



# Проблемы обработки Excel файлов в Apache NiFi и как их решать

Технологии. Стратегии. Консалтинг. Аутсорсинг.



# Обо мне



## Шемаров Игорь

- Middle Java разработчик в Ахених
- Пишу на Java 3 года
- Окончил ТвГУ

# План доклада

## #1

Что такое  
Apache NiFi

## #2

Зачем он  
используется

## #3

Описание  
задачи  
обработки  
.xlsx файлов

## #4

Способы  
обработки в  
NiFi

## #5

Проблема  
обработки  
файлов .xlsx в  
NiFi

## #6

Решение  
проблемы



# Что такое Apache NiFi

**Apache NiFi – это современный мощный open source инструмент для автоматизации процессов сбора, агрегации и передачи данных**

APACHE  
nifi

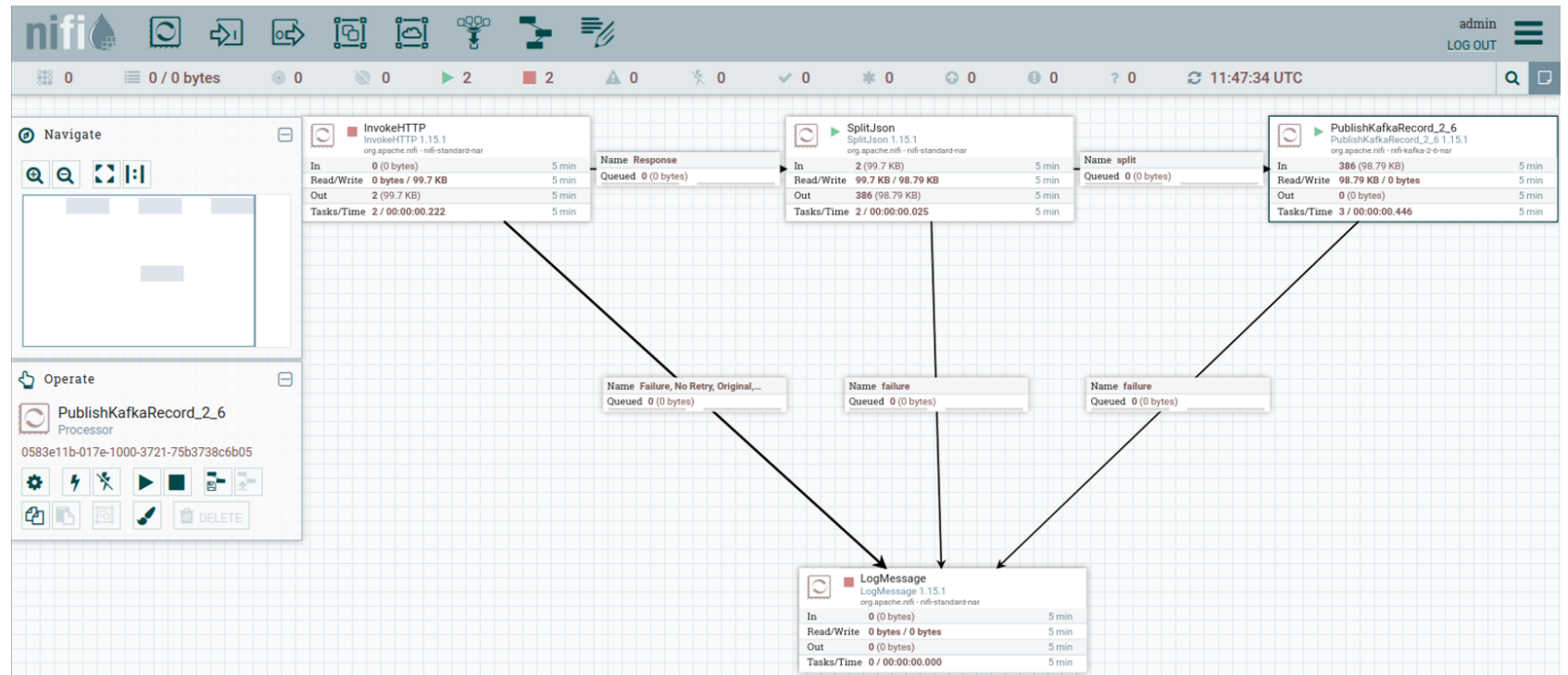




# Зачем он используется

Apache NiFi  
предоставляет  
интуитивно понятный  
веб-интерфейс для  
проектирования и  
мониторинга потоков  
данных

## Визуальное программирование

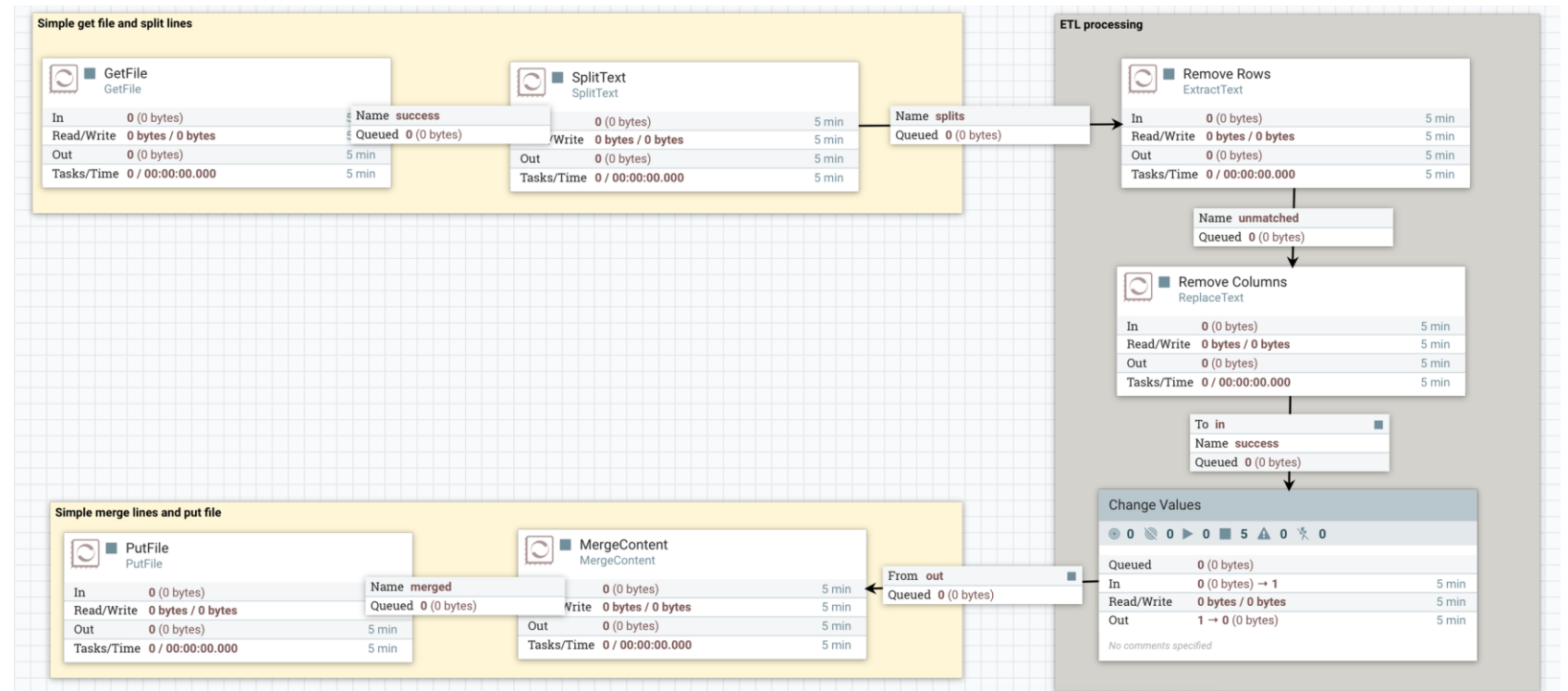




# Зачем он используется

Apache NiFi используется для сбора данных из различных источников

## ETL процесс



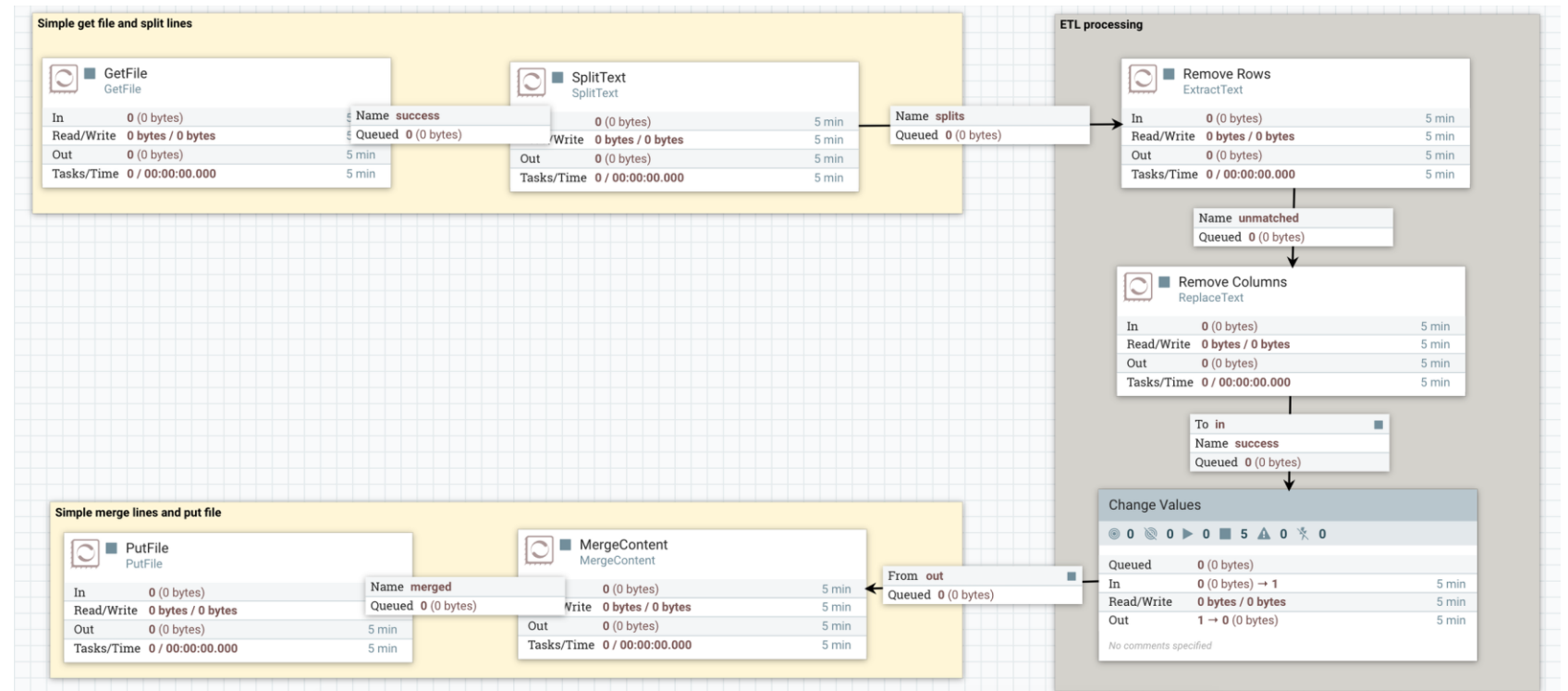


# Зачем он используется

Apache NiFi используется для сбора данных из различных источников

Он предоставляет средства для преобразования данных в реальном времени

## ETL процесс





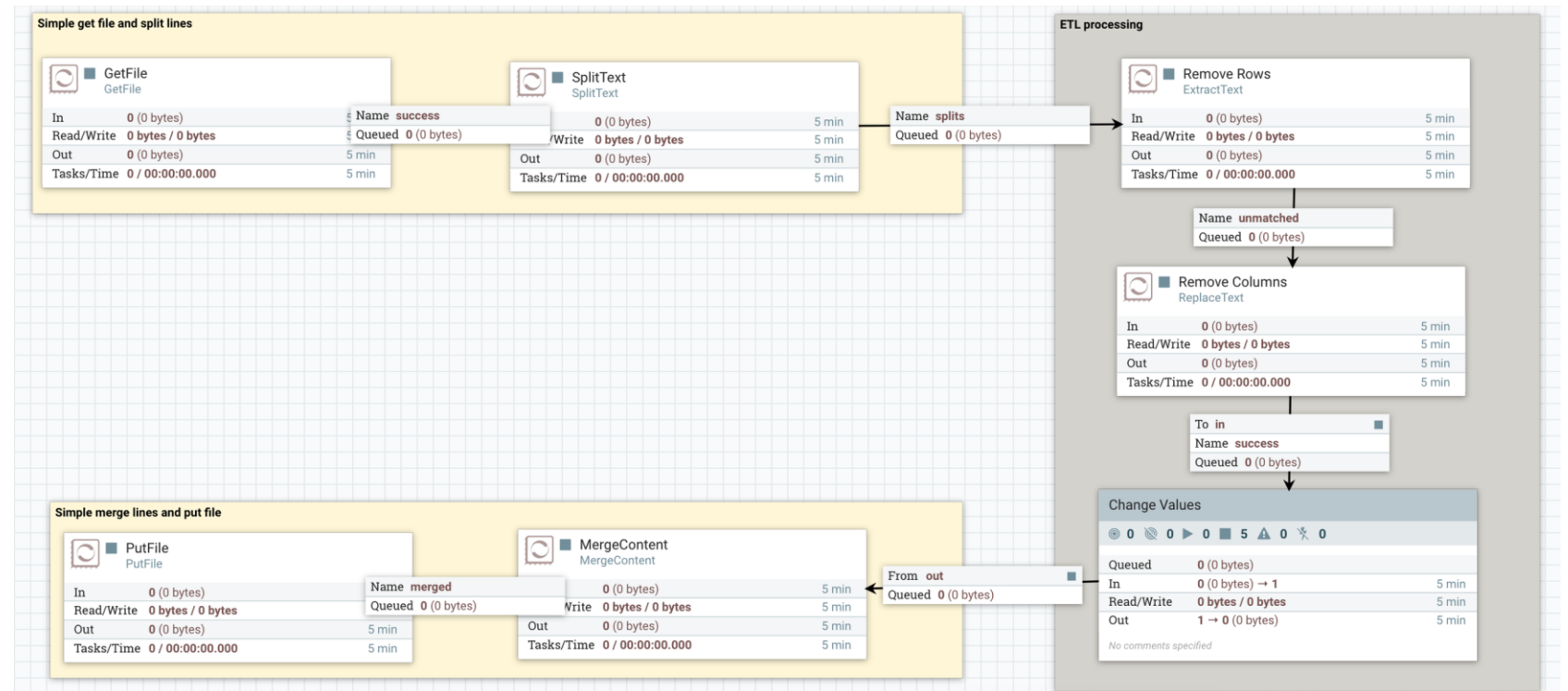
# Зачем он используется

Apache NiFi используется для сбора данных из различных источников

Он предоставляет средства для преобразования данных в реальном времени

Позволяет загружать данные в хранилища или другие внешние системы

## ETL процесс



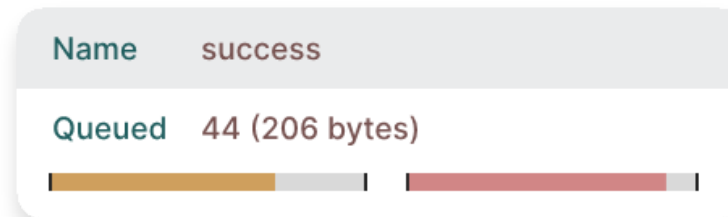




# Зачем он используется

**NiFi** дает возможность пользователям управлять потоком данных в реальном времени

## Управление данными



Available Prioritizers

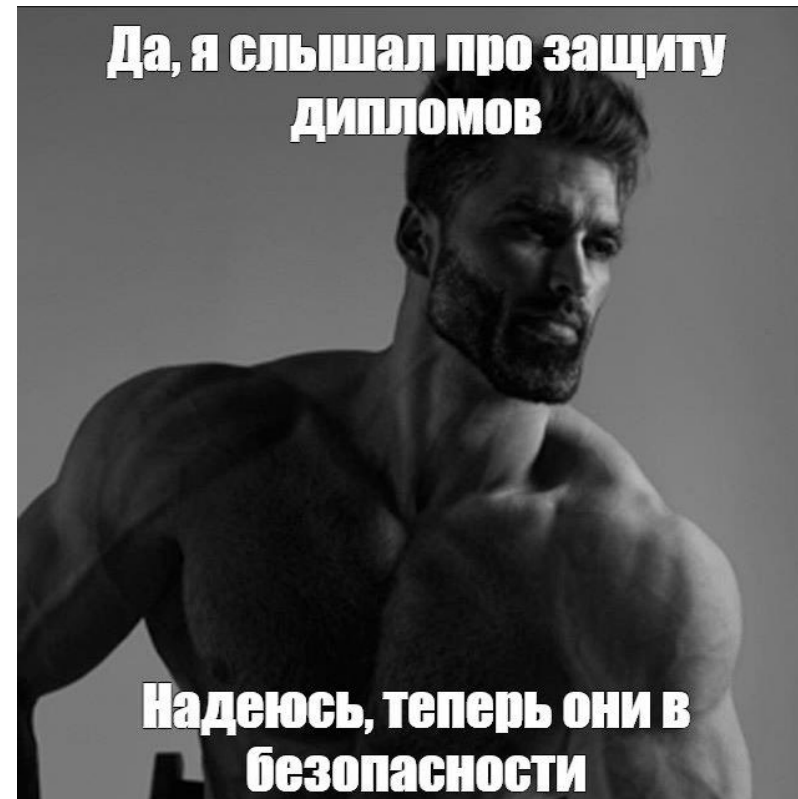
- FirstInFirstOutPrioritizer
- NewestFlowFileFirstPrioritizer
- OldestFlowFileFirstPrioritizer



# Зачем он используется

NiFi обеспечивает  
устойчивую и надежную  
обработку данных

## Устойчивость и надежность

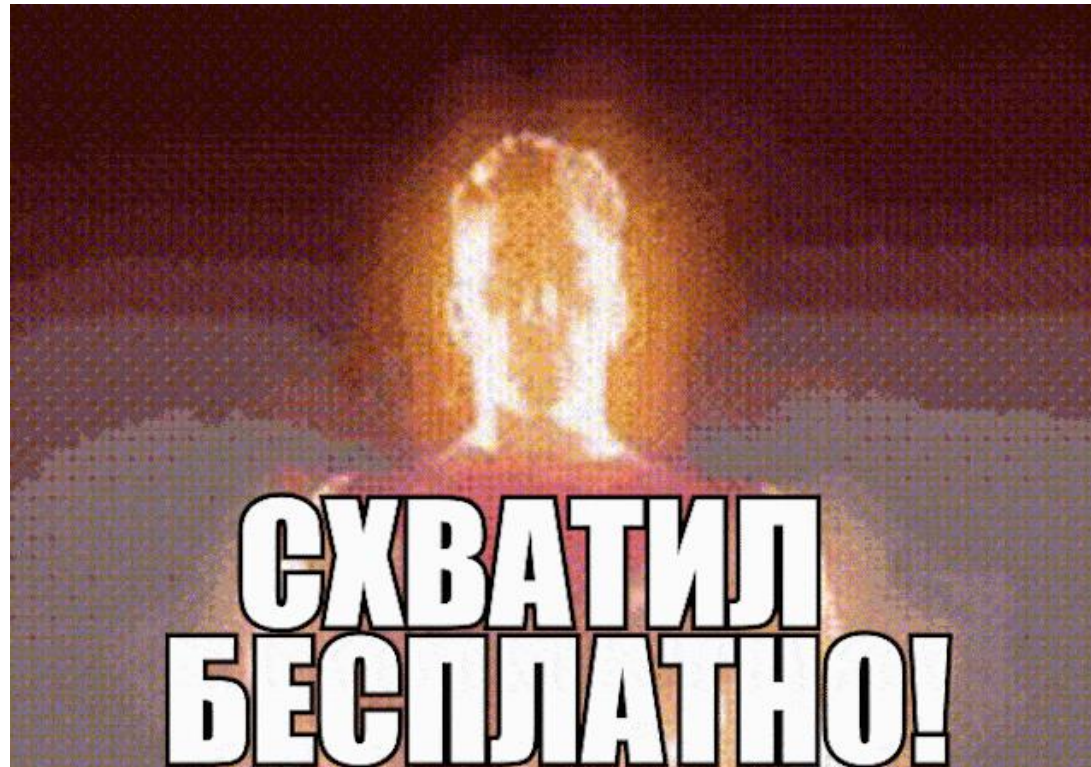




# Зачем он используется

## Снижение затрат

Apache NiFi является open source инструментом





# Зачем он используется

Apache NiFi может взаимодействовать с различными технологиями и поддерживает интеграцию с множеством приложений и систем.

## Поддерживаемые системы





# Зачем он используется

## Пользователи

**Apache NiFi часто используется в крупных средах, где существует необходимость в масштабируемых и гибких средствах обработки и управления данными**





# Обработка Excel-файлов

---

Обработка Excel-файлов в Apache NiFi может быть необходима в ситуациях, когда нужно:



# Обработка Excel-файлов

---

Обработка Excel-файлов в Apache NiFi может быть необходима в ситуациях, когда нужно:

- 1) Перенести данные из Excel в другую систему, такую как база данных или облачное хранилище.



# Обработка Excel-файлов

---

Обработка Excel-файлов в Apache NiFi может быть необходима в ситуациях, когда нужно:

- 1) Перенести данные из Excel в другую систему, такую как база данных или облачное хранилище.
- 2) Обработать большие объемы данных Excel и производить с ними различные операции, такие как фильтрация, трансформация, агрегация и т.д.





# Обработка Excel-файлов

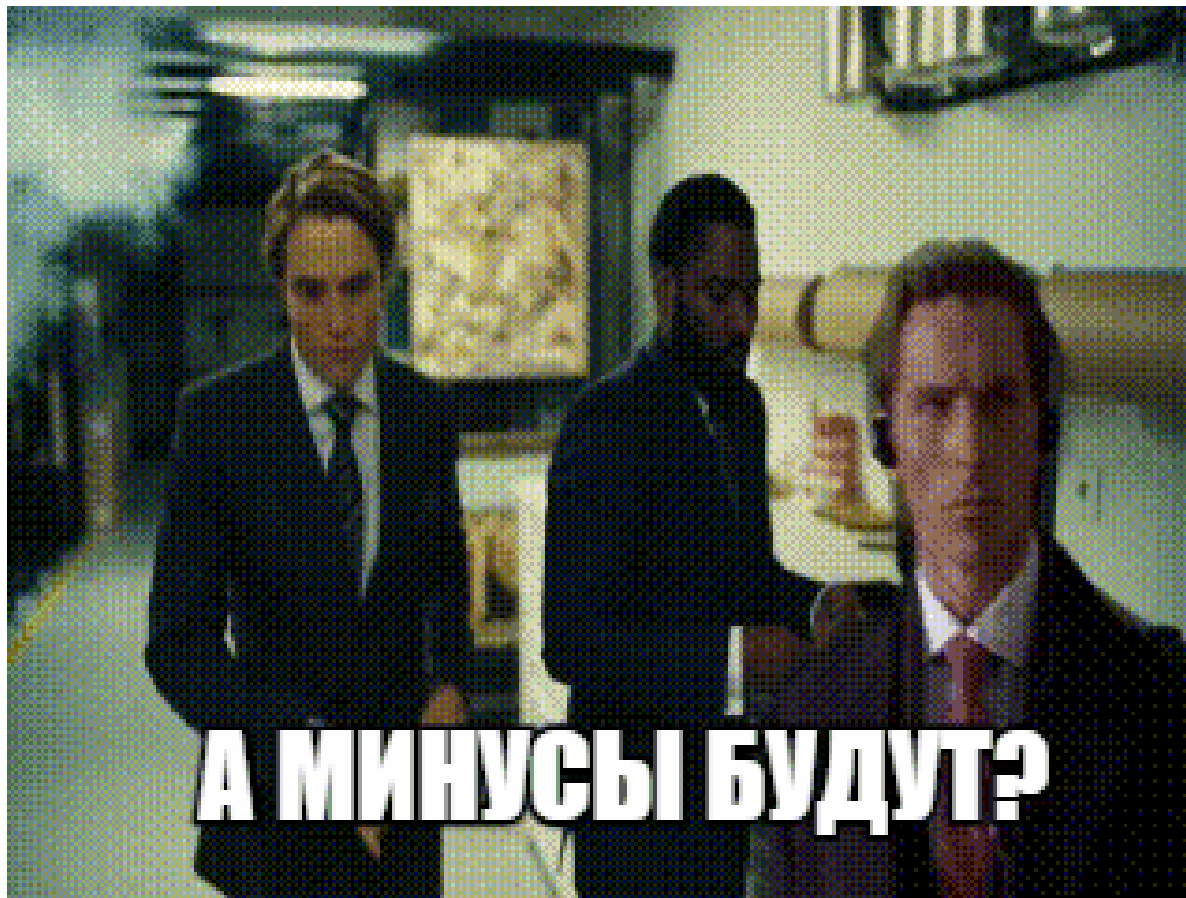
---

Обработка Excel-файлов в Apache NiFi может быть необходима в ситуациях, когда нужно:

- 1) Перенести данные из Excel в другую систему, такую как база данных или облачное хранилище.
- 2) Обработать большие объемы данных Excel и производить с ними различные операции, такие как фильтрация, трансформация, агрегация и т.д.
- 3) Автоматизировать процесс обработки данных Excel, особенно если он регулярно повторяется.





# Обработка Excel-файлов





# Проблема обработки файлов .xlsx в NiFi



- Apache NiFi не может напрямую взаимодействовать с файлами Excel

	 <b>ConvertExcelToCSVProcessor</b> ConvertExcelToCSVProcessor 1.14.0 org.apache.nifi - nifi-poi-nar	
In	<b>0</b> (0 bytes)	5 min
Read/Write	<b>0 bytes / 0 bytes</b>	5 min
Out	<b>0</b> (0 bytes)	5 min
Tasks/Time	<b>0 / 00:00:00.000</b>	5 min



# Проблема обработки файлов .xlsx в NiFi



- Apache NiFi не может напрямую взаимодействовать с файлами Excel
- .xlsx (XSSF 2007 OOXML)

	 <b>ConvertExcelToCSVProcessor</b> ConvertExcelToCSVProcessor 1.14.0 org.apache.nifi - nifi-poi-nar	
In	<b>0</b> (0 bytes)	5 min
Read/Write	<b>0 bytes / 0 bytes</b>	5 min
Out	<b>0</b> (0 bytes)	5 min
Tasks/Time	<b>0 / 00:00:00.000</b>	5 min



# Проблема обработки файлов .xlsx в NiFi



- Apache NiFi не может напрямую взаимодействовать с файлами Excel
- .xlsx (XSSF 2007 OOXML)
- .xls (HSSF '97(-2007))

	 <b>ConvertExcelToCSVProcessor</b> ConvertExcelToCSVProcessor 1.14.0 org.apache.nifi - nifi-poi-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



# Проблема обработки файлов .xlsx в NiFi



- Apache NiFi не может напрямую взаимодействовать с файлами Excel
- .xlsx (XSSF 2007 OOXML)
- .xls (HSSF '97(-2007))
- Проблемы при обработке сложных файлов

	 <b>ConvertExcelToCSVProcessor</b> ConvertExcelToCSVProcessor 1.14.0 org.apache.nifi - nifi-poi-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



# Проблема обработки файлов .xlsx в NiFi

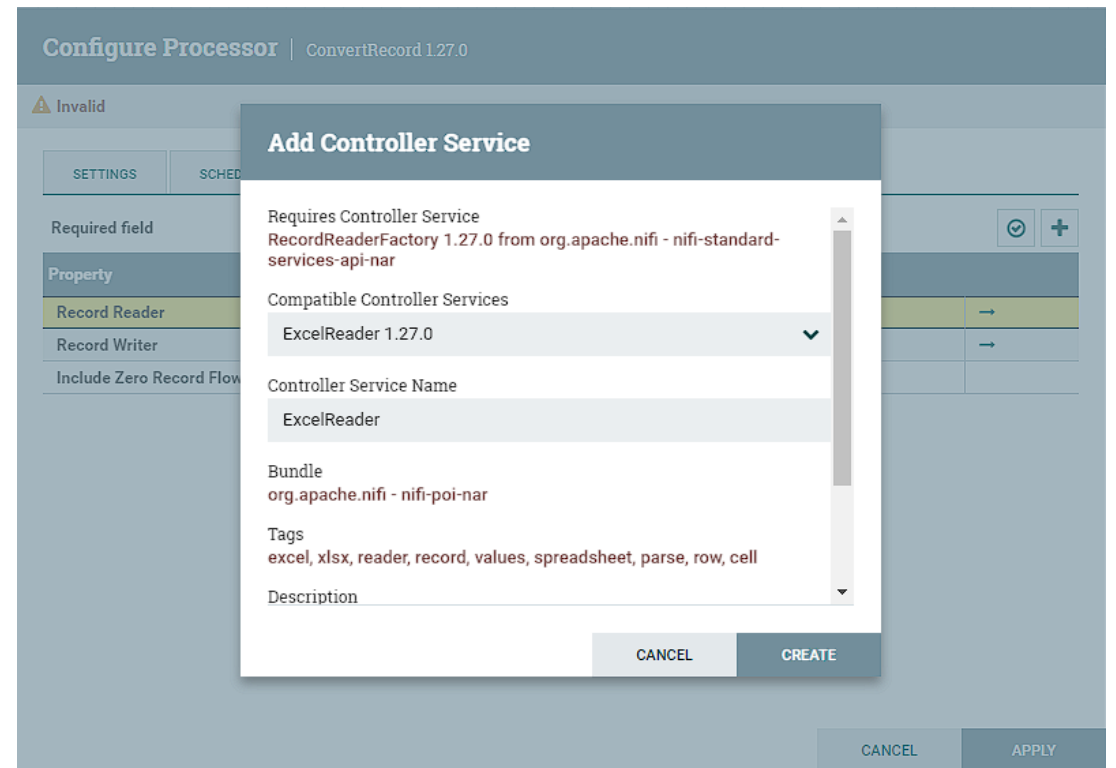
- Apache NiFi не может напрямую взаимодействовать с файлами Excel
- .xlsx (XSSF 2007 OOXML)
- .xls (HSSF '97(-2007))
- Проблемы при обработке сложных файлов
- Удален из NiFi 2.0.0

	 <b>ConvertExcelToCSVProcessor</b> ConvertExcelToCSVProcessor 1.14.0 org.apache.nifi - nifi-poi-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



# Проблема обработки файлов .xlsx в NiFi

- ExcelReader добавлен в версии 1.23.0

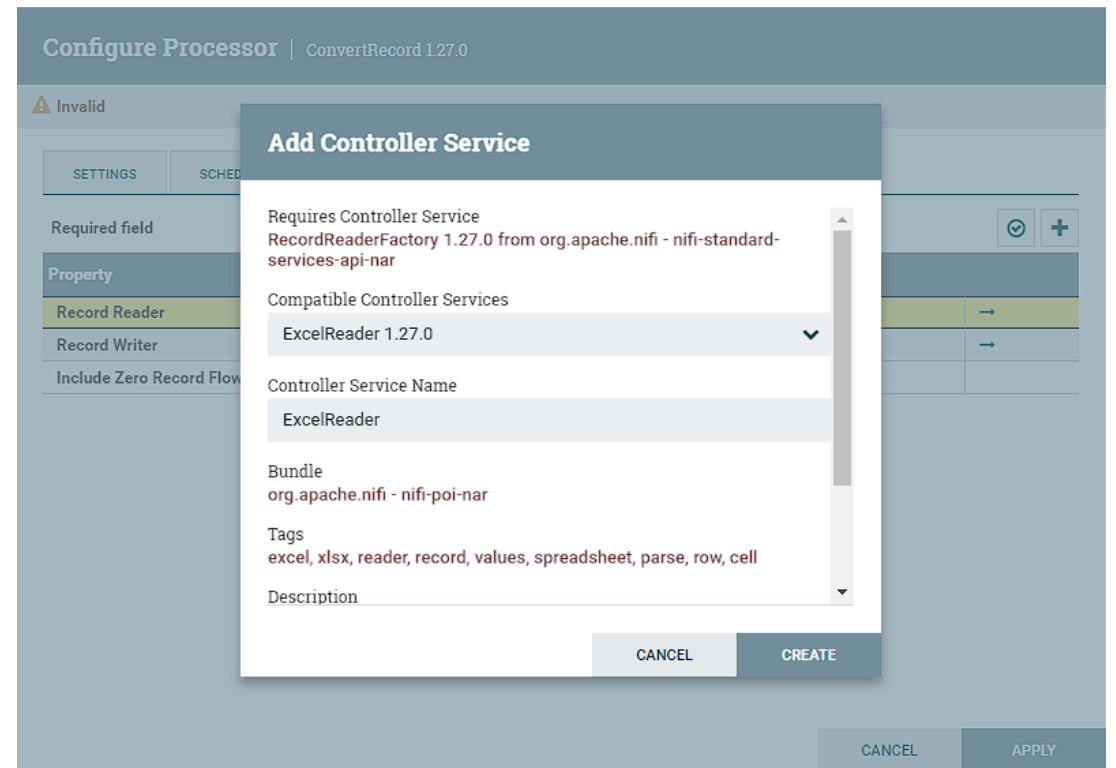






# Проблема обработки файлов .xlsx в NiFi

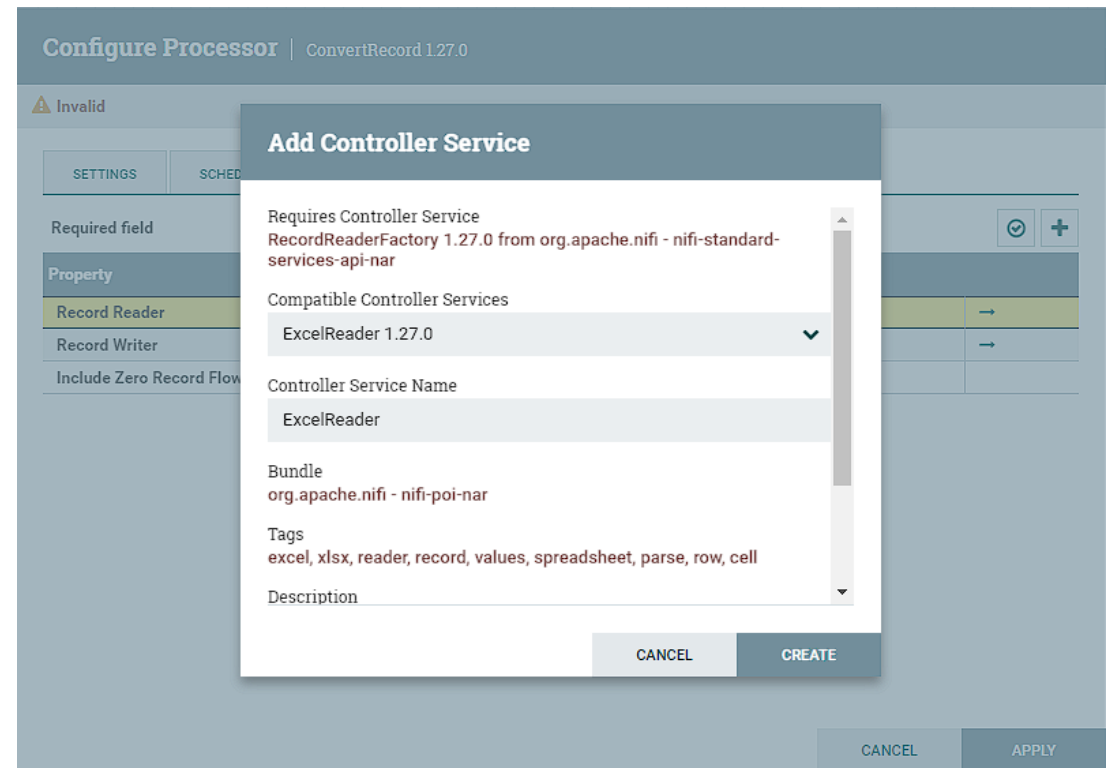
- ExcelReader добавлен в версии 1.23.0
- Те же проблемы и условия, что и с ConvertExcelToCSVProcessor





# Проблема обработки файлов .xlsx в NiFi

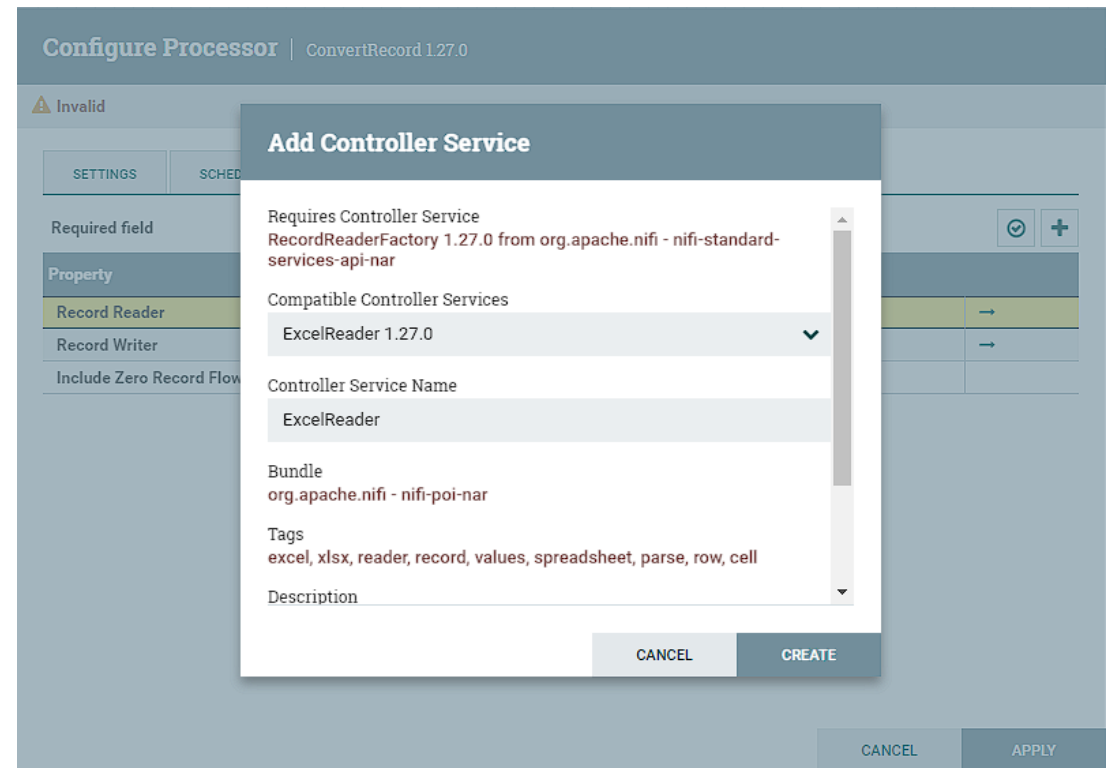
- ExcelReader добавлен в версии 1.23.0
- Те же проблемы и условия, что и с ConvertExcelToCSVProcessor
- Не может пропускать ненужные столбцы





# Проблема обработки файлов .xlsx в NiFi

- ExcelReader добавлен в версии 1.23.0
- Те же проблемы и условия, что и с ConvertExcelToCSVProcessor
- Не может пропускать ненужные столбцы
- Не работает с формулами





# Проблема обработки файлов .xlsx в NiFi

```
18:38:13 MSK ERROR org.apache.nifi - nifi-poi-nar  
ConvertRecord[id=ad27208a-0191-1000-4be4-9f19af50779e] Failed to process  
FlowFile[filename=test.xlsx]; will route to failure: java.lang.IllegalStateException:  
This cell has a shared formula and it seems setReadSharedFormulas has been set  
to false or the formula can't be evaluated 5 min
```



Apache NiFi / NIFI-13418

## Add support to ExcelReader to handle Excel spreadsheets with shared formulas

### Details

Type:	Bug	Status:	<b>RESOLVED</b>
Priority:	Major	Resolution:	Fixed
Affects Version/s:	None	Fix Version/s:	1.28.0, 2.0.0-M5
Component/s:	None		
Labels:	None		

### Description

Currently if an Excel spreadsheet has a shared formula, the ExcelReader throws the following exception

```
java.lang.IllegalStateException: This cell has a shared formula and it seems setReadSharedFormulas has been set to false or the formula can't be evaluated
```

NOTE: This issue was mentioned in [NIFI-11167](#) but I could not find data which exhibited this issue. I have now found data which exhibits the issue (see attached).

### Attachments

SaleData.xlsx	16 kB	08/Jul/24 18:47	...
---------------	-------	-----------------	-----

### Issue Links

fixes

[NIFI-12510](#) Excel Record Reader | Extend Functionality **RESOLVED**

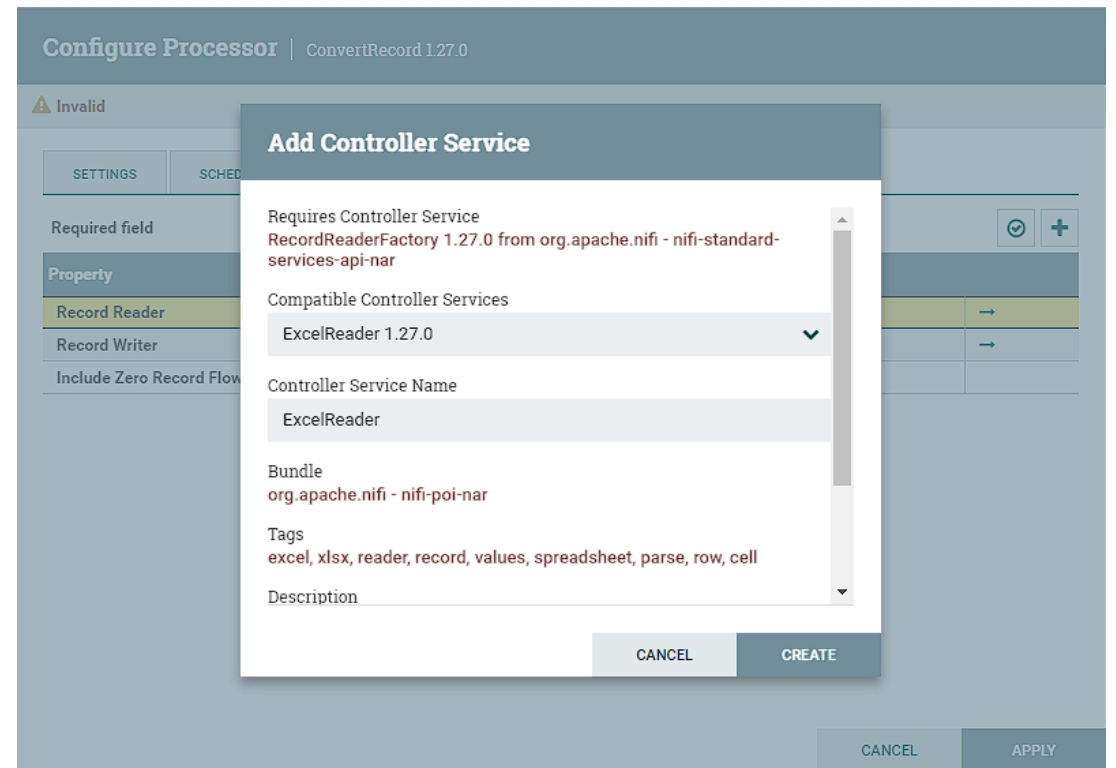
links to

[Web Link GitHub Pull Request #9063](#)



# Проблема обработки файлов .xlsx в NiFi

- ExcelReader добавлен в версии 1.23.0
- Те же проблемы и условия, что и с ConvertExcelToCSVProcessor
- Не может пропускать ненужные столбцы
- Не работает с формулами
- Сликает все листы в один файл





# Проблема обработки файлов .xlsx в NiFi

- **SplitExcel** добавлен в версиях **1.27.0** и **2.0.0-M4**



**SplitExcel**  
SplitExcel 1.27.0  
org.apache.nifi - nifi-poi-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



# Проблема обработки файлов .xlsx в NiFi

- **SplitExcel** добавлен в версиях **1.27.0** и **2.0.0-M4**
- **Нужен для разбиения файла по листам**

	<b>SplitExcel</b> SplitExcel 1.27.0 org.apache.nifi - nifi-poi-nar	
In	<b>0</b> (0 bytes)	5 min
Read/Write	<b>0 bytes / 0 bytes</b>	5 min
Out	<b>0</b> (0 bytes)	5 min
Tasks/Time	<b>0 / 00:00:00.000</b>	5 min



# ConvertExcelToCSVProcessor

## Исходный файл

period	01-20.01.2024					21-31.01.2024					
name	x1	x2	x3	x4	sum	x1	x2	x3	x4	sum	
y1			211	311	411	933		212	312	412	936
y2			221	321	421	963		222	322	422	966
sum	0		432	632	832	1896	0	434	634	834	1902

### Configure Processor

Stopped

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field +

Property	Value
Sheets to Extract	No value set
<b>Number of Rows to Skip</b>	<b>13</b>
Columns To Skip	No value set
Format Cell Values	false
CSV Format	Custom Format
Value Separator	,
Include Header Line	true
Quote Character	"
Escape Character	\
Comment Marker	No value set
Null String	No value set
Trim Fields	true

CANCEL APPLY





# ConvertExcelToCSVProcessor

## Исходный файл

period	01-20.01.2024					21-31.01.2024					
name	x1	x2	x3	x4	sum	x1	x2	x3	x4	sum	
y1			211	311	411	933		212	312	412	936
y2			221	321	421	963		222	322	422	966
sum	0		432	632	832	1896	0	434	634	834	1902

## Конвертированный файл

View as: original

```
1 period,01-20.01.2024,,,,,21-31.01.2024
2 name,x1,x2,x3,x4,sum,x1
3 y1,,211,311,411,933,
4 y2,,221,321,421,963,
5 sum,0,432,632,832,1896,0
6
```



# ConvertExcelToCSVProcessor

## Исходный файл

period	01-20.01.2024					21-31.01.2024				
name	x1	x2	x3	x4	sum	x1	x2	x3	x4	sum
y1			211	311	411	933	212	312	412	936
y2			221	321	421	963	222	322	422	966
sum	0	432	632	832	1896	434	634	834	1902	

## Конвертированный файл

View as: original

```
1 period,01-20.01.2024,,,,,21-31.01.2024
2 name,x1,x2,x3,x4,sum,x1
3 y1,,211,311,411,933,
4 y2,,221,321,421,963,
5 sum,0,432,632,832,1896,0
6
```



# Множественная обработка файла

## Файл с заголовком

View as: original

1	period,01-20.01.2024,,,,,21-31.01.2024
2	name,x1,x2,x3,x4,sum,x1
3	y1,,211,311,411,933,
4	y2,,221,321,421,963,
5	sum,0,432,632,832,1896,0
6	


## Файл с данными таблицы

View as: original

1	name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
2	y1,,211,311,411,933,,212,312,412,936
3	y2,,221,321,421,963,,222,322,422,966
4	sum,0,432,632,832,1896,0,434,634,834,1902
5	



# Множественная обработка файла



**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

**Invalid**

SETTINGS | SCHEDULING | **PROPERTIES** | COMMENTS


Required field +

Property	Value
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
header	(\A.*)

CANCEL APPLY




# Множественная обработка файла



**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	(?s)(^.*\$)
Replacement Value	\${header}
Character Set	UTF-8
Maximum Buffer Size	1 MB
Replacement Strategy	Regex Replace
Evaluation Mode	Entire text
Line-by-Line Evaluation Mode	All

CANCEL APPLY



# Множественная обработка файла

## Результат обработки

View as: original



```
1 name, x1, x2, x3, x4, sum, x1, x2, x3, x4, sum
2 y1, , 211, 311, 411, 933, , 212, 312, 412, 936
3 y2, , 221, 321, 421, 963, , 222, 322, 422, 966
4 sum, 0, 432, 632, 832, 1896, 0, 434, 634, 834, 1902
5 period, 01-20.01.2024, , , , 21-31.01.2024
```



# Множественная обработка файла

## Результат обработки

View as: original



```
1 name, x1, x2, x3, x4, sum, x1, x2, x3, x4, sum
2 y1, , 211, 311, 411, 933, , 212, 312, 412, 936
3 y2, , 221, 321, 421, 963, , 222, 322, 422, 966
4 sum, 0, 432, 632, 832, 1896, 0, 434, 634, 834, 1902
5 period, 01-20.01.2024, , , , 21-31.01.2024
```

## Результат обработки 2

View as: original



```
1 period, 01-20.01.2024, , , , 21-31.01.2024 name, x1, x2, x3, x4, sum, x1, x2, x3, x4, sum
2 y1, , 211, 311, 411, 933, , 212, 312, 412, 936
3 y2, , 221, 321, 421, 963, , 222, 322, 422, 966
4 sum, 0, 432, 632, 832, 1896, 0, 434, 634, 834, 1902
5
```

*просчитался,  
но...  
где?*





# Множественная обработка файла

Это происходит из-за того, что в стандартной конфигурации процессора MergeContent используется Merge Strategy = Bin-Packing Algorithm, которая сливает данные в произвольном порядке. Изменим её на Defragment

### Configure Processor

Invalid

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field +

Property	
Merge Strategy	Bin-Packing Algorithm
Merge Format	Bin-Packing Algorithm
Attribute Strategy	Defragment
Correlation Attribute Name	Reference parameter...
Minimum Number of Entries	1000
Maximum Number of Entries	0 B
Minimum Group Size	No value set
Maximum Group Size	No value set
Max Bin Age	5
Maximum number of Bins	Do Not Use Delimiters
Delimiter Strategy	

CANCEL APPLY





# Множественная обработка файла

- 1) `fragment.identifier` – все файлы с одинаковым значением этого атрибута будут объединены вместе

### Configure Processor

Invalid

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field +

Property	
Merge Strategy	<input type="text" value="Bin-Packing Algorithm"/>
Merge Format	<input type="text" value="Bin-Packing Algorithm"/>
Attribute Strategy	<input type="text" value="Defragment"/>
Correlation Attribute Name	<input type="text" value="Reference parameter..."/>
Minimum Number of Entries	<input type="text" value="1000"/>
Maximum Number of Entries	<input type="text" value="0 B"/>
Minimum Group Size	<input type="text" value="No value set"/>
Maximum Group Size	<input type="text" value="No value set"/>
Max Bin Age	<input type="text" value="5"/>
Maximum number of Bins	<input type="text" value="Do Not Use Delimiters"/>
Delimiter Strategy	<input type="text" value=""/>

CANCEL APPLY



# Множественная обработка файла

- 1) `fragment.identifier` – все файлы с одинаковым значением этого атрибута будут объединены вместе
- 2) `fragment.index` – этот атрибут указывает порядок, в котором должны быть собраны фрагменты

### Configure Processor

Invalid

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field +

Property		
Merge Strategy		
Merge Format		
Attribute Strategy		
Correlation Attribute Name	?	
Minimum Number of Entries	?	
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

Bin-Packing Algorithm

Bin-Packing Algorithm

Defragment

Reference parameter...

CANCEL APPLY



# Множественная обработка файла

- 1) `fragment.identifier` – все файлы с одинаковым значением этого атрибута будут объединены вместе
- 2) `fragment.index` – этот атрибут указывает порядок, в котором должны быть собраны фрагменты
- 3) `fragment.count` – значение этого атрибута указывает, сколько файлов следует ожидать для слияния

### Configure Processor

Invalid

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field +

Property		
Merge Strategy		
Merge Format		
Attribute Strategy		
Correlation Attribute Name	?	
Minimum Number of Entries	?	
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

Bin-Packing Algorithm

Bin-Packing Algorithm

Defragment

Reference parameter...

CANCEL APPLY



# Множественная обработка файла

- 1) `fragment.identifier` – все файлы с одинаковым значением этого атрибута будут объединены вместе
- 2) `fragment.index` – этот атрибут указывает порядок, в котором должны быть собраны фрагменты
- 3) `fragment.count` – значение этого атрибута указывает, сколько файлов следует ожидать для слияния
- 4) `segment.original.filename` – значение этого атрибута будет использоваться в качестве имени файла после слияния

### Configure Processor

Invalid

SETTINGS SCHEDULING PROPERTIES COMMENTS

Required field

Property		
Merge Strategy		
Merge Format		
Attribute Strategy		
Correlation Attribute Name		
Minimum Number of Entries		
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	

Bin-Packing Algorithm

Bin-Packing Algorithm

Defragment


Reference parameter...

CANCEL APPLY















# Множественная обработка файла


## Configure Processor

 Invalid

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field +

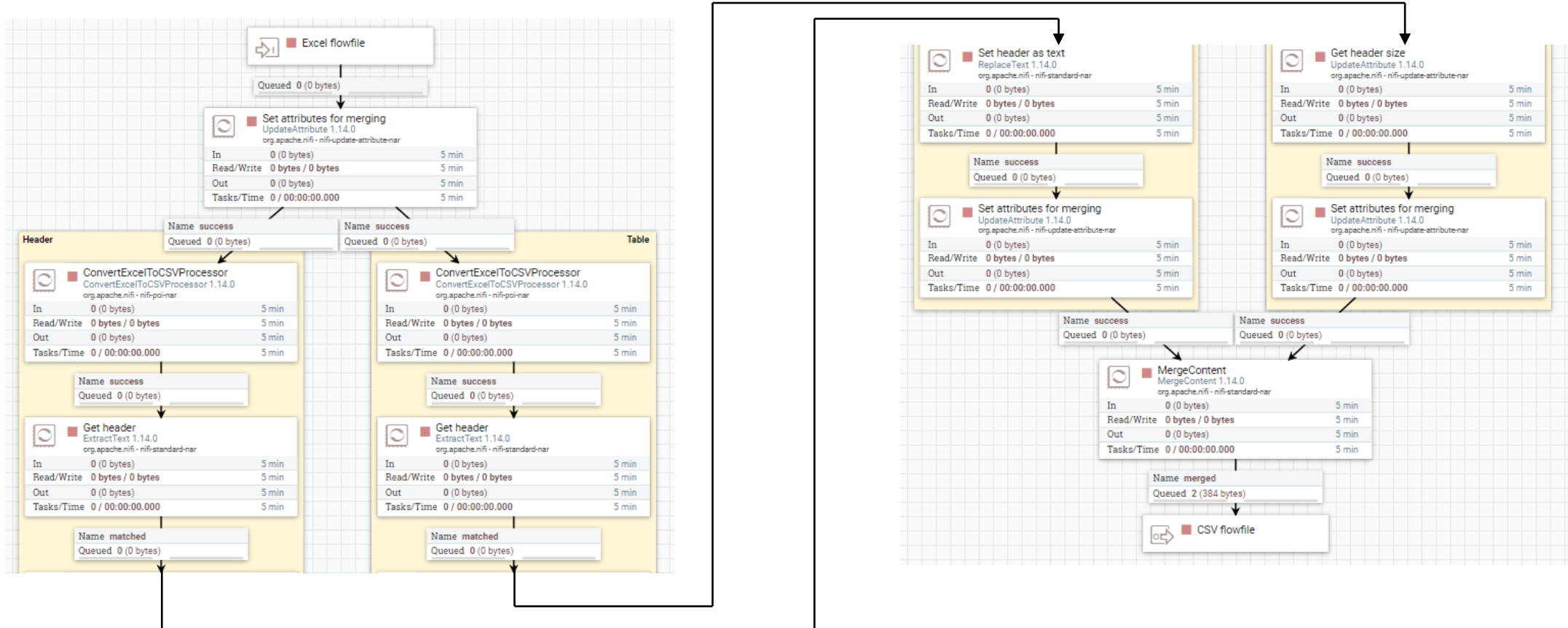
Property		Value	
Delete Attributes Expression		No value set	
<b>Store State</b>		<b>Do not store state</b>	
Stateful Variables Initial Value		No value set	
<b>Cache Value Lookup Cache Size</b>		100	
fragment.count		2	
fragment.identifier		\${uuid}	
segment.original.filename		\${sourcefilename}	
fragment.index		1(заголовок) 2(таблица)	

 **ADVANCED** CANCEL APPLY



# Множественная обработка файла

## Цепочка процессоров обработки





# Множественная обработка файла

## Результат обработки

View as: original

1	period,01-20.01.2024,,,,,21-31.01.2024
2	name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
3	y1,,211,311,411,933,,212,312,412,936
4	y2,,221,321,421,963,,222,322,422,966
5	sum,0,432,632,832,1896,0,434,634,834,1902
6	



# Множественная обработка файла

## Результат обработки

View as: original

1	period, 01-20.01.2024, , , , 21-31.01.2024
2	name, x1, x2, x3, x4, sum, x1, x2, x3, x4, sum
3	y1, , 211, 311, 411, 933, , 212, 312, 412, 936
4	y2, , 221, 321, 421, 963, , 222, 322, 422, 966
5	sum, 0, 432, 632, 832, 1896, 0, 434, 634, 834, 1902
6	



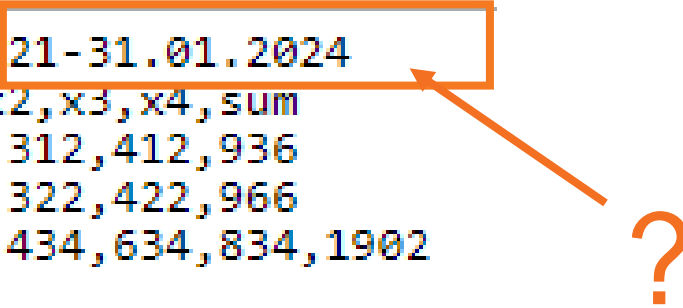


# Множественная обработка файла

## Результат обработки

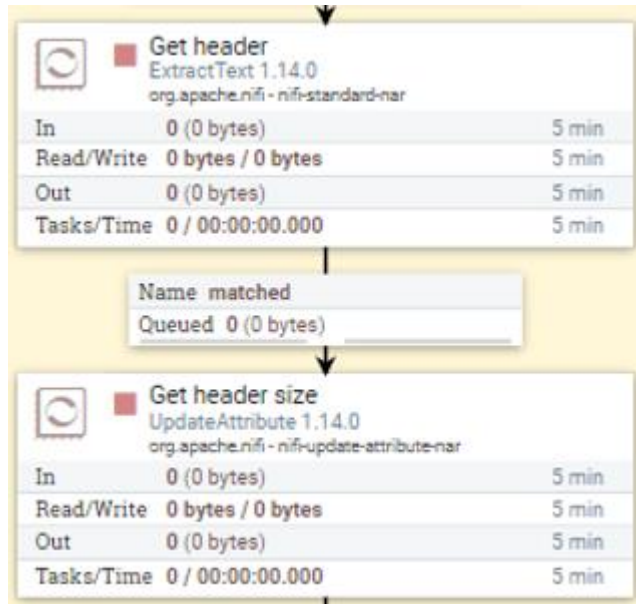
View as: original

1	period,01-20.01.2024,,,,21-31.01.2024
2	name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
3	y1,,211,311,411,933,,212,312,412,936
4	y2,,221,321,421,963,,222,322,422,966
5	sum,0,432,632,832,1896,0,434,634,834,1902
6	





# Обработка файла CSV



## Configure Processor

Invalid

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
headerSize	<code>\${allDelineatedValues("\${header}", ","):count()}</code>

ADVANCED

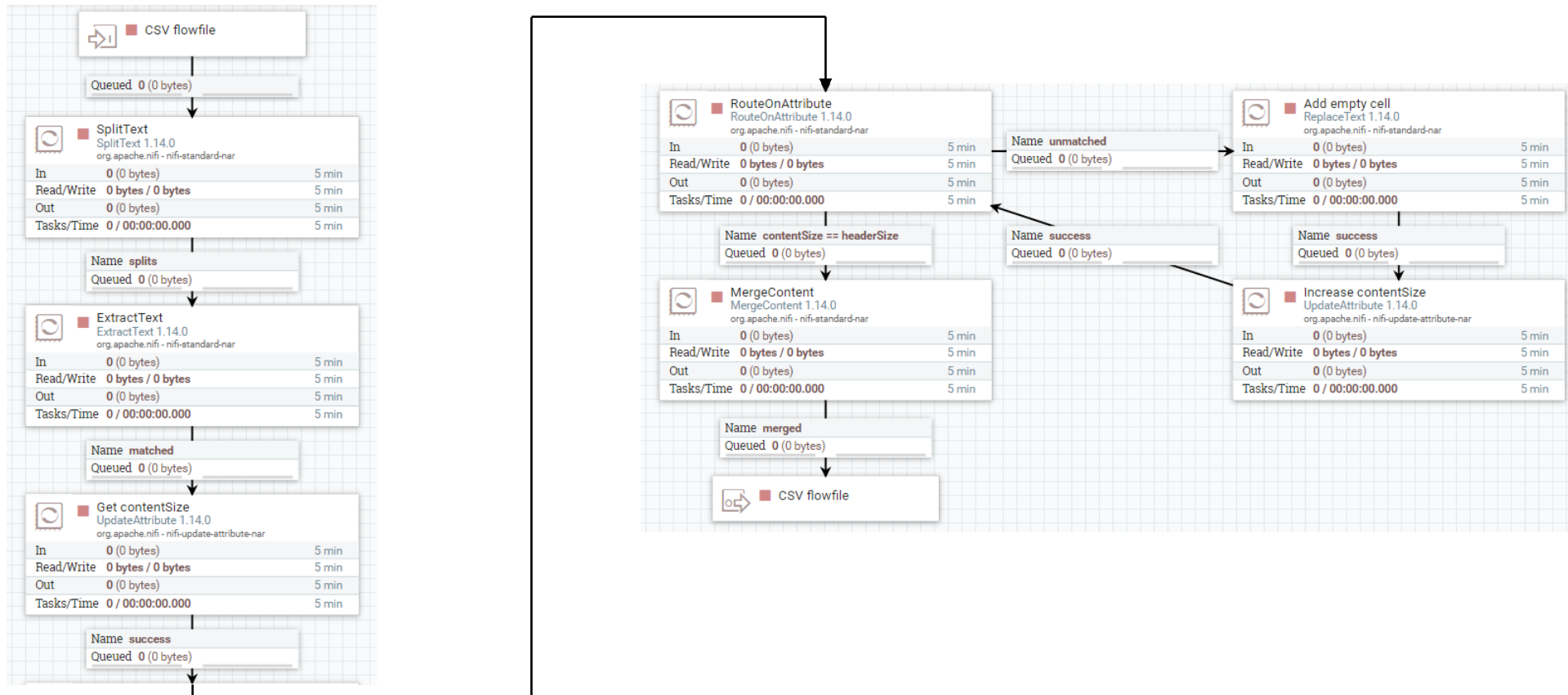
CANCEL

APPLY



# Обработка файла CSV

## Цепочка процессоров обработки





# Обработка файла CSV

## Результат обработки

View as: original



1	period,01-20.01.2024,,,,,21-31.01.2024,,,,
2	name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
3	y1,,211,311,411,933,,212,312,412,936
4	y2,,221,321,421,963,,222,322,422,966
5	sum,0,432,632,832,1896,0,434,634,834,1902



# Обработка файла CSV

## Результат обработки

View as: original



```
1 period,01-20.01.2024,,,,,21-31.01.2024,,,,
2 name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
3 y1,,211,311,411,933,,212,312,412,936
4 y2,,221,321,421,963,,222,322,422,966
5 sum,0,432,632,832,1896,0,434,634,834,1902
```

View as: original



```
1 period,01-20.01.2024,01-20.01.2024,01-20.01.2024,01-20.01.2024,01-20.01.2024,21-31.01.2024,21-31.01.2024,21-31.01.2024,21-31.01.2024
2 name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
3 y1,,211,311,411,933,,212,312,412,936
4 y2,,221,321,421,963,,222,322,422,966
5 sum,0,432,632,832,1896,0,434,634,834,1902
```



# Обработка файла CSV

## Результат обработки

View as: original

```
1 period,01-20.01.2024,,,,,21-31.01.2024,,,,
2 name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
3 y1,,211,311,411,933,,212,312,412,936
4 y2,,221,321,421,963,,222,322,422,966
5 sum,0,432,632,832,1896,0,434,634,834,1902
```

View as: original

```
1 period,01-20.01.2024,01-20.01.2024,01-20.01.2024,01-20.01.2024,01-20.01.2024,21-31.01.2024,21-31.01.2024,21-31.01.2024,21-31.01.2024
2 name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
3 y1,,211,311,411,933,,212,312,412,936
4 y2,,221,321,421,963,,222,322,422,966
5 sum,0,432,632,832,1896,0,434,634,834,1902
```

View as: formatted

```
1 [ {
2   "period" : "name",
3   "01-20.01.2024" : "x1",
4   "21-31.01.2024" : "x2"
5 } , {
6   "period" : "y1",
7   "01-20.01.2024" : null,
8   "21-31.01.2024" : 211
9 } , {
10  "period" : "y2",
11  "01-20.01.2024" : null,
12  "21-31.01.2024" : 221
13 } , {
14  "period" : "sum",
15  "01-20.01.2024" : 0,
16  "21-31.01.2024" : 432
17 } ]
```



# Обработка файла CSV

## Результат обработки

View as: original

```
1 period,01-20.01.2024,,,,,21-31.01.2024,,,,
2 name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
3 y1,,211,311,411,933,,212,312,412,936
4 y2,,221,321,421,963,,222,322,422,966
5 sum,0,432,632,832,1896,0,434,634,834,1902
```

View as: formatted

```
1 [ {
2   "period" : "name",
3   "01-20.01.2024" : "x1",
4   "21-31.01.2024" : "x2"
5 } ], {
6   "period" : "y1",
7   "01-20.01.2024" : null,
8   "21-31.01.2024" : 211
9 } ], {
10  "period" : "y2",
11  "01-20.01.2024" : null,
12  "21-31.01.2024" : 221
13 } ], {
14  "period" : "sum",
15  "01-20.01.2024" : 0,
16  "21-31.01.2024" : 432
17 } ]
```

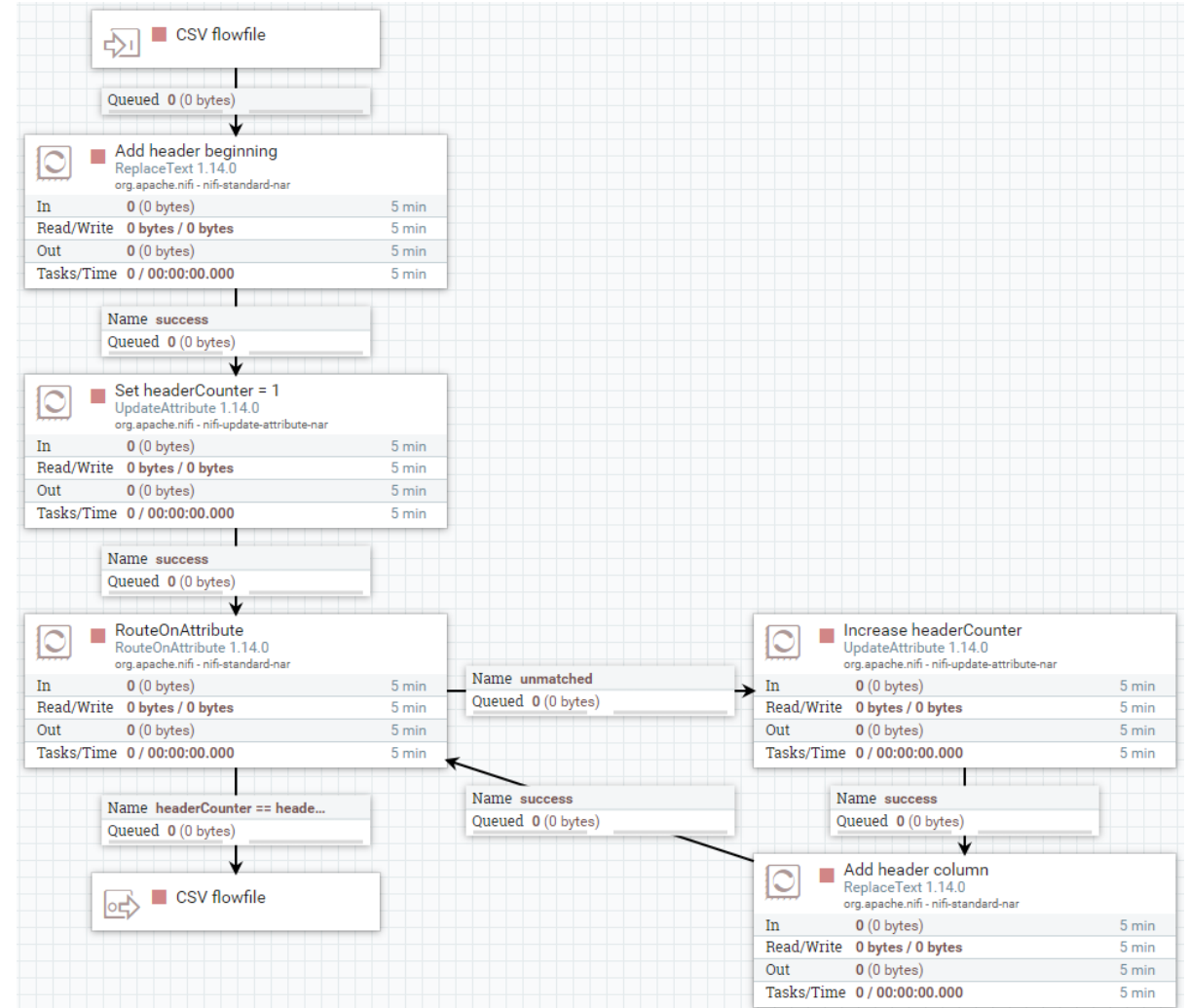
View as: original

```
1 period,01-20.01.2024,01-20.01.2024,01-20.01.2024,01-20.01.2024,01-20.01.2024,21-31.01.2024,21-31.01.2024,21-31.01.2024,21-31.01.2024,21-31.01.2024
2 name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
3 y1,,211,311,411,933,,212,312,412,936
4 y2,,221,321,421,963,,222,322,422,966
5 sum,0,432,632,832,1896,0,434,634,834,1902
```



# Обработка файла CSV

Сделаем свой заголовок





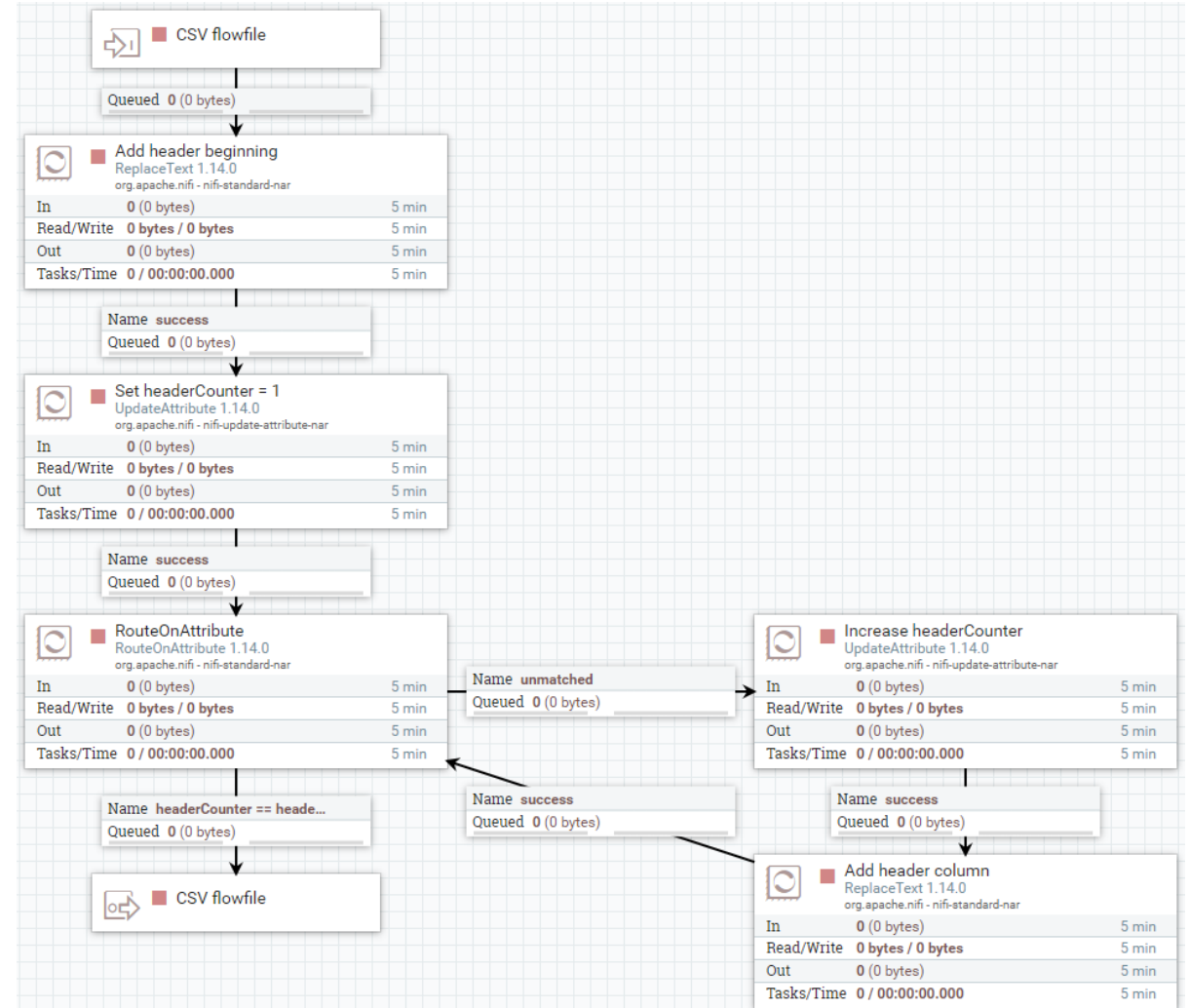


# Обработка файла CSV

## Результат обработки

View as: original

```
1 h1,h2,h3,h4,h5,h6,h7,h8,h9,h10,h11
2 period,01-20.01.2024,01-20.01.2024,01-20.01.2024,0
3 name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
4 y1,,211,311,411,933,,212,312,412,936
5 y2,,221,321,421,963,,222,322,422,966
6 sum,0,432,632,832,1896,0,434,634,834,1902
```





# Обработка файла CSV

View as: original

```
1 h1,h2,h3,h4,h5,h6,h7,h8,h9,h10,h11
2 period,01-20.01.2024,01-20.01.2024,01-20.01.2024,(
3 name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
4 y1,,211,311,411,933,,212,312,412,936
5 y2,,221,321,421,963,,222,322,422,966
6 sum,0,432,632,832,1896,0,434,634,834,1902
```

View as: formatted

```
1 [ {
2   "h1" : "period",
3   "h2" : "01-20.01.2024",
4   "h3" : "01-20.01.2024",
5   "h4" : "01-20.01.2024",
6   "h5" : "01-20.01.2024",
7   "h6" : "01-20.01.2024",
8   "h7" : "21-31.01.2024",
9   "h8" : "21-31.01.2024",
10  "h9" : "21-31.01.2024",
11  "h10" : "21-31.01.2024",
12  "h11" : "21-31.01.2024"
13 }, {
14   "h1" : "name",
15   "h2" : "x1",
16   "h3" : "x2",
17   "h4" : "x3",
18   "h5" : "x4",
19   "h6" : "sum",
20   "h7" : "x1",
21   "h8" : "x2",
22   "h9" : "x3",
23   "h10" : "x4",
24   "h11" : "sum"
25 }, {
26   "h1" : "y1",
27   "h2" : null,
28   "h3" : 211,
29   "h4" : 311,
30   "h5" : 411,
31   "h6" : 933,
32   "h7" : null,
33   "h8" : 212,
34   "h9" : 312,
35   "h10" : 412,
```



# Обработка файла CSV

View as: original

```
1 h1,h2,h3,h4,h5,h6,h7,h8,h9,h10,h11
2 period,01-20.01.2024,01-20.01.2024,01-20.01.2024,(
3 name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
4 y1,211,311,411,933,,212,312,412,936
5 y2,,221,321,421,966,,222,322,422,966
6 sum,0,432,632,832,1896,0,434,634,834,1902
```

View as: formatted

```
1 [ {
2   "h1" : "period",
3   "h2" : "01-20.01.2024",
4   "h3" : "01-20.01.2024",
5   "h4" : "01-20.01.2024",
6   "h5" : "01-20.01.2024",
7   "h6" : "01-20.01.2024",
8   "h7" : "21-31.01.2024",
9   "h8" : "21-31.01.2024",
10  "h9" : "21-31.01.2024",
11  "h10" : "21-31.01.2024",
12  "h11" : "21-31.01.2024"
13 }, {
14   "h1" : "name",
15   "h2" : "x1",
16   "h3" : "x2",
17   "h4" : "x3",
18   "h5" : "x4",
19   "h6" : "sum",
20   "h7" : "x1",
21   "h8" : "x2",
22   "h9" : "x3",
23   "h10" : "x4",
24   "h11" : "sum"
25 }, {
26   "h1" : "y1",
27   "h2" : null,
28   "h3" : 211,
29   "h4" : 311,
30   "h5" : 411,
31   "h6" : 933,
32   "h7" : null,
33   "h8" : 212,
34   "h9" : 312,
35   "h10" : 412,
```



# Обработка файла CSV

View as: original

```
1 h1,h2,h3,h4,h5,h6,h7,h8,h9,h10,h11
2 period,01-20.01.2024,01-20.01.2024,01-20.01.2024,(
3 name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
4 y1,,211,311,411,933,,212,312,412,936
5 y2,,221,321,421,963,,222,322,422,966
6 sum,0,432,632,832,1896,0,434,634,834,1902
```



View as: formatted

```
1 [ {
2   "h1" : "period",
3   "h2" : "01-20.01.2024",
4   "h3" : "01-20.01.2024",
5   "h4" : "01-20.01.2024",
6   "h5" : "01-20.01.2024",
7   "h6" : "01-20.01.2024",
8   "h7" : "21-31.01.2024",
9   "h8" : "21-31.01.2024",
10  "h9" : "21-31.01.2024",
11  "h10" : "21-31.01.2024",
12  "h11" : "21-31.01.2024"
13 }, {
14   "h1" : "name",
15   "h2" : "x1",
16   "h3" : "x2",
17   "h4" : "x3",
18   "h5" : "x4",
19   "h6" : "sum",
20   "h7" : "x1",
21   "h8" : "x2",
22   "h9" : "x3",
23   "h10" : "x4",
24   "h11" : "sum"
25 }, {
26   "h1" : "y1",
27   "h2" : null,
28   "h3" : 211,
29   "h4" : 311,
30   "h5" : 411,
31   "h6" : 933,
32   "h7" : null,
33   "h8" : 212,
34   "h9" : 312,
35   "h10" : 412,
```

```
1 {
2   "period" : "21-31.01.2024",
3   "name" : "sum",
4   "y1" : 936,
5   "y2" : 966
6 }
```





# Обработка файла JSON

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



# Обработка файла JSON

  <b>JoltTransformJSON</b> JoltTransformJSON 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

```
[
  {
    "operation": "shift",
    "spec": {
      "*": {
        "*": "[#1]"
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "0": null,
      "*": {
        "*": "[&1].@(2,[0].[&])"
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "0": null,
      "*": "[]"
    }
  },
  {
    "operation": "remove",
    "spec": {
      "*": {
        "sum": ""
      }
    }
  }
]
```



# Обработка файла JSON

```
1 [
2   ["period", "name", "y1", "y2", "sum"],
3   ["01-20.01.2024", "x1", null, null, 0],
4   ["01-20.01.2024", "x2", 211, 221, 432],
5   ["01-20.01.2024", "x3", 311, 321, 632],
6   ["01-20.01.2024", "x4", 411, 421, 832],
7   ["01-20.01.2024", "sum", 933, 963, 1896],
8   ["21-31.01.2024", "x1", null, null, 0],
9   ["21-31.01.2024", "x2", 212, 222, 434],
10  ["21-31.01.2024", "x3", 312, 322, 634],
11  ["21-31.01.2024", "x4", 412, 422, 834],
12  ["21-31.01.2024", "sum", 936, 966, 1902]
13 ]

1 [null, {
2   "period": "01-20.01.2024",
3   "name": "x1",
4   "y1": null,
5   "y2": null,
6   "sum": 0
7 }, {
8   "period": "01-20.01.2024",

1  [{
2   "period": "01-20.01.2024",
3   "name": "x1",
4   "y1": null,
5   "y2": null,
6   "sum": 0
7 }, {
8   "period": "01-20.01.2024",

1  [{
2   "period": "01-20.01.2024",
3   "name": "x1",
4   "y1": null,
5   "y2": null
6 }, {
7   "period": "01-20.01.2024",

[
  {
    "operation": "shift",
    "spec": {
      "*": {
        "*": "[#1]"
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "0": null,
      "*": {
        "*": "[&1].@(2,[0].[&])"
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "0": null,
      "*": "[]"
    }
  },
  {
    "operation": "remove",
    "spec": {
      "*": {
        "sum": ""
      }
    }
  }
]
```



# Обработка файла JSON

```
View as: formatted
1 [ {
2   "period" : "01-20.01.2024",
3   "name" : "x1",
4   "y1" : null,
5   "y2" : null
6 }, {
7   "period" : "01-20.01.2024",
8   "name" : "x2",
9   "y1" : 211,
10  "y2" : 221
11 }, {
12  "period" : "01-20.01.2024",
13  "name" : "x3",
14  "y1" : 311,
15  "y2" : 321
16 }, {
17  "period" : "01-20.01.2024",
18  "name" : "x4",
19  "y1" : 411,
20  "y2" : 421
21 }, {
22  "period" : "01-20.01.2024",
23  "name" : "sum",
24  "y1" : 933,
25  "y2" : 963
26 }, {
27  "period" : "21-31.01.2024",
28  "name" : "x1",
29  "y1" : null,
30  "y2" : null
31 }, {
32  "period" : "21-31.01.2024",
```

```
[
  {
    "operation": "shift",
    "spec": {
      "*": {
        "*": "[#1]"
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "0": null,
      "*": {
        "*": "[&1].@(2,[0].[&])"
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "0": null,
      "*": "[]"
    }
  },
  {
    "operation": "remove",
    "spec": {
      "*": {
        "sum": ""
      }
    }
  }
]
```





# Обработка файла JSON

```
View as: formatted
1 [ {
2   "period" : "01-20.01.2024",
3   "name" : "x1",
4   "y1" : null,
5   "y2" : null
6 }, {
7   "period" : "01-20.01.2024",
8   "name" : "x2",
9   "y1" : 211,
10  "y2" : 221
11 }, {
12  "period" : "01-20.01.2024",
13  "name" : "x3",
14  "y1" : 311,
15  "y2" : 321
16 }, {
17  "period" : "01-20.01.2024",
18  "name" : "x4",
19  "y1" : 411,
20  "y2" : 421
21 }, {
22  "period" : "01-20.01.2024",
23  "name" : "sum",
24  "y1" : 933,
25  "y2" : 963
26 }, {
27  "period" : "21-31.01.2024",
28  "name" : "x1",
29  "y1" : null,
30  "y2" : null
31 }, {
32  "period" : "21-31.01.2024",
```

```
1 {
2   "resource": "y2",
3   "value": 321,
4   "period": "01-20.01.2024",
5   "name": "x3"
6 }
```



# Обработка файла JSON

```
View as: formatted
1 [ {
2   "period" : "01-20.01.2024",
3   "name" : "x1",
4   "y1" : null,
5   "y2" : null
6 }, {
7   "period" : "01-20.01.2024",
8   "name" : "x2",
9   "y1" : 211,
10  "y2" : 221
11 }, {
12  "period" : "01-20.01.2024",
13  "name" : "x3",
14  "y1" : 311,
15  "y2" : 321
16 }, {
17  "period" : "01-20.01.2024",
18  "name" : "x4",
19  "y1" : 411,
20  "y2" : 421
21 }, {
22  "period" : "01-20.01.2024",
23  "name" : "sum",
24  "y1" : 933,
25  "y2" : 963
26 }, {
27  "period" : "21-31.01.2024",
28  "name" : "x1",
29  "y1" : null,
30  "y2" : null
31 }, {
32  "period" : "21-31.01.2024",
```

```
[
  {
    "operation": "shift",
    "spec": {
      "period": "period",
      "name": "name",
      "**": "result[.&"
    }
  },
  {
    "operation": "shift",
    "spec": {
      "result": {
        "**": {
          "@": "&",
          "@(2,period)": "&.period",
          "@(2,name)": "&.name"
        }
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "**": "[]"
    }
  },
  {
    "operation": "shift",
    "spec": {
      "**": {
        "period": "[&1].period",
        "name": "[&1].name",
        "**": {
          "$": "[&2].resource",
          "@": "[&2].value"
        }
      }
    }
  }
]
```



# Обработка файла JSON

```
1 {
2   "period": "21-31.01.2024",
3   "name": "sum",
4   "result": [{
5     "y1": 936
6   }, {
7     "y2": 966
8   }]
9 }

1 {
2   "0": {
3     "y1": 936,
4     "period": "21-31.01.2024",
5     "name": "sum"
6   },
7   "1": {
8     "y2": 966,
9     "period": "21-31.01.2024",
10    "name": "sum"
11  }
12 }

1 [{
2   "y1": 936,
3   "period": "21-31.01.2024",
4   "name": "sum"
5 }, {
6   "y2": 966,
7   "period": "21-31.01.2024",
8   "name": "sum"
9 }]

1 [{
2   "resource": "y1",
3   "value": 936,
4   "period": "21-31.01.2024",
5   "name": "sum"
6 }, {
7   "resource": "y2",
8   "value": 966,
9   "period": "21-31.01.2024",
10  "name": "sum"
11 }]

[
  {
    "operation": "shift",
    "spec": {
      "period": "period",
      "name": "name",
      "**": "result[].&"
    }
  },
  {
    "operation": "shift",
    "spec": {
      "result": {
        "**": {
          "@": "&",
          "@(2,period)": "&.period",
          "@(2,name)": "&.name"
        }
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "**": "[]"
    }
  },
  {
    "operation": "shift",
    "spec": {
      "**": {
        "period": "[&1].period",
        "name": "[&1].name",
        "$": "[&2].resource",
        "@": "[&2].value"
      }
    }
  }
]
```



# Обработка файла JSON

**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	Text
Header	[
Footer	]
Demarcator	,

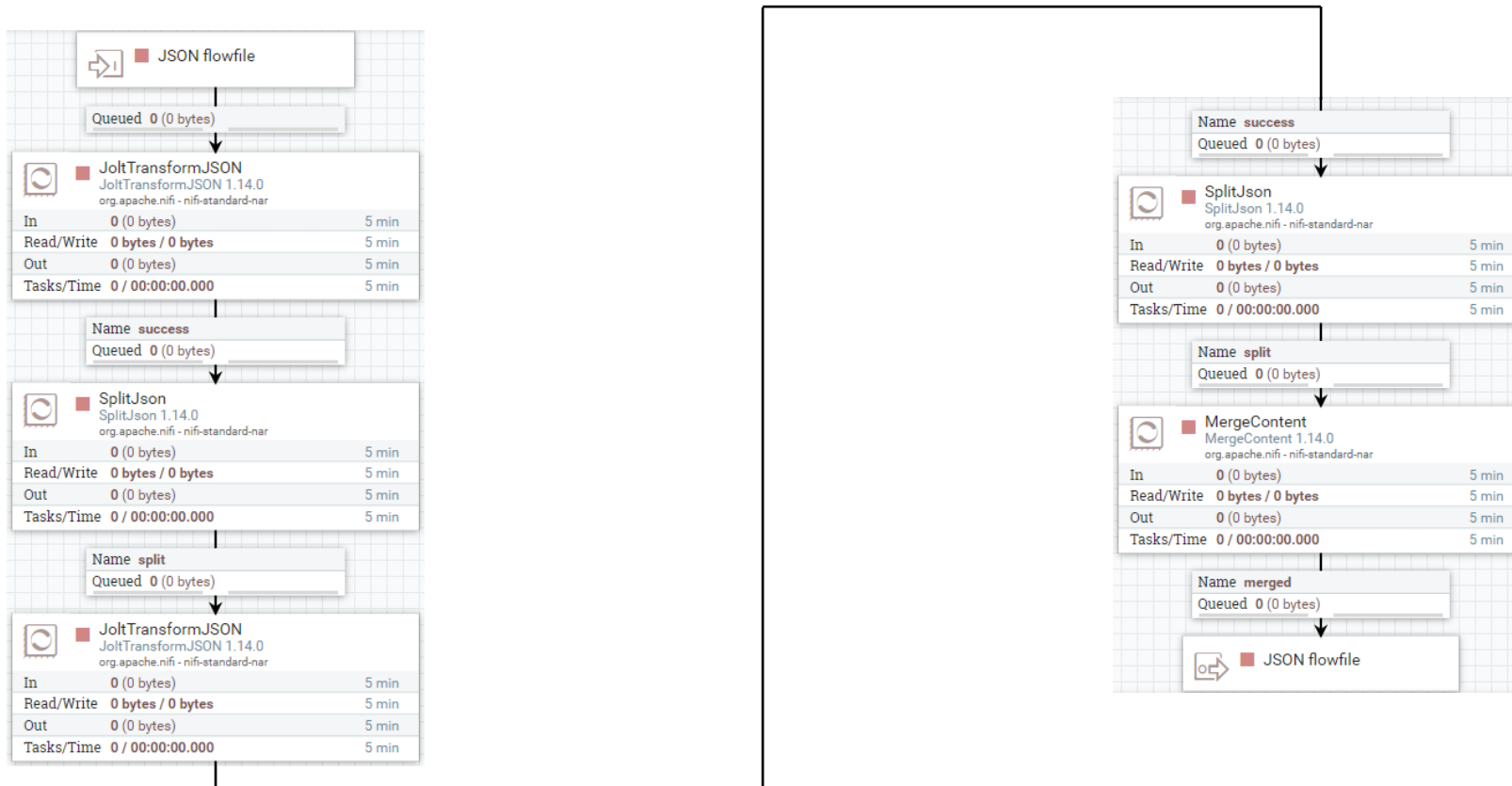
CANCEL APPLY

```
1 [ {
2   "resource" : "y1",
3   "value" : 312,
4   "period" : "21-31.01.2024",
5   "name" : "x3"
6 }, {
7   "resource" : "y2",
8   "value" : 322,
9   "period" : "21-31.01.2024",
10  "name" : "x3"
11 }, {
12  "resource" : "y1",
13  "value" : 936,
14  "period" : "21-31.01.2024",
15  "name" : "sum"
16 }, {
17  "resource" : "y2",
18  "value" : 966,
19  "period" : "21-31.01.2024",
20  "name" : "sum"
21 }, {
22  "resource" : "y1",
23  "value" : 411,
24  "period" : "01-20.01.2024",
25  "name" : "x4"
26 }, {
27  "resource" : "y2"
```



# Обработка файла JSON

## Цепочка процессоров обработки





# Проблема с обработкой больших значений после запятой

H17    X    ✓    fx    421,012345678987

	C	D	E	F	G	H	I	J	K	L	M	N	O
13													
14		period	01-20.01.2024					21-31.01.2024					
15		name	x1	x2	x3	x4	sum	x1	x2	x3	x4	sum	
16		y1		211	311	411	933		212	312	412	936	
17		y2		221	321	421,0123	963,0123		222	322	422	966	
18													
19													
20		sum	0	432	632	832,0123	1896,012	0	434	634	834	1902	
21													



# Проблема с обработкой больших значений после запятой

H17    X    ✓    fx    421,012345678987

	C	D	E	F	G	H	I	J	K	L	M	N	O
13													
14		period	01-20.01.2024					21-31.01.2024					
15		name	x1	x2	x3	x4	sum	x1	x2	x3	x4	sum	
16		y1		211	311	411	933		212	312	412	936	
17		y2		221	321	421,0123	963,0123		222	322	422	966	
18													
19													
20		sum	0	432	632	832,0123	1896,012	0	434	634	834	1902	
21													

View as: original

```
1 h1,h2,h3,h4,h5,h6,h7,h8,h9,h10,h11
2 period,01-20.01.2024,01-20.01.2024,01-20.01.2024,01-20.01.2024,01-20.01.2024,21-31.01.2024,21-31.01.2024,21-31.01.2024,21-31.01.2024,21-31.01.2024
3 name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
4 y1,,211,311,411,933,,212,312,412,936
5 y2,,221,321,421,012345678987,963.012345678987,,222,322,422,966
6 sum,0,432,632,832,012345678987,1896.012345678987,0,434,634,834,1902
```



# Проблема с обработкой больших значений после запятой

H17    421,012345678987

	C	D	E	F	G	H	I	J	K	L	M	N	O
13													
14		period	01-20.01.2024					21-31.01.2024					
15		name	x1	x2	x3	x4	sum	x1	x2	x3	x4	sum	
16		y1			211	311	411	933		212	312	412	936
17		y2			221	321	421,0123	963,0123		222	322	422	966
18													
19													
20		sum	0	432	632	832,0123	1896,012	0	434	634	834	1902	
21													

```
35 "h10" : 412,
36 "h11" : 936
37 }, {
38 "h1" : "y2",
39 "h2" : null,
40 "h3" : 221,
41 "h4" : 321,
42 "h5" : 421.01236,
43 "h6" : 963.0123,
44 "h7" : null,
45 "h8" : 222,
46 "h9" : 322,
47 "h10" : 422,
48 "h11" : 966
49 }, {
50 "h1" : "sum",
51 "h2" : 0,
52 "h3" : 432,
53 "h4" : 632,
54 "h5" : 832.0123,
55 "h6" : 1896.0123,
56 "h7" : 0,
57 "h8" : 434,
58 "h9" : 634,
59 "h10" : 834,
60 "h11" : 1902
61 } ]
```

View as: original



```
1 h1,h2,h3,h4,h5,h6,h7,h8,h9,h10,h11
2 period,01-20.01.2024,01-20.01.2024,01-20.01.2024,01-20.01.2024,01-20.01.2024,21-31.01.2024,21-31.01.2024,21-31.01.2024,21-31.01.2024,21-31.01.2024
3 name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
4 y1,,211,311,411,933,,212,312,412,936
5 y2,,221,321,421.012345678987,963.012345678987,,222,322,422,966
6 sum,0,432,632,832.012345678987,1896.012345678987,0,434,634,834,1902
```





# Проблема с обработкой больших значений после запятой

Чтобы это исправить нужно в CSVReader, который выбран как Record Reader в процессоре ConvertRecord, изменить параметр Schema Access Strategy с Infer Schema на Use String Fields From Header. Тогда все значения при конвертации будут строками и значения после запятой сохранятся

	 ConvertRecord ConvertRecord 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 Controller Service

SETTINGS | PROPERTIES | COMMENTS

Required field +

Property	Value
Schema Access Strategy	Use String Fields From Header
CSV Parser	Apache Commons CSV
Date Format	No value set
Time Format	No value set
Timestamp Format	No value set
CSV Format	Custom Format
Value Separator	,
Record Separator	\n
Treat First Line as Header	false
Ignore CSV Header Column Names	false
Quote Character	"
Escape Character	\
Comment Marker	No value set

CANCEL APPLY



# Проблема с обработкой больших значений после запятой

```
35     n10  : 412 ,
36     "h11" : "936"
37   }, {
38     "h1" : "y2",
39     "h2" : null,
40     "h3" : "221",
41     "h4" : "321",
42     "h5" : "421.012345678987",
43     "h6" : "963.012345678987",
44     "h7" : null,
45     "h8" : "222",
46     "h9" : "322",
47     "h10" : "422",
48     "h11" : "966"
49   }, {
50     "h1" : "sum",
51     "h2" : "0",
52     "h3" : "432",
53     "h4" : "632",
54     "h5" : "832.012345678987",
55     "h6" : "1896.012345678987",
56     "h7" : "0",
57     "h8" : "434",
58     "h9" : "634",
59     "h10" : "834",
60     "h11" : "1902"
61   } ]
```



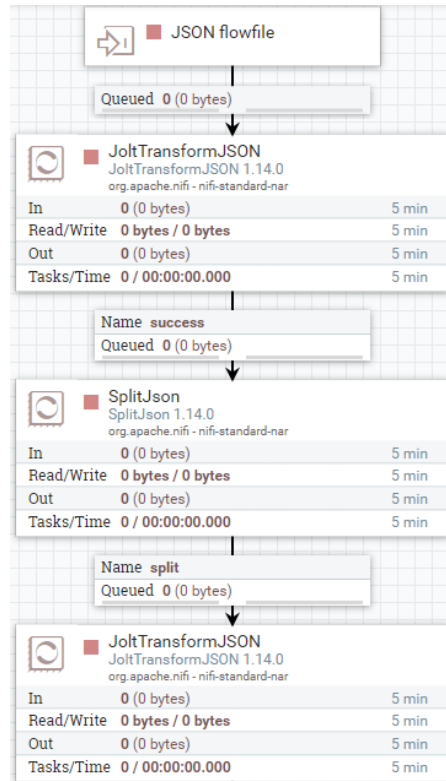
# Проблема с обработкой больших значений после запятой

```
35     n10  : 412 ,
36     "h11" : "936"
37   }, {
38     "h1"  : "y2",
39     "h2"  : null,
40     "h3"  : "221",
41     "h4"  : "321",
42     "h5"  : "421.012345678987",
43     "h6"  : "963.012345678987",
44     "h7"  : null,
45     "h8"  : "222",
46     "h9"  : "322",
47     "h10" : "422",
48     "h11" : "966"
49   }, {
50     "h1"  : "sum",
51     "h2"  : "0",
52     "h3"  : "432",
53     "h4"  : "632",
54     "h5"  : "832.012345678987",
55     "h6"  : "1896.012345678987",
56     "h7"  : "0",
57     "h8"  : "434",
58     "h9"  : "634",
59     "h10" : "834",
60     "h11" : "1902"
61   } ]
```

```
[
  {
    "operation": "shift",
    "spec": {
      "**": {
        "0": "#1"
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "0": null,
      "**": {
        "0": "[&1].@(2,[0].[&])"
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "0": null,
      "**": "[]"
    }
  },
  {
    "operation": "remove",
    "spec": {
      "**": {
        "sum": ""
      }
    }
  },
  {
    "operation": "modify-overwrite-beta",
    "spec": {
      "**": {
        "period": "=toString",
        "name": "=toString",
        "**": "=toDouble"
      }
    }
  }
]
```



# Проблема с обработкой больших значений после запятой



```
[
  {
    "operation": "shift",
    "spec": {
      "**": {
        "*": "[#1]"
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "0": null,
      "**": {
        "*": "[&1].@(2,[0].[&])"
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "0": null,
      "**": "[]"
    }
  },
  {
    "operation": "remove",
    "spec": {
      "**": {
        "sum": ""
      }
    }
  },
  {
    "operation": "modify-overwrite-beta",
    "spec": {
      "**": {
        "period": "=toString",
        "name": "=toString",
        "**": "=toDouble"
      }
    }
  }
]
```



# Проблема с обработкой больших значений после запятой

```
16 }, {
17   "period" : "01-20.01.2024",
18   "name" : "x4",
19   "y1" : 411.0,
20   "y2" : 421.012345678987
21 }, {
22   "period" : "01-20.01.2024",
23   "name" : "sum",
24   "y1" : 933.0,
25   "y2" : 963.012345678987
26 }, {
27   "period" : "21-31.01.2024",
```

```
{
  {
    "operation": "shift",
    "spec": {
      "**": {
        "sum": "[#1]"
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "0": null,
      "**": {
        "sum": "[&1].@(2,[0].[&])"
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "0": null,
      "**": "[]"
    }
  },
  {
    "operation": "remove",
    "spec": {
      "**": {
        "sum": ""
      }
    }
  },
  {
    "operation": "modify-overwrite-beta",
    "spec": {
      "**": {
        "period": "=toString",
        "name": "=toString",
        "**": "=toDouble"
      }
    }
  }
}
```



# Проблема с обработкой больших значений после запятой

```
16 }, {
17   "period" : "01-20.01.2024",
18   "name" : "x4",
19   "y1" : 411.0,
20   "y2" : 421.012345678987
21 }, {
22   "period" : "01-20.01.2024",
23   "name" : "sum",
24   "y1" : 933.0,
25   "y2" : 963.012345678987
26 }, {
27   "period" : "21-31.01.2024",
```

```
63   "value" : 411.0,
64   "period" : "01-20.01.2024",
65   "name" : "x4"
66 }, {
67   "resource" : "y2",
68   "value" : 421.012345678987,
69   "period" : "01-20.01.2024",
70   "name" : "x4"
71 }, {
72   "resource" : "y1",
73   "value" : 933.0,
74   "period" : "01-20.01.2024",
75   "name" : "sum"
76 }, {
77   "resource" : "y2",
78   "value" : 963.012345678987,
79   "period" : "01-20.01.2024",
80   "name" : "sum"
81 }, {
82   "resource" : "y1",
83   "value" : null
```

```
{
  {
    "operation": "shift",
    "spec": {
      "**": {
        "": "[#1]"
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "0": null,
      "**": {
        "": "[&1].@(2,[0].[&])"
      }
    }
  },
  {
    "operation": "shift",
    "spec": {
      "0": null,
      "**": "[]"
    }
  },
  {
    "operation": "remove",
    "spec": {
      "**": {
        "sum": ""
      }
    }
  },
  {
    "operation": "modify-overwrite-beta",
    "spec": {
      "**": {
        "period": "=toString",
        "name": "=toString",
        "": "=toDouble"
      }
    }
  }
}
```



# Проблема с обработкой нескольких страниц в одном файле

H17    X    ✓    fx    421,012345678987

	C	D	E	F	G	H	I	J	K	L	M	N	O
13													
14		period	01-20.01.2024					21-31.01.2024					
15		name	x1	x2	x3	x4	sum	x1	x2	x3	x4	sum	
16		y1		211	311	411	933		212	312	412	936	
17		y2		221	321	421,0123	963,0123		222	322	422	966	
18													
19													
20		sum	0	432	632	832,0123	1896,012	0	434	634	834	1902	
21													

year	month
------	-------

	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	
19																																	
20																																	
21		date	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29		
22		x1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29		
23		x2	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	50	52	54	56	58		
24		x3	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60	63	66	69	72	75	78	81	84	87		
25		x4	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64	68	72	76	80	84	88	92	96	100	104	108	112	116		
26																																	
27		sum	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200	210	220	230	240	250	260	270	280	290		
28																																	
29																																	



# Проблема с обработкой нескольких страниц в одном файле

success

Displaying 2 of 2 (182.00 bytes)

The destination of this queue is currently running. This listi...

	Position	UUID	Filename	File Size	Queued Duration	Lineage Duration	Penalized	
	1	7e2aa01f-5a8f-4a...	test_year.csv	182.00 bytes	00:00:05.142	00:00:05.170	No	
	2	ffbbf424-b513-42...	test_month.csv	0.00 bytes	00:00:05.140	00:00:05.170	No	





# Проблема с обработкой нескольких страниц в одном файле

## success

Displaying 2 of 2 (182.00 bytes)

The destination of this queue is currently running. This listi...

	Position	UUID	Filename	File Size	Queued Duration	Lineage Duration	Penalized	
	1	7e2aa01f-5a8f-4a...	test_year.csv	182.00 bytes	00:00:05.142	00:00:05.170	No	
	2	ffbbf424-b513-42...	test_month.csv	0.00 bytes	00:00:05.140	00:00:05.170	No	

## success

Displaying 2 of 2 (541.00 bytes)

The destination of this queue is currently running. This listi...

	Position	UUID	Filename	File Size	Queued Duration	Lineage Duration	Penalized	
	1	a849dee9-76a0-4...	test_year.csv	0.00 bytes	00:00:28.595	00:00:28.623	No	
	2	657200c9-0ff7-45...	test_month.csv	541.00 bytes	00:00:28.593	00:00:28.623	No	



# Проблема с обработкой нескольких страниц в одном файле

**Configure Processor**

Stopped

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field +

Property	Value
Sheets to Extract	year
Number of Rows to Skip	13
Columns To Skip	No value set
Format Cell Values	false
CSV Format	Custom Format
Value Separator	,
Include Header Line	true
Quote Character	"
Escape Character	\
Comment Marker	No value set
Null String	No value set
Trim Fields	true

CANCEL APPLY

**Configure Processor**

Stopped

SETTINGS SCHEDULING **PROPERTIES** COMMENTS

Required field +

Property	Value
Sheets to Extract	month
Number of Rows to Skip	20
Columns To Skip	No value set
Format Cell Values	false
CSV Format	Custom Format
Value Separator	,
Include Header Line	true
Quote Character	"
Escape Character	\
Comment Marker	No value set
Null String	No value set
Trim Fields	true

CANCEL APPLY



# Проблема с обработкой нескольких страниц в одном файле

## Connection

Displaying 2 of 2 (723.00 bytes)

The source of this queue is currently running. This listing m...

	Position	UUID	Filename	File Size	Queued Duration	Lineage Duration	Penalized	
	1	49dff5c9-7a21-4e...	test_month.csv	541.00 bytes	00:00:03.538	00:00:03.584	No	
	2	c40b4f4c-ca3e-4e...	test_year.csv	182.00 bytes	00:00:03.538	00:00:03.584	No	

View as: original



Filename: test\_month.csv

Content Type: text/csv

```
1 date,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29
2 x1,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29
3 x2,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40,42,44,46,48,50,52,54,56,58
4 x3,3,6,9,12,15,18,21,24,27,30,33,36,39,42,45,48,51,54,57,60,63,66,69,72,75,78,81,84,87
5 x4,4,8,12,16,20,24,28,32,36,40,44,48,52,56,60,64,68,72,76,80,84,88,92,96,100,104,108,112,116
6 sum,10,20,30,40,50,60,70,80,90,100,110,120,130,140,150,160,170,180,190,200,210,220,230,240,250,260,270,280,290
7
```

View as: original



```
1 period,01-20.01.2024,,,,,21-31.01.2024
2 name,x1,x2,x3,x4,sum,x1
3 y1,,211,311,411,933,
4 y2,,221,321,421.012345678987,963.012345678987,
5 sum,0,432,632,832.012345678987,1896.012345678987,0
6
```



## Исходный файл

H17 421,012345678987

period	01-20.01.2024						21-31.01.2024					
name	x1	x2	x3	x4	sum	x1	x2	x3	x4	sum		
y1		211	311	411	933		212	312	412	936		
y2		221	321	421,0123	963,0123		222	322	422	966		
sum	0	432	632	832,0123	1896,012	0	434	634	834	1902		

date	1	2	3	4	5	6	7	8	9	10	11
x1	1	2	3	4	5	6	7	8	9	10	11
x2	2	4	6	8	10	12	14	16	18	20	22
x3	3	6	9	12	15	18	21	24	27	30	33
x4	4	8	12	16	20	24	28	32	36	40	44

Configure Controller Service | ExcelReader 1.27.0

DISABLED

SETTINGS | PROPERTIES | COMMENTS

Required field

Property	Value
Schema Access Strategy	Infer Schema
Starting Row	14
Required Sheets	No value set
Protection Type	Unprotected
Date Format	No value set
Time Format	No value set
Timestamp Format	No value set

CANCEL APPLY



# Проблема обработки файлов .xlsx в NiFi

- **18:38:13 MSK ERROR** SplitExcel 1.27.0  
org.apache.nifi - nifi-poi-nar  
ConvertRecord[id=ad27208a-0191-1000-4be4-9f19af50779e] Failed to process FlowFile[filename=test.xlsx]; will route to failure: java.lang.IllegalStateException: This cell has a shared formula and it seems setReadSharedFormulas has been set to false or the formula can't be evaluated 5 min



## Исходный файл

period	01-20.01.2024					21-31.01.2024				
name	x1	x2	x3	x4	sum	x1	x2	x3	x4	sum
y1		211	311	411	933		212	312	412	936
y2		221	321	421,0123	963,0123		222	322	422	966

date	1	2	3	4	5	6	7	8	9	10	11
x1	1	2	3	4	5	6	7	8	9	10	11
x2	2	4	6	8	10	12	14	16	18	20	22
x3	3	6	9	12	15	18	21	24	27	30	33
x4	4	8	12	16	20	24	28	32	36	40	44

### Configure Controller Service | ExcelReader 1.27.0

DISABLED

SETTINGS PROPERTIES COMMENTS

Required field

Property	Value
Schema Access Strategy	Infer Schema
Starting Row	14
Required Sheets	No value set
Protection Type	Unprotected
Date Format	No value set
Time Format	No value set
Timestamp Format	No value set

CANCEL

APPLY



## Результат обработки

View as: original



```
1 column_0,column_1,column_2,column_3,column_4,column_5,column_6,column_7,column_8,column_9,column_10,column_11,colu
2 ,,,period,01-20.01.2024,,,,,21-31.01.2024,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
3 ,,,name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
4 ,,,y1,,211,311,411.0,933,,212,312,412,936,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
5 ,,,y2,,221,321,421.01236,963.0123457,,222,322,422,966,,,,,,,,,,,,,,,,,,,,,
6 ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
7 ,,,,,,,,,,date,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29
8 ,,,,,,,,,,x1,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29
9 ,,,,,,,,,,x2,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40,42,44,46,48,50,52,54,56,58
10 ,,,,,,,,,,x3,3,6,9,12,15,18,21,24,27,30,33,36,39,42,45,48,51,54,57,60,63,66,69,72,75,78,81,84,87
11 ,,,,,,,,,,x4,4,8,12,16,20,24,28,32,36,40,44,48,52,56,60,64,68,72,76,80,84,88,92,96,100,104,108,112,116
12
```



## Лист year

**Controller Service Details** | ExcelReader 1.27.0

▶ **ENABLED** ⚙️ DISABLE & CONFIGURE

- SETTINGS
- PROPERTIES
- COMMENTS

Required field

Property		Value
Schema Access Strategy	🔍	Infer Schema
Starting Row	🔍	14
Required Sheets	🔍	year
Protection Type	🔍	Unprotected
Date Format	🔍	No value set
Time Format	🔍	No value set
Timestamp Format	🔍	No value set

OK

## Лист month

**Controller Service Details** | ExcelReader 1.27.0

▶ **ENABLED** ⚙️ DISABLE & CONFIGURE

- SETTINGS
- PROPERTIES
- COMMENTS

Required field

Property		Value
Schema Access Strategy	🔍	Infer Schema
Starting Row	🔍	21
Required Sheets	🔍	month
Protection Type	🔍	Unprotected
Date Format	🔍	No value set
Time Format	🔍	No value set
Timestamp Format	🔍	No value set

OK





# ExcelReader

## Лист year

View as: original

1	column_0,column_1,column_2,column_3,column_4,column_
2	,,period,01-20.01.2024,,21-31.01.2024,,
3	,,name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
4	,,y1,,211,311,411.0,,212,312,412,
5	,,y2,,221,321,421.01236,,222,322,422,

## Лист month

View as: original

1	column_0,column_1,column_2,column_3,column_4,column_5,column_6,column_7,column_8,column_9,column_10,column
2	,,date,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29
3	,,x1,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29
4	,,x2,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40,42,44,46,48,50,52,54,56,58
5	,,x3,3,6,9,12,15,18,21,24,27,30,33,36,39,42,45,48,51,54,57,60,63,66,69,72,75,78,81,84,87
6	,,x4,4,8,12,16,20,24,28,32,36,40,44,48,52,56,60,64,68,72,76,80,84,88,92,96,100,104,108,112,116
7	



# ExcelReader

## Лист year

```
View as: original
1 column_0,column_1,column_2,column_3,column_4,column_
2 ,, ,period,01-20.01.2024,,, ,21-31.01.2024,,,
3 ,, ,name,x1,x2,x3,x4,sum,x1,x2,x3,x4,sum
4 ,, ,y1,,211,311,411.0,,,312,312,412,
5 ,, ,y2,,221,321,421.01236,,,222,322,422,
```

	A	B	C	D
13				
14				period
15				name
16				y1
17				y2
18				
19				
20				sum
21				

## Лист month

```
View as: original
1 column_0,column_1,column_2,column_3,column_4,column_5,column_6,column_7,column_8,column_9,column_10,column
2 ,, ,,,,,,date,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29
3 ,, ,,,,,,1,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29
4 ,, ,,,,,,2,2,4,6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40,42,44,46,48,50,52,54,56,58
5 ,, ,,,,,,3,3,6,9,12,15,18,21,24,27,30,33,36,39,42,45,48,51,54,57,60,63,66,69,72,75,78,81,84,87
6 ,, ,,,,,,4,4,8,12,16,20,24,28,32,36,40,44,48,52,56,60,64,68,72,76,80,84,88,92,96,100,104,108,112,116
7
```

	A	B	C	D	E	F	G	H	I	J	K	L
19												
20												
21												date
22												x1
23												x2
24												x3
25												x4
26												



# Avro schema

**Configure Controller Service** | ExcelReader 1.27.0

DISABLED

SETTINGS PROPERTIES COMMENTS

Required field

Property	Value
Schema Access Strategy	Use 'Schema Name' Property
Schema Registry	AvroSchemaRegistry
Schema Name	\${schema.name}
Schema Version	No value set
Schema Branch	No value set
Starting Row	14
Required Sheets	No value set
Protection Type	Unprotected
Date Format	No value set
Time Format	No value set
Timestamp Format	No value set

CANCEL APPLY

**Configure Controller Service** | AvroSchemaRegistry 1.27.0

DISABLED

SETTINGS PROPERTIES COMMENTS

Required field

Property	Value
Validate Field Names	true
year	

```
1 {
2   "type": "record",
3   "name": "ExcelRecord",
4   "fields": [
5     {"name": "Column1", "type": "string"},
6     {"name": "Column2", "type": "string"},
7     {"name": "Column3", "type": "string"},
8     {"name": "Column4", "type": "string"},
9     {"name": "Column5", "type": "string"},
10    {"name": "Column6", "type": "string"},
11    {"name": "Column7", "type": "string"},
12    {"name": "Column8", "type": "string"},
13    {"name": "Column9", "type": "string"},
14    {"name": "Column10", "type": "string"},
15    {"name": "Column11", "type": "string"},
16    {"name": "Column12", "type": "string"},
17    {"name": "Column13", "type": "string"}
18  ]
19 }
20
```

Set empty string

CANCEL OK



## Результат обработки

View as: original

1	Column1,Column2,Column3,Column4,Column5,Column6,Column7,C
2	,,,period,01-20.01.2024,,,,,21-31.01.2024,,,
3	,,,name,x1,x2,x3,x4,sum,x1,x2,x3,x4
4	,,,y1,,211.0,311.0,411.0,,212.0,312.0,412.0
5	,,,y2,,221.0,321.0,421.012345678987,,,222.0,322.0,422.0
6	



## Результат обработки

View as: original

1	Column1,Column2,Column3,Column4,Column5,Column6,Column7,C
2	,,,period,01-20.01.2024,,,,,21-31.01.2024,,,
3	,,,name,x1,x2,x3,x4,sum,x1,x2,x3,x4
4	,,,y1,,211.0,311.0,411.0,,,212.0,312.0,412.0
5	,,,y2,,221.0,321.0,421.012345678987,,,222.0,322.0,422.0
6	

?



## Результат обработки

View as: original



1	Column1,Column2,Column3,Column4,Column5,Column6,Column7,C
2	,,,period,01-20.01.2024,,,,,21-31.01.2024,,,
3	,,,name,x1,x2,x3,x4,sum,x1,x2,x3,x4
4	,,,y1,,211.0,311.0,411.0,,,212.0,312.0,412.0
5	,,,y2,,221.0,321.0,421.012345678987,,,222.0,322.0,422.0
6	





## Результат обработки

View as: original

1	Column1,Column2,Column3,Column4,Column5,Column6,Column7,C
2	,,,period,01-20.01.2024,,,,,21-31.01.2024,,,
3	,,,name,x1,x2,x3,x4,sum,x1,x2,x3,x4
4	,,,y1,,211.0,311.0,411.0,,,212.0,312.0,412.0
5	,,,y2,,221.0,321.0,421.012345678987,,,222.0,322.0,422.0
6	

View as: original

1	Column1,Column2,Column3,Column4,Column5,Column6,Column7,Column8,Column9,Column10,Column11,Column12,Column13
2	,,,period,01-20.01.2024,01-20.01.2024,01-20.01.2024,01-20.01.2024,01-20.01.2024,21-31.01.2024,21-31.01.2024,21-31.01.2024
3	,,,name,x1,x2,x3,x4,sum,x1,x2,x3,x4
4	,,,y1,,211.0,311.0,411.0,,,212.0,312.0,412.0
5	,,,y2,,221.0,321.0,421.012345678987,,,222.0,322.0,422.0



# Обработка файла JSON

## Результат обработки

View as: **formatted** ▼

```
1 [ {
2   "Column1" : null,
3   "Column2" : null,
4   "Column3" : null,
5   "Column4" : "period",
6   "Column5" : "01-20.01.2024",
7   "Column6" : "01-20.01.2024",
8   "Column7" : "01-20.01.2024",
9   "Column8" : "01-20.01.2024",
10  "Column9" : "01-20.01.2024",
11  "Column10" : "21-31.01.2024",
12  "Column11" : "21-31.01.2024",
13  "Column12" : "21-31.01.2024",
14  "Column13" : "21-31.01.2024"
15 }, {
16  "Column1" : null,
17  "Column2" : null,
18  "Column3" : null,
19  "Column4" : "name",
20  "Column5" : "x1",
21  "Column6" : "x2",
22  "Column7" : "x3",
23  "Column8" : "x4",
24  "Column9" : "sum",
```



**JoltTransformJSON**  
JoltTransformJSON 1.27.0  
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





# Обработка файла JSON

## Результат обработки

View as: formatted

```
1 [ {
2   "Column1" : null,
3   "Column2" : null,
4   "Column3" : null,
5   "Column4" : "period",
6   "Column5" : "01-20.01.2024",
7   "Column6" : "01-20.01.2024",
8   "Column7" : "01-20.01.2024",
9   "Column8" : "01-20.01.2024",
10  "Column9" : "01-20.01.2024",
11  "Column10" : "21-31.01.2024",
12  "Column11" : "21-31.01.2024",
13  "Column12" : "21-31.01.2024",
14  "Column13" : "21-31.01.2024"
15 }, {
16  "Column1" : null,
17  "Column2" : null,
18  "Column3" : null,
19  "Column4" : "name",
20  "Column5" : "x1",
21  "Column6" : "x2",
22  "Column7" : "x3",
23  "Column8" : "x4",
24  "Column9" : "sum",
```

```
{
  "operation": "shift",
  "spec": {
    "*": {
      "*": "[#1]"
    }
  }
},
```

```
1 [null, null, null, ["period", "name", "y1", "y2"],
2 ["01-20.01.2024", "x1", null, null],
3 ["01-20.01.2024", "x2", "211.0", "221.0"],
4 ["01-20.01.2024", "x3", "311.0", "321.0"],
5 ["01-20.01.2024", "x4", "411.0", "421.012345678987"],
6 ["01-20.01.2024", "sum", null, null],
7 ["21-31.01.2024", "x1", null, null],
8 ["21-31.01.2024", "x2", "212.0", "222.0"],
9 ["21-31.01.2024", "x3", "312.0", "322.0"],
10 ["21-31.01.2024", "x4", "412.0", "422.0"]
11 ]
```



# Обработка файла JSON

## Результат обработки

View as: formatted

```
1 [ {
2   "Column1" : null,
3   "Column2" : null,
4   "Column3" : null,
5   "Column4" : "period",
6   "Column5" : "01-20.01.2024",
7   "Column6" : "01-20.01.2024",
8   "Column7" : "01-20.01.2024",
9   "Column8" : "01-20.01.2024",
10  "Column9" : "01-20.01.2024",
11  "Column10" : "21-31.01.2024",
12  "Column11" : "21-31.01.2024",
13  "Column12" : "21-31.01.2024",
14  "Column13" : "21-31.01.2024"
15 }, {
16  "Column1" : null,
17  "Column2" : null,
18  "Column3" : null,
19  "Column4" : "name",
20  "Column5" : "x1",
21  "Column6" : "x2",
22  "Column7" : "x3",
23  "Column8" : "x4",
24  "Column9" : "sum",
```

```
{
  "operation": "shift",
  "spec": {
    "*": {
      "*": "[#1]"
    }
  }
},
```

```
1 [null, null, null, ["period", "name", "y1", "y2"],
2 ["01-20.01.2024", "x1", null, null],
3 ["01-20.01.2024", "x2", "211.0", "221.0"],
4 ["01-20.01.2024", "x3", "311.0", "321.0"],
5 ["01-20.01.2024", "x4", "411.0", "421.012345678987"],
6 ["01-20.01.2024", "sum", null, null],
7 ["21-31.01.2024", "x1", null, null],
8 ["21-31.01.2024", "x2", "212.0", "222.0"],
9 ["21-31.01.2024", "x3", "312.0", "322.0"],
10 ["21-31.01.2024", "x4", "412.0", "422.0"]
11 ]
```

```
1 [
2   ["period", "01-20.01.2024", "01-20.01.2024", "01-20.01.2024", "01-20.01.2024", "01-20.01.2024", "01-20.01.2024", "01-20.01.2024", "01-20.01.2024"],
3   ["name", "x1", "x2", "x3", "x4", "sum", "x1", "x2", "x3", "x4"],
4   ["y1", null, "211.0", "311.0", "411.0", null, null, "212.0", "312.0", "412.0"],
5   ["y2", null, "221.0", "321.0", "421.012345678987", null, null, "222.0", "322.0", "422.0"]
6 ]
```



# Обработка файла JSON

## Результат обработки

```
1 [
2   ["period", "name", "y1", "y2"],
3   ["01-20.01.2024", "x1", null, null],
4   ["01-20.01.2024", "x2", "211.0", "221.0"],
5   ["01-20.01.2024", "x3", "311.0", "321.0"],
6   ["01-20.01.2024", "x4", "411.0", "421.012345678987"],
7   ["01-20.01.2024", "sum", null, null],
8   ["21-31.01.2024", "x1", null, null],
9   ["21-31.01.2024", "x2", "212.0", "222.0"],
10  ["21-31.01.2024", "x3", "312.0", "322.0"],
11  ["21-31.01.2024", "x4", "412.0", "422.0"]
12 ]

{
  "operation": "shift",
  "spec": {
    "*": {
      "*": "[#1]"
    }
  }
},
```

```
1 [null, null, null, ["period", "name", "y1", "y2"],
2   ["01-20.01.2024", "x1", null, null],
3   ["01-20.01.2024", "x2", "211.0", "221.0"],
4   ["01-20.01.2024", "x3", "311.0", "321.0"],
5   ["01-20.01.2024", "x4", "411.0", "421.012345678987"],
6   ["01-20.01.2024", "sum", null, null],
7   ["21-31.01.2024", "x1", null, null],
8   ["21-31.01.2024", "x2", "212.0", "222.0"],
9   ["21-31.01.2024", "x3", "312.0", "322.0"],
10  ["21-31.01.2024", "x4", "412.0", "422.0"]
11 ]
```

```
1 [
2   ["period", "01-20.01.2024", "01-20.01.2024", "01-20.01.2024", "01-20.01.2024", "01-20.01.2024"],
3   ["name", "x1", "x2", "x3", "x4", "sum", "x1", "x2", "x3", "x4"],
4   ["y1", null, "211.0", "311.0", "411.0", null, null, "212.0", "312.0", "412.0"],
5   ["y2", null, "221.0", "321.0", "421.012345678987", null, null, "222.0", "322.0", "422.0"]
6 ]
```

## Исходный файл

H17 421,012345678987

	C	D	E	F	G	H	I	J	K	L	M	N	O
13													
14		period	01-20.01.2024				21-31.01.2024						
15		name	x1	x2	x3	x4	sum	x1	x2	x3	x4	sum	
16		y1		211	311	411	933		212	312	412	936	
17		y2		221	321	421,0123	963,0123		222	322	422	966	
18													
19													
20		sum	0	432	632	832,0123	1896,012	0	434	634	834	1902	
21													

	K	L	M	N	O	P	Q	R	S	T	U	V	W
20													
21		date	1	2	3	4	5	6	7	8	9	10	11
22		x1	1	2	3	4	5	6	7	8	9	10	11
23		x2	2	4	6	8	10	12	14	16	18	20	22
24		x3	3	6	9	12	15	18	21	24	27	30	33
25		x4	4	8	12	16	20	24	28	32	36	40	44
26													

Configure Controller Service | ExcelReader 1.27.0

DISABLED

SETTINGS PROPERTIES COMMENTS

Required field

Property	Value
Schema Access Strategy	Infer Schema
Starting Row	14
Required Sheets	No value set
Protection Type	Unprotected
Date Format	No value set
Time Format	No value set
Timestamp Format	No value set

CANCEL APPLY



## Результат обработки

View as: original



```
1 Column1,Column2,Column3,Column4,Column5,Column6,Column7,Column8,Column9,Column10,Column11,Column12,Column13
2 ,,period,01-20.01.2024,,,,,21-31.01.2024,,
3 ,,name,x1,x2,x3,x4,sum,x1,x2,x3,x4
4 ,,y1,,211.0,311.0,411.0,,,212.0,312.0,412.0
5 ,,y2,,221.0,321.0,421.012345678987,,,222.0,322.0,422.0
6 ,,,,,,,,,,
7 ,,,,,,,,,,
8 ,,sum,,,,,,,,,
9
```



## Выводы

---

- Apache NiFi предоставляет большой набор процессоров для обработки различных файлов, с которыми достаточно просто научиться работать
- Несмотря на простоту использования NiFi, обработка Excel-файлов может быть достаточно трудной задачей



# СПАСИБО!

## О компании AXENIX

AXENIX (ex-Accenture) – российская компания, предоставляющая широкий спектр профессиональных услуг в области цифровых сервисов, облачных технологий и решений для обеспечения информационной безопасности. В офисах и центрах разработки в Москве, Твери, Ростове-на-Дону, Краснодаре, Санкт-Петербурге и Алматы работают около 2 000 сотрудников. Благодаря сочетанию уникальных знаний, опыта и компетенций более чем в 40 отраслях, предлагает услуги в области стратегии и бизнес-консалтинга, технологических решений и других операций, направленных на цифровизацию бизнеса. Подробнее на [axenix.pro](https://axenix.pro)

