

The proposal of Data Warehouse test scenario

Tanuška, Pavol
Verschelde, Werner
Kopček, Michal

15th November 2007

“Slovak University of Technology” Institute of Applied Informatics, automation & mathematics, Paulínska 16, SK-917 01 Trnava, pavol.tanuska@stuba.sk

“KaHo Sint Lieven” Department of Informatics

G.Desmetstraat.1, B-9000 GENT, werner.verschelde@kahosl.be

Abstract

Analysis of relevant standards and guidelines refer, that there is no declared actions and activities which has connection to data warehouse testing. The absence of complex data warehouse test methodics seems to be a cardinal problem in the phase of the data warehouse implementation. The particular aim of this article will be the proposal of basic attributes of data warehouse test scenario (this scenario will be a part of the DW test methodics). The next very important attributes that are connected to test scenario proposal can include: prerequisite of testing, testing philosophy, traceability metrics, test case specification, test procedure specification.

Keywords: *data warehouse, test scenario, methodology, ISO standards*

1 Introduction

Institute of IT, Automation and Mathematics in cooperation with the educational support office of the faculty is developing small data warehouse since 2006 (the scheme of data warehouse is shown on Fig. 1).

It should to serve for processing information about candidates for studying on this faculty, actual students and its graduates. We will be able to use acquired information from the data warehouse (by OLAP and Data Mining) for example for the need of intake of adequate students (in term of the faculty trend) for each study program on the faculty.

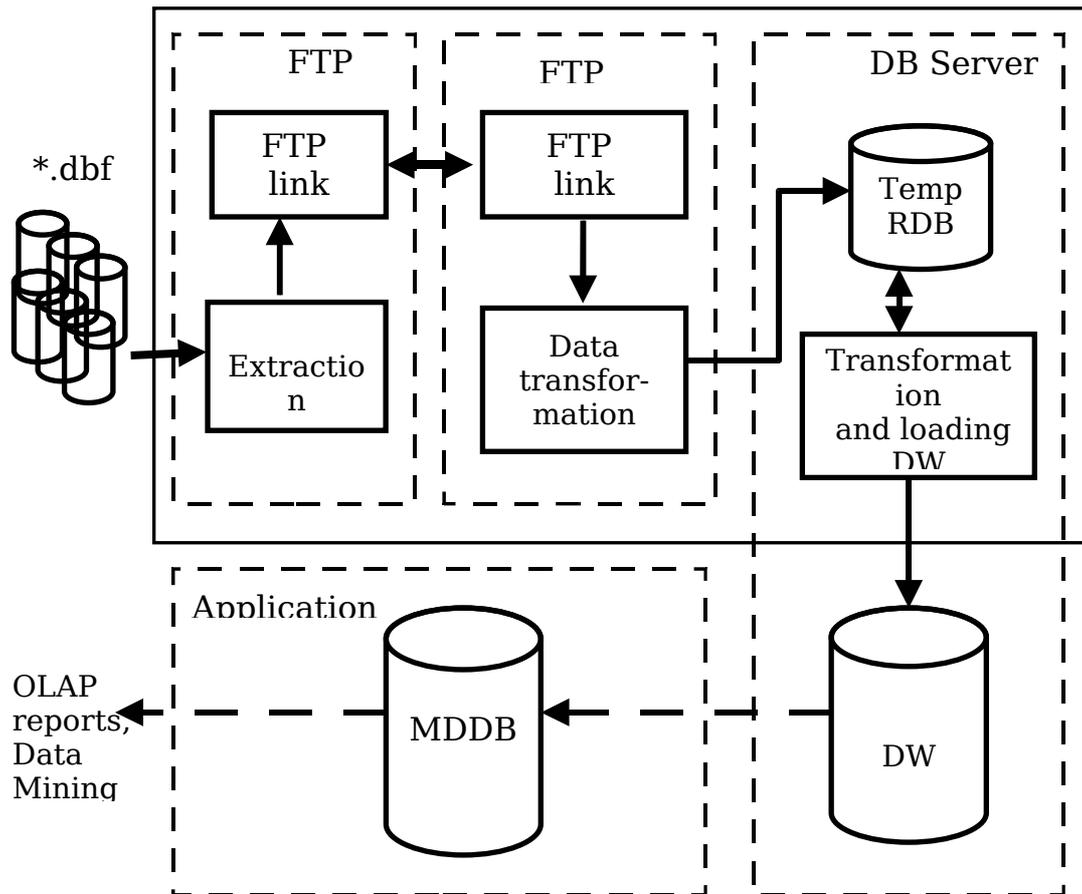


Fig. 1 The scheme of data warehouse

Another plumbless asset of using the data warehouse will be the possibility of online evaluating of study results (of students, departments and faculty) according to miscellaneous and variable parameters, which will surely contribute to increase the quality of pedagogic process.

We use the star scheme for the design of the structure of the data warehouse (Fig. 2), which is the best following the input data and required functionality of the data warehouse. The fundamentals of the star scheme are that the primary keys from dimension tables are transformed as foreign keys in the fact table. The center is the fact table FAKT_STUDIUM, which contains reference from dimension tables. Dimension tables consist of tables including the data about students, because students are just the one key factor of the data warehouse. Next dimension tables contain data about study, secondary schools and regions. The used dimension tables are Dimenzia_LOKALITA (Dimension Locality), Dimenzia_PREDMET (Subject), Dimenzia_SKOLA (School), Dimenzia_STUDENT, Dimenzia_STUDIUM (Study), Dimenzia_CAS (Time) and Dimenzia_UCITEL (Teacher). We take into account the hierarchical data arrangement while designing dimensions, which is characteristic for every dimension. [1, 3]

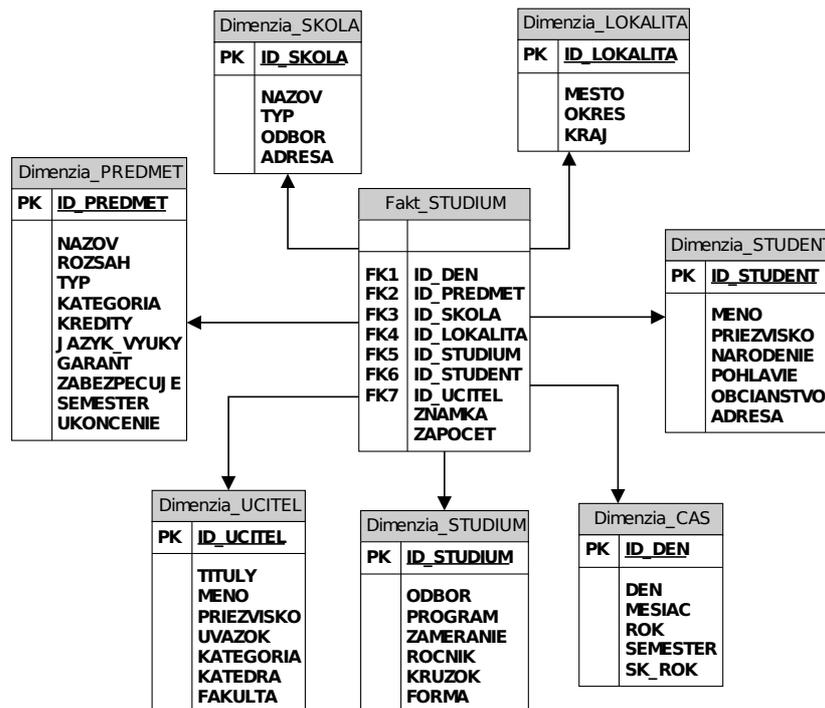


Fig. 2 Star scheme of the designed data warehouse

The absence of complex data warehouse test methodics seems to be a cardinal problem in the phase of the data warehouse implementation. Because this problem overreaches the size of this article, the particular aim will be the design of basic attributes of data warehouse test scenario (this scenario will be a part of the DW test methodics).

Unlike from transactional information systems, the data warehouse combines information from various data sources, including various data warehouses, various applications, files and even worksheets from MS Excel. The main risk, which carries building of data warehouse, is in the first place storing of wrong data. The part of testing is working-out the test planes and test requirements.

2 Test scenario

First of all we have to say, that we emanate from IEEE 829 standard and we respect it as a basis.

This standard defines 8 document types, which are divided into a three categories.

I. Preparation of Tests

Test Plan: Plan how the testing will proceed.

Test Design Specification: Decide what needs to be tested.

Test Case Specification: Create the tests to be run.

Test Procedure: Describe how the tests are run.

Test Item Transmittal Report: Specify the items released for testing.

II. Running the Tests

Test Log: Record the details of tests in time order.

Test Incident Report: Record details of events that need to be investigated.

III. Completion of Testing

Test Summary Report: Summarize and evaluate tests.

We emanate from existing methods, which we get from results of analysis of specifications and standards, during designing the scenario. [2] We eke out specifications with activities, which are not a part of ordinary tests of information systems.

The scenario (for the test or for the group of tests) should contain following attributes:

- The test name and identification. Clear test identification according to the reference format, which is established in the tests specification.
- The test purpose.
- Instructions. A step by step procedure and activities, which have to be made by a testing person.
- The name of module, or activity of the data warehouse, which will be tested.
- A control specification reference of tests execution (deducibility).
- Here comes a notice to the document and appropriate section, which defines the tested requirement e. g. URS or FS.
- Identification of performance requirements.
- Specification of test types, which will be demanded.
- A selection of suitable method and test technique.
- Applied tools.
- Assembling of all items, which are needed before the beginning of the test procedure.
(documentation, sources, software, interfaces, data etc.)
- The test configuration.
- Instructions for the execution of a test or a group of tests.
- Input parameters definition.
- Definition of acceptance criterions, which are specific for data warehouses. Actually, it is a defined collection of expected results, which have to be achieved by the test, in order to the test allowed to be successful.
- Identification of data, which have to be recorded. Definition of data, which are related with performed test and which should be accumulated and recorded. This could be inputs, outputs and descriptive data, which have to contain serial number of every used test device and calibration certificate, wherever it is necessary.
- Exact records of all tests and its results in the form of test protocols, which are designed for individual parts of data warehouses test.
- Design of activities, which are related to the test completion. This part shows in more detailed view such activities, which bring the tested system to a well known and defined state. An example could be a reset (readjustment) of process parameters, getting the system into a save state or the hot or cold restart of the system.

In order to eliminate useless duplicity is recommended to mention references on the appropriate information during the test design. This information is included in the specification of the test execution control everywhere, where it is possible.

The form of the test scenario could be written so, that the test results will be recorded directly to the approved copy of the test specification. It follows that, in the corresponding forms have to be reserved places for the test results notation, for the date of the test execution, for the name of the test executor and the name of the test execution witness. [4]

To another very important attributes, which are related with the test scenario, we could include these:

- The test prerequisites.
- The test philosophy.
- The deducibility matrix.
- The test scenario specification.
- The test procedure specification.

3 Conclusion

It is obvious from tenths of monographs and hundreds of papers, which were dedicated to the data warehouses in last few years, that the data warehouses became a non-detachable part of many institutions. However, theirs practical usage brings not only many problems, which are well-known from implementation and performance of the information systems, but moreover many new problems, which are specific exactly for the data warehouses.

The test phase by itself, as one of the information system life cycle stages, is very important, because the costs, which are spend to the elimination of potential mistakes or defections by the implemented data warehouse, are non-comparable higher compared to the costs of elimination of the defections, which were found during the tests. It is possible to test the data warehouse physically correct as good as information system only if the database, which will be used by it, is filled with data.

Only a few data in tables are not enough for the complete examination of data warehouse. It is ideal to approximate the quantity of test data to the real conditions in which will be the data warehouse working. It is also necessary to know the structure of the multidimensional database, the structure of the metadata, or the ETT process. The main aim of presented paper was to design the basic attributes of the test scenario as an initial part of the data warehouse test methodics.

The authors are very grateful for the support of the grant VEGA 1/4078/07.

References

- [1] Inmon, William (2002) "Building the Data Warehouse". Wiley Computer Publishing.
- [2] Mudrončík, Dušan (2003) "Validation of process control systems based on GAMP". 14th conference on Process Control.
- [3] Ponniah, P (2001) "Data Warehouse Fundamentals – Comprehensive Guide" . London: John Willey and Sons.
- [4] PQG - Quality Assurance (2004) "GAMP4".