



Работа с потоком в Kafka

Цель: Научиться обрабатывать большие файлы в стриминговом потоке Kafka. Задача в загрузке файла для его дальнейшего использования.

Описание задания: На данном этапе вам понадобится первоначальный CSV файл `customs_data.csv` с данными лога таможни. Вам необходимо создать топик, а также написать Kafka consumer и producer. Producer будет считывать в топик файл (можно использовать библиотеки `pandas` или `io` для оптимизации считывания файла). Необходимо также описать функцию сериализации данных, которую производит `KafkaProducer` перед отправкой в топик. Consumer должен выводить на печать в терминале десериализованные строки файла

Ответом на задание является: файлы `consumer.py`, `producer.py`, описание топика (команда для его инициализации), чтобы при запуске запускался поток с содержимым файла, который понадобится вам в следующем задании

Инструменты, которые пригодятся для выполнения: Python, Apache Kafka

Критерии, по которым будет оцениваться задание: Файл корректно считывается в топик Kafka и выводится через consumer на печать в терминале. Оцениваться будет время вывода всего содержимого файла в режиме реального времени, оптимально чтобы оно не превышало 12000 секунд, отличным считается решение выводящее построчно файл менее чем за 9000 секунд. Оценку можно производить библиотекой `time` в Python.

[Далее >>](#)