

# Envisioning Skills for Adopting, Managing, and Implementing Big Data Technology in the 21<sup>st</sup> Century

**Luis Emilio Alvarez-Dionisi**

Professor, Faculty of Social and Economic Science (FACES), School of Accounting,  
Santa Mar á University, Barinas, Venezuela  
E-mail: dr.luis.alvarez@outlook.com

**Abstract**—The skills for big data technology provide a window of new job opportunities for the information technology (IT) professionals in the emerging data science landscape. Consequently, the objective of this paper is to introduce the research results of suitable skills required to work with big data technology. Such skills include Document Stored Database; Key-value Stored Database; Column-oriented Database; Object-oriented Database; Graph Database; MapReduce; Hadoop Distributed File System (HDFS); YARN Framework; Zookeeper; Oozie; Hive; Pig; HBase; Mahout; Sqoop; Spark; Flume; Drill; Programming Languages; IBM Watson Analytics; Statistical Tools; SQL; Project Management; Program Management; and Portfolio Management. This paper is part of an ongoing research that addresses the link between economic growth and big data.

**Index Terms**—Big Data, Skills, NoSQL Databases, Hadoop.

## I. INTRODUCTION

The skill is the capability to perform something well. This is true because of knowledge, familiarity, talent, awareness, understanding, or practical experience.

There is no doubt that skills heighten the proficiency of professionals in the big data world. That is why this paper was developed.

Consequently, if you are familiar with the big data skills, then you can develop your resources toward achieving big data expertise.

Therefore, this research brings forefront knowledge to practitioners and scholars to further develop big data know-how.

According to Angel Gurr á from the Organisation for Economic Cooperation and Development (OECD), “Skills have become the global currency of the 21<sup>st</sup> century.” He highlights that “without proper investment in skills, people languish on the margins of society, technological progress does not translate into economic growth, and countries can no longer compete in an increasingly knowledge-based global society” [1].

As a matter of fact, the skills are the strategic assets

required to deliver goods and services.

As a result, Green visualizes the notion of skill as the personal quality embraced by the following features [2].

**Productive:** The skill deals with productive value;

**Expandable:** The skill is boosted by development and training; and

**Social:** The skill is an attribute that is socially determined.

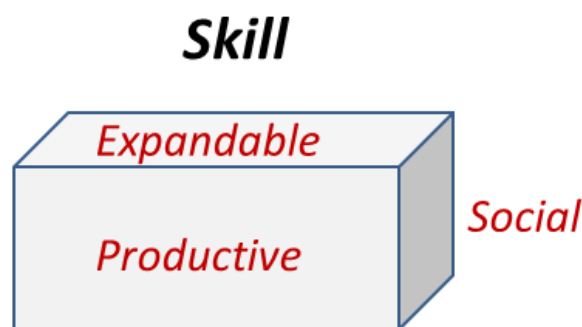


Fig.1. PES Model

By the same token, the idea of “PES” (Productive, Expandable, and Social) is embedded in the concept of skill and it is advocated accordingly. Fig.1. shows a graphical representation of the PES Model.

Therefore, the world economic downturn has produced an unemployment condition that affects the quality of life of many people across the globe. This situation has created the need to invest in new skills to find a better quality of life [1].

In that sense, this effort can be translated into achieving prosperity, good health, happiness, proper education, food, housing, transportation, freedom of association and religion, family stability, job stability, financial resources, and many other features that define a better life.

Green also indicates that many of the skill levels have significant economic effects on individuals, organizations, geographic locations, and even entire countries. However, there is not consensus when economists, sociologists, and psychologists try to define the meaning of skill. It seems to be that they are talking about different things. However, the idea of skill has numerous of synonyms such as

aptitude, ability, knack, talent, and competence [2].

Nevertheless, the arrival of the big data technology has contributed to the upgrade of many of the skills of the IT professionals.

Consequently, the growing environment of data volume, data variety, data velocity, data value, and complexity [3] has produced a new big data technological and economic trend.

If we look around us, big data is found almost everywhere: financial and banking institutions, manufacturing companies, healthcare organizations, military units, IT security services, weather forecast projections, time route maps analysis and many other areas.

Three particular examples of big data applications are: (1) the American drone aircraft that was capable of sending the equivalent of 24 years' worth of video footage involving its mission. (2) the supply chains management firm that was capable of visualizing in its network 100 gigabytes of information per day from its clients back in 2010. Similarly, starting in 2011 until today, this firm was able to generate more than ten times as many information per day [4]. (3) Parkinson prediction as highlighted by Shamli and Sathiyabhama in 2016 in their paper Parkinson's Brain Disease Prediction Using Big Data Analytics [5].

Similarly, prospective area for further application of big data technology is in biometric systems as underlined by Srivastava, Agrawal, Mishra, Ojha, and Garg in 2013 in their paper Fingerprints, Iris and DNA Features based Multimodal Systems: A Review [6].

As a result, we have entered into the big data age [7].

For that reason, this paper was crafted to present the basic literature needed to understand the essential skills required for adopting, managing, and implementing big data technology.

Accordingly, the rest of this paper is organized into the following sections: (2) Introducing the Big Data Skills Model; (3) Skill Perspectives; (4) NoSQL Databases Proficiency; (5) Hadoop Ecosystem Proficiency; (6) Forefront Tools Proficiency; (7) Project Management Proficiency; (8) About Developing Big Data Skills; and (9) Conclusion and Recommendations.

## II. INTRODUCING THE BIG DATA SKILLS MODEL

Looking at the evolution of computer science, communication technology, and information systems during the last three decades, we can appreciate the breadcrumbs that we leave behind us every time we use the Internet, send text messages, use mobile phones, and access our employment and electronic health records [8].

In fact, our personal and professional records are scattered everywhere. As a result, our data is progressively captured and visualized as information coming from big data repositories. Therefore, big data technology has changed our lives.

Fig.2. depicts the Big Data Skills Model that was created to show some of the skills necessary for working with big data environments.

This model is an abstraction of the reality necessary for embracing, managing, and kicking off the installation and putting to work the big data technology.

Therefore, the skills indicated in the model are grouped into the following proficiency sections.

**NoSQL Databases Proficiency:** This section includes the following NoSQL databases: Document Stored Database; Key-value Stored Database; Column-oriented Database; Object-oriented Database; and Graph Database.

**Hadoop Ecosystem Proficiency:** This section was structured into the following components: MapReduce; HDFS; YARN Framework; Zookeeper; Oozie; Hive; Pig; HBase; Mahout; Sqoop; Spark; Flume; and Drill.

**Forefront Tools Proficiency:** In this section, we include: Programming Languages; IBM Watson Analytics; Statistical Tools; and SQL.

**Project Management Proficiency:** This section underlines the key ideas of Project Management, Program Management, and Portfolio Management.

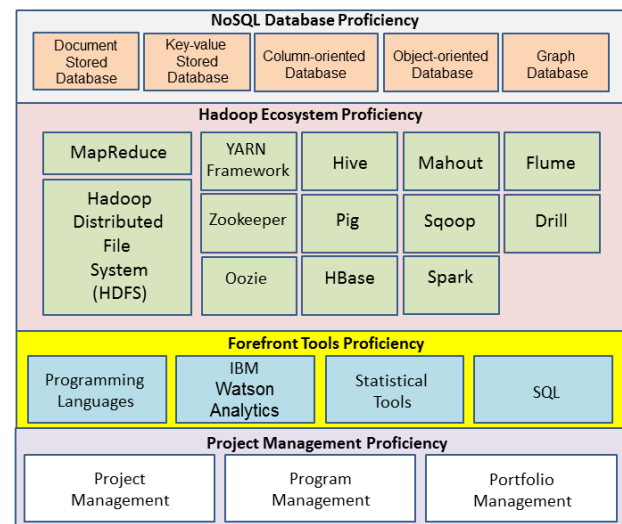


Fig.2. Big Data Skills Model

From the next section onward, the above-mentioned skills will be further highlighted toward the end of this paper.

## III. SKILL PERSPECTIVES

While skill is a polysemy term with different meanings from many writers with different disciplines, the term is considered the pivotal body for modern social and economic life.

As highlighted by Becker in 1964 and Mincer in 1974, in neoclassical economics, skill is one of the main components of "human capital", along with health. In that sense, education and training are investments used to develop skill/human capital. Accordingly, human capital is the value of the individual's stream of existing and forthcoming earnings discounted to the present.

On the other hand, Attewell underlined in 1990 that in contrast to neoclassical economics, sociology looks at the

concept of skill from the production process perspective. Consequently, in order to execute a greater skill, it is necessary to perform a more complex activity. Therefore, professionals perform sophisticated tasks with greater complexity as long they have acquired the corresponding skills.

The sociologists emphasize on complexity as being the real center of skill and using social construction theory they partially establish skill's association with value. In contrast, psychology is keen on measuring and classifying skills.

Nevertheless, psychology joins forces with sociology and evaluates the social context in which skills are learned and used [2].

#### IV. NOSQL DATABASES PROFICIENCY

The term NoSQL stands for “Not Only SQL.” However, it is important to highlight that NoSQL databases, along with the Hadoop Ecosystem, are used to create, manipulate, and analyze big data management storage.

As a result, NoSQL databases are those databases that are non-relational and horizontally scalable [9]. Likewise, several NoSQL databases are open source.

While the NoSQL databases do not guarantee the ACID properties (Atomicity, Consistency, Isolation, and Durability), they actually guarantee the BASE properties (Basically Available, Soft state, Eventual consistency). Similarly, the NoSQL databases are in compliance with the CAP (Consistency, Availability, Partition tolerance) theorem [10].

According to Cattell, NoSQL databases have the following six key features [11].

**Feature 1:** Capability to horizontally scale “simple operation” throughput over several servers.

**Feature 2:** Capability to replicate and distribute (partition) data across several servers.

**Feature 3:** Ability to perform a simple call level interface or protocol (in contrast to SQL binding).

**Feature 4:** Less strong concurrency model than the ACID properties of most relational (SQL) database systems.

**Feature 5:** Efficient usage of distributed indexes and RAM for data storage.

**Feature 6:** Capability to dynamically include new attributes to data records.

Consequently, learning one of the NoSQL databases is recommended for those professionals using this type of technology.

Fig.3. displays a graphical representation of the NoSQL Databases.

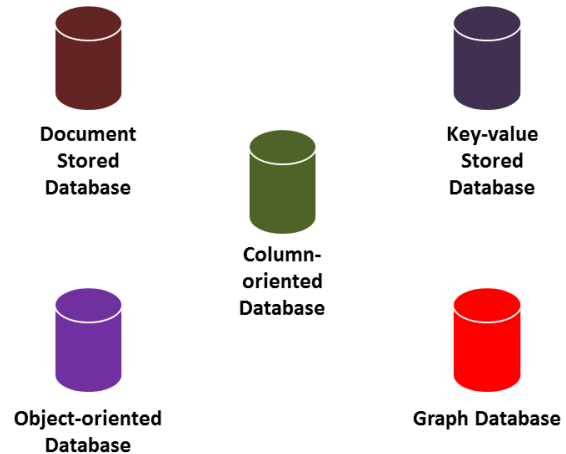


Fig.3. NoSQL Databases

Therefore, the following NoSQL databases' skills are available.

**Understanding Document Stored Database:** This skill involves understanding databases that store data in the form of documents. These documents are encoded in a standard data exchange format such as JSON (Javascript Option Notation), XML, and BSON (Binary JSON) in addition to PDF formats and many other layouts.

Examples of Document Stored Databases are MongoDB and CouchDB.

**Grasping Key-value Stored Database:** This skill involves understanding databases that store data in a key-value format. In practice, the data in this model is composed of two portions: 1) a string, which represents the key and 2) the actual data, which is referred as value.

Examples of Key-value Stored Databases are Amazon DynamoDB and Riak.

**Understanding Column-oriented Database:** This skill deals with understanding databases that employ a distributed, column-oriented data structure, which accommodates multiple attributes per key. Therefore, this model provides high scalability of data storage.

Examples of Column-oriented Databases are BigTable, HBase, and Cassandra.

**Grasping Object-oriented Database:** This skill has to do with databases that store data or information in an object format.

An example of an Object-oriented Database is db4o (database for objects).

**Understanding Graph Database:** This kind of skill deals with databases that allow users to store data in the form of a graph. Graph databases provide schema free and efficient storage of semi-structured data.

An example of a Graph Database is Neo4j.

## V. HADOOP ECOSYSTEM PROFICIENCY

Having proficiency in one of the pieces of the Hadoop Ecosystem is recommended for working with big data environments running under the Hadoop technology.

In that sense, Hadoop runs a distributed storage and computation framework across a cluster of computers. In fact, Hadoop is an open source environment licensed by Apache Software Foundation and it was written in Java programming language.

Similarly, Hadoop was intended to scale up from a single server to thousands of computers in where each machine offers local computation and storage. Likewise, it allows storing and processing big data in a distributed environment across clusters of machines.

Therefore, the following skills are available for those professionals using Hadoop.

**MapReduce:** This skill deals with managing the computational level of Hadoop. MapReduce is a parallel programming model for developing distributed applications, originally created at Google for managing large amounts of data with multi-terabytes and clusters with thousands of nodes.

**HDFS:** This skill has to do with managing the storage level of Hadoop. HDFS is based on the Google File System (GFS) and offers a distributed file system that is intended to run on commodity hardware.

**YARN Framework:** This type of skill deals with operating the framework for job scheduling and cluster resource management [12]. YARN stands for Yet Another Resource Negotiator.

**Zookeeper:** This skill has to do with managing the distributed coordination service, which provides distributed locks for building distributed applications [13].

**Oozie:** Represents the skill used for managing a service in charge of running and scheduling workflows of Hadoop jobs [13].

Oozie is the “Hadoop task management” [14].

**Hive:** This skill deals with managing the distributed data warehouse “Hive.” As a matter of fact, Hive provides ad-hoc query, data summarization, and analysis of big datasets. In that sense, Hive uses an SQL-like language named HiveQL (HQL).

Likewise, Hive simplifies the integration between Hadoop and business intelligence and visualization tools [15].

**Pig:** This type of skill has to do with managing the Pig data flow language and execution environment for exploring big datasets. Correspondingly, Pig operates on HDFS and MapReduce clusters [13].

Pig is a “data flow scripting language for analyzing unstructured data” [12].

**HBase:** While HBase is a Column-oriented Database

running on top of HDFS [13]; in this paper, we are including HBase as part of the Hadoop Ecosystem skills.

Though, it is important to point out that this type of skill is used for understanding the application and management of HBase from the Hadoop perspective and operation.

Likewise, such skill allows you to understand the technical integration and functionality of HBase with Hadoop.

**Mahout:** This skill embraces machine-learning libraries (such as classification and clustering algorithms) on top of Hadoop [13].

**Sqoop:** Represents a skill used for transferring bulk data between structured data stores (such as relational databases) and HDFS [13]. Similarly, Sqoop imports data from external sources either directly into HDFS or into systems like Hive and HBase. Likewise, Sqoop is also used to get data from Hadoop and export it to external artifacts such as relational databases and data warehouses [15].

Sqoop is used for “ingesting data from external datastores” [14].

**Spark:** This skill is oriented toward managing the Spark technology, which is a general-purpose data processing engine used for producing interactive queries across big data sets, treating streaming data coming from sensors & financial systems, and machine learning tasks. While Spark is usually used alongside with HDFS, it could be integrated with some of the NoSQL products. In a nutshell,

Spark is the “data processing engine for clusters computing” [14].

**Flume:** This skill embraces Flume, which is a service for gathering, aggregating, and moving big amounts of log data. In that sense, Flume allows log data coming from several sources, such as web servers [15].

Flume is a “tool for ingesting streaming data into your cluster” [14].

**Drill:** This skill deals with managing Drill, which is a low latency query engine developed for structured, semi-structured, and nested data [14].

## VI. FOREFRONT TOOLS PROFICIENCY

While big data tools are very broad, it is important to point out the rapid growth of some of the tools used for interacting with the big data technology. Therefore, learning one of the following tools is recommended for those professionals working in big data environments.

**Programming Languages:** These skills are related to programming languages such as Java, C, C++, C#, Python, and Scala.

**IBM Watson Analytics:** Mastering IBM Watson Analytics is a skill that provides you with cutting-edge and sophisticated analytics without the complexity of computer programming. IBM Watson Analytics is structured into the following four options: explore, predict, assemble, and refine. Due to the fact that many companies are using IBM Watson Analytics as the first step before they try their analytics with other solutions, this is a recommended skill to have.

**Statistical Tools:** Suggested skills for big data professionals that are required to perform computational statistics include tools such as IBM SPSS, R, SAS, Matlab, and Stata.

**SQL:** Structured Query Language (SQL) is normally used as the standard relational database language.

While SQL is oriented toward relational database management systems, new query languages (based on SQL) have been developed for big data environments. An example of this is the previously mentioned language HQL. As a result, SQL is also a recommended skill to have.

VII. PROJECT MANAGEMENT PROFICIENCY

Having proficiency in project management, program management, and portfolio management are critical success factors for implementing big data solutions.

From the organizational project management point of view, we visualize the following three levels of knowledge and expertise in managing big data related technology projects: project, program, and portfolio.

Accordingly, Fig.4. displays the three levels of the Organizational Project Management Pyramid. However, these levels are linked to three particular skills (project management, program management, and portfolio management).

Nevertheless, it is important to highlight that implementing big data technology is a very complex effort. Therefore, project management, program management, and portfolio management are adopted to successfully accomplish such effort.

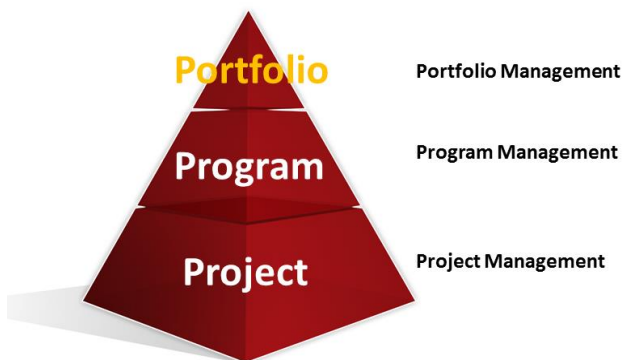


Fig.4. Project Management Pyramid

Consequently, the following skills are available for those professionals implementing big data technology.

**Project Management:** This skill deals with adopting project management for implementing big data. In that sense, project management from the big data perspective is the temporary organization that you put together in order to scope, plan, monitor, control, and deliver a big data solution within the constraints of time, cost, scope, big data technology, and quality.

Fig.5. displays a graphical representation of the Big Data Project Constraints. In that sense, this diamond shape has the following sides: time, cost, scope, and big data technology, having quality in the middle. Consequently, the project manager will be responsible for juggling with time, cost, scope, and big data technology around quality. Therefore, in this paper, we have called this model the diamond constraints of big data. This paradigm encapsulates what the project manager needs to do in order to achieve success in the execution of a big data project.

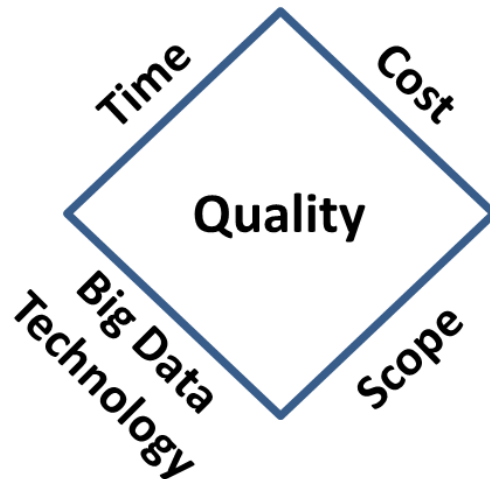


Fig.5. Big Data Project Constraints

Additionally, project management requires the management of risks, issues, and scope changes.

Likewise, project management embraces the following skills: project governance, benefits management, human resource management, communications management, project implementation lifecycle management, stakeholder management, and purchasing management.

**Program Management:** Depending on the scope of work, it is possible going one step further and organize the big data work into a program. In this case, the work is structured into projects, having project managers assigned to each constitute project.

Nevertheless, you will need a program manager to coordinate all the project managers, along with the interdependency of the constitute projects and the realization of benefits.

In summary, this skill has to do with adopting program management for implementing the big data work.

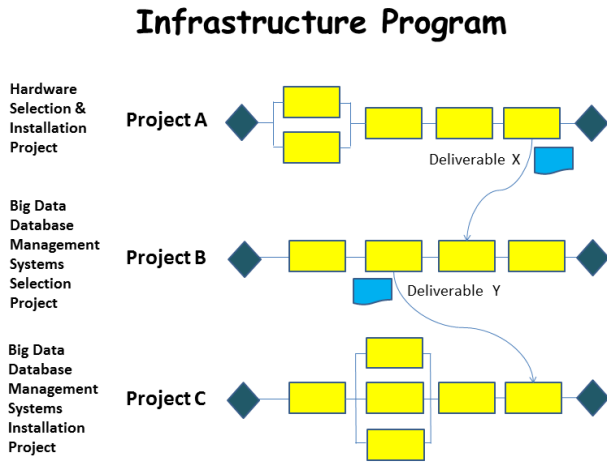


Fig.6. Interdependency of Constitute Projects

Fig.6. depicts a chart of the Interdependency of Constitute Projects corresponding to the infrastructure program. In this example, Project A (Hardware Selection & Installation Project) sends the Deliverable X to Project B (Big Data Database Management Systems Selection Project). Likewise, Project B (Big Data Database Management Systems Selection Project) sends the Deliverable Y to Project C (Big Data Database Management Systems Installation Project). In that sense, the program manager is responsible for coordinating the interaction of the interfaces (interdependencies) of the constitute projects of the program.

**Portfolio Management:** This skill deals with adopting portfolio management for selecting the appropriate big data projects and/or programs.

Since project management and program management have to do with doing the big data projects and/or programs right. Portfolio management deals with selecting the right big data projects and/or programs.

Consequently, with portfolio management you can select, prioritize, monitor, control, and authorize project and/or program investments.

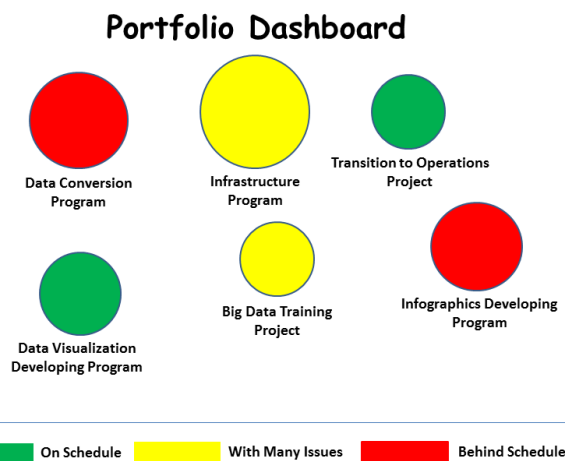


Fig.7. Big Data Portfolio Dashboard

Fig.7. displays a sample of the Big Data Portfolio Dashboard. In that sense, this dashboard includes the following investments: Data Conversion Program; Data Visualization Developing Program; Infrastructure Program; Big Data Training Project; Transition to Operations Project; and Infographics Developing Program.

In this example, the size of the bobble indicates the investment size of the project or program. If the bobble is in green then the project or program is on schedule. If the bobble is in yellow then the project or program has many issues that need to be solved right away. Finally, if the bobble is in red then the project or program is behind schedule.

### VIII. ABOUT DEVELOPING BIG DATA SKILLS

Adapting the work performed by the OECD on Better Skills, Better Jobs, Better Lives: A Strategic Approach to Skills Policies [1], organizations can encourage and enable people to learn big data as follow:

**Promote Equity in Education and Training Opportunities:** The idea of this section is to highlight that people without proficiency in IT, statistical tools, basic computer programming, SQL fundamentals, big data tool essentials, basic analytics, and project management fundamentals, require going through a progressive learning process in order to raise their knowledge level on big data.

Therefore, education and training equality should be applied to all the affected stakeholders (people without proficiency and people with proficiency).

**Gather and Use Intelligence on the Demand for Skills:** As the technological environment and structure of the IT employment has changed, so has the demand for big data skills.

As a result, new technical and project management skills are desired than ever before.

Therefore, these new big data skills need to be identified, structured, and converted into applicable high-tech curricula and programs.

**Design Efficient and Effective Education and Training Programs:** Organizations have the option to design their own internal education and training programs for big data technology. In that sense, they link together the worlds of a workplace, technology, and learning.

Likewise, they can achieve a suitable learning environment by blending hard skills with soft skills as part of their learning approach. This is applicable to experience and young people. Therefore, developing an efficient and effective big data education and training solution is necessary for achieving success.

**Engage Partners in Designing and Delivering Big Data Education and Training Programs:** A second option available to organizations is to engage partners in designing and delivering the big data education and training programs.

In that sense, they should evaluate and select partners such as big data's management storage companies, hardware vendors, training companies, universities; and implementation consulting firms. As a result, suitable education and training programs will be available to learn big data.

**Raise the Quality of Education and Training:** Excellence in education and training is a key success factor for learning big data.

Therefore, organizations need to assess their own learning environment in order to identify their big data quality level.

If the quality level is not up to high-quality standards, then they need to raise the quality at all levels so that investment in skills improvement is increased.

**Ensure Sharing Cost and Investment in Learning:** Organizations can establish supporting learning mechanisms and invest in learning of their employees throughout their professional career.

Examples of this effort include the design of financial incentives and sharing cost absorption methods.

**Maintain a Long-term Perspective on Skills Development:** Even during economic crises, organizations should follow their long-term skills development plan.

This plan should be based on long-term cost/benefit ratios of investments. On these grounds, a strong case could be made in maintaining learning of big data skills.

**Hire Skilled Professionals:** To close big data skills gaps it might be necessary to recruit external big data skilled professionals.

Therefore, organizations should have in place entry mechanisms for such big data professionals.

**Establish University Scholarship Programs:** To develop big data skilled professionals, organizations should establish university scholarships for big data students. Such students should be awarded with scholarships and hired after graduation.

## IX. CONCLUSION AND RECOMMENDATIONS

We have completed an IT research on exploring the suitable skills required for working with big data technology.

Therefore, this paper was organized into the following

sections: Introduction; Introducing the Big Data Skills Model; Skill Perspectives; NoSQL Databases Proficiency; Hadoop Ecosystem Proficiency; Forefront Tools Proficiency; Project Management Proficiency; About Developing Big Data Skills; Conclusion and Recommendations.

As a result, we conclude that big data skills provide the necessary know-how for adopting, managing, and implementing big data technology.

Accordingly, the following ideas have emerged as a ground base for further research on big data skills: develop case studies of the application of project management to big data environments; study available professional certifications related with MongoDB, CouchDB, Amazon DynamoDB, Riak, BigTable, HBase, Cassandra, db4o, and Neo4j; develop a corporate training and education program for Document Stored Databases, Key-value Stored Databases, Column-oriented Databases, Object-oriented Databases, and Graph Databases; develop a skill model for big data technology management; develop a corporate training and education program for MapReduce, HDFS, YARN Framework, Zookeeper, Oozie, Hive, Pig, HBase, Mahout, Sqoop, Spark, Flume, and Drill; study new Big Data skills associated with infographic and data visualization; study available professional certifications for Java, C, C++, C#, Python, and Scala; prepare a corporate training program for IBM Watson Analytics; develop case studies of the application of program management to big data environments; study new statistical tools for big data; study new query languages (based on SQL) for big data environments; study available professional certifications related with MapReduce, HDFS, YARN Framework, Zookeeper, Oozie, Hive, Pig, HBase, Mahout, Sqoop, Spark, Flume, and Drill; develop case studies of big data skills application; develop a skill model for big data technology implementation; perform a comparison analysis of statistical tools; perform a comparison analysis of Hadoop pieces; study the integration of NoSQL databases with Hadoop; perform a comparison analysis of forefront tools for big data; conduct a comparison analysis of big data degree programs; study available big data training certification programs; evaluate new skills required for data science; and develop case studies of the application of portfolio management to big data environments.

## ACKNOWLEDGMENT

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

## REFERENCES

- [1] OECD, *Better Skills, Better Jobs, Better Lives: A Strategic Approach to Skills Policies*, OECD Publishing, 2012.
- [2] F. Green, *What is Skill? An Inter-Disciplinary Synthesis*, Centre for Learning and Life Chances in Knowledge Economies and Societies, 2011.
- [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] The Economist, “The Data Deluge,” 2010.
- [5] N. Shamli and B. Sathiyabhama, “Parkinson’s Brain Disease Prediction Using Big Data Analytics,” *I. J. Information Technology and Computer Science*, vol. 6, pp. 73–84, 2016.
- [6] P. C. Srivastava, A. Agrawal, K. N. Mishra1, P. K. Ojha, and R. Garg, “Fingerprints, Iris and DNA Features based Multimodal Systems: A Review,” *I. J. Information Technology and Computer Science*, vol. 2, pp. 88–111, 2013.
- [7] X. Liu, X. Wang, S. Matwin, and N. Japkowicz, “Meta-MapReduce for scalable data mining,” *Journal of Big Data*, vol. 2, pp. 1–23, July 2015.
- [8] L. Einav and J. Levin, “The Data Revolution and Economic Analysis,” National Bureau of Economic Research, pp. 1–24, 2014.
- [9] V. Sharma and M. Dave, “SQL and NoSQL Databases,” *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, pp. 20–27, August 2012.
- [10] A. Nayak, A. Poriya, and D. Poojary, “Type of NOSQL Databases and its Comparison with Relational Databases,” *International Journal of Applied Information Systems*, vol. 5, pp. 16–19, March 2013.
- [11] R. Cattell, “Scalable SQL and NoSQL Data Stores,” *SIGMOD Record*, vol. 39, pp. 12–27, December 2010.
- [12] Tutorialspoint, HADOOP: Big data analysis framework, Tutorial Point (I) Pvt. Ltd., pp. 1–57, 2014.
- [13] T. White, *Hadoop: The Definitive Guide, Third Edition*, O’Reilly Media, pp. 1–629, 2012.
- [14] MapR Academy, *Hadoop Essentials*, Lesson 3: Hadoop Ecosystem, MapR Technologies, 2014.
- [15] Hortonworks, Apache Hadoop Basics, Hortonworks Inc., pp. 1–16, 2013.

**How to cite this paper:** Luis Emilio Alvarez-Dionisi, “Envisioning Skills for Adopting, Managing, and Implementing Big Data Technology in the 21<sup>st</sup> Century”, *International Journal of Information Technology and Computer Science (IJITCS)*, Vol.9, No.1, pp.18-25, 2017. DOI: 10.5815/ijitcs.2017.01.03

## Authors’ Profiles



**Luis Emilio Alvarez-Dionisi** received his Ph.D. in Strategy, Programme & Project Management from SKEMA Business School, Lille, France. He earned his MS in Computer Information Systems from Boston University, Boston, Massachusetts, USA. Likewise, he 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.