An investigation of software management utilizing agile principles and big data analytics

Dr THARINI BENARJI
Professor in Department of CSE, Indur Institute of Engineering and Technology, siddipet, TS
Email: tharinibenarji@gmail.com

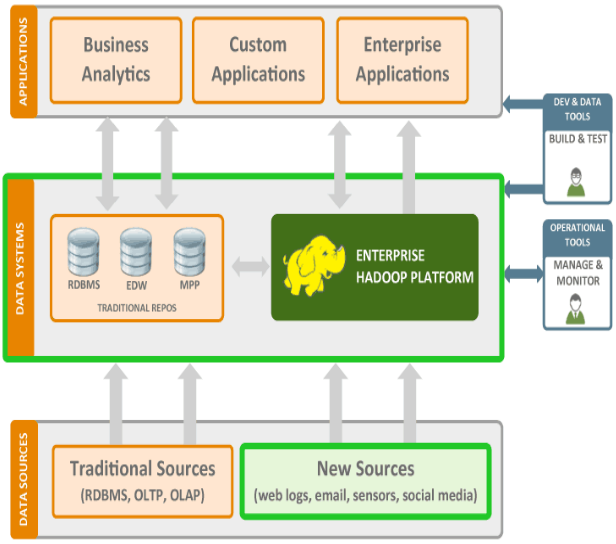
***Abstract*: Big data and big data analytics has become one of the important research frontiers. Big data and its emerging technologies including big data analytics have been not only making big changes in the way the e-commerce and e-services operate but also making traditional data analytics and business analytics bring new big opportunities for academia and enterprises. Big data analytics is an emerging big data technology, and has become a mainstream market adopted broadly across industries, organizations, and geographic regions and among individuals to facilitate data-driven decision making for business. The aim of this work is to document agile approach to project management and suggests ways of using it in projects related to Big Data management. User involvement is generally considered to contributing to user satisfaction and project success and is central to Agile software development. In theory, the expectations about user involvement, such as the PO's, are quite demanding in this Agile way of working. But what are the expectations seen in practice, and are the expectations of user involvement aligned among the development team and users.Any misalignment could contribute to conflict and miscommunication among stakeholders that may result in ineffective user involvement. The proposed approach in this paper might facilitate the research and development of business analytics, big data analytics, and business intelligence as well as intelligent agents.**

Index Terms—Big Data Analytics, Business Analytics-Commerce, Agile Software Development

I. INTRODUCTION

Big data is transforming the business landscapes and is "possibly the most significant 'tech' disruption in business and academic ecosystems since the meteoric rise of the Internet and the digital economy" . Big data analytics is the process of examining big data to uncover hidden patterns, unknown correlations and other useful information to provide "value"—the 5th V of big data—for organizations to increase operational efficiency, inform strategic direction, develop new products and services, identify new customers and markets, make better decisions, and become more innovative. Competitive pressures require rapid delivery of actionable predictions discovered from big data. Data scientists need support for value exploration, model development, testing and rapid deployment of new models while software engineers need to know how data scientists are going to use the data to design big data systems, orchestrating different types of storage, tools for different types of analytics, different types of visualizations, and so on.

Agile principles are being pursued for managing the risk associated with such challenges. Although agile analytics for traditional structured Relational data warehouse is already a familiar concept, agile development of big data analytics is relatively new and requires careful adaptation as big data analytics are dramatically different from "small" data analytics. Big data analytics requires capabilities beyond traditional relational data warehouses. Traditionally, a data warehouse integrates data from various operational data stores, through an ETL (extract, transform, load) process which cleans and transforms data and loads it into



pre-specified data cubes (based on star or snowflake schemas) for later manipulation by OLAP (Online Analytical Processing) tools, data mining, and visualization tools. The data is "at rest" and this process is batch-oriented.

Figure 1 : Traditional Vs Big Data Analytics

Thus, agile development for "small" data analytics projects concentrates on requirement analysis, ensuring that users get the features they want. The architectures of such systems are standardized and rarely change.

On the contrary, big data analytics focuses on value exploration and discovery, requiring integration of voluminous data of many types, often from unknown sources. As a result, data lake and data refinery concepts, for example, have emerged to replace or augment data warehouses. Raw data of various types from everywhere are ingested into a data lake in a high-capacity data store, such as Hadoop HDFS. A process called "schema on read" is employed when certain data are needed (data-in-use). For data-in motion or real time stream data, data are analyzed on the fly and then stored after actions have been taken for so-called "two second advantages" in the big data world. Many No SQL databases are available for storing unstructured data of various types: key value, column-oriented, document, graph, etc. Cloud storage is also a prevalent option. Many organizations still want to keep their enterprise data warehouses (EDW) and thus integrating the old with the new poses tremendous technical challenges. The emphasis on value exploration and rapid delivery of changing big data models necessitates agile principles to be applied to the entire development life cycle.

These challenges and examples just scratch the surface for agile big data analytics development. Our research questions rest on (1) how should a big data system be designed and developed to support advanced analytics effectively? and (2) how should agile processes be adapted for big data analytics development?

II. SOFTWARE DEVELOPMENT –AN AGILE APPROACH

What does the term "agile" mean this term is not in the context of software development clearly defined? The origin of the term of agility in manufacturing is defined as " the ability to successfully sell low-cost, high-quality products with short delivery times and in different capacities that by the adaptation to the needs of the user provide increased value for customers ". A key precondition for agility in software development is the freedom to adapt procedures and methods to the needs of a particular project." Success of practicing agile development is confirmed by the latest research State of Agile sponsored by Version One, which a supplier of business software in the field of agile is handling. Of the more than 3,500 respondents surveyed, mainly in the field of software development, up to 88% using agile development. An annual increase of 2 percent compared to the previous two years also reflects the growing popularity of agile approaches.

Agile methods are "methods that are trying to focus on the primary objective of effective software development, ie. The creation of working software (without defects)"

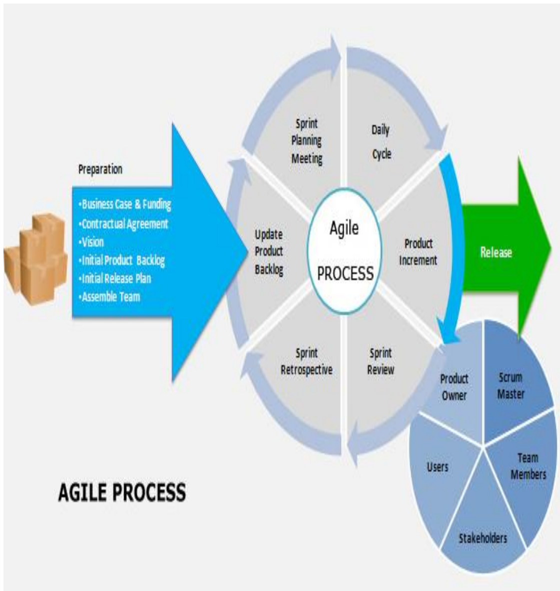


Figure 2: Transition to Agile Development

Effective software development means that we are developing the software "sufficiently" for a given situation, assuming that the requirements will vary The term "sufficient" thus indicates that "the development is not necessary to do anything more than is agreed with the customer and sufficient for a given iteration"."Agile centers on responding to change over following a plan, on what's the most important thing to you right now," said Abramson. "Agile practices can help developers bring real-time predictions, insight-driven optimization and unified views to big data reporting tools."Interest in agile is high, but our 2009 Agile Trends survey found that many organizations still follow traditional development practices. When asked which development processes their organizations employed, the number of respondents for agile had increased over the previous year, with 56% reporting they used Agile. More and more teams are switching to agile development.It's no longer just for innovators and early adopters, nor is it just for small projects. An agile methodology—such as Scrum—is a lighter weight approach to software development than many of the traditional approaches.

Agile methodologies feature self-organized teams that are empowered to achieve specific business objectives. Agile methodologies focus on rapid and frequent deliverables of partial solutions that can be evaluated and used to determine next steps.

In this way, solutions are built in an iterative and incremental manner. Agile methodologies have been shown to deliver higher quality products in less time, resulting in improved customer satisfaction.

III. AGILE ANALYTICS METHODOLOGIES IN BIG DATA

Big data analytics and management tools are commonly used today for business strategy planning, inside user-facing applications such CRM and ERP, and in support systems such as data warehouses and recommendation systems. Today, big data tools are bringing together information, both structured (as in SAP) and unstructured (think Twitter), to augment business leaders' knowledge about the company, their customers and how to do business well. Big data helps businesses understand how they communicate and consume in order to drive actions and choices in the future. Leadership and collaboration are the biggest hurdles to integration and innovation in big data technologies. Businesses need reports that effect change and are actionable to reach the golden mean, business value. High quality and effectiveness are hard to achieve in data analytics, largely because information and its sources are varied and often uncertified To get closer to that truth, business intelligence (BI) and data analytics have to be integrated with business processes.

Benefits of Agile Methodology

When applied to Big Data, agile brings collaboration to development and delivery. Agile methodology allows cross-functional teams in a company to work together to generate reliable insights into the company's customers so that the leadership can figure out how to address its biggest challenges in selling its products and services.Adopting agile methodologies in the context of big data allows the extraction of valuable information from available data in a precise and quick way. Essentially, the data is converted to knowledge that can be immediately used to make path-altering decisions for better business performance. With this approach, business leaders can integrate big data into the existing business strategy and give their company a competitive advantage.

The ideal agile methodology platform for big data should have the following capabilities:

1) Must be able to handle data in real time and in a continuous stream
2) Must have the capacity to dynamically construct and maintain several agile views of available data so as to satisfy the set application requirements
3) The agile views need to maintain the integrity of the data
4) The agile platform should simplify the application code by isolating the complexity of the schema from the application developers
5) The big data infrastructure needs to be reliable and fully scalable. It should have the capacity to stream both dynamic and processing views
6) The agile big data platform should have the capacity to handle event processing
7) It should be possible to access all views via advanced dashboards and in real time
8) The practices and principles of agile methodology focus on confirming assumptions the earliest possible during a project delivery lifecycle. Doing so reduces the prevailing risk exposure in the undertaking of the project.

Some of the benefits of agile methodology in big data include:

1) It helps in de-cluttering a company's information landscape
2) Makes it easy to access and combine data from several business units, functions and databases
3) The rapid iterations of MVPs provide immediate value to the business
4) The deep insights into the business data creates new growth opportunities for the business
5) Helps IT organizations in prioritizing digital and data transformation initiatives

6) Brings unparalleled visibility into business data to all decision-makers in the business organization

IV. BIG DATA TECHNOLOGIES WITH AGILE ANALYTICS METHODOLOGIES

There is a strong need today to tackle projects that allow companies to change their strategies more quickly in order to anticipate the market and the competition. With new technologies, companies have found themselves thrown into situations in which they have to act quickly in order to set themselves apart from their competitors and to offer new services and products adapted to the tastes and demands of potential clients, meaning they need to have a deep understanding of what those clients are looking for. Some of the questions that company directors are now asking themselves include what type of products and services their clients would like, what opinion they have of the company, whether their clients are satisfied or whether they will leave them for the competition.Thanks to new technologies and advanced analytical techniques, both predictive and prescriptive, these and other questions can now be answered. To do so, it's necessary to analyze data from various sources: client data, social networks, open data, data from surveys, information published on websites, navigation logs, user calls to the call center, sensor data, wearables, geolocation data, etc…

The volume and structure of all of these data are varied. The data can be structured, semi-structured or unstructured, and new technologies that allow for the storage and processing of all of the information are required to analyze them with new analytical techniques. One example of data requiring these new capabilities is data from social networks. Every day, millions of users publish opinions, comments, etc. on social networks, providing a huge volume of valuable, unstructured information in real time that allows for better customer knowledge.In order to obtain value from the data, such as learning the reasons for user dissatisfaction with a particular brand, a methodology is required that can manage the fluid interaction of a perfectly coordinated multidisciplinary team composed of new big data roles (data architects, data engineers, data scientists and visualization experts). This team can then provide an analytical methodology that will allow for a thorough analysis of the available data, as well as its modelling, producing results aligned with the business.

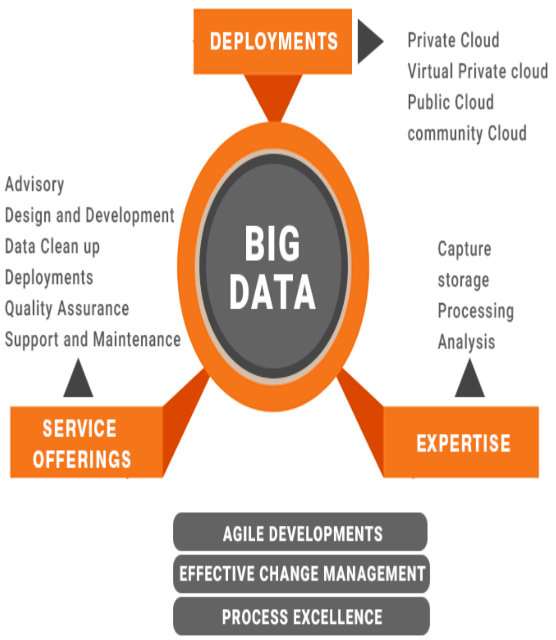


Figure 3: Big Data for Agile Development

An agile analytic methodology consists of the association of the management of the agile methodology as a method for work management based on short-term results and the analytical methodology as a strategy for obtaining adequate results in a massive data analysis. This union allows for a work setting that makes it possible to shift the knowledge needs of the company to the acquirement of adequate results for the company in a way that is both efficient and adapted to the changing and complex world we live in.All of this is the reason why the adoption of agile analytic methodologies together with new big data technologies and advanced analysis allowing value to be obtained from the data quickly and precisely is so necessary, because it converts the data into knowledge that can be used by company directors to make immediate decisions. Thus, companies are able to integrate big data as part of their strategy to transform the business and to obtain competitive advantages in the new era of information.

V. PRACTICAL RECOMMENDATIONS FOR AGILE WITH BIG DATA ANALYTICS

Data analysis, like other pursuits, is a balancing act. The rise of big data ratchets up the pressure on the traditional enterprise data warehouse (EDW) and associated software tools to handle rapidly evolving sets of new demands posed by the business. Companies want their EDW systems to be more flexible and more user friendly — without sacrificing processing speeds, data integrity, or overall reliability.In today's ultra-competitive markets, business operates far too quickly for traditional software development timetables, and most companies are rightfully wary of relying on outside vendors to solve ongoing business challenges. Luckily, agile and other "lean" software development methods offer repeatable processes for harnessing creativity and expertise, at a pace that's swift enough to keep the business happy.The difference between the "abstractedness" of traditional methods and the "concreteness" of agile might seem like a matter of semantics, but it's highly relevant in a world of hyper-turbulent markets and fickle customers. Abstractions are tolerable when you have plenty of time and money, but when time and money are tight, you need concrete results in a hurry. The whole point of agile is finding out quickly whether your idea works or not. Agile embodies the concept of "failing fast," a phrase that sums up the ethos of modern innovation.

But here's the rub: adopting agile methodologies is not the same as being agile. Agile is both a process and a mindset. It embraces both discipline and flexibility. Agile is not some kind of free-form anarchy; it's a structured way of producing usable software without having a pre-written script, like improvisational theatre.

The first rule of improve is always to say, "Yes, and…" Everything an actor does on stage is complementary; nothing is exclusionary. New information emerges unexpectedly, and continual change is a given, but established premises remain in place.

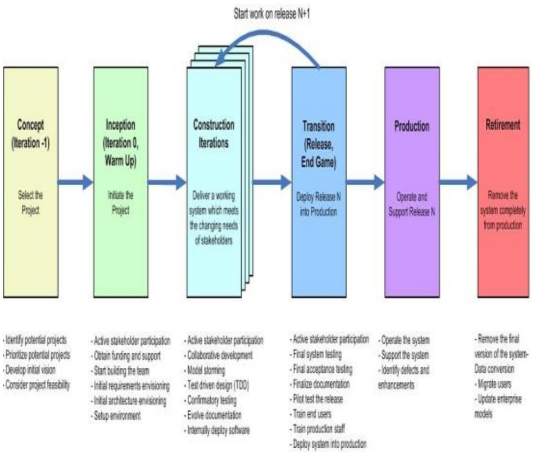


Figure 4: Agile Practical Life Cycle

Agile is optimally suited for situations in which the ability to generate lightning-fast results can produce genuine competitive advantages. Experimenting with subsets of customer data, testing new product categories, evaluating the usability of a web page design, gauging the appeal of a smartphone app — those are the best scenarios for getting the most value from agile.It seems clear that agile offers a reasonably quick and cost-effective way to bring flexibility to the EDW and to integrate newer open-source technologies, such as Hadoop, with existing database systems. Flexibility translates into greater speed and cost savings; integration promises a significantly wider range of analytic capabilities, which creates more opportunities for the business to pursue.As we move forward into an era of bigger, faster, and more effective data analytics, it seems logical to assume that database systems architectures will include both traditional and open-source components. In hybrid computing environments, the real challenge is optimizing the relationships between all of the various people, processes, and technologies required to get the job done quickly and efficiently.

Instead of picking fights over platforms, we should search for mutually beneficial scenarios in which we use available resources to help businesses gain advantages in competitive markets.

VI. BIG DATA ANALYTICS ONTOLOGY

Ontology is a formal naming and definition of a number of concepts and their interrelationships that really or fundamentally exist for a particular domain of discourse. Then, ontology of big data analytics is an investigation into a number of concepts and their interrelationships that fundamentally exist for big data analytics. Based on the above discussion, we propose ontology of big data analytics, as illustrated in the following Figure. In this ontology, big data analytics is at the top while big data and data analytics are at the bottom. Big data descriptive analytics, big data predictive analytics, and big data prescriptive analytics are at the middle level as the core parts of any big data analytics.

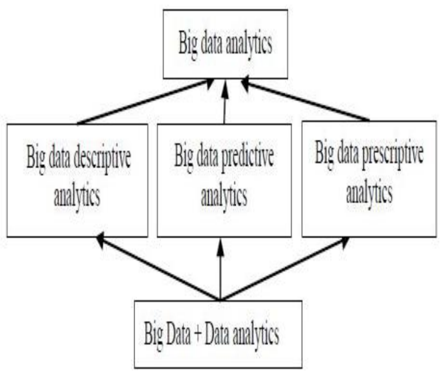


Figure 5: Big Data Analytics Ontology

It should be noted that the above-proposed ontology of big data analytics is still simple. We will extend it by adding another level between the second level and the third level in the above Figure. This added level will elaborate big data descriptive, predictive and prescriptive analytics taking into account the corresponding real-world examples, methods and techniques. In the above Figure, data analytics refers to as a method or technique that uses data, information, and knowledge to learn, describe and predict something. In brief, data analytics can be then considered as data-driven discoveries of knowledge, intelligence and communications. More generally, data analytics is a science and technology about examining, summarizing, and drawing conclusions from data to learn, describe and predict something.

The fundamentals of big data analytics consists of mathematics, statistics, engineering, human interface, computer science and information technology. The techniques for big data analytics encompass a wide range of mathematical, statistical, and modeling techniques . Big data analytics always involves historical or current data and visualization. This requires big data analytics to use data mining (DM) to discover knowledge from a data warehouse (DW) or a big dataset in order to support decision making, in particular in the text of big business and management. DM employs advanced statistical tools to analyze the big data available through DWs and other sources to identify possible relationships, patterns and anomalies and discover information or knowledge for rational decision making. DW extracts or obtains its data from operational databases as well as from external open sources, providing a more comprehensive data pool including historical or current data . Big data analytics also uses statistical modeling (SM) to learn something that can support decision making. Visualization techniques as an important part of big data analytics make knowledge patterns and information for decision making in a form of figure or table or multimedia. In summary, big data analytics can facilitate business decision making and realization of business objectives through analyzing current problems and future trends, creating predictive models to forecast future threats and opportunities, and analyzing/optimizing business processes based on involved historical or current data to enhance organizational performance using the mentioned techniques. Therefore, big data analytics can be represent

BIG DATA ANALYTICS = BIG DATA +
DATA ANALYTICS +
DATA WAREHOUSE +
DATA MINING +
SIMULATION

Where "+" can be explained as "and"". This representation reveals the fundamental relationship between big data, data analytics and big data analytics, that is, big data analytics is based on big data and data analytics, as illustrated in the above figure. It also shows that computer science and information technology play a dominant role in the development of big data analytics through providing sophisticated techniques and tools of DM, DW, ML and visualization [2]. SM and optimization still plays a fundamental role in the development of big data analytics, in particular in big data prescriptive analytics.It should be noted that the above equation is a concise representation for the technological components of big data analytics whereas the proposed ontology of big data analytics in this Section is to look at what big data analytics constitutes at a relatively high level.

VII RELATIONSHIP BETWEEN BIG DATA AND WEB SEVICES

Big data analytics can be considered a part of BI, because it ―supports business decision making with valuable data, information and knowledge. Both BI and big data analytics are common in emphasizing either valuable data or information or knowledge. BI involves interactive visualization for data exploration and discovery, for them Tableau, QlikView and Tibco's Spotfire are BI tools for interactive visualization for data exploration and discovery. These BI tools are also considered as the tools of big data analytics. This implies that BI and big data analytics share some common tools to support business decision making.

Currently, BI is based on four cutting-age technology pillars of cloud, mobile, big data and social technologies, each of these pillars corresponds to a special kind of web services, that is, cloud services, mobile services, big data services and social networking services; all these constitute modern web services. Each of these services has been supported by analytics services and technologies. They are effectively supported also by big data analytics as a service and technology, as shown in the following Figure.

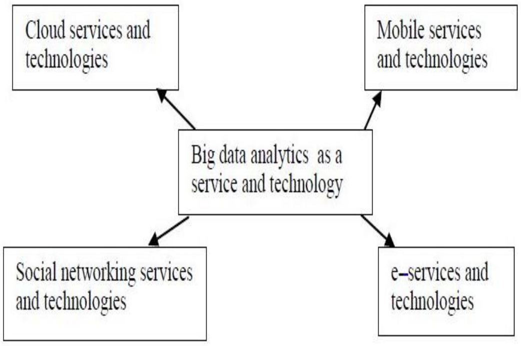


Figure 6: Relationship between BDA & WS

It should be noted that for the state-of-art web services, Sun and Yearwood explores that web services mainly consist of mobile services, analytics services, cloud services, social networking services, and service as a web service. In reality, each of them involves sophisticated ICT technologies. Then, technologies are added to mobile services, analytics services, cloud services and social networking services in. Here we emphasize big data analytics as a service and technology at the center to support cloud services and technologies, social networking services and technologies, mobile services and technologies, e-services and technologies to reflect the big data and analytics as an emerging new service and technology.

VIII. IN CONCLUSION

In this paper, we have presented the result of an exploratory case study to investigate the alignment between the expectations of Agile software development team.The Comparison based on the Triple Constraint of the project we can determine that for management of Big Data projects is the preferable agile approach. This is because of the constantly changing requirements posed by the giving of new questions. In implementing of Big Data project, it is recommended to start with small use case, accept small failures that move us forward and continue iterative approach. The problem arises if the company does not want or cannot use hosted or cloud platform provider. The reason as safety requirements in the domain such as banking or government must be specified in detail and analyzed at the start of the project. The findings from this paper contribute to the body of empirically-based knowledge related to agile software development practices. This paper deepens the understanding of factors related to alignment of expectation of user involvement,provides empirical evidence for the strength of the alignment of those expectations in practice across team roles. In addition the study contributes to a clear and consistent conceptualization of the meaning of “effective user involvement”. Practitioners will benefit from a deeper understanding and awareness of the differences and similarities of expectations of various development roles and user roles with respect to user involvement.

VIII REFERENCES

[1]C. P. Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," Information Sciences, vol. 275 , p. 314–347, 2014.
[2]Z. Sun, K. Strang and J. Year wood, "Analytics service oriented architecture for enterprise information systems," in Proceedings of iiWAS2014, CONFENIS 2014, 4 - 6 Dec 14, Hanoi, 2014.
[3]Z. Sun, S. Firmin and J. Yearwood, "Integrating online social networking with e-commerce based on CBR," in The 23rd ACIS 2012 Proceedings, 3-5 Dec, Geelong, 2012.
[4] Hunt J. Agile software construction. 2006.
[5]. Ceschi M, et al. Project management in plan-based and agile companies. Software, IEEE, 2005, 22.3: 21-27.
[6].Rouse M. Big data management. In SearchDataMnagement [online]. 2013 [cit. 2015-01-25]. Online:http://searchdatamanagement.techtarget.com/definition/big-data-management.
[7]. Accenture. Big Success With Big Data. Accenture Analytics survey [online]. 2013 [cit. 2015-01-25]. Online:http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-Big-Data-POV.PDF.
[8]. Stamford C. Gartner Survey Reveals That 64 Percent of Organizations Have Invested or Plan to Invest in Big Data in 2013. In GartnerNewsroom [online]. 2013 [cit. 2015-01-25]. Online: http://www.gartner.com/newsroom/id/2593815.
[9]. Stamford C. Gartner Survey Reveals That 73 Percent of Organizations Have Invested or Plan to Invest in Big Data in the Next Two Years. InGartner Newsroom [online]. 2014 [cit. 2015-01-25]. Online: http://www.gartner.com/newsroom/id/2848718.
Iivari, N., “Constructing the users” in open source software development: An interpretive case study of user participation.Information Technology & People, 2009. 22(2): p. 132-156.
[10]. Ives, B. and M.H. Olson, User involvement and MIS success:A review of research. Management science, 1984. 30(5): p.586-603.

[11]. Procaccino, J.D. and J.M. Verner, Software developers' views of end-users and project success. Communications of the ACM, 2009. 52(5): p. 113-116.
[12]. McKeen, J.D., T. Guimaraes, and J.C. Wetherbe, The relationship between user participation and user satisfaction: an investigation of four contingency factors. MIS quarterly, 1994: p. 427-451.
[13]G. Marakas, Modern Data Warehousing, Mining. And Visualization – Core Concepts, Prentice Hall, 2003.
[14]S. Suh, Practical Applications of Data Mining, Jones & Bartlett Learning, 2012.
[15]J. Han, M. Kamber, and J. Pei. Data Mining: Concepts and Techniques. Morgan Kaufmann, 3rd ed. 2011.
[16] B. Liu, Web Data Mining, Springer 2006.