August 2025: TOP 10 Read Articles in Database Management Systems Research Articles

International Journal of Database Management Systems (IJDMS)

*** WJCI Indexed***

https://airccse.org/journal/ijdms/index.html

Citations, h-index, i10-index

Citations 3689 h-index 29 i10-index 81

MAPPING COMMON ERRORS IN ENTITY RELATIONSHIP DIAGRAM DESIGN OF NOVICE DESIGNERS

Rami Rashkovits[1] and Ilana Lavy[2]

[1]Department of Management Information Systems, Peres Academic Center, Israel

[2] Department of Information Systems, Yezreel Valley College, Israel

ABSTRACT

Data modeling in the context of database design is a challenging task for any database designer, even more so for novice designers. A proper database schema is a key factor for the success of any information systems, hence conceptual data modeling that yields the database schema is an essential process of the system development. However, novice designers encounter difficulties in understanding and implementing such models. This study aims to identify the difficulties in understanding and implementing data models and explore the origins of these difficulties. This research examines the data model produced by students and maps the errors done by the students. The errors were classified using the SOLO taxonomy. The study also sheds light on the underlying reasons for the errors done during the design of the data model based on interviews conducted with a representative group of the study participants. We also suggest ways to improve novice designer's performances more effectively, so they can draw more accurate models and make use of advanced design constituents such as entity hierarchies, ternary relationships, aggregated entities, and alike. The research findings might enrich the data body research on data model design from
the students' perspectives.

KEYWORDS

Database, Conceptual Data Modelling, Novice Designers

REFERENCES

[1] Moody, D. L. & Shanks, G. G. (1998). "What Makes a Good Data Model? A Framework for Evaluating and Improving the Quality of Entity Relationship Models," Australian Computer Journal, vol. 30, pp. 97-110

[2] Codd, E. F. (1970). A relational model of data for large shared data banks. Communications of the ACM, 13(6), 377-387.

[3] Codd, E. F. (1979). Extending the database relational model to capture more meaning. ACM Transactions on Database Systems (TODS), 4(4), 397-434.

[4] Frederiks, P. J., & Van der Weide, T. P. (2006). Information modeling: The process and the required competencies of its participants. Data & Knowledge Engineering, 58(1), 4-20.

[5] Chen, P. P. S. (1976). The entity-relationship model—toward a unified view of data. ACM Transactions on Database Systems (TODS), 1(1), 9-36.

[6] Teorey, T.J., Yang, D., and Fry, J.F. (1986). A logical design methodology for relational databases using the extended entity-relationship model. Computing Surveys, Vol. 18, No. 2, pp. 197-222.

[7] Ram, S. (1995). "Deriving Functional Dependencies from the Entity Relationship Model," Communications of the ACM. Vol. 38, No. 9, pp. 95-107.

[8] Batra, D. and Antony, S (1994). Novice errors in database design. European Journal of Information Systems, Vol. 3, No. 1, pp. 57-69.

[9] Antony, S. R., & Batra, D. (2002). CODASYS: a consulting tool for novice database designers. ACM Sigmis Database, 33(3), 54-68.

[10] Batra, D. (2007). Cognitive complexity in data modeling: Causes and recommendations. Requirements Engineering 12(4), 231–244.

[11] Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Allyn & Bacon.

[12] Biggs, J. B., & Collis, K. F. (1982). Evaluation the quality of learning: the SOLO taxonomy (structure of the observed learning outcome). Academic Press.

[13] Bloom, B. S. (1956). Taxonomy of educational objectives. Vol. 1: Cognitive domain. New York: McKay, 20, 24.

[14] Rashkovits, R. & Lavy, I. (2020). Students difficulties in identifying the use of ternary relationships in data modeling. The International Journal of Information and Communication Technology Education (IJICTE), Vol. 16, Issue 2, 47-58.

[15] Balaban, M., & Shoval, P. (1999, November). Resolving the ―weak status‖ of weak entity types in entity-relationship schemas. In International Conference on Conceptual Modeling (pp. 369-383). Springer Berlin Heidelberg.

[16] Or-Bach, R., & Lavy, I. (2004). Cognitive activities of abstraction in object-orientation: An empirical study. The SIGCSE Bulletin, 36(2), 82-85.

[17] Liberman, N., Beeri, C., Ben-David Kolikant, Y., 2011). Difficulties in Learning Inheritance andPolymorphism. ACM Transactions on Computing Education, 11, (1), Article 4, 1-23.

[18] Chick, H. (1998). Cognition in the formal modes: Research mathematics and the SOLO taxonomy. Mathematics Education Research Journal, 10(2), 4-26.

[19] Huang, I. L. (2012). An empirical analysis of students' difficulties on learning conceptual data modeling. Journal of Management Information and Decision Sciences, 15(2), 73.

[20] Leung, F., & Bolloju, N. (2005). Analyzing the quality of domain models developed by novice systems analysts. In System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on (pp. 188b-188b). IEEE.

[21] Kesh, S. (1995). Evaluating the quality of entity relationship models. Information and Software Technology, 37(12), 681-689.

[22] Topi, H., Valacich, J. S., Wright, R. T., Kaiser, K., Nunamaker Jr, J. F., Sipior, J. C., & de Vreede, G. J. (2010). IS (2010): Curriculum guidelines for undergraduate degree programs in information systems. Communications of the Association for Information Systems, 26(1), 18.

[23] Lindland, O. I., Sindre G., and Solvberg A., (1994). "Understanding quality in conceptual modeling," IEEE Software, vol. 11, pp. 42-49.

[24] Taylor, S.J. & Bogdan, R. (1998). Introduction to Qualitative Research Methods. New York: John Wiley & Sons.

[25] Neuendorf, K. A.(2002). The Content Analysis Guidebook. Thousand Oaks, CA: Sage Publications.

[26] Batra, D., & Davis, J. G. (1992). Conceptual data modelling in database design: similarities and differences between expert and novice designers. International journal of man-machine studies, 37(1), 83-101

ALGORITHM FOR RELATIONAL DATABASE NORMALIZATION UP TO 3NF

Moussa Demba

Department of Computer Science & Information, Aljouf University
Sakaka, Kingdom of Saudi Arabia

ABSTRACT

When an attempt is made to modify tables that have not been sufficiently normalized undesirable side effects may follow. This can be further specified as an update, insertion or deletion anomaly depending on whether the action that causes the error is a row update, insertion or deletion respectively. If a relation R has more than one key, each key is referred to as a candidate key of R. Most of the practical recent works on database normalization use a restricted definition of normal forms where only the primary key (an arbitrary chosen key) is taken into account and ignoring the rest of candidate keys.

In this paper, we propose an algorithmic approach for database normalization up to third normal form by taking into account all candidate keys, including the primary key. The effectiveness of the proposed approach is evaluated on many real world examples.

KEYWORDS

Relational database, Normalization, Normal forms, functional dependency, redundancy

For More Details : http://airccse.org/journal/ijdms/papers/5313ijdms03.pdf

Volume Link : https://airccse.org/journal/ijdms/current2013.html

REFERENCES

[1] Thomas, C., Carolyn, B. (2005) Database Systems, A Practical Approach to Design, Implementation, and Management, Pearson Fourth edition .

[2] Bahmani A., Naghibzadeh, M. and Bahmani, B. (2008) "Automatic database normalization andprimary key generation", Niagara Falls Canada IEEE.

[3] Beynon-Davies, P. (2004) Database systems, Palgrave Macmillan, Third edition, ISBN 1–4039—1601–2.

[4] Dongare,Y. V., Dhabe,P. S. and Deshmukh, S. V. (2011) RDBNorma: "A semi-automated tool forrelational database schema normalization up to third normal form", International Journal of Database Management Systems, Vol.3, No.1.

[5] Vangipuram, R., Velputa, R., Sravya, V. (2011) "A Web Based Relational database design Tool toPerform Normalization", International Journal of Wisdom Based Computing, Vol.1(3).

[6] Codd, E.F. (1972) "Further normalization of the data base relational model", In Database Systems,Courant Inst. Comptr. Sci. Symp. 6, R. Rustin, Ed., Prentice-Hall, Englewood Cliffs, pp. 33—64.

[7] Elmasri, R., Navathe, S.B. (2003) Fundamentals of Database Systems, Addison Wesley, fourthEdition.

[8] Date, C.J. (2004) Introduction to Database Systems (8th ed.). Boston: Addison-Wesley. ISBN 978-0- 321- 19784-9.

[9] Ullman, J.D. (1982) Principe of Database Systems. Computer Science Press, Rockville, Md.

[10] Maier, D. (1983) The Theory of relational Databases. Computer Science Press, Rockville, Md.

[11] Bernstein, P.A. (1976) "Synthesizing third normal form relations from functional dependencies", ACM Transactions on database Systems, vol.1, No.4, pp.277—298.

[12] Diederich, J., Milton, J. (1988) "New Methods and Fast Algorithms for Database Normalization", ACM Transactions on database Systems, Vol.13, No.3, pp. 339—365

HIGH CAPACITY DATA HIDING USING LSB STEGANOGRAPHY AND ENCRYPTION

Shamim Ahmed Laskar[1] and Kattamanchi Hemachandran[2]

Department of Computer ScienceAssam University, Silchar, Assam, India

ABSTRACT

The network provides a method of communication to distribute information to the masses. With the growth of data communication over computer network, the security of information has become a major issue. Steganography and cryptography are two different data hiding techniques. Steganography hides messages inside some other digital media. Cryptography, on the other hand obscures the content of the message. We propose a high capacity data embedding approach by the combination of Steganography and cryptography. In the process a message is first encrypted using transposition cipher method and then the encrypted message is embedded inside an image using LSB insertion method. The combination of these two methods will enhance the security of the data embedded. This combinational methodology will satisfy the requirements such as capacity, security and robustness for secure data transmission over an open channel. A comparative analysis is made to demonstrate the effectiveness of the proposed method by computing Mean square error (MSE) and Peak Signal to Noise Ratio (PSNR). We analyzed the data hiding technique using the image performance parameters like Entropy, Mean and Standard Deviation. The stego images are tested by transmitting them and the embedded data are successfully extracted by the receiver. The main objective in this paper is to provide resistance against visual and statistical attacks as well as high capacity.

KEYWORDS

Steganography, Cryptography, plain text, encryption, decryption, transposition cipher, Least Significant Bit, Human Visual System, Mean square error and Peak Signal to Noise Ratio.

For More Details : https://airccse.org/journal/ijdms/papers/4612ijdms05.pdf

Volume Link : https://airccse.org/journal/ijdms/current2012.html

REFERENCES

[1] Anderson , R. J. and Petitcolas, F. A.P. (1998) "On The Limits of Steganography", IEEE Journal of Selected Areas in Communications, Vol.16 No.4, pp.474-481, ISSN 0733-8716.

[2] Petitcolas, F.A.P., Anderson, R. J. and Kuhn, M.G. (1999) "Information Hiding -A Survey", Proceedings of the IEEE, Special issue on Protection of Multimedia Content, vol. 87, no. 7, pp.1062-1078. International Journal of Database Management Systems (IJDMS) Vol.4, No.6, December 2012 67

[3] Johnson, N.F. and Jajodia, S. (1998) "Exploring Steganography: Seeing the Unseen", IEEE, Computer, vol. 31, no. 2, pp. 26-34.

[4] Raphael, A. J. and Sundaram, V. "Cryptography and Steganography – A Survey", Int. J. Comp. Tech. Appl., Vol 2 (3), pp. 626-630 , ISSN:2229-6093.

[5] Gutte, R. S. and Chincholkar, Y. D. (2012) "Comparison of Steganography at One LSB and Two LSB Positions", International Journal of Computer Applications, Vol.49,no.11, pp.1-7.

[6] Laskar, S.A. and Hemachandran, K. (2012), "An Analysis of Steganography and Steganalysis Techniques", Assam University Journal of Sscience and Technology, Vol.9, No.II, pp.83-103, ISSN: 0975-2773.

[7] Younes, M.A.B. and Jantan, A. (2008), "Image Encryption Using Block-Based Transformation Algorithm," International Journal of Computer Science, Vol. 35, Issue.1, pp.15-23.

[8] Walia, E., Jain, P. and Navdeep. (2010), " An Analysis of LSB & DCT based Steganography", Global Journal of Computer Science and Technology, Vol. 10 Issue 1 , pp 4-8.

[9] Khare, P., Singh, J. and Tiwari, M. (2011), "Digital Image Steganography", Journal of Engineering Research and Studies, Vol. II, Issue III, pp. 101-104, ISSN:0976-7916.

[10] Sokouti, M., Sokouti, B. and Pashazadeh, S. (2009), "An approach in improving transposition cipher system", Indian Journal of Science and Technology, Vol.2 No. 8, pp. 9-15, ISSN: 0974- 6846.

[11] Kharrazi, M., Sencar, H. T. and Memon, N. (2006), "Performance study of common image steganography and steganalysis techniques", Journal of Electronic Imaging, SPIE Proceedings Vol. 5681.15(4), 041104 pp.1-16.

[12] R., Chandramouli, and Nasir Memon.(2001), "Analysis of LSB based image steganography techniques." In Image Processing, 2001. Proceedings. 2001 International Conference on, IEEE, vol. 3, pp. 1019-1022.

[13] Giddy, J.P. and Safavi- Naini, R. (1994), " Automated Cryptanalysis of Transposition Ciphers", The Computer Journal, Vol.37, No.5, pp. 429-436.

[14] Johnson, N. F. and Katzenbeisser, S. (2000), "A survey of steganographic techniques", In Information Hiding, Artech House, Norwood, MA, pp. 43-78.

[15] Chandramouli, R. and Menon, N. (2001), "Analysis of LSB based image steganography techniques", IEEE Proceedings on Image Processing, Vol.3, pp.1019-1022.

[16] Carvajal-Gamez , B.E., Gallegos-Funes, F. J. and Lopez-Bonilla, J. L. (2009), " Scaling Factor for RGB Images to Steganography Applications", Journal of Vectorial Relativity, Vol. 4, no. 3, pp.55-65.

[17] Ulutas, G., Ulutas, M. and Nabiyev,V. (2011), "Distortion free geometry based secret image sharing", Elsevier Inc, Procedia Computer Science 3, pp.721–726.

[18] Tiwari, N. and Shandilya, M. (2010), "Evaluation of Various LSB based Methods of Image Steganography on GIF File Format", International Journal of Computer Applications (0975 – 8887) Vol. 6, no.2, pp.1-4.

[19] Rabah, K. (2004), "Steganography – The Art of Hiding Data", Information Technology Journal, Vol.3, no.3, pp. 245-269.

[20] Deshpande, N., Kamalapur, S. and Daisy, J. (2006), "Implementation of LSB steganography and Its Evaluation for Various Bits", 1st International Conference on Digital Information Management, pp.173-178.

[21] Karen, Bailey, and Kevin Curran.(2006) "An evaluation of image based steganography methods" Multimedia Tools and Applications, Springer Vol.30, no. 1, pp. 55-88.

[22] Celik, M. U., Sharma, G., Tekalp, A.M. and Saber, E. (2005), "Lossless Generalized-LSB Data Embedding", IEEE Transaction on Image Processing, Vol. 14, No. 2, pp. 253-266.

[23] Huang, Y. S., Huang, Y. P., Huang, K.N. and Young, M. S. (2005), "The Assessment System of Human Visual Spectral Sensitivity Curve by Frequency Modulated Light", Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, pp. 263-265. International Journal of Database Management Systems (IJDMS) Vol.4, No.6, December 2012 68

[24] Chan, Chi-Kwong, and L. M. Cheng. (2004), "Hiding data in images by simple LSB substitution." Pattern Recognition Vol. 37, no. 3, pp. 469-474.

[25] Brisbane, G., Safavi-Naini, R. and Ogunbona, P. 2005. "High-capacity steganography using a shared colour palette", IEEE Proceedings on Vision, Image and Signal Processing, Vol.152, No.6, pp.787- 792.

[26] Curran, K. and Bailey, K. (2003), "An Evaluation of Image Based Steganography Methods", International Journal of Digital Evidence Fall 2003, Volume 2, Issue 2, www.ijde.org.

[27] Dickman, S.D. (2007), "An Overview of Steganography", JMU-INFOSEC-TR-2007-002, http:// citeseerx.ist.psu.edu/viewdoc/summary? doi=10.1.1.137.5129. [28] Dunbar, B. (2002). "A detailed look at Steganographic Techniques and their use in an Open-Systems Environment", SANS Institute 2002, pp.1-9, http://www.sans.org.

[29] Lee, Y-K. ; Bell, G., Huang, S-Y., Wang, R-Z. and Shyu, S-J. (2009), "An Advanced LeastSignificant-Bit Embedding Scheme for Steganographic Encoding",PSIVT 2009, LNCS 5414, Springer, pp. 349–360.

[30] Smith, C. (2001), "Basic Cryptanalysis Techniques", SANS Institute 2001, GSEC Version 1.2f, http://www.sans.org.

[31] Kaur, R., Singh, B. and Singh, I. (2012), "A Comparative Study of Combination of Different Bit Positions In Image Steganography", International Journal of Modern Engineering Research, Vol.2, Issue.5, pp-3835-3840.

[32] Kruus, P., Caroline, S., Michael, H. and Mathew, M. (2002), "A Survey of Steganographic Techniques for Image Files", Advanced Security Research Journal, Network Associates Laboratories, pp.41-51.

[33] Kharrazi, M., Sencar, H. T. and Memon, N. (2004), "Image Steganography: Concepts and Practice", WSPC/Lecture Notes Series: 9in x 6in, pp.1-31.

[34] B, Li., J, He. and J, Huang. (2011), "A Survey on Image Steganography and Steganalysis", Journal of Information Hiding and Multimedia Signal Processing, Vol. 2, No. 2, pp. 142-172.

[35] Friedman, W.F. (1967), "Cryptology", Encyclopedia Britannica, Vol. 6, pp. 844-851, 1967.

[36] Kahate, A. (2008), "Cryptography and Network Security", 2nd Edition, Tata McGraw-Hill.

[37] Gonzalez, R. C. and Woods, R. E. (2002), "Digital Image Processing", 2nd edition, Prentice Hall, Inc

AUTHORS

Shamim Ahmed Laskar received his B.Sc. and M.Sc. degrees in Computer Science in 2006 and 2008 respectively from Assam University, Silchar, where he is currently doing his Ph.D. His research interest includes Image Processing, Steganography, Information Retrieval and Data Security

Prof. Kattamanchi Hemachandran obtained his M.Sc. Degree from Sri Venkateswara University, Tirupati and M.Tech and Ph.D Degrees from Indian School of Mines, Dhanbad. Presently, he is serving as Head, Department of Computer Science, Assam University, Silchar. He is associated with this department since 1998. He is supervising many research scholars. His areas of research interest are Image Processing, Software Engineering and Dist ributed Computing.

DESIGNING A FRAMEWORK TO STANDARDIZE DATA WAREHOUSE DEVELOPMENT PROCESS FOR EFFECTIVE DATA WAREHOUSING PRACTICES

Deepak Asrani[1] and Renu Jain[2]

[1]Department of Computer Science Engineering, Teerthanker Mahaveer University Moradabad, U. P., India
[2]Department of Computer Science Engineering, University Institute of Engineering & Technology, Kanpur, U.P., India

ABSTRACT

Data warehousing solutions work as information base for large organizations to support theirdecision making tasks. With the proven need of such solutions in current times, it is crucial to effectively design, implement and utilize these solutions. Data warehouse (DW) implementation has been a challenge for the organizations and the success rate of its implementation has been very low. To address these problems, we have proposed a framework for developing effective data warehousing solutions. The framework is primarily based on procedural aspect of data warehouse development and aims to standardize its process. We first identified its components and then worked on them in depth to come up with the framework for effective implementation of data warehousing projects. To verify effectiveness of the designed framework, we worked on National Rural Health Mission (NRHM) project of Indian government and designed data warehousing solution using the proposed framework.

KEYWORDS

Data warehousing, Framework Design, Dimensional modelling, Decision making, Materialized View

REFERENCES

[1] Lee, S.hyun. & Kim Mi Na, (2008) "This is my paper", ABC Transactions on ECE, Vol. 10, No. 5, pp120-122.[

[2] Chan. G. K. Y., Li. Q., and Fenf. L., "Design and Selection of Materialized Views in a Data Warehousing Environment: A Case Study", DOLAP' 99, Kanas City MO USA, ACM 1999 1- 58113-220-4/99/11

[3] Colleen Cunningham and Il-Yeol Song, "A Taxonomy of Customer Relationship Management Analyses for Data Warehousing", Twenty-Sixth International Conference on Conceptual Modelling ER 2007 Tutorials

[4] Connor, D. (2003). Report: Data Warehouse Failures Commonplace. Network World, 20(3), 24.

[5] Ewen. E. F., Medskar. C. E., and Dusterhoft. L. E., "Data Warehousing in an Integrated Health System;Building the Business Case" DOLAP '98 Washington DC USA, ACM 1999 l-581 13-120- 8/98/l 1

[6] Finnegan, P., Murphy, C., and O'Riordan, J. (1999). "Challenging the Hierarchical Perspective on Information Systems: Implications from External Information Analysis," Journal of Information Technology, Forthcoming.

[7] P. Giorgini et al. "Goal-Oriented Requirement Analysis for Data Warehouse Design", in "Eighth ACM International Workshop on Data Warehousing and OLAP", ISBN: 1-59593-162-7, ACM Press, NY, 2005...

[8] Hao Fan, Alexandra Poulovassilis "Using AutoMed Metadata in Data Warehousing Environments" DOLAP'03, November 7, 2003, New Orleans, Louisiana, USA.

[9] Hwang, M.I. & Xu, H., (2005, fall). "A Survey of Data Warehouse Success Issues", Business Intelligence Journal, 10(4), 7-13.

[10] Inmon W. H. "Building the Data Warehouse, 3rd Edition, Willey, 2002.

[11] James F. Kimpel "Critical Success Factors for Data warehousing: A Classical Answer to A Modern Question" Issues in Information Systems, Volume 14, Issue 1, pp. 376-384, 2013.

[12] Jindal. R. & Taneja. S., "Comparative Study of Data warehouse Approaches: A Survey",
International Journal of Database Management Systems (IJDMS) Vol. 4, No. 1, February 2012

[13] M.E. Jones and I-Y. Song. "Dimensional Modelling:Identifying, Classifying, and Applying Patterns ", in "Eighth ACM International Workshop on Data Warehousing and OLAP", ISBN: 1-59593-162-7, ACM Press, NY, 2005..

[14] Kim. H., Lee. T., Lee. S., and Chun. J., "Automated Data Warehousing for Rule-based CRM Systems", (ADC 2003), Fourteenth Australian Database Conference, Adelaide Australia.

[15] Kimball R., Reeves. L., Ross. M., and Thronthwaire. W. "The Data Warehouse Life Cycle Kit" Willey. New York, 1998.

[16] Kimball R., & Ross. M., "The Data Warehousing Tool Kit" Willey, New York, 2005."

[17] L. Bellatrach et al. "A Personalization Framework for OLAP Queries", in "Eighth ACM International Workshop on Data Warehousing and OLAP", ISBN: 1-59593-162-7, ACM Press, NY, 2005...

[18] J-N. Mazón et al. "Applying MDA to the Development of Data Warehouses", in "Eighth ACM International Workshop on Data Warehousing and OLAP", ISBN: 1-59593-162-7, ACM Press, NY, 2005..

[19] MET/S Methodology Documentation Book Release 2.3, Cork: Data Warehouse Network.

[20] Ordonez. C., Song. Y., & Alvarado. C. G., "Relational versus Non-Relational Database Systems for Data Warehousing", DOLAP'10, October 30, 2010, Toronto, Ontario, Canada.ACM 978-1-4503- 0383-5/10/10.

[21] J. M. Pérez et al. "A Relevance-Extended Multidimensional Model for a Data Warehouse Contextualized with Documents", in "Eighth ACM International Workshop on Data Warehousing and OLAP", ISBN: 1-59593-162-7, ACM Press, NY, 2005..

[22] K. Q. Pu "Modelling, Querying and Reasoning about OLAP Databases: a Functional Approach", in "Eighth ACM International Workshop on Data Warehousing and OLAP", ISBN: 1-59593-162-7, ACM Press, NY, 2005...

[23] Sammon, D. & Finnegon, P., "The Ten Commandments of Data warehousing", "The DATA BASE for Advances in Information Systems - Fall 2000 (Vol. 31, No. 4)"

[24] Sen, A., Ramamurthy, K. (Ram), & Sinha, A.P. (2012) "A Model of Data Warehouse Process Maturity", IEEE Transactions on Software Engineering", 38(2), 336-353. doi:http://dx.doi.org/10.1109/TSE.2011.2

[25] Sen. A, and Sinha. A.P. "A Comparison of Data warehousing Methodologies", " Communications of the ACM, March 2005, Vol. 48, No. 3"

[26] Trujillo. J. & Song. Y., "A Report on the Eighth ACM International Workshop on Data Warehousing and OLAP (DOLAP' 05), SIGMOD Records, Vol. 36, No. 4, December 2006.

[27] Business Intelligence: A managerial approach. (2011) (2nd ed.) Boston: Prentice Hall.

[28] http://nrhm.gov.in

[29] http://nrhm-mcts.nic.in

AUTHORS

Deepak Asrani is a Ph. D scholar of Teerthanker Mahaveer University, Moradabad, U.P., India. He is a B.E. (Electronics and Telecommunication), and M.S (Software Systems) from BITS, Pilani, Rajasthan, India. He has Software industry and academic experience of 24 years in various organizations and academic institutions.

Dr. Renu Jain is Director of University Institute of Engineering & Technology, Kanpur, U.P, India. She has worked as Head in the Department of Computer Science Engineering in the same institutions. She is an eminent personality in academics and holds Ph. D in Computer Science from BITS, Pilani, Rajasthan, India. She has rich academic experience of more than 30 years and has published around 50 research papers in various reputed national and international journals.

A STUDY ON CHALLENGES AND OPPORTUNITIES IN MASTER DATA MANAGEMENT

Tapan kumar Das[1] and Manas Ranjan Mishra[2]

[1]SITE, VIT University, Vellore, TN, India
[2] IBM India Pvt .Ltd, Bangalore, India

ABSTRACT

This paper aims to provide a data definition of one master data for cross application consistency. The concepts related to Master data management in broader spectrum has been discussed. The current challenges companies are facing while implementing the MDM solutions are outlined. We have taken a case study to highlight why Master Data Management is imperative for the enterprises in optimizing their business Also we have identified some of the long term benefits for the enterprises on implementing MDM..

KEYWORDS

Data quality, Information system, Unstructured, Transactional data

For More Details : https://airccse.org/journal/ijdms/papers/3211ijdms09.pdf

Volume Link : https://airccse.org/journal/ijdms/current2011.html

REFERENCES

[1] Berson, A. and Dubov, L. (2007), Master Data Management and Customer Data Integration for aGlobal Enterprise, McGraw-Hill, New York, NY.

[2] Boyd, M. (2006), "Product information management – forcing the second wave of data quality",available at: ww.thecopywritingpro.com/pages/samples_assets/2nd-waveDQ.pdf(accessed 27 April 2010).

[3] Breuer, T. (2009), "Data quality is everyone's business – designing quality into your datawarehouse – part 1", Journal of Direct, Data and Digital Marketing Practice, Vol. 11,pp. 20-9.

[4] Butler,David., Stackowiak,Bob., "Master Data Management", Oracle Corporation.available at www.oracle.com.

[5] Dayton, M. (2007), "Strategic MDM: the foundation of enterprise performance management",Cutter IT Journal, Vol. 20 No. 9, pp. 13-17.

[6] Dreibelbis,Allen , Hechler,Eberhard, Milman ,Ivan(2009), "Enterprise Master Data Management",Pearson Education.

[7] Dumas, M., Aalst, W. and Ter Hofstede, A. (2005), Process-aware Information Systems: BridgingPeople and Software Through Process Technology, Wiley, Hoboken, NJ.Managing onemaster data161

[8] Gartner MDM Summit,(2011),UK International Journal of Database Management Systems (IJDMS), Vol.3, No.2, May 2011 139

[9] Knolmayer, G. and Ro¨thlin, M. (2006), "Quality of material master data and its effect on theusefulness of distributed ERP systems", Lecture Notes in Computer Science, Vol. 4231,pp. 362-71.

[10] Lee, Y.W., Pipino, L.L., Funk, J.D. and Wang, R.Y. (2006)," Journey to Data Quality", MIT Press, Cambridge, MA.

[11] Loshin, D. (2009), Master Data Management, Morgan Kaufmann, Burlington,MA.

[12] Malcolm, Chisholm(Dec 2010),"The Governance Challenge for Master Data Management", Data Governance Conference, Orlando, Florida

[13] "MDM Fundamentals and Best practices",www.elearningcurve.com

[14] Moss, L. (2007), "Critical success factors for master data management", Cutter IT Journal, Vol. 20No. 9, pp. 7-12.

[15] Rachuri, S., Subrahmanian, E., Bouras, A., Fenves, S., Foufou, S. and Sriram, R. (2008),"Information sharing and exchange in the context of product lifecycle management: role ofstandards", Computer-Aided Design, Vol. 40 No. 7, pp. 789-800.

[16] Rittman,Mark."Introduction to Master Data Management".www.rittmanmead.com

[17] Russon,Philip. (2006), Master Data Management – TDWI Best practice Report a vailable at www.tdwi.org.

[18] Schumachar, Scott (Oct 2010),"MDM: Realizing the same benefits through different implementations", www.initiatesystems.com

[19] Toronto MDM Summit.(2008),"MDM Challenges and solutions from the real world" available at www.adastracorp.com

[20] White, A., Newman, D., Logan, D. and Radcliffe, J. (2006), "Mastering master datamanagement", available at: http://kona.kontera.com/IMAGE_DIR/pdf/MDM_gar060125_MasteringMDMB.pdf (accessed 12 April 2010).

[21] Wolter,Roger.,Haselden,Kirk.(2006),A White paper on MDM, Microsoft Corporation.

[22] Yang, X., Moore, P.R., Wong, C.-B., Pu, J.-S. and Chong, S.K. (2007), "Product lifecycle information acquisition and management for consumer products", Industrial Management & Data Systems, Vol. 107 No. 7, pp. 936-56.

TOP NEWSQL DATABASES AND FEATURES CLASSIFICATION

Ahmed Almassabi[1] , Omar Bawazeer and Salahadin Adam[2]

[1]Department of Computer Science, Najran University, Najran, Saudi Arabia

[2]Department of Information and Computer Science, King Fahad University of Petroleum

and Mineral, Dhahran, Saudi Arabia

ABSTRACT

Versatility of NewSQL databases is to achieve low latency constrains as well as to reduce cost commodity nodes. Out work emphasize on how big data is addressed through top NewSQL databases considering their features. This NewSQL databases paper conveys some of the top NewSQL databases [54] features collection considering high demand and usage. First part, around 11 NewSQL databases have been investigated for eliciting, comparing and examining their features so that they might assist to observe high hierarchy of NewSQL databases and to reveal their similarities and their differences. Our taxonomy involves four types categories in terms of how NewSQL databases handle, and process big data considering technologies are offered or supported. Advantages and disadvantages are conveyed in this survey for each of NewSQL databases. At second part, we register our findings based on several categories and aspects: first, by our first taxonomy which sees features characteristics are either functional or non-functional. A second taxonomy moved into another aspect regarding data integrity and data manipulation; we found data features classified based on supervised, semi-supervised, or unsupervised. Third taxonomy was about how diverse each single NewSQL database can deal with different types of databases. Surprisingly, Not only do NewSQL databases process regular (raw) data, but also they are stringent enough to afford diverse type of data such as historical and vertical distributed system, real-time, streaming, and timestamp databases. Thereby we release NewSQL databases are significant enough to survive and associate with other technologies to support other database types such as NoSQL, traditional, distributed system, and semirelationship to be as our fourth taxonomy-based. We strive to visualize our results for the former categories and the latter using chart graph. Eventually, NewSQL databases motivate us to analyze its big data throughput and we could classify them into good data or bad data. We conclude this paper with couple suggestions in how to manage big data using Predictable Analytics and other techniques.

KEYWORDS

NewSQL, NoSQL, RDBMs. FF, Non-FF, and Big data.

REFERENCES

[1] Ismail, M., Gebremeskel, E., Kakantousis, T., Berthou, G., Dowling, J. (2017, June). Hopsworks: Improving User Experience and Development on Hadoop with Scalable, Strongly Consistent Metadata. In Distributed Comput-ing Systems (ICDCS), 2017 IEEE 37th International Conference on(pp. 2525-2528). IEEE.

[2] Sangtani, M., D'souza, S. M., Harsh, M., Chander, S., Vijaya, P. IN-TERNATIONAL JOURNAL OF ENGINEERING SCIENCES RESEARCH TECHNOLOGY IMPLEMENTATION CHALLENGESINVOLVED IN BIG DATA ANALYTICS.

[3] Kobielus, J. (2012). Hadoop: Nucleus of the next-generation big data ware-house. IBM Data Management Magazine.

[4] Lightstone, S., Ohanian, R., Haide, M., Cho, J., Springgay, M., Steinbach, T. (2017, April). Making
Big Data Simple with dashDB Local. In Data Engi-neering (ICDE), 2017 IEEE 33rd International Conference on (pp. 1195-1205). IEEE.

[5] Santos, M. Y., Costa, C., Galv~ao, J., Andrade, C., Martinho, B. A., Lima, F. V., Costa, E. (2017, July). Evaluating SQL-on-hadoop for big data warehous-ing on not-so-good hardware. In Proceedings of the 21st International Database Engineering Applications Symposium (pp. 242-252). ACM.

[6] Ismail, M., Gebremeskel, E., Kakantousis, T., Berthou, G., Dowling, J. (2017, June). Hopsworks: Improving User Experience and Development on Hadoop with Scalable, Strongly Consistent Metadata. In Distributed Com-puting Systems (ICDCS), 2017 IEEE 37th International Conference on (pp. 2525-2528). IEEE.

[7] Thusoo, A., Shao, Z., Anthony, S., Borthakur, D., Jain, N., Sen Sarma, J., Liu, H. (2010, June). Data
warehousing and analytics infrastructure at face-book. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of data (pp. 1013-1020). ACM.

[8] Barkhordari, M., Niamanesh, M. (2017). Atrak: a MapReduce-based data warehouse for big data. The Journal of Supercomputing, 1-15.

[9] Tankard, C. (2012). Big data security. Network security, 2012(7), 5-8.

[10] Corbett, J. C., Dean, J., Epstein, M., Fikes, A., Frost, C., Furman, J. J., Hsieh, W. (2013). Spanner:
Google's globally distributed database. ACM Transactions on Computer Systems (TOCS), 31(3), 8.

[11] Song, L., Smola, A., Gretton, A., Borgwardt, K. M., Bedo, J. (2007, June). Supervised feature selection via dependence estimation. In Proceedings of the 24th international conference on Machine learning (pp. 823-830). ACM.

[12] Huang, S. H. (2015). Supervised feature selection: A tutorial. Arti cial Intelligence Research,4(2),

[13] Erturk, E., Jyoti, K. (2015). Perspectives on a Big Data Application: What Database Engineersand IT Students Need to Know. Engineering, Technology Applied Science Research, 5(5), pp-850.

[14] Davenport, T. H., Barth, P., Bean, R. (2012). How big data is di erent. MIT Sloan Management Review, 54(1), 43.

[15] [Book] Iafrate, F. (2015). From big data to smart data (Vol. 1). John Wiley Sons.

[16] Russom, P. (2013). Managing big data. TDWI Best Practices Report, TDWI Research, 1-40
.
[17] Cox, M., Ellsworth, D. (1997, August). Managing big data for scienti c visualization. In ACM Siggraph (Vol. 97, pp. 21-38).

EXPERIMENTAL EVALUATION OF NOSQL DATABASES

Veronika Abramova[1], Jorge Bernardino[1,2] and Pedro Furtado[2]

[1] Polytechnic Institute of Coimbra - ISEC / CISUC, Coimbra, Portugal
[2]University of Coimbra – DEI / CISUC, Coimbra, Portugal

Relational databases are a technology used universally that enables storage, management and retrieval of varied data schemas. However, execution of requests can become a lengthy and inefficient process for some large databases. Moreover, storing large amounts of data requires servers with larger capacities and scalability capabilities. Relational databases have limitations to deal with scalability for large volumes of data. On the other hand, non-relational database technologies, also known as NoSQL, were developed to better meet the needs of key-value storage of large amounts of records. But there is a large amount of NoSQL candidates, and most have not been compared thoroughly yet. The purpose of this paper is to compare different NoSQL databases, to evaluate their performance according to the typical use for storing and retrieving data. We tested 10 NoSQL databases with Yahoo! Cloud Serving Benchmark using a mix of operations to better understand the capability of non-relational databases for handling different requests, and to understand how performance is affected by each database type and their internal mechanisms

KEYWORDS

NoSQL databases, SQL databases, performance evaluation, database models, YCSB

For More Details : https://airccse.org/journal/ijdms/papers/6314ijdms01.pdf

Volume Link : https://airccse.org/journal/ijdms/current2014.html

REFERENCES

[1] Fayech, I. and Ounalli, H.: Towards a Flexible Database Interrogation. International Journal of Database Management Systems (IJDMS) Vol.4, No.3, June 2012 .

[2] http://nosql-database.org/. International Journal of Database Management Systems (IJDMS) Vol.6, No.3, June 2014 15

[3] Vimala, S., Khanna Nehemiah, H., Bhuvaneswaran, R. S., and Saranya, G.: Design Methodology for Relational Databases: Issues Related to Ternary Relationships in Entity-relationship Model and Higher Normal Forms.

[4] Stonebraker, M.: SQL databases vs. NoSQL databases. Communications of the ACM, Vol. 53 No. 4, Pages 10-11.

[5] Gajendran, S.: A Survey on NoSQL Databases, 2012, http://ping.sg/story/A-Survey-on-NoSQLDatabases---Department-of-Computer-Science.

[6] Cooper, B., Silberstein, A., Tam, E., Ramakrishnan, R., and Sears, R.: Benchmarking cloud serving systems with YCSB. In Proceedings of the 1st ACM Symposium on Cloud Computing (SoCC '10). ACM, New York, NY, USA, 143-154.

[7] Strozzi, C.: NoSQL – A relational database management system, 2013, http://www.strozzi.it.

[8] Chang, F., Jeffrey, D., Ghemawat, S., Hsieh, W., Wallach, D., Burrows, M., Chandra, T., Fikes, A. and Gruber, R.: Bigtable: A Distributed Storage System for Structured Data. ACM Transactions on Computer Systems, 26(2), Article 4.

[9] Decandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., Sivasubramanian, S., Vosshall, P., and Vogels, W.: Dynamo: amazon's highly available key-value store. In Proceedings of twenty-first ACM SIGOPS Symposium on Operating Systems principles (SOSP '07). ACM, New York, NY, USA, 205-220.

[10] Notes from the NoSQL Meetup, 2013, http://developer.yahoo.com/blogs/ydn/posts/2009/06/nosql_meetup.

[11] Hecht, R. and JABLINSKI, S.: NoSQL Evaluation A Use Case Oriented Survey. Proceedings International Conference on Cloud and Service Computing, pp. 12-14.

[12] Han, J.: Survey on NOSQL Databases. Proceedings 6th International Conference on Pervasive Computing and Applications, pp. 363-366.

[13] Leavitt, N.: Will NoSQL Databases Live up to Their Promise?. Computer Magazine, Vol. 43 No. 2, pp. 12-14.

[14] Floratou, A., Teletia, N., Dewitt, D., Patel, J. and Zhang, D.: Can the elephants handle the NoSQL onslaught?. Proc. VLDB Endow. 5,1712-1723.

[15] Tudorica, B.G. and Bucur, C.: A comparison between several NoSQL databases with comments and notes. Roedunet International Conference (RoEduNet), pp.1-5.

[16] Pritchett, D.: BASE: An Acid Alternative. ACM Queue 6(3), 48-55.

[17] Cook, J. D.: ACID versus BASE for database transactions, 2009, http://www.johndcook.com/blog/2009/07/06/brewer-cap-theorem-base.

[18]   Browne,   J.:   Brewer's   CAP   Theorem, 2009, http://www.julianbrowne.com/article/viewer/brewers-captheorem.

[19] Indrawan-Santiago, M.: Database Research: Are We at a Crossroad? Reflection on NoSQL. NetworkBased Information Systems (NBiS), 15th International Conference on Network-Based Information Systems, pp.45-51.

[20] Zhang, H. and Tompa, F.W.: Querying XML documents by dynamic shredding. InProceedings of the 2004 ACM symposium on Document engineering (DocEng '04). ACM, New York, NY, USA, 21-30.

[21] Crockford, D.: JavaScript: The Good Parts. Sebastopol, CA: O'Reilly Media.

[22]   Lamb,   C.:   Oracle   NoSQL   Database   in   5   minutes,   2013, https://blogs.oracle.com/charlesLamb/entry/oracle_nosql_database_in_5.

[23] Armstrong, T., Ponnekanti, V., Dhruba, B., and Callaghan, M.: LinkBench: a database benchmark based on the Facebook social graph. In Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data (SIGMOD '13). ACM, New York, NY, USA, 1185-1196.

[24] Dayarathna, M. and Suzumura, T.: XGDBench: A benchmarking platform for graph stores in exascale clouds. Cloud Computing Technology and Science (CloudCom), IEEE 4th International Conference on, Taipei, 363 – 370.

AUTHORS

Veronika Abramova is currently a researcher at CISUC – Centre for Informatics and Systems of the University of Coimbra. Previously she has received a bachelor degree at Instituto Superior Engenharia de Coimbra (ISEC). She is currently working in an industrial project, with an electricity sector company, focused on transferring part of the company's data to the non- relational storage. She has evaluated and studied different NoSQL databases, focusing on their performance comparison and characteristics. Her main research fields are business intelligence, big data, database knowledge management, NoSQL and SQL databases performance evaluation.

Jorge Bernardino received the PhD degree in computer science from the University of Coimbra in 2002. He is a Coordinator Professor at ISEC (Instituto Superior de Engenharia de Coimbra) of the Polytechnic of Coimbra, Portugal. His main research fields are big data, data warehousing, business intelligence, open source tools, and software engineering, subjects in which he has authored or co-authored dozens of papers in refereed conferences and journals.

Jorge Bernardino received the PhD degree in computer science from the University of Coimbra in 2002. He is a Coordinator Professor at ISEC (Instituto Superior de Engenharia de Coimbra) of the Polytechnic of Coimbra, Portugal. His main research fields are big data, data warehousing, business intelligence, open source tools, and software engineering, subjects in which he has authored or co-authored dozens of papers in refereed conferences and journals. Jorge Bernardino has served on program committees of many conferences and acted as referee for many international conferences and journals. He was President of ISEC from 2005– 2010. Actually, he is serving as General Chair of IDEAS conference and visiting professor at Carnegie Mellon University (CMU).

Pedro Furtado is Professor at University of Coimbra, Portugal, where he teaches courses in both Computer and Biomedical Engineering. His main research interests are data scalability and bigdata, data mining, service-oriented systems and real time systems. Lately, his research has focused both on scalable and real time warehousing, and also on middleware for wireless sensors in industrial and healthcare applications. He has more than 100 papers published in international conferences and journals, books published and several research collaboratio ns with both industry and academia. Besides a PhD in Computer Engineering from University Coimbra (UC) in 2000, Pedro Furtado also holds an MBA from Universidade Catolica Portuguesa (UCP)

A THEORETICAL EXPLORATION OF DATA MANAGEMENT AND INTEGRATION IN ORGANISATION SECTORS

Chisom E. Offia and Malcolm Crowe

[1]**SITE, VIT University, Vellore, TN, India**
[2] **IBM India Pvt .Ltd, Bangalore, India**

ABSTRACT

Big data development is a disturbing issue that will affect enterprise across various sectors. The increase of data volume, high speed of data generation and increasing rate of different data from heterogeneous sources have led to difficulties in data management. This paper first reviews different aspects of big data management, including data integration and traditional data warehouse, and their associated challenges.The problems include increase of redundant data, data accessibility, time consumption in data modelling and data movement from heterogeneous sources into a central database, especially in the big data environment. We then propose a logical data management approach using RESTview technology to integrate and analyse data, without fully adopting traditional ETL processes. Data that for governance, corporate, security or other restriction reasons cannot be copied or moved, can easily be accessed, integrated and analysed, without creating a central repository. Data can be kept in its original form and location, eliminating the movement of data, significantly speeding up the process and allowing for live data interrogation. It may not be the practical solution for every situation but, it is a feasible solution that is comparably cost effective.

KEYWORDS

Big Data; Data Integration; Data warehouse; RESTView

For More Details : https://aircconline.com/ijdms/V11N1/11119ijdms03.pdf

Volume Link : https://airccse.org/journal/ijdms/current2019.html

REFERENCES

[1] Bansal, S. and Kagemann, S. (2015). Integrating Big Data: A Semantic Extract-Transform-Load Framework. Computer, 48(3), pp.42-50.

[2] Saxena, S. and Kumar Sharma, S. (2016). Integrating Big Data in "e-Oman": opportunities and challenges. info, 18(5), pp.79-97.

[3] Khan, M., Wu, X., Xu, X. and Dou, W. (2017). Big data challenges and opportunities in the hype of Industry 4.0. 2017 IEEE International Conference on Communications (ICC).

[4] Karafiloski, E. and Mishev, A. (2017). Blockchain solutions for big data challenges: A literature review. IEEE EUROCON 2017 -17th International Conference on Smart Technologies.

[5] Crowe, M., Begg, C., Laiho, M. and Lau, F. (2016). Data validation for Big Live data.[online] Available at: https://www.researchgate.net/publication/315686427_Data_Validation_for_Big_Live_Data

[6] Chauhan, S., Agarwal, N. and Kar, A. (2016). Addressing big data challenges in smart cities: a systematic literature review. info, 18(4), pp.73-90.

[7] Al Nuaimi, E., Al Neyadi, H., Mohamed, N. and Al-Jaroodi, J. (2015). Applications of big data to smart cities. Journal of Internet Services and Applications, 6(1).

[8] Rabiul, I., Islam, R., Musfiqur, R. and Abiduzzaman, R. (2016). Big Data Characteristics, Value Chain and Challenges.

[9] Almeida, F. and Calistru, C. (2013). The main challenges and issues of big data management. International Journal of Research Studies in Computing, 2(1).

[10] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). Big data: the next frontier for innovation, competition, and productivity. McKinsey Global Institute Reports, 5, 15-36.

[11] Sivarajah, U., Kamal, M., Irani, Z. and Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. Journal of Business Research, 70, pp.263-286.

[12] De Oliveira Veras, A., de Sá, P., da Costa Pinheiro, K., Barh, D., Azevedo, V., Jucá Ramos, R. and da Costa da Silva, A. (2018). Computational Techniques in Data Integration and Big Data Handling in Omics. Omics Technologies and Bio-Engineering, pp.209-222.

[13] Mishra, S., Dhote, V., S. Prajapati, G. and Shukla, J. (2015). Challenges in Big Data Application: A Review. International Journal of Computer Applications, 121(19), pp.42-46.

[14] TOLE, A. (2013). Big Data Challenges. Database Systems Journal, vol. IV, p.no. 3.

[15] Trifu, M. and Ivan, M. (2014). Big Data: present and future. Database Systems Journal, 5(1), pp.32- 41.

[16] Mehmood, R., Meriton, R., Graham, G., Hennelly, P. and Kumar, M. (2017). Exploring the influence of big data on city transport operations: a Markovian approach. International Journal of Operations & Production Management, 37(1), pp.75-104.

[17] CEBR (2012). Data equity: unlocking the value of big data. Centre for Economics and Business Research White Paper, 4, 7-26.

[18] McNulty, E. and Freeman, H. (2014). Understanding Big Data: The Seven V's - Dataconomy. [online] Dataconomy. Available at: http://dataconomy.com/2014/05/seven-vs-big-data/.

[19] Chaudhuri S. What next?: a half-dozen data management research goals for big data and the cloud. In Proceedings of the 31st symposium on Principles of Database Systems. ACM; 2012. pp. 1–4.

[20] Lenzerini, M. (2002). Data integration. Proceedings of the twenty-first ACM SIGMOD-SIGACTSIGART symposium on Principles of database systems - PODS '02.

[21] Merelli, I., Pérez-Sánchez, H., Gesing, S. and D'Agostino, D. (2014). Managing, Analysing, and Integrating Big Data in Medical Bioinformatics: Open Problems and Future Perspectives. BioMed Research International.

[22] George, G., Osinga, E., Lavie, D. and Scott, B. (2016). Big Data and Data Science Methods for Management Research. Academy of Management Journal, 59(5), pp.1493-1507.

[23] Wang, L. and Alexander, C. (2015). Big Data in Distributed Analytics, Cybersecurity, Cyber Warfare and Digital Forensics. Science and Education Publishing, Vol. 1, pp.pp 22-27.

[24] Khan, N., Alsaqer, M., Shah, H., Badsha, G., Abbasi, A. and Salehian, S. (2018). The 10 Vs, Issues and Challenges of Big Data. Proceedings of the 2018 International Conference on Big Data and Education - ICBDE '18.

[25] Smith, J., Bernstein, P., Dayal, U., Goodman, N., Landers, T., Lin, K. and Wong, E. (1981). Multibase. Proceedings of the May 4-7, 1981, national computer conference on - AFIPS '81.

[26] Ruggles, S., Hacker, J. and Sobek, M. (1995). General Design of the Integrated Public Use Microdata Series. Historical Methods: A Journal of Quantitative and Interdisciplinary History, 28(1), pp.33-39.

[27] Ray, S., Bandyopadhyay, S. and Pal, S. (2009). Combining Multisource Information Through Functional-Annotation-Based Weighting: Gene Function Prediction in Yeast. IEEE Transactions on Biomedical Engineering, 56(2), pp.229-236.

[28] Mireku Kwakye, M. (2011). A Practical Approach to Merging Multidimensional Data Models. [online] Dx.doi.org. Available at: http://dx.doi.org/10.20381/ruor-5072.

[29] Dong, X. and Srivastava, D. (2015). Big Data Integration. Synthesis Lectures on Data Management, 7(1), pp.1-198.

[30] Kadadi, A., Agrawal, R., Nyamful, C. and Atiq, R. (2014). Challenges of data integration and interoperability in big data. 2014 IEEE International Conference on Big Data (Big Data).

[31] Ziegler, P. and Dittrich, K. (2008). Data Integration — Problems, Approaches, and Perspectives.

[32] Halevy, A., Rajaraman, A. and Ordille, J. (2006). Data integration: the teenage years. VLDB '06 Proceedings of the 32nd international conference on Very large data bases, pp.9 - 16.

[33] Knoblock, C.A. and Szekely, P., 2013, November. Semantics for big data integration and analysis. In 2013 AAAI Fall Symposium Series.

AUTHORS

Offia Chisom Ernesther studied her bachelor's degree on BSc computer Engineering and graduated with Second class Upper Division in 2015 at Ghana Telecom University College, Accra Ghana. She furthered her Education in Teesside University; Middlesbrough, England United Kingdom on MSc Computer Security and Networks and graduated with Merit. She started her PhD degree in the year 2017 at the University of the West of Scotland, Paisley Scotland with the research focus on achieving logical data warehouse in the process of data analytics (Data Integration).

Malcolm Crowe was born in Dublin 13 January 1948. His first and only employment was at Paisley College in 1972, now the University of the West of Scotland, where he is an Emeritus Professor. He received his D.Phil from Oxford University in 1978 with a thesis entitled "The Connective k-theory of the Infinite Symmetric Group". Computing rather than Mathematics is his real passion. During the 1980s and 1990s he worked on several ESPRIT projects and co-authored several books on CSCW, Information Systems, and Interdisciplinary Research.Since 2006 he has been developing new ideas for how to improve the implementation of relational DBMS, where he regards the maintenance of the transaction log as the only guarantee of accountability and transactional behaviour. He developed a strongly typed optimistic RDBMS called Pyrrho (www.pyrrhodb.com) to try out several technical additions to relational technology including role-based namespaces and semantic and distributed databases. In 2016 Malcolm began to focus on "Big Live Data" (Virtual Data Warehousing), and his collaborators on this concept include Fritz Laux who presented a paper on this concept at DBKDA 2017.

COMPARATIVE STUDY OF DATA WAREHOUSE DESIGN APPROACHES: A SURVEY

Rajni Jindal[1] and Shweta Taneja[2]

[1] Associate Professor, Dept. of Computer Engineering, Delhi Technological University Formerly Delhi College of Engineering (DCE), Bawana Road, Delhi-42.

[2] Research Scholar, Dept. of Computer Engineering, Delhi Technological University Formerly Delhi College of Engineering (DCE), Bawana Road, Delhi-42.

ABSTRACT

The process of developing a data warehouse starts with identifying and gathering requirements, designing the dimensional model followed by testing and maintenance. The design phase is the most important activity in the successful building of a data warehouse. In this paper, we surveyed and evaluated the literature related to the various data warehouse design approaches on the basis of design criteria and propose a generalized object oriented conceptual design framework based on UML that meets all types of user needs.

KEYWORDS

Data warehouse design, Multidimensional modelling, Unified Modelling Language

For More Details : https://airccse.org/journal/ijdms/papers/4112ijdms04.pdf

Volume Link : https://airccse.org/journal/ijdms/current2012.html

REFERENCES

[1] Inmon, W.H., Hackathorn, and R.D (1994) Using the data warehouse. Wiley-QED Publishing, Somerset, NJ, USA.

[2] June 1999,UML Modelling Language Specification. Version 1.3, Available at http://www.rational.com/uml/resources/documention / (March 2009).

[3] Booch G., Rumbaugh J., and Jacobson I.(1999) The Unified Modelling Language User Guide, Addison- Wesley Longman, p.482.

[4] Vassiliadis P. and Sellis, T.,(1999) "A Survey of Logical Models for OLAP Databases". SIGMOD Record 28(4),pp 64–69.

[5] S. Rizzi, A. Abelló, J. Lechtenbörger, J. Trujillo(2006) "Research in data warehouse modelling and design: dead or alive?" DOLAP, ACM, , pp. 3–10.

[6] A. Abelló, J. Samos, and F. Saltor (2001) "A Framework for the Classification and Description of Multidimensional Data Models" In Proceedings of the 12th International Conference on Database and Expert Systems Applications (DEXA'01).

[7] M. Blaschka, C. Sapia, G. Höfling, and B. Dinter,(1998) " Finding your way through ultidimensional data models" In Proceedings of the 9th International Conference on Database and Expert Systems Applications DEXA'98, volume 1460 of Lecture Notes in Computer science, pp 198–203, Vienna, Austria, August 1998. Springer-Verlag.

[8] Stefano Rizzi, Matteo Golfarelli. (1998) "A Methodological Framework for Data Warehouse Design". DOLAP 98 Washington DC USA.Copyright ACM 1999 l-581 13-120-8/98/l 1...$5.00.

[9] Juan Trujilio,E.Medina and S.Lujan Mora (2002) ,"A Web Oriented Approach to manage Multidimensional Models through XML Schemas and XSLT '' EDBT 2002 Workshops, LNCS 2490, pp. 29–44, 2002. Springer-Verlag Berlin Heidelberg.

[10] S.Lujan Mora and I.Song (2002),"Multidimensional Modeling with UML Package Diagrams " In Proc. of the 21st Int. Conf. on Conceptual Modeling.Lecture Notes in Computer Science pp 199- 213,Finland,October 7-11,2002, . Springer-Verlag

[11] Stefano Rizzi, Matteo Golfarelli,D.Maio (1998) "The Dimensional Fact Model:A Conceptual Model for Data Warehouses." International Journal of Cooperative Information Systems(IJC IS),7(2- 3):215-247.''

[12] Lujan Mora and Juan Trujilio (2003) "A Comprehensive Method for Data Warehouse Design."in Proceedings of 5th International Workshop on Design and Management of Data Warehose(DMDW'03), pp 1.1-1.14.

[13] Juan Trujillo and Sergio LujánMora (2004) " Physical Modeling of Data Warehouses using UML" DOLAP'04, Washington, DC, USA.Copyright 2004 ACM 1581139772/04/0011 ...$5.00.

[14] Sergio Luján-Mora1, Panos Vassiliadis and Juan Trujillo. (2004) " Data Mapping Diagrams for Data Warehouse Design with UML"in Proceedings of 23 rd International Conference on Conceptual Modeling (ER 04),volume 3288 of LNCS,China,Springer

[15] Lujan Mora and Juan Trujilio (2006) ."Physical Modeling of Data warehouses by using UML Component and Deployment Diagrams:Design and implementation issues." Journal of Database Management 17(1)

[16] Rizzi,Trujilio,Abello. (2006) " Research in Data Warehouse Modeling and Design:Dead or Alive?". DOLAP'06, Arlington, Virginia, USA.Copyright 2006 ACM 1-59593-530-4/06/0011 ...$5.00.

[17] Deepti Mishra, Ali Yazici, Beri, Pinar Başaran. (2008) " A Casestudy of Data Models in Data Warehousing." 978-1-4244-2624-9/08/$25.00 ©2008 IEEE.

[18] Kamal Alaskar1 and Akhtar Shaikh. (2009)" Object Oriented Data Modeling for Data Warehousing(An Extension of UML approach to study Hajj pilgrim's private tour as a Case Study). International Arab Journal of e-Technology, Vol. 1, No. 2.

[19] Hui Ma,Yiping Yang and Fan Zhang (2009) "The Anti-standardized Design Research of Data Warehouse". IEEE.

[20] Fernandez Medina et al. (2010)" Model Driven Development of Secure XML Data Warehouses: A Case Study" EDBT 2010, Lausanne, Switzerland. Copyright 2010 ACM 978-1-60558-945-9/10/0003 $10.00.

[21] Francois Pinet et. al. (2010) "The use of UML to design agricultural data warehouses." AgEng 2010,International Conference on Agricultural Engineering.France

[22] Jesús Pardillo and Jose-Norberto Mazón. (2011) " Using Ontologies for the Design of Data Warehouses." International Journal of Database Management Systems (IJDMS), Vol.3, No.2.

[23] Payal Pahwa and Shweta Taneja. (2011) "Design of a Multidimensional model using Object Oriented Features in UML." IARS International journal.

[24] L. Cabibbo and R.Torlone(1998) "A Logical Approach to Multidimensional Databases" in Proceedings of 6th International Conference on Extending Database Technology EDBT 98 ,Volume 1337 of LNCS,pp 183-197, Spain,Springer.

[25] J. Trujillo, M. Palomar, J. Gómez, I.-Y. Song(2001)," Designing data warehouses with OO conceptual models", IEEE Comput. 34 (12) (2001) 66– 75.

[26] N.Tryfona,F.Busborg and J.G.Chriastiansen(1998) ,"StarER:A Conceptual Model for Data Warehouse Design",in Proceedings of the ACM 2 nd International Workshop on Data Warehousing and OLAP , DOLAP 99,pp 3-8 .

[27] Luj´n-Mora S., Trujillo J., and Song, I.(2002), "Multidimensional modeling with UML package diagrams warehouses, " in Proceedings of 21 st International Conference on

Conceptual Modeling, ER 02 ,Volume 2503 of LNCS,pp 199-213,Finland ,Springer. [28] Golfarelli, M., & Rizzi, S. (2001)." WanD: A CASE Tool for Data Warehouse Design". In Demo Proceedings 17th International Conference on Data Engineering (ICDE 2001), Heidelberg, Germany, 7-9.

[29] Luj´n-Mora S., Trujillo J., and Song, I., "A UML profile for multidimensional modeling in data warehouses," Data Knowl. Eng. 59(3) 725–769

[30] Anjana Gosain,Suman Mann, (2010)"Object Oriented Multidimensional Model for a Data Warehouse with Operators", International Journal of Database Theory and Application Vol. 3, No. 4.

[31] Rodolfo Villarroel, Emilio Soler, Eduardo Fernández-Medina, Juan Trujillo4,and Mario Piattini (2006)," Using UML Packages for Designing Secure Data Warehouses", ICCSA 2006, LNCS 3982, pp. 1024 – 1034 .© Springer-Verlag Berlin Heidelberg.

[32] Payal Pahwa, Shweta Taneja and Garima Thakur(2011)" Uclean: A Requirement based Object Oriented ETL Framework", International Journal of Computer Science & Engineering Survey (IJCSES) Vol.2, No.4, November 2011

[33] Sarkar, A., Choudhury, S., Chaki, N. & Bhattacharya, S, (2009) "Conceptual Level Design of Object Oriented Data Warehouse: Graph Semantic Based Model", INFOCOMP Journal of Computer Science, pp. 60-70.

[34] Ponniah, P, (2001) Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals, pp 402.

[35] Elmasri, R. & Navathe, S.B, (2000) Fundamentals of Database Systems, Addison Weasely PubCo. ISBN 0201542633.

[36] Nazri, Mior Nasir Mior,Noah, Shahrul Azman Mohd,Hamid, Zarinah (2008)," Automatic data warehouse conceptual design ",International Symposium on Information Technology,Malaysia.

[37] Karen C. Davis Sandipto Banerjee (2007)," Teaching and Assessing a Data Warehouse Design Course", 24th British National Conference on Databases (BNCOD'07) 0-7695-2912-7/07 $20.00 © 2007 IEEE.

[38] Mayank Sharma, Navin Rajpal and B.V.R.Reddy (2010)," Physical Data Warehouse Design using Neural Network" . International Journal of Computer Applications 1(3):86–94, February 2010. Published By Foundation of Computer Science.

[39] Abraham Silberschatz, Henry F. Korth, and Sudarshan (2002). Database System Concepts pp 445- 489. 4th Edition, McGraw Hill.

[40] Marotta, A., Ruggia, R.(2002)," Data warehouse design: a schema-transformation approach", Computer Science Society, 2002. SCCC 2002. Proceedings. 22nd International Conference of the Chilean.

AUTHORS

Dr. (Mrs.) Rajni Jindal is working as an Associate Professor at Delhi College of Engineering(Now Delhi Technological University. She received her M.E. (Computer Technology & Applications) degree from Delhi college of Engineering. She completed her PhD (Computer Engineering) from Faculty of Technology, Delhi University in the area of Data Mining. She possesses a work experience of around 21 years in research and academics. Her major areas of interest are Database Systems, Data Mining & Data Warehouse and Operating systems. She has authored around 40 research papers and articles for various national and international journals/conferences. She has also authored 3 books. She is a life member of professional bodies like Computer Society of India (CSI) and senior member of Institute of Electrical Engineers (IEEE), USA.

Shweta Taneja is a research scholar in Computer and Engineering Department at Delhi College of Engineering (Now Delhi Technological University. She received her M.Tech(Information Systems) degree from Netaji Subash Institute of Technology, Delhi University. Her areas of interest are Data Warehousing, Data Mining and Database Management Systems

A Typical Case of Recommending The Use Of A Hierarchical Data Format

Elisa Margareth Sibarani[1], Paul Wagenaars[2], Guido Bakema[3]

[1]Del Polytechnic of Informatics, Toba Samosir, North Sumatra, Indonesia
[2]KEMA, Arnhem, the Netherlands
[3]HAN University of Applied Sciences, Arnhem and Nijmegen, The Netherlands

ABSTRACT

Storing large amounts of data in an efficient way from the point of view of very fast retrieval is an important requirement for many industries like KEMA, a company offering consulting, testing and certification in the energy business. Because their relational database could not cope fast enough with the large amounts of data involved, an alternative way for data storage was proposed, called Hierarchical Data Format 5 (HDF5). HDF5 is a data model, a library, and file format for storing and managing data. Four hierarchical designs for storing and retrieving the large amounts of data involved were investigated. A benchmark was carried out in order to know which hierarchical structure would perform best. Eventually, a benchmark between HDF5 and MS SQL Server was carried out. It could be shown that HDF5 performs four (4) times better for inserting and even 200 times better for retrieving data than the MS. SQL Server.

KEYWORDS

Hierarchical Data Format (HDF), Relational Database, Hierarchical Design, Benchmarking

REFERENCES

[1] Fred Steennis, Ad Kerstens, Theo van Rijn, Jan Mosterd, Leon Bokma, Piet Soepboer, Alfred Arts, Nico van Donk and Branko Carli, "On-line Partial Discharge Monitoring for MV Cables with SCG – Practical Experiences in the Period 2007 to 2010", 8th Int. Conf. on Insulated Power Cables (Jicable), Versailles, France, 19-23 June 2011.

[2] Sibarani, Elisa Margareth, Simanjuntak, Humasak T.A. & Lumbantoruan, Rosni, (2010) "A Solution to Partial Discharges Online (PD-OL) System in KEMA", Thesis Report in Hogeschool Van Arnhem en Nijmegen (HAN) University of Applied Sciences, Arnhem, The Netherlands.

[3] Why HDF? www.hdfgroup.org. [Online] September 23, 2009. [Cited: July 4, 2010.] http://www.hdfgroup.org/why_hdf/.

[4] Oskam, Jaco (2010) PD-Online Architectural Design Document.

[5] Treinish, Lloyd A. Scientific data models for large-scale applications. www.research.ibm.com. [Online] [Cited: September 9, 2010.] http://www.research.ibm.com/people/l/lloydt/dm/DM.htm.

[6] DBMS. www.scribd.com. [Online] [Cited: June 4, 2010.] http://www.scribd.com/doc/6947984/DBMS.

[7] Products. www.hdfgroup.org. [Online] September 23, 2009. [Cited: July 4, 2010.] http://www.hdfgroup.org/products/.

[8] The HDF Levels of Interaction. www.hdfgroup.org. [Online] September 23, 2009. [Cited: July 5, 2010.] http://www.hdfgroup.org/products/hdf4/whatishdf.html.

[9] Scientific Data Management in the Coming Decade. Gray, Jim, et al. Redmond : s.n., 2005.

[10] Chapter 1 The HDF5 Data Model and File Structure. www.hdfgroup.org. [Online] March 12, 2010. [Cited: July 8, 2010.] http://www.hdfgroup.org/HDF5/doc/UG/UG_frame03DataModel.html.

[11] Introduction to HDF5. www.hdfgroup.org. [Online] February 6, 2006. [Cited: July 6, 2010.] http://www.hdfgroup.org/HDF5/doc/H5.intro.html.

[12] Group, Hierarchical Data Format (HDF). HDF5 Tutorial. University of Illinois at Urbana-Champaign (UIUC) : National Center for Supercomputing Applications (NCSA), October 1999.

[13] SQL Server Replication. msdn.microsoft.com. [Online] [Cited: August 2, 2010.] http://msdn.microsoft.com/en-us/library/ms151198.aspx.

[14] Steed, Chad A., Braud, James E and Koehler, Kim A. VGRID: A Generic, Dynamic HDF5 Storage Model for Georeferenced, Grid Data. aser.ornl.gov. [Online] http://aser.ornl.gov/steed/personal/publications/papers/2002_Steed-etal_VGRID_OCEANS.pdf.

[15] Improving Access to Multi-dimensional Self-describing Scientific Datasets. Nam, Beomseok and Sussman, Alan. p. 10.

[16] Comparison of relational database management systems. en.wikipedia.org. [Online] [Cited: August 5, 2010.] http://en.wikipedia.org/wiki/Comparison_of_relational_database_management_systems.

[17] Mainzer, John. RFC: Metadata Journaling to Improve Crash Survivability. www.hdfgroup.org. [Online] August 22, 2008. [Cited: August 1, 2010.] http://www.hdfgroup.org/pubs/rfcs/Metadata_Journaling_RFC.pdf.

[18] SQL Server 2008 R2 Pricing. www.microsoft.com. [Online] [Cited:August 9, 2010.] http://www.microsoft.com/sqlserver/2008/en/us/pricing.aspx.

[19] SQL Server 2008 Troubleshooting and Support. technet.microsoft.com. [Online] [Cited: August 9, 2010.] http://technet.microsoft.com/en-us/sqlserver/bb895929.aspx.

[20] Helpdesk and Mailing Lists. www.hdfgroup.org. [Online] [Cited: August 8, 2010.] http://www.hdfgroup.org/services/support.html.

[21] HDFView. www.hdfgroup.org. [Online] [Cited: July 1, 2010.] http://www.hdfgroup.org/hdf-java-html/hdfview/.

[22] PyTables Getting the most *out* of our data. www.pytables.org. [Online] [Cited: July 8, 2010.] http://www.pytables.org/moin.

[23] Järvinen, Jani. The Basics of Manipulating File Access Control Lists with C#. www.developer.com. [Online] [Cited: August 3, 2010.] http://www.developer.com/net/article.php/10916_3701811_1/The-Basics-of-Manipulating-File-Access-Control-Lists-with-C.htm.

[24]   Microsoft, MSDN. Partitioning. MSDN Library. [Online] http://msdn.microsoft.com/en-us/library/ms178148.aspx.

AUTHORS

Elisa Margareth SIBARANI got a Master of Information Systems Development in HAN University of Applied Sciences in The Netherlands. At the moment she is head of and lecturer at the Informatics Engineering Study Program at Del Polytechnic Institute of Informatics in Indonesia. Her research interests are mainly about information systems development, software engineering, database performance and business intelligence.

Paul WAGENAARS received the MSc. degree in electrical engineering from the Eindhoven University of Technology in The Netherlands in 2004. In 2010, he received a PhD. degree from the same university for his research on online partial discharge monitoring of medium-voltage cable systems. Since then he joined KEMA as specialist on power cable technology.

Guido BAKEMA is a graduated theoretical mathematician from Leiden University and a informatician from Nijmegen University, both in The Netherlands. Early 2011 he retired as a professor in information systems development at HAN University of Applied Sciences in The Netherlands. In 2010 he supervised the first author during her final project and master thesis writing at the end of her master study on Information Systems Development in The Netherlands.