A multidimensional data model for the analysis of learning management systems under different perspectives

Vanessa Araujo Borges
University of Sao Paulo
Sao Carlos (SP), Brazil
va.borges@icmc.usp.br

Bruno Magalhaes Nogueira
Federal University of Mato Grosso do Sul
Campo Grande (MS), Brazil
bruno@facom.ufms.br

Ellen Francine Barbosa
University of Sao Paulo
Sao Carlos (SP), Brazil
francine@icmc.usp.br

***Abstract*—The decision-making process in the educational context has been widely investigated as an effective mechanism to support educators and administrators in distance and blended learning. Pattern discovery in data produced by Learning Management Systems (LMS) elicits important information about the educational process, especially when considering correlations between stakeholders interacting in different perspectives of the LMS like grades, access, forum or message communications. In this paper, we propose a data and metadata model for Learning Analytics applications using data from LMS. This model is LMS independent and supports Business Intelligence and Data Mining applications for decision-making in the educational context. Our model considers the use of historical data from LMS treated under three perspectives that can be analyzed individually or correlated. The model also allows analysis considering different stakeholders (student, educator, etc.) and levels of granularity (institution, department, course, discipline, etc.). In this sense, the proposed model allows consistent and correlated analysis of data from LMSs. As a result, our model effective collaborates with the educational process through an efficient decision-making.**

I. Introduction

The discovery of potentially useful knowledge in Learning Management Systems (LMS) such as *Moodle* [1], *Sakai* [2], *Edx* [3], *TidiaAe* [4], among others, has been widely investigated in recent years. These research projects have as objective to aid educators and researchers in better managing these environments and, thus, in improving the educational process.

Converting brute data from educational environments into information capable of generating useful knowledge is a hard task, mainly due to data heterogeneity. In these environments, data can be analysed considering different stakeholders (students, educators, tutors, etc), under different perspectives (access and usage of environment, stakeholder features, means of communication, etc) and according to different levels of granularity (institution, course, discipline, etc) [5].

Learning Analytics (LA) is an emerging research area in this context. LA is grounded in two other traditional research areas: Business Intelligence (BI) and Data Mining (DM) [6]. According to Bienkowski *et al.* [7], LA refers to interpreting a huge amount of data produced by educational environments with the objective of evaluating the academic progress, forecasting future performance and detecting specific problems.

Data used in LA applications can be collected from explicit or implicit student actions. Examples of explicit student actions are completing tasks and doing exams. Implicit actions, on the other hand, are actions such as online social interaction, extracurricular activities, posting in discussion forums and other activities can not be directly evaluated as part of the education progress of the student [8].

Methods and tools have been proposed to support decision-making in the educational context using LA approaches. These methods and tools have diverse goals, such as [9]–[11]: to analyse access and usage in educational environments, to anticipate student dropout and to analyse student behaviour. However, these research projects explore specific educational contexts and most of them do not consider multiple data sources or different perspectives and data granularities. Thus, each of these methods and tools has its own data and metadata model, designed to deal with a single data resource (in general, a single LMS) and meet the needs of its specific application.

These applications can not be directly replicated in different scenarios, which restricts their usage. Moreover, the fact of not exploring different perspectives makes impracticable to these applications to manage and to discover novel knowledge about the entire institution and the different stakeholders involved in the teaching and learning process. These problems would be minimized by the adoption of reference models with data and metadata standards that enable the analysis of distinct perspectives and adequate scenarios for each stakeholder.

Motivated by this context, a data and metadata reference model for Learning Analytics is presented in this work. The main purpose is to provide a consistent model that can be used in LA processes that fulfil three requirements: (i) data from LMS stored in the model can be explored under different granularity levels (from less detailed to more detailed data); (ii) this model must enable the analysis of data involving different stakeholders (students, educators, etc); and (iii) the analysis performed in this model can use data from different perspectives (communication, grades, etc). The resulting data and metadata model is LMS independent and may be used in different contexts. This model corresponds to the core of a reference model that will also contemplate knowledge discovery using Data Mining methods. So, through the proposed model, we intend to transform data into relevant information to decision-making in a unified data and metadata model to effectively collaborate with the educational process.

This paper is organized as follows. In Section II, we provide a brief overview of the Learning Analytics, Learning Management Systems and its main concepts. Then, in Section III, we present some of the main related work. In Section IV, we define the methodology used to define the data model presented in this work, as well as the tools used to automatize this process. Following, in Section V, we present the data and metadata model considering its application to the *Moodle* LMS, provided by the University of Sao Paulo. Finally, in Section VI, we conclude and point out some future work.

II. Background

In this section, we briefly address some essential concepts of Learning Analytics and Learning Management Systems, which represent the background of our research.

A. Learning Analytics

Learning Analytics (LA) can be defined as the usage of data and analysis models to discover information and social connections with the objective of improving the teaching and learning process [12]. Data used in LA applications can be collected from explicit student actions, such as completing tasks and exams, or from implicit actions, such as online social interaction, extracurricular activities, posts on discussion forums and other activities that are not directly evaluated as part of the education progress of a student [8]. The adoption of LA process benefits different stakeholders, helping educators, students and administrators in performing their tasks [13].

The LA process is based on two research fields, Business Intelligence (BI) [14] and Data Mining (DM) [6]. BI focuses on searching for computational tools to aid decision-making through the efficient fusion of data collected in different functionalities of a system or from different systems. DM, on the other hand, search for new and potentially knowledge in a large amount of data.

LA provides a series of powerful tools based on BI and DM to help the different stakeholders of the educational process. Tasks as real-time feedback to student performance and recommendation of learning activities and materials help students to detect deficiencies, self-evaluate and improve their performance. Also, educators and administrators are able to better understand the students and to evaluate the teaching and learning process [13]. Some interesting applications of LA processes in recent years can be seen in [15], [16].

B. Learning Management Systems

The main data source for LA applications is the Learning Management Systems (LMS). LMS are computational platforms that act as an intermediate of the teaching and learning process. These platforms allow students and educators to communicate through, for example, forums, chats and messages, enabling collaborative learning. Moreover, LMS allow educators to produce and share content material, exams and exercises, facilitating the information sharing [15].

Data generated by the different stakeholders are stored by the LMS, from access logs to forums messages and content material. In general, LMS process transactions in relational datasets (OLTP – On-line Transaction Processing), with tables that represent the different entities and their relationships. This kind of storage can be is an operational-level representation as it simply stores transaction data and allows queries [17].

In LA applications, however, it could be interesting to store data in order to facilitate the analysis process. A data representation through an analytical processing (OLAP – On-line Analytical Processing) better meets these requirements. OLAP is used to represent data in informational or strategical levels. This representation allows structuring data such that information analysis under multiple perspectives or dimensions can be performed, i.e., a dimensional data structure.

Barbieri [17] identified differences between OLTP (operational) and OLAP (informational) data, described in Table I.

TABLE I: Main features of operational and informational data models [17].

Feature	Operational	Informational
Content	Current values	Summarized values, calculated and merged from diverse sources
Data organization	Application-driven / information systems	Topic-driven
Data nature	Dynamic	Static until programmed update, which occurs from time to time
Structure form	Relational, designed for transational computation	Dimensional, designed for analytical activities
Usage	Structured in tables; repetitive processing	Structured in facts and dimensions; analytical/predictive processing
Response time	Optimized for times below 1 second	Complex analysis and larger response time

The dimensional data structure of the informational data model was developed to respond queries from a Data Warehouse (DW) [18], [19]. According to Inmon (2005) [20], a DW is a data repository that preserves the following features:

1) Topic-driven data storage;
2) Data consolidated in a single source;
3) Variable along the time;
4) Non-volatile.

The dimensional model is composed of key tables, called **fact**, and a set of detailing tables called **dimension**. Generally, the dimensions have natural hierarchies and descriptive attributes that describe the characteristics of a fact. Among the main dimensional modeling techniques, we can cite [19]:

- **Star Schema**: in this model, the dimension tables are directly related to a central table (fact table). It is a simple and efficient model that does not take into consideration storage space economy and data normalization. On the other hand, it respects the precept of rapid information.
- **Snowflake Schema**: in this model, the dimensions are normalized and the hierarchies are separated. These features reduce the number of redundancies and allow maintenance agility.

In other words, a DW centralizes in a single repository of structured and unstructured data types that are relevant to the analysis. This repository contains current and historical data, which are necessary to the decision-making process.

In the educational context, the features of a DW are essential to solving problems inherent to educational data heterogeneity, considering its multiple perspectives, topics and granularities. Thus, it is possible, for example, to perform an analysis of data from an LMS considering the perspective of communication among the different stakeholders (number of posts in forums, chats, etc) associated with the student's grade, access statistics, among others. Moreover, it is possible to analyse such perspectives according to different stakeholders (educators, students, tutors, etc.). Finally, the DW storage makes easier to extract patterns using Data Mining approaches, since it provides data treatment in the ETL step and stores rich historical data in different levels of granularity.

In this paper, we present a data and metadata model based on DW storage. The steps performed to define this model instantiated in a BI process are presented in Section IV.

III. Related Work

It is possible to identify in the literature methods and tools proposed to support the decision-making process through LA approaches. These initiatives analyze relevant data to the educational process through BI analysis and pattern recognition in DM processes. These works use their own data model, designed specifically to meet the objective of the application.

Morris *et al.* [21] investigated student engagement in online and asynchronous courses. The analysis used data relative to participation frequency in the LMS (for example, the number of page views and publications) and relative to the duration of this participation (for example, the number of hours spent in the visualization of discussion and content pages). Results show that the time spent and participation frequency are important for determining success in the learning process.

In this empirical study, the authors used only access data about three semesters of three undergraduate courses. Thus, the analysis is limited to only one perspective (access perspective) without considering historical data.

Similarly, Ajmi *et al.* [22] investigated DM and DW techniques in their application considering only the access perspective. The patterns investigated were related to the usage mining in *Moodle*. The contribution of this research consists in using historical data and structured granularities in a DW. The DW model developed by the authors uses web services to create log formatted in an adequate way to pattern extraction through DM techniques.

In the research of Akintola *et al.* [23] the authors use historical data and time and course granularities in the grade perspective. The objective was to answer questions like "what is the trend of student admission, studentâs enrollments, Lecturers room schedule and the number of annual enrollment".

The usage of granularities and historical data in the grade perspective was analysed through n approach based on BI and DW in data from the Federal University of Technology Akure Nigeria. As a result, the researchers could show the effectiveness of using LA approaches for improving the decision-making in educational contexts.

The research of Mutean *et al.* [24] presents a dimensional model for *Moodle* and two other data sources. These data were related to only one semester of the Economic Informatics course, from the Faculty of International Business and Economics. The main objective of this model is to obtain estimations about the following questions: (i) *"How often students have accessed the e-learning platform?"*; (ii) *"Which resources are used more?"*; and (iii) *"Which educators are more active in terms of usage time?"*. The perspectives of access, communication and grades were used in this research considering a specific course and time contexts. Comparing this model to the models previously presented in this section, it has the advantage of using a diverse data source (*Moodle* and two other data sources). However, the database was related to just one semester of a course.

Guitart and Conesa [25] proposed a BI model to analyse data from the computer-science degree and computer-science master of Open University of Catalonia. This model explores the perspectives of access, communication/interaction and grades grouped into discrete categories according to student performances. The researchers highlight that the model, besides being restricted to this case study, allowed the reduction of time spent by educators in the information interpretation and management. The model, however, is very specific to the application and lacks of generalization and reuse aspects.

Finally, in the research of Gomex-Aguilar *et al.* [26], the authors analysed user interaction associated with the student success, using the grade perspective. The researchers used data clustering to this objective. In the end, the extracted patterns showed recurrent student behaviour in terms of usage frequency and academic performance along the different courses.

Considering the works cited in this section, it is possible to observe that great part of them have a specific purpose resulting the analysis using few volume data, specifics perspectives, and little or no granularity. Thus, it is not possible to analyse other correlated data in the educational process. This problem could be smoothed by the adoption of data and metadata reference models. Such models allow the analysis of different perspectives and scenarios adequate for different stakeholders of the educational process. As a consequence, reference data and metadata models for educational environments make solutions more suitable for different applications and easily replicable.

Next, we present our proposal of reference data and metadata model for LA applications instantiated in a BI process. In the next section, we introduce the BI approach used in this work to extract OLTP data from LMS to OLAP data, which is adequate to LA processes. Then, in Section V, we introduce our reference model for data and metadata representation.

IV. Business Intelligence approach for data transformation

Decision support systems (DSS), such as BI platforms, are frequently used to access and explore data stored in Data Warehouse in order to improve the decision-making process.

Usually, BI use different information sources to support decision-making and to define business strategies. These processes use data structures represented by traditional databases, Data Warehouse and Data Marts (in a lower scale) to dimensionally treat the information. Also, these approaches employ Data Mining techniques for correlating information and discovering patterns with implicit information [17], [18].

Additionally, BI contemplates a set of Extract Transform Load (ETL) tools that are essential for the transformation of operational resources into informational resources. In this sense, BI acts in a transformation layer that aims at producing correlated information that will result in knowledge at the end of the process.

In Figure 1, we show the steps of a BI architecture used in the scope of this work. The first step, **(1) External Sources**, refers to the usage of multiple data sources represented by transactional systems (OLTP) from an LMS, such as *Moodle*, or different data sources (academic systems, sheets, etc).

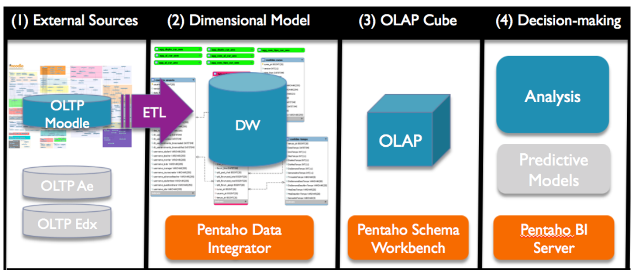

Fig. 1: BI architecture adopted in this work.

In the following step, **(2) Dimensional Model**, the operational data are transformed in informational data through an ETL process. The resulting representation is stored in a DW (dimensional structure). Then, in step **(3) OLAP Cube**, the process creates metadata necessary to translate the dimensional model in OLAP engines.

Finally, in step **(4) Decision-making**, the analysis generated in the analytical process is performed using the OLAP cube developed in the previous step. This cube contains metadata that represent the logical mapping of the physical model (DW). In this step, it is also possible to perform predictive analysis using DM methods.

In order to perform the BI process, we have adopted the set of tools available in the Pentaho platform. Pentaho is a BI suite available in two versions, Community and Enterprise. The Pentaho BI Community version[1] is an open source solution with some components that were used in this work:

- Pentaho Data Integrator (PDI) – in order to perform the ETL process;
- Pentaho Aggregation Design – for creation of aggregation tables;
- Pentaho Schema Workbench – for creation of the OLAP cube; and
- Pentaho BI Server or Pentaho User Console (PUC) – for creation and sharing of the performed analysis.

[1]Pentaho Community: http://community.pentaho.com/

Considering the BI process described, in the next section we discuss in details the transformation of the operational data into informational data. This translation process is the step of the process that demands a major analysis of the transactional model and its peculiarities. Thus, it requires both great human and computational efforts.

V. Proposal of a Data and Metadata Reference Model Learning Analytics

In this section, we present our data and metadata reference model for Learning Analytics applications. We highlight three aspects: the dimensional data model, the metadata from the logical model (the OLAP cube) and the dashboards and analysis.

A. *Dimensional data model*

In this research, the star schema model was chosen as the most adequate due to its modelling simplicity and fast response to queries. In Figure 2, we present the star schema (dimensional model) created. This model contains three Fact tables, four Dimensions and eight Aggregation tables.

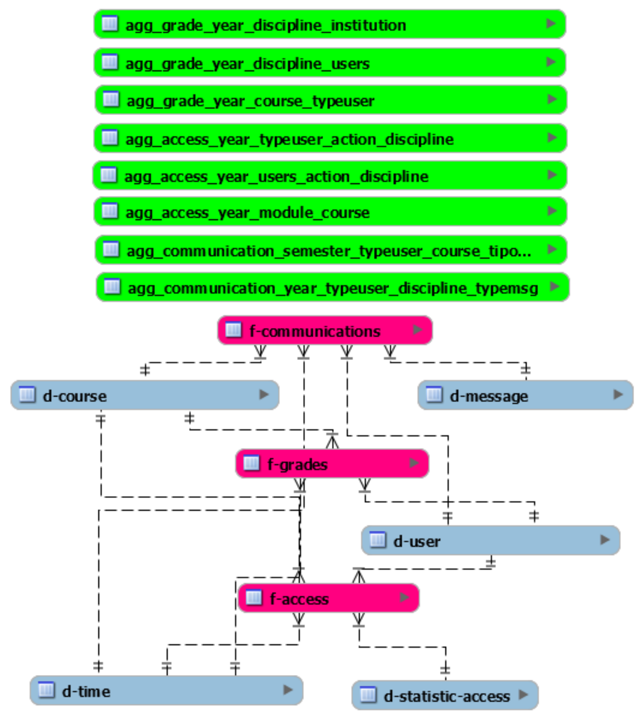

Fig. 2: Star Schema proposed in this work.

The Fact tables represent the quantitative information of the model. In this sense, it is simple to correlate data from the three distinct perspectives (communication, grades and access) considering their detailing in the Dimension tables.

The Aggregation tables, such as *agg_access_year_module_course* and *agg_grade_year_discipline_user*, have summarized information at some level of granularity. For example, the *agg_communication_semester_typeuser_course_typemsg* groups different types of user (student, educator, tutor, etc) for a semester (it could be for year, trimester, month, week, day, etc), per course (it could be for discipline or group), per type of message (post or read).

Aggregation tables allow a better performance in analysis with a greater level of granularities or data summarization. They also allow the direct application of DM processes that need data summarization at some level of granularity.

The model proposed herein allows associating, in the same analysis, data from communication functionalities (such as forums, chat and messages), grade history and access data from stakeholders. These data come from, respectively, Fact tables *f-communication*, *f-grade* and *f-access*. The detailing or qualitative information of these Fact tables are obtained through the following Dimension tables:

- *d-user*: contains information about stakeholders (students, educators, tutors, etc);
- *d-course*: describes the information about course characteristics considering course prerequisites and hierarchies (course, discipline, group, etc);
- *d-message*: describes the message and their characteristics (chat, forum, email, etc);
- *d-statistic-access*: contains information about stakeholders access; and
- *d-time*: describes time-oriented data (day, week, month, trimester, semester, year, etc).

In order to validate this model, in this case study we used data from *Moodle* provided by the University of Sao Paulo. It is important to highlight that besides our model being proposed to deal with different LMS, the database shown in this work refers to a preliminary validation of our model. For the sake of space and simplicity, this preliminary validation does not include other LMS and does not explore all of its interoperability properties. However, we have successfully transformed to our model data from different databases (academic systems, LMS) and different versions of *Moodle* (1.9.8, 2.6 and 2.7).

Moodle (Modular Object-Oriented Dynamic Learning Environment) [1] is an open source and free to use LMS. *Moodle* has diverse modular functionalities that can be freely optimized by the institutions.

Moodle has 203 tables that store data for every course, user, access, activity, exam, among other data. These tables were mapped from the original transactional model to the dimensional model proposed in this paper. We analysed 6,367 courses and 62,009 users divided in some categories (59,808 *students*, 2,203 *editingteacher*, 71 *advancedteacher*, 632 *teacher*, 2,496 *monitor*, 188 *manager*, 50 *coursecreator*, 3 *questionsharer*), from 2010 to 2015. These data are summarized in Table II.

B. Metadata from the logical model - OLAP Cube

Data granularity is one main factor to be considered in data modelling regardless the used architecture and implementation.

Granularity refers to the level of elements summarization and details available. The more detailed data, the lower is the granularity level. Having a low granularity level implies having higher detailing, but this may reduce the computational performance. On the other hand, having high granularity level implies processing fewer data and better performance in analysis, but this can make it impossible to respond to all necessary questions. Thus, a clear and coherent definition of granularity level is essential for the development of a project.

TABLE II: Amount of data in each dimension and fact table in the used *Moodle*.

Dimensions and fact tables	Number of tuples	Size	Obs.
f-communication	30726,748	6.6 GB	198 chat, 1,785 forum, 11,861 email
f-grade	588,830	157.3 MB	
f-access	11700,331	2.4 GB	
d-course	272,992	92.7 MB	6,367 courses
d-user	66,547	15.5 MB	62,023 users
d-message	7764,388	775.1 MB	
d-statistic-access	46672,839	8.1 GB	
d-time	99,099	16.5 MB	Dimension standard

In order to navigate through hierarchies and their different levels of granularities, it is possible to use operations like *drill-down* and *roll-up*. These operations allow performing aggregation, summarization and consolidation in the data. For example, it is possible to navigate in the time granularity from “year” (less data) to “day” (more data) passing through intermediate granularities (semester, trimester, month and week).

Moreover, for performing analysis from a DW mapped into OLAP cubes it is possible to navigate through data using *slicing* and *dicing* operations. These operations are important to reduce datasets in appropriated subsets for analysis or to examine the data from different perspectives. In a cube dimensional view, these operations allow reducing a data cube in smaller sets / cubes. These smaller cubes provide perspectives that may be more appropriate for the analysis than analysing the entire data cube. For example, it may be possible to analyse the user perspective (larger perspective) or the student perspective (a subset from the user perspective).

In the educational context, these operations and properties are essential for efficient LA applications considering the various perspectives and levels of granularity.

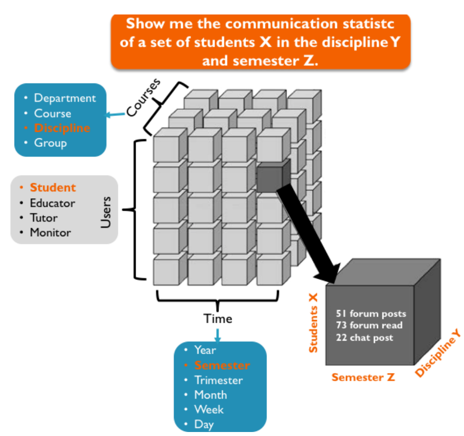

Fig. 3: Communication statistics cube with its granularities and stakeholders.

In order to understand the granularity and perspectives, in Figure 3 we present a simple analysis that returns communication metrics for a *set of students X* that attended *discipline Y* in *semester Z*. This query ran through the hierarchies of user, course and time dimensions.

The data model proposed in this work has tree data cubes - *communication*, *grades* and *statistic-access* cube, as described in Figure 4. In general, data cubes represent a logical model (metadata) that acts like a pointer to the physical model (dimensional model). Thus, each data cube of our model has references for quantitative (present in fact tables) and qualitative information (present in dimension tables).

OLAP cubes are able to store aggregation operation, like arithmetic and logical formulas, that must be applied to the physical data. These operations are responsible for creating the adequate measures for analysing the data. Thus, all logical operation may be centralized in one layer of the BI process - the OLAP cube. The OLAP cube proposed have the following measures associated to each of the defined perspectives:

- **Communication perspective**: an amount of messages posted in discussion forums; amount of posts read in discussion forums; amount of chat posts; amount of messages sent by email; amount of email messages read; amount of disciplines in which enrolled students use communication tools; amount of total communication (forum, chat and email) sent from the LMS.
- **Access statistics perspective**: an amount of accessed disciplines; an amount of different IPs (Internet Protocol) that accesses the LMS; an amount of accesses to the LMS; an amount of user that accessed the LMS.
- **Grades perspective**: amount of published grades; average of final grades; an amount of users with grades published in the LMS; an amount of courses with grades published in the LMS; an amount of disciplines with grades published in the LMS.

The correlation of these three data cubes is possible through a fourth and central OLAP data cube, a ***virtual cube***. In other words, the virtual cube is a reference for the other three cubes, each of them defining a perspective of the data.

In our application in the *Moodle* environment, this fourth cube is called ***virtual moodle*** and is presented in Figure 4.

Considering these four OLAP cubes, it is possible to map quantitative data and perform correlations that are independent of perspective from these data. Moreover, these quantitative data can be correlated with information from dimension tables detailed in different levels.

C. Dashboards and analysis

Visual data analysis, correlating stakeholders and perspectives, may be decisive during decision-making in educational context. The visual analysis of LMS brute data, however, may be a very complex task. This brute data is untreated and heterogeneous. On the other hand, when using a dimensional model with well-defined structure (dimensions and facts) and granularities mapping, the correlated data analysis is an easy task. This task may be carried out through analytical visualization (OLAP) tools.

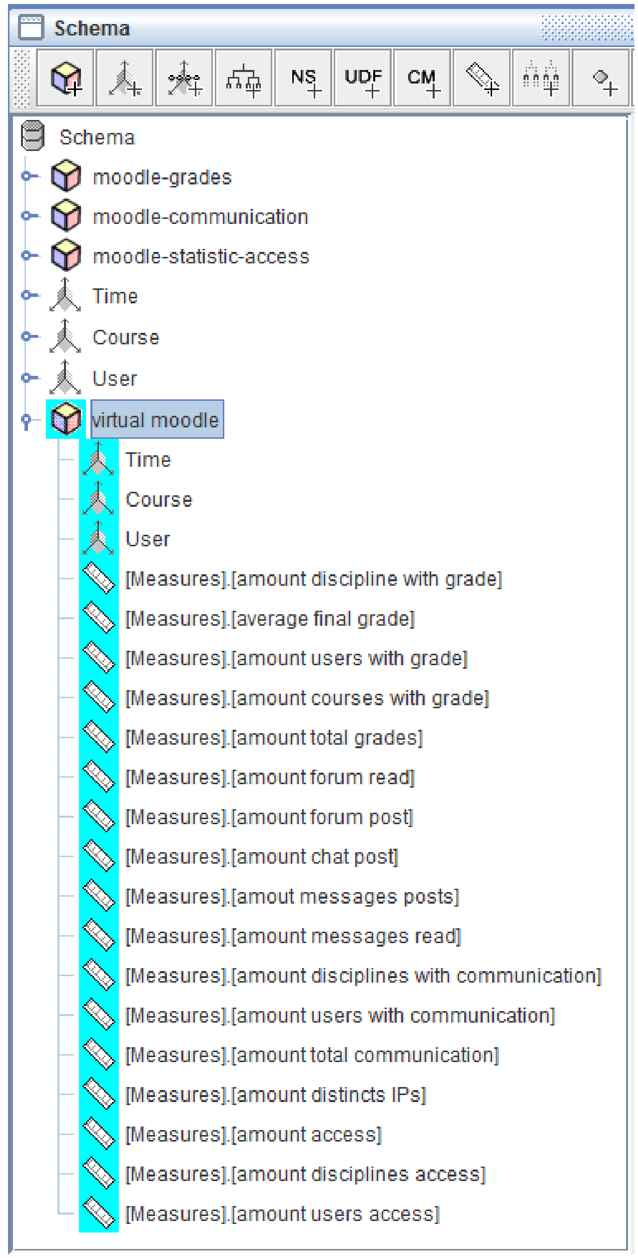

Fig. 4: OLAP cubes defined for the three analysed perspectives.

In this sense, a small effort is required from a user to generate visualizations (such as graphs) from our proposed model, since it has a well defined dimensional model.

In our model validation using *Moodle*, considering the quantitative data and their possible correlations, we defined two initial dashboards that translate these data to interactive graphs to support the decision-making process. These dashboards allow to analysing the perspectives of learning management via LMS and the educator perspective, as shown in Figure 5. These dashboards were created using the construction and visualization tool from the Pentaho suite (Pentaho User Console) and show the following information:

- **The dashboard of the administrative perspective**: it contains dynamic filters of year and type of user.

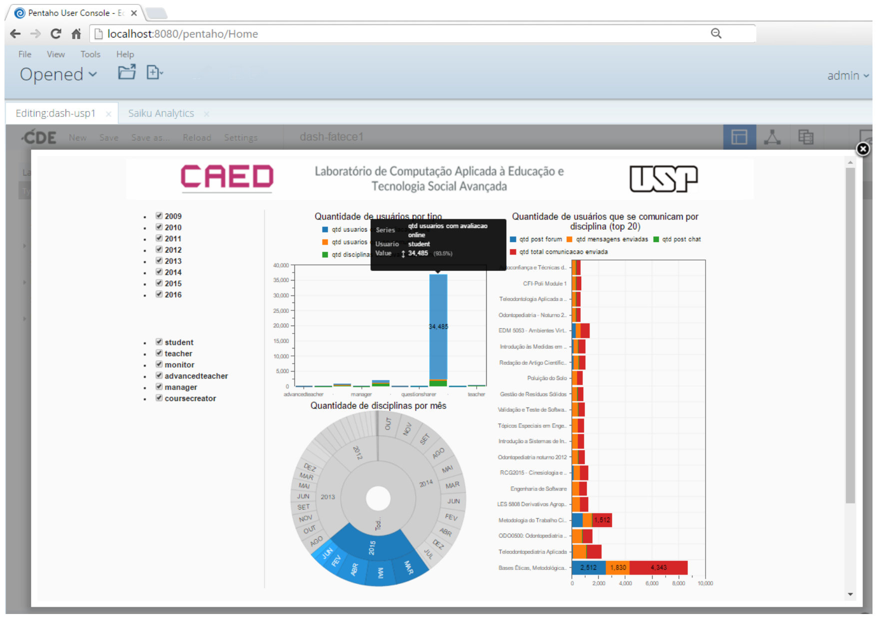

Fig. 5: Administrative dashboard created using Pentaho User Console.

It shows information about the amount of users that access the LMS; amount of users that have grades; and amount of users that communicate with other users per user type (students, educators, etc); the top 20 disciplines with more communication considering chat, forum and emails; and the amount of access per action (login, view, etc).

- **The dashboard of the educator perspective**: it has a dynamic filter of year and professor name. It shows information about an amount of courses and disciplines; the amount of communication per course, discipline and group; a list of students that never accessed the courses; a list of the top 20 students and courses that have more access to the LMS; and the grade average per discipline.

We highlight that it is possible to create different dashboards and analysis other than the examples we showed here. These new dashboards and analysis could address different perspectives and could be easily created from our data model. Moreover, it is important to highlight that our model is platform-independent and can be used in different BI platforms.

VI. Conclusion and Future Work

In this work we presented a dimensional model for LA applications that enables the analysis of three different perspectives, considering different stakeholders and different levels of granularity. These characteristics are innovative and collaborate to knowledge discovery and management in the teaching and learning process.

Our model was inserted in an LA process using BI process and tools. This process was automatic, which enables the ETL process to be performed in time intervals (hour, day, week). Thus, it is possible to create information in an almost online way considering historical data.

An initial validation was carried out using data from a *Moodle* LMS (version 2.6) from the University of Sao Paulo, containing data from 2010 to 2015. The performed analysis show that it is possible to extract correlated information from different perspectives and granularities through our model. The results obtained until this moment show that our model can facilitate the management of LMS and their different stakeholders providing intuitive and easy-to-create analysis based on BI approaches.

As future work, we intend to use data from others LMS and academic systems in our model. Besides, we are interested in patterns that could be extracted using statistical models and DM algorithms in order to generate a descriptive and predictive analysis from our model.

Finally, we intend to present the generated dashboards to specialist evaluation (educators, coordinators, etc). The goal is to obtain insights about the information that would be helpful to educational and decision-making processes.

VII. Acknowledgment

The authors would like to thank the Brazilian funding agencies - FAPESP (Process 2014/03389-9), CAPES (Procad 071/2013) and CNPq.

References

[1] D. G. Moodle, "Moodle – a free, open source course management system for online learning," 2006. [Online]. Available: http://moodle.org/

[2] C. Counterman, G. Golden, R. Gollub, M. Norton, C. Severance, and L. Speelmon, "Technical Report Sakai Project," may 2004. [Online]. Available: http://sakaiproject.org/

[3] MIT and Harvard University, "Edx – online education," may 2013. [Online]. Available: https://www.edx.org/

[4] Tidia, "Ae: Aprendizado eletrônico na internet avançada," 2005. [Online]. Available: http://www.larc.usp.br/conteudo/me/TIDIA_Ae_WP_v05.pdf

[5] C. Romero and S. Ventura, “Data mining in education,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 1, pp. 12–27, 2013.

[6] S. Buckingham Shum and R. Ferguson, “Social learning analytics.” *Educational Technology & Society*, vol. 15, no. 3, pp. 3–26, 2012.

[7] M. Bienkowski, M. Feng, and B. Means, “Enhancing teaching and learning through educational data mining and learning analytics: An issue brief,” *Washington, DC: SRI International*, pp. 1–77, 2012.

[8] L. F. Johnson, A. Levine, R. Smith, and S. Stone, “The horizon report: 2010 edition,” 2010. [Online]. Available: http://www.nmc.org/pdf/2010-Horizon-Report.pdf

[9] A. Ingram, “Using web server logs in evaluating instructional web sites,” *Journal of Educational Technology Systems*, vol. 28, no. 2, pp. 137–57, 1999.

[10] R. Baker and K. Yacef, “The state of educational data mining in 2009: A review and future visions,” *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.

[11] A. J. C. Kampff, “Mineração de dados educacionais para geração de alertas em ambientes virtuais de aprendizagem como apoio à prática docente,” Ph.D. dissertation, Universidade Federal do Rio Grande do Sul, 2009.

[12] G. Siemens, “What are learning analytics,” *Retrieved March*, vol. 10, 2010.

[13] M. Vahdat, A. Ghio, L. Oneto, D. Anguita, M. Funk, and M. Rauterberg, “Advances in learning analytics and educational data mining,” in *ESANN ’15: Proceedings of the 2015 European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2015, pp. 297–306.

[14] H. J. Watson and B. H. Wixom, “The current state of business intelligence,” *Computer*, vol. 40, no. 9, pp. 96–99, 2007.

[15] C. Romero and S. Ventura, “Educational data mining: A survey from 1995 to 2005,” *Expert Systems with Applications*, vol. 33, no. 1, pp. 135–146, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417406001266

[16] A. Peña-Ayala, “Educational data mining: A survey and a data mining-based analysis of recent works,” *Expert Systems with Applications*, vol. 41, no. 4, Part 1, pp. 1432–1462, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417413006635

[17] C. Barbieri, *BI2 – Business intelligence: Modelagem & Qualidade*. Elsevier, 2011.

[18] R. Kimball and M. Ross, *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons, 2011.

[19] R. Kimball, *The data warehouse lifecycle toolkit*. John Wiley & Sons, 2008.

[20] W. H. Inmon, *Building the data warehouse*. John wiley & sons, 2005.

[21] L. V. Morris, C. Finnegan, and S.-S. Wu, “Tracking student behavior, persistence, and achievement in online courses,” *The Internet and Higher Education*, vol. 8, no. 3, pp. 221 – 231, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1096751605000412

[22] M. F. AlAjmi, S. Khan, and A. Sharma, “Studying data mining and data warehousing with different e-learning system,” *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 4, no. 1, 2013.

[23] K. Akintola, A. Adetunmbi, and O. Adeola, “Building data warehousing and data mining from course management systems: A case study of futa course management information systems,” *International Journal of Database Theory and Application*, vol. 4, 2011.

[24] M. Muntean, A. Bologa, R. Bologa, and A. Florea, “Business intelligence systems in support of university strategy,” *Recent Researches in Educational Technologies*, pp. 118–123, 2011.

[25] I. Guitart and J. Conesa, “Analytic information systems in the context of higher education: Expectations, reality and trends,” in *Intelligent Networking and Collaborative Systems (INCOS), 2015 International Conference on*, Sept 2015, pp. 294–300.

[26] D. A. Gómez-Aguilar, Ángel Hernández-García, F. J. García-Peñalvo, and R. Therón, “Tap into visual analysis of customization of grouping of activities in elearning,” *Computers in Human Behavior*, vol. 47, pp. 60 – 67, 2015, learning Analytics, Educational Data Mining and data-driven Educational Decision Making. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563214005779