We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,300

Open access books available

172,000

International authors and editors

190M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Introductory Chapter: Data Integrity and Quality

Santhosh Kumar Balan

1. Introduction

The significance of data could be employed to escalate the income incision expenses or both. The data integrity related software is one of the numbers of an assessment tool for evaluating the information. It permits the users to assess the confidentiality of the information is rising regularly. Due to these diverse analyses are performed on confidentiality safeguarded data integrity the design of fresh scheme permits the mining data while attempting to safeguard the confidentiality of the users. Many of these schemes are intended for user's confidentiality but still, others are intended on the confidentiality of the organization. It is broadly termed as data exploration in repositories which is the significant mining of hidden, conventionally indefinite and probably needful data from the information in the repositories. The data exploration is required to make logic and usage of information. Even though the data exploration in repositories are regularly regarded as replacements, and it is crucially a segment of the data exploration process. Normally the data integrity, for instance, the information or the data exploration is the process of assessing information from diverse viewpoints and abstracting them into needful data from diverse angles, classification and abstraction of recognized association.

Precisely it is the process of locating synchronization or prototypes among dozens of domains in immense relational repositories. Though, it is relatively fresh domain the technology still remains the same. The organizations are making use of potent computers for straining through immense of supermarket scanner information and assess market trends over years. Therefore, the constant improvements in processing the computational power, disk storages and arithmetic software are significantly escalating the precision of the assessment while governing the expenses. The data integrity of fresh and potential prototypes in immense information sets is a domain in the ignition. The major feature is to enhance the safety for instance identification of interferences. The subsequent feature is the possible safety risks imposed on the opponent has its abilities. The confidentiality related risks have gained the care of multimedia, legislators, government firms, trade and confidential promoters. The data integration edges resume to develop, there are diverse prevailing issues examined. The assembly might choose to contemplate in relation to the planning and inaccuracy. These problems comprise but are not restricted to the quality of the information, interoperability and stealing of assignment and confidentiality. Along with the added features, the technology-based abilities are crucial where other features might also govern the victory of the results [1].

2. Data quality

The Data Quality which we are maintaining in the data integrity comprises the usage of cultured information assessment tools to explore conventionally unfamiliar, lawful prototypes and associations in immense information sets. These tools could comprise arithmetic prototypes, numerical routines and machine learning schemes. Therefore, it comprises gathering, arranging and preserving information which comprises assessment and forecast. It could be accomplished on information denoted in measurable, text-based, visual, image or hypermedia patterns. The applications could make use of selective metrics to assess the information. They comprise relationship orders or route assessment, categorization, grouping and estimation. Diverse firms gathers and perfectly voluminous extents of information. The schemes could be used quickly on the conventional software and hardware platforms for improving the values of the prevailing resources and could be combined with the fresh products and systems due to their availability online. The repositories and information repositories are becoming more and more attractive and make use of the immense volume of information which requires being assessed efficiently. The data exploration in repositories could be entailed as the exploration of attractive, hidden and conventionally unfamiliar data from the immense repositories.

The data integrity repositories might be reasonable instead than a physical subgroup of the information depository offered that the information depository Database Management Systems which could aid the supplementary supply requirements of information mining. If it is possible, it is better to leave a distinct it's repository. In general usage, the terms data integrity and data quality are used interchangeably. However, they often have few significant differences between each other. Data integrity validates that the data and ensures that it remains unaltered throughout its life cycle. Numerous operations such as storing, retrieving, updating, etc., are performed very often on data. The techniques ensure that, irrespective of all the operations performed, the data is maintained, just as how it was inputted. The data encryption, backup, access controls, validation are few practices that maintain data integrity. On the other hand, data is labeled as quality data if it is relevant and complete and is suitable to the intended purpose. As per the standards, data quality is defined in three different perspectives such as from a consumer's perspective, from a business perspective, and from a standard-based perspective.

The Data quality is a multi-layered problem which symbolizes the immense disputes in data integrity. The quality of information states the precision and fullness of the information. The quality of the information could also be bothered based on the framework and reliability of the information which is being assessed. The existence of redundant reports, the missing information policies, the properness of revisions and human faults could crucially influence the efficiency of more intricate it's schemes which are delicate to the elusive variations which might prevail over the information. In order to enhance the quality of information, it is roughly mandatory to refine the information which comprises the eradication of redundant reports, standardizing the values employed to symbolize data in the repository [2].

3. Confidentiality safeguarded data integrity

In relation to the quality of the information, the problem is the synchronization of various repositories and its software. The synchronization denotes the capability of a computational system or information to work with other systems or the

information employing usual principles or operations. Synchronization is a crucial segment of the immense determination to enhance the linked association and data distribution using e-government and native confidentiality edges. For Data Integrity, the synchronization of repositories and software is crucial to allow the exploration and assessment of diverse repositories consequently and aids in assuring the comforts of its actions of diverse firms. It attempts to partake the benefits of the prevailing inherited repositories or that are opening the initially shared attempts with other firms or extents of government might practice synchronization issues. Likewise, as the firms progress onward with the generation of fresh repositories and data distribution attempts, they will require resolving the synchronization problems during their phases of implementation to better assure the efficiency of their schemes [3–6].

The Data Integrity has influenced crucial attention, especially over the past years with its immense varieties of applications. In terms of safety concerns, it is considered advantageous in challenging diverse sorts of risks to the computational system. Therefore, the similar methodologies could be employed to generate probable risks related to safety. Moreover, the collection of data and assessment attempted by the government firms and trade elevates the anxieties related to confidentiality which inspires the confidentiality safeguarding in data integrity. The feature of confidentiality safeguarding is that it shall be capable to make use of various schemes without monitoring the values of private information. But still, the disputes are being explored. An additional feature is that the use of its schemes, the opponent can gain access private data which cannot be attained using request tools risking the confidentiality of peoples. Diverse preliminary analyses are available in confidentiality safeguarded Data Integrity. Conversely, there are several problems which require more analyses in the conception of data integrity from both confidentiality and safety initiatives [7, 8].

4. Applications

The analytical illustration offers trade buying system greatest of the products from the preceding year one could forecast the level of products which requires goods for the impending periods. The authentication could verify on the ailments such as viral with the exception that it is probable to locate the acknowledgment and withdrawal identification in terms of scams. It is employed for diverse objectives in both the private and public firms. The organization like banking, insurance, medicals and purchasing normally make use of data integrity to minimize the expenses, improve analysis and escalates trades. Consider the insurance and banking organization employing data integrity applications for identifying scams and aid in threat evaluation. The usage of user-related information gathered over the present periods the firms could design prototypes which forecasts the threats prevailing to the users in terms of credits or regarding the privileges during accident might be false and shall be inspected more carefully.

The medical society roughly makes use of data integrity to aid the analysis of the efficiency of the scheme or medicines. The medical firms make use of data integrity of the chemical substances and genetic components to aid the governance of studies on fresh management for ailments. The vendors could employ the data gathered using attraction programs to evaluate the efficiency of choosing the items and position related choices, voucher offers and the frequency of items bought regularly. The firms like telephone service suppliers and music clubs could make use to generate a segment assessment to examine which users are probable to continue as users and which ones are probably to migrate to the opponent [9, 10].

IntechOpen

IntechOpen

Author details

Santhosh Kumar Balan

Department of Computer Science and Engineering, Guru Nanak Institute of Technology, Hyderabad, Telangana, India

*Address all correspondence to: bsanthosh.csegnit@gniindia.org

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

- [1] Tina Hui, Ensuring Data Quality and Integrity in Financial Management Reporting for Medical Imaging Operations, *Journal of Medical Imaging and Radiation Sciences*, Volume 50, Issue 3, Supplement, 2019, Page S13, ISSN 1939-8654, doi:10.1016/j.jmir.2019.06.034.
- [2] T. Hongxun et al., "Data quality assessment for on-line monitoring and measuring system of power quality based on big data and data provenance theory, " 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 2018, pp. 248-252, doi:10.1109/ICCCBDA.2018.8386521.
- [3] I. Taleb, M. A. Serhani and R. Dssouli, "Big Data Quality: A Survey," 2018 IEEE International Congress on Big Data (BigData Congress), 2018, pp. 166-173, doi: 10.1109/BigDataCongress.2018.00029.
- [4] MI Svanks, Integrity analysis: Methods for automating data quality assurance, *Information and Software Technology*, Volume 30, Issue 10, 1988, Pages 595-605, ISSN 0950-5849, [https://doi.org/10.1016/0950-5849\(88\)90116-4](https://doi.org/10.1016/0950-5849(88)90116-4).
- [5] Alan R. Simon, Steven L. Shaffer, Chapter 9 - Data Quality and Integrity Issues, Editor(s): In the Morgan Kaufmann Series in Data Management Systems, Data Warehousing And Business Intelligence For e-Commerce, Morgan Kaufmann, 2002, Pages 193-208, ISBN 9781558607132, <https://doi.org/10.1016/B978-155860713-2/50012-7>.
- [6] Nikita R. Nikam, Priyanka R. Patil, et. al., "Data Integrity: An Overview", *International Journal of Recent Scientific Research*, Vol. 11, Issue, 06 (A), pp. 38762-38767, June, 2020
- [7] World Health Organization, working document QAS/19.819, Guideline on Data Integrity, October 2019, 14-16.
- [8] Boritz, J. IS practitioners views on Core concepts of information integrity, *Int. J. Account. Inf. Syst*, Elsevier.
- [9] <https://www.veracode.com/blog/2012/05/what-is-dataintegrity>.
- [10] <https://digitalguardian.com/blog/what-data-integrity-dataprotection-101>