Assuring Information Systems' Effectiveness Through Data Integrity: Essential Guidelines for Information Systems Databases

Eghosa Ugboma
Department of Computer Science and Mathematics
Florida Memorial College
Miami, Florida 33054, U.S.A.
eugboma@fmc.edu

Abstract

This paper discusses data integrity and how it assists in making information systems reliable systems. The paper also shows that data integrity is one of the primary means of establishing, enforcing, and ensuring the effectiveness of information systems. In addition, the paper addresses the different categories of data integrity that help in ensuring information systems' effectiveness. This class of effectiveness is implemented through information systems databases. Information systems facilitate both the distribution of information (information sharing) and decision-making practices to accomplish users' goals. Information systems are used to (a) fulfill users' information needs, (b) control organizations current activities, and (c) predict future management expectations. The reliability of information systems is in most part controlled by the databases from which the systems' information is generated. The databases on which information systems rely must have integrity for the information systems to be deemed effective to the accuracy and correctness of the information they produce. Databases without data integrity are bound to store erroneous data, therefore making the information systems they serve undependable. Data integrity is a measure for enforcing data consistency, accuracy, and completeness within the databases of information systems. In other words, data integrity establishes and maintains correctness of data in databases. Data integrity ensures that the integrities of the databases are enforced and that users' information requirements are satisfactorily met. Databases must meet the data integrity requirements for the information systems they serve to be productive. Data integrity is all about the correctness of data rather than security measures.

Keywords: database, data file, data integrity, decision making, effectiveness, Information Systems, users' control

1. INTRODUCTION

The major concern of this paper is to communicate how data integrity helps in assuring the effectiveness of information systems and to define each data integrity category that can be applied to information systems databases. Information systems – whether close or open – are systems that manipulate data contained in their databases to provide information needed for both problem-solving tasks and decision-making activities. The databases on which information systems depend must maintain integrity for the information generated through the manipulation of their data to be valuable and useful. Data integrity addresses constraints (rules) governing data of databases rather than the relationships among databases. The constraints help retain the usefulness of data in the databases.

One of the functions of databases' design phase is to model data that are consistent

and that meet the requirement of users – individuals or organizations. For good databases to be developed, their design must contain the necessary details needed for data in the databases to be accurate and complete, thus, keeping the databases in rule-implemented or consistent states. This is where data integrity comes into play especially in determining the correctness of the data in the databases. Data integrity monitors and streamlines data entering the databases of information systems to eliminate data-entry errors and unnecessary duplications of data.

Data must be modeled to meet the needs of users and to produce the required information. Data in databases must have integrity to retain their worthiness and the reliability of the information systems they support.

Data integrity is established during the databases' design phase of information systems' development and the integrity is implemented through constraints. The constraints govern the authenticity of the databases data and administrate the validity of data to ensure correctness of the databases in terms of design, implementation, and manipulation.

2. INFORMATION SYSTEMS

Numerous authors, in the past, have defined information systems in many different fashions. Oz (1998) defines an information system as "aggregate of components that work together to process data and produce information." Stair and Reynolds (2001) define an information system as "set of interrelated elements or components that collect, manipulate, and disseminate data and information and provide a feedback mechanism to meet an objective." Wiederhold (1994) points out that an information system must be able to deal with forward projections – costs and benefits of todays and tomorrows actions into the futures.

The above definitions emphasize that information systems are created and maintained not only to store and retrieve information, but also to convert data into useful information for users' control. For the purpose of this paper, information systems are simply defined as systems, among other capabilities, that enable users to make future predictions of management and administrative variables. This definition gears toward the effective (productive) applications of information systems.

Effectiveness

The fundamental role of information systems is to disseminate information for effective problem-solving tasks and sound decision-making processes. Therefore, this section discusses different information systems characteristics that constitute effectiveness.

The applications of information systems in users' environments can generate required information, increase productivity, improve work performance, and gain control over decision outcomes. With objective-aimed information systems, decision-making processes become more effective, rendering of services becomes better, information interpretation and understanding becomes easier, and sharing of information is improved.

In addition, goal-oriented information systems can provide alternative information to facilitate the making of quality decisions and to reduce risk if used to its capacity, and network of information for effective management.

3. DATABASES

Rob and Coronel (2002) stress that well-designed databases facilitate data management and make the databases valuable information generators. They add that databases that are poorly designed are likely to lead to bad decisions. According to Rob and Coronel statements, one can argue that for databases to maintain their integrity and the integrity of the data they contain, they must be engineered to serve their useful purpose – generating needed information.

Good databases must have data integrity in order for their data to be consistent and valid. Data integrity must be defined during information systems' databases design phase to correctly and consistently support the information systems.

Good information from information systems can lead to sound decisions that might improve the preparedness and actions of users. Information is gotten from data and infor-

mation systems with poorly designed databases are ineffective in terms of the results they generate. The underlying rationale of databases is to store data that are manipulated by information systems to generate needed information. As such, good databases must produce the relevant data that information systems use to generate the correct information that will assist users in making good decisions and solving problems.

Information systems without reliable databases are weak and useless. Creating effective information systems must include good databases that store consistent and stable data for the information systems. Databases are the driving force of information systems, and as a result, they must maintain data integrity for the information systems that they support to generate the necessary information for effective decision-making practices. The validity and the integrity of the data within databases make databases credible.

4. DATA INTEGRITY

Data integrity provides both preventive and detective measures that ensure that valid data that fall within the data validation range of specified data types enter the databases. The preventive measure makes certain that valid data are entering the information systems databases while the detective measure guarantees that data fall within the validation range defined for the databases.

Kroenke (2003) defines data integrity as "the state of a database in which all constraints are fulfilled." In his article entitled Ensuring Data Integrity Through the use of Prevent and Detect Controls, Part I, Wu (2004) defines data integrity as "information that adheres to a strict standard of value and completeness." Pratt and Adamski (2002) point out that data integrity constraints assist to ensure the accuracy and consistency of individual field values of a database table.

For the purpose of this paper, data integrity is defined as rules that must be applied to and maintained by databases for the data contained in the databases to be justifiable and reliable to assure their usefulness.

Data integrity uses constraints to affirm restrictions, as required, to data contained in databases. In other words, data integrity asserts rules that data in databases must obey. These rules are implemented through the databases of information systems.

5. CATEGORIES OF DATA INTEGRITY

Data integrity categories discussed in this article includes key integrity (entity, referential), unique integrity, not null integrity, default integrity, data type integrity, domain integrity, and format integrity. These categories assist in assuring that information systems are effective as the data contained in the information systems databases are complete and consistent. This helps put the databases in consistent or perfect states.

Data Type Integrity

The data type integrity defines the types of data, such as date or numeric, columns of databases data files (users' data) must accept and store. The data entered for columns must meet the data types' definitions for the columns and data that are not of the types defined are rejected. For example, date columns must allow entry in date formats that are recognized by databases. The data type integrity is mandatory for columns in databases data files.

Default Integrity

The default integrity allows beginning values or initial values to be established for columns of databases data files (users' data). Beginning values are permitted to remain as the columns data when no data entries operations are performed on the columns. Data entered as default values must satisfy the data type requirements of the columns. The default integrity specifies the default values, if any, columns in data files must store at the time of databases' creation or their alteration.

Domain Integrity

The domain integrity declares the permissible values and controls the legitimacy of values that must enter columns of data files (users' data) even if the values satisfy the columns data types. The domain integrity allows only acceptable values that fall within

specified data ranges or lists of values of the defined data types in columns. For example, columns that are set up to accept numeric data might also be declared to accept values ranging from 10 to 25 only or to accept list of values such as 2.0, 3.25, and 7.50. Numeric values outside the specified ranges or lists are rejected even though the values meet the requirements of the columns data types. The domain integrity is used at the column level of data files to establish data consistency.

Entity Integrity

The entity integrity is one of the two integrities that constitute the key integrity. The entity integrity is implemented through the primary keys of databases data files and it ensures (a) that values in primary keys columns must not be duplicated, except where the columns are part of composite primary keys, and (b) that the columns must never be null. Entity integrity is used to distinguish one row (record) from another in data files, thereby making each row unique.

The entity integrity implicitly establishes both the unique integrity and the not null integrity for columns that serve as the primary keys of databases data files. The entity integrity is used at the row level of data files to establish data consistency.

Format Integrity

The format integrity specifies the types of input and output formats for data files columns. For example, the format integrity might direct that alphabets of character data be entered into columns in uppercases. For numeric columns, the format integrity might direct that monetary data be displayed using currency symbols such as the "$" symbol.

Not Null Integrity

The not null integrity defines which columns must always receive data when data entry operations are performed on databases data files. This integrity ensures that values must be entered into columns. The integrity also permits the columns to accept duplicate values.

Referential Integrity

The referential integrity is the other integrity that constitutes the key integrity. The referential integrity is implemented through the foreign keys of databases data files and the foreign keys columns must either contain data or be null. If the foreign keys columns contain data, the data must match some values in the primary keys columns of the data files (users' data) the foreign keys are referencing.

The referential integrity implicitly establishes the link that exists among data files in databases. If data files are to be associated in databases, columns that serve as foreign keys must exist either in all the data files or in some of the files. The referential integrity is used at the data file level of databases to establish data consistency.

Unique Integrity

The unique integrity defines that columns of data files must either contain values that are not duplicated or store null values. The unique integrity directs databases to reject values that have previously been entered into the specified columns.

6. DATA INTEGRITY VIOLATIONS

Although data integrity guarantees correctness of data within databases to ensure databases' worthiness, there might be situations where complex data integrity definitions violate the integrity rules defined for particular columns. This section explains the combination of data integrity categories that might cause such violations or conflicts.

Setting default values for primary keys columns or for unique columns for databases data files are permissible. Allowing such columns to store the default values more than once for multiple rows cause conflicts and violate the entity integrity or the unique integrity. Both integrities determine that values stored in columns defined with entity integrity or unique integrity must not be duplicated. Assigning default values to primary keys columns or to unique columns are not beneficial and do not improve databases performances. Instead the operations result in the duplications of values and therefore violate the already defined integrity.

Assigning default values to foreign keys columns are also permitted in databases. If the values do not match some values in the primary keys columns of the data files (users data) that the foreign keys columns reference, conflicts and violations of referential integrity occur. Again such operations do not enhance databases performances as matches in values of the primary keys columns and the foreign keys columns are non-existence. Referential integrity indicates that non-null values in foreign keys columns of databases data files must be values that exist in the primary keys columns of the data files the foreign keys columns are referencing.

7. CONCLUSIONS

This paper communicates how the effectiveness of information systems can be maintained through the use of data integrity. Information systems can facilitate decision-making actions if the data contained in their databases are accurate, consistent, and complete. Databases without integrity cause failures to the information systems they support in terms of the data and information generated.

Data integrity is both preventive and detective measures that assure databases generate the data required by the information systems they support. Data integrity assists in reducing the chance of entering erroneous data into information systems databases and prevents information systems from generating unreliable information. Data integrity helps to make information systems dependable.

Data integrity is an important factor when considering information systems' effectiveness. This is due to the fact that it assists in putting information systems in objective-oriented dependable states.

Therefore, it should be noted that information systems' effectiveness is not only achieved through security measures such as password authentications and user identification encryptions, but also through the systems databases' integrity.

REFERENCES

Center for Technology in Government – University at Albany/SUNY, 1997, A Practical Guide to State-Local Information systems. http://www.ctg.albany.edu /resources/pdfrpwp/iis1.pdf.

Connolly, M. Thomas and Carolyn E. Begg, 2002, Database Systems: A Practical Approach to Design, Implementation, and Management (3rd ed.). Addison-Wesley, New York.

Elmasri, Ramez and Shamkant B. Navathe, 2002, Fundamental of Database Systems (3rd ed.). Addison-Wesley, New York.

Frick, D. R. & Co, 2004, Data Integrity. http://www.frick-cpa.com/ss7/Theory _DataIntegrity.asp.

Kroenke, M. David, 2003, Database Concepts (2nd ed.). Prentice Hall, New Jersey.

Modell, Martin, 2002, The Various Types of Information Systems Analysis Projects. http://www.dai-sho.com/pgsa2 /pgsa02.html.

Oz, Effy, 1998, Management Information systems. Course Technology, Cambridge.

Pratt J. Pratt and Joseph J. Adamski, 2002, Concepts of Database Management (4th ed.). Course Technology, Cambridge.

Rob Peter and Carlos Coronel, 2002, Database Systems – Design, Implementation, and Management (4th ed.). Course Technology, Cambridge.

Shasha, Dennis and Philippe Bonnet, 2003, Database Tuning: Principles, Experiments, and Troubleshooting Techniques. Morgan Kaufmann Publisher, San Francisco.

Stair, M. Ralph and George W. Reynolds, 2001, Fundamentals of Information Systems. Course Technology, Cambridge.

Swanson, Marianne and Barbara Guttman, 1996, Generally Accepted Principles and Practices for Securing Information Sys-

tems. http://csrc.nist.gov/publications /nistpubs/800-14/800-14.pdf.

Wiederhold, Gio, 2004, From Data Engineering to Information Engineering. http://www-db.stanford.edu/pub/gio /1994/inf-eng-abstract.html.

Wu, Jonathan, 2004, Ensuring Data Integrity through the Use of Prevent and Detect Controls, Part I. http://www.evaltech .com/wpapers/ensuringdataintgrity.htm.