Achieving Effective Data Analytics: An Agile Framework

Ms. Lymari Castro
lcastr2@radium.ncsc.mil
Department of Defense
Ft. Meade, MD 20755

Dr. Rashmi Jain
jainra@mail.montclair.edu
Department of Information Management and Business Analytics
School of Business
Montclair State University
Montclair, New Jersey, 07043

Abstract

The world of Information Systems is being redefined by the complexity of Big Data, whereby volume, veracity, and velocity needs to be delivered efficiently and effectively. Organizations rely on effective data analytics to manage large volumes and variety of data, and deliver value for decision making in a timely manner. This research introduces an agile framework for effective data analytics.

Keywords: Systems Engineering, Agile, Big Data, Analytics, Business Intelligence, Framework

1. INTRODUCTION

In today's dynamic environment, companies rely on the use of data analytics and business intelligence systems to achieve value and address changing business needs to remain competitive in the market.

Big Data deals with voluminous data objects that are diverse in nature: structured, unstructured, or semi-structured, including sources internal and external to an organization, and generated at a high degree of velocity, with an uncertainty pattern, that does not fit neatly into traditional, structured, and relational data stores (Haque & Hacid, 2014). The fast growing and complexity of the datasets makes it difficult to manage Big Data using traditional database management concepts and tools. Big Data systems can be decomposed into four sequential modules: data generation, data acquisition, data storage, and data analytics (Yin & Kaynak, 2015).

The objective of Big Data is to achieve a fault-free and cost effective running of the process while achieving the desired performance levels. It requires the use of high performance, computing platform, and analytical capabilities to capture, process, transform, discover, and derive business insights and value within a reasonable timeframe. Hence there is an increased demand for effective agile data analytics that can be used to analyze information and discover patterns in the data to quickly allow an organization to make decision needs, which is the biggest challenge of Big Data (Stamford, 2011).

Data analytics cannot be treated like traditional projects with defined outcomes, required tasks, and detailed plans (Marchand Donald A., 2013). Data analytics projects are small, short term efforts that require the upfront, rapid feedback cycle of test and learn with stakeholders and customers by posing business questions and describing available datasets (Krishnaswamy, 2013). Project stakeholders work together as a team to narrow the scope by stating hypotheses and understanding the available data before implementation begins.

Effective data analytics can be achieved by the use of an agile framework that reduces the time to market while balancing the five dimensions of Big Data: volume, variety, veracity, velocity, and value. The agile framework for the development of effective data analytics requires tailoring agile development methodologies, agile project management methodologies, agile infrastructure, cloud technology, and IT organization to the specific needs of the project.

This research introduces the concept of an agile framework for effective data analytics within the complexity of Big Data.

2. WHAT IS EFFECTIVE DATA ANALYTICS?

Effective data analytics is defined as an analytics that brings benefits to an organization by managing multiple sources and types of data to quickly answer business questions while providing high quality relevant information to solve a problem.

3. AN AGILE FRAMEWORK FOR DATA ANALYTICS

The pillars of an agile framework for the development of effective big data analytics rely on the implementation of best practices from the agile development methodology, agile project management methodology, agile infrastructure, IT organization, and cloud technology (Bruni, 2011) to manage the five dimensions of Big Data or the five V's: value, volume, variety, veracity, and velocity (Yin & Kaynak, 2015) in a reasonable timeframe to bring the most benefit to an organization. Figure 1 illustrates and summarizes our research findings and proposed framework. Agility is a key factor in the development of effective data analytics. Agile data analytics utilizes an agile development style for building data warehouse, data marts, business intelligence applications, and analytic applications that focuses on the early and continuous delivery of business value throughout the development lifecycle (Collier, 2011). Agile data analytics consists of a set of technical, project management, and customer collaboration practices that can be tailored to fit the project demands of a given organization. The main objective is to deliver high-quality and high-value data analytics in a short timeframe.

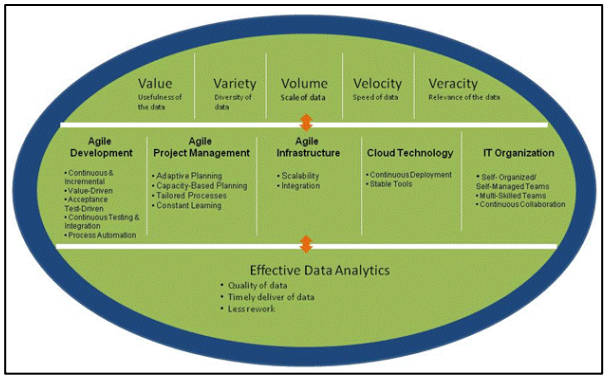

Figure. 1. Agile Framework for Effective Data Analytics

Agile Development Methodology

Agile development techniques have been successfully applied to speed the time to market of business intelligence systems. The methodologies utilize an iterative incremental process to shorten the development lifecycle. A critical factor in Big Data is how fast the data is produced and how fast it can be processed to meet the business demand. Effective data analytics needs to process the data and make it available for access and delivery in a timely manner. In Big Data the frequency of data generation is too rapid to be managed using traditional approaches (Stamford, 2011). By using an agile development approach, an effective analytics ensures that large volumes of data are managed in a reasonable timeframe for the organization to deliver value to the customer and for leaders to make important business decisions in an informed manner to remain competitive in the market. An effective data analytics applies and demonstrates the following agile development techniques:

Continuous and Incremental Delivery of Working Features: An incremental development approach to break-up large functions into smaller manageable pieces allows the team to focus on the core highly prioritized functions (Jain, Chandrasekaran, Castro, & VanLeer, 2011).

Incremental delivery of capability makes information more rapidly available to the customer and the stakeholders to solve a business problem or make business decisions. Each iteration delivers working features that have been tested by the stakeholders and adapted in further iterations to better suit the organization's needs (Preimesberger, 2013). Every iteration must produce at least one new user-valued feature (Collier, 2011). Streaming real-time data and in large volumes requires continuous and incremental delivery. Hadoop, MongoDB kinds of infrastructures support this feature to be delivered.

Value-Driven Development: Contrary to traditional projects where progress and performance is measured by task completion and schedule, in agile data analytics progress and performance is defined by the value that the capability delivers and its quality. Value driven data analytics goes above and beyond facts and intuition. Value from data comes not from the data in its raw form, but from its processing and analysis and the insights, decisions, products, and services that emerge.

Acceptance Test-Driven Development: Acceptance test-driven development is similar to test-driven development but it occurs at a higher level of functionality. While test-driven development is done at the unit level to validate the implementation of code, acceptance test-driven development is done at a higher level of functional testing and tests the features and behaviors of the system that are observable by the user (Cobb, 2015). This technique requires writing the acceptance tests for the functionality that must be provided prior to starting development. This helps the team to build a concise and common understanding about the functionality to be delivered; removes the need to write detailed functional specifications; and simplifies the requirements definition process. Data analytics needs to be designed based on the users' needs. What kinds of questions does the business needs answers for - should be driving the data analytics not the other way round. An approach that focuses on the testing of such relevant hypothesis suits the world of Big Data. Organizations that might be sitting on piles of data but cannot formulate the right questions cannot deliver agile data analytics.

Continuous Testing and Integration: The desire for agility in data analytics is driven by the need to serve end users while reducing the time-to-market. This is achieved by incorporating testing and integration as a continuous process throughout the development lifecycle. Agile data analytics would involve continuous integration and testing frequently in test systems pre-production or demo environments. Data integration is the most important issue for any analytics initiative. Major companies are unable to leverage analytics due to this limitation. Legacy data integration requires continuous testing. Developers are expected to plan for rigorous testing during the development process.

Process Automation: To accelerate development of agile analytics it is recommended that routine processes (i.e. processes that have to be performed more than once) be automated to allow developers to focus their efforts on the development of user features. For instance, the automation of testing enables developers to frequently revalidate features and ensure they are working as expected. In addition, the automation of regression testing allows running tests repeatedly and efficiently on a frequent basis and provides the capability to check the entire system as new functionality is being added. By utilizing build automation, developers frequently build a version of the data analytics in a demo or pre-production environment, which reduces the time to release a new version to production. Equally important are the processes of configuration, migrations, change control to accomplish agility in data analytics. Open platforms like Hadoop, have led to the growth of several vendor solutions that efficiently automate these important development processes.

Agile Project Management

An effective data analytics that retains its veracity and provides value requires continuous planning and execution with constant customer feedback. Agility is achieved by doing planning at the beginning of each cycle and use customer feedback to change the scope any time during the development phase. Data analytics is effective when the data is relevant and can be trusted. Effective data analytics uses agile project management approaches to remove uncertainty in the data by removing inconsistencies and ambiguities caused by lack of understanding of customer requirements and the future needs of the changing market environment. Such information is important in providing information superiority if its timely, relevant, and accurate. The following agile project management techniques are conducive to effective data analytics:

Adaptive Planning: The core purpose of big data is to find key insights upon which an organization can pivot (Preimesberger, 2013) Developers, customers, and stakeholders work together to plan each iteration using information learned from previous iterations and prioritizing iteration requirements based on their risk level and business value (Jain, Chandraserakan, & Castro, 2015). The team is constantly learning from what end users want, tests, and business conditions, and they use that information to effectively plan subsequent iterations.

Capacity-Based Planning: Iterations should be planned in a way that they do not exceed the team's demonstrated capacity (i.e. do not commit to deliver more than what the team can effectively deliver). Scoping out iterations of data analytics within the information systems project life cycles is an important element of capacity planning.

Tailored Processes: In traditional projects, requirements, specification, design, development and testing are discrete phases of the project lifecycle. In agile data analytics, these are not discrete phases; instead, the team continuously cycles between all phases multiple times to redefine requirements and the design. The delivery of a working feature is the best proof of requirements understanding and design. Processes should be tailored to support the iterative, incremental, and evolutionary development of data analytics.

Constant Learning: Agile data analytics projects utilize constant customer and test feedback to gain knowledge that is used to plan future iterations and quickly adapt to change. Constant feedback from stakeholders helps maintain the veracity and value of the data. The frequent acquisition of knowledge substantially increases the team's agility and keeps the organization ahead of its competition. Valuable insights through pattern and trend analysis are a constant learning opportunity.

Agile Infrastructure

Effective data analytics should have virtualization and horizontal scaling capability that will provide the flexibility to modify the infrastructure and be able to support near-real-time business intelligence. The growth of world data is exponential (Deloitte, 2013). The fast increase of data volumes can represent storage issues and too much data can represent analytics issues. Effective data analytics should provide the capability to search and analyze large volumes of data, while providing the ability to discover patterns, consolidate information, visualization of the data, and reporting. An agile infrastructure provides the capability to manage large volumes of diverse data in a reasonable time. The following are the characteristics displayed by an effective data analytics infrastructure:

Scalability: Agile data analytics require an infrastructure that is flexible to manage extreme volumes, velocity, and variety of data. Scalability in agile data analytics infrastructure is essential to improve the customer experience and to provide value.

Integration: Virtual integration of data allows stakeholders to visualize the data earlier in the development lifecycle which helps refine the requirements. This helps to ensure the veracity and value of the data during the development cycle. In addition, agile data analytics displays the integration of commercial technologies, data modeling and manipulation, and custom code. As well as it needs to be able to support the transfer of legacy systems to newer technologies. The use of an agile infrastructure that utilizes virtualization facilitates agility by allowing the quick rollback to a baseline instance in seconds and providing developers the ability to mimic distributed systems on a single workstation.

Cloud Technology

Effective data analytics are achieved by leveraging the benefits of cloud technology. Cloud technology provides a cheaper alternative to store and transfer large volumes of variety of data. It provides an efficient way to move large volumes of data across the network and store constantly changing data. Big Data consists of a variety of data types: structured, unstructured, and semi-structured and it deals with large amounts of data types such as video, email, text, audio, and log files, among others. An effective agile data analytics needs to provide the capability to translate diverse data sets in such a way that the business leader can make important business decisions in a reasonable time. Cloud technology provides a stable architecture with proven tools that reduces development waste efforts caused by the use of unproven technologies. Effective data analytics is achieved by the following:

Continuous Deployment: The use of cloud technology provides the ability to quickly and frequently deploy capabilities and new features

to production. With cloud services, software can be provisioned as a service; which means that for an organization that has problems with in-house systems, they can use the cloud to avoid having to upgrade hardware or software. Architectures such as Amazon Web Services (AWS) are doing an excellent job of using cloud services.

Stable Tools: Effective data analytics achieves agility by using tools and technologies that have been tested and have been demonstrated to be robust. The use of stable tools significantly reduces potential errors and the rework associated with fixing them. As the data analytics tools evolve and get refined with the availability of the right technology, this will provide more stable infrastructures for data analytics.

IT Organization

Effective data analytics is able to manage velocity, veracity, and value by having an IT team that is knowledgeable of the business it serves and understands the business problems. The IT team is a cohesive team that includes the customer and in most instances is collocated. Value is achieved by the interdisciplinary cooperation of the team. Effective data analytics uses methods and techniques that require the frequent cooperation of stakeholders (i.e. researchers, engineers, developers, managers, and customers) to improve current processes and foster new ideas. An IT organization environment that promotes effective data analytics displays the following:

Self-Organized and Self- Managed Teams: Effective data analytics require the creation of highly skilled teams that take ownership of the final product. Teams are empowered to make decisions regarding how much work can be completed during the iteration and they hold themselves accountable. Teams are composed by the developers, project manager, and the customer. To achieve agility, the project manager role transitions from the traditional task-based approach to a more supportive team-management role that is responsible for providing the tools needed by the team, and facilitates collaboration with customers and team members to increase team productivity (Suitling, Mansor, & Widyarto, 2014).

Multi-Skilled Teams: Effective data analytics requires the integration of team members that are knowledgeable not only of the technical and management fields, but who also understand the cognitive and behavioral sciences. This will allow the team to better understand how customers use the system and derive additional information to plan future work and address business needs. Teams should be comprised of individuals who know the business and are inquisitive about how users' use the information provided and are not afraid of providing hypothesis to find new ways to solve old problems.

Continuous Collaboration: Effective data analytics projects are comprised of multiple members: project managers, technical experts, business stakeholders, end users, their advocates, and others. Frequent collaboration between the technical experts, business personnel and customers is essential for project success (Marbach, Rosser, Osvalds, & Lempia, 2015). The rapid communication between team members is key to achieve agility. The creation of collocated teams is conducive to a collaborative workspace which promotes agility (Wells, Darren, & Hedley, 2015). In the development of agile data analytics, it is expected that there will be daily coordination within team members to identify issues and accomplishments, which is critical for project success. Agility is achieved by having the customer partner with developers to test features and provide feedback on a frequent basis. It is crucial to promote a collaborative culture in which team members feel free and are motivated to share their ideas and knowledge (Marchand Donald A., 2013).

4. CONCLUSIONS

Management of the complexity of Big Data requires the use of an agile framework that promotes the delivery of effective data analytics by addressing volume, veracity, variety, velocity, and value. For the agile framework to succeed it is imperative to leverage best practices from agile methodologies along with the benefits offered by cloud technology and robust data analytics infrastructures. Effective data analytics can only be achieved by implementing the five pillars of the agile framework. Failure to implement the elements of the proposed agile framework could lead to a lack of agility and risk of not meeting the business needs of the organization and customers in a timely and accurate manner.

5. FUTURE WORK

We have been working with several companies informally to share our research and opportunities for validation. However, we have still not collected data from the industry to validate our proposed framework. These companies are still busy putting their data analytics strategy and a governance structure to support it. They are still not ready to share information for research. We will soon be able to collect data from the industry to validate our proposed framework.

6. REFERENCES

Bruni, M. (2011, June 13). *InformationWeek Analytics.* Retrieved November 2015, from InformationWeek.com: www.informationweek.com

Cobb, C. (2015). *The Project Management Guide to Mastering Agile: Principles and Practices for an Adaptive Approach.* John Wiley and Sons.

Collier, K. (2011, October). *Agile Analytics: A Value-Driven Approach to Business Intelligence and Data Warehousing.* Addison-Wesley Professional. Retrieved November 2015, from ThoughtWorks.com: https:\\www.thoughtworks.com/insights/blog/introducing-agile-analytics

Deloitte. (2013). Big Data Challenges and Sucess Factors. *Deloitte Analytics*.

Haque, R., & Hacid, M.-S. (2014). Blinked Data: Concepts, Characteristics, and Challenges. *2014 IEEE 10th World Congress on Services* (pp. 426-433). IEEE Computer Society.

Jain, R., Chandrasekaran, A., Castro, L., & VanLeer, M. (2011). Feasibility of a rapid systems engineering framework: an exploratory study. *International Journal of Industrial and Systems Engineering*, Vol. 7, No.1,45-65.

Jain, R., Chandraserakan, A., & Castro, L. (2015). Identifying Suitable Projects for Rapid Development: Some Proposed Selection Criteria. *International Journal of Management and Information Systems*, Vol. 19, No. 2, 51-59.

Krishnaswamy, A. (2013, November 22). *The Benefits of Agile Analytics Development and How to Do It Right.* Retrieved November 2015, from Data Informed: http://data-informed.com/benefits-agile-analytics-development-right/

Marbach, P., Rosser, L., Osvalds, G., & Lempia, D. (2015). Principles for Agile Development. *INCOSE International Symposium Vol. 25, Issue 1* (pp. 524-537). Wiley.

Marchand Donald A., P. J. (2013). Why IT Fumbles Analytics. Tech projects should focus less technology and more on information. *Hardvard Business Review*, 1-9.

Preimesberger, C. (2013, July 26). *Agile Analytics: What It Is, and 10 Best Practices for Using It.* Retrieved November 2015, from eweek: http://www.eweek.com/cloud/slideshows/agile-analytics-what-it-is-and-10-best-practices-for-using-it

Stamford. (2011, June 27). *Gartner Says Solving 'Big Data' Challenge Involves More than Just Managing Volumes of Data*. Retrieved January 2016, from www.gartner.com: www.gartner.com/newsroom/id/1731916

Suitling, K., Mansor, Z., & Widyarto, S. (2014). Agile Project Manager Behavior: The Taxonomy. *8th Malaysian Software Engineering Conference (MySEC).*

Wells, H., Darren, D., & Hedley, S. (2015). The adoption of agile management practices in a traditional project enviroment: An IT/IS Case Study. *48th Hawaii International Conference on System Sciences* (pp. 4446-4452). IEEE Computer Society.

Yin, S., & Kaynak, O. (2015). Big Data for Modern Industry: Challenges and Trends. *Proceedings of the IEEE Vol.103, No. 2, February 2015*, 143-146.