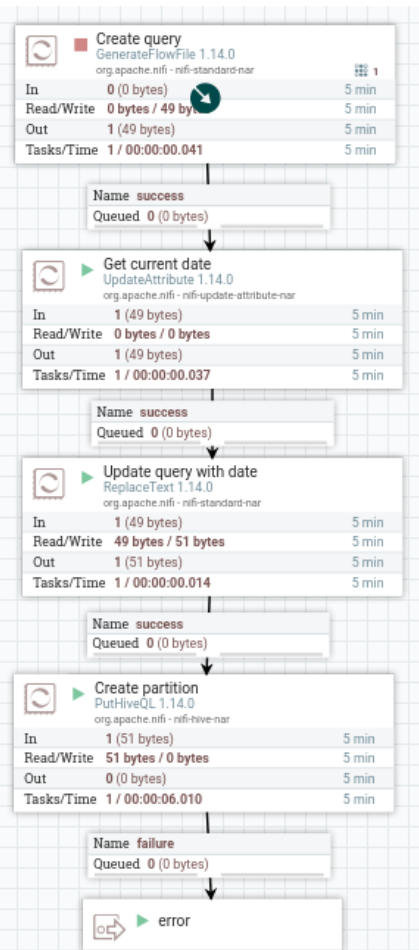


## Raport z testów

Komponent	Kroki	Oczekiwany wynik	Potwierdzenie rezultatów
Tworzenie partycji w nifi	<ol style="list-style-type: none"><li>Uruchomienie flow w NiFi <i>Create new partition</i></li><li>Ręczne wywołanie <i>GenerateFlowFile - Create query</i></li></ol>	W wyniku zapytania <code>show partitions trams;</code> powinna się znaleźć partycja z dzisiejszą datą (2022-02-12).	<p>Przed uruchomieniem flow:</p> <pre>hive&gt; show partitions trams; OK day=2021-12-26 day=2022-01-08 2 rows selected (1.878 seconds)</pre> <p>Flow:</p>



Wynik:

			<pre>hive&gt; show partitions trams; OK day=2021-12-26   --nie rezultatów day=2022-01-08 day=2022-01-12 3 rows selected (0.279 seconds)</pre>
<p>Pobieranie informacji o lokalizacji pojazdów i wrzucanie na Kafkę</p>	<ol style="list-style-type: none"> <li>1. Uruchomienie consumera na topicu input-trams</li> <li>2. Uruchomienie flow <i>Get trams &amp; put into kafka &amp; HDFS</i></li> </ol>	<p>W terminalu z consumerem powinny pojawić się JSONy z informacjami o lokalizacji tramwajów</p>	<p>Consumer:</p> <pre>java -Dlog4j2.configurationFile=/usr/local/kafka\$ ./bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic input-trams --group test</pre> <pre>{   "Lines": "4",   "Lon": 21.012999,   "VehicleNumber": "1123+1122",   "Time": "2022-01-12 23:27:11",   "Lat": 52.311127,   "Brigade": "015" }</pre> <pre>{   "Lines": "22",   "Lon": 21.05114,   "VehicleNumber": "1227+1228",   "Time": "2022-01-12 23:27:11",   "Lat": 52.239037,   "Brigade": "3" }</pre> <pre>{   "Lines": "17",   "Lon": 20.992174,   "VehicleNumber": "1229+1230",   "Time": "2022-01-11 16:29:06",   "Lat": 52.180542,   "Brigade": "017" }</pre> <pre>{   "Lines": "26",   "Lon": 20.979766,   "VehicleNumber": "1255+1256",   "Time": "2022-01-12 23:27:08",   "Lat": 52.236706,   "Brigade": "015" }</pre> <pre>{   "Lines": "22",   "Lon": 20.92926,   "VehicleNumber": "1265+1266",   "Time": "2022-01-12 23:27:12",   "Lat": 52.289944,   "Brigade": "4" }</pre> <pre>{   "Lines": "26",   "Lon": 20.99287,   "VehicleNumber": "1279+1280",   "Time": "2022-01-10 13:17:49",   "Lat": 52.24177,   "Brigade": "10" }</pre> <pre>{   "Lines": "22",   "Lon": 20.980774,   "VehicleNumber": "1281+1282",   "Time": "2022-01-12 23:27:11",   "Lat": 52.25231,   "Brigade": "6" }</pre> <pre>{   "Lines": "18",   "Lon": 20.972689,   "VehicleNumber": "1285+1286",   "Time": "2022-01-12 23:27:11",   "Lat": 52.273102,   "Brigade": "11" }</pre> <pre>{   "Lines": "24",   "Lon": 21.065001,   "VehicleNumber": "1295+1296",   "Time": "2022-01-12 23:27:07",   "Lat": 52.24455,   "Brigade": "017" }</pre> <pre>{   "Lines": "33",   "Lon": 20.929848,   "VehicleNumber": "1299+1300",   "Time": "2022-01-12 23:27:08",   "Lat": 52.292496,   "Brigade": "11" }</pre> <pre>{   "Lines": "35",   "Lon": 20.92285,   "VehicleNumber": "1309+1310",   "Time": "2022-01-12 23:27:12",   "Lat": 52.2847,   "Brigade": "016" }</pre> <p>Flow:</p>

<p>Dumpowanie pobieranych danych o lokalizacji pojazdów</p>	<p>1. Uruchomienie flow <i>Get trams &amp; put into kafka &amp; HDFS</i></p>	<p>Nowe niepuste pliki znalazły się w lokalizacji <code>/data/in</code></p>	<pre>vagrant@node1: /usr/local/kafka\$ hdfs dfs -ls /data/in SLF4J: Class path contains multiple SLF4J bindings. SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.7.6/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar/] SLF4J: Found binding in [jar:file:/usr/local/apache-tez-0.9.1-bin/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class] SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation. SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory] Found 662 items -rw-r--r-- 1 root supergroup 11246 2022-01-08 13:55 /data/in/007ae456-0be7-41cf-bca7-c968855e4247 -rw-r--r-- 1 root supergroup 11246 2022-01-08 12:53 /data/in/0090ea8c-1c0f-4b6f-92ce-7a05522aca23 -rw-r--r-- 1 root supergroup 9216 2022-01-12 21:09 /data/in/0140df24-3aaa-49f2-8594-3d0d00fdddec5 -rw-r--r-- 1 root supergroup 6026 2022-01-12 22:04 /data/in/016b0c4d-182b-4473-87e4-4d19d821339c -rw-r--r-- 1 root supergroup 11246 2022-01-08 14:33 /data/in/01884678-a8b8-42cf-9e4c-9c0442cca216 -rw-r--r-- 1 root supergroup 11142 2022-01-08 12:22 /data/in/01c88364-305e-4d66-b3df-c812dd943ac1 -rw-r--r-- 1 root supergroup 5267 2022-01-12 22:19 /data/in/01da7001-723c-4661-97aa-c97345a38da2</pre>
<p>Flink - informacje o szybkich tramwajach</p>	<p>1. Uruchomienie flow <i>Get trams &amp; put into kafka &amp; HDFS</i>  2. Uruchomienie consumera na topicu <code>fast-trams</code>  3. Uruchomienie joba flinkowego</p>	<p>W terminalu z consumerem powinny pojawić się JSONy z informacjami o tramwajach i ich prędkości</p>	<pre>vagrant@node1: /usr/local/kafka\$ ./bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic fast-trams --group test {"Lon":20.978687,"Time":"2022-01-12 23:24:38","velocity":829.0316486492674,"VehicleNumber":"3158","Lat":52.236397,"Lines":"26","Brigade":"3"} {"Lon":20.940779,"Time":"2022-01-12 23:24:36","velocity":78.92382781442227,"VehicleNumber":"3199","Lat":52.300953,"Lines":"1","Brigade":"022"}</pre>

<p>Flink - deduplikacja z wrzucaniem na Kafkę</p>	<ol style="list-style-type: none"><li>1. Uruchomienie flow <i>Get trams &amp; put into kafka &amp; HDFS</i></li><li>2. Uruchomienie consumera na topicu output-trams</li><li>3. Uruchomienie joba flinkowego</li></ol>	<p>W terminalu z consumerem powinny pojawiać się JSONy z informacjami o tramwajach</p>	<p>Consumer:</p> <pre>agrzeta@w0401:~/usr/local/kafka\$ ./bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic output-trams --group test</pre> <pre>{   "Lon":21.012997,"Time":"2022-01-12 23:28:06","VehicleNumber":"1123+1122","Lat":52.311134,"Lines":"4","Brigade":"015"} {"Lon":21.050026,"Time":"2022-01-12 23:28:06","VehicleNumber":"1227+1228","Lat":52.24118,"Lines":"22","Brigade":"3"} {"Lon":20.97521,"Time":"2022-01-12 23:28:03","VehicleNumber":"1255+1256","Lat":52.23546,"Lines":"26","Brigade":"015"} {"Lon":20.9302,"Time":"2022-01-12 23:28:05","VehicleNumber":"1265+1266","Lat":52.290894,"Lines":"22","Brigade":"41"} {"Lon":20.981632,"Time":"2022-01-12 23:28:06","VehicleNumber":"1281+1282","Lat":52.254044,"Lines":"22","Brigade":"6"} {"Lon":20.968704,"Time":"2022-01-12 23:28:06","VehicleNumber":"1285+1286","Lat":52.273,"Lines":"18","Brigade":"11"} {"Lon":21.088823,"Time":"2022-01-12 23:27:54","VehicleNumber":"1295+1296","Lat":52.245045,"Lines":"24","Brigade":"017"} {"Lon":20.933258,"Time":"2022-01-12 23:28:07","VehicleNumber":"1299+1300","Lat":52.296997,"Lines":"33","Brigade":"11"} {"Lon":20.926647,"Time":"2022-01-12 23:27:59","VehicleNumber":"1309+1310","Lat":52.287804,"Lines":"35","Brigade":"016"} {"Lon":20.982431,"Time":"2022-01-12 23:28:07","VehicleNumber":"1313+1314","Lat":52.25486,"Lines":"35","Brigade":"12"} {"Lon":20.924517,"Time":"2022-01-12 23:28:10","VehicleNumber":"1377+1378","Lat":52.281242,"Lines":"26","Brigade":"16"}</pre> <p>Job:</p> <p>DeduplicationJob <span>RUNNING</span> 2</p> <p>ID: 63d8abe3ece5d0f3cbbc8b03fa2b153   Start Time: 2022-01-12 22:17:00   Duration: 1h 12m 32s <a href="#">Cancel Job</a></p> <p><a href="#">Overview</a> <a href="#">Exceptions</a> <a href="#">TimeLine</a> <a href="#">Checkpoints</a> <a href="#">Configuration</a></p> <pre>graph LR     A[Source: Custom Source -&gt; Map Parallelism: 1 Backpressured (max): 0% Busy (max): N/A] -- "map" --&gt; B[KeyedProcess -&gt; Map -&gt; Sink: Unnamed Parallelism: 1 Backpressured (max): 0% Busy (max): 0%]</pre> <table><tr><th>Name</th><th>Status</th><th>Bytes Received</th><th>Records Received</th><th>Bytes Sent</th><th>Records Sent</th><th>Parallelism</th><th>Start Time</th><th>Duration</th><th>Tasks</th></tr><tr><td>Source: Custom Source -&gt; Map</td><td><span>RUNNING</span></td><td>0 B</td><td>0</td><td>1.22 MB</td><td>20,614</td><td>1</td><td>2022-01-12 22:17:01</td><td>1h 12m 32s</td><td><span>1</span></td></tr><tr><td>KeyedProcess -&gt; Map -&gt; Sink: Unnamed</td><td><span>RUNNING</span></td><td>1.22 MB</td><td>20,614</td><td>0 B</td><td>0</td><td>1</td><td>2022-01-12 22:17:01</td><td>1h 12m 32s</td><td><span>1</span></td></tr></table>	Name	Status	Bytes Received	Records Received	Bytes Sent	Records Sent	Parallelism	Start Time	Duration	Tasks	Source: Custom Source -> Map	<span>RUNNING</span>	0 B	0	1.22 MB	20,614	1	2022-01-12 22:17:01	1h 12m 32s	<span>1</span>	KeyedProcess -> Map -> Sink: Unnamed	<span>RUNNING</span>	1.22 MB	20,614	0 B	0	1	2022-01-12 22:17:01	1h 12m 32s	<span>1</span>
Name	Status	Bytes Received	Records Received	Bytes Sent	Records Sent	Parallelism	Start Time	Duration	Tasks																								
Source: Custom Source -> Map	<span>RUNNING</span>	0 B	0	1.22 MB	20,614	1	2022-01-12 22:17:01	1h 12m 32s	<span>1</span>																								
KeyedProcess -> Map -> Sink: Unnamed	<span>RUNNING</span>	1.22 MB	20,614	0 B	0	1	2022-01-12 22:17:01	1h 12m 32s	<span>1</span>																								
<p>Wysyłanie maila z informacją o szybkich tramwajach</p>	<ol style="list-style-type: none"><li>1. Uruchomienie flow <i>Get trams &amp; put into kafka &amp; HDFS</i></li><li>2. Uruchomienie joba flinkowego</li><li>3. Uruchomienie flow <i>Alerts with fast trams</i></li></ol>	<p>Po jakimś czasie powinny zacząć przychodzić maile o szybkich tramwajach</p>	<p>Maile:</p> <p>Szybkie tramwaje w twojej okolicy <a href="#">Odebrane x</a></p> <p>artystadziura@gmail.com</p> <pre>{   "Lon":20.933113,"Time":"2022-01-12 23:45:16","velocity":75.63209895020613,"VehicleNumber":"3153","Lat":52.296677,"Lines":"23","Brigade":"8"}</pre> <p>Flow:</p>																														

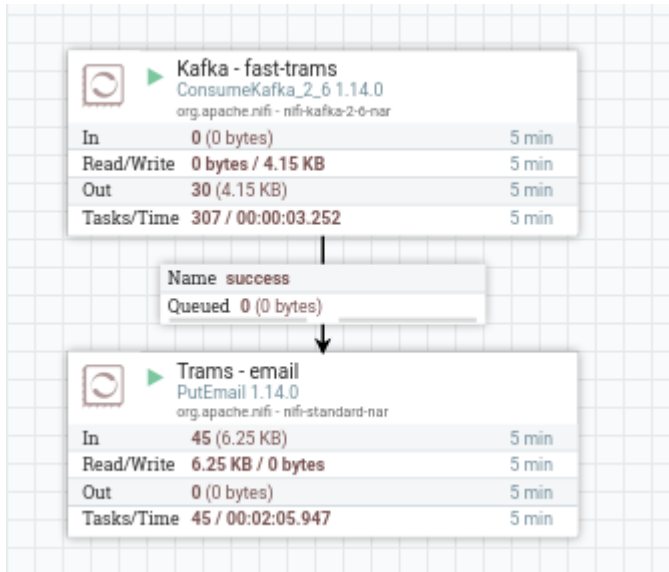
			 <p>The screenshot shows a NiFi flow with two processors. The first processor, 'Kafka - fast-trams' (ConsumeKafka_2_6 1.14.0), has an input of 0 bytes and outputs 30 (4.15 KB) over 5 minutes. The second processor, 'Trams - email' (PutEmail 1.14.0), has an input of 45 (6.25 KB) and outputs 0 bytes over 5 minutes. The flow is connected by a 'Name success' relationship, with 'Queued 0 (0 bytes)' shown between them.</p>
Zapisywanie danych o lokalizacji tramwajów na HDFSie	<ol style="list-style-type: none"> <li>1. Uruchomienie flow <i>Get trams &amp; put into kafka &amp; HDFS</i></li> <li>2. Uruchomienie joba flinkowego</li> <li>3. Uruchomienie flow <i>Insert into hive table</i></li> </ol>	W /data/trams/ w odpowiedniej partycji pojawiają się nowe dane	<p>Zawartość katalogu:</p> <pre> program@node1:~/local/kafka\$ hdfs dfs -ls /data/trans/day=2022-01-12 SLF4J: Class path contains multiple SLF4J bindings. SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.7.6/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/ SLF4J: Found binding in [jar:file:/usr/local/apache-tez-0.9.1-bin/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLogger SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation. SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory] Found 134 items -rw-r--r-- 1 root supergroup 8045 2022-01-12 21:33 /data/trans/day=2022-01-12/06f071ce-cb54-4795-b065-4b1f67bede5f -rw-r--r-- 1 root supergroup 5292 2022-01-12 22:08 /data/trans/day=2022-01-12/0b495e83-a112-45ab-b792-05a9bd970eff -rw-r--r-- 1 root supergroup 6074 2022-01-12 21:32 /data/trans/day=2022-01-12/0bcd8252-d188-4b35-a2cd-8887bfe5e9a2 -rw-r--r-- 1 root supergroup 5990 2022-01-12 21:38 /data/trans/day=2022-01-12/11684d05-be03-4438-9d2c-5ce0e67952d0 -rw-r--r-- 1 root supergroup 4961 2022-01-12 22:16 /data/trans/day=2022-01-12/13f988c2-9b7d-4804-927e-981ee19e5964 -rw-r--r-- 1 root supergroup 6183 2022-01-12 21:26 /data/trans/day=2022-01-12/142edc9a-b442-44d5-b1f8-67b390f54b2c -rw-r--r-- 1 root supergroup 6170 2022-01-12 21:28 /data/trans/day=2022-01-12/18553aee-0771-4b4a-b8cd-547c33ab34bb -rw-r--r-- 1 root supergroup 5258 2022-01-12 22:10 /data/trans/day=2022-01-12/199bdc1e-1587-4680-9565-67220620fc76 -rw-r--r-- 1 root supergroup 6062 2022-01-12 21:41 /data/trans/day=2022-01-12/1ab17db1-5733-431d-9ece-7ae4a56dded4 -rw-r--r-- 1 root supergroup 6106 2022-01-12 21:18 /data/trans/day=2022-01-12/1f16fae6-6ed6-448f-804a-b19260318ad2 -rw-r--r-- 1 root supergroup 5450 2022-01-12 22:01 /data/trans/day=2022-01-12/22c28cf4-6558-4f63-928b-117538ddf5c4 </pre>

Tabela trams na Hive zawiera nowe dane

1. Uruchomienie flow *Get trams & put into kafka & HDFS*
2. Uruchomienie joba flinkowego
3. Uruchomienie flow *Insert into hive table*

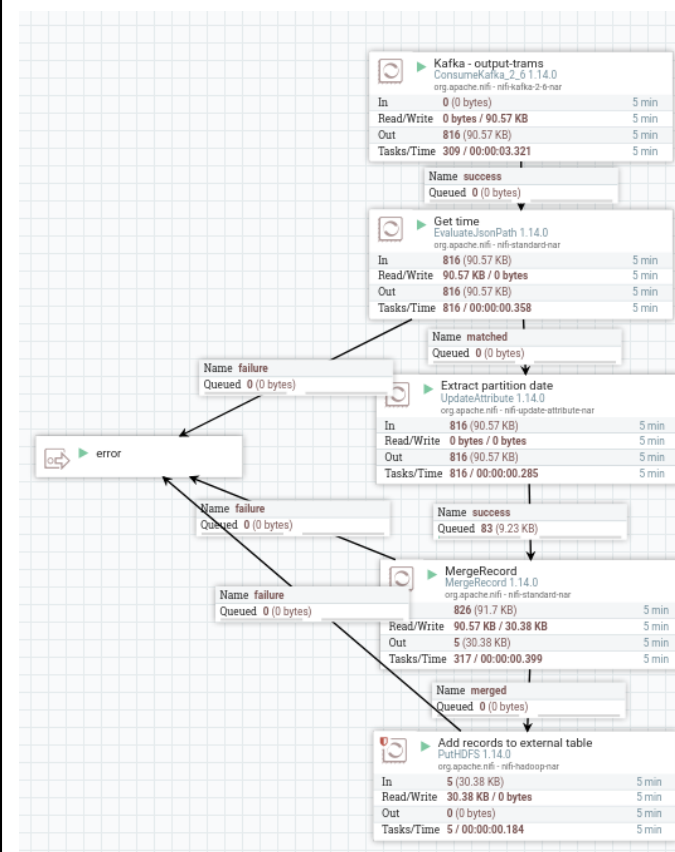
```
select * from trams
where day='<dzien>'
limit 5; zwraca rekordy
```

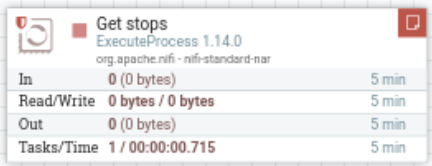
```
0: jdbc:hive2://localhost:10000> select * from trams where day='2022-01-12' limit 5;
```

trans.lines	trans.lon	trans.lat	trans.vehiclenumber	trans.brigade	trans.time	trans.day
6	21.108454	52.241116	1381	015	1642026774000	2022-01-12
28	20.9786	52.262447	3001	8	1642026774000	2022-01-12
25	21.050844	52.239616	3008	6	1642026770000	2022-01-12
25	21.02913	52.281624	3010	5	1642026773000	2022-01-12
22	20.986105	52.22841	3013	7	1642026773000	2022-01-12


5 rows selected (3.84 seconds)

Flow:



<p>Po uruchomieniu <i>Get stops</i> pojawia się plik stops.json</p>	<ol style="list-style-type: none"> <li>1. Pojedyncze uruchomienie procesora <i>Get stops</i></li> </ol>	<p>W folderze ~/trams/nifi powinien pojawić się plik stops.json</p>	<p>Przed uruchomieniem:</p> <pre>vagrant@node1:~/trams/nifi\$ ls stops.sh</pre> <p>Flow:</p>  <p>(NiFi za warningi przyjął logi ze skryptu)</p> <p>Po uruchomieniu:</p> <pre>vagrant@node1:~/trams/nifi\$ ls -la total 2528 drwxrwxr-x 2 vagrant vagrant 4096 Jan 12 22:18 . drwxrwxr-x 6 vagrant vagrant 4096 Jan 12 18:30 .. -rw-r--r-- 1 root root 2572578 Jan 12 22:18 stops.json -rwxrwxr-x 1 vagrant vagrant 170 Jan 12 18:08 stops.sh</pre>
<p>Flow <i>Reload stops</i> aktualizuje dane o przystankach</p>	<ol style="list-style-type: none"> <li>1. Pojedyncze uruchomienie procesora <i>Get stops</i></li> <li>2. Uruchomienie flow <i>Reload stops</i></li> </ol>	<p>Powinien zaktualizować się plik w /data/stops</p>	<p>Przed:</p> <pre>vagrant@node1:~/trams/nifi\$ hdfs dfs -ls /data/stops SLF4J: Class path contains multiple SLF4J bindings. SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.7.6/share/hadoop/common/lib/slf4j- SLF4J: Found binding in [jar:file:/usr/local/apache-tez-0.9.1-bin/lib/slf4j-log4j12-1.7 SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation. SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory] Found 1 items -rw-r--r-- 1 root supergroup 617974 2022-01-13 18:52 /data/stops/stops.json.avro</pre> <p>Po:</p> <pre>vagrant@node1:~/trams/nifi\$ hdfs dfs -ls /data/stops SLF4J: Class path contains multiple SLF4J bindings. SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.7.6/share/hadoop/common/lib/slf4j- SLF4J: Found binding in [jar:file:/usr/local/apache-tez-0.9.1-bin/lib/slf4j-log4j12-1.7 SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation. SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory] Found 1 items -rw-r--r-- 1 root supergroup 617974 2022-01-13 18:53 /data/stops/stops.json.avro</pre> <p>Flow:</p>






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

In	0 (0 bytes)	5 min
Read/Write	2.45 MB / 2.45 MB	5 min
Out	1 (2.45 MB)	5 min
Tasks/Time	36 / 00:00:00.080	5 min


Name success  
Queued 0 (0 bytes)



**SplitStops**  
Split.Json 1.14.0  
org.apache.nifi - nifi-standard-nar

In	1 (2.45 MB)	5 min
Read/Write	2.45 MB / 2.42 MB	5 min
Out	7,652 (2.42 MB)	5 min
Tasks/Time	1 / 00:00:00.757	5 min

Name split  
Queued 0 (0 bytes)



**Get Values**  
Evaluate.JsonPath 1.14.0  
org.apache.nifi - nifi-standard-nar

In	7,652 (2.42 MB)	5 min
Read/Write	2.42 MB / 0 bytes	5 min
Out	7,652 (2.42 MB)	5 min
Tasks/Time	7,652 / 00:00:04.891	5 min


Name matched  
Queued 0 (0 bytes)



**ValuesToCSV**  
AttributesToCSV 1.14.0  
org.apache.nifi - nifi-standard-nar

In	7,652 (2.42 MB)	5 min
Read/Write	0 bytes / 589.56 KB	5 min
Out	7,652 (589.56 KB)	5 min
Tasks/Time	7,652 / 00:00:05.182	5 min


Name success  
Queued 0 (0 bytes)



**Add newline to each line**  
ReplaceText 1.14.0  
org.apache.nifi - nifi-standard-nar


In	7,652 (589.56 KB)	5 min
Read/Write	589.56 KB / 597.03 KB	5 min
Out	7,652 (597.03 KB)	5 min
Tasks/Time	7,652 / 00:00:04.746	5 min

↓

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

In	7,652 (597.03 KB)	5 min
Read/Write	597.03 KB / 597.03 KB	5 min
Out	1 (597.03 KB)	5 min
Tasks/Time	48 / 00:00:01.443	5 min

Name merged  
Queued 0 (0 bytes)

**Add header**  
ReplaceText 1.14.0  
org.apache.nifi - nifi-standard-nar


In	1 (597.03 KB)	5 min
Read/Write	597.03 KB / 597.11 KB	5 min
Out	1 (597.11 KB)	5 min
Tasks/Time	1 / 00:00:00.017	5 min

Name success  
Queued 0 (0 bytes)

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


In	1 (597.11 KB)	5 min
Read/Write	597.11 KB / 603.49 KB	5 min
Out	1 (603.49 KB)	5 min
Tasks/Time	1 / 00:00:00.085	5 min

Name success  
Queued 0 (0 bytes)

**Change filename**  
UpdateAttribute 1.14.0  
org.apache.nifi - nifi-update-attribute-nar

In	1 (603.49 KB)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	1 (603.49 KB)	5 min
Tasks/Time	1 / 00:00:00.002	5 min

Name success  
Queued 0 (0 bytes)

**Add records to external table**  
PutHDFS 1.14.0  
org.apache.nifi - nifi-hadoop-nar

In	1 (603.49 KB)	5 min
Read/Write	603.49 KB / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	1 / 00:00:00.009	5 min



