ISSN: 2348 - 2117

International Journal of Engineering Technology and Computer Research (IJETCR)
Available Online at www.ijetcr.org
Volume 4; Issue 6; November-December; 2016; Page No. 29-33

Employing the Agile Approach in Big Data Analytics using Iterative and Adaptive Model Technique (ITAM)

S Sai Akshaya Pillai[1], Manjula R[2]

[1,2] VIT University, School of Computer Science and Engineering, Vellore, Tamil Nadu

Abstract

Colossal quantities of data get generated every day, which can yield valuable business information. Big Data analytics projects are aimed to leverage this data generated and glean it of business insights. As more and more businesses start to invest in big data analytics, project management techniques tailored for analytics projects may be devised, to save cost, time and effort. Consequently, a new branch called Agile big data analytics has developed which focuses on using agile techniques on analytics projects involving Big Data or otherwise. Thus, the agile approach which can be extended to managing projects may also be used to efficiently manage analytics projects. Iterative discovery is the key to lending agility to such Analytics and Intelligence based projects. In this article, we explore a method to extend the agile approach to big data called ITAM- Iterative and Adaptive Model Technique. Here the projects are completed in phases which consist of ideation, development of proof of concepts; analysis of the model; implementation and scaling wherein the models proposed undergo iterative improvements to provide greater project success.

Keywords: *Agile Analytics; Big Data Management; Agile Manifesto, Business Intelligence, Data warehousing; big data modeling*

Introduction

Agile has become the tech industry's buzzword with more and more companies taking to agile strategies for software development activities. Not just for development operations, agile can be used in various other scenarios also, one of them being analytics, which forms the subject matter of this article. First, we would like to expound a little on the agile strategy, analytics, and big data.

The Agile Strategy: The crux of the agile approach is about failing fast. Agile, basically involves the fast generation of working code that sufficiently satisfies the objectives of the particular iteration. The emphasis is on delivery over documentation and people over process. In the software development scenario, failing fast means that developers cannot spend months developing and testing a project only to discover that it does not meet the needs of the clients. [1][4] The applications ability to satisfy the business needs must be determined as soon as possible and this requires agility. Agility lends the ability to determine the applicability of the development and saves precious time and money and effort.

Although agility is also about tailoring the process to suit the project, a typical agile development scenario mostly has a defined pattern for the lifecycle. Unlike the traditional software development life cycle models such as waterfall model etc., an agile development process' first phase is not about gathering detailed and comprehensive requirements. Instead of striving for perfection, agile aims for the minimum viable product. In this manner agile produces concrete results faster than traditional development methods. Thus agile is suitable in situations when funds and time are not available aplenty.

Although agile might make the development cycle, sound like a free-for-all playground as agility is about improvisation and adapting to changing needs, it is actually quite a structured approach. The freedom to change and improvise should involve adherence to discipline and should not translate into a chaotic and unanswerable fashion of development.

Corresponding author: S Sai Akshaya Pillai

As development processes become agile, it is only natural for agile strategies to be tried and employed in other business value generating units. One such important process is the analysis of data in organizations. The insights and the impact of the insights offered by the data are what lend the data business value. Agile intervention into data analytics can help time-critical processes and result in timely delivery. [2][4]

Agile Analytics has gained much interest in the industry in the past few months. Like agile development, agile analytics is about iterating quickly and eliminating tracks that lead to non-delivery early. Lending this agility to the data can be done in various ways. This article discusses ways of lending agility to enterprise data warehousing systems and analytics [7]

Big Data Analytics: The term big data refers to such quantities of data as that which cannot be processed by traditional data processing architecture and systems. The datasets in big data analytics are typically so large and diverse that they are difficult to store, process and analyze. Examples of big data are spatiotemporal stream data, continuous satellite feeds etc. In order to analyze such vast quantities of a distributed architecture can be used where processing of the raw data can be done node-wise and the results of the nodal processing then put together to give the final outcome of the analysis. [8]

In Big Data analytics systems as such the open-source Hadoop framework, the fundamental idea is to take the processing to the data instead of bringing the data to the processing component as in traditional data processing systems. [6]

II. NEED FOR AGILITY IN BIG DATA ANALYTICS

When Hadoop was first adopted into practice, it was thought to not require project management as it was very simple to set up and running. But since then big data analytics has come to incorporate so many other technologies and the scope of such analytics ventures has expanded to a large extent. Ventures could involve as many as 100 engineers working. Such projects usually have major business impact and involve setting up of expensive infrastructure. With this upward trend in the scope of big data projects, project managers must form a vital part of any major big data venture. [5] [9]

Often big data ventures are unsuccessful due to a variety of reasons. It is reported that about 65-100% of Big Data Analytics projects fail. These reasons can be traced back to mis-management or under-management of the projects.

The reasons cited are schedule and budget overruns, incomplete project, inability to implement analytics and low quality or unsatisfactory results.

Existing project management techniques must be adapted for data analytics projects. Agile methodologies may be leveraged to increase the success rate of such projects. These methods are well-suited because of the iterative nature of the discovery process especially in frameworks such as SCRUM. [5] As agile methods have begun to be adopted in areas other than software development, Big Data Analytics must also adapt to become agile. As business needs evolve, the ability to satisfy them must also become tailored to suit them. [11]

III. A PRACTICAL IMPLEMENTATION

When determining the best methodology to manage big data projects, one must take into consideration all the factors that make big data project different from other analytics projects. The size and architecture of data, the method of scaling all influence the management of such projects. [3] Here we propose a methodology to adapt agile analytic practices to big data projects based on an iterative model called ITAM (Iterative and Adaptive Model Technique). The whole process can be spread over several iterations. Each iteration shall comprise of some activities.

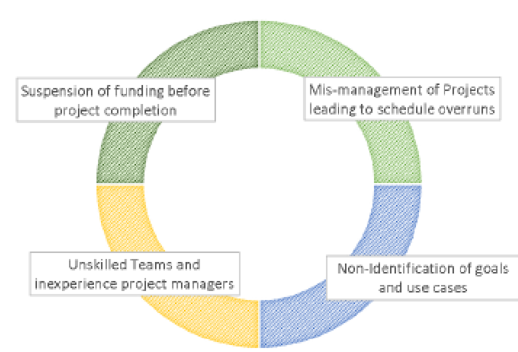

Fig 1 Reasons Why Big Data Analytics Projects Fail

Method of scaling all influence the management of such projects. [3] Here we propose a methodology to adapt agile analytic practices to big data projects based on an iterative model called ITAM (Iterative and Adaptive Model Technique). The whole process can be spread over several iterations. Each iteration shall comprise of some activities.

Iteration 0: Task 1 Identification of Goals: The initial iteration involves the identification of the stakeholders and the solidification of the business goals. For example, the stakeholders for a typical big data project could be a mix of managers, data scientists, reviewers, business analysts, B2Bcustomers, B2C customers and so on. Business goals to be met can be reducing risk, bringing about operational excellence, improving brand image, improving service, increasing sales and revenue etc. After the stakeholders and business goals have been determined, each business goal is explained using a set of user stories. Based on the user stories a set of quality attributes and system constraints are decided upon for every business goal.

Task 2: Lay out a roadmap. Schedules and teams for each phase such as pre-production, pilot, testing etc. must be laid out. Estimations of various parameters such as system volumes, load balancing and capacity require expertise as Big Data Analytics systems process data differently Expertise and experience must be leveraged to create an effective roadmap.

Iteration 1: Task 1 Identification of Big Data Model: One of the big data models may be identified such as kappa, lambda or polyglot persistence etc. Lambda architecture is a data-processing architecture that is designed to handle massive quantities of data by taking advantage of both batch- and stream-processing methods. This step i.e. selection of appropriate architecture is key to achieving agility. Collaboration is an important to solve the design problem at this stage. Designs must not be based on negotiation but collaboration.

Task 2 Instantiate the model: Various technologies both open source and proprietary may be combined as per the needs and budget of the organization. This step also depends on the training of the team working on the project. Once again expertise in a particular technology must be taken into account but of more importance is the choosing a more well-suited technology and training the team accordingly. Appropriate buffer periods for such activities may be allocated during iteration 0 for this purpose.

Iteration 2: Task 1 Decide on the learning/analyzing strategy and perform a POC-level analysis on sample data: Processing or storage of data alone cannot bring the business value expected of the application**.** A strategy for learning may be decided upon but can be kept flexible with respect to any hardware or technological constraint. POCs and Pilots play an essential role in determining the feasibility of a particular strategy. These may be early indicators of a flaws in the proposed model and must be performed with full diligence **Task 2 Review the pilot results to determine if the model adds to the business value**: Proof of concepts and token analysis of the data will indicate whether the model/models brings any significant addition to the business value or helps in achieving the business goal. The results must be jointly reviewed by every stakeholder to determine the next course of action.

Iteration 3: Task 1 Rebuild model architecture according to results of POC-level analysis: The results of the pilot phase shall determine whether the model has to be rebuilt or if any other modifications have to be made. Fine tuning of the model may need to be done such as replacing a certain technology or modifying a learning method incase the system involves it. This is important to make sure the project gives the results expected. The model can be changed at this stage to suit the budget or schedule.

Task 2 Evaluate model again: If the model is modified or changed in this iteration, then it must be evaluated again and especially checked for eventualities like schedule overrun or cost overrun etc.

Iteration 4 Fail Fast: Agility is about failing fast. This helps us to rule out methods that don't reap the desired results. But in a Big Data Analytics scenario where expensive infrastructure and many petabytes of data may be involved, failing fast translates as failing better and failing cheaply. Big Data initiatives are different from other projects as they are more of an exploration exercise. In order to fail fast, the project teams must perform this exploration in a faster and more efficient manner. For example, instead of waiting for the data generated in the day to day operations to be warehoused and made ready for analysis, they can tap the data at the source to perform analysis and speedily determine the suitability and success of the method followed

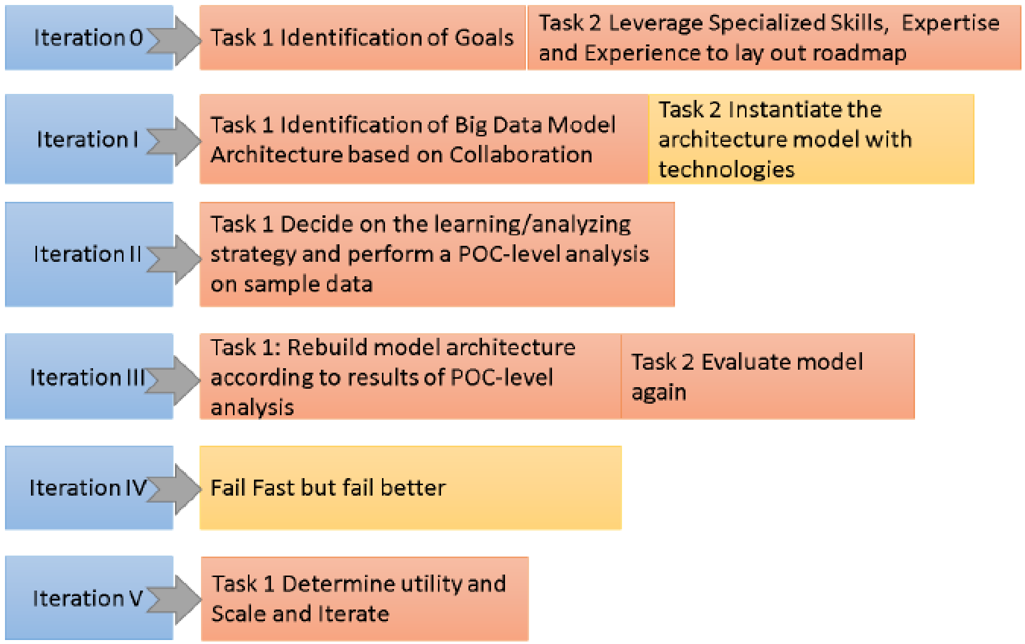

Fig 2 Lending Agility to Big Data through Iterative Adaptive Model Technique (ITAM)

Iteration 5: Determine utility, scale and iterate

If the model is found to be appropriate and the results generated in iteration 3 and 4 in tandem with the business goals, then preparations can be made to scale the proposed model. The iterative model discussed above results in the development of a suitable model for implementation and schedule map for the build phase and subsequent scaling.

IV. BENEFITS, PITFALLS AND CHALLENGES

Agile approaches lead to higher project success rates. This is because of the iterative discovery process. Agile practices help to overcome scope inflexibilities, schedule and budget constraints. The project starts with a small environment and is then scaled to suit the requirements. However, the iterative method proposed has some challenges of its own too. [12]

The benefits that come from following an incremental model such as ITAM is that it provides better risk management, prevents schedule and cost overruns to a degree and leads to greater chances of customer satisfaction and project success. Also, it Since customer input is solicited at regular intervals the project is more accountable and hence better managed too.

For all the successes of agile methodologies, there are scenarios where they shall fail or deliver sub-optimal results. The following are some of these cases.

Lack of training or skills: Tools and methods developed for regular software development projects may not work when applied to an analytics project. If the project team lacks and exposure to the technologies to be used in the project, then this may lead to schedule overruns. The management will have to choose a less optimal option to which the team possesses sufficient training over a more appropriate alternative because of lack of time for training.

Unrealistic Goals: The hype around big data analytics and can sometimes result in some unrealistic goals being set for the project. Most big data based analytics projects require time to reap high value results.

Unclear Business Use Case: Without a clear use case being defined in the initial phases, the project suffers in the downstream iterations. As there is no clear end goal in mind, the technologies to be used in the implementation may be chosen randomly and there may be no clear direction to proceed in at several junctions in the project. Additionally, because of the absence of a clear use case, there might be no way to evaluate results obtained at a stage.

V. CONCLUSION

Agile Analytics is all about striking a balance between discipline and flexibility. In fact, agile helps balance the right amount of structure and formality against a sufficient amount of flexibility, with a constant focus on building the right solution. [10] Big Data analytics must also become agile as various processes in the industry adapt to become agile. An agile methodology that is adapted to the big data scenario that iteratively improves the model in order to achieve the project goal can be implemented with success in cases when the business objectives are clear and the team possess sufficient skills and training. ITAM is one such technique that assists the naturally iterative discovery process in a big data analytics project. The model architecture evolves through many iterations before it is deemed sufficiently suitable to realize the business expectations set for the project. This is a largely plan based method which delivers impact mainly by eliminating risks.

VI. REFERENCES

1. S. Bellomo, I. Gorton, and R. Kazman, "Insights from 15 Years of ATAM Data: Towards Agile Architecture", IEEE Software, to appear, 2015.
2. Boehm BW. A spiral model of software development and enhancement. Computer, 1988, 21.5: 61-72.
3. Hong-Mei Chen ; Rick Kazman ; Serge Haziyev, Agile Big Data Analytics Development: An Architecture-Centric Approach DOI 10.1109/ TBDATA.2016.2564982, IEEE Transactions on Big Data
4. Ken Collier, Agile analytics: A Value-Driven Approach to Business Intelligence and Data Warehousing
5. Tiffani Crawford, Big Data Analytics Project Management: Methodologies, Caveats and Considerations
6. Seth Earley , Agile Analytics in the Age of Big Data 1520-9202/14/ © 2014 IEEE
7. Patrícia Franková, Martina Drahošová, Peter Balco, 2016 Agile Project Management Approach and its Use in Big Data Management Procedia Computer Science 83 (2016) 576 – 583
8. Manjul Gupta, Joey F. George, Towards the developement of big data analytics capability, Information & Management ISSN 0378-7206
9. David Jonker, Scott Langevin, Peter Schretlen, Casey Canfield , Agile Visual Analytics for Banking Cyber "Big Data" IEEE Symposium on Visual Analytics Science and Technology 2012 October 14 - 19, Seattle, WA, USA 978-1-4673-4753-2/12/$31.00 ©2012 IEEE
10. Russell Jurney, Agile Data Science ISBN 9781449326265
11. Nan Li, Anthony Escalona, Yun Guo, Jeff Offutt A Scalable Big Data Test Framework 978-1-4799-7125-1/15 IEEE
12. Cristiano Xaviera, Fernando Moreira, Agile ETL, Conference on ENTERprise Information Systems / PRojMAN 2013 - International Conference on Project Management / HCIST 2013 - International Conference on Health and Social Care Information Systems and Technologies