



Стриминговая загрузка данных из Kafka

По результатам данного задания вы получите навык работы со стриминговыми данными в NiFi с использованием распределённого программного брокера сообщений Apache Kafka

Инструменты, которые пригодятся для выполнения:

Apache NiFi, база данных Redis, брокер сообщений Apache Kafka, опционально: MS Excel.

Описание задания

В одном из тестовых заданий вы уже залили данные из файла в брокер Kafka. Теперь предстоит построить поток, который заберёт эти данные и отфильтрует только те, которые за XX/YYYY дату (выберите любой диапазон, который не пересекается с вашими коллегами-студентами). Эти данные надо обогатить описанием по коду из Redis и выгрузить в csv файл, который можно открыть в MS Excel.

После того, как записи пришедшие из Kafka, будут обработаны и обогащены, запишите всё в один файл с использованием скрипта, как описано тут: [nifi-scripts/Files/flowfile_to_file_append_at_main · Shadilan/nifi-scripts \(github.com\)](https://github.com/Shadilan/nifi-scripts/blob/main/Files/flowfile_to_file_append_at_main)

Обратите внимание:

1. На сайте Ni-Fi для скачивания доступны две версии сервиса: 1.9.X и 2.0.0-M2. На данный момент, используйте, пожалуйста, исключительно 1.9.XX.
Доступный для скачивания 2.0.0-M4 является версией для ознакомления (для энтузиастов) Milestone — это что-то вроде pre alpha.
2. Как подтянуть записи из Redis: сейчас там хранится n-ое количество записей, каждая под своим уникальным ключом. В качестве ключа используйте составной ключ год+месяц.

Для проверки

Для проверки задания, выгрузите группу с этим потоком в темплейт. Получившийся в результате CSV файл тоже. Оба файла разместите в своём репозитории на GitLab и предоставьте ссылку для проверки.