Департамент образования города Москвы

Государственное автономное образовательное учреждение

высшего образования города Москвы

«Московский городской педагогический университет»

Институт цифрового образования

Департамент информатики, управления и технологий

ОТЧЕТ

по дисциплине «Проектный практикум по разработке ETL-решений»

Направление подготовки 38.03.05 – бизнес-информатика

Профиль подготовки «Аналитика данных и эффективное управление»

(очная форма обучения)

Бизнес кейс «Rocket»

Выполнил:

Студент группы

st\_89

Руководитель:

Москва  
2025

**Цель:** Автоматизировать процесс извлечения, преобразования и загрузки (ETL) данных в рамках бизнес-кейса «Rocket» с использованием Apache Airflow.

(Вариант 2)

**Задачи:**   
1) Спроектировать верхнеуровневую архитектуру аналитического решения задания Бизнес-кейса «Rocket»

2) Спроектировать архитектуру DAG Бизнес-кейса «Rocket»

3) Диаграмма Ганта работы DAG в Apache Airflow

**Ход работы:**

1. Верхнеуровневая архитектура аналитического решения задания Бизнес-кейса «Rocket»

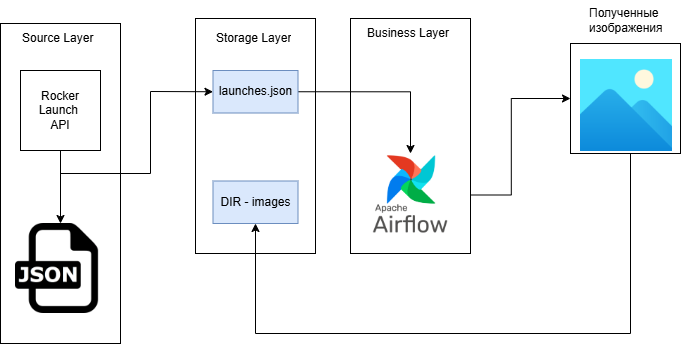


Рис. 1 – Верхнеуровневая архитектура

1. Архитектура DAG

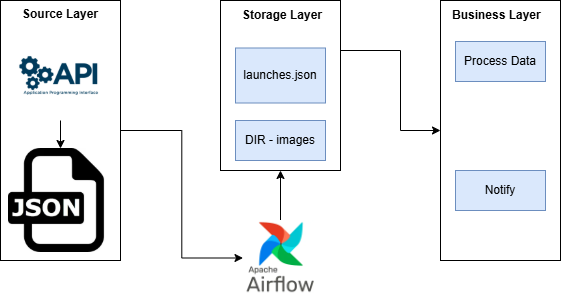


Рис. 2 – Архитектура DAG

1. Анализ JSON

В JSON присутствуют основные поля:

"rocket": {

}

"mission": {

}

"pad": {

}

1. Создание файла .sh для копирования

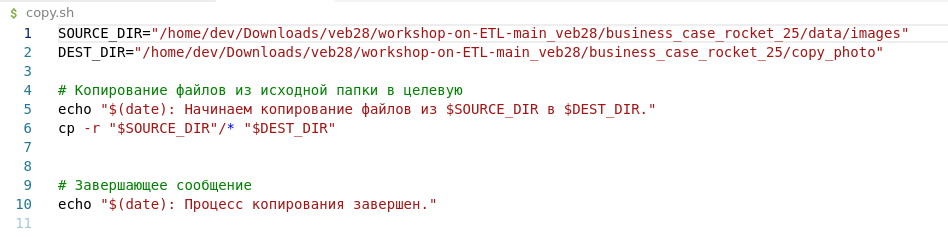


Рис. 3 - Копирование фото

1. Далее производим запуск контейнера с измененным DAG-ом по заданию:
   1. Разработать метод уведомлений при ошибке в скачивании изображений
   2. Настроить DAG для обработки данных о стартах ракет с ежедневным запуском

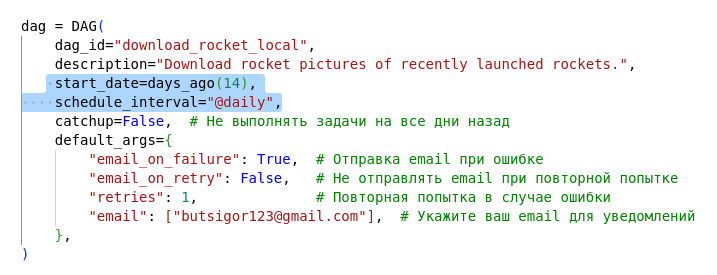


Рис. 4 – Изменение времени запуска по времени

Для того, чтобы корректно работала рассылка на email при ошибочном скачивании, необходимо добавить в airflow.cfg следующие значения

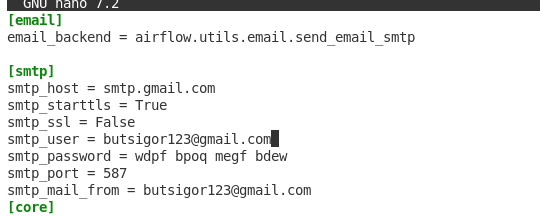


Рис. 5 – Добавление установок для конфига

Так же на Рис. 4 добавлены строки для того, что сообщения при ошибке отправлялось на email.

И так же добавлена функция для отправки сообщения

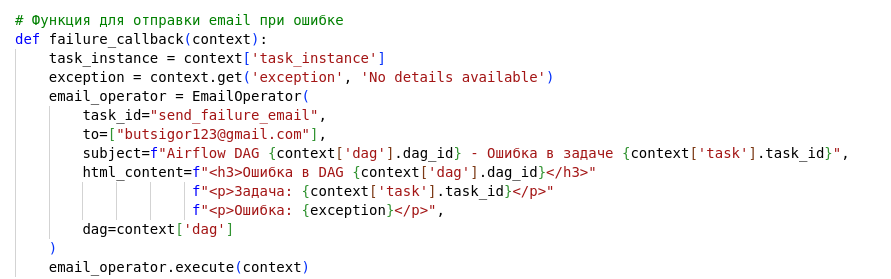


Рис. 6 – Функция отправки

Помимо шагов выше было проведено изменения файла .yml, DockerFile

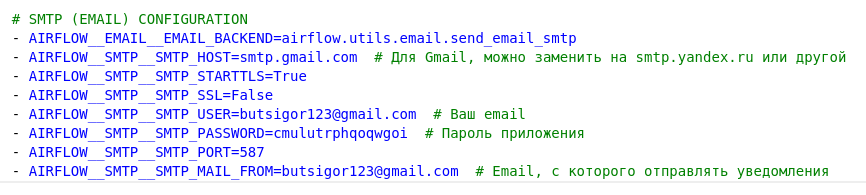


Рис. 7 - Файл .yml

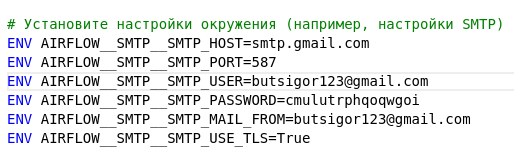


Рис. 8 - Настройки DockerFile

Даже после этих настроек отправку сообщения на почту при ошибке в DAG не удалось реализовать из-за сложности реализации.

1. DAG в виде Диаграммы Ганта

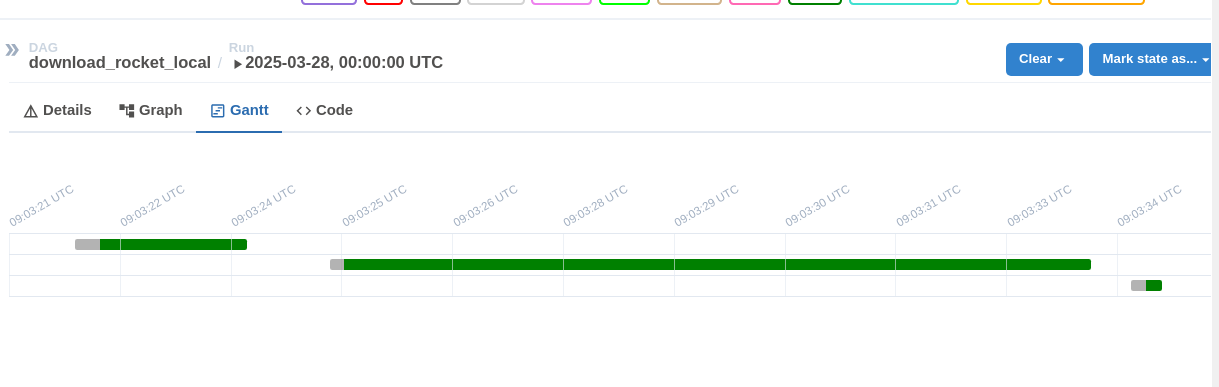


Рис. 9 – Диаграмма Ганта DAG’а

**Экономический эффект**

1. Снижение затрат на труд: Автоматизация работы с данными позволяет значительно сократить время, которое сотрудники тратят на ручную обработку данных. Это снижает потребность в дополнительных человеческих ресурсах.
2. Увеличение производительности: DAGs позволяют запускать задачи без вмешательства человека, что улучшает производительность и уменьшает вероятность задержек.
3. По мере того, как бизнес будет расширяться, количество данных увеличится, а с ним и нагрузка на ручные процессы. Автоматизация этих процессов обеспечит значительный рост продуктивности без пропорционального увеличения затрат на рабочую силу.
4. При этом, использование автоматической отправки сообщения на email принесет более автоматизированную систему обработки данных, но она имеет определенные сложности в настройке.