Software Engineering
2018; 6(1): 1-6
http://www.sciencepublishinggroup.com/j/se
doi: 10.11648/j.se.20180601.11
ISSN: 2376-8029 (Print); ISSN: 2376-8037 (Online)

Development of Data Warehouse Structure Model for Educational Process Management

Mykola Pasyeka

Department of Software of Automated Systems, Ivano-Frankivsk National Technical University of Oil and Gas, Ivano-Frankivsk, Ukraine

Email address:
pms.mykola@gmail.com

To cite this article:
Mykola Pasyeka. Development of Data Warehouse Structure Model for Educational Process Management. *Software Engineering*. Vol. 6, No. 1, 2018, pp. 1-6. doi: 10.11648/j.se.20180601.11

Received: November 23, 2017; **Accepted**: December 1, 2017; **Published**: January 16, 2018

Abctract: This article considers the development ways of data warehouse structure model for the management of the educational process in higher educational establishments. A model of analytical processing of structured and unstructured data is highlighted and the scheme of data files position in the data warehouse is visually presented. The analysis of modern methods and means of development data warehousing structures for building of informational systems with the ability for the analytical study of available information is being done. The result of this study is developed architecture of of data warehouse system and conclusions about use features while building of the information-analytical system of higher educational institution.

Keywords: Data Warehouse, Analytical Data Processing, Data Storage Model, Information-Analytical System, Decision-Making System

1. Introduction

Problem Formulation lies in the features of the development and operation of automated systems with use of data warehouse for management, which provide opportunities for operational tuning and management of data flow interactions and exchange of electronic documents within automated complex for the effective provision of educational services.

Study purpose is to develop a data warehouse structure model for the educational process administration and models of analytical processing of structured and unstructured data as an element of the quality improvement of educational services in Ukrainian higher educational institutions. Formulated aim requires solutions for such tasks:

- system analysis of methods and tools of data warehouse structure development for building of informational systems with the ability of the analytical operation of available information;
- analysis of methods and means of data warehouse structure development for building informational systems with the ability of the analytical study of available information;
- development of infological model of data warehouse structure;
- development of data warehouse scheme for informational-analytical systems of higher educational institution management.

Study methods are system analysis in data warehouses, algorithms and data structures that were used to develop a model of the efficient management in higher educational institutions.

2. Analysis of Recent Research and Publications of Domestic and Foreign Scientists

The rapid development of information technology in education is characterized by a modern trend in the use of data warehouses and is an important stepof informational data streams for further processing by means of OLAP (Online analytical processing, analytical processing in real time).

After defining priorities and needs urgency for the improvement of computer systems in informational-analytical administration support, based on data warehouse structure model.

This issue was studied by scientist Kalinichenko LS which

examined data space and options for their application to build unified data flow diagrams with further processing of user requests. As per his opinion, the user, while choosing the needed attributes for processing of built data warehouse scheme, refers to information sources as to its own analytical system. This theory needs further investigation and clarification for the canonical model development of analytical processing of structured and unstructured data streams. Since the idea of constructing models for informational flow visualization in canonical form was used for the construction of a consolidated data warehouse as data integrational process from different sources.

Continuing the Kalinichenko L. S. ideas there exists a Tuzlovskiy A. F. method, which requires processing exclusively structured information sources and is based on a modified Kalinichenko L. S. model. The disadvantage of this approach is the use of a single query language SPARQL (Protocol and RDF Query Language), that requires the user to a particular qualification. Based on the facts above it is possible to affirm that on the level of data warehouse processing it is appropriate to use traditional approaches to integration, and on levels of space data analysis it is appropriate to use semantic integration and extension of the traditional methods with the initial determination of the data warehouse structure of information sources and means of its access.

In scientific publications of Shakhovskaya N. B. and Pasichnyk V. V. there were reviewed scientific fundamentals and principles of development and operation of informational-analytical systems based on data warehouses, as well as were identified the problems which they should be applied for. The authors proposed a formal model of informational data space as a new abstraction approach to data management.

In scientific work «The Data Warehouse Toolkit. Second Edition» by R. Kimball an approach to building a multidimensional data warehouse based on a design approach to the «bottom up» is proposed. Such approach is characterized by the fact that the design starts from the lowest level, namely data marts and ends up with consolidated analytical processing of data warehouse. Proposed model of multidimensional data warehouse architecture makes it possible to efficiently handle user requests to aggregated data, metadata and store them. Using this architecture makes it possible to obtain consistent data of independent of each other windows that use the default set of measurements which provides the possibility of implementing a parallel windows design for different departments and facilitate the design of all informational-analytical system as a whole.

In the works of Thanh N. Huynh «Metadata for Object-Relational Data Warehouse» the design of object-relational data warehouse based on the logical architecture using objective-oriented approach and metadata is reviewed. In this model the architecture of data information flows collect from a variety of structured and unstructured operational databases that are processed, summarized, aggregated and transferred to the data warehouse with further OLAP processing.

Having analyzed scientific publications on the subject, namely the use of data warehousing to informational-analytical systems of higher education institutions requires different approaches and methods for processing data streams to various mathematical tools. In publications various models of data warehouse architecture are proposed, however a unified approach that would define the infological model development of data warehouse structures for higher education was not offered.

3. Theoretical Substantiation of Data Warehouse Structure

Table 1. *The definition of "data warehousing"*

Authors	Interpretation Degree
Bill Inmon (1992)	Determined kind of databases as a data warehouse. Data Warehouse - is a special form of database structure that is visually oriented, integrated, invariable, supports chronology of data sets to support management decisions in the form of aggregated information, that is obtained on the basis of the available data from various information systems and external sources.
Edgar Frank Codd (1993)	Formulated 12 rules for designing data warehouses that should enable parameterization of data information flow on various grounds and foundations for multivariate analysis. The term OLAP (On-Line Analytical Processing) was propose - a processing technology of multidimensional structured and unstructured data warehouse (Data Warehouse), tend to be aggregated, and that is the result of processing the data stream, consisting of a tables plurality. Data Warehouse is a trend to use in information systems with extra-large data sets and enables multidimensional analysis to support management decision making.
Hackathorn Richard D. (1999)	Founder of data warehousing concepts, that are presented in the form of multidimensional information structures called "Star" or "Snowflake" and provide management units "the only way the current reality of data processing." Thus the data warehouse is used for storage and analytical processing of information about the company.
Ralph Kimlall	Proposed spatial data warehouse based on tires, which was the alternative for "stars" or "snowflakes" architecture. By definition of R. Kimball, data warehouse is not only a physical repository of information as stated B. Inmon but "virtual" storage space for further processing of analytical information to support management decision making.
Douglas Hackney	Conducted the concept of "hybrid architecture" or "matching data marts" of data warehouses that combines features of relational and multidimensional models for correct storage of information data flows in order to fulfill user requests.

The use of data warehouses in the information-analytical systems are important design decision as they are better suited for the accumulation of large amounts of information with further analytical processing. Architecture of data warehouses has developed methods and tools for integrating data from various sources and consolidates information flows which are aggregated as necessary for upload to data warehouse. For more reasonable data warehouse design in the development of information systems with the ability to analytical information processing will determine approaches

to the "data warehouse" that are listed in Table 1.

After a systematic analysis of the use of data warehouses has been done, there was proposed to use hybrid information system architecture during the development that will ensure both operational and analytic processing of data to support decision-making in higher education institutions to improve the quality of educational services. For decision-making in higher education institutions (HEIs), analytical data to provide educational services is required, as well as data streams from various external factors and forecasting of the labor market in the medium term. To effectively solve this class of problems, statistics data that is generated for several years with the better cope storage is needed. So based on the above stated, in our opinion it is better to build information systems basing on data warehouse models and algorithms of analytical data processing.

3.1. Infological Data Warehouse Structure Model for HEI

The most common specification for building infological models and abstract representation of data warehousing is ER-model or the model "entity - relationship." Initially this specification was introduced by Peter Chen in 1976, it consists of three types of elements: a) attribute; b) bonds; c) a plurality of entities. Will build infological mode “entity - relationship” for informational-analytical educational management system. To do this, first we need to define the basic nature and links between them:

Group. Among the attributes of the essence Group are: ID, name. The user can choose from a list his group and look at the schedule. Schedule access is not limited, i.e any user can review information about the time and procedure of classes without prior registration.

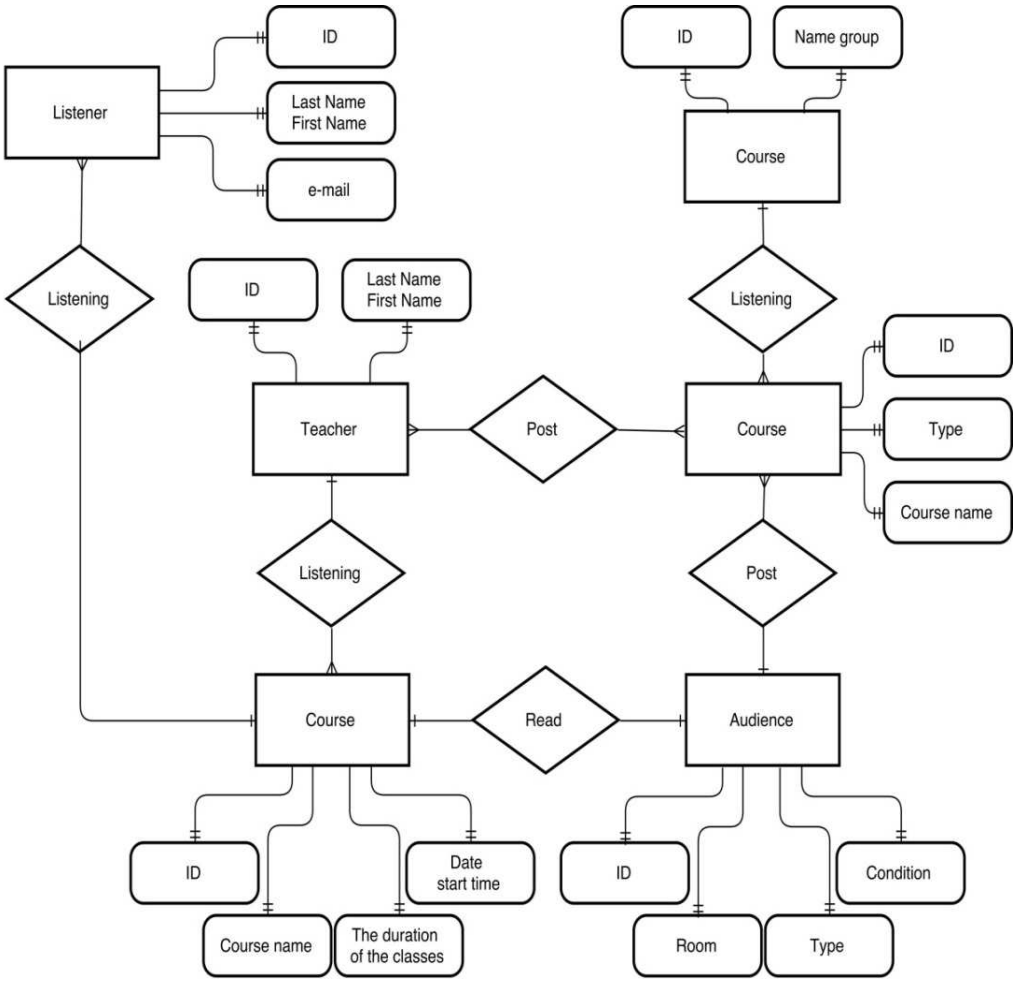


Figure 1. *Infological model of educational service providing.*

1. Tutor. Entity Tutor can be defined with attributes such as ID number, full name. Tutors teach subjects and courses, that’s why the essence of tutor is associated with entities Subject and Courses though many-to-many and one-to-many links respectively. One subject can be taught by several tutors for different groups; one tutor can conduct several courses.
2. Subject. The main attributes of this entity are ID, name, type. The essence of the Subject is associated with the Group, Audience and Tutor entities. Depending on the class type, some audiences for its implementation are defined. Labs - for laboratory exercises, lecture halls - for lectures. The tutor teaches a subject, the group listens to the subject matter.
3. Course. The attributes of the Course are identification number, name, duration, start date (time). The course is

linked to the entity Listener. Few listeners will hear a course. Tutor will present the course to a certain audience, so the relationship one-to-one is used.

4. Audience. The attributes of this entity are: id, number of audience, type and condition. Number of audience may consist of numeric and symbolic characters (A11, 1419). Audience type provides division on lectures, laboratory and practical training for the audience. Attribute Status displays if classes or other activities take place that specific moment in this audience - otherwise it is free.
5. Listener. The attributes of this entity are: identification number, name, e-mail. The entity Listener is associated with the entity Course with many-to-one link. Listeners may either belong to different groups and not be university students.

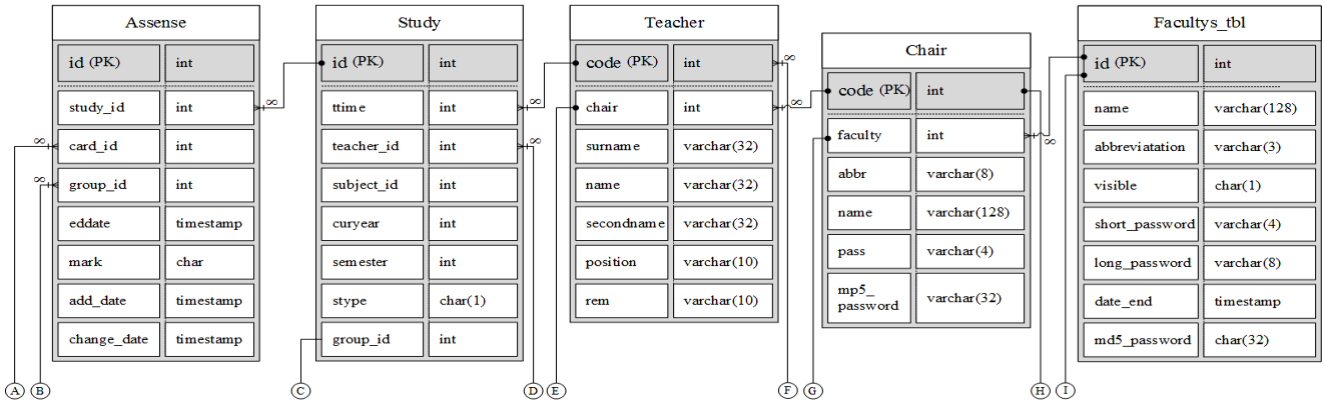
Developed infological model of providing educational services of HEI is displayed at Figure 1.

During the development process the basic software entities, and the links between them were defined. This will further simplify the modeling, construction of data warehouse circuit and provide efficient project coding.

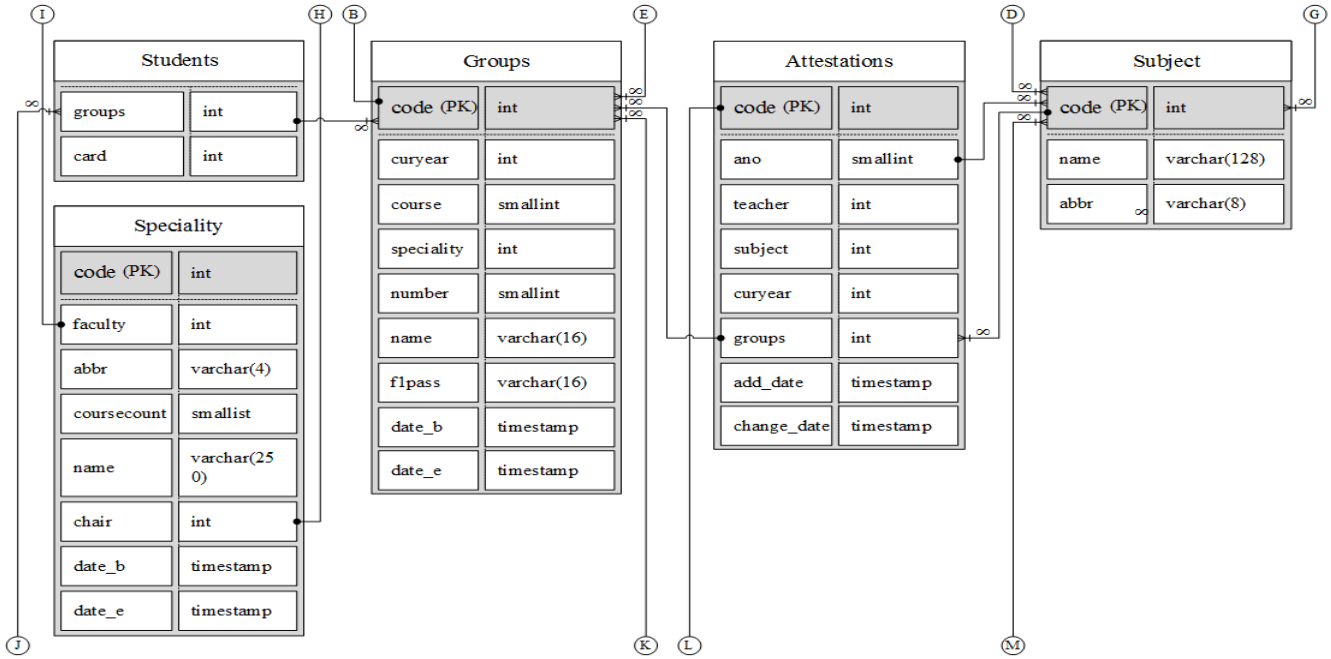
3.2. Model of HEI Relational Data Warehouse Structure

For the design of integrated analytical system with distributed architecture according to HEI regional specialties will develop the structure of relational data warehouse. This structure will ensure horizontal and vertical linkages, providing integration and further development of current software applications to create new means of analytical processing for data information streams.

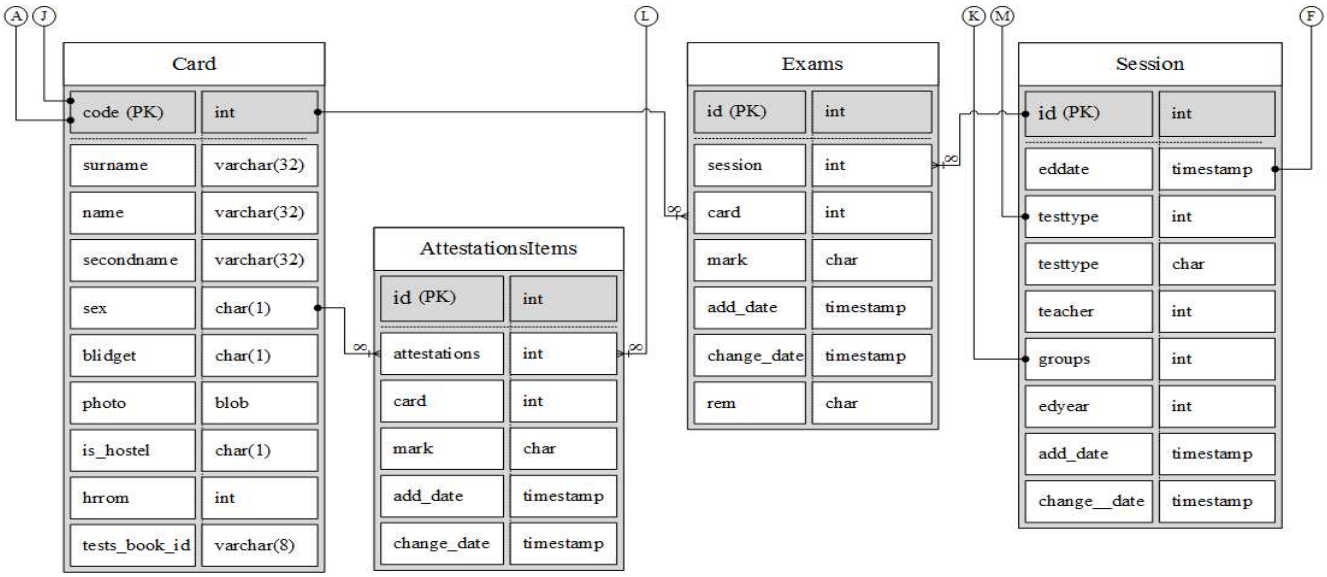
The development of domain relational data warehouse was conducted in terms of the data warehouse relational model and by approaching the required set tuples relationship as shown in Figure 2.



a) the first part of the model



b) second part of the model



c) third part of the model

Figure 2. *The structure of data warehouse relational model.*

However, the data warehouse relational model has the typical shortcomings of relational data warehouse management systems (RDWMS), namely they embede algorithms of processing data flows with domain area. Use a relational data warehouses is only effective for small volume processing records scope and considerable intensity for modification of the available data. These models are effectively used in operating systems of HEI administration while providing education services.

4. Conclusion

There was developed the infological model of information system providing educational services with further development of the data warehouse relational model structure, that will assist administering universities, providing an opportunity to set up and control the interaction of data flow and exchange of electronic documents within the automated system for the effective provision of educational services.

Referenses

[1] Sikora L. B+-tree in database design for decision-making information system / Lubomyr Sikora; Yuliya Miyushkovych/- 2010 Proceedings of VIth International Conference on Perspective Technologies and Methods in MEMS Design: 2010 – P. 185–186.

[2] Chesanovskyy M. The Formal Structuring of Subject Domain for oil and Gas Industry IT Applications /M. S. Chesanovskyy, V. I. Sheketa, V. M. Yurchyshyn, T. R. Styslo// Modern problems of radio engineering, telecommunications, and computer science (TCSET'2016): the XIII International Theoretical and Practical Conference, (Lviv-Slavske, 23-26 february, 2016). – Lviv, 2016. – P. 503–505.

[3] Sheketa V. The Formally Stated Model for Technological Process Operator Queries Interpretation / V. I. Sheketa, M. M. Demchyna, R. B. Vovk, Y. L. Romanyshyn // Modern problems of radio engineering, telecommunications, and computer science (TCSET'2016): the XIII International Theoretical and Practical Conference, (Lviv-Slavske, 23-26 february 2016). – Lviv, 2016. – P. 476–479.

[4] Sheketa V. The Construction of Technological Problems Cases for the Purpose of Intelligible Control / V. I. Sheketa, V. D. Melnyk, Y. L. Romanyshyn, M. S. Chesanovskyy // Perspective technologies and methods in MEMS design (MemsTech'2016) the XIIth International Conference, (Lviv-Polyana, 20-24 th April 2016). – Lviv, 2016. – P. 96–99.

[5] Riznyk O. Recovery schemes for distributed computing based on BIB-schemes / O. Riznyk, Yu. Kynash, O. Povshuk, V. Kovalyk // First International Conference on Data Stream Mining & Processing (DSMP) – 2016. – P. 134–137.

[6] Pasyeka N. Construction of multidimensional data warehouse for processing students' knowledge evaluation in universities / Pasyeka, N., Pasyeka, M. // Modern Problems of Radio Engineering, Telecommunications and Computer Science, Proceedings of the 13th International Conference on TCSET 2016 - P. 822–824.

[7] Rashkevych Y. Optimization search process in database of learning system /Rashkevych, Y., Peleshko, D., Pasyeka, M // Proceedings of the 2nd IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2003 - P. 358–361.

[8] Fedir E. Invariant operations on discrete neural functions over Galois field / Fedir E. Geche, Anatolii. Batyuk, V. Buchok // First International Conference on Data Stream Mining & Processing (DSMP) – 2016. – P. 112–116.

[9] Mulesa O. Information technology for determining structure of social group based on fuzzy c-means / O. Mulesa, F. Geche, A. Batyuk // Xth International Scientific and Technical Conference "Computer Sciences and Information Technologies" (CSIT) – 2015. – P. 60–62.

[10] Batyuk A. Apache storm based on topology for real-time processing of streaming data from social networks / A. Batyuk, V. Voityshyn // First International Conference on Data Stream Mining & Processing (DSMP) – 2016. – P. 345–349.

[11] Martsyshyn R. Technology of speaker recognition of multimodal interfaces automated systems under stress / R. Martsyshyn, M. Medykovskyy, L. Sikora, Y. Miyushkovych, N. Lysa, B. Yakymchuk // 12th International Conference on the Experience of Designing and Application of CAD Systems in Microelectronics (CADSM) – 2013. – P. 447–448.

[12] Sikora L. Using of n-ary taxonomic trees for saving and searching knowledge in the management process of technical systems / L. Sikora, Y. Miyushkovych // 10th International Conference - The Experience of Designing and Application of CAD Systems in Microelectronics – 2009. –P. 261–262.