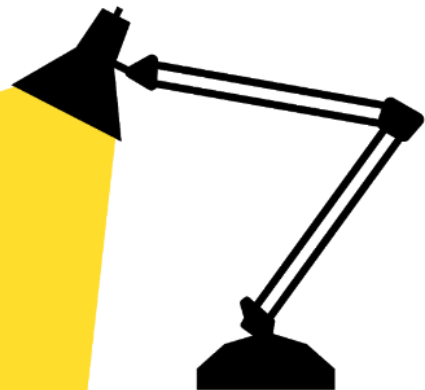


Скрипты как элемент Less-code в NIFI

О себе



Работаю в IT с 2004

В разработке с 2005

Архитектор в Тинькофф

ETL платформой и слабоструктурированные данные

Слабоструктурированные данные



ETL платформа.



О NiFi в Тинькофф

Запустили NiFi в 2018 году

Обрабатываем терабайты сырой информации ежедневно

Создали более 200 бизнес-потоков

Создаем собственные процессоры

Кастомизируем ядро NiFi (Но сегодня не про это)

О чем поговорим

О концепции Less-code

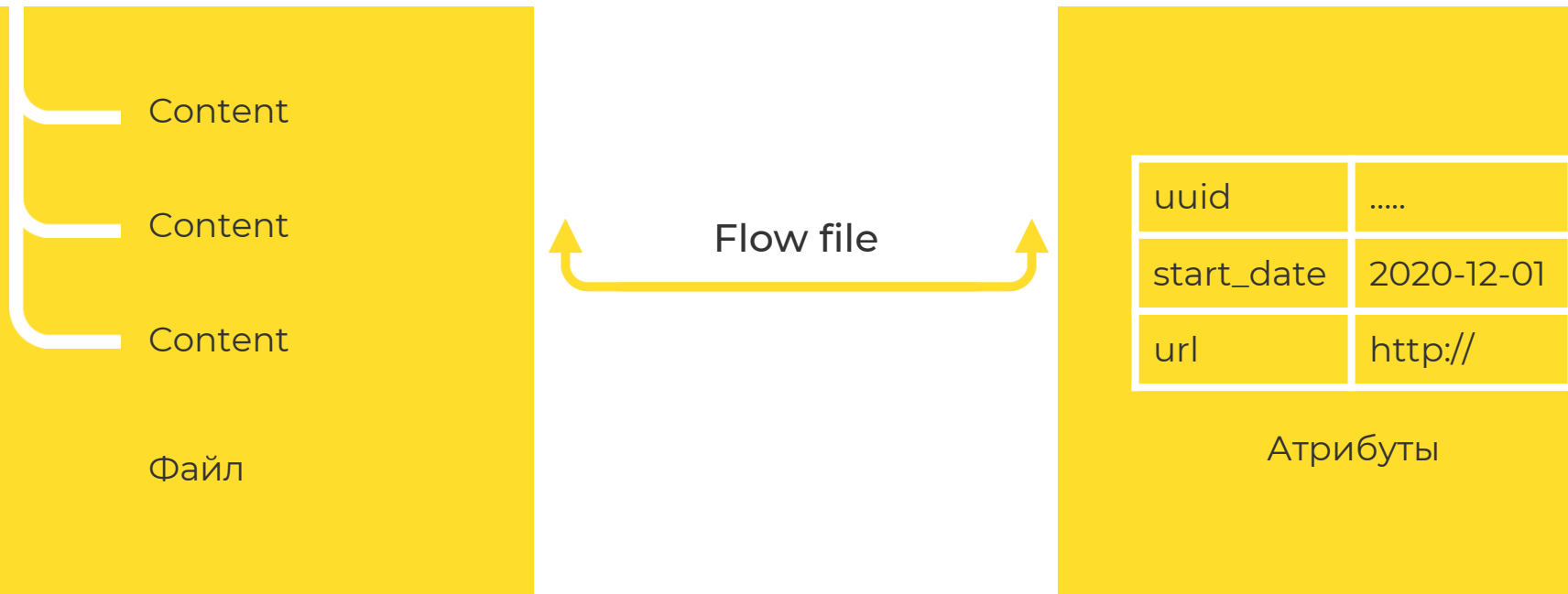
О расширениях возможностей NIFI

О типах скриптов в NIFI

И о том какие минусы есть в скриптах



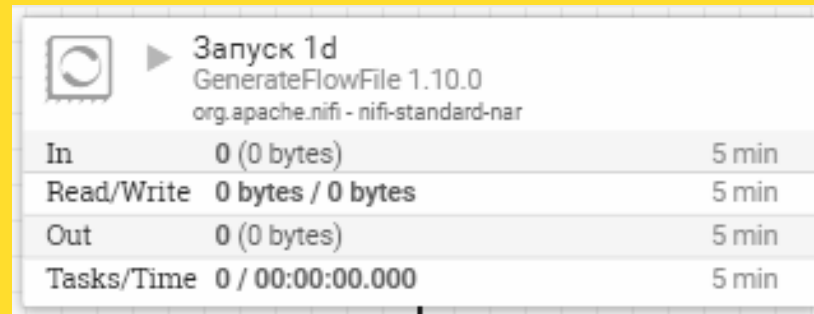
Немного терминологии



Файл - часто будет иметься ввиду FlowFile

Немного терминологии

Процессор - элемент отвечающий за обработку данных



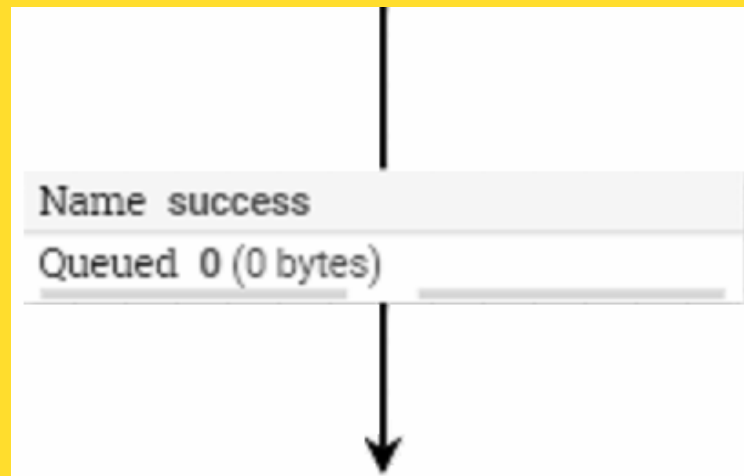
The screenshot shows a configuration window for a NiFi processor. The processor is named 'Запуск 1d' (Run 1d) and is of type 'GenerateFlowFile 1.10.0'. The version is 'org.apache.nifi - nifi-standard-nar'. The window displays performance metrics for the processor over a 5-minute interval.

Metric	Value	Time
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

Немного терминологии

Очередь - элемент связывающий процессоры и хранящий файлы

Отношение – Исходящий поток данных из процессора в очередь



Немного терминологии

Группа процессоров - набор процессоров для решения одной задачи

Частный случай группы процессоров это - Поток (процесс)



The screenshot shows the control panel for a 'NIFI TEST GROUP'. At the top, there are several status icons: a target icon with '0', a refresh icon with '0', a play icon with '4', a square icon with '0', a warning triangle icon with '0', and a scissors icon with '1'. Below these is a table with the following data:

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

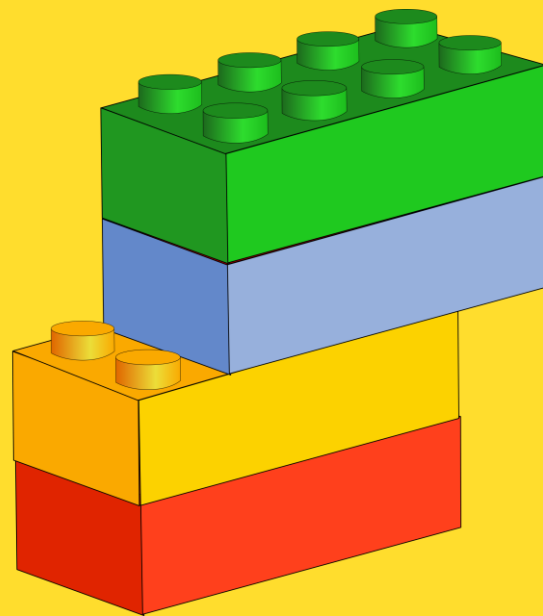
At the bottom of the panel, there are more status icons: a checkmark icon with '0', an asterisk icon with '0', a refresh icon with '0', a warning triangle icon with '0', and a question mark icon with '0'.

Что такое Less-code

Графический интерфейс

Инструментарий

Визуализация алгоритма



Что дает Less-code



Простой онбординг

Визуализация процесса

Контроль за процессом

Чем мешает в Less-code

Урезанные возможности

Производительность

Классические практики
разработки



Почему NIFI Less-code, но не No-code

Визуализация процесса

Богатый инструментарий

Контроль

Собственные модули

Вызов внешнего кода

Скриптовые движки

Как без скриптов

Реализация собственных Processors

- Java framework
- Maven прототипы
- Широкие возможности



NiFi Developers Guide



Доклад про кастомизацию

Как без скриптов

Вызов внешнего приложения

- Комманда OS
- Приложение установленное в OS
- Собственные приложения/скрипты

ExecuteProcess

- Генератор данных
- STDOUT в Content файла
- STDERR в Content файла
- Последовательный вывод

ExecuteStreamCommand

- Трансформация данных
- Content в STDIN
- STDOUT в атрибут
- STDOUT в Content

Что дают скрипты



API NIFI

Без компиляции

В интерфейсе NIFI

Сконцентрируемся на скриптах



Почему Groovy

Функционал

Копирование кода

Легкое подключение библиотек

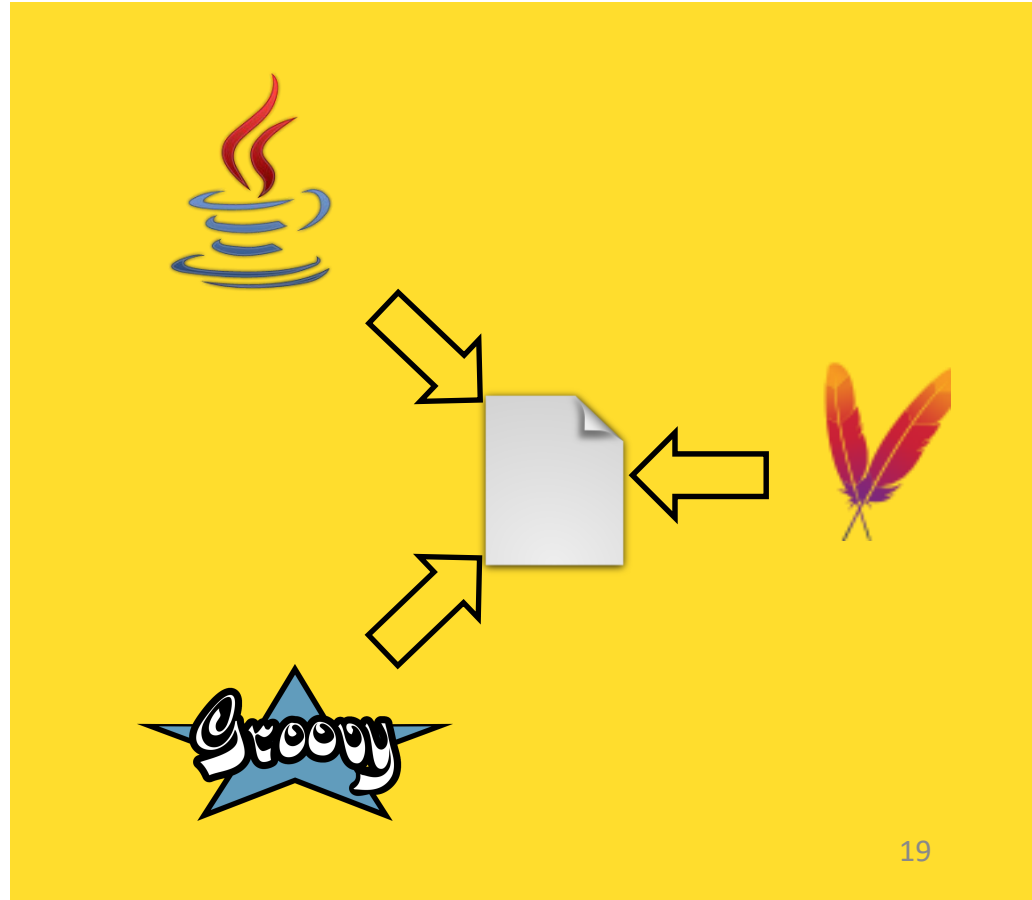


Импорт модулей

jar файлы

Grape loader

GroovyClassLoader



Варианты использования скриптов

A thick yellow vertical line on the left side of the slide, with three horizontal bars extending to the right, acting as a bracket for the list items.

Работа с Record

Обогащение данных

Универсальные трансформации

Record процессоры

Файл как набор записей

Множество стандартных трансформаций

QueryRecord

LookupRecord

JoltTransformRecord

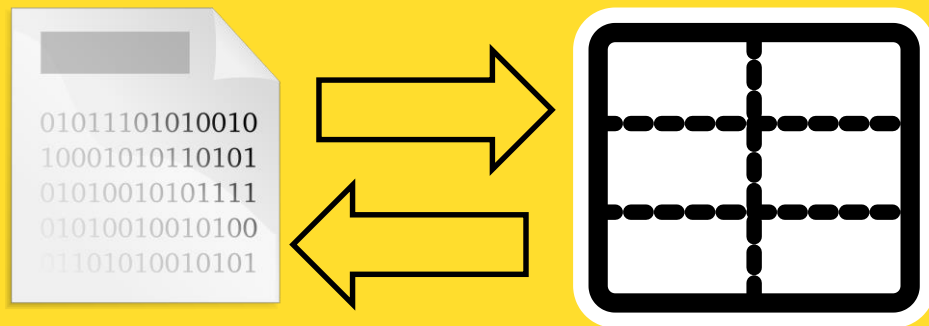
ConvertRecord

UpdateRecord

Работа с Record через скрипты

ScriptedReader

ScriptedRecordSetWriter



Record процессоры и Скрипты

ScriptedFilterRecord - фильтрация

ScriptedValidateRecord – проверка по условию

ScriptedPartitionRecord - группировка

ScriptedTransformRecord - преобразование

Обогащение данных

Lookup

- Обращение к внешнему источнику для обогащения данных

ScriptedLookupService

- Получает на вход Map с данными
- Возвращает объект
- Может использоваться для LookupRecord

SimpleScriptedLookupService

- Получает на вход Map с данными
- Возвращает строку
- Может использоваться для LookupRecord и LookupAttribute

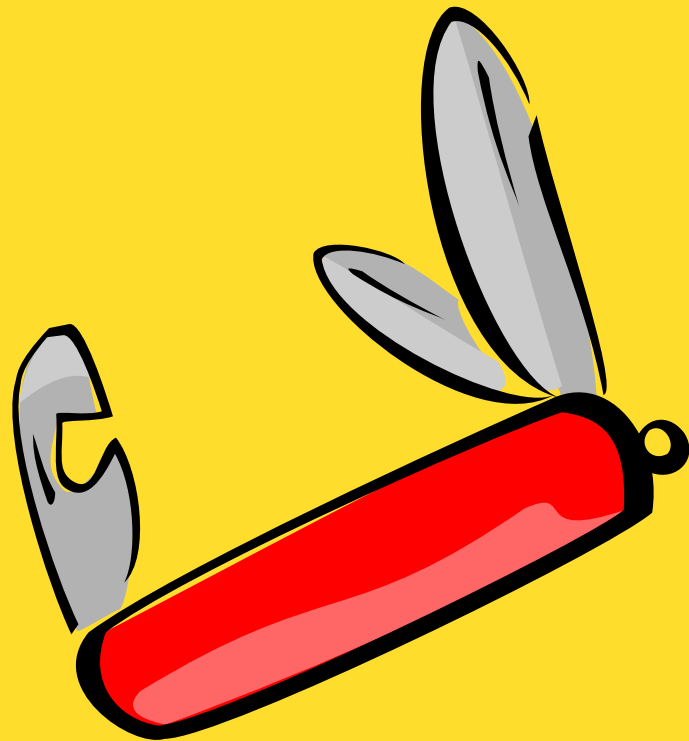
Универсальные скрипты

ExecuteScript/ExecuteGroovyScript

Любые операции

Больше кода

Больше ответственности



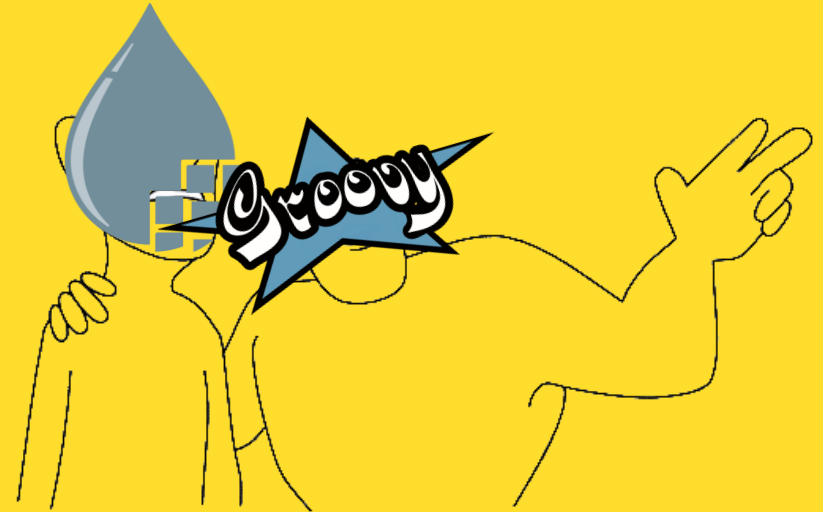
Почему бы не использовать всегда?

Уровень входа

Сопровождаемость

Производительность

Безопасность



Если все-таки скрипт

Review скриптов

Стандартизация

Доп. модули

1.

2.

.....

n. profit

NiFi Mockscript



NiFi Scripts

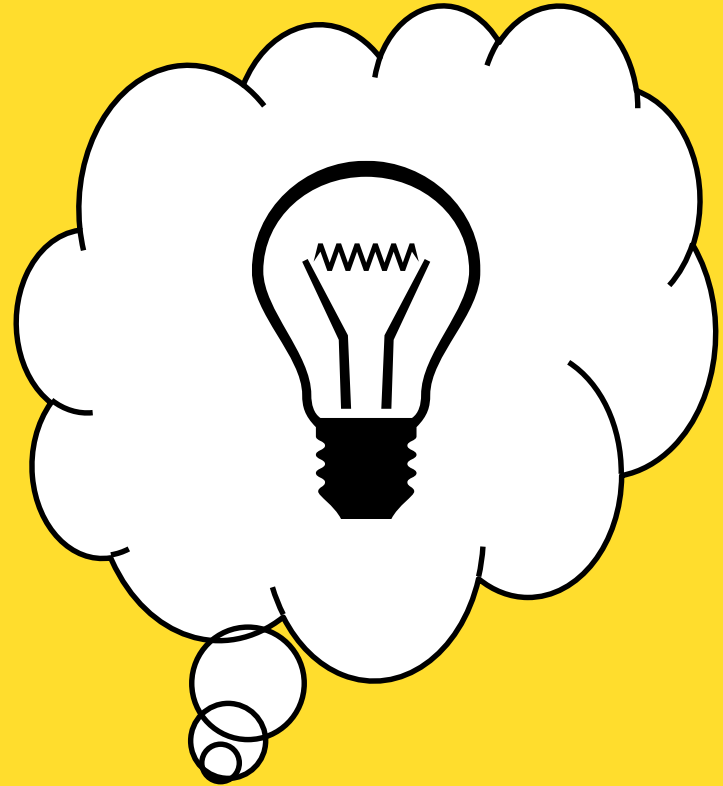


Преимущества ScriptedProcessor

Реализовать Start/Stop

Реализовать Relations

Переход к разработке



Итоги



NiFi может вызывать внешний код

NiFi может исполнять Scripts внутри

Скрипты медленнее

Скрипты опаснее

Но они могут помочь

Ваши вопросы

Можно задать сейчас

Или позже в группе



nifusers

