

Отчет по лабораторной работе №1

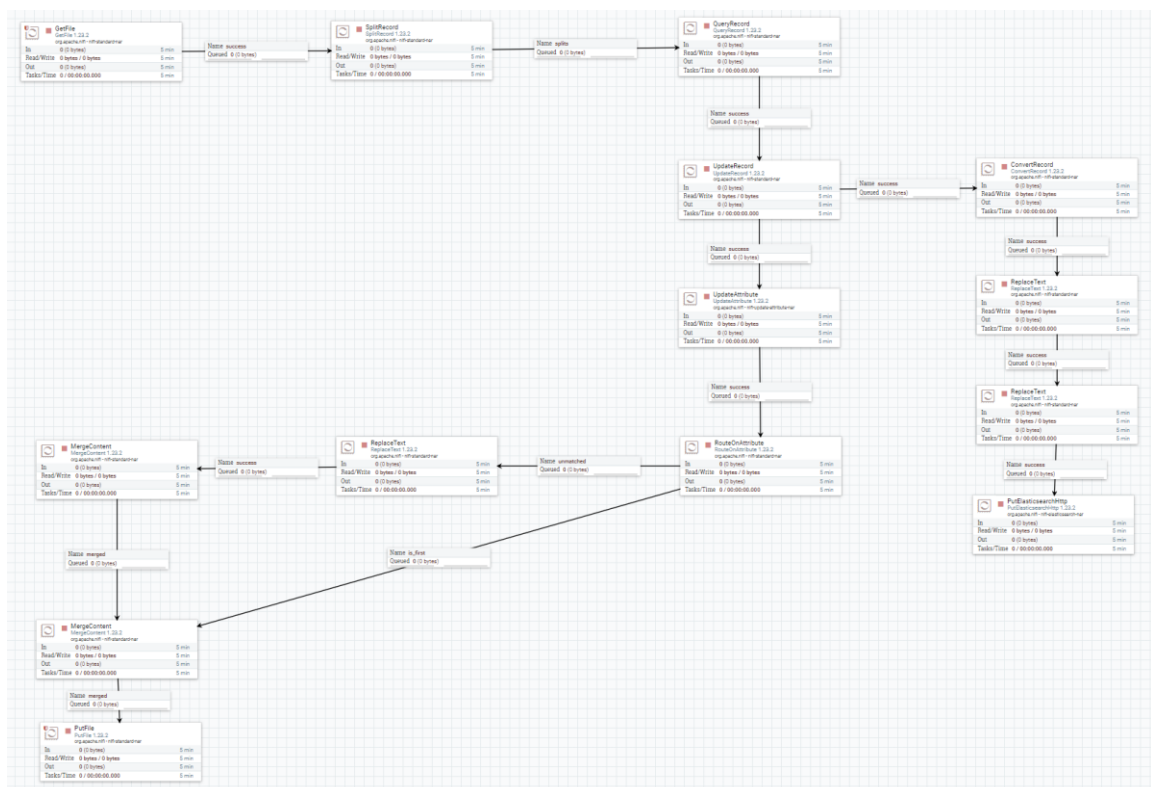
Выполнил: Казаркин Тимофей, гр. 6232

Apache NiFi

Во время выполнения задания в Apache NiFi я столкнулся с рядом проблем:

- Несмотря на интуитивно понятное визуальное программирование, документация нод оставляет желать лучшего;
- Небольшое количество информации в интернете приводит к тому, что на некоторые вещи, полезные при работе, натыкаешься случайно в процессе бесконечных попыток запустить обработку правильно;
- Оказалось, что нода MergeContent не умеет совмещать много документов сразу, из-за чего пришлось делить поток по 1000 элементов и потом их ещё раз совмещать.

Схема нод в NiFi:



Apache Airflow

Здесь уже было интереснее, т.к. все задачи можно писать на Python, а общий принцип работы с DAG прост и понятен, однако возникли некоторые проблемы:

- Изначально было непонятно как хранить информацию между вызовами задач, я выбрал вариант с Variable, но она хранит только строки, так что было необходимо преобразовать DataFrame в json;
- Также столкнулся с проблемой отправки данных в Elasticsearch, большой json не отправлялся, были ошибки с маппингом, поэтому отправил каждую строчку отдельно в цикле, что заняло определённое время;

В Elasticsearch получил гистограмму с ценами на напиток относительно оценок дегустаторов:

