

Лабораторная работа № 8

Использование ETL инструментов на примере NiFi

Задание

1. Разработать приложение, генерирующее файл в формате JSON/XML/CSV с данными, соответствующими теме БД. С частотой раз в 5 минут необходимо создавать новый файл, имя которого соответствует разработанной маске. Маска должна включать в себя как минимум:
 1. идентификатор файла
 2. имя таблицы, в которую загружаются данные из этого файла
 3. дату и время формирования файла
2. Установить NiFi следуя инструкции - <https://nifi.apache.org/docs/nifi-docs/html/getting-started.html>
3. Разработать NiFi flow для загрузки файлов в БД. Предполагается, что в загружаемых файлах хранятся только новые записи (только INSERT). Обновление и удаление записей не производится. Flow должен включать в себя как минимум следующие процессоры:
 1. GetFile - для считывания файла
 2. ConvertRecord (Reader и Writer необходимо выбирать в зависимости от расширения входных файлов) - для приведения файлов к типу JSON
 3. SplitJSON - для разбиения JSON-массива на отдельные элементы
 4. ConvertJSONToSQL - для преобразования JSON к инструкции SQL
 5. PutSQL - для записи строки в базу данных
 6. LogAttribute - для логирования возможных ошибок
4. Разработать подход к логированию загруженных файлов.
5. Разработать подход к проверке на возможные ошибки входных JSON/XML/CSV файлов.