

Отчет по курсовой работе

Александр Кричевцов

Цель работы: освоить навыки создания OLTP и OLAP схем, создания процесса ETL, создание визуализации отчета, используя Power BI.

Были реализованы:

- OLTP база данных для работы Bike Rental приложения,
- OLAP база данных для анализа данных и последующего внесения изменений в работу бизнеса Аренды Велосипедов,
- ETL процесс для OLAP базы данных,
- Визуальное представление отчета в Power BI за 2024 год с возможностью фильтрации по кварталам и возможностью просматривать данные по отдельным станциям города.

К отчету также прилагаются файлы со скриптами для OLTP и OLAP, Power BI файл, *.csv файлы, которые были использованы для реализации проекта, OLAP и OLTP схемы (сгенерированы в pgAdmin).

Порядок запуска скриптов:

1. Создание OLTP схемы (файл: **CREATE TABLES OLTP SCRIPT**)
2. Загрузка *.csv файлов (файл: **UPLOAD_CSVs_SCRIPTS**)
3. Создание OLAP схемы (файл: **CREATE TABLES OLAP SCRIPT**)
4. Запуск ETL процесса (файл: **RUN ETL**)

Также можно отдельно открыть файл в Power BI

Aleksandr_Krichevstov_Course_Work_Bike_Rental_App, чтобы увидеть визуальный результат работы.

Скрипты по загрузке *.csv файлов находятся в одном файле. Их необходимо запускать через psql окно. Перед запуском данных скриптов убедитесь, что путь к файлам корректен и доступен для PostgreSQL сервера.

Текстовое описание таблиц OLTP Базы данных

Таблица **users**

- **Назначение:** Хранение данных о зарегистрированных пользователях системы.
- **Первичный ключ:** **user_phone** (уникальный идентификатор пользователя).
- **Атрибуты:**
 - **user_phone:** Номер телефона пользователя (основной уникальный естественный идентификатор).
 - **name:** Имя пользователя.
 - **email:** Электронная почта пользователя (уникальна, но не обязательна).
 - **birth_date:** Дата рождения пользователя.
 - **registration_date:** Дата регистрации пользователя в системе.

- **Ограничения:**
 - `user_phone` обязателен для заполнения, уникален и является первичным ключом.
 - `email` уникален, если указан.

Таблица `bikeTypes`

- **Назначение:** Хранение типов велосипедов и соответствующих им почасовых тарифов.
- **Первичный ключ:** `bike_type`.
- **Атрибуты:**
 - `bike_type`: Тип велосипеда.
 - `hourly_rate`: Почасовой тариф аренды для данного типа велосипеда.
- **Ограничения:**
 - `bike_type` обязателен и уникален.

Таблица `bikes`

- **Назначение:** Хранение информации о велосипедах.
- **Первичный ключ:** `bike_serial_number`.
- **Атрибуты:**
 - `bike_serial_number`: Уникальный серийный номер велосипеда.
 - `bike_type`: Тип велосипеда, ссылается на `bikeTypes(bike_type)`. При удалении типа велосипеда все связанные записи также удаляются (`ON DELETE CASCADE`).
 - `model`: Модель велосипеда.
 - `status`: Состояние велосипеда (например, "Доступен", "В ремонте").
- **Ограничения:**
 - `bike_serial_number` обязателен, уникален и является первичным ключом.

Таблица `stations`

- **Назначение:** Хранение данных о станциях аренды.
- **Первичный ключ:** `station_name`.
- **Атрибуты:**
 - `station_name`: Название станции.
 - `location`: Местоположение станции (уникально).
 - `capacity`: Вместимость станции (максимальное количество велосипедов).
- **Ограничения:**
 - `station_name` обязателен, уникален и является первичным ключом.
 - `location` уникально.

Таблица `rentals`

- **Назначение:** Хранение данных об аренде велосипедов.
- **Первичный ключ:** `rental_id`.
- **Атрибуты:**
 - `rental_id`: Уникальный идентификатор аренды.
 - `user_phone`: Ссылка на пользователя, который взял велосипед в аренду (`users(user_phone)`).
 - `bike_serial_number`: Ссылка на арендованный велосипед (`bikes(bike_serial_number)`).
 - `station_name`: Ссылка на станцию, велосипед был взят в аренду (`stations(station_name)`).
 - `start_time`: Время начала аренды.
 - `end_time`: Время окончания аренды (может быть `NULL` до завершения аренды).
 - `total_price`: Итоговая стоимость аренды (рассчитывается по почасовому тарифу).
- **Ограничения:**
 - При удалении пользователя, велосипеда или станции связанные записи об аренде удаляются (`ON DELETE CASCADE`).

Таблица `maintenance`

- **Назначение:** Хранение информации о техническом обслуживании велосипедов.
- **Первичный ключ:** `maintenance_id`.
- **Атрибуты:**
 - `maintenance_id`: Уникальный идентификатор задачи по техобслуживанию.
 - `bike_serial_number`: Ссылка на велосипед, проходящий обслуживание (`bikes(bike_serial_number)`).
 - `station_name`: Ссылка на станцию, где проводится обслуживание (`stations(station_name)`).
 - `description`: Описание проблемы или выполняемых работ.
 - `date_start`: Дата начала обслуживания.
 - `date_end`: Дата завершения обслуживания (может быть `NULL`, если обслуживание еще не завершено).
 - `cost`: Стоимость обслуживания.
- **Ограничения:**
 - При удалении велосипеда или станции записи об обслуживании удаляются (`ON DELETE CASCADE`).

Таблица `feedbacks`

- **Назначение:** Хранение отзывов пользователей об аренде велосипедов.
- **Первичный ключ:** `feedback_id`.
- **Атрибуты:**

- `feedback_id`: Уникальный идентификатор отзыва.
- `rental_id`: Ссылка на аренду, к которой относится отзыв (`rentals(rental_id)`).
- `user_phone`: Ссылка на пользователя, оставившего отзыв (`users(user_phone)`).
- `rate`: Оценка аренды (от 1 до 5).
- `comment`: Комментарий пользователя (опционально).
- **Ограничения:**
 - Проверка, чтобы оценка (`rate`) была в пределах от 1 до 5.
 - При удалении аренды или пользователя связанные отзывы удаляются (`ON DELETE CASCADE`).

Таблица `payments`

- **Назначение:** Хранение данных об оплатах за аренду.
- **Первичный ключ:** `payment_id`.
- **Атрибуты:**
 - `payment_id`: Уникальный идентификатор оплаты.
 - `rental_id`: Ссылка на аренду, за которую произведена оплата (`rentals(rental_id)`).
 - `amount`: Сумма оплаты.
 - `date`: Дата и время оплаты.
 - `method`: Метод оплаты.
- **Ограничения:**
 - При удалении аренды связанные платежи удаляются (`ON DELETE CASCADE`).

Связи между таблицами

1. **`users` и `rentals`:**
 - Пользователь (`users.user_phone`) связан с арендой велосипеда (`rentals.user_phone`).
2. **`bikeTypes` и `bikes`:**
 - Тип велосипеда (`bikeTypes.bike_type`) связан с конкретными велосипедами (`bikes.bike_type`).
3. **`bikes` и `rentals/maintenance`:**
 - Велосипед (`bikes.bike_serial_number`) связан с арендой (`rentals.bike_serial_number`) и техобслуживанием (`maintenance.bike_serial_number`).
4. **`stations` и `rentals/maintenance`:**
 - Станция (`stations.station_name`) связана с арендой (`rentals.station_name`) и техобслуживанием (`maintenance.station_name`).

5. **rentals** и **feedbacks/payments**:

- Аренда (**rentals.rental_id**) связана с отзывами (**feedbacks.rental_id**) и платежами (**payments.rental_id**).

Текстовое описание таблиц OLAP Базы данных

Таблица **Dim_Users**

- **Назначение:** Содержит информацию о пользователях, включая телефон, имя, электронную почту и даты регистрации.
- **Ключи и ограничения:**
 - **user_id**: Первичный ключ (генерируется автоматически).
 - **user_phone**: Уникальное значение, обязательное для каждого пользователя.
 - Поля **is_current**, **start_date**, **end_date** необходимы для реализации SCD - медленно изменяющиеся измерения.
- **Связи:** Используется в качестве ссылки в таблице фактов **Fact_Rentals**.

Таблица **Dim_bike_types**

- **Назначение:** Содержит список типов велосипедов с их уникальными идентификаторами и почасовыми тарифами.
- **Ключи и ограничения:**
 - **bike_type_id**: Первичный ключ (генерируется автоматически).
 - **bike_type**: Уникальное название типа велосипеда.
- **Связи:** Используется в таблице **Dim_Bikes** для описания типа велосипеда.

Таблица **Dim_Bikes**

- **Назначение:** Содержит данные о велосипедах, включая серийный номер, тип, модель, статус и временные поля для отслеживания изменений.
- **Ключи и ограничения:**
 - **bike_id**: Первичный ключ (генерируется автоматически).
 - **bike_serial_number**: Уникальный серийный номер велосипеда.
 - **bike_type_id**: Внешний ключ, связанный с таблицей **Dim_bike_types** (указывает тип велосипеда).
- **Связи:** Используется в таблицах фактов **Fact_Rentals** и **Fact_Maintenance**.

Таблица **Dim_Stations**

- **Назначение:** Хранит данные о станциях аренды велосипедов, включая их названия, местоположения и вместимость.
- **Ключи и ограничения:**
 - **station_id**: Первичный ключ (генерируется автоматически).
 - **station_name**: Уникальное имя станции.
- **Связи:** Используется в таблицах фактов **Fact_Rentals** и **Fact_Maintenance**.

Таблица **Dim_Time**

- **Назначение:** Временное измерение, содержащее данные о дате, дне, месяце, году, квартале и днях недели. Также указывает, является ли день выходным.
- **Ключи и ограничения:**
 - **time_id:** Первичный ключ (генерируется автоматически).
 - **date:** Уникальная дата.
- **Связи:** Используется в таблицах фактов **Fact_Rentals** и **Fact_Maintenance**.

Таблица **Fact_Rentals**

- **Назначение:** Хранит данные об аренде велосипедов с ссылками на соответствующие измерения (пользователи, велосипеды, станции, время).
- **Ключи и ограничения:**
 - **rental_id:** Первичный ключ (генерируется автоматически).
 - **user_id, bike_id, station_id, time_id:** Внешние ключи, ссылающиеся на таблицы измерений.
- **Агрегированные поля:**
 - **total_rentals_count:** Общее количество аренд.
 - **total_rentals_duration:** Общая продолжительность аренды (в минутах).
 - **total_rentals_price:** Общая стоимость аренды.
- **Связи:** Использует данные из измерений для анализа аренд.

Таблица **Fact_Maintenance**

- **Назначение:** Содержит данные об обслуживании велосипедов, включая ссылки на велосипеды, станции и время.
- **Ключи и ограничения:**
 - **maintenance_id:** Первичный ключ (генерируется автоматически).
 - **bike_id, station_id, time_id:** Внешние ключи, ссылающиеся на таблицы измерений.
- **Агрегированные поля:**
 - **duration_days:** Длительность обслуживания (в днях).
 - **total_bikes_in_maintenance:** Общее количество велосипедов в обслуживании.
 - **total_maintenance_cost:** Общая стоимость обслуживания.
 - **total_maintenance_duration:** Общая продолжительность обслуживания.
- **Связи:** Анализирует затраты и длительность обслуживания.

Общие взаимосвязи и принципы схемы

1. **Dimension:**

- Таблицы измерений ([Dim_Users](#), [Dim_Bikes](#), [Dim_Stations](#), [Dim_Time](#), [Dim_bike_types](#)) содержат данные, описывающие факты, на основе которых происходит анализ.
- Поддержка SCD позволяет фиксировать изменения данных во времени.

2. **Fact:**

- Таблицы фактов ([Fact_Rentals](#), [Fact_Maintenance](#)) содержат количественные показатели (агрегированные поля) и ссылаются на измерения для получения детальной информации.
- Внешние ключи обеспечивают связи между фактами и измерениями.

3. **Аналитика:**

- Схема позволяет анализировать данные по времени, пользователям, станциям и типам велосипедов.
- Поля агрегатов в таблицах фактов облегчают расчет сумм, средних значений и других метрик для анализа.

Вывод: были освоены навыки создания OLTP и OLAP схем, создания процесса ETL, создание визуализации отчета, используя Power BI. Впервые за 5 лет у меня вернулся интерес к работе с Базами Данных и мне понравилось работать с Анализом данных! Буду рад, если сможете порекомендовать литературу, сайты, иные источники для освоения Баз данных и Анализа данных.