Big Data Approach and Dimensions for Educational Industry

Dorina Kabakchieva[*]

Kamelia Stefanova[**]

Summary:

The dynamic environment we live in now is not driven by any specific business or product – the real engine for change are the customers. Businesses are using constantly growing amounts of information. In this respect they need adequate storage resources and innovative methods and technologies for managing the data and information. The phenomenon Big Data appears to describe the next generation of new data sources that will support the next level of information management and analytics penetration in business, academia and government. These solutions will radically change the way organizations work and manage their business. Big Data approach appears because data processing and analytical needs today are not possible within the traditional environment.

Some of the Big Data source characteristics refer to the new types of data sources, the automatic generation by a machine and the absence of standardised and concretely designed way of streaming data.

Big Data approach grows with generated bulks of data, but it is not related only to data volume. Big Data definitions are generally accepted to be explained with the substance of the three Vs - velocity, variety, and volume. This paper presents the most exhaustive aspects of Big Data and presents additional four Vs - veracity, variability, visualization, and value.

It is hard to identify a unique reason for the explosive growth of data. The paper focuses on the four main directions of change in the world of data - business model transformation, globalization, personalization of services, and new sources of data.

The role of Big Data is changing the whole world. Big Data could give unique opportunities for the educational sector development as well, and improvements could scale – from a single school, to governmental directions and satisfaction of the labour market. Big Data could support changes in education that would revolutionize all educational aspects - the way students learn, teachers teach, administrators manage, employers select and develop staff.

Key words: Big Data, Business Intelligence Systems

JEL Classification: C63 ; I21

1. Introduction

The greatest challenge for companies for development and growth today is to identify the appropriate transformation of their business from product oriented to service oriented approach. The dynamic environment we live in now is not driven on by any specific business or product – the real engine for change are the customers. Customers need to be very well informed about the products and related services, about the company details and its customers' overall satisfaction.

[*] Chief Assistant Professor, Department Information Technologies and Communications, Faculty Applied Informatics and Statistics, UNWE, dkabakchieva@unwe.bg

[†*] Professor, Department Information Technologies and Communications, Faculty Applied Informatics and Statistics, UNWE, kstefanova@unwe.bg

Businesses are using constantly growing amounts of information. In this respect they need adequate storage resources and innovative methods and technologies for managing the data and information. The reasons for the need of massive amounts of information vary from business to business. Some reasons relate to the regulation requirements, others – to the need of preserving transactions, others – to the need of exhaustive analysis.

The phenomenon Big Data appears to describe the next generation of new data sources that will support the next level of information management and analytics penetration in business, academia and government. These solutions will radically change the way organizations work and manage their business. The analytics performed by Big Data enables decision-making process to be better informed and even could be very differently transformed from the way it is today. Big Data approach appears because data processing and analytical needs today are not possible within the traditional environment. The old way of performing analysis is not effective anymore. Big Data is coming with new requirements for the new data sources and their management in organisations. Additional scale of Big Data necessitates utilizing specific processes, tools, and technologies. The need for change to the next level advanced analytics is already there.

Advanced analytics has gained enough experience and figures out large impact on the expanding growth of new and powerful data sources. Each industry today has at least one completely new data source coming online.

Big Data expands dynamically everywhere and requires appropriate measures to be introduced accordingly in order organisations to get competitive advantage. Missing the opportunity to establish Big Data on time will put organizations at risk and cause competitive damages. Organizations should consistently follow identifying, capturing and analyzing the new data sources in order to get the right insights. Analytical processes should be thoroughly changed in order to satisfy the management needs. It is inevitable to invest much time and efforts to incorporate Big Data alongside, but future development without Big Data is almost impossible.

2. Big Data Substance and Main Characteristics

The main purpose for Big Data appearance is that existing hardware and software could not organise and process the vast amounts of different types of data being created and coming from different sources at so high speed. Big Data appears to be very complex and dynamic to be stored, processed, managed, and analysed with traditional data tools.

Big Data approach comes with respective hardware, instruments and methods developed in order to transform all available data into meaningful information. The main reason for providing new insights of the information is to improve the decision-making process, increase the efficiency, and enhance the competitiveness. There are many aspects of the Big Data wave, and this wave comes to impact all types of industries and all types of organisations.

It is difficult to identify consensus in publications about the definition of Big Data, but some consistent aspects could be outlined. The first attempts for defining Big Data appear in 2011. A definition from Gartner says, "Big Data exceeds the reach of commonly used hardware environments and software tools to capture, manage, and process it within a tolerable elapsed time for its user population." (Adrian, 2013). Another good definition is provided by the McKinsey Global Institute: "Big Data refers to data sets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse" (Manyika et al., 2011).

The main notion that these definitions refer to is that the scope and characteristics of Big Data wave will change over time as technology advances. What was considered

to be Big Data in the past is not Big Data today and will not be Big Data tomorrow. This aspect of the definitions of Big Data is one that is hard to be clearly taken. Another aspect that should be extracted from the definitions is that what constitutes Big Data can vary from industry to industry, even from organization to organization. Additionally, the methods, tools and technologies used could vary greatly in solutions and functionalities.

2.1. Big Data is not reflecting Only on Volume

Big Data approach grows with generated bulks of data, but it is not related only to data volume. Some of important Big Data characteristics are also velocity, complexity, and variety. Gartner Group describes that "big" data refers to several characteristics (Prentice, 2011). These aspects include increased volume, increased velocity, increased variety, and extra complexity. This means that the wave of Big Data is getting a lot of data that is coming very fast, in complex formats, and from a variety of sources.

Big Data creates challenges for organizations and their analytical processes, techniques, and intelligence systems. Big Data cannot be managed with traditional approaches and requires additional analytical processes, methods, and techniques in order to be effectively organized, analysed and appropriately used within the decision making.

Organizations are forced to clearly understand that Big Data is becoming critically important. The analyses that organizations perform based on Big Data philosophy should be combined with management actions taken to improve business and competitiveness.

Capturing a big source of data does not itself bring any additional value to management. Data that is collected but not analysed and used creates even additional expenditures and risks. The volume of data is irrelevant without being put into context and adequate use. The power of Big Data is in the effect of analysis and actions taken as a result, the changes introduced by the business.

It is not correct to believe that just because Big Data has high volume, velocity, and variety, it is more important than other data. It should be clearly stated that the importance of Big Data comes from the possibility to better inform the decision making and perform the right actions on time.

Ninety percent of all the data that is recorded in human history has been created in the last two years, but the Big Data need has been identified long before (Adrian, 2013).

2.2. Main Historical Big Data Aspects

Roger Mougalas from O'Reilly Media introduced the term "Big Data" in 2005, and the company created the term Web 2.0 during the next year (Magoulas, 2010). His main idea was to address the term Big Data and to focus on the large set of data that is almost impossible to be managed and processed using traditional Business Intelligence tools.

Some companies reacted very quickly with innovative solutions – Google developed MapReduce and Yahoo! created Hadoop on top of it in 2006 (Woods, 2011), (IBM website). Many organizations worldwide use the open-source Hadoop to organise massive data sets today.

With expansion of social network sites and Web 2.0 functionality development, huge amount of data has been created daily. Logically, the need for respective processing technology has been identified and different Big Data projects were initiated.

Eric Schmidt, Google Executive Chairman, stated in 2010 that "every two days now we create as much information as we did from the dawn of civilization up until 2003… That's something like five Exabytes of data…" (Woods, 2011).

The need for organising and analysing data has created new initiatives for establishing software companies that will gain knowledge and experience in supporting the Big Data phenomena. Their

main purpose is to help organizations manage and understand this explosion of Big Data. The Big Data revolution is still ahead and a lot will change in the coming years (Thiele, 2012). Big Data will completely change organizations and societies around the world. New terms have been created to describe the amount of data that is expected to be created in coming years (Figure 1). It is expected that the amount of data currently available will double every two years worldwide (McGaughey, 2011).

Megabytes	10^6
Gigabytes	10^9
Terabytes	10^{12}
Petabytes	10^{15}
Exabytes	10^{18}
Zettabytes	10^{21}
Yottabytes	10^{24}
Brontobytes	10^{27}

Fig. 1: From Megabytes to Brontobytes

2.3. Big Data Sources Unique Characteristics

It is very important to figure out the ways Big Data is different from traditional data sources in order to take the right approaches in managing it. There are many characteristics typical only for Big Data, but this is not required all of them to be met in every Big Data source. Most of Big Data sources have only some of these characteristics.

- Big Data is typically automatically generated by a machine. Traditional data sources that are used to be created always require a person to be involved and responsible for the data generation. Taking into consideration different transaction types – retail, banking, telephone call records, product logistics, invoice payments, etc. All of these data generation processes involve a person who is doing actions being supported by a data record creation. It is not the same with Big Data. Most of Big Data sources are generated without any human intervention at all. For example, a sensor embedded in an engine produces data even if nobody asks for it.
- Big Data appears to be an entirely new source of data. It is not simply an extended data collection. Many customers execute transactions - banking or retail online. In general, the transactions they fulfil are not different from those they have done traditionally. The difference is that they simply execute the transactions through a new channel. However, in fact capturing browsing behaviours while customers execute transactions creates fundamentally new data.
- Unfortunately, many Big Data sources are not standardised and concretely designed. For example text streams from social media sites. People use their own way of expressing their opinions - users do not follow any standards of grammar, or sentence ordering, or vocabulary. It is very difficult to process and analyse people's posting. In general, traditional data sources were specifically designed to be user-friendly. Systems used to capture transactions, provide data in a clean, formatted templates and as a result, data could be easily loaded and processed. Traditional data sources were very tightly defined. Every bit of data had a high level of value. With the almost negligible cost of storage space today, Big Data sources are not always tightly defined and typically capture everything that could be of use. This is what creates the challenges for Big Data - to cope with messy, junk-filled data when processing and analysing it.

3. Big Data 7Vs

Big Data definitions are generally accepted to be explained with the substance of the three Vs - velocity, variety, and volume. During the past few years, researchers have identified additional Big Data characteristics that could be expressed by additional Vs - veracity, variability, visualization, and value. All 7Vs, their measures and main substance are presented on Figure 2.

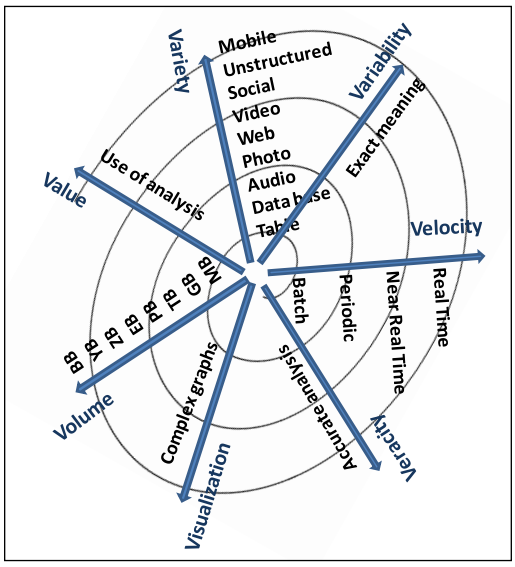

Fig. 2: Big Data Characteristics

Volume

The data volume growth is enormous, and the largest contributor to the ever-expanding digital world is the "Internet of Things" - with sensors and devices creating data all over the world in every second. At the rate data is now being created, the amount will double every two years (Taylor, 2011).

Organisations could not even imagine what problems would have caused such Big Data today without introducing appropriate processes and technologies. Nowadays, with decreasing storage costs, better storage solutions, such as Hadoop, and the algorithms to create meaning from all this data, storage is not such a big problem anymore. Researchers only constantly introduce new bigger measure for data storage.

Value

All the available data should create a lot of value for organizations, societies, and consumers. Big Data means big business, and every industry could get its benefits. For example McKinsey's report (Manyika et al., 2011) mentions that Big Data has a potential annual value of €250 billion to Europe's public sector administration.

The important focus is that, data itself is not valuable at all. The value is in the analyses done on that data and how the data is turned into information and, eventually, knowledge and wisdom. The value is in how organizations use that data to create information-centric companies that base their decision making on insights derived from data analyses.

Variety

All data in the past was structured and placed into columns and rows of respected databases. But this is over today. Now, 90 percent of data is unstructured (Taylor, 2011). Data now comes in many different formats, including structured, semi structured, unstructured, and even complex structured.

Analytical specialists are required to identify and use different types of analysis and different tools to interpret for each type of data. Social media could provide insights on what customers think of concrete brand, service, or product, sensor data could provide information on how a product or machine is used and how to improve it.

STRUCTURED DATA

Flat files in record format; GEO location data; HTML5 micro data; Legacy data; Log files; Micro formats; Sensor data; Spreadsheets.

SEMISTRUCTURED DATA

Documents containing metadata tags; Electronic Data Interchange (EDI) documents; Rich Site Summary (RSS) feeds; Extended markup language (xml) objects.

UNSTRUCTURED DATA

Binary Large Objects (BLOBs); Business records requiring control; Content management data; Digital assets; Dynamic content (multiple users); Email, text messages, chat; Intellectual property data; Social data; Specialized content (web data); Static documents; Taxonomies/ontologies; Voice recognition data.

COMPLEX STRUCTURED DATA

Hierarchically structured data (xml-based MISMO) - XML datasets related to complex financial products.

Variability

Big Data is extremely variable. Brian Hopkins, defines variability as the "variance in meaning" (Popescu & Bacalu, 2011). Variability is often confused with variety. Variety is related to the different kind of objects, but if the characteristics of a single object are changing, this refers to variability.

Variability is thus very relevant in performing sentiment analyses. Variability means that the definition is changing (rapidly) (Gualtieri, 2012). To perform a proper sentiment analysis, algorithms need to identify the exact meaning of a word in its context. This is a process still reported to be very difficult to perform.

Velocity

Velocity is the main expression of speed – speed of data creation, storage, analysis, visualization. Batch processing was the common practice in the past, when database updates were done every month, week, or in the best case, evening. Processing was a time consuming activity that needed to be planned carefully according to the computers and servers functionalities. The difference in the Big Data era is the real time creation of data. All devices, machines connected in Internet can pass on data in the very moment it is created.

Human imagination could not even understand the speed at which data is currently created – e.g. about 100 hours of video is uploaded to YouTube every minute, over 200 million emails are sent every minute, around 20 million photos are viewed, almost 300000 tweets are sent, and almost 2.5 million queries on Google are performed (Welch, 2013), (Temple, 2012), (Tsukayama, 2013).

The biggest challenges for organizations today are to cope with the enormous speed at which data is created and to introduce appropriate actions for using referred data in real time.

Veracity

The purpose of data collection, processing and analysis is to support in the best way the decision making. So, generation, collection, processing of Big Data at high speed would be worthless if that data is not the correct and right one. Incorrect and not appropriate data could cause significant problems for organizations and for consumers (Gualtieri, 2012). In order to build the Big Data strategy, organizations should ensure that all data and its analysis are correct. Therefore, Big Data strategy requires a strong focus on data correctness and accuracy of its analysis.

Visualization

Visualisation is a very hard part of Big Data approach. It requires identifying the right user needs, to figure out the right analysis to be performed and to select the right technology and manner that vast amount of data could be transformed in an easy to read and understand form. With the right visualizations, raw data could be better understood and appropriately used. Today, visualization is not a simple graph or pie chart presentation. Visualisations are getting more and more complex in graphs that could include many variables of data, and extracting new understanding and insights for the business and processes.

Visualization is not the most technologically difficult task, but it is the most challenging one. Visualisation requires business and management knowledge on one hand, and on the other, detailed understanding of analytical tools functionality. The use of a graph to explain a complex case is very difficult and critical for the management. Luckily, there are more and more Big Data start-ups that focus on this aspect of the challenge. In the end, visualizations will make the difference.

4. Big Data Phenomenon Details

It is hard to identify a unique reason for the explosive growth of data. One cause is definitely innovation. Innovation is a phenomenon itself that transformed the way businesses and societies function. Four are the main directions of change in the world of data - business model transformation, globalization, personalization of services, and new sources of data.

- *Business model transformation* - business models have been heavily influenced by connectivity. Companies have changed their performance from being product-oriented to service-oriented, where the value of the organization in its customers' view is measured by service effectiveness and not only by product usefulness. This change influenced every business to publish additional data for products and services to serve variety of customer segments and channels, and to collect and analyse as much data from each customer access point - direct feedback, social media, forums, surveys, call centers, research, etc., as possible. The amount of data produced and used by every organization today is practically exploding. The data related to the business functionality itself remains, but the supporting data that is needed today and was not available or accessible is expanding and getting accessible through multiple channels.
- *Globalization* - is a key and multi-aspect trend that has drastically changed the world of commerce - from manufacturing to customer service. This process was supported by complex and fundamental change in variety and formats of data. The whole information technology was put under heavy pressure to change and serve the growing business needs.
- *Personalization of services* - business transformation's maturity index is measured by the extent of personalization of services and the value perceived by their customers from such transformation. This model is one of the primary causes for the velocity of data that is generated.
- *New sources of data* – the main challenge for collecting data is constantly getting harder. New technology tipping points are created in the last decade - data is floating around in social media, mobile devices, sensor networks, and new media much more than ever before. The new data sources are producing content within the corporation now that was never used in decision-making process from a Business Intelligence and Analytics perspective.

The establishment of newer business models and the aggressive growth of technology capabilities over the last decade have outlined the way for integrating all the data across the enterprise into one holistic platform to create a meaningful and contextualized platform for supporting the business decision-making process.

Big Data is fundamentally about applying innovative and cost-effective techniques for solving existing and future business problems that resource requirements exceed the capabilities of traditional computing environments which currently are configured within the enterprise. The important dimensions that Big Data is influencing the business environments in are shown on Figure 3.

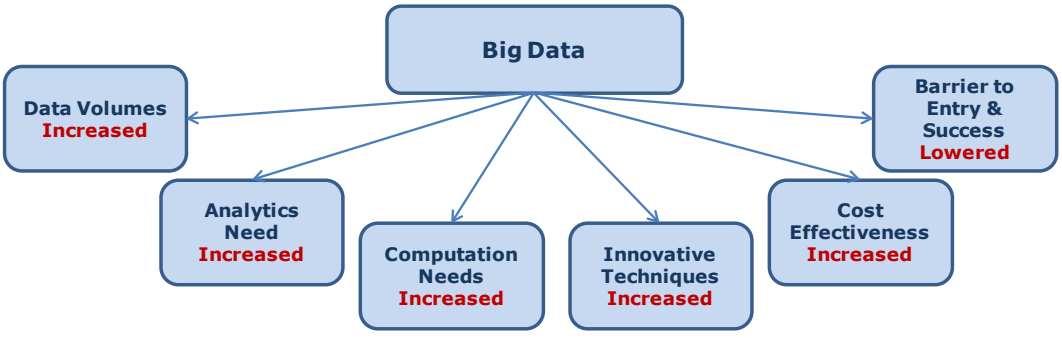

Fig. 3. Big Data Performance Dimensions

5. Big Data Main Directions for Improving Educational Industry

Innovative technologies could equip universities, colleges, and schools with tools and instruments to analyse absolutely everything that happens within their environment - students, teachers, and employees. They could receive additional insights about student behaviour, about testing results, about career development, about educational needs reflecting the society changes. Much of this data has been already collected and processed by statistical analysis to serve the government institutions.

Big Data could be used for the whole sector improvements – from a single school, to governmental directions and satisfaction of the labour market. With more and more distance learning and online courses, all that data acquires a completely new meaning. Big Data could support changes in education that would revolutionize all educational aspects - the way students learn, teachers teach, administrations manage, employers select and develop staff.

- *Improving student results* - the main objective of implementing Big Data within the educational system should be addressed to improve student results. The better the students are prepared for society, governments, and organizations, the better the educational institutions fulfil their role. Currently, student performance is only measured by the answers to assignments and exams. What stay hidden behind the formal results are the student behaviour changes that could create an exhaustive record of their development. Collecting data from different sources, for different students, for different processes and for all participants, could figure out better all the players, processes and relations. This data could be analysed in real time to deliver an optimal view of the whole learning environment and give a better understanding of individual student behaviour and all contributing components. Within such an analytical environment with all data collected, it is now possible to monitor all student actions – to analyse how long they need to answer a question, what sources they prefer to use, which questions they skip, which teaching resources work best for which student, what is the relation between teaching materials and questions, etc.

Additionally, Big Data could support creation of more productive groups of students. Students often work in groups where the members do not complement each other. By appropriate data and analysis, it would be possible to determine the strengths and weaknesses of each student and create stronger groups that could allow students to have a growing learning curve and deliver better results.

- *Creating mass student-centric programs* - the collected data related to the different educational process aspects could be a basis for creating customized programs for each student, even if a college or university has tens of thousands of enrolees. Such data could be captured from different points of student access delivered by the learning and management platform. According to the restricted educational requirements and given opportunities for personal choices, students could design their own personalized program, following classes from their interest and working at their own pace, while having the possibility for directed communication with professors, administrators and colleagues. In order Big Data analysis to support the educational process, the right tools are required to process, store, analyse, and visualize all the data involved. At the moment, most of the online courses are still mass produced, but in the future they could be mass customized.
- With many students participating in training, universities create big amounts of data that are still not appropriately used for approaching the right students and support their study in the best way.

In the future, universities would have the opportunity to find the best students when making scholarship decisions. This will increase the overall ranking of a university.

- *Improving the learning experience in real time* – the learning process is very dynamic and the reaction for the appropriate study tailoring should be very fast. Students today should be motivated to work independently in supported but personally customized learning programs. Students should learn how to teach themselves and be able to customize their courses. Using innovative Big Data analytical tools, the professor could easily monitor students in real time and direct their interest in deeper topics of choice. This approach would give students an opportunity to gain a better understanding of the subjects and be confident of constructing individual approach. When students are monitored in real time, digital textbooks and course outlines can be improved. Special algorithms could monitor how the students read the texts, including identification of which parts are difficult, which are easy, and which are unclear. Changes could be based on analysis of how often a text is read, how long it takes a text to be read, how many questions are asked about a specific topic, how many links are clicked when looking for more information, and how many and which sentences and paragraphs are underlined. If this information is provided in real time, professors could adapt their textbooks to meet the needs of students, in this way improving the overall education process results.

Additionally, Big Data could give insights on how each student learns. This is very important for improving the learning effectiveness and results. Some students are very concentrated and learn very efficiently, while others may be slow and extremely inefficient. Delivering course materials online, allows the whole learning process to be monitored and analysed, and on the other hand, the academic administration procedures could be very well organised and managed.

- *Increasing results and reducing drop-outs* - all analyses on educational Big Data is logically approached to improve student results and reduce drop-out rates. Closely monitoring students according to the outlined key performance indicators and receiving instant feedback about their personal needs, and delivering guidelines accordingly, could help to reduce the number of drop-outs that would benefit not only educational institutions, but the society as well.

Educational institutions should start using predictive analytics on all collected data in order to gain deeper insights into future student performance. Predictions in educational process could be used to build curricula according to the current society needs, to change a program if negative results are predicted, to change administrative procedures if bottlenecks are identified or even to run scenario analyses on a program before it starts. Universities and colleges would become more efficient in developing programs if they are structured and organised based on internal educational and management processes analysis, and on external societal evaluation and concrete needs.

Big Data educational analysis should continue after student graduation. Monitoring should be approached to analyse how students perform in their jobs. Opening this information publicly would give future students better insights how to choose the appropriate university.

Big Data will deeply change the whole learning industry in the coming years. More and more universities and colleges are already turning to Big Data to improve their overall performance and serve better the society and business.

6. Educational Big Data Governance Requirements

Analytical processes would build value to the university performance, only

when the managing body puts in place appropriate managing procedures for using the analytical results into practice. A very important aspect of analytical results usage is how to present them in the best and easy way, to the right place for decision taking and on time. Without using analytics, universities will not leverage them to the full possible extent for improving educational and administrative processes.

Data governance (Loshin, 2013) describes the processes for defining corporate data policies, defining processes for operationalizing observance of those policies, along with the organizational structures that include data governance councils and data stewards put in place to monitor, and hopefully ensure compliance with those data policies.

Related to universities, the objective of data governance should be focussed on establishing the right levels of data organisation and control, in order to achieve the following outcomes:

- *Alerting*: Identifying data issues that could cause negative impact on university – could be internal processes or external ones;
- *Sorting*: Prioritizing those issues that correspond to university value drivers – putting in priority those projects that could contribute to the problem solving or to the opportunity development;
- *Remediating*: Introducing data stewards who know what are the proper actions to be taken when alerting is flagged on – university analytical processes should be supported by concrete responsibilities and procedures to communicate with the original data owners or to direct data intervention.

In order to realise the data governing procedures, it becomes critical scoring processes to be implemented and utilized by a wide range of users and applications.

There are some methods that could be applied for developing embedded scoring processes:

- User-defined functions are the most native approach to embed a scoring function in the database.
- Another option could be SQL - the native language of the database. This is suitable for models like decision trees, linear regression, or logistic regression. Coding score systems in SQL is very appropriate for such models.
- Predictive modelling mark-up language is a new way to build a model in one system and pass information about the model to another system. Passed information enables the receiving system to generate scoring code automatically.
- Embed scoring processes that allow analytic tools to be run in the database directly, without the requirement for additional translation from analytic tool languages to other languages.

Universities are very sensitive structures that depend very much on data to be well organised, analysed and used. Five main principal components could be defined that are required for effective management of all big data governance of universities development – big data entry sets for analytics, analytical model design, model validation and reporting, scoring model, and technology (Figure 4).

Fig. 4. Big Data Governance

Big Data Entry Sets

The importance of this component is focused on tracking the details for each university analytic data set that feeds into an analytics process. All variety of information about the data sets and the technical facts related to creation and storing them should be tracked. The scope of these data sets should be very well defined in order to collect all academic analytical data sets, students' data sets for the process, and all related external data sets. The related technical information includes:

- User-defined function, embedded process, name of the SQL script, stored procedure, view that will provide the data set to the user;
- Parameters for the analytic data set processes – data range or object filter should be specified;
- Views which analytical processes will create supported by the metrics to be followed;
- Relationships between analytic data sets and analytic processes - analytic data sets could enter into one or more processes.

Analytical Model Design

To build this model, it is necessary to define each process and to track a variety of information about it. The model could be used for prediction or other analytic objectives, that for universities could be directed at analysis of student performance, analysis of potential candidates, analysis of administration effectiveness, etc. that needs to be utilized on a regular basis. Tracked information includes:

- The way the model would be used, the concrete educational or administrative issue addressed and the designed analytical scenarios;
- Details for model design development – when was created, by whom and revisions done;
- Model descriptions – algorithms used and methods applied;
- Details for the model scoring function - user-defined function, embedded process, name of the SQL script, stored procedure, view that will provide scores back to the user.
- Defining model input variables - the specific variables from the input analytic data sets, their variety and required metrics.

Model Validation and Reporting

In order to effectively manage the models and processes it is necessary to design a series of reports to cover a range of topics and purposes. Tracked information includes:

- Reports showing the analytical results and practical usage;
- Summary statistics or validations - diagrams or charts reviewing the scoring results;
- Model comparisons or variables summaries;
- Reporting could be automatically generated or run upon request. Reports are important for monitoring the critical steps of the performance. All models should be updated accordingly and reports should help identifying the moment.

Scoring Model

This model should track the scoring results - actual scores generated for every entity - student, program, administration procedure, etc. Tracked information includes:

- Generating and presenting the score value about related identifier – student, program, administration procedure, etc.;
- Historical scores record for a defined period of time for additional analysis in

order to understand the insights of big data analytics governance.

Technology

There are commercially available tools to help with model and score management, or a custom solution could be built to address the university specific needs.

Big data governance and related modelling and scoring management systems support universities within the difficult and dynamic processes of change for transformation to the Big Data era. A model and score management system will help universities to ensure the realisation of such a big growth and will guarantee the appropriateness of analytical processes definition and their updates. The lack of specifically devoted efforts to track models and analytic processes development could put universities at risk of running wrong models or being incorrectly used.

7. Conclusions

Enabling universities to take advantages of analytics in many new and innovative ways, becomes a very challenging and important issue. The expanding universe of created information and the increased need of its analysis for management processes, require a very critical change for understanding and reflecting on Big Data phenomena.

It is important to remember that Big Data is not only a new technology. Universities should accordingly reflect on the new requirements related to the business processes design and modelling for meeting the need to introduce analytics for all the aspects of management.

Universities should start devoting additional efforts today for understanding the characteristics of Big Data in general, and to figure out the specific dimensions of their own environment. They should reconsider the whole data generation, collection, analysis and use processes.

Logically, the desire for receiving measurable high quality information within the Big Data era requires from universities to define and establish Big Data governance strategy. It is naive to believe that Big Data governance could be reached with traditional approaches to data collection, analysis and use.

Big data analytics and governance is a challenge for each institution, because there are no standards and recipes universally adaptable to the conventional approaches for data quality and data governance. Vendors, system integrators, and consultants are referring to Big Data by stressing on the need for Big Data governance. The methods and tools used to monitor, review, and correct data streaming into a Big Data platform should be a reflection of the individual organisation definition of its Big Data governance strategy in order to meet the expected competitive development objectives.

References

Adrian, M., 2013. Big Data. Teradata Magazine. Available at: http://www.teradatamagazine.com/v11n01/Features/Big-Data/

Buytendijk, F., Laney, D., 2014. Information 2020: Beyond Big Data. Available at: https://www.gartner.com/doc/2681316?srcId3132930041#a183310972

Dumbill, E., 2012. What is big data? An introduction to the big data landscape. Available at: http://radar.oreilly.com/2012/01/what-is-big-data.html

Gartner Inc. Big Data Management & Analytics. Available at: http://www.gartner.

com/technology/topics/big-data.jsp

Gualtieri, M., 2012. What's Your Big Data Score? Available at: http://blogs.forrester.com/mike_gualtieri/12-05-17-whats_your_big_data_score

Henschen, D., 2014. 16 Top Big Data Analytics Platforms. Available at: http://www.informationweek.com/big-data/big-data-analytics/big-data-analytics/1113609

IBM Website. What is MapReduce? Available at: http://www-01.ibm.com/software/data/infosphere/hadoop/mapreduce/

Loshin, D., 2013. Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph. Morgan Kaufmann Publishers.

Magoulas, R., 2010. Roger Magoulas on Big Data. Radar Insight, Analysis and Research about Imerging Technologies, 15 January 2010. Available at: http://strata.oreilly.com/2010/01/roger-magoulas-on-big-data.html

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung Byers, A., 2011. Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute, May 2011. Available at: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

McGaughey, K., 2011. Worl' Data More Than Doubling Every Two Years - Driving Big Data Opportunity, New IT Roles. EMC Press Release, 28 June 2011. Available at: www.emc.com/about/news/press/2011/20110628-01.htm

Popescu, A., Bacalu, A., 2011. BigData: Volume, Velocity, Variability, Variety. Available at: http://nosql.mypopescu.com/post/6361838342/bigdata-volume-velocity-variability-variety

Prentice, S., 2011. CEO Advisory: "Big Data" Equals Big Opportunity. Gartner Inc. website, 31 March 2011. Available at: https://www.gartner.com/doc/1614215/ceo-advisory-big-data-equals

Taylor, J., 2011. IBM's Big Data Platform and Decision Management, Decision Management Solutions. Available at: www-01.ibm.com/software/data/bigdata/

TDWI. Big Data Analytics. Available at: http://tdwi.org/portals/big-data-analytics.aspx

Temple, K., 2012. What Happens in an Internet Minute? Available at: http://scoop.intel.com/what-happens-in-an-internet-minute/

Thiele, M., 2012. Big Data adoption issues – What's the big deal? Available at: http://gigaom.com/cloud/big-data-adoption-issues-whats-the-big-deal/

Tsukayama, H., 2013. Twitter turns 7: Users send over 400 million tweets per day. Available at: http://www.washingtonpost.com/business/technology/twitter-turns-7-users-send-over-400-million-tweets-per-day/2013/03/21/2925ef60-9222-11e2-bdea-e32ad90da239_story.html

Welch, C., 2013. YouTube Users Now Upload 100 Hours of Video Every Minute. Available at: http://www.theverge.com/2013/5/19/4345514/youtube-users-upload-100-hours-video-every-minute

Woods, D., 2011. Explaining Hadoop to Your CEO. Forbes website. Available at: www.forbes.com/sites/danwoods/2011/11/03/explaining-hadoop-to-your-ceo/

Xpert Service Wiki. Big data analytics. Available at: http://www.xpertservice.com/services/big-data-analytics.html