

SKaMP. Tests

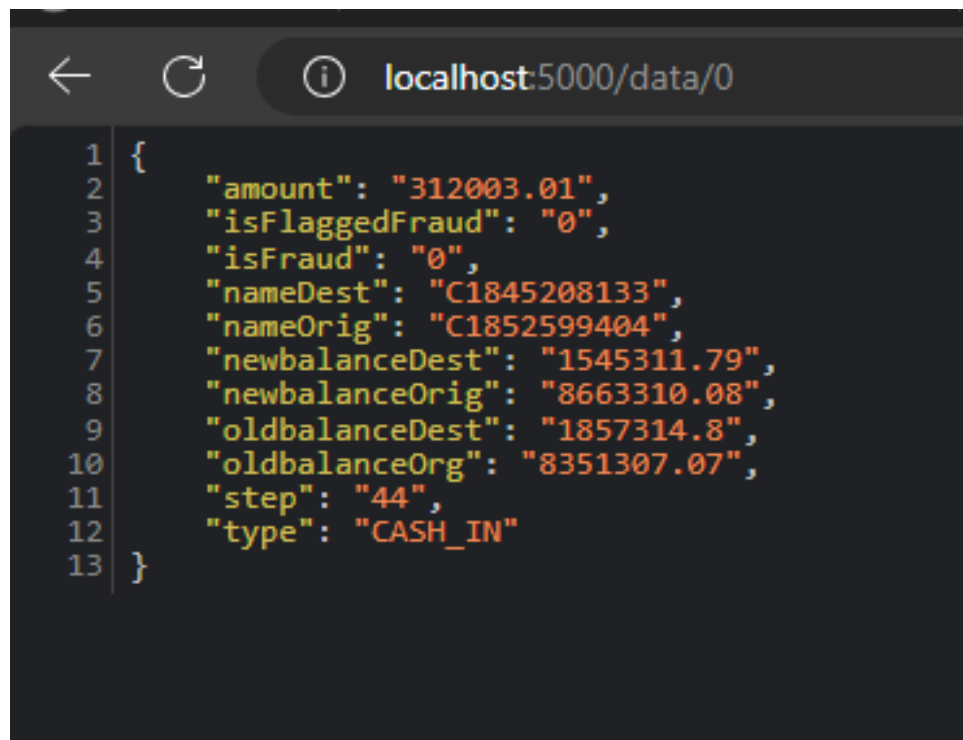
The goal of this report is to provide information on the performed testing of data acquisition and data pre-processing.

GitHub repository: <https://github.com/salveendutt/Big-Data-Analytics>.

1 Data acquisition

Test objective	Steps	Expected Result	Actual Result
Verify data incoming from stream API	1. Start the server using start_containers.bat; 2. Navigate to http://localhost:5000	Incoming data is available on /data/0	Passed. The screenshot is provided in Fig.1 and Fig.2
Verify correct setup of the stream	Run 'pytest' from the root folder	Data stream is configured as expected; Incoming data is not null; Returned status code - 200	PASSED. The screenshot is provided in Fig. 3

Table 1: Data acquisition tests



```
1 {  
2   "amount": "312003.01",  
3   "isFlaggedFraud": "0",  
4   "isFraud": "0",  
5   "nameDest": "C1845208133",  
6   "nameOrig": "C1852599404",  
7   "newbalanceDest": "1545311.79",  
8   "newbalanceOrig": "8663310.08",  
9   "oldbalanceDest": "1857314.8",  
10  "oldbalanceOrg": "8351307.07",  
11  "step": "44",  
12  "type": "CASH_IN"  
13 }
```

Figure 1: Data incoming via the stream

Provenance

Oldest event available: 11/25/2024 09:01:24 UTC

Showing the events that match the specified query. [Clear Search](#)

Filter filter by component name

Filter matched 12 of 12

Event Time ↓	Type	FlowFile UUID	File Size	Component Name	Component Type
11/25/2024 09:01:56.936 UTC	CLONE	138ca4b5-14a5-4179-8b07-727a893aa1...	253 bytes	Funnel	Funnel
11/25/2024 09:01:56.934 UTC	CLONE	5af3616e-6111-48a0-83f1-435a6740023d	270 bytes	Funnel	Funnel
11/25/2024 09:01:56.932 UTC	CLONE	9bc0c60e-8872-4dc6-b31f-b4f206433...	278 bytes	Funnel	Funnel
11/25/2024 09:01:46.934 UTC	CLONE	6392e646-a82d-44bb-bf9e-9efa85c63...	260 bytes	Funnel	Funnel
11/25/2024 09:01:46.933 UTC	CLONE	380784a9-9f26-48b0-a477-21408a48...	286 bytes	Funnel	Funnel
11/25/2024 09:01:46.921 UTC	CLONE	709c4d42-574d-4a09-b98b-b270d12fc...	280 bytes	Funnel	Funnel
11/25/2024 09:01:36.922 UTC	CLONE	f9bfcdf3-b2ac-49b6-9a6b-668bf7a790...	252 bytes	Funnel	Funnel
11/25/2024 09:01:36.921 UTC	CLONE	abe7160e-9b0e-4c65-8b49-2c7313af6...	282 bytes	Funnel	Funnel
11/25/2024 09:01:36.919 UTC	CLONE	d827d474-e467-48f1-b12b-2ad90068d...	287 bytes	Funnel	Funnel
11/25/2024 09:01:24.239 UTC	CLONE	7c648c6d-59d6-4d21-a380-b4d590cd...	253 bytes	Funnel	Funnel
11/25/2024 09:01:24.238 UTC	CLONE	0e67298c-d1da-4b5c-bb18-3893d1ea9...	262 bytes	Funnel	Funnel
11/25/2024 09:01:24.235 UTC	CLONE	51f98930-f1b6-4254-b39b-59465fbb...	279 bytes	Funnel	Funnel

Figure 2: NiFi Data Provenance

2 Data pre-processing

Test objective	Steps	Expected Result	Actual Result
Verify correct data pre-processing of dataset 1	Run 'pytest' from the root folder	Feature 'type' is correctly transformed into numeric value (5 cases); Feature 'is-Merchant' is correctly prepared (2 cases)	PASSED. The screenshot is provided in Fig. 3
Verify correct data processing of dataset 2	Run 'pytest' from the root folder	Numeric boolean values are transformed to int from float (4 cases)	PASSED. The screenshot is provided in Fig. 3
Verify correct data pre-processing of dataset 3	Run 'pytest' from the root folder	Feature 'entry_mode' is correctly transformed into numeric value (4 cases); Unnecessary features are omitted.	PASSED. The screenshot is provided in Fig. 3
Verify correct data pre-processing of dataset 4	Run 'pytest' from the root folder	Features 'Amount', 'Class' are renamed to 'amount' and 'is-Fraud'; Extra features are removed	PASSED. The screenshot is provided in Fig. 3

Table 2: Data pre-processing tests

Unit testing is included in the CI/CD pipeline on GitHub and must be successful before any merge into the main branch.

```

===== test session starts =====
platform win32 -- Python 3.13.0, pytest-8.3.3, pluggy-1.5.0 -- C:\ProgramFiles\Anaconda3\envs\bigdata13\python.exe
cachedir: .pytest_cache
rootdir: C:\home\WUT\Semester_3\BigData\Big-Data-Analytics
collected 12 items

services/streaming_simulation/test_streaming_simulation.py::StreamingSimulationTestCase::test_data_stream PASSED [ 8%]
tests/data_utils/test_utils.py::test_preprocess_1_payment PASSED [ 16%]
tests/data_utils/test_utils.py::test_preprocess_1_cash_in PASSED [ 25%]
tests/data_utils/test_utils.py::test_preprocess_1_cash_out PASSED [ 33%]
tests/data_utils/test_utils.py::test_preprocess_1_debit PASSED [ 41%]
tests/data_utils/test_utils.py::test_preprocess_1_unknown PASSED [ 50%]
tests/data_utils/test_utils.py::test_preprocess_row_2 PASSED [ 58%]
tests/data_utils/test_utils.py::test_preprocess_3_contactless PASSED [ 66%]
tests/data_utils/test_utils.py::test_preprocess_3_chip PASSED [ 75%]
tests/data_utils/test_utils.py::test_preprocess_3_swipe PASSED [ 83%]
tests/data_utils/test_utils.py::test_preprocess_3_unknown PASSED [ 91%]
tests/data_utils/test_utils.py::test_preprocess_row_4 PASSED [100%]

===== 12 passed in 0.57s =====

```

Figure 3: Unit testing result

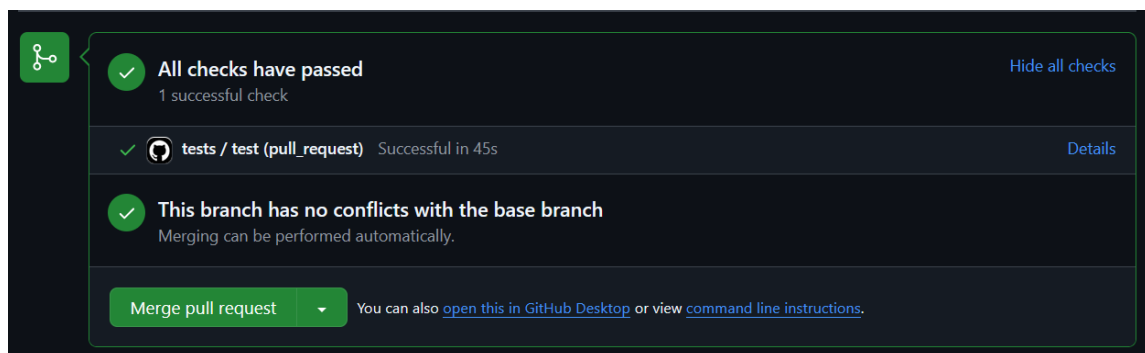


Figure 4: GitHub checks before merge