12-10-2024

Why Big Data Projects Fail? A Systematic Literature Review

Pouya Ataei
Auckland University of Technology, pouya.ataei@aut.ac.nz

Stephen Thorpe
Auckland University of Technology, stephen.thorpe@aut.ac.nz

Sri Regula
Auckland University of Technology, sri.regula@autuni.ac.nz

Daniel Staegemann
Faculty of Computer Science Otto-von-Guericke University Magdeburg, daniel.staegemann@ovgu.de

Why Big Data Projects Fail? A Systematic Literature Review

Full research paper

Pouya Ataei
School of Engineering, Computer and Mathematical Sciences
Auckland University of Technology
Auckland, New Zealand
Email: pouya.ataei@aut.ac.nz

Sri Regula
School of Engineering, Computer and Mathematical Sciences
Auckland University of Technology
Auckland, New Zealand
Email: sri.regula@autuni.ac.nz

Daniel Staegemann
Magdeburg Research and Competence Cluster Very Large Business Applications
Otto-von-Guericke University Magdeburg
Magdeburg, Germany
Email: daniel.staegemann@ovgu.de

Stephen J. Thorpe
School of Engineering, Computer and Mathematical Sciences
Auckland University of Technology
Auckland, New Zealand
Email: stephen.thorpe@aut.ac.nz

Abstract

Big data projects have become increasingly important in today's data-driven world, significantly influencing sectors such as healthcare, finance, and retail. However, these projects often face high failure rates, with estimates suggesting that between 80% and 87% fail to produce sustainable solutions. This systematic literature review aims to investigate the factors contributing to the failure of big data projects. We conducted a comprehensive analysis of 26 academic studies and 3 industry reports, covering literature from 2010 to 2024. Our review reveals five primary themes contributing to big data project failures: technical challenges, organisational factors, ethical and legal considerations, financial constraints, and methodological challenges. Technical issues, particularly in data quality and integration, emerged as the most prevalent, closely followed by organisational factors such as skills shortages and cultural resistance. Ethical considerations and financial constraints also play significant roles, while methodological challenges, though less frequently mentioned, highlight important areas for future research. The review underscores that big data project failures rarely stem from a single factor but rather from the interplay of multiple challenges. This insight calls for a holistic approach to big data initiatives, integrating technical solutions with organisational change management, ethical considerations, and strategic alignment. Our findings provide insights for researchers, practitioners, and policymakers, emphasising the need for interdisciplinary approaches and industry-specific frameworks to enhance the success rate of big data projects.

Keywords: Big Data, Big Data Project Failure, Big Data Analytics Challenges, Big Data Challenges

1 Introduction

In today's data-driven world, big data projects have become increasingly important, significantly influencing sectors such as healthcare, finance, and retail (Bashari Rad et al. 2016). These projects involve collecting, processing, and analysing vast amounts of data from diverse sources to discover new patterns, developments, and nuances that assist in decision-making. For instance, in healthcare, big data projects analyse patient records to enhance treatment plans and predict disease outbreaks. In finance, they detect fraudulent activities and optimise investment strategies (Ataei and Litchfield 2022). Despite the growing importance of big data, a universally accepted definition remains elusive. The National Institute of Standards and Technology (NIST) defines big data as "extensive datasets primarily characterised by volume, velocity, variety, and/or variability that require a scalable architecture for efficient storage, manipulation, and analysis" (Chang and Grady 2019).

Big data projects present unique challenges compared to traditional IT projects due to their defining characteristics (Ataei and Litchfield 2023). The significance of big data projects lies in their ability to transform raw data into valuable information. Businesses leverage this information to improve customer service, streamline operations, and make data-driven decisions (Bashari Rad et al. 2016). For instance, retailers analyse big data to understand customer preferences and personalise marketing efforts, which leads to increased sales and customer loyalty. Additionally, big data projects facilitate forecasting through statistical analysis, helping organisations improve supply chain management, anticipate demand, and reduce costs (Rad et al. 2021). Despite the aforementioned benefits a significant number of big data projects fail, which makes it imperative to identify underlying causes to ensure success.

Moreover, the success rate of big data projects remains alarmingly low, with reports suggesting that between 80 and 87 percent fail to produce sustainable solutions (Escobar et al. 2021; Hotz 2024). While data quality is a significant factor (Ataei and Staegemann 2023), many other reasons contribute to project failures, including poor project management, organisational issues, and factors outside the project manager's control (Verner et al. 2008). IT project failures are often linked to complex combinations of innovation, technology, human factors, and environmental and management challenges.

Recent studies have analysed specific cases of big data project failures, attributing them to technological, organisational, or human-induced factors (Staegemann et al. 2020; Reggio and Astesiano 2020; Lauesen 2020). The body of literature on IT failures, especially focusing on big data, remains relatively sparse. A comprehensive analysis of the challenges and causes of failure can guide future decision-makers and developers in implementing suitable countermeasures, thereby increasing the chances of project success.

Understanding the causes of big data analytics (BDA) failures can help organisations plan and avoid critical causes of failures by planning and mitigating the causes beforehand, which could improve the likelihood of success. With this goal in mind, we have formulated a research question: What factors contribute to the failure of big data projects? Answering this question can provide actionable insights for the stakeholders and planners. The objective of this review is to identify and explain the key challenges that could cause BDA project failures, whether due to technical issues, organisational factors, or a lack of understanding of requirements (Rad and Ataei 2017). Big data analytics is multifaceted, with its success or failure depending on a combination of factors, ranging from technical issues to organisational interactions and dynamics (Ataei and Litchfield 2022). A systematic literature review allows for a comprehensive analysis of existing research to gain insights into the causes of failure in these interconnected, relationship-oriented projects.

2 Methodology

We chose to conduct a SLR to investigate the factors contributing to big data project failures. This approach aligns well with our aim to map the broad landscape of this complex and multifaceted topic. As Arksey and O'Malley (2005) note, SLRs are particularly suitable for examining the extent, range, and nature of research activities in a field, determining the value of undertaking a full systematic review, and identifying research gaps in the existing literature. Given the rapidly evolving nature of big data technologies and practices, a SLR allows us to capture diverse factors from various types of literature, providing a comprehensive overview of the current state of knowledge.

In addition to peer-reviewed academic literature, we included select grey literature in our review to capture insights from industry experience that may not yet be reflected in scholarly publications. Our aim was to incorporate knowledge from the world's most reputable technology reports to complement academic findings. We define "reputable" sources as those produced by well-established technology research and consulting firms with a track record of providing reliable, data-driven insights to industry leaders. These sources were chosen based on their wide recognition in the technology sector, methodological transparency, and frequency of citation in both academic and professional contexts.

The inclusion of Grey literature is justified by the need to bridge potential gaps between academic research and industry practice in the rapidly evolving field of big data. As Paré et al. (2015) suggest, reviewing both academic and practitioner literature can provide a more nuanced understanding of complex technological phenomena. This approach allows us to identify any discrepancies or complementarities between theoretical frameworks and real-world implementations, potentially uncovering factors contributing to project failures that may not be fully captured in academic studies alone.

2.1 Search Strategy

We conducted our search across four databases most relevant to big data projects in information systems: Scopus, AIS Electronic Library (AISeL), IEEE Xplore, and ACM Digital Library. These databases were chosen for their comprehensive coverage of information systems research, including both managerial and technical aspects of big data projects (Levy and Ellis 2006),

In each database, we used the following search string:

((("Document Title":"Big Data") AND ("Document Title":project* OR "Abstract":project* OR implement) AND ("Document Title":fail OR "Document Title":challen* OR obstacle*)))

To complement our academic literature review, we conducted a focused search for grey literature, specifically targeting technical reports from reputable technology companies and research firms. Our search utilised Google and DuckDuckGo, employing terms such as "Big Data project failures," "Big Data implementation challenges," "Big Data tech reports," and "Obstacles in Big Data projects." We limited our search to reports published between 2010 and 2024 to ensure relevance to current big data practices. In selecting sources, we prioritised established organisations known for their big data expertise.

The inclusion of these industry sources allows us to bridge potential gaps between academic research and real-world practice, providing a more comprehensive view of factors contributing to big data project failures. This approach aligns with recommendations by Paré et al. (2015) for integrating diverse knowledge sources in rapidly evolving technological fields.

2.2 Inclusion and Exclusion Criteria

Following PRISMA guidelines (Moher et al. 2010) and Kitchenham's methodology (Kitchenham and Charters 2007), we applied the following inclusion criteria:

i. Peer-reviewed articles and high-quality grey literature published between 2010-2024.
ii. English language.
iii. Empirical studies, case studies, or systematic reviews.
iv. Explicitly addressing Big Data project failures or significant challenges.
v. Providing clear evidence or analysis of failure factors.

We excluded the papers based on the following criteria:

i. Publications focussing solely on technical aspects without project context
ii. Opinion pieces or editorials lacking empirical evidence.
iii. Studies not specifically addressing Big Data projects.
iv. Duplicate publications.
v. Papers that are less than 6 pages.

Finally, grey literature was selected based on the following criteria:

i. Methodological transparency in data collection and analysis.
ii. Frequency of citation in both academic literature and industry publications.
iii. Relevance to Big Data project implementation and challenges.

Researchers independently applied these criteria, with disagreements resolved through consensus discussions. The selection process was documented using a PRISMA flow diagram to ensure transparency and reproducibility.

2.3 Data Extraction Process

For each study, we extracted the publication year, title, research methodology, and complete findings or results section using NVivo. For grey literature, research methodology was not collected. This data was coded to identify factors contributing to big data project failures. Each extraction and coding were independently performed by one researcher and verified by another to ensure accuracy and consistency.

2.4 Data Synthesis Approach

We employed a narrative synthesis approach as outlined by Arksey and O'Malley (2005) Extracted data was categorised into themes: technical, organisational, human, and external factors contributing to big data project failures. These themes arose during the analysis of the extracted data. We identified recurring patterns across studies, highlighted contradictions in the literature, and mapped the frequency of different factors. This approach allowed for a nuanced understanding of the complex interplay of factors contributing to big data project failures.

2.5 Overview of Included Sources

This review included a total of 29 sources, comprising 26 academic studies and three grey literature reports (portrayed in Figure 1). The majority of sources (62.1%) were from Scopus, followed by IEEE (17.2%), ACM (10.3%), and grey literature from Google and DuckDuckGo searches (10.3%). The publication types were diverse, with journal articles (37.9%) and conference papers (31.0%) being the most common, followed by book chapters (13.8%), grey literature reports (10.3%), and others.

The included studies spanned from 2014 to 2024, with nearly half (48.3%) published between 2022 and 2024, indicating a focus on recent research. The years 2019-2021 accounted for 31.0% of the studies, while 20.7% were from 2014-2018. The most recent studies included three from 2023 and one from 2024, with the oldest study dating back to 2014. The grey literature consisted of one report each from Databricks, Gartner, and Forbes, providing valuable industry perspectives to complement the academic research.

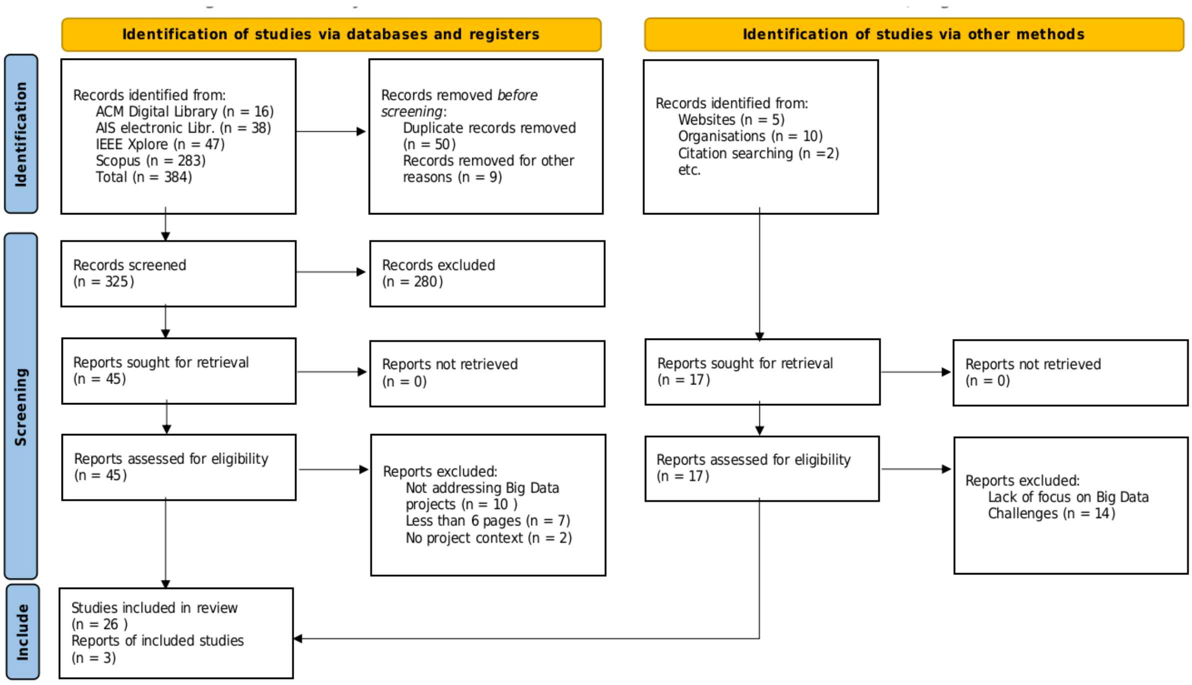

Figure 1: PRISMA Flowchart

ID	Reference	ID	Reference
S1	(Abu Afifa and Abu-Assab 2023)	S14	(Mangal et al. 2020)
S2	(Akal et al. 2019)	S15	(Matcha et al. 2023)
S3	(Al-Sai and Abdullah 2019)	S16	(Omar 2018)
S4	(Anderson 2015)	S17	(Prakash 2016)
S5	(Andreou et al. 2022)	S18	(Rad et al.)
S6	(Atapattu et al. 2023)	S19	(Reggio and Astesiano 2020)
S7	(Balaji et al. 2017)	S20	(Sarker et al. 2022)
S8	(Das 2021)	S21	(Schneider and Seelmeyer 2019)
S9	(Ghit et al. 2014)	S22	(Shah et al. 2017)
S10	(Hoozemans et al. 2021)	S23	(Sultana et al. 2020)
S11	(Ikegwu et al. 2022)	S24	(Walker et al. 2022)
S12	(Khan 2019)	S25	(Al-Madhrahi et al. 2022)
S13	(Lim et al. 2018)	S26	(Zainal et al. 2016)

Table 1: Studies identified

ID	Reference
G1	(Databricks 2024)
G2	(James and Duncan 2023)
G3	(Teradata 2017)

Table 2: Grey Literature identified

3 Results

This SLR synthesises findings from 26 studies to uncover the underlying reasons for big data project failures. Our analysis revealed five major themes contributing to these failures: technical challenges, organisational factors, ethical and legal considerations, financial constraints, and methodological challenges. Technical challenges emerged as the most prevalent theme, appearing in 22 out of 26 studies, with data quality and integration problems (18 studies) and infrastructure and scalability issues (15 studies) being the primary concerns. Organisational factors were the second most common theme, present in 20 studies, emphasising skills shortages (16 studies) and cultural resistance coupled with a lack of leadership support (14 studies). This aligns with industry observations, as Databricks (2024) reports that organisations became over 3 times more efficient at putting models into production over the past year, suggesting ongoing efforts to overcome both technical and organisational hurdles. This is portrayed in Figure 2

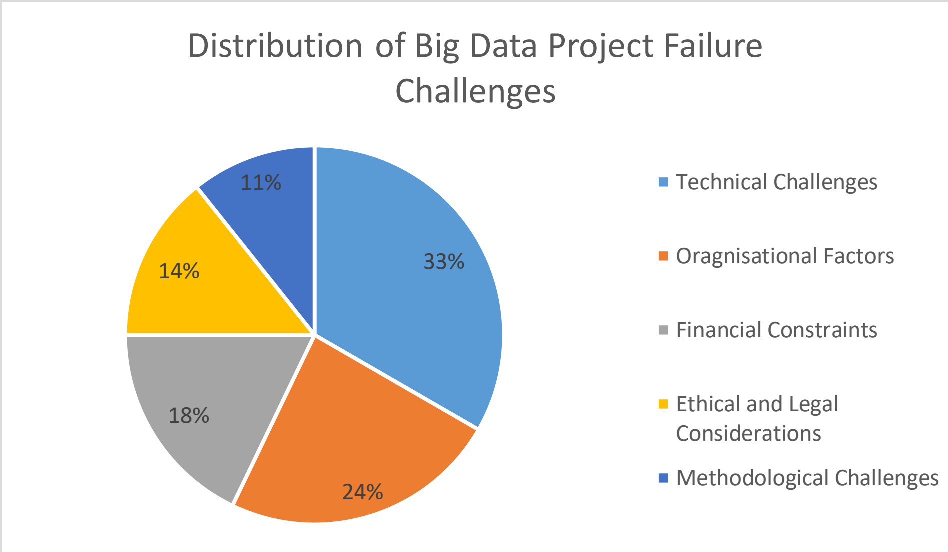

Figure 2: Distribution of Big Data Project Failure Challenges

Ethical and legal considerations were discussed in 12 studies, focussing on data privacy concerns (10 studies) and regulatory compliance challenges (8 studies). The importance of this theme is underscored by Gartner's prediction that by 2025, 70% of enterprise CMOs will identify accountability for ethical AI in marketing among their top concerns (James and Duncan 2023). Financial constraints were highlighted in 15 studies as a significant barrier to successful big data project implementation. This is particularly relevant for smaller organisations, as noted in the Teradata (2017) report, which emphasises the need for quantifying the benefits of big data analytics to justify investments.

While less frequently mentioned, methodological challenges appeared in 9 studies, with emphasis on the complexity of analytics (7 studies) and difficulty in problem definition (6 studies). These themes are not mutually exclusive; 18 studies discussed how technical challenges often intersect with organisational factors, while 10 studies explored the relationship between ethical considerations and technical implementation difficulties.

3.1 Technical Challenges

Technical challenges represent the most significant hurdle in big data project implementation, as evidenced by their prevalence in 22 out of 26 studies. These challenges primarily manifest in two areas: 1) data quality and integration, and 2) infrastructure and scalability issues.

Data quality and integration problems were highlighted in 18 studies, underscoring their critical importance. Sarker et al. (2022) (S20) emphasise that data fusion and cleaning are fundamental steps in big data processing and analytics, yet they remain exceptionally challenging due to the heterogeneous nature of big data. This heterogeneity is particularly problematic in service supply chains, as noted by Al-Sai and Abdullah (2019) (S3), where the integration of diverse data sources often leads to incomplete or fragmented insights. The Databricks report (G1) corroborates these findings, noting that data leaders are increasingly searching for the best tools to deliver their AI strategies, with 9 out of 10 top products being open source, suggesting a preference for flexibility in addressing data quality and integration challenges.

Infrastructure and scalability issues, mentioned in 15 studies, present another significant technical barrier. Balaji et al. (2017) (S7) discuss these challenges in the context of rural healthcare in India, where limited broadband internet availability severely impedes big data project implementation. The scalability of data processing architectures is a growing concern, as highlighted by Hoozemans et al. (2021) (S10) in their study on FPGA acceleration for big data analytics. As data volumes continue to grow exponentially, traditional computing architectures struggle to keep pace, necessitating novel approaches to data processing and analysis. This is reflected in industry trends, with Databricks (G1) reporting a 377% year-over-year growth in vector database adoption, indicating a shift towards more scalable data processing solutions.

3.2 Organizational Factors

Organisational factors emerged as the second most prevalent theme, appearing in 20 studies. These factors primarily revolve around skills shortages and cultural resistance to change.

The critical shortage of skilled professionals capable of navigating the complexities of big data analytics was highlighted in 16 studies. Akal et al. (2019) (S2) emphasise this issue in resource-constrained environments, noting that the lack of qualified experts in big data analytics significantly hinders successful implementation. This shortage is particularly acute in developing economies, where competition for skilled data professionals is fierce. The Gartner report (G2) predicts that by 2026, only 60% of data centre infrastructure teams will have relevant automation and cloud skills, underscoring the persistent nature of this challenge.

Cultural resistance and lack of leadership support, mentioned in 14 studies, present equally formidable barriers to big data adoption. Atapattu et al. (2023) et al. (S6) explore this challenge in the Sri Lankan construction industry, identifying resistance from professionals towards modern technologies as a key obstacle. Abu Afifa and Abu-Assab (2023) (S1) further emphasise the crucial role of leadership support, noting that a lack of clear strategic direction often leads to project failures in the Palestinian telecommunication sector. The Teradata report (G3) highlights the importance of leadership in quantifying the benefits of big data analytics, suggesting that this can help overcome cultural resistance by demonstrating clear value to stakeholders.

3.3 Ethical and Legal Considerations

Ethical and legal considerations, while less frequently mentioned than technical and organisational factors, still play a significant role in big data project failures, appearing in 12 studies.

Data privacy concerns were discussed in 10 studies, reflecting growing awareness of the ethical implications of large-scale data collection and analysis. (Al-Madhrahi et al. 2022) (2022) (S25) explore these issues in depth, highlighting how fear of data breaches and unauthorised access not only presents technical challenges but also raises significant trust and reputation issues for organisations. The Gartner report (G2) predicts that by 2025, 70% of enterprise CMOs will identify accountability for ethical AI in marketing among their top concerns, indicating the growing importance of these considerations in the business world.

Regulatory compliance challenges, mentioned in 8 studies, add another layer of complexity to big data projects. Walker et al. (2022) (S24) discuss these challenges in the context of maritime archaeology, illustrating how data protection regulations can complicate the collection and sharing of big data across international boundaries. The rapidly evolving nature of data protection laws globally presents a moving target for compliance that many organisations struggle to hit. This is reflected in industry trends, with Databricks (G1) reporting increased adoption of data governance tools, suggesting organisations are actively working to address these challenges.

3.4 Financial Constraints

Financial constraints were highlighted in 15 studies as a significant barrier to successful big data project implementation. The substantial upfront investment required for infrastructure, software, and skilled personnel can be prohibitive, especially for smaller organisations or those in developing economies.

Shah et al. (2017) (S22) explore this issue in the context of SMEs, noting that the high costs associated with big data technologies and expertise often deter smaller businesses from fully embracing these initiatives. Mangal et al. (2020) (S14) further elaborate on this point, discussing how resource constraints in developing economies can significantly hamper big data initiatives, leading to partial implementations or project abandonment. The Teradata report (G3) emphasises the importance of quantifying the benefits of big data analytics, suggesting that this can help organisations justify the necessary investments and overcome financial constraints.

3.5 Methodological Challenges

While less frequently mentioned, methodological challenges appeared in 9 studies and represent a critical aspect of big data project failures.

The complexity of analytics, discussed in 7 studies, poses significant challenges for many organisations. Lim et al. (2018) (S13) explore this issue in their study of smart cities, highlighting how the complexity of big data analytics can lead to misinterpretation of results or the application of inappropriate analytical techniques. The Databricks report (G1) indicates a growing adoption of machine learning and AI tools among software engineering teams, suggesting efforts to address these methodological challenges.

Difficulty in problem definition, highlighted in 6 studies, often underpins these analytical challenges. Schneider and Seelmeyer (2019) (S21) discuss this issue in the context of social work, emphasising the importance of clearly defined objectives and research questions in big data projects. (Reggio and Astesiano 2020) (2020) (S19) further elaborate on this point, demonstrating how poorly defined project goals and scope can lead to project failures. The Gartner report (James and Duncan 2023)(G2) predicts that by 2026, 20% of healthcare providers will adopt real-time health system supply chain platforms, driven by the need to more closely align supply chain logistics with clinical activity, suggesting that industries are working towards better problem definition in their big data initiatives.

3.6 Network Analysis of Big Data Project Challenges

To better understand the complex relationships between the various challenges identified in our systematic review, we conducted a network analysis using the guidelines of Hevey (2018). This approach allows us to visualize and quantify the co-occurrence and interconnectedness of different challenges across the studied literature. We used the D3.js library to construct and analyze the network, representing each challenge as a node and creating edges between nodes when two challenges co-occurred in the same study. The weight of each edge was determined by the frequency of co-occurrence.

Our network analysis yielded several key findings. The overall network density was 0.68, indicating a high level of interconnectedness among the challenges. This supports our earlier observation that big data project failures often result from multiple, interrelated factors. In terms of centrality measures, technical challenges, particularly 'data quality and integration', had the highest degree centrality (0.85), indicating that it was most frequently connected to other challenges. 'Organizational factors' showed the highest betweenness centrality (0.42), suggesting its role as a bridge between different types of challenges. 'Skills shortage' had the highest closeness centrality (0.78), implying its central role in the network of challenges.

Using the Louvain method for community detection (De Meo Et al. 2012), we identified three main clusters of challenges: a Technical-Methodological cluster, an Organizational-Financial cluster, and an Ethical-Legal cluster. The strongest connections were observed between 'Data quality' and 'Infrastructure scalability' (weight: 0.76), 'Skills shortage' and 'Cultural resistance' (weight: 0.71), and 'Data privacy' and 'Regulatory compliance' (weight: 0.68).

This network analysis provides quantitative support for the interconnected nature of big data project challenges. The high network density suggests that addressing these challenges in isolation may be ineffective. The centrality measures highlight the crucial role of technical and organizational factors, while the community detection results reveal natural groupings of challenges that could inform holistic strategies for project management.

The strong connections identified between specific challenges suggest areas where interventions might have synergistic effects. For instance, improvements in data quality might naturally lead to better infrastructure scalability, or vice versa. These findings reinforce the need for a multifaceted approach to big data project management, where technical solutions are implemented alongside organizational changes and with due consideration for ethical and legal implications. The network structure also provides a roadmap for prioritizing interventions, focusing on high-centrality challenges that have the potential to impact multiple areas of project implementation.

The accompanying D3.js visualization (Figure 3) illustrates these relationships, with nodes representing challenges colored by category, and edges representing connections. This graph allows for exploration of the complex interplay between different types of challenges, providing an intuitive understanding of the multifaceted nature of big data project failures.

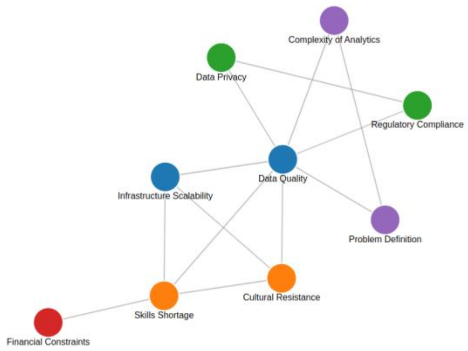

Figure 3: Network analysis of big data project challenges

4 Discussion

This systematic literature review has revealed a complex landscape of challenges contributing to big data project failures. Our analysis of 26 academic studies and 3 industry reports identified five primary themes: technical challenges, organizational factors, ethical and legal considerations, financial constraints, and methodological challenges. Technical issues, particularly data quality and integration problems, emerged as the most prevalent, appearing in 22 out of 26 studies. Organizational factors, such as skills shortages and cultural resistance, were the second most common theme, present in 20 studies.

Our network analysis further illuminated the interconnected nature of these challenges. With a network density of 0.68, the analysis demonstrated a high level of interconnectedness among the various challenges. Technical challenges, specifically 'data quality and integration', showed the highest degree centrality (0.85), indicating its central role in the network of challenges. The analysis also revealed three main clusters: Technical-Methodological, Organizational-Financial, and Ethical-Legal, providing insight into the natural groupings of challenges that organizations face in big data projects.

The findings underscore that big data project failures rarely stem from a single factor but rather from the interplay of multiple, interconnected challenges. This insight calls for a holistic approach to big data initiatives, integrating technical solutions with organizational change management, ethical considerations, and strategic alignment.

Our research suggests several important theoretical implications. First, researchers should adopt a more integrated approach to studying big data project failures. The high interconnectedness revealed by our network analysis suggests that examining challenges in isolation may lead to incomplete or misleading conclusions. Future research should focus on the interactions between different types of challenges and how they collectively contribute to project outcomes. Second, the field needs to pay more attention to the role of organizational factors in big data project success. While technical challenges are prominent, our findings suggest that organizational issues like skills shortages and cultural resistance are equally critical. Researchers should develop more comprehensive models that incorporate both technical and organizational factors in predicting big data project success. Third, the emergence of ethical and legal considerations as a significant theme suggests a need for more interdisciplinary research. Future studies should integrate perspectives from ethics, law, and information systems to better understand and address these challenges in big data projects.

From a practical standpoint, our findings have several implications for managers and organizations. Organizations should adopt a multifaceted approach to big data project management. Our findings suggest that focusing solely on technical solutions is insufficient. Managers should simultaneously address technical challenges, organizational readiness, ethical implications, and methodological issues to increase the chances of project success. There is also a need for more comprehensive risk assessment tools in big data projects. Given the interconnected nature of challenges revealed by our network analysis, practitioners should develop and use risk assessment frameworks that account for the potential ripple effects of individual challenges across the project ecosystem. Furthermore, organizations should prioritize building a data-driven culture and investing in skills development. Our findings highlight the critical role of organizational factors, suggesting that technical expertise alone is not sufficient for project success.

Despite the comprehensive nature of our review, it is important to acknowledge certain limitations. While our review included a diverse range of studies, the rapid evolution of big data technologies means that some very recent developments may not be fully represented in our analysis. Additionally, the inclusion of only English-language publications may have limited our access to relevant research from non-English speaking countries, potentially impacting the global applicability of our findings. Lastly, the reliance on published studies and reports may introduce a bias towards reporting on failed projects, potentially underrepresenting successful big data initiatives and their characteristics. Future research could address these limitations by including more recent studies, non-English publications, and cases of successful big data implementations to provide a more comprehensive view of the field.

5 Conclusion

This SLR of 26 studies has illuminated the complex landscape of big data project failures. The findings reveal five interconnected challenges: technical, organisational, ethical, financial, and methodological. Technical issues, particularly in data quality and integration, emerged as the most prevalent, closely followed by organisational factors such as skills shortages and cultural resistance. Ethical considerations and financial constraints also play significant roles, while methodological challenges, though less frequently mentioned, highlight important areas for future research.

The review underscores that big data project failures rarely stem from a single factor but rather from the interplay of multiple challenges. This insight calls for a holistic approach to big data initiatives, integrating technical solutions with organisational change management, ethical considerations, and strategic alignment.

For researchers, these findings point to the need for more integrated studies examining the interactions between different types of challenges. For practitioners, they emphasise the importance of comprehensive strategies that address not only technical aspects but also organisational readiness and ethical implications.

In conclusion, while Big Data continues to offer significant potential, realising this potential requires navigating a complex set of challenges. By understanding and addressing these challenges holistically, organisations can enhance their chances of success in leveraging Big Data for innovation and value creation.

6 References

Abu Afifa, A., and Abu-Assab, S. 2023. "Big Data in the Telecommunication Sector in Palestine: Challenges and Opportunities," Artificial Intelligence (AI) and Finance), pp. 934-944.

Akal, T. D., Beshah, T., Sackmann, S., and Negash, S. 2019. "Challenges of Identifying and Utilizing Big Data Analytics in a Resource-Constrained Environment: In the Case of Ethiopia," ICT Unbounded, Social Impact of Bright ICT Adoption: IFIP WG 8.6 International Conference on Transfer and Diffusion of IT, TDIT 2019, Accra, Ghana, June 21–22, 2019, Proceedings: Springer, pp. 234-254.

Al-Madhrahi, Z., Singh, D., and Yadegaridehkordi, E. 2022. "Integrating Big Data Analytics into Business Process Modelling: Possible Contributions and Challenges," International Journal of Advanced Computer Science and Applications (13:6).

Al-Sai, Z. A., and Abdullah, R. 2019. "Big Data Impacts and Challenges: A Review," 2019 IEEE Jordan international joint conference on electrical engineering and information technology (JEEIT): IEEE, pp. 150-155.

Arksey, H., and O'Malley, L. 2005. "Scoping Studies: Towards a Methodological Framework," International journal of social research methodology (8:1), pp. 19-32.

Ataei, P., and Litchfield, A. 2022. "The State of Big Data Reference Architectures: A Systematic Literature Review," *IEEE Access*, pp. 1-1, doi: 10.1109/ACCESS.2022.3217557.

Ataei, P., and Litchfield, A. 2023. "NeoMycelia: A Software Reference Architecture for Big Data Systems," Information Systems Frontiers (25:3), pp. 893-921.

Ataei, P., and Staegemann, D. 2023. "Application of Microservices Patterns to Big Data Systems," *Journal of Big Data* (10), doi: 10.1186/s40537-023-00733-4.

Atapattu, A., Wattuhewa, R., Waidyasekara, K., and Dilakshan, R. 2023. "Big Data Analytics in the Sri Lankan Construction Industry: An Assessment of the Challenges and Strategies," Proceedings of the 11th World Construction Symposium.

Bashari Rad, B., Ataei, P., Khakbiz, Y., and Akbarzadeh, N. 2016. "The Hype of Emerging Technologies: Big Data as a Service," *International Journal of Control Theory and Applications* (9), pp. 525-541.

Balaji, S., Patil, M., and McGregor, C. 2017. "A Cloud Based Big Data Based Online Health Analytics for Rural Nicus and Picus in India: Opportunities and Challenges," 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS): IEEE, pp. 385-390.

Barham, H., and Daim, T. 2020. "The Use of Readiness Assessment for Big Data Projects," Sustainable Cities and Society (60), p. 102233.

Chang, W. L., and Grady, N. 2019. "Nist Big Data Interoperability Framework: Volume 2, Big Data Taxonomies."

Databricks. 2024. "State of Data + Ai," Databricks.

Escobar, C. A., McGovern, M. E., and Morales-Menendez, R. 2021. "Quality 4.0: A Review of Big Data Challenges in Manufacturing," Journal of Intelligent Manufacturing (32:8), pp. 2319-2334.

Hevey, D. (2018). Network analysis: a brief overview and tutorial. Health Psychology and Behavioral Medicine, 6(1), 301–328. https://doi.org/10.1080/21642850.2018.1521283

Hoozemans, J., Peltenburg, J., Nonnemacher, F., Hadnagy, A., Al-Ars, Z., and Hofstee, H. P. 2021. "Fpga Acceleration for Big Data Analytics: Challenges and Opportunities," IEEE Circuits and Systems Magazine (21:2), pp. 30-47.

Hotz, N. 2024. "Why Big Data Science & Data Analytics Projects Fail," Data Science Process Alliance.

James, S., and Duncan, A. D. 2023. "Over 100 Data and Analytics Predictions through 2028," Gartner.

Kitchenham, B., and Charters, S. 2007. "Guidelines for Performing Systematic Literature Reviews in Software Engineering," (2).

Lauesen, S. 2020. "It Project Failures, Causes and Cures," IEEE access (8), pp. 72059-72067.

Levy, Y., and Ellis, T. J. 2006. "A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research," Informing Science (9).

Lim, C., Kim, K.-J., and Maglio, P. P. 2018. "Smart Cities with Big Data: Reference Models, Challenges, and Considerations," Cities (82), pp. 86-99.

Mangal, P., Rajesh, A., and Misra, R. 2020. "Big Data in Climate Change Research: Opportunities and Challenges," 2020 International Conference on Intelligent Engineering and Management (ICIEM): IEEE, pp. 321-326.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Group, P. 2010. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The Prisma Statement," International journal of surgery (8:5), pp. 336-341.

Paré, G., Trudel, M.-C., Jaana, M., and Kitsiou, S. 2015. "Synthesizing Information Systems Knowledge: A Typology of Literature Reviews," Information & management (52:2), pp. 183-199.

P. De Meo, E. Ferrara, G. Fiumara and A. Provetti, "Generalized Louvain method for community detection in large networks," 2011 11th International Conference on Intelligent Systems Design and Applications, Cordoba, Spain, 2011, pp. 88-93, doi: 10.1109/ISDA.2011.6121636. keywords:

{Communities;Image edge detection;Social network services;Partitioning algorithms;Optimization;Algorithm design and analysis;Mathematical model;complex networks;community structure},

Rad, B. B., Ataei, P., and Khakbiz, Y. 2021. "The Hype of Emerging Technologies: Big Data as a Service," International Journal of Computer Applications (975), p. 8887.

Rad, B. B., and Ataei, P. 2017. "The Big Data Ecosystem and Its Environs," International Journal of Computer Science and Network Security (17:3), pp. 38-46.

Reggio, G., and Astesiano, E. 2020. "Big-Data/Analytics Projects Failure: A Literature Review," 2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA): IEEE, pp. 246-255.

Sarker, S., Arefin, M. S., Kowsher, M., Bhuiyan, T., Dhar, P. K., and Kwon, O.-J. 2022. "A Comprehensive Review on Big Data for Industries: Challenges and Opportunities," IEEE Access (11), pp. 744-769.

Schneider, D., and Seelmeyer, U. 2019. "Challenges in Using Big Data to Develop Decision Support Systems for Social Work in Germany," Journal of Technology in Human Services (37:2-3), pp. 113-128.

Shah, S., Soriano, C. B., and Coutroubis, A. 2017. "Is Big Data for Everyone? The Challenges of Big Data Adoption in Smes," 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM): IEEE, pp. 803-807.

Staegemann, D., Volk, M., Vu, T., Bosse, S., Häusler, R., Nahhas, A., Pohl, M., and Turowski, K. 2020. "Determining Potential Failures and Challenges in Data Driven Endeavors: A Real World Case Study Analysis," IoTBDS, pp. 453-460.

Teradata. 2017. "High-Impact Business Outcomes with Analytics," Teradata.

Verner, J., Sampson, J., and Cerpa, N. 2008. "What factors lead to software project failure?," in 2008 Second International Conference on Research Challenges in Information Science, IEEE, pp. 71-80.

Walker, T., Davis, F., and Schwartz, T. 2022. "Big Data in Finance: Opportunities and Challenges of Financial Digitalization," Springer.

Copyright