



Задание №2

Загрузка данных из внешнего ресурса. Преобразование и фильтрация.

По результатам данного задания вы научитесь самостоятельно строить поток данных, запускать его, отлаживать и получать необходимые результаты.

Вы можете сдать это задание не ревью, по вашему желанию.

Инструменты, которые пригодятся для выполнения: Apache Nifi, база данных PostgreSQL и любой удобный для вас SQL инструмент.

Что сдавать на ревью:

Данные в полном объёме должны оказаться в целевой таблице. Предъявите DDL таблицы и результат выборки, приведённой выше и темплейт потока (в виде XML файла). Результаты залейте на GitLab и предоставьте ссылку. По этим данным и будет оцениваться задание.

Описание задания

- На сайте [Euro foreign exchange reference rates](#) представлены файлы в разделе Time series. Используйте [CSV \(.zip\)](#). Внутри архива находится файл eurofxref-hist.csv, который содержит обменный курс различных валют по отношению к EUR.
- Начните поток с процессора GenerateFlowfile. Задайте в одном из атрибутов актуальную на момент выполнения задания ссылку, например, <https://www.ecb.europa.eu/stats/eurofxref/eurofxref-hist.zip>
- Используя процессор InvokeHTTP, скачайте файл по ссылке
- Разархивируйте его с использованием процессора UnpackContent
- Используя процессор QueryRecord, извлеките данные за весь 2023 год, но только для валют USD, JPY, GBP, CNY.
- Сохраните данные в БД. Для этого создайте таблицу (назовите её ecb) с полями date (date), currency (char), rate (number). Используйте составной ключ date + currency. В NiFi используйте PutDatabaserecord процессор.

Желаю удачи!



Подсказка: используйте union all и date надо брать в двойные кавычки). Для CSV Reader контроллера подойдут дефолтные настройки с Schema Access Strategy = Infer Schema

Задайте Date Format = уууу-ММ-дд

Для Output Grouping поменяйте на One Line Per Object

Задайте Date Format = уууу-ММ-дд

```
select * from (  
  
select "date", 'USD' as currency, USD as rate from flowfile  
  
union all  
  
select "date", 'JPY' as currency, JPY from flowfile  
  
union all  
  
select "date", 'GBP' as currency, GBP from flowfile  
  
union all  
  
select "date", 'CNY' as currency, CNY from flowfile  
  
) t  
  
where "date" like '2023%'
```

Перед записью в БД flowfile будет иметь следующее содержимое:

```
{"date":"2023-04-28","currency":"USD","rate":"1.0981"}  
  
...  
  
{"date":"2023-04-28","currency":"JPY","rate":"149.35"}  
  
...  
  
{"date":"2023-01-02","currency":"CNY","rate":"7.3698"}
```

Ваше решение

Отправить на проверку

Сохранить как черновик

Далее >

