TOWARDS A DATA WAREHOUSE TESTING FRAMEWORK

Bharath kumar R[*1]
Nachiyappan.s[2]
[1*] VIT University Chennai, Tamil Nadu, India
[2]Faculty of SCSE School,VIT University Chennai, Tamil Nadu, India

ABSTRACT
This Data warehouse is a non-standard data collection with administrative support in object-oriented, integrated, time-variant and decision making process. Direct quarries and analyzes were collected in the warehouses. The quality of the test database system in the data warehouse is determined. Test modules include several test methods, including test testing, computer testing, unit testing, and test testing. Here we use the ETL function. Advanced testing methods considered to be the best practices are the effects of this process. The test process results in the passing phase of the main database. Finally, the test structure is archived through the ETL process; to increase the result of higher qualified data warehouse test. We need to get the data warehouse test structure.

Keywords:
Data Warehouse, Testing ETL procedure, Testing database, testing activites, methodology.

INTRODUCTION

The data warehouse includes the title, application tools, structures, information service, and communications infrastructure, to integrate useful information from diverse distributed data sources. For this reason, vendors agree that DWs cannot be stacked, but should be designed and optimized for the customer's attention.
Many DW life cycle approaches were given in the literature to discuss how the DW system was built. In those approaches, building design is one of the early and major stages of developing DW systems. On the other hand, no experiments have been proposed in proposed life cycle approaches as it has always been considered in the approaches of all well-known life cycles, such as water and spindle samples.
Initially, the data stored in data sources (DS) is taken, converted and loaded in the database warehouse (DW). Sometimes this DW is specialized in specific groups of data databases, each containing data that is targeted to the source target.
The data passes through several changes and integration levels before the DW is loaded into DW on DS, all databases, all of the information we have created, will be later converted to DW.
The most common and comprehensive DW system includes elements commonly used in a DW program.
Although most test operations are carried out before the use of common software systems, the data warehouse continues after the computer output.

ETL AUTOMATION BENEFITS:

1. Reducing consumption time in testing phases and accelerating the implementation of automated testing experimental events.
2. The restructuring of the test, since saved tests can be used multiple times.
3. Make use of automated tools to generate test results.
4. Using experimental events generated by automated tools.

LITERATURE SURVEY

Two studies have introduced the classification of experimental operations: what was tested and how it was tested. "What" integration is the need for accurate checking of data mounted by ETL practices and concerns in data quality tests that are accessed by ranking tools. The second coordinate "how" is achieved by defining seven categories in addition to the essential characteristics of enormous information to break the components affecting the performance of large data applications. In large data applications, rapid development of small processing and computer customers, cell phones, changing information, and information receipt are taking place very rapidly with the increase in standard communication.
ETL testing may be a more complex and critical test phase because it avoids data quality directly. Since ETL is heavily code-based, most standard technologies for common software system testing are reused here.

TESTING THE DATABASE

Logic has already been verified during logical system testing and we consider ETL experiments are responsible for all the data related to data quality. Later, the database testing aims to test database performance mainly using standard (performance test) or heavy (pressure test) workloads. Like ETL, the amount of test databases and their data distribution should be discussed with designers and database administrators. Performance tests can be performed on a database, including real data or a pseudo database, but the database size should be compatible with the average expected data.

TESTING THE ETL PROCEDURES

The ETL processes are aimed at correctly extracting, cleaning, changing, and transferring data into the data. The best approach here is like a variety of tests. What is the problem of complex cleaning and change, and how to use the program implemented for developers. In particular, the key features to consider during the loading test will depend on both the dimensional tables. ETL procedures should be tested.

ETL errors are the most common reasons of misinformation in planning the right strategy. Generic techniques for handling the given types of errors \ Automatic Poor Data, \ "Wrong Data Rejected \", "Misuse Data for Data Management Administrator".

This phase by just checking that the test results, The three main phases of testing are

1. Create a test plan. The test plan describes the tests to be planned and their expected safety of system requirements.
2. Prepare test cases. Test cases have been implemented in the testing program by describing their test results as test results.
3. Execute tests. Each test resulted in a test execution and consequently.

TEST ACTION STEP

Once tested in selected items, successful testing events will be sent directly to the next study. In case of wrong experimental case (s), you can use several techniques regarding ETL errors defined during the essential analytics grid: false data that is maliciously cleaned, incorrect data is rejected, and data is misinterpreted for data administrator.

TESTING ACTIVITIES

To test different experiments better, we check and test what is being tested and how it is being tested.

The quality of data quality is mainly accessed by the exact precision test and front-end tools of the databases mounted by the ETL procedures. However, we need the opportunity to perform a useful test, which requires the database to be routed, and any standard is expected.

The test is not a man's activity. The test team must have testers, developers, designers, database administrators, and end users, and must be set up during the planning planning phase.

The data warehouse systems are often based on data. A successful test must be based on actual data, but it must contain pseudo-data to reproduce most common error environments that can be encountered in ETL. Precisely one of the most important activities carried out during the testing plan is to prepare the appropriate data set.

One of the basic components of the software is testing. Research and development departments cannot start the software without finding all possible defects. A test that checks whether the software is properly designed to properly implement one of the most important functions of the software process.

Test activities to ensure that data warehouse test operations are implemented, to achieve a better result.

The fundamental problem is that the size of the database is not equal to the amount of information stored. Only working properly when the working database was loaded.

All result reports prepared by the system showed that seed errors were detected, which demonstrates its performance in testing ETL level data quality.

By analyzing the result, the causes of these faults should the following points.

1. The process of searching does not work correctly because there are many fake values.
2. Sometimes we have created that wrong database and created. Finally will also get wrong decision.

IMPLEMENTATION DESCRIPTION

In this step, first place the notes in databases, source databases and our data warehouse. Secondly, ETL is involved in this operation. It contains source database, source table name, source program name, and change. Finally, a link to ETL practices.

EXISTING DW TESTING APPROACHES

There have been many trials for the DW test process. DW test process and from a different perspective. We're worried about the types of experiments to test when testing DW.

Introduced a DW test and validation technique. He broke the test and verification process into four well-defined and top-level processes

1. Integration testing
2. System testing
3. Data validation
4. Acceptance testing

Approach I: tends to follow the data from the source to the target warehouse.

Approach II: tends to follow the source through the Extraction Transformation Loading (ETL) process then into the target warehouse.

The divided the data warehouse testing into:

1. Requirement testing
2. Unit testing
3. Integration testing
4. Acceptance testing

These test types are Existing DW system should be used:

1. Functional test
2. Usability test
3. Performance test
4. Stress test
5. Recovery test
6. Security test
7. Regression test

We had to provide quality by checking our stock rules on the project. We met this challenge by deciding what the ETL process is.

Six types of testing should be conducted during the data warehouse operation with its ETL tools and applications. These tests are:

1. ETL Testing
2. Functional testing
3. Performance testing
4. Security testing
5. User acceptance testing
6. End to end testing

The data warehousing system focuses on how these tests are conducted in the data warehouse environment. Different tools like this test can use querysurge, sql server tools.

The data warehouse has several tests that require testing process. These tests must be classified by sql queries.

PROPOSED DQ TESTING FRAMEWORK SYSTEM

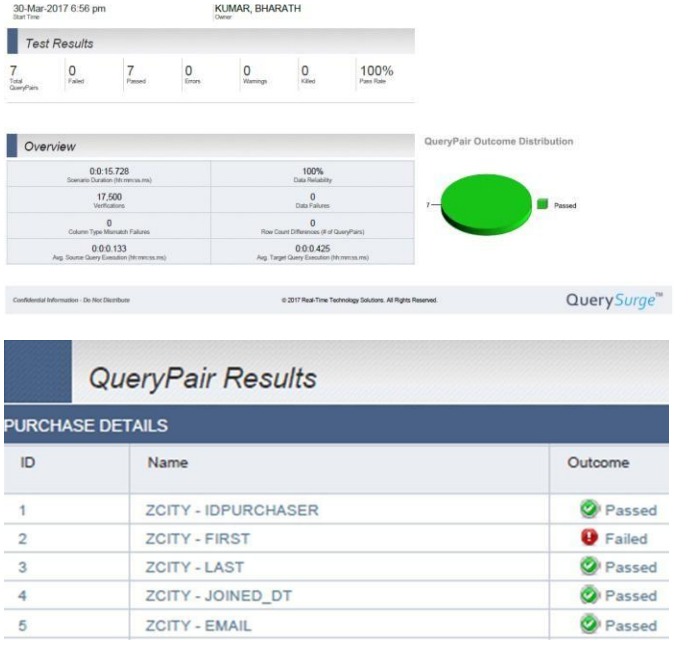
The proposed data quality framework main goal is to automate tests that check data quality in ETL process by automating the creation and execution of these tests.

The proposed framework that used to test for checking the data quality needed to achieve results reports.

Source to target checking if the data is passes successfully executed or else it might be fail.

The proposed framework consists finally we need to achieve 100% data validation or not we are checking on that if that data error its showing that data is fail.

SAMPLE OUTPUT OF THIS SCREENSHOT



CONCLUSION

In this paper, we have proposed testing strategies for automating ETL testing for data quality has been proposed. The proposed framework delivers a data quality The proposed framework delivers a data quality of database. As a result finally we should check 100% data validation or not.

REFERENCES

1) M.Golfarelli and S.Rizzi, “A comprehensive approach to data warehouse testing,” in ACM 12th international workshop on data warehousing and olap(dolap 09) china,2009. .
2) C.Bateman, “where are the articles on data warehouse testing and validation strategy?,” in www.information-management.com,2002..
3) K.Brahmkshatriya, “Data warehouse testing,” in 2007.
4) V.Rainardi, ”testing your data warehouse,” in building a data warehouse with example in sql server.2008..
5) inergy, “automated ETL testing in data warehouse environment.”,2007.
6) P.Tanuska, O.Moravcik,P.Vazan, and F.Miksa, “The proposal of the essential strategies of data warehouse testing,” in 19th central European conference on information and intelligent system(CECIIS),2008,pp.63-67..
7) R.K.Sharma, “Test automation: in data warehouse project,” in 2007.
8) Tarek M.Mahmoud, Sara B.Dakrory, and Abdelmgeid A.Ali, “Automated ETL testing on the data quality of a data warehouse,” international journal of computer application Volume 131-No.16, December 2015.