

Ревью и деплой потоков NiFi через Git

Попова Клавдия
Инженер данных, Сибур Диджитал

Сентябрь 2026

О себе

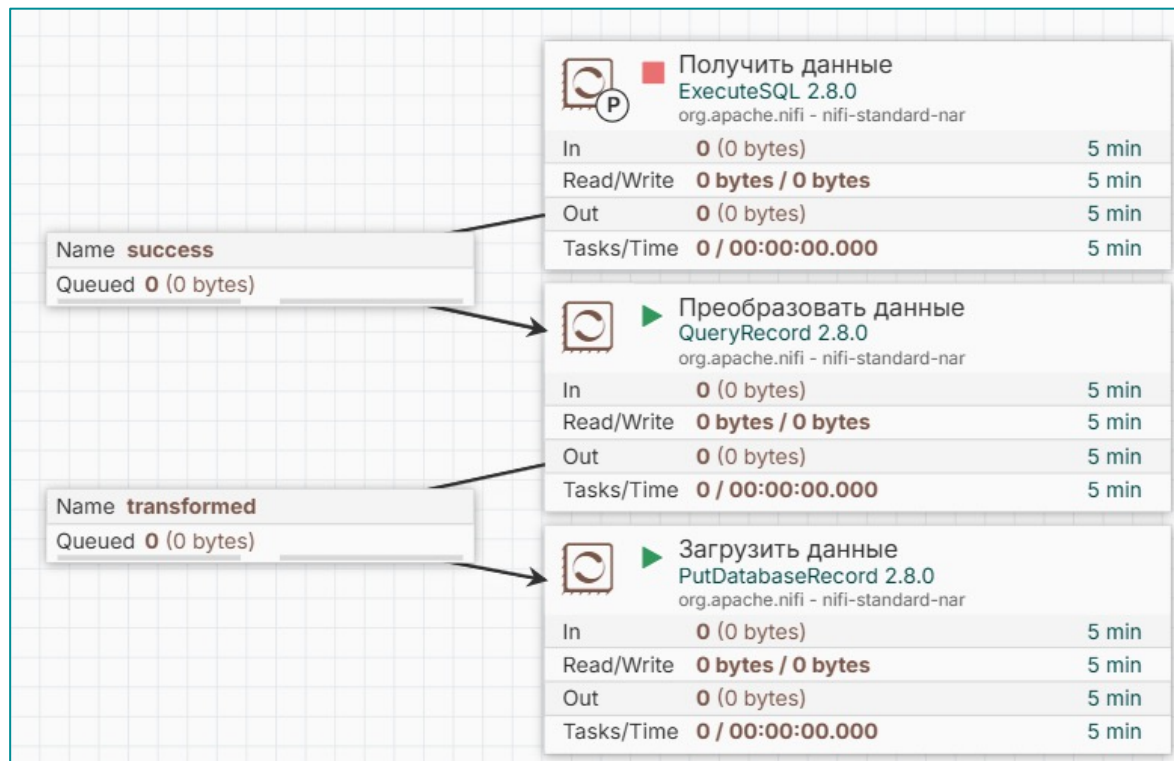
- Работаю с NiFi с 2020 года
- Инженер данных в Сибур Диджитал
- Потоки, процессоры, мониторинг, lineage
- mail: popovaks@sibur.ru, tg: [@klavoch](https://www.t.me/klavoch)



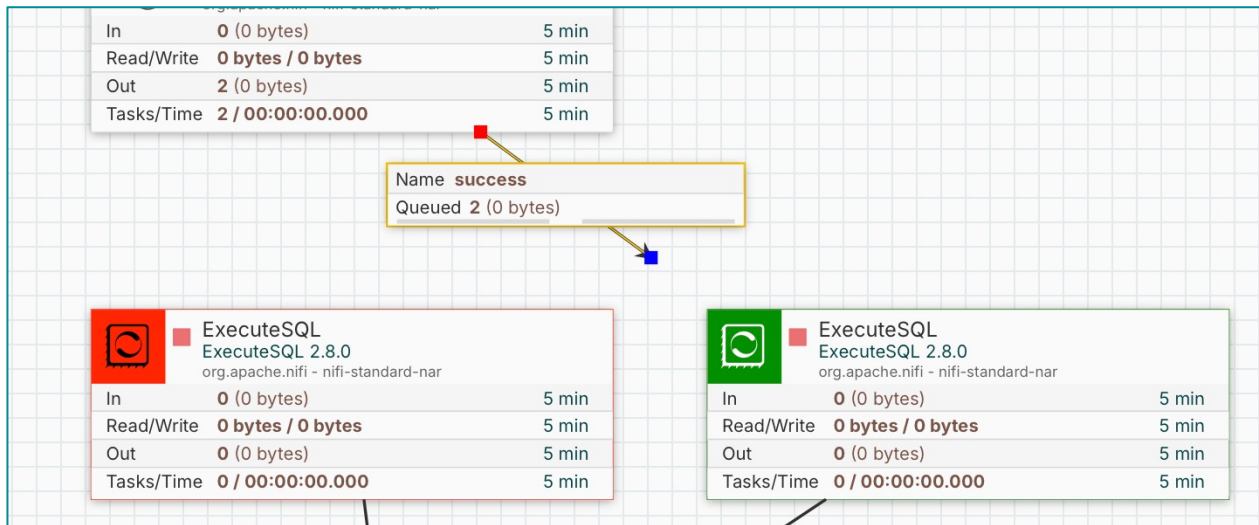
План доклада

1. Проблематика ревью и деплоя в NiFi
2. NiFi Registry — решение проблем контроля версий
3. 2.x: почему не подходит только `Git*RegistryClient`
4. Кастомный git-репозиторий: читаемый вид
5. Ограничения подхода
6. Выводы

Что такое NiFi



Прелести визуального редактора

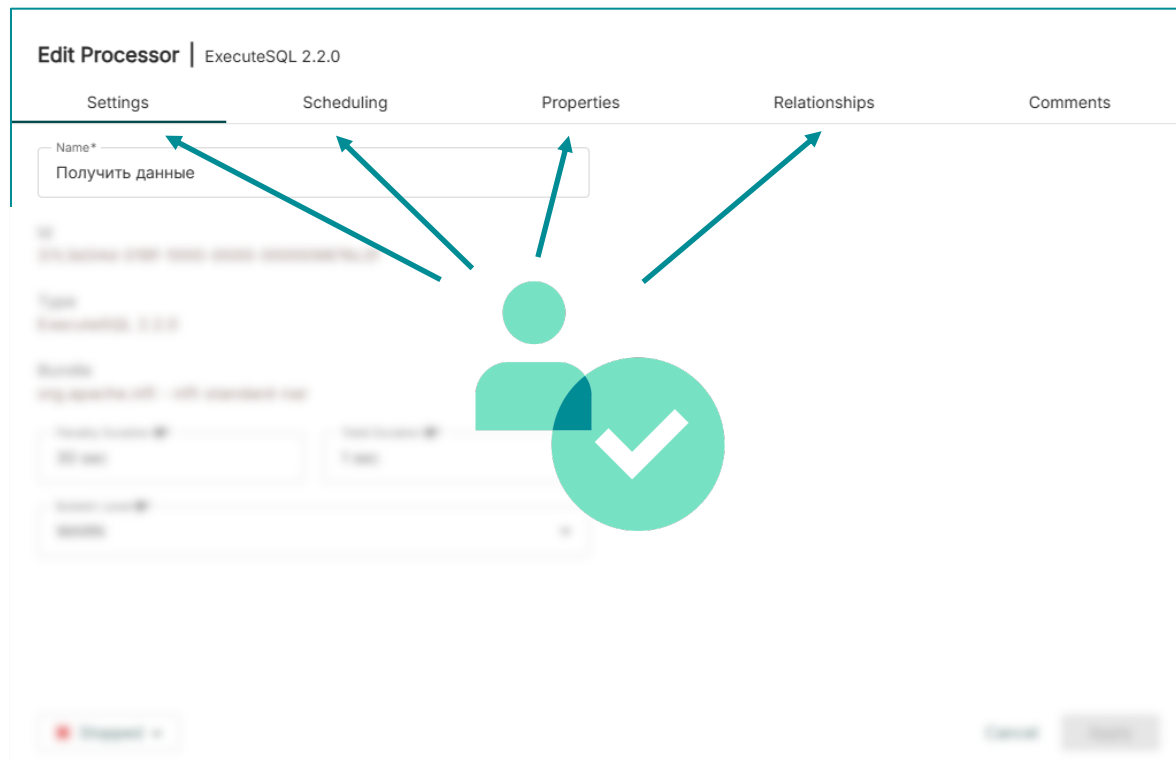


Зачем нужно ревью?

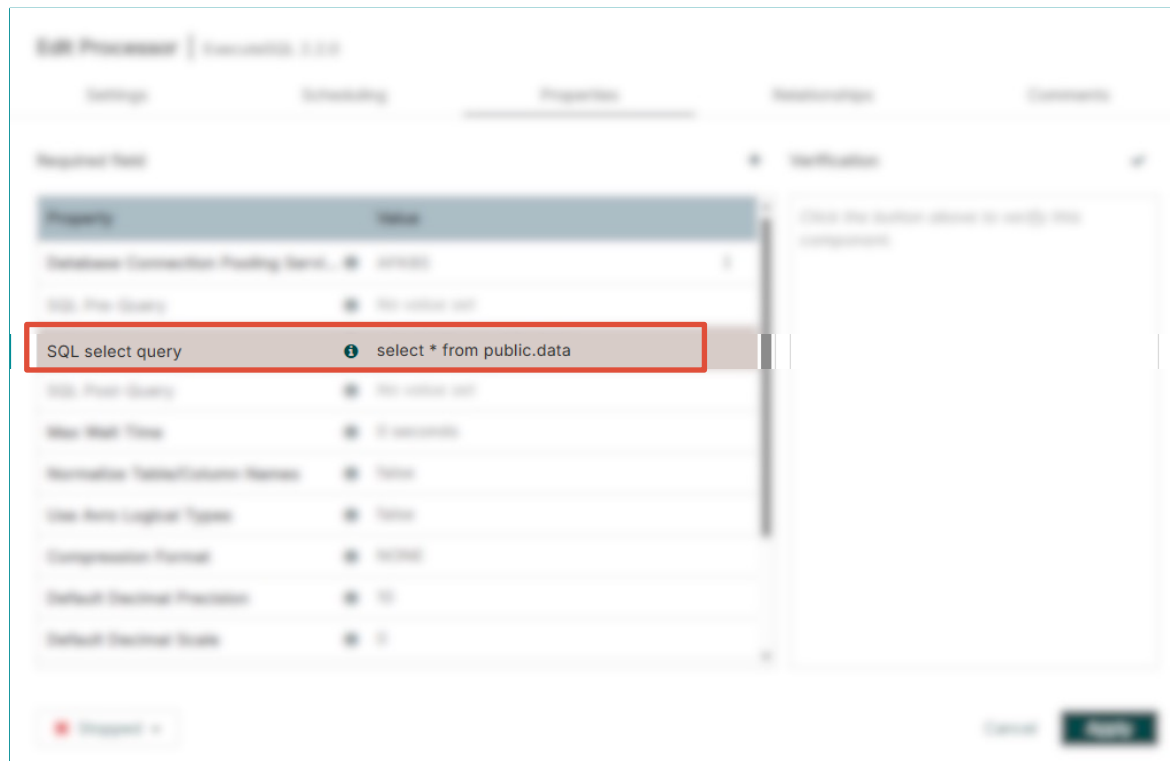


Проблематика ревью

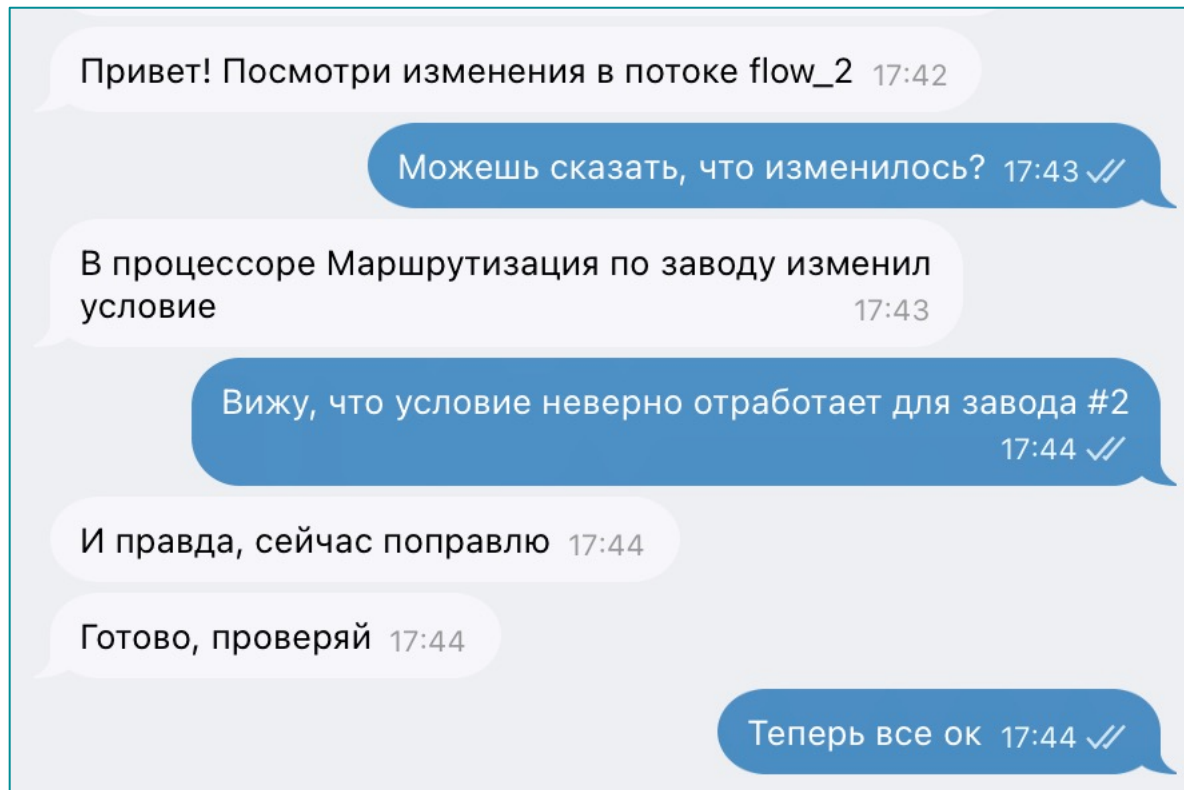
Проблематика ревью: визуальный редактор



Проблематика ревью: актуальное состояние



Проблематика ревью: ведение дискуссии



Проблематика деплоя: перенос через templates

KLIK_OTIPB		
 0	 0	 0
 3	 2	 0
Queued	0 (0 bytes)	
In	0 (0 bytes) → 0	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 → 0 (0 bytes)	5 min
 0	 0	 0
 0	 0	 0

Проблематика деплоя: много ручных действий

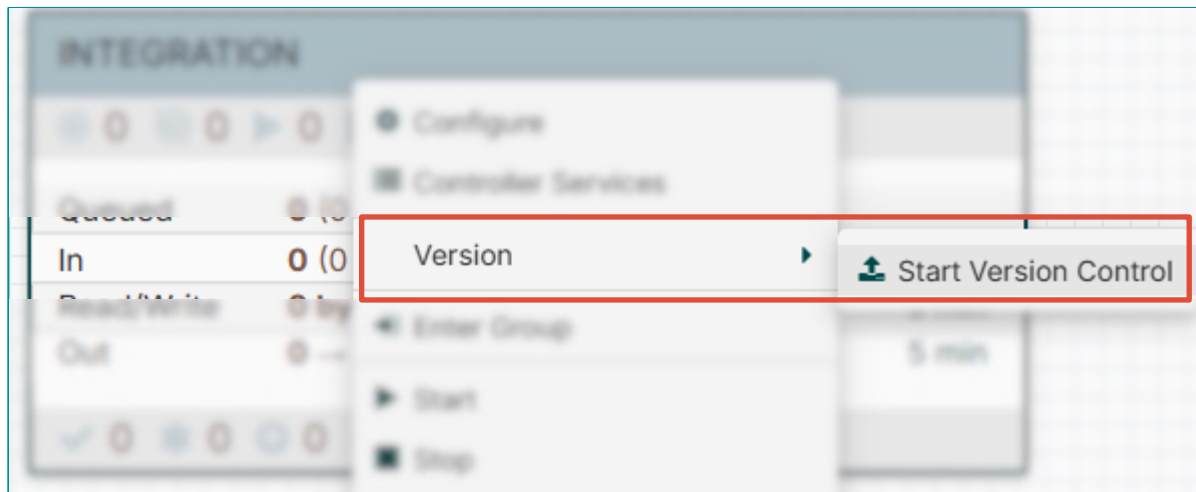


Проблематика подхода

- × Актуальное состояние
- × Ручные действия
- × Визуальное ревью
- × Общение в личных переписках

NiFi Registry client

Что такое NiFi Registry



Проблематика ревью: NiFi Registry, детализация изменений

Show Local Changes

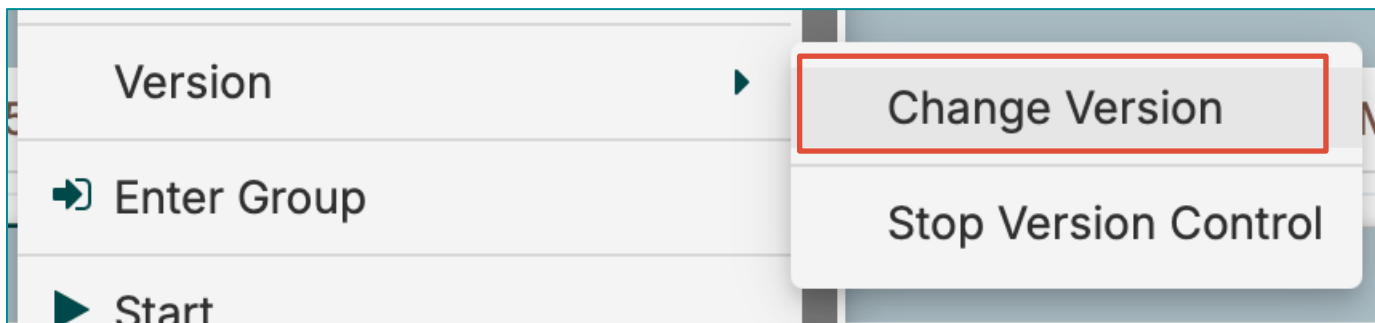
The following changes have been made to UNITED_ETL_VERTICAL_DM_KAFKA (Version 4).

Filter

Displaying 2 of 2

Component Name ↑	Change Type	Difference	
	Component Added	Label was added	...
Ручной запуск	Property Added	Property 'Custom Text' was added	...

Деплой стал легче



NiFi Registry: решённые проблемы

- ✓ Актуальное состояние
- ✓ Ручные действия
- ✗ Визуальное ревью
- ✗ Общение в личных переписках

Opensource интеграция с git

NiFi Registry flow version – json config

```
{
  "snapshotMetadata": {
    "bucketIdentifier": "4d5e6f7a-1111-2222-3333-abcdef123456",
    "flowIdentifier": "7b8c9d0e-4444-5555-6666-abcdef654321",
    "version": 1,
    "timestamp": 1785718400000,
    "author": "admin",
    "comments": "Initial production data ingest baseline."
  },
  "flowContents": {
    "identifier": "root-process-group-id",
    "name": "S3 Data Ingestion Flow",
    "comments": "Fetches raw files, appends staging metadata, and forwards payload.",
    "position": {
      "x": 0.0,
      "y": 0.0
    }
  },
}
```

NiFi Registry + git / Git*FlowRegistryClient

 MTOIR_TOIR_POSTGRESQL_ODS

 0  0  61  0  0  1

Queued	0 (0 bytes)	
In	0 (0 bytes) → 0	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 → 0 (0 bytes)	5 min

 0  0  0  0  0



 1 file +5110 -30

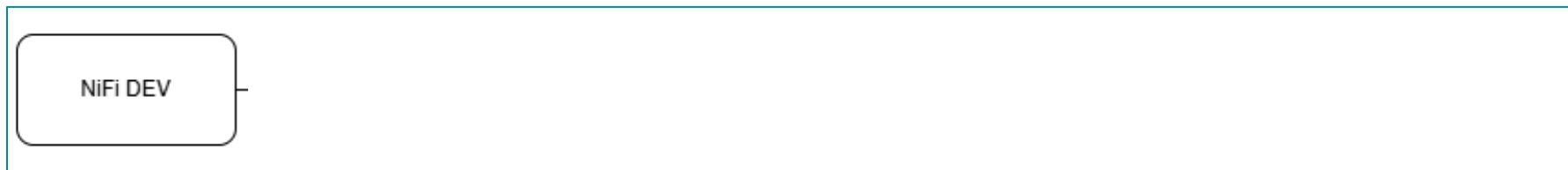
GitHub Action

```
### Executing Snowflake Flow Diff for flow: `MyExample`  
  
#### Checkstyle Violations  
> [!CAUTION]  
> - Processor named `UpdateAttribute` is configured with 4 concurrent tasks  
  
#### Flow Changes  
- The destination of a connection has changed from `UpdateAttribute` to `InvokeHTTP`  
- A self-loop connection `[success]` has been added on `UpdateAttribute`  
- A connection `[success]` from `My Generate FlowFile Processor` to `UpdateAttribute` has been added  
- A Processor of type `GenerateFlowFile` named `GenerateFlowFile` has been renamed from `GenerateFlowFil`  
- In Processor of type `GenerateFlowFile` named `GenerateFlowFile`, the Scheduling Strategy changed from  
- A Parameter Context named `Test Parameter Context` has been added  
- The Parameter Context `Test Parameter Context` with parameters `{addedParam=newValue}` has been added  
- A Processor of type `UpdateAttribute` named `UpdateAttribute` has been removed  
- In Processor of type `GenerateFlowFile` named `GenerateFlowFile`, the Run Schedule changed from `1 min`  
- A Parameter Context named `Another one to delete` has been added  
- A Processor of type `UpdateAttribute` named `UpdateAttribute` has been added with the configuration [  
  - `Store State` = `Do not store state`  
  - `canonical-value-lookup-cache-size` = `100`  
  
#### Bundle Changes  
- The bundle `org.apache.nifi:nifi-standard-nar` has been changed from version `2.1.0` to version `2.2.0`
```

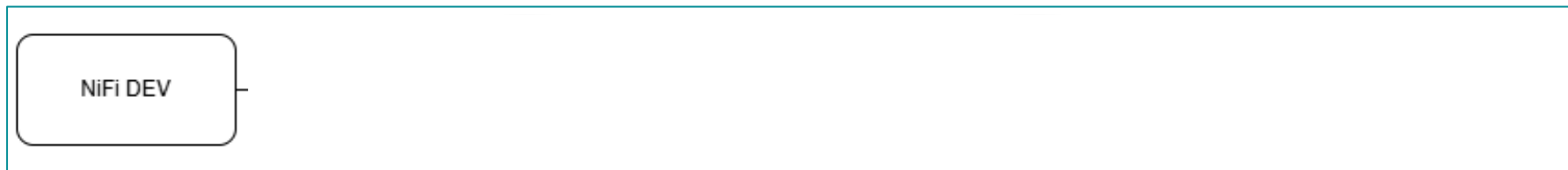
Snowflake-Labs/snowflake-flow-diff — GitHub Action для сравнения различий в определениях потоков Apache NiFi

Целевое решение

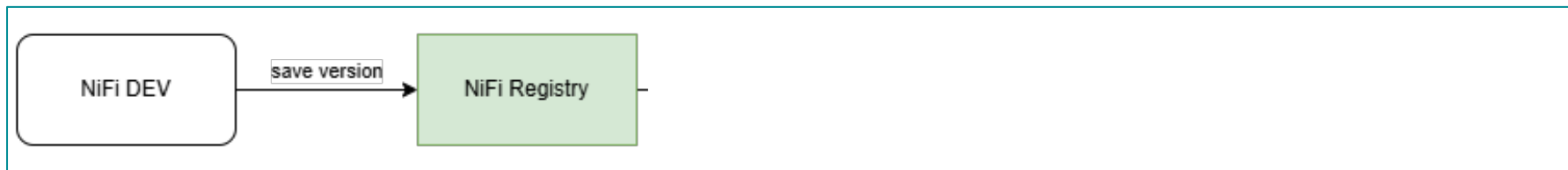
Целевое решение



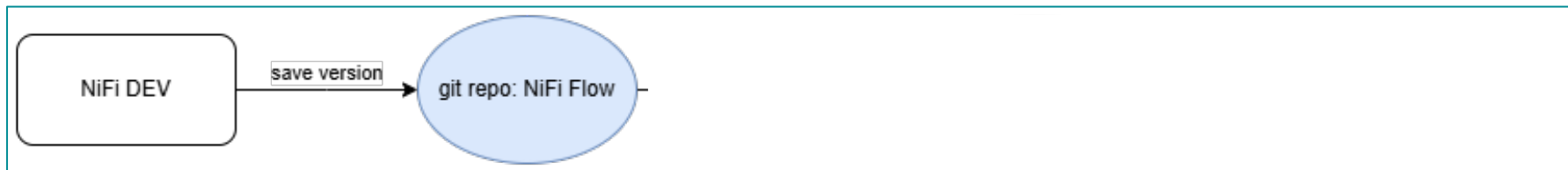
или



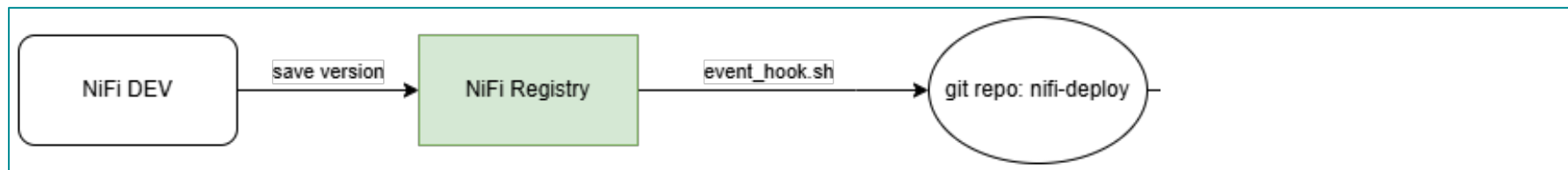
Целевое решение



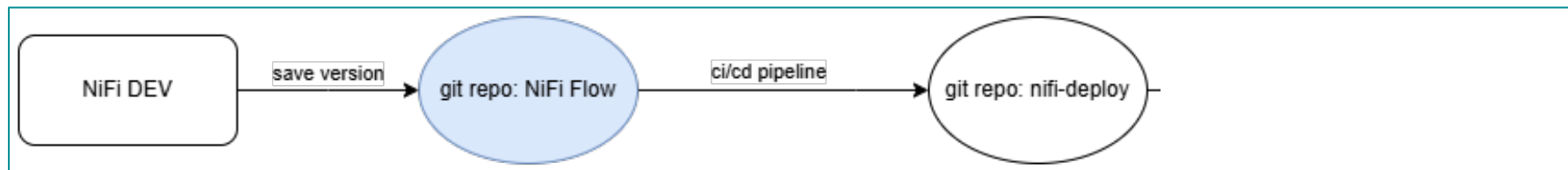
или



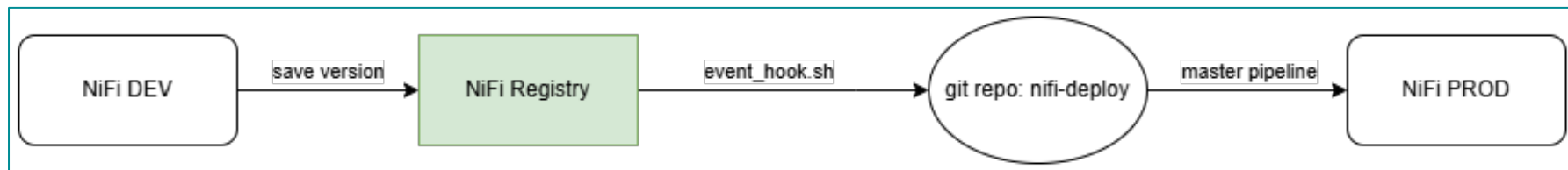
Целевое решение



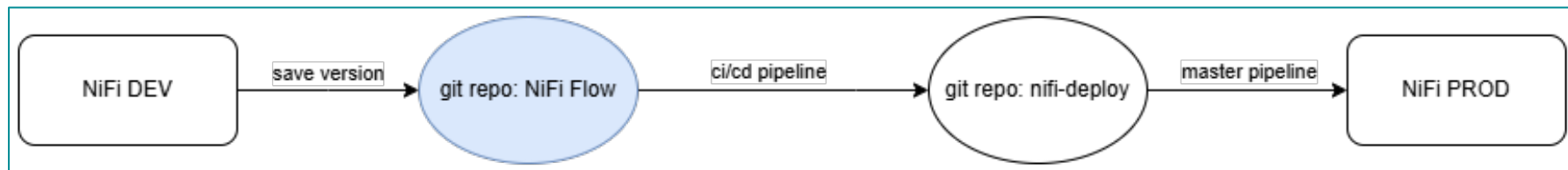
или



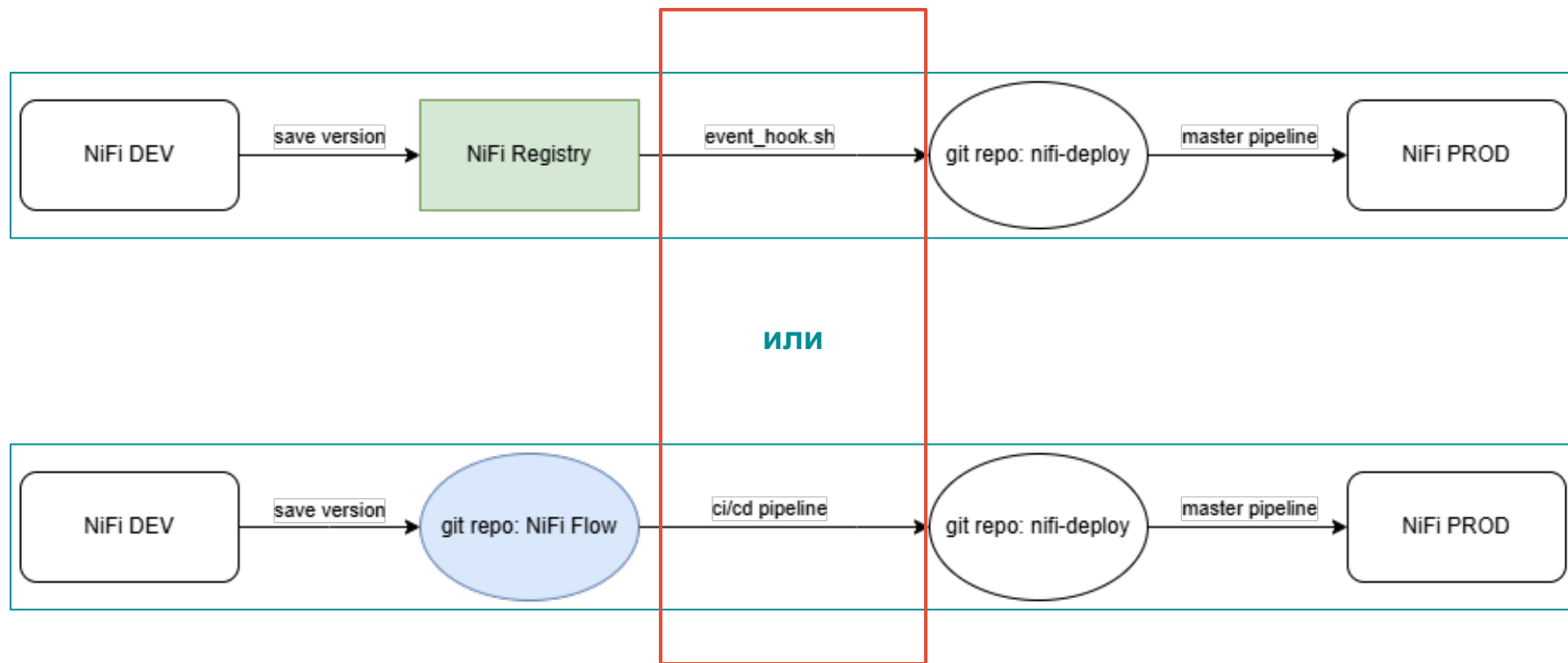
Целевое решение



или

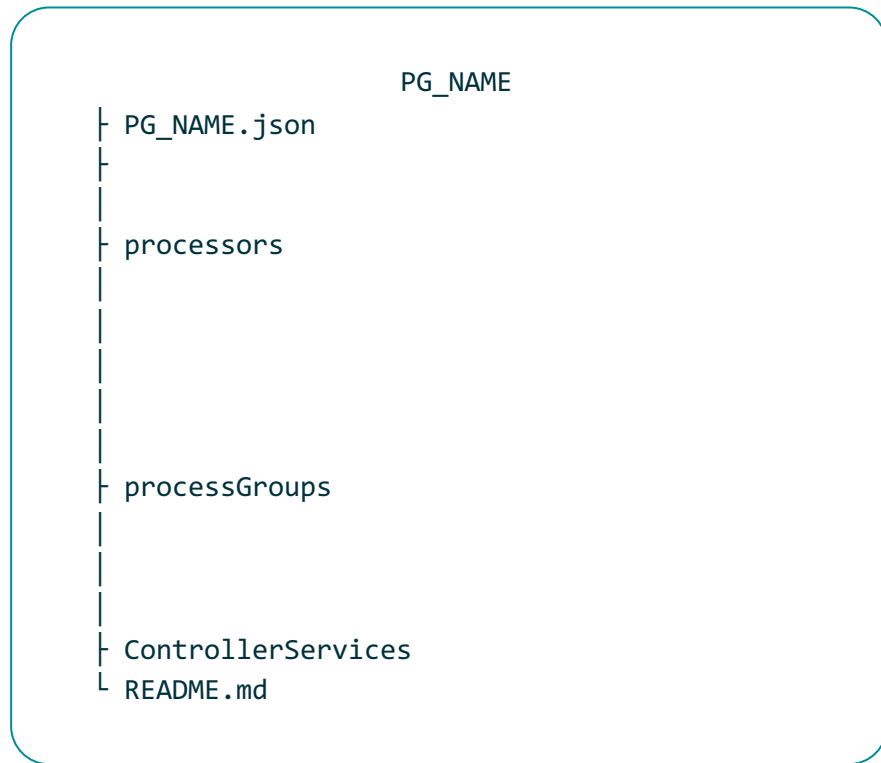


Целевое решение



Декомпозиция json

Читаемый вид: декомпозиция Flow



Читаемый вид: декомпозиция Flow

```
PG_NAME
├ PG_NAME.json
├ connections
│   └ validation_result.json
├ processors
│
│
├ processGroups
│
│
├ ControllerServices
└ README.md
```

Читаемый вид: декомпозиция Flow

```
PG_NAME
├ PG_NAME.json
├ connections
│   └ validation_result.json
├ processors
│   ├── processor_1_id.json
│   ├── processor_2_id.json
│   │
│   ├── processor_3_id.json
│   └ ...
├ processGroups
│
└ ControllerServices
    └ README.md
```

Читаемый вид: декомпозиция Flow

```
PG_NAME
├ PG_NAME.json
├ connections
│   └ validation_result.json
├ processors
│   ├── processor_1_id.json
│   ├── processor_2_id.json
│   │
│   ├── processor_3_id.json
│   └ ...
├ processGroups
│   ├── SUBPG_1_NAME
│   ├── SUBPG_2_NAME
│   └ SUBPG_3_NAME
├ ControllerServices
└ README.md
```

Читаемый вид: декомпозиция Flow

```
PG_NAME
├ PG_NAME.json
├ connections
│   └ validation_result.json
├ processors
│   ├── processor_1_id.json
│   ├── processor_2_id.json
│   ├── processor_2_id_sql_query.sql
│   ├── processor_3_id.json
│   └ ...
├ processGroups
│   ├── SUBPG_1_NAME
│   ├── SUBPG_2_NAME
│   └ SUBPG_3_NAME
├ ControllerServices
└ README.md
```

Читаемый вид: многострочные property

 c6d3f0aa-9281-3aa9-b185-3b18fab1a1d1.json 436 B

```
1 {
2   "validation_result": {
3     "isValid": 0,
4     "errorCode": 0
5   },
6   "identifier": "c6d3f0aa-9281-3aa9-b185-3b18fab1a1d1",
7   "name": "Сервис в формате json с sql-запросом",
8   "properties": {
9     "Database Connection Pooling Service": "DBCP",
10    "SQL select query": "c6d3f0aa-9281-3aa9-b185-3b18fab1a1d1_sql_query.sql"
11  },
12  "type": "org.apache.nifi.processors.standard.ExecuteSQL"
13 }
```

 c6d3f0aa-9281-3aa9-b185-3b18fab1a1d1_sql_query.sql 198 B

```
1 select top 1
2 Id,
3 Code,
4 PrimitiveEntityDataStateId,
5 CreationDate,
6 CreationUserId,
7 LastChangeDate,
8 LastChangeUserId,
9 Name_ru,
10 FullAddress,
11 Name_en,
12 Code_NSI
13 from ${schema_name_src}.${table_name_src}
```

Читаемый вид: ControllerService id

 c6d3f0aa-9281-3aa9-b185-3b18fab1a1d1.json 436 B

```
1  {
2    "controllerService": {
3      "id": "SVT2",
4      "name": "Database Connection Pooling Service",
5      "type": "org.apache.nifi.processors.standard.ExecuteSQL",
6      "properties": {
7        "Database Connection Pooling Service": "SVT2",
8        "SQL select query": "select * from nifi.processors.standard.ExecuteSQL"
9      }
10   }
11   "type": "org.apache.nifi.processors.standard.ExecuteSQL"
12 }
13 }
```

Очистка json

Читаемый вид: Проверил — удалил

```
1 1 {
2   - "autoTerminatedRelationships": [
3     - "unmatched"
4   ],
5   - "backoffMechanism": "PENALIZE_FLOWFILE",
6   - "bulletinLevel": "WARN",
7   - "comments": "",
8   - "componentType": "PROCESSOR",
9   - "concurrentlySchedulableTaskCount": 1,
10  - "executionNode": "ALL",
11  2 + "validation_result": {
12    + "Invalid": {},
13    + "Uncommon": {}
14  },
15  6 "identifier": "b781c93c-8de6-32bf-8523-f02bd8be42f5",
16  - "maxBackoffPeriod": "10 mins",
17  7 "name": "Запись названия таблицы",
18  - "penaltyDuration": "30 sec",
19  8 "properties": {
20    - "Character Set": "UTF-8",
21    - "Enable Canonical Equivalence": "false",
22    - "Enable Case-insensitive Matching": "false",
23    - "Enable DOTALL Mode": "false",
24    - "Enable Literal Parsing of the Pattern": "false",
25    - "Enable Multiline Mode": "false",
26    - "Enable Unicode Predefined Character Classes": "false",
27    - "Enable Unicode-aware Case Folding": "false",
28    - "Enable Unix Lines Mode": "false",
29    9 "Include Capture Group 0": "true",
30    10 "Maximum Buffer Size": "1 MB",
31    11 "Maximum Capture Group Length": "1024",
32    - "Permit Whitespace and Comments in Pattern": "false",
33    - "extract-text-enable-named-groups": "false",
34    - "extract-text-enable-repeating-capture-group": "false",
35    12 "table_name": "(?s)(^.*$)"
36  },
37  13 "retriedRelationships": [],
38  - "retryCount": 10,
39  - "runDurationMillis": 0,
40  - "scheduledState": "ENABLED",
41  - "schedulingPeriod": "0 sec",
42  - "schedulingStrategy": "TIMER_DRIVEN",
43  - "style": {},
44  - "type": "org.apache.nifi.processors.standard.ExtractText",
45  - "yieldDuration": "1 sec",
46  14 + "type": "org.apache.nifi.processors.standard.ExtractText"
47  15 }
```

Правила очистки: поля, которые удаляются всегда

```
{
  "autoTerminatedRelationships": [
    "failure"
  ],
  "backoffMechanism": "PENALIZE_FLOWFILE",
  "bulletinLevel": "WARN",
  "comments": "",
  "componentType": "PROCESSOR",
  "concurrentlySchedulableTaskCount": 1,
  "executionMode": "ALL",
  "identifier": "bb76ae714-4aeb-3451-ba0f-ef61a82581fe",
  "maxBackoffPeriod": "10 mins",
  "name": "Get_table_schema",
  "penaltyDuration": "30 sec",
  "retriedRelationships": [],
  "retryCount": 10,
  "runDurationMillis": 0,
  "scheduledState": "ENABLED",
  "schedulingPeriod": "0 sec",
  "schedulingStrategy": "TIMER_DRIVEN",
  "style": {},
  "type": "org.apache.nifi.processors.standard.ExecuteSQL",
  "yieldDuration": "1 sec"
}
```

Правила очистки: критичные поля

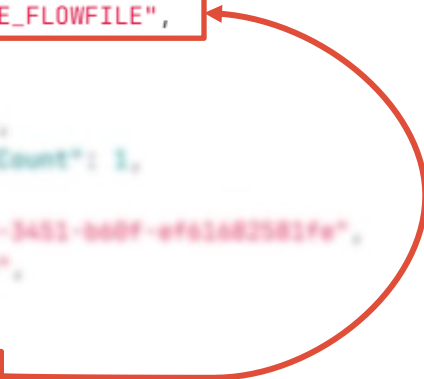
```
{
  "autoTerminatedRelationships": [
    "failure"
  ],
  "backoffMechanism": "PENALIZE_FLOWFILE",
  "bulletinLevel": "WARN",
  "comments": "",
  "componentType": "PROCESSOR",
  "concurrentlySchedulableTaskCount": 1,
  "executionNode": "ALL",
  "identifier": "8076ef14-4aeb-3451-ba0f-e961e82581fe",
  "maxBackoffPeriod": "10 mins",
  "name": "Get_table_schema",
  "penaltyDuration": "30 sec",
  "retriedRelationships": [],
  "retryCount": 10,
  "runDurationMillis": 0,
  "scheduledState": "ENABLED",
  "schedulingPeriod": "0 sec",
  "schedulingStrategy": "TIMER_DRIVEN",
  "style": {},
  "type": "org.apache.nifi.processors.standard.ExecuteSQL",
  "yieldDuration": "1 sec"
}
```

Правила очистки: некритичные поля

```
{  
  "autoTerminatedRelationships": [  
    "failure"  
  ],  
  "backoffMechanism": "PENALIZE_FLOWFILE",  
  "bulletinLevel": "WARN",  
  "comments": "",  
  "componentType": "PROCESSOR",  
  "concurrentlySchedulableTaskCount": 1,  
  "executionMode": "ALL",  
  "identifier": "8076ef14-4aeb-3451-ba0f-ef61a82581fe",  
  "maxBackoffPeriod": "10 mins",  
  "name": "Get_table_schema",  
  "penaltyDuration": "30 sec",  
  "retriedRelationships": [],  
  "retryCount": 10,  
  "runDurationMillis": 0,  
  "scheduledState": "ENABLED",  
  "schedulingPeriod": "0 sec",  
  "schedulingStrategy": "TIMER_DRIVEN",  
  "style": {},  
  "type": "org.apache.nifi.processors.standard.ExecuteSQL",  
  "yieldDuration": "1 sec"  
}
```

Правила очистки: зависимые поля

```
{  
  "autoTerminatedRelationships": [  
    "failure"  
  ],  
  "backoffMechanism": "PENALIZE_FLOWFILE",  
  "bulletinLevel": "WARN",  
  "comments": "",  
  "componentType": "PROCESSOR",  
  "concurrentlySchedulableTaskCount": 1,  
  "executionMode": "ALL",  
  "identifier": "8076ef14-4aeb-3451-ba0f-ef61e82581fe",  
  "maxBackoffPeriod": "10 mins",  
  "name": "Get_table_schema",  
  "penaltyDuration": "30 sec",  
  "retriedRelationships": [],  
  "retryCount": 10,  
  "runDurationMillis": 0,  
  "scheduledState": "ENABLED",  
  "schedulingPeriod": "0 sec",  
  "schedulingStrategy": "TIMER_DRIVEN",  
  "style": {},  
  "type": "org.apache.nifi.processors.standard.ExecuteSQL",  
  "yieldDuration": "1 sec"  
}
```



Правила очистки: результат

```
{
  "validation_result": {
    "Invalid": {},
    "Uncommon": {}
  },
  "identifier": "bb76af14-4aeb-3451-b60f-ef61682581fe",
  "name": "Get_table_schema",
  "properties": {
    "Database Connection Pooling Service": "MTOIR TOIR",
    "SQL Query": "SELECT * FROM \"${schema_name_src}\".\"${table_name_src}\" LIMIT 1;"
  },
  "type": "org.apache.nifi.processors.standard.ExecuteSQL"
}
```

Сохранение нарушений

Читаемый вид: validation_result

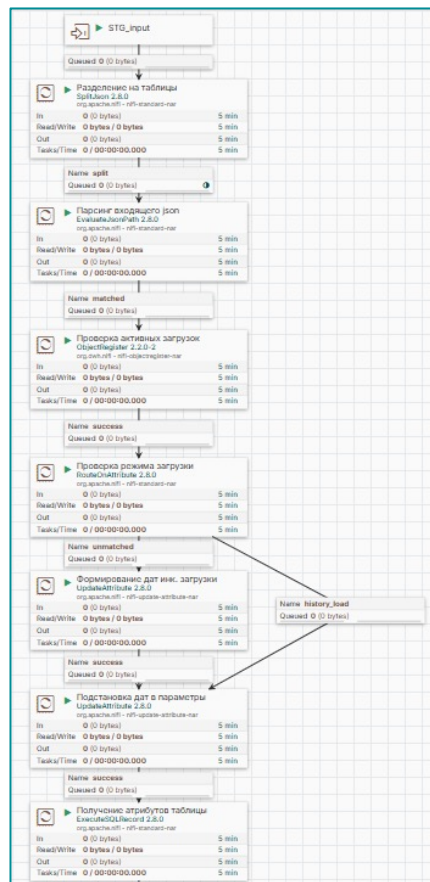
```
{
  "validation_result": {
    "Invalid": {
      "scheduledState": "DISABLED"
    },
    "Uncommon": {}
  },
  "identifier": "a837aadb-d52f-37e5-aabb-9b46d10bacc3",
  "name": "SetInputTables.json",
  "properties": {
    "generate-ff-custom-text": "${sap.input_json}"
  },
  "schedulingPeriod": "0000000000000000 sec",
  "type": "org.apache.nifi.processors.standard.GenerateFlowFile"
}
```

Читаемый вид: validation_result.json

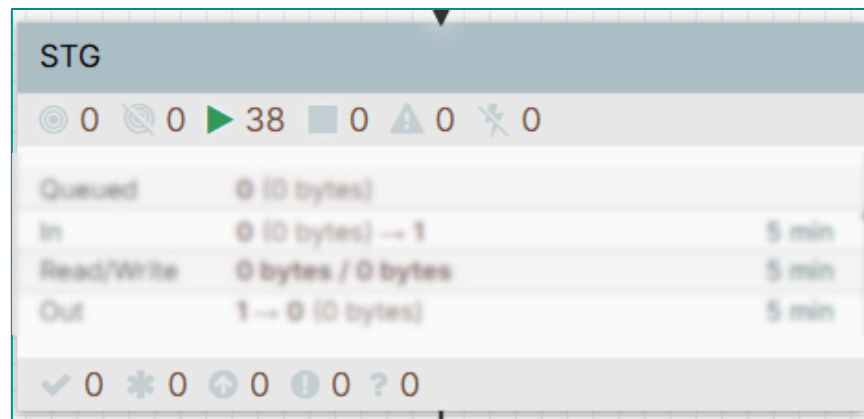
```
validation_result.json 116 B
1 {
2   "Invalid": {
3     "3aecd460-a7ab-3c7c-91e3-d42691731a91_tech_is_dead_end": true
4   },
5   "Uncommon": {}
6 }
```

Восстановление логики потока

Читаемый вид: README.md



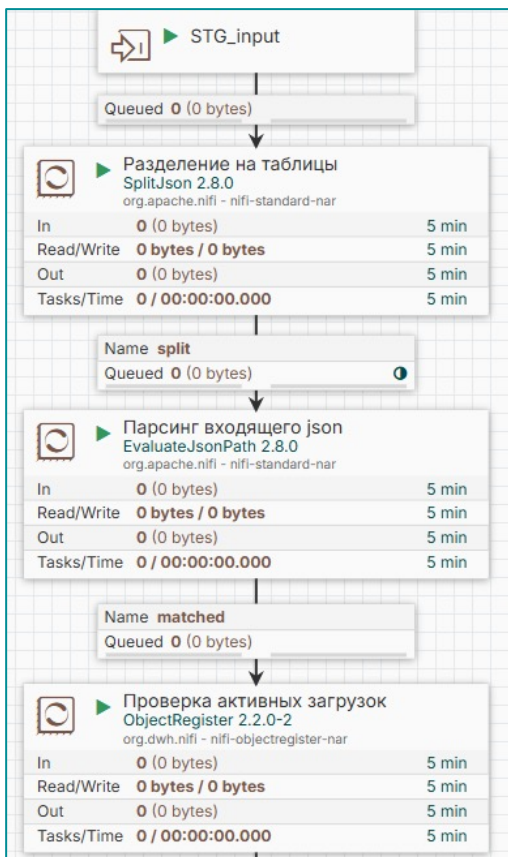
Читаемый вид: README.md



Processor Group: STG

- Общее кол-во процессоров в текущей ПГ: 36
- Кол-во процессоров в текущей ПГ: 36
- Кол-во ПГ: 0

Читаемый вид: README.md



Детализация основной группы 1

1.1. STG_input (INPUT_PORT)

1.2. from_input_port Разделение на таблицы (SplitJson)

1.3. split Парсинг входящего json (EvaluateJsonPath)

1.4. matched Проверка активных загрузок (ObjectRegister)

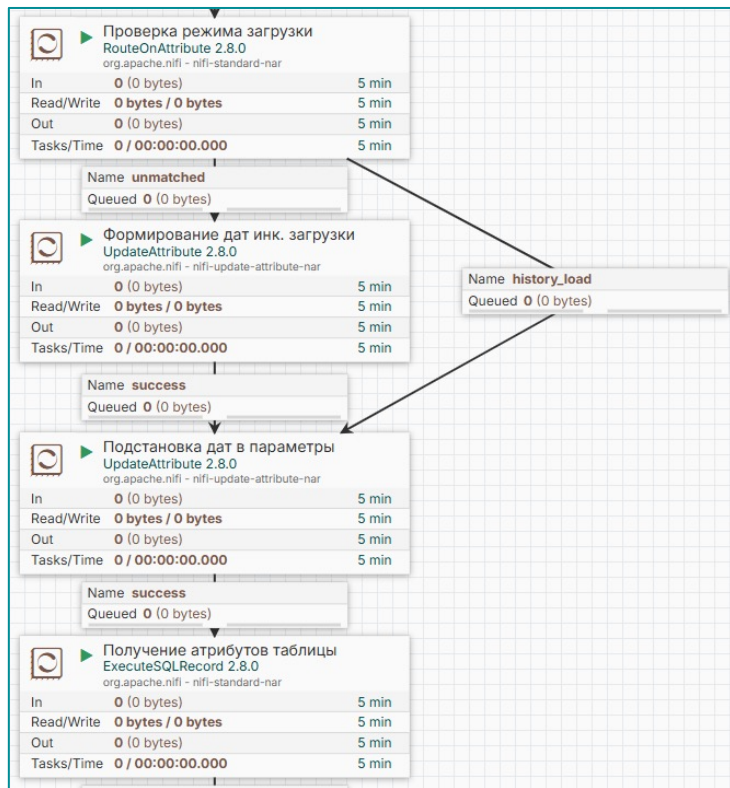
1.5. matched Проверка режима загрузки (RouteOnAttribute) (1.7. History_load)

1.6. matched Формирование дат ик. загрузки (UpdateAttribute)

1.7. matched Подготовка дат в параметры (UpdateAttribute)

1.8. matched Получение атрибутов таблицы (ExecuteSQLRecord)

Читаемый вид: README.md



Детализация основной группы 1

1.1. **STD_inpt** (INPUT_PORT)

1.2. **from_input_port** Разделение на таблицы (SplitJoin)

1.3. **split** Парсинг входных json (EvaluateJsonPath)

1.4. **matched** Проверка активных загрузок (ObjectRegister)

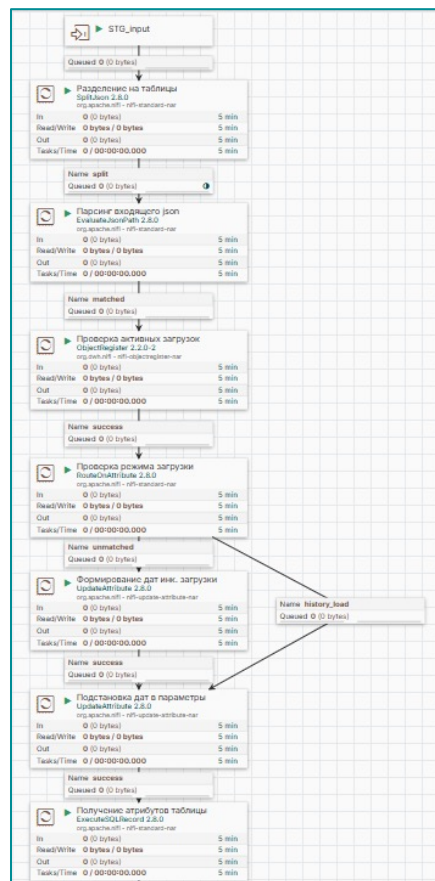
1.5. **success** Проверка режима загрузки (RouteOnAttribute). {1.7: history_load}

1.6. **unmatched** Формирование дат инк. загрузки (UpdateAttribute)

1.7. **success** Подстановка дат в параметры (UpdateAttribute)

1.8. **success** Получение атрибутов таблицы (ExecuteSQLRecord)

Читаемый вид: README.md



Processor Group: STG

- Общее кол-во процессоров в текущей ПГ: 36
- Кол-во процессоров в текущей ПГ: 36
- Кол-во ПГ: 0

Детализация основной группы 1

1.1. STG_input (INPUT_PORT)

1.2.from_input_port Разделение на таблицы (SplitJson)

1.3.split Парсинг входящего json (EvaluateJsonPath)

1.4.matched Проверка активных загрузок (ObjectRegister)

1.5.success Проверка режима загрузки (RouteOnAttribute). {1.7: history_load}

1.6.unmatched Формирование дат инк. загрузки (UpdateAttribute)

1.7.success Подстановка дат в параметры (UpdateAttribute)

1.8.success Получение атрибутов таблицы (ExecuteSQLRecord)

Читаемый вид: README.md*

✓ 1 STG

✓ processors

{...} 1.02 Разделение на таблицы.json

{...} 1.03 Парсинг входящего json.json

{...} 1.04 Проверка активных загрузок.json

{...} 1.05 Проверка режима загрузки.json

{...} 1.06 Формирование дат инк. загрузки.json

{...} 1.07 Подстановка дат в параметры.json

{...} 1.08 Получение атрибутов таблицы.json

☰ 1.08 Получение атрибутов таблицы.sql

Читаемый вид: итог

```
3 3      "enabled": {  
4 4      "process": {  
5 5      },  
6 +      "backoffMechanism": "PENALIZE_FLOWFILE",  
7 7      "identifier": "abcF0b43-1a70-3723-abc-c1f4d7c4c4Ba",  
8 8      "name": "Load to STD",  
9 +      "penaltyDuration": "30 sec",  
10 10     "properties": {  
11 11         "Column Options": "${column_to_insert}",  
12 12         "Database User": "${vertex_loader_username}",  
13 13         "Max Bin Age": "ads",  
14 14         "Maximum Number of Entries": "20000",  
15 15         "Maximum Number of Entries": "1000",  
16 16         "Parser Type": "V_COMPATIBLE",  
17 17         "Password": "${vertex_loader_password}",  
18 18         "Schema Name": "${stg_schema_name}",  
19 19         "Table Name": "${table_name}",  
20 20         "Vertex Connection URL": "${vertex_loader_url}"  
21 21     },  
22 +     "retriedRelationships": [  
23 +         "failure"  
24 +     ],  
25 +     "retryCount": 3,  
26 26     "schedulingPeriod": "1 sec",  
27 27     "type": "org.apache.nifi.processors.standard.CopyVertex.CopyVertex"  
28 28 }
```

Результат декомпозиции и очистки

- ✓ Актуальное состояние
- ✓ Ручные действия
- ✓ Визуальное ревью
- ✗ Общение в личных переписках

Git: одна точка взаимодействия

Git repo: commit



Version 12 - 15 days ago
by `popovaks@sibur.local`

delete extra json reader
Apr-23-2026 at 12:34 PM



Apr 23, 2026

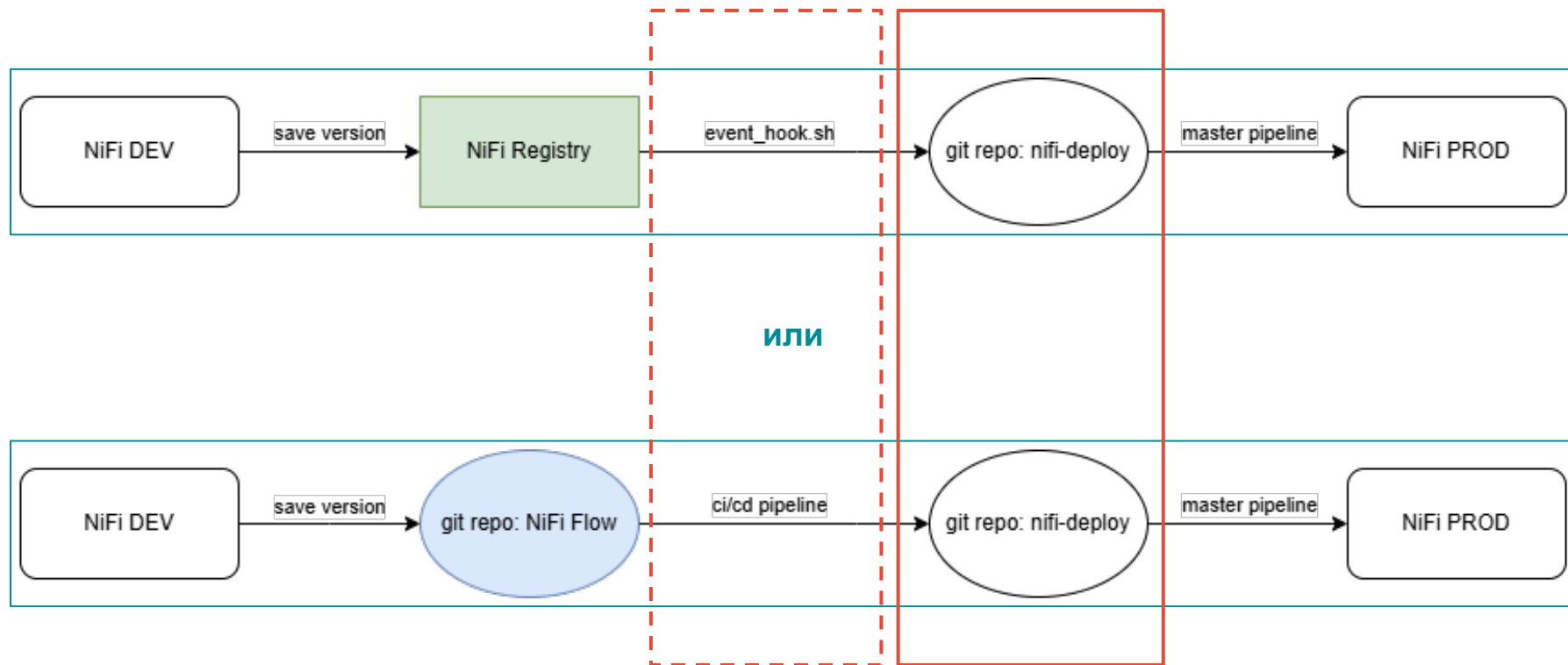


popovaks: delete extra json reader
a001-dwh-buildrobot authored 3 months ago

Git repo: validation_result

```
{
  "validation_result": {
    "Invalid": {
      "scheduledState": "DISABLED"
    },
    "Uncommon": {}
  },
  "identifier": "b837ad0b-d52f-37e5-a886-9b44d10baeb3",
  "name": "SetInputTables.json",
  "properties": {
    "generate-flow-custom-text": "${sap.input_json}"
  },
  "schedulingPeriod": "000000000000000000 sec",
  "type": "org.apache.nifi.processors.standard.GenerateFlowFile"
}
```

Git репо: пространство взаимодействия



Git репо: автоматические треды

SAP_SM_API_ODS/processors/e1cd29e1-680f-3b21-aa13-d09959b0002c.json: Значение DISABLED не прошло автоматическую проверку.



```
1 + {  
2 +   "validation_result": {  
3 +     "Invalid": {  
4 +       "scheduledState": "DISABLED"
```

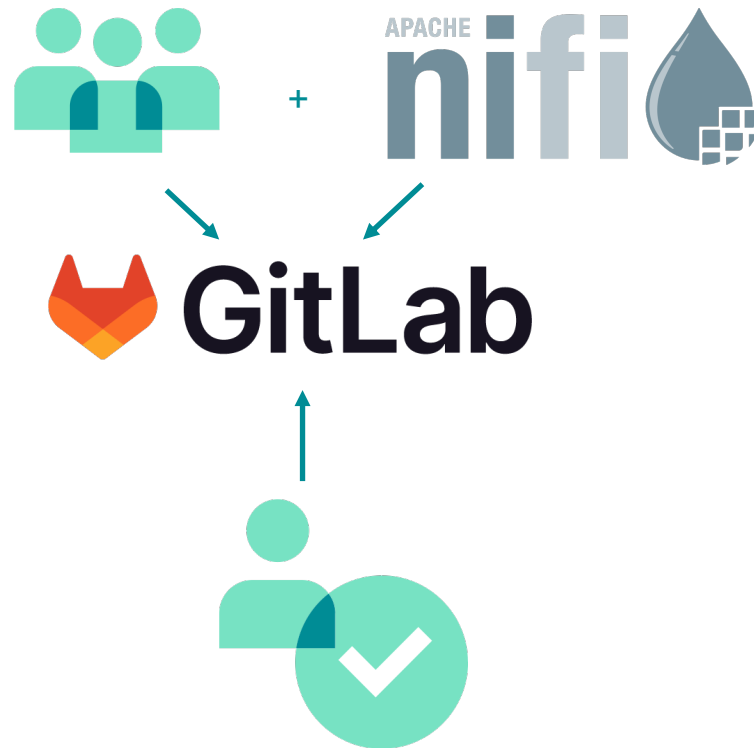
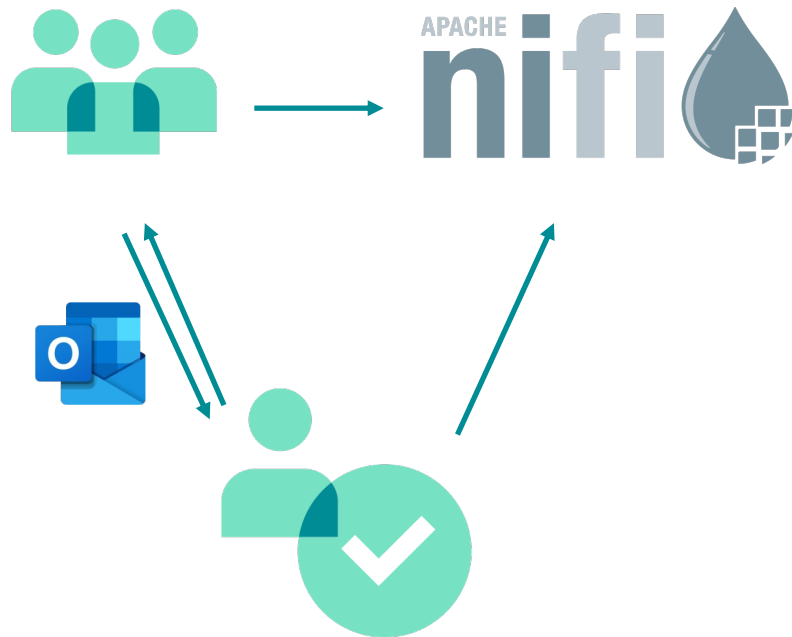


a001-dwh-buildrobot @a001-dwh-buildrobot Jul 27, 2024

Значение DISABLED не прошло автоматическую проверку.

- Попова Клавдия Сергеевна changed this line in [version 2 of the diff](#) Jul 27, 2024

Git repo: было/стало

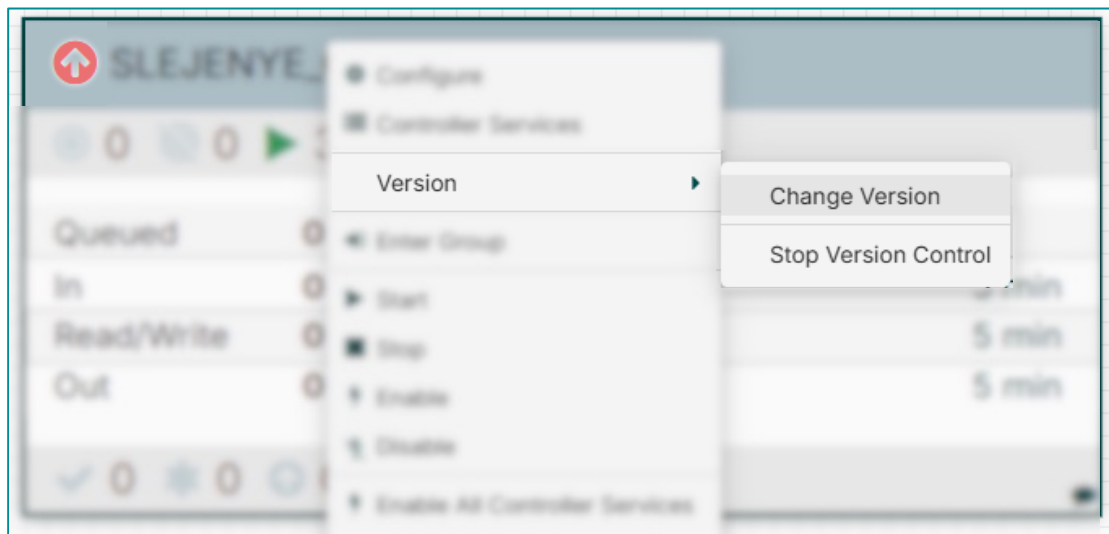


Проблемы ревью решены!

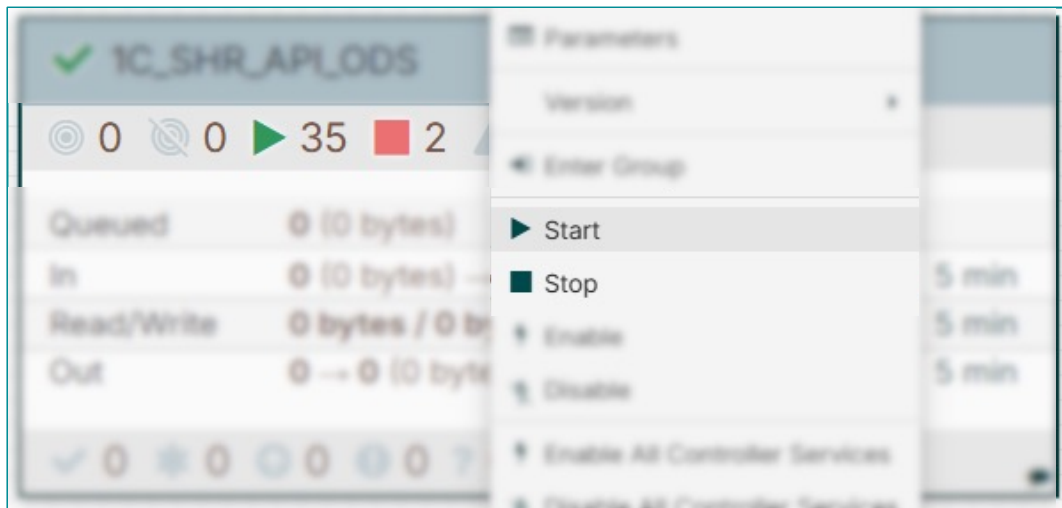
- ✓ Актуальное состояние
- ✓ Ручные действия
- ✓ Визуальное ревью
- ✓ Общение в личных переписках

Автоматизация деплоя

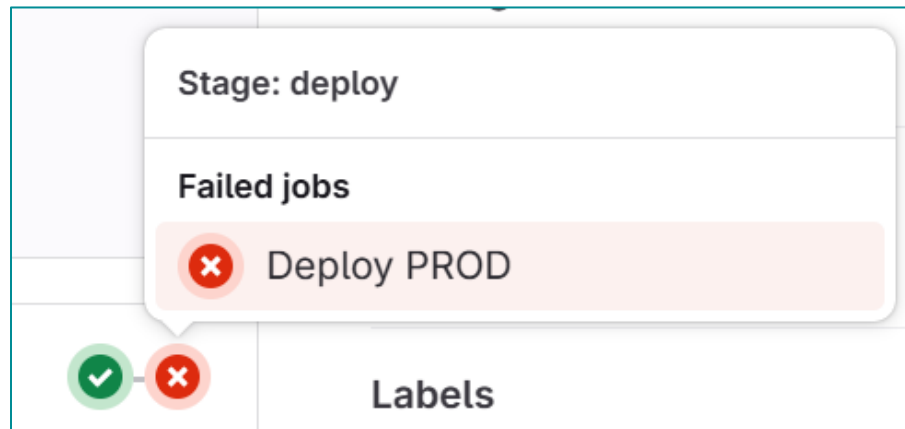
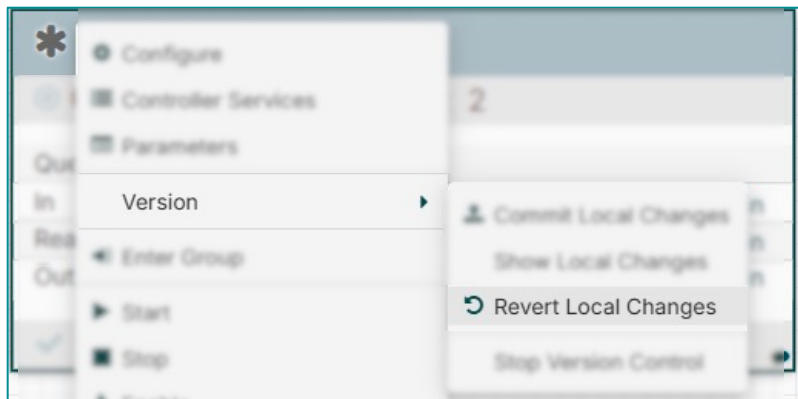
Деплой: поток возможно обновить



Деплой: запустить/остановить поток



Деплой: откат изменений



Ограничения

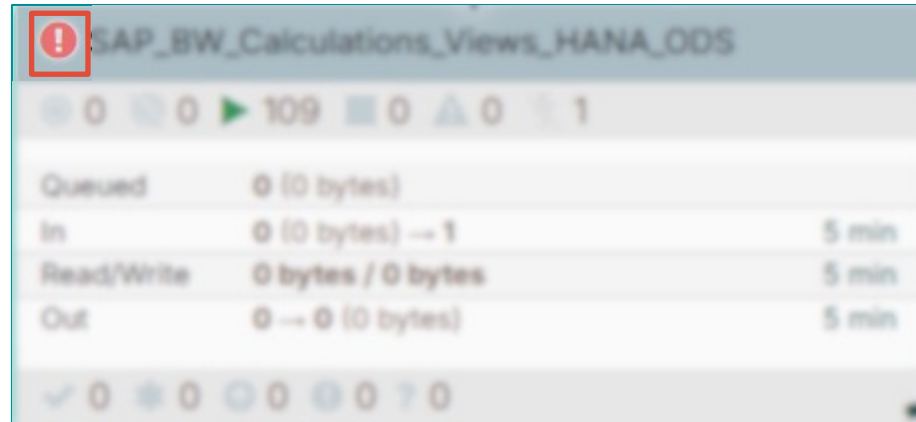
Ограничения: вложенные потоки



Ограничения: долго работающие задачи

✓ IIOT_DATA_KAFKA_ODS			 1
 0  0  8  0  0  0			
Queued	1,152 (773.35 KB)		
In	0 (0 bytes) → 0	5 min	
Read/Write	44.39 MB / 44.44 MB	5 min	
Out	0 → 0 (0 bytes)	5 min	
✓ 0  0  0  0  0 ? 0			

Ограничения: ручные изменения в проде

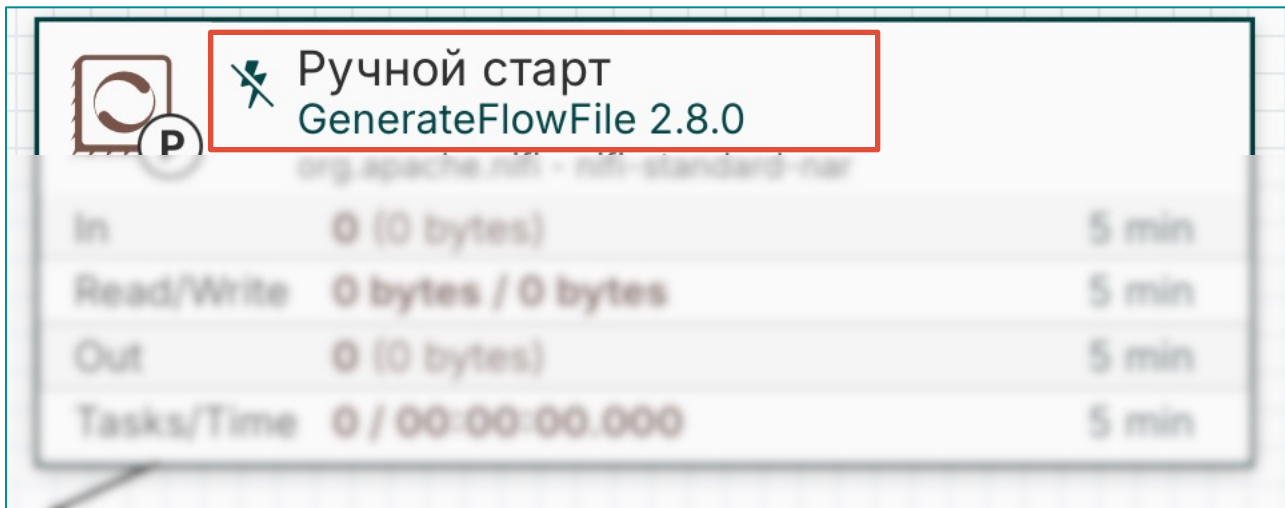


The screenshot shows the SAP BW Calculations Views HANA ODS interface. A red square with a white exclamation mark icon is positioned at the top left of the interface, indicating a warning or error. Below the title bar, there is a status bar with various icons and numbers: a blue circle with 0, a green circle with 0, a green play button with 109, a blue square with 0, a blue triangle with 0, and a red circle with 1. Below this, there is a table with the following data:

Queued	0 (0 bytes)	
In	0 (0 bytes) → 1	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 → 0 (0 bytes)	5 min

At the bottom of the interface, there is another status bar with icons and numbers: a green checkmark with 0, a blue circle with 0, a green circle with 0, a blue circle with 0, and a red circle with 0.

Разрешено менять “ручной старт”



org.apache.nifi - nifi-standard-hat

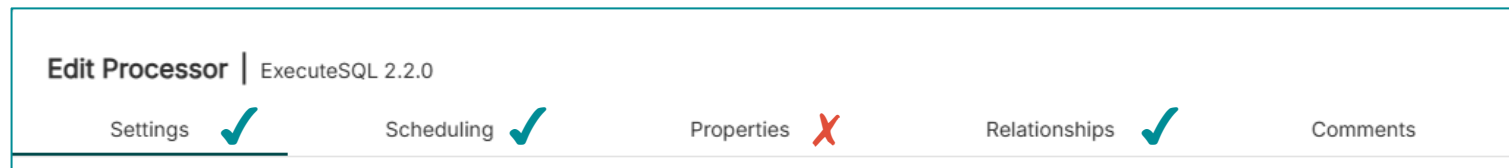
In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Выводы

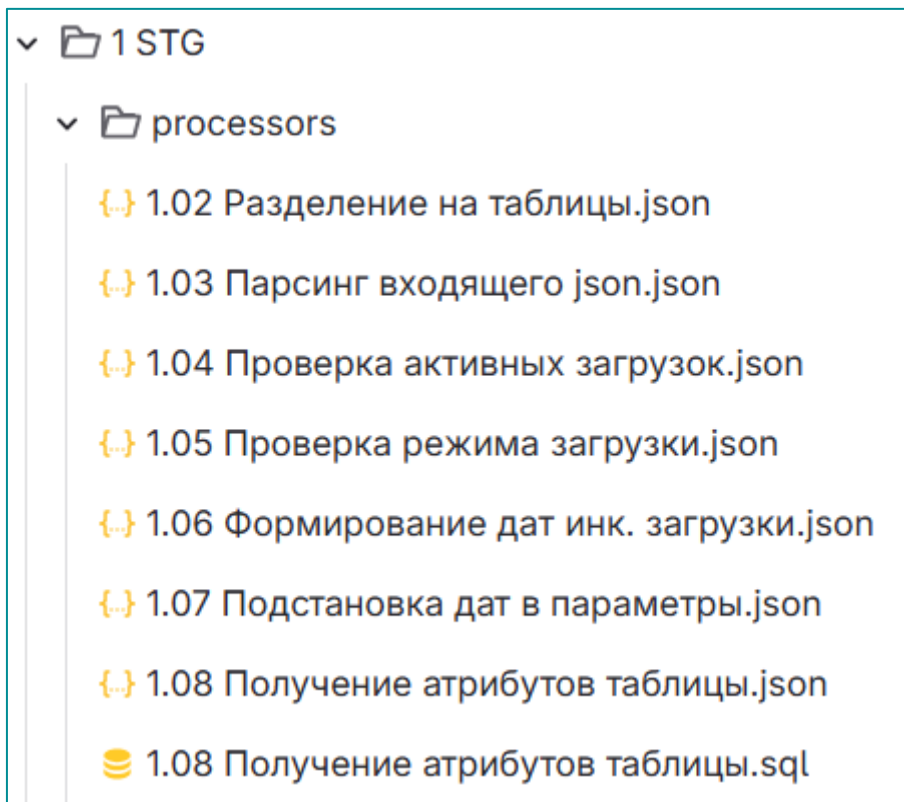
Итог: Registry Client



Итог: автопроверки параметров



Итог: последовательность процессоров



Итог: одна точка взаимодействия



Итог: автоматизация через NiFi API



Вопросы?