1. В ETL-процесс были добавлены шаги, которые вызывают три проверки качества данных. Это не изменило последовательность шагов загрузки и обработки данных, но условий выполнения шагов стало больше.

2. После шага `get_files` появилась группа из четырёх проверок `dq_rule_1_check_files`. Эта группа проверяет наличие всех файлов, и проверка критичная: если хотя бы один из четырёх отсутствует, загрузка остановится.

После загрузки данных в staging-область добавились ещё четыре шага, вызывающие проверки качества данных.

- `dq_rule_3_check_customer_id_order_log` и `dq_rule_4_check_customer_id_activity_log`, которые проверяют записи на наличие значений `NULL`;

- `dq_rule_2_check_test_data_order_log` и `dq_rule_5_check_test_data_activity_log`, которые проверяют наличие тестовых данных.

Последние две проверки тоже критичные, и если данные их не пройдут, то процесс не сможет продолжить работу.

1. Общее правило при поиске ошибок — вернуться на шаг назад и проверить логи предыдущего этапа. Первичную информацию о процедурах можно найти в логах Airflow, но обычно, чтобы понять первопричину ошибки, нужно пойти в логи той программы, где были созданы шаги.

2. Проверка `dq_rule_1_check_files` должна показать наличие всех четырёх файлов.
   a. Первым делом нужно убедиться, что все они есть в каталоге для загрузки.
   b. Если там нет хотя бы одного из файлов, проверяем логи или журналы предыдущего шага —`get_files`. Наша задача — убедиться, что все четыре файла были найдены и переписаны в локальный каталог на ETL-сервере.
   c. Если файлов в каталоге не обнаружилось, то нужно проверить, где в шаге `get files` произошла ошибка. Чтобы решить, что делать дальше, отталкиваемся от информации об ошибках из лога `get_files`.

3. Проверка `dq_rule_2_check_test_data_order_log` определяет наличие тестового набора данных. 1. Если проверка показала, что данные не качественные, нужно проверить, есть ли в таблице с результатами работы проверок `dq_checks_results` запись с результатом этой проверки.

   Результат не соответствует нашим требованиям, но нужно убедиться, что проблема действительно в качестве данных. Поэтому следующий шаг — проверить данные в таблице `user_order_log`. Для этого нужно написать запрос для подсчета количества клиентов:

   ```
   Select count(distinct(customer_id)) from user_order_log
   ```

   a. Если значение окажется меньше трёх, значит поступил тестовый набор данных. Проверка показала верный результат — теперь остаётся разбираться с тем, откуда изначально пришли файлы с тестовыми данными.
   b. Если значение больше трёх, нужно просмотреть логи выполнения этой проверки по ключевому слову «error» и искать причины проблемы, как и в случае с первой проверкой.