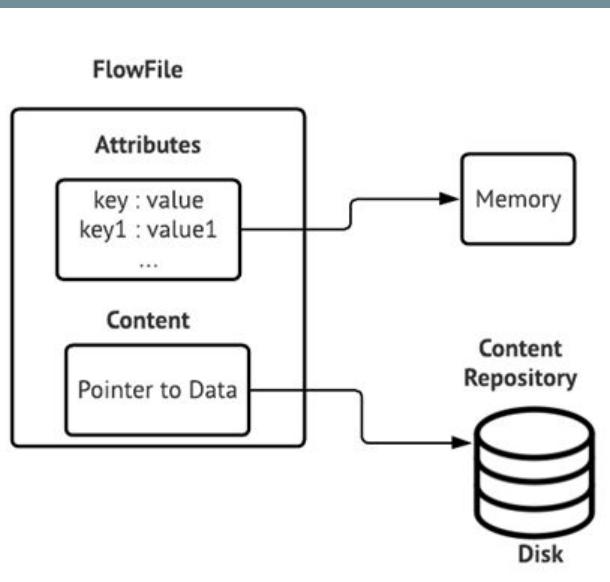
Unit of Work that moves between Processors

Made up of **Attributes + Content**

Input FlowFile

Processor

Output Flowfile



Attributes

Are set of key-value pairs

Available in memory as strings (Java HashMap)

Useful for quick decision making and routing

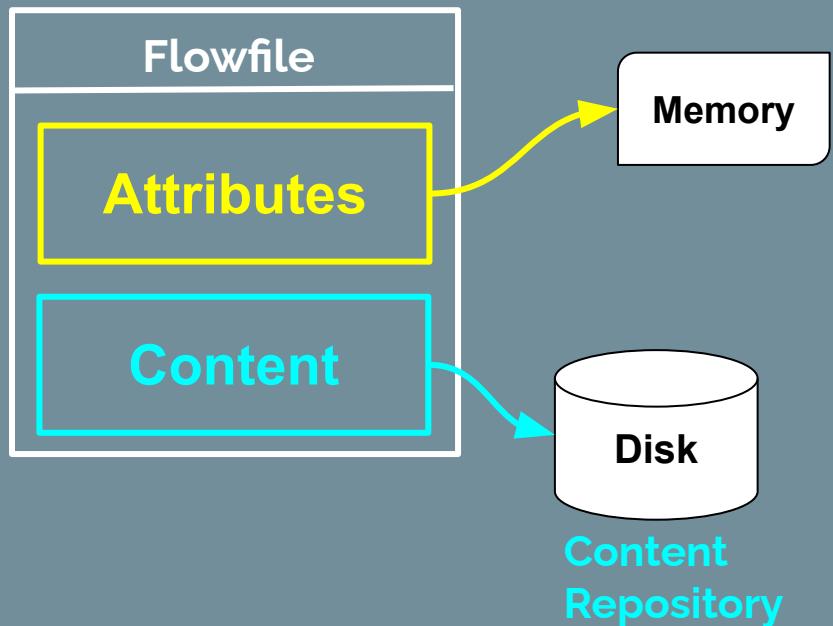
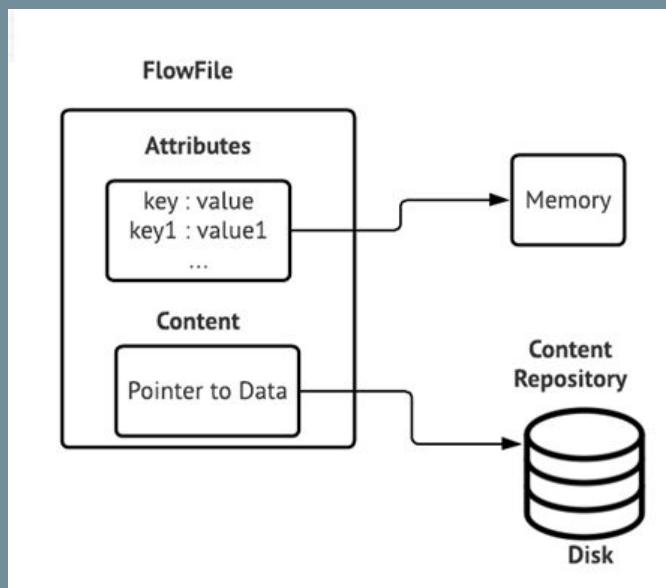
Made up of **Attributes + Content**

Content

FlowFile is a pointer to the content in content repository

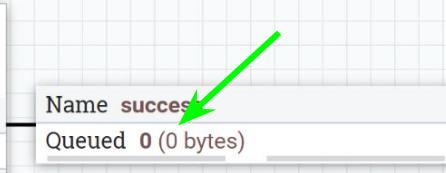
Content is only accessed when processor needs it

Could pass through multiple processors without accessing the content



Right clicking the Connection Queue

	GetFile	GetFile 1.14.0	org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)	5 min	
Read/Write	0 bytes / 0 bytes	5 min	
Out	0 (0 bytes)	5 min	
Tasks/Time	0 / 00:00:00.000	5 min	



	PutFile	PutFile 1.14.0	org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)	5 min	
Read/Write	0 bytes / 0 bytes	5 min	
Out	0 (0 bytes)	5 min	
Tasks/Time	0 / 00:00:00.000	5 min	

 **Configure**

 **View status history**

 **List queue**

 **Go to source**

 **Go to destination**

 **Bring to front**

 **Create template**

 **Empty queue**

 **Delete**

FlowFile

DETAILSATTRIBUTES

FlowFile Details

UUID
a991de0b-5e42-40ba-87e0-8f7f579390b4

Filename
data.txt

File Size
36 bytes

Queue Position
No value set

Queued Duration
00:00:10.139

Lineage Duration
00:00:10.169

Penalized
No

Content Claim

Container
default

Section
1

Identifier
1634312852493-1

Offset
0

Size
36 bytes

 DOWNLOAD VIEW

View the content
of FlowFile

OK

FlowFile

DETAILS ATTRIBUTES

Attribute Values **Attribute Name**

absolute.path ←
C:\Users\212601209\Desktop\GetFile/ ← **Attribute Value**

file.creationTime
2021-10-15T20:20:45+0530

file.lastAccessTime
2021-10-15T21:17:32+0530

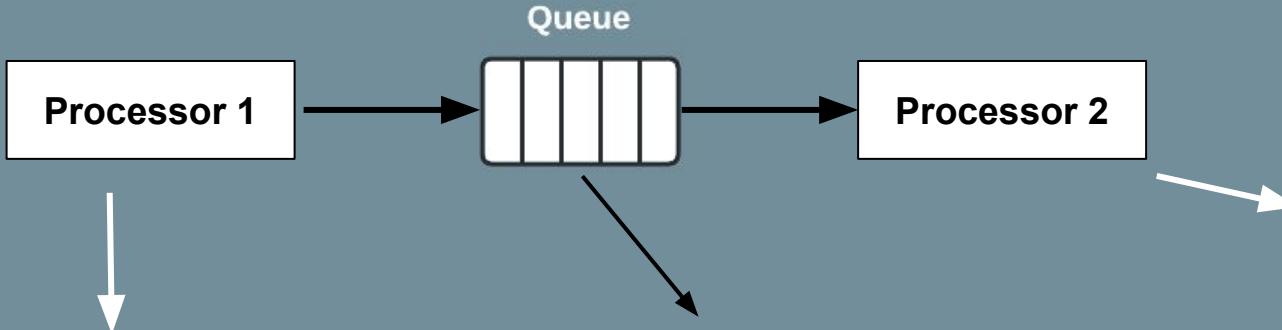
file.lastModifiedTime
2021-10-15T20:21:25+0530

file.owner
LOGON\212601209

filename
data.txt

path

OK



Schedule

Penalty Duration
(delay flowfile)

Yield Duration
(retry delay on error)

Level of Logging

Size

No of Flowfiles

When to expire
Flowfiles

How to prioritize
Flowfiles



Right clicking the Connection Queue

	GetFile	GetFile 1.14.0	org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)	5 min	
Read/Write	0 bytes / 0 bytes	5 min	
Out	0 (0 bytes)	5 min	
Tasks/Time	0 / 00:00:00.000	5 min	



	PutFile	PutFile 1.14.0	org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)	5 min	
Read/Write	0 bytes / 0 bytes	5 min	
Out	0 (0 bytes)	5 min	
Tasks/Time	0 / 00:00:00.000	5 min	

 **Configure**

 **View status history**

 **List queue**

 **Go to source**

 **Go to destination**

 **Bring to front**

 **Create template**

 **Empty queue**

 **Delete**

Create Connection

DETAILS SETTINGS

From Processor
GetFile
GetFile

Within Group
NiFi Flow

For Relationships
 success

To Processor
PutFile
PutFile

Within Group
NiFi Flow

Success Connection

CANCEL

ADD

Create Connection

DETAILS SETTINGS

Name

Id
No value set

FlowFile Expiration ?

0 sec

Back Pressure

Object Threshold ?

10000

Size Threshold ?

1 GB

Load Balance Strategy ?

Do not load balance

Available Prioritizers ?

FirstInFirstOutPrioritizer

NewestFlowFileFirstPrioritizer

OldestFlowFileFirstPrioritizer

PriorityAttributePrioritizer

Selected Prioritizers ?

Priority Pass for Data

CANCEL

ADD

Right clicking the Processor

 **GetFile**
GetFile 1.14.0
org.apache.nifi - nifi-standard-nar

In 0 (0 bytes)
Read/Write 0 bytes / 0 bytes
Out 0 (0 bytes)
Tasks/Time 0 / 00:00:00.000



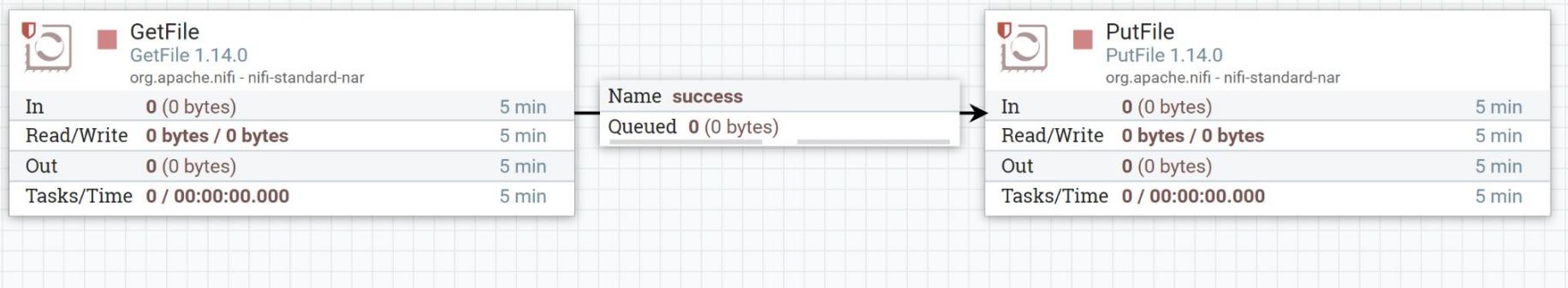
-  Configure
-  Disable
-  View data provenance
-  View status history
-  View usage
-  View connections
-  Center in view
-  Change color
-  Group
-  Create template
-  Copy
-  Delete

Configure Processor

⚠ Invalid

SETTINGS	SCHEDULING	PROPERTIES	COMMENTS
Name GetFile	<input checked="" type="checkbox"/> Enabled	Automatically Terminate Relationships <small>?</small> <input type="checkbox"/> success All files are routed to success	
Id 841ce36f-017c-1000-290f-d1beb890e478			Unique Processor ID
Type GetFile 1.14.0			Penalizes output for 30 secs before next processor can take it
Bundle org.apache.nifi - nifi-standard-nar			
Penalty Duration <small>?</small> 30 sec		Yield Duration <small>?</small> 1 sec	How long processor should pause before continuing on error
Bulletin Level <small>?</small> WARN			Default Log level of processor

CANCEL **APPLY**



Understanding Course Material

The screenshot shows a GitHub repository page for 'naddym/nifi-course-abm'. The repository is public. The main tab is selected. There is one branch ('main') and zero tags. The repository contains several files:

- installation (highlighted)
- monitoring (highlighted)
- productionizing (highlighted)
- templates (highlighted)
- .gitignore
- README.md
- best-practices-and-general-guideli... (highlighted)

Each file has a 'Add course material' button next to it.

—

Lab : Moving Files

+

Changing filename

– Moving File + Changing filename

get-file



put-file



/home/nadeem/Documents

Local
FileSystem

Processor Properties Types

Configure Processor | GetFile 1.15.3

⚠ Invalid

Required field

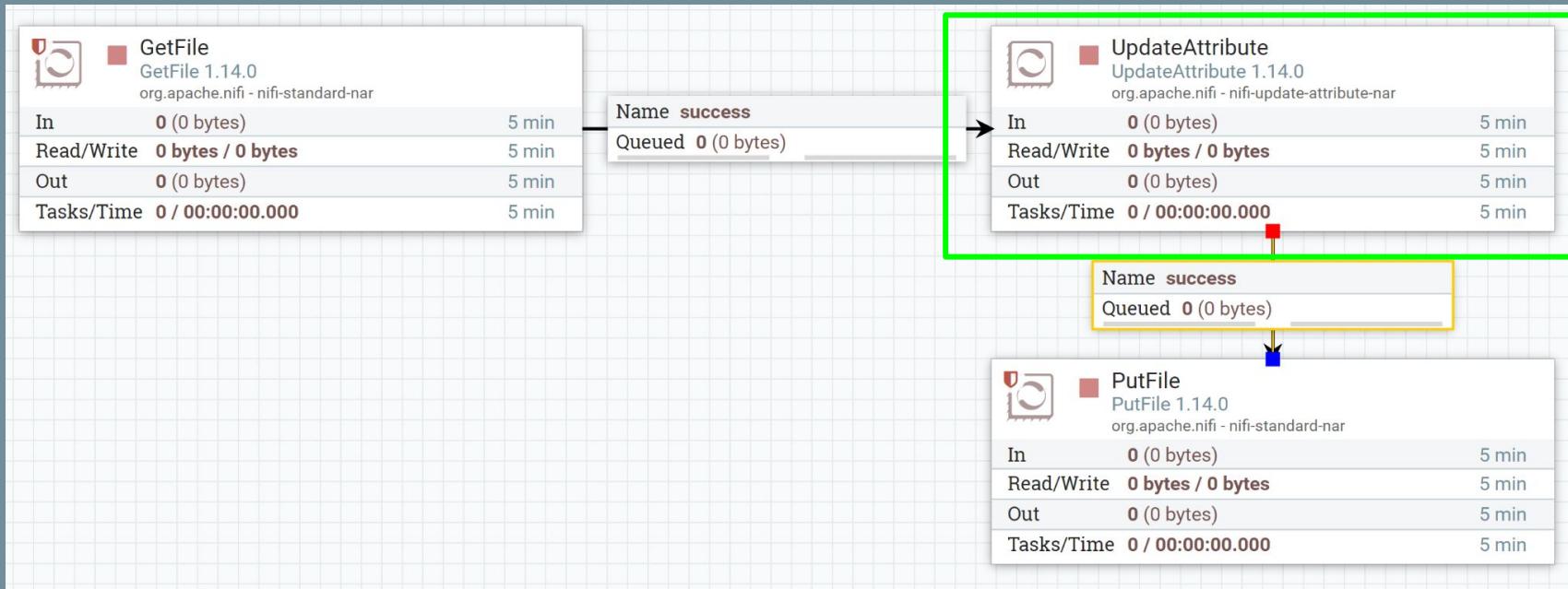
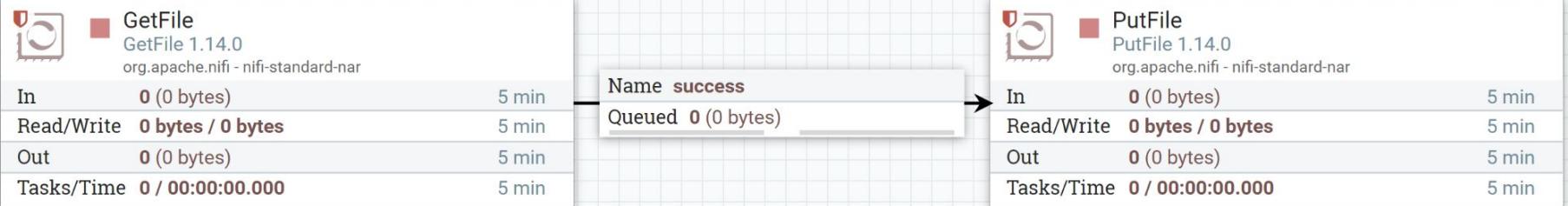
Property	Value	
Input Directory	?	No value set
File Filter	?	[^\\.]*
Path Filter	?	No value set
Batch Size	?	10
Keep Source File	?	false
Recurse Subdirectories	?	true
Polling Interval	?	0 sec
Ignore Hidden Files	?	true
Minimum File Age	?	0 sec
Maximum File Age	?	No value set
Minimum File Size	?	0 B
Maximum File Size	?	No value set

Verify Properties

Dynamic Properties
Or
User Defined Properties

Static Properties

CANCEL APPLY



FlowFile

DETAILS ATTRIBUTES

Attribute Values

absolute.path
C:\Users\212601209\Desktop\GetFile/

file.creationTime
2021-10-15T20:20:45+0530

file.lastAccessTime
2021-10-15T21:17:32+0530

file.lastModifiedTime
2021-10-15T20:21:25+0530

file.owner
LOGON\212601209

filename ← **Attribute Name**

data.txt ← **Attribute Value**

path

OK

Configure Processor

Stopped

SETTINGS SCHEDULING

PROPERTIES

COMMENTS

Required field



Dynamic Property
For adding new property

Property	Value
Delete Attributes Expression	?
No value set	
Store State	?
Do not store state	
Stateful Variables Initial Value	?
No value set	
Cache Value Lookup Cache Size	?
100	

ADVANCED

CANCEL

APPLY

Configure Processor

Stopped

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Delete Attributes Expression	No value set
Store State	Do not store state
Stateful Variables Initial Value	No value set
Cache Value Lookup Cache Size	...
filename	<input type="text" value="1 processed-data.txt"/> EL ✓ PARAM ✓ <input type="button" value="Delete"/>

Property Name

Property Value

1 processed-data.txt

Set empty string

CANCEL OK

ADVANCED CANCEL APPLY

FlowFile

DETAILS

ATTRIBUTES

Attribute Values

absolute.path
C:\Users\212601209\Desktop\GetFile/

file.creationTime
2021-10-15T20:20:45+0530

file.lastAccessTime
2021-10-15T21:36:33+0530

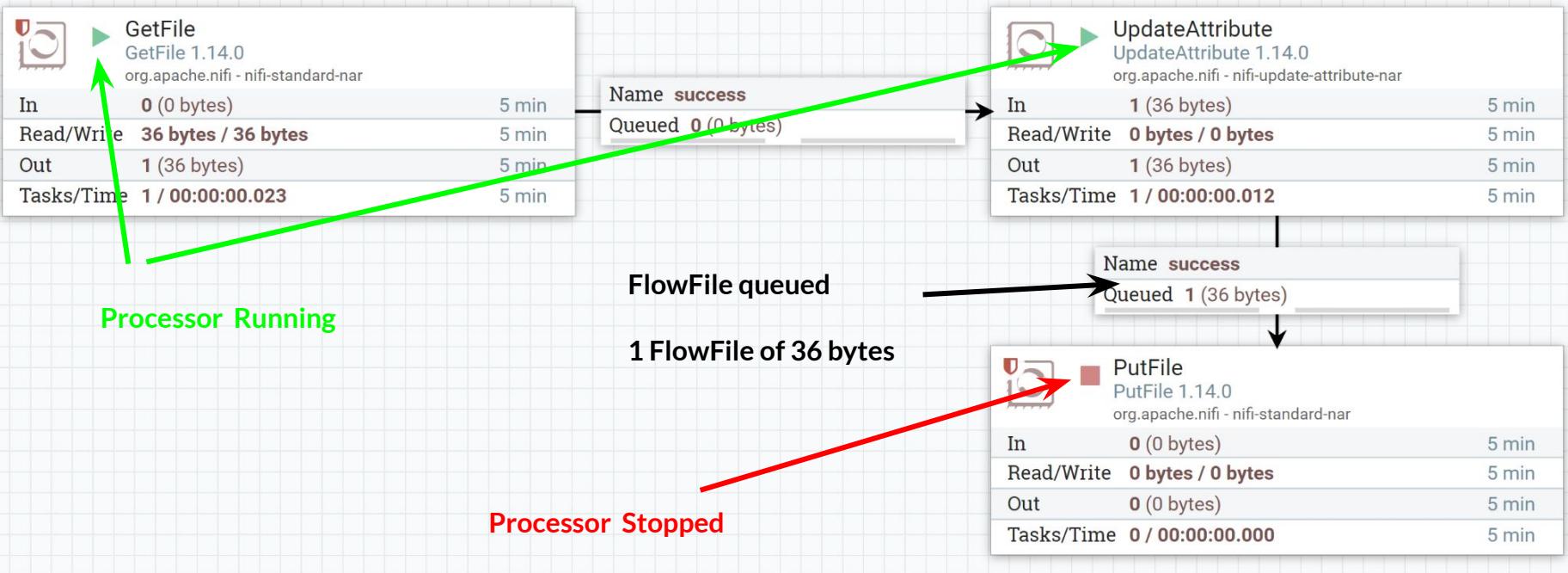
file.lastModifiedTime
2021-10-15T20:21:25+0530

file.owner
LOGON\212601209

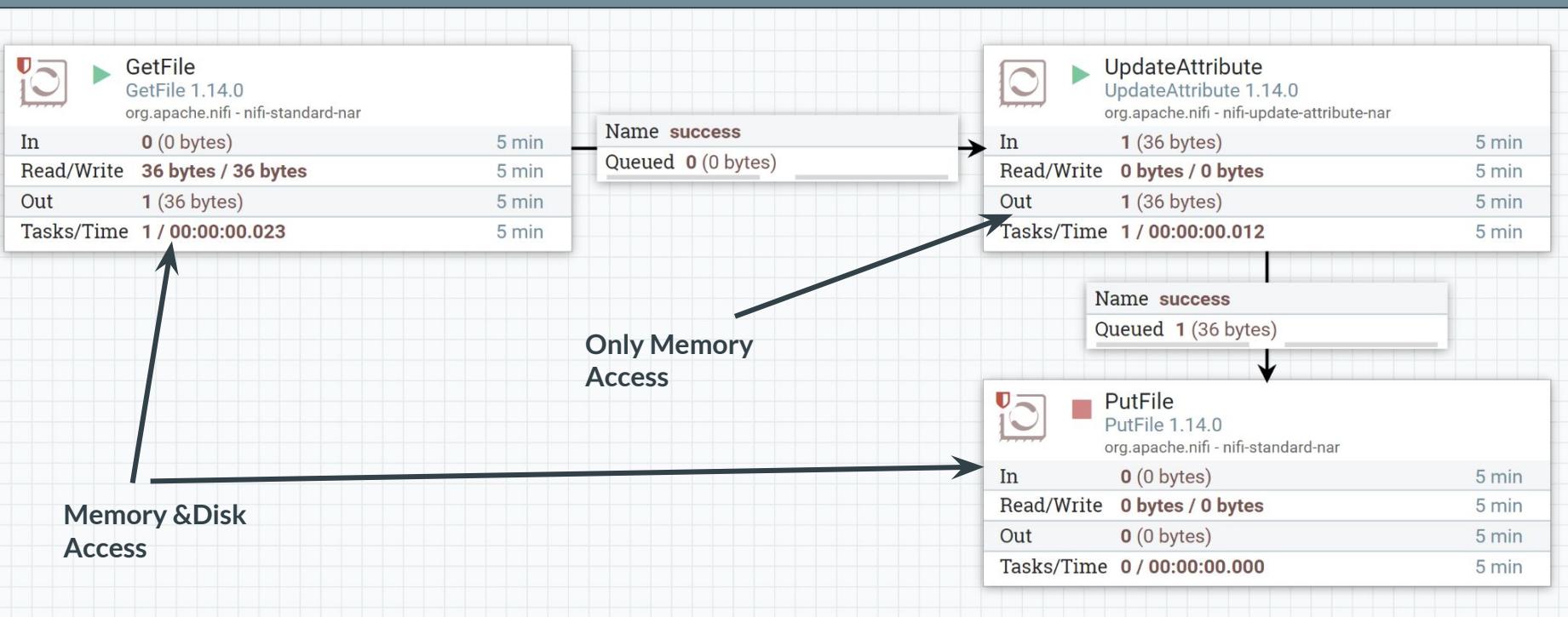
filename
processed-data.txt 

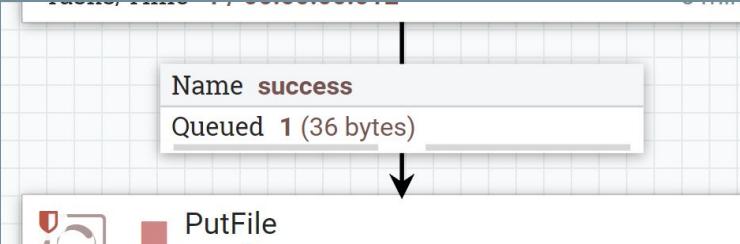
path

OK



From Resources Perspective





Create Connection

DETAILS SETTINGS

Name

Available Prioritizers [?](#)

- FirstInFirstOutPrioritizer
- NewestFlowFileFirstPrioritizer
- OldestFlowFileFirstPrioritizer
- PriorityAttributePrioritizer

Id No value set

FlowFile Expiration [?](#)
0 sec

When to expire the flowfile

Back Pressure Object Threshold [?](#) 10000

Size Threshold [?](#) 1 GB

Selected Prioritizers [?](#)

Load Balance Strategy [?](#)
Do not load balance

Maximum size of the queue

Number of flowfiles to hold in the queue maximum

CANCEL ADD

This screenshot shows the 'Create Connection' dialog box. It has tabs for 'DETAILS' and 'SETTINGS', with 'SETTINGS' currently selected. The 'DETAILS' tab includes fields for 'Name' (which is empty), 'Id' (set to 'No value set'), and 'FlowFile Expiration' (set to '0 sec'). The 'SETTINGS' tab includes fields for 'Back Pressure Object Threshold' (set to '10000') and 'Size Threshold' (set to '1 GB'). A dropdown menu for 'Selected Prioritizers' is open, listing four prioritizer types: 'FirstInFirstOutPrioritizer', 'NewestFlowFileFirstPrioritizer', 'OldestFlowFileFirstPrioritizer', and 'PriorityAttributePrioritizer'. Below these settings are 'Load Balance Strategy' options, with 'Do not load balance' selected. Three green arrows point from annotations to specific settings: one from the annotation 'When to expire the flowfile' to the 'FlowFile Expiration' field; one from the annotation 'Maximum size of the queue' to the 'Size Threshold' field; and one from the annotation 'Number of flowfiles to hold in the queue maximum' to the 'Back Pressure Object Threshold' field. At the bottom right are 'CANCEL' and 'ADD' buttons.



■ GetFile
GetFile 1.14.0
org.apache.nifi -

In 0 (0 bytes)
Read/Write 0 bytes / 0
Out 0 (0 bytes)
Tasks/Time 0 / 00:00:00

- ⚙ Configure
- ▶ Start ← Starting of processor
- ▶ Run Once ← Run the processor only once
- ✗ Disable ← Disable, when no longer required for use
- ⠇ View data provenance ← Data Provenance for this Processor
- 📈 View status history ← Flowfile statistics for this processor
- 📄 View usage ← Processor documentation
- ⌚ View connections
- 📍 Center in view
- 🖌 Change color
- ⬇ Create template ← Template -> for exporting to another NiFi instance
- _COPY Copy
- DELETE Delete

Process Group



Process Group

The Apache NiFi interface shows the 'Process Group' icon in the status bar. A green arrow points from the toolbar area down to the status bar. Below the status bar, the 'Add Process Group' dialog is open, prompting for a 'Process Group Name'. The dialog also includes 'CANCEL' and 'ADD' buttons at the bottom.

Process Group

On dragging

Add Process Group

Process Group Name

Name

CANCEL ADD



Groups all processors

Folder / Workspace for placing all processors in it.

A detailed monitoring box for the 'Process Group Name' showing various metrics: Queued (0), In (0), Read/Write (0 bytes / 0 bytes), Out (0), and other counters like CANCEL (0), ERROR (0), and UNKNOWN (0).

Process Group Name

Queued 0 (0 bytes)

In 0 (0 bytes) → 0 5 min

Read/Write 0 bytes / 0 bytes 5 min

Out 0 → 0 (0 bytes) 5 min

CANCEL 0 ERROR 0 UNKNOWN 0

Move File

0 0 0 0 3 0 0

Queued 2 (72 bytes)

In 0 (0 bytes) → 0 5 min

Read/Write 0 bytes / 0 bytes 5 min

Out 0 → 0 (0 bytes) 5 min

✓ 0 * 0 ⬆ 0 ! 0 ? 0

Inside

GetFile
GetFile 1.14.0
org.apache.nifi - nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Name success
Queued 0 (0 bytes)

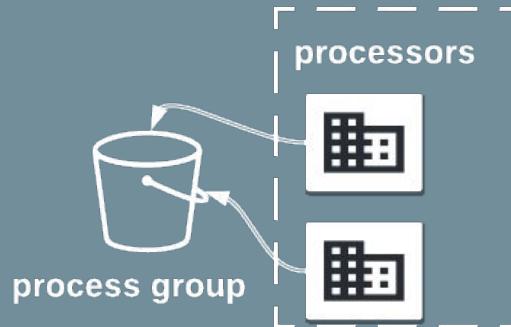
UpdateAttribute
UpdateAttribute 1.14.0
org.apache.nifi - nifi-update-attribute-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Name success
Queued 2 (72 bytes)

PutFile
PutFile 1.14.0
org.apache.nifi - nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min





Queued 2 (72 bytes) ← Flowfiles Queued

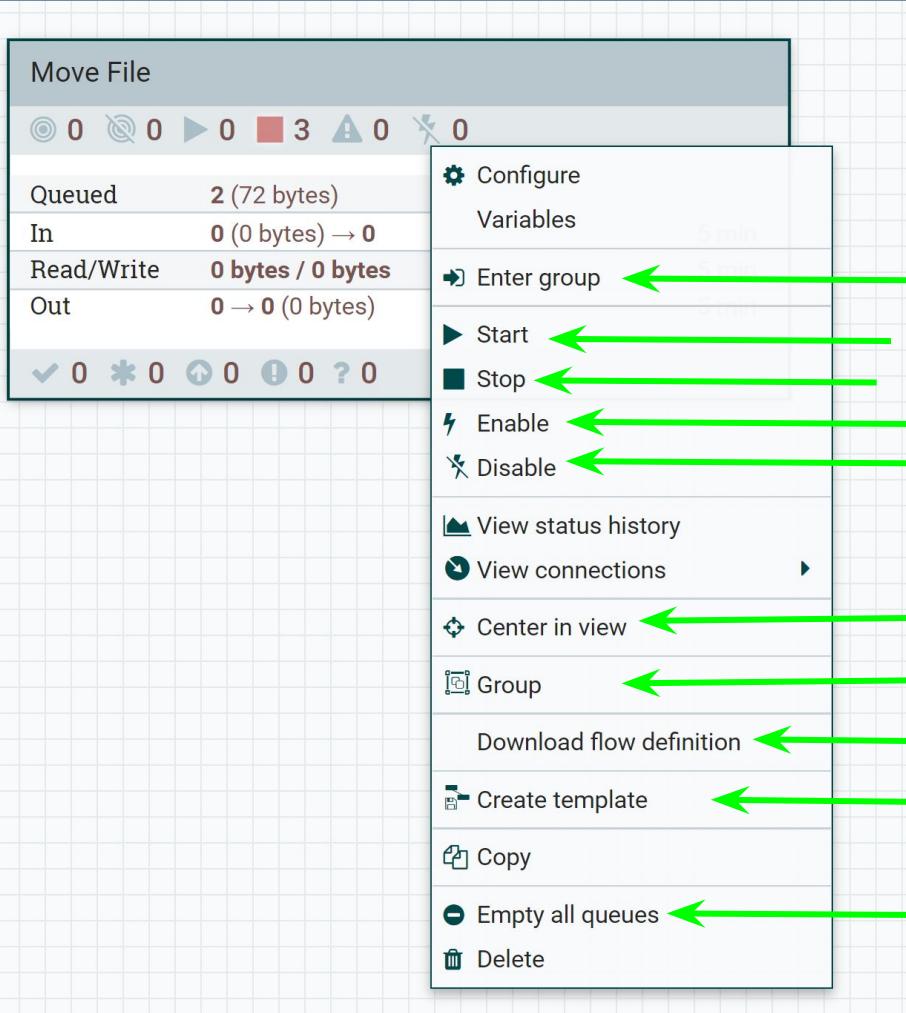
In 0 (0 bytes) → 0 ← No of Input flowfiles 5 min

Read/Write 0 bytes / 0 bytes ← Flowfiles data read 5 min

Out 1 → 0 (0 bytes) ← No of flowfiles Outputted by this process group 5 min



All the statistics in last 5 minutes



Enter Process Group or double click

Starting all processors within it

Stopping all processors within it

Enabling all processors within it

Disabling all processors within it

Center its position

Put this process group into another parent group

Download flow definition

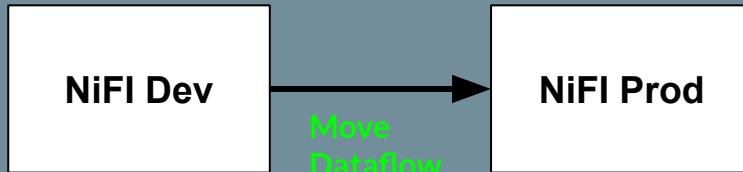
To create template to be exported to another nifi instance

Empty all connection queues (i.e flowfiles)

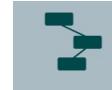
Templates



Templates



Templates



Makes it easier to **reuse** and **distribute** dataflows to another instance of NiFi

Export from one NiFi and put into another NiFi



Importing
Templates
to Canvas



Instantiate Template

No templates have been loaded into this NiFi.

OK

– Template Operations



Create Template

Upload Template

Instantiate Template

Create Template



0

0 / 0 bytes

0

Navigate



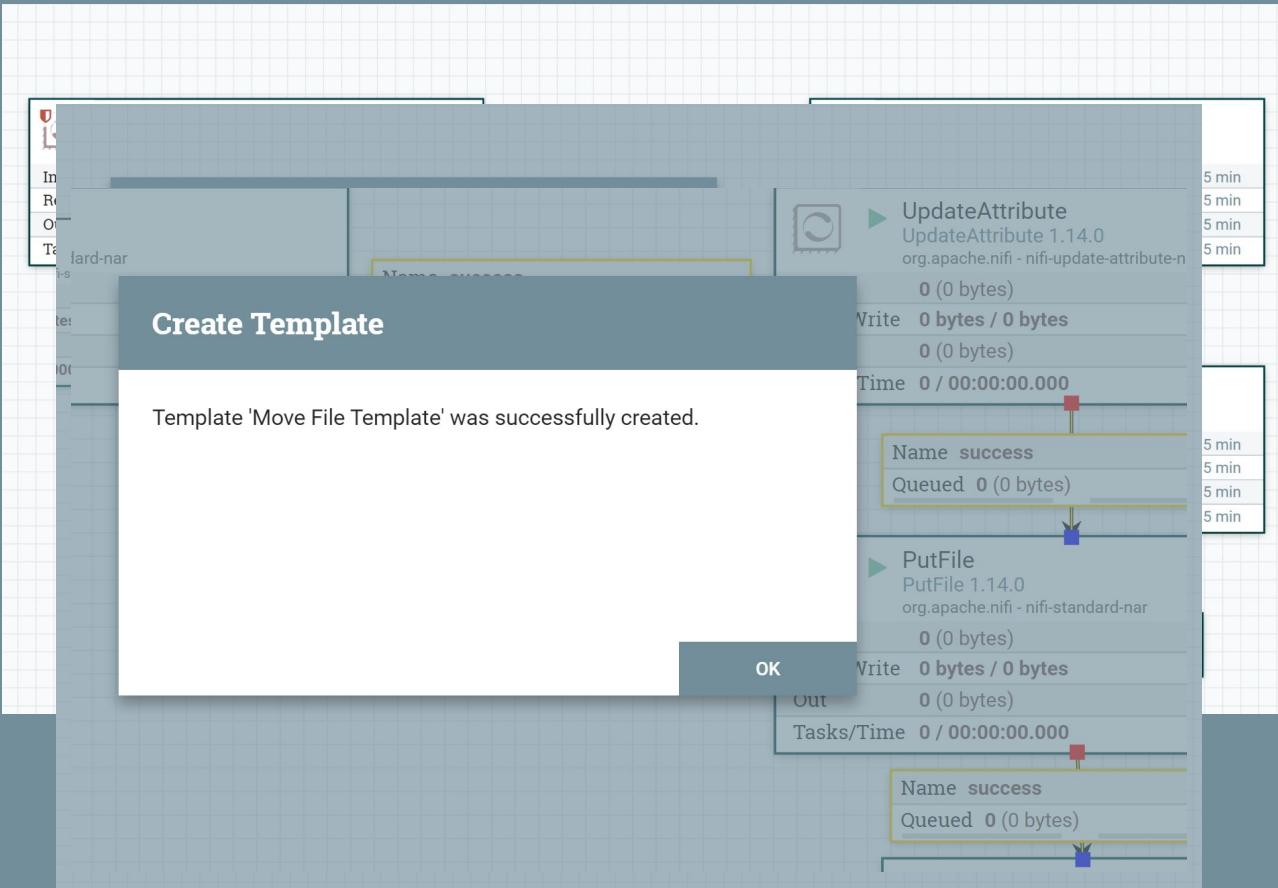
Operate

NiFi Flow
Process Group

04737759-017c-1000-0eca-cf6dae988919



Create Template



Upload Template



0

0 / 0 bytes

0

Navigate



Operate

 NiFi Flow
Process Group

04737759-017c-1000-0eca-cf6dae988919



Upload Template



4ebba3dcc



DELETE



GetFile

GetFile 1.14.0
org.apache.nifi - nifi-standard-nar

In 0 (0 bytes)

Read/Write 0 bytes / 0 bytes

Out 0 (0 bytes)

Tasks/Time 0 / 00:00:00.000

Name: success

Upload Template

Select Template



Search through
Templates for
Upload form your
local file system

CANCEL

UPLOAD

Success

Template successfully imported.

OK

Instantiate Template



0

0 / 0 bytes

0

0

2

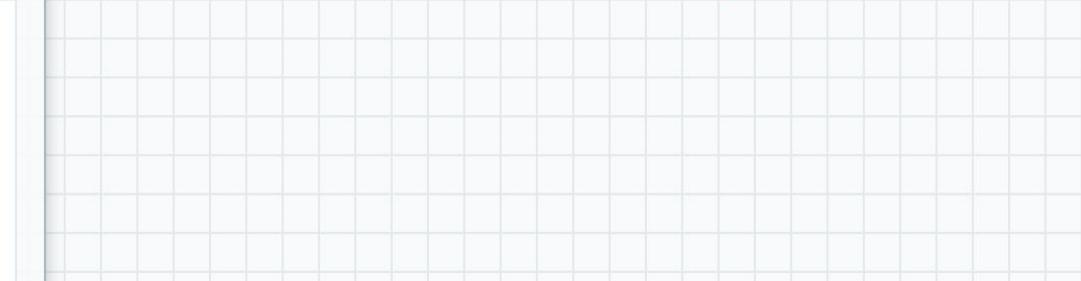
1

3

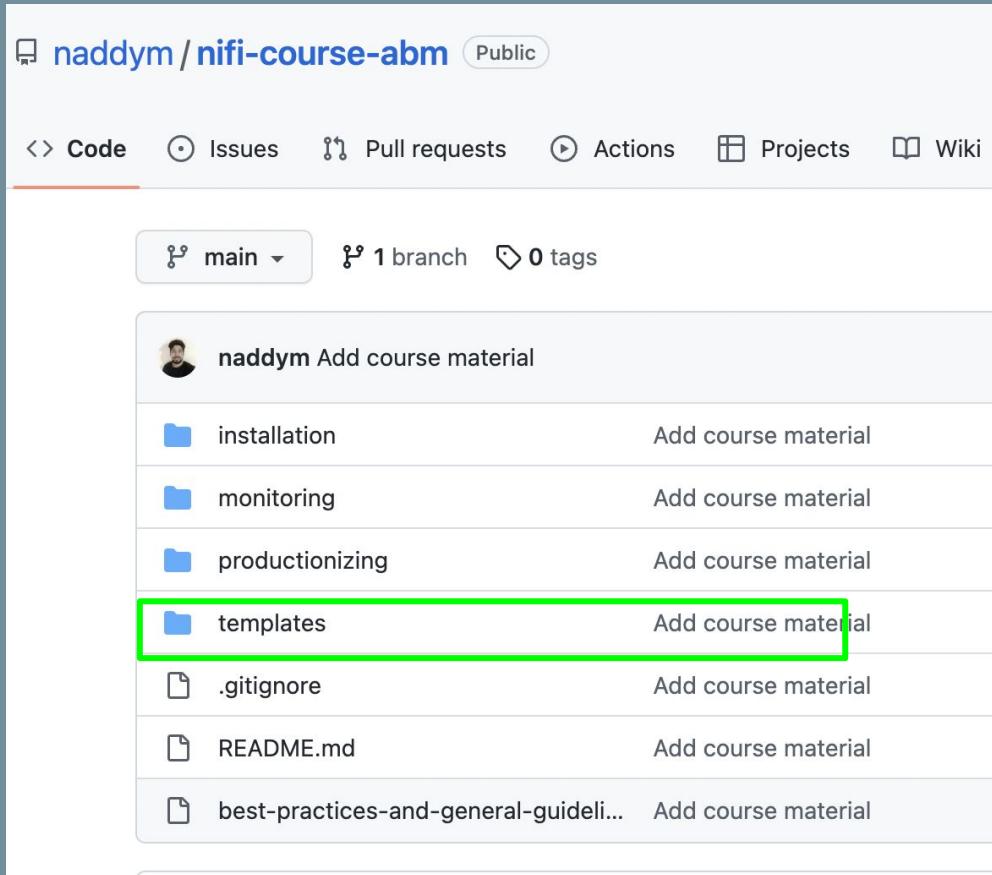
Navigate



Importing
Templates to Canvas



Understanding Course Material



The screenshot shows a GitHub repository page for 'naddym/nifi-course-abm'. The repository is public and has one branch ('main') and zero tags. The main content area displays a list of files:

- installation Add course material
- monitoring Add course material
- productionizing Add course material
- templates** Add course material
- .gitignore Add course material
- README.md Add course material
- best-practices-and-general-guideli... Add course material

The file 'templates' is highlighted with a green border.

File	Action
installation	Add course material
monitoring	Add course material
productionizing	Add course material
templates	Add course material
.gitignore	Add course material
README.md	Add course material
best-practices-and-general-guideli...	Add course material

Input/Output Ports

Move File

0 0 0 0 3 0 0

Queued 2 (72 bytes)

In 0 (0 bytes) → 0 5 min

Read/Write 0 bytes / 0 bytes 5 min

Out 0 → 0 (0 bytes) 5 min

✓ 0 * 0 ⌂ 0 ! 0 ? 0

Kafka Push

0 0 0 0 1 0 0

Queued 0 (0 bytes)

In 0 (0 bytes) → 0 5 min

Read/Write 0 bytes / 0 bytes 5 min

Out 0 → 0 (0 bytes) 5 min

✓ 0 * 0 ⌂ 0 ! 0 ? 0

Move File

0 0 0 0 3 0 0

Queued 2 (72 bytes)

In 0 (0 bytes) → 0

Read/Write 0 bytes / 0 bytes

Out 0 → 0 (0 bytes)

✓ 0 * 0 ⌂ 0 ! 0 ? 0

Connection Configuration

'Move File' does not have any local output ports.

'Kafka Push' does not have any local input ports.

OK



Input Port

Output Port

Input Port



Port of this type **accepts incoming connection from data sender process group (Ingress)**

Output Port



Port of this type **transfer data as outgoing connection to receiver process group (egress)**

GetFile
GetFile 1.14.0
org.apache.nifi - nifi-standard-nar

In	0 (0 bytes)	
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	

UpdateAttribute
UpdateAttribute 1.14.0
org.apache.nifi - nifi-update-attribute-nar

In	0 (0 bytes)	
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0	

Add Output Port

Output Port Name

Send To

Local connections

CANCEL ADD

Local connections

Local connections

Remote connections (site-to-site)

CANCEL ADD

	GetFile	GetFile 1.14.0
		org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Name **success**

Queued **0 (0 bytes)**

	UpdateAttribute	UpdateAttribute 1.14.0
		org.apache.nifi - nifi-update-attribute-nar
In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Name **success**

Queued **2 (72 bytes)**

Move File	
	
Queued	2 (72 bytes)
In	0 (0 bytes) → 0
Read/Write	0 bytes / 0 bytes
Out	1 → 0 (0 bytes)
	

Inside

Output Port

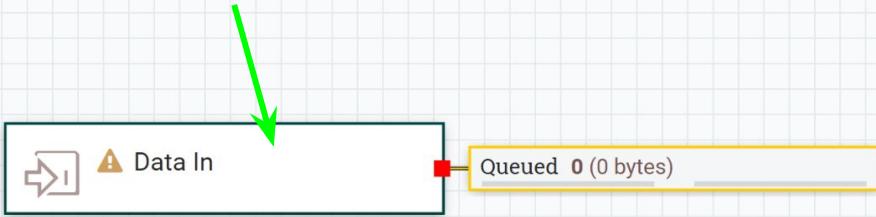
	PutFile	PutFile 1.14.0
		org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Name **success**

Queued **0 (0 bytes)**



Input Port



	PublishKafkaRecord_1_0 PublishKafkaRecord_1_0 1.14.0 org.apache.nifi - nifi-kafka-1-0-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

Inside

Kafka Push		
⟳ 0	⟳ 0	▶ 0
queued 0 (0 bytes)		
In 0 (0 bytes) → 0		5 min
Read/Write 0 bytes / 0 bytes		5 min
Out 0 → 0 (0 bytes)		5 min
✓ 0	* 0	⌚ 0
!	?	0

Funnel



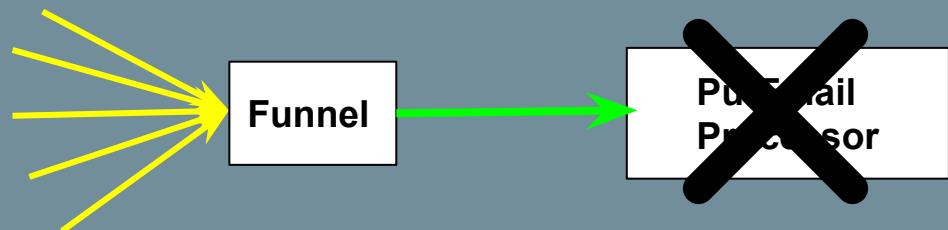
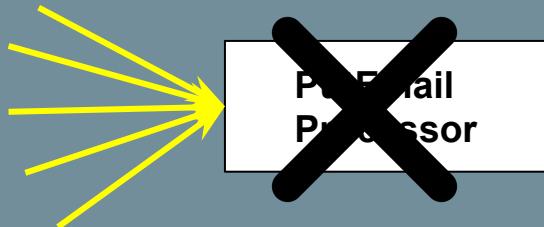
Funnel

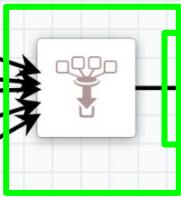
Funnel



Combines several input connections into one output connection

Makes it easy to replace the destination when lot of components are together





Queued 0 (0 bytes)



PutEmail

PutEmail 1.13.2

org.apache.nifi - nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Label



Label

Label



Used for labelling
Components such as
providing detailed
description, beautify
etc..

Label

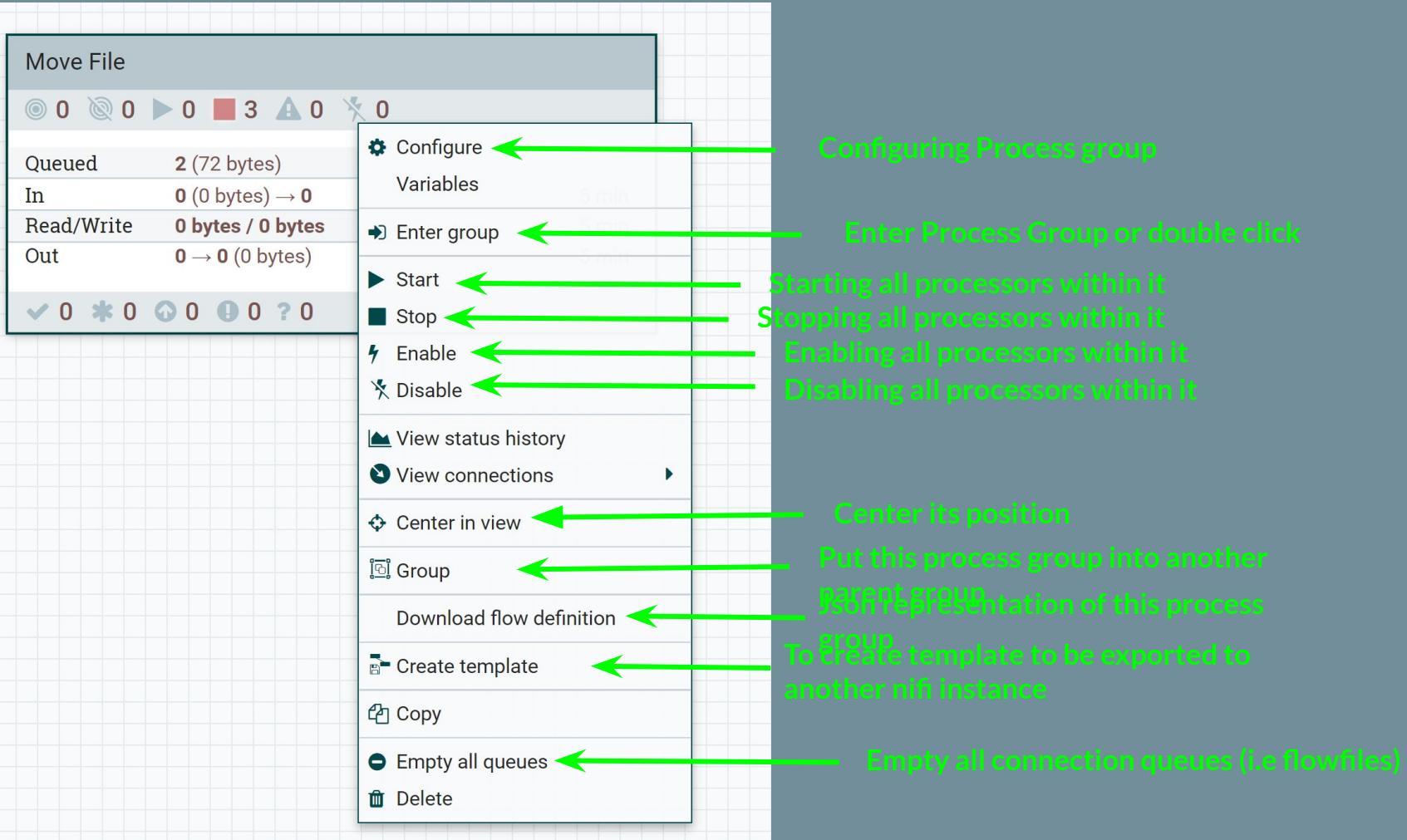
Move File

⌚ 0	⌚ 0	▶ 2	🔴 1	⚠ 1	✖ 0
Queued	0 (0 bytes)				
In	0 (0 bytes) → 0		5 min		
Read/Write	0 bytes / 0 bytes			5 min	
Out	1 → 0 (0 bytes)			5 min	
✓ 0	* 0	⌚ 0	⚠ 0	⌚ 0	?

Moves a file from source to destination

Move File

⌚ 0	⌚ 0	▶ 2	🔴 1	⚠ 1	✖ 0
Queued	0 (0 bytes)				
In	0 (0 bytes) → 0		5 min		
Read/Write	0 bytes / 0 bytes			5 min	
Out	1 → 0 (0 bytes)			5 min	
✓ 0	* 0	⌚ 0	⚠ 0	⌚ 0	?





Move File Configuration

GENERAL CONTROLLER SERVICES

Controller Service Tab

Process Group Name

Move File



Name of Process Group

Process Group Parameter Context

No parameter context



List of parameter context

Process Group Comments



PG description

Process Group FlowFile Concurrency

Unbounded



Controlling flowfile output

Single FlowFile Per Node



When to route flowfile

Stream When Available



Stream When Available



Batch Output



How long to hold flowfiles

Default Back Pressure Object Threshold

10000



When to start throttling

Default Back Pressure Data Size Threshold

1 GB



When to start throttling
(Size)

APPLY

FlowFile Concurrency Policy

Single FlowFile per Node



Batch Output



Stream When Available

flowfiles



Move File Configuration

GENERAL CONTROLLER SERVICES

Process Group Name
Move File

Process Group Parameter Context
No parameter context

Process Group Comments

Process Group FlowFile Concurrency
Unbounded

Process Group Outbound Policy
Stream When Available

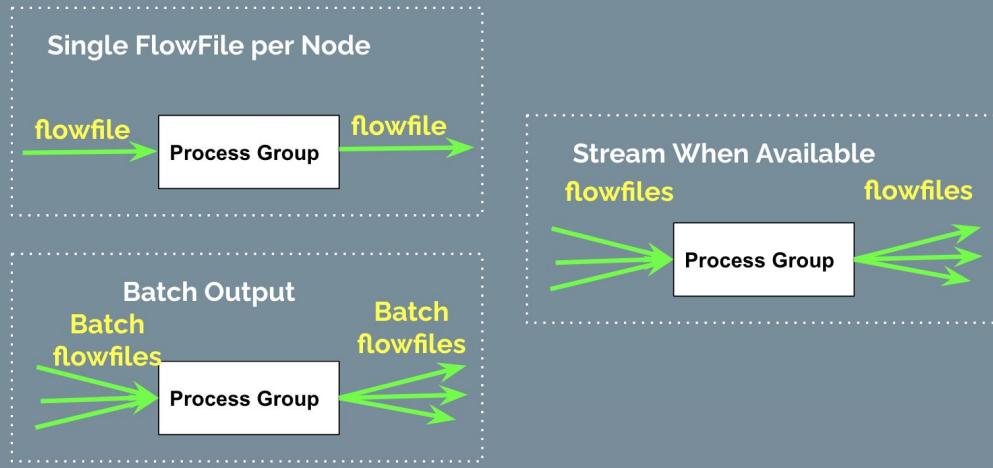
Default FlowFile Expiration
0 sec

Default Back Pressure Object Threshold
10000

Default Back Pressure Data Size Threshold
1 GB

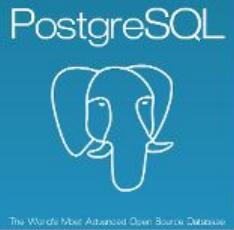
APPLY

FlowFile Concurrency Policy



Lab : Reading from Database

– Reading from Database



```
postgres=# CREATE TABLE EMPLOYEE(  
    ID INT PRIMARY KEY      NOT NULL,  
    NAME          VARCHAR(50) NOT NULL,  
    COMPANY       VARCHAR(50) NOT NULL  
);  
CREATE TABLE
```

ID	NAME	COMPANY
1	John	Apache
2	Simon	NiFi
3	Mark	Apache

Executing Select query

```
postgres=# select * from EMPLOYEE;
 id | name           | company
----+-----+-----+
  1 | John          | Apache
  2 | Simon         | NiFi
  3 | Mark          | Apache
(3 rows)
```



0



0 / 0 bytes



0



0



0



1



1

ExecuteSQL
ExecuteSQL 1.13.2
org.apache.nifi - nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Configure Processor

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

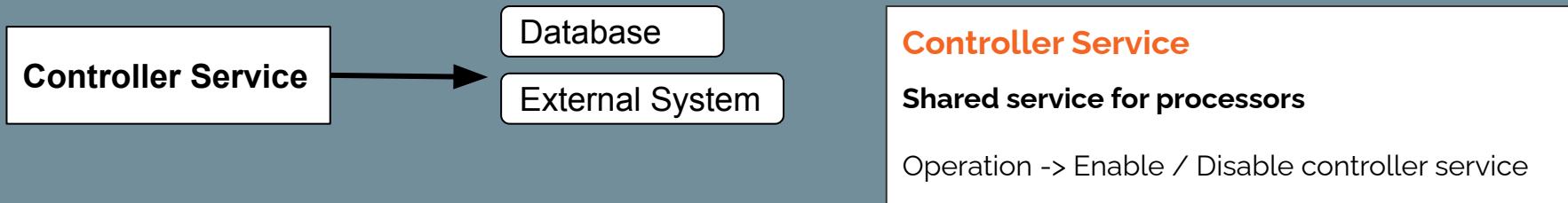
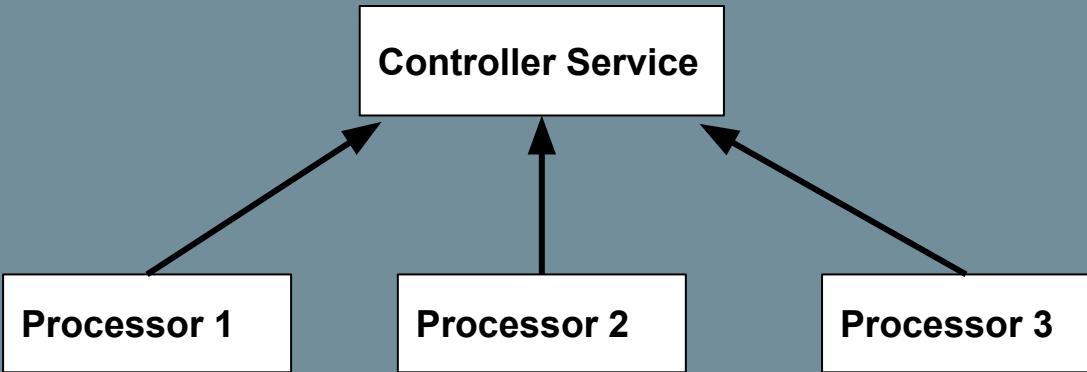


Property	Value
Database Connection Pooling Service	No value set The Controller Service that is used to obtain connection to database
SQL Pre-Query	
SQL select query	
SQL Post-Query	
Max Wait Time	
Normalize Table/Column Names	
Use Avro Logical Types	
Compression Format	NONE
Default Decimal Precision	10
Default Decimal Scale	0
Max Rows Per Flow File	0
Output Batch Size	0

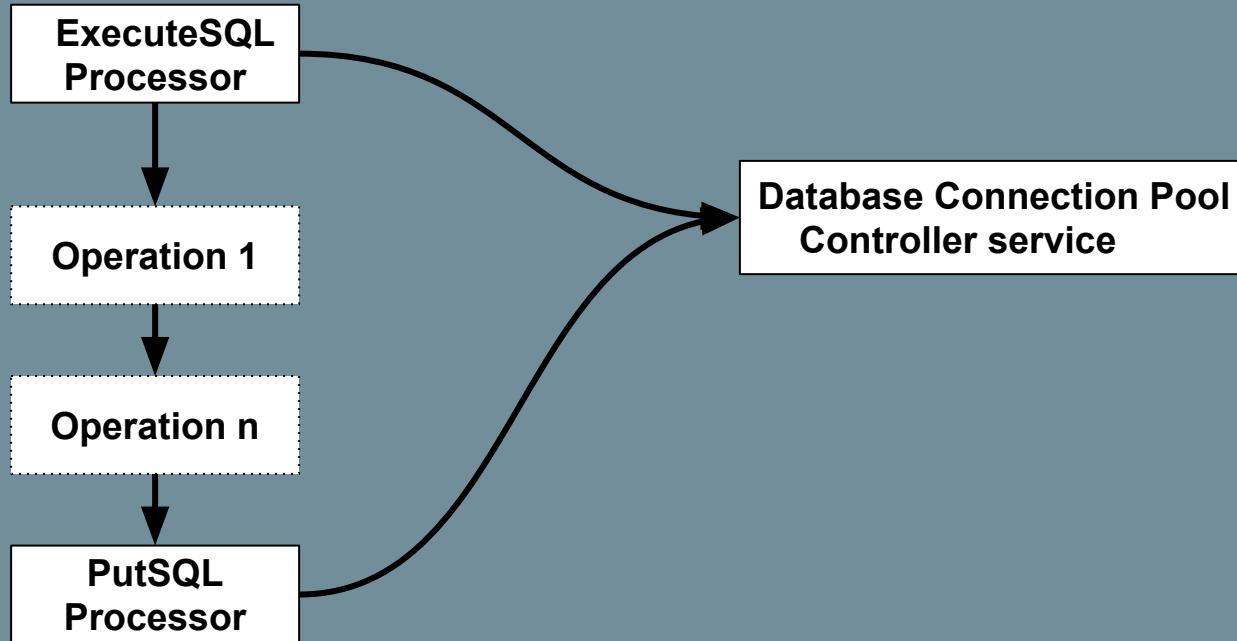
CANCEL

APPLY

Controller Service



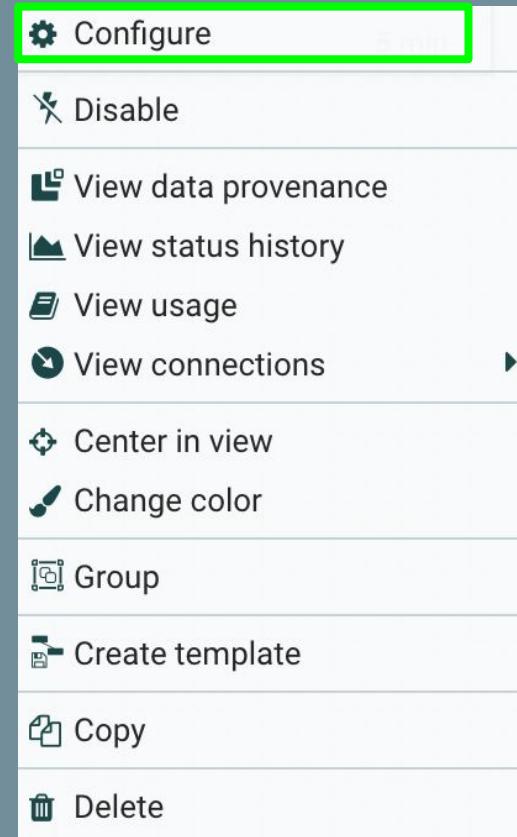
Analogy of Controller Service use



– Reading from Database

	ExecuteSQL
	ExecuteSQL 1.13.2
	org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

Right click the processor



Creating Controller Service

Configure Processor | ExecuteSQL 1.15.3

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property

Database Connection Pooling Service

No value

Create new service...

SQL Pre-Query

SQL select query

SQL Post-Query

Max Wait Time

Normalize Table/Column Names

Use Avro Logical Types

Compression Format

Default Decimal Precision

Default Decimal Scale

Max Rows Per Flow File

Output Batch Size

CANCEL APPLY

Property	Value
Database Connection Pooling Service	No value
SQL Pre-Query	
SQL select query	
SQL Post-Query	
Max Wait Time	0 seconds
Normalize Table/Column Names	false
Use Avro Logical Types	false
Compression Format	NONE
Default Decimal Precision	10
Default Decimal Scale	0
Max Rows Per Flow File	0
Output Batch Size	0

Add Controller Service

Source Displaying 100 of 100 Filter

all groups Type Version Tags

Type	Version	Tags
CassandraDistributedMapC...	1.14.0	cache, cassandra, distributed, ...
CassandraSessionProvider	1.14.0	database, pooling, cassandra, ...
ConfluentSchemaRegistry	1.14.0	schema, registry, confluent, ka...
CouchbaseClusterService	1.14.0	database, couchbase, connect...
CouchbaseKeyValueLooku...	1.14.0	lookup, couchbase, enrich, val...
CouchbaseMapCacheClient	1.14.0	cluster, cache, couchbase, dist...
CouchbaseRecordLookupS...	1.14.0	lookup, couchbase, enrich
DBCPConnectionPool	1.14.0	database, pooling, dbcp, jdbc, ...
DBCPConnectionPoolLook...	1.14.0	database, pooling, dbcp, jdbc, ...
DatabaseRecordLookupSer...	1.14.0	lookup, cache, database, rdbm...
DatabaseRecordSink	1.14.0	database, record, jdbc, connec...
DistributedMapCacheClient...	1.14.0	cluster, cache, distributed, stat...

cache connection credentials database distributed enrich invoke javascript jruby js jython key lookup luaj map parse python reader record recordset restricted row script value writer

DBCPConnectionPool 1.14.0 org.apache.nifi - nifi-dbc-pool-service-nar

Provides Database Connection Pooling Service. Connections can be asked from pool and returned after usage.

Brief description

Controller Service Name

Adding Controller Service

CANCEL ADD

Move File Configuration

GENERAL		CONTROLLER SERVICES			
Name ▾	Type	Bundle	State	Scope	
DBCPConnectionPool	DBCPConnectionPool 1.14.0	org.apache.nifi - nifi-dbcn-service-nar	Invalid	Move File	

Documentation



Configure Delete

Configure Controller Service

SETTINGS PROPERTIES COMMENTS

Name

DBCPConnectionPool

Referencing Components [?](#)

No referencing components.

Id

e1213959-017c-1000-4e68-58eabf2b79af

Type

DBCPConnectionPool 1.14.0

Bundle

org.apache.nifi - nifi-dbcn-service-nar

Supports Controller Service

- DBCPService 1.14.0 from org.apache.nifi - nifi-standard-services-api-nar

Name of controller service

All processors referencing this Controller service

Controller service build information

CANCEL

APPLY

Configure Controller Service

SETTINGS PROPERTIES COMMENTS

Required field

Property	Value	
Database Connection URL	?	No value set
Database Driver Class Name	?	No value set
Database Driver Location(s)	?	No value set
Kerberos Credentials Service	?	No value set
Kerberos Principal	?	No value set
Kerberos Password	?	No value set
Database User	?	No value set
Password	?	No value set
Max Wait Time	?	500 millis
Max Total Connections	?	8
Validation query	?	No value set
Minimum Idle Connections	?	0
Max Idle Connections	?	8
Max Connection Lifetime	?	-1

CANCEL **APPLY**

GENERAL CONTROLLER SERVICES

Configure Controller Service

SETTINGS PROPERTIES COMMENTS

Name PostgresConnectionPool

Referencing Components ⓘ
No referencing components.

Id e9fe78eb-017c-1000-05a4-9938a91f248c

Type DBCPConnectionPool 1.13.2

Bundle org.apache.nifi - nifi-dbc-p-service-nar

Supports Controller Service

- DBCPService 1.13.2 from org.apache.nifi - nifi-standard-services-api-nar

CANCEL APPLY



Custom name for the Controller service

Configure Controller Service

SETTINGS	PROPERTIES	COMMENTS
Required field		
Property	Value	
Database Connection URL	jdbc:postgresql://postgres:5432/postgres	
Database Driver Class Name	org.postgresql.Driver	
Database Driver Location(s)	/opt/nifi/nifi-current/postgresql-42.2.2.jar	
Kerberos Credentials Service	No value set	
Kerberos Principal	No value set	
Kerberos Password	No value set	
Database User	postgres	
Password	Sensitive value set	
Max Wait Time	500 millis	
Max Total Connections	8	
Validation query	No value set	
Minimum Idle Connections	0	
Max Idle Connections	8	
Max Connection Lifetime	-1	

CANCEL **APPLY**

NiFi Flow Configuration

GENERAL		CONTROLLER SERVICES			
Name	PostgresConnectionPool	Type	DBCPConnectionPool 1.13.2	Bundle	org.apache.nifi - nifi-dbcn-service-nar
				State	Disabled

 **Enable Controller Service**

NiFi Flow Configuration

GENERAL		CONTROLLER SERVICES			
Name	PostgresConnectionPool	Type	DBCPConnectionPool 1.13.2	Bundle	org.apache.nifi - nifi-dbcn-service-nar
				State	Enabled

Configure Processor



Invalid

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field



Property

Database Connection Pooling Service

PostgresConnectionPool



PostgresConnectionPool

SQL Pre-Query

SQL select query

SQL Post-Query

Max Wait Time

Normalize Table/Column Names

Use Avro Logical Types

Compression Format

Default Decimal Precision

Default Decimal Scale

Max Rows Per Flow File

Output Batch Size

Create new service...

CANCEL

APPLY

Configure Processor

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field



Property	Value
Database Connection Pooling Service	PostgresConnectionPool
SQL Pre-Query	No value set
SQL select query	No value set
SQL Post-Query	No value set
Max Wait Time	0 seconds
Normalize Table/Column Names	false
Use Avro Logical Types	false
Compression Format	NONE
Default Decimal Precision	10
Default Decimal Scale	0
Max Rows Per Flow File	0
Output Batch Size	0

CANCEL

APPLY

```
postgres=# select * from EMPLOYEE;
 id | name
----+-----
 1 | John
 2 | Simon
 3 | Mark
(3 rows)
```

name

company

Configure Processor

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Database Connection Pooling Service	PostgresConnectionPool
SQL Pre-Query	
SQL select query	1 Select * from Employee
SQL Post-Query	
Max Wait Time	
Normalize Table/Column Names	
Use Avro Logical Types	
Compression Format	
Default Decimal Precision	
Default Decimal Scale	
MAX ROWS PER FLOW FILE	0
Output Batch Size	0

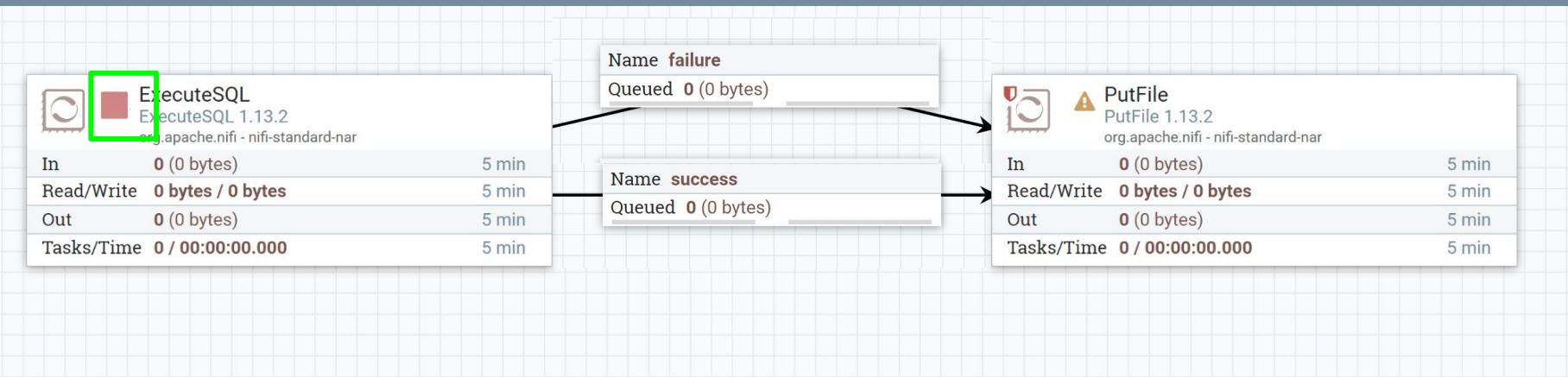
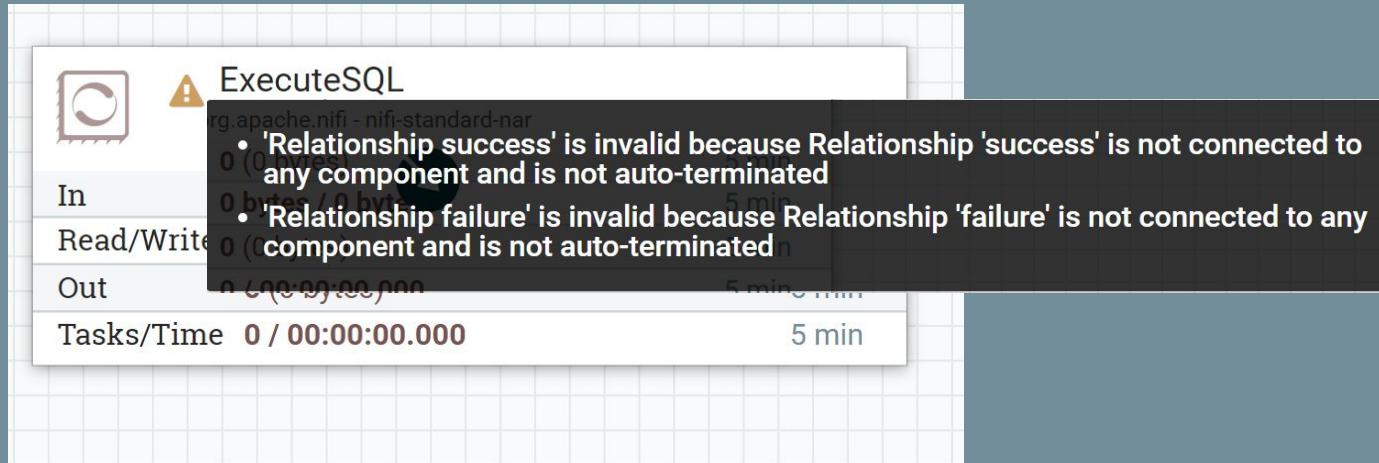
EL ✓ PARAM ✓

1 Select * from Employee

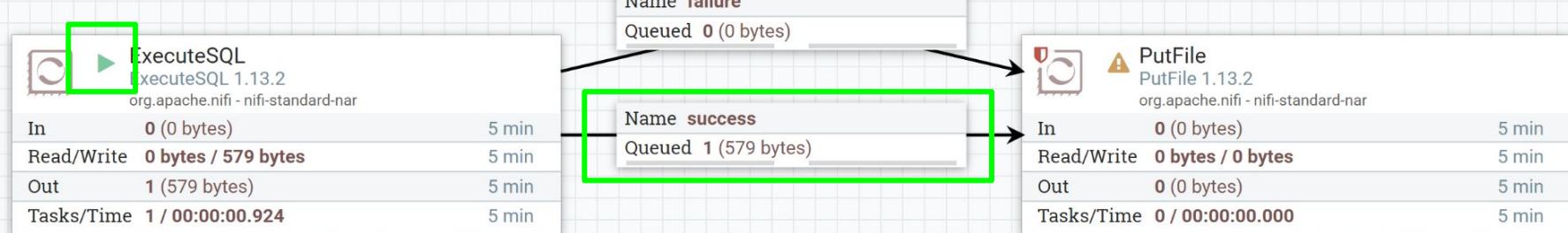
Set empty string

CANCEL OK

CANCEL APPLY



Running ExecuteSQL Processor



The screenshot shows a browser interface displaying the contents of an Avro file. The file is named `f284fcbb8-f717-414d-89e8-241c84338ed1` and has a Content Type of `application/avro-binary`. The file is viewed in "original" mode. The content of the file is a single record:

```
1 Obj...avro.schema{"type": "record", "name": "employee", "namespace": "any.data", "fields": [{"name": "id", "type": ["null", "int"]}, {"name": "name", "type": ["null", "string"]}, {"name": "company", "type": "string"}]}
```

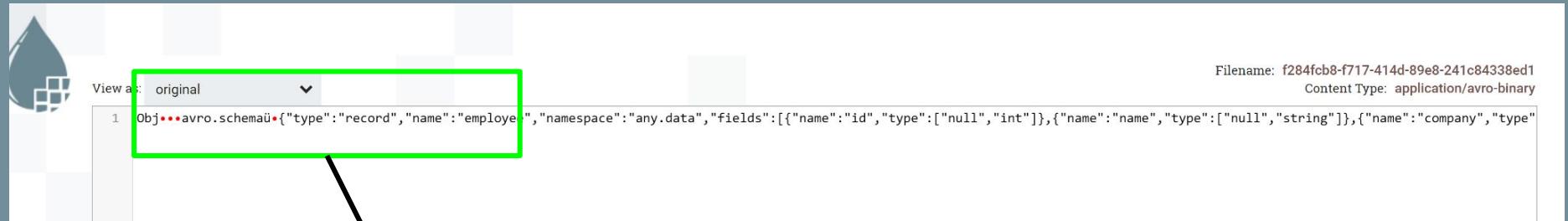
- What is Avro

It's a Binary Format.

Requires less size compared to JSON/CSV

**Data is serialized and deserialized easily
with schema**

Very much efficient to pull data



A screenshot of a file viewer interface. At the top, there's a navigation bar with a back button, a file icon, and a search bar containing the text "f284fc...". Below the navigation bar, the file information is displayed: "Filename: f284fc8-f717-414d-89e8-241c84338ed1" and "Content Type: application/avro-binary". The main area shows the file content with a dropdown menu labeled "View as: original". A green rectangular box highlights the first few lines of the code. A black arrow points from this highlighted area down to the text "Need Conversion from Avro to some other text format" located below the screenshot.

```
1 Obj***avro.schema***{"type": "record", "name": "employee", "namespace": "any.data", "fields": [{"name": "id", "type": ["null", "int"]}, {"name": "name", "type": ["null", "string"]}, {"name": "company", "type": "string"}]}
```

Need Conversion from Avro to
some other text format

Converting Avro to JSON

The image shows the NiFi user interface. On the left, there is a blue vertical bar with a water droplet icon. To its right is a list of processors. The first processor is highlighted with a green box and labeled "ConvertAvroToJSON". Below it, the processor details are shown:

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Below the processor list, there is a "View as" dropdown menu set to "original". A green box highlights the dropdown and the schema definition:

```
1 Obj***avro.schema***{"type": "record", "name": "employee", "namespace": "any.data", "fields":
```

Source as Avro

The image shows the JSON output view. At the top, it says "View as: formatted". The JSON data is displayed as follows:

```
1 [ {  
2   "id" : 1,  
3   "name" : "John",  
4   "company" : "Apache"  
5 }, {  
6   "id" : 2,  
7   "name" : "Simon",  
8   "company" : "NiFi"  
9 }, {  
10  "id" : 3,  
11  "name" : "Mark",  
12  "company" : "Apache"  
13 } ]
```

Target as JSON



ExecuteSQL

ExecuteSQL 1.13.2

org.apache.nifi - nifi-standard-nar

In 0 (0 bytes) 5 min

Read/Write 0 bytes / 579 bytes 5 min

Out 1 (579 bytes) 5 min

Tasks/Time 1 / 00:00:00.924 5 min



ConvertAvroToJson

ConvertAvroToJson 1.13.2

org.apache.nifi - nifi-avro-nar

In 0 (0 bytes) 5 min

Read/Write 0 bytes / 0 bytes 5 min

Out 0 (0 bytes) 5 min

Tasks/Time 0 / 00:00:00.000 5 min



PutFile

PutFile 1.13.2

org.apache.nifi - nifi-standard-nar

In 0 (0 bytes) 5 min

Read/Write 0 bytes / 0 bytes 5 min

Out 0 (0 bytes) 5 min

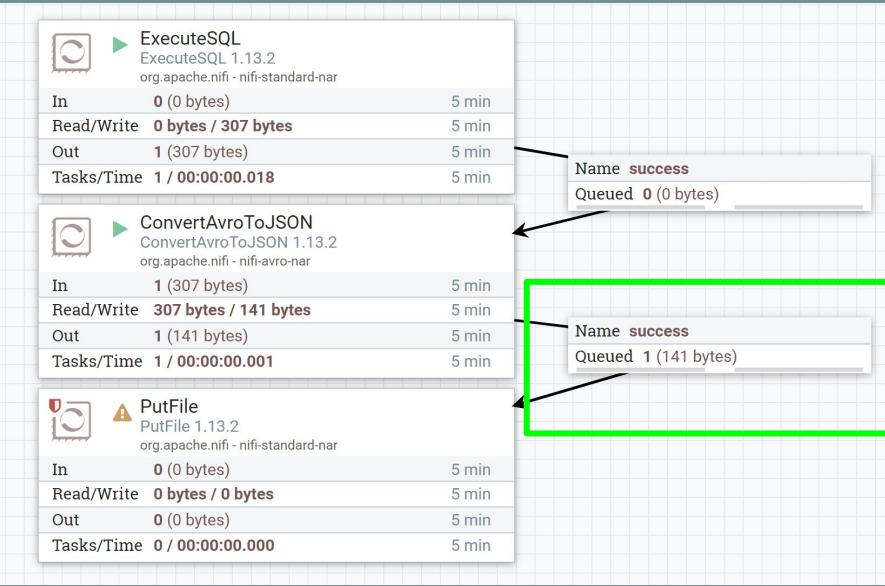
Tasks/Time 0 / 00:00:00.000 5 min

Name success

Queued 1 (579 bytes)

Name success

Queued 0 (0 bytes)



Name success
Queued 0 (0 bytes)

Name success
Queued 1 (141 bytes)

View as: formatted ▾

```
[ {  
  "id" : 1,  
  "name" : "John",  
  "company" : "Apache"  
}, {  
  "id" : 2,  
  "name" : "Simon",  
  "company" : "NiFi"  
}, {  
  "id" : 3,  
  "name" : "Mark",  
  "company" : "Apache"  
} ]
```

Digging into Canvas

Bread Crumbs (Navigating Path)



0 2 / 72 bytes

0

0

0

0

3

3

0

0

0

0

0

0

23:27:54 IST

Q □

Navigate

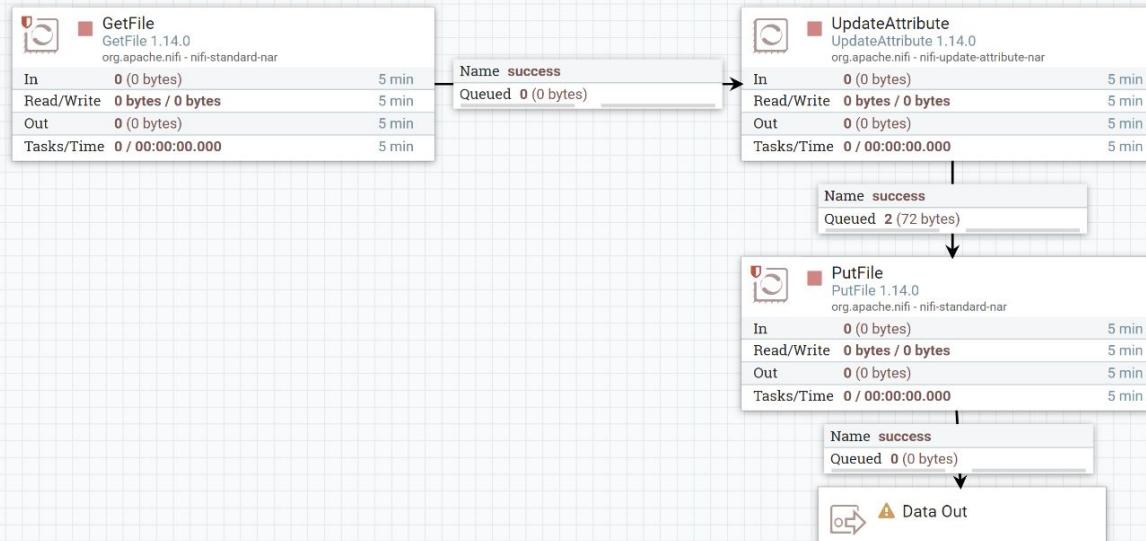


Operate

Move File

Process Group

8520a736-017c-1000-ed4c-8a14ebba3dcc



Bread Crumbs

Bread Crumbs

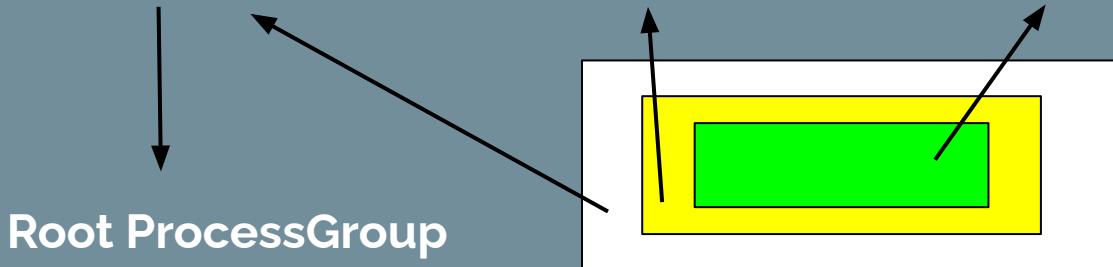


Bread Crumbs

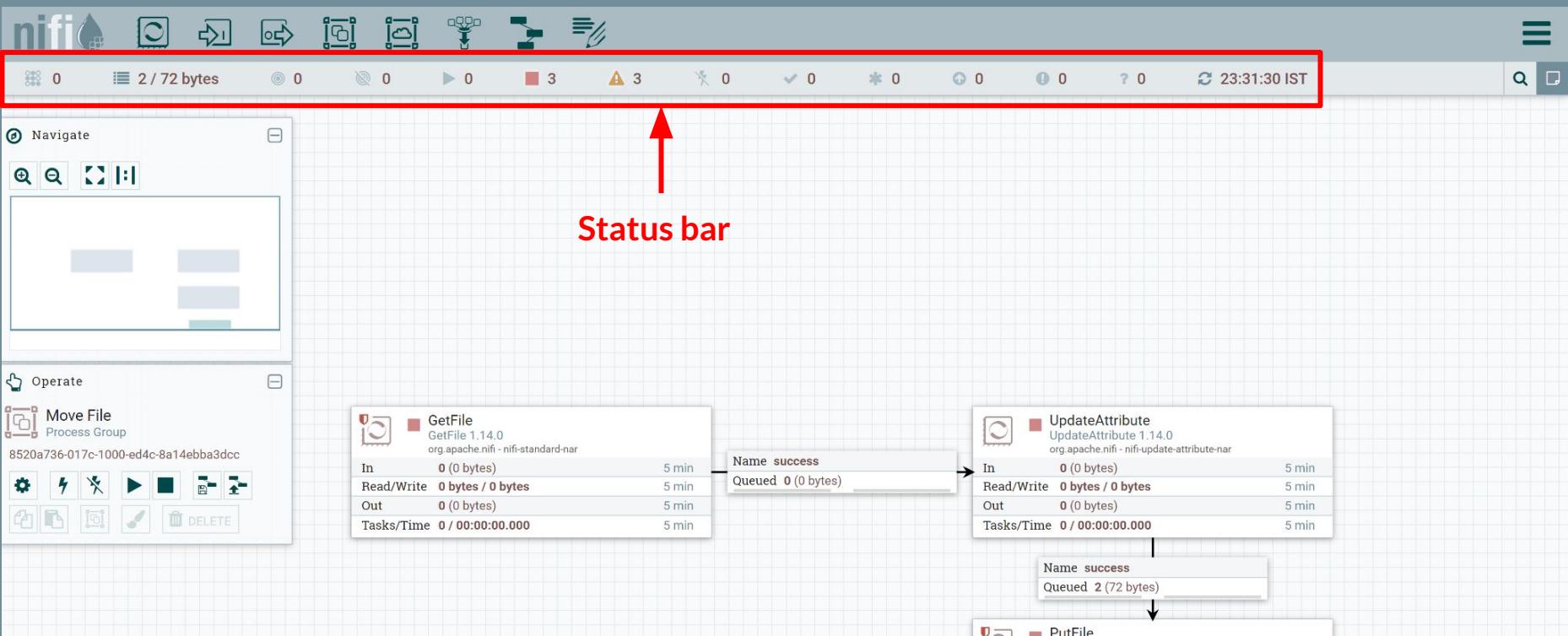
NiFi Flow » Move File

Example

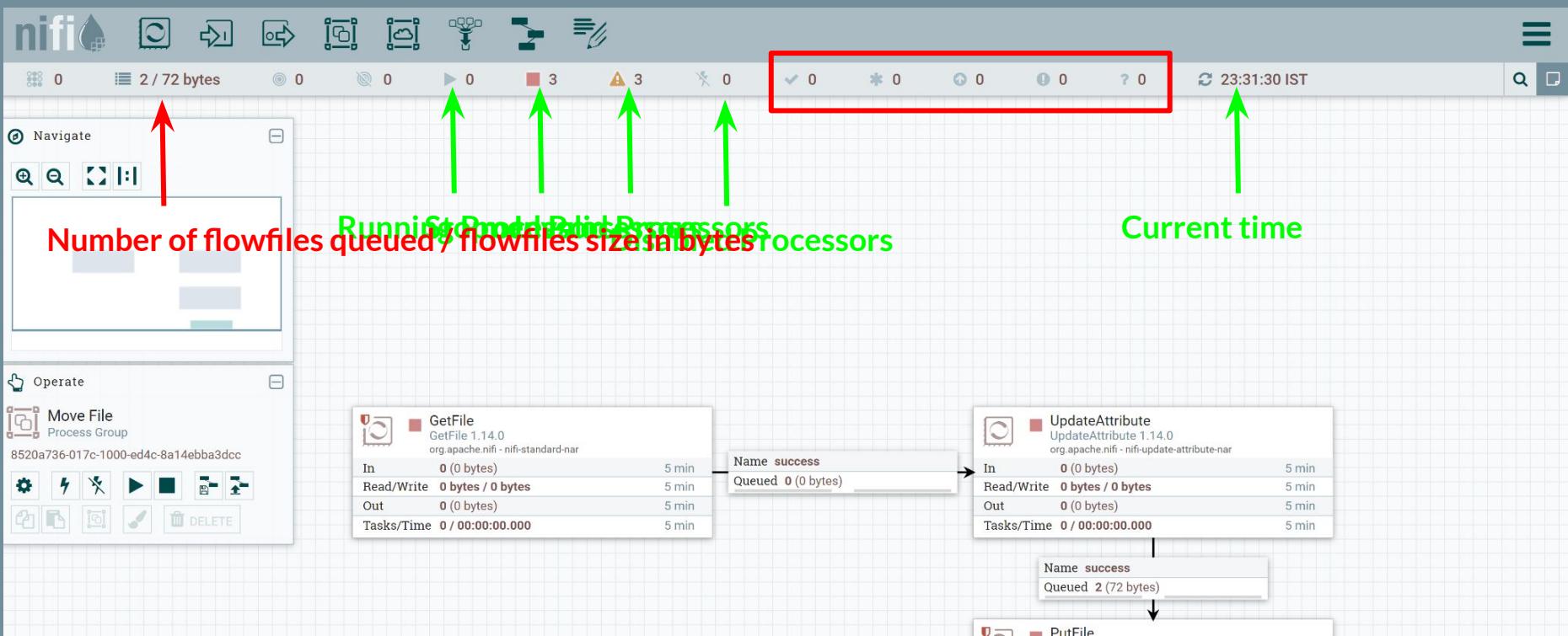
NiFi Flow / ProcessGroup A / ProcessGroup B



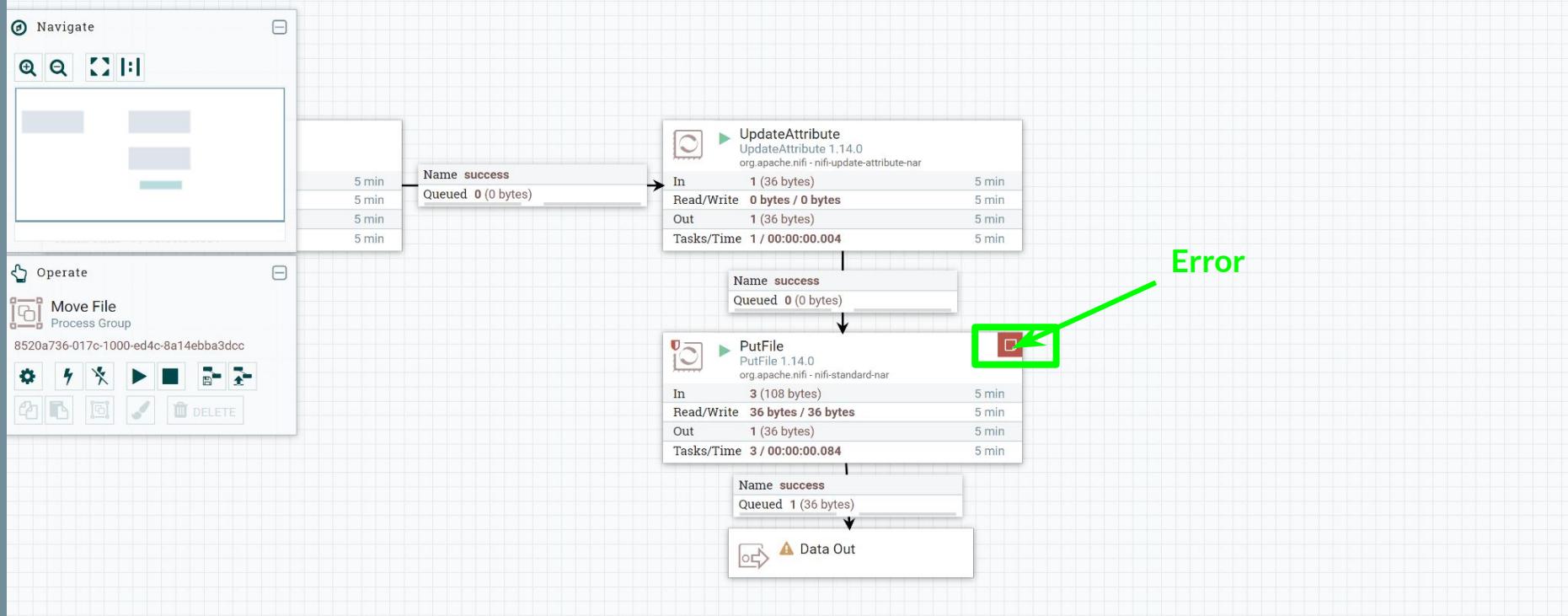
Status Bar

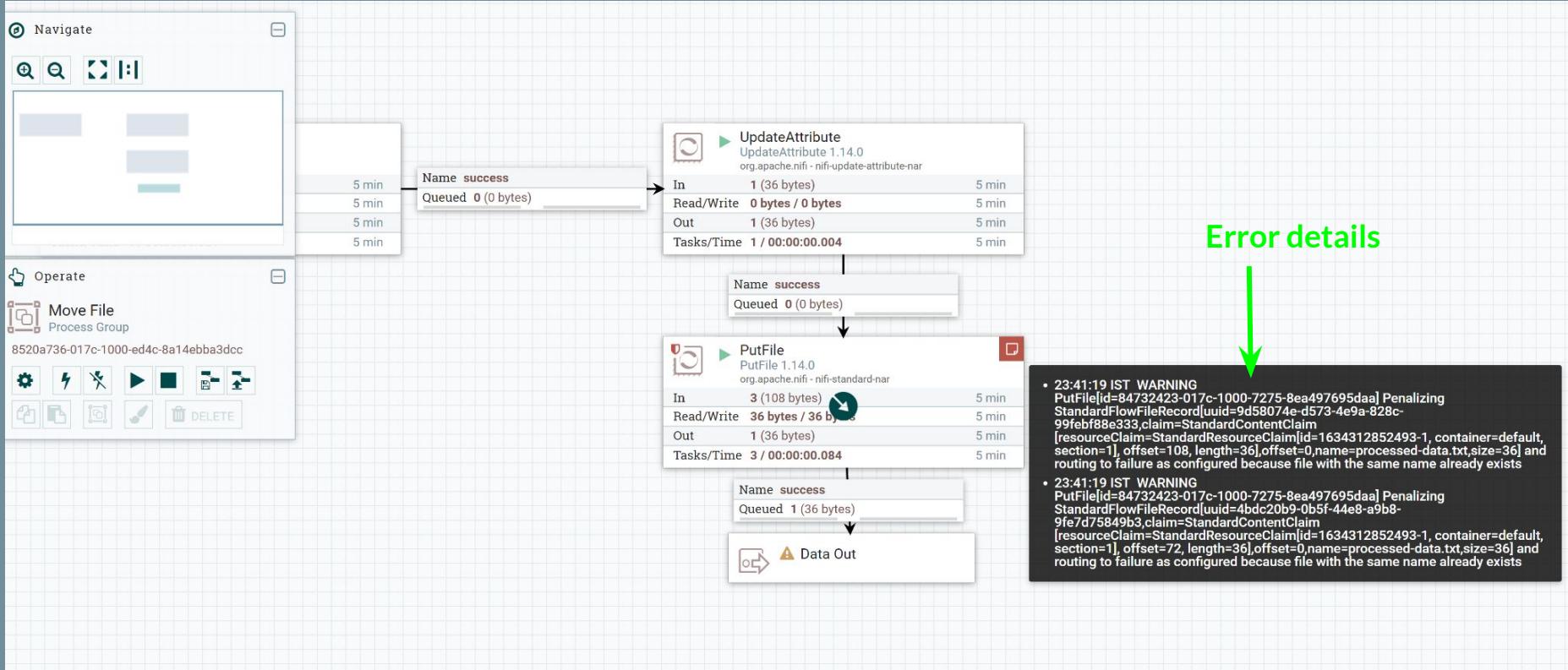


Status bar

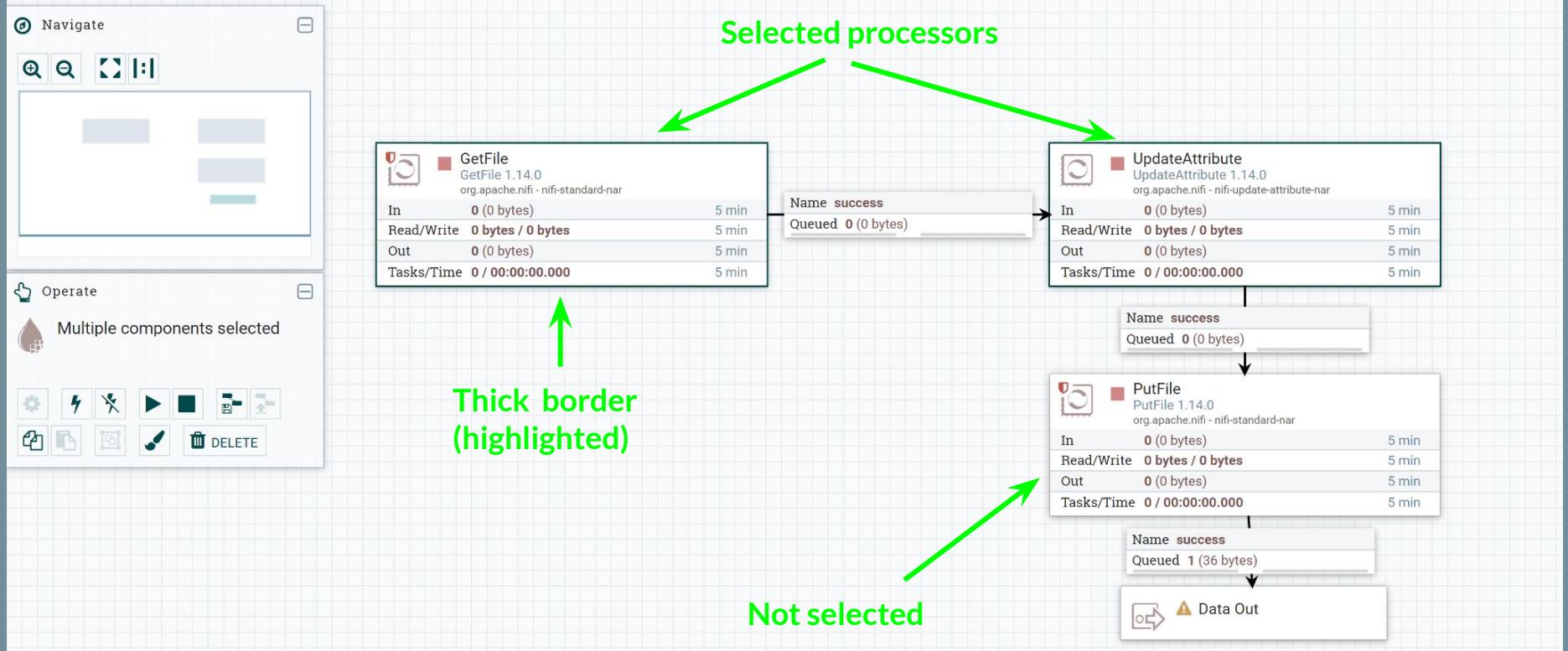


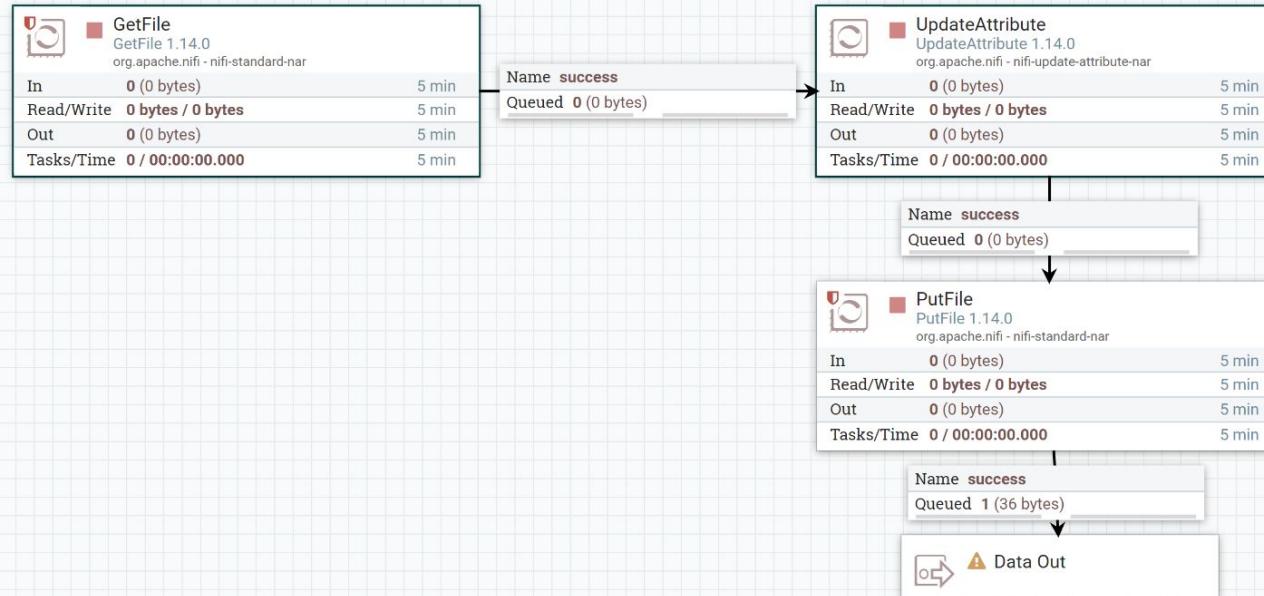
Processor Errors



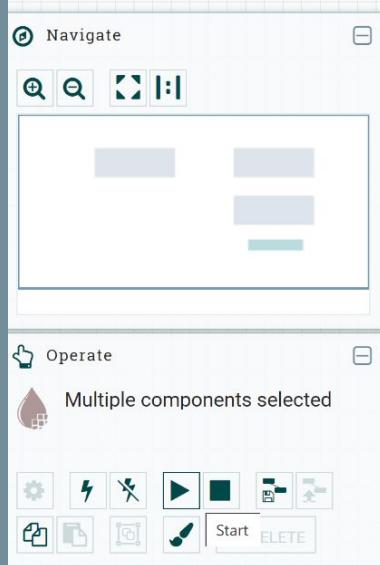


Selecting Processors





Start
Selected
processors



state running

Started processors



Name success
Queued 0 (0 bytes)



Name success
Queued 0 (0 bytes)



Name success
Queued 1 (36 bytes)



active threads

– Difference between Processors & Controller Services

Controller Service	Processor
Shared by processors	No sharing
Enable / Disable	Start / Stop
No Scheduling strategy	Scheduling strategy

Another way to creating
Controller Service

Move File Configuration

GENERAL	CONTROLLER SERVICES				
Name	Type	Bundle	State	Scope	



Adding Controller Service



Last updated: 19:22:04 IST

Listed services are available to all descendant Processors and services of this Process Group.

Working with Processors

Splitting and Routing

Splitting and Routing

id	ident	type	name	latitude_deg	longitude_deg	elevation_ft	continent	iso_country	iso_region	municipality	scheduled_service	gps_code	iata_code	local_code	home_link
6523	00A	heliport	Total Rf Heliport	40.07080078125	-74.93360137939453	11	NA	US	US-PA	Bensalem	no	00A		00A	
6524	00AK	small_airport	Lowell Field	59.947733	-151.692524	450	NA	US	US-AK	Anchor Point	no	00AK		00AK	
6525	00AL	small_airport	Epps Airpark	34.86479949951172	-86.77030181884766	820	NA	US	US-AL	Harvest	no	00AL		00AL	
6526	00AR	closed	Newport Hospital & Clinic Heliport	35.6087	-91.254698	237	NA	US	US-AR	Newport	no				
6527	00AZ	small_airport	Cordes Airport	34.305599212646484	-112.16500091552734	3810	NA	US	US-AZ	Cordes	no	00AZ		00AZ	
6528	00CA	small_airport	Goldstone (GTS) Airport	35.35474	-116.885329	3038	NA	US	US-CA	Barstow	no	00CA		00CA	
6529	00CO	closed	Cass Field	40.622202	-104.344002	4830	NA	US	US-CO	Briggsdale	no				
6531	00FA	small_airport	Grass Patch Airport	28.64550018310547	-82.21900177001953	53	NA	US	US-FL	Bushnell	no	00FA		00FA	
6532	00FD	heliport	Ringhaver Heliport	28.846599578857422	-82.34539794921875	25	NA	US	US-FL	Riverview	no	00FD		00FD	

Splitting and Routing

EL ✓ PARAM ✓

```
1 6523,"00A","heliport", "Total Rf Heliport",40.07080078125,-74.93360137939453,11,"NA","US","US-PA","Bensalem","no","00A","","00
2 323361,"00A","small_airport", "Aero B Ranch Airport",38.704022,-101.473911,3435,"NA","US","US-KS","Leoti","no","00AA","","00A
3 6524,"00AK","small_airport", "Lowell Field",59.947733,-151.692524,450,"NA","US","US-AK","Anchor Point","no","00AK","","00AK",
4 6525,"00AL","small_airport", "Epps Airpark",34.86479949951172,-86.77030181884766,820,"NA","US","US-AL","Harvest","no","00AL"
5 6526,"00AR","closed", "Newport Hospital & Clinic Heliport",35.6087,-91.254898,237,"NA","US","US-AR","Newport","no",,,,,"00A
6 322127,"00AS", "small_airport", "Fulton Airport",34.9428028,-97.8180194,1100,"NA","US","US-OK","Alex","no","00AS","","00AS",
7 6527,"00AZ", "small_airport", "Cordes Airport",34.305599212646484,-112.16500091552734,3810,"NA","US","US-AZ","Cordes","no","0
8 6528,"00CA", "small_airport", "Goldstone (GTS) Airport",35.35474,-116.885329,3038,"NA","US","US-CA","Barstow","no","00CA","","0
9 324424,"00CL","small_airport", "Williams Ag Airport",39.427188,-121.763427,87,"NA","US","US-CA","Biggs","no","00CL","","00CL",
10 322658,"00CN","heliport", "Kitchen Creek Helibase Heliport",32.7273736,-116.4597417,3350,"NA","US","US-CA","Pine Valley","no
11 6529,"00CO","closed", "Cass Field",40.622202,-104.344002,4830,"NA","US","US-CO","Briggsdale","no",,,,,"00CO"
12 6531,"00FA","small_airport", "Grass Patch Airport",28.64550018310547,-82.21900177001953,53,"NA","US","US-FL","Bushnell","no"
13 6532,"00FD","heliport", "Ringhaver Heliport",28.846599578857422,-82.34539794921875,25,"NA","US","US-FL","Riverview","no","00
14 6533,"00FL","small_airport", "River Oak Airport",27.230899810791016,-80.96920013427734,35,"NA","US","US-FL","Okeechobee","no
15 6534,"00GA","small_airport", "Lt World Airport",33.76750183105469,-84.06829833984375,700,"NA","US","US-GA","Lithonia","no",
16 6535,"00GE","heliport", "Caffrey Heliport",33.889245,-84.73793,957,"NA","US","US-GA","Hiram","no","00GE","","00GE",
17 6536,"00HI","heliport", "Kaupulehu Heliport",19.832715,-155.980233,43,"NA","US","US-HI","Kailua-Kona","no","00HI","","00HI",
18 6537,"00ID","small_airport", "Delta Shores Airport",48.145301818847656,-116.21399688720703,2064,"NA","US","US-ID","Clark For
19 322581,"00IG", "small_airport", "Goltl Airport",39.724028,-101.395994,3359,"NA","US","US-KS","McDonald","no","00IG","","00IG",
20 6538,"00II", "closed", "Bailey Generation Station Heliport",41.644501,-87.122803,600,"NA","US","US-IN","Chesterton","no",,,,,
21 6539,"00IL", "small_airport", "Hammer Airport",41.97840118408203,-89.5604019165039,840,"NA","US","US-IL","Polo","no","00IL",
22 6540,"00IN","heliport", "St Mary Medical Center Heliport",41.51139831542969,-87.2605972290039,634,"NA","US","US-IN","Hobart"
23 6541,"00IS","small_airport", "Hayenga's Cant Find Farms Airport",40.02560043334961,-89.1229019165039,820,"NA","US","US-IL",
24 6542,"00KS", "small_airport", "Hayden Farm Airport",38.72779846191406,-94.93049621582031,1100,"NA","US","US-KS","Gardner","no
25 6543,"00KY", "small_airport", "Robbins Roost Airport",37.409400939941406,-84.61969757080078,1265,"NA","US","US-KY","Stanford"
26 45437,"00LA","heliport", "Shell Chemical East Site Heliport",30.191944,-90.980833,15,"NA","US","US-LA","Gonzales","no","00LA
27 6544,"00LL","heliport", "Ac & R Components Heliport",39.66529846191406,-89.70559692382812,600,"NA","US","US-IL","Chatham","n
```

Set empty string

CANCEL OK

Splitting and Routing



Data Generator

	GenerateFlowFile	
	GenerateFlowFile 1.14.0	
	org.apache.nifi - nifi-standard-nar	
In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Data Splitting

	SplitText	
	SplitText 1.14.0	
	org.apache.nifi - nifi-standard-nar	
In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

	RouteOnContent	
	RouteOnContent 1.14.0	
	org.apache.nifi - nifi-standard-nar	
In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Data Routing

Splitting and Routing

Data Generator

	GenerateFlowFile GenerateFlowFile 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes) 5 min
Read/Write	0 bytes / 0 bytes 5 min
Out	0 (0 bytes) 5 min
Tasks/Time	0 / 00:00:00.000 5 min

Data Splitting

	SplitText SplitText 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes) 5 min
Read/Write	0 bytes / 0 bytes 5 min
Out	0 (0 bytes) 5 min
Tasks/Time	0 / 00:00:00.000 5 min

	RouteOnContent RouteOnContent 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes) 5 min
Read/Write	0 bytes / 0 bytes 5 min
Out	0 (0 bytes) 5 min
Tasks/Time	0 / 00:00:00.000 5 min

Data Routing

Splitting and Routing

Data Generator

	GenerateFlowFile GenerateFlowFile 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

Configure Processor

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

EL ✓ PARAM ✓

1	6523,"00A","heliport","Total Rf Heliport",40.07080078125,-74.9336
2	323361,"00AA","small_airport","Aero B Ranch Airport",38.704022,-1
3	6524,"00AK","small_airport","Lowell Field",59.947733,-151.692524,
4	6525,"00AL","small_airport","Epps Airpark",34.86479949951172,-86.
5	6526,"00AR","closed","Newport Hospital & Clinic Heliport",35.6087
6	322127,"00AS","small_airport","Fulton Airport",34.9428028,-97.818
7	6527,"00AZ","small_airport","Cordes Airport",34.305599212646484,-
8	6528,"00CA","small_airport","Goldstone (GTS) Airport",35.35474,-1
9	324424,"00CL","small_airport","Williams Ag Airport",39.427188,-12
10	322658,"00CN","heliport","Kitchen Creek Helibase Heliport",32.727
11	6529,"00CO","closed","Cass Field",40.622202,-104.344002,4830,"NA"
12	6531,"00FA","small_airport","Grass Patch Airport",28.645500183105
13	6532,"00FD","heliport","Ringhaver Heliport",28.846599578857422,-8
14	6533,"00FL","small_airport","Piney Oak Airport",37.32280001070412
15	

Set empty string

CANCEL OK CANCEL APPLY

The screenshot shows the 'Configure Processor' dialog for a 'GenerateFlowFile' processor. The 'PROPERTIES' tab is selected. A table lists 15 entries of generated flowfile content. The last entry's value is partially visible as '080078125,-74...'. A green box highlights this row, and a red box highlights the last digit '5' of the coordinate. Below the table is a checkbox for 'Set empty string'. At the bottom are 'CANCEL' and 'OK' buttons, with 'CANCEL' also appearing at the very bottom.

Splitting and Routing

Data Splitting

	SplitText	SplitText 1.14.0
		org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Configure Processor

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Line Split Count	No value set
Maximum Fragment Size	No value set
Header Line Count	0
Header Line Marker Characters	No value set
Remove Trailing Newlines	true

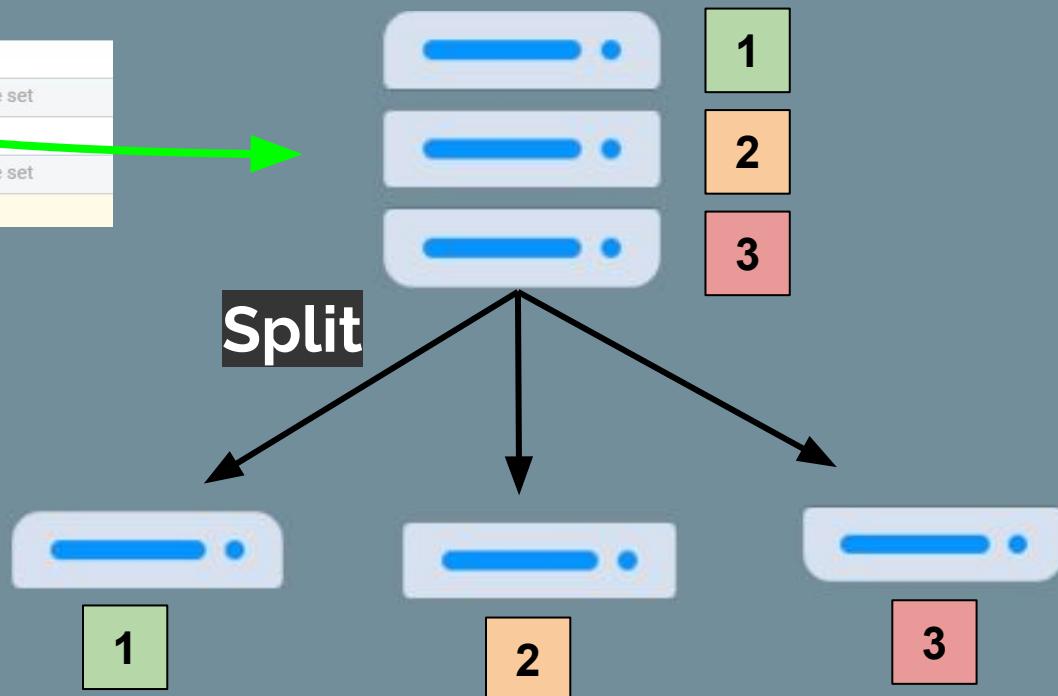
CANCEL APPLY

Splitting Records

Required field

Property	Value
Line Split Count	1
Maximum Fragment Size	No value set
Header Line Count	0
Header Line Marker Characters	No value set
Remove Trailing Newlines	false

3 CSV Records



Splitting and Routing

Data Splitting

	⚠ SplitText SplitText 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)
Read/Write	0 bytes / 0 bytes
Out	0 (0 bytes)
Tasks/Time	0 / 00:00:00.000

Configure Processor

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field +

Property	Value
Line Split Count	1
Maximum Fragment Size	No value set
Header Line Count	0
Header Line Marker Characters	No value set
Remove Trailing Newlines	false

CANCEL APPLY

Splitting and Routing

Data Routing

 ⚠ RouteOnContent	RouteOnContent 1.14.0	org.apache.nifi - nifi-standard-nar
In 0 (0 bytes)	5 min	
Read/Write 0 bytes / 0 bytes	5 min	
Out 0 (0 bytes)	5 min	
Tasks/Time 0 / 00:00:00.000	5 min	

Configure Processor

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

+ 

Property	Value
Match Requirement	content must match exactly
Character Set	content must match exactly
Content Buffer Size	content must contain match
	Reference parameter...

CANCEL APPLY

Splitting and Routing

id	ident	type	name	latitude_deg	longitude_deg	elevation_ft	continent	iso_country	iso_region	municipality	scheduled_service	gps_code	iata_code	local_code	home_link
6523	00A	heliport	Total Rf Heliport	40.07080078125	-74.93360137939453	11	NA	US	US-PA	Bensalem	no	00A		00A	
6524	00AK	small_airport	Lowell Field	59.947733	-151.692524	450	NA	US	US-AK	Anchor Point	no	00AK		00AK	
6525	00AL	small_airport	Epps Airpark	34.86479949951172	-86.77030181884766	820	NA	US	US-AL	Harvest	no	00AL		00AL	
6526	00AR	closed	Newport Hospital & Clinic Heliport	35.6087	-91.254698	237	NA	US	US-AR	Newport	no				
6527	00AZ	small_airport	Cordes Airport	34.305599212646484	-112.16500091552734	3810	NA	US	US-AZ	Cordes	no	00AZ		00AZ	
6528	00CA	small_airport	Goldstone (GTS) Airport	35.35474	-116.885329	3038	NA	US	US-CA	Barstow	no	00CA		00CA	
6529	00CO	closed	Cass Field	40.622202	-104.344002	4830	NA	US	US-CO	Briggsdale	no				
6531	00FA	small_airport	Grass Patch Airport	28.64550018310547	-82.21900177001953	53	NA	US	US-FL	Bushnell	no	00FA		00FA	
6532	00FD	heliport	Ringhaver Heliport	28.846599578857422	-82.34539794921875	25	NA	US	US-FL	Riverview	no	00FD		00FD	

Splitting and Routing

Configure Processor

⚠ Invalid

SETTINGS	SCHEDULING	PROPERTIES	COMMENTS
Required field			
Property	Value		
Match Requirement	ⓘ	content must match exactly	
Character Set	ⓘ	UTF-8	
Content Buffer Size	ⓘ	1 MB	

+

CANCEL APPLY

Configure Processor

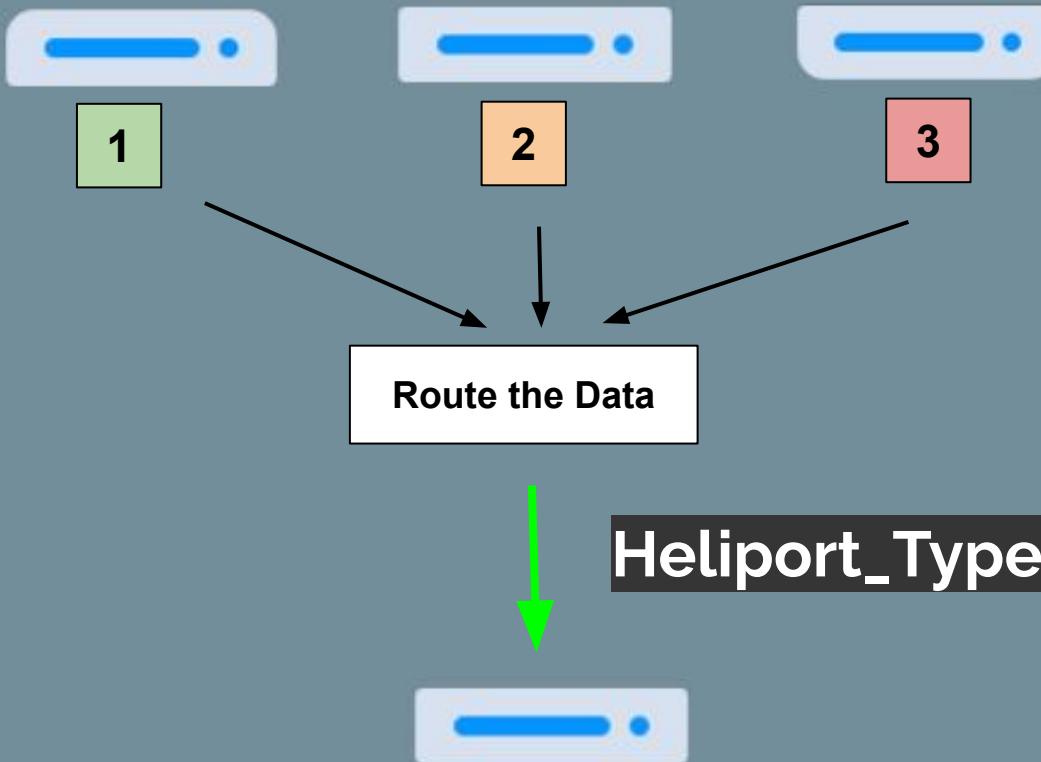
⚠ Invalid

SETTINGS	SCHEDULING	PROPERTIES	COMMENTS
Required field			
Property	Value		
Match Requirement	ⓘ	content must contain match	
Character Set	ⓘ	UTF-8	
Content Buffer Size	ⓘ	1 MB	
Heliport_Type	ⓘ	heliport	

+

CANCEL APPLY

Filtering Records



Splitting and Routing

CSV Data

	GenerateFlowFile GenerateFlowFile 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes) 5 min
Read/Write	0 bytes / 0 bytes 5 min
Out	0 (0 bytes) 5 min
Tasks/Time	0 / 00:00:00.000 5 min

Split line by line

	SplitText SplitText 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes) 5 min
Read/Write	0 bytes / 0 bytes 5 min
Out	0 (0 bytes) 5 min
Tasks/Time	0 / 00:00:00.000 5 min

	RouteOnContent RouteOnContent 1.14.0 org.apache.nifi - nifi-standard-nar
In	0 (0 bytes) 5 min
Read/Write	0 bytes / 0 bytes 5 min
Out	0 (0 bytes) 5 min
Tasks/Time	0 / 00:00:00.000 5 min

Route based on Heliport