

Дата-инженер

Построение потока данных

Николай Акимов



План занятия

- 1 Построение потока данных
- 2 Запуск потока
- 3 Изучение результатов
- 4 Контроль ошибок
- 5 Сохранение потока данных
- 6 Использование Data Provenance



Загрузка и обработка данных

Прочитаем данные из внешнего источника и преобразуем их

Входящие данные

```
<?xml version="1.0" encoding="UTF-8"?>
<gesmes:Envelope xmlns:gesmes="http://www.gesmes.org/xml/2002-08-01" xmlns="
http://www.ecb.int/vocabulary/2002-08-01/eurofxref">
  <gesmes:subject>Reference rates</gesmes:subject>
  <gesmes:Sender>
    <gesmes:name>European Central Bank</gesmes:name>
  </gesmes:Sender>
  <Cube>
    <Cube time='2022-11-08'>
      <Cube currency='USD' rate='0.9996' />
      <Cube currency='JPY' rate='146.25' />
      ...
      <Cube currency='THB' rate='37.220' />
      <Cube currency='ZAR' rate='17.8397' />
    </Cube>
  </Cube>
</gesmes:Envelope>
```



Загрузка и обработка данных

Итоговые данные

time, currency, rate

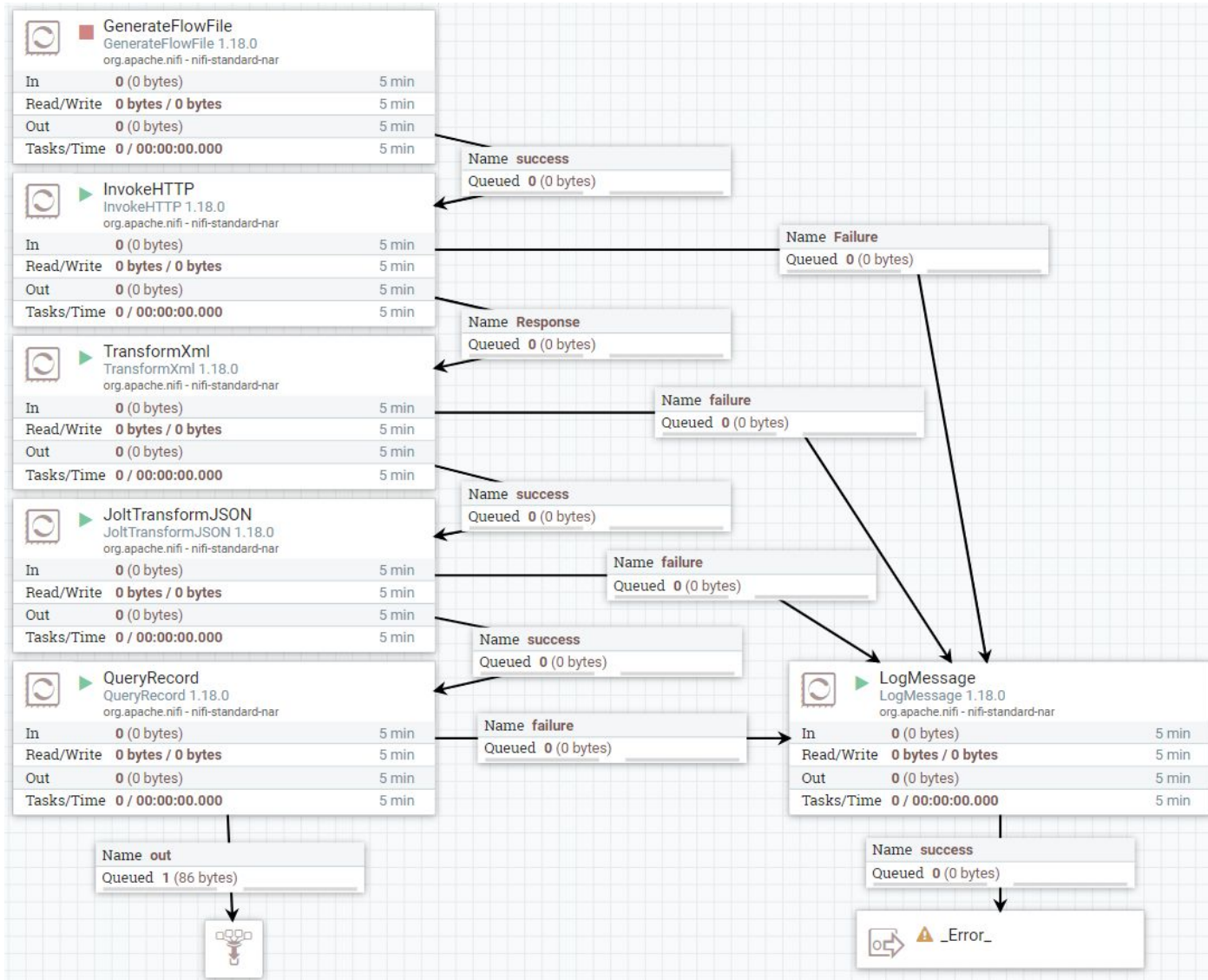
2022-11-08, USD, 0.9996

2022-11-08, JPY, 146.25

2022-11-08, GBP, 0.87378

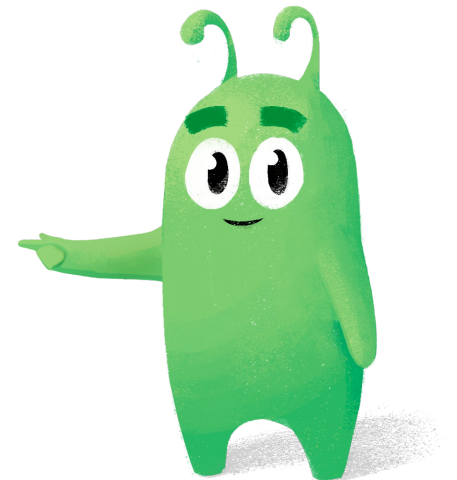


Загрузка и обработка данных



Итоги

- 1 Построили рабочий поток
- 2 Научились им управлять
- 3 Научились сохранять поток
- 4 Научились мониторингу ошибок
- 5 Научились контролю результатов
- 6 Отработали использование Data Provenance





**Спасибо
за внимание!**

