The Partial Proposal of Data Warehouse Testing Task

Pavol Tanuska [1+], Pavel Vazan [1] and Peter Schreiber [1]

[1] Slovak University of Technology, Faculty of Material Science and Technology,
Paulínska 16, 91724 Trnava, Slovakia

Abstract. The aim of this article is to suggest the partial proposal of datawarehouse testing methodology. The analysis of relevant standards and guidelines proved the lack of information on actions and activities concerning data warehouse testing. The absence of the complex data warehouse testing methodology seems to be crucial particularly in the phase of the data warehouse implementation.

Keywords: Data warehouse, testing, methodology, UML.

1. Introduction

In the phase of testing methodology designing for the process of the data warehouse validation, it is necessary to consider the results and knowledge acquired in the analysis of the existing standards and guidelines. The analysis of valid European standards and guidelines related to software products evaluation has been done by authors as an opening part of proposed testing methodology. There has been analyzed many standards that deal with software life cycle processes (e.g. ISO/IEC 12207, ISO/IEC 15288, ISO/IEC TR 12182, ISO/IEC TR 15271, ISO/IEC TR 16326), software products operating reliability evaluation (e.g. DIN/IEC 56/363 (458,478, 575)), with general evaluation of software products and tools (e.g. ISO/IEC 14598, ISO/IEC 14102) and with systems of software quality assurance (e.g. ISO 9001:2000, ISO 25000, DIN ISO/IEC 12119, DIN 66272) and also standards that deals with metrics (e.g. ISO 9126-2, ISO 9126-3, ISO 9126-4). Special part was an analysis of guidelines Gamp4 and TickIT. All this selection represents about 150 standards and guidelines, which cover the area of the software evaluation. The analysis result is, that the software testing is standardized in details, but without special cases (like e.g. data warehouses) specification. The data warehouses have been used massively only in the last few years and they are specific with some properties and processes like data pump (ETT process), metadata management, multidimensional database design and management, special OLAP-reports design, etc. [1]

In the first stage of the methodology design, it is necessary to focus on attributes of testing scenario.

2. Identification of the essential DW test scenario attributes

When proposing the scenario attributes, the authors used the test procedures acquired from the results of analysis of relevant standards and guidelines. The outcome is the statement that testing is inevitable and important. It is worth to state that in testing software systems (including data warehouse) the designers and testers appear from following scheme shown on fig.1.

There are numerous testing methods and approaches, yet individual standards do not provide detailed information, thus offering room for various disinformation. As for the design of data warehouses, it is worth

---

[+] Corresponding author. Tel.: + 421 918 646 061; fax: +421 33 5511758.
  E-mail address: *pavol.tanuska@stuba.sk.*

to state that the current standards and guidelines do not cover the activities relevant for building a data warehouse. [4]

Neither particular procedure nor activities regarding the process of a multidimensional databases proposal, ETT process or optimisation of scripts for OLAP reports are available.
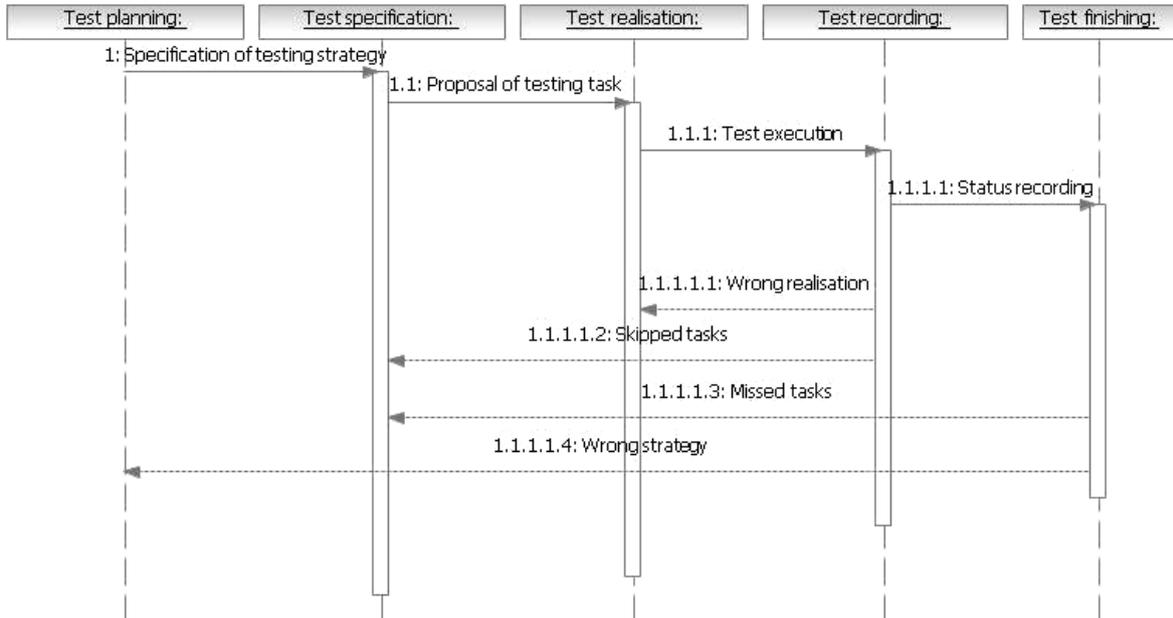

Fig. 1: UML sequence diagram of testing process.

While designing the proposal of DW test attributes, we therefore enhanced the standard testing methods by the activities which are generally not involved in testing standard information systems.

The scenario (for a test or a set of tests) should comprise the following attributes [2]:

- To unambiguously describe the purpose of testing, including instructions defining step-by-step procedure as well as the activities that should be performed by the testing person or team.
- To provide a link to the specifications of the tests administration check (traceability), specifying the document and section which defines the test requirements, e.g. URS or FS.
- To identify specific performance requirements and define the necessary test types including the selection of appropriate testing methods and techniques, particularly regarding the specifications of ETT process and meta-data testing.
- To complete all the items necessary prior to testing, such as documentation, sources, software, interfaces, data etc.
- To build and subsequently verify specific testing configuration.
- To define the input parameters, particularly those regarding specific activities such as data pump implementation, mainly exact definition of parameters related to the building and loading a multi-dimensional database. The whole process entails transformation of data into de-normalised structures and involves data selection, cleaning, integration, derivation and following de-normalisation.
- To define the acceptance criteria specific for data warehouses, i.e. a defined set of anticipated results, which the test should achieve in order to be considered successful.
- To identify the data to be recorded, i.e. to define the data related with the administered test, which should be sensed and recorded, such as input, output, descriptions comprising the identification of every equipment used, and calibration certificates, wherever necessary.
- To exactly record all the tests and their results in the form of test protocols, designed for individual stages of testing. All the test protocols have to be approved by an associated test manager and must comprise all relevant items described in International standards, while the new items, not specified in standards and guidelines, must be clearly marked.

- To design the activities associated with the test completion. This section provides detailed information on the activities which will set the testing system into defined, e.g. resetting the process parameters, setting the system in a safe state, or warm/cold restart.

In order to remove redundant duplicity in the process of designing the test, it is recommended to provide links to relevant information comprised in the specification of the test administration check where applicable. Format of the test scenario may allow recording the test results directly into a copy of the approved test specification. Related forms should therefore bear the names of the test administrator and witness, their signatures as well as the date of the test execution.

Other very important attributes related to test scenario should include: testing prerequisites, proposal of traceability matrix and specification of test procedures.

3. The testing task as a part of testing scenario

The logical continuation of data warehouse testing methodology is the proposal of realization of data warehouse test scenario, what are concrete principles that are necessary to do in the process of data warehouse testing.

Preliminary to inscribe the mode of test scenario realization, is necessary to mention problem of testing strategy, too. [5] Data Warehouse testing strategy determines the project's approach to testing. The strategy looks at the characteristics of the system to be built, and plans the breadth and depth of the testing effort. The testing strategy will influence tasks related to test planning, test types, test script development, and test execution.

In Data Warehouse building efforts, one of the most important critical success factors is to have consistent and timely data in the Data Warehouse, allowing users to have consistent and timely answers to their queries. Unlike information systems, DW unifies information from the various data sources including different database systems, different applications, flat files and even Excel sheets. The unification process contains extraction of this data from the original source, performing necessary massaging and placing into the DW. This is the most complex and time consuming part of the whole DW implementation which contains number of custom built programs, scripts as well as commercial software specifically designed for ETT. Verification of ETT process by applying necessary tests is an important step of the DW implementation process. This testing includes, but is not limited to, stand alone tests of each script, program and modules, integrity testing of theses which will work together and of course the consistency of data which finally loaded to DW.

Accuracy of the data and performance of the DW are also subject to testing before DW goes into production. Because queries run against a DW differ in type than a typical OLTP application, the performance expectations are also different. The data volume stored in the Warehouse is also a factor of rationalizing a more realistic performance expectation.

In the process of testing there is necessary to reflect on following attributes (represented by set of important activities), that has influence on methodology proposal and also on testing progress:

- testing plan, technology and tools,
- identification of testing constraints,
- design requirements,
- specification of test-cases series,
- identification of possible problems,
- creation of testing procedures,
- technical architecture,
- completion of criteria,
- time scheduling,
- error tracing and message allocation,
- plan authorizing.

The special part is creation of testing tasks, which will be performed in the process of testing implementation. The testing tasks that must be implemented in the process of the testing can be split into four logical objects plus two actors.[3] In the real testing system these objects can represent four autonomous tests tasks. Whole process of testing task execution is designed on fig. 2 in UML sequence diagram.
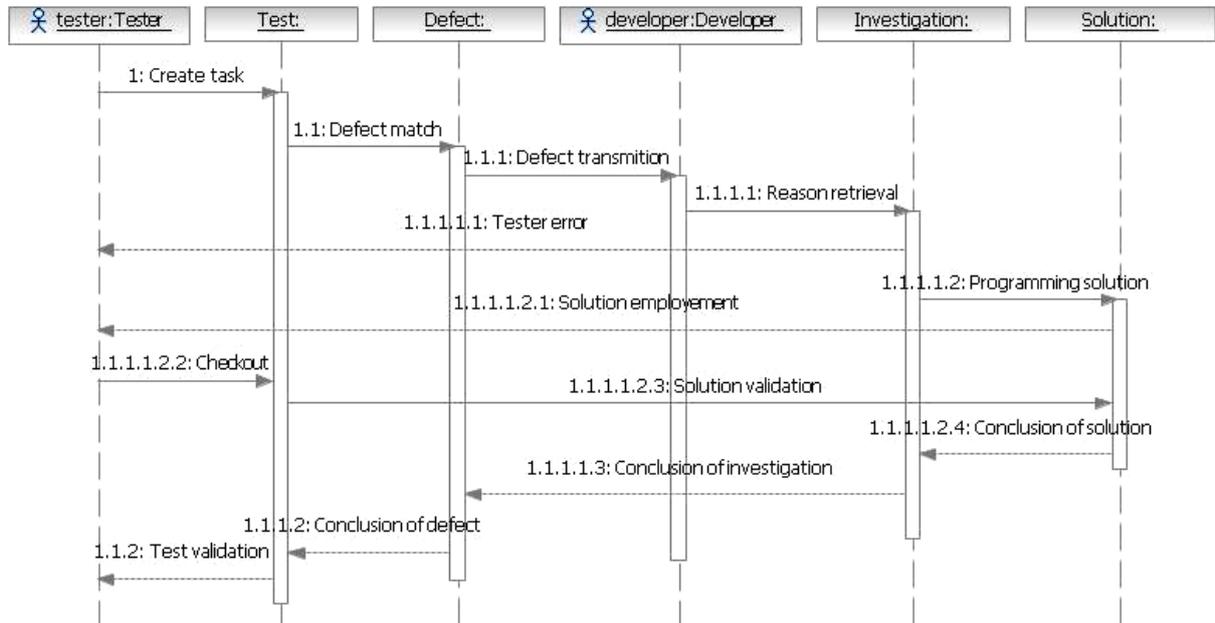

Fig. 2: Process of testing task execution.

The diagram on fig. 3 shows possible set of status testing task in process of testing. Testing starts by creation of task in status "New". If tester begins to edit anything in task records, the status will change into "In process".
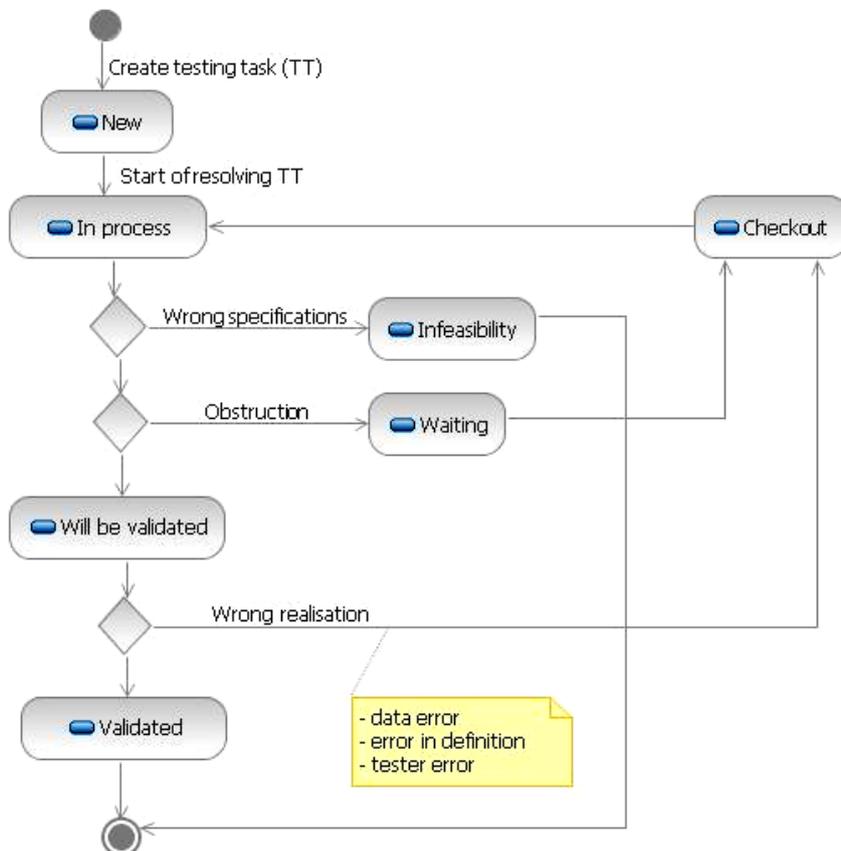

Fig. 3: UML State machine diagram of testing task.

This status represents realization of testing task. If tester during the testing finds that task has wrong specification, status will change into "Infeasibility" and task finish. The status "Waiting" arise if some obstructions appear in testing task solution. The tester knows that continuation with that task is possible after obstruction removal. When the obstruction disappear the testing task change status into "Checkout" and process can continue in realisation.

Having successfully finished the testing task the tester change the status into "Will be validated" and will send task to validator. The validator will check the task realization step by step. If finds some errors in realization (data error, error in definition, tester error…) he will send task back for checkout. In other case change task status into "Validated" and process will finish.

4. Acknowledgements

The testing phase as one of the stages of DW development lifecycle is very important, since the cost depleted for the elimination of a potential error or defect in a running data warehouse is much higher. A data warehouse as well as an information system can be physically correctly tested only when the working database is loaded. The aim of this article has been to suggest partial proposal of datawarehouse testing methodology.

5. References

[1] Kebísek, Michal - Makyš, Peter: Knowledge discovery in database and possibilities of they usage in industrial process control. In: *Process Control 2004: Proceedings: 6th. International Scientific - Technical Conference. Kouty nad Desnou, Czech Republic*, University of Pardubice, 2004. - ISBN 80-7194-662-1.

[2] Tanuška P., Verschelde, W., Kopček M.: The proposal of Data Warehouse test scenario. *Proceedings of III. International conference ECUMICT Gent, Belgium*, ISBN 9-78908082-553-6, 2008.

[3] Zobok, Maroš: Čiastkový návrh smernicepre testovanie veľkých softvérových systémov. *Final thesis MTF STU Trnava,* In Slovak. 2009.

[4] Tanuška P., Schreiber P., Zeman J.: The realization of data warehouse test scenario. In: III. Meždunarodnaja naučno-techničeskaja konferencia – Infokomunikacionie technologii v nauke i technike. Stavropoľ, Russia 2008. UDK 004 (06) BBK 32.81

[5] Strémy, Maximilián - Eliáš, Andrej - Važan, Pavol - Michaľčonok, German: Virtual Laboratory Implementation. In: Ecumict 2008: Proceedings of the Third European Conference on the Use of Modern Information and Communication Technologies. Gent, 13-14 March 2008. -: Nevelland v.z.w., 2008. - ISBN 9-78908082-553-6