

# Лекция 4. Практическое задание

Курс: разработчик ХД



Copyright © 2018–2025 by Datatech. All Rights Reserved.





## Лекция 4: Практическое задание (1 из 4)

На данном этапе проекта Заказчик предоставил спецификации на источники данных и доступ к ним.

- **Источник 1:** реляционная база данных.

- customer – покупатели с картами лояльности.
- item – номенклатура товаров.
- item\_inventory – история остатков товаров в точках продаж (еженедельно).
- item\_price\_history – история изменения цен товаров.
- sales\_transaction – финансовая информация о покупках клиентов.
- sales\_transaction\_line – детальная информация о покупках клиентов (чеки).

- **Спецификация источника:**

- Провайдер данных: JDBC.
- Сервер: 10.4.107.31.
- Порт: 5432.
- Пользователь и пароль источника: studentXX (ваш обычный номер, ранее предоставленный куратором).
- База данных источника: mwdo\_base.



## Лекция 4: Практическое задание (2 из 4)

- **Источник 2:** текстовые файлы:

- osh\_commodity – индекс премиальности товара.
- osh\_grocery\_basket – продуктовые корзины.
- osh\_item\_subclass – категории товаров.
- osh\_item\_type – типы товаров.
- t\_all\_divisions – орг. структура самого верхнего уровня.
- t\_division – орг. структура 2 уровня.
- t\_region – орг. структура 3 уровня.
- t\_district – орг. структура 4 уровня.
- t\_location – точки продаж (орг. структура самого нижнего уровня).
- t\_location\_type – типы точек продаж.
- t\_brand – названия торговых марок.
- t\_channel – каналы продаж.
- t\_price\_change\_reason – причины изменения цен товаров.
- t\_vendor – поставщики товаров.

- **Спецификация источника:**

- Тип файла: CSV в кодировке UTF-8.
- Разделитель колонок: запятая.
- Заголовок: присутствует (первая строка).
- Сервер: 10.4.107.31.
- Порт: 8082.



## Лекция 4: Практическое задание (3 из 4)

Полный атрибутный состав таблиц-источников приведен в **раздаточных материалах** к лекции.

1. Создайте в своей схеме **внешние таблицы**, смотрящие на представления или файлы **источников**, используя доступ через протоколы gpfdist или PXF. Примеры задания значения атрибута LOCATION для подключения к источникам приведены в раздаточных материалах.
2. Создайте в своей схеме **таблицы Staging** для последующего импорта данных из источников.
  - **Обязательное условие:** если на источнике колонка не заполнена (IS NULL) или пустая (= "") для всех строк таблицы, она исключается из Staging и последующей обработки.
3. Напишите SQL-скрипты загрузки данных в **промежуточные** таблицы хранилища из источников.
  - **Подсказка:** объекты всех слоев мы создаем в одной и той же схеме Greenplum! Используйте приставку src\_ для источников и stg\_ для таблиц Staging, чтобы различать объекты разных слоев.
  - **Подсказка:** не применяйте ограничения ссылочной **целостности** к таблицам-источников и Staging!
  - Результаты практического задания лекции 4 принимаются в файлах форматов .sql или .txt.

## Лекция 4: Практическое задание (4 из 4)

### ОБРАТИТЕ ВНИМАНИЕ!

- Получив спецификации на источники, **не старайтесь** переделать ваши логические и физические модели детального слоя и/или витрин, чтобы добавить все возможные атрибуты из источников! Они не обязательно будут значимы для удовлетворения бизнес-требованиям. Используйте то, что соответствуют вашей ЛМД и добавьте только то, что считаете действительно важным – что вы упустили из виду.
- **Не старайтесь** переделать ваши логические и физические модели детального слоя и/или витрин, если на источниках нет нужной вам сущности или отдельных атрибутов! Реляционная ЛМД гибкая и расширяемая, и она будет готова поддержать гипотетические доработки на источниках. Используйте возможности СУБД для **скрытия** избыточных колонок от конечных пользователей (например, через представления).



Значение атрибута LOCATION внешней таблицы необходимо задавать строго одной строкой (без переносов)!