2014 International Conference on Future Internet of Things and Cloud

From Big Data to Big Projects: a Step-by-step Roadmap

Hajar Mousanif
LISI Laboratory, FSSM
Cadi Ayyad University
Marrakesh, Morocco
mousannif@uca.ma

Hasna Sabah, Yasmina Douiji, Younes Oulad Sayad
OSER research team, FSTG
Cadi Ayyad University
Marrakesh, Morocco
{hasna.sabah; yasmina.douiji; younes.ouladsayad}
@ced.uca.ma

***Abstract* – while technologies to build and run big data projects have started to mature and proliferate over the last couple of years, exploiting all potentials of big data is still at a relatively early stage. In fact, building effective big data projects inside organizations is hindered by the lack of a clear data-driven and analytical roadmap to move businesses and organizations from an opinion-operated era where humans skills are a necessity to a data-driven and smart era where big data analytics plays a major role in discovering unexpected insights in the oceans of data routinely generated or collected. This paper provides a solid and well-founded methodology for organizations to build any big data project and reap the most rewards out of their data. It covers all aspects of big data project implementation, from data collection to final project evaluation. In each stage of the process, we introduce different sets of platforms and tools in order to assist IT professionals and managers in gaining a comprehensive understanding of the methods and technologies involved and in making the best use of them. We also complete the picture by illustrating the process through different real-world big data projects implementations.**

Keywords: big data, advanced analytics, big data project, big data technologies.

I. Introduction

Every time we visit a website, “like” or “follow” a social page, and share our experiences, thoughts, feelings, and opinions on the Internet, we make already “big” data even “bigger”! Every day we collectively generate mountains of data that is waiting to be processed and analyzed. As an example, almost 500 terabytes of data is uploaded each day to Facebook servers [1], while Youtubers upload 100 hours of video every minute [2], and over 571 new websites are created every minute of the day [3]. Yet, using big data is not about collecting or generating massive amounts of data, but more about making sense of it. In fact, big data is absolutely worthless if it is not actionable and most importantly smart. Hence, what companies and organizations gather about customers, suppliers, transactions, and such, may be of no use if no insights are smartly and timely extracted from it.

How to establish an effective and rewarding big data solution is the major concern of any company or organization willing to embark on the big data adventure. Throughout this paper, we will show how business leaders and directors can leverage their data in the most efficient way possible through a clear and analytical methodology where descriptive, inquisitive, predictive and prescriptive analytics enter into action to improve results, support mission-critical applications, and drive better decision-making. While doing so, this paper attempts to provide satisfying answers to the following fundamental questions:

- Where does big data come from?
- What is/are the appropriate system(s) to capture, cure, store, explore, share, transfer, analyze, and visualize data?
- What is the size range of data and its implications in term of storage and retrieval?
- How could big data be used to determine market opportunities and seize them? And how could it contribute in making forecasts?
- And finally, how to take into account people’s expectations of privacy and bake it in advance into the big data project design?

The remainder of this paper will be organized as follows: Section 2 introduces some existing methodologies for implementing big data projects in today’s enterprise, and highlights our contribution. In section 3, we describe a clear roadmap for building smart and effective big data projects within organizations and illustrate the stages of the process through different sets of platforms and tools, as well as real-world big data projects use cases. Conclusions and directions for future work are given in section 4.

II. Related work

A recent survey of Gartner showed that companies are now more aware of the opportunities offered by analyzing larger amounts of data and are increasingly investing or planning to invest in big data projects, from 58% in 2012 to 64% last year [4]. This trend is accompanied by an increase in the need of global model or roadmap, that assist IT departments not only in implementing a Big data project, but in making the best of it in order to meet business objectives.

The Gartner approach in [5] introduces a roadmap to succeed big data solutions adoption, starting from the stage of company unawareness of the necessity of Big data in facing today’s business objectives, to the final stage of data-driven enterprise.

The other example is that of the U.S. Census Bureau, which has implemented a big data project to conduct a head count of all the people in the U.S. The life cycle of the US

978-1-4799-4357-9/14 $31.00 © 2014 IEEE
DOI 10.1109/FiCloud.2014.66

census bureau big data project includes three fundamental steps: 1°) Data collection using a multi-mode model, 2°) Data analysis to explore technology solutions based on methodological techniques, and 3°) Data dissemination by implementing new platforms for integrating census and survey data with other Big data [6].

In [7], ASE consulting provides a well-considered approach for building big data projects and which consists of six steps that are not committing to any particular technology or tool, ranging from understanding the scope of the project by identifying business problems and opportunities, to evaluating the big data project while providing insights into what worked well and what did not.

IBM in their recent report [8] introduced a three-phases approach for building big data projects: planning, execution and post-implementation, and which mainly consists in understanding the business and legal policies, communication between IT departments and the project stakeholders, and conducting an impact analysis at the end of the implementation.

With respect to all related literature presented above, the existing efforts either fail to cover some fundamental aspects of big data project setup, or limit their approach to providing basic guidelines for big data projects implementation without further insights into the technologies and platforms involved. The present work comes to overcome such limitations by:

- Providing a holistic approach to building big data projects, which tackles all implementation challenges a company or organization may face in each stage of their big data project setup: from strategy elaboration, to final project evaluation.
- Assisting companies and organizations, willing to establish an effective and rewarding big data solution, in gaining a comprehensive understanding of the technologies involved and in making the best use of them.
- Baking in advance people's expectations of privacy and security into the big data project design.
- Illustrating the proposed process through different real-world big data projects implementations.

III. ROADMAP FOR BUILDING SMART BIG DATA PROJECTS

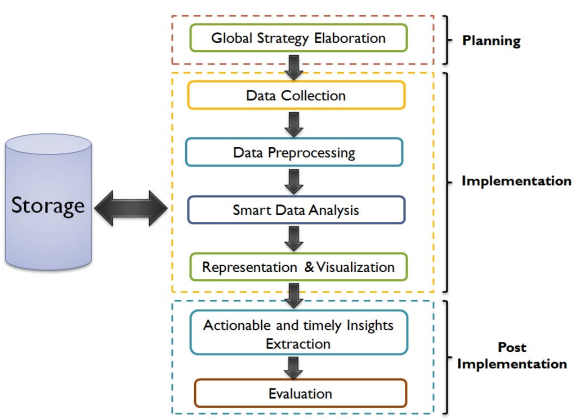

Fig. 1. Big data project workflow

In this section, we explore the design of the proposed methodology and provide a set of useful tips to follow in each phase of the big data project setup. The suggested approach, as shown in Figure1, consists of three major phases: 1°) elaboration of the global strategy, 2°) implementation of the project, and 3°) post-implementation.

A. Global Strategy Elaboration

Starting a big data project requires big changes and new investments. The changes particularly include the establishment of a new technological infrastructure, and a new way to process and harness data. Here are a few points to consider before undertaking any changes:

1) Why a big data project?

To answer this question, companies have to find the problems that need a solution, and decide whether they could be solved using new technologies or just with available software and techniques. Those problems could be: volume challenges, real time analytics, predictive analytics, customer-centric analytics, among others. The other thing to consider is to define business priorities, by focusing on the most important activities that form the greatest economic leverage in the business.

2) What data should the organization consider?

Once priorities are defined, business leaders and IT practitioners must target the data that will yield most value. This important phase is defined by IBM as Data exploration [9], as it is about exploring both internal and external data available to the company, to ensure that it can be accessed to sustain decision making and everyday operations. IBM InfoSphere Data Explorer [10] is one example of tools that allows to perform such a task. It allows federated discovery, navigation and inquiry over a wide extend of data sources and types, both inside and outside the organization, to help companies start big data initiatives.

3) How to protect data?

Securing a huge amount of continuously evolving data can be very complicated, considering that firms' servers cannot store all the needed data. Moreover, the fact that big data is most of the time processed in real-time induces even more security challenges. The Cloud Security Alliance (CSA), is one of the few organizations that took care of this issue by providing, in their recent report [11], a set of solutions and methods to win every privacy or security challenge. Similarly, the Enterprise Strategy Group (ESG) shows in [12] the most significant obstacles facing the implementation of a security policy in a big data environment, and gives valuable tips for CIOs to enter the big data security analytics Era, whereby companies would not only be able to monitor the traffic coming into and getting out of their systems to detect threats, but also predict cyber-attacks even before they happen.

We consider that there are three main features to focus on when planning to implement a new security management solution:

- Protection of sensitive data: by controlling the access to the data or by providing encryption solutions or both. The chosen solution for this purpose must be easy to integrate within the current system and consider performance issues. Vortmetric Encryption [13] and Voltage [14] are examples of such solutions.

- Network security management: by monitoring the local network, analyzing data coming from security devices and network end-points, timely detecting intrusions and suspicious traffic and reacting to it, without impeding the main objective of the big data project. Among available platforms for this purpose, we find LogRhythm SIEM 2.0 [15] and Fortscale [16].
- Security intelligence: by providing actionable and comprehensive insight that reduces risk and operational effort for any size organization using data generated by users, applications and infrastructure. InfoSphere Streams and InfoSphere BigInsights [9] are some of big data technologies that offer security intelligence features.

4) What to avoid?

Below are the most common mistakes companies may make whether before or while undertaking a big data initiative:

- Technology is not the goal of a big data project, it is rather a mean to be seriously thought about once business objectives are fixed.
- There is no ever-lasting technological solution for implementing the whole cycle of a big data project. As big data solutions proliferate, it becomes difficult to predict which platforms, applications or methods will better work in the future. Hence, companies should stay open to any new big data solution.
- Avoid the warehouse-or-Hadoop trick, it is imperative to use both of them, as they work well alongside and complement each other.

B. Data collection

Data collection is the first technical step of a big data project setup. This section will shed light on some major big data sources. We will cover sensors, open data, social media and crowdsourcing.

1) Internet of things

Sensors are a major source of big data. They are increasingly deployed everywhere: smart phones and other daily life devices, commercial buildings or transportation systems. With an expected population of 1 trillion by 2015 [17], sensors allow collecting various types of data including: body related metrics, location, movement, temperature and sounds. Coupled with ubiquitous wireless networks, sensors are driving myriad of smart innovations in the context of Internet of Things (IoT), for example: smart buildings where lightening and air-conditioning are optimized, smart transportation and traffic management system that monitor both vehicular and pedestrian traffic for better flow and better evacuation in emergencies [18], and smart phones that automatically recognize our emotional states and appropriately respond to them [19].

2) Open data

Public institutions, organizations and a growing number of private companies are making some of their datasets available for public. A major contributor to open data initiative is the World Bank [20]. The catalog includes macro, financial and sector databases. In addition to searching datasets on the World Bank site and downloading tables, users can also access the data via different APIs [21].

The open data catalog [22] maintains a list of worldwide open data and shows an increase in governments' contribution. A number of companies are also making parts of their data available through download or via APIs like Yelp [23] which gives academics access to its business rating database.

3) Social networks

Social networks are storing huge amounts of data. This data is generally accessible via proposed APIs or through special grants. For example, Twitter proposes its own APIs [24] that give access to fractions from all tweets. The company also established a certified partners program in 2012: partners, such as Gnip [25] are given a deeper access to twitter data, which they process to offer custom datasets and services. Finally, Twitter announced in 2014 a data grant project that will give selected research institutions access to its public and historical data [26].

4) Crowdsourcing

Collecting massive amounts of data can be quite challenging, especially when it has to be done in a large scale. Crowdsourcing is a great solution for data collection and emerged in the last decade as an efficient way to harness the creativity and intelligence of crowds. Recently, researchers sought to apply crowdsourcing to human subject research [27]. Technical University Munich (TUM)'s ProteomicsDB and the International Barcode of Life projects are two good examples of collecting and gathering data using crowdsourcing [28]. Amazon Mechanical Turk (Mturk) [29] is one crowdsourcing framework among others in which assignments are distributed to a population of many unknown workers for fulfillment.

C. Data preprocessing

It is important to lay the ground for data analysis by applying various preprocessing operations to address various imperfections in raw collected data. For example, data can have different formats as multiple sources might be involved. It can also contain noise (errors, redundancies, outliers and others). Finally, it may simply need to fit requirements of analysis algorithms. Data preprocessing includes a range of operations [30]:

- Data cleaning eliminates the incorrect values and checks for data inconsistency.
- Data integration: combines data from databases, files and different sources.
- Data transformation: converts collected data formats to the format of the destination data system. It can be divided into two steps: 1°) data mapping which links data elements from the source data system to the destination data system, and 2°) data reduction which elaborates the data into a structure that is smaller but still generates the same analytical results.
- Data discretization: can be considered as part of the data reduction, yet it has its own particular importance, and it refers to the process of partitioning or converting constant properties, characteristics or variables to nominal variables, features or attributes.

Most data mining and business intelligence platforms include data preprocessing tools such as the open source WEKA [31][31] and Data cleaner for Pentaho [32].

D. Smart data analysis

Extracting value from a huge set of data is the ultimate goal of many firms. One efficient approach to achieve this is to use advanced analytics, which provides algorithms to perform complex analytics on either structured or unstructured data. There are four types of advanced analytics:

- Descriptive analytics: answers the question: *what happened* in the past? Knowing that in this context, the past could mean one minute ago or a few years back [33]. It uses descriptive statistics such as sums, counts, averages, min, and max to provide meaningful results about the analyzed dataset. Descriptive analytics is typically used in social analytics and recommendation engines, such as Netflix recommendation system.
- Inquisitive analytics: also called diagnostic analytics, it answers the question *why something is happening*? By validating or rejecting business hypotheses, Explorys, a leader in healthcare big data, used inquisitive analytics to find out that an unexplained variation in the evaluation of patients' weight was mainly due to some documentation gap [34].
- Predictive analytics: consists in studying the data we have, to predict data we do not have, such as future outcomes, in a probabilistic way [35], answering thereby the question "what is likely to happen?".
- Prescriptive analytics: or Optimization analytics consists in guiding decision making by answering the question "*so what?*" or "*what we must do now?*" It can be used by companies to optimize their scheduling, production, inventory and supply chain design [33].

To guide companies in their software analytics choice Gartner published the first magic quadrant (MQ) for advanced analytics [36], which presents 16 analytics platforms divided into four areas: leaders, challengers, visionaries and niche players, based on two criteria, which are the completeness of vision and the ability to execute. The three top leaders of Gartner MQ are IBM [37], SAS Visual analytics [38] and Knime [39].

E. Representation and Visualization

Visualization guides the analysis process and presents results in a meaningful way. For the simple depiction of data, most software packages support classical charts and dashboards. The choice is generally dictated by the type of desired analytics [40]. Working at big data scale brings multiple technical issues [41]. Firstly, there are challenges related to the volume such as processing time, memory limitations, and the need to fit different display types. Different approaches are explored to scale to big data, for example at Intel Science & Technology Center for Big Data, projects work on various techniques such as visual summaries, caching and prefetching to hide data store latency, query steering and large scale parallelism [42].

Secondly, deriving meaning from semi-structured and unstructured data requires adequate visualizations that are variably supported by software. Examples are word clouds, association trees, and network diagrams.

Concerning the market of data visualization platforms, a study by Forrester Inc [43] highlights market's diversity and the importance of both technology and visual design quality. It also shows that main technical differentiation factors are the performance of the in-memory engine, the quality of the graphical user interface, and the comprehensiveness of data exploration and discovery tools. Leaders' board includes Tableau Software, Tibco Spotfire and SAS BI. A white paper by Tableau Inc [44] identifies, as shown in TABLE 1, seven key features to assess a visual analytics application.

TABLE I. KEY ELEMENTS OF VISUAL ANALYTICS APPLICATION.

Key elements	**Description & importance**	
	Description	***Importance***
Comprehensive visual exploration process	Integrates data exploration and querying	Ease of use and better focus
Augmented human perception	Effective visual properties and well designed graphics	Fostering visual thinking
Visual expressiveness	Depth, flexibility and multi-dimensional expressiveness	Simple displaying of complex problems
Automatic visualization	Automatic suggestion of effective visualizations	Assisting the analysis process
Visual perspective shifting	Easily shifting between alternative visualizations of data	Assisting the analysis process
Visual perspective linking	Visual correlation of information in different visualizations	Assisting the analysis process
Collaborative visualization	Ease of sharing and distributed collaboration	Enhancing productivity

F. Actionable and timely insights Extraction

Many sectors have already grasped big data regardless of whether the information comes from private or open sources. Below are examples of extracted insights from big data, illustrated through successful big data projects implementations:

- *Manufacturing:* Companies specialized in consumer products or retail organizations are using social networks to get insight into customer behavior, preferences, and product perception. Duke Energy is a case in point [45].
- *Finance:* The Oversea-Chinese Banking Corporation (OCBC), as an example, analyzes customer data to

determine customers' preferences. It designed an event-based advertising procedure that focused on a large volume of coordinated and customized marketing communications over numerous channels [46].

- *Health:* An example of the use of big data in health-care is a drug company in Seattle which uses data related to the genetics of cells to test cancer drug effectiveness [47].

G. Evaluation of big data projects

To evaluate a Big data project, we need to consider a range of diverse data inputs, their quality, and expected results. To develop procedures for Big Data evaluation, we first need to answer the following questions [48]:

- *Does the project allow stream processing and incremental computation of statistics?* To get answers in real time about what is happening in the business, data streams are required. Examples of technologies for queuing data streams include TIBCO [49], ZeroMQ [50] and Esper [51].
- *Does the project parallelize processing and exploit distributed computing?* Working with distributed data requires distributed processing, so that data will be processed in a reasonable amount of time.
- *Does the project easily integrate with visualization tools?* Once the implementation of the project is done, it must be able to integrate multiple visualization tools.
- *Does the project perform summary indexing to accelerate queries on big datasets?* Summary indexing is the procedure of making a pre-calculated summary of data to accelerate running queries.

H. Storage

Companies handling large amounts of data, such as Google and Amazon, were first to experience the limitations of traditional database management systems. The accelerated growth in data size requires horizontal scaling, which is the ability to extend the database over additional servers. But it turns out that managing sharding and replication with SQL databases is difficult and slow, as separate applications are required to handle these tasks. Furthermore, managing rapidly changing data needs greater flexibility in schema definition that is not available in classical databases, where structure and data types must be defined in advance. Finally, unstructured data is poorly supported and the transaction mode can penalize performance for some big data analysis.

Several alternatives have been developed and are loosely grouped within the Nosql family [52]. These products are dissimilar as they fit different needs but most natively support horizontal scaling thanks to automatic replication and auto-sharding. They also support dynamic schemas allowing for transparent real-time application changes.

On another aspect, distributed massive storage naturally raises computational problems. Google MapReduce paradigm [53] is an efficient solution that distributes extremely large problem across extremely large computing cluster, allowing programs to automatically parallelize. Most NoSQL databases adopted the MapReduce model and one of the most popular open source implementation of MapReduce is Hadoop [54].

Hadoop was originally created at Yahoo, by building upon Google MapReduce and Google File System papers. It is now a comprehensive framework for big data projects, maintained by the Apache Software foundation. Hadoop is essentially a batch processing system but can integrate with other projects for real-time analysis such as Apache Storm.

Many vendors offer customized and extended Hadoop distributions for specific needs, according to an assessment published by Forrester research [55] that covers nine Hadoop solutions. Top three leaders are Amazon Web Services, Cloudera, and HortonWorks. There are also solutions outside Hadoop ecosystem, such as Spark and Disco which are appreciated by engineers for their speed and real-time support [56].

Choosing the right solution must take into consideration company's needs, required technical competence and the characteristics of existing solutions. Decisions then should be made in capacity planning, infrastructure, tradeoffs and complexity of various processes [57]:

- Impact on existing infrastructure: as companies generally have an existing storage infrastructure, it must be decided whether the big data solution will replace it or complement it.
- Capacity planning: it is important to decide on potential required hardware, storage types and network configurations.
- Tradeoffs: the company must align its project priorities with the characteristics of big data solutions by deciding on tradeoffs between availability, consistency and partition tolerance, and between scalability, elasticity and availability.
- Latency: whether the company needs batch processing of data or real-time analysis has different impact in term of infrastructure.
- Complexity: there should be a clear vision on the flow and potential difficulties of the deployment, maintenance, monitoring and management processes.

IV. Conclusion and Future Work

In this paper, we presented a clear and step-by-step roadmap that covers the whole life cycle of a big data project setup. We tried to provide answers to some fundamental questions any company, willing to embark on the big data adventure and reap the most rewards out of its data, would ask. Future work will mainly consist in applying the proposed methodology to build a big data project within our institution and which mainly consist in using reality mining to engineer a smarter healthcare system.

References

[1] "Facebook is collecting your data — 500 terabytes a day — Tech News and Analysis." [Online]. Available: http://gigaom.com/2012/08/22/facebook-is-collecting-your-data-500-terabytes-a-day/. [Accessed: 24-Feb-2014].

[2] "Statistics - YouTube." [Online]. Available: http://www.youtube.com/yt/press/statistics.html. [Accessed: 24-Feb-2014].

[3] "The Big List of Big Data Infographics". [Online]. Available: http://wikibon.org/blog/big-data-infographics/, visited: [Accessed: 24-Feb-2014].

[4] S. Sicular, "The Road Map for Successful Big Data Adoption," 2013. [Online]. Available: https://www.gartner.com/doc/2576226/road-map-successful-big-data. [Accessed: 23-Mar-2014].

[5] S. Sicular, "The Roadmap for Successful Big Data Adoption Twitter: @ Sve _ Sic," 2013.

[6] W. G. Bostic, E. Programs, and U. S. C. Bureau, "Is the Census Bureau working on Big Data projects? If so, how are these different from other projects?" 2013.

[7] C. King, "Navigating through the Big Data Space," pp. 1–12, 2013.

[8] K. C. Desouza, "Realizing the Promise of Big Data Implementing Big Data Projects" 2014.

[9] IBM Software, "The top five ways to get started with big data," 2013.

[10] IBM, "IBM - InfoSphere Data Explorer," 06-Jun-2014. [Online]. Available: http://www-03.ibm.com/software/products/en/dataexplorer. [Accessed: 02-Mar-2014].

[11] Cloud Security Alliance, "Expanded Top Ten Big Data Security and Privacy Challenges," 2013.

[12] J. Olstik and ESG, "the big data security analytics era is here," 2013.

[13] Vortmetric, "Big Data Security Intelligence Solution" [Online]. Available: http://www.vormetric.com/data-security-solutions/applications/big-data-security. [Accessed: 08-Mar-2014].

[14] Voltage Security, "Voltage Enterprise Security for Big Data, Solution brief" 2013.

[15] LogRythm, "Log & Event Management, Log Analysis." [Online]. Available: http://www.logrhythm.com/. [Accessed: 08-Mar-2014].

[16] Fortscale, "Big Data Security Analytics." [Online]. Available: http://www.fortscale.com/. [Accessed: 08-Mar-2014].

[17] G. P. Khanzode, "Insights Internet of Things: Endless Opportunities," 2012.

[18] H. Mousannif, I. Khalil, S. Olariu, (2012) 'Cooperation as a Service in VANET: Implementation and Simulation results', Mobile Information Systems Journal, Vol. 8 No. 2, pp.153-172

[19] H. Mousannif, I. Khalil (2014), The human face of mobile, ICT-EurAsia, Lecture Notes in Computer Science, vol. 8407, pp. 1–20

[20] The World Bank, "Open Data Catalog" [Online]. Available: http://datacatalog.worldbank.org/. [Accessed: 18-Mar-2014].

[21] The World Bank, "the world bank for Developers." [Online]. Available: http://data.worldbank.org/developers. [Accessed: 18-Mar-2014].

[22] Datacatalogs, "datacatalogs.org." [Online]. Available: http://datacatalogs.org/. [Accessed: 18-Mar-2014].

[23] Yelp, "Yelp's Academic Dataset." [Online]. Available: https://www.yelp.com/academic_dataset. [Accessed: 18-Mar-2014].

[24] Twitter Developpers, "History of the REST & Search API." [Online]. Available: https://dev.twitter.com/docs/history-rest-search-api. [Accessed: 16-Mar-2014].

[25] Gnip, "The Source for Social Data." [Online]. Available: http://gnip.com/. [Accessed: 16-Mar-2014].

[26] Raffi Krikorian, "Introducing Twitter Data Grants". (February 5, 2014). [Online]. Available: https://blog.twitter.com/2014/introducing-twitter-data-grants. [Accessed: March 16, 2014].

[27] L. A. Schmidt, "Crowdsourcing for Human Subjects Research," 2010.

[28] STRATA, "Collecting Massive Data via Crowdsourcing: Strata 2014 - O'Reilly Conferences, February 11 - 13, 2014, Santa Clara, CA," 2012. [Online]. Available: http://strataconf.com/strata2014/public/schedule/detail/33541. [Accessed: 05-Feb-2014].

[29] J. Ross, A. Zaldivar, L. Irani, and B. Tomlinson, "Who are the Turkers? Worker Demographics in Amazon Mechanical Turk," pp. 1–5, 2010.

[30] S. B. Kotsiantis, D. Kanellopoulos, and P. E. Pintelas, "Data Preprocessing for Supervised Leaning," vol. 1, no. 2, pp. 111–117, 2006.

[31] S. Singhal and M. Jena, "A Study on WEKA Tool for Data Preprocessing, Classification and Clustering," no. 6, pp. 250–253, 2013.

[32] DataCleaner, "Reference documentation." [Online]. Available: http://datacleaner.org/resources/docs/3.5.10/html/. [Accessed: 19-Mar-2014].

[33] BigData-startups, "Descriptive, Predictive and Prescriptive Analytics Helps Understanding Your Business," 2013. [Online]. Available: http://www.bigdata-startups.com/understanding-business-descriptive-predictive-prescriptive-analytics/. [Accessed: 12-Mar-2014].

[34] Explorys, "Explorys categorizes analytics as descriptive, diagnostic, predictive and prescriptive," 2013. [Online]. Available: https://www.explorys.com/results/news-results/2013/12/09/explorys-categorizes-analytics-as-descriptive-diagnostic-predictive-and-prescriptive. [Accessed: 15-Mar-2014].

[35] W. Mike, "Big Data Reduction 2: Understanding Predictive Ana... - Lithium Community," 2013. [Online]. Available: http://community.lithium.com/t5/Science-of-Social-blog/Big-Data-Reduction-2-Understanding-Predictive-Analytics/ba-p/79616. [Accessed: 12-Mar-2014].

[36] Gartner. "Magic Quadrant for Advanced Analytics Platforms", 2014. [Online]. Available: https://www.gartner.com/doc/2667527. [Accessed: 16-Mar-2014].

[37] IBM, "Get to Know the IBM SPSS Product Portfolio." 2013.

[38] SAS, "Sas visual analytics". [Online]. Available: http://www.sas.com/en_us/software/business-intelligence/visual-analytics.html. [Accessed: 17-Mar-2014].

[39] Knime, "Open for Innovation." [Online]. Available: http://www.knime.org/. [Accessed: 17-Mar-2014].

[40] D. Hom, R. Perez, and L. Williams, Tableau, "Which chart or graph is right for you?"

[41] SAS, "Data Visualization Techniques," 2013.

[42] Intel Science and Technolology Center for Big data, "Making Big Data Visualization More Accessible". [Online]. Available: http://istc-bigdata.org/index.php/making-big-data-visualization-more-accessible/. [Accessed: 24-Mar-2014].

[43] B. Evelson and N. Yuhanna, Forrestor, "The Forrester Wave ™: Advanced Data Visualization (ADV) Platforms, Q3 2012," 2012.

[44] P. Hanrahan, C. Stolte and J. Mackinlay. Tableau Software, "Selecting a Visual Analytics Application".

[45] Michele Nemschoff, "How Big Data Can Improve Manufacturing Quality", 04 January 2014. [Online]. Available: http://smartdatacollective.com/michelenemschoff/176666/how-big-data-can-improve-manufacturing-quality. [Accessed: 06-Apr-2014].

[46] IBM and Said Business School, "Analytics: The real-world use of big data in financial services".

[47] Alan S. Horowitz, "Let's Play Moneyball: 5 Industries That Should Bet on Data Analytics", 06 November 2013. [Online]. Available: http://plotting-success.softwareadvice.com/moneyball-5-industries-bet-on-analytics-1113/. [Accessed: 06-Apr-2014].

[48] M. Driscoll, "Big Data Technology Evaluation Checklist". [Online]. Available: http://www.forbes.com/sites/danwoods/2011/10/21/big-data-technology-evaluation-checklist/. [Accessed: 06-Apr-2014].

[49] TIBCO. [Online]. Available: http://www.tibco.com/. [Accessed: 06-Apr-2014].

[50] ZEROMQ. [Online]. Available: http://zeromq.org/. [Accessed: 06-Apr-2014].

[51] ESPER. "Event Series Intelligence: Esper & NEsper". [Online]. Available: http://esper.codehaus.org/. [Accessed: 06-Apr-2014].

[52] NOSQL Database. [Online]. Available: http://nosql-database.org/. [Accessed: 20-Mar-2014].

[53] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," pp. 1–13.

[54] Apache Hadoop. [Online]. Available: http://hadoop.apache.org/. [Accessed: 20-Mar-2014].

[55] M. Gualtieri and N. Yuhanna, "The Forrester Wave ™: Big Data Hadoop," pp. 4–12, 2014.

[56] Byte-minning, "Hadoop Fatigue -- Alternatives to Hadoop", 2011. [Online]. Available: http://www.bytemining.com/2011/08/hadoop-fatigue-alternatives-to-hadoop/. [Accessed: 10-Mar-2014].

[57] Open Data Center Alliance, "Big Data Consumer Guide," 2012.