

Отчет по лабораторной работе №2

Выполнила Агафонова Елизавета, группа 6233

1 часть

Создадим подключение к airflow.

The screenshot shows the Airflow web interface. The top navigation bar includes links for Airflow, DAGs, Cluster Activity, Datasets, Security, Browse, Admin, and Docs. The 'List Connection' screen displays a table with one connection:

Conn Id	Conn Type	Description	Host	Port	Is Encrypted	Is Extra Encrypted
Conn_lab2	docker	Connection to docker	tcp://docker-proxy	2375	False	False

The 'Edit Connection' screen shows the following fields:

- Connection Id: Conn_lab2
- Connection Type: Docker
- Description: Connection to docker
- Registry URL: tcp://docker-proxy
- Username: (empty)
- Password: (empty)
- Port: 2375
- Extra:

```
{  "reauth": false}
```
- Reauthenticate: (checkbox unchecked)
- Email: (empty)

At the bottom, there are buttons for 'Save', 'Test', and a back arrow. The footer shows the version: v2.7.0 and the Git commit hash: c08c82e9dd0e4a3baa5121519819a636df635210.

Получим api с hugging face.

Access Tokens

User Access Tokens

+ Create new token

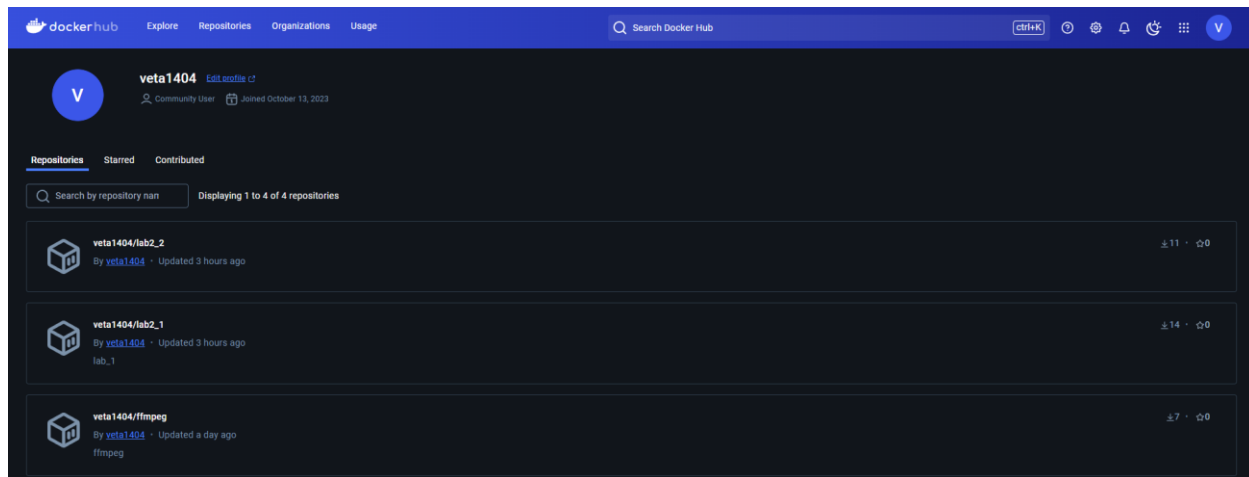
Access tokens authenticate your identity to the Hugging Face Hub and allow applications to perform actions based on token permissions.

Do not share your Access Tokens with anyone; we regularly check for leaked Access Tokens and remove them immediately.

Name	Value	Last Refreshed Date	Last Used Date	Permissions
Lab2	hf_...xylq	1 day ago	-	WRITE

Создадим докер образы для 1 и 2 части

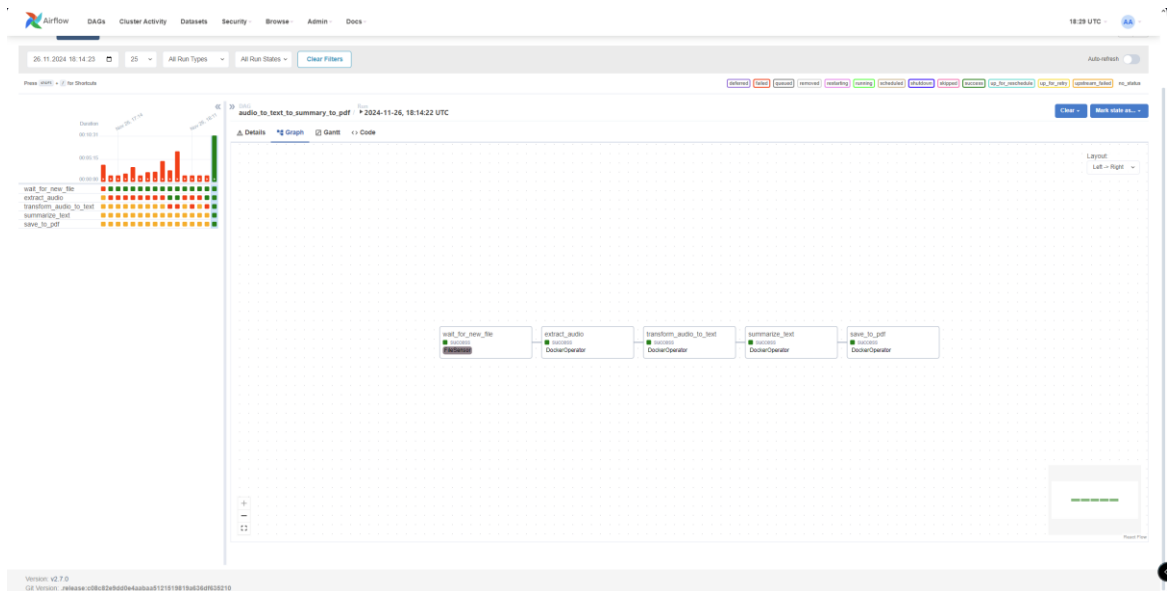
<https://hub.docker.com/u/veta1404>



Создадим DAG файл DAGFile_interference.py.

- wait_for_new_file – ожидание файла. Для этого было создано connection с airflow.
- extract_audio – извлечение аудио из видео (образ ffmpeg-image)
- transform_audio_to_text – перевод аудио в текст при помощи модели whisper-small (образ lab2_1- image)
- summarize_text – перевод аудио в текст при помощи модели bart-summarization (образ lab2_1- image)
- save_to_pdf – сохранение текстового файла в pdf (образ lab2_1- image)

Выполненный DAG.



В результате работы пайплайна:

- считан файл input.mp4
- извлечен звук audio.wav
- звук переведен в текст full.txt

```
full.txt
1 Obi-Wan never told you what happened with your father. He told me enough! He told me you killed him. No. I am your father.
```

- создано саммари для текста summary.txt

```
summary.txt
1 Obi-Wan never told his son what happened with his father. Obi-Wan is your father. He told him that you killed his father, so he doesn't want to tell him what happened. Obi-Wan is not afraid to tell his son.
```

- текст сохранен в pdf summary.pdf

Obi-Wan never told his son what happened with his father. Obi-Wan is your father. He told him that you killed his father, so he doesn't want to tell him what happened.

Obi-Wan is not afraid to tell his son.

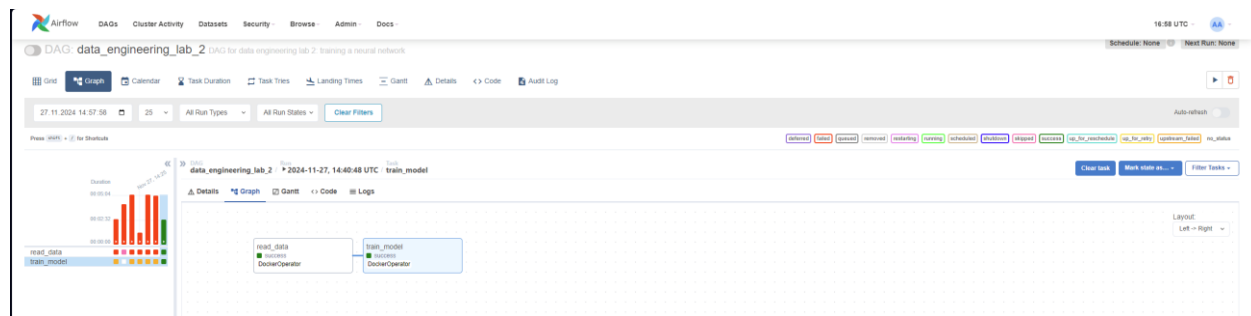
2 часть

Создадим DAG файл DAGFile_train.py.

- `read_data` – чтение датасета, кодирование категориальных данных и разделение датасета на обучающие и тестовые наборы данных (образ lab2_2- image)
- `train_model` – создание простейшей полносвязной нейронки, ее компилирование и обучение модели (образ lab2_2- image)

Я обучала созданную простейшую нейронную сеть на датасете Housing.csv (Нейронка дает плохие результаты с мизерной точностью, но точность никак не получилось улучшить()

Выполненный DAG.



Процесс инференса модели

```
WARNING:absl:Compiled the loaded model, but the compiled
2/2 ----- 0s 30ms/step
```

Classification Report:				
	precision	recall	f1-score	support
Class 0	0.02	1.00	0.04	1
Class 1	0.00	0.00	0.00	10
Class 2	0.00	0.00	0.00	34
Class 3	0.00	0.00	0.00	10
accuracy			0.02	55
macro avg	0.00	0.25	0.01	55
weighted avg	0.00	0.02	0.00	55