Portfolio Task 3 – Big Data Failures

Word Count -788

# Introduction

Taylor [1] describes big data as data that grows exponentially over time, so that traditional data management tools cannot process it efficiently. However, big data projects have seen some failures since its adoption. Gartner’s initial prediction that 60% of big data projects would fail now revised upwards to 85%, according to several sources [2, 3, 4]. Firstly, this report aims to identify what big data failures have transpired in the past. Secondly, analysing why these failures came to fruition. Finally, to learn from past failures to avoid future ones.

# People

Burman [5, p. 321] describes one reason for failure in big data projects, are people placed on them. Despite this, Burman does not seem to delve further, aside from a description of the wrong human resources brought onto the project by wrong people. ITPro [6] mentions a study from the Royal Society back in 2019 [7] found that demand for big data roles such as Scientists and Engineers tripled over five years, but not enough candidates to fill these roles. Hotz [8] attempts to examine the difficulty of hiring people with a correct combination of technical and soft skills required to conduct such a project. Saltz [9], agreed mostly by Dietrich, Heller, and Yang [10, pp. 26–27] explains further cross-disciplinary roles required for a Big Data project to have the most opportunity to be successful, such as Data Scientists, Engineers, Architects, Analysts, and Product Owners to name a few. Miller [11] concurs, stating whilst Data Scientists are important for these projects, they are not the only skills desired. Incentives such as training of internal staff on big data practices [12] or recruiting qualified specialists in these fields [13] could help fill this gap.

# Bad Data

Another reason for failures found within big data projects, is bad data. Berman expresses this as inaccurate, underrepresented, or underspecified data [14, p. 322]. Again, this does not go further at this point to help specify what that means for a big data project. An example comes from Hillier [15] where the misuse of data caused a self-driving Uber vehicle to kill a woman, due to the AI not understanding the data the classifies a person, when not by the sidewalk, or crosswalk [16].For obvious reasons, this is something that cannot be acceptable in a big data project to be classed as successful, where the improper use or incorrect data has a negative outcome; especially where big data can come from such diverse range of sources including photos and videos, medical devices, smart devices notably [17, pp. 15–16]. Another term connotated with bad data, refers to a quote used by Kilkenny and Robertson [18] “Garbage in – garbage out”. They allude to concerns ensuring data is accurate, validated, complete, and available, before use. A lesson to learn as a result, should ensure adequate data, that the data can describe itself, only use data that is relevant to the project, perform external validation before application [19, pp. 330–332]. Whilst big data is bringing a new age of opportunities, having more data helps, but having the right data is still key [20, p. 35].

# Planning

A further element relates to poor planning. Ailments can include setting unrealistic goals and expectations, poor project management, and not understanding the business problem [21]. Aggarwal and Narisi [22, 23] argue that companies underestimate the amount of time, resources, scope, required to complete big data projects. An example of poor planning comes from a case study of the Google Flu Trends, with the aim to predict flu outbreaks. However, the logic was flawed, inaccurate, and missed real flu epidemics found [24]. Whilst claims that this case study could be an example of both a failure of people and data, bad planning is also accountable. Google should have been able to plan this project ahead with the right milestones place, ensuring the service is to an acceptable standard, before getting rolled out for use. The lessons to learn for future big data projects include ensuring the right people with the right skills are in place, the correct time allocated to the projects, the correct data evaluated as fit for purpose, and the project continuously validates from external parties before a roll out.

# Conclusion

Although it may be obvious that technology will have a large factor to play in terms of the success or failure of a big data project, there are many influences that will have an impact. Items discussed above are not an exhaustive list of what to consider. People, data, and planning, however, play fundamental roles in contributing towards a successful big data project. Lessons are learnable from previous big data projects that assist in the knowledge of what should and should not execute going forward.

# References

[1] D. Taylor, “What is BIG DATA? Introduction, Types, Characteristics, Example,” www.guru99.com, Jan. 22, 2022. https://www.guru99.com/what-is-big-data.html#6 (accessed Mar. 27, 2022).

[2] M. Henrion, “Why most big data analytics projects fail,” pubsonline.informs.org, Dec. 04, 2019. https://pubsonline.informs.org/do/10.1287/orms.2019.06.08/full/ (accessed Mar. 27, 2022).

[3] M. Asay, “85% of big data projects fail, but your developers can help yours succeed,” TechRepublic, Nov. 10, 2017. https://www.techrepublic.com/article/85-of-big-data-projects-fail-but-your-developers-can-help-yours-succeed/ (accessed Mar. 27, 2022).

[4] B. O’Neill, “Failure rates for analytics, AI, and big data projects = 85% – yikes! | Designing for Analytics (Brian T. O’Neill),” Designingforanalytics.com, Jul. 23, 2019. https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/ (accessed Mar. 27, 2022).

[5] J. J. Berman, Principles and practice of big data preparing, sharing, and analyzing complex information, 2nd ed. London Cambridge, Ma Academic Press, An Imprint Of Elsevier, 2018, p. 321.

[6] ITPro, “Data scientist jobs: Where does the big data talent gap lie?,” IT PRO, Sep. 17, 2019. https://www.itpro.co.uk/careers/28929/data-scientist-jobs-where-does-the-big-data-talent-gap-lie (accessed Mar. 27, 2022).

[7] R. Society, “Growing demand for data science leaves Britain vulnerable to skills shortages | Royal Society,” Royalsociety.org, May 09, 2019. https://royalsociety.org/news/2019/05/data-science-skills-shortages/ (accessed Mar. 27, 2022).

[8] N. Hotz, “Why Big Data Science & Data Analytics Projects Fail,” Data Science Project Management, Feb. 13, 2021. https://www.datascience-pm.com/project-failures/ (accessed Mar. 27, 2022).

[9] J. Saltz, “8 Key Data Science Team Roles,” Data Science Process Alliance, Jan. 29, 2020. https://www.datascience-pm.com/8-key-roles-within-a-data-science-project/ (accessed Mar. 27, 2022).

[10] D. Dietrich, B. Heller, and B. Yang, Data science & big data analytics : discovering, analyzing, visualizing and presenting data. Indianapolis, In: Wiley, 2015, pp. 26–27.

[11] S. Miller, “Collaborative Approaches Needed to Close the Big Data Skills Gap,” papers.ssrn.com, Apr. 10, 2014. https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=2458930. (accessed Mar. 27, 2022).

[12] E. R. Tauchman, “4 Big Data Challenges and How to Overcome Them,” Default, May 15, 2019. https://www.comptia.org/blog/4-big-data-challenges-and-how-to-overcome-them (accessed Mar. 27, 2022).

[13] nix, “Big Data Challenges and Best Ways Out – NIX United,” NIX United – Custom Software Development Company in US, Jan. 19, 2022. https://nix-united.com/blog/12-big-data-issues-growing-companies-face/ (accessed Mar. 27, 2022).

[14] J. J. Berman, Principles and practice of big data preparing, sharing, and analyzing complex information, 2nd ed. London Cambridge, Ma Academic Press, An Imprint Of Elsevier, 2018, p. 322.

[15] W. HIllier, “Is Big Data Dangerous?,” careerfoundry.com, Sep. 20, 2021. https://careerfoundry.com/en/blog/data-analytics/is-big-data-dangerous/ (accessed Mar. 27, 2022).

[16] P. McCausland, “Self-driving Uber car that hit and killed woman did not recognize that pedestrians jaywalk,” NBC News, Nov. 09, 2019. https://www.nbcnews.com/tech/tech-news/self-driving-uber-car-hit-killed-woman-did-not-recognize-n1079281 (accessed Mar. 27, 2022).

[17] D. Dietrich, B. Heller, and B. Yang, Data science & big data analytics : discovering, analyzing, visualizing and presenting data. Indianapolis, In: Wiley, 2015, pp. 15–16.

[18] M. F. Kilkenny and K. M. Robinson, “Data quality: ‘Garbage in – garbage out,’” Health Information Management Journal, vol. 47, no. 3, pp. 103–105, May 2018, doi: 10.1177/1833358318774357.

[19] J. J. Berman, Principles and practice of big data preparing, sharing, and analyzing complex information, 2nd ed. London Cambridge, Ma Academic Press, An Imprint Of Elsevier, 2018, pp. 330–332.

[20] J. D. Kelleher and B. Tierney, Data science. Cambridge, Massachusetts ; London, England: The Mit Press, 2018, p. 35.

[21] S. Koegler, “Don’t Let These Examples of Big Data Failures Trip You Up - Big Data & Analytics Tech Brief,” www.bdatechbrief.com, Nov. 30, 2020. https://www.bdatechbrief.com/index.php/bigdata/item/6894-don-t-let-these-examples-of-big-data-failures-trip-you-up (accessed Mar. 27, 2022).

[22] S. Aggarwal, “Why the Majority of Big Data Projects Fail,” Qubole, Nov. 30, 2021. https://www.qubole.com/blog/big-data-projects-fail/ (accessed Mar. 27, 2022).

[23] S. Narisi, “Why 55% of Big Data projects fail – and what IT can do about it,” IT Manager Daily, Feb. 13, 2013. https://www.itmanagerdaily.com/big-data-projects-fail/ (accessed Mar. 27, 2022).

[24] N. D’Mello, “5 cases where big data was a big flop,” CIO, Jun. 14, 2016. https://www.cio.com/article/218157/5-cases-where-big-data-was-a-big-flop.html (accessed Mar. 27, 2022).