

Лекция 5. Практическое задание

Курс: разработчик ХД



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



Лекция 5: Практическое задание (1 из 4)

Разработайте **регулярный** ETL/ELT-процесс наполнения данными таблиц интеграционного (детального) слоя ХД из промежуточного слоя, загруженного в предыдущем практическом задании.

1. Начните с написания SQL-запросов чтения таблиц Staging и вставки/обновления детального слоя.

- Для вашего справочника **изменения стоимости товаров** в DDS на основе данных Staging.item_price_history реализуйте обновление таблицы согласно **SCD Типа 2** (интервалы действия записей «с – по»). Начало «времен» = 01.01.2000, конец «времен» = 31.12.2999. Интервалы не должны пересекаться!
- Для прочих **справочных** таблиц определите кандидатов на SCD Тип 0 или SCD Тип 1. Реализуйте соответствующие запросы загрузки и обновления данных.



Обращайте внимание на полноту данных, но не обращайте внимания на их качество! Вы работаете с **синтетическими** учебными данными (пример: Якутск в Астраханской области).



Лекция 5: Практическое задание (2 из 4)

2. Реализуйте **разовую** загрузку **фактических** данных из Staging в детальный слой. В проекте нет отдельного требования по реализации – реализация на ваше усмотрение (самый простой вариант: INSERT-SELECT).
- **Подсказка:** если у вас есть слой ODS и исторические данные для вашего ХД не требуются для генерации суррогатных ключей и т.п., загружайте историю напрямую из Staging в DDS.
- **Обратите внимание:** фактические данные на источнике (схема mwdo_base) для item_inventory, sales_transaction, sales_transaction_line заканчиваются на 30.12.2023. Используйте для загрузки истории ограничение сверху 30.11.2023.
- Данные за декабрь 2023 в дальнейшем нужно будет загружать **инкрементально** (инкремент = 1 день) при помощи пользовательских **функций**.



Лекция 5: Практическое задание (3 из 4)

3. Оберните SQL-запросы наполнения Staging (из предыдущей лекции), промежуточного (если есть) и детального слоя ХД в пользовательские функции. Требования:
- a) Все справочники **промежуточного** слоя должны наполняться одной **общей** функцией.
 - b) Каждая таблица фактов **промежуточного** слоя должна наполняться собственной **отдельной** функцией.
 - c) Каждая таблица **детального** слоя должна наполняться собственной **отдельной** функцией.
 - d) Для всех функций должны быть явно заданы **атрибуты** категории изменчивости и места исполнения.
Атрибуты подбирайте, следя рекомендациям лекции. Явное задание других атрибутов на ваше усмотрение.
 - e) Функции не должны возвращать **результат** в виде набора данных (должны возвращать псевдотип VOID).
 - f) Функции загрузки **справочников** не должны иметь аргументов.
 - g) Функции загрузки **фактов** и событий (как для SRC->STG, так и для STG->DDS) должны иметь один входящий аргумент типа date – дата инкремента.
 - h) Для каждого приемника определите и реализуйте стратегию обновления (вставка только новых строк, очистка предыдущей порции и повторная вставка, MERGE и т.д.). Функции должны поддерживать **многократный** запуск с сохранением корректного результата на приемнике (без задублиивания данных и т.п.).
 - i) Функции загрузки детального слоя должны **очищать** свои источники данных (Staging) по окончании работы.
Реализуйте, но **отключите** эту возможность в функциях до тех пор, пока полностью их не отладите!



Лекция 5: Практическое задание (4 из 4)

- **Подсказка:** Используйте приставки `stg_`, `dds_` и т.п., чтобы отличать функции разных слоев.
 - 4. Определите такую **последовательность** выполнения функций, которая обеспечивает полноту и непротиворечивость данных в детальном слое ХД. **Обязательное условие:** загрузка данных в следующий слой ХД не должна начинаться, пока полностью не завершится загрузка предыдущего.
 - 5. Выполните функции в определенной последовательности.
 - **Обратите внимание:** функции при вызове будут возвращать отсутствие значения `NULL`, согласно параметру `RETURNS`. Контролируйте результат работы функций `SELECT`-запросами.
 - Результаты практического задания лекции 5 принимаются в файлах форматов `.sql` или `.txt`.
- Результаты должны включать в себя:**
- Скрипты исторической загрузки фактов Задания 2;
 - Код создания пользовательских функций с учетом всех требований Задания №3;
 - Запросы вызова функций регулярной загрузки в заданной автором последовательности заданий №№4-5.