

A Study on Human Centric Agile Methodologies with Big Data & Predictive Analytics in Software Development

T. Sasi Vardhan^{1*}, C. S. R. Prabhu² and V. Anitha³

¹Research Scholar, Dept. of CSE, UOT, Jaipur, Rajasthan, India. Email: sasi.mrecw@gmail.com

²Director General (Retd), NIC, Hyderabad, Telangana, India.

³Assistant Prof., Dept. of CSE, BVRIT, Narsapur, Telangana, India.

*Corresponding Author

Abstract: This paper proposes an agile model-based systems engineering (SE) methodology to engineer the contemporary large, complex, and interdisciplinary systems of systems. This paper introduces the reader the background of Big Data Analytics and how efficiently Agile methodology can be applied to achieve the business goal. The journal focus on giving background of Big Data and how using Agile practices such as iterative, incremental, and evolutionary style of development can be applied for Big Data Analytics. This methodology brings in the advantage of involving business community during development and continuous delivery of working user features. The Agile uses a universal and intuitive SE base process, reducing the complexity and intricacy of the base methods, emphasizing the agile principles such as continuous communication, feedback and stakeholders' involvement, short iterations, and rapid response, and rousing the utilization of a coherent system model developed through the benchmark systems graphical modeling languages. The Agile methodology also includes a supporting graphical tool that aims to be an agile instrument to be used by systems engineers in a model-based development environment.

Keywords: Agile, Big data analytics, Big data, Data analyst, Model-Based System Engineering (MBSE), Software engineering.

I. INTRODUCTION

The Modern world is characterized by large interdisciplinary complex socio technical systems made of other system, personnel, hardware, software, information, processes, and facilities. The Model-Based SE (MBSE) paradigm is an emerging approach in the SE field. This approach relies on the development of a unified coherent system model that should act like a shared working platform and should reflect the

stakeholders' ideas and positions in order to lead to a resulting successful system. An MBSE environment includes several interacting elements (e.g., standards, formalisms, methods, modeling tools, applications, etc.) that must work together to achieve the main result, which corresponds to the development of a successful system. Business establishment of all sizes now realized the need of Big Data analytics to help various business key decision such as sales forecast, to predict raw material need and also to help companies understand their customers better. Most of the time companies them self able to gather data from different source filter unwanted and analyse data to take key decisions. But the scope of data analysed may be very little and hence use of big data has become the need for modern data analysis. Therefore, a quick, less risky and better approach is needed in analysing big data. Agile methodology will add value by bringing the concept of think big, start small, move fast and deliver useful product as soon as possible.

II. BACKGROUND STUDY

The amount of digital data is set to grow to 44 times its current volume by 2020. Big data arrives through different velocity, volume and variety. 2.7 Zeta bytes of data exist in the digital universe today. IBM estimates 2.5 quintillion bytes of data are generated each day. Ninety percent of the data in the world is less than two years old. Facebook stores, accesses, and analyses 30+ Pet bytes of user generated data. More than 5 billion people are calling, texting, tweeting and browsing on mobile phones worldwide. YouTube users upload 48 hours of new video every minute of the day. 571 new websites are created every minute of the day. 100 terabytes of data uploaded daily to Facebook. 100 terabytes of data uploaded daily to Facebook. Exponential growth in the number of data sources will continue to make gathering trusted information a challenge. Most of data generated today is unstructured or in a semi-structured data format that makes it more difficult to manage and process. The increased size of unstructured data and increased complexity

managing to it is a difficult task requiring enhancements of how data is managed and analysed which can be delivered, even if there is only 3 iterations worth of code vs. no code and no value. Thus there is early ROI, increased control and reduced risk, improved communication, and accelerated time to market. Generally customer satisfaction is improved as a result and are better able to manage business through metrics.

A methodology can be seen as “a set of related activities, techniques, and conventions that implement one or more processes and is generally supported by a set of tools”. The process is the set of interacting activities that transform the inputs into outputs (the WHAT activities to perform), the method specifies the techniques for performing the tasks of the process (the HOW to execute), and the tools are the resources applied to the method in order to improve the efficiency of the tasks (enhancing the WHAT and the HOW). An MBSE methodology gathers all these pieces, implementing a given process, which is supported by a given method, which is enhanced by a set of tools.

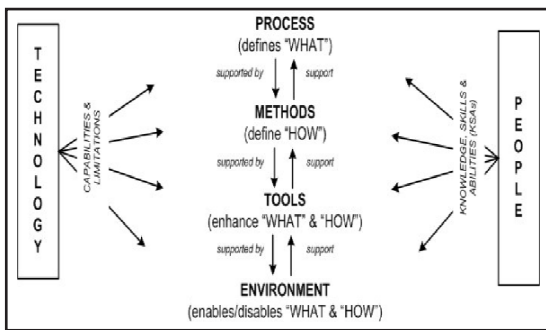


Fig. 1

III. USAGE OF BIG DATA

To determine more accurate information validated from different reliable sources.

- To improve business decisions based on the real time data to increase revenue.
- To improve marketing strategy and targeting right people.
- To increase customer base.
- To adapt to latest technology or trend.
- To increase the security and to avoid data leakage.



Fig. 2

IV. AGILE APPROACH

The fundamental research lines in the MBSE field will be related with the following: 1) the development of simple and agile MBSE methodologies and 2) the effective utilization of graphical modeling languages able to support collaborative development environments and successful stakeholders' communication / interactions, thus successful systems. As referred, a successful MBSE environment uses a given methodology to develop the model of a system and transform this model into a real thing. The methodology consists of a SE process, a method, and a set of modeling tools that enhance the process and the method. The main artifact of an MBSE methodology is the system model.

This new methodology, the AGILE (Agile Systems Modeling Engineering), intends to be a more agile methodology using a universal and intuitive SE process, reducing the complexity and intricacy of the supporting method, emphasizing the agile principles such as continuous communication, feedback, and stakeholders' involvement, short iterations, and rapid response, and rousing the utilization of a coherent system model developed through the benchmark SE graphical modeling languages (SysML and OPM).

SE Process (WHAT to Do): The process is the set of activities to perform in order to transform the inputs into outputs. The proposed AGILE methodology encloses a revised and contextualized (in the framework of the international SE process standard) version of this SIMILAR process model.



Fig. 3

Agile Modelling: Agile Modelling (AM) is a practice-based methodology for modelling and documentation of software based systems. It is intended to be a collection of values, principles, and practices for modelling software that can be applied on a software development project in a more flexible manner than traditional modelling methods. The aim is to keep the amount of models and documentation as low as possible.

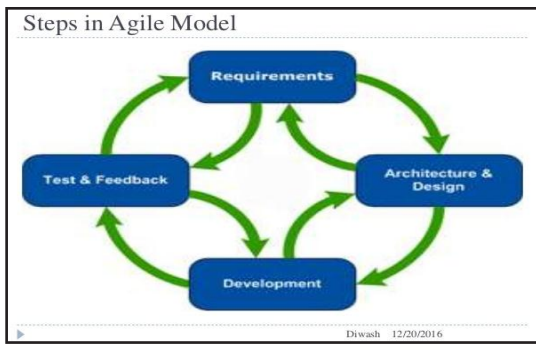


Fig. 4

V. BIG DATA ANALYTICS APPROACH

The following are the cases for Big Data Analytics:

- **Business Need:** The Business environment today is more complex than ever before. Businesses are expected to be diligently responsive to the increasing demands of customers, various stakeholders and even regulators.
- **Solution:** Organizations have been turning to the use of analytics. More than 83% of Global CIOs surveyed by IBM in 2010 singled out Business Intelligence and Analytics as one of their visionary plans for enhancing competitiveness.
- **Goal:** In most cases the primary objective of an organization that seeks to turn to analytics is Revenue / Profit Growth, Optimize expenditure.
- Support a mix of structured, semi-structured, and unstructured data without complex, time-consuming IT engineering efforts. Process data quickly and efficiently to return answers before the business opportunity is lost. Present the business user with an interface that doesn't require extensive IT knowledge to operate. Unlocking the value in data is the key to providing value.

A. Role of Business Owner

The Business Owner will be responsible for taking key decisions with the analysed data presented to them.

B. Role of Big Data Analyst

Data Analysis is an art of science and all Big data may not contain direct information, but sometimes the interpretation of data will give us what we need. Therefore, the Role of Data Analyst is very important in Big Data Analytics. Data analyst knows understand the whole business context and will able to help where to locate big data and how to interpret the data into more meaningful to support business decisions.

VI. BIG DATA SOURCE

- Internal Company Database or Data warehouse.
- Internal documents such as word document, excel sheet, csv files etc.
- External data that flows into the company such as invoices, proposals, information from third party vendors etc.
- Social websites and data feeds from twitter, facebook etc.
- Message forums, public website, Google ads etc.
- Customer information from different public websites.
- Data from mobile apps.

VII. BIG DATA ANALYTICS USING AGILE MODEL

Roles & Responsibilities

Agile Role	Stake Holder	Responsibilities
Product Owner	Business Owner	Responsible for writing requirement as stories.
Developer	Data Analyst	Responsible for analysis of data.

A. Planning Phase

Planning phase involves activities related to Requirement analysis & elicitation, preparation of stories for Product Backlog. Identification of all stakeholders in one of the key activity in this phase. Kick off meeting will be held among the stakeholders. The product owner will be responsible for writing small stories to achieve the final goal. All stories written will be stored as the product backlog. Stories are planned and prioritized in a way to deliver individual functionality that can be tested and delivered independently. The product owner will also identify the appropriate tools needed for Big Data Analytics in this phase. Alteryx is one of the tools that can be planned to be used.

B. Development Phase

Stories planned in product backlog will be prioritized for iteration and developed in this phase. The data collected from different sources are analysed in this phase. Data collect from multiple data sources will be blended into a single worksheet. The data will be joined on common dimensions and presented in a single view. The data analyst will filter unwanted data. It is very important to have high collaboration between all stakeholders such as business owners, data analyst, technical team and product owner. One of the key features of Agile is that requirement keep changing to achieve what is needed for business today. Stories will be planned and prioritized

accordingly by the product owner. Tools that are planned during the planning phase will be used for the development. Generic algorithm will also be developed in this phase. Fine tuning of algorithm will happen in each iteration. Use of predictive analytics tools can also be used to predict or forecast the data. Frequent reviews are also in this phase. Testing frequently will ensure the data that is being used is correct. Integrating continuously and testing the system in different environment throughout the project will not bring any surprise at the end of development.

C. Closure Phase

Closure occurs when all requirements are met. This is the last phase before releasing the product with all documentation. Final tests are done at this point and at the end - the release itself is final. In this phase the algorithm developed and the data delivered by this will be used to take key decision for the business. The predictive analysis tool will also assist in taking business decisions for the future. The implementation and post implementation support will happen in this phase.

Steps Involved in Big Data Analytics:

1. Know your goal.
2. Collect from right source.
3. Filter unwanted.
4. Verify often.
5. Generate Algorithm.
6. User Predictive Analytics Tools.

VIII. COMPARISON STUDY

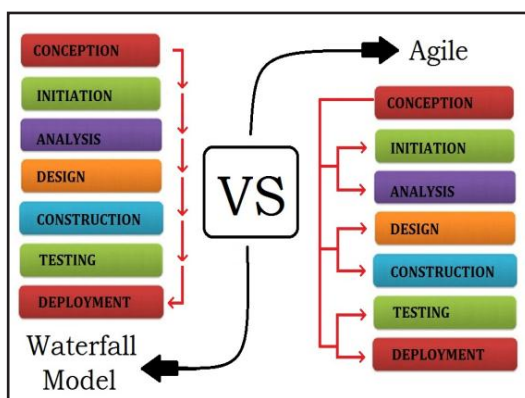


Fig. 5

IX. CONCLUSION

The development of an integrated methodology with simple, lean, and customizable processes and methods is of paramount

importance to enable the widened utilization of MBSE practices. The SE process (WHAT) must be intuitive, logical, universal, and easy to use and to tailor. The ISO/IEC 15288 process standard requires some integration that can be provided by the SIMILAR process model. The MBSE method specifies the HOW to execute the process and relies on the development of a coherent system model. This Journal helps the readers to understand how to apply Agile Methodology for Big Data Analytics Projects. Big Data provides more accurate real time data to improve business decisions on time. Many companies started investing to help their business by taking key decisions. Customers will get benefit by adopting Agile Methodology on Big Data Analytics Projects. Agile brings early time-to-market, good collaboration, and high customer satisfaction and brings risk to light faster through iterative development and prioritization approach. The comparison study between traditional waterfall model and agile will help the readers to understand the advantage of Agile Approach.

REFERENCES

- [1] D. H. Rhodes, "Addressing systems engineering challenges through collaborative research," in *SEARI-Systems Engineering Advancement Research Initiative*, Cambridge, MA: MIT Press, 2008.
- [2] S. Sheard, and A. Mostashari, "Principles of complex systems for systems engineering," *Systems Engineering*, vol. 12, no. 4, pp. 295-311, September 2009.
- [3] A. L. F. A. Ramos, J. V. Ferreira, and J. Barceló, "Revisiting the similar process to engineer the contemporary systems," *Journal of Systems Science and Systems Engineering*, vol. 19, no. 3, pp. 321-350, September 2010.
- [4] S. Friedenthal, A. Moore, and R. Steiner, *A Practical Guide to SysML: The Systems Modeling Language*, Burlington, MA: OMG Press, 2008.
- [5] J. O. Grady, "Universal architecture description framework," *Systems Engineering*, vol. 12, no. 2, pp. 91-116, May 2009.
- [6] R. Jurney, *Agile Data Science*, 2013.
- [7] M. Poppendieck, and T. Poppendieck, *Lean Software Development: An Agile Toolkit*, Addison-Wesley, Boston, 2003.
- [8] V. Mayer-Schonberger, and K. N. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, 2013.
- [9] M. Poppendieck, and T. Poppendieck, *Leading Lean Software Development: Results Are not the Point*, Addison-Wesley, Upper Saddle River, NJ, 2010.