Big Data Failures and Preventions

Marcelo Fernandes
Software Engineer
Wellington, New Zealand
marceelofernandes@gmail.com

I. Introduction

The term "Big Data" is deluging our society and businesses. Big Data solutions have achieved success in the most complex and challenging problems of contemporaneity, from geography [1], to healthcare [2], social studies [3], and more [4,5]. In contrast, Big Data projects can become too ambitious and fail to achieve their goals. In 2019, Gartner claimed that 80% of analytic insights would fail to deliver business outcomes throughout 2022, and that throughout 2020, 80% of AI projects (often built on top of Big Data resources), "will remain alchemy, run by wizards"[6]. This failure to deliver can be caused by management resistance, undefined goals, lack of middle management, undermined client relationship, high-investment-long-turn-over issues, and more [7,8]. Some of these causes of failure are discussed in this paper, along with their possible solutions.

November 29, 2020

II. Lack of Resources and Skills

Building a team with the necessary skillset in Big Data is the first challenge. This challenge will involve investing in skills outside the business area of expertise and orchestrating the money and time resource available for the project. Amazon Web Services has stated that the costs involved in running an in-house data warehouse range from between $19,000 and $25,000 per terabyte per year [9]. This means that for a modest warehouse containing 50TB of data, costs can reach up to 1.2 million dollars per year. Big Data projects are costly, and they don't guarantee financial return. DeepMind, for instance, lost more than $1 billion dollars between 2016 and 2018 [10]. In 2015, Gardner predicted that 60% of Big Data projects would be abandoned [12]; one of the reasons is poor investment in tools and skills. In order to mitigate problems due to lack of resources and skills, an enterprise should take into account:

- *Outsourcing the work and buying app solutions when lacking analytics expertise*: Organisations may think that a centralised BI team will extend the use of their tools and build advanced analytics tools themselves. However, this would involve a high initial cost, whereas contracting out or using packaged app solutions might be good enough to achieve small wins without compromising the budget and resources.
- *Deciding whether to build the necessary skills and tools internally*: If advanced analytics is not a differentiator in the business or if analytics won't benefit multiple business use-cases, then it does not make sense to build internally [11].

III. Management and Planning

Companies that achieve digital maturity are 26% more profitable than their peers [12]. This creates high demand for digital analytics. Executives are often pressured by vendors to incorporate Big Data analytics before even identifying which problems can be solved, or even which problems need solving in the first place. This pressure culminates in a lack of direction, and without the information about which areas of the business can effectively be improved by Big Data analytics, the chances of success are small.

Another problem affecting Big Data projects is management failure. Senior managers often cannot effectively communicate with the data scientists and engineers running the project. A study presented by David Becker on the leading causes of Big Data project failures [13] was further categorised [14] exposing that 62% of Big Data failures were caused by project management issues. In the UK, one instance that is considered the worst mismanagement in IT ever seen, cost more than £10 billion to taxpayers [15]. The patient record system that aimed to unify all the data in a central database was entirely abandoned before its completion.

Management and planning issues can be fixed by applying simple but tough measures. Upper business executives must elaborate objective problem statements. These need to go through a data-driven process to thrash out the planning and priorities. The next step is to initiate a top-down approach to make the organisation "data-ready", which would facilitate the strategy for taking advantage of quick wins and also aid long-term projects to be conducted effectively.

IV. Conclusion

With the complexity of Big Data projects and vendors' pressure for companies to implement Big Data analytics, the forecasted failure rates for Big Data projects will remain high. Absence of a clear business case, ineffective team coordination, poor communication and incorrect budgeting are problems that fall into the mismanagement category. These problems are responsible for the majority of failures [14] related to Big Data projects and are relatively straightforward to solve by applying good management practices and with the correct leadership. Incorrect use of technology, incorrect

infrastructure, lack of Big Data and analytics skills are problems that fall into the resources and skillset category. These problems can be more difficult to resolve, but outsourcing and understanding when to build the tooling and skills internally can affect the project success and minimise the impact of technical issues.

REFERENCES

[1] R. Kitchin, "Big data and human geography", Dialogues in Human Geography, vol. 3, no. 3, pp. 262-267, 2013. Available: 10.1177/2043820613513388 [Accessed 29 November 2020].

[2] J. Andreu-Perez, C. Poon, R. Merrifield, S. Wong and G. Yang, "Big Data for Health", IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 4, pp. 1193-1208, 2015. Available: 10.1109/jbhi.2015.2450362 [Accessed 29 November 2020].

[3] L. Harlow and F. Oswald, "Big data in psychology: Introduction to the special issue.", Psychological Methods, vol. 21, no. 4, pp. 447-457, 2016. Available: 10.1037/met0000120 [Accessed 29 November 2020].

[4] L. Harlow and F. Oswald, "Big data in psychology: Introduction to the special issue.", Psychological Methods, vol. 21, no. 4, pp. 447-457, 2016. Available: 10.1037/met0000120 [Accessed 29 November 2020].

[5] E. Ewing, V. Kimmerly and S. Ewing-Nelson, "Look Out for 'La Grippe': Using Digital Humanities Tools to Interpret Information Dissemination during the Russian Flu, 1889–90", Medical History, vol. 60, no. 1, pp. 129-131, 2015. Available: 10.1017/mdh.2015.84 [Accessed 29 November 2020].

[6] "Our Top Data and Analytics Predicts for 2019 - Andrew White", Andrew White, 2020. [Online]. Available: https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/. [Accessed: 29- Nov- 2020].

[7] [7]"Why most big data analytics projects fail", Volume 46, Number 6, December 2019, 2019. Available: 10.1287/orms.2019.06.08 [Accessed 29 November 2020].

[8] A. Patrizio, "4 reasons big data projects fail—and 4 ways to succeed", InfoWorld, 2020. [Online]. Available: https://www.infoworld.com/article/3393467/4-reasons-big-data-projects-failand-4-ways-to-succeed.html. [Accessed: 29- Nov- 2020].

[9] "Amazon Redshift – The New AWS Data Warehouse — Amazon Web Services", Amazon Web Services, 2020. [Online]. Available: https://aws.amazon.com/blogs/aws/amazon-redshift-the-new-aws-data-warehouse/. [Accessed: 29- Nov- 2020].

[10] [10]G. Marcus, "DeepMind's Losses and the Future of Artificial Intelligence", Wired, 2020. [Online]. Available: https://www.wired.com/story/deepminds-losses-future-artificial-intelligence/. [Accessed: 29- Nov- 2020].

[11] "Gartner Says Business Intelligence and Analytics Leaders Must Focus on Mindsets and Culture to Kick Start Advanced Analytics", Gartner, 2020. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2015-09-15-gartner-says-business-intelligence-and-analytics-leaders-must-focus-on-mindsets-and-culture-to-kick-start-advanced-analytics. [Accessed: 29- Nov- 2020].

[12] "Digitally Mature Firms are 26% More Profitable Than Their Peers — MIT Initiative on the Digital Economy", Ide.mit.edu, 2020. [Online]. Available: http://ide.mit.edu/news-blog/blog/digitally-mature-firms-are-26-more-profitable-their-peers. [Accessed: 29- Nov- 2020].

[13] D. Becker, "Predicting outcomes for big data projects: Big Data Project Dynamics (BDPD): Research in progress", 2017 IEEE International Conference on Big Data (Big Data), 2017. Available: 10.1109/bigdata.2017.8258186 [Accessed 29 November 2020].

[14] "Data Science Project Failures - Data Science Project Management", Data Science Project Management, 2020. [Online]. Available: https://www.datascience-pm.com/project-failures/. [Accessed: 29- Nov- 2020].

[15] "Abandoned NHS IT system has cost £10bn so far", the Guardian, 2020. [Online]. Available: https://www.theguardian.com/society/2013/sep/18/nhs-records-system-10bn. [Accessed: 29- Nov- 2020].