Testing Improvement in Business Intelligence Area

Sani Kumar[1], Dr. Hari Om Sharan[2]

[1](Research Scholar), Department Computer Science & Engineering Rama University, Kanpur 209217, India
[2](Associate Professor), Department Computer Science & Engineering Rama University, Kanpur 209217, India

---***---

Abstract - *In present, testing techniques have gradually involved from the practice of single programmers or small development teams into a systematic, managed engineering discipline. Not only have there been numerous researches on testing techniques, but also more and more considerable industry practices. There are testing classes taught in universities. There have been special testing teams, test managers, and tester job positions open to professional testers; there have been training programs and complete procedures for testing in large enterprises; and there are increasing number of companies and vendors doing testing work for other companies.*

However, despite the numerous research results (quite a lot of them are really sound) testing remains an awkward, time - consuming, cost-ineffective chunk of work that is always not very satisfying in most industry practices. Only a small number of the research results have been utilized successfully in industry practices so that the test process can be greatly improved or automated. The most common testing exercises in industry are static analysis including code inspections, peer reviews, walkthroughs. Not enough testing tools can be applied directly on industry projects and products without being largely modified and even re-developed. Test plans are still written by hand, while test environment remained simple and crude. This inconsistency of testing research and practice has been called the "testing gap."

Fundamental researches in techniques need to:

- Demonstrate effectiveness of existing techniques
- Address the need in new areas
- Create new adaptive techniques
- Facilitate transferring technology to industry

General Terms: Data warehousing, data cleansing, quality data, dirty data

Keywords: Data inconsistency, identification of errors, organization growth, ETL, data quality

Introduction

Let's takes the example of the Amazon retail store. The offices of that store present in many country let's say India, USA, Australia and the data regarding all of these store can be present in any of the formats, can be present with text file within the database within the CSV file, XML file or within the database. It can be stored SQL server format within oracle or within mysql. Now if the data we have for each and every country. Now the things is we may ask the question which of the country is having the maximum sales and what is the total number of laptop that are being sold or many more questions can be asked. This question when we will answer such question is called as unanswered so the unanswered one is going to do answer that collected data. How we can do that because we have different different variety of files the variety of source file so now for that we need the ETL. Here comes to ETL in the picture lie we have files, we have the DBMS, we have the OLAPS. Now from all of these source files what we will do we will pass this data to the target database or we call this as the target data ware house. It comes under the process ETL. So first letter of ETL is E that is for the extraction that is fetch data from different different sources and the second process is the transformation. Transformation means to make some changes according to business requirements like if I take some example in the month of January in the last year which of the product is being sold maximum so we have the transformation rule. We want the filter condition on the basis of the month filter on the month of January so this is one of the filter condition that cones under the transformation and the third process comes under ETL Loading so in the end we load our data to the target database and within ETL we call the target database as the Data ware house.

In this paper, we focus on the technology maturation of testing techniques, including these functional and structural techniques that have been influential in the academic world and widely used in practice. We are going to examine the growth and propagation of the most established strategy and methodology used to

Select test cases and analyses test results. Research in software testing techniques can be roughly divided into two branches: theoretical and methodological, and the growth in both branches pushes the growth of testing technology together. Inhibitors of maturation, which explains why the in-depth research hasn't brought revolutionary advantage in industry testing practice, are also within our scope of interest. There are many other interesting areas in software testing. We limit the scope of our study within the range of testing techniques, although some of the areas maybe inseparable from our study.

Specifically, we are **not** going to discuss:

1. *How testing is involved in the software development cycle?*

2. *How different levels of testing are performed?*

3. *Testing process models*

4. *Testing policy and management responsibilities, and Stop criteria of testing and software testability*

ETL Data Movement Flow
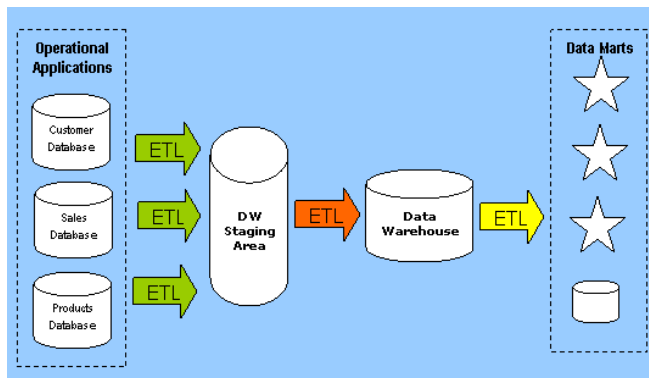
Comparison And Analysis.

Review of Work already done on the subject of proposed Research Study:

Banks get large amounts of information from countless sources. Banking domain also has to handle variety of information like every day transactions, holding information, sweep information, credit information, sock information etc. In spite of this they have to store past account information to report the user to provide the current and past information maintain varieties of handling this large amount of data using existing technologies reduce the performance.

I) To meet an increasing commercial demand for well-trained researchers experienced in testing techniques.

II) To develop and implement new quantitative and econometric methods for empirical finance and risk management with large and complex datasets.

To achieve the objectives, the emphasis is put on exploiting testing techniques to manage and use datasets that are too large and complex to process with conventional method Banks and other financial institutions must be able to manage, process, and use massive heterogeneous data sets in a fast and robust manner for successful risk management .Compared to the USA, Europe is still at an early stage of adopting testing techniques and methods.

The proposed research is relevant both academically and practically, because the program is built around real challenges faced both by the academic and private sector partners.

Following are the key issues need to consider.

- One of the lookup table called LTP_RTAES having 63 million plus records is used in Lookup transformation of ETL mapping.
- Some of the tables are not properly indexed and required proper database table partition mechanism to perform the faster query processing.
- Some of the big tables are fragmented and having stale database, schema statistics as a result taking full table scans instead of indexed based scans..
- The current Informatica mapping need re-structuring of those mappings to replace the LTP_RATES lookup table with performing the outer join on LTP_RATES table.

Create the two mapping flow for each mapping, In the first flow insert the source data which is coming from various source systems into respective temporary tables

Material & Methods:

- To perform the research following Tools, software, hardware components are used to support research assignment.

- The data warehousing & RDBMS concepts

- The dimensional and relational modeling concept
- Microsoft SQL Server 2008, Help from SQL Server DBA team
- UNIX and Windows Operating System and Help from UNIX administrator also need graphical tool to access the UNIX environment.
- The Informatica Power center 9.0.1 ETL tool
- The Lesson learned documents of past projects
- Organization process assets available to support the project

Outcomes

In general, the research on testing techniques can be roughly divided into two categories: *theoretical* and *methodological*. Software testing techniques are based in an amalgam of methods drawn from graph theory, programming language, reliability assessment, reliable-testing theory, etc. In this paper, we focus on those significant theoretical research results, such as test data adequacy and testing criteria, which provide a sound basis for creating and refining methodologies in a rational and effective manner. Given a solid theoretical basis, a systematic methodology seeks to employ rational techniques to force sequences of actions that, in aggregate, accomplish some desired testing-oriented effect.

We are going to start with the major technical contributions of theoretical researches as well as milestone methodologies of testing techniques. In next section, the riddle maturity model will be used to illustrate how testing techniques have

matured from an intuitive, ad hoc collection of methods into an Integrated, systematic discipline.

Conclusions

Testing has been widely used as a way to help engineers develop high-quality systems, and the techniques for testing have evolved from ad hoc activities means of small group of programmers to an organized discipline in software engineering. However, the maturation of testing techniques has been fruitful, but not adequate. Pressure to produce higher-quality software at lower cost is increasing and existing techniques used in practice are not sufficient for this purpose. Fundamental research that addresses the challenging problems, development of methods and tools, and empirical studies should be carried out so that we can expect significant improvement in the way we test software. Researchers should demonstrate the effectiveness of many existing techniques for large industrial software, thus facilitating transfer of these techniques to practice. The successful use of these techniques in industrial software development will validate the results of the research and drive future research. The pervasive use of software and the increased cost of validating it will motivate the creation of partnerships between industry and researchers to develop new techniques and facilitate their transfer to practice. Development of efficient testing techniques and tools that will assist in the creation of high-quality software will become one of the most important research areas in the near future.

The ultimate goal of software testing is to help designers, developers, and managers construct systems with high quality. Thus research and development on testing aim at efficiently performing effective testing – to find more errors in requirement, design and implementation, and to increase confidence that the software has various qualities.

References

[1] Chinta Someswara Rao, J Rajanikanth, V Chandra Sekhar, Bhadri Raju MSVS (2012) "Data Cleaning Framework for Robust Data Quality in Enterprise Data Warehouse", IJCST e-ISSN: 0976-8491 p. ISSN: 2229-4333, Vol. 3, Issue 3, pp. 36-41

[2] K. Srikanth, n.v.e.s Murthy, J. Anitha (2013) "Data Warehousing Concept Using etl Process For scd Type-3" International Journal of Emerging Trends & Technology in Computer Science (ijettcs) Issn: 2276-6856, Vol. 2, Issue 5, pp. 142-145.

[3] Kabiri A.; Chiadmi D. (2013) "Survey on ETL Processes", Journal of Theoretical and Applied Information Technology. Vol. 54, No. 2

[4] P andey K.Rahul (2014). Data Quality in Data warehouse: problems and solution.IOSR-Journal Of Computer Engineering, Volume 16, Issue 1, pp. 18-24.

[5] Rahm, E., Do, H.H. (2000). Data Cleaning: Problems and Current Approaches. IEEE Data Engineering Bull. Vol 23 No. 4, pp. 3-13

[6] Rodi´c J.; Baranovi´c M. (2009) "Generating Data Quality Rules and Integration into ETL Process", DOLAP'09 ACM

[7] Sakshi Agarwal '*Reasons of Data Quality Problems in Data Warehousing*' International Journal of Computer, Information Technology & Bioinformatics (ijcitb) issn: 2278-7593, Volume-1, Issue-4 ieee & ieee Computational Intelligence Society, 2013.

[8] Saravanan p. (2014) "An Iterative Estimator for Predicting the Heterogeneous Data Sets", Weekly Science Research Journal issn: 2321-7871, Volume 1, Issue 27, pp. 1-15.