Enforcement of Partial Referential Integrity in a Database System

Jaja, Azunna Ibimina & E. O. Bennett
Department of Computer Science
Rivers State University,
Port Harcourt.
Nigeria.
jibimina@yahoo.com, bennett.okoni@ust.edu.ng

Abstract
A lack of Referential Integrity in a database can lead Relational Databases to return incomplete data, usually with no indication of an error thus it results in records been lost in the database because they are not returned in queries or reports but the presence of Referential Integrity checks prevents orphan table that no longer has a relationship with its parent table by deleting the orphan row, updating the cells or preventing the action from be performed. However, Simple Referential Integrity cannot check for multiple referential relationships as such this research aims at developing a module that will enforce Partial Referential Integrity checks for MySQL database. The Structured System Analysis and Design Methodology (SSADM) was used in the design and implemented with Java programming language. Meanwhile the tools used for this research are Stored Procedures which encapsulates sets of operations or queries to execute on the database server which is supported by most Relational database systems. Furthermore, MySQL Workbench which is a Graphical User Interface based tool was used to run the queries. Therefore the module will reject or validate data that passes the partial referential integrity check and a database was used to test the module hence the results show that the execution time of the system without the module is less than the execution time of the system with the module because the module ensures that data queried are valid and also, the integrity of the system that uses the module shows that the data in the system meets all constraints enforced and the work done by the database server is minimized.

Keywords*: Database, Integrity, Referential Integrity and Partial Referential Integrity.*

1. Introduction

When the term "referential integrity" is mentioned Relational Database Management Systems (RDBMSs) usually comes to mind meanwhile the term is broader and encompasses databases in general as well as programming languages and modelling. The connotations of "referential integrity" that arise from relational DBMSs are those of an implementation mechanism. However, the deeper meaning is that of dependencies among objects. In this broader sense, referential integrity is an integral aspect of how we think about problems and represent them via models. Consequently, referential integrity must be dealt with regardless of the implementation platform RDBMS, Object Oriented Database Management System (OO-DBMS), or programming language (Blaha & Premerlani, 1998). The adjective "referential" explains the actions that a foreign key performs, referring to a link field in another table. In other words, Referential Integrity is a guarantee that the target it **refers** to will be found. But under simple semantics, only total foreign key values must be matched by some referenced key values. Under partial semantics, total and partial foreign key values must be matched by some referenced key values ie, Partial Referential Integrity is valid if at least one foreign key

column is Null and the rest of the Not Null columns match those of the referenced table. Or all foreign key columns are Not Null and match the referenced table (Blaha, 2010).
Referential Integrity is the relationships between the tables of a database that is the data of one table does not contradict the data of another table (Melton, 2003). Specifically, all foreign key value in a table must have a matching primary key value in the related table. This is the most common type of integrity constraint that is used to manage the relationships between primary and foreign keys.

If a table has integrity constraint, all data in the table must agree to the corresponding rule. And when a Structured Query Language (SQL) statement that modifies data in a table is issued, Oracle Database ensures that the new data satisfies the integrity constraint within the program, so there will be no need to do any checking. Referential Integrity is essential for data processing and data quality (Blaha, 2001). The SQL standard offers different semantics under which referential integrity can be practically enforced. In simple semantics, only total foreign key values must be matched by some referenced key values but in partial semantics, total and partial foreign key values must be matched by some referenced key values. The support for simple semantics is vast and widespread across different database management systems but astonish; partial semantics does not enjoy any native support in known or recognized systems. This research proposes to design a database to ensure its authenticity and integrity.

2. Review of Related Works

Literature Review describes, summarizes, evaluates and clarifies the literature to a selected area of study. It pulls previous Researchers together and explains how it connects to the proposed research. This information could be gotten either on the internet or in academic library. Referential Integrity is a property of data and if it is satisfied, requires all value of one attribute (column) of a relation (table) to exist as a value of another attribute in a different (or the same) relation. In order for referential integrity to exist in a relational database, whichever field that is declared a foreign key in a table can contain either a null value, or only values from a parent tables primary or a candidate key (Thalheim, 1991). In essence, if a foreign key value is used, it must reference a valid and existing primary key in the parent table. For example, lf a record that contains a value referred to by a foreign key in another table is deleted, and then referential integrity is broken or violated. Databases are usually designed to handle large amounts of information or data by inputting, storing, retrieving and managing the information. Databases also use table form which is rows and columns to store their information and every single information entered in a row forms a record. Once these records are created in the database, they can be sorted and manipulated in several ways that are basically restricted by the software been used.
Databases are comparable to spread sheets, but are more powerful than spread sheets because of their ability to manipulate data and it is possible to do so many functions with a database which is difficult to do with a spread sheet.

Referential Integrity

For referential integrity to be enforced in a Relational database management system (RDBMS), deleting the foreign key rows can maintain the integrity, not performing the delete or return an error and not performing the delete. Any method used may be determined by a referential integrity constraint that is defined in a data dictionary. As proposed by the SQL standard (Melton, 2003), the different semantics of referential integrity have not received much attention from neither academia nor practice. (Turker & Gertz, 2001) Observed that there are no database management systems that offer built-in support for

enforcing partial referential integrity while all database management system does has built-in support for enforcing simple referential integrity. (Harder & Reinert, 2008) examined the functional requirements for preserving simple and partial referential integrity.

Partial Referential Integrity

(Memari, 2013) Partial Referential Integrity ensures the consistency of data between database tables. The Partial Referential Integrity SQL standard suggests several semantics to tackle partial information under referential integrity. Simple semantics neglects tuples with nulls and enjoys built-in support by commercial database systems. Partial semantics checks tuples with nulls, but does not enjoy in-built support. Truly, insight is gained in the understanding between cleaner data in partial semantics and the competency of checking simple semantics. The cost for referential integrity checking is calculated for various dataset sizes, indexing structures and degrees of cleanliness. And the cost of partial semantics supersedes that of simple semantics but their performance trends follow similar patterns under growing database sizes. The application of multiple index structures and exploiting appropriate validation mechanisms enhances the efficiency of checking partial semantics. Normally, if the term "referential integrity" is mentioned, only relational DBMSs come to mind. Although the term is actually broader and encircles databases in general as well as modelling and programming languages. The significance of referentia1 integrity" that comes from relational DBMSs are those of an implementation mechanism. Through semantic support for associations, Referential Integrity also arises with the Unified Modeling Language (UML). When constructing UML class models, it is important to show the dependencies of objects categorically with associations instead of hiding them as attributes. (Codd, 1970) Instigated the principles of entity and referential integrity as two fundamental cruxes of the relational model of data and over 100 classes of relational integrity constraints have been scrutinized. Relational database management systems only offers native support for primary keys and foreign keys, which enforce entity and referential integrity, respectively (Thalheim, 1991). (Melton, 2003) The different semantics of referential integrity, as proposed by the SQL standard have not gotten much attention from neither academia nor practice. (Turker & Gertz, 2001) Observed that there are no database management systems that offer built-in support for enforcing or implementing partial referential integrity but virtually all database management system offers in-built support for enforcing simple referential integrity. (Harder and Reinhart, 1996) investigated the functional requirements for keeping simple and partial referential integrity. Indeed, they determined the number and kinds of searches necessary for referential integrity maintenance, without implementation considerations. Their main result was that a combined access path structure is the most appropriate for checking simple semantics, while partial procedures. The avoidance of the use of Match Partial at all was their best advice. Although if required, they advised the use of an index for each of the key columns on the referenced table and a compound index on the foreign key columns of the referencing relation. They also investigate the performance of multi-dimensional access paths by considering grid file structures. Here, the access costs for partial match queries are more expensive remarkably than their suggested index option. The main reason is that grid files retrieve all matching tuples while partial referential integrity requires only one matching tuple.

3. Methodology

The Qualitative research method was adopted for this research because the data were collected through participant observation and interview. The Constructive research method which is a type of Qualitative Methodology was used because it is mostly used in software engineering and computer science research by constructing diagrams, models, plans, etc. The

design methodology used is the Structured Systems Analysis and Design Method (SSADM). This method divides or decomposes a system into modules in order to develop better quality models. The major goal of Design Technique is to produce models of the system that can easily be analysed and reviewed not only by users but by other parties involved in the development of the system.

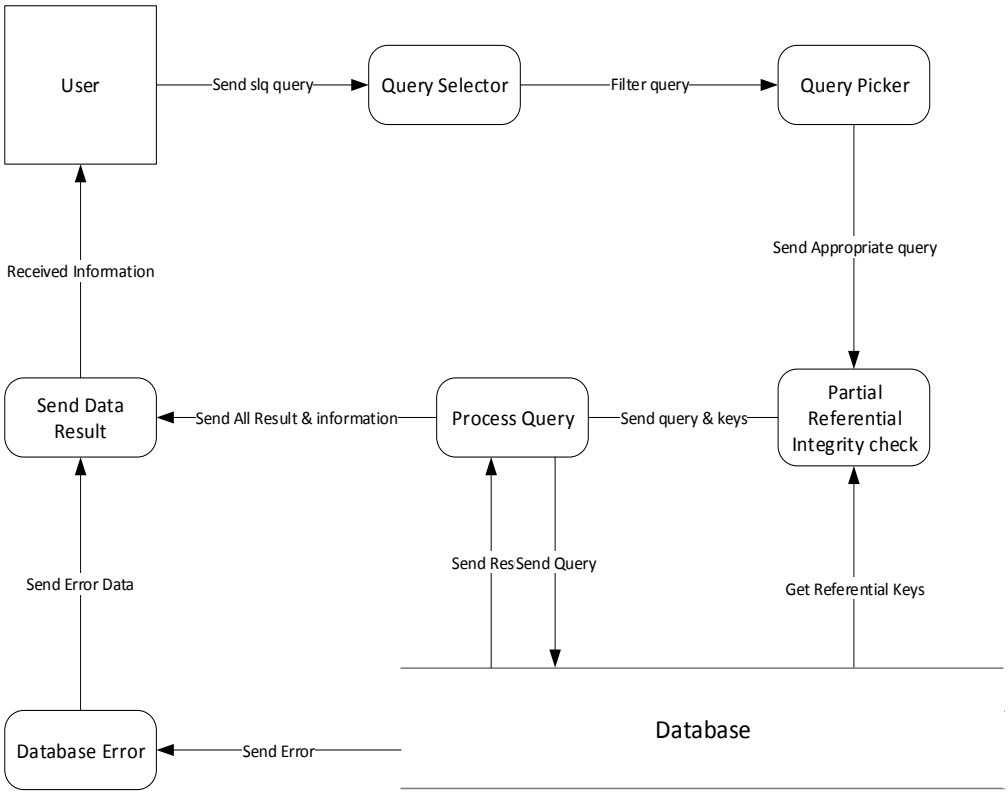

Figure 1: Data flow diagram of Partial Referential Integrity System

When the user sends an SQL request, the query selector selects a single record and sends it to the Query Picker which displays the existing information for the Partial Referential Integrity check of the record from the Database. If the record matches the Database Structure, then the query is processed and the result sent back to the User. Which shows that the result is valid otherwise, the system returns incomplete data by sending an error message.

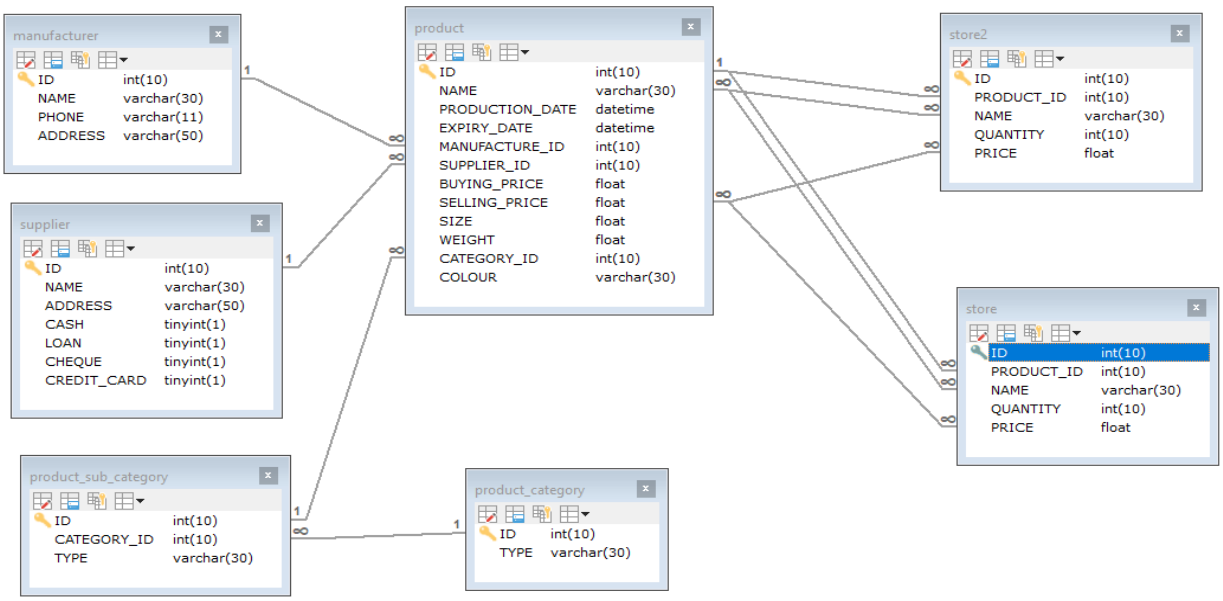

Figure 2: Database Schema of a Partial Referential Integrity System

This is a skeletal structure that shows the logical view of the database and it describes the organisation of data and how it is displayed as database objects such as tables, fields and relationships. The arrows indicate their relationships to the parent table with the Primary keys of the Child Relations as Foreign keys to the Parent Table.

4. Experimental Results

System Requirements

This includes the hardware requirement, software requirement and the setup of the system. These tools are discussed below.

NetBeans which is a Java-based Integrated Development Environment (IDE) was used for this research. The term also refers to the IDEs underlying application platform framework. The IDE was designed to limit coding errors and facilitate error connection with tools such as the NetBeans Find Bugs to locate and fix common Java coding problems and debugger to manage complex code with field watches, breakpoints and execution monitoring for designing the system while the data were stored using MySQL. While Java programming Language was used for the implementation. The System uses the module to Check and maintain the Integrity of the database by strictly rejected Null Tuples and getting correct data as structured in the database.

Result

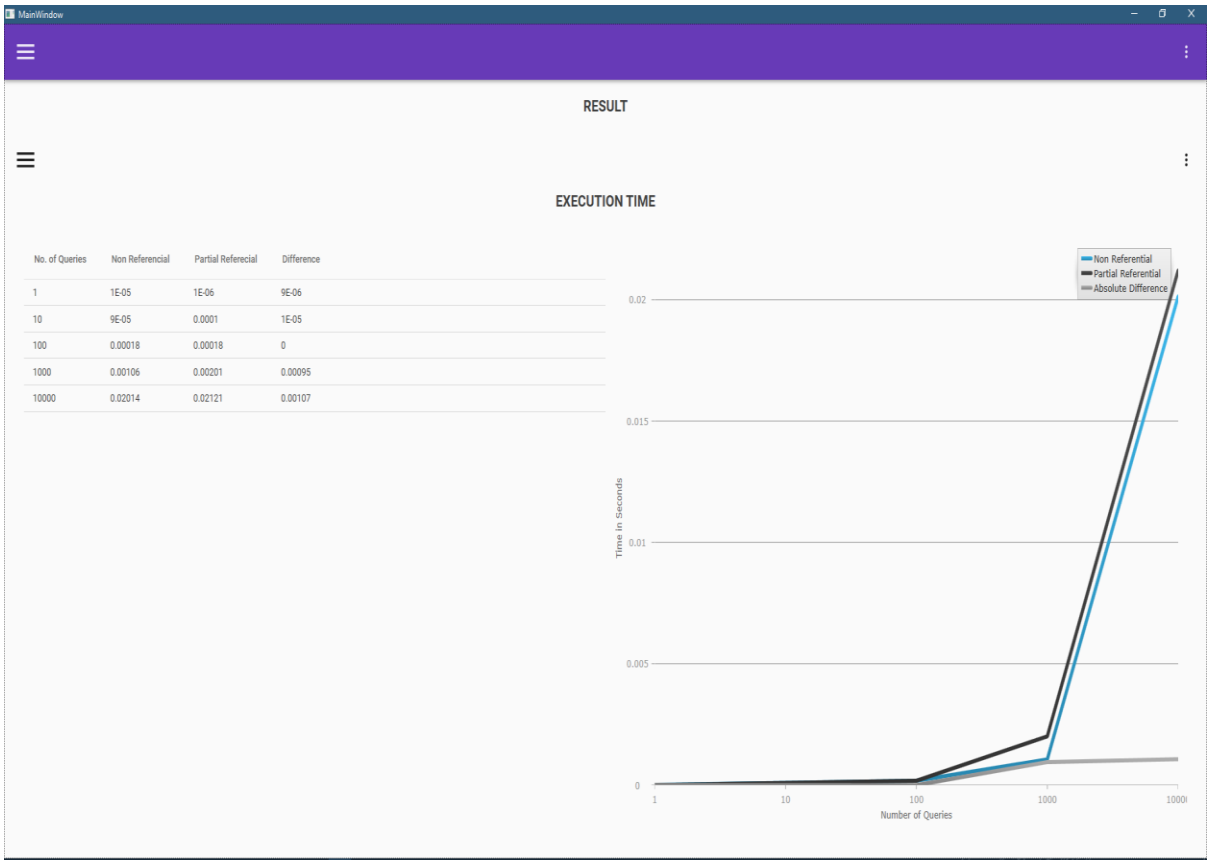

Figure 2: Result showing the execution time of the system

5. Discussion

The graph is plotted on number of queries against time and displayed using a line graph. The plot represents the execution time it takes data to be queried at different times. Hence the result shows that it takes more time for the Partial Referential Integrity module to execute because it checks each data with the database structure to confirm its relationship and validity before displaying the results. Thus, the difference at each time data is queried is seen as the Absolute Difference.

6. Conclusion

Based on the results gotten, Partial Referential Integrity has been tested using both the foreign key and a module to avoid data redundancy in the Database. Despite the efficiency of the system, this Research did not cover all Integrity constraints of a relational database, so future Researchers can also work on the execution time of the database to be more minimal.

7 Future Work

Based on the knowledge gotten from this research, I recommend that future Researchers employ an Execution time of complex queries to be reduced to the barest minimum time to enhance Partial Referential Integrity Checks in Database Systems.

References

Blaha, M, (2001). *A Restospective on Industrial Database Reverse Engineering Projects Part 2.* Eight Working Conference on Reverse Engineering Stuttgart, Germany. 47-53.

Blaha, M. & Premerlani, W. (1998). *Object-Oriented Modeling and Design for Database*

Applications. Upper Saddle River, New Jersey, USA, 140-149.

Codd, E.F., (1970). *Relational Model of Data for Large Shared Data Banks.* Communications of the ACM United Kingdom. 377-387.

Harder, T & Reinert J., (2008). Access Path Support for Referential Integrity in SQL2. *The VLDB Journal, vol 5(3)269-279.*

Melton, J. (2003). *SQL Foundation.* 2nd Edition, ISO Standard, Roberton Publishing, California, USA, 466-469.

Memari, M. (2013). *Static Analysis of Partial Referential Integrity for Better Quality SQL Data.* Benjamin Commy Pub Co, Auckland, New Zealand. 343-347.

Thalheim, B. (1991). Dependencies in Relational Database System. *System Journal, vol 3(5), 46-58.*

Turker, C & Gertz, M. (2001). Semantic Integrity Support in SQL and Commercial (Object) Relational DBMS. . *International Journal for Scientific and Research Publiucation,* Vol 10(4), 1-12.