

Datawarehouse Automated Test Case Generation(A-Tcg) Framework Towards Data Quality Analysis

K.Mahalakshmi

*Assistant Professor, Department of Computer Science, Sankara College of Science and Commerce
Saravanampatti, Coimbatore Tamil Nadu, India*

Abstract: *Data Quality analysis is essential task to perform any transaction based operation particularly in huge data environment. Test case Generation is predefined process and these case can applied as specific requirement. The specification can differ based on test coverage criteria. Automated test cases based on source of the program and its design artifacts, in terms of data ware house it's difficult to test due to the multiple type of data format. A Specific framework is needed to overcome this challenge. This research work described a different dimension to Test case generation also in data analytics scenario.*

Keywords: *test case generation;data quality test;TCG framework;data transformation testing ;Data Accuracy;*

I. Introduction

Random testing accepts resultant input values also, it generate test cases by far. DART and JCUTE are random fundamental testing tools. Search-based Software Engineering (SBSE) is an optimization technique. Hill Climbing search based algorithm can also used to generate automated test cases in iterative approach. Genetic algorithm used to reduce the test time and improve the test efficiency to a certain extent a search-based approach to automatically generating test cases for object oriented software. It depends on a tree-based illustration of technique call sequences [3].

The MA used for testing of object-oriented containers. Tabu Search (TS) is a metaheuristic search technique based on the premise that, in order to qualify as intelligent, problem solving should incorporate adaptive memory and responsive exploration. Thus, the algorithm of Tabu Search is based on that of the next k neighbors, whereas maintaining a Tabu list (memory) of visited neighbors that tabu. The Tabu Search rule features a range of parameters that have to be chosen on the idea of the problem to be solved: the objective operate (fitness function) that has to measure the cost of a solution, an appropriate Candidate list strategy (to try to select good neighbor candidates that goes beyond a local optimum without exploiting the examination of parts within the neighborhood) and, it's additionally necessary to outline short - term memory and memory and their several methods Simulated annealing is one type of global search technique it is similar in principle to Hill Climbing. To improve the result it recombines result after sampling the whole area. [2]. Compared to conventional software testing, a test process of big data based applications primarily focuses on their unique features, such as oracle problems, learning capability, and timeliness testing. Big data quality validation [5] and big data-based application system quality assurance becomes a critical concern and research subject [3]

A classification approach to program testing usually involves two steps: a) training a classifier to distinguish failures from successful cases on a selected subset of results, and then b) applying the trained classifier to identify failures in the main set of results. Metamorphic (MT) operates by checking whether a program under test behaves according to an expected set of properties known as metamorphic relations. .Data-Model Based Testing automatically generate test cases for a scientific program having many input parameters with dependencies. Rule-based software testing –This approach could be used in testing rule-based or knowledge based systems. The basic idea is to design test cases based on the rules specified in an expert system. Conventional black-box software testing –To assure the system performance and other related QoS parameters of big data applications, engineers could use convention black-box approaches to controlling their quality. Typical examples include decision table testing, equivalence partitioning, boundary value analysis, cause-effect graph, and use case testing, and so on. [4].

II. Automated Tcg-Framework

The objective of this test case generation frame work is the demand in test huge data and its application system requires necessary support of validation and software engineering process[5]. Modular architecture of different Specialized components comparing data from different source systems is needed with rule-based data quality evaluations [6].

- The test case generation focused on data based functional problems.
- Different data models are implementing in current scenario and holding large amount of test input.
- The input may be in structured or unstructured. Its executions also on demand, and timeline oriented.
- Test coverage required to focus more in this type of processing. Particular data warehouse contains different type of data models.

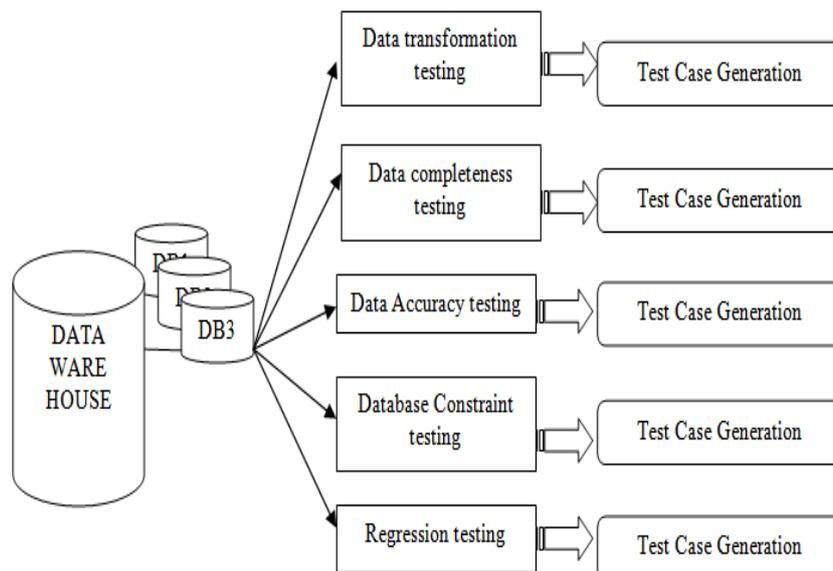


Figure 1: A-TCG Frame Work

Figure.1 describes the basic steps to process the automatic test case generation framework. It initiated with the data sources, includes data transformation, Completeness, Accuracy, Constraint testing processes. Appropriate test cases will generate simultaneously.

Step 1: Initialize DATA SOURCE;
Step 2: Evaluate DATA;
Step 3: IF TEST Criteria Not Satisfied REPEAT step 4 to 6
Step 4: Select DATA SOURCE for reproduction;
Step 5: GENERATE test cases A-TCG;
Step 6: Evaluate & Recover the Errors;

Figure 2: A-TCG Frame Work Execution Steps.

The above Framework also follows certain process steps to generate test cases includes initialization, evaluation with test condition, check the assigned test criteria, which depends on set of rules previously assigned by various tests. If the test criteria is not satisfies repeat step 4 & 6, if test criteria is satisfied, generate the test cases successfully.

III. Data Transformation Teting

Data transformation testing is used to validate all targeted data is moved properly to mentioned protocol [1].Table 1. Indexed the essential test cases in data warehousing to test individual tables of database and mandatory tests during data transformation. The table1 represents the basic test cases involved during transforming data.

Table1. Data transformation-test cases

Test Case #	Test Field	Steps To Test	Resutes From Exection
1	Valid Field (Column To Test)	Test Table Contents Are Valid	True / False
2	Aggrigation & Transformation	Test Data Movement (Both Aggrigate Transformation Tables)	True / False
3	Trasformation Column	Varify Test Souce (Single Source Targeted Query?)	True / False
4	All Table Contents	Test Lowest Transformation (Change No Of Records)	True/False
5	Resultant Column	Returned Expected Result?	True/False

IV. Data Completeness & Accuracy Testing

Data completeness testing is essential step to any data warehouse testing, which is a conformation phrase that all expected data is loaded[1].

Table2. Data completeness & acceptance -test cases

Test Case #	Test Field	Steps To Test	Resutes From Exection
1	Saource Data	Identify Rejected Data	True / False
2	Sourcedata Column & Data Loaded Column	Compare And Verify Of Possible Data Errors	True / False
3	Source Data Set & Target Data Set	Compate The Data Sets And Identify Data Anomalies	True / False
4	All Table Contents (Each Field)	Identify Trunction Occurred In Any Step In Complteness Tesing Process	True/False
5	All Table Contents (Each Field)	Test Data Base Limitations (Test Boundary Values)	True/False

V. Database Constraint & Regression Testing

The properties of the database is created with certain constrains, its properties are apply or compared with targeted table or schema. constraints and indexes remain valid when their tables, and the columns in these, are renamed because the rename command automatically updates the constraint and index metadata[7]. Regression testing tests an existing functionality of data and it test cases may need to execute frequently to identify bugs and transformations. To avoid these frequent testing risks, retest the source and destination dataset compare it with set targeted rules. This strategy used to conclude to validate the complete execution [1].

Table3. Data constraint & regression -test cases

Test Case #	Test Field	Steps To Test	Resutes From Exection
1	Any Column (Required To Test)	Compare Columns	True / False
2	All Table Contents (Each Field)	Indentify Column Changed Recently	True / False
3	Last Updated Column	Compare The Report Data Results	True / False
4	Any Column (Required To Test)	Identify Priority & Measure Production Usge	True / False
5	Any Column (Required To Test)	Capture Average Runtime To Analyse Table Vreacity	True / False

VI. Conclusion

In this paper described the basic frame work to generate automated test cases and improving analytical quality.Also Reduce and Simplifies Data analytical test and Fulfills the domain specific needs. A-TCG frame work is abasic idea to implement software engineering concepts into data alytical engineeringGive a complete solution about end data. The improvization of this frame work will provides an effective approach.

References

- [1]. http://dbmanagement.info/Books/MIX/75899021-DW-Master-Test-Plan2_Informatica.pdf
- [2]. Pranali Mahadik, Debnath Bhattacharyya, Hye-jin Kiml, “Techniques for Automated Test Cases Generation: A Review,” International Journal of Software Engineering and Its Applications Vol. 10, No. 12 (2016), pp. 13-20.
- [3]. Akshat Sharma1 , Rishon Patani1 and Ashish Aggarwal1, “SOFTWARE TESTING USING GENETIC ALGORITHMS” International Journal of Computer Science & Engineering Survey (IJCSSES) Vol.7, No.2, April 2016 .

- [4]. Chuanqi Tao, Jerry Gao “Quality Assurance for Big Data Application– Issues, Challenges, and Needs” DOI reference number: 10.18293/SEKE2016-166.
- [5]. <https://www.infosys.com/IT-services/validation-solutions/case-studies/Documents/optimal-data-test-coverage.pdf>
- [6]. Malburg, J., Fraser, G.: Search-based testing using constraint-based mutation. *J. Softw. Test. Verif. Reliab.* **24**(6), 472–495 (2014)
- [7]. <http://www.oracle.com/technetwork/database/availability/edition-based-redefinition-1-133045.pdf>