A Primer to SQL

November 14th, 2019

CSUF; CPSC 311

Professor Peralta

Research Project

Amman Sandhu 889488383

Outline

1 Abstract

This paper aims to give a brief introduction into what SQL and MySQL are, and how each of them relate to each other. We will first overview the history of SQL and how it has become today's SQL, then we will explain some of the design choices for SQL, and how they were arrived at, then we will get an overview of the general syntax of SQL so that we can understand the general structure for any RDBMS, then we will review some of the common criticisms of SQL, then we will look at what MySQL is, how it relates to SQL, and explore what the difference is between the two, then we will attempt to understand how MySQL works in a practical sense while also taking a look at some of its advantages and disadvantages, finally we will look at how to optimize a database in MySQL from a high level and from a low-level perspective. Finally, we will conclude with a brief overview of every topic that has been covered relating to this subject.

2 Introduction

To understand MySQL, we must first take a look at SQL and understand what it is. Structured Query Language (SQL) is a language used for managing data in a Relational Database Management Systems (RDBMS) (Dol 1). Most RDBMS' use SQL as the standard database query language such as MySQL. SQL defines several ways to manipulate data in RDBMS. Some of the things SQL can do is execute queries, insert records, update records, create new databases, and so on. We will be looking into SQL and one particular RDBMS known as MySQL.

3 History

Before diving into more information about SQL, we will briefly take a look at the history of its development. SQL was developed by Donald D. Chamberlin, and Raymond F. Boyce in the 1970s after a paper on the relational model from Ted Codd was published. The first version of it was called Structured English Query Language (SEQUEL), but this name was already trademarked, which lead to the name SQL in 1979. Now the American National Standards Institute (ANSI) and International Organization for Standardization (ISO) have declared SQL as the standard language for database management. To this date, SQL is still the standard language for database management, although most vendors may modify their version of SQL.

4 SQL's Design

SQL was designed to be used on a console. To be able to communicate with an RDBMS, we must use a query. Queries will tell the server what you want but will not disclose how to do it. The first two things the server does when it receives a query, is parse and optimize. After parsing and optimization, the next step is execution, also known as the Query Execution Plan. Queries have a logical and physical processing stage.

4.1 Logical Processing Stage

During the logical processing stage, the order that the different SQL clauses are processed is determined. Each step of a clause would generate a temporary virtual table that is used to perform the next step of processing. This temporary virtual table is only available to the caller of the function.

4.2 Physical Processing Stage

During the physical processing stage, two things occur. First, a QL server engine must generate a query execution plan, and second, the SQL Server storage engine must take the query execution plan as an input and perform actions based upon it that will return the desired output.

4.3 Summary

In summary, the overall stages of an SQL query involve 3 phases: parsing, optimization, and a query execution plan. Splitting up the process into 3 phases allows for SQL to operate more efficiently and effectively.

5 Syntax

SQL, just like any programming language his it's own set of rules and guidelines, also known as its syntax. SQL syntax is made up of keywords, SQL literals, operators, operator precedence, and so on. The most fundamental operations to a database are reading data, inserting new data, updating data, and removing data. SQL provides access to performing each of these operations. With the SELECT, INSERT, UPDATE, and DELETE keywords. It is important to note that SQL is case insensitive such that

UPDATE and update both carry the same meaning. However, MySQL does make differences in table names.

Briefly, SQL is composed of several elements that defined the entire language, such as clauses, queries, statements, predicates, expressions, and whitespace. Clauses would include things such as SELECT, INSERT, and UPDATE. Queries would be what is used to retrieve data from a server-based on certain criteria. Statements are used for control structures and program flow. Predicates are used to create conditions or alter program flow. Expressions are solely scalar values. Finally, whitespace is generally ignored in SQL and used just for code organization.

6 Criticism

Although SQL is a reliable database management system, that does not leave it free from criticism. Some of the criticisms of SQL include a lack of orthogonality, a formal definition, and missing functions. It is argued that the language should provide for each of the classes supported, a constructor function, a method of comparison among objects, a method of assignment to objects, and a selector function (C.J. Date 12). The formal definition of SQL is not precise enough. When there is a host language for SQL, there are many distinctions between the host language and SQL, and many of these distinctions server no purpose (C.J. Date 40).

7 MySQL

MySQL is a free, open-source database management system and offers an excellent introduction to modern database technology. It started out using the mSQL database system but later realized that it could not run fast enough and was not flexible enough, resulting in the creation of MySQL, a new SQL interface that would be more flexible and faster and later become one of the most widely used RDBMS.

7.1 Main features of MySQL

Internally MySQL has been written in C and C++ and tested with a wide range of compilers. Clients are able to connect to MySQL using several protocols, including TCP/IP, named pipes on windows, and on Unix clients can connect with Unix domain socket files. MySQL also can support large databases with up to 5,000,000,000 rows, and there is now support for 64 indexes per table, allowing each index to consist of 1 to 16 columns (Axmark and Widenius 1.3.2). In relation to security, MySQL provides a privilege and password system, also referred to as privileged credential management. Also, when connecting to a server, MySQL provides password security by encryption of all traffic when connecting to a server allowing for safe transfer of data (Axmark and Widenius 1.3.2).

Some of the other features of MySQL include

- Data types FLOAT, DOUBLE, CHAR, TIME, DATE, ENUM, SET, etc. (Axmark and Widenius 1.3.2)
- Operator and function support within a select list (Axmark and Widenius 1.3.2)
- Support for MySQL-specific SHOW statements that are able to get information about databases, tables, and indexes (Axmark and Widenius 1.3.2)
- An EXPLAIN statement that explains how the optimizers resolves a query (Axmark and Widenius 1.3.2).

- It implements in-memory hash tables as the virtual temporary tables (Axmark and Widenius 1.3.2)
- Uses a highly optimized library works fast (Axmark and Widenius 1.3.2)
- APIs exist for other programming languages such as C, C++, Eiffel, Java, Perl, PHP, Python, Ruby, and Tcl (Axmark and Widenius 1.3.2).
- Includes several Client and Utility programs such as mysqldump and mysqladmin (Axmark and Widenius 1.3.2).

8 How MySQL works

MySQL is based on what is known as the client-server model, relying heavily on what is known as MySQL server. MySQL server is a separate program used to handle database instructions and must be installed on the server machine. To send commands to the MySQL server, the MySQL client must be used, which would be installed on a client computer.

8.1 Difference between SQL and MySQL

SQL and MySQL used to be almost exactly the same except that MySQL was able to operate on more platforms. MySQL gives the option to choose a different storage engine for a given table, whereas SQL has its own storage system. MySQL was designed with the idea of being compatible with many systems.

8.2 Advantages

Some of the advantages of MySQL include:

- MySQL uses caching so it will, therefore, run faster since it can handle repeated actions efficiently
- It is known to be simple to use. MySQL is easy to install and setup which makes it a lot more accessible than many other RDBMS
- Portable. MySQL is able to un on every platform without having to install some sort of package or environment
- Open Source. Since MySQL is open source, it undergoes severe vetting allow it to be a more secure
- Inexpensive. Since MySQL is open source under GPL, it is free
- Uses OVER, which is a solution for a problem regarding selecting the top N row for each group

8.3 Disadvantages

Some of the disadvantages of MySQL include:

- Debugging stored procedures. MySQL does not provide a way to debug stored procedures
- Oracle, SQL Server, and PostGRE have a better set of JOIN algorithms which means that JOIN will performed faster on those implementations
- Maintenance. It is difficult to maintain MySQL because the stored procedures generally need a high level of skill that most developers do no possess

- Inefficiency. Although MySQL is theoretically able to handle an infinite amount of data. It has been known to come to a stop when multiple operations are being performed at the same time
- Indexes will take up disk space and slow down the speed of writing queries
    - MySQL uses pointers to inserted rows in a data file, so performance will drop if every time a record changes because indexes will have to be changed.

9 Database Optimization

With a database, it is important to minimize the number of operations taking place to lower CPU and I/O operations.

9.1 High Level

Optimizing a database at a high level involves taking a look at the design of the database such as the tables being structured properly, having indexes in place, using the correct storage engine for each table, and all of the memory areas being used for caching sized correctly (Axmark and Widenius 8.1)

9.2 Low Level

Optimizing a database at the low level involves tinkering with the hardware. Every database may or may not eventually hit its hardware limit. When the hardware becomes the bottleneck, there are usually several sources that cause it.

- Disk Seeks. This refers to the time the storage device takes to acquire a piece of data. The way to optimize this is to distribute the data onto multiple disks (Axmark and Widenius 8.1)
- CPU cycles. Depending on how much memory the system has, the size of the table may be affecting the speed. A simple solution is to use smaller tables (Axmark and Widenius 8.1).

10 Conclusion

MySQL, an implementation of SQL, has become a highly recognized RDBMS software due to its availability, compatibility, and optimization. The creators of SQL, Donald D. Chamberlin and Raymond F. Boyce, researched and developed SQL using the relational model which would provide structure and modularity to SQL. SQL is designed using 3 phases: parsing, optimization, and a query execution plan. SQL also has its own set of rules and guidelines, also known as its syntax. Although SQL is so widely used, it does have its criticisms. MySQL implements SQL to create an RDBMS that is accessible to many people and maintain a fast and flexible design providing a fair trade-off for advantages to disadvantages. There are also many ways that MySQL can be optimized to perform better with large scale databases. Overall, MySQL is an RDBMS implemented with SQL that is highly accessible and flexible.

Works Cited

Dol, Sunita. Gandhmal, Dattatry. “Use of ‘Basic SQL-The Online Beginner’s Guide’ Site to Give Hands on Experience of SQL to Students.” (2018): 14 Nov. 2019.

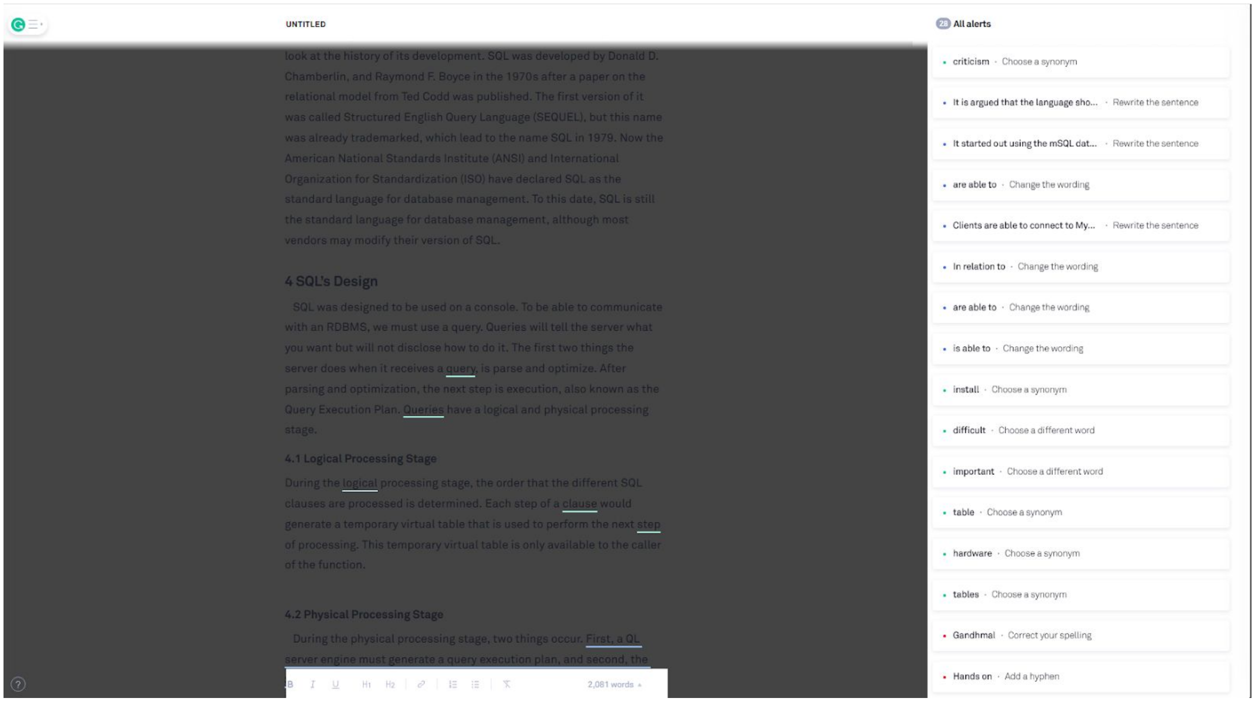
Taipalus, Toni. Siponen, Mikko. Vartiainen Tero. “Errors and Complications in SQL Formulation.” ACM Transactions on Computing Education 18.3 (2018): 14 Nov. 2019.

Date C. J. “A Critique of The SQL Database Language.” (1983): 14 Nov. 2019.

MySQL. (1995). My*SQL 8.0* Retrieved from https://dev.mysql.com/doc/refman/8.0/en/

Oracle Corporation. (1979). Database SQL Tuning Guide Retrieved from https://docs.oracle.com/database/121/TGSQL/tgsql_sqlproc.htm#TGSQL175

Grammarly



look at the history of its development. SQL was developed by Donald D. Chamberlin, and Raymond F. Boyce in the 1970s after a paper on the relational model from Ted Codd was published. The first version of it was called Structured English Query Language (SEQUEL), but this name was already trademarked, which lead to the name SQL in 1979. Now the American National Standards Institute (ANSI) and International Organization for Standardization (ISO) have declared SQL as the standard language for database management. To this date, SQL is still the standard language for database management, although most vendors may modify their version of SQL.

4 SQL's Design

SQL was designed to be used on a console. To be able to communicate with an RDBMS, we must use a query. Queries will tell the server what you want but will not disclose how to do it. The first two things the server does when it receives a query, is parse and optimize. After parsing and optimization, the next step is execution, also known as the Query Execution Plan. Queries have a logical and physical processing stage.

4.1 Logical Processing Stage

During the logical processing stage, the order that the different SQL clauses are processed is determined. Each step of a clause would generate a temporary virtual table that is used to perform the next step of processing. This temporary virtual table is only available to the caller of the function.

4.2 Physical Processing Stage

During the physical processing stage, two things occur. First, a QL server engine must generate a query execution plan, and second, the

- criticism · Choose a synonym
- It is argued that the language sho... · Rewrite the sentence
- It started out using the mSQL dat... · Rewrite the sentence
- are able to · Change the wording
- Clients are able to connect to My... · Rewrite the sentence
- In relation to · Change the wording
- are able to · Change the wording
- is able to · Change the wording
- install · Choose a synonym
- difficult · Choose a different word
- important · Choose a different word
- table · Choose a synonym
- hardware · Choose a synonym
- tables · Choose a synonym
- Gandhmal · Correct your spelling
- Hands on · Add a hyphen

The outline should have working links.

Some errors in the grammarly are related to words that grammarly does not know for SQL and the authors name

Another error is related to the title of a book I cited. It says I need to add a hypen but the book name does not have a hypen.