

# Toward Grasping the Dynamic Concept of Big Data

**Luis Emilio Alvarez-Dionisi**

INSISOC - Social System Engineering Centre of the University of Valladolid, Valladolid, Spain

E-mail: [dr.luis.alvarez@outlook.com](mailto:dr.luis.alvarez@outlook.com)

**Abstract**—The idea of big data represents a growing challenge for companies such as Google, Yahoo, Bing, Amazon, eBay, YouTube, LinkedIn, Facebook, Instagram, and Twitter. However, the challenge goes beyond private companies, government agencies, and many other organizations. It is actually an alarm clock that is ringing everywhere: newspapers, magazines, books, research papers, online, offline, it is all over the world and people are worried about it. Its economic impact and consequences are of unproportioned dimensions. This research outlines the fundamental literature required to understand the concept of big data. Additionally, the present work provides a conclusion and recommendations for further research on big data. This study is part of an ongoing research that addresses the link between economic growth and big data.

**Index Terms**—Database, Data Science, Big Data, Software Engineering, Software Architecture, Business Analytics.

## I. INTRODUCTION

Whenever you perform an Internet search, your inquiries and successive clicks are usually recorded. Likewise, every time you shop on Amazon or eBay, your purchase and every click is caught and logged; in a nutshell, your customer behavior is most likely recorded [1].

As a result, data is exponentially increasing, getting bigger and bigger.

On the other hand, the literature on big data continues to expand as this paper is being crafted. Big data is a polysemy term with multiple definitions and meanings.

Therefore, as highlighted by Diebold in 2012 in his paper *On the Origin(s) and Development of the Term “Big Data”*, the term big data “spans computer science and statistics/econometrics, probably originated in a lunch-table conversations at Silicon Graphics Inc. (SGI) in the mid 1990s” [2].

Likewise, big data “applies to information that can’t be processed or analyzed using traditional processes or tools. Furthermore, big data is referred to “data volumes in the range of exabytes ( $10^{18}$ ) and beyond. Such volumes exceed the capacity of current on line storage systems and processing systems. Data, information, and knowledge are being created and collected at a rate that is rapidly approaching the exabyte/year range. But, its creation and

aggregation are accelerating and will approach the zettabyte/year range within a few years. Volume is only one aspect of big data; other attributes are variety, velocity, value, and complexity” [3].

Big data represents the state-of-the-art techniques and technologies used to capture, store, allocate, manage and perform analysis of large datasets (petabyte- or larger-size) with high-velocity and different structures that traditional data management systems are not capable of handling [4].

As a matter of fact, credit for the concept of big data must be shared; in particular with John Mashey and many others at SGI, responsible for producing “highly-relevant (unpublished, non-academic) work in the mid-1990s” [2].

Similarly, big data has “demonstrated the capacity to improve predictions, save money, boost efficiency and enhance decision-making in fields as disparate as traffic control, weather forecasting, disaster prevention, finance, fraud control, business transaction, national security, education, and health care” [4].

Big data is a notion that is associated with data science.

Data science is the study of the transformation of the data into information. For that reason, data science embraces big data analytics.

As a result, big data analytics is the process for analyzing and mining big data. Consequently, big data analytics “can produce operational and business knowledge at an unprecedented scale and specificity” [5].

Nonetheless, there is a need to receive a formal education in data science. Correspondingly, universities across the United States of America are now offering degrees in the data science area to fill the gap.

For example, you can obtain a master’s degree in Data Science from New York University; Illinois Institute of Technology; Indiana University, Bloomington; Southern Methodist University; University of Virginia; Columbia University; and many other schools.

Nevertheless, there are academic programs that emphasize on decision support systems, visualization design, statistical packages, business analytics, and other academic orientations.

Each university has its own specific approach to deliver data science education.

In the same way, you can earn a master’s degree in Applied Business Analytics Management from Boston University. This graduate degree is very attractive because it could be done on campus and online.

Similarly, Spain has many academic institutions with big data related degrees, some examples of such institutions and their degree programs are: (1) Universitat Oberta de Catalunya (UOC) that currently offers an online master's degree in Business Intelligence and Big Data; (2) CIFF Business School of the University of Alcalá that offers a master's degree in Big Data and Business Analytics; (3) Universidad Internacional de la Rioja (UNIR) that runs an online university master's degree in Visual Analytics and Big Data; (4) Instituto de Empresa (IE) that offers a master's degree in Business Analytics and Big Data; (5) University of Valladolid (UVA) that is currently offering a degree in Informatics Engineering where participants learn about the details of the big data technological and management field; and (6) Universidad de Málaga that runs a master's degree in Advanced Analytics on Big Data. While the list is very long, it is out of the scope of this research to mention all the academic institutions that are currently delivering big data education in Spain.

Though, the concept of big data represents a big challenge for many organizations across the globe.

However, such challenge is currently knocking the main door of private and government institutions worldwide and it needs to be addressed right away.

In "a few years more data will be available in a year than has been collected since the dawn of man" [3].

As stated by IBM, "human beings now create 2.5 quintillion bytes of data per day. The rate of data creation has increased so much that 90% of the data in the world today has been created in the last two years alone" [5].

Consequently, big data requires immediate actions; otherwise, its impact could produce a domino effect on the organization's information assets, generating a technological and economic outrage.

In practice, organizations are putting together big data project teams to tackle the big data phenomenon.

In that sense, a typical core team includes the following roles: project director (a top senior business management executive), chief information officer (CIO), project manager, a database administrator (DBA), IT security manager, online marketing manager, system architect, and business analytics manager.

Based on the organization's requirements, a big data project management implementation lifecycle (BDPMIL) needs to be defined by the core team in order to put into practice the big data technology and finally achieve success.

The BDPMIL provides the necessary phases required to accomplish the big data Project, along with the essential tasks needed to fulfill the implementation effort.

In the same way, the BDPMIL includes all the milestones and deliverables needed to achieve the project, which is constrained by cost, time, scope, big data technology, around the quality constraint

Likewise, a project governance organization has to be established to oversee the entire big data Project.

In consequence, once the big data Project is successfully completed, many of the team members are usually moved into a new data science organizational unit

under the company's IT department.

As a result, big data is a global technological fact with massive economic implications.

Due to the fact that implementing big data technology is a complex assignment, it is possible to breakdown the big data Project into subprojects, producing a big data Program, which is constrained by the dependency among the constituent projects, the realization of the program benefits, and the management of the deliverables between projects.

In some cases, organizations go one step further and structure their big data implementation effort into a big data Project Portfolio, which is constrained by each of the project investments.

Nevertheless, the review of the current literature has revealed a limited gap in producing a research based framework for visualizing the origin and technical foundation of big data. Therefore, this article was written to introduce the fundamental literature review of such important topic.

The rest of this paper is organized into the following sections: (2) Developing the Big Data Equation; (3) Going from Megabyte to Yottabyte; (4) About the Big Deal with Big Data; (5) About Types of Data; (6) Adopting Big Data; (7) Introducing Big Data Solutions; (8) About Major Risks; (9) A Brief on Big Data 1.0 and Big Data 2.0; (10) Introducing the Fourth Paradigm; (11) Presenting Big Data Sample Applications; and (12) Conclusion and Recommendations.

## II. DEVELOPING THE BIG DATA EQUATION

Using the ideas of Kaisler, Armour, Espinosa, and Money [3], along with the fundamentals of discrete mathematics, the following equation was developed to further understand the concept of big data.

$$\text{Big Data} = \{ \text{Data Volume, Data Variety, Data Velocity, Data Value, Complexity} \} \quad (1)$$

In where, Big Data is defined as a mathematical set that includes the following features.

**Data Volume:** This feature measures the amount of data open to an organization. The idea is for the organization to have access to the proper data. However, it is possible that such data is owed by somebody else. Though, it is important to point out that it is probable that as data volume increases, the value of several data records will decrease in relation to type, age and quantity among other elements [3].

**Data Variety:** This characteristic measures the richness of the data format (e.g., text, audio and video). From the big data analytics viewpoint, it is perhaps a key challenge to effectively using a large amount of data. This is because some of the data are coming in different layouts and generated from multiple data sources [3].

**Data Velocity:** This feature measures how rapid the data is created, streamed, and aggregated. Data velocity

management goes beyond bandwidth management. However, it is also important to highlight that e-commerce capability has promptly improved the speed and richness of data velocity.

**Data Value:** This feature measures the usefulness of the data in the decision-making process. Specifically, it is associated with the data value that is provided to the organization [3].

**Complexity:** This characteristic measures the degree of interdependence and interconnectedness of big data structures [3].

### III. GOING FROM MEGABYTE TO YOTTABYTE

In this section, we will introduce some illustrations of the units of measurement of data volume.

In practice, the idea of data volume is essential to grasp the concept of big data.

As depicted in Fig.1., we have adapted the overview of data scale from megabytes to yottabytes (log scale) to provide an example of data volume [6].

The mentioned figure indicates that a Megabyte is 1,000,000 bytes ( $10^6$  bytes); one Gigabyte is equivalent to 1,000,000,000 bytes ( $10^9$  bytes); a Terabyte is 1,000,000,000,000 bytes ( $10^{12}$  bytes); a Petabyte is 1,000,000,000,000,000 bytes ( $10^{15}$  bytes); one Exabyte is 1,000,000,000,000,000,000 bytes ( $10^{18}$  bytes); a Zettabyte is 1,000,000,000,000,000,000,000 bytes ( $10^{21}$  bytes); and one Yottabyte is equal to 1,000,000,000,000,000,000,000,000 bytes ( $10^{24}$  bytes).

Unit	Numeric Representation	Exponential Representation
Megabyte	1,000,000 bytes	$10^6$ bytes
Gigabyte	1,000,000,000 bytes	$10^9$ bytes
Terabyte	1,000,000,000,000 bytes	$10^{12}$ bytes
Petabyte	1,000,000,000,000,000 bytes	$10^{15}$ bytes
Exabyte	1,000,000,000,000,000,000 bytes	$10^{18}$ bytes
Zettabyte	1,000,000,000,000,000,000,000 bytes	$10^{21}$ bytes
Yottabyte	1,000,000,000,000,000,000,000,000 bytes	$10^{24}$ bytes

Fig.1. Going from Megabyte to Yottabyte (Adapted from Figure 1 of Harris [6])

### IV. ABOUT THE BIG DEAL WITH BIG DATA

According to Einav and Levin [1], a summary of the big deal with big data is encapsulated in the following paragraph.

During the last thirty years, technology has changed our lives. Looking at the Internet, cell phones, text messaging, electronic health records, and employment records, all these footsteps are part of the trails that we leave behind. Evidently, data is everywhere about us. However, what is new about it is that the data is now retrieved faster, has larger exposure and scope, and

includes new types of remarks and observations that previously were not there. Therefore, data is now available on a large scale. At present, data sets come with masses of distinct observations and an enormous number of comments. For that reason, it is real big and the discipline of big data was born. Likewise, it is available in real time. The magnificence of real-time capability in term of capturing and processing data is that it becomes vital for many big data business applications. Similarly, it is also available in all type of formats. A great amount of data is now being logged in events that previously were almost impossible to detect. Consider, for example, geolocation records showing where the persons are situated, along with social networks connections. In the same way, data is less structure and highly dimensionally. Nowadays, data can be loaded in all shapes and forms, for instance: databases, worksheets, videos, emails, websites, and graphics.

Nevertheless, it is important to highlight that the amount of information and knowledge that can be extracted from big data keeps growing as users come up with new ways to control and process the data found in big data environments [3].

Big data is important because it captures the essence of the innovative phenomenon of science and business produced as part of the IT evolution [2].

As a result, big data “will transform business, government, and other aspects of the economy” [1].

Definitely, it can be stated that the arrival of big data will change the way we do business. Therefore, big data is here to stay and we need to cope with it.

### V. ABOUT TYPES OF DATA

In order for you to comprehend big data, it is necessary to understand the following three types of data [7].

**Structured Data:** This is the data stored in traditional database structures. An example of this type of data is the data kept in relational database management systems (RDBMS).

**Unstructured Data:** This is the data with no formatting. Examples of this type of data are emails, PDF files, and online documents in general.

**Semi-structured Data:** This is the data that has been processed to some extent. Examples of this type of data are HTML or XML-tagged texts.

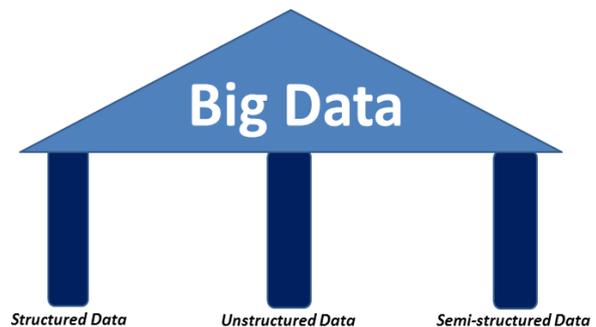


Fig.2. Big Data Columns

Consequently, Fig.2. shows the three columns necessary to support the big data environment. In fact, structured data, unstructured data, and semi-structured data represent the pillars sustaining the big data management information roof. Without proper management of these three columns, big data will be at risk.

VI. ADOPTING BIG DATA

The main three reasons for adopting big data are introduced in the following paragraphs [5].

**Decreasing Storage Cost:** The cost of storage has extensively decreased during the last few years. Therefore, big data applications are typically built to retain widely historical tendencies in storage devices that have reduced storage cost.

**Variety of Formats:** Big data tools allow you to load structured, unstructured, and semi-structured data in a variety of formats without defining schemas ahead of time.

**Innovative Data Management Solutions:** Solutions such as NoSQL databases and Hadoop increase the processing speed and queries of big data analytics.

Accordingly, Fig.3. depicts the interaction of above reasons for adopting big data.

Furthermore, in order to benefit from big data, innovative storage mechanisms and groundbreaking analysis methods need to be implemented [7]. Therefore, big data represents a significant investment in terms of hardware, software, management, and technological skills for many organizations.

Equally, big data technologies are organized into two groups: batch processing and stream processing. In that sense, batch processing are analytics on data at rest and stream processing are analytics on data in motion [5].

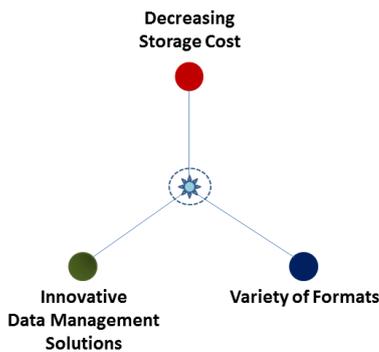


Fig.3. Reasons for Adopting Big Data

VII. INTRODUCING BIG DATA SOLUTIONS

The big data solutions outlined in this paper are NoSQL databases and Hadoop.

Some of the NoSQL databases are Amazon DynamoDB; BigTable; Neo4j; db4o; MongoDB;

Cassandra; Oracle NoSQL Database; CouchDB; and Riak.

Fig.4. shows the aforementioned NoSQL databases and their respective database types.

NoSQL Databases	Types
Amazon DynamoDB	Key-value Stored Database
BigTable	Column-oriented Database
Neo4j	Graph Database
db4o	Object-oriented Database
MongoDB	Document Stored Database
Cassandra	Column-oriented Database
Oracle NoSQL Database	Key-value Stored Database
CouchDB	Document Stored Database
Riak	Key-value Stored Database

Fig.4. NoSQL Databases

Amazon DynamoDB service is provided by Amazon. Similarly, BigTable was built on the Google File System (GFS). BigTable is not available outside Google. By the same token, Neo4j was developed by Neo Technology. Equally, db4o (database for objects) was developed by Actian. Similarly, MongoDB was developed by MongoDB Inc. In the same way, Cassandra was developed by Apache Software Foundation. Likewise, Oracle NoSQL Database is an Oracle Corporation product. Equally, CouchDB was developed by Apache Software Foundation. And finally, Riak was developed by Basho technologies.

Additionally, HBase (a column-oriented database) is also introduced in this paper. In that sense HBase was developed by Apache Software Foundation using Java.

HBase runs on top of Hadoop Distributed File System (HDFS) and was modeled after Google's BigTable.

On the other hand, Hadoop is an open source ecosystem licensed by Apache Software Foundation. Hadoop has two core layers: (1) a computational layer, which is MapReduce; and (2) a storage layer, which is HDFS. Likewise, Hadoop includes the YARN Framework. YARN stands for Yet Another Resource Negotiator.

Moreover, it is applicable to mention the following tools used by Hadoop: Flume; Sqoop; Spark; Hive; Oozie; Pig; Mahout; and Drill. Some authors include HBase as a Hadoop tool. Therefore, HBase was included in this paper as well.

As a result, Fig.5. shows a representative sample of some of the technological pieces of the Hadoop solution.



Fig.5. Hadoop Solution

VIII. ABOUT MAJOR RISKS

A risk is the “probability or threat of damage, injury, liability, loss, or any other negative occurrence that is caused by external or internal vulnerabilities, and that may be avoided through preemptive action” [8].

Therefore, the three major risks that are usually affecting big data environments are described as follow [4].

**Privacy and Cybersecurity:** Are perhaps the highest risks for big data applications. Privacy denotes personally identifiable information (PII) that is used to detect a person. Privacy becomes a risk for big data; especially whenever sensitive data is massive collected about people with-or-without their knowledge.

A further case becomes the profiling of such people, which can definitely result in the wrong conclusions about who someone really is.

On the other hand, cybersecurity deals with protecting big data applications against an Internet attack.

**Making False Decisions:** Big data analytics allow users to identify patterns from different data sources and forecast the relationships between variables.

However, big data analytics face the risk of misusing or misinterpreting the relationships between variables, originating false results.

**Over-dependence on Data:** This represents a significant risk to organizations because people become complete dependent of data for all the decisions they make.

Consequently, important things may be ignored and incorrect decisions may be made. The data diversity, size, and speed can meet only the essential but not appropriate conditions to solve the problem or assist with the decision-making process.



Fig.6. Big Data Risks

Given the relevancy of the big data risks, Fig.6. shows a visualization of these risks.

However, it is important to underline that big data

Risks need to be managed properly in order to mitigate the threats against big data technology.

As a result, big data Risk Management is a critical success factor that has to be in place in order to achieve big data accomplishment.

IX. A BRIEF ON BIG DATA 1.0 AND BIG DATA 2.0

In term of big data context, big data can be referred as Big Data 1.0 and Big Data 2.0.

The best way to visualize this classification is using Internet technologies (e.g., Web 1.0 and Web 2.0) to embrace the right version of the big data’s name.

For instance, in Web 1.0, organizations engaged themselves in setting the basic Internet technologies up in place so that they could establish a presence on the Internet, create electronic commerce functionality and increase operating efficiency and effectiveness. Some organizations can think themselves as being in the era of Big Data 1.0, they are currently engaged in creating capabilities to manage a large amount of data. This result in the right infrastructure required for supporting the big data environment in order to improve current operations. On the other hand, in Web 2.0, organizations started to exploit the interactive capability of the web using social networking such as Facebook, Twitter, and Instagram. Correspondingly, in the Big Data 2.0 era, the rise of the consumer’s “voice” become very important and it is used to evaluate products and services [9].

While many organizations are currently in the era of Big Data 1.0, others have achieved the level of Big Data 2.0.

X. INTRODUCING THE FOURTH PARADIGM

As displayed in Fig.7., the Pyramid of Paradigms shows the pathway toward the Fourth Paradigm for science on a large amount of data and intensive computing.

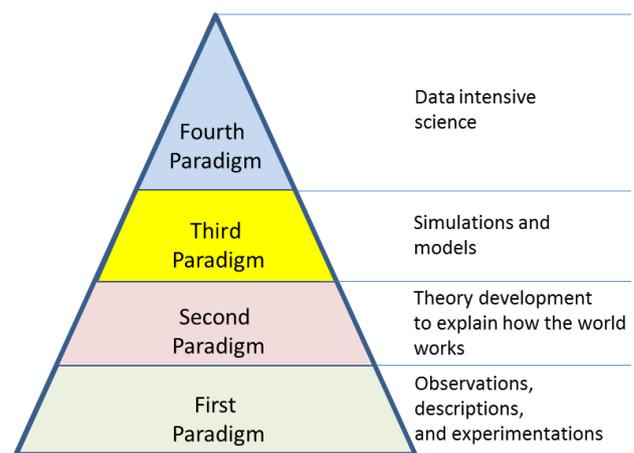


Fig.7. Pyramid of Paradigms

Accordingly, the First Paradigm is characterized by the participation of early scientists and pioneers and their

observations, descriptions, and experimentations. Similarly, the Second Paradigm deals with the development of theories to explain the way the world works. Likewise, the Third Paradigm takes into account and further develops the previous theories in order to create extensive simulations and models. Finally, the Fourth Paradigm has to do with data intensive scientific discovery that is available to science and society [6, 10]. As a result, the Fourth Paradigm deals with the world of big data.

## XI. PRESENTING BIG DATA SAMPLE APPLICATIONS

The best approach to visualize big data is through real life practical applications. For that reason, in the subsequent paragraphs, big data sample applications are introduced.

**Health Informatics:** The discipline of health informatics gathers medical data from molecular, tissue, patient, and population levels of the human being in order to assist the understanding of medicine and medical practice procedures. Data such as clinical-scale, human-scale biology, and epidemic-scale are also captured via health informatics [11].

**Advanced Persistent Threats Detection:** The advanced persistent threat (APT) is a direct attack perpetrated against a physical system or a high-value asset. APT maneuvers in a low-and-slow mode approach [5]. Therefore, APT management is a classic application of big data implementation.

**Enterprise Events Analytics:** On a daily basis, organizations collect a large amount of security data such as people action events, network events, and software application events. Consequently, the analytical techniques which are not big data oriented, do not perform very well at this scale; they can actually generate a large number of false positives and their efficacy could be affected. Of course, things become unmanageable as organizations move to a cloud architecture environment and collect further data [5].

**Manufacturing:** The incorporation of manufacturing automation, engineering, IT, and big data analytics are key successful factors in the manufacturing industry of today. Evolving technologies such as Internet of things (IoT), big data and cyber physical systems (CPS) make a significant contribution to measuring and monitoring real-time big data from the factory environment [12].

**Biomedical Big Data Initiative:** The objectives of this National Institutes of Health (NIH) initiative are to facilitate the use of biomedical big data, along with the development and distribution of analysis methods and software for such data, and enhancing the training for biomedical big data. Likewise, an additional objective is to create centers of excellence to study biomedical big data [4].

**Relationship Between Genes and Cancers:** The Frederick National Laboratory has been using big data applications “to cross-reference the Cancer Genome Atlas (TCGA) gene expression data from simulated 60 million

patients and miRNA expression for a simulated 900 million patients” [4].

**Using NetFlow Monitoring to Identify Botnets:** This application allows users to analyze large quantities of netFlow data in order to identify infected hosts participating in the botnet effort [5].

**Network Security:** This application allows users to perform a frequency network security analysis of events. This includes data mining of security information coming from firewalls, website traffic, security devices, and many other sources [5].

**Expanding Evidence Approaches for Learning in a Digital World:** The U.S. Department of Education has developed several big data learning applications to explore how people learn [4].

**Molecular Simulation:** The molecular simulation (MS) is a prevailing tool for studying chemical and physical characteristics of large systems in engineering and scientific areas. MS generates a very large number of atoms required to study their spatial and temporal links necessary to perform a scientific study. Consequently, MS applications need appropriate big data access and intensive process mechanism [13].

**Government Search Engine:** USASearch.gov is a site-search service provided by GSA that allows users to perform a “public’s search on federal, state, local, tribal, and territorial government websites” [4].

Similarly, prospective areas of application of big data technology are: (1) airline traffic control; (2) marketing analysis; (3) electronic encyclopedia summarization system as highlighted by Hatipoglu and Omurca in 2016 in their paper A Turkish Wikipedia Text Summarization System for Mobile Devices [14]; (4) agro-industrial engineering projects; (5) stock exchange predictions; (6) spam detection mechanisms as underlined by Iqbal, Abid, Ahmad, and Khurshid in 2016 in their paper Study on the Effectiveness of Spam Detection Technologies [15]; political campaigns; economic research; drug discovery; molecular analysis; weather forecast; route maps analysis; music creation; colonoscopy analysis; and many others.

## XII. CONCLUSION AND RECOMMENDATIONS

We have completed a basic literature search on big data.

Therefore, this paper was organized into the following sections: Introduction; Developing the Big Data Equation; Going from Megabyte to Yottabyte; About the Big Deal with Big Data; About Types of Data; Adopting Big Data; Introducing Big Data Solutions; About Major Risks; A Brief on Big Data 1.0 and Big Data 2.0; Introducing the Fourth Paradigm; Presenting Big Data Sample Applications; and Conclusion and Recommendations.

As a result, we conclude that big data provides the strategic input required to make critical decisions in organizations.

Consequently, the following ideas have emerged as a ground base for additional research on big data: study the application of big data to a space program; support

of political campaigns with big data; support of electronic encyclopedia with big data; use of big data to support the weather forecast analysis; review of SAS with big data; study the application of big data to economic research; study the application of big data to airline traffic control; study the application of big data to marketing analysis; support of stock exchange predictions with big data; use big data technology to collect and analyze span data; collect and analyze drug discovery data; use of big data to support molecular analysis; perform a real-time route maps analysis; create music with big data technology; apply big data to a colonoscopy analysis; perform a review of big data statistical applications; study big data skills; understand big data hardware, software, and methods; study big data project management practices; develop a NoSQL taxonomy; study the application of agile methods to big data implementation projects; study big data database management systems; evaluate the application of the Stata to big data; study Hadoop attributes; assess the economic and business impact of big data; study Spark technology; create a big data theory; understand the link of NoSQL databases with legacy systems; study document stored databases; study graph databases; review key-value stored databases; review column-oriented databases; study object-oriented databases; perform an analysis of the YARN framework; develop a visualization equation; study visualization techniques with storytelling; create a data science theory; study interactive tabletops with visualization; analyze the application of visualization to the Madden-Julian Oscillation (MJO) phenomenon; analyze big flow datasets; perform a comparison analysis of explanatory visualization versus exploratory visualization; use of visualization to support business intelligence analysis; support of agro-industrial engineering projects with big data; study the application of Matlab to big data; use of visualization to analyze geospatial data; study the visualization of social media data; study navy, army, air force, and national guard data with visualization; study the application of infographic; develop an agriculture big data model; study big data visualization philosophy; understand machine learning with big data; identify big data issues; analyze the use of IBM SPSS with big data;

conduct big data simulations; study the philosophy of HDFS; study the application of the statistical package R to big data; study the data conversion process from RDBMS to NoSQL databases; study the interfaces of Hadoop with NoSQL databases; understand the link of Hadoop with legacy systems; review big data security, ownership, privacy, and governance; study the data conversion process from RDBMS to Hadoop; and study big data analytics tools and techniques.

Likewise, it is recommended to group big data and data science practitioners and scholars in a Data Science Institute (DSI) in order to create a Data Science Body of Knowledge (DSbok); set up several levels of professional certifications in big data and data science; and build a progressive suite of big data and data science education and training programs.

It is equally important to create an accreditation program for universities that are offering data science and big data degrees.

#### ACKNOWLEDGMENT

I'd like to thank Aurilu Rivas for editing the manuscript of this paper. Thank you so much!

#### REFERENCES

- [1] L. Einav and J. Levin, "The Data Revolution and Economic Analysis," National Bureau of Economic Research, pp. 1–24, 2014.
- [2] F. Diebold, "On the Origin(s) and Development of the Term Big Data," Penn Institute for Economic Research, PIER Working Paper 12-037, Department of Economics, University of Pennsylvania, 2012.
- [3] S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big Data: Issues and Challenges Moving Forward," in *46th Hawaii International Conference on System Sciences*, 2013, pp. 995–1004.
- [4] J. Yan, "Big Data, Bigger Opportunities - Data.gov's roles: Promote, lead, contribute, and collaborate in the era of big data," President Management Council Inter-agency Rotation Program, Cohort 2, 2013.
- [5] Big Data Working Group, Big Data Analytics for Security Intelligence, Cloud Security Alliance, pp. 1–22, 2013.
- [6] R. Harris, "ICSU and the Challenges of Big Data in Science," *Research Trends: Special Issue on Big Data*, pp. 11–12, 2012.
- [7] B. Purcell, "Emergence of "Big Data" Technology and Analytics," *Journal of Technology Research*, vol. 4, pp. 1–7, July 2013.
- [8] Business Dictionary. Internet: <http://www.businessdictionary.com/definition/risk.html>, April 07, 2015.
- [9] F. Provost and T. Fawcett, "Data Science and its relationship to Big Data and Data-driven Decision Making," *Big Data*, vol. 1, pp. 51–59, March 2013.
- [10] T. Hey, S. Tansley, and K. Tolle, *The Fourth Paradigm: Data-intensive Scientific Discovery*, Microsoft Corporation, 2009.
- [11] M. Herland, T. M. Khoshgoftaar, and R. Wald, "A review of data mining using big data in health informatics," *Journal of Big Data*, vol. 1, pp. 1–35, June 2014.
- [12] P. O'Donovan, K. Leahy, K. Bruton, and D. O'Sullivan, "Big data in manufacturing: a systematic mapping study," *Journal of Big Data*, vol. 2, pp. 1–22, September 2015.
- [13] A. Kumar, V. Grupcev, M. Berrada, J. C. Fogarty, Y. Tu, X. Zhu, S. A. Pandit, and Y. Xia, "DCMS: A data analytics and management system for molecular simulation," *Journal of Big Data*, vol. 2, pp. 1–22, November 2014.
- [14] A. Hatipoglu and S. I. Omurca, "A Turkish Wikipedia Text Summarization System for Mobile Devices," *I. J. Information Technology and Computer Science*, vol. 1, pp. 1–10, 2016.
- [15] M. Iqbal, M. M. Abid, M. Ahmad, and F. Khurshid, "Study on the Effectiveness of Spam Detection Technologies," *I. J. Information Technology and Computer Science*, vol. 1, pp. 11–21, 2016.

## Authors' Profiles



**Luis Emilio Alvarez-Dionisi** earned his Ph.D. in Strategy, Programme & Project Management from SKEMA Business School, Lille, France. He received his MS in Computer Information Systems from Boston University, Boston, Massachusetts, USA and obtained his BS in Computer Science, Cum Laude from Central New England College of Technology, Worcester, Massachusetts, USA. His research interests include big data applications, global project management trends, agile project management, and information technology governance. He has worked on management consulting engagements in the USA, Switzerland, England, France, Trinidad and Tobago, St. Maarten, Aruba, Mexico, China, Costa Rica, Venezuela, Puerto Rico, Russian Federation, Pakistan, Sri Lanka, Singapore, Philippines, Vietnam, Malaysia, Hong Kong, and Indonesia. Dr. Alvarez-Dionisi was born in Venezuela. He has taught Computer Science, Information Systems, and Project, Program and Portfolio management across the globe. Currently, he provides big data consulting services; strategic project management advisory services; and corporate training to senior executives worldwide.

**How to cite this paper:** Luis Emilio Alvarez-Dionisi, "Toward Grasping the Dynamic Concept of Big Data", *International Journal of Information Technology and Computer Science (IJITCS)*, Vol.8, No.7, pp.8-15, 2016. DOI: 10.5815/ijitcs.2016.07.02