

Data Warehouse Testing and Security: A Conspectus

Sonali Mathur, Vikram Bali, S.L. Gupta, Payal Pahwa

Abstract: Data warehouse is a central storage facility that stores information from many sources which can be in structured or unstructured format, queries this information for retrieval based on certain input facts and delivers the outcome analysis to many analysts, to meet decision support and business intelligence requirement. Not much research has been carried out in this research area in the past few years. In this research paper, we are discussing the data warehouse architecture and the testing techniques that are used for best suited to be used for the data warehouses. Literature for the testing techniques is integrated at one place and the outcome is to focus on security issues while performing data warehouse testing.

Index Terms: Database, Data Warehouse, Data Warehouse Testing, Database Testing, Software Testing

I. INTRODUCTION

Data warehouse are decision support systems used for reporting and data analysis to enable the users (Executives, Managers and Analysts) to make effective and efficient decisions faster. Data warehouse technology has been successfully deployed in many industries for eg: financial services (for claims analysis, risk analysis, credit card analysis, and fraud detection), manufacturing industry (for order shipment and customer support analysis), retail industry (for user profiling and inventory management), transportation industry (for fleet management), telecommunications (for call analysis and fraud detection), utilities (for power usage analysis) and healthcare (for outcomes analysis).

A data warehouse consolidates the historical data, which helps the organization to analyze its business and predict the outcomes based on the trend analysis. It helps the executives to organize, understand and use their data to take strategic decisions. Data warehouse systems help in the integration of diversity of application systems.

The banking sector over the years, has accumulated large amount of customer data and operational data which is a valuable asset of the banks and this data when analysed efficiently can be used for report generation and for performing statistical analysis for various meta data collected to meet the needs of the bank customer analysis [1-2]. Thus the banking data warehouse technology shall help the banks to expand their business scope, improve customer service levels, strengthen internal management, help them further in development of healthy banks and also help them to be successful as compared to their competitors [3].

Revised Manuscript Received on June 15, 2019.

Sonali Mathur, Vikram Bali, JSSATE, UP, India

S.L. Gupta, OMAN

Payal Pahwa, BPIT, Rohini, India.

II. PROCEDURE FOR PAPER SUBMISSION

Databases are designed to make transactional systems run efficiently. Databases basically are called as OLTP (Online Transaction Processing Systems). They are not used for analytics. On the other hand, the primary usage of data warehouses are they are used for analytics. Therefore, data warehouse thus, can be called OLAP (Online Analytical Processing) database. The data warehouse takes the data from all the databases and creates a layer dedicated to data analytics. The databases are therefore designed to handle only transactions whereas data warehouses are designed to handle analytical processing of data.

Database consist of collection of data that is organized for storage, retrieval and accessibility of operations. Data warehouse on the other hand, integrates copies of transaction data from various sources and uses them for analytical purpose.

Each database is constrained to run on a single application whereas data warehouse accommodates data storage for any number of applications thereby stating that one data warehouse consists of infinite applications and infinite databases.

OLTP databases must be up every time as failure in the system will lead to chaos. The database is linked to the front-end application. OLAP databases are scalable and not linked to the front-end databases. Data is refreshed from the source system as needed and results in historical trend analytics and help in making business decisions.

OLTP database are optimized for performing read write operations for single point transactions. Performing large analytical queries is not preferred. OLAP databases are optimized for efficiently and effectively reading and retrieving large data sets and aggregating the data. The data warehouse are designed to handle large analytical queries.

The structure of the OLTP database is very complex as it consists of large number of joins because the data is normalized, and it is structures as there is no data duplication. On the other hand, the structure of OLAP database is not very complex and denormalized to enhance analytical query response time to specially facilitate analysis and reporting. Submit your manuscript electronically for review.

III. COMPARISON TABLE

DATABASE	DATAWAREHOUSE
They are OLTP (Online Transaction Processing Systems)	They are OLAP (Online Analytical Processing Systems)

It consists of collection of data	It consists of collection of copies of transaction data
It stores data on single application	It stores data from number of applications
Data is used for storage, retrieval and accessibility of operations	Data is used for analytical processing
They are linked to front-end applications; thus failure needs to be handled properly	They are not linked to front-end applications and are scalable
They are designed to perform single point transactions	They are designed to handle large analytical queries
The structure is complex and normalized	The structure is not complex and denormalized
Real time data processing is done	Data is refreshed from source systems when required

IV. DATA WAREHOUSE ARCHITECTURE

A data warehouse is a “subject-oriented, integrated, timevarying, non-volatile collection of data that is used primarily in organizational decision making [4]. A data warehouse is used by organizations to provide efficient data analysis to decision makers [5]. Figure 1 shows a typical three tier architecture of data warehouse where at the first level the data is gathered from various data sources (any organization database.e.g: CRM, Bank, ERP Systems, etc.) and transformed into the desired format through Extract, Transform and Load (ETL) Process at the second level. Then the desired data is stored in the data warehouse. At the third level, data analysis is performed using decision support systems and reports are generated for the users for the purpose of strategic decision making.

The three-tier architecture of data warehouse [6] can be explained as:

1. Enterprise Warehouse – It is the bottom tier of a data warehouse consisting of database server consisting of relational database system. The data is gathered from operational data sources and other external databases and data is extracted using application program interfaces known as gateways. A gateway is database management system and allows the client program to execute code. It consists of details comprising of all information about subjects related to the entire organization.

2. Data Mart – It is the middle tier of a data warehouse consisting of OLAP Server performing OLAP operations. The data mart is a subset of data warehouse and contains information related to a specific department or group of users.

3. Virtual Warehouse – It is the top tier of a data warehouse which consists of query, reporting, analysis or data mining tools. Analytical analysis is carried out and reports are generated for the users to forecast strategic business decisions.

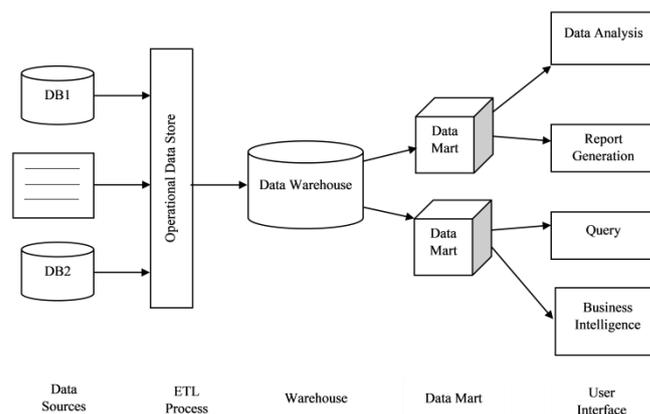


Figure 1. Architecture of Data warehouse

V. SOFTWARE TESTING VS DATA WAREHOUSE TESTING

In a data warehouse, the data passes through several stages, wherein each stage certain changes are made in the data before it finally reaches the user in the form of a chart, table or a report. Thus, at every stage of the data warehouse data should be tested to guarantee accuracy and preserve the quality of data. Data warehouse consisting of high-quality data is termed as a good data warehouse. Thus, before we see what testing techniques can be best applied to the data warehouse, let us see how software testing is different from data warehouse testing.

The difference between Software testing and Data warehouse testing [8] is represented in Table 1as shown below:

SOFTWARE TESTING	DATA WAREHOUSE TESTING
It is a carried out prior to deployment of software.	It is carried out post deployment of the application.
It focuses on testing of each use case which contains various test cases [7].	It is focused on querying the test data loaded by the ETL process.
It is source code specific.	It is content specific.
Focus is on generation and testing of test cases.	Focus is on query triggered operation in testing.
Tester bridges the gap between informal specifications and formal verification in software testing [7].	As volume of data is large it is not possible in data warehouse testing.
It is carried out through user interface.	It is primarily carried out on business logic
Instant or overnight result is generated by execution of transactions.	Execution of transactions take a long time due to large volume of data.
It is user triggered as the input is given by the user and individual transactions are processed.	It is system triggered due to Extraction, Transformation and Loading process [9].
Software defects found later in development lifecycle increases the development cost.	data warehouse focuses on correctness of data based on which critical business decisions shall be taken.



The testing process ends with the development life cycle.	The testing process continues after system delivery.
Either manual or automated testing is carried out on the test data.	Only automated testing is done on the test data.

Thus, it can be seen that software testing basically focuses on see and test whereas data warehouse testing is an automated testing where smart scripts are written to validate business. Thus, due to volume of data, processing may take time but the outcomes are efficient and effective as the reports are used for decision making purposes.

VI. DATA WAREHOUSE TESTING TECHNIQUES

With data being the most important factor in driving critical business decisions, testing the data warehouse becomes a critical process. Data in the data warehouse is integrated from numerous sources. The data source affects the data quality, so data profiling and data cleaning must be carried out. Data Warehouse testing consists of Extract, Transform and Load (ETL) testing to validate that the data has been transformed and loaded as expected and Business Intelligence (BI) testing to validate that the data shown in the reports is accurate [10]. Data Warehouse testing is very important because the business relies upon this data to make key decisions. Bad data and issues in Data Warehouse can lead to failures of the Data Warehouse projects because lack of trust in the data and consequently low usage.

Unlike software testing where a software is tested with the intent of finding errors, data warehouse testing is carried out to ensure that the data has been loaded from the source to the destination after business transformation is accurate. It also involves the verification of data at various middle stages that are being used between source and destination.

Much literature is available on what testing technique are best suited for the various stages of the data warehouse testing. In this section, we shall be integrating all the testing techniques to help us analyse the essential techniques that should be carried out at various stages of the data warehouse.

Reference	Published date	Testing Technique Proposed at Backend (DS -> ODS -> DW)	Testing Technique Proposed at Frontend (DW -> DM -> UI)	Conclusion
[12]	2007	Requirements Testing Unit Testing	Integration Testing User acceptance Testing	Unit testing is essential to be done at the backend to check the integration of data along with User Acceptance testing at the frontend to see that query analysis is as per the user requirements
[17]	2007	Unit Testing. Technical Shakedown Testing System Testing	Integration Testing Operation readiness Testing User acceptance Testing	Unit testing is essential to be done at the backend to check the integration of data along with User

				Acceptance testing at the frontend to see that query analysis is as per the user requirements
[13]	2008	Unit Testing	Integration Testing User acceptance Testing Performance Testing	Unit testing is essential to be done at the backend to check the integration of data along with User Acceptance testing at the frontend to see that report generated are satisfactory.
[18]	2008	Data Validation Testing Oneshot/Retrospective Testing View Testing	Prospective Testing Regression Testing	Data validation testing is essential to be done at the backend to check the integrity of the data along with Regression testing at the frontend to check the data functions correctly after the changes has been made.
[11]	2009	Functional Testing Usability Testing Stress Testing	Performance Testing Security Test Recovery Test Regression Test	Functional testing is essential to be done at the backend to check the data is compliant as per the business requirements along with Security testing at the frontend to check the data is protected and functionality is maintained.
[19]	2009	Multidimensional database testing Data pump (ETL) testing	Metadata Testing OLAP testing.	ETL testing is essential to be done at the backend to check the data validation along with OLAP testing at the frontend for report and query analysis



DATA WAREHOUSE TESTING AND SECURITY: A CONSPECTUS

[9]	2010	Extraction Testing Transformation Testing Load Testing Stress and Volume Testing	Adhoc Query Testing End to End Integrated Testing Security Testing Parallel Testing	ETL testing is essential to be done at the backend to check the data validation along with Security testing at the frontend to check the data is protected and functionality is maintained.
[15]	2010	Unit Testing	Integration Testing System Testing User acceptance Testing	Unit testing is essential to be done at the backend to check the integration of data along with User Acceptance testing at the frontend to see that query analysis is as per the user requirements
[16]	2011	End to End Testing Row Count Testing Field Size Testing	Report Testing	End to End testing is essential to be done at the backend to check the data has been loaded along with Report testing at the frontend for report and query analysis.
[20]	2012	Functional Test Usability Test	Security Test Performance Test Maintainability Test	Functional testing is essential to be done at the backend to check the data is compliant as per the business requirements along with Security testing at the frontend to check the data is protected and functionality is maintained.
[21]	2013	New Data warehouse Testing Migration Testing Change Request	Report Testing	New data warehouse testing is essential to be done at the backend as complete transformed data is

				loaded along with Report testing at the frontend for report generation and query analysis as per user business requirements .
[22]	2015	ETL Testing	Security Testing Regression Testing	ETL testing is essential to be done at the backend to check the data validation as bank database is voluminous along with Security testing at the frontend to check the data is protected and functionality is maintained.
[23]	2015	ETL Testing	Automated Testing	Testing ETL (Extract, Transform, and Load) procedures is an important and vital phase during testing of Data warehouse as it affects the quality of data, so automated testing improves data quality in less time, cost and attains good data quality.
[25]	2016	New Data warehouse Testing Migration Testing Change Request	Report Testing	New data warehouse testing is essential to be done at the backend as complete transformed data is loaded along with Report testing at the frontend for report generation and query analysis as per user business requirements



[24]	2018	End to End Testing	Regression Testing	End to end testing at the backend helps to verify that the data warehouse system meets its design specifications and other requirements and Regression Testing at the frontend to support agile and iterative development processes.
------	------	--------------------	--------------------	--

Contrary, on the other hand, data warehouse testing when compared with the software testing techniques that are carried out for testing of the software, the literature available is integrated in the table given below.

Reference	Published date	Software Testing Technique Proposed at Backend (Database)	Software Testing Technique Proposed at Frontend (GUI)	Conclusion
[26]	2010	Correctness Testing Performance Testing Security Testing Stress Testing Reliability Testing	Unit Testing Regression Testing	Performance and Security Testing are essential to be performed for checking data validation at the backend whereas Regression testing is performed at the frontend to check the application performance .
[27]	2012	Correctness Testing Performance Testing Security Testing Reliability Testing	Manual Testing Unit testing Integration testing Acceptance / Validation testing System testing	Security Testing aims to verify that protection mechanisms built into a system thus it is essential to be performed at backend whereas to check the functionality of a system System Testing is performed at the front end.
[28]	2014	Performance Testing Security Testing Reliability	User Acceptance Testing	Security Testing aims to verify that protection mechanisms

		Testing		built into a system thus it is essential to be performed at backend whereas to check the User acceptance Testing is performed at the front end so that the software is completely accepted by the user.
[29]	2016	Performance Testing Security Testing Reliability Testing System testing	Unit testing Integration testing	Security Testing aims to verify that protection mechanisms built into a system thus it is essential to be performed at backend whereas to check the functionality of a system System Testing is performed at the front end.
[30]	2018	Performance Testing Security Testing Load Testing Stress Testing	Unit Testing Fuzz Testing Regression Testing System Testing	Security Testing aims to verify that protection mechanisms built into a system thus it is essential to be performed at backend whereas to check the functionality of a system System Testing is performed at the front end.



VII. DISCUSSION

In software systems, Front-end Testing is testing or verifying the frontend functionality, GUI, and Usability. The main aim of Frontend testing to make sure that every user is well-protected from bugs. Figure 2 represents the most frequently software testing techniques used at the frontend. User Acceptance testing is done at the front end so check that the application made is as per the user requirements.

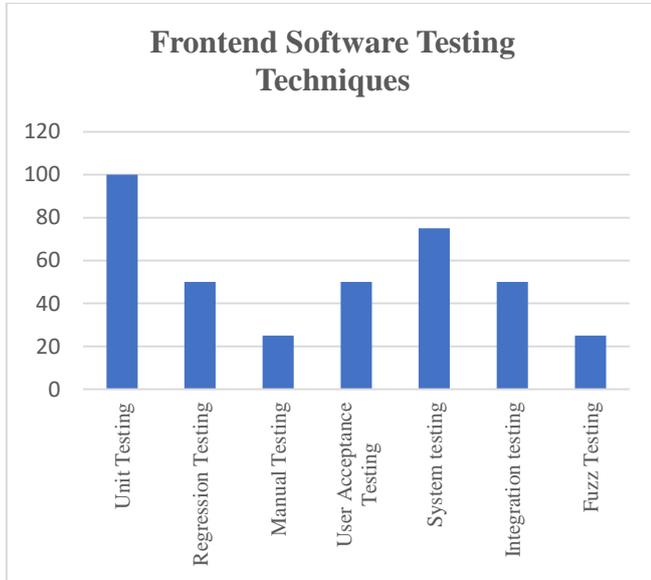


Figure 2.Frequently used Software Testing Techniques at Frontend

On the other hand, Backend testing is defined as a type of testing that checks the server side or Database. It is also known as Database Testing. The data entered in the front end will be stored in the back-end database. From the literature gathered, it can be concluded that Performance and Security Testing are performed at the backend as shown in Figure 3 which also represent the frequently used backend software testing techniques.

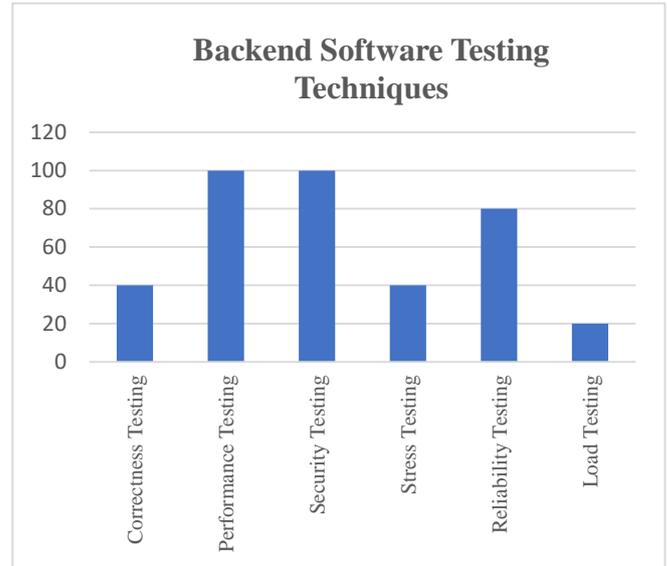


Figure 3.Frequently used Software Testing Techniques at Backend

In data warehouse systems, in the frontend testing, regression, security or report testing takes place to check the generated query or report analysis is as per the customer requirements. Figure 4 represents the most frequently used front end testing techniques for a data warehouse.

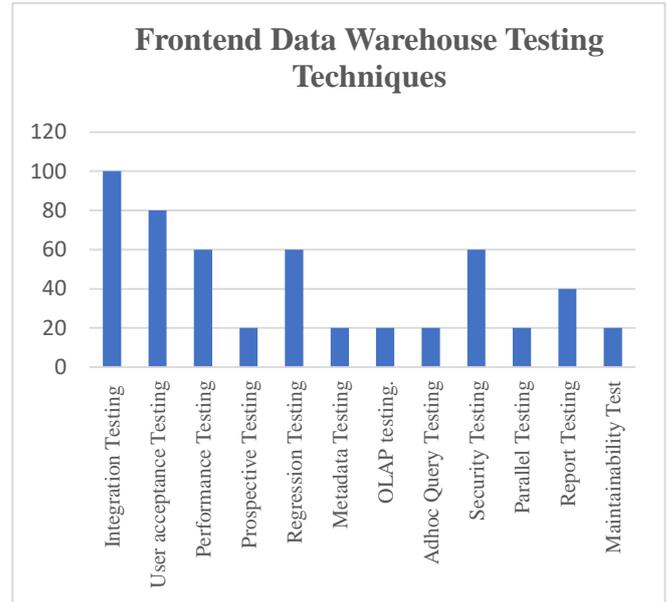


Figure 4.Frequently used Data Warehouse Testing Techniques at Frontend

The backend testing data warehouse is tested using mainly ETL Testing to check for data validation and there is no loss in quality of data during the ETL process. Figure 5 represents the most



frequently used testing techniques at the backend testing of a data warehouse.

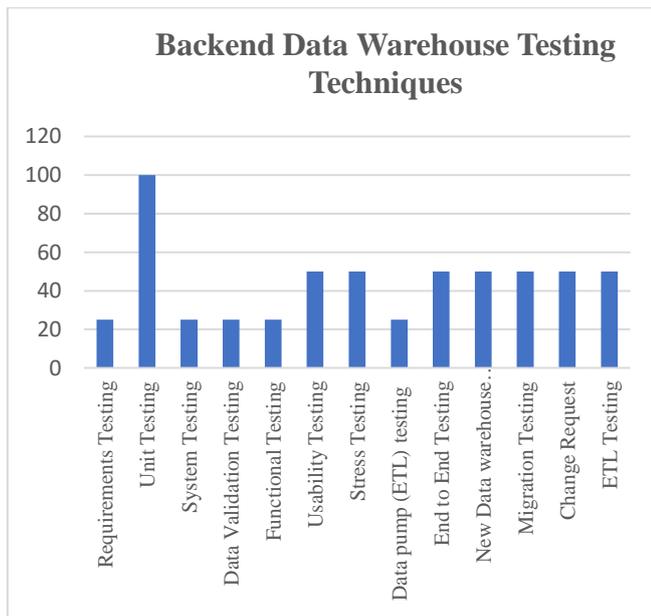


Figure 5. Frequently used Data Warehouse Testing Techniques at Backend

VIII. CONCLUSION

In this paper, we have integrated the literature available over the years, about the testing techniques which are being used for testing in the data warehouse. As the architecture of the data warehouse has evolved, thus the testing techniques will also evolve according to the architecture of the data warehouse. In future, we would like to apply the frequently used techniques in a proposed architecture featuring more on the security aspects of data in data warehouse testing. Apart from maintaining high quality of data in a data warehouse, the security of the data should also be focused to get a secure data at the target in the data warehouse.

REFERENCES

1. Färber F, Cha S K, Primsch J, et al. SAP HANA database, "Data management for modern business applications[J]", ACM Sigmod Record, 2012, 40(4): 45-51.
2. Chen C L P, Zhang C Y, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data[J]", Information Sciences, 2014, 275: 314-347.
3. Ranjan J, "Business intelligence: Concepts, components, techniques and benefits[J]", Journal of Theoretical and Applied Information Technology, 2009, 9(1): 60-70.
4. W. H. Inmon, "Building the data warehouse" (the 2nd edition) [M], New York: Wiley Computer Pub., 1996.
5. J. A. Senn, "Information technology in business: principles, practices, and opportunities", Prentice-Hall, Inc., 2000
6. [Online] http://dataminingzone.weebly.com/uploads/6/5/9/4/6594749/c-h4_dw_architecture.pdf, (2013) April 9.
7. Sneed M. Harry, "Testing a Data Warehouse – an Industrial Challenge", in proceedings of the Testing: Academic & Industrial Conference on Practice and Research Techniques, IEEE Computer, p. 203-210, 2006
8. Gupta S L, Pahwa P, and Mathur S, "Classification Of Data Warehouse Testing Approaches", in International Journal Of Computers & Technology, vol. 3, no. 3a. pp. 381–386, 2012.
9. Executive-MiH, "Data Warehouse Testing is different", 2010
10. online <https://panoply.io/data-warehouse-guide/data-warehouse-testing-101/>

11. M. Golfarelli and S. Rizzi, "A Comprehensive Approach to Data Warehouse Testing," in ACM 12th international workshop on Data warehousing and OLAP (DOLAP '09) Hong Kong, China, 2009.
12. K. Brahmshatriya, "Data warehouse testing", <http://www.stickyminds.com>, 2007.
13. A. van Bergenhenegouwen, "Data warehouse testing", <http://www.ti.kviv.be>, 2008.
14. R. Cooper and S. Arbuckle, "How to thoroughly test a data warehouse", In Proc. STAREAST, Orlando, 2002.
15. Breue T., "Turning data into Dollars", XLNT Consulting, 2010
16. Ceres E., 2011, "Data Warehouse Testing: Why QA projects need Automation", white paper published for Real-Time Technology Solutions.
17. Mookerjea A. and Malisetty P., 2008, "Data Warehouse/ETL Testing: Best Practices", www.pureconferences.com.
18. S. Bhat, "Data Warehouse Testing - Practical," in www.Stickminds.com, 2007.
19. P. Tanuška, O. Moravčík, P. Važan, and F. Miksa, "The Proposal of Data Warehouse Testing Activities," in 20th Central European conference on Information and Intelligent Systems, Varaždin, Croatia, 2009, pp. 7-11.
20. Golfarelli, M. and Rizzi, "Data Warehouse Testing: A Prototype based Methodology", in Information and Software Technology Journal, Elsevier, Vol 53, Issue 11, Pg. 1183-1198, 2011
21. K. Vishnu Vandana, V. Sujatha, "ETL Testing in Data Warehousing", in International Journal of Engineering Sciences Research, Vol 04, Special Issue 01, 2013
22. C. Wenzhe, "The Application of Data Warehouse Technology in Modern Finance", in International Conference on Advances in Mechanical Engineering and Industrial Informatics (AMEII), 2015
23. S. B. Dakrory, T. M. Mahmoud, A. A. Ali, "Automated ETL Testing on the Data Quality of a Data Warehouse", in International Journal of Computer Applications (IJCA), Vol 131 – No.16, 2015
24. Online <https://dzone.com/articles/building-an-end-to-end-data-warehouse-testing-strategy>
25. P. Kumar, Dr. A. Saxena, "Advance Techniques Procedure and Challenges for Data Warehouse or ETL Testing", in International Journal Of Core Engineering & Management (IJCEM) Volume 2, Issue 10, 2016
26. Mohd. E. Khan, "Different Forms of Software Testing Techniques for Finding Errors", in International Journal of Computer Science (IJCSI) Issue, Vol. 7, Issue 3, No 1, 2010
27. A. A. Sawant, P. H. Bari, P. M. Chawan, "Software Testing Techniques and Strategies", in International Journal of Engineering Research and Applications (IJERA), Vol. 2, Issue 3, 2012
28. A. Bansal, "A Comparative Study of Software Testing Techniques", in International Journal of Computer Science and Mobile Computing, Vol.3 Issue.6, 2014
29. S. R. Jan, S. T. U. Shah, Z. U. Johar, Y. Shah, F. Khan, "An Innovative Approach to Investigate Various Software Testing Techniques and Strategies", in IJSRSET, Vol 2, Issue 2, 2016
30. A. Takanen, J. Demott, C. Miller, A. Kettunen, "Fuzzing for software security testing and quality assurance", 2018.
31. G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
32. W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
33. H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.
34. B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.
35. E. H. Miller, "A note on reflector arrays (Periodical style—Accepted for publication)," *IEEE Trans. Antennas Propagat.*, to be published.
36. J. Wang, "Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication)," *IEEE J. Quantum Electron.*, submitted for publication.
37. C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.
38. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces (Translation Journals style)," *IEEE Transl. J. Magn. Jpn.*, vol. 2, Aug. 1987, pp. 740–741 [*Dig. 9th Annu. Conf. Magnetism Japan*, 1982, p. 301].
39. M. Young, *The Technical Writers Handbook*. Mill Valley, CA: University Science, 1989.
40. (Basic Book/Monograph Online Sources) J. K. Author. (year, month, day). *Title* (edition) [Type of medium].



- Volume(issue). Available: [http://www.\(URL\)](http://www.(URL))
41. J. Jones. (1991, May 10). Networks (2nd ed.) [Online]. Available: <http://www.atm.com>
 42. (Journal Online Sources style) K. Author. (year, month). Title. *Journal* [Type of medium]. Volume(issue), paging if given. Available: [http://www.\(URL\)](http://www.(URL))
 43. R. J. Vidmar. (1992, August). On the use of atmospheric plasmas as electromagnetic reflectors. *IEEE Trans. Plasma Sci.* [Online]. 21(3). pp. 876—880. Available: <http://www.halcyon.com/pub/journals/21ps03-vidmar>

AUTHORS PROFILE



Dr. Vikram Bali has received his B.Tech (CSE) from REC, Kurukshetra, M.E. (CSE) from NITTTR, Chandigarh and Ph.D from Banasthali Vidyapith, Rajasthan. He has more than 18 years of rich academic experience. He is a Professor & Head of Department (CSE) at JSS Academy of Technical Education, Noida. He is life time member of Indian Society for Technical Education (ISTE), Computer Society of India (CSI) and Institution of Engineers (IE). He has contributed 21 Research papers in International Journal and 7 Research papers in National Conferences/ proceedings and Edited Books. He has also attended Faculty Enablement programme organised by Infosys and NASSCOM. He has been the member of board of studies of different Indian Universities and member of organizing committee for various national and international seminar/conferences. He has written books on Fundamental of "Cyber Security and Laws", "Software Engineering" and "Operating System". He is reviewer to many International Journals of repute like Inderscience and IGI Global. His research interest includes Software Engineering, Cyber Security, Automata Theory, CBSS and ERP.



She is B.Tech (CSE) M.E. (CSE). She has more than 10 years of rich academic experience. He is a faculty at JSS Academy of Technical Education, Noida. She is life time member of Indian Society for Technical Education (ISTE), Computer Society of India (CSI) and Institution of Engineers (IE). she has contributed 15 Research papers in International Journal and 7 Research papers in National Conferences/ proceedings.

