# A Reference Architecture for Educational Data Mining

Francisco A. de Almeida Neto, Alberto Castro
PPGI – Informatics Postgraduate Program
Institute of Computing, Federal University of Amazonas (IComp/UFAM)
Av. Rodrigo Otávio, 6200, CEP: 69080-900, Manaus–AM, Brazil
francisco.neto@icomp.ufam.edu.br, alberto@icomp.ufam.edu.br

***Abstract*— In this paper we present a reference architecture for ETL stages of EDM and LA that works with different data formats and different extraction sites, ensuring privacy and making easier for new participants to enter into the process without demanding more computing power. Considering scenarios with a multitude of virtual environments hosting educational activities, accessible through a common infrastructure, we devised a reference model where data generated from interaction between users and among users and the environment itself, are selected, organized and stored in local "baskets". Local baskets are then collected and grouped in a global basket. Organization resources like item modeling are used in both levels of basket construction. Using this reference upon a client-server architectural style, a reference architecture was developed and has been used to carry out a project for an official foundation linked to Brazilian Ministry of Education, involving educational data mining and sharing of 100+ higher education institutions and their respective virtual environments. In this architecture, a client-collector inside each virtual environment collects information from database and event logs. This information along with definitions obtained from item models are used to build local baskets. A synchronization protocol keeps all item models synced with client-collectors and server-collectors generating global baskets. This approach has shown improvements on ETL like: parallel processing of items, economy on storage space and bandwidth, privacy assurance, better tenacity, and good scalability.**

***Keywords*— *educational data mining; learning analytics; distance education***

## I. INTRODUCTION

An increase in computation power and quality of systems used for online education resulted in an increase of information recorded by education systems about interactions between the system and its users and between users themselves. This increase in information quantity has pushed forward the fields of Learning Analytics (LA) and Educational Data Mining (EDM). Yet, to get to higher levels, EDM and LA need an architecture that can handle differences between databases of several virtual learning environments and to improve the data gathering process.

LA and EDM require development of specific tools and knowledge discovery to help teachers and students in the learning process, both present or at distance. LA and EDM can help teachers to answer questions like the following:

- What sequence of topics is more effective for a specific student?
- When exactly are the students ready to move to the next topic?
- What student actions are associated with better learning and better grades?
- When exactly a student is in risk of failing a course?
- What actions are indications of satisfaction and engagement?
- What grade a student will probably get without intervention?
- What environment qualities translate to better learning?
- Should a student be sent to a counselor?

EDM and LA have been seen as important resources to the future of learning and advances in these fields should be quite relevant to Education. The work described here shows a way to handle different databases and to improve data mining process in EDM and LA. We reduced the amount of data to be transferred as well as the time needed to finish data gathering and preprocess stages.

## II. CONTEXT

The following are some concepts presented under the perspective of our proposal.

### A. *Educational Data Mining - EDM*

A description of EDM with key topics and objectives of the field can be found in [1]. EDM is defined as a research field with focus in development of methods to explore sets of data gathered from virtual learning environments. Its objective is to get new information about the relationship between educational data in a way we can produce new knowledge or find scientific discoveries such as: a better understanding about the student behavior, or a better understanding about the learning process in virtual environments. To process big volumes of data, EDM is developing and using expertise from several subfields of Computer Science.

### B. *Learning Analytics - LA*

Elias *et al* [2] describe LA as a field that uses sophisticated analysis tools to improve the learning process. LA uses expertise of other research fields like: Business Intelligence,

978-1-5090-5920-1/17/$31.00 ©2017 IEEE

Web Analytics, EDM and Action Analytics. LA aims to use its skills to make models to predict behavior, to react according to predicted behaviors and to provide information to the learning process in a way to always improve the process and his behavior models. Although LA is still a new field, UNESCO has already published recommendations to researchers [15].

### C. *Market-Basket Model*

The Market-Basket Model is inspired by a scenario of shopping in a market. The market offers a large amount of items and every time a client buys he gets a little subset of the total items in her basket. Could a market chain, with a lot of markets and a lot of clients, find patterns of shopping by its clients so it can improve its sales?

Rajaraman and Ullman [3] state that data Market-basket model is a commonly used description for many-to-many data relationships. At one end we have items and in the other we have baskets (database transactions). Each basket is formed by an item set and the average number of items in a basket is very small when compared with the total number of items. Other point is that the total number of items is very large (not possible to store in RAM) so it is assumed that the data is stored in a file in the form of a sequence of baskets.

A typical example for using Market-basket model in the context of Virtual Learning Environments is the student-grades relationship. Items in the basket are the grades of a student in each assignment and a basket is a set of grades that a specific student got (we have a basket for each student) in a specific time interval (for instance an academic year).

### D. *ETL Process*

ETL (Extract, Transform, Load) is a central stage in educational data mining process. In this stage the process gets raw data (from databases, logs, etc.) as input, and produces data where the mining process will be carried out (baskets). It is during this stage that inconsistencies are removed. Examples of data inconsistencies are:

- Same information stored as different data
- Same information stored with different formats

In ETL is also common to do computation (sums, averages, etc.) needed to get the data requested.

Normally during ETL all the raw data is gathered in a place before an ETL tool do its job.

### E. *Reference Model*

A reference model, as said in [4], is an abstract framework that describes key entities and key relationships between entities of an environment. It allows us to create patterns and consistent specifications that can be used to make specific references while developing a system, or to build a more concrete architecture (like a reference architecture). So, a reference model needs to have the following characteristics:

- To be ‘abstract’. It needs to have a minimal set of unifying concepts that describes entities types that exists inside the problem domain. It should not describe specifics entities.
- To have a minimal set of axioms about the problem domain.
- To describe a minimal set of relationships between the concepts of the problem domain. It should describe how the concepts work together and if there is any common element between the concepts.
- To be independent of patterns, technologies and implementations or any other more concrete details.

The objective of a reference model is to be used as a common place where various implementations can be anchored.

### F. *Reference Architecture*

Brown *et al* [5] says that a reference architecture models the abstract elements of an architecture that exists inside a problem domain, however, this is still made independently of technologies, protocols or products that would be used in an implementation stage. It differs from a reference model because it is a clearer image of the solution since it states all entities in the model.

We can say that the objective of a reference architecture is to identify abstract solutions to the problem domain by mapping a reference model to software elements. Thus, a reference architecture can be applied to all problems of a class.

## III. PROBLEM DEFINITION

The problem we tackle here refers to optimization of ETL stage in educational data mining process with various sources for raw data to be used with methods that adopt Market-Basket Model.

We would like to answer the following questions:

1. Can we create a reference model to be used by other researchers?
2. Can we create a reference architecture to be used in the development of suitable solutions?

Both fields - EDM and LA - use methods of data mining as tools for their research. Both have the same problems with data gathering and preprocess stages of data mining. Thus, to better contextualize the problem we will describe which stages in EDM and LA could be improved through our proposal.

### A. *EDM Context*

In EDM, as said in [11], the use of data mining methods follows the same cyclic pattern used in general data mining:

- Data collection
- Data preprocessing
- Application of data mining algorithms
- Results interpretation, evaluation and deployment
- New data mining cycle

In their work, Romero *et al* [11], show an example of using EDM with Moodle as the Virtual Learning Environment. Their example is show in Figure 1.

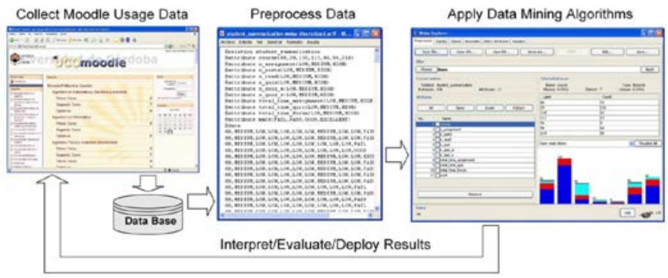


Figure 1 – EDM with Moodle [11]

We propose here an improvement on collection and preprocessing phases.

### *B. LA Context*

Elias and Lias [2] show the model for a LA process shown in Figure 2. They used several models and frameworks (Knowledge Continuum, Five Steps of Analytics, Web Analytics Objectives e Collective Applications Model).

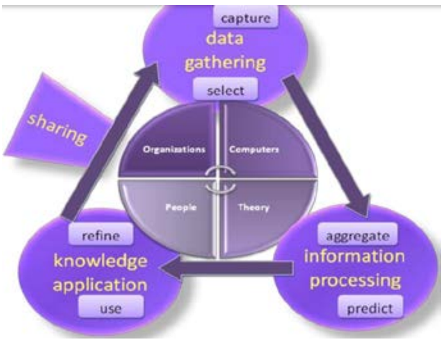


Figure 2 – LA process [2]

Just like EDM, LA process is a cyclic improvement process. In this paper we propose an improvement to be applied over 'select' and 'capture' phases of LA process.

### *C. Not in context*

The reference model and reference architecture proposed in this work are not to be used to explain or to define the metadata collected. Our work ignores what the metadata (the items in the basket) are about. It focuses on the needs for collecting and transmitting the metadata to a central place where someone will use the baskets as input for a data mining process.

### *D. Why ETL stage needs optimization?*

The first problems traditional ETL comes up against in a Big Data context are storage amount and bandwidth needed to collect raw data. In addition to this, source environments of raw data are not homogeneous. Other points to add are information privacy and security of data transferred (because the sources don't have a unique owner). All this considered, we might conclude that the traditional way is not good in a Big Data context.

Another improvement to ETL has to do with the process not being fully detailed about EDM or LA in large scale, it is only cited as a stage in a larger process. Examples of this type of description can be seen in [6][7][8][10]. Yet at this point, even if a paper does not describe details about its ETL stage it is still possible to use our proposal with little modifications inside other platforms or frameworks like those in [16] and [17].

In reports where there are more details about ETL stage, they usually do not consider multiple sources for the raw data. This is the case of the work described in [9].

Next, we discuss two of the biggest problems on traditional ETL in data mining with virtual learning environments: storage space and management issues.

### *E. The Storage Space Problem*

The growth of online courses over the Internet made the number of virtual learning environments explode. Every one of these environments has lots of tools and each tool generates a significant quantity of information about students and their learning process. This information can be aggregated and mined so we can find new knowledge to improve the online learning process. As an anchor to our discussion about storage space needs we will use Moodle as an example because it is a very popular and open-source virtual learning environment.

Taking as first point the size of events record, we have in Moodle a table called 'logs'. This table stores all the action done by users inside Moodle. In Figure 3 we show the description of this table in Moodle version 3.1.3.

| # | Nome | Tipo | Colação | Atributos | Nulo | Padrão | Extra |
|---|---|---|---|---|---|---|---|
| 1 | id | bigint(10) | | | Não | Nenhum wrap (padrão: none) | AUTO_INCREMENT |
| 2 | eventname | varchar(255) | utf8_general_ci | | Não | | |
| 3 | component | varchar(100) | utf8_general_ci | | Não | | |
| 4 | action | varchar(100) | utf8_general_ci | | Não | | |
| 5 | target | varchar(100) | utf8_general_ci | | Não | | |
| 6 | objecttable | varchar(50) | utf8_general_ci | | Sim | NULL | |
| 7 | objectid | bigint(10) | | | Sim | NULL | |
| 8 | crud | varchar(1) | utf8_general_ci | | Não | | |
| 9 | edulevel | tinyint(1) | | | Não | Nenhum wrap (padrão: none) | |
| 10 | contextid | bigint(10) | | | Não | Nenhum wrap (padrão: none) | |
| 11 | contextlevel | bigint(10) | | | Não | Nenhum wrap (padrão: none) | |
| 12 | contextinstanceid | bigint(10) | | | Não | Nenhum wrap (padrão: none) | |
| 13 | userid | bigint(10) | | | Não | Nenhum wrap (padrão: none) | |
| 14 | courseid | bigint(10) | | | Sim | NULL | |
| 15 | relateduserid | bigint(10) | | | Sim | NULL | |
| 16 | anonymous | tinyint(1) | | | Não | 0 | |
| 17 | other | longtext | utf8_general_ci | | Sim | NULL | |
| 18 | timecreated | bigint(10) | | | Não | Nenhum wrap (padrão: none) | |
| 19 | origin | varchar(10) | utf8_general_ci | | Sim | NULL | |
| 20 | ip | varchar(45) | utf8_general_ci | | Sim | NULL | |
| 21 | realuserid | bigint(10) | | | Sim | NULL | |

Figure 3 – Logs from Moodle

Using the description shown in Figure 3 we can calculate the lowest size of a record in this table (best case is when 'varchar' has 1 byte prefix + 1 byte data). Doing the maths, we have: 10 *bigint* fields, 8 *varchar* fields and 2 *tinyint* fields, that sums up to 10*8+8*2+2*1 = 98 bytes per record. Considering now the worst case (we consider *longtext* as a 30 byte field - we do this because its max size should dominate all the maths and in a database we had access its average was 29.5 bytes) we have a record size of 773 bytes. Suppose we have 100 active uses with average of 2 active hours by day and an average of 1 action by minute, we would have a size of 1,1MB per day in the best case and 8,8MB in the worst case. If we collect data from 100 environments (we have carry out tests with project UAB – the Brazilian Open University that have a lot more than that) we should have 112MB in the best case and 884MB in the worst case. In a year (365 days) we should have 40GB in the best case and 315GB in the worst one.

Not all the information we want is inside the logs table, we need information that is in other tables of the database, like: what messages from users are answered, what the impact a certain user have in the forums they participate, etc. This information is distributed between several tables of Moodle (just like in other virtual environments), so as a possible solution we could copy the entire database from that environment. However, a current Moodle environment should have near 200 tables and these tables when summed up could result in numbers near the same or greater than the ones from the log table. In that case our estimate should at least be doubled.

As illustrated by this typical scenario, potential data volume for mining process can easily match a Big Data context.

Another problem derived from data volume is to transfer this data to a central place where it would be mined. This should demand a good bandwidth and management of transfer process so it would not overload an Internet link while transfer happens.

### *F. The Management Issue*

Although storage space is an important issue, it could be tackled by buying enough storage, computer power and bandwidth to do all data transfer and preprocessing. Since nowadays we have commercial solutions for this (again, given enough money) we could reduce the problem to a money issue.

The central problem is not a place to put the data but the fact that data is collected from various sources, possibly from different owners. Each owner has its autonomy and specific policy and still it must attend to general policies and regulations. To add another level on complexity, one should remember that raw data is about people and so it will involve privacy and security issues.

Data owners might feel uneasy allowing external access to their databases. There is a natural resistance to it because of sensitive and personal information (from teachers, admin staff and students) stored there, and once shared, usage would go out of owner's control. Thus, to negotiate a deal involving several owners is not an easy nor quick matter to deal with.

### *G. On Improvements*

The solution presented here has shown some improvements upon current approaches:

- To work with baskets and not with original databases lowers storage space and bandwidth
- The ability to work with raw data under different formats and storage choices is increased and ETL process stays more database independent, so a change in database organization would not have a high impact upon ETL.
- It is possible to add new databases as sources to ETL process. An independency from structure of databases eases the use of ETL in a Big Data context.
- Finally, privacy and security issues are better dealt with since data is transferred in baskets and not taken directly from a database.

## IV. Related Work

Although we did not found any work presenting a solution in the same way we show here. The following papers describe related proposals involving all data mining process:

DeFreitas and Bernard [6] show a framework for use in development of flexible educational data mining environments where planned tools can be used either by experts or beginners in data mining.

Kechadi [7] talks about Distributed Data Mining (DDM) and points out that traditional methods are limited and new ones are necessary to handle the problem of data distributed by different environments. It is speculated that this might happen because this area still haven't enough researchers.

Lepouras *et al* [8] describe a vision of a platform for learning analytics that can be used by all stakeholders in the educational process (students, teachers, admin staff and policy makers). The platform is designed to use various sources for raw data and different data mining methods.

There are proposals dealing with more specific aspects that are closer to ours:

Krüger *et al* [9] show a data model to structure and export educational data that are mined from the interaction between users and the learning objects in the environment, working upon the preprocess stage of data mining. Only one environment and only data from students are considered.

Ngo *et al* [10] tackle some of the same questions about privacy and information security we have done in our work. They present a framework to extract and aggregate information to be used in analytics research about higher education in USA and there are issues on how to replicate its structure.

Butoianu [16] and Wolpers [17] while explaining the Attention Metadata Approach, describe the necessity of gathering information from various sources. Because their approach is focused on the student and his behavior in others contexts besides the VLE, it does not gather any information that is not directly about the student. Although their approach is more general than the one used in the framework described here, the latter could still be used with our reference model and reference architecture in their ETL stage.

## V. Designing a Solution

The development of our reference model was based in an analogy of the gathering process that happens with farms. Each farm in a region is responsible by its goods. At gathering time, each farm gathers its goods and puts them in silos. After a while, trucks gather the goods from farm silos and transport it to a central silo.

Like the farms, each virtual environment should produce its baskets of items and after baskets are ready, they could be send to a central server where the data mining process is done.

Next, we describe in detail our reference model and reference architecture.

## *A. The Reference Model*

To represent the reference model we choose to use a concept map, shown in Figure 4. The concepts are described below.

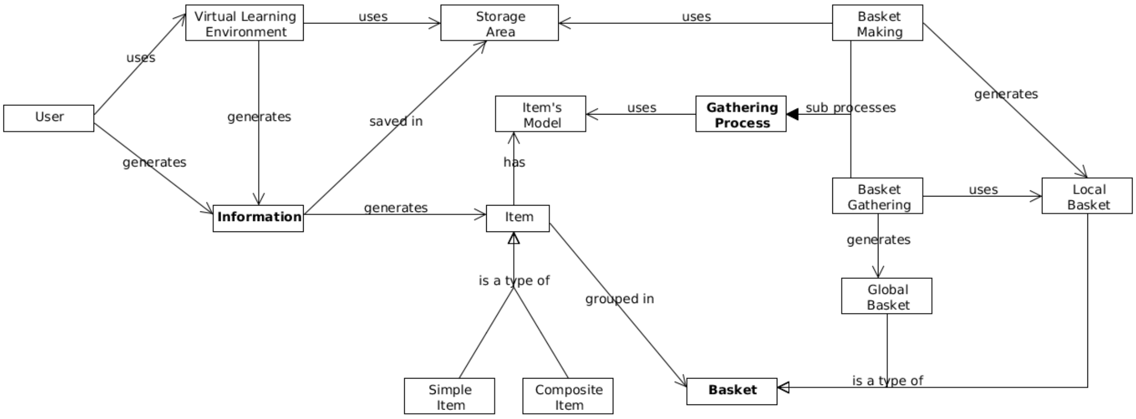


Figure 4 – A Reference Model to ETL Process

### *1) Information*

We borrowed a definition of information from Knowledge Continuum in the same way used by [13] *apud* [2] and described in [14].

**Axiom 1:** Data is an observable fact of the virtual environment. It can be generated by users or by the environment itself. For example: a forum access, a grade in an assignment, a status in a course.

**Axiom 2:** Information is the junction of data and its meaning. Information can answer one or more of the five following question: Who? What? When? Where? How many? For example: John accessed the forum two times. William got grade 5 in the assignment. Elias was approved in the course.

**Axiom 3:** Information cannot be an answer to questions Why? or How?

Information can be generated by users or by the environment itself and it is saved in a storage area so it be persisted.

### *2) Virtual Environment*

A virtual environment is a place where not only the learning process happens but where users can interact with each other or with the environment itself. The interaction between users includes not only the relationship student/teacher but also all other relationships (teacher/tutor, tutor/student, student/manager, etc.).

Virtual environments use storage areas.

### *3) User*

Users are all participants (both human and non-human) within the virtual environment, including: teachers, managers, students, chatbots, etc.

### *4) Storage Area*

The storage area holds all data and information produced during normal use of the environment, some can be easily accessed and other might require additional processing. For example, to compute the average grade of a group of students would require to know the grade of every student and carry out the necessary maths.

### *5) Basket*

Our proposal is focused on data mining methods that use the Market-Basket Model. These methods are exploratory and aimed at finding interesting patterns inside data. Further analysis and study of these patterns can result at discovery of new knowledge.

Thus, the basket in our context is made of information about a specific student inside a specific course. If the same student coursed two courses this means we would have to create two baskets.

**Axiom 4:** Every basket is unique for a student-course relationship. The basket is an item set that is a subset from the set formed by all the items in the environment.

**Axiom 5:** Every item can be in all baskets.

**Axiom 6**: Every item can be present only one time in a basket.

To gather all the baskets the system executes a gathering process.

#### 6) *Item*

An item is the basic component of a Basket. Every item represents a piece of information in the Basket. An item can be classified as a Simple Item or a Composite Item.

##### a) *Simple Item*

An item is simple when to get its value one only needs to verify the storage area. For example: Quantity of times a user accesses a forum.

**Axiom 7:** A Simple Item is an Item that does not use any other Item in its structure.

##### b) *Composite Item*

An Item is considered composite when to get its value we need to compute or combine two or more simple items. For example: Percentage of posts of a user that got an answer.

**Axiom 8**: Every Composite Item can be split in two or more simple items and one or more math operation.

**Axiom 9:** A Composite Item can use other composite items in its structure.

#### 7) *Item's Model*

The Item's Model has the specification of how to create a certain item. It is formed by:

- A unique name identifying the model
- A description about a target Item of the model. The description goal is to make clear to stakeholders of data mining, the meaning of the item.
- Instructions on how to get an item. The algorithm can be different according to each environment, but the same Item must be returned.

#### 8) *Gathering Process*

The gathering process explains how the work is split between various environments e what actions must be taken in each environment. It also should specify how to handle conflicts between various subprocesses and when all baskets are stored in a central place.

The gathering process is split in two subprocesses: Basket Making and Basket Gathering.

##### a) *Basket Making*

The Basket Making process is analogous to the process of seeding and harvesting in a farm. This process defines and explains needed actions inside the environments to get the item information and to create the baskets.

The baskets created by this process are the Local Baskets.

*Local Basket* – A Local Basket is an item setting where the items are collected inside a virtual learning environment for a relationship student-course. Throughout the next phases local baskets will be collected and aggregated in an only place.

##### b) *Basket Gathering*

The basket gathering process is like the process that a farm uses to send its goods to a central storage. This process should explain how local baskets are harvested and how the global basket is made. Restrictions for harvesting of local basket (like volume, velocity and time) need be explained and dealt with.

While a global basket is still been built, basket gathering can create composite items to be added to the global basket. This is normally used when computing needs information about more than one environment. An example of a composite item where this would be necessary is the ranking of environments by using total use of chat rooms.

*Global Basket* – A Global Basket is the goal of the work described here. It is the union of all local baskets of all environments plus the composite items computed during the basket gathering process.

### B. *A Reference Architecture*

Using the reference model proposed here and a client-server architectural style, we created the reference architecture shown in Figure 5, where each environment is a client and the place where the global basket will be stored is the server.

Next, we describe the main elements of this architecture and possible advantages of it.

Virtual Environment, Local Basket, Global Basket and User are as described for the Reference Model.

#### 1) *BD + Logs*

A virtual environment stores data needed to normal working in a database and logs may or may not be stored in the same database. In some cases, logs are split and stored in various places. In any case, the collector-client needs to have access to both the DB and the logs so it can harvest all the information needed.

#### 2) *Collector-Client*

The collector-client is a piece of software that will be executed inside the virtual environment and will collect the information needed to create local baskets and later send them to collector-server. A key function of collector-client is to synchronize the item's model to be sure it stays updated during all ETL process.

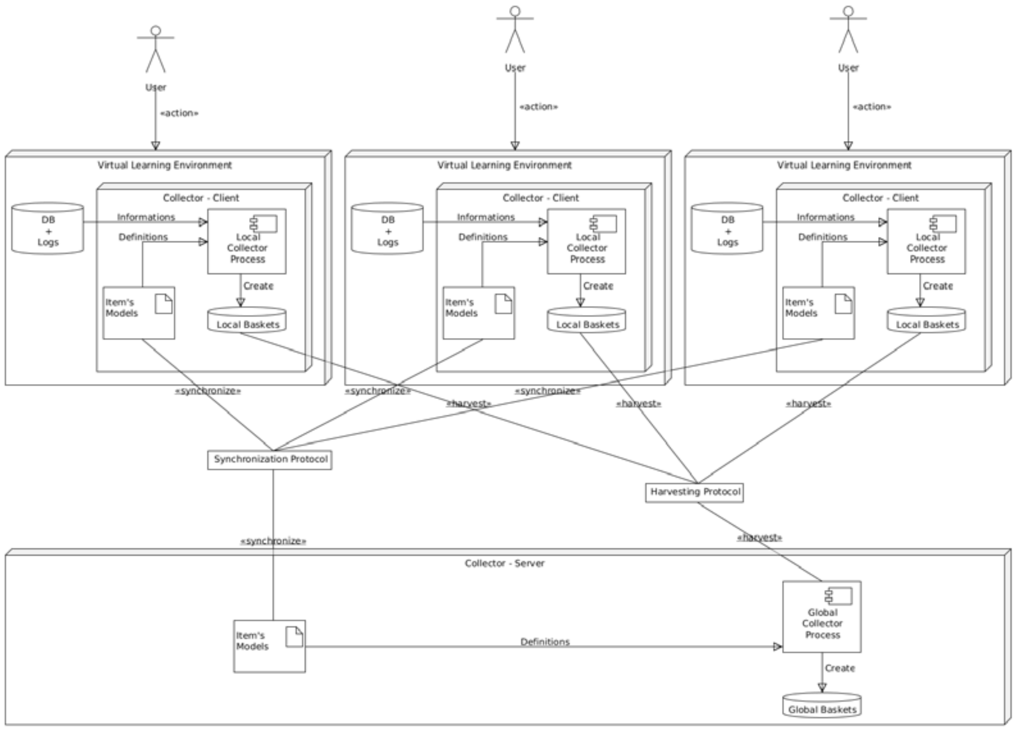


Figure 5 – A Reference Architecture to ETL Process

*3) Local Collector Process*

The local collector process uses definitions stored in the item's models to process data from the virtual environment and to create the local baskets.

*4) Item's Model*

The item's model is a shared repository between all collector-clients and the collector-server. It contains the definitions of items used to create baskets. The repository is synced using a synchronization protocol. The model should differentiate items that can be created by the collector-client and items that can be created by collector-server.

Every collector-client is responsible to implement locally the extraction of items defined in the repository. In this way is possible to deal with environments and storage options through different client implementations. For example, lets say that one of the items to be used to make a basket is named "Total Emails". In environment A this data is inside table 'users', in environment B this data is inside table 'students' and in environment C we need to combine tables 'people' and 'people_emails' to get this data. With three implementations we could solve the data format problem for this item yet maintaining its original sense.

*5) Synchronization Protocol*

This protocol should explain the procedures to create, update, remove and distribute the item's definitions that are saved in the item's models.

*6) Harvesting Protocol*

The harvesting protocol should explain how the local baskets would be harvested by the collector server. It should handle questions about priorities, time windows for harvesting and basket partitions.

*7) Collector Server*

The collector server is responsible for harvesting local baskets from virtual environments and for building global baskets. To do this, it would execute the harvesting protocol.

*8) Global Collector Process*

The global collector process is responsible for organizing the harvesting of local baskets and for carrying out any last needed preprocessing (using the item's model) in the items before creation of global baskets.

## VI. CONCLUSION

The reference model and the reference architecture proposed here can be used not only for implementing solutions that improve traditional ETL process, but also for studying this domain and explain or compare specific implementations.

Since local baskets are calculated independently at each client, we have parallel processing of items and a reduction on time needed to harvest all baskets (a delay in one environment would not imply in a delay in the others environments). Another point is the possibility for clients to carry out an anonymization process while building their baskets.

Because only the items in baskets are transferred to a central place, we might:

- Reduce the probability of sensitive information be transferred
- Save storage needed for mining place
- Save bandwidth between clients and central server

Saving bandwidth adds one more improvement, given that it would increase probability of successful download even through poor connections with the Internet, especially when one or more virtual environments are connected through an intranet.

Finally, systems implementing the architecture proposed here would be easily expanded, given that both clients and server have access to the repository with the item's model. So, always a new item is added to the repository it will be shared between all clients and server, and always a new client is added, it will have a copy of the repository.

## REFERENCES


[1] Baker, R., Isotani, S. and Carvalho, A. Mineração de Dados Educacionais: Oportunidades para o Brasil [Educational Data Mining: Opportunities for Brazil]. *Revista Brasileira de Informática na Educação*, v. 19, n. 2, p. 3–13.

[2] Elias, T. and Lias, T. E. (2011). Learning Analytics : Definitions , Processes and Potential. *Learning Analytics: The Definitions, the Processes, and the Potential*. http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf.

[3] Rajaraman, A. and Ullman, J. D. (2013). Mining of Massive Datasets. *Lecture Notes for Stanford CS345A Web Mining*, v. 67, p. 328.

[4] MacKenzie, C. M., Laskey, K., McCabe, F., Brown, P. F. and Metz, R. (2006). Reference Model for Service Oriented Architecture 1.0. *OASIS Standard*.

[5] Brown, P., Estefan, J. A., Laskey, K., McCabe, F. G. and Thornton, D. (2012). Reference Architecture Foundation for Service Oriented Architecture Version 1.0. *OASIS Standard*.

[6] DeFreitas, K. and Bernard, M. (2014). A Framework for Flexible Educational Data Mining. *Proceedings of the International Conference on Data Mining (DMIN)*, p. 1.

[7] Kechadi, M. T. (2010). The data wave: Data management and mining. *Proceedings of the Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, WETICE*, p. 7–11.

[8] Lepouras, G., Katifori, A., Vassilakis, C., Antoniou, A. and Platis, N. (2014). Towards a learning analytics platform for supporting the educational process. *IISA 2014 - 5th International Conference on Information, Intelligence, Systems and Applications*, n. page 2, p. 246–251.

[9] Krüger, A., Merceron, A. and Wolf, B. (2010). A Data Model to Ease Analysis and Mining of Educational Data. *International Conference on Educational Data Mining (EDM)*, p. 131–140.

[10] Ngo, L. B., Dantuluri, V., Stealey, M., Ahalt, S. and Apon, A. (2012). An architecture for mining and visualization of U.S. higher educational data. *Proceedings of the 9th International Conference on Information Technology, ITNG 2012*, p. 783–789.

[11] Romero, C., Ventura, S. and García, E. (aug 2008). Data Mining in Course Management Systems: Moodle Case Study and Tutorial. *Computers & Education*, v. 51, n. 1, p. 368–384.

[12] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[13] Baker, B. M. (2007). *A conceptual framework for making knowledge actionable through capital formation*. University of Maryland University College.

[14] Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, v. 33, n. 2, p. 163–180.

[15] Shum, S. B. (2012). IITE Policy Brief - Learning Analytics. *IITE Policy Brief*. http://iite.unesco.org/pics/publications/en/files/3214711.pdf.

[16] Butoianu, V. et al. User context and personalized learning: a federation of contextualized attention metadata. Journal Of Universal Computer Science, v. 16, n. 16, p. 2252--2271, 2010.

[17] Wolpers, M. et al. Tracking Actual Usage: The Attention Metadata Approach. Educational Technology and Society, v. 10, n. 3, p. 106–121, 2007.