

PRÁCTICA APACHE NIFI

Primero, creamos los directorios que llamamos “Practica_nifi” y “Pratica2_nifi” para Flujo 1 y Flujo 2 respectivamente. Dentro de cada uno, creamos los subdirectorios “Data”, “DataIn” y “DataOut” para almacenar los archivos de entrada y salida.

Luego, copiamos el fichero de “UAM_Clientes_Completo.csv” a cada directorio “DataIn” dentro de “Practica_nifi” y “Pratica2_nifi”.

```
umaster@ibmuamdocker:~/nifi-2.0.0-M2
File Edit View Search Terminal Help
[umaster@ibmuamdocker ~]$ cd nifi-2.0.0-M2/
[umaster@ibmuamdocker nifi-2.0.0-M2]$ mkdir Practica_nifi
[umaster@ibmuamdocker nifi-2.0.0-M2]$ mkdir Practica_nifi/Data
[umaster@ibmuamdocker nifi-2.0.0-M2]$ mkdir Practica_nifi/Data/DataIn
[umaster@ibmuamdocker nifi-2.0.0-M2]$ mkdir Practica_nifi/Data/DataOut
[umaster@ibmuamdocker nifi-2.0.0-M2]$ ls
bin           extensions      NOTICE          run
conf          flowfile_repository Practica_nifi state
[umaster@ibmuamdocker nifi-2.0.0-M2]$ cp ~/UAM_Clientes_Completo.csv Practica_nifi/Data/DataIn/
[umaster@ibmuamdocker nifi-2.0.0-M2]$ ls Practica_nifi/Data/DataIn/
UAM_Clientes_Completo.csv
[umaster@ibmuamdocker nifi-2.0.0-M2]$ ls Practica_nifi/Data/DataOut/
[umaster@ibmuamdocker nifi-2.0.0-M2]$ █
```

```
[umaster@ibmuamdocker nifi-2.0.0-M2]$ mkdir Practica2_nifi
[umaster@ibmuamdocker nifi-2.0.0-M2]$ mkdir Practica2_nifi/Data
```

```
[umaster@ibmuamdocker nifi-2.0.0-M2]$ mkdir Practica2_nifi/Data/DataIn
[umaster@ibmuamdocker nifi-2.0.0-M2]$ mkdir Practica2_nifi/Data/DataOut
[umaster@ibmuamdocker nifi-2.0.0-M2]$ ls
bin           docs          LICENSE        Practica_nifi      run
conf          extensions    logs          provenance_repository state
content_repository flowfile_repository NOTICE        python        work
database_repository lib          Practica2_nifi   README
[umaster@ibmuamdocker nifi-2.0.0-M2]$ ls Practica2_nifi/
Data
[umaster@ibmuamdocker nifi-2.0.0-M2]$ ls Practica2_nifi/Data/
DataIn  DataOut
```

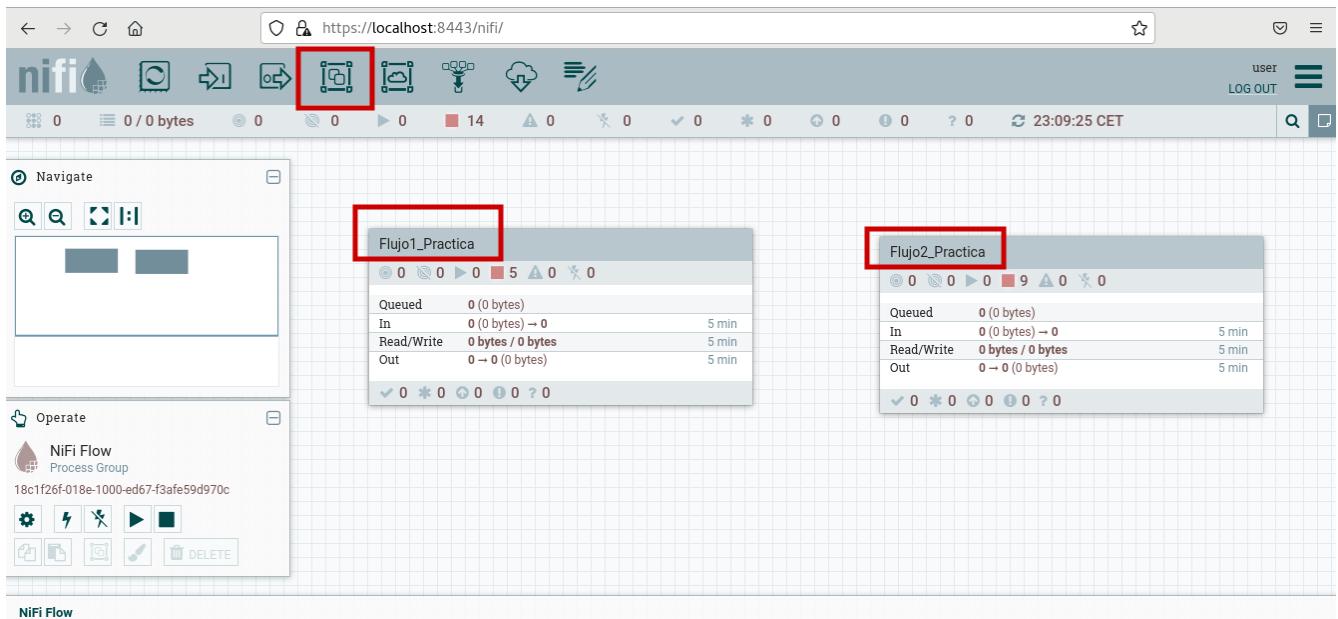
```
[umaster@ibmuamdocker nifi-2.0.0-M2]$ cp ~/UAM_Clientes_Completo.csv Practica2_nifi/Data/DataIn
[umaster@ibmuamdocker nifi-2.0.0-M2]$ ls Practica2_nifi/Data/DataIn/
UAM_Clientes_Completo.csv
[umaster@ibmuamdocker nifi-2.0.0-M2]$ █
```

Parte Obligatoria

FLUJO 1: Dividir el fichero de entrada en registros, convertir cada registro en un fichero y dejar todos los ficheros en un directorio.

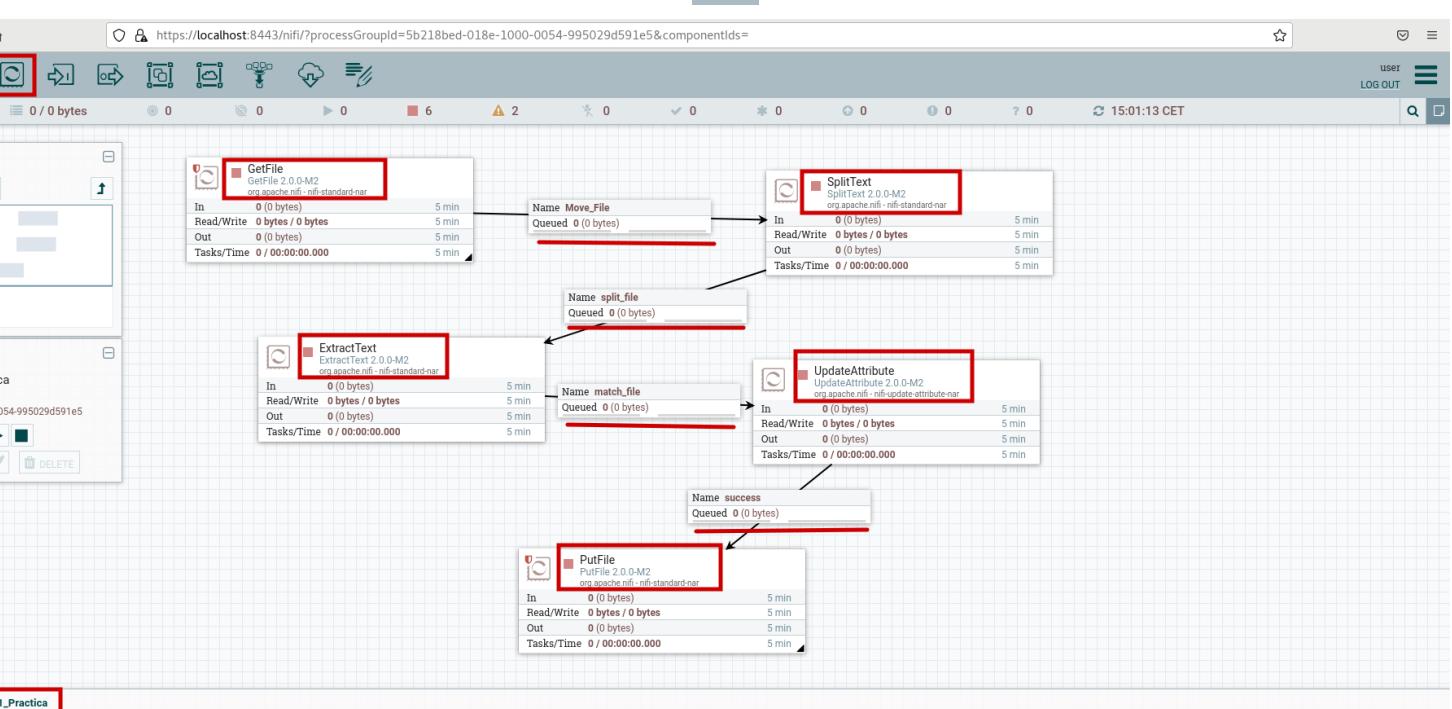
Primero, creamos grupos de flujos que llamamos Flujo1_Practica y Flujo2_Practica con el símbolo

Process Group 



Luego, en Flujo1_Practica creamos procesadores llamados GetFile, SplitText, ExtractText,

UpdateAttribute y PutFile con el símbolo Processor 



GetFile Configuración

Mantenemos el nombre de GetFile por ser un nombre descriptivo y fácil de recordar.

Configure Processor | GetFile 2.0.0-M2

Stopped

| | | | | |
|----------|------------|------------|---------------|----------|
| SETTINGS | SCHEDULING | PROPERTIES | RELATIONSHIPS | COMMENTS |
|----------|------------|------------|---------------|----------|

Name
GetFile Enabled

Id
5beb7bc5-018e-1000-8f1d-b333c7d2836d

Type
GetFile 2.0.0-M2

Bundle
org.apache.nifi - nifi-standard-nar

Penalty Duration ?
30 sec

Yield Duration ?
1 sec

Bulletin Level ?
WARN

CANCEL APPLY

Configure Processor | GetFile 2.0.0-M2

Stopped

| | | | | |
|----------|------------|------------|---------------|----------|
| SETTINGS | SCHEDULING | PROPERTIES | RELATIONSHIPS | COMMENTS |
|----------|------------|------------|---------------|----------|

Scheduling Strategy ?
Timer driven

Concurrent Tasks ?
1

Run Schedule ?
0 sec

Execution ?
All nodes

CANCEL APPLY

Indicamos la ruta del archivo DataIn y desactivamos la opción "terminate" para mantener la conexión.

Configure Processor | GetFile 2.0.0-M2

■ Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Required field

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

CANCEL APPLY

Configure Processor | GetFile 2.0.0-M2

■ Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Automatically Terminate / Retry Relationships ?

success
 terminate retry

All files are routed to success

CANCEL APPLY

Añadimos una descripción.

Configure Processor | GetFile 2.0.0-M2

■ Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

This is the first processor, which takes the data from the DataIn directory.

CANCEL APPLY

SplitText Configuración

Configuramos Line Split Count en 1 para dividir cada fila.

Configure Processor | SplitText 2.0.0-M2

■ Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Required field

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

CANCEL APPLY

Conexión entre GetFile y SplitText

Creamos una conexión llamada "Move_File" para conectar GetFile y SplitFile.

Configure Connection

DETAILS **SETTINGS**

| | |
|---|-----------------|
| From Processor | To Processor |
| GetFile | SplitText |
| GetFile | SplitText |
| Within Group | Within Group |
| Flujo1_Practica | Flujo1_Practica |
| For Relationships | |
| <input checked="" type="checkbox"/> success | |

CANCEL **APPLY**

Configure Connection

DETAILS **SETTINGS**

| | |
|--|---|
| Name | Available Prioritizers <small>?</small> |
| Move_File | FirstInFirstOutPrioritizer |
| Id | NewestFlowFileFirstPrioritizer |
| 5bf323cb-018e-1000-a60c-443936d24ec2 | OldestFlowFileFirstPrioritizer |
| FlowFile Expiration <small>?</small> | PriorityAttributePrioritizer |
| 0 sec | |
| Back Pressure Object Threshold <small>?</small> | Selected Prioritizers <small>?</small> |
| 10000 | |
| Size Threshold <small>?</small> | |
| 1 GB | |
| Load Balance Strategy <small>?</small> | |
| Do not load balance | |

CANCEL **APPLY**

ExtractText Configuración

Agregamos una nueva propiedad llamada "firstword" para encontrar la primera palabra en cada fila.

Configure Processor | ExtractText 2.0.0-M2

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Required field

| Property | Value |
|---|-------|
| Minimum Capture Group Length | 0 |
| Enable Canonical Equivalence | false |
| Enable Case-insensitive Matching | false |
| Permit Whitespace and Comments in Pattern | 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 Classes | false |
| Enable Unix Lines Mode | false |
| Include Capture Group 0 | true |
| Enable repeating capture group | false |
| Enable named group support | false |

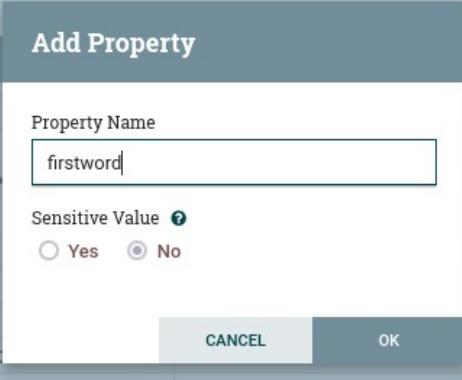
Add Property

Property Name: firstword

Sensitive Value: Yes No

CANCEL OK

CANCEL APPLY



Configure Processor | ExtractText 2.0.0-M2

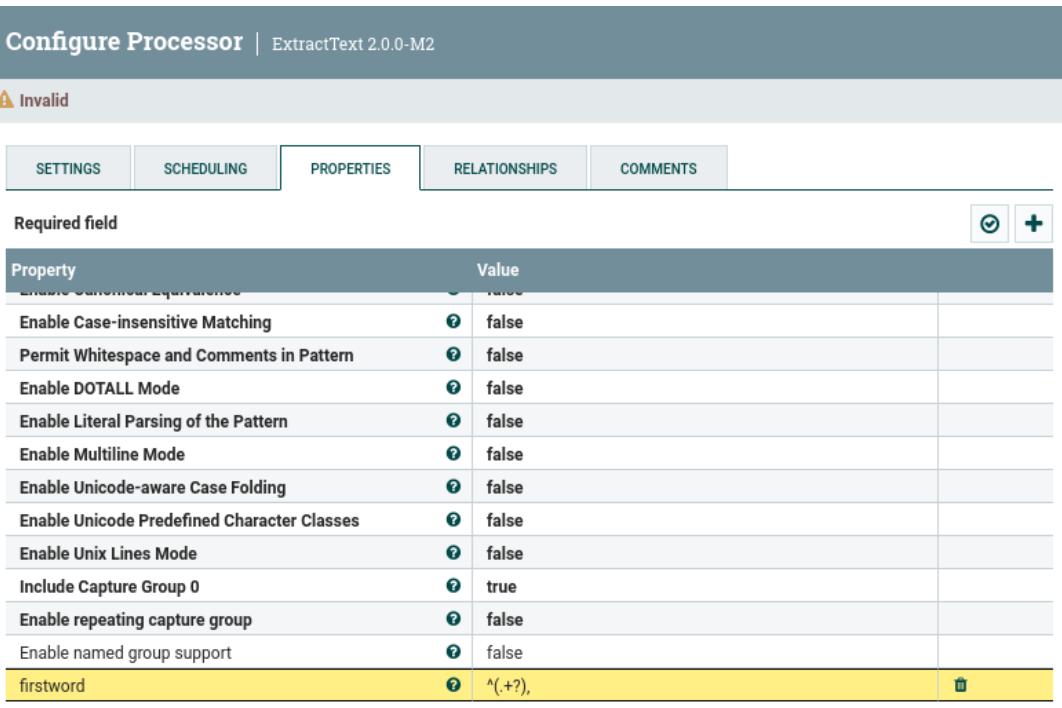
⚠ Invalid

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Required field

| Property | Value |
|---|---------|
| Enable Canonical Equivalence | false |
| Enable Case-insensitive Matching | false |
| Permit Whitespace and Comments in Pattern | 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 Classes | false |
| Enable Unix Lines Mode | false |
| Include Capture Group 0 | true |
| Enable repeating capture group | false |
| Enable named group support | false |
| firstword | ^(.+?), |

CANCEL APPLY



No seleccionamos la opción "match terminate" para continuar con la coincidencia de archivos.

Configure Processor | ExtractText 2.0.0-M2

■ Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Automatically Terminate / Retry Relationships ?

matched
 terminate retry

FlowFiles are routed to this relationship when the Regular Expression is successfully evaluated and the FlowFile is modified as a result

unmatched
 terminate retry

FlowFiles are routed to this relationship when no provided Regular Expression matches the content of the FlowFile

CANCEL APPLY

Conexión entre SplitText y ExtractText

Creamos una conexión llamada "split_file" para conectar SplitText y ExtractText.

Configure Connection

DETAILS SETTINGS

From Processor
SplitText
SplitText
Within Group
Flujo1_Practica

To Processor
ExtractText
ExtractText
Within Group
Flujo1_Practica

For Relationships
 failure
 original
 splits

CANCEL APPLY

| | |
|---|--|
| DETAILS | SETTINGS |
| <p>Name split_file</p> <p>Id 5bfc7d48-018e-1000-49fa-01c5c2a95b8b</p> <p>FlowFile Expiration ? 0 sec</p> <p>Back Pressure Object Threshold ? 10000</p> <p>Size Threshold ? 1 GB</p> | <p>Available Prioritizers ?</p> <ul style="list-style-type: none"> FirstInFirstOutPrioritizer NewestFlowFileFirstPrioritizer OldestFlowFileFirstPrioritizer PriorityAttributePrioritizer <p>Selected Prioritizers ?</p> |
| <p>Load Balance Strategy ?</p> <p>Do not load balance</p> | |

UpdateAttribute Configuración

Añadimos una nueva propiedad llamada "filename" para determinar el nombre de fichero de salida.

Configure Processor | UpdateAttribute 2.0.0-M2

Stopped

| | | | | |
|----------|------------|------------|---------------|----------|
| SETTINGS | SCHEDULING | PROPERTIES | RELATIONSHIPS | COMMENTS |
|----------|------------|------------|---------------|----------|

Required field

Add Property

Property Name
filename

Sensitive Value ?
 Yes No

CANCEL OK

ADVANCED CANCEL APPLY

El nombre del fichero de salida estará formado por la primera palabra de la fila y un identificador único para evitar nombres similares.

Configure Processor | UpdateAttribute 2.0.0-M2

Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Required field

| Property | Value |
|----------------------------------|---|
| Delete Attributes Expression | No value set |
| Store State | Do not store state |
| Stateful Variables Initial Value | No value set |
| Cache Value Lookup Cache Size | |
| filename | <input type="text" value="\${firstword} . \${uuid}"/> <input checked="" type="checkbox"/> Set empty string |

EL PARAM OK CANCEL

ADVANCED CANCEL APPLY

Conexión entre ExtractText y UpdateAttribute

Creamos una conexión llamada "match_file" para conectar ExtractText y UpdateAttribute.

Configure Connection

DETAILS SETTINGS

| From Processor | To Processor |
|-----------------|-----------------|
| ExtractText | UpdateAttribute |
| ExtractText | UpdateAttribute |
| Within Group | Within Group |
| Flujo1_Practica | Flujo1_Practica |

For Relationships
 matched
 unmatched

CANCEL APPLY

Configure Connection

DETAILS SETTINGS

| | |
|--|---|
| Name match_file | Available Prioritizers ? FirstInFirstOutPrioritizer |
| Id 5bfca0f8-018e-1000-7d8b-b6f703702050 | NewestFlowFileFirstPrioritizer |
| FlowFile Expiration ? 0 sec | OldestFlowFileFirstPrioritizer |
| Back Pressure Object Threshold ? 10000 | PriorityAttributePrioritizer |
| Size Threshold ? 1 GB | Selected Prioritizers ? |
| Load Balance Strategy ? Do not load balance | |

CANCEL APPLY

PutFile Configuración

Mantenemos el nombre de PutFile por ser un nombre descriptivo y fácil de recordar.

Configure Processor | PutFile 2.0.0-M2

■ Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

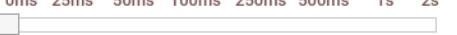
| | |
|---|---|
| Name PutFile | <input checked="" type="checkbox"/> Enabled |
| Id 5beb99fb-018e-1000-11ed-09392eef79f9 | |
| Type PutFile 2.0.0-M2 | |
| Bundle org.apache.nifi - nifi-standard-nar | |
| Penalty Duration ? 30 sec | Yield Duration ? 1 sec |
| Bulletin Level ? WARN | |

CANCEL APPLY

| | |
|-------------------------------|-------|
| Read/Write: 0 bytes / 0 bytes | 5 min |
| Out: 0 (0 bytes) | |

Configure Processor | PutFile 2.0.0-M2

■ Stopped

| SETTINGS | SCHEDULING | PROPERTIES | RELATIONSHIPS | COMMENTS |
|--|---|------------|---------------|----------|
| Scheduling Strategy ? Timer driven | Run Duration ? 0ms 25ms 50ms 100ms 250ms 500ms 1s 2s  Lower latency Higher throughput | | | |
| Concurrent Tasks ? 1 | Run Schedule ? 0 sec | | | |
| Execution ? All nodes | | | | |

CANCEL **APPLY**

Indicamos la ruta del archivo DataOut.

Configure Processor | PutFile 2.0.0-M2

■ Stopped

| SETTINGS | SCHEDULING | PROPERTIES | RELATIONSHIPS | COMMENTS | | | | | | | | | | | | | | | | | | |
|---|----------------------------|------------|---------------|----------|----------|-------|-----------|----------------------------|------------------------------|------|----------------------------|------|--------------------|--------------|--------------------|--------------|-------------|--------------|-------|--------------|-------|--------------|
| Required field | | | | | | | | | | | | | | | | | | | | | | |
| <table border="1"> <thead> <tr> <th>Property</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>Directory</td> <td>Practica_nifi/Data/DataOut</td> </tr> <tr> <td>Conflict Resolution Strategy</td> <td>fail</td> </tr> <tr> <td>Create Missing Directories</td> <td>true</td> </tr> <tr> <td>Maximum File Count</td> <td>No value set</td> </tr> <tr> <td>Last Modified Time</td> <td>No value set</td> </tr> <tr> <td>Permissions</td> <td>No value set</td> </tr> <tr> <td>Owner</td> <td>No value set</td> </tr> <tr> <td>Group</td> <td>No value set</td> </tr> </tbody> </table> | | | | | Property | Value | Directory | Practica_nifi/Data/DataOut | Conflict Resolution Strategy | fail | Create Missing Directories | true | Maximum File Count | No value set | Last Modified Time | No value set | Permissions | No value set | Owner | No value set | Group | No value set |
| Property | Value | | | | | | | | | | | | | | | | | | | | | |
| Directory | Practica_nifi/Data/DataOut | | | | | | | | | | | | | | | | | | | | | |
| Conflict Resolution Strategy | fail | | | | | | | | | | | | | | | | | | | | | |
| Create Missing Directories | true | | | | | | | | | | | | | | | | | | | | | |
| Maximum File Count | No value set | | | | | | | | | | | | | | | | | | | | | |
| Last Modified Time | No value set | | | | | | | | | | | | | | | | | | | | | |
| Permissions | No value set | | | | | | | | | | | | | | | | | | | | | |
| Owner | No value set | | | | | | | | | | | | | | | | | | | | | |
| Group | No value set | | | | | | | | | | | | | | | | | | | | | |
| <input checked="" type="checkbox"/> <input type="button" value="+"/> | | | | | | | | | | | | | | | | | | | | | | |

CANCEL **APPLY**

Activamos la opción "terminate" para finalizar la conexión y añadimos una descripción.

Configure Processor | PutFile 2.0.0-M2

■ Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Automatically Terminate / Retry Relationships ⓘ

failure terminate retry

Files that could not be written to the output directory for some reason are transferred to this relationship

success terminate retry

Files that have been successfully written to the output directory are transferred to this relationship

CANCEL APPLY

Configure Processor | PutFile 2.0.0-M2

■ Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Put the input data in the DataOut directory

CANCEL APPLY

Conexión entre UpdateAttribute y PutFile

Creamos una conexión llamada "success" para conectar UpdateAttribute y PutFile.

Configure Connection

DETAILS **SETTINGS**

| | |
|---|-----------------|
| From Processor | To Processor |
| UpdateAttribute | PutFile |
| UpdateAttribute | PutFile |
| Within Group | Within Group |
| Flujo1_Practica | Flujo1_Practica |
| For Relationships | |
| <input checked="" type="checkbox"/> success | |

CANCEL **APPLY**

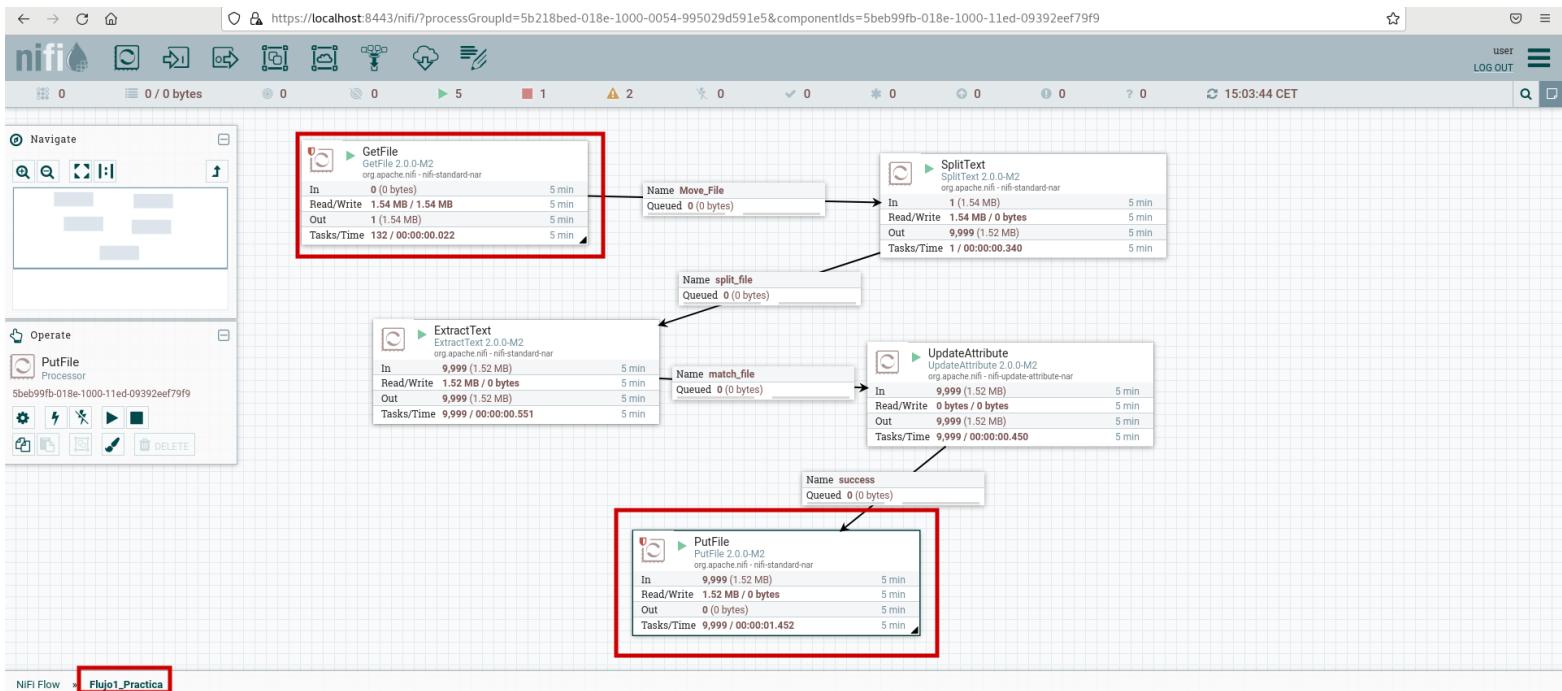
Configure Connection

DETAILS **SETTINGS**

| | |
|--|---|
| Name | Available Prioritizers ? |
| success | FirstInFirstOutPrioritizer |
| Id | NewestFlowFileFirstPrioritizer |
| 5bfccb225-018e-1000-e7b7-2f374536474c | OldestFlowFileFirstPrioritizer |
| FlowFile Expiration ? | PriorityAttributePrioritizer |
| 0 sec | |
| Back Pressure Object Threshold ? | Size Threshold ? |
| 10000 | 1 GB |
| Load Balance Strategy ? | Selected Prioritizers ? |
| Do not load balance | |

CANCEL **APPLY**

Se observa que el "Flujo 1" funciona correctamente al ejecutar todos los procesadores.



Realizamos una verificación para determinar si los ficheros han sido exportados a DataOut y si el total es igual a 9999.

Observamos que todos los archivos han sido exportados con nombres diferentes y el total es 9999.

```
[umaster@ibmuamdocker nifi-2.0.0-M2]$ cd Practica_nifi/Data/DataOut/
[umaster@ibmuamdocker DataOut]$ ls
CRM.000d0865-1e15-472c-af88-e0781cbb81c5 CRM.f7f150a2-57b3-4a5a-b9bd-cf6582bcb095
CRM.000e0808-6cdd-47ac-9e72-ce2cd45dd768 CRM.f7f4e961-2bd2-4c6d-a3fe-6636c6e9798f
CRM.000e812b-4069-49be-b42d-81c820b2dca5 CRM.f7f912d3-a8e8-47ce-bc4b-7cecf4d4447f
CRM.001874c8-abfa-48b8-8fea-8e046b819380 CRM.f7fe732b-4f6e-4396-8567-3716e7881d17
CRM.001faf64-f271-4afc-badd-e907b205f63e CRM.f8033812-f19e-4f32-ac65-5b47ea5092a1
CRM.002c6818-b997-4f4b-bc9b-596f0a363c19 CRM.f81f40a8-5b82-4897-965b-b72eee85ee41
CRM.004494eb-f992-411d-ba6d-222b6568c6fa CRM.f83a3b5d-c2df-48a0-aa97-3a28d42b6836
CRM.00663df4-3ddc-41b5-9d09-2fab8fed2b14 CRM.f84a6dfe-1d60-4c64-ace2-28621db9b656
CRM.006a9bb3-354f-4916-95c0-37cc9d80da1f CRM.f84b6c5c-9c11-4309-8d86-3f43d3d24e09
CRM.00847122-8657-4563-87a1-f143017357e2 CRM.f8540582-e4fc-4ad2-a726-dcd48d77bdda
CRM.00980bd6-9054-4306-84a3-7e893b135f91 CRM.f87987ee-6134-4f81-b22a-47882aec168
CRM.00982c1e-62f3-4b58-819d-ee8876c8fe41 CRM.f88dcf92-b623-4b59-8983-c2794230e406
CRM.009e25b3-7926-4afe-b734-41ae9c936601 CRM.f8a3134a-2dfb-4f4d-b8c3-9af8ffdb3d3b
CRM.00a09414-e84a-4c12-b151-bedc2a41694b CRM.f8be3da5-2ea6-4514-8196-6d0303251b61
CRM.00a22319-aa3e-4593-86bb-6f0dea2e3291 CRM.f8eca374-e89b-44bf-b8df-c23fa0cab390
CRM.00a853ed-d0b7-46fe-b3ca-e4d8b7aae81e CRM.f8f12d6a-d02f-48ee-bfce-0ee4fadfff14
CRM.00af2cf8-29fe-4bba-9dd3-33c5838ddd0f CRM.f8ffb393-b6d8-4c44-9ecb-63d191fe27fc
CRM.00b3dba9-30db-42f5-8e6b-023913f9f91e CRM.f8ffff006-4a0f-4b1c-af89-a4d246a29b81
CRM.00ce9134-dda7-4367-9b7e-0a5ed7d2a854 CRM.f900620c-3bc4-4309-b546-0c2a26f95ded
CRM.00e56a95-a1be-4c43-8e4e-5a3cf4961a9b CRM.f90f25f4-2799-4e7e-93dc-5dcc32aa1f72
CRM.0107c813-0f6f-4836-a058-a42318250923 CRM.f913c7c2-8d29-44bd-ad46-5bb05b06a3de
CRM.01109304-d246-4100-b562-4fb7a7bbdec3 CRM.f93d461e-360f-4384-b735-2c5b75c436a0
CRM.0113f434-99cd-4881-a772-19538c77ab5c CRM.f95f1a3d-de36-45f1-af09-be950113fff3
CRM.011ba8bc-016e-44d3-b933-8a8f11ec46b5 CRM.f96b731f-5160-4854-8a64-0d32319d0343
CRM.012422b5-9348-4ffd-80c1-ed752078a5d1 CRM.f971d75c-4104-4125-8976-49f368b2eeb8
CRM.013acc1c-30b5-47e7-b487-d5ee4da097cc CRM.f9720821-d229-4c5e-b1c7-e009706d4a43
CRM.013fe337-98d1-4669-b629-4c3cf62aa482 CRM.f980e280-42f3-46eb-b3cf-19855134a7dd
CRM.014451f0-4ff6-4ff7-86b9-da92e2e318b1 CRM.f9831bb9-debd-4a46-b7e9-457e6f1a0a69
```

| Recent | Name | Size | Modified |
|-----------|---------|-------------|----------|
| Home | Dataln | 0 items | 15:01 |
| Documents | DataOut | 9,999 items | 15:03 |

| Name | Size | Modified |
|---|-------------------------------|----------|
| CRM.0a2aed2a-ba53-4c5e-a1ec-2fdaeec79... | 167 bytes | 15:03 |
| CRM.0a2e40d3-da8e-42cf-9884-e329fbf93... | 161 bytes | 15:03 |
| CRM.0a3f9e98-f95d-450b-9daf-a066bca15... | 164 bytes | 15:03 |
| CRM.0a4b0113-5196-4538-9350-328731a4... | 149 bytes | 15:03 |
| CRM.0a5badca-bf73-4202-8995-359a40de... | 155 bytes | 15:03 |
| CRM.0a9f982e-1b8a-48ef-9aa5-bbce4dc8d... | 155 bytes | 15:03 |
| CRM.0a18df6f-dbf3-4f77-806f-3f59cba79fe9 | 157 bytes | 15:03 |
| CRM.0a076fab-8892-47c9-8d4a-e0b6bde2... | 162 bytes | 15:03 |
| CRM.0a86c698-2439-48b1-9bd7-98f6fe80... | 155 bytes | 15:03 |
| CRM.0a87bb24-a88f-4d24-ab35-d35ec529... | 144 bytes | 15:03 |
| CRM.0a379f93-5410-4c73-a854-c2f1d3472... | 158 bytes | 15:03 |
| CRM.0a619ebf-eddd-44d9-8455-665ceb95... | 150 bytes | 15:03 |
| CRM.00a853ed-d0b7-46fe-b3ca-e4d8b7aa... | 157 bytes | 15:03 |
| CRM.0a930e9c-b0e1-45c5-a8a1-cef4e77e0... | 168 bytes | 15:03 |
| CRM.0a966b66-c7a3-4392-81e2-8765962b... | 152 bytes | 15:03 |
| CRM.00a09414-e84a-4c12-b151-bedc2a41... | 145 bytes | 15:03 |
| CRM.00a22319-aa3e-4593-86bb-6f0dea2e... | 156 bytes | 15:03 |
| CRM.0a63077f-14c6-4393-96aa-1447b607... | 152 bytes | 15:03 |
| CRM.0a83226f-d5bf-4f39-a0f3-38d0ef3975... | 165 bytes | 15:03 |
| CRM.0aa56405-d052-4c65-afd3-2a391737... | 161 bytes | 15:03 |
| CRM.0aae87e2-804a-41c1-9e62-1fda2e71c... | 168 bytes | 15:03 |
| CRM.0ab6d1be-a114-42c4-bdb9-177e0224... | 147 bytes | 15:03 |
| CRM.0abf32ca-0b8b-4400-b893-34d1d1d6... | 163 bytes | 15:03 |
| CRM.0ac6b898-5c33-423d-8144-e317ad | 9,999 items selected (1.6 MB) | |

NiFi Data Provenance

Podemos revisar data provenance of PutFile, UpdateAttribute, ExtractText, SplitText y GetFile.
Dado que ejecutamos los procesadores varias veces, las horas pueden ser diferentes.

NiFi Data Provenance

| Showing the most recent 1,000 of 1,000+ events, please refine the search. | | | | | | |
|---|-----------------------------|------|--------------------------------------|-----------|----------------|----------------|
| Filter | by component name | Type | FlowFile Uuid | Size | Component Name | Component Type |
| 1 | 03/20/2024 19:28:38.863 CET | DROP | e8ae7d9c-6738-4395-81c5-9f90d818b44c | 153 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.863 CET | SEND | e8ae7d9c-6738-4395-81c5-9f90d818b44c | 153 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.863 CET | DROP | adc31393-ebf1-4ea4-8d6e-511700f51595 | 153 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.863 CET | SEND | adc31393-ebf1-4ea4-8d6e-511700f51595 | 153 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.863 CET | DROP | 4c6d690e-7d84-48ea-8450-933c82cb76db | 156 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.863 CET | SEND | 4c6d690e-7d84-48ea-8450-933c82cb76db | 156 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.863 CET | DROP | 22371d49-98c4-49c6-a253-39a5db12c168 | 149 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.863 CET | SEND | 22371d49-98c4-49c6-a253-39a5db12c168 | 149 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.863 CET | DROP | 5bc8a04b-ba67-48bc-b977-9599c6c4178d | 138 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.863 CET | SEND | 5bc8a04b-ba67-48bc-b977-9599c6c4178d | 138 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.863 CET | DROP | a8e5d3b3-586d-4d8b-94a9-8ecc6216cac | 156 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.863 CET | SEND | a8e5d3b3-586d-4d8b-94a9-8ecc6216cac | 156 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.863 CET | DROP | 1403418-1fcf-4392-b3dc-50ca7ee2b5b | 142 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.863 CET | SEND | 1403418-1fcf-4392-b3dc-50ca7ee2b5b | 142 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.863 CET | DROP | 7fb6125-6706-45f9-4ab1-cf3f5f4df21 | 151 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.862 CET | SEND | 7fb6125-6706-45f9-4ab1-cf3f5f4df21 | 151 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.862 CET | DROP | 9effe997-7425-40c2-bd29-8285003d79fb | 148 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.862 CET | SEND | 9effe997-7425-40c2-bd29-8285003d79fb | 148 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.862 CET | DROP | 737a086e-8c23-4a54-838c-6dbc8ba1c297 | 165 bytes | PutFile | PutFile |
| 1 | 03/20/2024 19:28:38.862 CET | SEND | 737a086e-8c23-4a54-838c-6dbc8ba1c297 | 165 bytes | PutFile | PutFile |

Last updated: 20:28:40 CET

NiFi Data Provenance

| Showing 1,000 of 1,000+ events that match the specified query, please refine the search. Clear search | | | | | | |
|---|-----------------------------|---------------------|--------------------------------------|-----------|-----------------|-----------------|
| Filter | by component name | Type | FlowFile Uuid | Size | Component Name | Component Type |
| 1 | 03/20/2024 19:28:37.951 CET | ATTRIBUTES_MODIFIED | e8ae7d9c-6738-4395-81c5-9f90d818b44c | 153 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.951 CET | ATTRIBUTES_MODIFIED | adc31393-ebf1-4ea4-8d6e-511700f51595 | 153 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.951 CET | ATTRIBUTES_MODIFIED | 4c6d690e-7d84-48ea-8450-933c82cb76db | 156 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.951 CET | ATTRIBUTES_MODIFIED | 22371d49-98c4-49c6-a253-39a5db12c168 | 149 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.951 CET | ATTRIBUTES_MODIFIED | 5bc8a04b-ba67-48bc-b977-9599c6c4178d | 138 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.951 CET | ATTRIBUTES_MODIFIED | a8e5d3b3-586d-4d8b-94a9-8ecc6216cac | 156 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.951 CET | ATTRIBUTES_MODIFIED | 1403418-1fcf-4392-b3dc-50ca7ee2b5b | 142 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.951 CET | ATTRIBUTES_MODIFIED | 7fb6125-6706-45f9-4ab1-cf3f5f4df21 | 151 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.951 CET | ATTRIBUTES_MODIFIED | 9effe997-7425-40c2-bd29-8285003d79fb | 148 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.950 CET | ATTRIBUTES_MODIFIED | 737a086e-8c23-4a54-838c-6dbc8ba1c297 | 165 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.950 CET | ATTRIBUTES_MODIFIED | 1625f05d-82ff-41f5-91f2-8af393b285fd | 159 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.950 CET | ATTRIBUTES_MODIFIED | 8e43208e-77bc-40a4-b712-76c1026fe4f | 161 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.950 CET | ATTRIBUTES_MODIFIED | e084aff1-2b75-4e73-b4b7-b17a26968264 | 156 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.950 CET | ATTRIBUTES_MODIFIED | 7ef6d009-3f96-4e82-9cb4-ed09573d422 | 167 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.950 CET | ATTRIBUTES_MODIFIED | c2509bab-402d-49fe-904e-58bb77a20db | 141 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.950 CET | ATTRIBUTES_MODIFIED | fd80ff2-6dd6-4651-9ff2-e704ead0fc0f | 148 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.950 CET | ATTRIBUTES_MODIFIED | fa7436a9-8119-4fb8-bc95-5c0370d71723 | 154 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.950 CET | ATTRIBUTES_MODIFIED | 7ebcc1ed-d859-4131-bc86-562471ffedfb | 163 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.950 CET | ATTRIBUTES_MODIFIED | 1299cf6c-59e5-4e40-9d4e-f44a1e9d665 | 170 bytes | UpdateAttribute | UpdateAttribute |
| 1 | 03/20/2024 19:28:37.950 CET | ATTRIBUTES_MODIFIED | 3da40f59ff142e6-8144-8e7b33073cd3 | 170 bytes | UpdateAttribute | UpdateAttribute |

Last updated: 20:29:13 CET

NiFi Data Provenance

Displaying 1,000 of 1,000
Oldest event available: 03/20/2024 11:49:27 CET

Showing 1,000 of 1,000+ events that match the specified query, please refine the search. Clear search

| Filter | by component name | | | | | | |
|-----------------------------|---------------------|---------------------------------------|-----------|----------------|----------------|--|--|
| Date/Time | Type | FlowFile Uuid | Size | Component Name | Component Type | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | e8ae7d9c-6738-4395-81c5-9f90d81b844c | 153 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | adc31393-ebf4-4ea4-8d6e-511700f51595 | 153 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | 4cd690e-7d84-48ee-8450-933c82cb76d4 | 156 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | 22371049-98c4-49c6-a253-39a5bd12c168 | 149 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | 5bc8a04b-ba67-48bc-b977-9599c6c4178d | 138 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | a8e5d5b3-586d-4d8b-944a-8ec0cc21ca0c | 156 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | 1403f418-1fc1-4992-8d3c-50ca71e2c05b | 142 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | 7fb6125-6706-4659-94b1-c3f554d21 | 151 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | 9effe997-7425-402-bd29-8285003d79fb | 148 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | 737a066e-8c23-4a54-838c-6dbccba1c297 | 165 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | 162505d-82f1-41f5-911c-8af3d9b285f1 | 159 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | 8e43208e-776c-400a-b024-76c1026efef4f | 161 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | e084df11-2b75-4e73-b4b7-b17a26968264 | 156 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | 7e6ffd09-3f9c-4e82-9cb4-e095c73d422 | 167 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | c2509bab-402d-496e-904a-58bb77a720d8 | 141 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | f8b0fe12-6d6f-4651-9827-e7d04ead0fcf | 148 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | f7a436a9-8119-4fbd-bc95-5c0370d71723 | 154 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | 7ebc61ed-d859-4131-bc86-d5247f1feefbf | 163 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.947 CET | ATTRIBUTES_MODIFIED | 1299c6f2-5965-4e40-9b64-f44a1e9de665 | 170 bytes | ExtractText | ExtractText | | |
| 03/20/2024 19:28:37.946 CET | ATTRIBUTES_MODIFIED | 3da840f3-9ff1-42ee-8146-87b33073c3d | 170 bytes | ExtractText | ExtractText | | |

Last updated: 20:29:54 CET

NiFi Data Provenance

Displaying 1,000 of 1,000
Oldest event available: 03/20/2024 11:49:27 CET

Showing 1,000 of 1,000+ events that match the specified query, please refine the search. Clear search

| Filter | by component name | | | | | | |
|-----------------------------|-------------------|--------------------------------------|-----------|----------------|----------------|--|--|
| Date/Time | Type | FlowFile Uuid | Size | Component Name | Component Type | | |
| 03/20/2024 19:28:36.686 CET | DROP | ce6cbc88-866f-4fdb-a69d-c24439b5521e | 1.54 MB | SplitText | SplitText | | |
| 03/20/2024 19:28:36.643 CET | FORK | ce6cbc88-866f-4fdb-a69d-c24439b5521e | 1.54 MB | SplitText | SplitText | | |
| 03/20/2024 15:01:52.431 CET | DROP | 82da7cdd-4cb2-42f3-a934-cc65f50f708a | 1.54 MB | SplitText | SplitText | | |
| 03/20/2024 15:01:52.401 CET | FORK | 82da7cdd-4cb2-42f3-a934-cc65f50f708a | 1.54 MB | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | 8485654d-dc42-4530-bd3d-58bb79822f2d | 161 bytes | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | ce7aa9f3-7f23-4c32-43c6-12ed2a254e6 | 156 bytes | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | 0b24288e-e731-46a4-bbde-2c9df47f07fd | 156 bytes | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | a8837e12-f0b2-4321-9409-b54c1c0203bd | 159 bytes | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | c689eb4a-4845-4479-8d09-d4d66b0797f2 | 177 bytes | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | 367c13df-ad6e-4656-9d62-57769b083ad | 157 bytes | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | f2d6495e-5430-4b19-ba77-6a7c7396932c | 171 bytes | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | 0087c492-8701-49d1-aa5a-0f8ceef647c | 159 bytes | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | 8281d8de-8d69-4dee-9ee9-85d2b250840 | 162 bytes | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | 5038823f-c955-4b92-935e-e51e77049191 | 167 bytes | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | a8d7cfaa-d6f1-49ce-a80e-96f61c120ef | 167 bytes | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | 0d06ab82-f0d7-450c-b329-b8e931a5f6 | 140 bytes | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | 2178287f-a11a-4c6f-a5ce-4ba4e919b22d | 156 bytes | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | df724720-0faf-450a-9785-fa13a7cd60bf | 160 bytes | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | da95278d-9764-4e50-9b6d-b6d68a304782 | 155 bytes | SplitText | SplitText | | |
| 03/20/2024 14:02:21.836 CET | DROP | 09c8bd1d-9037-4b7b-a5d4-4e9202844eb5 | 159 bytes | SplitText | SplitText | | |

Last updated: 20:30:30 CET

NiFi Data Provenance

Displaying 3 of 3
Oldest event available: 03/20/2024 11:49:27 CET

Showing the events that match the specified query. Clear search

| Filter | by component name | | | | | | |
|-----------------------------|-------------------|---------------------------------|---------|----------------|----------------|--|--|
| Date/Time | Type | FlowFile Uuid | Size | Component Name | Component Type | | |
| 03/20/2024 19:28:36.637 CET | RECEIVE | ce6cbc88-866f-4fdb-a69d-c244... | 1.54 MB | GetFile | GetFile | | |
| 03/20/2024 15:01:33.440 CET | RECEIVE | 82da7cdd-4cb2-42f3-a934-cc65... | 1.54 MB | GetFile | GetFile | | |
| 03/20/2024 14:01:49.912 CET | RECEIVE | d5e86250-376a-4dd4-aad8-007f... | 1.54 MB | GetFile | GetFile | | |

Provenance Event

DETAILS

ATTRIBUTES

CONTENT

rwxrwx---

filename

CRM.e8ae7d9c-6738-4395-81c5-9f90d818b44c

UAM_Clientes_Completo.csv (previous)

firstword

CRM

firstword.0

CRM,

firstword.1

CRM

fragment.count

9999

fragment.identifier

1d8d8610-19cc-46b6-a244-8dd378dea154

fragment.index

9999

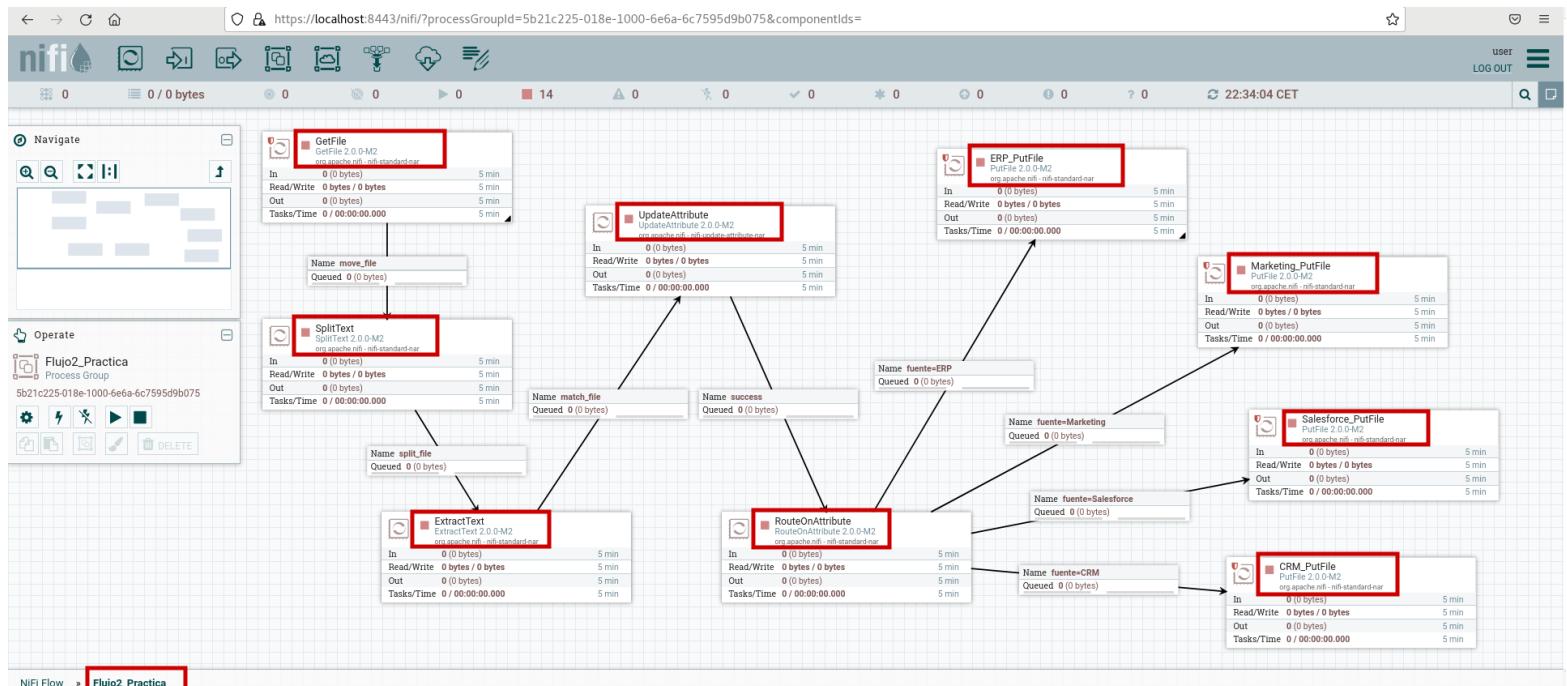
fragment.size

OK

FLUJO 2: Dejar en un directorio todos los ficheros que vengan de una misma fuente. Es decir, en un directorio se encontrarán todas las entradas de datos que procedan de ERP, en otro directorio todas las entradas que vengan de Marketing, un tercer directorio con todas las entradas que vengan de CRM y un último con los registros de Salesforce.

Primero, en Flujo2_Practica creamos procesadores llamados GetFile, SplitText, ExtractText,

UpdateAttribute, RouteOnAttribute y cuatro PutFile con el símbolo Processor



Configuración de GetFile

Mantenemos el nombre de GetFile por ser un nombre descriptivo y fácil de recordar. Indicamos la ruta del archivo DataIn.

Configure Processor | GetFile 2.0.0-M2

SETTINGS
SCHEDULING
PROPERTIES
RELATIONSHIPS
COMMENTS

⚠ Invalid

| Required field | |
|------------------------|----------------------------|
| Property | Value |
| Input Directory | Practica2_nifi/Data/DataIn |
| File Filter | [^\].* |
| Path Filter | No value set |
| Batch Size | 10 |
| Keep Source File | false |
| Recurse Subdirectories | true |
| Polling Interval | 0 sec |
| Ignore Hidden Files | true |
| Minimum File Age | 0 sec |
| Maximum File Age | No value set |
| Minimum File Size | 0 B |
| Maximum File Size | No value set |

CANCEL
APPLY

Desactivamos la opción "terminate" para mantener la conexión y añadimos una descripción.

Configure Processor | GetFile 2.0.0-M2

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Automatically Terminate / Retry Relationships ⓘ

success

terminate retry

All files are routed to success

CANCEL APPLY

Configure Processor | GetFile 2.0.0-M2

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Este es el primer procesador, que coge los datos del directorio DataIn

CANCEL APPLY

Configuración de SplitText

Configuramos Line Split Count en 1 para dividir cada fila.

Configure Processor | SplitText 2.0.0-M2

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Required field

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

Cancel Apply

No seleccionamos “splits terminate” para dividir los archivos.

Configure Processor | SplitText 2.0.0-M2

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Automatically Terminate / Retry Relationships ?

failure

terminate retry

If a file cannot be split for some reason, the original file will be routed to this destination and nothing will be routed elsewhere

original

terminate retry

The original input file will be routed to this destination when it has been successfully split into 1 or more files

splits

terminate retry

The split files will be routed to this destination when an input file is successfully split into 1 or more split files

Cancel Apply

Conexión entre GetFile y SplitText

Creamos una conexión llamada "move_file" para conectar GetFile y SplitText.

Create Connection

DETAILS **SETTINGS**

| | |
|---|-----------------|
| From Processor | To Processor |
| GetFile | SplitText |
| GetFile | SplitText |
| Within Group | Within Group |
| Flujo2_Practica | Flujo2_Practica |
| For Relationships | |
| <input checked="" type="checkbox"/> success | |

CANCEL **ADD**

Create Connection

DETAILS **SETTINGS**

| | |
|-----------------------------------|---------------------------------------|
| Name | Available Prioritizers |
| move_file | FirstInFirstOutPrioritizer |
| Id | NewestFlowFileFirstPrioritizer |
| No value set | OldestFlowFileFirstPrioritizer |
| FlowFile Expiration | PriorityAttributePrioritizer |
| 0 sec | |
| Back Pressure Object Threshold | Selected Prioritizers |
| 10000 | |
| Size Threshold | |
| 1 GB | |
| Load Balance Strategy | |
| Do not load balance | |

CANCEL **ADD**

Configuración de ExtractText

Configure Processor | ExtractText 2.0.0-M2

■ Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Name
ExtractText Enabled

Id
5c3906f7-018e-1000-f562-c88f2992e4b6

Type
ExtractText 2.0.0-M2

Bundle
org.apache.nifi - nifi-standard-nar

Penalty Duration ?
30 sec

Yield Duration ?
1 sec

Bulletin Level ?
WARN

CANCEL APPLY

Agregamos una nueva propiedad llamada "firstword" para encontrar la primera palabra en cada fila.

Configure Processor | ExtractText 2.0.0-M2

■ Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Required field

| Property | Value | | |
|---|---------|--|--|
| Enable Unicode Equivalence | false | | |
| Enable Case-insensitive Matching | false | | |
| Permit Whitespace and Comments in Pattern | 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 Classes | false | | |
| Enable Unix Lines Mode | false | | |
| Include Capture Group 0 | true | | |
| Enable repeating capture group | false | | |
| Enable named group support | false | | |
| firstword | ^(.+?), | | |

CANCEL APPLY

No seleccionamos la opción "match terminate" para continuar con la coincidencia de archivos.

Configure Processor | ExtractText 2.0.0-M2

Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Automatically Terminate / Retry Relationships ⓘ

matched
 terminate retry

FlowFiles are routed to this relationship when the Regular Expression is successfully evaluated and the FlowFile is modified as a result

unmatched
 terminate retry

FlowFiles are routed to this relationship when no provided Regular Expression matches the content of the FlowFile

CANCEL APPLY

Conexión entre SplitText y ExtractText

Creamos una conexión llamada "split_file" para conectar SplitText y ExtractText.

Create Connection

DETAILS SETTINGS

From Processor To Processor

| | |
|-----------------|-----------------|
| SplitText | ExtractText |
| SplitText | ExtractText |
| Within Group | Within Group |
| Flujo2_Practica | Flujo2_Practica |

For Relationships

failure
 original
 splits

CANCEL ADD

Create Connection

| DETAILS | SETTINGS |
|--|---|
| Name split_file | Available Prioritizers ? FirstInFirstOutPrioritizer NewestFlowFileFirstPrioritizer OldestFlowFileFirstPrioritizer PriorityAttributePrioritizer |
| Id No value set | |
| FlowFile Expiration ? 0 sec | Selected Prioritizers ? |
| Back Pressure Object Threshold ? 10000 | Size Threshold ? 1 GB |
| Load Balance Strategy ? Do not load balance | |
| CANCEL ADD | |

Configuración de UpdateAttribute

Añadimos una nueva propiedad llamada "filename" para determinar el nombre de fichero de salida. El nombre del fichero de salida estará formado por la primera palabra de la fila y un identificador único para evitar nombres similares.

Configure Processor | UpdateAttribute 2.0.0-M2

■ Stopped

| SETTINGS | SCHEDULING | PROPERTIES | RELATIONSHIPS | 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 | | | | |
| filename | ? \${firstword}-\${uuid} | | | | |

Conexión entre ExtractText y UpdateAttribute

Creamos una conexión llamada "match_file" para conectar ExtractText y UpdateAttribute.

Create Connection

DETAILS **SETTINGS**

| | |
|-----------------|-----------------|
| From Processor | To Processor |
| ExtractText | UpdateAttribute |
| ExtractText | UpdateAttribute |
| Within Group | Within Group |
| Flujo2_Practica | Flujo2_Practica |

For Relationships

matched

unmatched

Create Connection

DETAILS **SETTINGS**

| | |
|--|---------------------------------------|
| Name | Available Prioritizers ? |
| match_file | FirstInFirstOutPrioritizer |
| Id | NewestFlowFileFirstPrioritizer |
| No value set | OldestFlowFileFirstPrioritizer |
| FlowFile Expiration ? | PriorityAttributePrioritizer |
| 0 sec | |
| Back Pressure Object Threshold ? | Selected Prioritizers ? |
| 10000 | |
| Size Threshold ? | |
| 1 GB | |
| Load Balance Strategy ? | |
| Do not load balance | |

Configuración de RouteOnAttribute

Añadimos las nuevas propiedades "fuentes=CRM", "fuentes=ERP", "fuentes=Marketing" y "fuentes=Salesforce". Cuando la primera palabra de una fila coincide con el nombre de un fichero de salida, se dirige a ese fichero.

Configure Processor | RouteOnAttribute 2.0.0-M2

■ Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Required field

| Property | Value | |
|-------------------|---------------------------------------|--|
| Routing Strategy | Route to Property name | |
| fuente=CRM | \$(firstword.0>equals('CRM')) | |
| fuente=ERP | \$(firstword.0>equals('ERP')) | |
| fuente=Marketing | \$(firstword.0>equals('Marketing,')) | |
| fuente=Salesforce | \$(firstword.0>equals('Salesforce,')) | |

Seleccionamos la opción "unmatched terminate" para que el proceso pueda seguir coincidiendo.

Configure Processor | RouteOnAttribute 2.0.0-M2

■ Stopped

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Automatically Terminate / Retry Relationships

fuente=CRM
 terminate retry

fuente=ERP
 terminate retry

fuente=Marketing
 terminate retry

fuente=Salesforce
 terminate retry

unmatched
 terminate retry

FlowFiles that do not match any user-defined expression will be routed here

Conexión entre UpdateAttribute y RouteOnAttribute

Creamos una conexión para conectar UpdateAttribute y RouteOnAttribute.

Create Connection

DETAILS **SETTINGS**

| | |
|--|--|
| From Processor UpdateAttribute UpdateAttribute Within Group Flujo2_Practica | To Processor RouteOnAttribute RouteOnAttribute Within Group Flujo2_Practica |
| For Relationships <input checked="" type="checkbox"/> success | |

Create Connection

DETAILS **SETTINGS**

| | |
|---|--|
| Name | Available Prioritizers ? |
| Id | FirstInFirstOutPrioritizer |
| No value set | NewestFlowFileFirstPrioritizer |
| FlowFile Expiration ? | OldestFlowFileFirstPrioritizer |
| 0 sec | PriorityAttributePrioritizer |
| Back Pressure Object Threshold ? | Selected Prioritizers ? |
| 10000 | |
| Size Threshold ? | |
| 1 GB | |
| Load Balance Strategy ? | |
| Do not load balance | |

Configuración de ERP_PutFile, Marketing_PutFile, Salesforce_PutFile y CRM_PutFile

Establecemos el nombre de PutFile como "ERP_PutFile", "Marketing_PutFile", "Salesforce_PutFile" y "CRM_PutFile"

Configure Processor | PutFile 2.0.0-M2

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Required field

| Property | Value |
|------------------------------|---|
| Directory | Practica2_nifi/Data/DataOut/\${firstword} |
| Conflict Resolution Strategy | fail |
| Create Missing Directories | true |
| Maximum File Count | No value set |
| Last Modified Time | No value set |
| Permissions | No value set |
| Owner | No value set |
| Group | No value set |

CANCEL APPLY

Activamos la opción "terminate" para finalizar la conexión.

Configure Processor | PutFile 2.0.0-M2

⚠ Invalid

SETTINGS SCHEDULING PROPERTIES RELATIONSHIPS COMMENTS

Automatically Terminate / Retry Relationships [?](#)

failure
 terminate retry

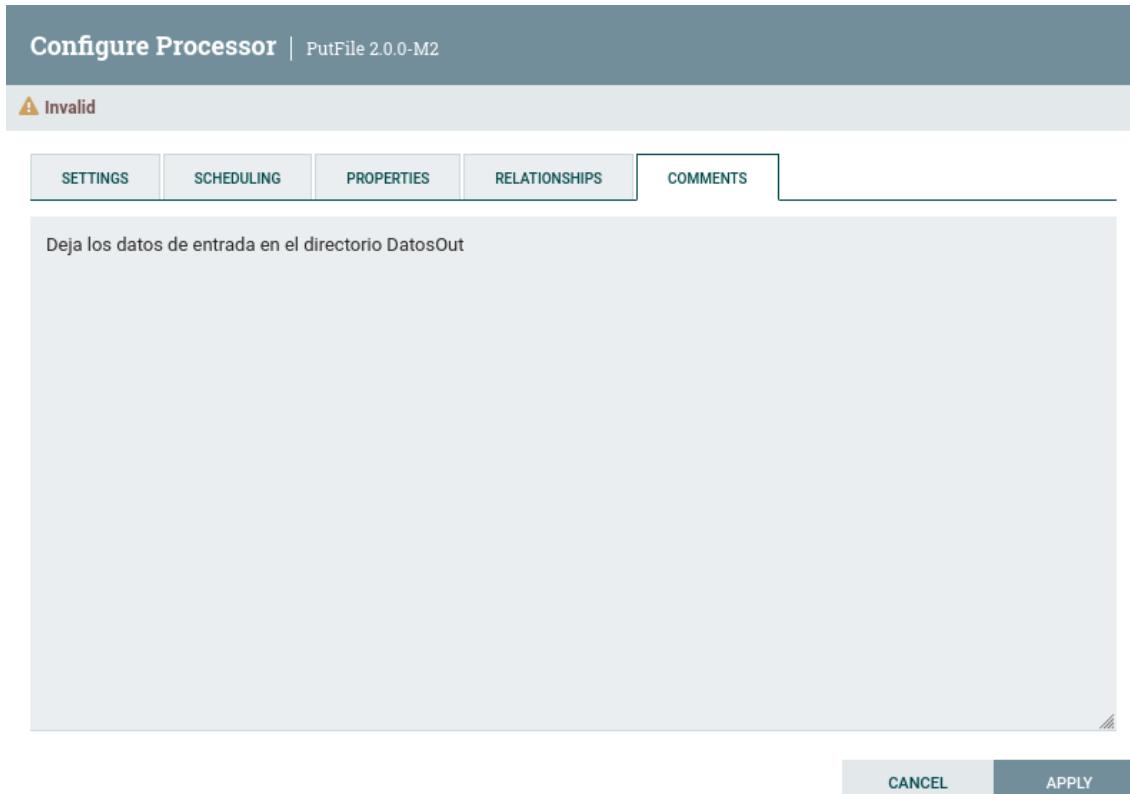
Files that could not be written to the output directory for some reason are transferred to this relationship

success
 terminate retry

Files that have been successfully written to the output directory are transferred to this relationship

CANCEL APPLY

Añadimos una descripción.



Conexión entre RouteOnAttribute y PutFile

Creamos una conexión para conectar RouteOnAttribute y PutFile y seleccionamos con qué archivo queremos que coincida.



Create Connection

DETAILS

SETTINGS

From Processor

RouteOnAttribute

RouteOnAttribute

Within Group

Flujo2_Practica

To Processor

Marketing_PutFile

PutFile

Within Group

Flujo2_Practica

For Relationships

fuente=CRM

fuente=ERP

fuente=Marketing

fuente=Salesforce

unmatched

Create Connection

DETAILS

SETTINGS

From Processor

RouteOnAttribute

RouteOnAttribute

Within Group

Flujo2_Practica

To Processor

Salesforce_PutFile

PutFile

Within Group

Flujo2_Practica

For Relationships

fuente=CRM

fuente=ERP

fuente=Marketing

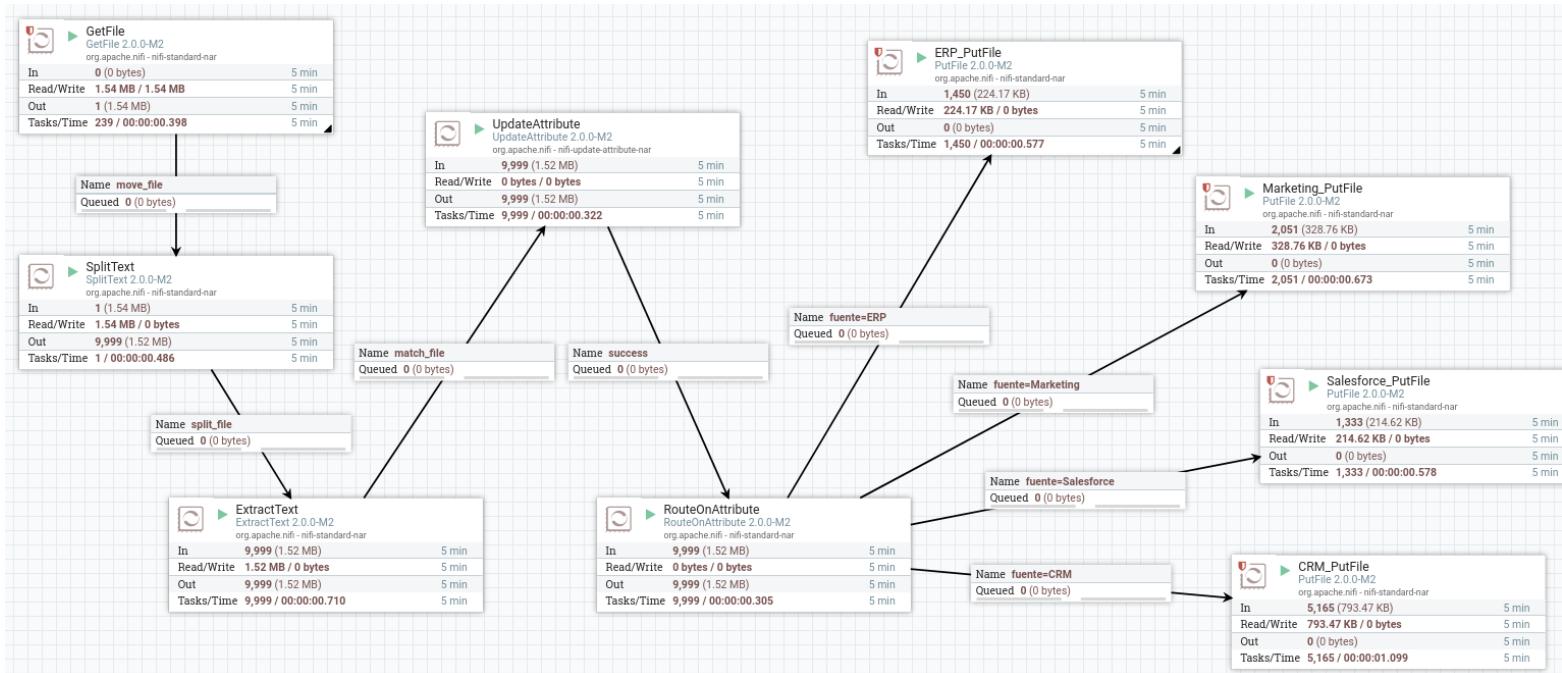
fuente=Salesforce

unmatched

Create Connection



Se observa que el "Flujo 2" funciona correctamente al ejecutar todos los procesadores.



NiFi Data Provenance

Podemos revisar data provenance of PutFile, RouteOnAttribute o UpdateAttribute.
Dado que ejecutamos los procesadores varias veces, las horas pueden ser diferentes.

NiFi Data Provenance

| Showing 1,000 of 1,000 events that match the specified query, please refine the search. Clear search | | | | | | |
|--|-----------------------------|-------|---------------------------------------|-----------|------------------|------------------|
| Filter | Date/Time | Type | FlowFile Uuid | Size | Component Name | Component Type |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | 48f35e58-a01f-44f7-bf66-4194b4a2a8ea | 153 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | e34bcbe05-d31b-4b98-b68b-c3a20f196a8 | 153 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | 2a2ebca0-fa60-48a2-97d1-e27ae02f883 | 156 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | db213ecd-8765-4232-9b85-54de9f00c2c2 | 149 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | 73838a5d-b5e4-4d2f-b20f-be9a90e7d513 | 138 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | fb10a798-2b97-470f-b0d7-abe571e886fa | 156 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | d3ed005-3a19-44f7-a1e9-84985a332871 | 142 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | 20332414-759d-4aaa-8bad-cfb005237531 | 151 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | e4a7df46-829c-495c-bbf8-60c33161af7e | 148 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | 7ec487a0-7979-4061-9af7-3750a74dd1d21 | 165 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | fb3c8c08-5d06-44f7-8e4d-2f18aa9fb58 | 159 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | 56008678-1362-4720-be06-1331526489d7 | 161 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | 91beb20-de71-482c-98df-1690cd91bf6d | 156 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | c12a220d-3902-4703-a0f7-3aa9f98b1241 | 167 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | 3f00b42e-3b4d-4d1d-b2b5-8f2b7971d41a | 141 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | 9a962c56-7599-4a9c-b55a-a8d047ab6253 | 148 bytes | RouteOnAttribute | RouteOnAttribute |
| 0 | 03/20/2024 22:22:06.714 CET | ROUTE | 75cfaf83-9a3f-4ccb0-ac7-ab56f5e29be | 154 bytes | RouteOnAttribute | RouteOnAttribute |

Provenance Event

| DETAILS | ATTRIBUTES | CONTENT |
|---|------------|---------|
| Type ROUTE | | |
| FlowFile Uuid 48f35e58-a01f-44f7-bf66-4194b4a2a8ea | | |
| File Size 153 bytes | | |
| Component Id 5c39c63d-018e-1000-5ab7-9093fe16debe | | |
| Component Name RouteOnAttribute | | |
| Component Type RouteOnAttribute | | |
| Relationship fuente=CRM | | |
| Details No value set | | |

OK

El nombre del fichero puede parecer diferente aquí porque hemos cambiado el nombre del fichero después.

Provenance Event

DETAILS ATTRIBUTES CONTENT

file.lastModifiedTime
2018-11-04T11:59:49+0100

file.owner
umaster

file.permissions
rwxrwx---

filename
48f35e58-a01f-44f7-bf66-4194b4a2a8ea-CRM-UAM_Clientes_Completo.csv

firstword
CRM

firstword.0
CRM,

firstword.1
CRM

fragment.count
9999

Provenance Identifier

Provenance Event

DETAILS ATTRIBUTES CONTENT

< 1ms

Lineage Duration
00:00:01.152

Type
DROP

FlowFileUuid
8a3d2c50-b91a-480f-8643-d5f179d1736f

File Size
170 bytes

Component Id
5c3a64eb-018e-1000-a8a3-51985a54bac5

Component Name
ERP_PutFile

Component Type
PutFile

Details
Auto-Terminated by success Relationship

Child FlowFiles (0)
No children

OK

Provenance Event

| DETAILS | ATTRIBUTES | CONTENT |
|---|------------|--|
| /home/umaster/nifi-2.0.0-M2/Practica2_nifi/Data/DataIn/ | | |
| file.creationTime | | 2018-11-04T11:59:49+0100 |
| file.group | | umaster |
| file.lastAccessTime | | 2024-03-20T22:22:05+0100 |
| file.lastModifiedTime | | 2018-11-04T11:59:49+0100 |
| file.owner | | umaster |
| file.permissions | | rwxrwx--- |
| filename | | 8a3d2c50-b91a-480f-8643-d5f179d1736f-ERP-UAM_Clientes_Completo.csv |
| firstword | | ERP |

NiFi Data Provenance

Displaying 1,000 of 1,000
Oldest event available: 03/20/2024 11:49:27 CET

Showing 1,000 of 1,000+ events that match the specified query, please refine the search. [Clear search](#)

| Filter | by component name | ▼ | Search | | |
|-----------------------------|-------------------|--------------------------------------|-----------|----------------|----------------|
| Date/Time | Type | FlowFileUuid | Size | Component Name | Component Type |
| 03/20/2024 22:22:06.530 CET | DROP | 8a3d2c50-b91a-480f-8643-d5f179d1736f | 170 bytes | ERP_PutFile | PutFile |
| 03/20/2024 22:22:06.530 CET | SEND | 8a3d2c50-b91a-480f-8643-d5f179d1736f | 170 bytes | ERP_PutFile | PutFile |
| 03/20/2024 22:22:06.530 CET | DROP | 4ed1aea-0370-4536-819a-e09263f35477f | 157 bytes | ERP_PutFile | PutFile |
| 03/20/2024 22:22:06.530 CET | SEND | 4ed1aea-0370-4536-819a-e09263f35477f | 157 bytes | ERP_PutFile | PutFile |
| 03/20/2024 22:22:06.530 CET | DROP | b74d172e-9d29-48e3-95e7-3732776b508a | 152 bytes | ERP_PutFile | PutFile |
| 03/20/2024 22:22:06.530 CET | SEND | b74d172e-9d29-48e3-95e7-3732776b508a | 152 bytes | ERP_PutFile | PutFile |
| 03/20/2024 22:22:06.524 CET | DROP | 106926af-c660-4d92-9df8-8226db7720ba | 152 bytes | ERP_PutFile | PutFile |
| 03/20/2024 22:22:06.524 CET | SEND | 106926af-c660-4d92-9df8-8226db7720ba | 152 bytes | ERP_PutFile | PutFile |
| 03/20/2024 22:22:06.522 CET | DROP | c62f9cb0-be02-4a13-8f2b-63d7a5b1cc1d | 157 bytes | ERP_PutFile | PutFile |
| 03/20/2024 22:22:06.522 CET | SEND | c62f9cb0-be02-4a13-8f2b-63d7a5b1cc1d | 157 bytes | ERP_PutFile | PutFile |
| 03/20/2024 22:22:06.522 CET | DROP | 616d184f-3190-47e2-a540-a8c52ee81f96 | 157 bytes | ERP_PutFile | PutFile |
| 03/20/2024 22:22:06.522 CET | SEND | 616d184f-3190-47e2-a540-a8c52ee81f96 | 157 bytes | ERP_PutFile | PutFile |
| 03/20/2024 22:22:06.521 CET | DROP | d6313c4a-9b15-464a-9c9c-e5ff34bcd429 | 150 bytes | ERP_PutFile | PutFile |

Realizamos una verificación para determinar si los archivos han sido exportados a DataOut. Observamos que todos los archivos han sido exportados en archivos diferentes. Comprobamos cuántos archivos hay dentro de cada uno.

| Name | Size | Modified |
|------------|-------------|----------|
| CRM | 5,165 items | 23:02 |
| ERP | 1,450 items | 23:02 |
| Marketing | 2,051 items | 23:02 |
| Salesforce | 1,333 items | 23:02 |

```
[umaster@ibmuamdocker nifi-2.0.0-M2]$ cd Practica2_nifi/  
[umaster@ibmuamdocker Practica2_nifi]$ ls  
Data  
[umaster@ibmuamdocker Practica2_nifi]$ cd Data/  
[umaster@ibmuamdocker Data]$ ls DataIn/  
[umaster@ibmuamdocker Data]$ ls DataOut/  
CRM ERP Marketing Salesforce
```

También podemos ver el número de archivos que solo consisten en ERP, CRM, Salesforce y Marketing.

```
[umaster@ibmuamdocker DataOut]$ cd ERP/  
[umaster@ibmuamdocker ERP]$ ls ERP* | wc -l  
1450  
[umaster@ibmuamdocker ERP]$ cd ..  
[umaster@ibmuamdocker DataOut]$ cd CRM/  
[umaster@ibmuamdocker CRM]$ ls CRM* | wc -l  
5165  
[umaster@ibmuamdocker CRM]$ cd ..  
[umaster@ibmuamdocker DataOut]$ cd Salesforce/  
[umaster@ibmuamdocker Salesforce]$ ls Salesforce* | wc -l  
1333  
[umaster@ibmuamdocker Salesforce]$ cd ..  
[umaster@ibmuamdocker DataOut]$ cd Marketing/  
[umaster@ibmuamdocker Marketing]$ ls Marketing* | wc -l  
2051  
[umaster@ibmuamdocker Marketing]$ █
```

| Name | Size | Modified |
|--|-----------|----------|
| CRM-0a0a45eb-23dc-4cfe-b2a5-374f755331f4 | 144 bytes | 23:02 |
| CRM-0a01ba1a-310e-4fd5-835c-2b0d00913cb2 | 135 bytes | 23:02 |
| CRM-0a2b04b6-140d-4db7-949b-e952c5beb2e0 | 153 bytes | 23:02 |
| CRM-0a4e5a9e-8af0-4f3a-9ac8-30cb8ec4145d | 158 bytes | 23:02 |
| CRM-0a05c9ab-3b9e-45bc-a3ac-176d582d8c87 | 155 bytes | 23:02 |
| CRM-0a42ac2a-7308-47d1-9f76-7ad260d9a43a | 157 bytes | 23:02 |
| CRM-0a73bdf8-fe8a-401c-91b2-5fecae074d0 | 165 bytes | 23:02 |
| CRM-0a86f08c-ab7e-4d04-952b-79d3bc27e9f7 | 131 bytes | 23:02 |
| CRM-0a89a3cf-da57-465f-a2ba-f3b9c153578d | 142 bvtex | 23:02 |

| Name | Size | Modified |
|--|-----------|----------|
| ERP-00a8c919-6f15-49e6-b700-a204d9eb67c1 | 172 bytes | 23:02 |
| ERP-0a0403be-efb5-4aaa-bf1d-00eb415f2978 | 162 bytes | 23:02 |
| ERP-0a613d9c-78ec-4f7f-842b-9f511615bba9 | 155 bytes | 23:02 |
| ERP-0a8348b6-de25-4ee1-991e-1076320eaf01 | 150 bytes | 23:02 |
| ERP-0ae366fc-8e33-4ba7-8f3d-ad638c79958b | 169 bytes | 23:02 |
| ERP-0afab484-d65f-4860-92d2-16469ead2593 | 155 bytes | 23:02 |
| ERP-0b0d7d8a-2df9-4232-a44b-d6086e8af511 | 151 bytes | 23:02 |
| ERP-0b3edfd1-179e-462f-9939-9bcf838532bd | 169 bytes | 23:02 |
| ERP-0b22c1cc-ac98-4177-9d13-cf336ed67ef3 | 153 bytes | 23:02 |

| Name | Size | Modified |
|--|-----------|----------|
| Marketing-0a2e6c1d-b48e-4842-ae1-466c75721001 | 168 bytes | 23:02 |
| Marketing-0a5d6f1e-ea14-4d3f-b8b6-ad29385556a1 | 164 bytes | 23:02 |
| Marketing-0a5f8a1d-a14b-4abb-9057-3d939509c46c | 159 bytes | 23:02 |
| Marketing-0a148e72-1d44-473b-a486-7a984175f68b | 166 bytes | 23:02 |
| Marketing-0a356b2a-ec75-4060-a6a1-22a6ba40687d | 173 bytes | 23:02 |
| Marketing-0a809c45-7d5a-4ca1-b634-6aa24fde87a8 | 155 bytes | 23:02 |
| Marketing-0a299750-ce5a-4c4e-b97d-9ee6710740f7 | 169 bytes | 23:02 |
| Marketing-0ada7ed1-9f9e-422d-88c4-57b5396743de | 165 bytes | 23:02 |
| Marketing-0b1b7395-5b6f-4459-b0f2-e1e9bf1a7005 | 172 bytes | 23:02 |
| Marketing-0b2a2050-4acb-4643-84a2-31dc6f9e3e65 | 144 bytes | 23:02 |

| Name | Size | Modified |
|---|-----------|----------|
| Salesforce-0a162af8-b7f1-4da7-a3ba-7cfb06c09134 | 148 bytes | 23:02 |
| Salesforce-0ab52bd7-02ca-477d-8993-ce7d5f9273cf | 157 bytes | 23:02 |
| Salesforce-0ad3b33a-fba5-4b7f-866d-6741c2d96fcc | 160 bytes | 23:02 |
| Salesforce-0af72bde-f832-45e5-9ec7-0ac9bae71b26 | 166 bytes | 23:02 |
| Salesforce-0b50f4ff-0f51-4356-b0dc-8f3022437e70 | 166 bytes | 23:02 |
| Salesforce-0bb1a224-28b2-4521-854b-b8d1f35758b9 | 176 bytes | 23:02 |
| Salesforce-0bf0b73c-6eb1-479e-856f-db7cd084d8d9 | 177 bytes | 23:02 |
| Salesforce-0bf4312a-2fb5-49f3-ba14-2141a42df66e | 165 bytes | 23:02 |
| Salesforce-0c2b7af7-7cce-4854-8dde-51132a6248fd | 164 bytes | 23:02 |
| Salesforce-0c3f6330-d9b5-4f1d-ae6d-b49e1b33d440 | 174 bytes | 23:02 |
| Salesforce-0cfc49d4-0r9d-4ah6-9216-rff8a3adfe8h | 166 bvtex | 23:02 |

CRM-0a0a45eb-23dc-4cfe-b2a5-374f755331f4
~/nifi-2.0.0-M2/Practica2_nifi/Data/DataOut/CRM

Save - x

CRM,JUNE,GIBBY,5244 NORTH RUE FERMAT,HAYNESVILLE,VA,22472,927-1898,1445 WEST CALHOUN STREET,HARTFORD,CT,6101,475,319-9639,,489-26-0125,1/26/62,F

Plain Text ▾ Tab Width: 8 ▾ Ln 1, Col 1 ▾ INS

ERP-00a8c919-6f15-49e6-b700-a204d9eb67c1
~/nifi-2.0.0-M2/Practica2_nifi/Data/DataOut/ERP

Save - x

ERP,TIMOTHY,REASER,9289 SOUTHEAST JOSEPH E. LOWERY, BECKLEY,WV,25802,573-1345,2574 SOUTHEAST JOHNSON ST.,EL MONTE,CA,91732,626,741-5039,662-65-8547,152-00-6713,06/04/1959,M,

Plain Text ▾ Tab Width: 8 ▾ Ln 1, Col 1 ▾ INS

Marketing-0a2e6c1d-b48e-4842-aee1-466c75721001
~/nifi-2.0.0-M2/Practica2_nifi/Data/DataOut/Marketing

Save - x

Marketing,LEONARD,JOHNSON,9017 WEST ST. ADELINE ST.,MANVEL,ND,58256,361-2127,4912 WEST ST. ANDREW ST.,LAWLER,IA,52154,563,273-1802,188-05-6034,312-62-7256,05/01/1990,M,

Plain Text ▾ Tab Width: 8 ▾ Ln 1, Col 1 ▾ INS

Salesforce-0ab52bd7-02ca-477d-8993-ce7d5f9273cf
~/nifi-2.0.0-M2/Practica2_nifi/Data/DataOut/Salesforce

Save - x

Salesforce,YURI,ADKINS,3707 SOUTH PROTECTION ST.,SWINK,OK,74761,807-1304,5439 S YORK STREET,KEMPTON,PA,19529,610,198-3496,134-89-9269,186-22-3905,10/22/55,F,

Plain Text ▾ Tab Width: 8 ▾ Ln 1, Col 1 ▾ INS

Parte Opcional

1. Incorporar cada registro final a una base de datos

¿Qué procesadores utilizarías?

Utilizaría los procesadores llamados GetFile, SplitText, ExtractText, UpdateAttribute, RouteOnAttribute y PutFile o PutDatabaseRecord o PutElasticsearchRecord

¿Con qué orden los encadenarías?

Encadenaría con orden:

GetFile > SplitText > ExtractText > UpdateAttribute > RouteOnAttribute > PutFile

GetFile > SplitText > ExtractText > UpdateAttribute > RouteOnAttribute > PutDatabaseRecord

GetFile > SplitText > ExtractText > UpdateAttribute > RouteOnAttribute > PutElasticsearchRecord

¿Qué aspectos consideras especialmente relevantes?

Puede ser necesario definir "firstword" en propiedades de ExtractText y "filename" en properties de UpdateAttribute, como se hizo en los pasos anteriores, para que cada nombre de archivo sea único. También es importante definir la ruta de acceso a los archivos para "GetFile" y "PutFile" y el valor de "Line Split Count : 1 en properties de SplitText.

2. Plantear el caso en el que el campo discriminador de registros inicial no fuera la fuente de datos (que viene recogida en el primer campo del fichero) sino la fecha de nacimiento

¿Qué procesadores utilizarías?

Utilizaría también los procesadores llamados GetFile, SplitText, ExtractText, UpdateAttribute, RouteOnAttribute y PutFile.

¿Con qué orden los encadenarías?

Encadenaría con orden: GetFile > SplitText > ExtractText > UpdateAttribute > RouteOnAttribute > PutFile

¿Qué aspectos consideras especialmente relevantes?

En las propiedades de ExtractText, definimos la variable "firstword" para buscar y encontrar la primera palabra de cada fila. La expresión regular utilizada para "firstword" es " $^(.+?)$," que incluye todos los caracteres como cadenas de texto, números enteros o caracteres especiales como "/" o "-".

Si la fecha de nacimiento se encuentra en la primera columna, podemos utilizar "firstword" para obtenerla. Sin embargo, no es necesario utilizar "firstword" para asignar un nombre al archivo.

En las propiedades de UpdateAttribute, podemos definir el nombre del archivo utilizando la siguiente sintaxis:

filename: \${firstword}-\${uuid}-\${filename}: Esta opción utiliza "firstword", un UUID y el nombre original del archivo para generar el nombre final.

filename: \${uuid}-\${filename}: Esta opción solo utiliza un UUID y el nombre original del archivo para generar el nombre final.

Puede ser necesario definir "firstword" en propiedades de ExtractText y "filename" en properties de UpdateAttribute, como se hizo en los pasos anteriores, para que cada nombre de archivo sea único.

También es importante definir la ruta de acceso a los archivos para "GetFile" y "PutFile" y el valor de "Line Split Count : 1 en properties de SplitText.