

Домашнее задание 7. Построение пайплайнов и визуализация потоков данных в Airflow

— Зарегистрируйтесь в OpenWeatherApi (<https://openweathermap.org/api>)

— Создайте ETL, который получает температуру в заданной вами локации, и дальше делает ветвление:

- В случае, если температура больше 15 градусов цельсия — идёт на ветку, в которой есть оператор, выводящий на экран «тепло»;
- В случае, если температура ниже 15 градусов, идёт на ветку с оператором, который выводит в консоль «холодно».

Оператор ветвления должен выводить в консоль полученную от API температуру.

— Приложите скриншот графа и логов работы оператора ветвления.

```
from airflow import DAG
from airflow.operators.python import BranchPythonOperator
from airflow.operators.bash import BashOperator
from datetime import datetime
import requests

def get_weather():
    city_id = 523750
    appid = "df4ca991b6bdd4362be0546c97d931fb"
    try:
        res = requests.get('http://api.openweathermap.org/data/2.5/weather?',
                           params={'id': city_id, 'units': 'metric', 'lang': 'ru', 'APPID': appid})
        data = res.json()
        print(data['main']['temp'])
        return int(data['main']['temp'])
    except Exception as e:
        print("Exception (weather):", e)

def _choose_temp():
    accuracy = get_weather()
    if accuracy > 15:
        return 'accurate'
    return 'inaccurate'
```

```
with DAG('hw7', start_date=datetime(2024, 6, 18), schedule='@daily', catchup=False, tags=["homework"]):
    choose_best_model = BranchPythonOperator(
        task_id='choose_temp',
        python_callable=_choose_temp
    )

    accurate = BashOperator(
        task_id="accurate",
        bash_command="echo 'тепло'"
    )

    inaccurate = BashOperator(
        task_id="inaccurate",
        bash_command=" echo 'холодно'"
    )

choose_best_model >> [accurate, inaccurate]
```


▼

[Clear Filters](#)

shift

fo

r Sh

ortc

uts

fa

q

re

re

ru

SO



00:00:02

00:00:00

19.03.00

Run

Task

hw7 / ▶ 2024-06-19, 03:00:00 MSK / choose_temp

 Graph

< > Code

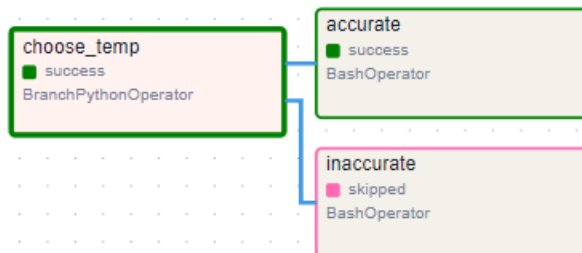
< > Code

≡ Logs

≡ Logs

⌚ Task Duration

⌚ Task Duration



19.06.2024 14:03:12

All Run Types

All Run States

Clear Filters

Press **shift** + **/** for Shortcuts

deferredfailedqueuedremove

Duration

00:00:04

00:00:02

00:00:00

Jun 19, 03:00

choose_temp

accurate

inaccurate

DAG Run Task

hw7 / 2024-06-19, 03:00:00 MSK / choose_temp

DetailsGraphGanttCodeAudit LogLogsXComTask Duration

(by attempts)

1

All Levels

All File Sources

MSI-notebook.
*** Found local files:
*** * /home/akst/airflow/logs/dag_id=hw7/run_id>manual__2024-06-19T11:03:10.168128+00:00/task_id=choose_temp/attempt=1.log
[2024-06-19, 14:03:13 MSK] {local_task_job_runner.py:120} ▶ Pre task execution logs
[2024-06-19, 14:03:13 MSK] {baseoperator.py:399} WARNING - BranchPythonOperator.execute cannot be called outside TaskInstance!
[2024-06-19, 14:03:13 MSK] {logging_mixin.py:188} INFO - 27.49
[2024-06-19, 14:03:13 MSK] {python.py:237} INFO - Done. Returned value was: accurate
[2024-06-19, 14:03:13 MSK] {branch.py:36} INFO - Branch into accurate
[2024-06-19, 14:03:13 MSK] {skipmixin.py:178} INFO - Following branch accurate
[2024-06-19, 14:03:13 MSK] {skipmixin.py:246} INFO - Skipping tasks [('inaccurate', -1)]
[2024-06-19, 14:03:13 MSK] {taskinstance.py:441} ▶ Post task execution logs

19.06.2024 14:03:12

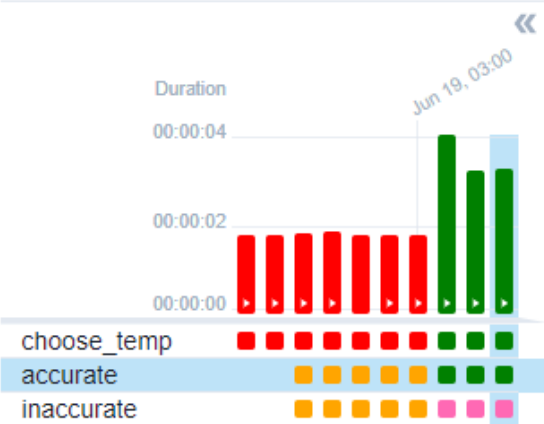
All Run Types

All Run States

Clear Filters

Press shift + / for Shortcuts

deferredfailedqueuedremovedrestarting



DAG Run Task
hw7 / 2024-06-19, 03:00:00 MSK / accurate

Details Graph Gantt Code Audit Log Logs XCom Task Duration

(by attempts)

1

All Levels All File Sources

MSI-notebook.
*** Found local files:
*** * /home/akst/airflow/logs/dag_id=hw7/run_id>manual__2024-06-19T11:03:10.168128+00:00/task_id=accurate/attempt=1.log
[2024-06-19, 14:03:15 MSK] {local_task_job_runner.py:120} ▶ Pre task execution logs
[2024-06-19, 14:03:15 MSK] {subprocess.py:63} INFO - Tmp dir root location: /tmp
[2024-06-19, 14:03:15 MSK] {subprocess.py:75} INFO - Running command: ['/usr/bin/bash', '-c', "echo 'tenno'"]
[2024-06-19, 14:03:15 MSK] {subprocess.py:86} INFO - Output:
[2024-06-19, 14:03:15 MSK] {subprocess.py:93} INFO - tenno
[2024-06-19, 14:03:15 MSK] {subprocess.py:97} INFO - Command exited with return code 0
[2024-06-19, 14:03:15 MSK] {taskinstance.py:441} ▶ Post task execution logs

19.06.2024 14:03:12

All Run Types

All Run States

Clear Filters

Press **shift** + **/** for Shortcuts

deferred

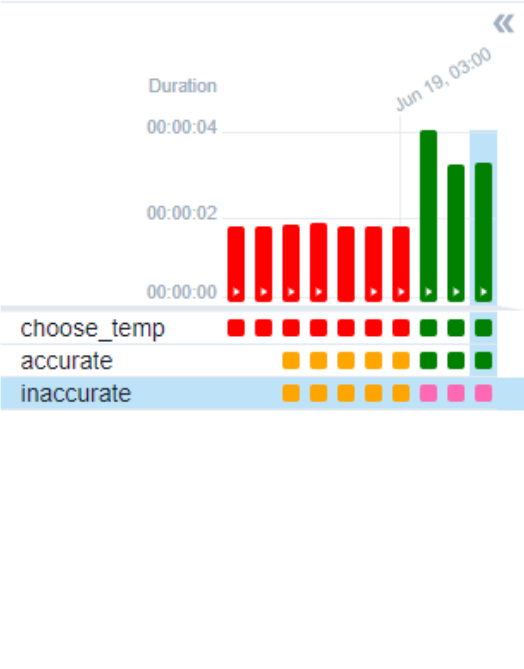
failed

queued

removed

restarting

running



DAG

Run

Task

hw7 / ▶ 2024-06-19, 03:00:00 MSK / inaccurate

Details

Graph

Gantt

Code

Audit Log

Logs

XCom

Task Duration

(by attempts)

All Levels

All File Sources

*** Could not read served logs: Request URL is missing an 'http://' or 'https://' protocol.