

Лабораторная работа 3

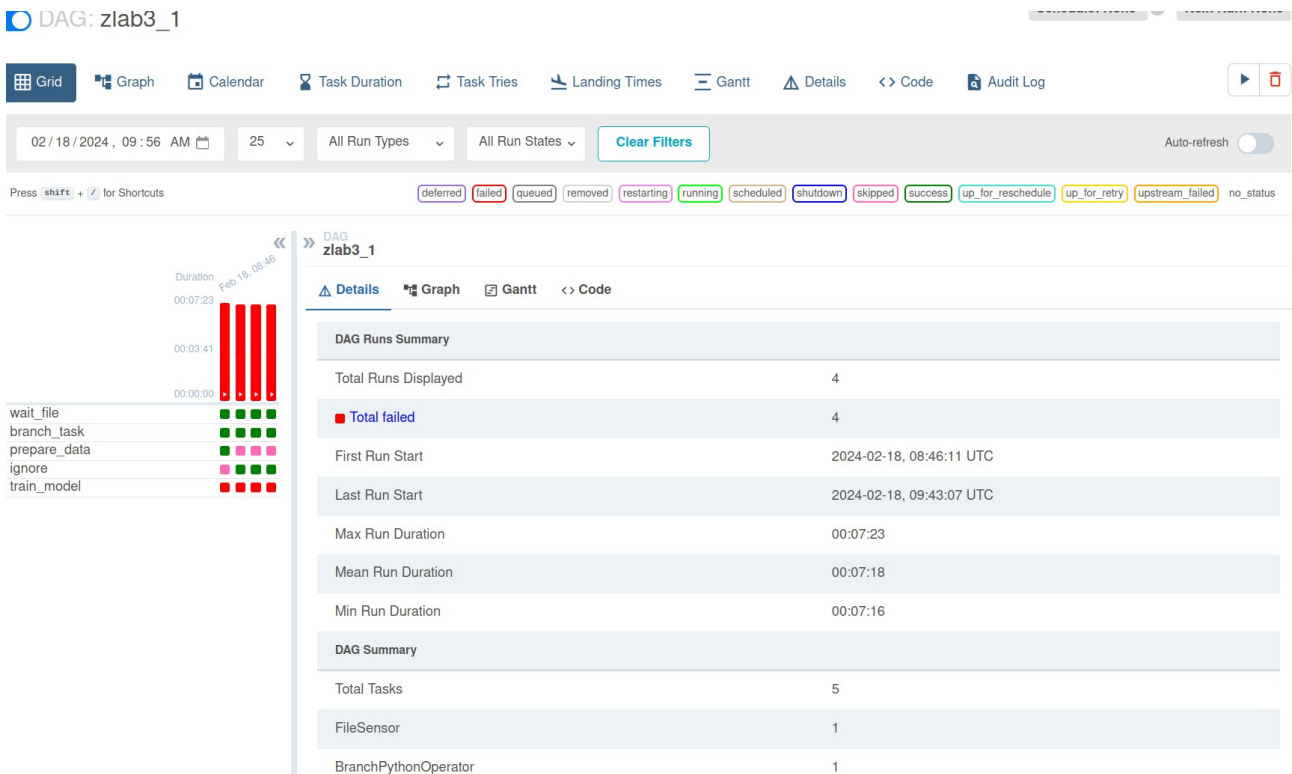
Был сконфигурирован пайплайн для обучения, с кодом представленным ниже:

```
zlab3_1.py
~/IP2/Pre/airflow/dags

1 from airflow import DAG
2 from datetime import datetime
3 from airflow.operators.bash import BashOperator
4 from airflow.operators.empty import EmptyOperator
5 from airflow.sensors.filesystem import FileSensor
6 from airflow.operators.python import BranchPythonOperator
7 from airflow.utils.trigger_rule import TriggerRule
8 import os
9
10 os.environ["AWS_ACCESS_KEY_ID"] = "minio"
11 os.environ["AWS_SECRET_ACCESS_KEY"] = "minio123"
12 os.environ["MLFLOW_S3_ENDPOINT_URL"] = "http://minio:9000"
13
14
15 def branch_func():
16     if (os.path.isfile('/opt/airflow/data/x_train.csv') and os.path.isfile('/opt/airflow/data/y_train.csv')
17         and os.path.isfile('/opt/airflow/data/x_val.csv') and os.path.isfile('/opt/airflow/data/y_val.csv')
18         and os.path.isfile('/opt/airflow/data/x_test.csv') and os.path.isfile('/opt/airflow/data/y_test.csv')):
19         return ignore.task_id
20     else:
21         return prepare_data.task_id
22
23 default_args = {
24     'owner': 'alexcey',
25     'retries': 1,
26 }
27
28 dag = DAG(
29     'zlab3_1',
30     default_args=default_args,
31     description='',
32     start_date = datetime(2024, 2, 10),
33     schedule_interval=None,
34 )
35
36 branch_op = BranchPythonOperator(
37     task_id='branch_task',
38     python_callable=branch_func,
39     trigger_rule="all_done",
40     dag=dag,
41 )
42
43 prepare_data = BashOperator(
44     task_id='prepare_data',
45     bash_command="python /opt/airflow/data/prepare_data.py",
46     dag=dag
47 )
48
49 wait_file = FileSensor(
50     task_id='wait_file',
51     poke_interval=5,
52     filepath='/opt/airflow/data/conf.json',
53     fs_conn_id='fs_conn_id',
54     dag=dag,
55 )
56
57 ignore = EmptyOperator(
58     task_id='ignore'
59 )
60
61 train_model = BashOperator(
62     task_id='train_model',
63     bash_command="python /opt/airflow/data/train.py",
64     trigger_rule=TriggerRule.NONE_FAILED_MIN_ONE_SUCCESS,
65     dag=dag
66 )
67
68 wait_file >> branch_op
69
70 branch_op >> ignore >> train_model
71 branch_op >> prepare_data >> train_model
```

Python ▾ Tab Width: 10 ▾ Ln 64, Col 30 ▾ INS

Результат работы в Airflow:



На участке кода с обучением модели получена ошибка подключения к серверу, как не пытался, победить не смог.

```
.[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -  
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - Traceback (most recent call  
last):  
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - File  
"/home/***/.local/lib/python3.8/site-packages/requests/adapters.py", line 486,  
in send  
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -     resp = conn.urlopen(  
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - File  
"/home/***/.local/lib/python3.8/site-packages/urllib3/connectionpool.py", line  
826, in urlopen  
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -     return self.urlopen(  
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - File  
"/home/***/.local/lib/python3.8/site-packages/urllib3/connectionpool.py", line  
826, in urlopen  
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -     return self.urlopen(  
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - File  
"/home/***/.local/lib/python3.8/site-packages/urllib3/connectionpool.py", line  
826, in urlopen  
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -     return self.urlopen(  
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - [Previous line repeated 2  
more times]  
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - File  
"/home/***/.local/lib/python3.8/site-packages/urllib3/connectionpool.py", line  
798, in urlopen  
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -     retries =  
retries.increment()
```

```
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - File
"/home/***/.local/lib/python3.8/site-packages/urllib3/util/retry.py", line 592,
in increment
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - raise
MaxRetryError(_pool, url, error or ResponseError(cause))
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -
urllib3.exceptions.MaxRetryError: HTTPConnectionPool(host='mlflow_server',
port=5000): Max retries exceeded with url: /api/2.0/mlflow/experiments/get?
experiment_id=0 (Caused by
NewConnectionError('<urllib3.connection.HTTPConnection object at
0x7fc049afe730>: Failed to establish a new connection: [Errno 111] Connection
refused'))
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - During handling of the
above exception, another exception occurred:
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - Traceback (most recent call
last):
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - File
"/home/***/.local/lib/python3.8/site-packages/mlflow/utils/rest_utils.py", line
108, in http_request
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - return
_get_http_response_with_retries(
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - File
"/home/***/.local/lib/python3.8/site-packages/mlflow/utils/request_utils.py",
line 212, in _get_http_response_with_retries
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - return
session.request(method, url, allow_redirects=allow_redirects, **kwargs)
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - File
"/home/***/.local/lib/python3.8/site-packages/requests/sessions.py", line 589,
in request
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - resp = self.send(prepare,
**send_kwargs)
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - File
"/home/***/.local/lib/python3.8/site-packages/requests/sessions.py", line 703,
in send
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - r =
adapter.send(request, **kwargs)
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - File
"/home/***/.local/lib/python3.8/site-packages/requests/adapters.py", line 519,
in send
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - raise
ConnectionError(e, request=request)
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -
requests.exceptions.ConnectionError: HTTPConnectionPool(host='mlflow_server',
port=5000): Max retries exceeded with url: /api/2.0/mlflow/experiments/get?
experiment_id=0 (Caused by
NewConnectionError('<urllib3.connection.HTTPConnection object at
0x7fc049afe730>: Failed to establish a new connection: [Errno 111] Connection
refused'))
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - During handling of the
above exception, another exception occurred:
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - Traceback (most recent call
last):
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - File
"/opt/***/data/train.py", line 25, in <module>
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -
mlflow.set_experiment(experiment_id="0")
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO - File
"/home/***/.local/lib/python3.8/site-packages/mlflow/tracking/fluent.py", line
150, in set_experiment
```

```
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -      experiment =
client.get_experiment(experiment_id)
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -      File
"/home/***/.local/lib/python3.8/site-packages/mlflow/tracking/client.py", line
487, in get_experiment
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -      return
self._tracking_client.get_experiment(experiment_id)
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -      File
"/home/***/.local/lib/python3.8/site-packages/mlflow/tracking/_tracking_service/
client.py", line 214, in get_experiment
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -      return
self.store.get_experiment(experiment_id)
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -      File
"/home/***/.local/lib/python3.8/site-packages/mlflow/store/tracking/
rest_store.py", line 111, in get_experiment
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -      response_proto =
self._call_endpoint(GetExperiment, req_body)
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -      File
"/home/***/.local/lib/python3.8/site-packages/mlflow/store/tracking/
rest_store.py", line 59, in _call_endpoint
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -      return
call_endpoint(self.get_host_creds(), endpoint, method, json_body,
response_proto)
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -      File
"/home/***/.local/lib/python3.8/site-packages/mlflow/utils/rest_utils.py", line
216, in call_endpoint
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -      response =
http_request(**call_kwargs)
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -      File
"/home/***/.local/lib/python3.8/site-packages/mlflow/utils/rest_utils.py", line
130, in http_request
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -      raise
MlflowException(f"API request to {url} failed with exception {e}")
[2024-02-18, 09:16:31 UTC] {subprocess.py:93} INFO -
mlflow.exceptions.MlflowException: API request to
http://mlflow\_server:5000/api/2.0/mlflow/experiments/get failed with exception
HTTPConnectionPool(host='mlflow_server', port=5000): Max retries exceeded with
url: /api/2.0/mlflow/experiments/get?experiment_id=0 (Caused by
NewConnectionError('<urllib3.connection.HTTPConnection object at
0x7fc049afe730>: Failed to establish a new connection: [Errno 111] Connection
refused'))
[2024-02-18, 09:16:31 UTC] {subprocess.py:97} INFO - Command exited with return
code 1
[2024-02-18, 09:16:31 UTC] {taskinstance.py:1943} ERROR - Task failed with
exception
```