

Т П И П И

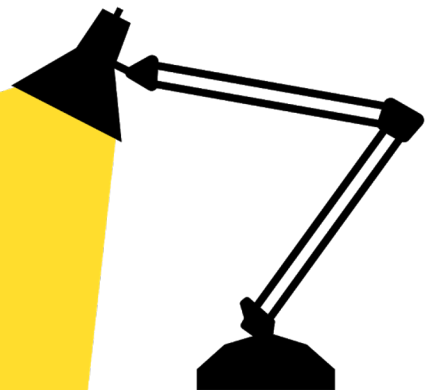
NiFi. Иницирующая загрузка

Бронислав

12.2020

К О Е Е

О себе



Работаю в IT с 2004

В разработке с 2005

Архитектор в Тинькофф

Платформа извлечения сырых данных

Платформа извлечения сырых данных



платформа извлечения сырых данных



О NiFi в Тинькофф

Запустили NiFi в 2018 году

Обрабатываем терабайты сырой информации ежедневно

Создали более 150 бизнес-потоков

Создаем собственные процессоры

Кастомизируем ядро NiFi

О чем поговорим

О NiFi

Об иницилирующей загрузке

И о том как упростить себе жизнь



Немного терминологии

Content

Content

Content

Файл

Flow file

uuid

.....

start_date

2020-12-01

url

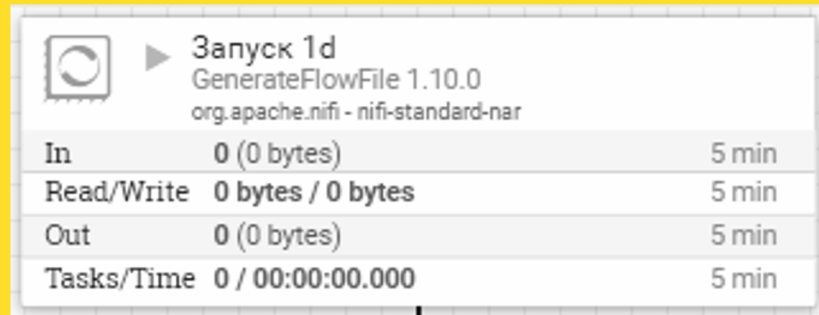
http://

Атрибуты

Файл - часто будет иметься ввиду FlowFile

Немного терминологии

Процессор - элемент отвечающий за обработку данных



The screenshot shows a processor status window for 'Запуск 1d' (Run 1d). The processor is 'GenerateFlowFile 1.10.0' from the 'org.apache.nifi - nifi-standard-nar' namespace. The status table shows 0 bytes of input and output, and 0 tasks completed over a 5-minute period.

Запуск 1d		
GenerateFlowFile 1.10.0		
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

Немного терминологии

Группа процессоров - набор процессоров для решения одной задачи

Частный случай группы процессоров это - Поток (процесс)



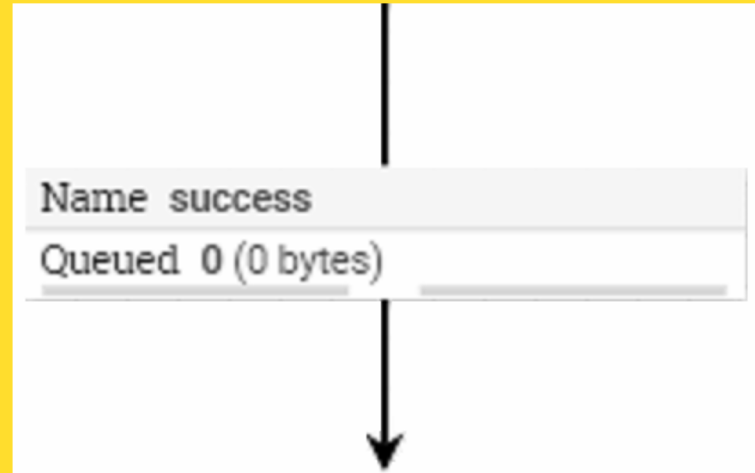
The screenshot shows the control panel for a 'NIFI TEST GROUP'. At the top, there are several status icons: a target icon with '0', a refresh icon with '0', a play icon with '4', a square icon with '0', a warning triangle icon with '0', and a scissors icon with '1'. Below these icons is a table with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

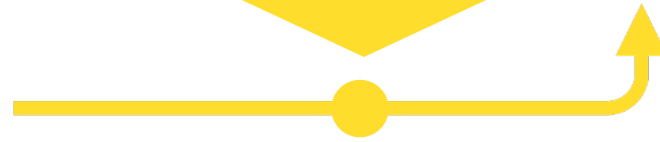
At the bottom of the panel, there are more status icons: a checkmark icon with '0', an asterisk icon with '0', a circular arrow icon with '0', an exclamation mark icon with '0', and a question mark icon with '0'.

Немного терминологии

Очередь - элемент связывающий процессоры и хранящий файлы



Что такое NiFi



2014 год



Что такое NiFi

Концепция FlowFiles

Управление у файлов а не у обработчиков.



Что такое NiFi



Что такое NiFi

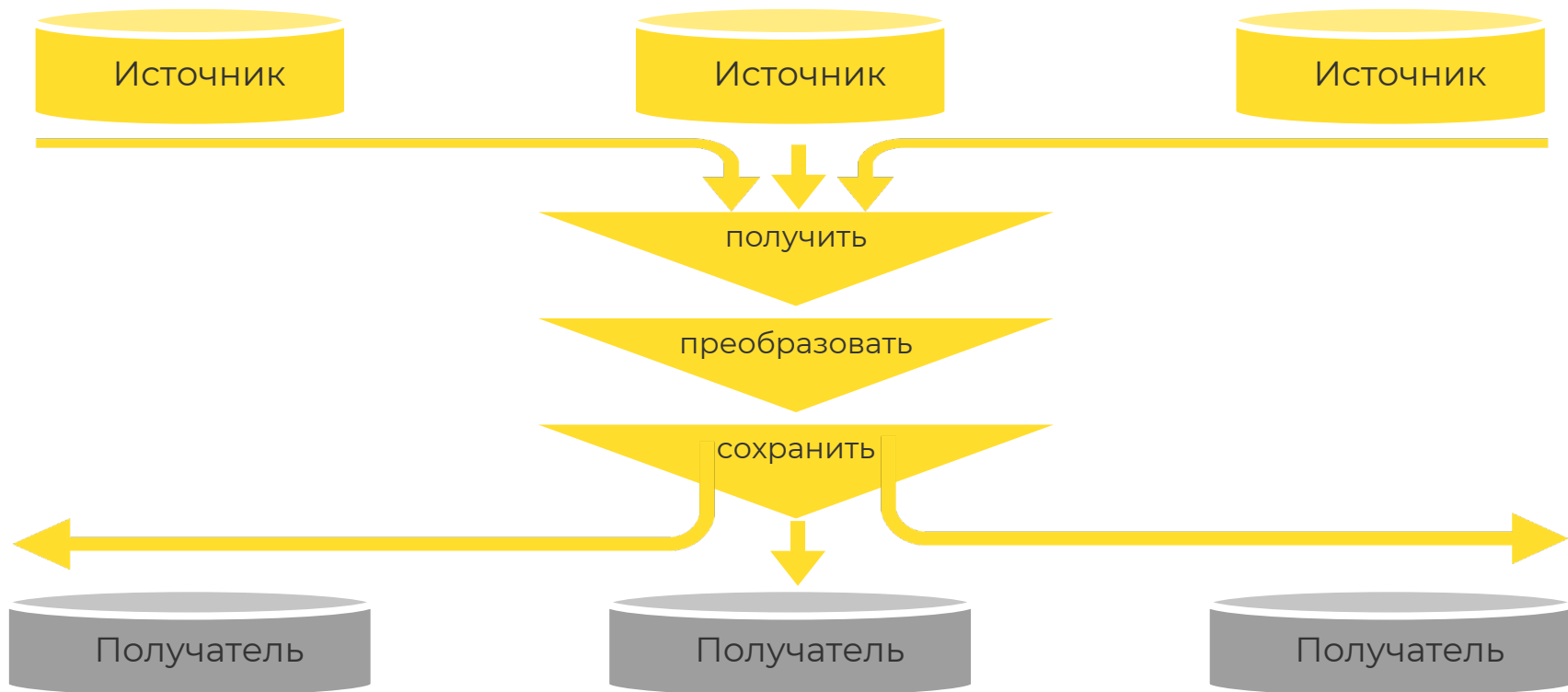


UI



Speed

Поток данных в NiFi



Почему NiFi?

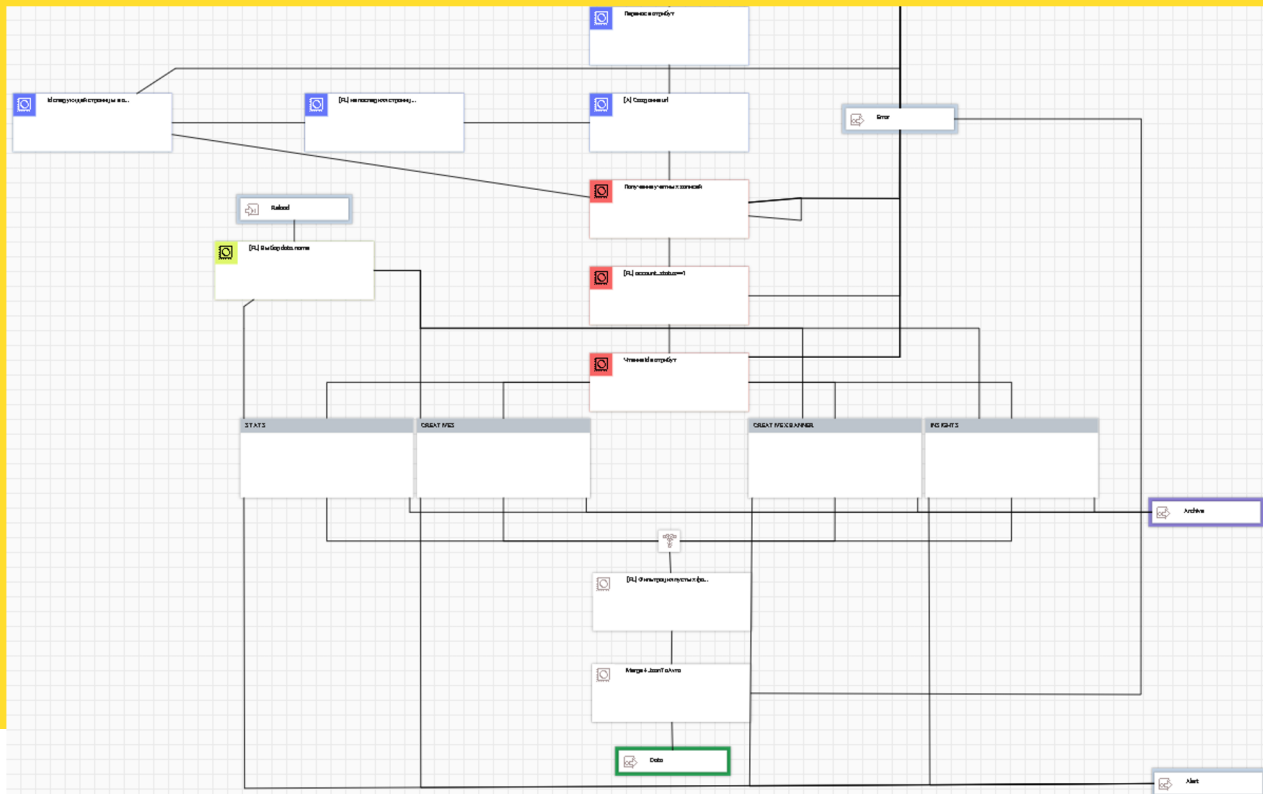
UI

Codeless

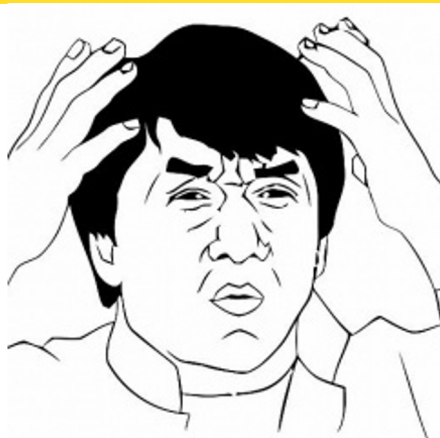
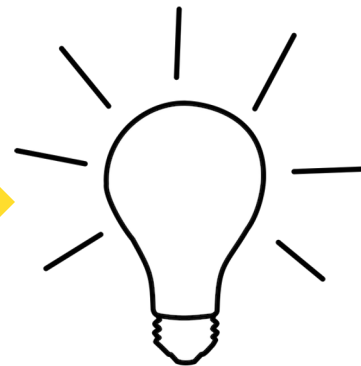


OpenSource

Почему NiFi?



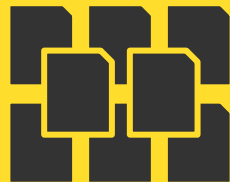
Иницилирующая загрузка



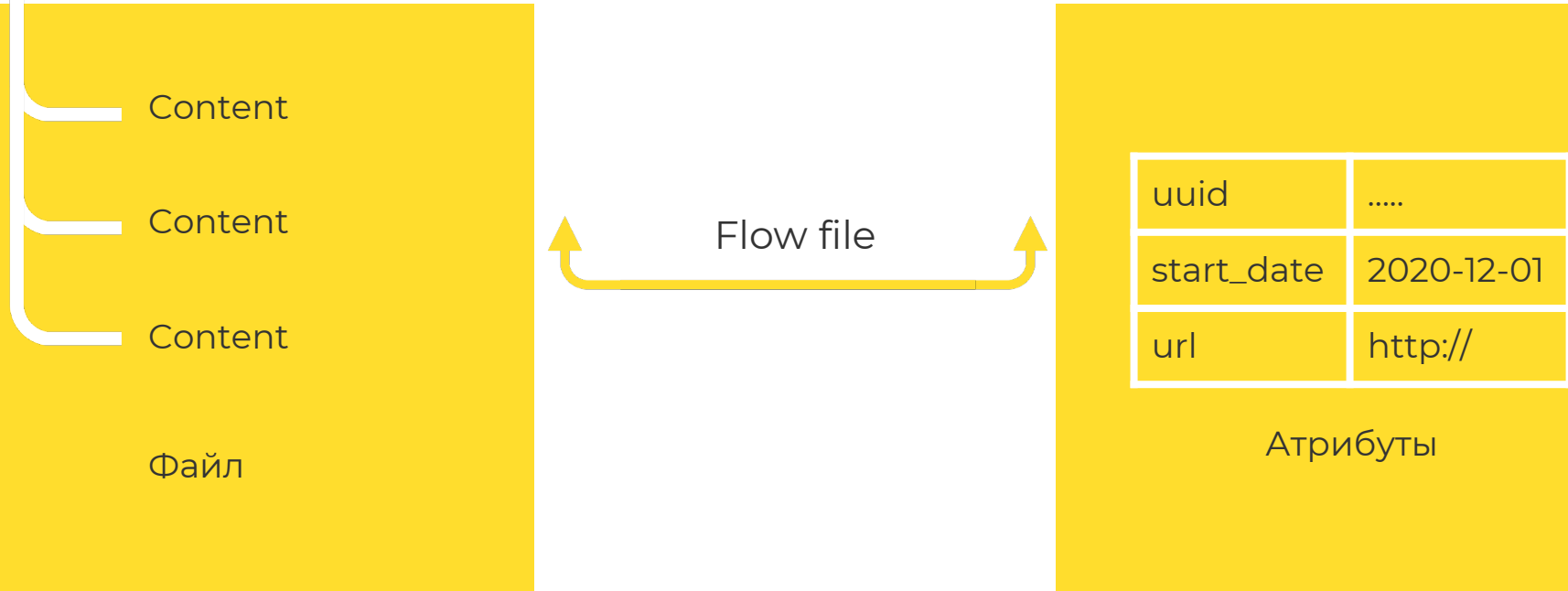
Регламентная загрузка



Иницилирующая загрузка



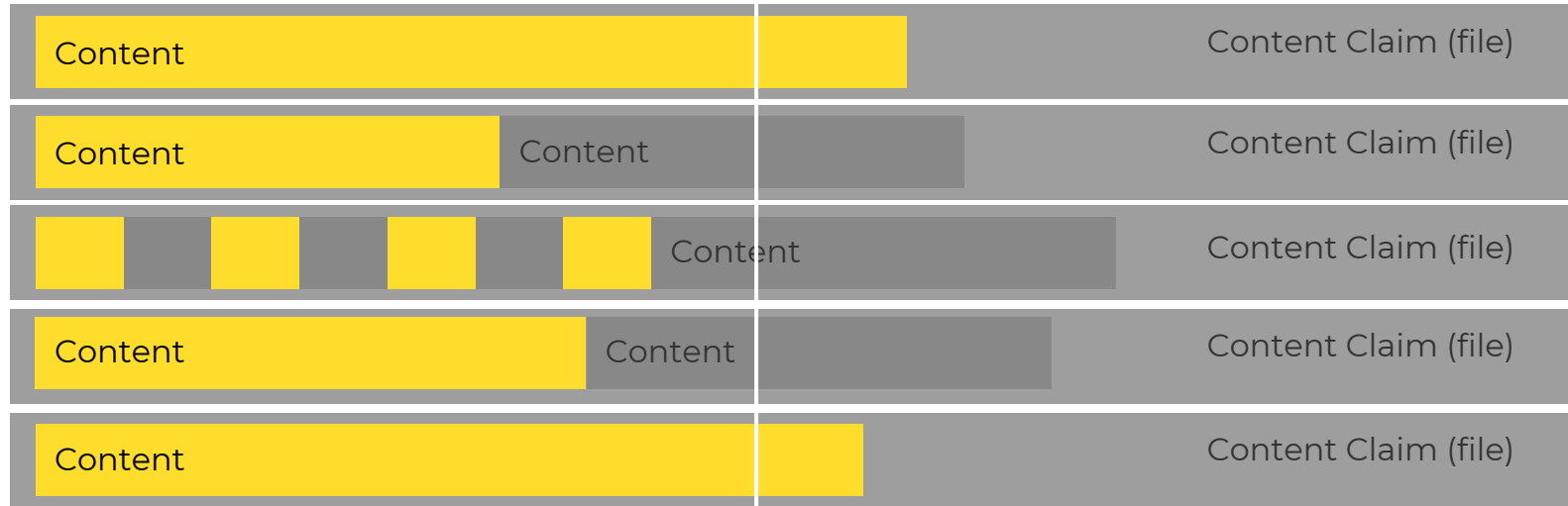
Структура хранения



Контент - ссылка на файл и на позицию в файле в которой размещен контент

Структура хранения

Content Repository



Граница

Совет 1

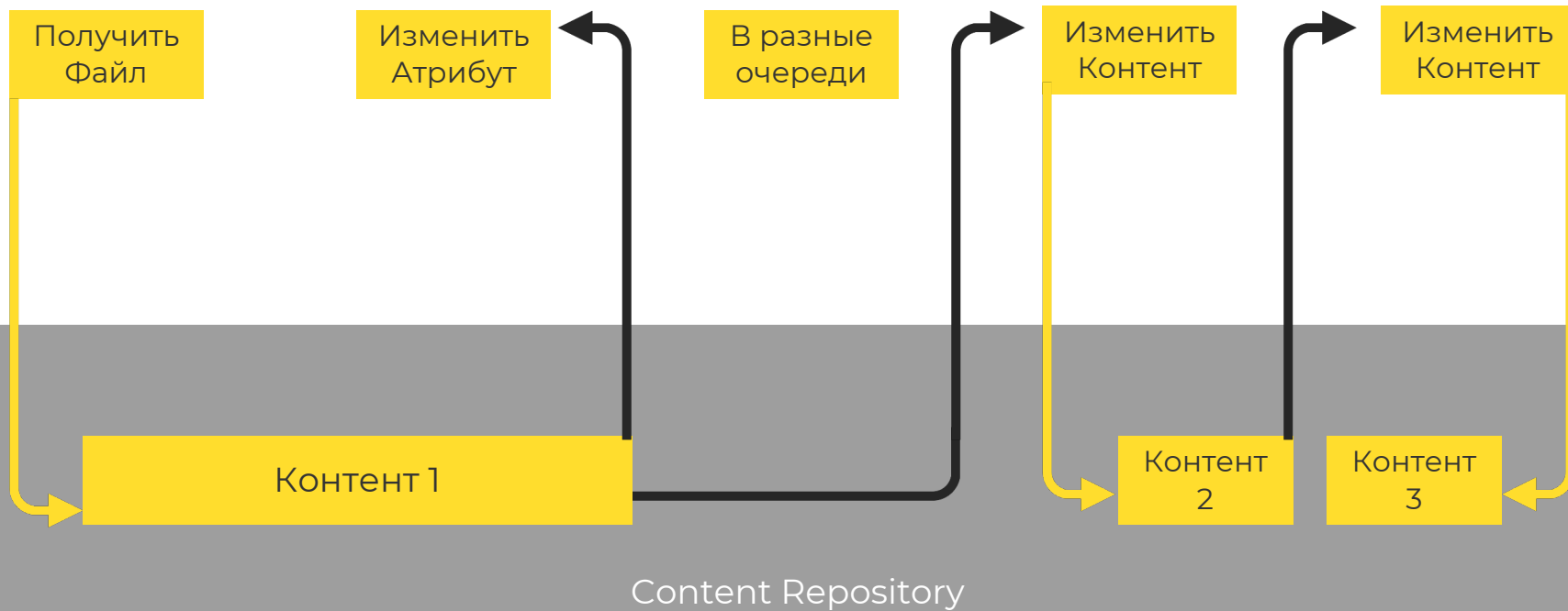


Храните только нужный контент

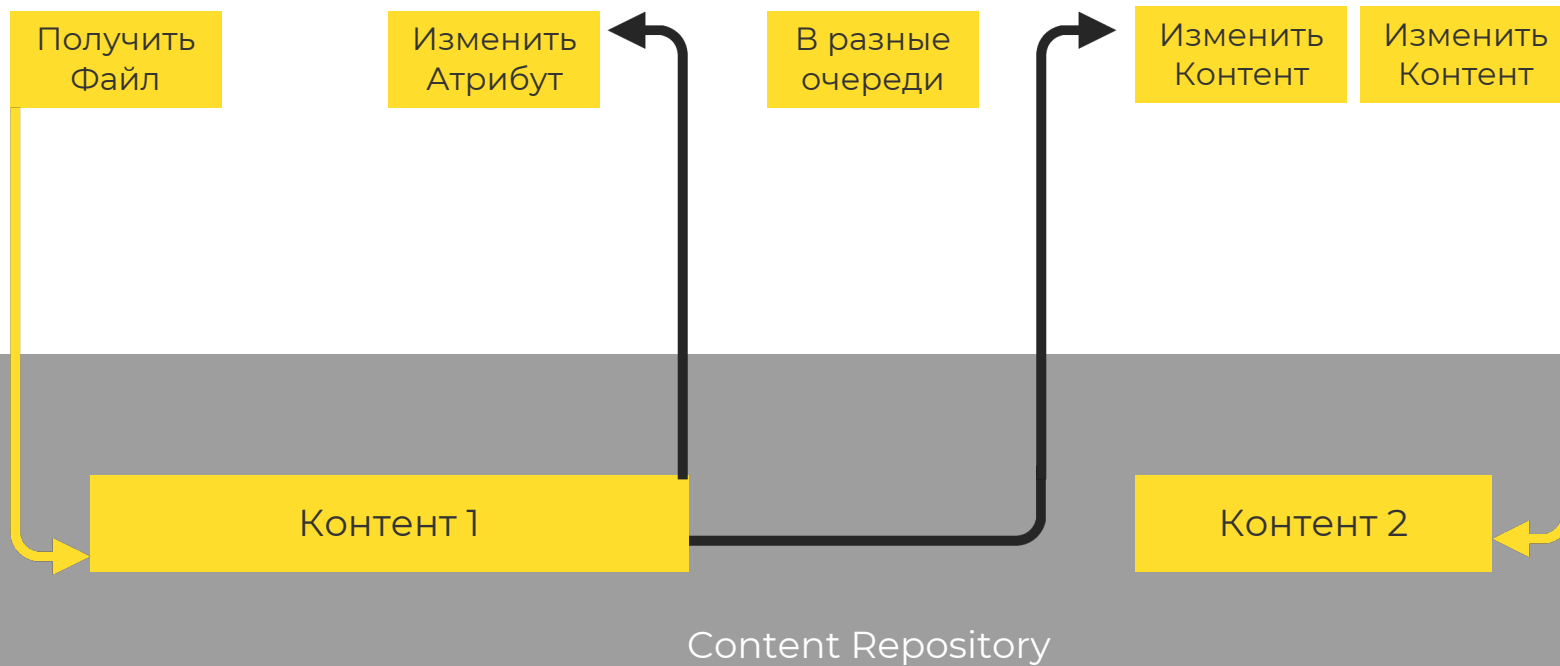


Обработывайте контент как можно быстрее

Работа с записью



Работа с записью



Совет 2



Минимизируйте изменение контента если есть такая возможность

Среда исполнения



Поток 1

Поток 2

Поток 3



Возможности
среды

Иницирующий поток

Среда исполнения



Среда исполнения



Поток 1

Поток 2

Поток 3



Среда для
регламентных потоков



Иницирующий поток



Среда для
иницирующих потоков

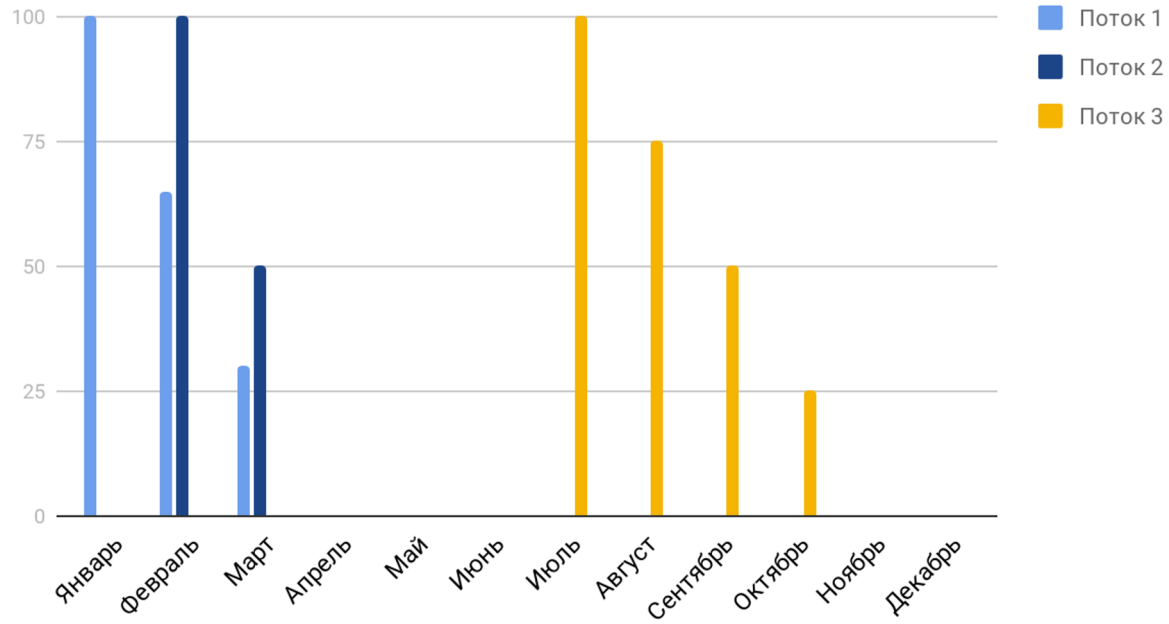
Совет 3



Разделите среды для исполнения

Виртуальные среды

Использование ресурсов



Совет 4



Используйте виртуальные среды для иницилирующей загрузки

Все потоки одинаковы



Ctrl+C
Ctrl+V



Совет 5



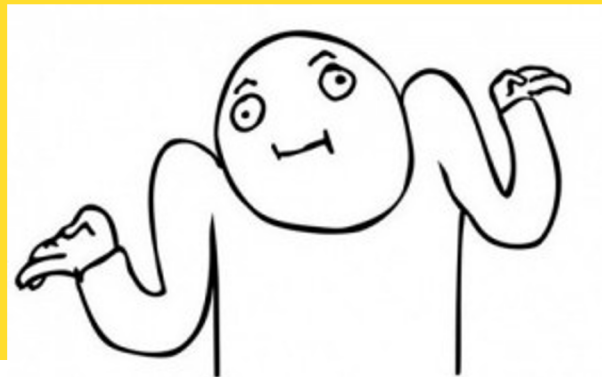
Стандартизируйте поток, это упростит создание инициирующего потока

И об ошибках

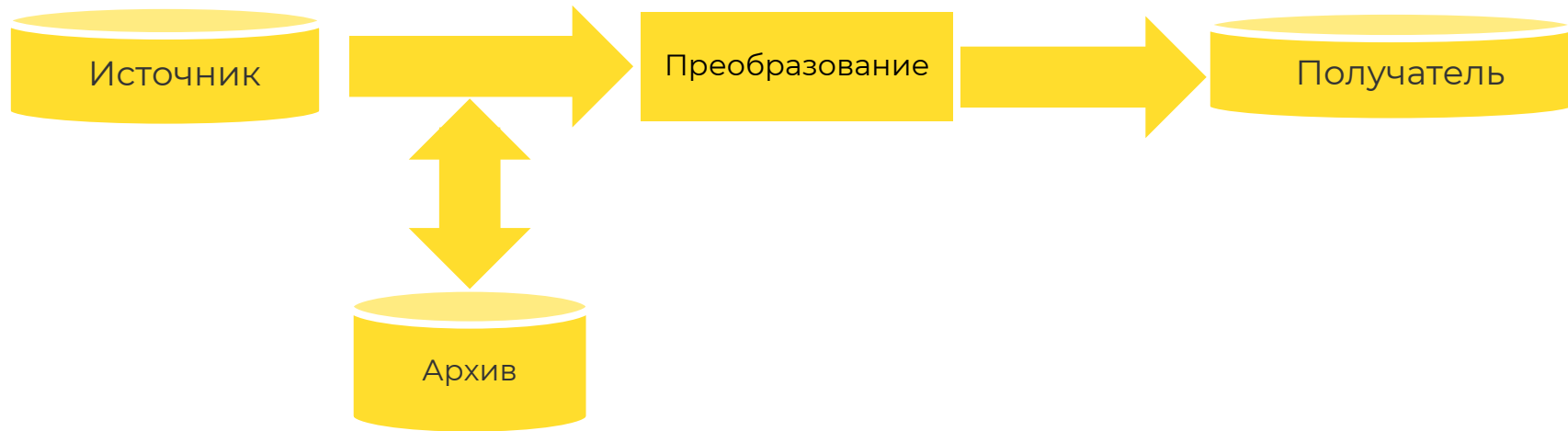
ошибки в обработке

ошибки в модели

изменения требований



Концепция резервирования потока



Концепция резервирования потока



Перегрузка потребует времени

Концепция резервирования потока



Хранение требует ресурсов

Итоги



- Храните только нужный контент
- Обрабатывайте контент быстрее
- Минимизируйте изменение контента
- Разделите среды исполнения
- Используйте виртуальные среды
- Стандартизируйте потоки
- Резервируйте данные источников

Ваши вопросы

Можно задать сейчас

Или позже в группе



nifusers

