Андреев Александр 6233

По шагам из репозитория загрузили и запустили все необходимые инструменты.

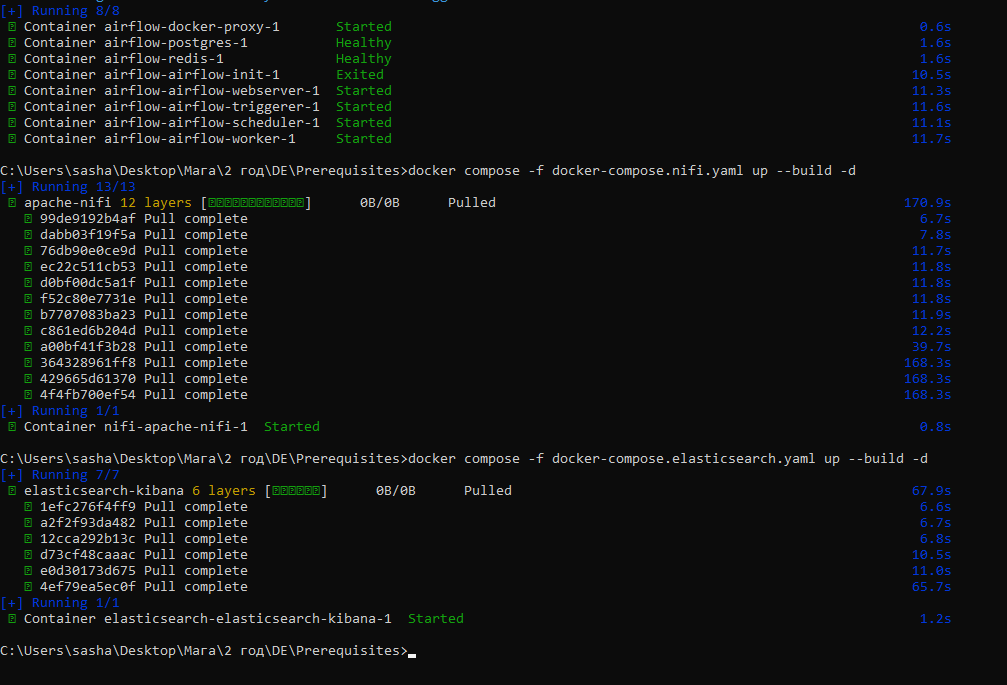


Рисунок 1 – Запуск сервисов

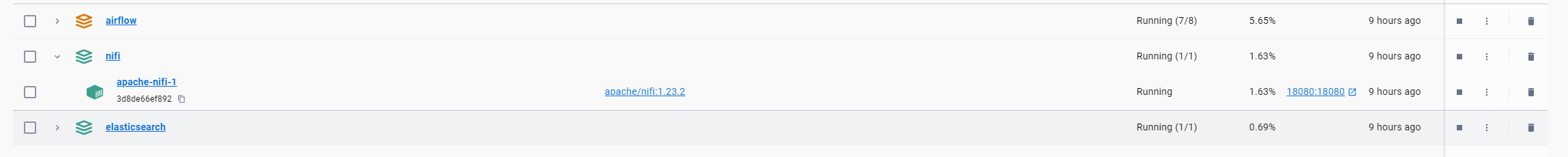


Рисунок 2 – Запущенные сервисы

Далее в VS Code был написан код на Python для обработки датасета и отправки его в ElasticSearch.

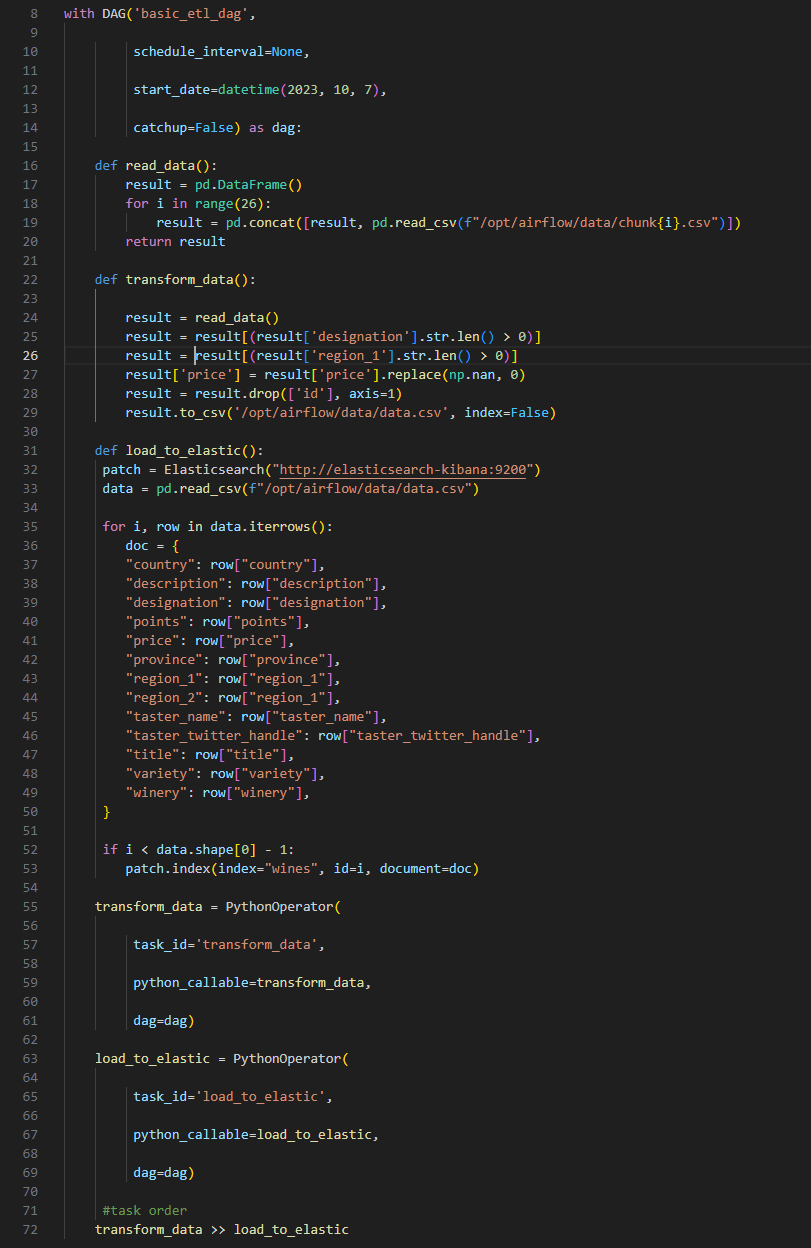


Рисунок 3 – Код Apache Airflow

С этой часть проблем не возникло.

В NiFi получилась следующая модель.

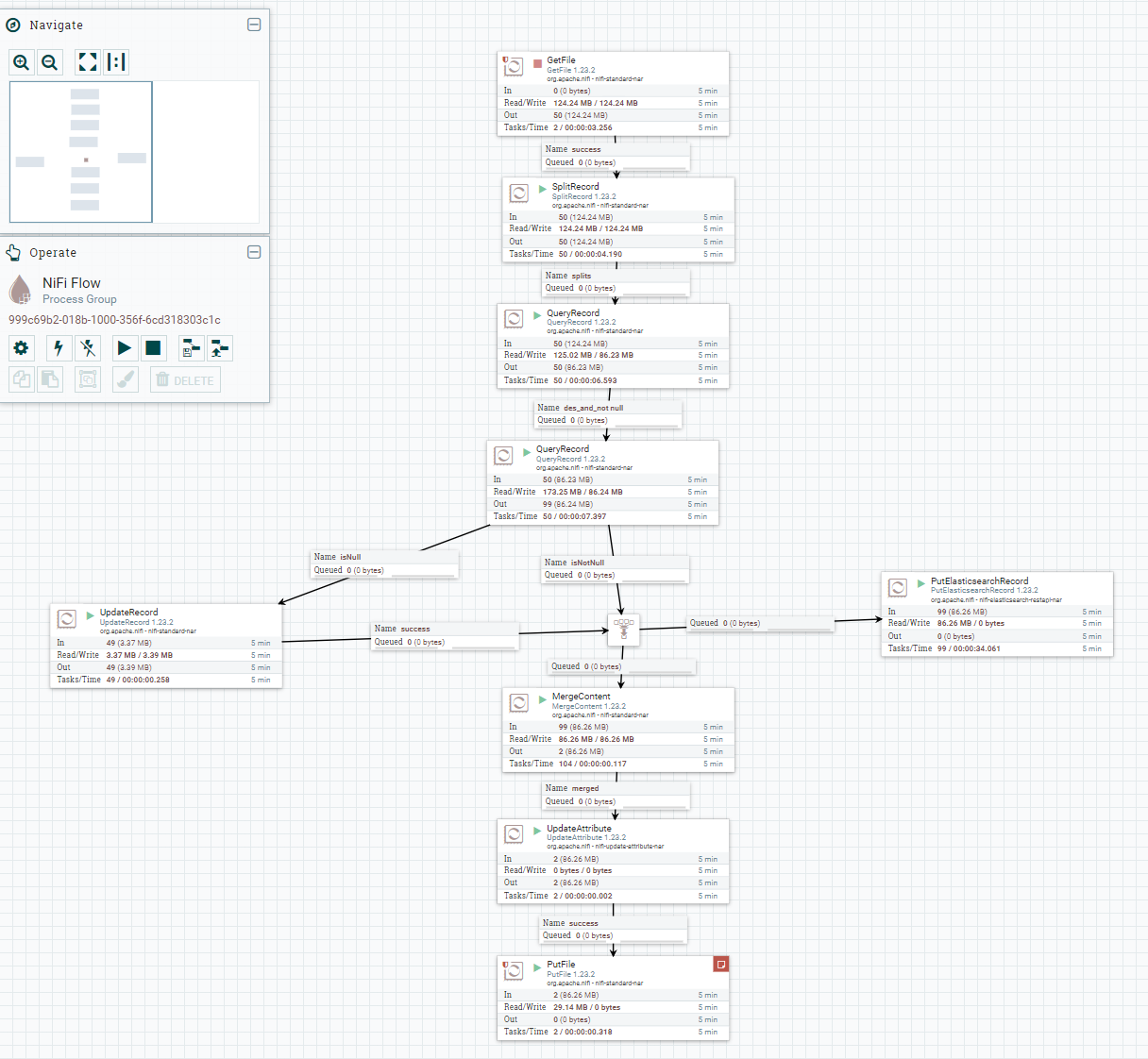


Рисунок 4 – Модель в NiFi

Тут первой проблемой стала обработка price, ибо непонятно было, как обрабатывать пустую строку. В итоге решилось все так:

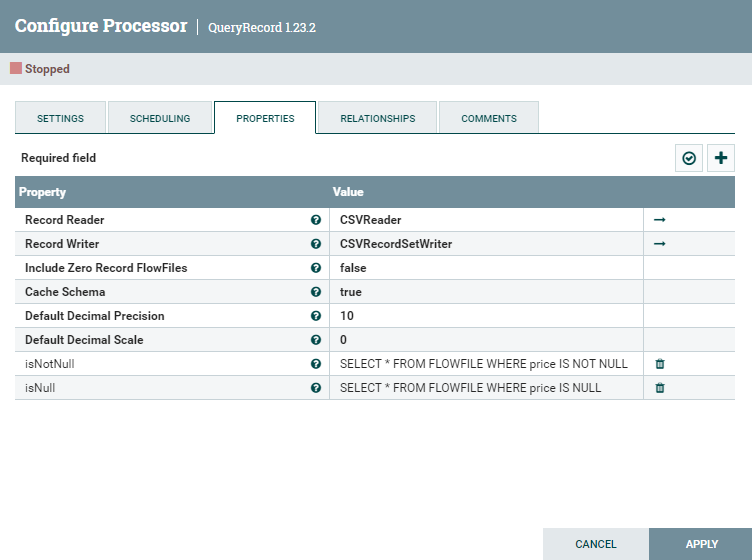


Рисунок 5 – Настройки QueryRecord

Еще одной проблемой стало то, что MergeContent мержит с заголовками. После нескольких часов поиска у меня не получилось найти, как можно пропускать первую строчку. Была идея сделать через RouteOnAttribute, но я так и не понял почему он не находит строки по такому условию ${attribute:contains('id'):not()}

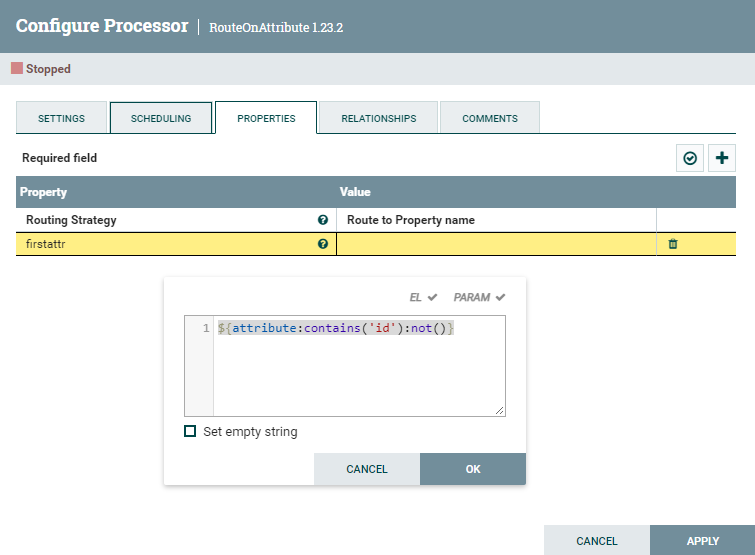


Рисунок 6 – Попытка убрать хедер

Однако несмотря на это отобразить графики получилось корректно.

Про настройки каждого процессора не вижу смысла говорить, там все стандартно и проблем с остальным не возникло.

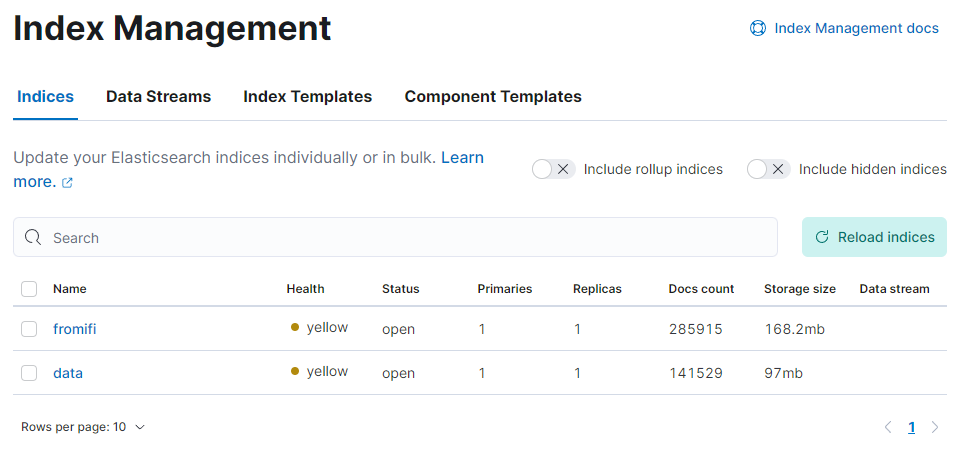


Рисунок 7 – Данные в Elastic

В Kibana создали паттерн и построили гистограмму.

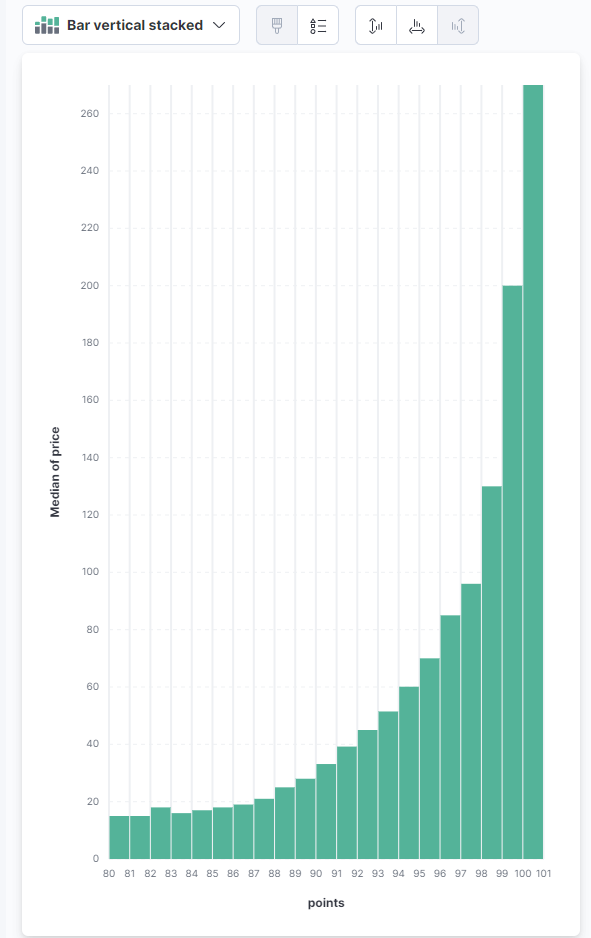


Рисунок 8 – Итоговая гистограмма

От себя пару слов: Apache NiFi ужасно неудобная программа. Я лично не понял зачем ей пользоваться, если для работы и обработки датасетов есть Python с более или менее хорошей документацией. С Apache Airflow работать оказалось куда приятнее.

Также в обоих случаях корректно все данные корректно сохранились компьютер.

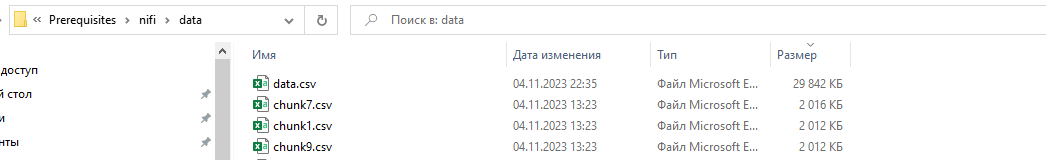


Рисунок 9 – Результирующие данные на компьютере