International Journal of Engineering Sciences Paradigms and Researches, Vol. 02, Issue 01, December 2012
An Indexed Referred and Impact Factor Journal
ISSN (Online): 2319-6564
www.ijesonline.com

Domain Level Analysis of Data Mart Quality Factors along with Best-fit Tests to Minimize Defects

Payal Pahwa[1] and Renu Miglani[2]

[1]Bhagwan Parshuram Institute of Technology, I.P University, Delhi
pahwapayal@gmail.com

[2] GVM Institute of Technology & Management, DCRUST Murthal, Sonepat, Haryana
rnkakkar@gmail.com

Abstract

In modern epoch Data Warehouse is a crucial business application and used extensively by management in various types of organizations to take strategic decisions. A huge amount of data is transformed, incorporated, structured, purified and assembled during the construction and expansion of DW. Decision making in an organization relies heavily on the accurate retrieval of facts. Any faults in these facts results in business loss. DW testing is an activity that is intended for evaluating quality of a DW and also for improving it, by identifying these faults. Therefore testing is a crucial phase of any DW development procedure to improve its quality. It is too difficult, too costly, too impolitic and requires too much more efforts to test the huge amount of data warehouse, for most of the organizations. Therefore the concept of data mart arises. Testing of a data mart is much easier and manageable process. Data Mart testing cannot test the quality of data directly instead it measures the related factors that give us the estimation of quality. In our paper we have discussed various quality factors related with different domains of DW/Data Mart and also walk through some of the basic tests to improve the quality. As a single test can measure more than one quality attributes, so any organization can select best fit test according to their requirements and importance of the particular quality factor.

Keywords*: Data mart, Data Warehouse, Data Mart testing, Data Mart quality, Data Mart quality factors.*

1. Introduction

A Data warehouse is a cohesive data model that defines the central data repository for an organization[1]. A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision-making process[2]. It is a composite and collaborated data model that captures the entire data of an organization[3]. It is the main repository of any organization's historical data. It contains the material and information for management's decision support system[4] whereas data mart, is a decision support system incorporating a subset of the enterprise's data focused on specific functions or activities of the enterprise. Data marts may incorporate substantial data, even hundreds of gigabytes, but they contain much less data than would a data warehouse developed for the same company. Since data marts are focused on relatively specific business purposes, system planning and requirements analysis are much more manageable processes, and consequently design, implementation, testing and installation are all much less expensive than for data warehouses. As Data Mart is constructed from heterogeneous sources, therefore quality is key concern. So, testing the quality of the resulting information will support the trust worthiness of the system[5]. Thus testing is an indispensable phase of Data Mart development process. It is an activity that is intended for estimating the quality of the system and also for improving and standardizing it, by identifying and fixing the deficiencies and tribulations in any of its domain. It is a process of verifying and validating that domains meet the business and technical requirements along with functionality as expected. As we know that quality cannot be tested directly but the measurement of related factors gives the estimation of quality. So Data Mart testing is not even test all the domains but also used to measure the software quality factors related with specific domains. Various quality factors related with Data Mart are correctness, usability, reliability, integrity, efficiency,

maintainability, flexibility, testability, portability, reusability, interoperability, scalability etc. The magnitude of any particular quality factor varies from application to application according to requirement of the business. Testing approach differs for different quality factors and good testing provides measures for all relevant factors. There are different types of tests to measure the quality factors. These attributes have been proposed to be tested using various types of tests such as Functional Test, Usability Test, Performance Test, Stress Test, Recovery Test, Security Test, Regression Test, Integration Test, Acceptance Test and Reliability Test. Each test cannot measure all the quality factors. So numbers of tests are required during various development phases to maximize the quality of data. They ensure that quality and testing goals are achieved to its maximum in a repeatable manner. Any organization can select best fit number of tests to measure the quality attribute according to their requirement to achieve their quality goal.

The layout of the paper is structured as follows: Section 1 gives a short-term introduction about Data Mart, Data warehouse, Data Mart Testing, quality factors and the type of tests that should be carried out to measure these factors for obtaining the high quality Data Mart. Section 2 discusses the Data Mart testing in brief and section 3 describes the different domains which are related to various quality factors. In section 4 a set of quality attributes has been defined. Section 6 specifies a variety of tests to the measure quality attributes to judge the quality of Data Mart. We explore the relationship among these tests and quality attributes in section 7 and the summarized result is concluded in section 8.

2. Data Mart Testing

Data Mart testing is an activity that is intended for evaluating quality of a Data Mart and also for improving it, by identifying fault and tribulations. It is also consider an investigation directed to provide information about the quality of data and Data Mart. As with any other piece of software a Data Mart implementation also undergoes the usual cycle of component testing, system testing, regression testing, integration testing and acceptance testing. However, unlike others there are no off the shelf testing products available for it[6]. It is launched at the requirement stage and continues during the development lifecycle to discover as many defects as possible in the system at earlier stages to diminish the instances of defects that may convert into failure. It is also used to test domains as well as measure the software quality factors related with specific domains like correctness, usability, reliability, integrity, efficiency, maintainability, flexibility, testability, portability, reusability, interoperability, scalability etc. Testing is mainly applied on these domains of Data Mart discussed in next section.

3. Data Mart Domains

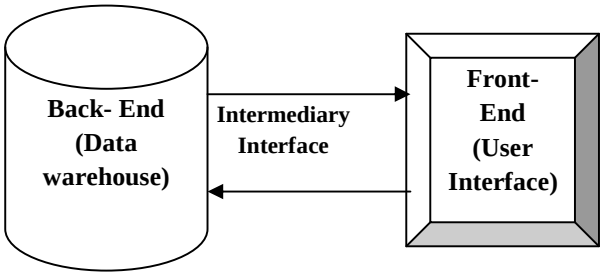
Data Mart testing will be appraised for three distinct domains.

Back- End: This is the lowest level in Data Mart framework that deals with the databases present in it. It may be Oracle, Microsoft SQL Server, IBM DB2 or any other database used for managing the Data Mart. It performs all the querying analysis and housekeeping tasks necessary for backup and maintenance of the Data Mart. The quality factors related with this domain are correctness, Usability, Maintainability, verifiability, Flexibility, Portability, Interoperability, testability, Reusability, integrity etc.

Front-End: This is the highest level of abstraction in Data Mart framework which hides out all the background details from the user. Front-End is the interface server between the user and the database. This domain is used for interaction with Back-End also known as user interface or user screen. It may be a CUI that accept input as a text string like SQL Query for accessing the information from the DW or it may be a GUI that interacts with the user using GUI features to manipulate the data. The quality factors related with this domain are Correctness, portability, Reusability, Integrity, Flexibility, Maintainability, Testability, Efficiency, Interoperability, Reliability, scalability etc.

Intermediary Interface: This domain consists of a layer of code along with some stored procedures sitting between Back-End and Front-End. It is necessary to bridge the database

records with the algorithms as both are implemented in different applications. This intermediary interface is of prime importance because it causes a smooth transition between the above two domains. Poor quality in intermediate products always proves harmful to the quality of the final product[7]. Therefore testing of interface is required as independent testing of Back-End and Front-End cannot detect most interface errors because the errors may generate during the interaction between these two modules. Interface testing is intended to discover defects in the interfaces of objects or modules. Tests are executed to make sure that code is really reading and writing the right data from the Data Mart[8]. All the quality factors are related with this domain.



Different quality factors related to above mentioned domains are important for the estimation of quality of Data Mart. All these factors are discussed in next section.

4. Data Mart quality factors

Quality cannot be tested straight forwardly although the correlated factors that make quality tangible can be tested. Each factor can be used for quality control, quality assurance as well as to measure the product performance. Some of the factors can be measured directly and some can be measured only indirectly like usability, maintainability etc. Software quality factors[9] are related with above mentioned domains has been described below

Correctness: Correctness is a mathematical property that establishes the equivalence between the Data Mart specification and fulfillment of requirement regarding accuracy, completeness, availability, standards, conciseness, consistency and up to datedness of information. This factor is measured directly. Data Mart is correct in terms of its intended functionality, calculations used internally and perfect navigation.

Usability: This factor is characterized by adequacy, learnability and robustness. It is measured in terms of ease of use. A Data Mart is user friendly if its interface is easy to operate. Input instructions should be adequate, clear, simple, flexible and the resulting information should be uniform and comprehensible for the user. Usability factor also measures the degree to which flexible data input is handled by putting plausibility checks on the input.

Flexibility: This attribute specifies the ability of Data Mart to add, alter or remove functionality without affecting the existing system. Extensibility is the subset of flexibility as it states the capability of system to add the functionality only. Functionality changes are required when there is a change in the objectives, goals, constraints, environment or any other aspect related to the organization. As changes are inevitable during the life time of system so this is one of the most important properties of quality Data Mart.

Reliability: Reliability is defined in terms of statistical behavior. It is defined as the probability that the Data Mart will operate correctly or as expected over a specified period of time for specified input conditions. It determine maximum system failure rate to provide service within specified time unit. The symbolization of reliability is comparative, if the consequence of a fault is not serious, the incorrect Data Mart may still be reliable. This factor is measured indirectly for correctness & availability.

Integrity: Integrity comes with security. A Data Mart is secured if its data is isolated and secured from an unauthorized access. Security is necessary to prevent illegal access of Data Mart functions, protection from virus attacks and privacy of data. It also determines the capability of Data Mart to withstand attacks.

Efficiency: This factor defines capability of a Data Mart to accomplish its intended purpose with the least utilization of all essential resources such as time, processor, buses, storage,

peripherals etc. It should utilize processor capacity, storage space, peripherals efficiently to produce maximum throughput.

Maintainability: Maintainability is little similar with flexibility but it emphasis on alterations about error corrections and minor function modifications not major function extensibilities. This factor depends on other attributes like readability, extensibility, flexibility and testability etc. It verifies that all the domains are easy to read, extend or modify and suitable for debugging. This factor is also measured indirectly.

Portability: Portability of a Data Mart is the ease with which it can be adapted to run in the environment other than the one for which it was designed. This factor is related with compatibility, adaptability and flexibility. It estimates the efforts required to transfer any of the domain from hardware or software system environment to another. A quality Data Mart executes on as much various platforms as it can.

Reusability: Reusability is corresponding to evolvability. During Data Mart evolution, a domain or module is modified to build the new version of the same object. It can be further reuse with minor changes to build another object. It is a cost efficient and time saving attribute. This attribute provides us the facility to develop new application with existing domain using only few efforts.

Testability: This attribute checks the suitability of all the domains regarding required functionality, fitness for debugging, ease to testing and locating defects. Testability regarding Data Mart is to confirm that it provides required information and it is suitable for extension and modification. A modular, well-structured Data Mart is more suitable for systematic, stepwise testing as compared to monolithic, unstructured DW.

Interoperability: Interoperability of one domain to another, one module to another and one procedure to another should be easy in a Data Mart. All of them should have a defined interface that is invoked by another object to communicate and should be integrated to achieve the common objective of Data Mart.

Scalability: This factor tests the ability of a Data Mart system to continue function well, when it is changed in size or volume in order to congregate an emergent need. This factor allows required modifications at the appropriate locations to be made without undesirable side effects. A scalable system responds user input in a tolerable amount of time even if load increases. Scalability depends on structure, readability and availability.

Verifiability: A Data Mart is verifiable if its attributes can be verified easily. Verifiability may be considered by integrated design, controlled development practices and use of appropriate technique. A common technique for improving verifiability is the use of software monitors i.e. code is inserted to monitor various attributes such as performance or correctness.

All the domains of Data Mart potentially contribute to errors, so it needs to drive test plans for every phase of Data Mart development process. Test plans specify the test to be conducted to measure the specific Data Mart quality attributes according to requirement of organization. The various types of tests are discussed in the next section.

5. Types of Tests

In the preceding section, a set of Data Mart quality factors were discussed for the measurement of Data Mart quality. In this section, we examine a pool of test that can be applied to it for the assessment of quality by calculating above mentioned quality factors. These tests can be used to develop an indication about the quality of Data Mart. Testing consists of different types of tests depending on the time the test is taking place[9]. Following types of tests[10] are best fit to measure the quality factors of data warehouse and provides the estimation of quality.

Functional Test: It verifies that data loaded in Data Mart meets external requirements and achieve goals of organization. It also verifies that a Data Mart does what it is supposed to do and does not do what it should not[7]. Functional test reads input from real system input files, exports results to real result files, then reads the result files back in and compares individual results to

expected values[11].

Usability Test: It evaluates the Data Mart by testing it on users and verifies that interface is easy to use and understandable as well as result is also comprehensible and meaningful to user. It commonly involves evaluating how well test material respond in four regions: efficiency, accuracy, recall, and emotional response.

Performance Test: Performance has always been a major concern and a driving force of Data Mart enlargement. It verifies that Data Mart performance is satisfactory under typical workload environment. Its assessment includes resource usage and throughput mainly. Performance testing is often done in conjunction with load and stress testing. Qualitative attributes such as reliability, scalability and interoperability may also be evaluated.

Security Test: A process to determine that a Data Mart protects data from unauthorized access to the system and information or resources within it. This test recognizes and eliminates any flaws that may lead to security violations and validate the effectiveness of security measures. It makes sure that the Data Mart is protected from virus attacks and protecting the privacy of data stored in it. It maintains the Data Mart functionality as intended according to organization need.

Stress Test: It determine the stability and level of performance of a Data Mart during peak and very heavy work load to ensure that it behaves decently and does not crash in conditions of insufficient computational resources. In this test the load is intentionally distorted to force the system into resource depletion and system is stressed to the breaking point in order to find potentially harmful bugs.

Integration Test: Two units that have already been tested are combined into a component and the interface between them is tested. This focuses on testing multiple modules or domains working together to achieve the common quality goal. The intention of integration testing is to confirm functional, performance, and reliability on key design objects

Recovery Test: It defines how well a Data Mart is able to recover from crashes, power shutdown, hardware failures, interface failure, insufficient memory and other similar problems. It checks the proficiency of recovering quickly with least human interference. A recoverable system maintains a log of activities happening before the crash and a record of messages generated during the failure and reactivation of system.

Reliability Test: Testing for reliability is about exercising the Data Mart so that failures are discovered and removed before it is deployed. It estimates present reliability and predict future reliability on the basis of failure data and an estimation model. It also determines utmost system failure rate to provide service within précised time unit.

Regression Test: This test is done to ensure that the older functionality of Data Mart has not been degraded with the new changes. If a piece of any domain is modified for any reason, testing needs to be done to ensure that it works as specified and it has not negatively impacted any functionality that Data Mart offered previously.

Acceptance Test: The process to obtain confirmation that a Data Mart meets mutually agreed requirements. This test checks system for unexpected interactions among the units, modules or domains and also evaluates it for compliance with the functional requirements. It is a final stage of testing that is performed prior to the system being delivered to a live environment including user.

All the above mentioned tests are used to measure different quality factors. The relation of these tests with the specified quality factors is described in next section.

6. Relation of Quality Factors with different Tests

As each test has been designed to examine the Data Mart quality factor but not any single test is capable to calculate all the related factors. So tester has to choose a group of tests to estimate all the factors. The relationship among quality factors and different tests is shown in table 1.

Table 1 : Relationship among Quality Factors and tests

Factors/ Tests	Correctness	Integrity	Usability	Reliability	Efficiency	Maintainability	Flexibility	Testability	Portability	Reusability	Verifiability	Interoperability	Scalability
Functional test	√	√						√			√	√	√
Usability test	√		√								√	√	
Performance test	√										√	√	
Stress test	√	√		√	√			√			√	√	
Recovery test	√	√		√		√		√			√	√	
Security test	√	√						√			√	√	√
Regression test	√	√		√		√	√	√	√	√	√	√	√
Integration test	√			√				√		√	√	√	
Acceptance test	√	√	√	√	√	√	√	√	√		√	√	√
Reliability test	√	√		√				√			√	√	√

This observation table is based upon my observation and understanding of the subject. The importance of any quality factor varies from application to application according to requirement of the organization. Good testing provides measures for all relevant factors. Any system where human lives are at stake must place extreme emphasis on reliability and integrity. In the typical business system usability and maintainability are the key factors, while for a one-time scientific program neither may be significant. Any organization can select any one test or the combination of tests best fit to their requirement and importance of quality factors. Our testing, to be fully effective, must be geared to measuring each relevant factor and thus forcing quality to become tangible and visible[7].

Lastly a few words of concern to end with, testing a Data Mart application should be done with a sense of utmost responsibility. A bug in a Data Mart traced at a later stage consequence in unpredictable losses and the task is even harder in the lack of any single end-to-end testing tool. So the strategies for testing should be systematically developed, refined and streamlined. This is also true since the requirements of a Data Mart are often changing dynamically, so the vital area of concern is test coverage and it has to be reviewed numerous times to make sure about completeness of testing.

7. Conclusion

Once Data Mart has been developed, it must be tested to uncover as many errors as possible before using it in real environment because error traced at later stages may cause huge losses to organization. So Data Mart testing is a critical element of Data Mart quality assurance. Its quality can be computed by using the measurement of various quality attributes related with different domains. To measure these attributes, a series of tests have a likelihood of finding errors and computing quality factors. In our paper we have given a brief introduction of Data Mart domains and defined various quality attributes related with different domains along

with a set of tests to measure these quality attributes. Any organization can select a single test or a bunch of tests best fit to organization requirement and importance of quality factor to improve the quality of Data Mart.

References

[1] Rebecca Cooper and Suellen Arbuckle, "How to Thoroughly Test a Data Warehouse", STAREAST Orlando, 2002.
[2] W. H. Inmon, "Building the Data Warehouse", John Wiley and Sons, Second Edition, March 27, 1996.
[3] Manoj Philip Mathen, " Data WarehouseTesting", , Developer IQ Magizine, Infosys Technologies Limited, Mar 2010,
[4] Anindy. Mookerjea and prasanthMalisetty, "Best practices in data warehouse testing", Proc. Test, Pure Conference Solution Pvt Ltd., 2008.
[5] NeveenElGamal,Ali El BastaWissy and GalalGalal-Edeen, "Towards a Data Warehouse Testing Framework" , Ninth International Conference on ICT and Knowledge Engineering, 2011
[6] Soumyajit Sarcar, "Testing Data Warehouse Applications− the ETL Perspective", 2008.
[7] Nasib Singh Gill, Software Engineering, Khanna Book Publishing Co.(P) Ltd., 2002
[8] http\\www.Dallaway.com
[9] J. Mc. Call, P. Richards, and G.Walters, "Factors in software quality", Technical Report, 1977.
[10] Matteo Golfarelli and Stefano Rizzi, "A comprehensive Approach to Data Warehouse Testing", DOLAP'09, November 2009.
[11] http\\www.c2.com