International Journal of Research in Advanced Engineering and Technology
Online ISSN: 2455-0876
www.engineersjournals.com
Volume 1; Issue 2; November 2015; Page No. 98-99

A brief literature study of data warehouse development techniques

Naveen Dahiya

Department of CSE, MSIT, New Delhi, India

Abstract
The paper discuss brief introduction to various techniques used in each phase of data warehouse development. Basically there are four phases of data warehouse design: requirement gathering, design, quality evaluation and data extraction. The techniques used in each phase of data warehouse development are introduced in this literature study.

Keywords: Data Warehouse, Requirement gathering, quality evaluation, data extraction.

1. Introduction

A information delivery system makes use of data warehouse for information delivery. A data warehouse is capable to provide strategic information about customer needs and preferences, emerging technologies. The process of data warehouse development follows an incremental approach starting from requirement gathering, design, quality evaluation and finally data extraction.

Requirement gathering is the initial phase of data warehouse development. The users are interviewed and joint application sessions conducted for segregation of initial requirements towards building of a data warehouse. Then conceptual designing techniques such as star ER, dimension fact, multidimensional ER, OODM are applied to data collected in requirement gathering phase towards building of a conceptual model. The conceptual design is the initial phase of data warehouse design. It is 100 times more economical to detect and rectify errors in the initial phase rather than at the end of any software development.

The next phase after design of conceptual model is the quality evaluation of conceptual models built in design phase. The quality evaluation of conceptual models can be predicted by quality metrics that are based on size and structural complexities. Serrano et al. (2008) [1] discussed two methods for validation of quality metrics namely theoretical and empirical validation. After quality evaluation, the model can be subjected to various techniques for information extraction such as data mining, complex querying and statistical analysis techniques.

2. Literature study

This section present various techniques used in each phase of data warehouse development namely requirement gathering, design, quality evaluation and information extraction. Table 1 show various techniques used in development phases of data warehouse development and the research papers showing relevant references.

Table 1: Data warehouse development techniques (Dahiya et al. 2013) [2]

Development Phases	Techniques	References
Requirement Gathering	1)Interview Techniques like open and closed ended interviews 2)Group discussions 3)Joint Application Development	Hofman(2011) [3], Haigh(2010) [4], Mellado et al.(2010) [5], Rodriguez et al.(2006) [6], Solar et al.(2008) [7]
Design	1)Star ER Model 2)Dimension Fact Model 3)Multidimensional ER Model 4)Object Oriented Dimensional Modeling 5)Extraction, Transformation, Loading Techniques	Bara et al.(2009) [8], Genero et al.(2007) [9], Haider and Kumar(2011) [10], Hendawi and Sappagh(2012) [11], Rifaie et al.(2009) [12]
Quality Evaluation	1)Theoretical Validation Techniques like Distance Framework, Zeuse Framework Approach 2)Empirical Validation Techniques like Surveys, Experiments, Questionnaires 3)Statistical Techniques like Correlation, Regression, Principal Component Analysis, Formal Concept Analysis, Fuzzy Logic	Batini et al.(2009) [13], Caballero et al.(2009) [14], Kefi and Koppel(2011) [15], Kpodjedo et al.(2011) [16], Blanco et al.(2008) [17], Moody(2008) [18]
Information Extraction	1)Data Mining Techniques like K-Means clustering, Neural Network based approaches 2)Querying 3) Statistical Techniques like Correlation, Regression, Principal Component Analysis, Formal Concept Analysis, Fuzzy Logic	Aggarwal et al.(2012) [19], Bobby and Lee(2009) [20], Bhamra et al.(2011) [21], Mojaveri et al.(2010) [22], Nejdar et al.(2009) [23]

3. Research Implications

The study of literature conducted will help the novice users to gain basic knowledge of data warehouse development and the techniques used in each phase of data warehouse development. The existing study will help the researchers to discover new dimensions towards efficient data warehouse development. On the basis of existing techniques new hybrid techniques can be designed for development of efficient information delivery systems.

4. Conclusion

The paper presents the study of literature conducted towards development of data warehouse development and the techniques used in each phase of data warehouse development. The relevant references to each of the techniques have been properly cited.

The paper can be extended by conducting a more detailed literature review in data warehouse development to explore newer dimensions of research and development of more efficient techniques for each phase of data warehouse development.

5. References

1. Serrano M, Calero C, Sahraoui H, Piattini M. 'Empirical studies to access the understandability of data warehouse schemas using structural metrics', Software Quality Journal. 2008; 16(01):79-106.
2. Dahiya N, Bhatnagar V, Singh M. 'Effective data warehouse for information delivery: a literature survey and classification,' International Journal of Networking and Virtual Organisations, 2013; 12(3):217-237.
3. Hofman R. Behavioral economies in software quality engineering', Journal of empirical software engineering. 2011; 16(02):278-293.
4. Haigh M. Software Quality, non-functional software requirements and IT-business alignment', Software Quality Journal. 2010; 18(03):361-385.
5. Mellado D, Blanco C, Sanchez L, Medina E. A systematic review of security requirements engineering, Computer Standards & Interfaces, 2010; 32(4):153-165.
6. Rodriguez A, Medina E, Piattini M. Security requirement with a UML 2.0 profile', 1st international conference on ARES, IEEE, 2006, 1-8.
7. Solar E, Stefanov V, Mazon J, Trujillo J, Medina E. Piattini M. Towards comprehensive requirement analysis for data warehouses: considering security requirements', 3rd international conference on ARES, IEEE, 2008, 104-111.
8. Bara A, Diaconita V, Lungu I, Velicanu M. Improving performance in integrated DSS with object oriented modeling', WSEAS Transactions on Computers, World Scientific and Engineering Academy and Society (WSEAS). 2009; 8(4):599-609.
9. Genero M, Manso, E, Visaggio A, Canfora G, Piattini M. Building measure based prediction models for UML class diagram maintainability', Journal of Empirical Software Engineering, 2007; (12):517-549.
10. Haider M, Kumar T. Materialized views selection using size and query frequency, Int. J. of Value Chain Management, 2011; 5(2):95-105.
11. Hendawi M, Sappagh S. 'EMD: entity mapping diagram for automated extraction, transformation, and loading processes in data warehousing', Int. J of Intelligent Information and Database Systems. 2012; 6(3):255-272.
12. Rifaie M, Kianmehr K, Alhajj R, Ridley M. Data modelling for effective data warehouse architecture and design, Int. J of Information and Decision Sciences. 2009; 1(3):282-300.
13. Batini C, Cappiello C, Francalanci C, Maurino A. Methodologies for data quality assessment and improvement, Computing Surveys, 2009; 41(3):1-52.
14. Caballero I, Vizcaino A, Piattini M. Optimal data quality in project management for global software developments' 4th international conference on COINFO, IEEE, 2009; 210-219.
15. Kefi H, Koppel N. Measuring data warehousing success: an empirical investigation applying the DeLone and McLean model' International Journal of Data Analysis Techniques and Strategies. 2011; 3(2):178-201.
16. Kpodjedo S, Ricca F, Galinier F, Gueheneuc Y, Antoniol, G. Design evolution metrics for defect prediction in object oriented systems', Empirical Software Engineering, 2011; 16(01):141-175.
17. Blanco C, Trujillo J, Fernandez-Medina E, Piattini M. Implementing multidimensional security in OLAP tools' 3rd international conference on ARES, IEEE, 2008, 1248-1253.
18. Moody D. Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions' Data & Knowledge Engineering, 2005; 55(3): 243-276.
19. Aggarwal N, Kumar A, Khatter H, Aggarwal V. Analysis of the effect of DM techniques on database', Advances in Engineering Software, 2012; 47(01):164-169.
20. Bobby D, Lee J. A framework for discovering relevant patterns using aggregation and intelligent data mining agents in telematics systems' Telematics and Informatics 2009; 26(4):343-352.
21. Bhamra G, Verma A, Patel R. Agent enriched distributed association rules mining: a review Proceedings of the 7th international conference on Agents and Data Mining Interaction, Springer Verlag, 2011, 30-45.
22. Mojaveri S, Mirzaeian E, Bornaee Z, Ayat S. New approach in data stream association rule mining based on graph structure' Proceedings of the 10th industrial conference on Advances in data mining: applications and theoretical aspects, Springer–Verlag, 2010; 1-12.
23. Nedjar S, Casali A, Cicchetti R, Lakhal L. Reduced representations of Emerging Cubes for OLAP database mining', Int. J of Business Intelligence and Data Mining. 2009; 4(3):267-300.