



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ Информатика и системы управления

КАФЕДРА Программное обеспечение ЭВМ и информационные технологии

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

НА ТЕМУ:

«Исследование совместимости диалектов SQL для решения задачи по конвертации запросов СУБД MySQL в PostgreSQL»

Студент группы ИУ7-56Б

(Подпись, дата)

Д.В. Варин

(И.О. Фамилия)

Руководитель

(Подпись, дата)

Д.А. Кузнецов

(И.О. Фамилия)

2022 г.

РЕФЕРАТ

Отчет содержит 18 стр., 3 табл., 12 источн.

СОДЕРЖАНИЕ

Введение	6
1 Анализ предметной области	8
2 Различия в диалектах SQL СУБД MySQL и PostgreSQL	10
2.1 Функциональные различия	10
2.2 Синтаксические различия	12
2.3 Различия типов данных	14
2.4 Анализ различий для решения задачи конвертации запросов	15
2.4.1 Конвертация с ограничениями	15
2.4.2 Полная конвертация запросов	15
Заключение	17
Список использованных источников	18

ОПРЕДЕЛЕНИЯ

В настоящей РПЗ применяют следующие термины с соответствующими определениями.

Microsoft SQL Server — система управления реляционными базами данных, разработанная корпорацией Microsoft.

MySQL — свободная реляционная система управления базами данных.

PostgreSQL — свободная объектно-реляционная система управления базами данных.

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

СУБД — Система управления базами данных — совокупность программных средств, обеспечивающих управление созданием и использованием баз данных.

SQL — Structured Query Language — язык структурированных запросов.

NoSQL — Not only Structed Query Language — термин, обозначающий ряд подходов, направленных на реализацию хранилищ баз данных, имеющих существенные отличия от моделей, используемых в традиционных реляционных СУБД с доступом к данным средствами языка SQL.

MVCC — Multiversion Concurrency Control — один из механизмов СУБД для обеспечения параллельного доступа к базам данных, заключающийся в предоставлении каждому пользователю так называемого «снимка» базы, обладающего тем свойством, что вносимые пользователем изменения невидимы другим пользователям до момента фиксации транзакции.

JSON — JavaScript Object Notation — текстовый формат обмена данными, основанный на JavaScript.

XML — Extensible Markup Language — расширяемый язык разметки, используемый для хранения и передачи данных.

ANSI — American National Standards Institute — объединение американских промышленных и деловых групп, разрабатывающее торговые и коммуникационные стандарты.

ISO — International Organization for Standardization — международная организация, занимающаяся выпуском стандартов.

ВВЕДЕНИЕ

Реляционная модель данных нужна для организации данных в виде таблиц (отношений) с помощью связей - полей таблиц, ссылающихся на другие таблицы, которые называются внешними ключами. Сегодня существуют другие модели данных, включая NoSQL, но системы управления реляционными базами данных (СУБД) остаются доминирующими для хранения и управления данными во всем мире [1].

Одними из самых популярных СУБД являются MySQL и PostgreSQL.

MySQL — свободная реляционная система управления базами данных.

PostgreSQL - свободная объектно-реляционная система управления базами данных. Postgres более функциональная СУБД, нежели MySQL хотя и сложнее для использования.

Часто, с развитием проекта, возникает необходимость миграции с одного движка СУБД на другой. Одной из причин является нехватка функциональных возможностей.

PostgreSQL является популярным выбором для функций NoSQL. Она изначально поддерживает большое разнообразие типов данных, включая JSON, hstore и XML [3].

Управление параллельным доступом посредством многоверсионности (MVCC) - это одна из главных причин, почему компании выбирают PostgreSQL. MVCC предоставляет одновременный доступ к базе данных множеству агентов на чтение и запись. Это устраняет необходимость каждый раз блокировать чтение-запись, когда кто-то взаимодействует с данными. Таким образом, значительно повышается эффективность управления СУБД и ее производительность.

MVCC обеспечивает такую функциональность через изоляцию снапшотов. Моментальные снимки (снапшоты) представляют состояние данных в определенный момент времени.

Цель работы — исследовать совместимость диалектов sql для СУБД MySQL и PostgreSQL для решения задачи по конвертации запросов. Таким образом, необходимо выполнить следующие задачи:

- выяснить совместимость СУБД с стандартом SQL;
- выявить различия диалектов PostgreSQL, MySQL;
- определить взаимозаменяемость конструкций;
- описать существующие решения по конвертации запросов MySQL в PostgreSQL;
- выяснить особенности и их преимущества;
- сформулировать выводы.

1 Анализ предметной области

Базы данных — это упорядоченный набор структурированной информации или данных, которые обычно хранятся в электронном виде в компьютерной системе. База данных обычно управляется системой управления базами данных (СУБД). Система управления базами данных — это программа, которая взаимодействует с базой данных [4]. СУБД позволяет контролировать доступ к базе данных, записывать данные, выполнять запросы и любые другие задачи, связанные с управлением базой данных.

База данных может быть любым набором данных, а не только хранящимся на компьютере, СУБД — это программное обеспечение, которое позволяет взаимодействовать с базой данных.

Большинство реляционных баз данных используют язык структурированных запросов (SQL) для управления данными и запросов к ним. Однако многие СУБД используют свой собственный диалект SQL, который может иметь определенные ограничения или расширения. Эти расширения обычно включают в себя дополнительные функции, которые позволяют пользователям выполнять более сложные операции, чем они могли бы делать, используя стандартный SQL.

Стандартный SQL — язык структурированных запросов, поддерживаемый организациями ANSI[6] и ISO[7]. В стандарте используется модульная структура, основная часть стандарта вынесена в раздел "SQL/Foundation". Для современных СУБД существует требование по соблюдению стандарта — они должны поддерживать основной модуль. Однако полной поддержки стандарта не предоставляет ни одна из систем. Так, стандарт SQL 2016 года содержит 177 функций в основном модуле. PostgreSQL реализует 170 из них [8], MySQL — чуть более 140 [9].

По статистике на январь 2022 года [10], MySQL занимает 2 место, Microsoft SQL Server 3 место в мире по популярности среди реляционных СУБД. PostgreSQL находится на 4 месте.

Исходя из больших функциональных возможностей и лучшей поддержкой управления параллельным доступом, переход с MySQL на PostgreSQL является популярной задачей разработки.

В научно-исследовательской работе предстоит выяснить, чем отличаются данные СУБД, совместимы ли они, как перейти с одной на другую.

2 Различия в диалектах SQL СУБД MySQL и PostgreSQL

Для полноценной конвертации запросов следует выявить особенности, которые есть в MySQL, но нет в PostgreSQL, различия в данных СУБД и привести способ адаптирования запросов MySQL в PostgreSQL. Ниже приведены сравнительные таблицы на основе данных, полученных в ходе знакомства с документацией СУБД, а также некоторых статей [11], [12].

2.1 Функциональные различия

Ниже приведена сравнительная таблица функциональных различий между СУБД. В первую очередь учитываются моменты, которые есть в MySQL, но нет в PostgreSQL. При конвертации запросов необходимо обратить особое внимание на описанные ниже различия и адаптировать их под новую систему.

Таблица 2.1 – Функциональные различия СУБД

Возможности	MySQL	PostgreSQL
Работа с датой	CURDATE() CURTIME() EXTRACT()	CURRENT_DATE() CURRENT_TIME() EXTRACT()
Autoincrement тип	auto_increment	serial
Столбец идентификации	Нет	Да
Аналитические функции	Нет	Да
Установка параметра по умолчанию	Константа	Константа и функция
ЯП для хранимых процедур	Синтаксис, удовлетворяющий SQL:2003	Ruby, Perl, Python, TCL, PL/pgSQL, SQL, JavaScript и др.
Индексы	Только индексы: B-Tree, R-Tree, Hash, Inverted	Индексы MySQL + Частичные, Bitmap, Expression, Функциональные
Условные конструкции	IF, IFNULL	CASE
Возможность удалить TEMP TABLE	Да	Нет

Во время поиска функциональных различий MySQL и PostgreSQL были получены следующие выводы: конвертация запросов с MySQL на PostgreSQL проще в реализации, чем наоборот, так как MySQL реализует меньшее количество стандартных SQL функций, чем PostgreSQL. Это накладывает трудности при конвертации запросов, следует выявить конкретные различия в используемых запросах и написать аналоги для PostgreSQL. MySQL не поддерживает следующие возможности:

- объединение таблиц FULL OUTER JOINS;
- пересечение - INTERSECT;
- разность - EXCEPT;
- невозможно выполнить операторы UPDATE или DELETE с SELECT подзапросом из той же таблицы;

- ограничение количества строк LIMIT не поддерживается в подзапросах;
- подзапрос LIMIT с IN/ALL/ANY/SOME;
- частичные индексы, битовые индексы и индексы выражений;
- оконные функции нельзя использовать как часть UPDATE или DELETE;
- ограничение на уникальность строк DISTINCT не поддерживается функциями окна;
- вложенные оконные функции не поддерживаются.

2.2 Синтаксические различия

Ниже приведена таблица синтаксических различий СУБД.

Таблица 2.2 – Синтаксические различия СУБД

MySQL	PostgreSQL	Комментарий
#	–	MySQL использует # для начала строки комментария; PostgreSQL использует двойное тире (стандарт ANSI); MySQL может использовать –, но требует пробела после –, в то время как в PostgreSQL это не обязательно.
'Vasiliy' и "Vasiliy"	'Vasiliy'	MySQL использует 'или " для заключения значений в кавычки. Это не стандарт ANSI для баз данных. PostgreSQL использует для этого только одинарные кавычки. Двойные кавычки используются для обозначения системных идентификаторов, имен полей имен таблиц и т. д.. С помощью SET sql_mode='ANSI_QUOTES' можно заставить MySQL использовать только одинарные кавычки
... WHERE lastname="car"	... WHERE lower(lastname)='car'	PostgreSQL чувствителен к регистру при сравнении строк. MySQL - нет.
'CaR' = 'car' а может и нет	"CaR" <>"car"	Имена баз данных, таблиц, полей и столбцов в PostgreSQL не зависят от регистра, если их не создали с двойными кавычками вокруг имени. В этом случае они чувствительны к регистру. В MySQL имена таблиц могут быть чувствительны к регистру или нет, в зависимости от используемой операционной системы.
'foo' 'bar' означает или	'foo' 'bar' означает конкатенацию	MySQL принимает операторы языка C для логики, SQL требует AND, OR. Решение - использовать стандартные ключевые слова SQL, обе базы данных понимают это.

2.3 Различия типов данных

Ниже приведена таблица сопоставления типов данных MySQL и PostgreSQL.

Таблица 2.3 – Таблица сопоставления типов данных СУБД

MySQL	PostgreSQL	Комментарий
TINY(SMALL)INT MEDIUMINT BIGINT	SMALLINT INTEGER BIGINT	Размер INTEGER в PostgreSQL 4 байта
TINYINT UNSIGNED SMALLINT UNSIGNED MEDIUMINT UNSIGNED INT UNSIGNED BIGINT UNSIGNED	SMALLINT INTEGER INTEGER BIGINT NUMERIC(20)	PostgreSQL соответствует стандарту, в нем нет беззнаковых типов
FLOAT(UNSIGNED)	REAL	PostgreSQL чувствителен к регистру при сравнении строк. MySQL - нет.
DOUBLE	DOUBLE PRECISION	
BOOLEAN	BOOLEAN	В MySQL тип BOOLEAN - это псевдоним для TINYINT(1). PostgreSQL не поддерживает автоматическую конвертацию bool в числовые типы
(TINY,MEDIUM, LONG)TEXT	TEXT	
(VAR)BINARY(n) (TINY)BLOB (MEDIUM,LONG)BLOB	BYTEA BYTEA BYTEA BYTEA	
ZEROFILL	Не поддерживается	
DATE TIME DATETIME TIMESTAMP	DATE TIME [NO TZ] TIMESTAMP [NO TZ] TIMESTAMP [NO TZ]	

2.4 Анализ различий для решения задачи конвертации запросов

При рассмотрении способов конвертации запросов, следует рассмотреть два варианта - с ограничениями и без них. Это необходимо, потому что каждая СУБД имеет свои уникальные функции, которых нет в других СУБД или их функционал не в точности повторяется.

2.4.1 Конвертация с ограничениями

Существующие различия между СУБД препятствуют реализации полной конвертации запросов.

Если бы стандарт SQL был реализован полностью каждой из рассматриваемых СУБД, или же были реализованы одни и те же функции, то одним из способов конвертации запросов(при условии, что СУБД использует только принятые стандартом SQL функций) был бы следующий: так как все используемые функции являются стандартными, то следует перевести синтаксис MySQL на PostgreSQL, тем самым решив задачу конвертации.

Однако использование каждой СУБД исключительно функций из стандарта SQL убирает их преимущества, отличающие каждую СУБД.

2.4.2 Полная конвертация запросов

Для решения задачи конвертации запросов, следует выполнить следующие действия:

- заменить несовместимые типы данных на аналоги в PostgreSQL;
- перевести запросы с синтаксиса MySQL на синтаксис PostgreSQL;
- заменить нестандартные функции на аналоги в PostgreSQL;
- СУБД MySQL нечувствительна к регистру, следует определить общий регистр;
- адаптировать отличающиеся типы.

В результате запросы для MySQL будут преобразованы в корректные запросы для PostgreSQL и задача конвертации будет решена.

ЗАКЛЮЧЕНИЕ

В научно-исследовательской работе были проанализированы различия и на их основе составлен перечень, на основе которого можно конвертировать запрос из СУБД MySQL в PostgreSQL.

Было выявлено, что диалекты несовместимы, так как они имеют отличающиеся функциональные возможности, которые реализованы сверх стандарта SQL. При этом PostgreSQL предоставляет больше возможностей по взаимодействию с данными и имеет более полный функционал. Также в результате исследования удалось выяснить различия между СУБД, определить взаимозаменяемые конструкции и привести сравнительные таблицы,

В результате работы было сформулировано решение для конвертации запроса из MySQL в PostgreSQL с использованием полученной в работе информации.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. DBMS popularity broken down by database model. — URL: https://db-engines.com/en/ranking_categories (online; accessed: 24.12.2021).
2. Стандартная общественная лицензия GNU (GPL). — URL: <https://www.gnu.org/licenses/gpl-3.0.html> (дата обращения: 24.12.2021).
3. Типы данных. — URL: <https://www.postgresql.org/docs/14/datatype.html> (дата обращения: 24.12.2021).
4. База данных. — URL: <https://www.oracle.com/ru/database/what-is-database/> (дата обращения: 24.12.2021).
5. Реляционная система управления базой данных. — URL: <https://clck.ru/a09co> (дата обращения: 24.12.2021).
6. ANSI SQL. — URL: https://docs.oracle.com/cd/B12037_01/server.101/b10759/ap_standard_sql001.htm (online; accessed: 25.12.2021).
7. ISO/IEC 9075-1:2016. — URL: <https://www.iso.org/obp/ui/#iso:std:iso-iec:9075:-1:ed-5:v1:en> (online; accessed: 25.12.2021).
8. PostgreSQL Conformance. — URL: <https://www.postgresql.org/docs/current/features.html> (online; accessed: 21.01.2021).
9. MySQL Standards Compliance. — URL: <https://dev.mysql.com/doc/refman/8.0/en/compatibility.html> (online; accessed: 21.01.2021).
10. DB-Engines Ranking of Relational DBMS. — URL: <https://db-engines.com/en/ranking/relational+dbms> (online; accessed: 25.12.2021).
11. Comparison of different SQL implementations. — URL: <https://troels.arvin.dk/db/rdbms/> (online; accessed: 26.12.2021).
12. PostgreSQL VS MySQL - What Are The Main Differences. — URL: <https://www.softwaretestinghelp.com/postgresql-vs-mysql/> (online; accessed: 26.12.2021).