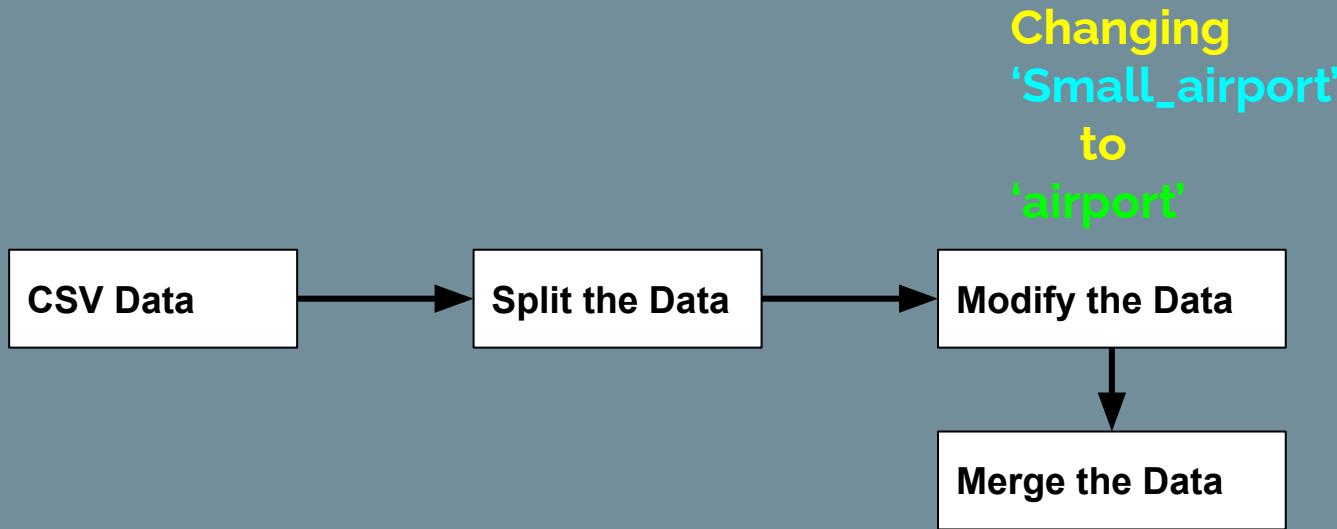


# Splitting, Transforming & Aggregating

# Splitting, Transformation, Aggregating



Changing  
'Small\_airport'  
to  
'airport'

# Splitting, Transformation, Aggregating

## 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

## Merge Data

 ■ MergeContent	MergeContent 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

## Modify Data

 ■ ReplaceText	ReplaceText 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



# Splitting, Transformation, Aggregating

## Modify Data



■ ReplaceText  
ReplaceText 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

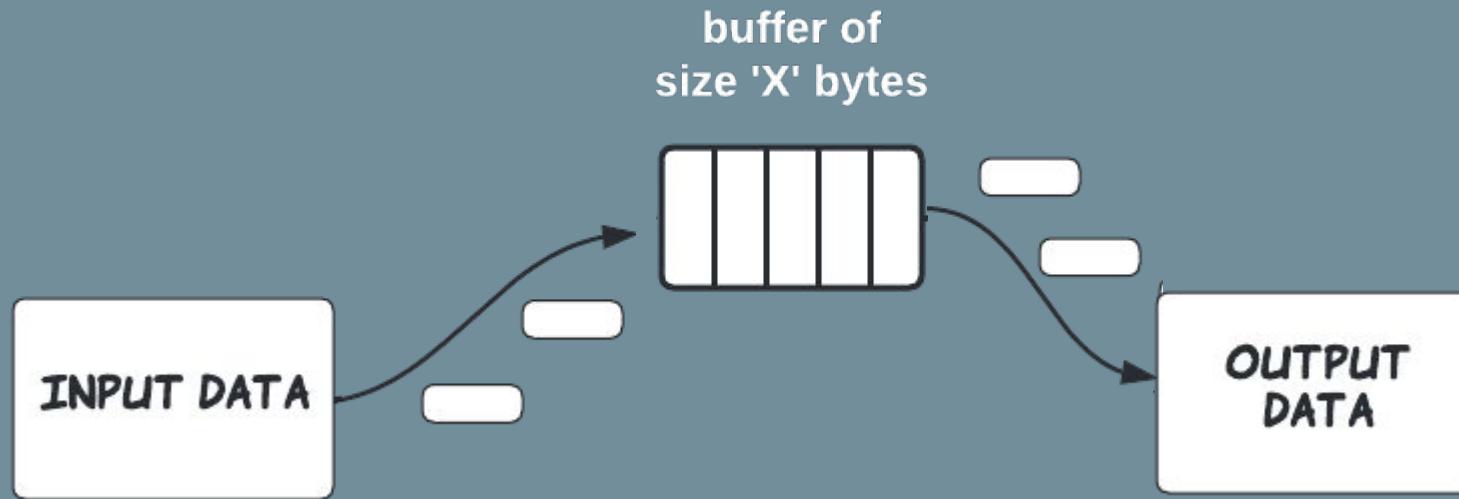
SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Search Value	(2s)(^ *\$)
Replacement Value	S1
Character Octet	UTF-8
Maximum Buffer Size	1 MB
Replacement Strategy	Regex Replace
Evaluation Mode	Line-by-Line
Line-by-Line Evaluation Mode	All

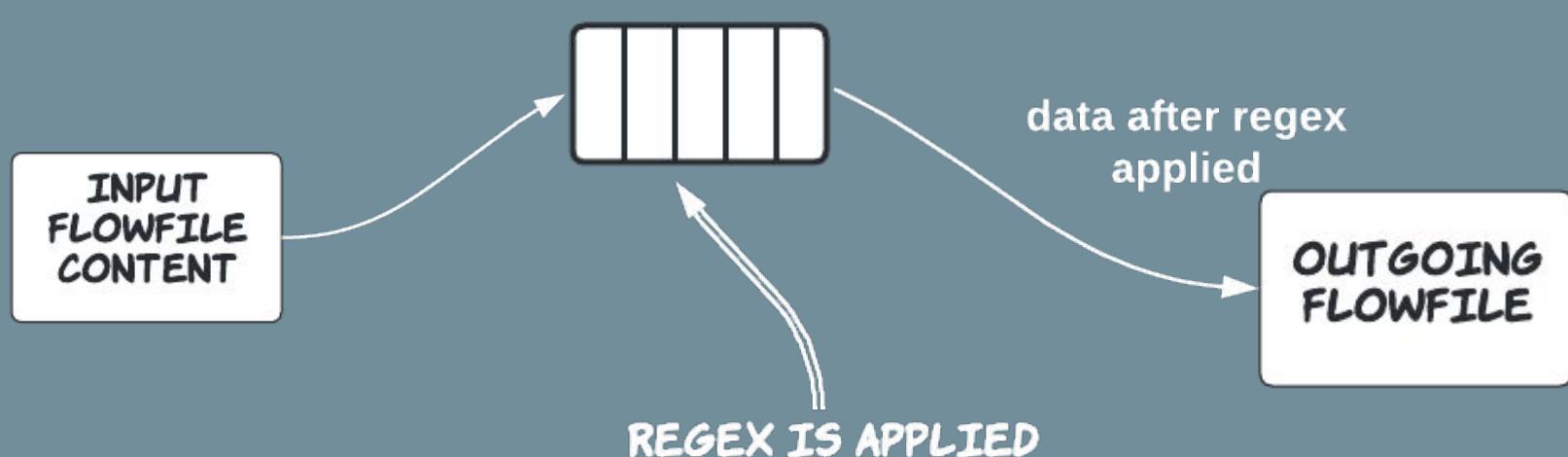
+ CANCEL APPLY

# Splitting, Transformation, Aggregating



# Splitting, Transformation, Aggregating

Let's say we have buffer size of **1MB**



# Splitting, Transformation, Aggregating

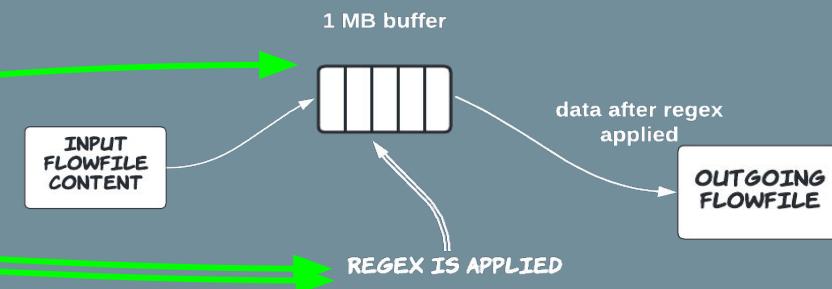
## Configure Processor

⚠ Invalid

Required field

Property	Value
Search Value	<input checked="" type="radio"/> (?s)(^.+\$)
Replacement Value	<input checked="" type="radio"/> \$1
Character Set	<input checked="" type="radio"/> UTF-8
Maximum Buffer Size	<input checked="" type="radio"/> 1 MB
Replacement Strategy	<input checked="" type="radio"/> Regex Replace
Evaluation Mode	<input checked="" type="radio"/> Line-by-Line
Line-by-Line Evaluation Mode	<input checked="" type="radio"/> All

**CANCEL** **APPLY**



# Splitting, Transformation, Aggregating

## Modify Data

	ReplaceText	ReplaceText 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

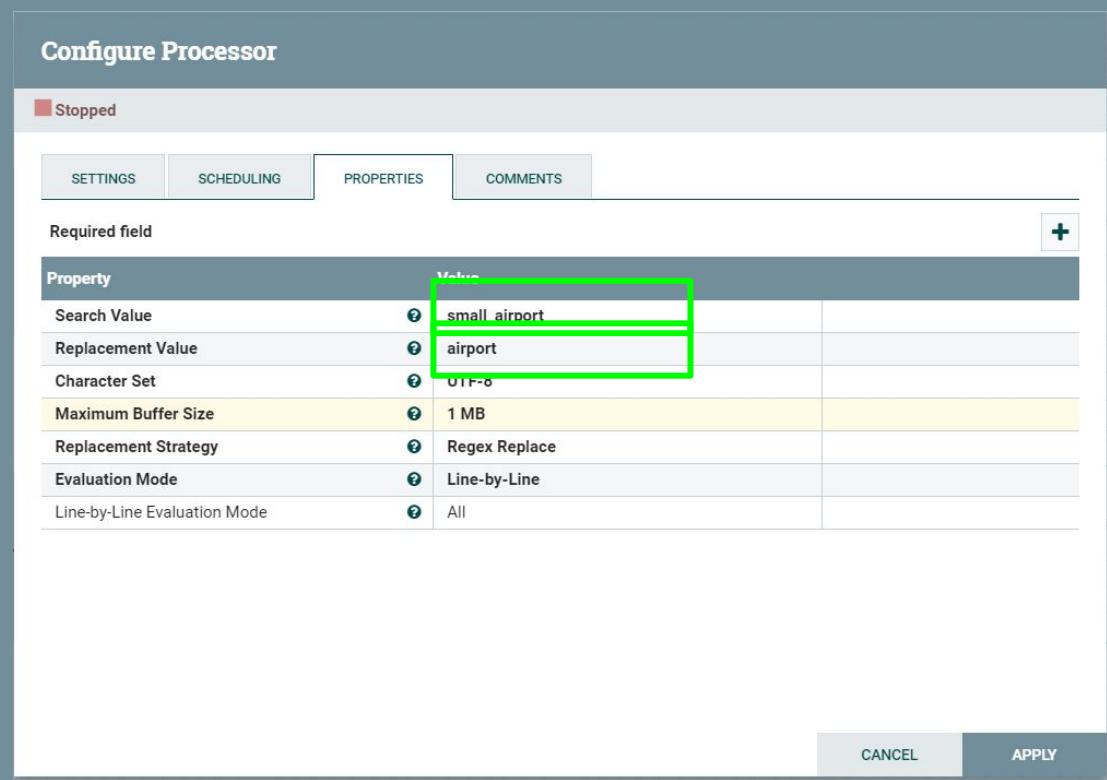
Stopped

SETTINGS SCHEDULING PROPERTIES COMMENTS

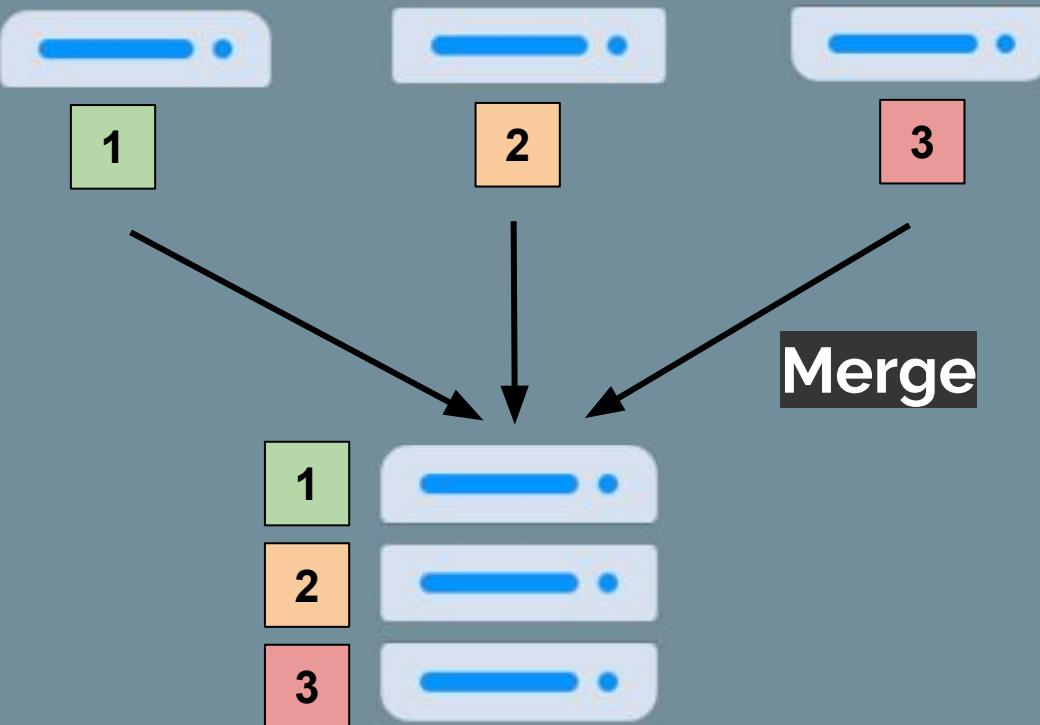
Required field

Property	Value
Search Value	small_airport
Replacement Value	airport
Character Set	UTF-8
Maximum Buffer Size	1 MB
Replacement Strategy	Regex Replace
Evaluation Mode	Line-by-Line
Line-by-Line Evaluation Mode	All

CANCEL APPLY



# Merging Data



# Splitting, Transformation, Aggregating

## Merge Data

	MergeContent	MergeContent 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

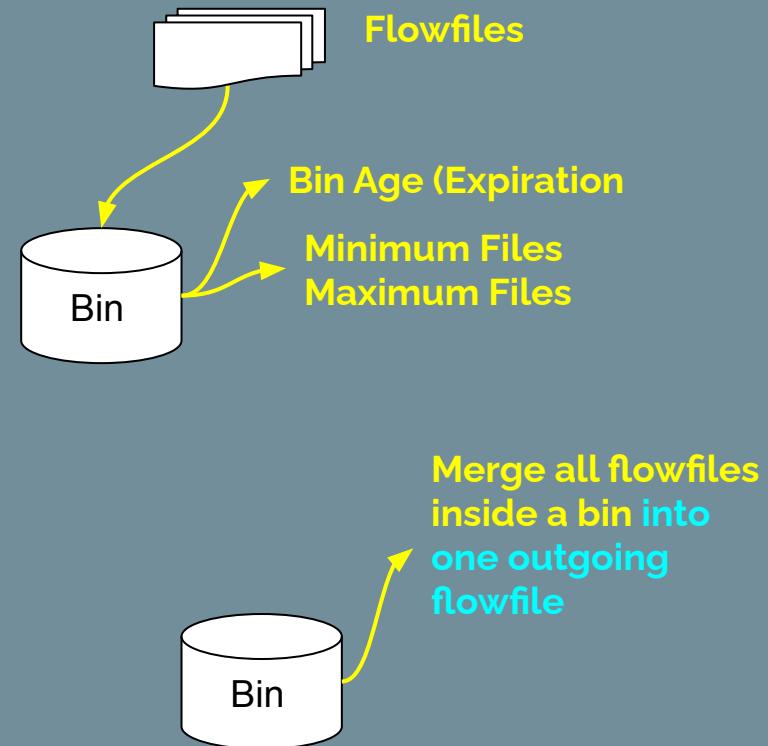
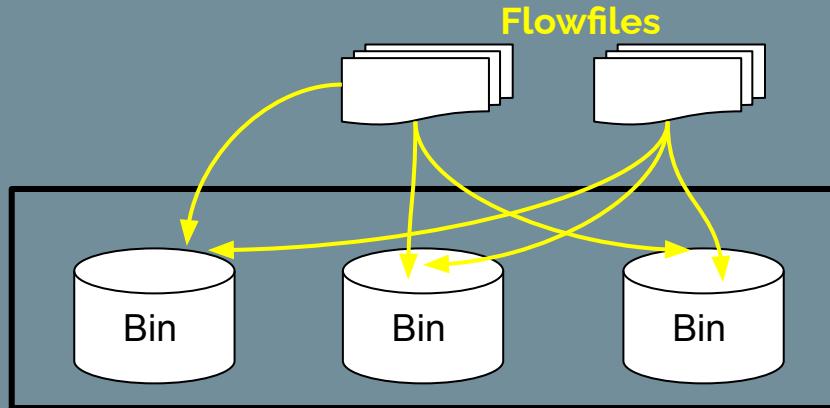
Required field

Property	Value
Merge Strategy	Bin-Packing Algorithm
Merge Format	Binary Concatenation
Attribute Strategy	Keep Only Common Attributes
Correlation Attribute Name	No value set
Minimum Number of Entries	1
Maximum Number of Entries	1000
Minimum Group Size	0 B
Maximum Group Size	No value set
Max Bin Age	No value set
Maximum number of Bins	5
Delimiter Strategy	Do Not Use Delimiters

CANCEL      APPLY

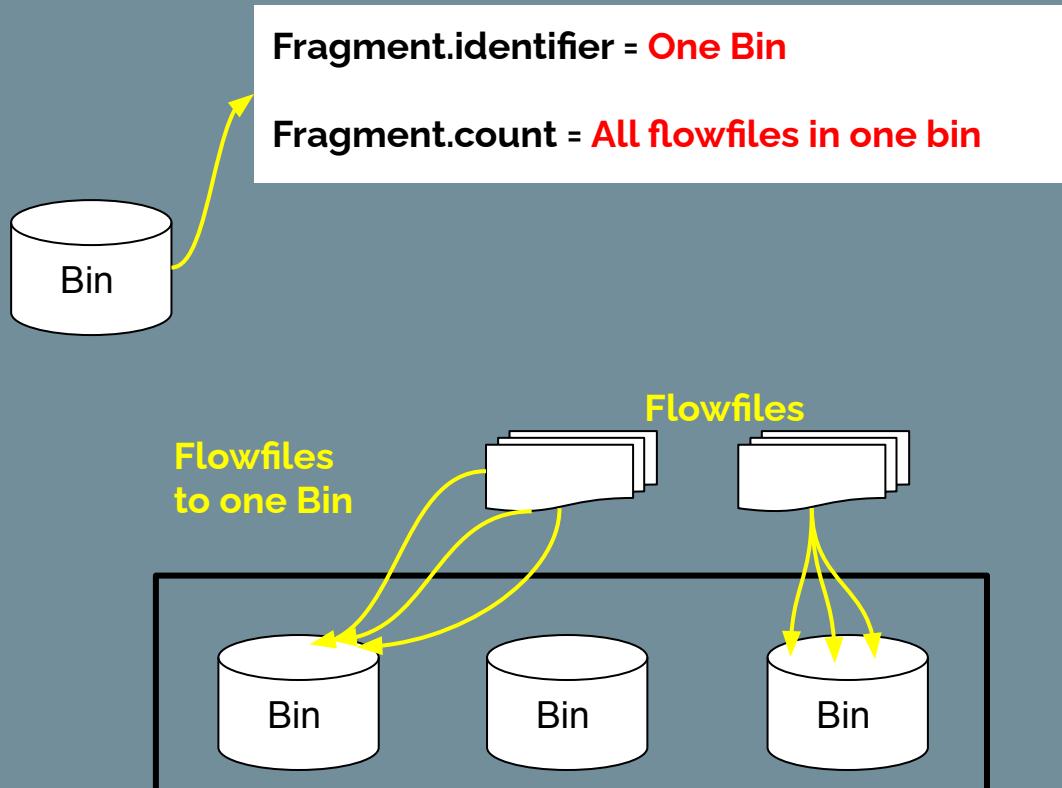
# Splitting, Transformation, Aggregating

## Merge Content Strategies



# Splitting, Transformation, Aggregating

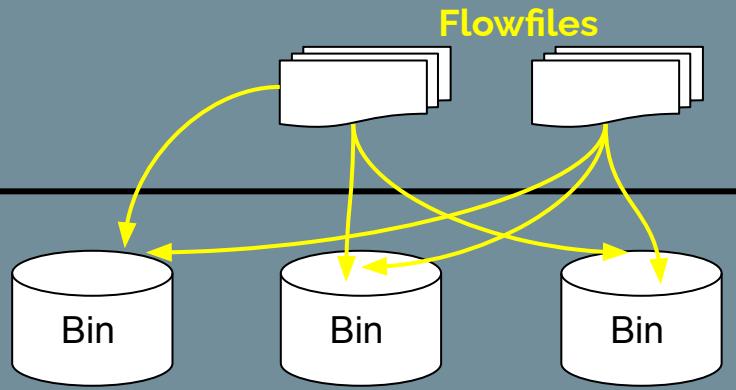
## Merge Content Strategies



# Splitting, Transformation, Aggregating

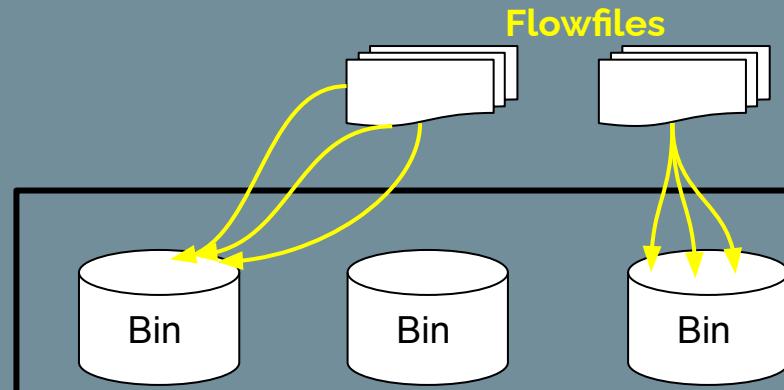
## Merge Content Strategy

### Bin Packing



Multiple Output Flowfiles

### Defragment

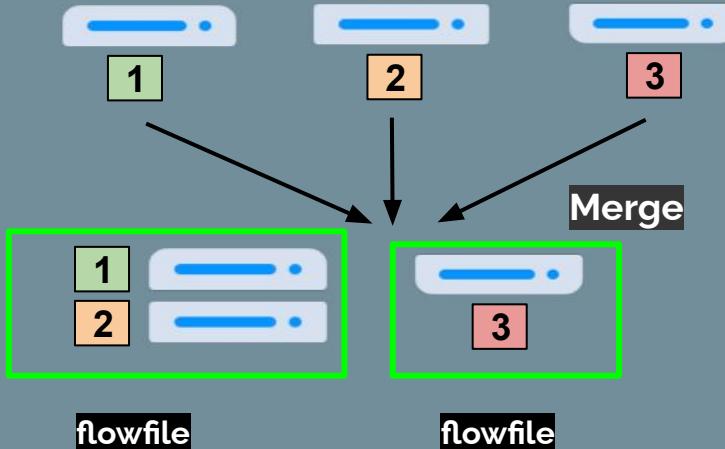


Single Output Flowfile for  
Each fragment identifier

# Splitting, Transformation, Aggregating

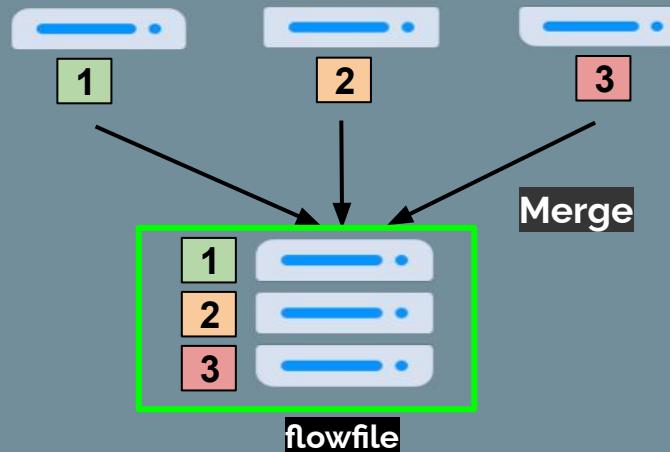
## Merge Content Strategy

### Bin Packing



Multiple Output Flowfiles

### Defragment



Single Output Flowfile for  
Each fragment identifier

# Splitting, Transformation, Aggregating

Splits flowfiles as  
fragments



	<b>SplitText</b>
	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

FlowFile

DETAILS ATTRIBUTES

Attribute Values

filename	fe24c22f-0ace-4b8f-8d0c-38fb93e35329
fragment.count	27
fragment.identifier	3a297ce9-07bc-40c8-b368-b3ef088aa91f
fragment.index	3
name	John
path	./
segment.original.filename	

OK

# Splitting, Transformation, Aggregating

## Merge Data

	MergeContent	MergeContent 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

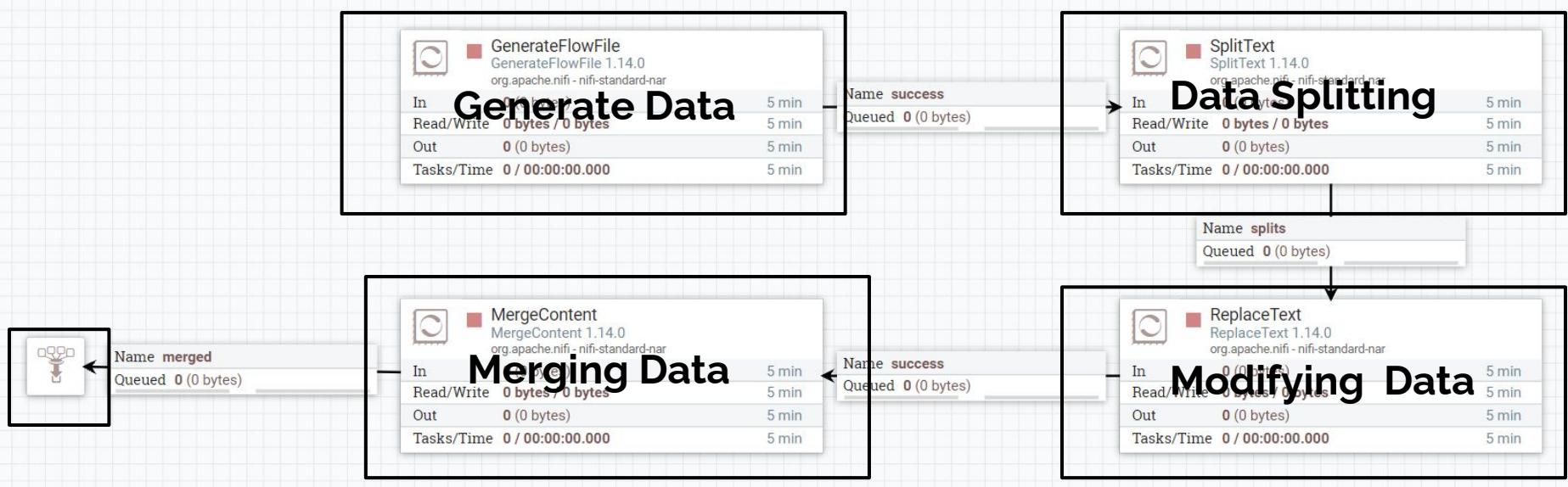
Required field

+ 

Property	Value
Merge Strategy	Bin-Packing Algorithm
Merge Format	Binary Concatenation
Attribute Strategy	Keep Only Common Attributes
Correlation Attribute Name	No value set
Minimum Number of Entries	1
Maximum Number of Entries	1000
Minimum Group Size	0 B
Maximum Group Size	No value set
Max Bin Age	No value set
Maximum number of Bins	5
Delimiter Strategy	Do Not Use Delimiters

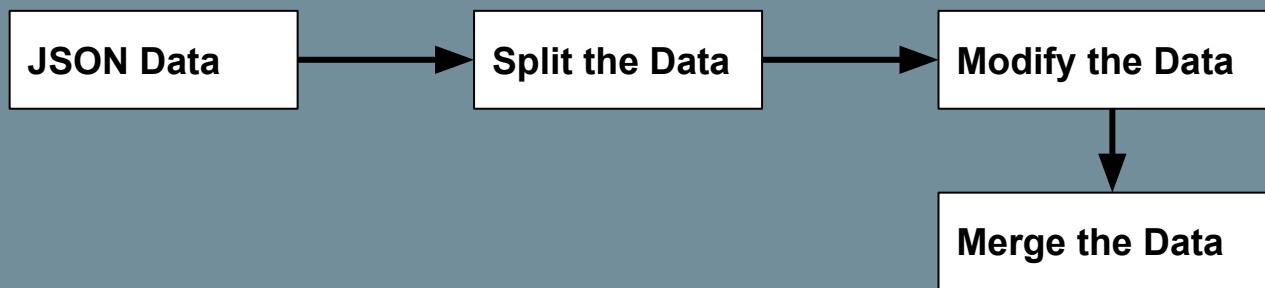
CANCEL APPLY

# Splitting, Transformation, Aggregating



# Splitting, Transformation, Aggregating (JSON)

Changing  
'Small\_airport'  
to  
'airport'



# Splitting, Transformation, Aggregating (JSON)

```
{  
  "id": 6523,  
  "ident": "00A",  
  "type": "heliport",  
  "name": "Total Rf Heliport",  
  "latitude_deg": 40.07080078125,  
  "longitude_deg": -74.93360137939453,  
  "elevation_ft": 11,  
  "continent": "NA",  
  "iso_country": "US",  
  "iso_region": "US-PA",  
  "municipality": "Bensalem",  
  "scheduled_service": "no",  
  "gps_code": "00A",  
  "iata_code": "",  
  "local_code": "00A",  
  "home_link": "",  
  "wikipedia_link": "",  
  "keywords": ""  
},  
[
```

# Splitting, Transformation, Aggregating (JSON)

## 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

	■ SplitJson	SplitJson 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	

## Merge Data

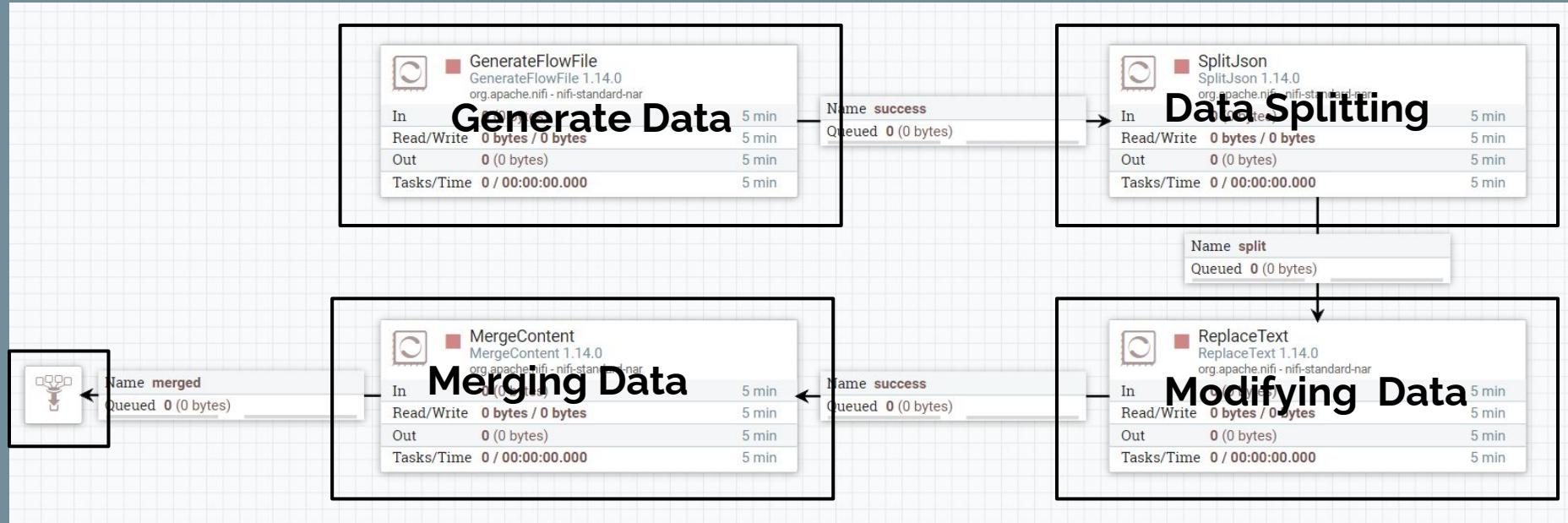
	■ MergeContent	MergeContent 1.14.0	org.apache.nifi - nifi-standard-nar
In	0 (0 bytes)	5 min	
Read/Write	0 bytes / 0 bytes	5 min	
Out	0 (0 bytes)	5 min	
Tasks/Time	0 / 00:00:00.000	5 min	

## Modify Data

	■ ReplaceText	ReplaceText 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	

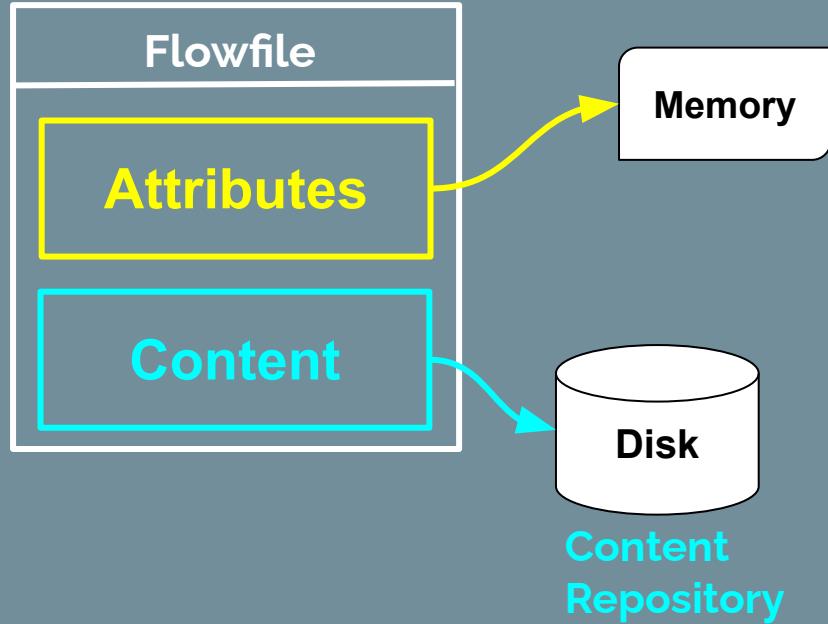
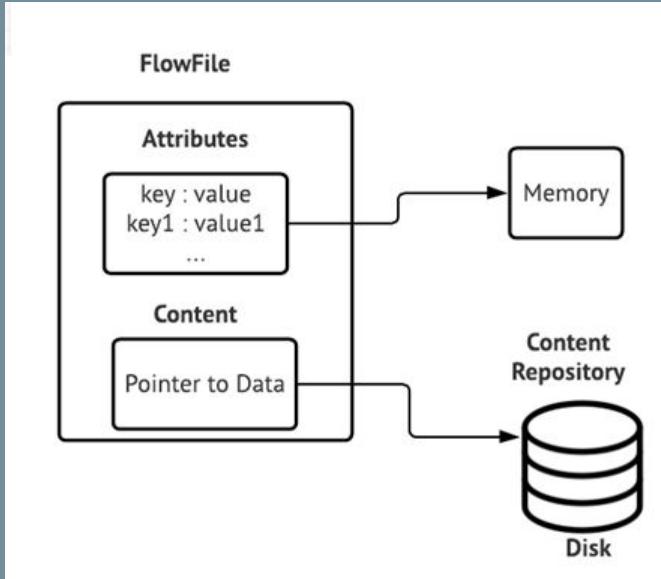


# Splitting, Transformation, Aggregating



# NiFi Expression Language

# – NiFi Expression Language



# – Why Expression Language

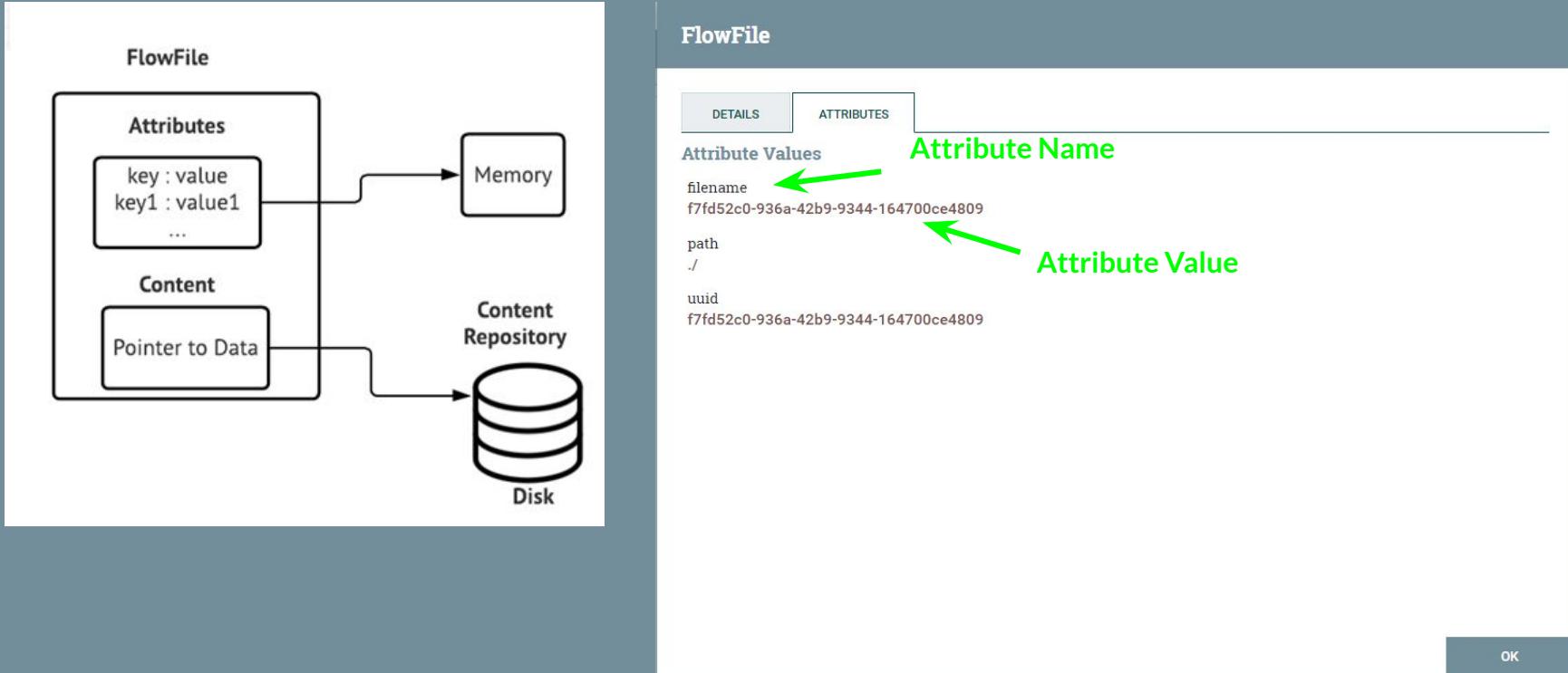
Language to access FlowFile Attributes

Provides wide variety of functions

In-memory processing, no disk access

Makes processing faster

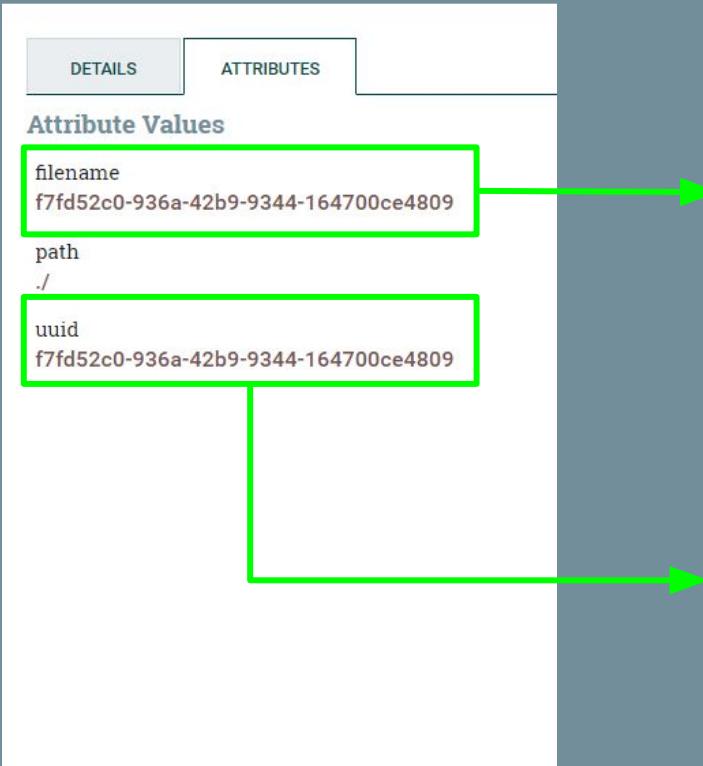
# – NiFi Expression Language



# Syntax of NiFi Expression language

# – NiFi Expression Language

FlowFile



Attribute Name  
\${ filename }  
Expression Syntax '\${ }'

\${ uuid }

General Syntax  
\${ **AttributeName** }

# – NiFi Expression Language

## General Syntax

`${AttributeName}`

`${ name }`



## FlowFile

DETAILS	ATTRIBUTES
<b>Attribute Values</b>	
filename	907061b8-a642-4ed9-97f0-d3ff214cbd46
name	John
path	./
uuid	907061b8-a642-4ed9-97f0-d3ff214cbd46

# – NiFi Expression Language

Expression	Result
<code> \${ name }</code>	<code>John</code>
<code> \${ name:toUpperCase() }</code>	<code>JOHN</code>
<code> \${ name:toLowerCase() }</code>	<code>john</code>
<code> \${ name:append('Doe') }</code>	<code>John Doe</code>
<code> \${ name:replace('John', 'Smith') }</code>	<code>Smith Doe</code>

Attribute Name

Function Name

General Syntax for using function

`${ AttributeName : functionName() }`

# Processors for directly playing with Attribute

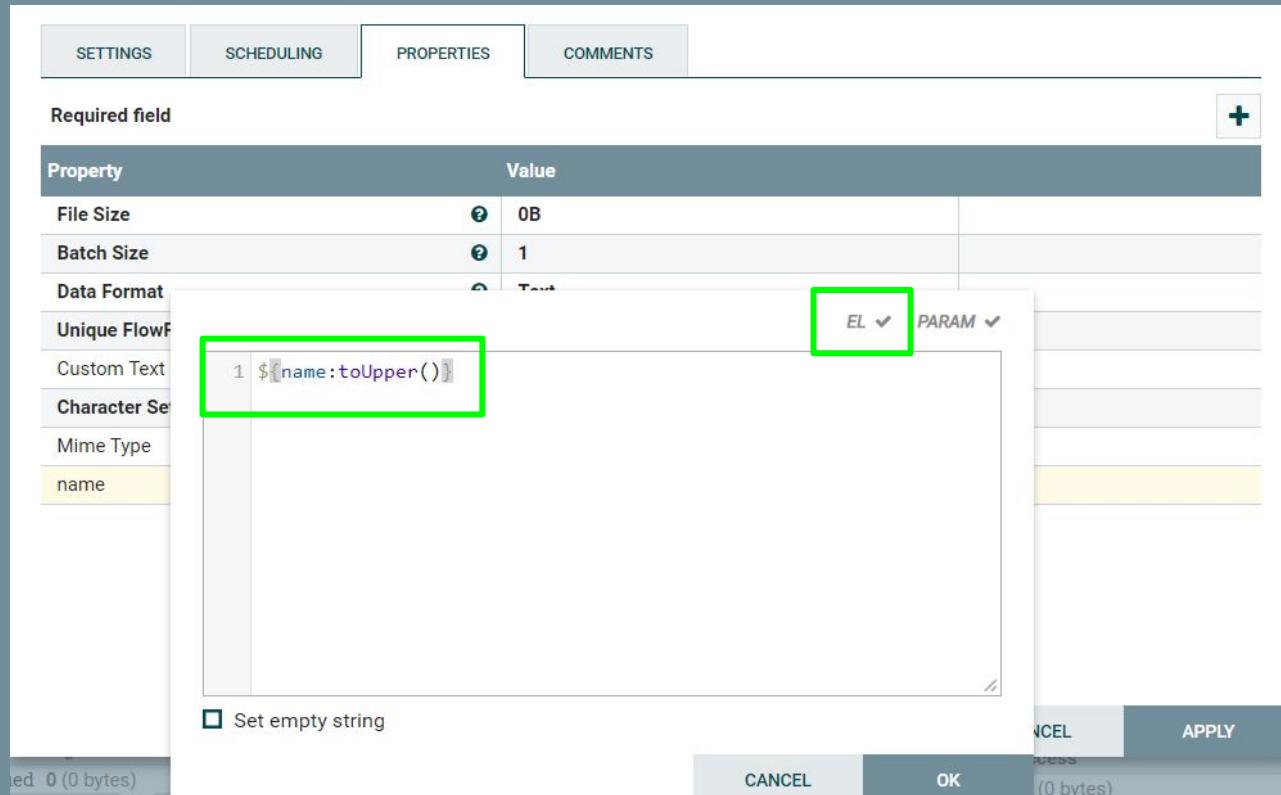
UpdateAttribute

RouteOnAttribute

AttributeToJson

LogAttribute

LookupAttribute



# - UpdateAttribute Processor Example

	■ 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

Stopped

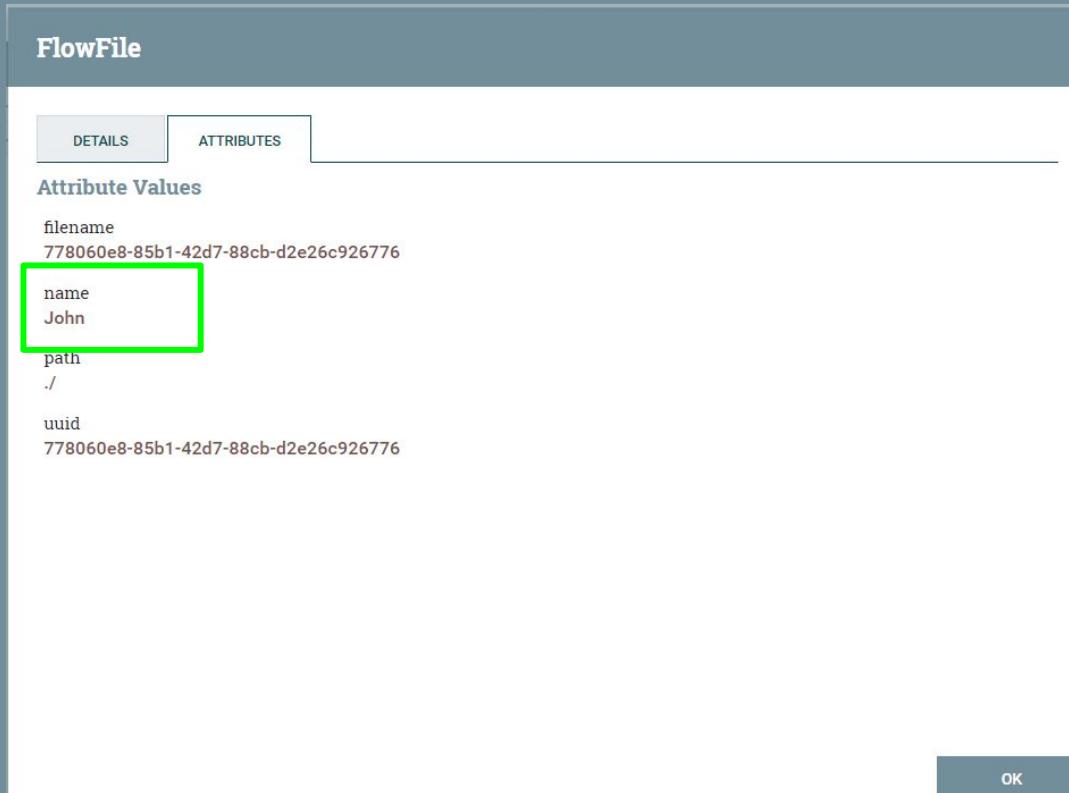
Required field

Property	Value
File Size	0B
Batch Size	1
Data Format	Text
Unique FlowFiles	false
Custom Text	No value set
Character Set	UTF-8
More Properties	...
<b>name</b>	<b>John</b>

**+ **

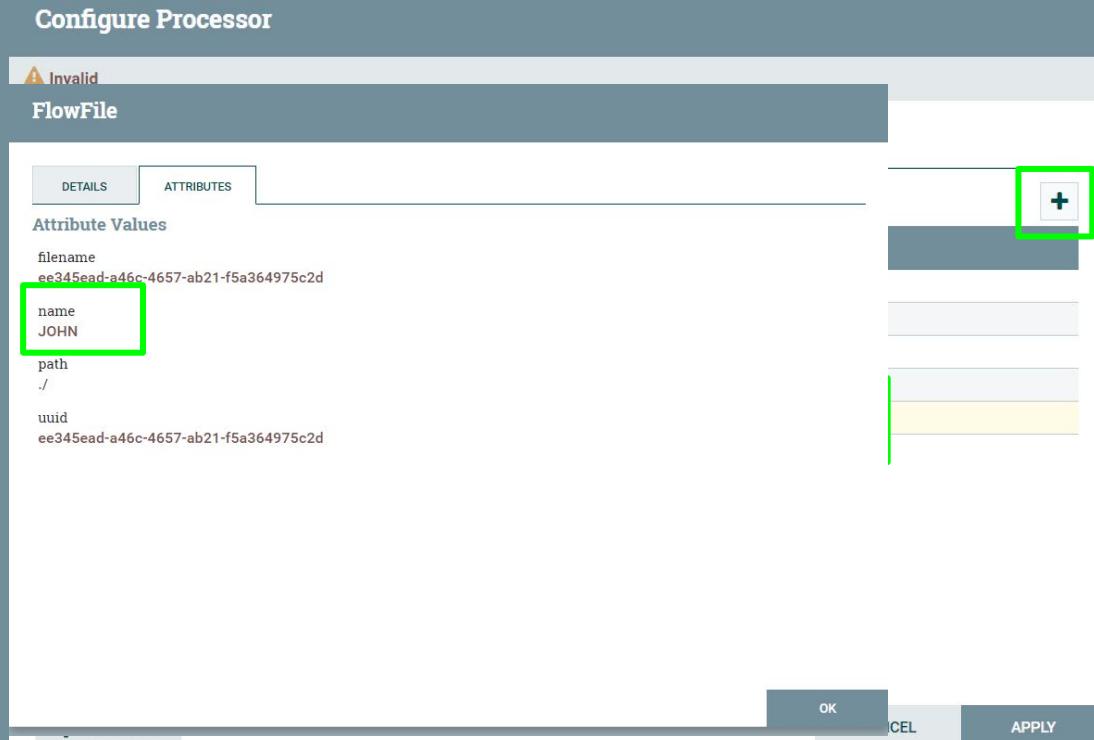
CANCEL    APPLY

# - UpdateAttribute Processor Example

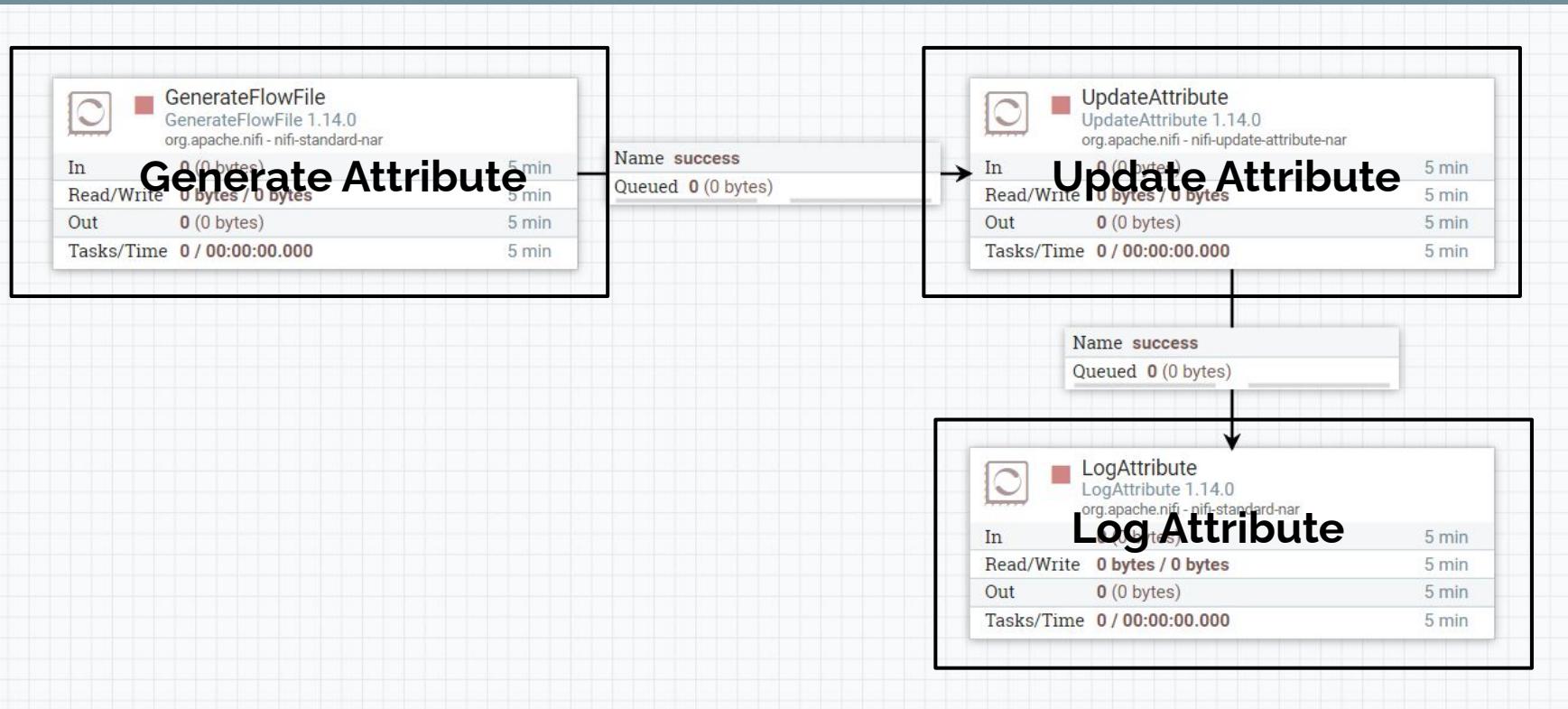


# - UpdateAttribute Processor Example

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



# - UpdateAttribute Processor Example



# Splitting, Transformation, Aggregating (With Attribute way)

## 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

## Merge Data

	■ MergeContent	MergeContent 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

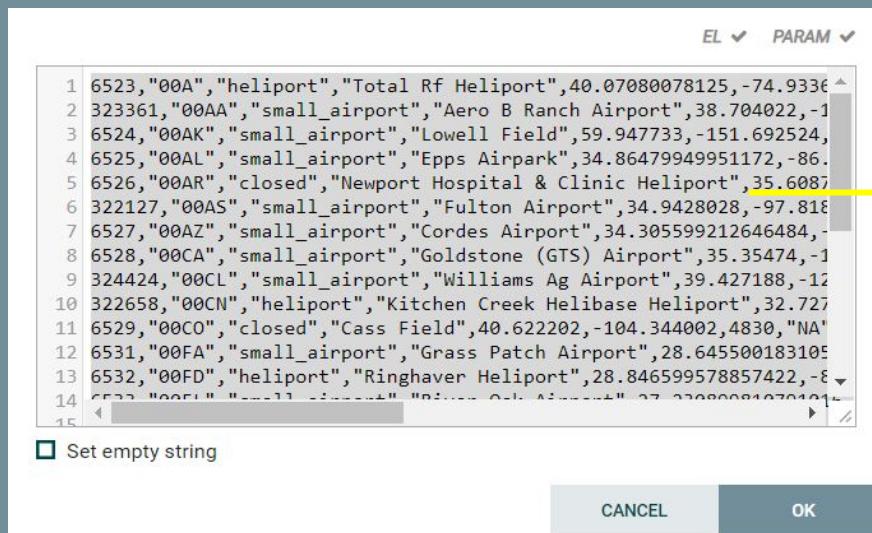
## Modify Data

	■ ReplaceText	ReplaceText 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



# Splitting, Transformation, Aggregating (With Attribute way)

## FlowFile Content



Perform the logic  
by loading  
content as  
attributes

# Splitting, Transformation, Aggregating (With Attribute way)

## 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

## Merge Data

	■ MergeContent	MergeContent 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

## Modify Data

	■ ReplaceText	ReplaceText 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



# Splitting, Transformation, Aggregating (With Attribute way)

## Modify Data

	■ ReplaceText	ReplaceText 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

Alternatively

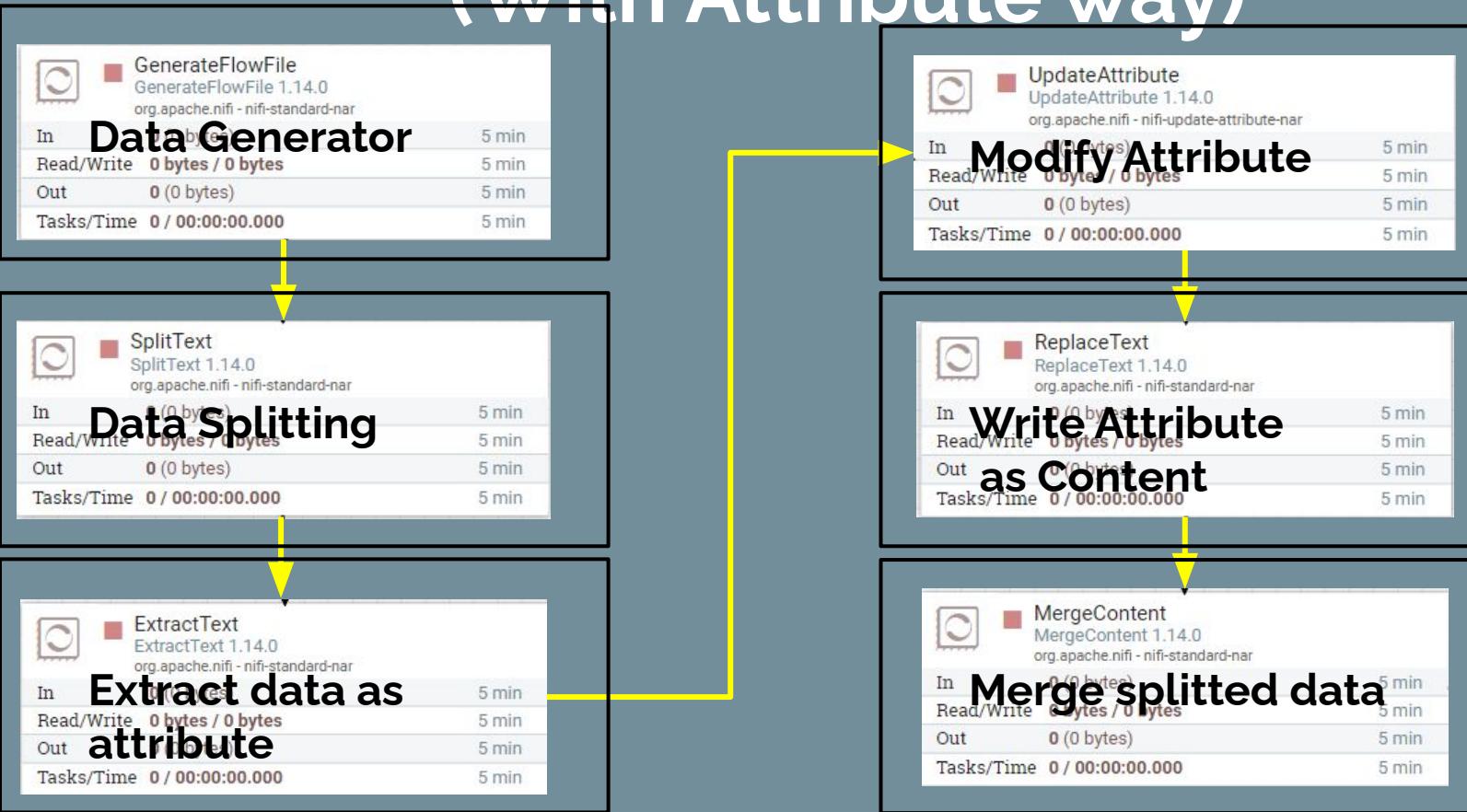


	■ ExtractText	ExtractText 1.14.0
org.apache.nifi - nifi-standard-nar		
In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

	■ 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

	■ ReplaceText	ReplaceText 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

# Splitting, Transformation, Aggregating (With Attribute way)



# Splitting, Transformation, Aggregating (With Attribute way)

## Extract data as attribute

	ExtractText	ExtractText 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

Stopped

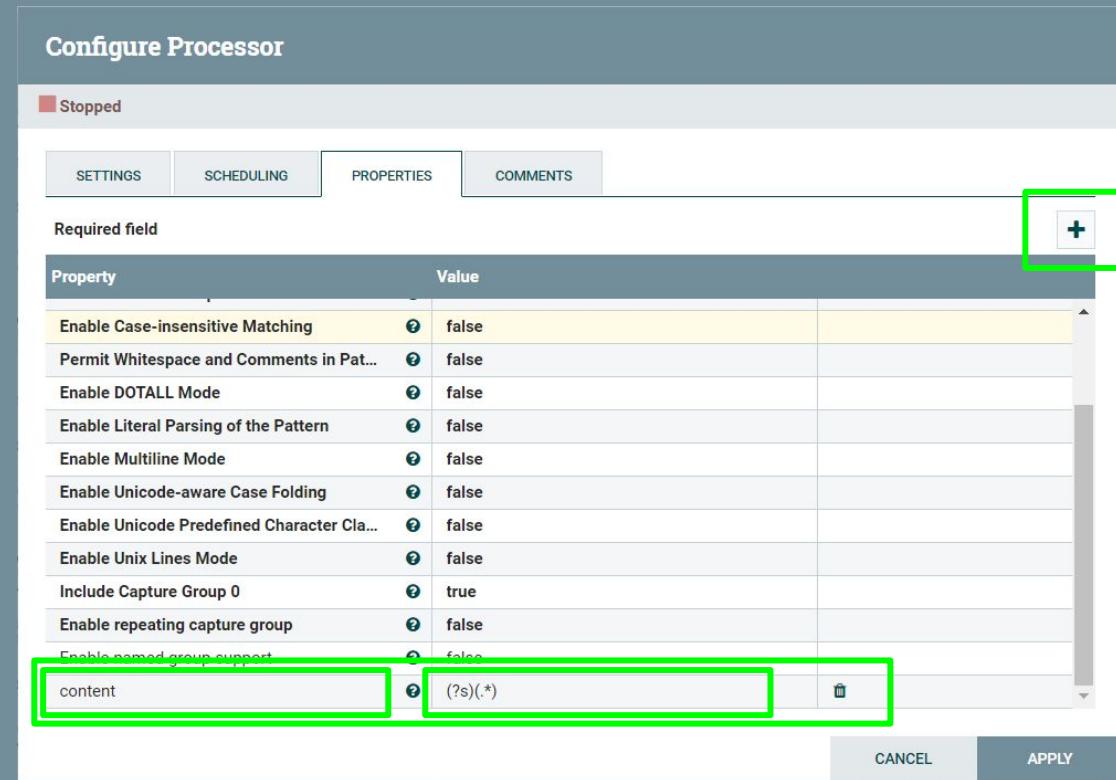
SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

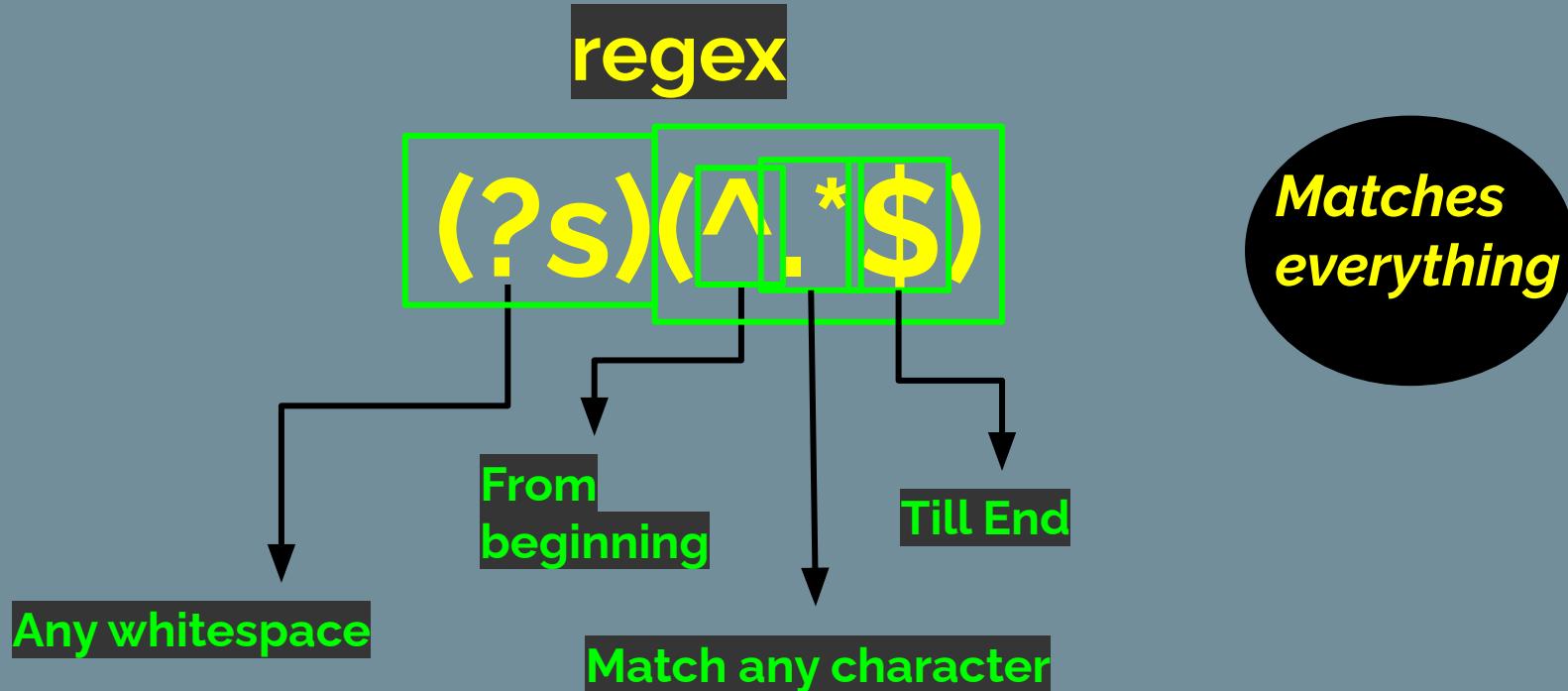
Property Value

Property	Value
Enable Case-insensitive Matching	false
Permit Whitespace and Comments in Pat...	false
Enable DOTALL Mode	false
Enable Literal Parsing of the Pattern	false
Enable Multiline Mode	false
Enable Unicode-aware Case Folding	false
Enable Unicode Predefined Character Cla...	false
Enable Unix Lines Mode	false
Include Capture Group 0	true
Enable repeating capture group	false
Enable named group support	false
content	(?s)(.*)

**CANCEL** **APPLY**



# Splitting, Transformation, Aggregating (With Attribute way)



# Splitting, Transformation, Aggregating (With Attribute way)

## Extract data as attribute

	<b>ExtractText</b>	ExtractText 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

Stopped

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Enable Case-insensitive Matching	false
Permit Whitespace and Comments in Pat...	false
Enable DOTALL Mode	false
Enable Literal Parsing of the Pattern	false
Enable Multiline Mode	false
Enable Unicode-aware Case Folding	false
Enable Unicode Predefined Character Cla...	false
Enable Unix Lines Mode	false
Include Capture Group 0	true
Enable repeating capture group	false
Enable named group support	false
content	(?s)(.*)

CANCEL APPLY

# Splitting, Transformation, Aggregating (With Attribute way)

## Modify attribute

	<b>UpdateAttribute</b> 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

Configure Processor

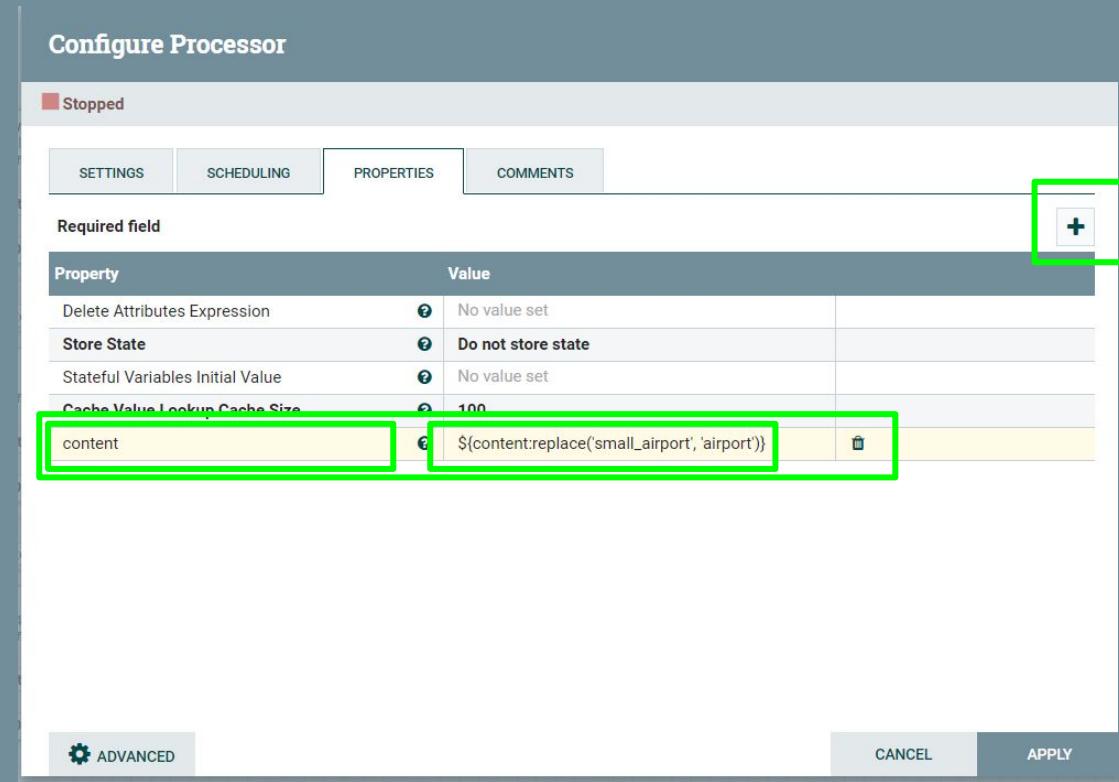
■ Stopped

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	100
content	<code>\$(content:replace('small_airport', 'airport'))</code>

**ADVANCED** CANCEL APPLY



# Splitting, Transformation, Aggregating (With Attribute way)

Write attribute as content

Processor Overview		
	ReplaceText	ReplaceText 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

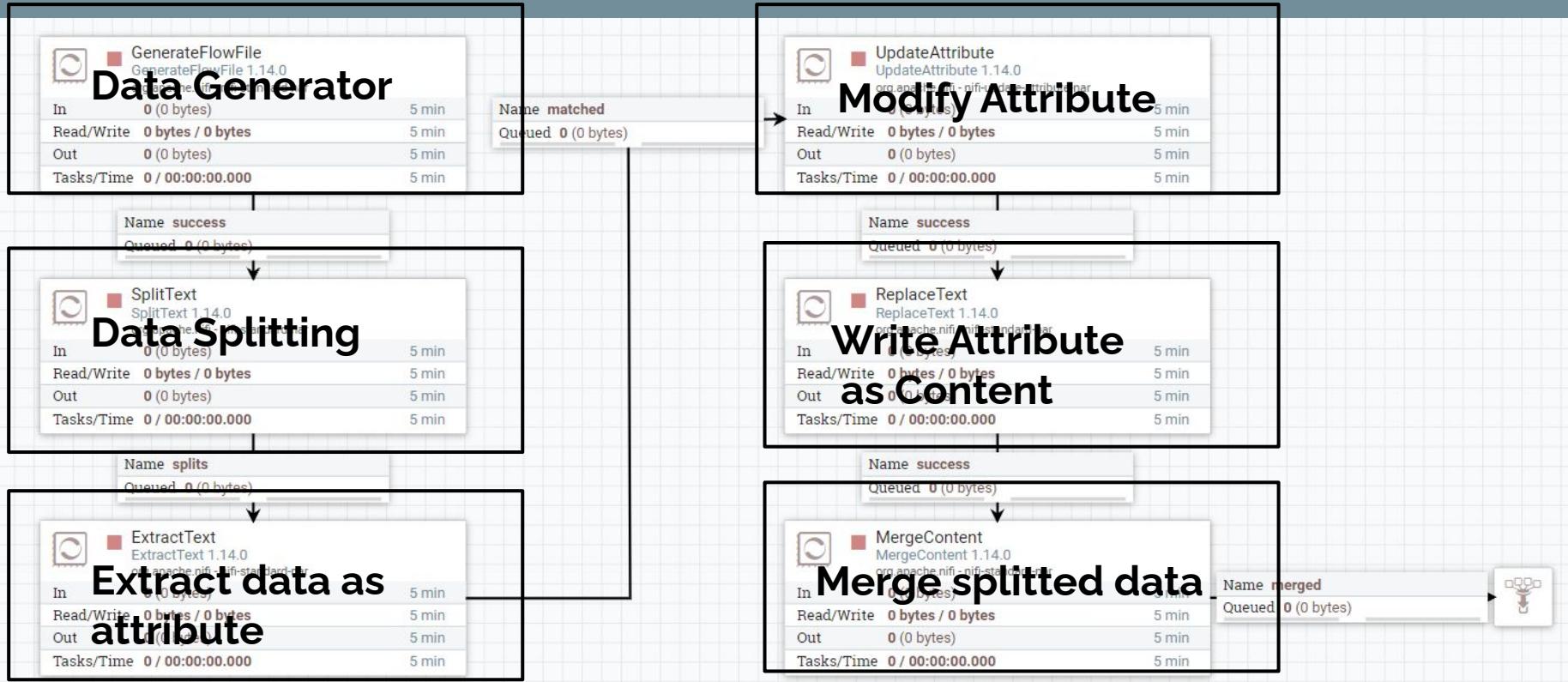
Stopped

Required field

Property	Value
Search Value	(?s)(^.*\$)
Replacement Value	\${content}
Character Set	UTF-8
Maximum Buffer Size	1 MB
Replacement Strategy	Always Replace
Evaluation Mode	Line-by-Line
Line-by-Line Evaluation Mode	All

CANCEL      APPLY

# Splitting, Transformation, Aggregating (With Attribute way)



# – NiFi Expression Language

Conditional Logic using expression language

`${ name }`

`John`

FlowFile

DETAILS

ATTRIBUTES

Attribute Values

filename  
907061b8-a642-4ed9-97f0-d3ff214cbd46

name  
John

path  
./

uuid  
907061b8-a642-4ed9-97f0-d3ff214cbd46

`${ name:equals('John') :ifElse('Found' , 'Not Found' ) }`

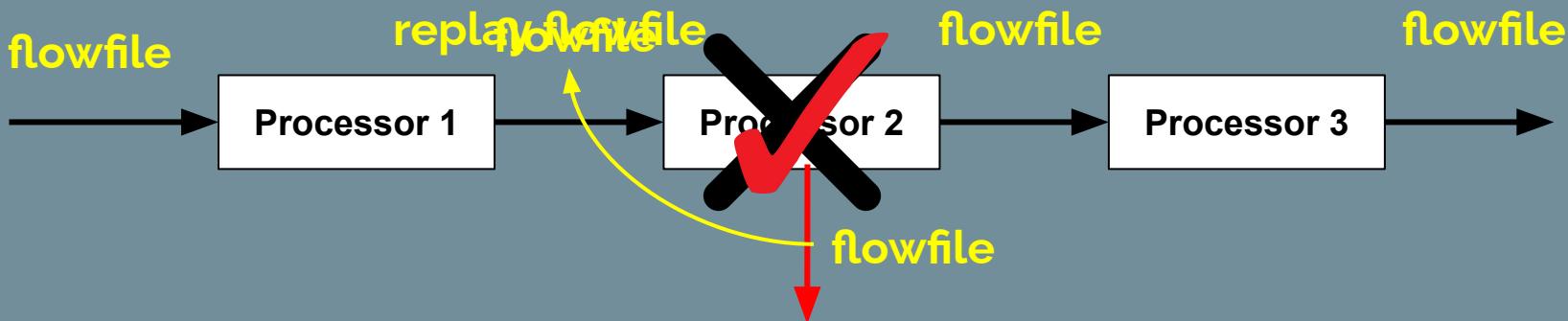
```
If(name == 'John') {  
    Found;  
} else {  
    Not Found;  
}
```

# Data Provenance

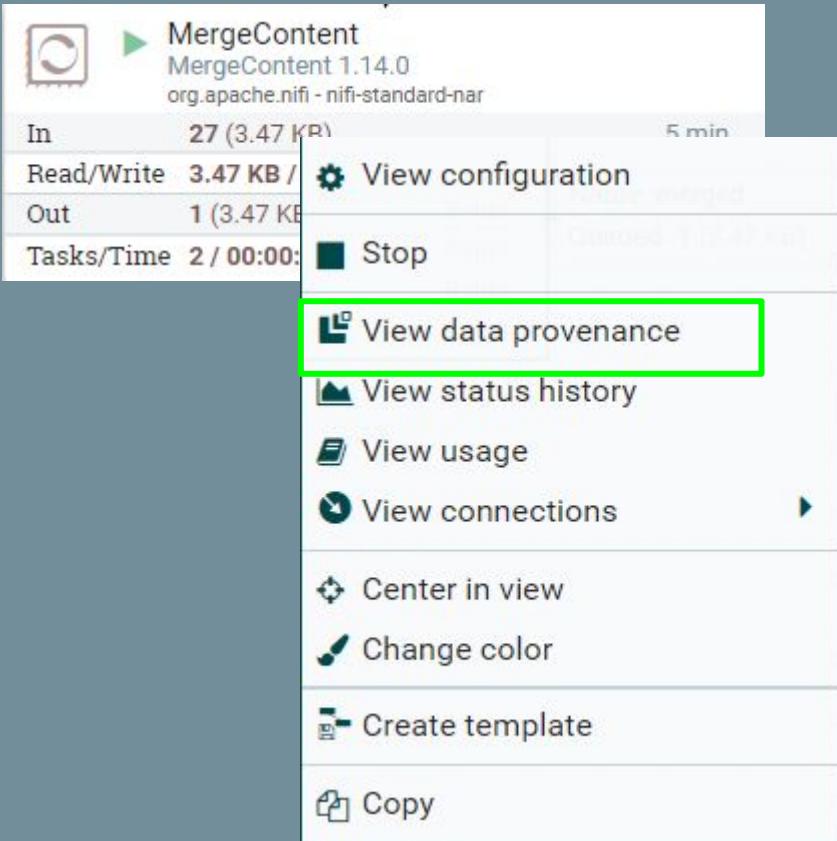
# - Data Provenance

NiFi logs and stores every bit of information in provenance repository

This repository host everything that system needs to replay the history (In other words, any flowfile can be replay from the beginning or from a specific point .



# - Data Provenance



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

In 27 (3.47 KB) 5 min

Read/Write 3.47 KB /

Out 1 (3.47 KB)

Tasks/Time 2 / 00:00:00 Stop

**View configuration**

**View data provenance** (highlighted)

**View status history**

**View usage**

**View connections**

**Center in view**

**Change color**

**Create template**

**Copy**

NiFi Data Provenance

Displaying 55 of 55  
Oldest event available: 10/16/2021 14:25:58 IST

Showing the events that match the specified query. Clear search

Date/Time	Type	FlowFile UUID	Size	Component Name	Component Type
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	05e91935-8d5f-4f14-e89a-2cf2eece90...	138 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	68547cc-338f-4819-a77a-b6e29fb04...	139 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	11a54db4-ac99-4053-9841-7915cb72...	137 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	ed4bc097-1161-4-e44b-02b7-0903778b...	110 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	15a046e2-8e1-4b4c-2b27-786dcda4a...	144 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	7e21b5e4-2974-42d8-b596-776bb696...	121 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	6f81e55-39b8-43d9-b890-dc342a42...	126 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	7239f346-db5f-499e-a1ab-040c593...	134 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	44de28f537b-4a24-a8d0-c7ad5cd...	118 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	80ec0254-94c3-4760-bea3-9f01911c...	127 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	f918493-3d3b-4060-b54b-c6596fc1...	130 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	741df93-8383-4a8f-9eb3-d35fc4085...	120 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	3459f84c-10c2-46a9-9d3a-30383460...	124 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	7ff7d921-1df5-4507-a1ea-846dd69e...	130 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	7ac9046f-1781-4ae8-892b-c249120d...	143 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	4616509b-268d-4522-8644-6198cc1...	137 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	01d0f546-0474-4650-873e-0c95b537...	142 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	f5ed616-714a-4758-a325-64d2d669f...	138 bytes	MergeContent	MergeContent

Last updated: 23:31:04 IST

# – Data Provenance

NiFi Data Provenance

Displaying 55 of 55  
Oldest event available: 10/16/2021 14:25:58 IST

Showing the events that match the specified query. Clear search

Date/Time	Type	FlowFile Uuid	Size	Component Name	Component Type	Actions
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	05e01925-8d5f-414a-89a9-ef20ee0a00	128 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	58f5476c-f388-48f9-976a-b6e29eb0a	139 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	11a54cb4-ac99-4053-9b41-7915cb72...	137 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	ed4bac9f-1161-4ca4-bbf2-090c378fb...	110 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	15a04e62-8ec1-4b4c-b2e7-86dcad4a...	144 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	7e2185e4-292f-42d8-95f6-776bb969...	121 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	6f81ec55-398c-43a9-b890-da342a42f...	126 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	72359346-db5f-498e-a1ab-04bc5693...	134 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	44de628f-537b-4a24-ad80-cf7ad6cdf...	118 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	80ecc25d-94c3-4760-bea3-a901911c...	127 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	f91845f3-3cd8-40b0-b54b-c6696fcf5...	130 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	741d1df9-83b3-4a8f-beb3-d35fc4085...	120 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	3459f84c-10c2-4e6a-8c03-e30834b0f...	124 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	7ff7daf2-1df5-45d7-a1ea-84ded68ea...	130 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	7ac9046f-1781-4ae8-892b-c24912cd...	143 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	461650a8-2e8d-4d32-8644-f619bcc1...	137 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	01d0f546-d474-4650-873e-0c9b5637...	142 bytes	MergeContent	MergeContent	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	f5ed5e16-7f4a-4758-a325-6fd2d669f...	138 bytes	MergeContent	MergeContent	

Last updated: 23:31:04 IST

# - Data Provenance

S.No.	Field Name	Description
1	Date/Time	Date and time of event.
2	Type	Type of Event like 'CREATE'.
3	FlowFileUuid	UUID of the flowfile on which the event is performed.
4	Size	Size of the flowfile.
5	Component Name	Name of the component which performed the event.
6	Component Type	Type of the component.
7	Show lineage	Last column has the show lineage icon, which is used to see the flowfile lineage as shown in the below image.

# - Data Provenance

NiFi Data Provenance

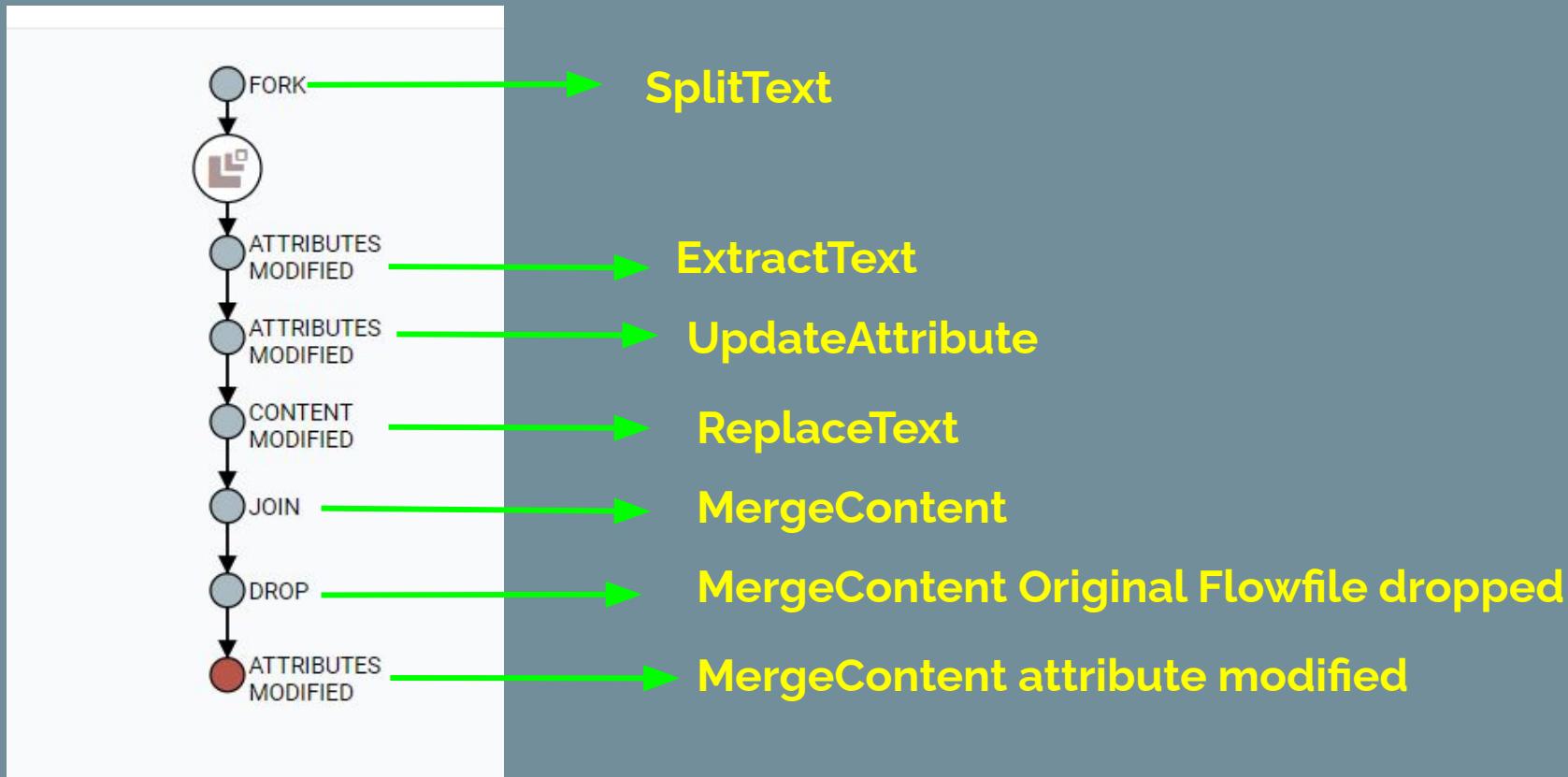
Showing the events that match the specified query. Clear search

Date/Time	Type	FlowFile Uuid	Size	Component Name	Component Type
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	05e91935-8d5f-414e-892c-22eeef90...	138 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	68f5476c-f384-8499-976a-b6e296b0...	139 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	11a54cb4-ac99-4053-9b41-7915ab72...	127 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	ed4ba9f1-1161-4e44-b05f-090378b...	110 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	15a04602-8e61-4bdc-b267-86dcda4...	144 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	7e2185e4-2924-42ab-9596-776bb699...	121 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	6f81e55-396c-43a9-8990-d342d42...	126 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	72359346-db5f-499e-a1a0-04bc593...	134 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	44de528f-537b-4a24-ad80-c7ad5cd...	118 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	80ec0254-9463-4760-bea3-001911c...	127 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	f918453-3cb0-40b0-b54b-c699fcf...	130 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	7414df9-8363-4a8f-beb3-d35fe405...	120 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	345984-10c2-4e6a-8033-e303840f...	124 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	7ff76a2-1df5-457a-81ea-84ded9e...	130 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	7a9046f1-7814-e8-892b-c24912cd...	143 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	46150a8-2e64-4d32-8644-61198cc...	137 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	01d0f548-d474-4650-873a-0c9b5637...	142 bytes	MergeContent	MergeContent
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	f5ed5e16-714a-4758-e325-6fd2d669f...	138 bytes	MergeContent	MergeContent

Last updated: 23:31:04 IST

The image shows a screenshot of the NiFi Data Provenance interface on the left and a process flow diagram on the right. The provenance interface displays a table of 55 events, with the last event being from 11/09/2021 at 23:29:17.487 IST. A green arrow points from the 'MergeContent' column header in the table to the 'JOIN' node in the process flow diagram. The process flow consists of the following steps: a 'FORK' node at the top, followed by three 'ATTRIBUTES MODIFIED' nodes, a 'CONTENT MODIFIED' node, a 'JOIN' node, a 'DROP' node, and finally another 'ATTRIBUTES MODIFIED' node at the bottom. All nodes are enclosed in a large green rectangular frame.

# – Data Provenance



# - Data Provenance

The timeline shows the following sequence of events:

- FORK
- ATTRIBUTES MODIFIED
- ATTRIBUTES MODIFIED
- CONTENT MODIFIED
- JOIN
- DROP
- ATTRIBUTES MODIFIED

A context menu is open at the FORK step, containing the following options:

- View details** (highlighted with a green border)
- Find parents
- Expand
- Collapse

**Provenance Event**

DETAILS	ATTRIBUTES	CONTENT
Time 11/09/2021 23:28:59.023 IST		
Event Duration 00:00:00.035		
Lineage Duration 00:00:01.187		
Type FORK		
FlowFile Uuid fbda3019-8c3e-438d-bad1-d124587afb40		
File Size 3.54 KB		
Component Id 000644d6-ad44-3c40-d2f3-e1da31950054		
Component Name SplitText		
Component Type		

Two sections of the event details are highlighted with green boxes:

- Parent FlowFiles (1)**  
fbda3019-8c3e-438d-bad1-d124587afb40
- Child FlowFiles (27)**  
A long list of 27 child flowfile UUIDs, starting with:  
01d0f546-d474-4650-873e-0c9b5637ac7c  
05e91935-8d5f-414e-89a2-cf2ece90325  
11a54cb4-ac99-4053-9b41-7915cb727554  
15a04e62-8ec1-4b4c-b2a7-86dcad4a11e1  
17bb70a6-0a6c-a553-a264-8303482ca5b  
1b5807e5-c3a4-4df0-a014-9be0e25d1190  
3459f84c-10c2-4e6a-8c03-30834ab0f725  
44de628f-537b-4a24-ad80-cf7ad6cdcf3a32  
461650a8-2e8d-4d32-8644-f6198cc12ec1  
68f5476c-f388-48f9-976a-b6e29eb0ad16  
6f81ec55-398c-43a9-b890-da342a421707  
72359346-db5f-498e-a1ab-04bc56936193  
741d1df9-83b3-4eef-beb3-d35fc4085ca8  
7ac9046f-1781-4e48-892b-c24912cd97a0  
7e2185e4-292f-42d8-95f6-776bb969c45f  
7f7da1f2-1df5-45d7-a1ea-84ded68ea90c  
80ecc25d-94c3-4760-bea3-a901911c536f

OK

# - Data Provenance

## Provenance Event

DETAILS ATTRIBUTES CONTENT

Container default

Section 43

Identifier 1636457213148-43

Offset 912178

Size 159 bytes

**REPLAY**

Connection Id 902009fc-bd57-3441-3d20-7b525d8627a8

**REPLAY**

OK

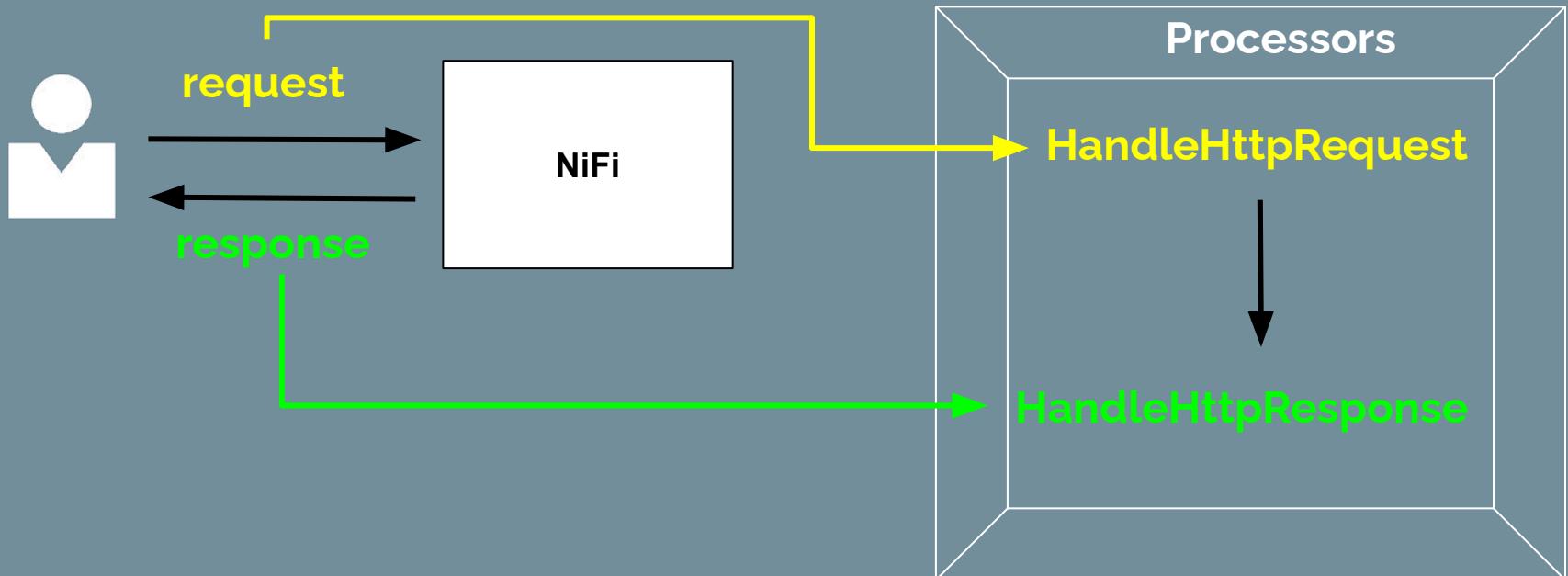
NiFi Data Provenance							
Filter		Showing the events that match the specified query. Clear search					
Time	Type	FlowFile Uuid	Size	Component Name	Component Type	Actions	
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	05e91935-8d5f-414e-89a2-cf2eece0...	138 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	68f5476c-f388-48f9-976a-b5e29eb0...	139 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	11a5464c-a99-4053-9e41-7915b072...	137 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	e4d8ac9f1161-4ca4-b4b2-990f37fb...	110 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	15d04e62-8ec1-4b4b-b2e7-88dcad4a...	144 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	7e218564-292f-4269-95f6-7768699...	121 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	6fb1ec55-39bc-43a9-b690-d3a42a42...	126 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	7239346-d5-49fe-e1ab-94de5693...	134 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	4466e28f5325-4a24-a480-774d60cf...	118 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	80ec225d-94c3-4760-be3a-9901911c...	127 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	919485f3-3c0b-40b0-a54b-c6696f0f...	130 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	741d10f9-82b3-4a8f-beb3-435c40f5...	120 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	3459f94c-10c2-4e65-8c03-c035340f...	124 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	7ff7da2f-21d5-45d7-a1ea-84de656e...	130 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	7ac9046f-1781-4ae8-897b-24912cd...	143 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	461650a8-2e8d-4d52-b244-616198c1...	137 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	01d0f546-d474-4650-873e-0c9b5637...	142 bytes	MergeContent	MergeContent		
11/09/2021 23:29:17.487 IST	ATTRIBUTES_MODIFIED	f5ed5e16-7f4a-4759-a325-6fd2dd669f...	138 bytes	MergeContent	MergeContent		

Last updated: 23:31:04 IST

# Creating Web Services using Processors

# Creating Web Services using Processors

Run a web server listening for requests from client



# Creating Web Services using Processors

To handle request

	<b>HandleHttpRequest</b>	HandleHttpRequest 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
Listening Port	80
Hostname	No value set
SSL Context Service	No value set
HTTP Context Map	No value set
Allowed Paths	No value set
Default URI Character Set	UTF-8
Allow GET	true
Allow POST	true
Allow PUT	true
Allow DELETE	true
Allow HEAD	false
Allow OPTIONS	false

CANCEL APPLY

# Creating Web Services using Processors

To handle response

	<b>HandleHttpResponse</b>	HandleHttpResponse 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

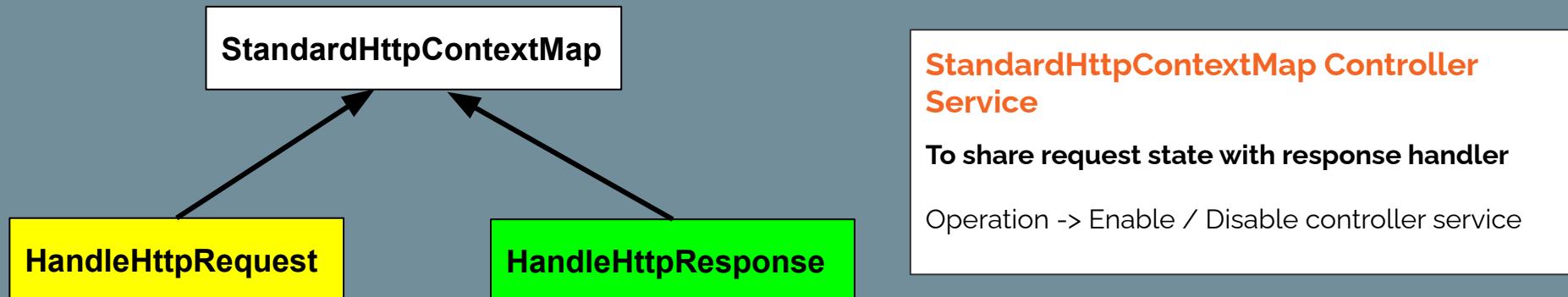
Property	Value
HTTP Status Code	No value set
HTTP Context Map	No value set
Attributes to add to the HTTP Response (Regex)	No value set

+ CANCEL APPLY

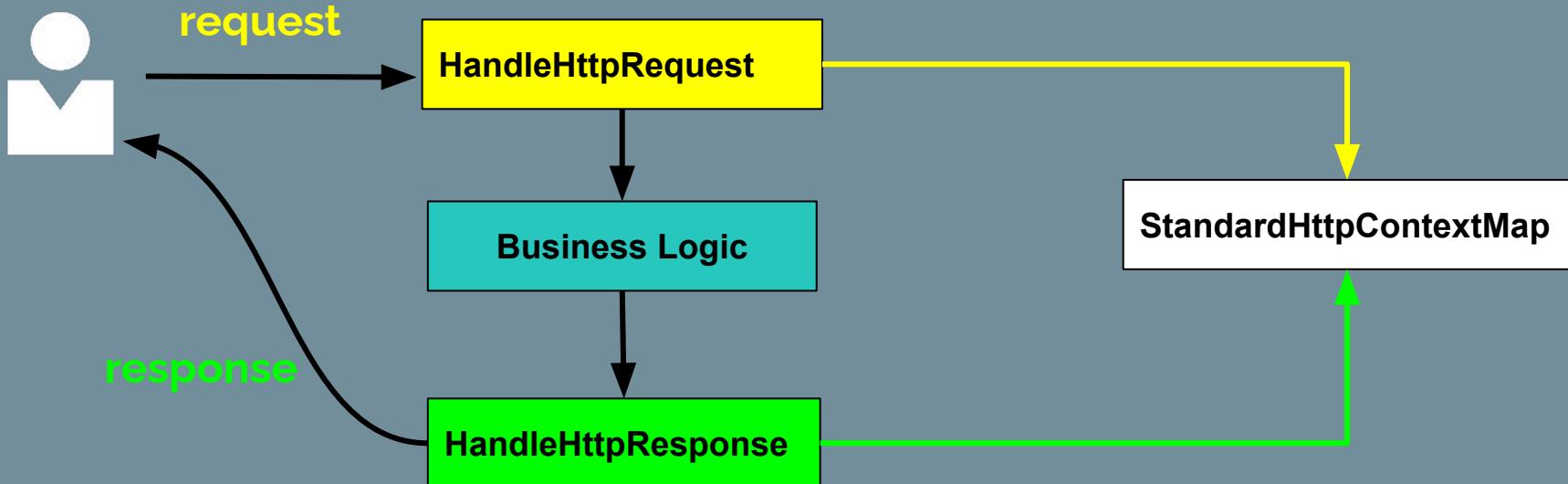
# Creating Web Services using Processors

HTTP Status Code	?	No value set
HTTP Context Map	?	No value set

StandardHttpContextMap Controller Service is used for sharing requests details to response handler



# Creating Web Services using Processors



# Creating Web Services using Processors

To share state

## Add Controller Service

Requires Controller Service  
HttpContextMap 1.14.0 from org.apache.nifi - nifi-standard-services-api-nar

Compatible Controller Services

- StandardHttpContextMap 1.14.0

Controller Service Name

- StandardHttpContextMap

Bundle

- org.apache.nifi - nifi-http-context-map-nar

Tags

- request, response, http

Description

CANCEL CREATE

### Configure Controller Service

SETTINGS PROPERTIES COMMENTS

Required field

Property	Value
Maximum Outstanding Requests	5000
Request Expiration	1 min

+ CANCEL APPLY

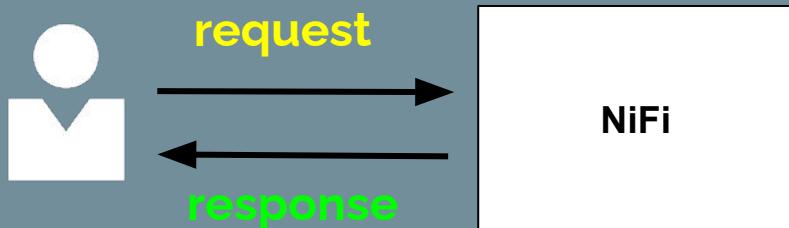
# Creating Web Services using Processors

To handle request

	<b>HandleHttpRequest</b> HandleHttpRequest 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

To handle response

	<b>HandleHttpResponse</b> HandleHttpResponse 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



# Creating Web Services using Processors

To handle request

	<b>HandleHttpRequest</b> HandleHttpRequest 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

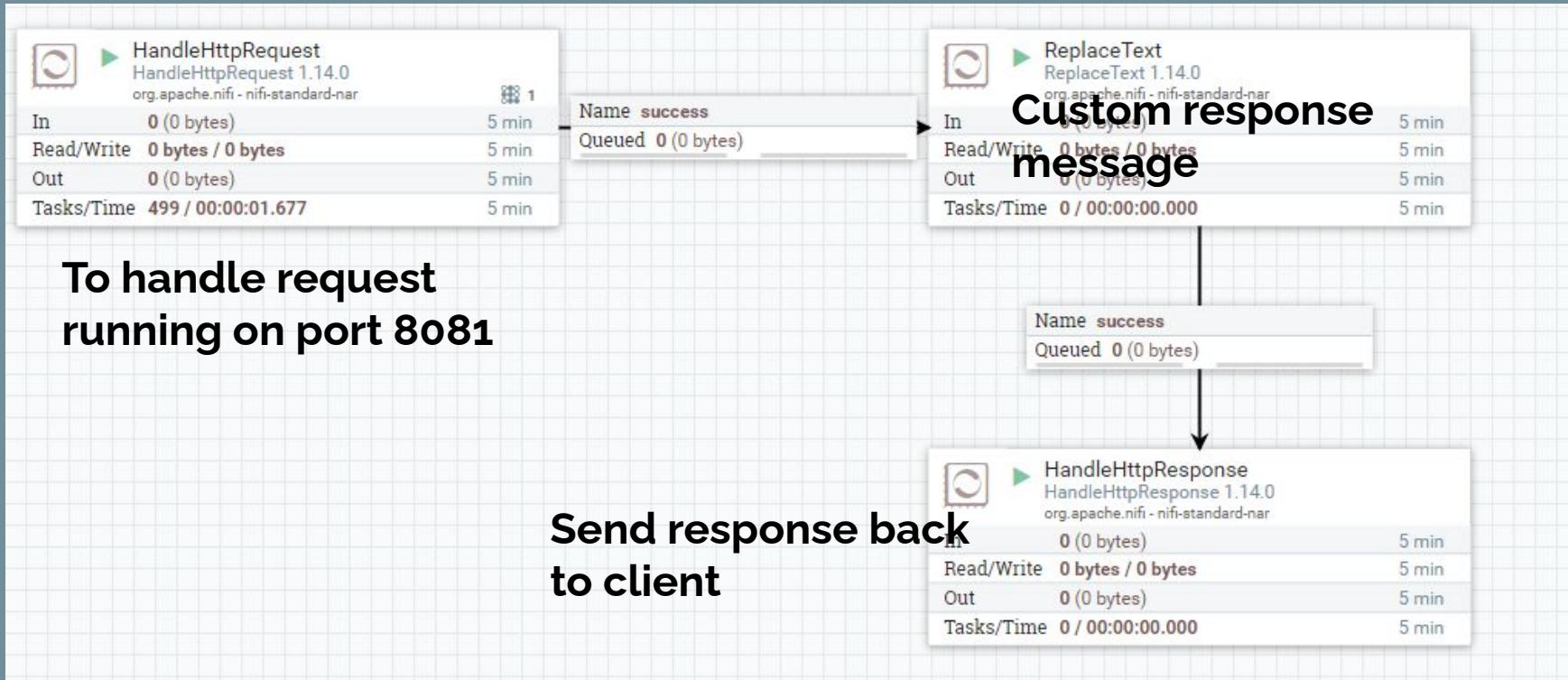
Custom Response

	<b>ReplaceText</b> ReplaceText 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

To handle response

	<b>HandleHttpResponse</b> HandleHttpResponse 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

# Creating Web Services using Processors



# Creating Web Services using Processors

The screenshot shows a POSTMAN interface with the following details:

- URL: `http://localhost:8081`
- Method: `GET`
- Request URL: `http://localhost:8081` (highlighted with a green box)
- Buttons: `Save`, `Send` (highlighted with a blue box)
- Tab navigation: Params (selected), Authorization, Headers (6), Body, Pre-request Script, Tests, Settings, Cookies
- Query Params table:

KEY	VALUE	DESCRIPTION	...	Bulk Edit
Key	Value	Description		
- Response status: Status: 200 OK, Time: 4.69 s, Size: 139 B
- Body tab selected, showing the response:

```
1 Hello from NiFi Course
```
- Other tabs: Cookies, Headers (3), Test Results
- Bottom right: Save Response, search icon

# Transforming JSON

# Evaluating JSON with processors

Source JSON

```
[  
  {"id" : 6523,  
   "ident" : "00A",  
   "type" : "heliport",  
   "name" : "Total Rf Heliport",  
   "latitude_deg" : 40.07080078125,  
   "longitude_deg" : -74.93360137939453,  
   "elevation_ft" : 11,  
   "continent" : "NA",  
   "iso_country" : "US",  
   "iso_region" : "Pennsylvania",  
   "country_name" : "United States"  
]
```

Target JSON

```
[  
  {"id" : 6523,  
   "ident" : "00A",  
   "type" : "heliport",  
   "name" : "Total Rf Heliport",  
   "iso_region" : "Pennsylvania",  
   "country_name" : "United States"  
]
```

# Evaluating JSON with processors

## Handling Json

EvaluateJsonPath		
EvaluateJsonPath 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
Destination	flowfile-content
Return Type	auto-detect
Path Not Found Behavior	ignore
Null Value Representation	empty string

CANCEL APPLY

# Evaluating JSON with processors

## Source JSON

```
[  
  "id" : 6523,  
  "ident" : "00A",  
  "type" : "heliport",  
  "name" : "Total Rf Heliport",  
  "latitude_deg" : 40.07080078125,  
  "longitude_deg" : -74.93360137939453,  
  "elevation_ft" : 11,  
  "continent" : "NA",  
  "iso_country" : "US",  
  "iso_region" : "Pennsylvania",  
  "country_name" : "United States"  
]
```

## Json Path

`$.id`

`$.type`

`$.longitude_deg`

`$.continent`

General Syntax

`$.fieldName`

# Evaluating JSON with processors

## Handling Json

EvaluateJsonPath		
EvaluateJsonPath 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

Stopped

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property	Value
Destination	flowfile-attribute
Return Type	auto-detect
Path Not Found Behavior	ignore
Null Value Representation	empty string
country_name	\$country_name
id	\$id
ident	\$ident
iso_region	\$iso_region
name	\$name
type	\$type

CANCEL APPLY

# Evaluating JSON with processors

## Creating json

	■ ReplaceText	ReplaceText 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

Stopped

EL ✓ PARAM ✓

```
1 {  
2   "id" : "${id}",  
3   "ident" : "${ident}",  
4   "type" : "${type}",  
5   "name" : "${name}",  
6   "iso_region" : "${iso_region}",  
7   "country_name" : "${country_name}"  
8 }
```

Set empty string

CANCEL APPLY

# Evaluating JSON with processors



■ 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

**Source Json**

Name success

Queued 0 (0 bytes)



■ EvaluateJsonPath  
EvaluateJsonPath 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

**Evaluating json**

Name matched

Queued 0 (0 bytes)



■ ReplaceText  
ReplaceText 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

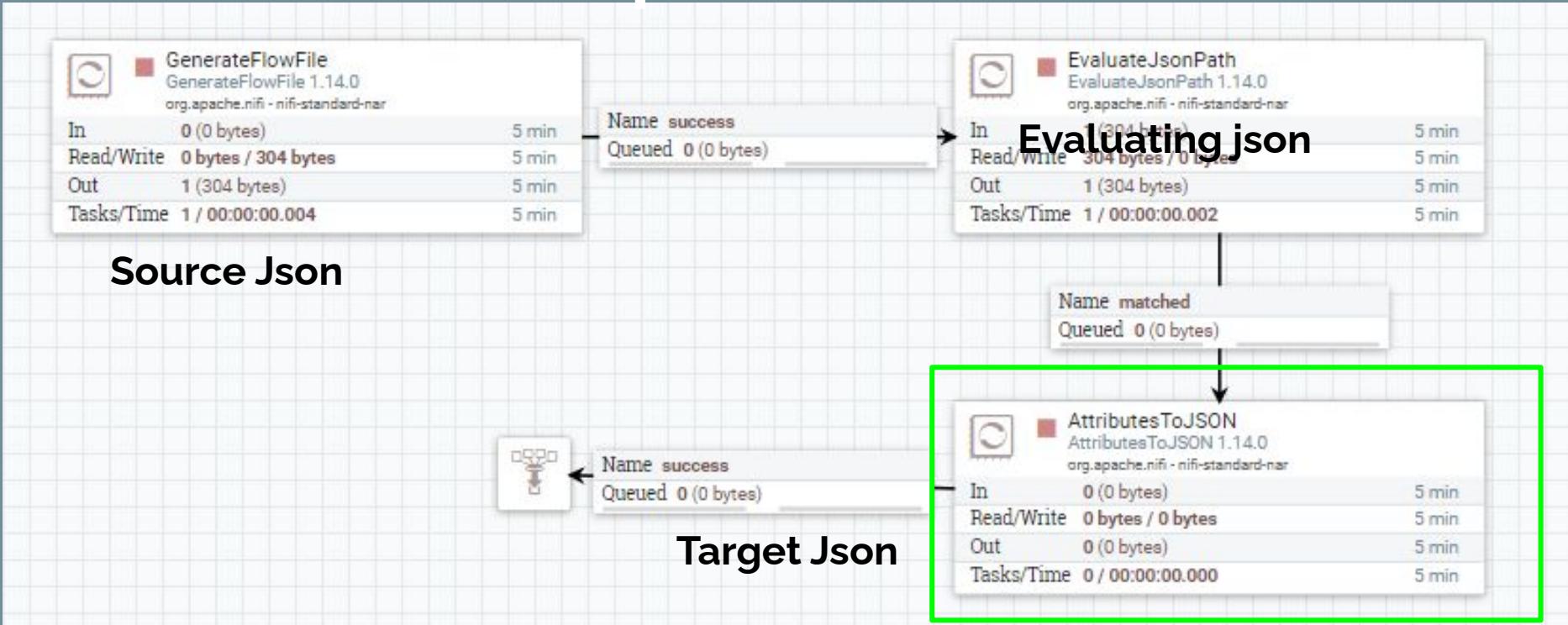


Name success

Queued 0 (0 bytes)

**Target Json**

# – Evaluating JSON with AttributeToJson processor



# – Evaluating JSON with AttributeToJson processor

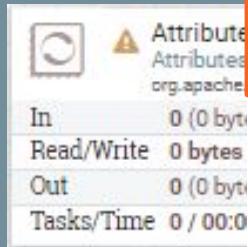
## Creating j

### FlowFile

A FlowFile is a logical notion that correlates a piece of data with a set of Attributes about that data. Such attributes include a FlowFile's unique identifier, as well as its name, size, and any number of other flow-specific values. While the contents and attributes of a FlowFile can change, the FlowFile object is immutable. Modifications to a FlowFile are made possible by the ProcessSession.

The core attributes for FlowFiles are defined in the `org.apache.nifi.flowfile.attributes.CoreAttributes` enum. The most common attributes you'll see are `filename`, `path` and `uuid`. The string in parentheses is the value of the attribute within the `CoreAttributes` enum and how it appears in the UI/API.

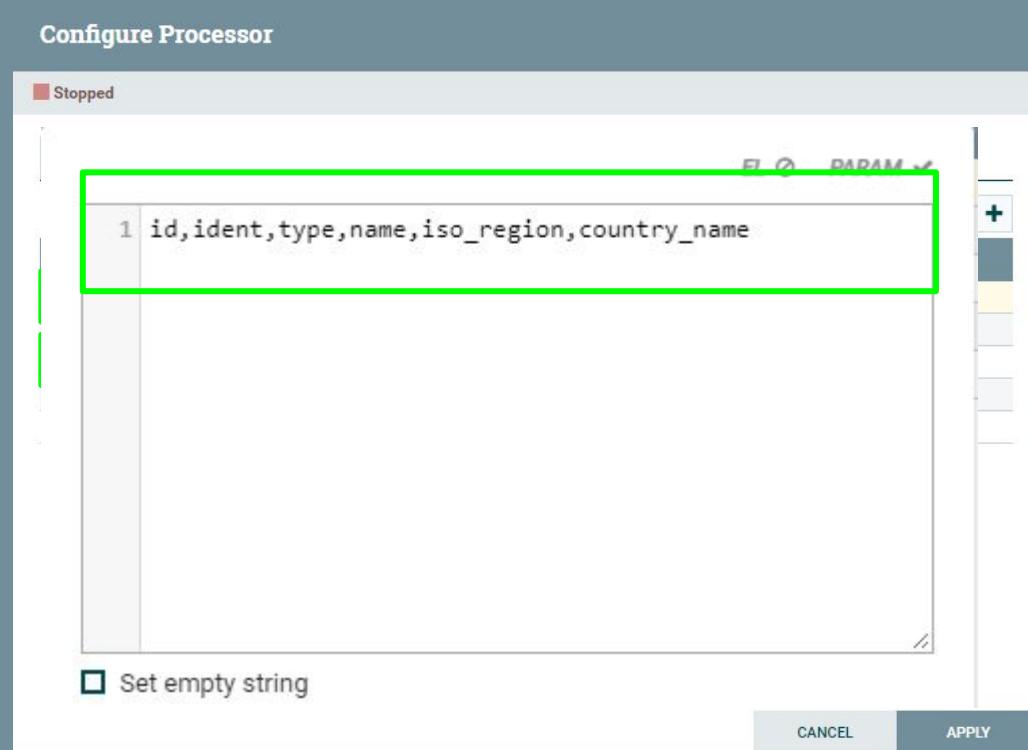
- **Filename (filename)**: The filename of the FlowFile. The filename should not contain any directory structure.
- **UUID (uuid)**: A Universally Unique Identifier assigned to this FlowFile that distinguishes the FlowFile from other FlowFiles in the system.
- **Path (path)**: The FlowFile's path indicates the relative directory to which a FlowFile belongs and does not contain the filename.
- **Absolute Path (absolute.path)**: The FlowFile's absolute path indicates the absolute directory to which a FlowFile belongs and does not contain the filename.
- **Priority (priority)**: A numeric value indicating the FlowFile priority.
- **MIME Type (mime.type)**: The MIME Type of this FlowFile.
- **Discard Reason (discard.reason)**: Specifies the reason that a FlowFile is being discarded.
- **Alternate Identifier (alternate.identifier)**: Indicates an identifier other than the FlowFile's UUID that is known to refer to this FlowFile.



# – Evaluating JSON with AttributeToJson processor

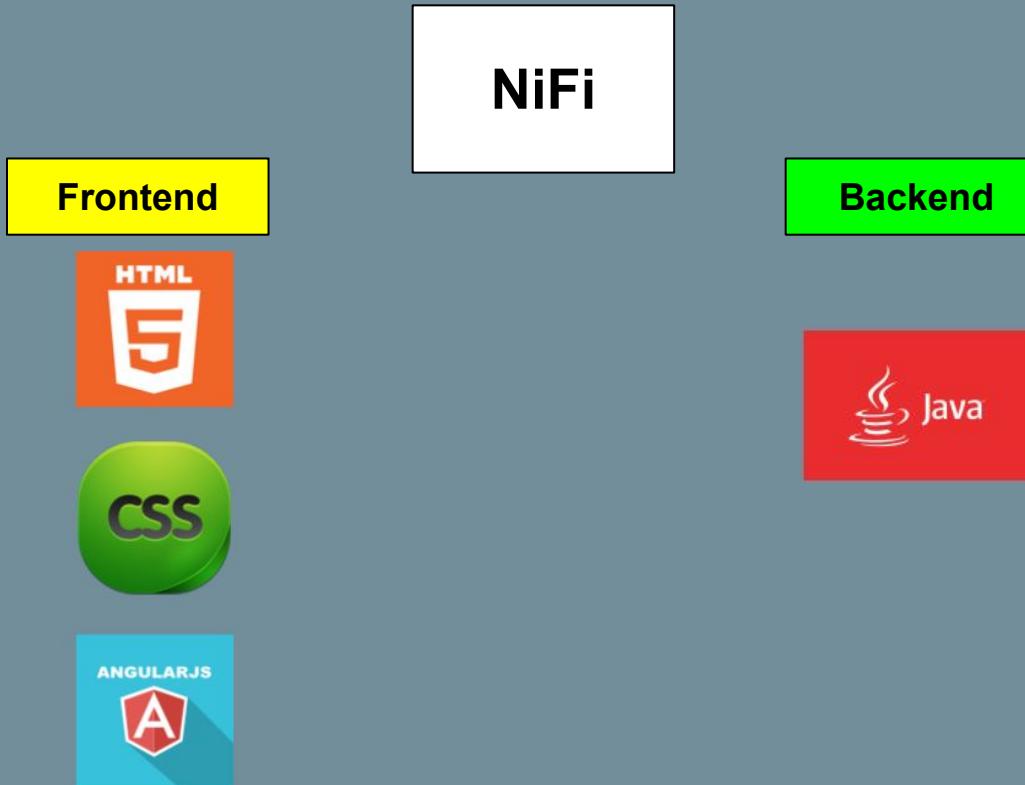
Creating json

	<b>AttributesToJson</b>	AttributesToJson 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

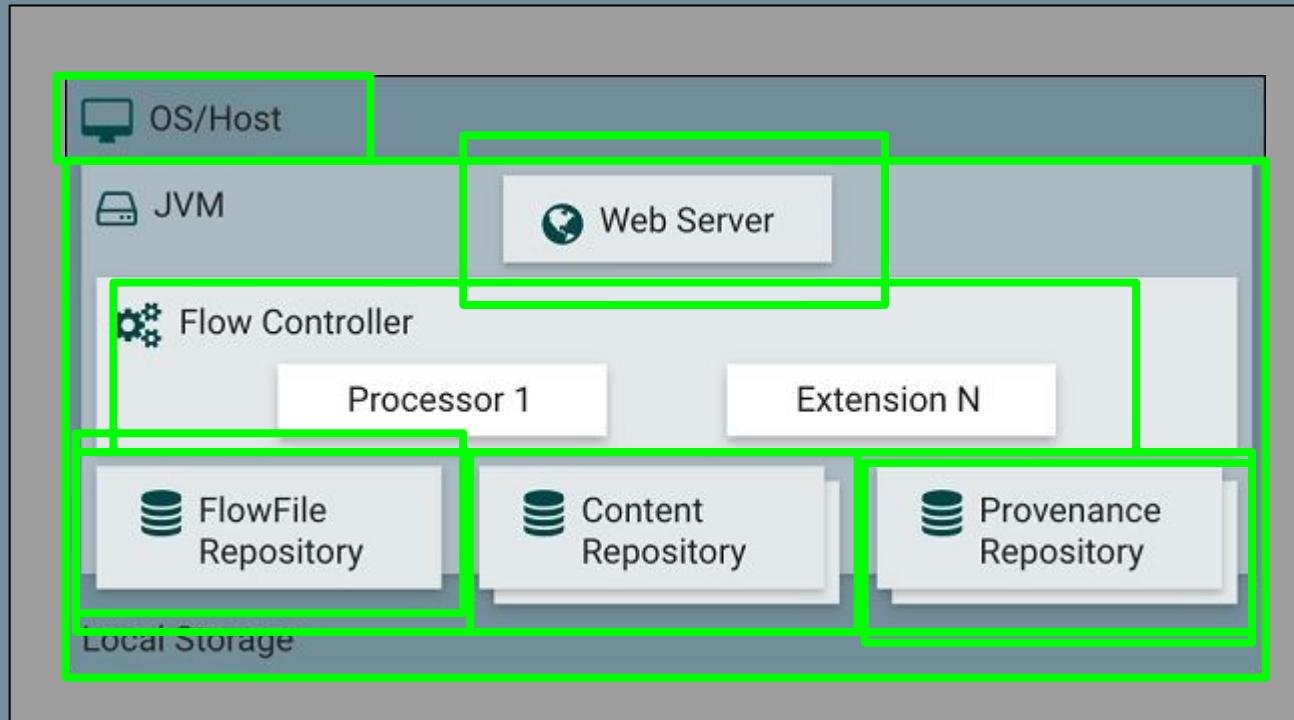


# NiFi Architecture

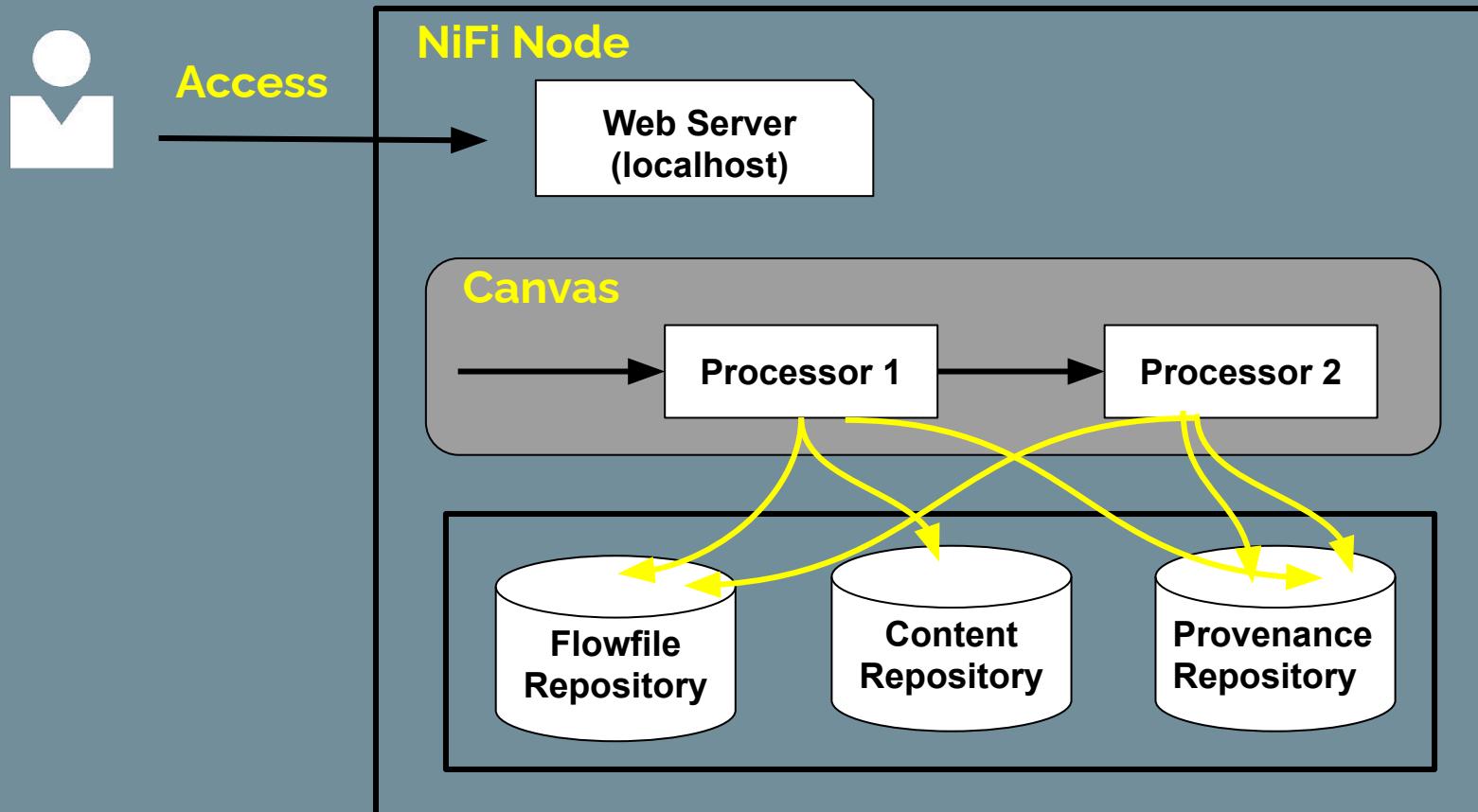
# – NiFi Technologies Used



# – NiFi Standalone Architecture



# – NiFi Standalone Architecture

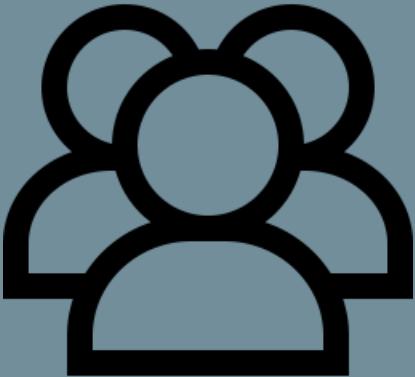


# Some terminologies

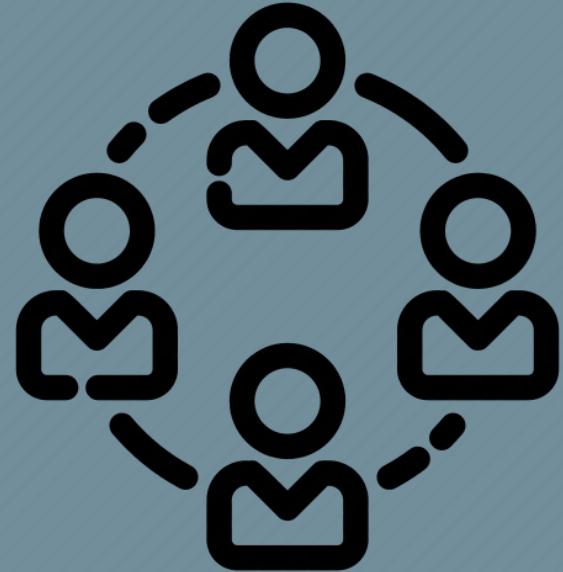
# - What is Cluster



Single



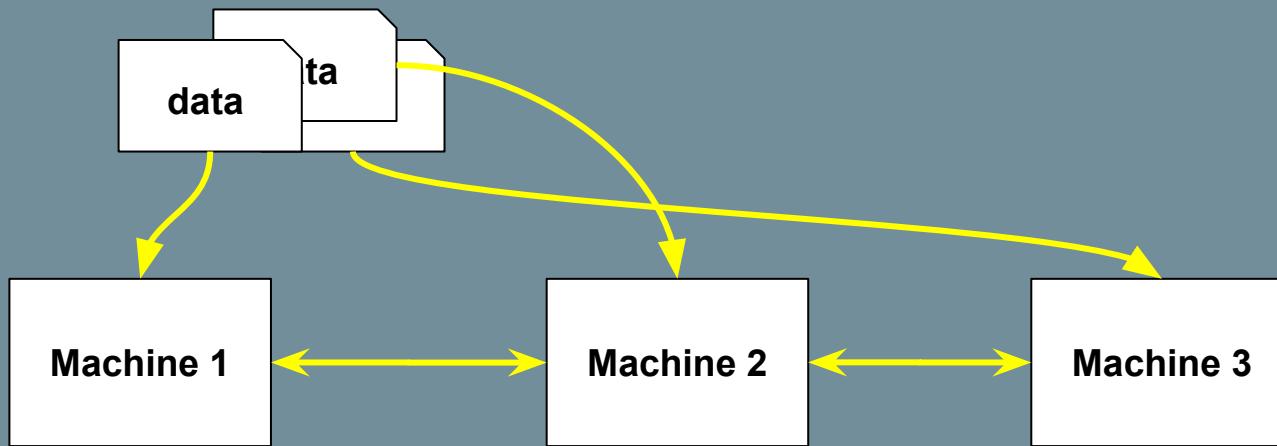
Group



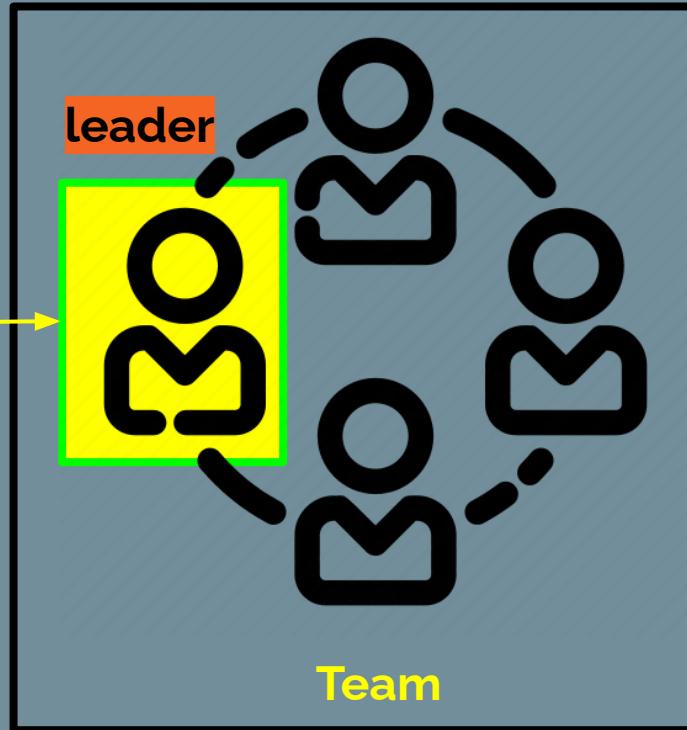
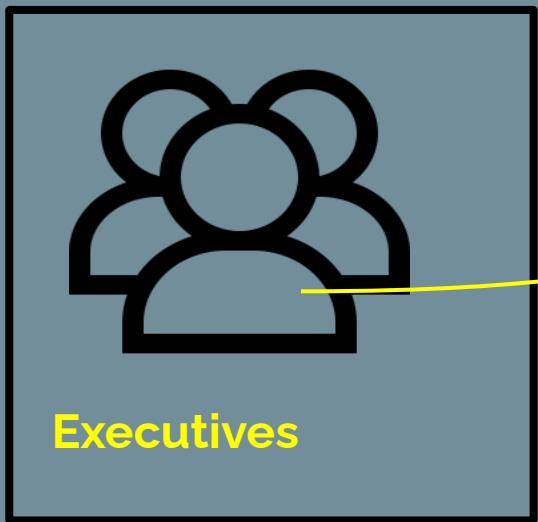
Cluster

# - What is Cluster

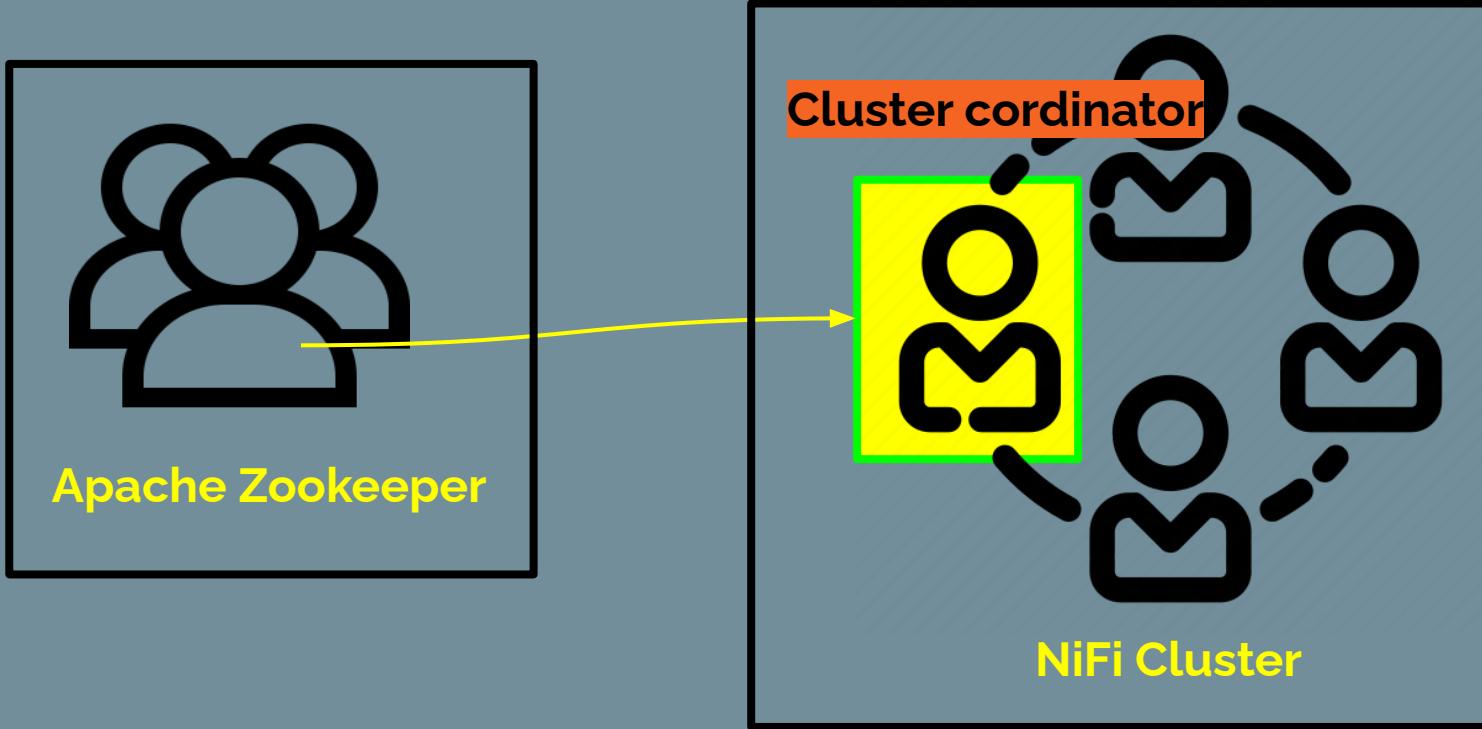
Group of commodities hardwares (machines) working together to achieve the common goal but however operate on different datasets



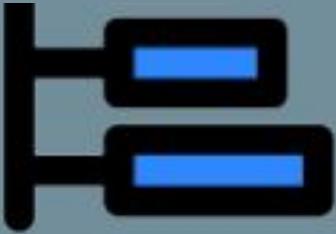
# - Electing Leader



# - Analogy in NiFi



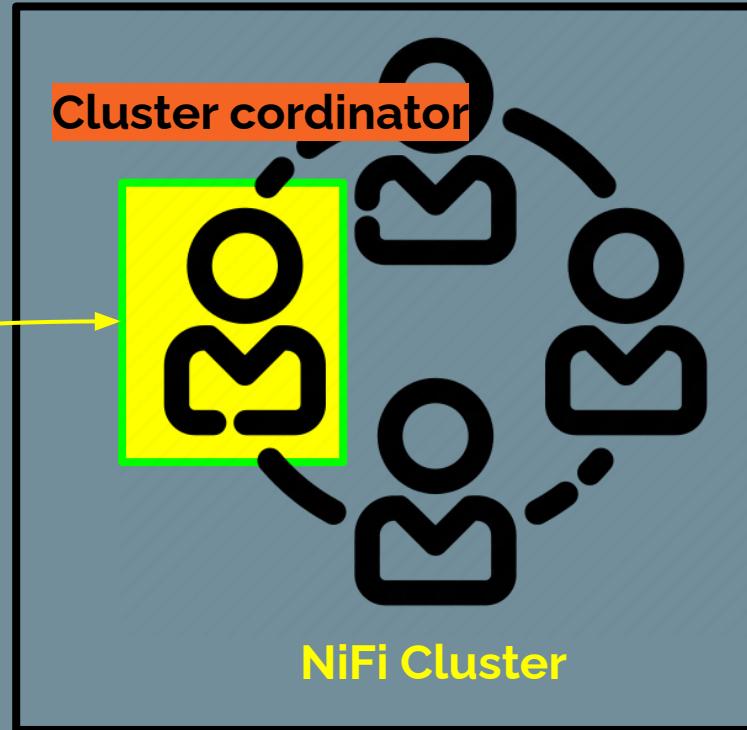
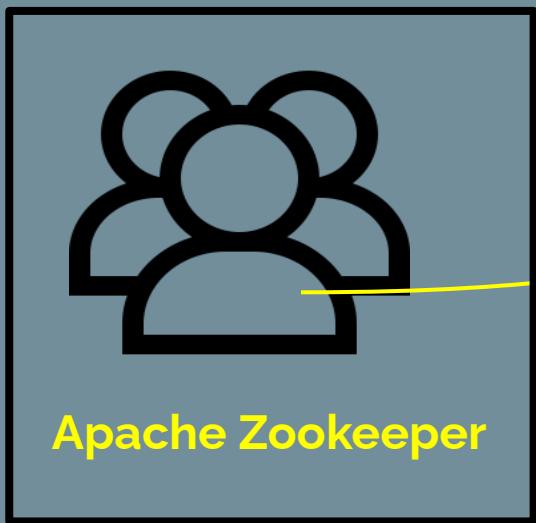
# – Zookeeper Provisioning



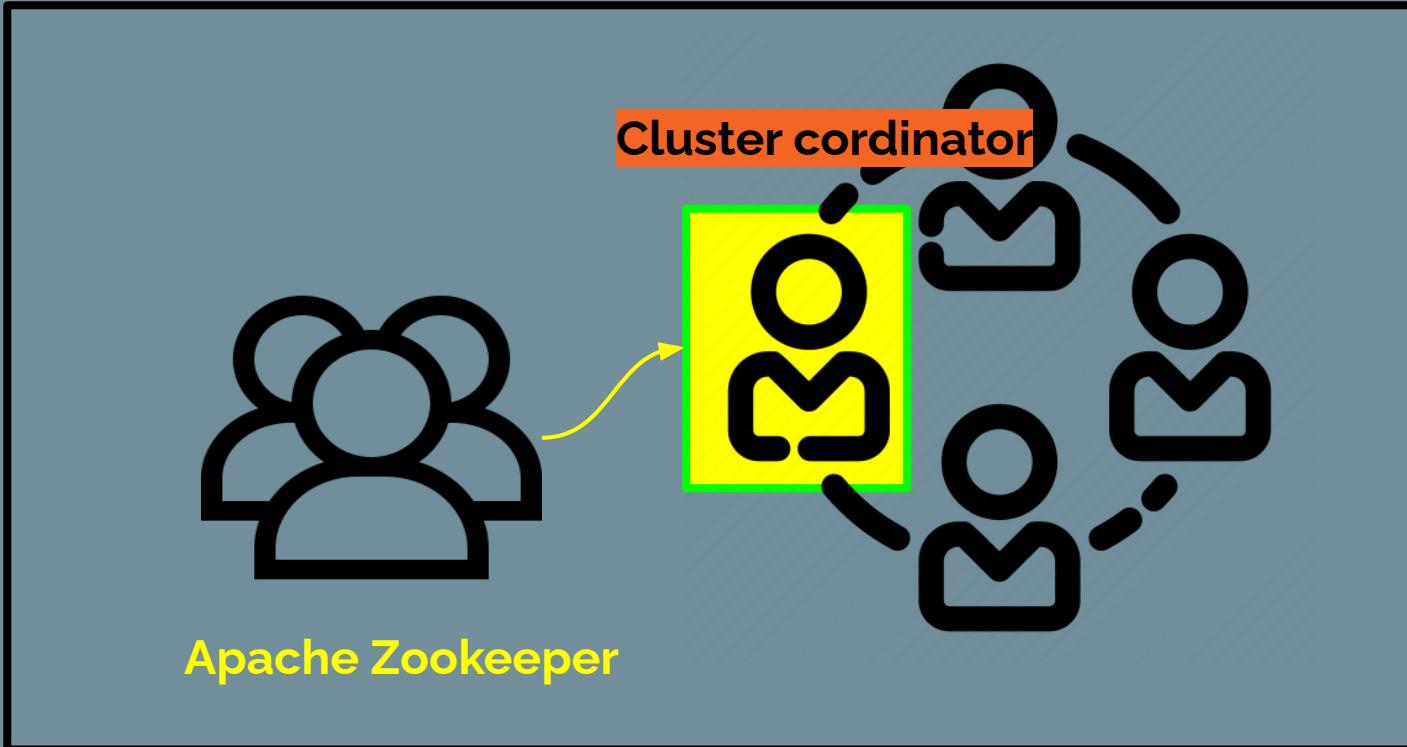
External Zookeeper

Embedded Zookeeper

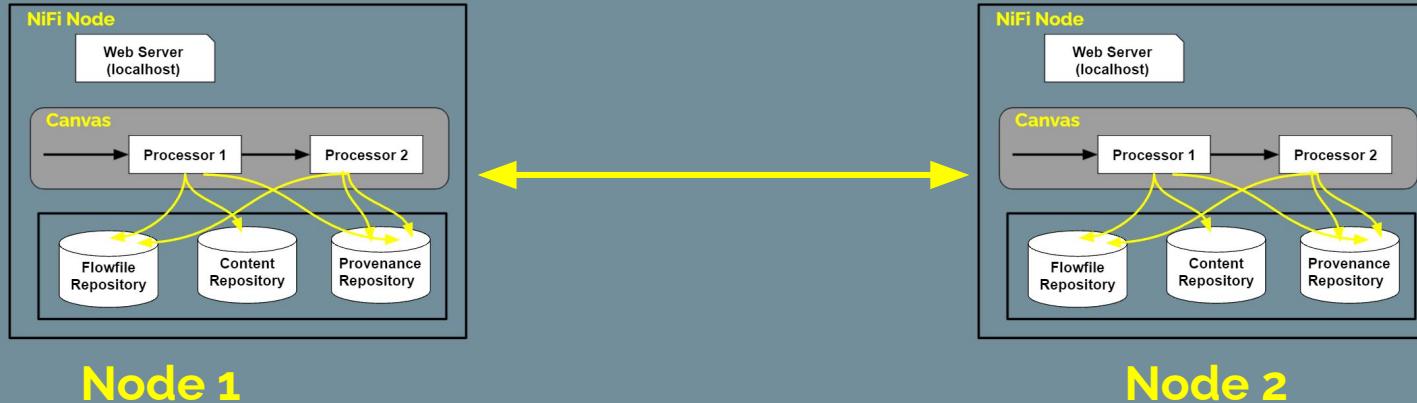
# - External Zookeeper



# Embedded Zookeeper NiFi

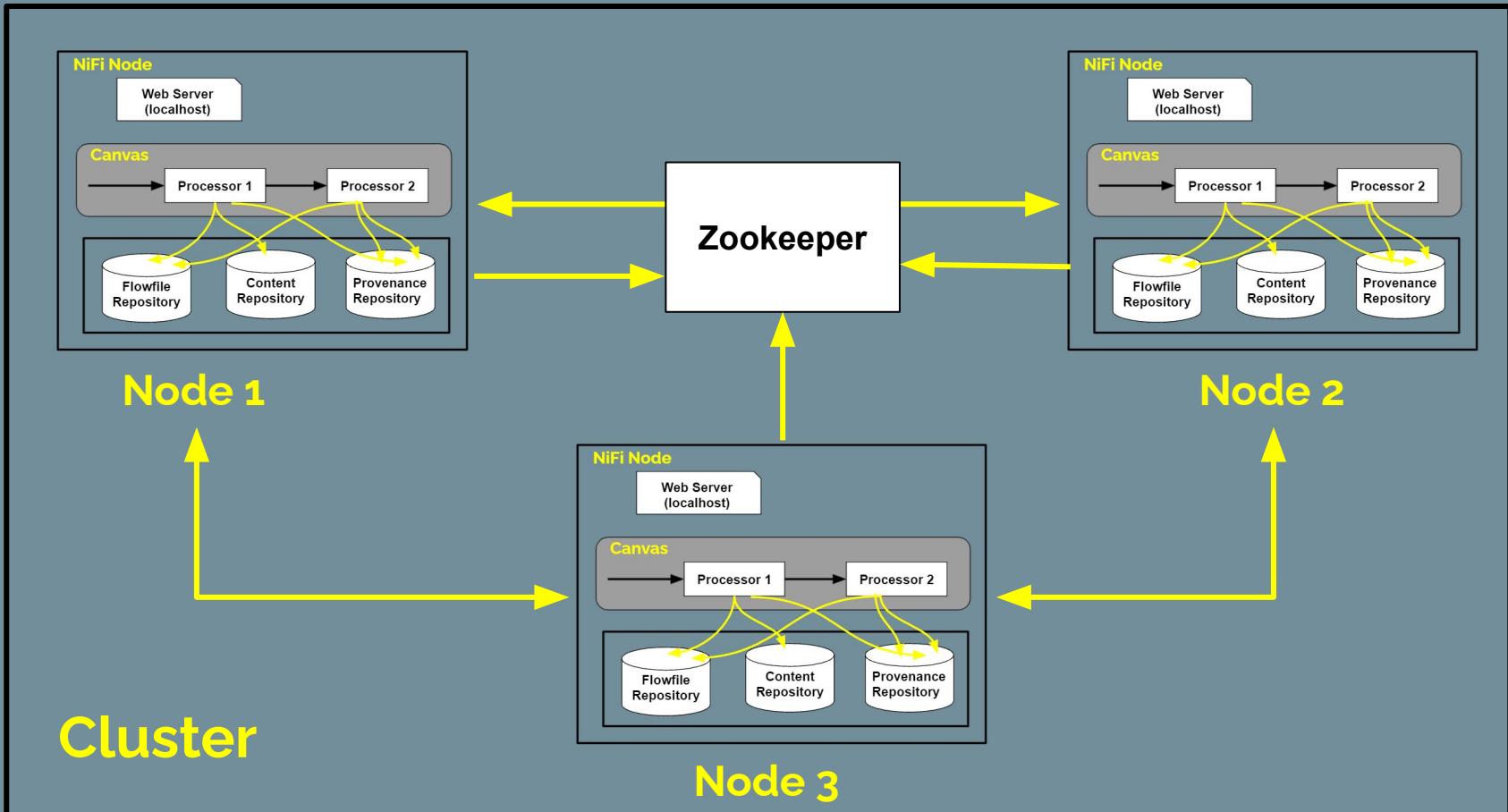


# – NiFi Cluster

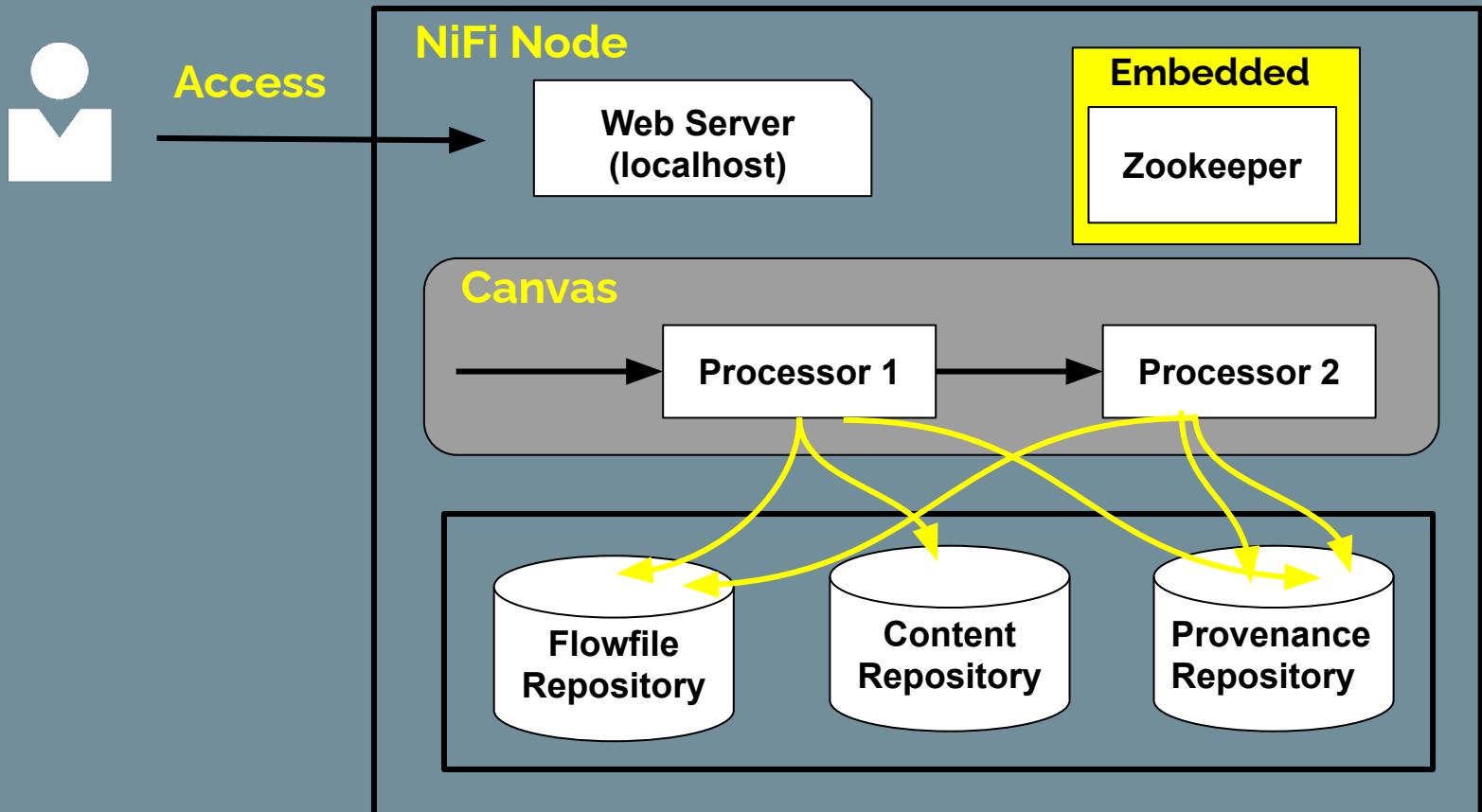


- External Zookeeper
- Embedded Zookeeper

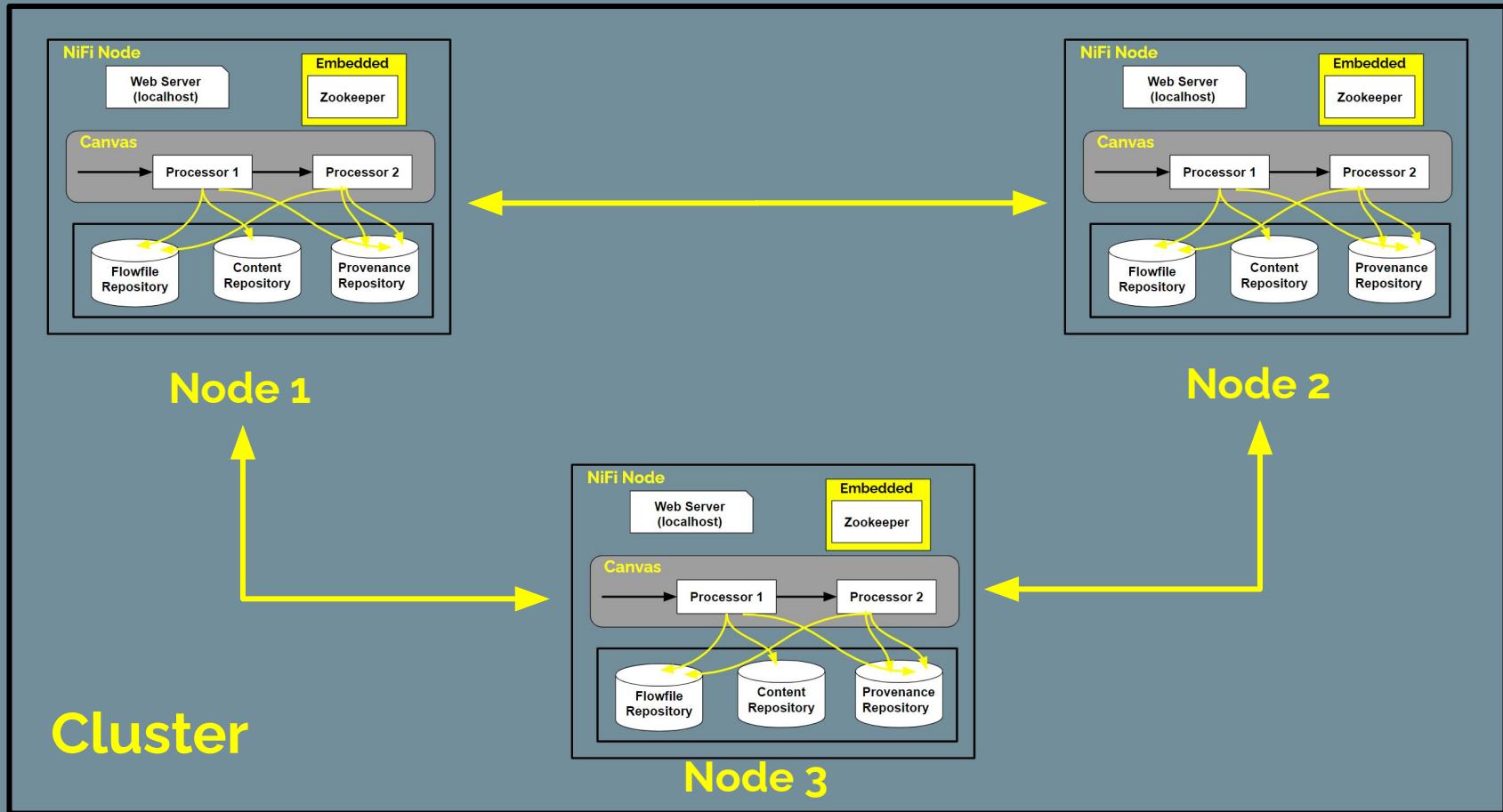
# NiFi Cluster Architecture (External Zookeeper)



# \_ Standalone with Embedded Zookeeper

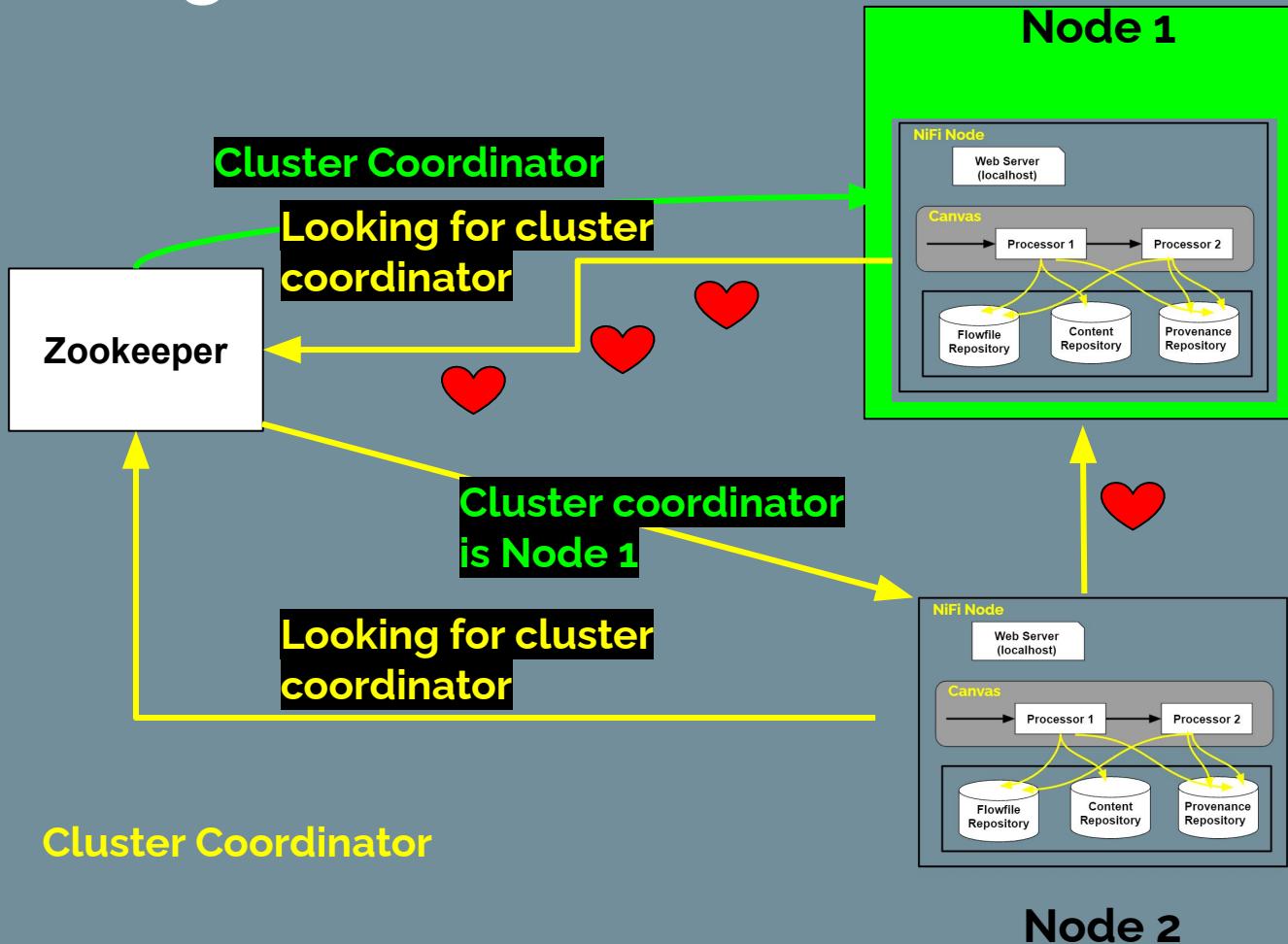


# NiFi Cluster Architecture (Embedded Zookeeper)



# Electing Cluster Coordinator

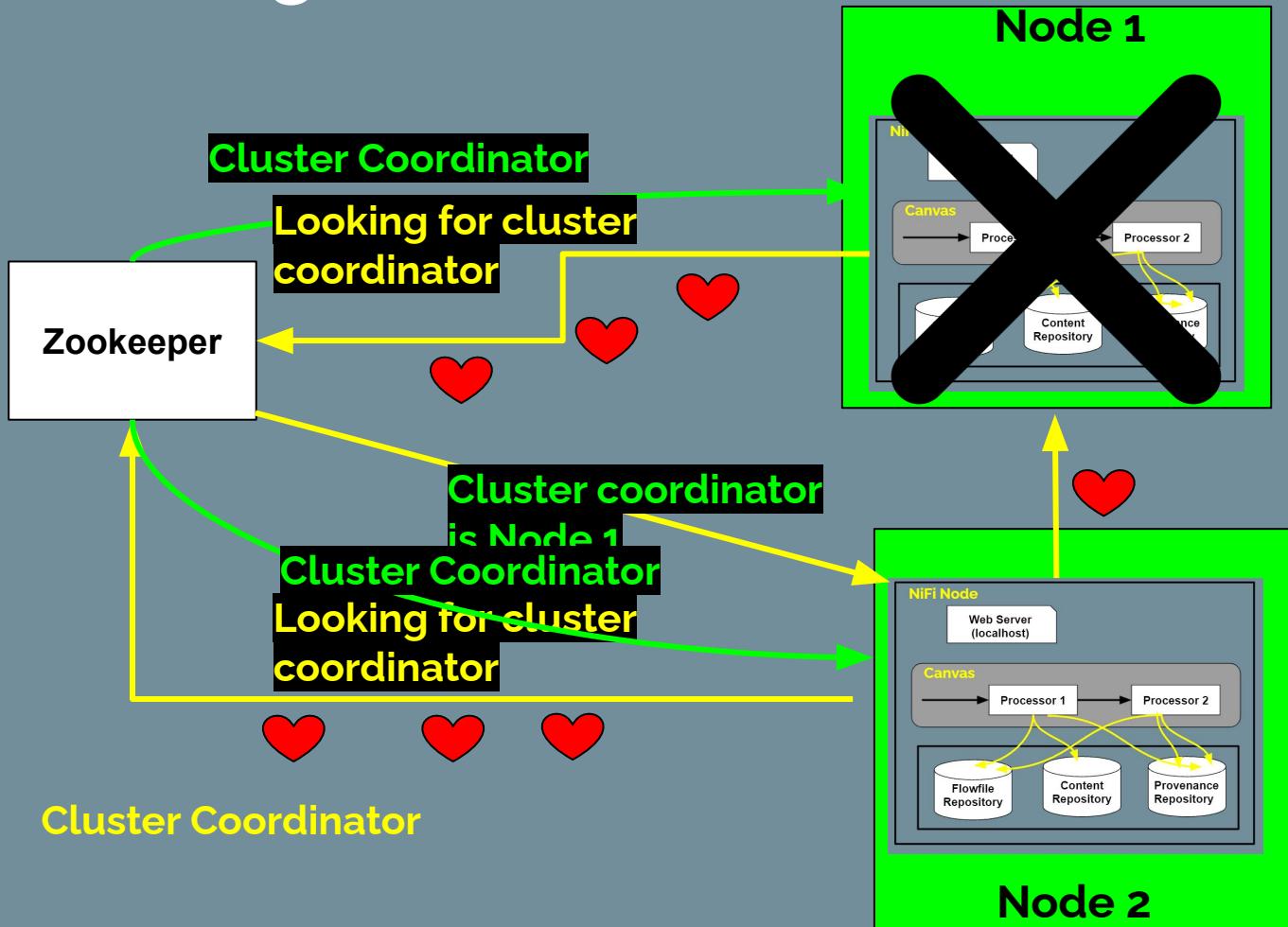
# Electing Cluster Coordinator



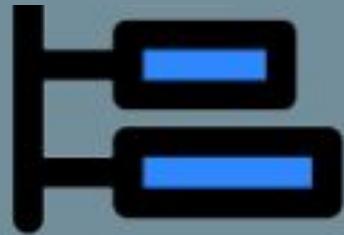
# What If Cluster Coordinator **DIES ?**

Zookeeper will perform  
**Re-election**

# Reelecting Cluster Coordinator



# Zookeeper Performs Two Elections?

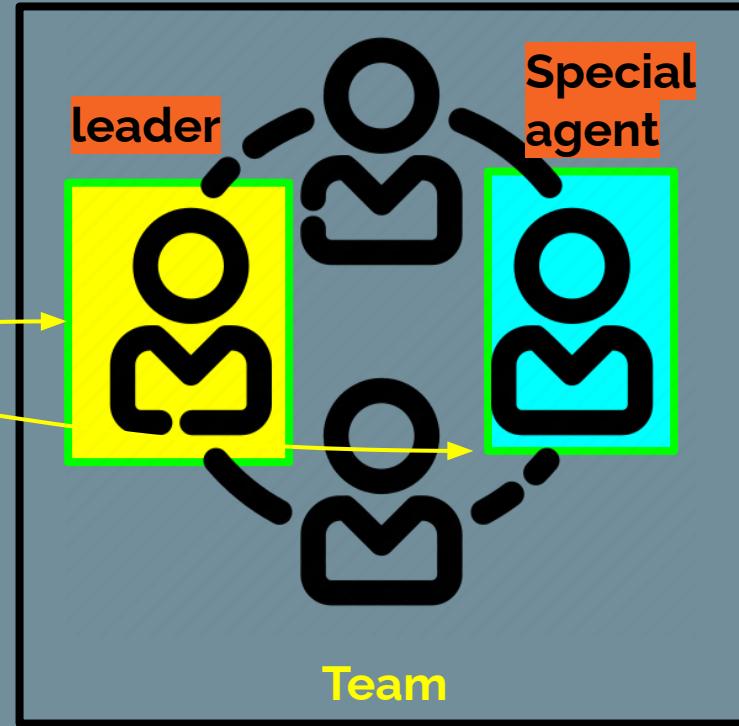
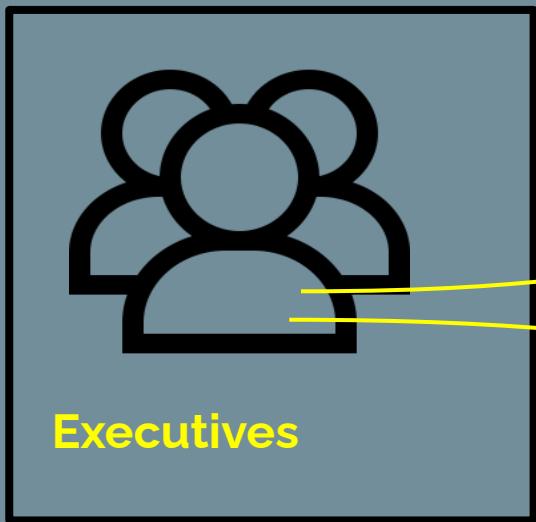


Electing Cluster Coordinator

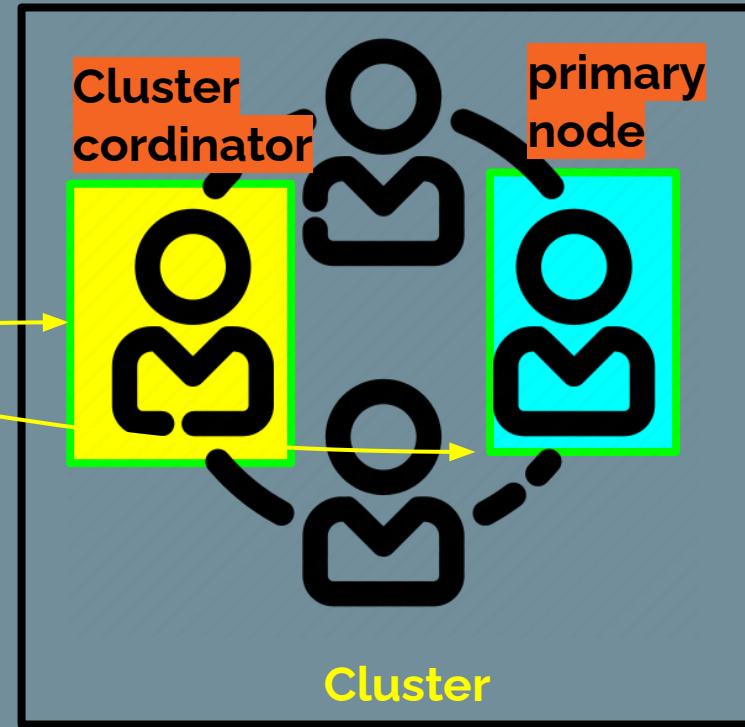
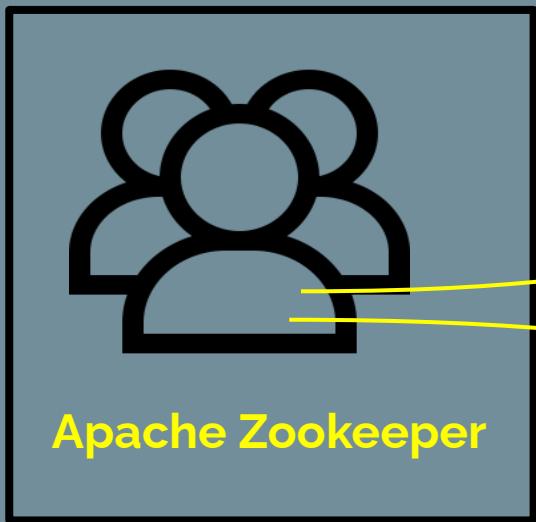
Electing Primary Node

# Electing Primary Node

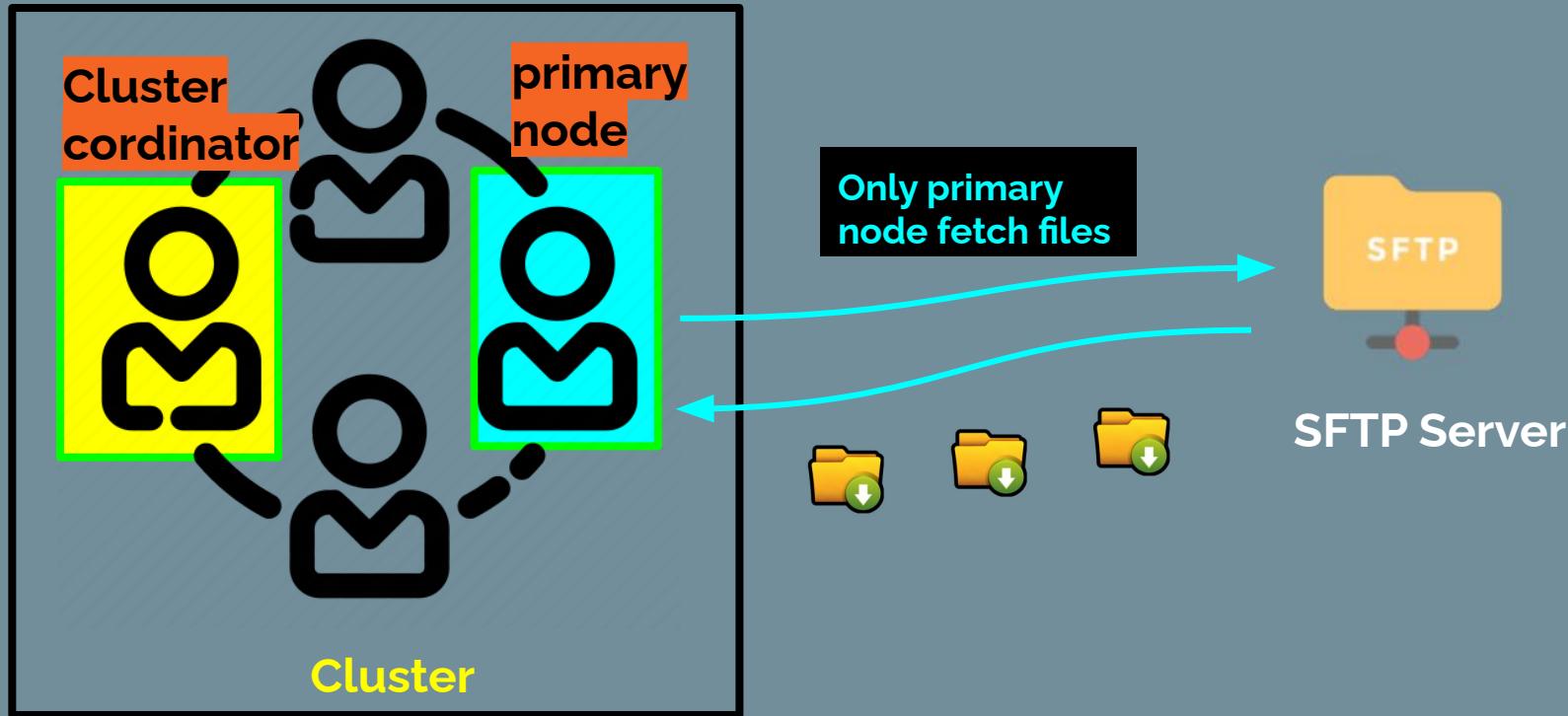
# - Electing Primary Node



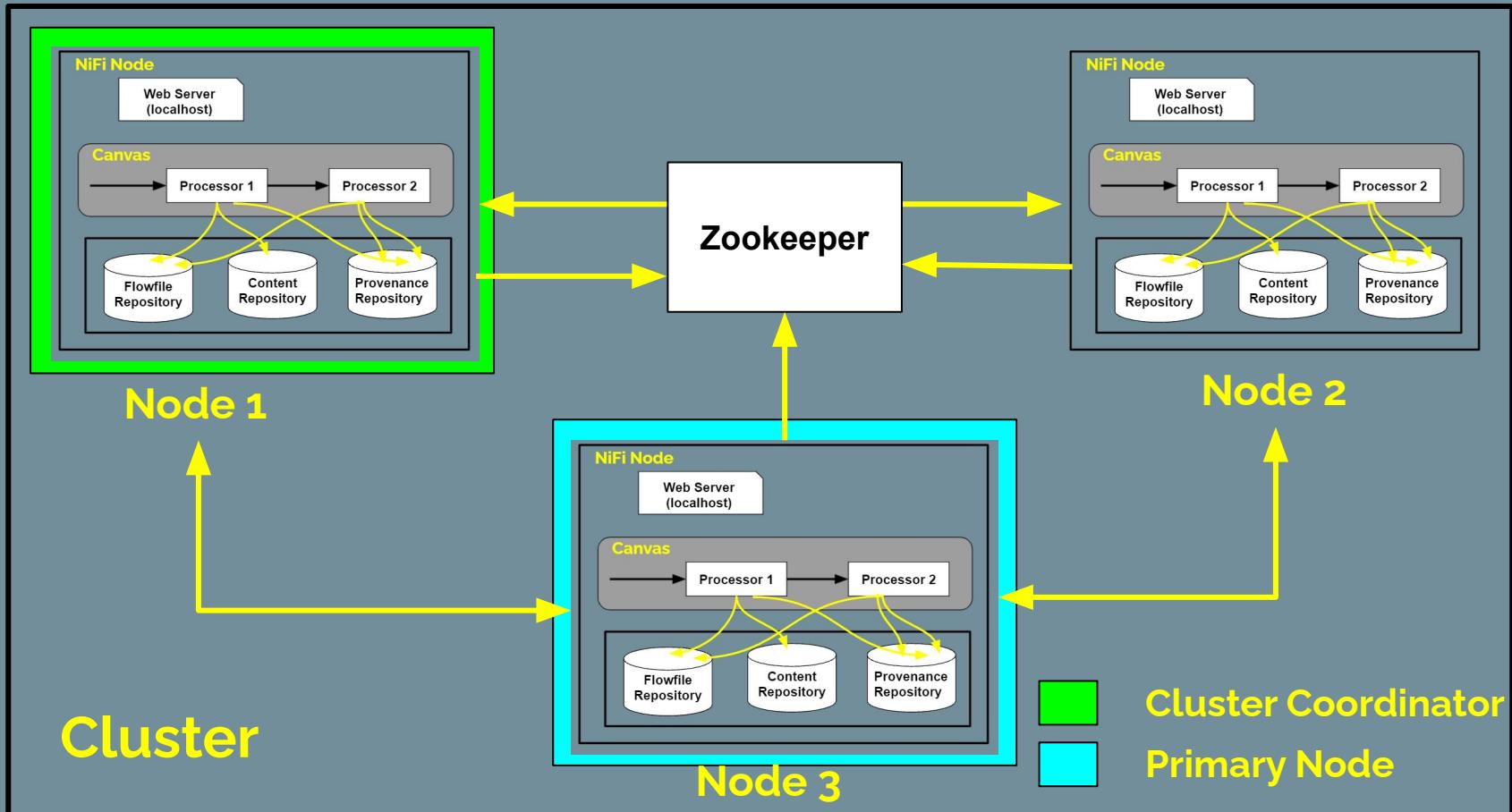
# – Electing Primary Node



# – Role of Primary Node



# Electing Primary Node



# NiFi Cluster View

NiFi Cluster

Displaying 2 of 2

Filter by address

Node Address	Active Thread Count	Queue / Size	Status	Started At	Last Heartbeat
node-1:8080	0	0 / 0 bytes	CONNECTED PRIMARY COORDINATOR	11/15/2021 08:18:43 UTC	11/15/2021 08:40:14 UTC
node-2:8080	0	0 / 0 bytes	CONNECTED	11/15/2021 08:18:45 UTC	11/15/2021 08:40:15 UTC

Last updated: 08:40:17 UTC

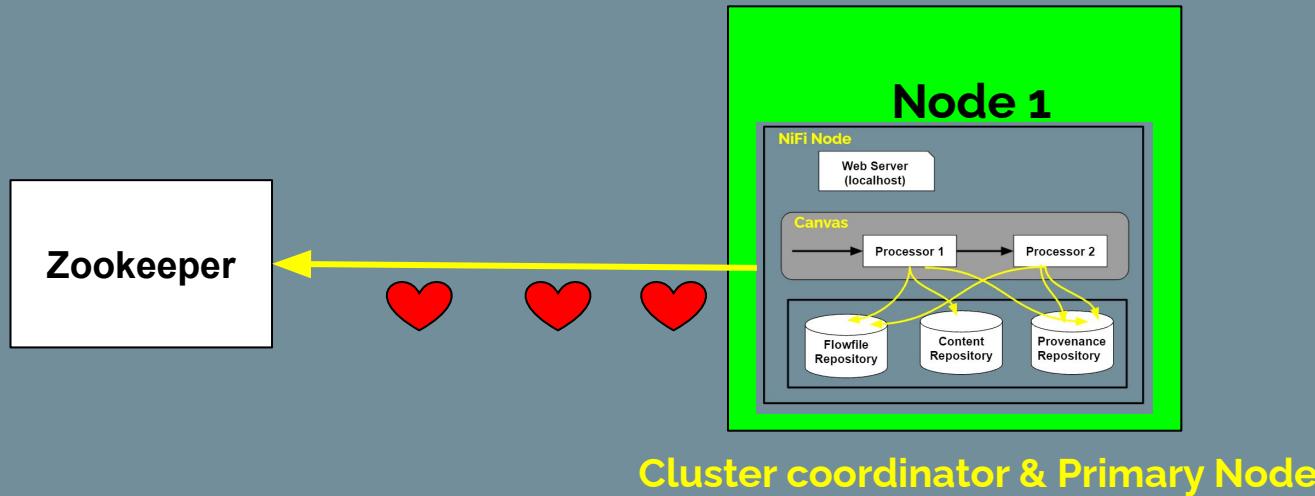
NIFI Flow

# What If Primary Node **DIES ?**

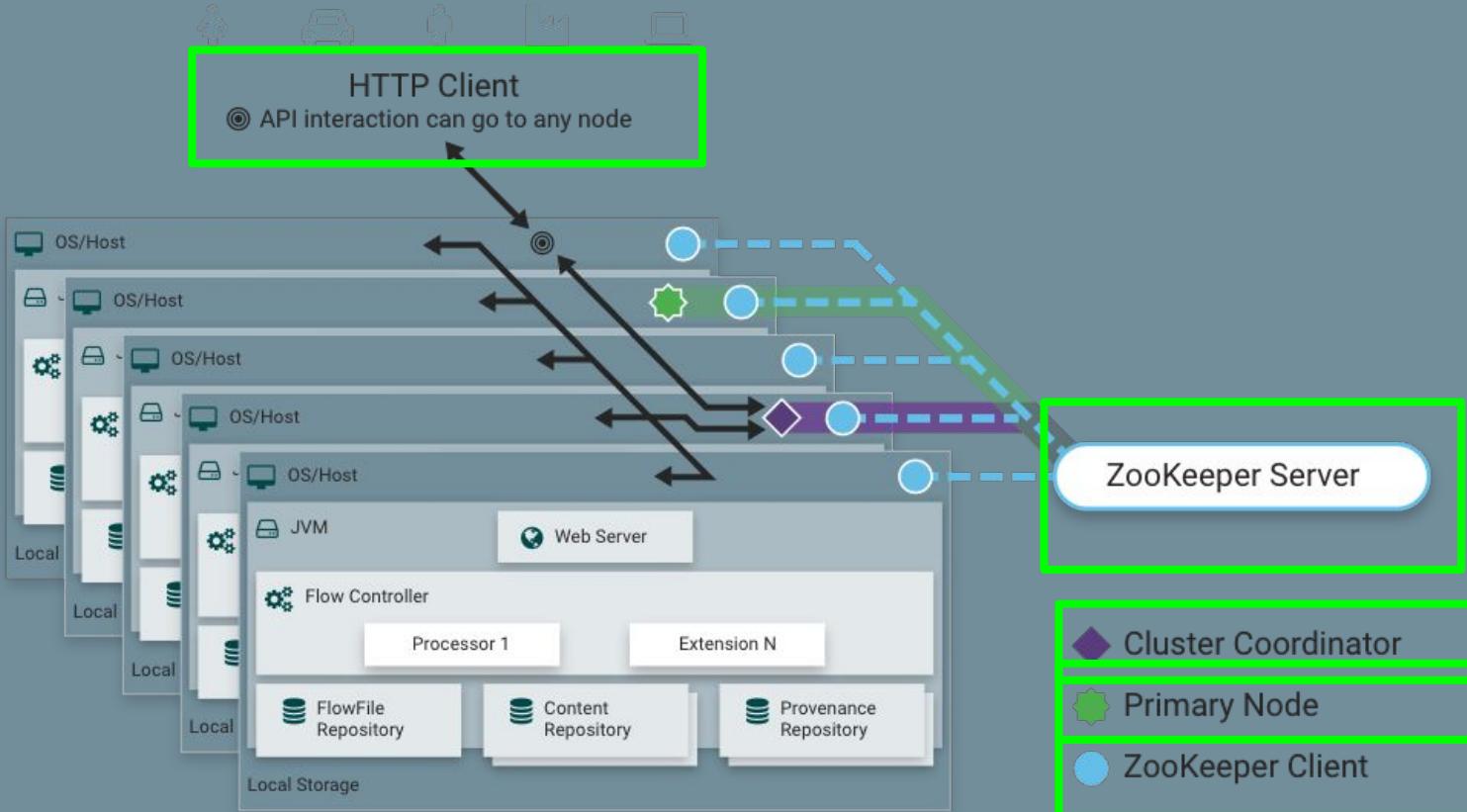
Zookeeper will perform  
**Re-election**

# - Role of Zookeeper

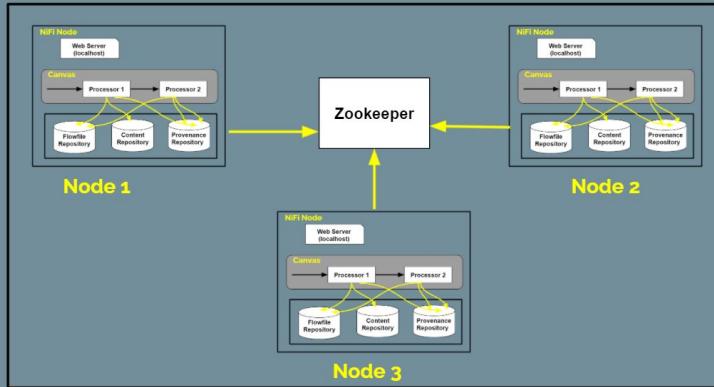
- Elects Primary and Cluster Coordinator
- Listens for heartbeat from them



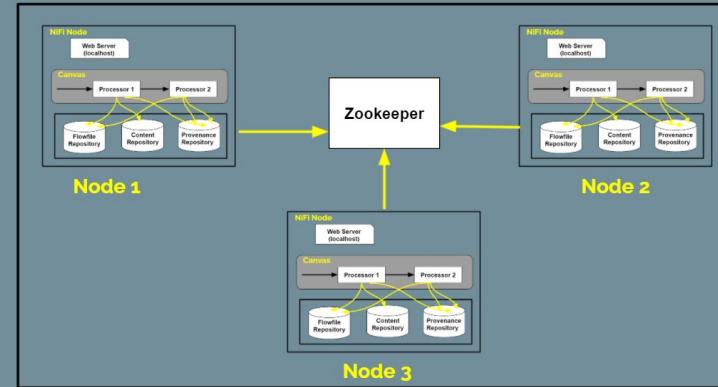
# – NiFi Cluster Architecture



# – NiFi Provisioning



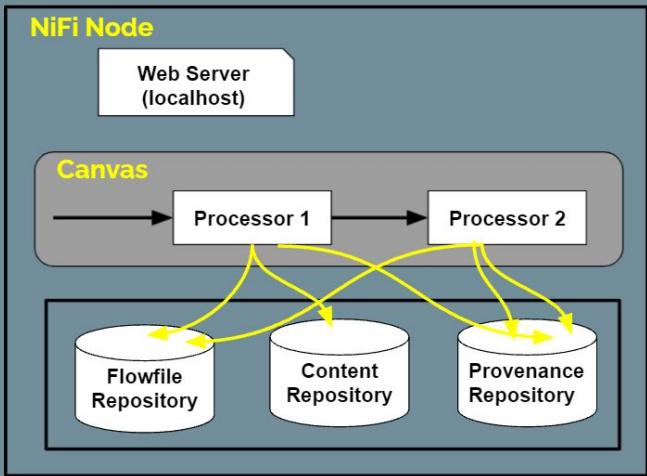
Dev Instance



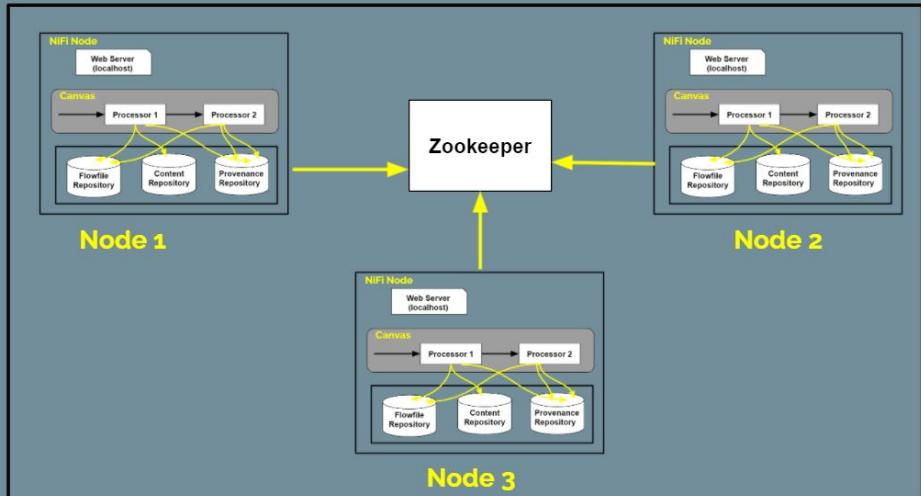
Prod Instance

# NiFi Installation

# – NiFi Installation



**Standalone**

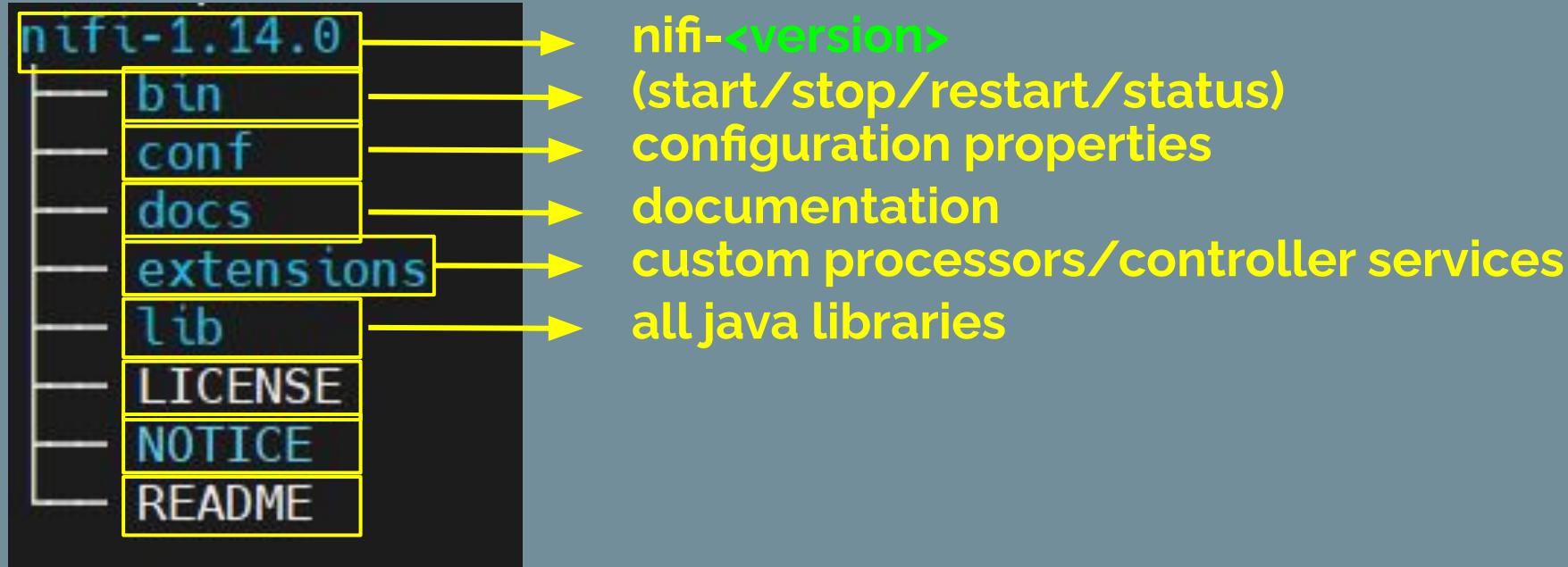


**Cluster**

# Let's understand **NiFi Directory Structure**

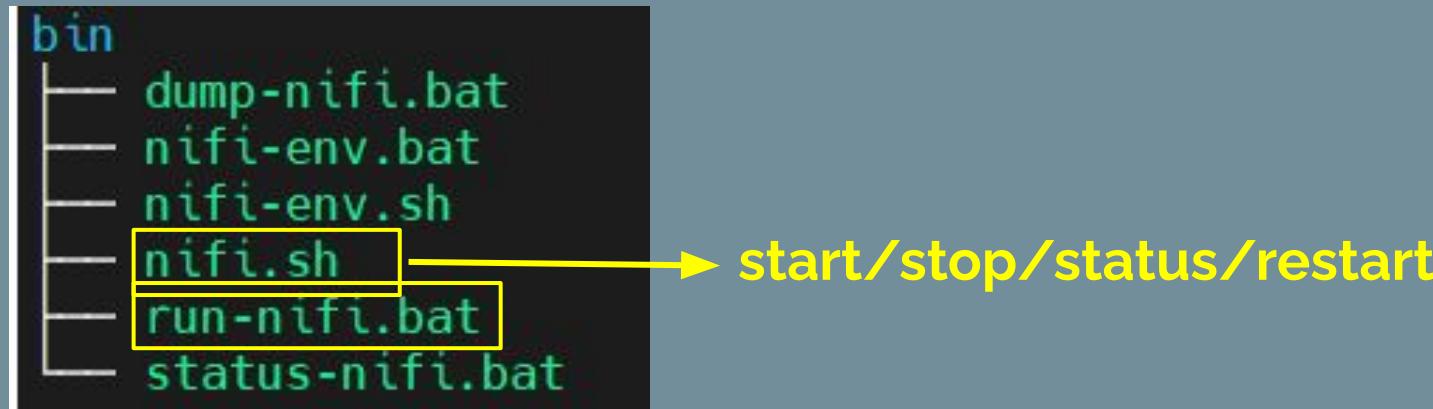
# - NiFi Directory Structure

## Inside NiFi Binary



# - NiFi Directory Structure

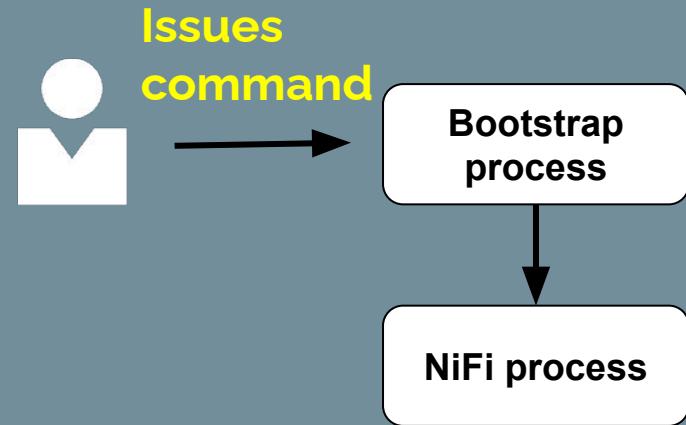
Inside bin directory



# - NiFi Directory Structure

## Inside conf directory

```
conf/
├── authorizers.xml
├── bootstrap-hashing-vault.conf
├── bootstrap-notification-services.xml
├── bootstrap.conf
├── logback.xml
├── login-identity-providers.xml
├── nifi.properties
├── state-management.xml
├── stateless-logback.xml
├── stateless.properties
└── zookeeper.properties
```



# NiFi Conf Files

nifi.properties

NiFi Users

users.xml

NiFi User Policies

authorization.xml

All Login Identity Providers

login-identity-providers.xml

who can access config

authorizers.xml

embedded zookeeper  
config

zookeeper.properties

nifi logging conf

logback.xml

# - NiFi Directory Structure

## Inside lib directory

```
lib/
├── aspectj
├── bootstrap
├── java11
├── javax.servlet-api-3.1.0.jar
├── jcl-over-slf4j-1.7.30.jar
├── jetty-schemas-3.1.jar
├── jul-to-slf4j-1.7.30.jar
├── log4j-over-slf4j-1.7.30.jar
├── logback-classic-1.2.3.jar
├── logback-core-1.2.3.jar
├── nifi-ambari-nar-1.14.0.nar
├── nifi-amqp-nar-1.14.0.nar
├── nifi-api-1.14.0.jar
├── nifi-avro-nar-1.14.0.nar
├── nifi-aws-nar-1.14.0.nar
├── nifi-aws-service-api-nar-1.14.0.nar
├── nifi-azure-nar-1.14.0.nar
├── nifi-azure-services-api-nar-1.14.0.nar
├── nifi-cassandra-nar-1.14.0.nar
├── nifi-cassandra-services-api-nar-1.14.0.nar
├── nifi-cassandra-services-nar-1.14.0.nar
├── nifi-ccda-nar-1.14.0.nar
├── nifi-cdc-mysql-nar-1.14.0.nar
├── nifi-confluent-platform-nar-1.14.0.nar
├── nifi-couchbase-nar-1.14.0.nar
├── nifi-couchbase-services-api-nar-1.14.0.nar
├── nifi-cybersecurity-nar-1.14.0.nar
├── nifi-datadog-nar-1.14.0.nar
└── nifi-dbcpservice-nar-1.14.0.nar
```

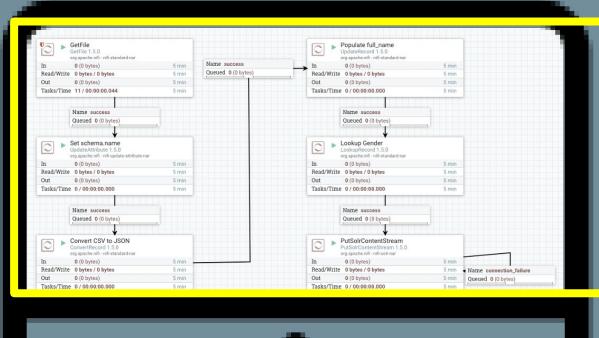
# - NiFi Directory Structure

## Additional Directories when NiFi is started

```
.  
├── bin  
├── conf  
├── content_repository  
├── database_repository  
├── docs  
├── extensions  
├── flowfile_repository  
├── lib  
├── logs  
├── provenance_repository  
├── run  
├── state  
└── work  
└── LICENSE  
└── NOTICE  
└── README
```

# Understanding **nifi.properties**

# \_ Conf Directory - nifi.properties



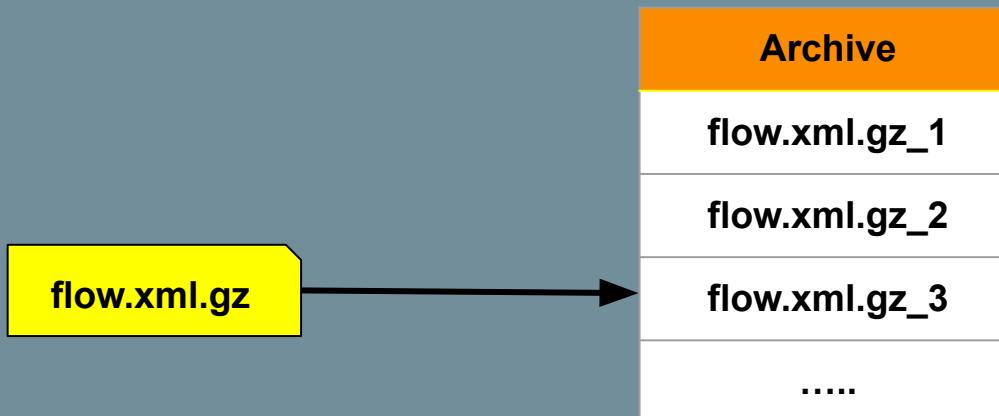
flow.xml.gz

## # Core Properties #

```
nifi.flow.configuration.file=./conf/flow.xml.gz  
nifi.flow.configuration.archive.enabled=true  
nifi.flow.configuration.archive.dir=./conf/archive/  
nifi.flow.configuration.archive.max.time=30 days  
nifi.flow.configuration.archive.max.storage=500 MB  
nifi.flow.configuration.archive.max.count=
```

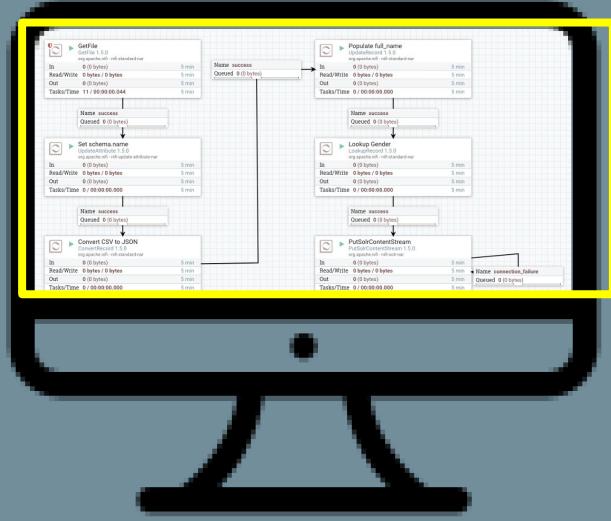
# \_ Conf Directory - nifi.properties

```
# Core Properties #
nifi.flow.configuration.file=./conf/flow.xml.gz
nifi.flow.configuration.archive.enabled=true
nifi.flow.configuration.archive.dir=./conf/archive/
nifi.flow.configuration.archive.max.time=30 days
nifi.flow.configuration.archive.max.storage=500 MB
nifi.flow.configuration.archive.max.count=
```



- Maximum Size
- How to long to keep

# – Conf Directory - nifi.properties



flow.xml.gz

Before  
Restart

Processor  
RUNNING

After  
Restart

Processor  
RUNNING

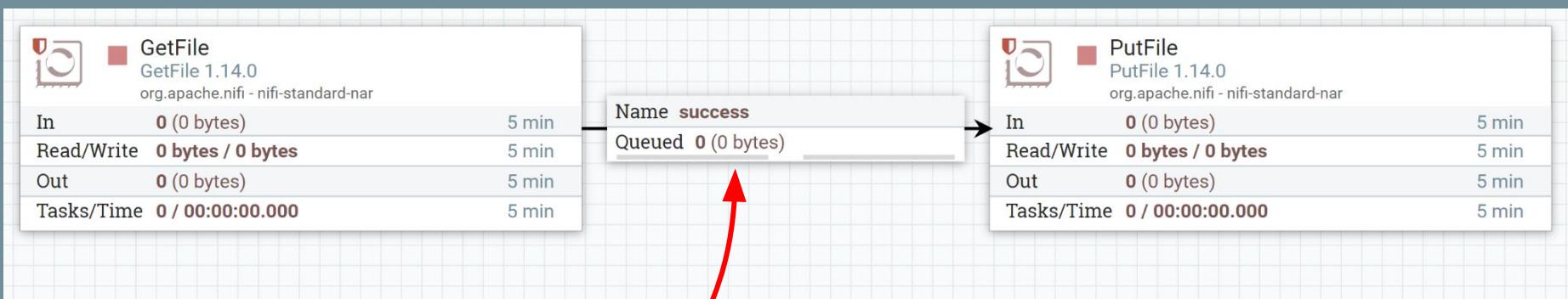
Processor  
STOPPED

nifi.flowcontroller.autoResumeState=true

nifi.flowcontroller.graceful.shutdown.period=10 sec

nifi.flowservice.writedelay.interval=500 ms

# \_ Conf Directory - nifi.properties

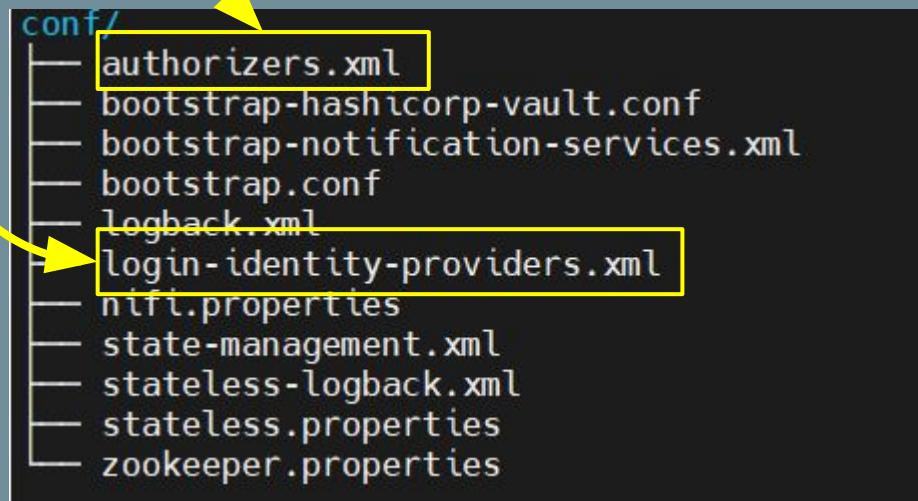


```
# If a component has no work to do (is "bored"), how long sh  
nifi.bored.yield.duration=10 millis  
nifi.queue.backpressure.count=10000  
nifi.queue.backpressure.size=1 GB
```

# \_ Conf Directory - nifi.properties

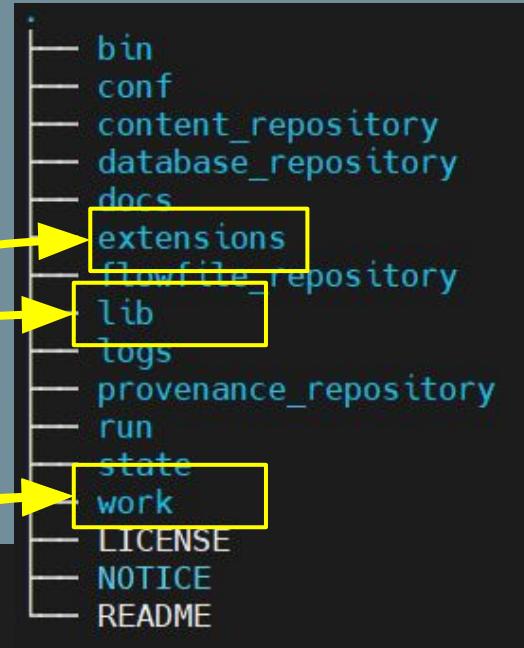
```
nifi.authorizer.configuration.file=./conf/authorizers.xml
```

```
nifi.login.identity.provider.configuration.file=./conf/login-identity-providers.xml
```



# \_ Conf Directory - nifi.properties

```
nifi.ui.autorefresh.interval=30 sec
nifi.nar.library.directory=./lib
nifi.nar.library.autoload.directory=./extensions
nifi.nar.working.directory=./work/nar/
nifi.documentation.working.directory=./work/docs/components
```



# \_ Conf Directory - nifi.properties

```
#####
# State Management #
#####

nifi.state.management.configuration.file=./conf/state-management.xml
# The ID of the local state provider
nifi.state.management.provider.local=local-provider
# The ID of the cluster-wide state provider. This will be ignored if NiFi is not clustered
nifi.state.management.provider.cluster=zk-provider
# Specifies whether or not this instance of NiFi should run an embedded ZooKeeper server
nifi.state.management.embedded.zookeeper.start=false
# Properties file that provides the ZooKeeper properties to use if using state management
nifi.state.management.embedded.zookeeper.properties=./conf/
```

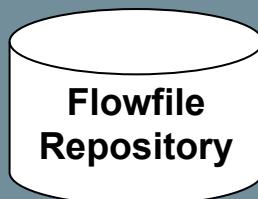
```
conf/
├── authorizers.xml
├── bootstrap-hashicorp-vault.conf
├── bootstrap-notification-services.xml
├── bootstrap.conf
├── logback.xml
├── login-identity-providers.xml
└── nifi.properties
```

nifi.properties

```
└── state-management.xml
└── stateless-logback.xml
└── stateless.properties
└── zookeeper.properties
```

# \_ Conf Directory - nifi.properties

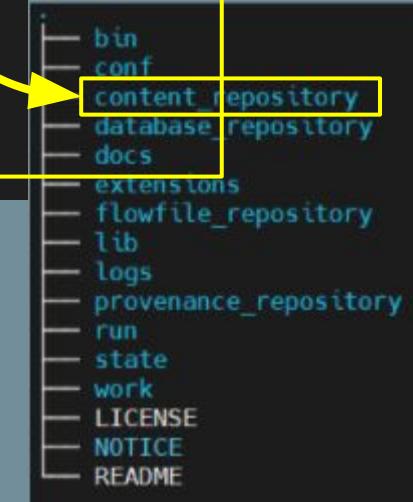
```
# FlowFile Repository
nifi.flowfile.repository.implementation=org.apache.nifi.controller.repository.WriteAheadFlowFileRepository
nifi.flowfile.repository.wal.implementation=org.apache.nifi.wal.SequentialAccessWriteAheadLog
nifi.flowfile.repository.directory=../flowfile_repository
nifi.flowfile.repository.checkpoint.interval=20 secs
nifi.flowfile.repository.always.sync=false
nifi.flowfile.repository.encryption.key.provider.implementation=
nifi.flowfile.repository.encryption.key.provider.location=
nifi.flowfile.repository.encryption.key.provider.password=
nifi.flowfile.repository.encryption.key.id=
nifi.flowfile.repository.encryption.key=
nifi.flowfile.repository.retain.orphaned.flowfiles=true
```



```
. bin
conf
content_repository
database_repository
docs
etc
lib
logs
provenance_repository
run
state
work
LICENSE
NOTICE
README
```

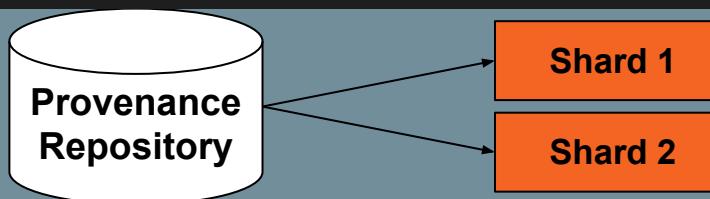
# \_ Conf Directory - nifi.properties

```
# Content Repository
nifi.content.repository.implementation=org.apache.nifi.controller.repository.FileSystemRepository
nifi.content.claim.max.appendable.size=1 MB
nifi.content.repository.directory.default=/content_repository
nifi.content.repository.archive.max.retention.period=7 days
nifi.content.repository.archive.max.usage.percentage=50%
nifi.content.repository.archive.enabled=true
nifi.content.repository.always.sync=false
nifi.content.viewer.url=../nifi-content-viewer/
nifi.content.repository.encryption.key.provider.implementation=
nifi.content.repository.encryption.key.provider.location=
nifi.content.repository.encryption.key.provider.password=
nifi.content.repository.encryption.key.id=
nifi.content.repository.encryption.key=
```



# Conf Directory - nifi.properties

```
# Persistent Provenance Repository Properties
nifi.provenance.repository.directory.default=./provenance_repository
nifi.provenance.repository.max.storage.time=30 days
nifi.provenance.repository.max.storage.size=10 GB
nifi.provenance.repository.rollover.time=10 mins
nifi.provenance.repository.rollover.size=100 MB
nifi.provenance.repository.query.threads=2
nifi.provenance.repository.index.threads=2
nifi.provenance.repository.compress.on.rollover=true
nifi.provenance.repository.always.sync=false
# comma-separated list of fields. Fields that are not indexed will not be searchable. Valid fields
# EventType, FlowFileUUID, Filename, TransitURI, ProcessorID, AlternateIdentifierURI, Relationship,
nifi.provenance.repository.indexed.fields=EventType, FlowFileUUID, Filename, ProcessorID, Relations
# FlowFile Attributes that should be indexed and made searchable. Some examples to consider are fi
nifi.provenance.repository.indexed.attributes=
# Large values for the shard size will result in more Java heap usage when searching the Provenance
# but should provide better performance
nifi.provenance.repository.index.shard.size=500 MB
# Indicates the maximum length that a FLOWFILE attribute can be when retrieving a Provenance Event
# the repository. If the length of any attribute exceeds this value, it will be truncated when the
nifi.provenance.repository.max.attribute.length=65536
nifi.provenance.repository.concurrent.merge.threads=2
```



Search Events

Event Type

Exclude from search results ?

FlowFile UUID

Exclude from search results ?

Filename

Exclude from search results ?

Component ID

Exclude from search results ?

Relationship

Exclude from search results ?

Details

Exclude from search results ?

Twitter.msg

Exclude from search results ?

Start Date ? Start Time (UTC) ?  
07/29/2016 00:00:00

End Date ? End Time (UTC) ?  
07/29/2016 23:59:59

Minimum File Size ? Maximum File Size ?

Search Location  
cluster

CANCEL SEARCH

# – Conf Directory - nifi.properties

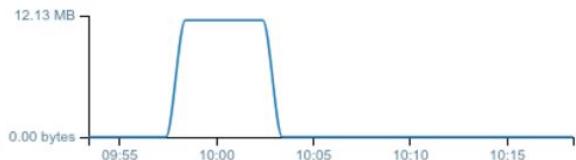
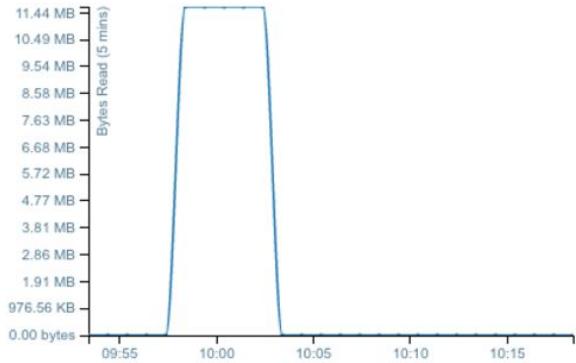
```
# Component and Node Status History Repository  
nifi.components.status.repository.implementation=nifi  
  
# Volatile Status History Repository Properties  
nifi.components.status.repository.buffer.size=14  
nifi.components.status.snapshot.frequency=1 min  
  
# QuestDB Status History Repository Properties  
nifi.status.repository.questdb.persist.node.days=1  
nifi.status.repository.questdb.persist.component=true  
nifi.status.repository.questdb.persist.location=/tmp/nifi-status  
  
# Site to Site properties  
nifi.remote.input.host=  
nifi.remote.input.secure=false  
nifi.remote.input.socket.port=4443  
nifi.remote.input.http.enabled=true  
nifi.remote.input.http.transaction.ttl=30 sec  
nifi.remote.contents.cache.expiration=30 secs
```

## Status History

### Status History

**Id**  
0159101d-c1a9-1856-9d74-fce1209ceaa6  
**Group Id**  
b0b5b671-0158-1000-8689-368ab328cf1e  
**Name**  
PutFile  
**Type**  
PutFile  
**Start**  
12/29/2016 09:53:22.946  
**End**  
12/29/2016 10:18:23.260  
**NiFi**  
**Min / Max / Mean**  
0.00 bytes / 11.66 MB / 2.24 MB  
 **NiFi**

Bytes Read (5 mins)



Last updated: 10:19:19 EST

CLOSE

NiFi 1

# \_ Conf Directory - nifi.properties

```
# web properties #
#####
# For security, NiFi will present the UI on 127.0.0.1 and only be accessible through this loopback interface.
# Be aware that changing these properties may affect how your instance can be accessed without any restriction.
# We recommend configuring HTTPS instead. The administrators guide provides instructions on how to do this.
```

```
nifi.web.http.host=
nifi.web.http.port=8080
nifi.web.http.network.interface.default=
```

HTTP NiFi

```
#####
nifi.web.https.host=
nifi.web.https.port=
nifi.web.https.network.interface.default=
nifi.web.jetty.working.directory= /work/jetty
nifi.web.jetty.threads=200
nifi.web.max.header.size=16 KB
nifi.web.proxy.context.path=
nifi.web.proxy.host=
nifi.web.max.content.size=
nifi.web.max.requests.per.second=30000
nifi.web.request.timeout=60 secs
nifi.web.request.ip.whitelist=
nifi.web.should.send.server.version=true
```

HTTPs NiFi

# \_ Conf Directory - nifi.properties

```
# security properties #
nifi.sensitive.props.key=
nifi.sensitive.props.key.protected=
nifi.sensitive.props.algorithm=NIFI PB
nifi.sensitive.props.provider=BC
nifi.sensitive.props.additional.keys=
```

Configure Controller Service

SETTINGS PROPERTIES COMMENTS

Required field

Property	Value
Database Connection URL	?
Database Driver Class Name	?
Database Driver Location(s)	?
Kerberos Credentials Service	?
Kerberos Principal	?
Kerberos Password	?
Database User	?
Password	?
Max Wait Time	500 millis
Max Total Connections	?
Validation query	?
Minimum Idle Connections	?
Max Idle Connections	?
Max Connection Lifetime	?

CANCEL

APPLY

# Understanding NiFi Secure Properties

# HTTPS (Secure Connection)

## What is SSL Certificate

Authentication mechanism used to identify the identity of who they say they are

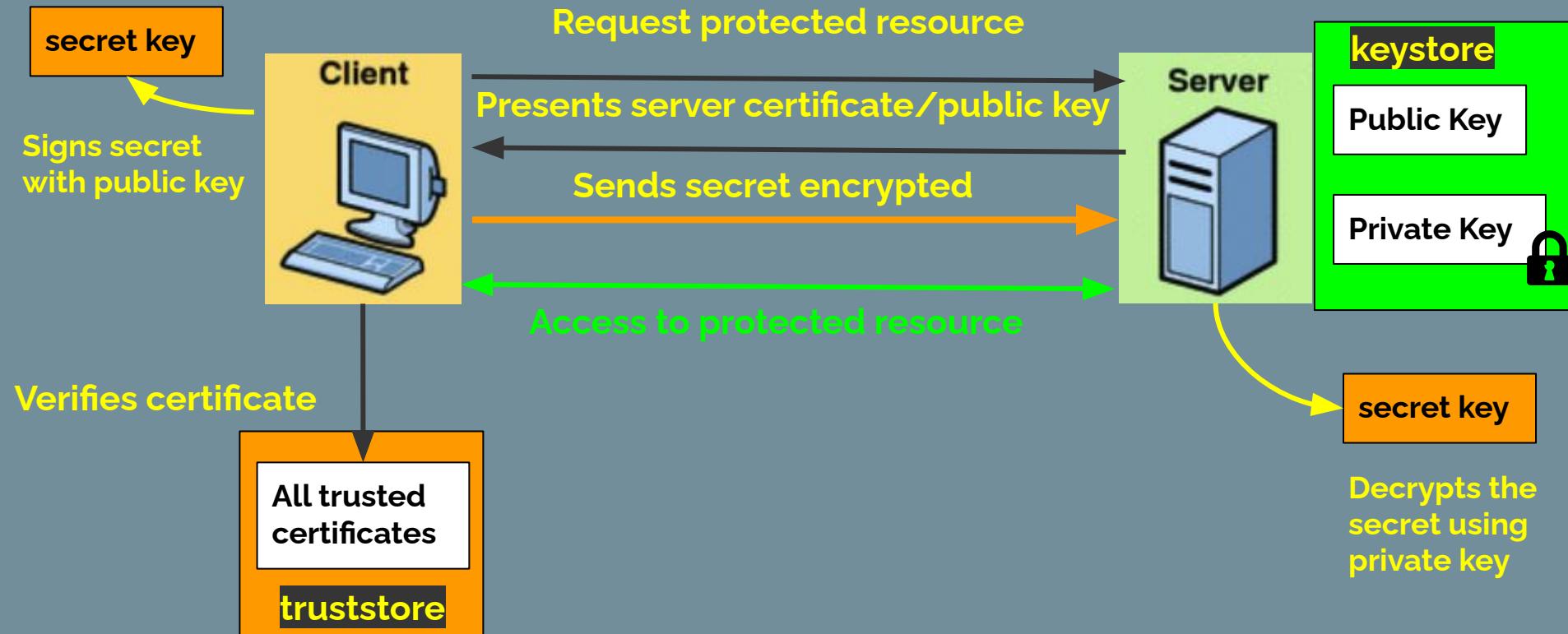
Generally, these certificate based authentication is for Web sites to provide their identity during SSL connection established

Digitally Signed by certificate authority (CA)

# HTTPS



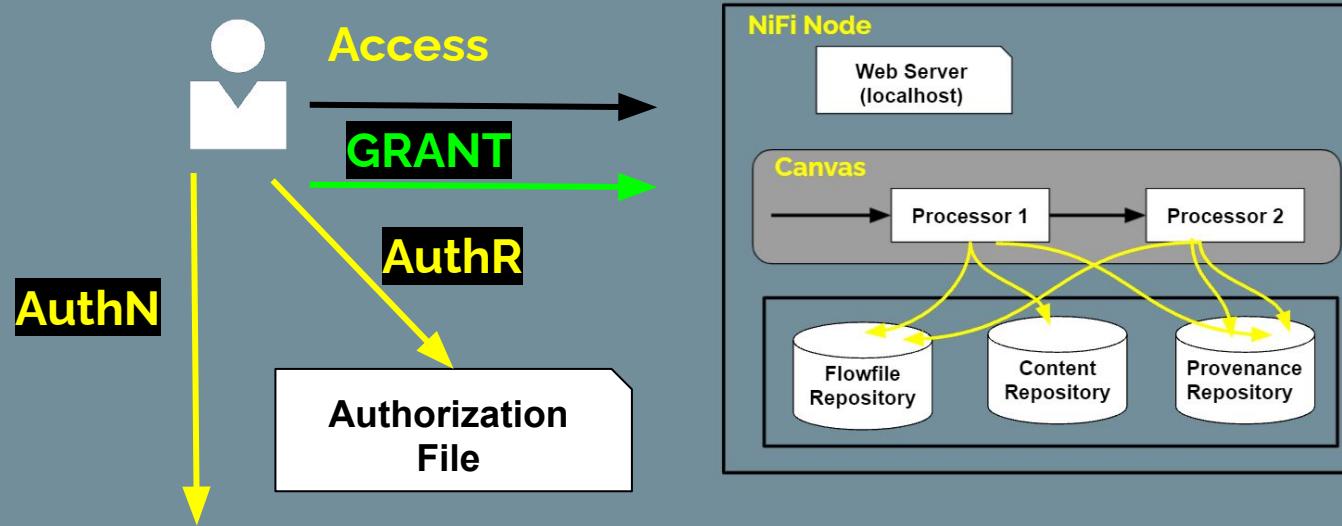
# HTTPS



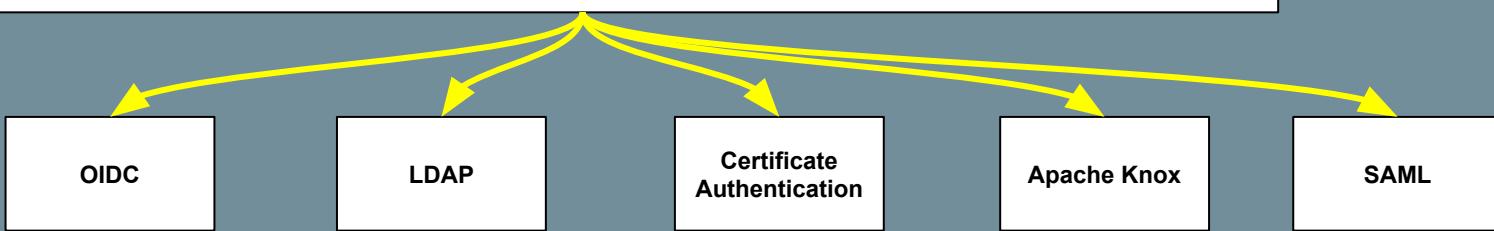
# \_ Conf Directory - nifi.properties

```
nifi.security.autoreload.enabled=false
nifi.security.autoreload.interval=10 secs
nifi.security.keystore=
nifi.security.keystoreType=
nifi.security.keystorePasswd=
nifi.security.keyPasswd=
nifi.security.truststore=
nifi.security.truststoreType=
nifi.security.truststorePasswd=
nifi.security.user.authorizer=
nifi.security.allow.anonymous.authentication=false
nifi.security.user.login.identity.provider=
nifi.security.ocsp.responder.url=
nifi.security.ocsp.responder.certificate=
```

# – NiFi Login Process



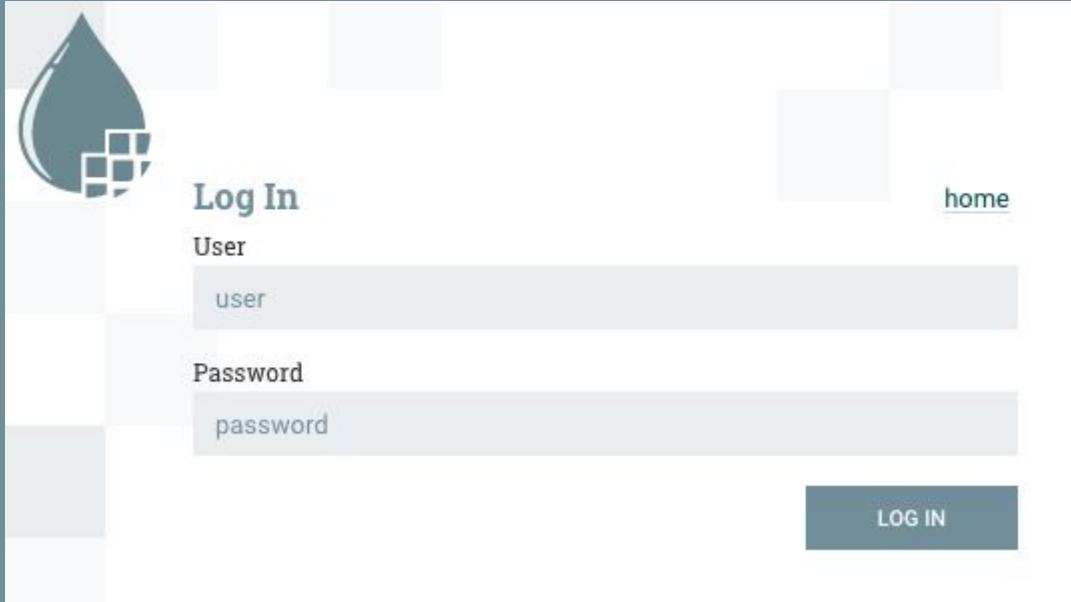
## Authentication Layer



# NiFi Authentication Types

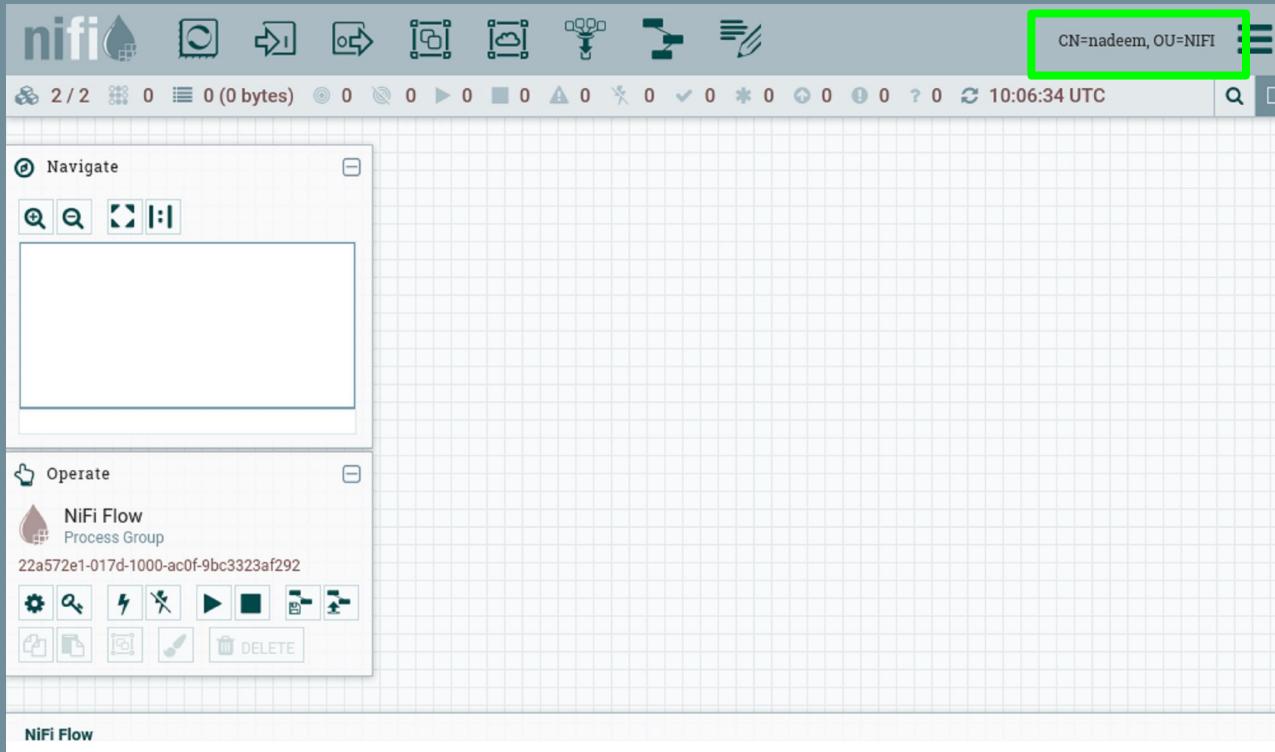
# NiFi Authentication Types

## Single User Credentials / LDAP User Authentication



# NiFi Authentication Types

## Certificate Authentication



# NiFi Authentication Types

- OIDC (OpenID Connect)
- Apache Knox
- SAML (Security Assertion Markup Language)

# \_ Conf Directory - nifi.properties

## OIDC (OpenID Connect) Authentication

```
# OpenId Connect SSO Properties #
nifi.security.user.oidc.discovery.url=
nifi.security.user.oidc.connect.timeout=5 secs
nifi.security.user.oidc.read.timeout=5 secs
nifi.security.user.oidc.client.id=
nifi.security.user.oidc.client.secret=
nifi.security.user.oidc.preferred.jwsalgorithm=
nifi.security.user.oidc.additional.scopes=
nifi.security.user.oidc.claim.identifying.user=
nifi.security.user.oidc.fallback.claims.identifying.user=
```

# \_ Conf Directory - nifi.properties

## Apache Knox Authentication

```
# Apache Knox SSO Properties #
nifi.security.user.knox.url=
nifi.security.user.knox.publicKey=
nifi.security.user.knox.cookieName=hadoop-jwt
nifi.security.user.knox.audiences=
```

# \_ Conf Directory - nifi.properties

## SAML (Security Assertion Markup Language)

```
# SAML Properties #
nifi.security.user.saml.idp.metadata.url=
nifi.security.user.saml.sp.entity.id=
nifi.security.user.saml.identity.attribute.name=
nifi.security.user.saml.group.attribute.name=
nifi.security.user.saml.metadata.signing.enabled=false
nifi.security.user.saml.request.signing.enabled=false
nifi.security.user.saml.want.assertions.signed=true
nifi.security.user.saml.signature.algorithm=http://www.w3.org/2001/04/xmldsig-more#rsa-sha256
nifi.security.user.saml.signature.digest.algorithm=http://www.w3.org/2001/04/xmlenc#sha256
nifi.security.user.saml.message.logging.enabled=false
nifi.security.user.saml.authentication.expiration=12 hours
nifi.security.user.saml.single.logout.enabled=false
nifi.security.user.saml.http.client.truststore.strategy=JDK
nifi.security.user.saml.http.client.connect.timeout=30 secs
nifi.security.user.saml.http.client.read.timeout=30 secs
```

# \_ Conf Directory - nifi.properties

## Cluster Protocol

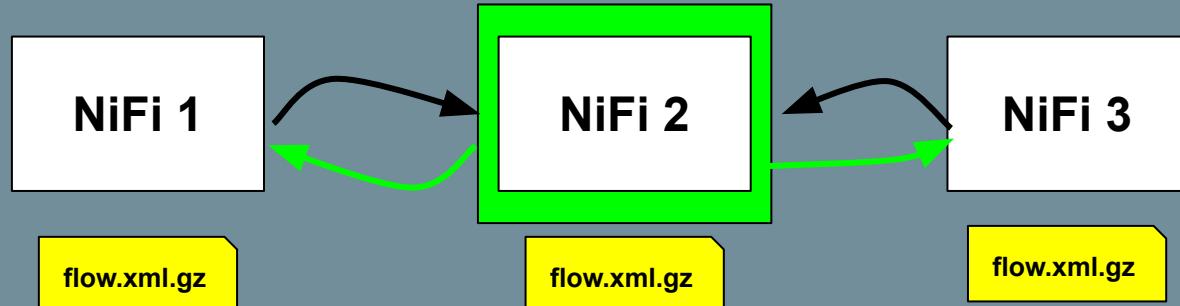
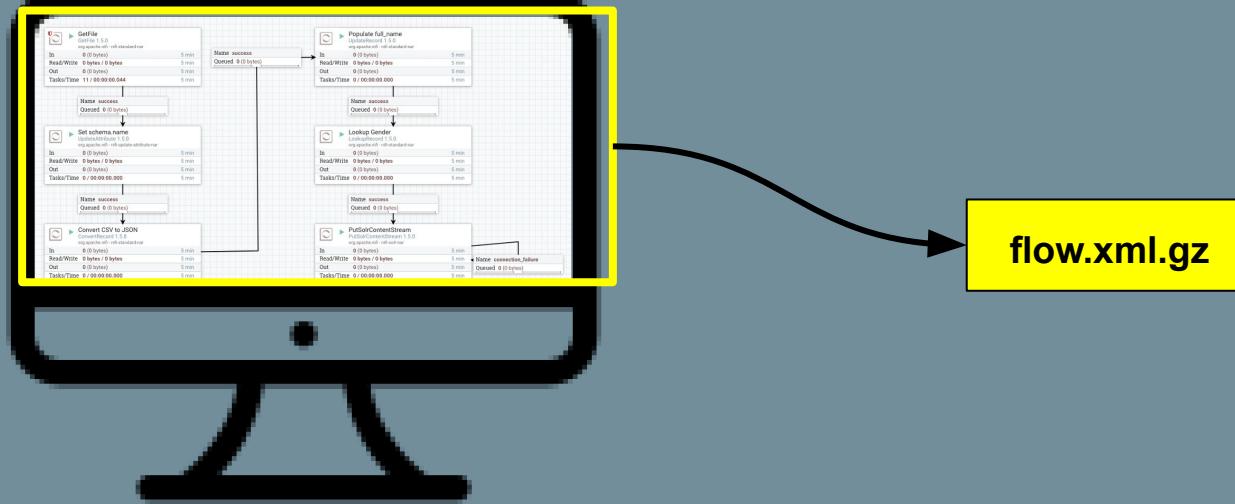
```
# cluster common properties (all nodes must have same values) #
nifi.cluster.protocol.heartbeat.interval=5 sec
nifi.cluster.protocol.heartbeat.missable.max=8
nifi.cluster.protocol.is.secure=false
```

# \_ Conf Directory - nifi.properties

## Cluster Node Properties

```
# cluster node properties (only configure for cluster nodes) #
nifi.cluster.is.node=false
nifi.cluster.node.address=
nifi.cluster.node.protocol.port=
nifi.cluster.node.protocol.threads=10
nifi.cluster.node.protocol.max.threads=50
nifi.cluster.node.event.history.size=25
nifi.cluster.node.connection.timeout=5 sec
nifi.cluster.node.read.timeout=5 sec
nifi.cluster.node.max.concurrent.requests=100
nifi.cluster.firewall.file=
nifi.cluster.flow.election.max.wait.time=5 mins
nifi.cluster.flow.election.max.candidates=
```

# - NiFi Flow Election



# \_ Conf Directory - nifi.properties

## Zookeeper Properties

```
# zookeeper properties, used for cluster management #
nifi.zookeeper.connect.string=
nifi.zookeeper.connect.timeout=10 secs
nifi.zookeeper.session.timeout=10 secs
nifi.zookeeper.root.node=/nifi
nifi.zookeeper.client.secure=false
nifi.zookeeper.security.keystore=
nifi.zookeeper.security.keystoreType=
nifi.zookeeper.security.keystorePasswd=
nifi.zookeeper.security.truststore=
nifi.zookeeper.security.truststoreType=
nifi.zookeeper.security.truststorePasswd=
```

# \_ Conf Directory - nifi.properties

## Kerberos Authentication

```
# kerberos #
nifi.kerberos.krb5.file=

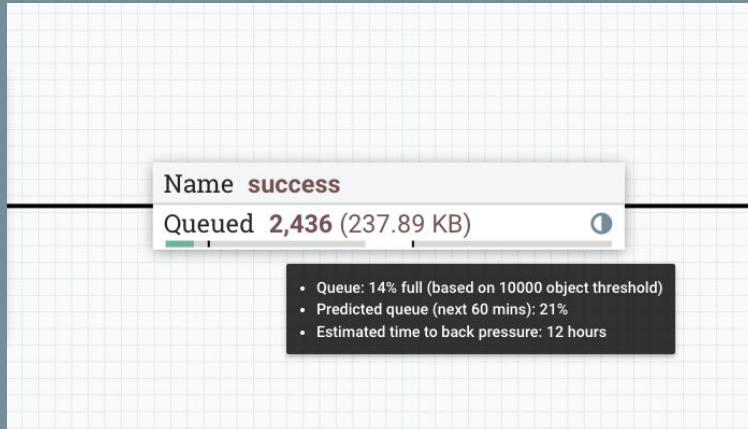
# kerberos service principal #
nifi.kerberos.service.principal=
nifi.kerberos.service.keytab.location=

# kerberos spnego principal #
nifi.kerberos.spnego.principal=
nifi.kerberos.spnego.keytab.location=
nifi.kerberos.spnego.authentication.expiration=12 hours
```

# \_ Conf Directory - nifi.properties

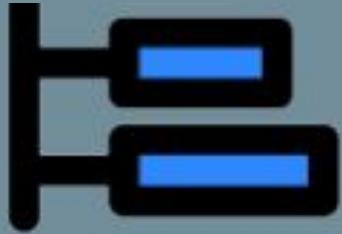
## Connection Queue Analytics

```
# analytics properties #
nifi.analytics.predict.enabled=false
nifi.analytics.predict.interval=3 mins
nifi.analytics.query.interval=5 mins
nifi.analytics.connection.model.implementation=org.apache.nifi.controller.status.analytics.models.OrdinaryLeastSquares
nifi.analytics.connection.model.score.name=rSquared
nifi.analytics.connection.model.score.threshold=.90
```



# Securing NiFi

# Securing NiFi



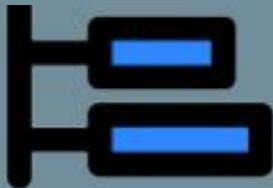
Standalone

Cluster

# Hands-On

## Securing NiFi

### Standalone



Single User Auth  
LDAP Auth  
(Multiple Users)

### Cluster



**2 node cluster**  
LDAP Auth  
(Multiple Users)

# Understanding Course Material

 [naddym / nifi-course-material](#) 

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Security](#) [Insights](#) [Settings](#)

 main  2 branches  0 tags [Go to file](#)

File	Description
 installation	Update slides and some improvisation
 monitoring	NiFi 1.15.3 support
 presentation-slides	Update slides and some improvisation
 productionizing	Add course material
 templates	Update material
 usecase	Add OpenID auth for secure cluster
 .gitignore	Update material
 README.md	Update material to remove bespoke changes
 best-practices-and-general-guideli...	Add course material

NiFi is Secure  
(by default)

Single User Auth

# Lab: Securing NiFi (Single User Auth)

# Lab: Securing NiFi (Single User Auth)

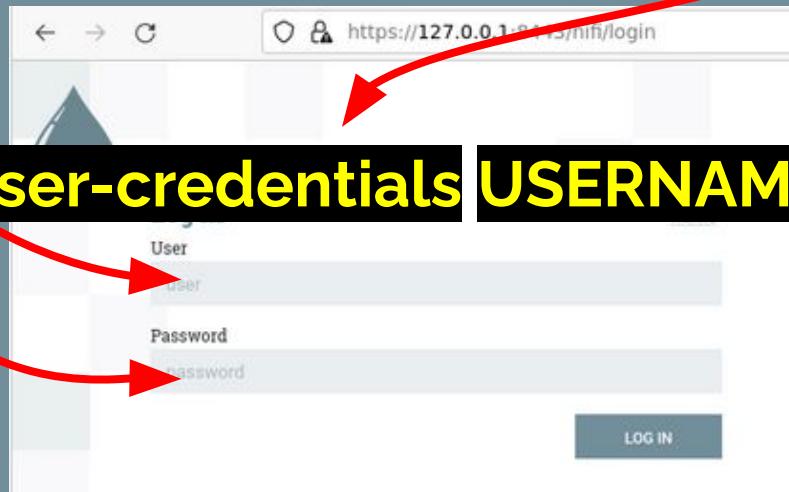
**Explanation**

# Single User Authentication

```
-the-world time = 14 milliseconds), max transaction ID -1
2021-12-19 10:09:09,894 INFO [main] o.a.n.c.r.WriteAheadFlowFileRepository Successfully restored 0 FlowFiles and 0 Swap Files
2021-12-19 10:09:09,940 INFO [main] o.apache.nifi.controller.FlowController Performed initial validation of all components in 1 milliseconds
2021-12-19 10:09:09,940 INFO [main] o.apache.nifi.controller.FlowController Starting 0 processors/ports/funnels
2021-12-19 10:09:09,940 INFO [main] o.apache.nifi.controller.FlowController Started 0 Remote Group Ports transmitting
2021-12-19 10:09:09,941 INFO [main] o.a.n.w.c.ApplicationStartupContextListener Flow Controller started successfully.
2021-12-19 10:09:09,962 INFO [main] o.a.n.a.s.u.SingleUserLoginIdentityProvider

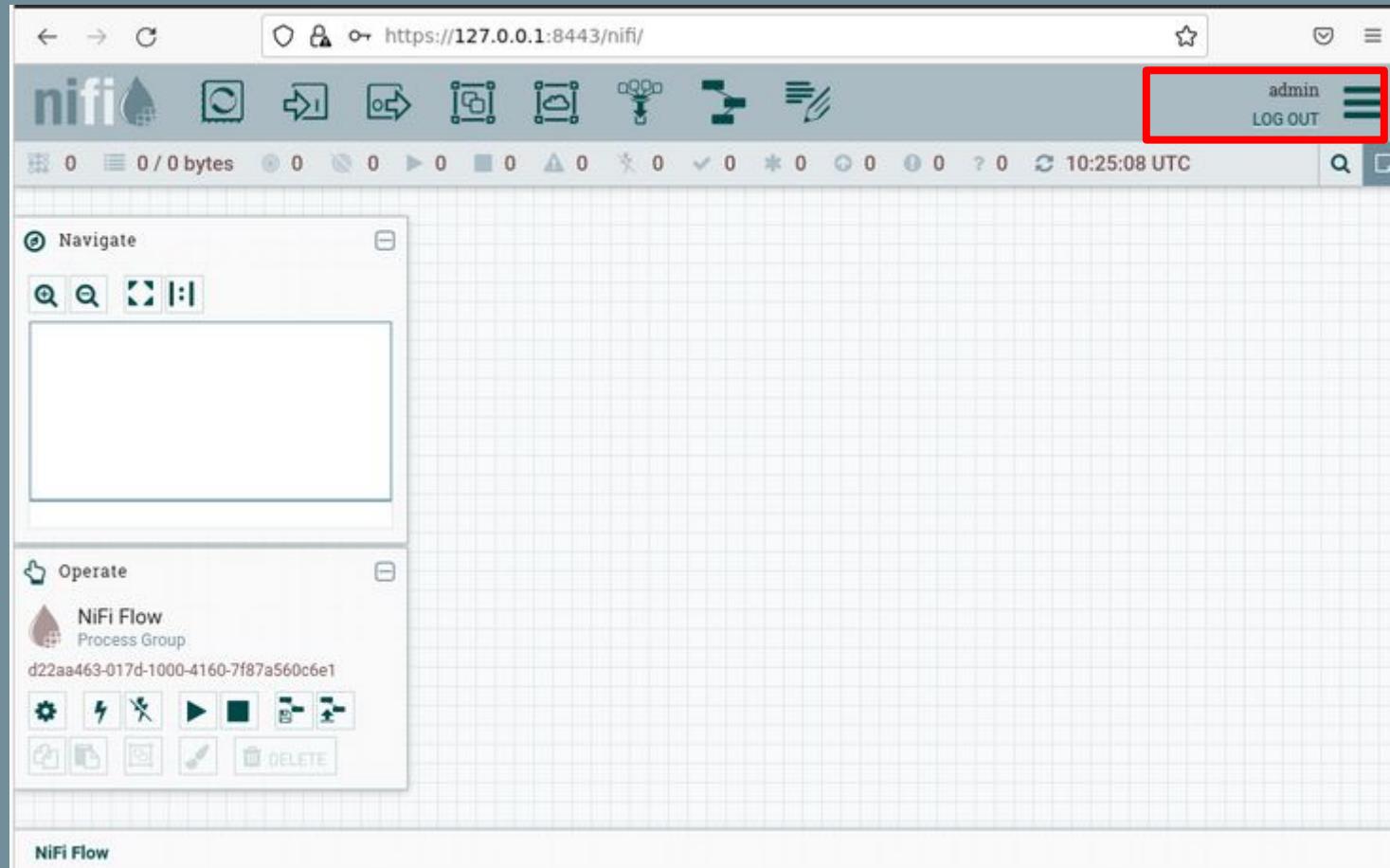
Generated Username [f40d380e-71c5-4083-825d-0e7886421002]
Generated Password [+d0bmVF0l+9pFAnbe8Va5IRDaj6tF4h0]

2021-12-19 10:09:09,963 INFO [main] o.a.n.a.s.u.SingleUserLoginIdentityProvider Run the following command to change credentials: nifi.sh set-single-user-credentials USERNAME PASSWORD
2021-12-19 10:09:10,438 INFO [main] o.a.n.a.s.u.SingleUserLoginIdentityProvider Updating Login Identity Providers Configuration [./conf/login-identity-providers.xml]
```



**nifi.sh set-single-user-credentials USERNAME PASSWORD**

# Single User Authentication



# Single User Authentication

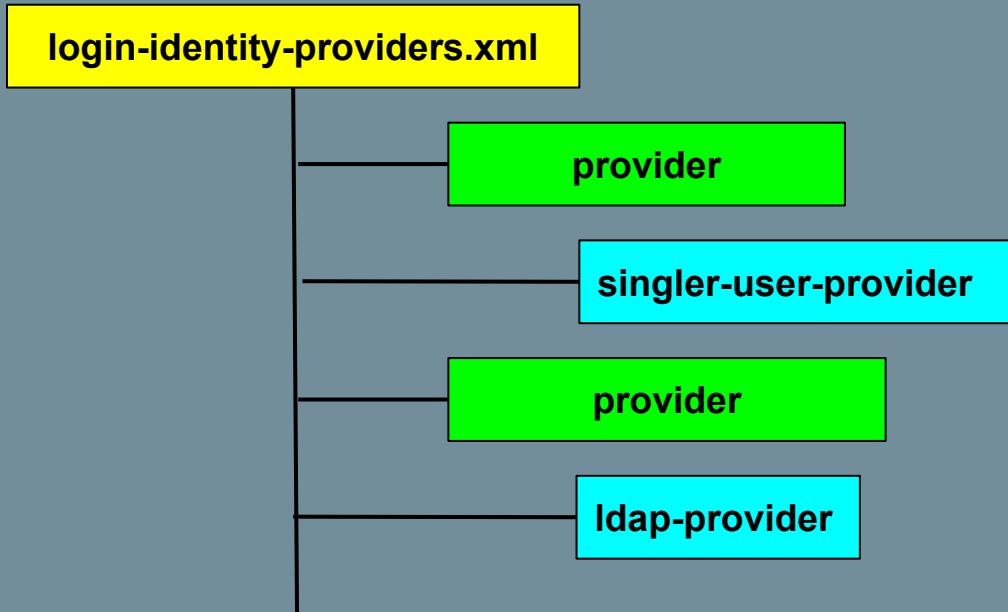
```
nifi.security.keystore=./conf/keystore.p12
nifi.security.keystoreType=PKCS12
nifi.security.keystorePasswd=ba3706c7b7267153437514a174542497
nifi.security.keyPasswd=ba3706c7b7267153437514a174542497
nifi.security.truststore=./conf/truststore.p12
nifi.security.truststoreType=PKCS12
nifi.security.truststorePasswd=ad0e74fdd42691363a63cd51b75e3cf8
nifi.security.user.authorizer=single-user-authorizer
nifi.security.allow.anonymous.authentication=false
nifi.security.user.login.identity.provider=single-user-provider
nifi.security.ocsp.responder.url=
nifi.security.ocsp.responder.certificate=
```



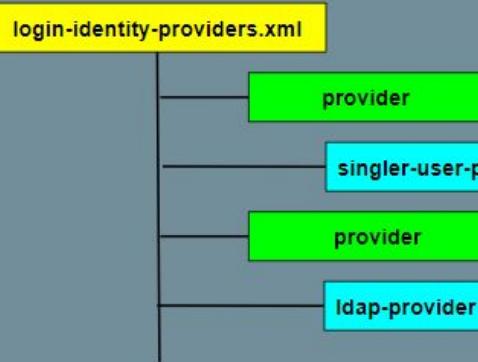
```
<authorizer>
    <identifier>single-user-authorizer</identifier>
    <class>org.apache.nifi.authorization.single.user.SingleUserAuthorizer</class>
</authorizer>
```

# Login Identity Providers

In NiFi, login via username/password is controlled by  
**login-identity-providers.xml**



# Login Identity Providers



**single user credentials**

```
<provider>
  <identifier>single-user-provider</identifier>
  <class>org.apache.nifi.authentication.single.user.SingleUserLoginIdentityProvider</class>
  <property name="Username">admin</property>
  <property name="Password">$2b$12$4CKSdgM3ZgemHd3.VGczGe1pHbgNulhqAqhQTmO.ia1sm5UChkNmy</property>
</provider>
```

# Single User Authentication

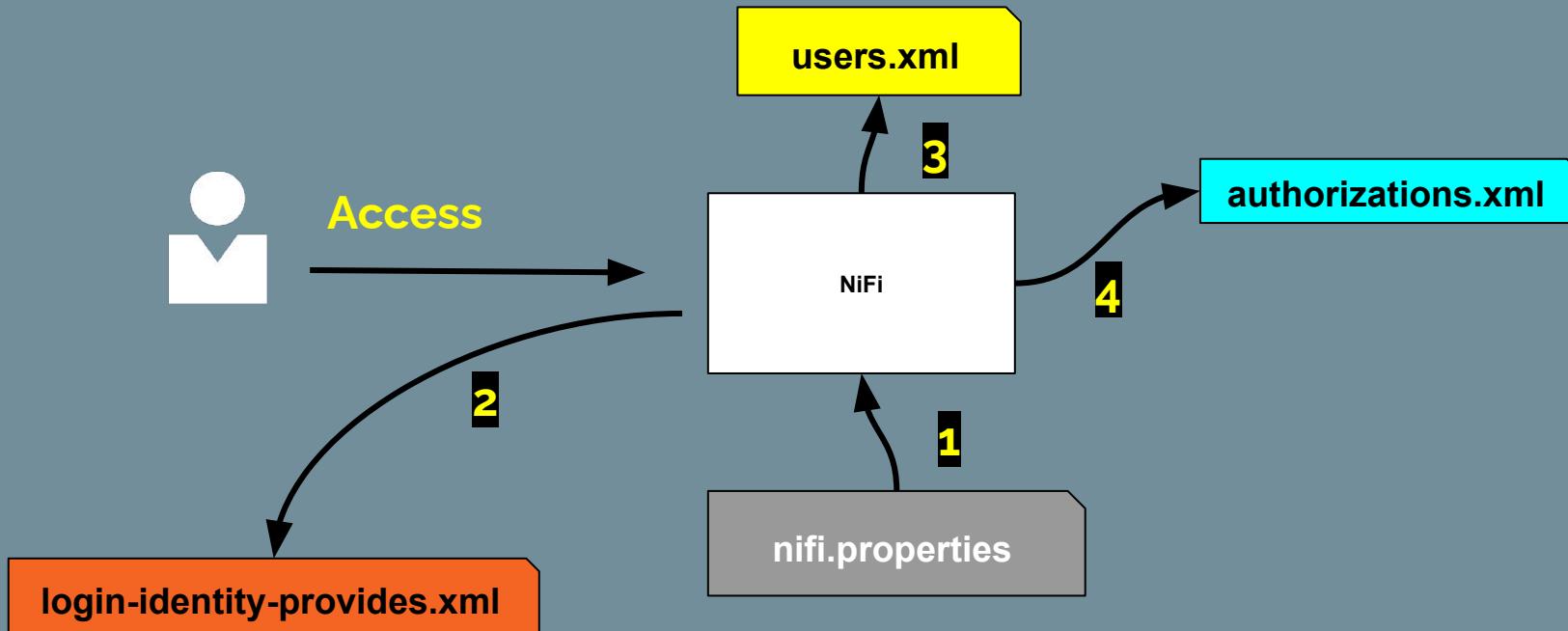
## nifi.properties

```
nifi.security.truststoreType=PKCS12  
nifi.security.truststorePasswd=ad0e74fdd42691363a63cd51b75e3cf8  
nifi.security.user.authorizer=single-user-authorizer  
nifi.security.allow.anonymous.authentication=false  
nifi.security.user.login.identity.provider=single-user-provider  
nifi.security.ocsp.responder.url=
```

## login-identity-providers.xml

```
<provider>  
    <identifier>single-user-provider</identifier>  
    <class>org.apache.nifi.authentication.single.user.SingleUserLoginIdentityProvider</class>  
    <property name="Username">admin</property>  
    <property name="Password">$2b$12$4CKSdgM3ZgemHd3.VGczGe1pHbgNulhqAqhQTmO.ia1sm5UCHkNmy</property>  
</provider>
```

# – How conf files get used during authentication & authorization



# Securing NiFi (LDAP Auth)

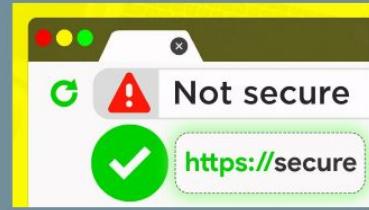
**Multi Users**

# – Steps to secure NiFi with LDAP

## Steps

1. Generate KeyStore and Truststore keys using NiFi Toolkit
2. Create users in LDAP Server
3. Configure LDAP Configuration in NiFi
4. Start NiFi

# Understanding Certificate Authority



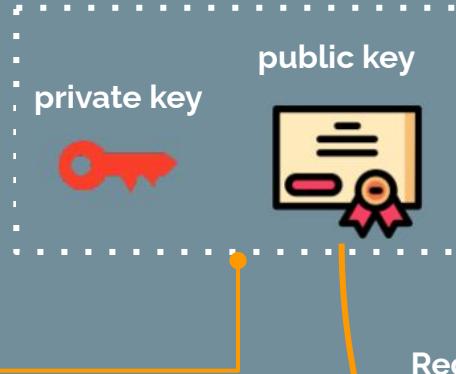
https:/

www.mysite.com

All trusted  
CA's



Certificate signed!!



Request to Sign the  
certificate

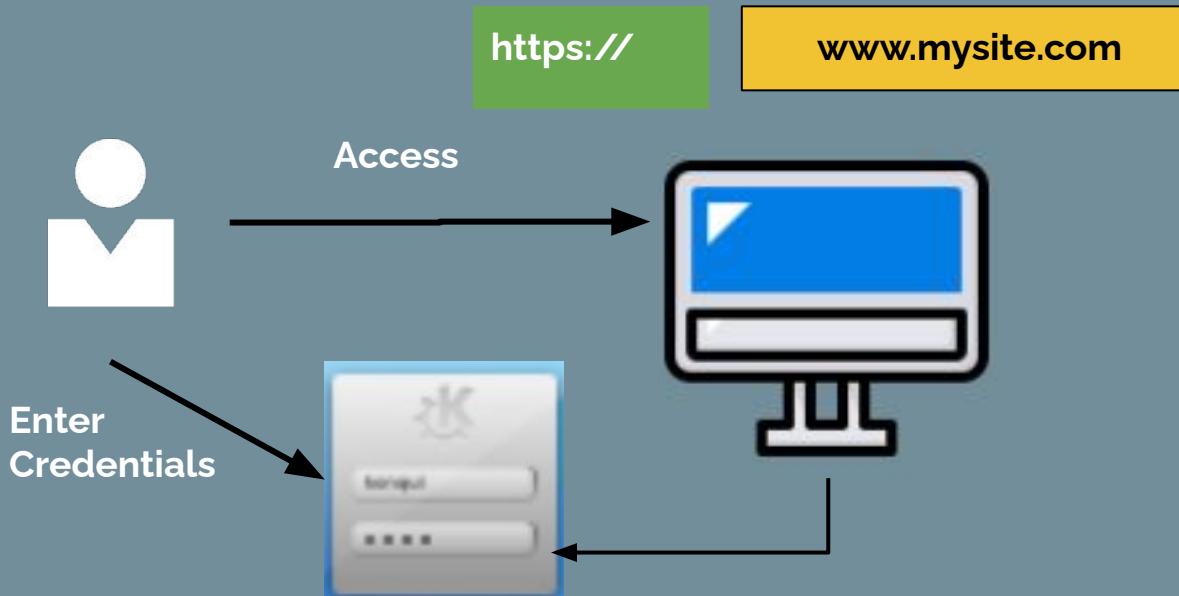


Certificate  
Authority  
(CA)

Certificate Issuer

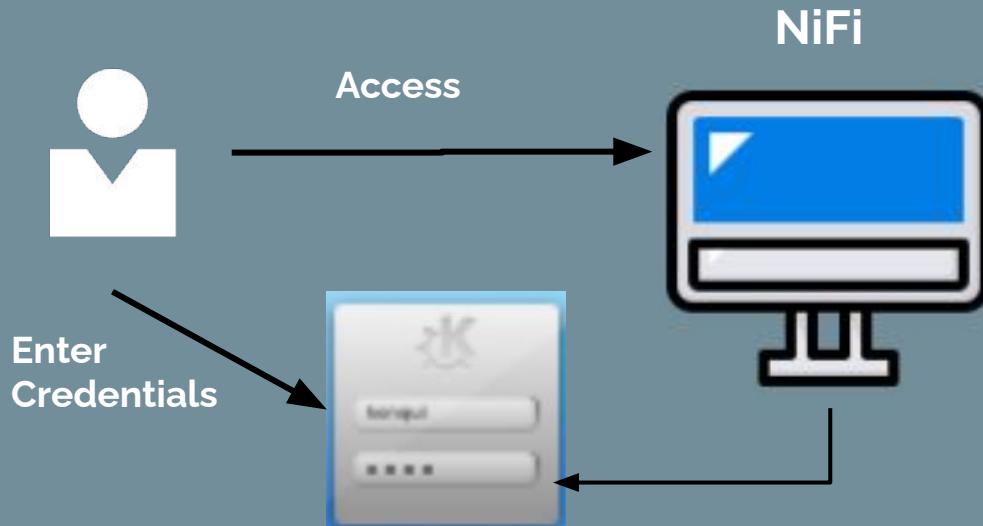
# \_ Understanding Certificate Authority

## User Authentication

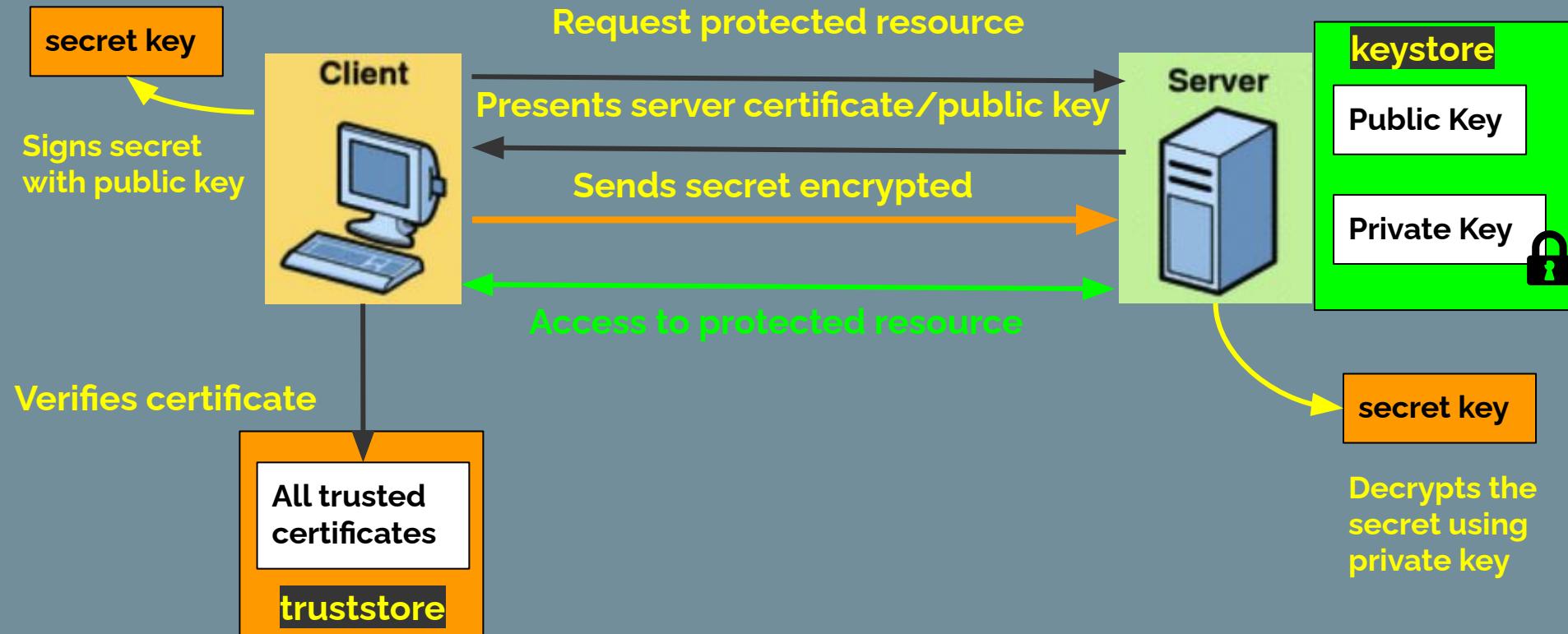


# \_ Understanding Certificate Authority

## LDAP Authentication



# HTTPS



# Generating Keystore/Trustore

1.Generate keystore(At server):

```
keytool -genkey -alias bmc -keyalg RSA -keystore KeyStore.jks -keysize 2048
```

2.Generate new ca-cert and ca-key:

```
openssl req -new -x509 -keyout ca-key -out ca-cert
```

3.Extracting cert/creating cert sign req(csr):

```
keytool -keystore KeyStore.jks -alias bmc -certreq -file cert-file
```

4.Sign the "cert-file" and cert-signed wil be the new cert:

```
openssl x509 -req -CA ca-cert -CAkey ca-key -in cert-file -out  
cert-signed -days 365 -Ccreateserial -passin pass:yourpass
```

5.importing the ca-cert to keystore file:

```
keytool -keystore KeyStore.jks -alias CARoot -import -file ca-cert
```

6.import cert-signed to keystore:

```
keytool -keystore KeyStore.jks -alias bmc -import -file cert-signed
```

7.Copy ca-cert into client machine and generate truststore: (At client)

```
keytool -keystore truststore.jks -alias bmc -import -file ca-cert-s
```

8.Copy ca-cert into client machine and generate truststore: (At server)

```
keytool -keystore truststore.jks -alias bmc -import -file ca-cert-c
```

# NiFi ToolKit

# NiFi ToolKit

## Why NiFi Toolkit

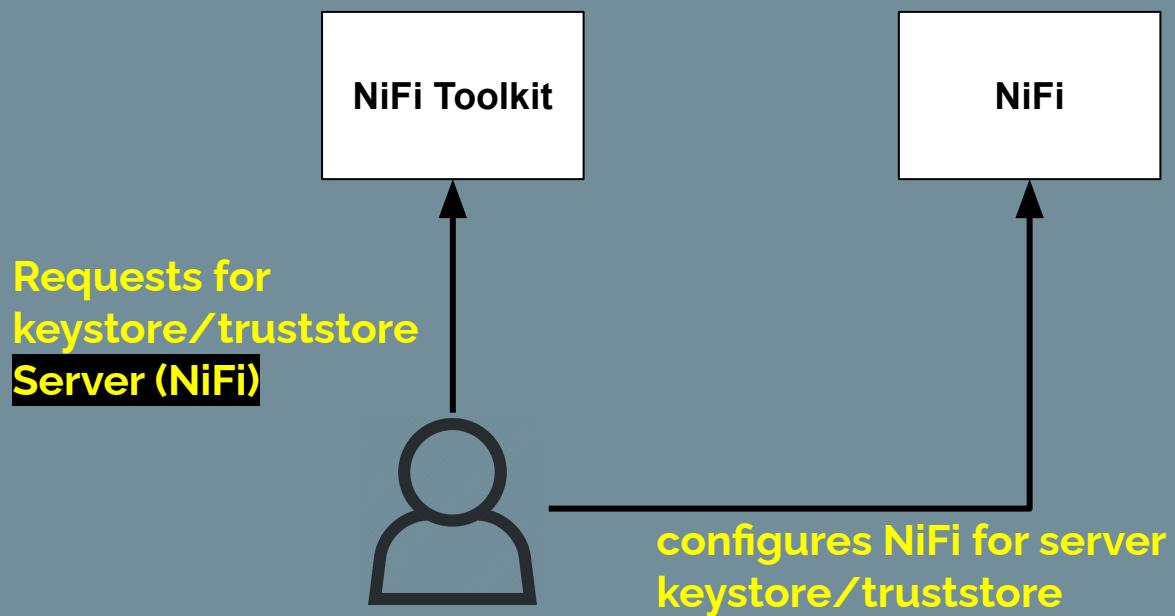
Helpful utility which contains necessary functionalities which helps user to better manage NiFi

**Flow Analyzer** : Sends reports to Admin about Max amount of data to be stored in queues.

**Node Manager** : Provides admins to manage nifi nodes such as connect/disconnect/offload/remove

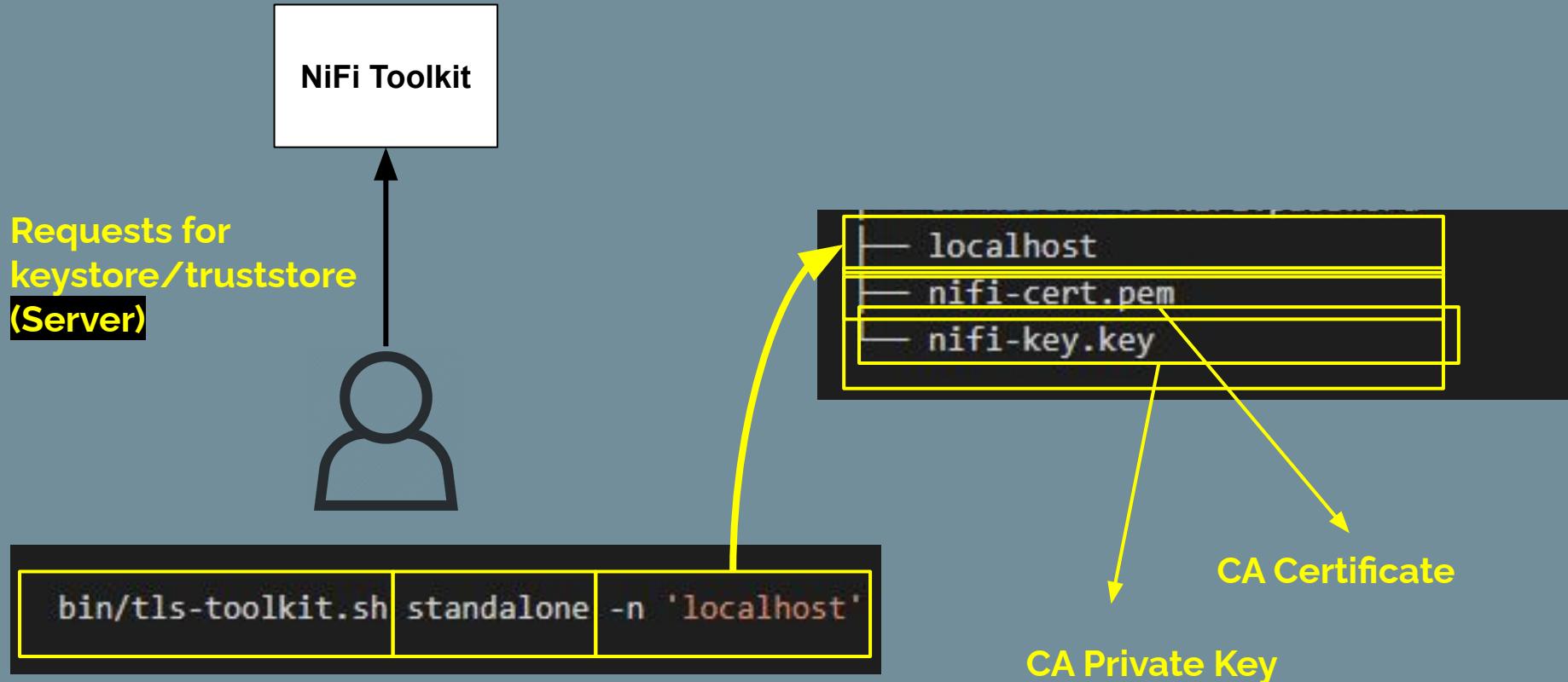
**Certificate Generation** : Provides quick way of generating certificates otherwise it would have been tedious

# Generating Keystore/Truststore for server



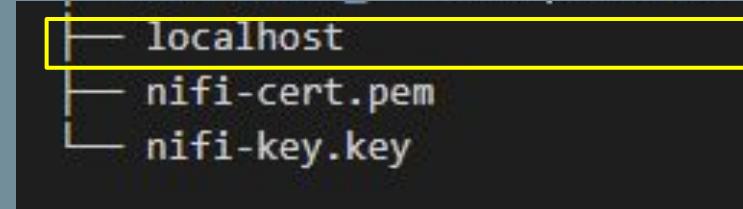
```
nifi.security.autoreload.enabled=false  
nifi.security.autoreload.interval=10 secs  
nifi.security.keystore=  
nifi.security.keystoreType=  
nifi.security.keystorePasswd=  
nifi.security.keyPasswd=  
nifi.security.truststore=  
nifi.security.truststoreType=  
nifi.security.truststorePasswd=  
nifi.security.user.authorizer=  
nifi.security.allow.anonymous.authentication=false  
nifi.security.user.login.identity.provider=  
nifi.security.ocsp.responder.url=  
nifi.security.ocsp.responder.certificate=
```

# Generating Keystore/Truststore for server



# Configure Keystore/Truststore for server

```
nifi.security.keystore=
nifi.security.keystoreType=
nifi.security.keystorePasswd=
nifi.security.keyPasswd=
nifi.security.truststore=
nifi.security.truststoreType=
nifi.security.truststorePasswd=
```



# – Steps to secure NiFi with LDAP

## Steps

1. Generate KeyStore and Truststore keys using NiFi Toolkit
2. Create users in LDAP Server
3. Configure LDAP Configuration in NiFi
4. Start NiFi

# LDAP

---

## What is LDAP

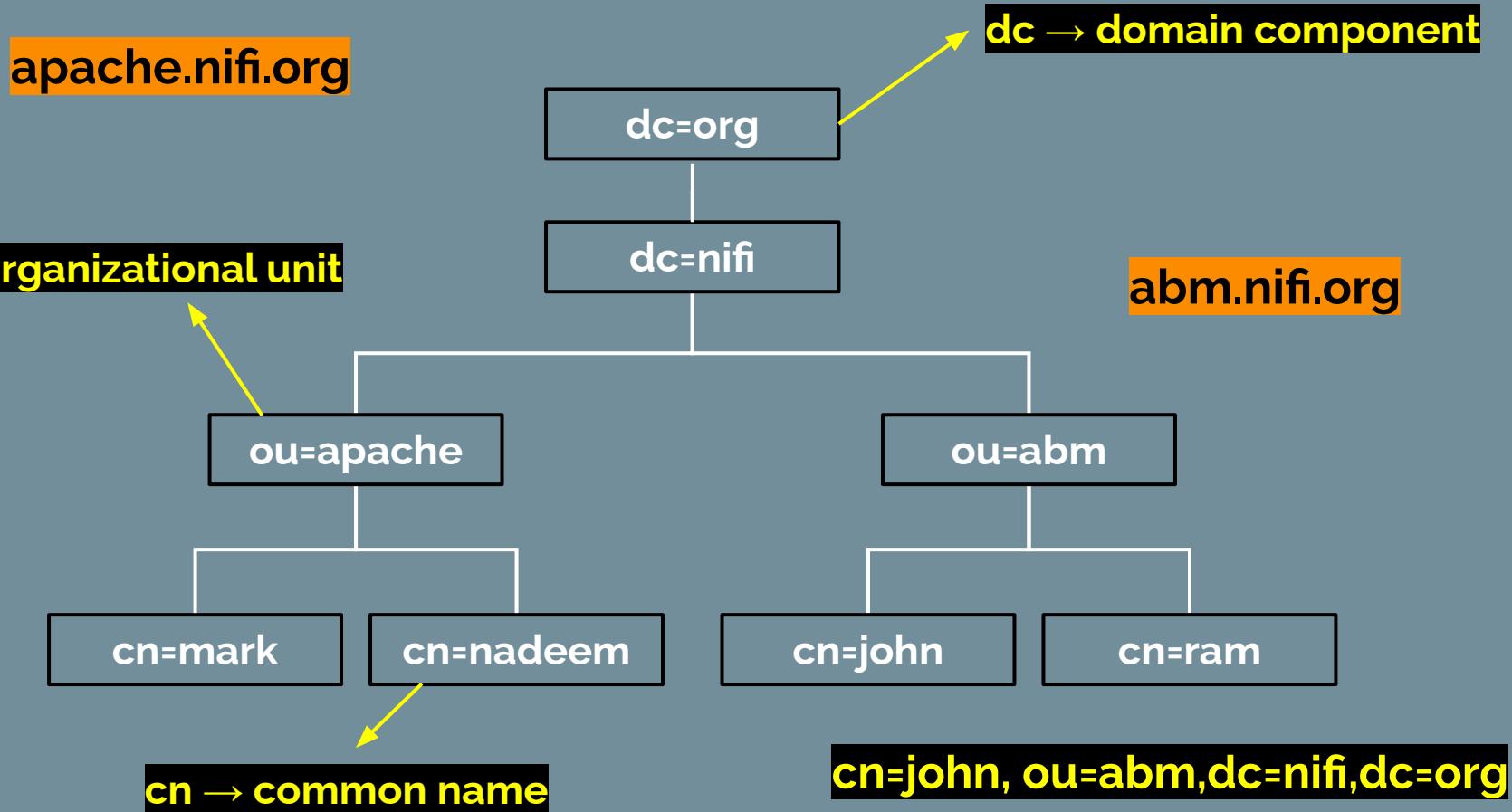
LDAP (Lightweight Directory Access Protocol) is a protocol used for directory service authentication

Used for validating users with username and password as authentication.

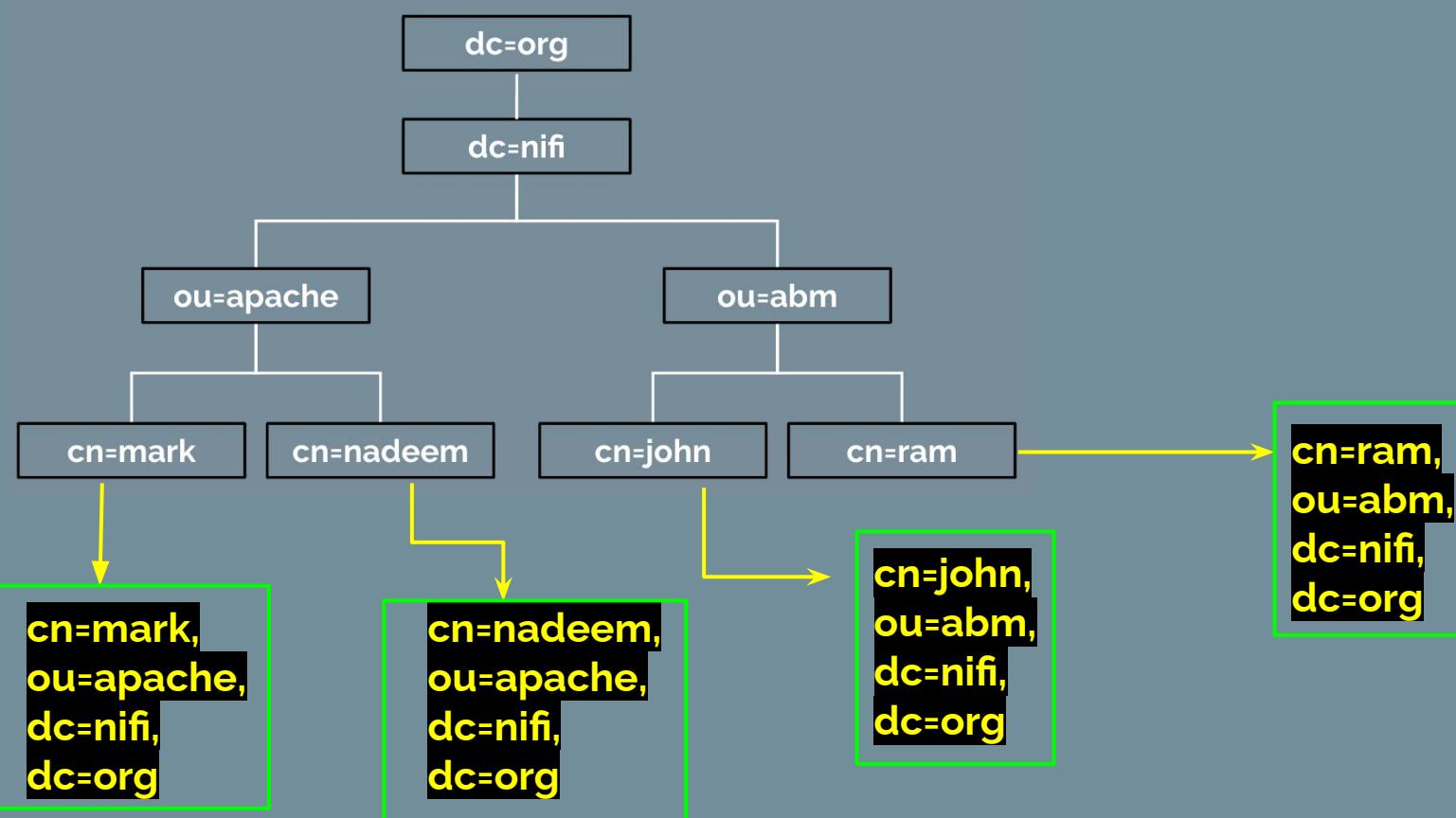
Credentials are stored in a directory structure which helps in consolidation of information at one place

I'm using OpenLDAP application to demonstrate LDAP server

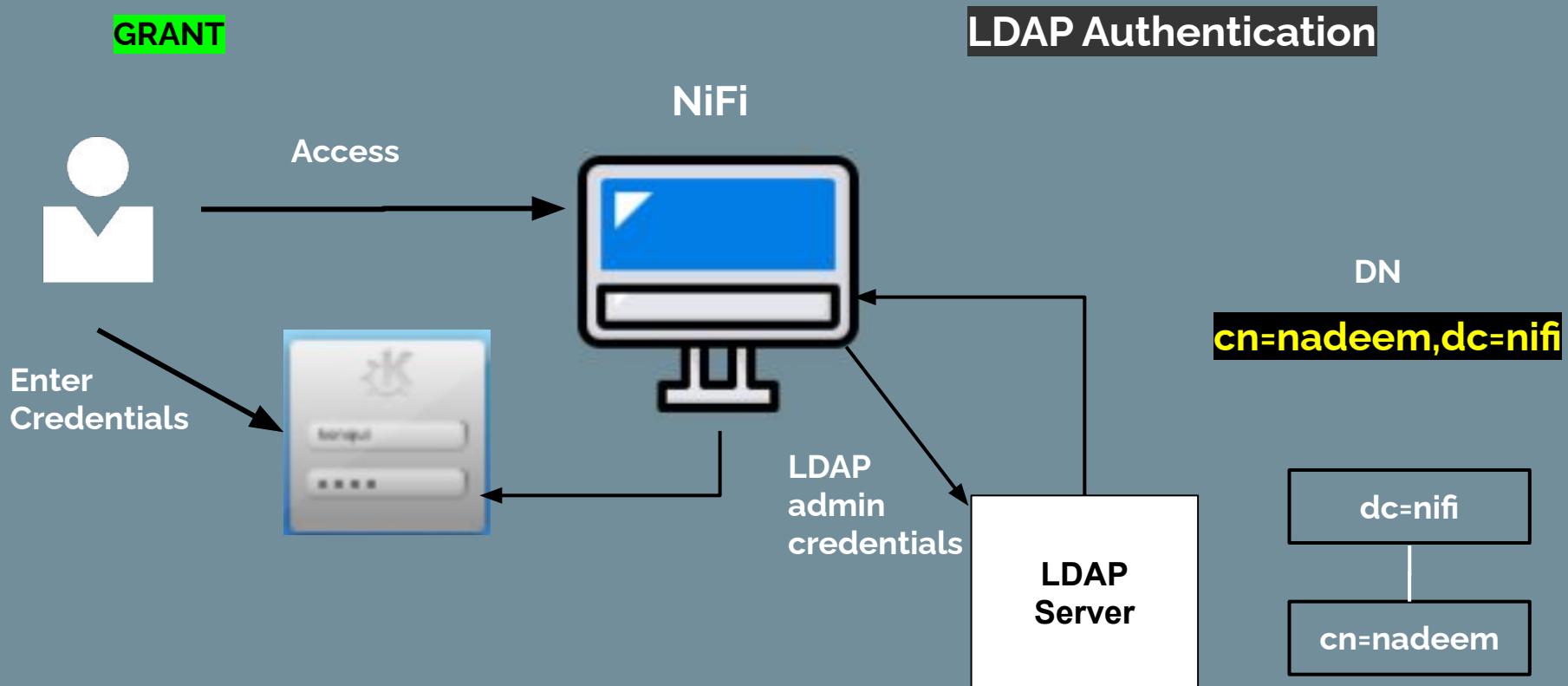
# How data is stored in LDAP



# DN = Distinguish Names



# \_ NiFi with LDAP Setup



# LDAP Authentication Configuration

# – Steps to secure NiFi with LDAP

## Steps

1. Generate KeyStore and Truststore keys using NiFi Toolkit
2. Create users in LDAP Server
3. Configure LDAP Configuration in NiFi
4. Start NiFi

# — LDAP Server

I'm using OpenLDAP application to demonstrate LDAP server

The OpenLDAP server is running on AWS.

# Configure LDAP Settings

login-identity-provider.xml file



DN

ldap://localhost:port

cn=nadeem,dc=nifi

dc=nifi

cn=nadeem

```
<provider>
    <identifier>ldap-provider</identifier>
    <class>org.apache.nifi.ldap.LdapProvider</class>
    <property name="Authentication Strategy">START_TLS</property>

    <property name="Manager DN"></property>
    <property name="Manager Password"></property>

    <property name="TLS – Keystore"></property>
    <property name="TLS – Keystore Password"></property>
    <property name="TLS – Keystore Type"></property>
    <property name="TLS – Truststore"></property>
    <property name="TLS – Truststore Password"></property>
    <property name="TLS – Truststore Type"></property>
    <property name="TLS – Client Auth"></property>
    <property name="TLS – Protocol"></property>
    <property name="TLS – Shutdown Gracefully"></property>

    <property name="Referral Strategy">FOLLOW</property>
    <property name="Connect Timeout">10 secs</property>
    <property name="Read Timeout">10 secs</property>

    <property name="Url"></property>
    <property name="User Search Base"></property>
    <property name="User Search Filter"></property>

    <property name="Identity Strategy">USE_DN</property>
    <property name="Authentication Expiration">12 hours</property>
</provider>
```

```
<provider>
  <identifier>ldap-provider</identifier>
  <class>org.apache.nifi.ldap.LdapProvider</class>
  <property name="Authentication Strategy">SIMPLE</property>
  <property name="Manager DN">cn=admin,dc=nifi</property>
  <property name="Manager Password">nifi</property>
  <property name="TLS – Keystore"></property>
  <property name="TLS – Keystore Password"></property>
  <property name="TLS – Keystore Type"></property>
  <property name="TLS – Truststore"></property>
  <property name="TLS – Truststore Password"></property>
  <property name="TLS – Truststore Type"></property>
  <property name="TLS – Client Auth"></property>
  <property name="TLS – Protocol"></property>
  <property name="TLS – Shutdown Gracefully"></property>
  <property name="Referral Strategy">FOLLOW</property>
  <property name="Connect Timeout">10 secs</property>
  <property name="Read Timeout">10 secs</property>
  <property name="Url">ldap://ec2-54-237-215-138.compute-1.amazonaws.com:389</property>
  <property name="User Search Base">dc=nifi</property>
  <property name="User Search Filter">cn={0}</property>
  <property name="Identity Strategy">USE_DN</property>
  <property name="Authentication Expiration">12 hours</property>
</provider>
```

cn=admin,dc=nifi

DN

cn=nadeem,dc=nifi

nadeem

```
nifi.security.user.authorizer=managed-authorizer  
nifi.security.allow.anonymous.authentication=false  
nifi.security.user.login.identity.provider=ldap-provider  
nifi.security.user.jws.key.rotation.period=PT1H  
nifi.security.ocsp.responder.url=  
nifi.security.ocsp.responder.certificate=
```

nifi.properties

```
<provider>  
    <identifier>ldap-provider</identifier>  
    <class>org.apache.nifi.ldap.LdapProvider</class>  
    <property name="Authentication Strategy">SIMPLE</property>
```

login-identity-provider.xml

```
        name="Manager DN">cn=admin,dc=nifi</property>  
        name="Manager Password">nifi</property>
```

```
        <property name="TLS – Keystore"></property>  
        <property name="TLS – Keystore Password"></property>  
        <property name="TLS – Keystore Type"></property>  
        <property name="TLS – Truststore"></property>  
        <property name="TLS – Truststore Password"></property>
```

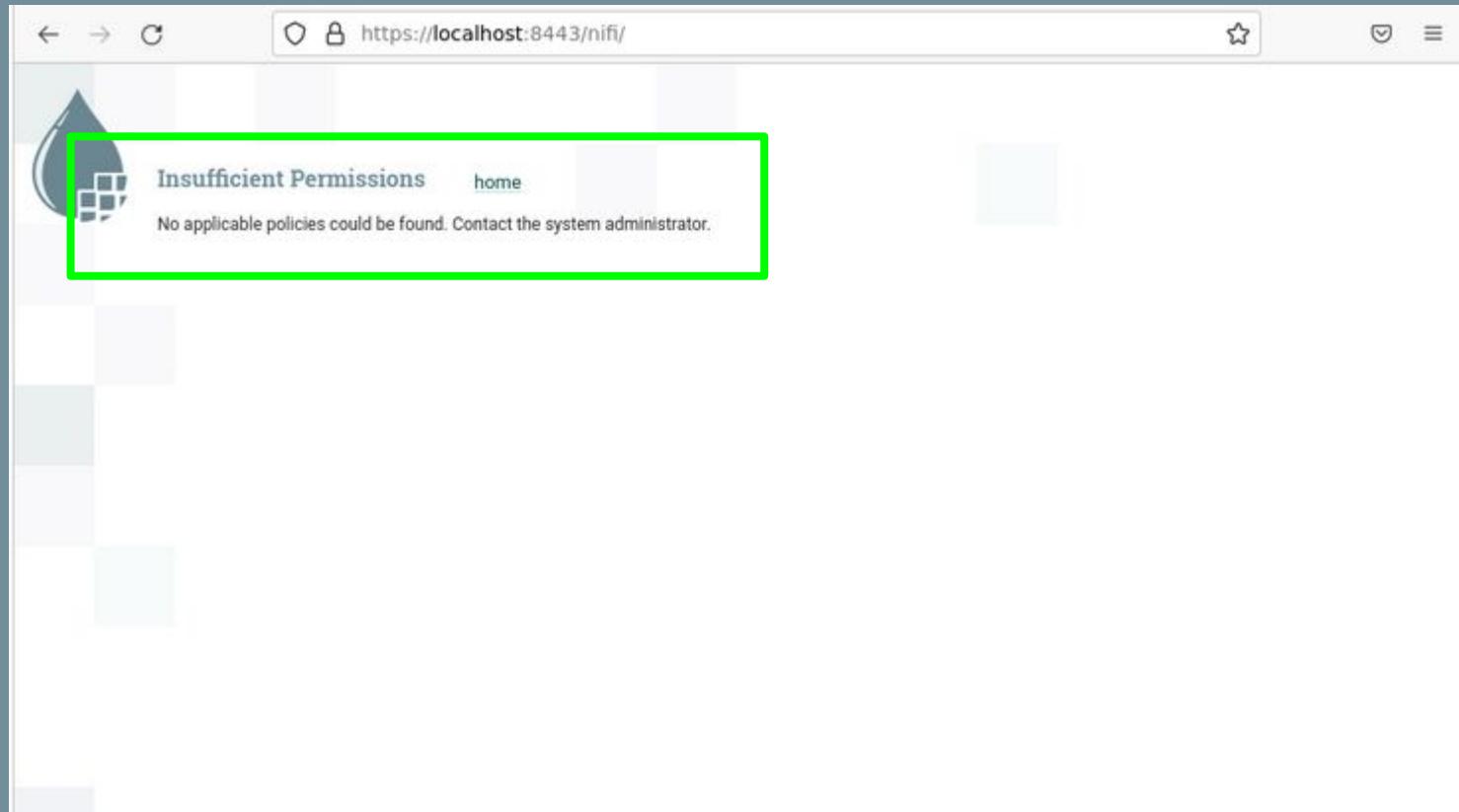
Phew!!! That's it **Start NiFi**

# – Steps to secure NiFi with LDAP

## Steps

1. Generate KeyStore and Truststore keys using NiFi Toolkit
2. Create users in LDAP Server
3. Configure LDAP Configuration in NiFi
4. Start NiFi

# No policies configured ERROR



# NiFi Authorization

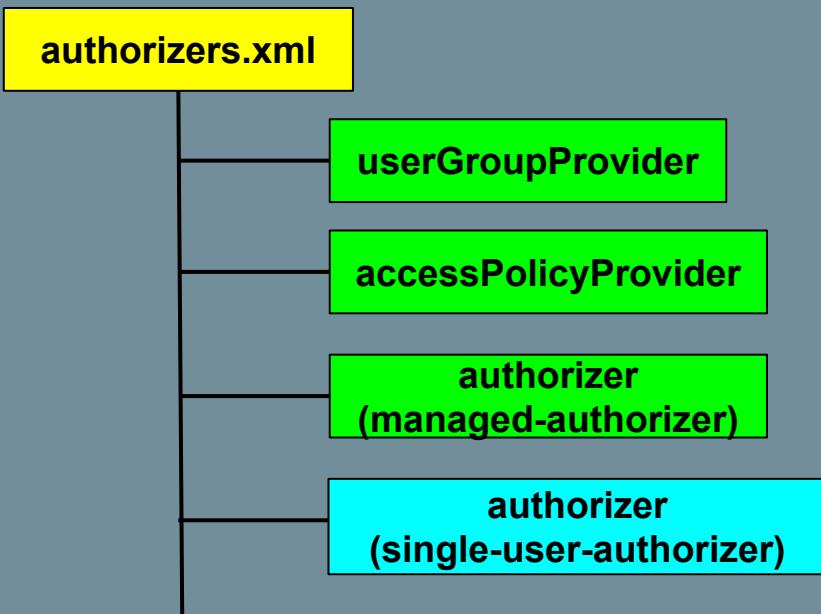
- **Steps to secure NiFi with OpenID Connect**

## Steps

1. Generate KeyStore and Truststore keys using NiFi Toolkit
2. Create users in LDAP Server
3. Configure LDAP Configuration in NiFi
4. Add initial admin user and assign Policies
5. Start NiFi

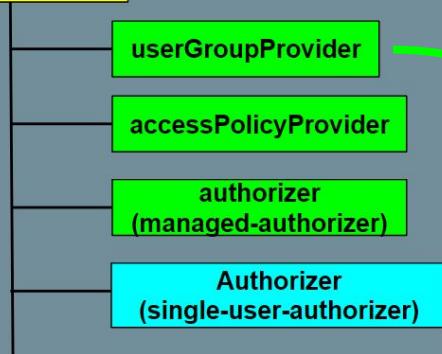
# Authorization

In NiFi, authorization is controlled by **authorizers.xml**



# Authorization

authorizers.xml

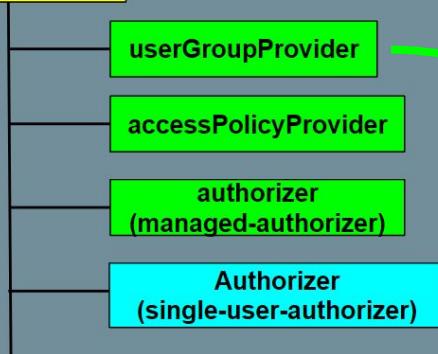


Related to  
NiFi Users

```
<userGroupProvider>
    <identifier>file-user-group-provider</identifier>
    <class>org.apache.nifi.authorization.FileUserGroupProvider</class>
    <property name="Users File">./conf/users.xml</property>
    <property name="Legacy Authorized Users File"></property>
    <property name="Initial User Identity 1"></property>
</userGroupProvider>
```

# Authorization

authorizers.xml



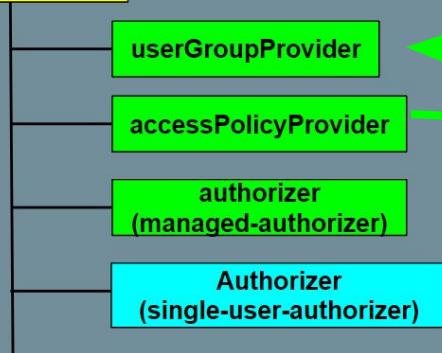
Related to  
NiFi Users

```
<userGroupProvider>
    <identifier>file-user-group-provider</identifier>
    <class>org.apache.nifi.authorization.FileUserGroupProvider</class>
    <property name="Users File">./conf/users.xml</property>
    <property name="Legacy Authorized Users File"></property>

    <property name="Initial User Identity 1">cn=nadeem,dc=nifi</property>
</userGroupProvider>
```

# Authorization

authorizers.xml

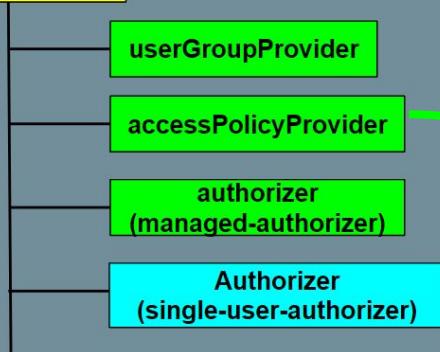


Related to  
**NiFi User policies**

```
<accessPolicyProvider>
    <identifier>file-access-policy-provider</identifier>
    <class>org.apache.nifi.authorization.FileAccessPolicyProvider</class>
    <property name="User Group Provider">file-user-group-provider</property>
    <property name="Authorizations File">./conf/authorizations.xml</property>
    <property name="Initial Admin Identity"></property>
    <property name="Legacy Authorized Users File"></property>
    <property name="Node Identity 1"></property>
    <property name="Node Group"></property>
</accessPolicyProvider>
```

# Authorization

authorizers.xml



Related to  
**NiFi User policies**

```
<accessPolicyProvider>
  <identifier>file-access-policy-provider</identifier>
  <class>org.apache.nifi.authorization.FileAccessPolicyProvider</class>
  <property name="User Group Provider">file-user-group-provider</property>
  <property name="Authorizations File">./conf/authorizations.xml</property>
  <property name="Initial Admin Identity">cn=nadeem,dc=nifi</property>
  <property name="Legacy Authorized Users File"></property>
  <property name="Node Identity 1"></property>
  <property name="Node Group"></property>
</accessPolicyProvider>
```

# Authorization

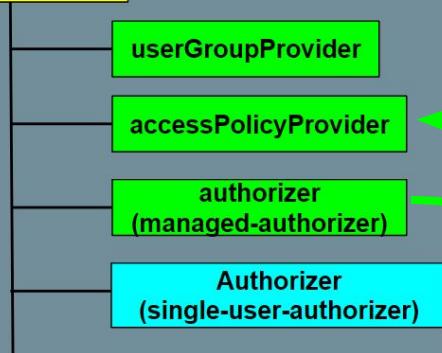
```
<userGroupProvider>
    <identifier>file-user-group-provider</identifier>
    <class>org.apache.nifi.authorization.FileUserGroupProvider</class>
    <property name="Users File">./conf/users.xml</property>
    <property name="Legacy Authorized Users File"></property>

    <property name="Initial User Identity 1">cn=nadeem,dc=nifi</property>
</userGroupProvider>
```

```
<accessPolicyProvider>
    <identifier>file-access-policy-provider</identifier>
    <class>org.apache.nifi.authorization.FileAccessPolicyProvider</class>
    <property name="User Group Provider">file-user-group-provider</property>
    <property name="Users Policies File">./conf/authorizations.xml</property>
    <property name="Initial Admin Identity">cn=nadeem,dc=nifi</property>
    <property name="Legacy Authorized Users File"></property>
    <property name="Node Identity 1"></property>
    <property name="Node Group"></property>
</accessPolicyProvider>
```

# Authorization

authorizers.xml



**authorization provider**

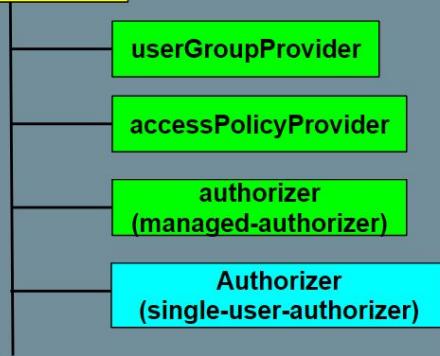
**Multiple users**

`nifi.security.user.authorizer=managed-authorizer`

```
<authorizer>
  <identifier>managed-authorizer</identifier>
  <class>org.apache.nifi.authorization.StandardManagedAuthorizer</class>
  <property name="Access Policy Provider">file-access-policy-provider</property>
</authorizer>
```

# Authorization

authorizers.xml



**authorization provider**

**Single user**

```
<authorizer>
  <identifier>single-user-authorizer</identifier>
  <class>org.apache.nifi.authorization.single.user.SingleUserAuthorizer</class>
</authorizer>
```



cn=nadeem,dc=nifi

LOG OUT



0

0 / 0 bytes

0

0

0

0

0

0

0

0

0

0

0

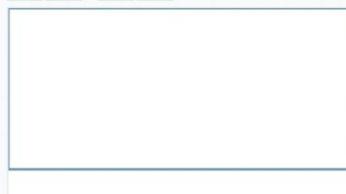
0

0

03:10:58 IST



## Navigate



## Operate



NiFi Flow

Process Group

efd5065f-0181-1000-11e9-ee5b0ebc1b71



—

# Let's say how user credentials are stored in LDAP

```
#  
# LDAPv3  
# base <dc=nifi> with scope subtree  
# filter: (objectclass=*)  
# requesting: ALL  
#
```

```
# nifi  
dn: dc=nifi  
objectClass: top  
objectClass: dcObject  
objectClass: organization  
o: nifiorg  
dc: nifi
```

```
# nadeem, nifi  
dn: cn=nadeem,dc=nifi  
cn: nadeem  
sn: 1  
objectClass: top  
objectClass: inetOrgPerson  
userPassword:: bmFkZWVtX3Bhc3M=
```

```
<property name="Url">ldap://ec2-54-237-215-138.compute-1.amazonaws.com:389</property>  
<property name="User Search Base">dc=nifi</property>  
<property name="User Search Filter">cn={0}</property>  
  
<property name="Identity Strategy">USE_DN</property>  
<property name="Authentication Expiration">12 hours</property>  
</provider>
```

—

# Demonstrate Multi-Tenant (User) Mechanism