

DATA SCIENCE, SOLUTION FOR MANAGERIAL DECISION WITHIN BUSINESS MODELS

Marian-Sorin IONESCU^{a}*

^a University Politehnica of Bucharest, Romania

ABSTRACT

The stages to be completed in the business processes approached using the data science concepts and paradigms have a very clearly defined functional structure.

The identification and evaluation by algorithms and automated procedures of the data models, underlying the informational approach, the development of creative processes generating higher added value are all are operational and strategic developments specific to the business models approached.

Understanding the processes generated in their informational integrity, but also in the depth of the functioning mechanisms is a factor that sets the projects for data extraction, transforming them into alternatives of system analysis, minimizing the risk and human decision, possibly erroneous and generating future errors.

Viewed as a whole or the process of data extraction, the final target of the fidding models is fragmented into a set precisely defined under tasks, especially useful in understanding the concepts and paradigms used in approaching information processes and analyzing data in their complexity and dimensionality.

In the business models operational and strategically implemented for the economic organizations the tasks of data extraction and analysis have common general characteristics but also particularities of the main object of economic activity rendered.

KEYWORDS: *business model, data mining, data science, decision.*

1. DESCRIPTION OF BUSINESS ISSUES

The concepts of data extraction and processing have a relatively high degree of ambiguity within the problems encountered in the economic ecosystem, their identification within the business models chosen and transposed operationally by the top management, they are the answer to the real and potential challenges that have arisen.

It is usual in practice to solve the business problem that the challenge arises to be the object of a reformulation process, with a certain prioritization of the met and used data and the design of a solution, possibly optimized or circumscribed by an optimization algorithm, within a broad iterative process of discovery (Russell & Norvig, 2003).

Understanding the business process is in a strong connection with the presentation of the initial formulation, this may require additional data for a complete understanding or allow for improved structures until optimization, in order to achieve this form, a process of successive iterations must be designed, implemented are operational.

Understanding the business process in its entirety, adopting the operational model and designating the appropriate strategic trends that can be developed within well-defined time horizons, requires a

* Corresponding author. E-mail address: sorin_ionescu_07@yahoo.com

unique process of managerial decision-making creativity from the human individuals involved, the decisive manager has at his disposal a set of powerful tools to solve all the challenges that come from the process of data extraction, the fundamental activity in the way of data mining.

In understanding the are operational of the business, the organizational task force must have a relevant degree of perception of the economic reality to conceive the optimal approach or an optimization process for the real challenges but also the intrinsic or extrinsic potentials, what is the final target to be reached, which are the intermediate targets through which the obligatory one must pass, the way of designing and realizing, how the complex process of data extraction is elaborated and implemented.

New concepts and paradigms are transformed into tools that help the decision-making process for business models, quantifying the predicted values is the option of systematic decomposition into exact responsibilities by extracting the data considered by the decision manager as having a degree of importance and relevance.

2. THE RELEVANCE OF THE DATA

The data identified as usable in the business model are those that generate the sustainability of the resolved challenges, the data have limitations but also strengths that generate strategic opportunities, they give as a whole the image of the business model that is the most appropriate operational for the problematic approach.

The transactions, the client portfolio, the market studies and the marketing analyzes cover within the databases induced sets that can have common parts of information and different degrees of the perception of the significant economic information, with a large weight in the elaboration of the operational, strategic decision.

The data used in the organizational business models have a variable cost, from free to significant financial values, the decisive manager is the one who can start a program for acquiring the data considered important for the economic organization or developing algorithms and procedures for identifying and extracting them through his own efforts, the operational current of the developed strategic projects is totally interdependent with the access and extraction of data with a degree required by relevance.

The financial allocation decided on the board of the economic organization, for the extraction, processing and subsequent storage of the data, aims to quantify the costs, the added value offered by them, the profits and the subsequent benefits offered by them, the extraction, processing and storage of data considered relevant for the economic organization are part of a complex process, which is in continuous progress, the solutions appearing having no coherence of the response directions.

A broad application of this innovative paradigm is in the financial banking world, in fraud detection processes, the processes of data exploitation know an extrapolation of their operation in this area so vulnerable, such as fraud detection, monitoring and monitoring in a continuous flow of data implies the understanding and application of all their specific tasks, punctual, in the case of credit card frauds, they are signed on each coding allocated to the client account, the tariffs with a certain suspicion fence are signaled to the economic organization, and subsequently by the card holder, if the card is checked with a certain periodicity the banking activity in the respective account (Provost & Fawcett, 2019).

These innovative paradigms make it possible to fully integrate fraud, receive a subsequent identification tag, the opposite process, crime, victim, is practically the final target of the surveillance technique.

By its financial dimensionality, billions of dollars of annualized damages, frauds that are characteristic of the U.S. health care system, are a huge problem for the federal government, apparently a simple fraud detection challenge, due to the significantly different business models used, this vulnerability has enormous informational complexity, the approach through the process of

identifying and extracting data, identifies healthcare providers, users of invoice issuing systems and even patients as potential fraudsters.

Fraudsters in the American medical system are mathematically assimilated to a subset of regular users, a set of operational implementable rules that defines the structure of the exact charge portfolio.

As a specificity of the data typical of medical applications, they have a low degree of reliability in fraud detection processes, developing a supervised learning process, specific to the algorithms in Artificial Intelligence, specifying frauds with credit card, is insufficient for the complexity of the problem, for this type of challenge, we consider that an unsupervised approach to detecting anomalies and fraudulent processes is more appropriate.

The problems of detecting fraud, with the help of the processes of identification, extraction, processing and storage of data have their own particularity, interpreting them in profusion but also in their quantity, dimensionality, is possible only by understanding the business model used for each economic organization, the priority of data extraction tasks is to justify the paradigms and innovative technologies within this innovative field, the reference is for the data mining tasks.

3. REQUIREMENTS ON THE DATA

Technologies specific to the processes of data preparation are extremely powerful and efficient tools for the elaboration of the managerial decision at the level of economic organization, their use is conditioned by certain requirements on the data used, it is obligatory for the data to be provided in a certain form, a process interactive between the provider and the data user is mandatory initiated, (Lemberger et al., 2019; Provost & Fawcett, 2019).

The preparation of the data that is extracted, studied and used operationally are part of the broader process of data preparation, a good perception of their informational content allowing also an activity of manipulation in order to obtain the best, even optimal results.

In operational practice the data are transformed into tabular format, a process of eliminating those considered irrelevant or deducting missing values and transforming them into different types is desired, we also identify a process of extracting data into symbolic data, sorted by sets of categories or only values numerical in a mathematical symbolism.

In the operational transposition of the numerical data they are either normalized or scaled, the possibility of comparison between them being therefore accessible, the specific processes of identification, extraction, processing and storage of data offering algorithms, specific technical procedures in this regard.

A particularly dangerous process for the operations in which the data is involved is that of "data leakage", it generates a situation in which the relevant historical economic data provided information on the targeted, intermediate and final targets, but becoming unavailable in the decision-making mechanisms management.

4. THE MODEL USED FOR DATA PROCESSING

In the study and application of data technology in the economic and business processes, the model creation stage is identified with the part in which all the data extraction techniques are gathered, the study, the understanding and the processing thereof.

In the modeling process, the first step to be taken is the development of an evaluation process, it is the one that evaluates the final results of the data evaluation, the degree of validity and trust offered by them, a rigorous control component is the determining factor in their choice.

The models developed for the chosen data must offer reliable, predictable, which are subject to the principles and rules, previously defined, in the evaluation stage, is verified and tested if it serves

correctly to the satisfaction of the initial objectives of the economic processes, of the desired business performance target.

The stakeholders of the analyzed economic organization are directly interested in evaluating the data processes both from a quantitative and qualitative analysis point of view, these directly influencing directly the strategic operational decisions that generate the added value of the economic organization's activity and its performance, stakeholders are interested in the applicability of business models in the smooth running of the processes that require data processing and exploitation, these models must reach a degree of satisfaction required.

The degree of understanding of the model used for working with the data, the simplification of the mathematical formulation in which it is presented and accessed by the decision manager comes within the responsibility of the scientist and represents an offer as attractive as possible, with the advantages deriving from here, of the development for the problematics. business of using algorithms and procedures for identifying, extracting, processing and storing data, interpreting them in accordance with the objectives of the economic organization benefiting from this new type of approach.

As a particular feature of the use of data extraction, study, processing and storage processes in economic issues, we mention the importance of creating and developing a more comprehensive framework, because the data on the actual and potential performances of the business models are very difficult to quantify and foresee strategically (Provost & Fawcett, 2019).

Assuming the risks, real and potential, especially those coming from the investment area and leading to their own financial efforts or external indebtedness, in certain time horizons and with mathematically quantified profit hopes, led the economic organizations to include in the decision board a complex but flexible teams of data science experts, they build surrounding information and economic eco-environments, in which the data most accurately reflect the organizational operational activity, but also serve as an argument in the elaboration of strategic forecasts as close to the economic reality, partners, competition, thus following a minimization, until to optimize the risks arising from the decision of the top organizational management.

Specific to the implementation of business models at the level of economic organizations are the emergence of systems with a high degree of hazard, random evolutions, therefore the extension of the evaluation in the economic eco-system for the developing environment is mandatory.

In the operational practice of the groups working in data science, laboratory tests are initially carried out, their results and conclusions are subsequently extrapolated and applied to real evolutions, for a problem and targets to be reached by the economic organization in its activity, the economic systems chosen in randomly, however, it always maintains a stable group of control clients, these experimental constructions are those that favor a good understanding of the complex activity of working with the data, the technologies developed later not fully covering the possible cases provided within the global economic eco-system.

The managerial decision systems structured on the results generated and interpreted from the data science provide evaluations that ensure the top organizational management that the external changes within the economic environment do not negatively influence the operational decisions transposed in implementation, but also the strategic ones that coordinate the future activity of the studied economic organization, the business model implemented, showing continued trust for all stakeholders, of the risks identified in this approach, we mention all types of fraud and spam, which affect in certain degrees the entire activity of the economic organization.

The exits offered to the top organizational management are decisive in an overwhelming degree of the totality of the input data, which are subject to a continuous process of change in depth but also quantitative, with all the subsequent influences arising from here, without the operational entity of the scientific experts, data within the economic organization aims to seize them, so without any reaction to the modification of the processing and subsequent interpretation of the data.

5. USE OF DATA MINING PROCESS RESULTS

The quantification of the final activity of any type of economic organization is done by evaluating the profitability of the financial allocations, and therefore of the return on investments, data mining techniques help in this regard in creating a predictive model in which an information system is already operational, with an existing economic history or a business model chosen by the top management.

The models for predicting the evolution of the data within the specific processes of the economic organizations are an integral part of the performance economy with the chosen business model, the latter development becomes intrinsically linked to the development of specific procedures and algorithms of data science, which foresees all the situations with which it is possible to confront the studied economic entity, but especially how to solve them, data science, having in the subsidiary or the technologies performed by the extraction, processing and storage of data, are already used in directing online marketing and promotion campaigns for products or sets of products, the process of implementation and operational transposition is an automatic one, also the models business activities involving production activities are tested and subject to the desired managerial and economic performance criteria (Provost & Fawcett, 2019).

The data exploitation system is more important than an approach to the problem faced by the economic organizations through accessing them in a trivial way, due to the change of the economic environment much faster than the adaptability of the data mining team intrinsic to the studied organization, the same observation. is made for the processes of detection and identification of fraud and unauthorized access to information, so we affirm that the chosen business model also has a component of adaptation and modeling of the team of people specialized in data science.

Positioning ourselves in this paradigm, we consider that the first approach of the team specialized in data science within the economic organization is the approach of the extraction, analysis, processing and storage of data for the processes of manufacture, production, the identification of real or potential anomalies must be anticipated, for a functioning in safe conditions of the process followed in its entirety.

Additional operations that require greater investment and financial efforts are needed for the implementation, operation and elaboration of a strategic forecast, related to the new production business model, the immediate advantages deriving from this are the increased speed of work, establishing adaptability connections with already existing systems, existing and verified from the point of view of the economic performances achieved (Russell & Norvig, 2003; Lemberger et al., 2019; Provost & Fawcett, 2019).

The systems for extracting, analyzing, processing and storing data for the specific problems of the production processes, require the re codification of the model, result in higher working speeds and ensure compatibility with the systems already implemented at the level of the economic organization, the effort of the economic organization for the data science team is one of high financial nature, due to the huge amount of information to be processed, the quality of the results pursued, but also the speed of work that has to be correlated with the answers to the challenges that the economic entity faces beneficiary of the final information.

The data science team most often has a variable degree of responsibility in developing, creating and developing a working prototype for the organizational business model that is operationally tested and offers arguments for justifying the integrity of top management decisions, it is particularly important that the organizational decision maker, possibly its CEO, understand clearly the processes regarding the data, they offer a broad, possibly exhaustive perspective, of the issues that need to be solved, of understanding the business complexity of the possible successes to be achieved but and of the dangers and vulnerabilities with which it is faced, operationally or strategically in a time perspective, precisely identifiable.

A series of successive iterations lead directly to the improvement of the solution, economic experiences already passed, the way of thinking the business model, the data relevant for the present and future process, the intermediate objectives to be achieved, the final objective of the entire activity, the managerial and the strategic trends of consolidation and development of the economy entity and even new projects are their direct consequences.

The importance of the evaluation stage consists in cataloging the results obtained for the subsequent implementation, defining the problems and the proposed business model it is possible to undergo adjustments and adaptations, these possibilities are a direct consequence of understanding the functioning of the business in its entirety and complexity.

It should be mentioned that there is the ease of returning each time to the previous stage, the process having affinities between successive stages, the projects developed by the top management decisive for the economic organization have a high degree of flexibility and adaptability, so that new developments in the processes working with the data contributes and justifies the revision and improvement that tends to the optimum of the steps previously taken.

6. SOFTWARE SKILLS VERSUS ANALYTICS SKILLS

Conceptually belonging to the engineering sciences, the projects specific to the data science are translated and taken over by the economic organizations as tools that improve the managerial decision in search of superior business performance.

This process of conceptual translation and innovative paradigms, has required the emergence of specialized software departments, experts in data science and business analytics management, the future pertaining to intelligent business, with decisions of top operational and strategic organizational management, rigorously substantiated and mathematically argued, generators of superior economic performance, quantified in financial results that offers satisfaction to the whole set of stakeholders.

The software, (Provost & Fawcett, 2019), used by the top management decision maker is the one who manages the projects regarding the extraction, analysis, processing and storage of data, following the operational structural form CRISP (Cross Industry Standard Process for Data Mining), from this diagram it is evident that the hysteria is practically a procedural rule, solving the given problem requiring more process cycle runs, developing a similar development cycle, there is the possibility of extrapolating this approach paradigm to the analysis projects.

The development and application of the CRISP cycle in economic issues is based on exploration and strategic forecasting, the results generated from one stage to another are those that define the problematic accuracy, as many stages successfully completed giving the managerial decision an exhaustive image of the positioning the economic organization within the macro-business environment and the challenges to be answered.

The strict approach from the point of view of the software used of the economic problems encountered is reduced to the ability of the data scientist team to design and write codes of relevant quality, in a certain quantity or bugs tickets closed.

The analytical approach of the business processes, prioritizes the stages to be followed in the processes given by science as follows: the correct formulation of the problem faced by the economic organization, the elaboration of prototypes in response rate regime, the issuance of reasonable working hypotheses in front of some problems with high degree of uncertainty and ambiguity in formulations, the capacity of experimentation that leads to quantifiable results, subsequently subjected to extensive analysis procedures, the conclusions drawn from here (figure 1).

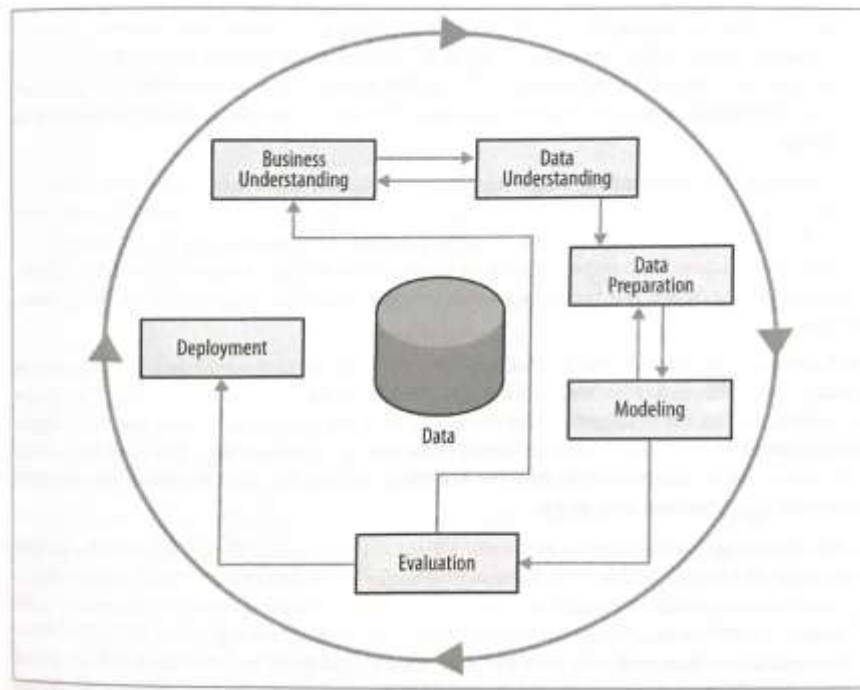


Figure 1. The stages to be followed in the processes

It is particularly important to know and understand the data analysis technologies, the facilities offered to achieve the decision objectives and the managerial practice, the advantages that are obtained within the eco-economic system over the competition, based on them the development of fast and highly efficient sustainable strategies (Provost & Fawcett, 2019).

We consider that according to the focus on the automatic search for knowledge, the requested models, the regular trends of data sets, they are identifiable as used in the management decision within the economic organizations the following groups of relevant analytical technologies:

- (i) Statistics;
- (ii) Data base querying;
- (iii) Data Warehousing;
- (iv) Regression analysis;
- (v) Machine learning & Data mining;
- (vi) Data Mining or KDD (Knowledge Discovery & Data Mining);

7. USE OF DATA SCIENCE TECHNIQUES FUNDAMENTAL QUESTIONS

In the analytical analysis of business, economic organizations, (Geffner & Bonet, 2013), regardless of the main field of activity, must consider and seek answers for a set of questions, considered relevant:

- (i) Identify the clients with the highest degree of profitability, those who bring the highest income to the economic organization;

Observation: the question requires the definition of profitability, standard query tools, the results obtained are sorted by the cumulative sum of the transaction that defines the profitability indicator;

- (ii) The difference between the most profitable customers and the average customers is followed
Observation: It is a hypothetical study direction with logical value of the conjecture, the processes of statistical analysis are those that quantify the probabilities that the coefficient of difference that appears after market analysis is correct;

- (iii) Identify and describe the category of profitable consumers.

Observation: The typology of the profitable consumer is described and possibly extracted from databases resulting from market studies and analyzes, using appropriate techniques such as database querying, continuing the process, the decisive management has a clear presentation of the differences between the profitable and the unprofitable clients;

(iv) What financial value must a new client assign to the products of the economic organization so that he / she is listed as a profitable client.

Observation: The problem can be solved with the help of data mining technology that examines its economic and financial history, generating predictive models for profitability, efficient allocation;

8. MAPREDUCE ALGORITHM, MODERN APPLICATION OF DATA SCIENCE IN MODERN BUSINESS

In the following we present a specific application to work with data science in business activities, called Map Reduce, this represents an intelligent structure associated with the generation and processing of big data structures, frequently encountered in economic applications, the chosen algorithm is a parallel type distributed in a cluster.

To begin with, we proceeded to a formal description of this specific data science approach, the subsequent focus is on concrete, operational examples within certain business models implemented. MapReduce, (Provost & Fawcett, 2019) is an algorithm developed by Google engineers as a solution with a high degree of generality of problems such as those encountered in data science.

In approaching data science for business models through MapReduce, the data is structured as a list of value pairs, similar to those stored in ECV type formulas (key value stores, persistent value pairs on a disk) or BDOD (ECV in which the values are semi-structured documents), the key value is an informational structure, a "blob", without a priori structure.

Applying MapReduce in the problematic data science for business models, we distinguish the following structure:

(i) At the beginning of the approach, a map is developed with the help of "mappers" for all the accessed data records, each call (invocation) implicitly generating a list of intermediate values of the calculation that must ensure a complete list.

(ii) In the second phase, after the operational tasks of the "mappers" have been achieved, the functions of type "reducers" executed by specialized structures, "reducers" are introduced, these agree in continuous flow the associated intermediate values using the same key.

Each entry in the list corresponds to the sale of a portfolio of articles, the only data considered in this approach as having a significant degree of relevance are the names of the articles, the price and the quantity sold there, the treatment of this typical Data Science problem with the help of MapReduce is to provide top organizational management with a comprehensive picture of total sales and number of products sold.

In a first approach, the solution of such a problem is reduced to a list of additions, from the point of view of the concrete are operational the difficulty is not a conceptual one but it is that of gathering the data positioned on a large number of servers.

The task of the "map" function is to extract for each sale, the list of products associated with the related price and the number of them effectively marketed, after a grouping of the products according to specific characteristics, such as, name, price, managed within a framework, the function "reduces" "is the one that calculates the two amounts of interest.

We propose in the following two variants of pseudo-code for the two operational functions, "map" and "reduce" (Provost & Fawcett, 2019).

```
// Function "map" read "command" and in the extract  
// The desired data to be aggregated, knowing the list  
// The quantities and the value of the products sold  
map(int orderID, Order order)
```



```
list intermediateKeyValueList;  
for each item in order  
intermediateKey := item.getItemKey( );  
value := item.getAmountAndQuantity( );  
intermediateKeyValueList.append(intermediateKey, value);  
return intermediateKeyValueList;  
// The "reduce" function consumes a list of intermediate values  
// associated with the same key, c. a – d. the same article, and we calculate  
// the sum of the sales and the quantities of products on this list  
reduce (int intermediateKey, ListintermediateListOfValues)  
totalAmount := 0;  
totalQuantity := 0;  
for each value in intermediateListOfValues  
totalAmount := totalAmount + value.getAmount( );  
totalQuantity := totalQuantity + value.getQuantity( );  
return (total Amount, totalQuantity):
```

MapReduce is possible to be presented much more succinctly but with a much higher degree of abstraction:

```
map : (k1, v1) → list(k2, v2)  
reduce : (k2, list(v2)) → aggregate(list(v2))
```

In strategic operational practice for business models there are no guarantees that the correct treatment through the concepts and paradigms of data science must be done in the previous way, in certain concrete situations this approach is practically impossible, in others internal resources must be allocated to the economic organization for a constant effort of conception and translation, under a non-MapReduce code, this is the price to be paid for the automatic implementation of parallelism, it is restricted to a particular set of treatments.

We make the observation that writing algorithms in MapReduce formalism is a difficult and delicate task, which requires highly qualified human resources and high financial allocations, in order to make this type of approach more flexible to all the demands and challenges received from the business environment, most often, the approach for complex tasks is through the development of a parallelism, the solution is made either by setting up a task force with specific tasks well defined, or through a chain of jobs with less complex tasks, whose mix defines the final solution.

Addressing the problem of business models through specific concepts and paradigms MapReduce is a mimicry similar to the one based on the NoSQL approach, the price paid for the achieved economic performance is quantified in the complexity of the lines of code that must be written for the "map" and "reduce" functions, which must be flexible and reliable enough to effectively respond to all the challenges arising from the business environment.

The advantages offered to this type of approach to the problem encountered are generated by the availability of SQL skills and the power generated by programming through MapReduce, the mix generated by these two software facilities, resulting in the managerial decision obtained.

9. MAPREDUCE, APPLICATIONS FOR BUSINESS MODELS

Within the business models implemented operationally, with the possibility of developing the strategic trends, we consider that an eloquent example in this sense comes from the area of statistical analysis, the MapReduce algorithm, using the Google search engine, the identification of keywords, the presentation of the resulting statistics from here.

For the text application, in the first phase we proceeded to a classification of the text, as follows: long text, more than ten letters, medium text, between five and nine letters, short text, less than five

letters, the operational algorithms provide that before a transfer of all files to mappers, a division into smaller informational sequences is required, each subdivision being assigned to a mapper (Provost & Fawcett, 2019).

Before transferring the list of files to mappers, it is the framework that develops a process of dividing them into much smaller batches, each subdivision thus generated is subsequently assigned to a separate mapper, in the case of dimensionality implication of thousands of mappers, each of these computer entities directly handles several thousand files involved in the studied process.

We make the working hypothesis, to simplify that each individual file is short enough to be handled with a single map-type operation, the map function is the one that gets key input value pairs, in the case of working with documents, the key is the name under which it is identified, the identity of the document, the content of which is the one that defines its value.

Map-type functions are those that count and order the three categories of words, classified by the number of letters, each call of a map implicitly provides a list of words associated with their occurrence in the document studied, at the end of this operational step is obtained a first list of key values, it is the framework that merges these intermediate lists, sorts them and splits them afterwards after new words are created new lots, each of which is assigned to a reducer.

The reduce function has at its entry a list of intermediate values associated with the same key and it is the one that calculates the desired aggregate value, each call to a reduction leads to the calculation of partial amounts for the analyzed words.

10. OPERATIONAL APPLICATION, THE COMBINATION BETWEEN TWO LARGE TABLES

Used quite frequently in modern business models, oriented towards creation of higher added value and economic performance, the combination of two large dimensional tables, the presentation and implementation of a relational model suitable for this type of applications, is an operation that requires allocation of significant financial and human resources for the economic organization studied, in order to avoid the difficulties and vulnerabilities generated by the Big Data context, operational practices are implemented models of data management and information flows, (Nilsson, 2009), offered by NoSQL.

The data positioned in the same registration are collocated, meaningful in this respect are the informational facilities offered by the BDOC and BDOD approaches, the redundancy in these applications must be increased, the model having reduced flexibility, a use process for the specificities of the chosen business model is mandatory developed.

The process of combining the two tables A and B is practically the concatenation of lists of records, defined by two keys and a single final table, the searched result is a new relationship in which the records are obtained as a concatenation of all the couples formed from a record A with a record B in which the keys are the same, that is to say coincident.

The fundamental idea in obtaining this result using the facilities offered by MapReduce is to perform a relatively simple operation before applying the map function, knowing that the records from the two tables should be grouped into a single list of the later lengths of records, for each record, not only the content but also the name of the table of origin must be highlighted.

This task is possible to be carried out only by means of the map function, for each registration that enters, we highlight a key value pair, in which the key is the one at the junction and the value is the content of the registration is the name of the original table.

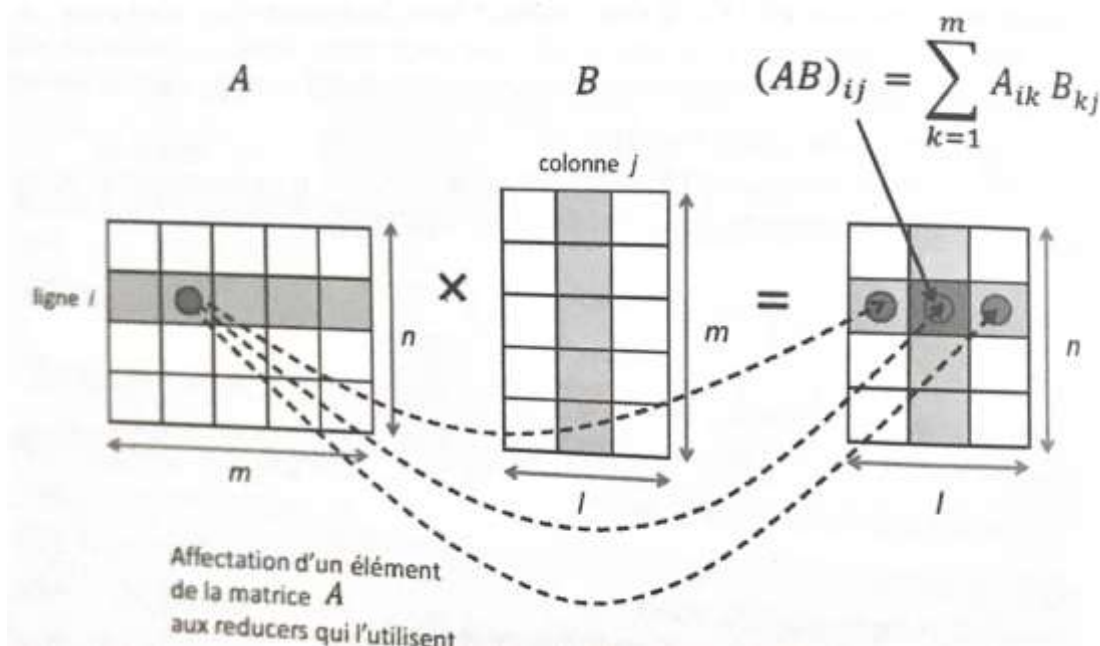
It is the framework that finally brings together the intermediate lists of registrations that lead to a sharing of the same junction key, each of these lists has allocated a reduction for the operations that need to be performed, the call to the operation reduces the production of the junction, thus all the possible concatenation operations are performed for a record in table A with a record in table B with the same key value at the moment.

11. HIGHLIGHTING THE PRODUCT OF TWO NULL TABLES

In the data science approach to large-scale business models, the situation is often encountered in which data tables contain thousands or even tens of thousands of attributes of which few are known, a typical situation for an economic organization that becomes an actor in an environment new business.

In this case, statistical treatment involves the manipulation of matrices with most elements having zero value except a few non-zero values, this type of matrix is generically referred to in the specialty literature as empty or null matrix (Bader et al., 2003).

The MapReduce computing facility shows its usefulness in the calculation of the product of two matrices of this type, we use the formalism of representing the null matrix, A with the help of a list of couples of the type, $\{i, j, a_{ij}\}$ where, a_{ij} is the value of the element present on the line, i and the column, j all the elements for which are omitted from this list, a_{ij} is null, a scalar product is made between the lines in A and the columns in B , expressible by a convolution, amounts of products. In order to parallelize the calculation with the help of the MapReduce facility, for each of the $(n \times l)$ elements in the matrix AB , use a reducer that calculates the sum of the requested products, which has practically unlimited access to all the elements in the matrices A and B that it needs.



This is practically the task of mappers to provide them, more explicitly the calculation using the two map functions and reduces (Geffner & Bonet, 2013):

- (i) **Reduce**, we consider that the reduce function calculates the element (i, j) of the product AB , this acting on the two lists of values, in one part the list of elements a_{ik} , for each $k = 1, m$, and the list of elements b_{kj} , for $k = 1, m$, these two lists associated with (i, j) considered as an intermediate key, are two lists of intermediate values, which the framework must deliver to the reduce function;
- (ii) **Map**, this function, must consequently deliver lists to reducers using the pair (i, j) as an intermediate key, each mapper goes through the set of elements of the arrays to which it was

assigned, the action of the map function is the one that looks like if an input element belongs to matrix A or matrix B ;

In the implicit development of the presented approach, we present some other facilities offer by the concepts and paradigms of specific data science tools, such as the MapReduce algorithm.

12. THE HADOOP FRAMEWORK

The use of the MapReduce algorithm in the case of the parallelism problem under different formulations depending on the business model chosen and implemented by the type of approach of the problem encountered, is sometimes not sufficient to achieve the desired performance, however, the map and reduce functions show their limits in the face of the calculation required for a parallelism with a high degree of complexity.

However, modern information technology offers the possibility of using a performance framework such as Hadoop which executes a very large range of tasks with a reasonable level, low complexity, in a precise order, with an optimal allocation of resources intrinsic to the economic organization.

The phenomena of coupling between abstracting and effective operational implementation are repeatable until an optimal form of the chosen model is reached, the usefulness of the managerial decision process.

The implementation and are operational of MapReduce is totally dependent on the business ecosystem in which it is located, in a specific Big Data context, the best solution is to use a server cluster.

Hadoop's operation is totally dependent on the most used framework to solve the challenges of MapReduce organizational business models, we identify the client application executable in Hadoop, as called job MapReduce, each this type of job is broken down into tasks that are structured on two levels map tasks and tasks reduce (Provost & Fawcett, 2019).

The set of jobs being executed by the framework is coordinated and supervised by a specialized processor called job tracker, it has the role of planning the execution of the tasks assigned to the subordinate processes, these are called task trackers.

We propose within the problem of business models, for a good delimitation, isolation even of the execution of the map tasks and reduce, in order to avoid all the possible shortcomings generated by the server used, that the tasktracker perform each map task and reduce within a Java Virtual Machine (Russell & Norvig, 2003; Provost & Fawcett, 2019).

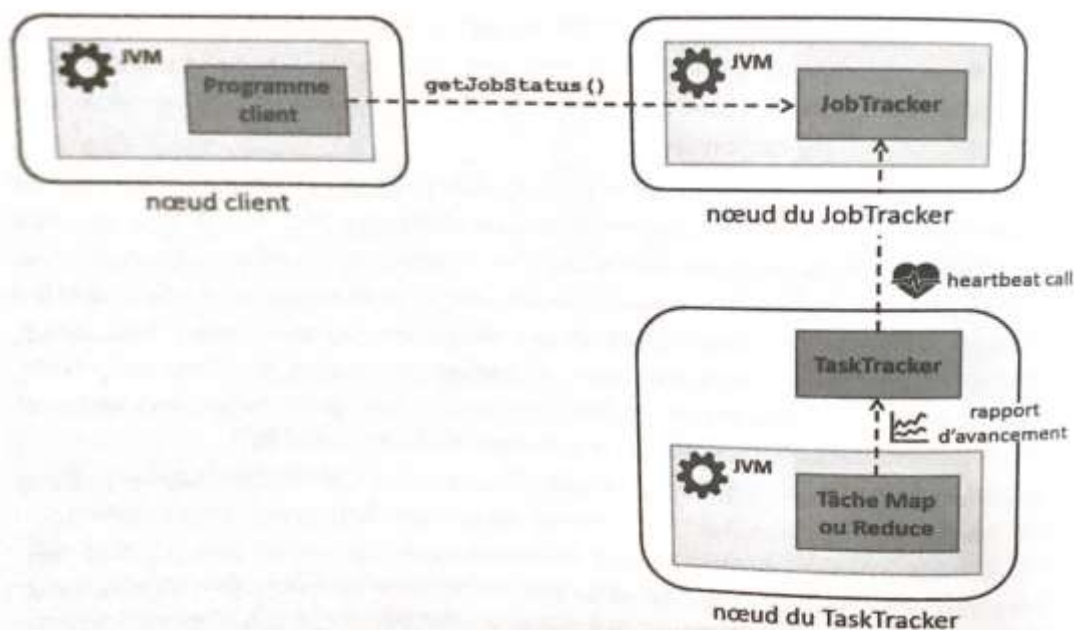


Figure 3.

13. CONCLUSIONS

In the specific applications of the business models approached with the help of the concepts and paradigms of data science, we consider that MapReduce is a real aid for the elaboration of managerial and strategic decisions.

This logical pattern allows the parallelization of the treatment for large volumes of data under the conditions of use for the necessary calculations of two map functions and reduces, which are active and operate on lists of key values, specific to the adopted business models.

The platform of execution of the MapReduce IT facility is standardized by the Hadoop application, transferred to the problem of organizational business models, due to the facilities for homogeneous clusters of the machines with optimal costs, it manages the essential of the low level problem regarding the distribution of tasks, the extraction, analysis and storage of data.

Hadoop is very well suited for batch processing and processing of large volumes of data, a very common case in the modern economic eco-system, also having the ability to develop real-time machine-type analysis, the only vulnerabilities of the MapReduce logic pattern are those reported by interactive treatments for business models.

REFERENCES

- Russell, S., & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*, 2nd Edition Prentice Hall.
- Lemberger, P., Batty, M., Morel, M., & Raffaelli, J. L. (2019). *Big Data et Machine Learning*, 3e edition Dunod.
- Provost, F., & Fawcett, T. (2019). *Data Science for Business*. O'Reilly.
- Nilsson, N. J. (2009). *The Quest for Artificial Intelligence*. Cambridge University Press.
- Bader, F., Calvanese, D., McGuinness, D.L., Nardi, D., & Pantel-Schneider, P.F. (2003). *The Description Logic Handbook: Theory, Implementation, Applications*, Cambridge University Press, U.K.
- OWL Web Ontology Language Reference (2005). World Wide Web Consortium.
- Geffner, H. & Bonet, B. (2013). *A Concise Introduction to Models and Methods for Automated Planning, Synthesis Lecture on Artificial Intelligence and Machine Learning*, Morgan & Claypool.