1. Что такое Airflow
2. Как он устроен, рассказать про DAG’и и таски
3. Запустить пару тестов
4. Структура проекта
5. Pipe
6. Декораторы внутри

Apache Airflow – это оркестратор процессов. Он написан на языке Python и имеет открытый исходный код. Тут имеются в виду процессы по работе с данными, например:

- загрузка данных в базу

- чтение данных и копирование в файл

- упаковка данных из базы в какой-нибудь другой формат или архив

Вообще все эти процессы используются для работы с big data, чтобы было удобно их подготовить и, например, проанализировать.

Airflow использует DAG’и – это направленные ациклические графы. То есть это такая сущность, в которой можно объединить несколько задач (тасков), настроить некоторое их взаимодействие, выполнить их как-то параллельно или последовательно. По сути таск – это питоновская функция, в которой прописана некая бизнес-логика. DAG’и можно запускать вручную, по событию или планировщиком. Для этого тоже используется одна из питоновских библиотек.

Собственно, провайдер написан на Питоне. Вообще суть провайдера с том, чтобы облегчить написание кода для выполнения бизнес-логики. То есть мы вынесли всякие повторяющиеся запросы в декораторы и сама по себе питоновская функция теперь выглядит компактно и понятно. Чтобы не прописывать каждый раз для внутри таска, что мы хотим подключиться к Clickhouse, например, и держать это подключение открытым. Если подключение не получится и вывалится ошибка, нужно либо переподключиться, либо как-то обработать ее.

Собственно, запустим тестовый DAG и посмотрим на примере, как он отработает.

Посмотрим на структуру провайдера. Значит, тут идет разбиение на модули. То есть мы можем в дальнейшем масштабировать программу.

Общая суть описана в модуле pipe.

…

Далее декораторы описаны в соответствующих модулях.

…

Итак, с помощью провайдера можно описать непосредственную логику работы с данными или логику запросов внутри Python функции, а сами запросы выполнять конвейерным способом. При этом код выглядит понятным, не перегруженным, все идет поэтапно.