Big Data in Industrial Manufacturing:A Case Study of General Electric's Predix™ Platform

Piyush Doifode[*1]
Student – PGDM – Business Analytics [ISMS, PUNE]

Prof. Avishek Das[*2]

Assistant Professor – Department of Analytics [ISMS, PUNE]

May 8, 2026

Author Note
https://orcid.org/0000-0000-0001-0001

How to Cite this Article:

Doifode, P. (2026). Big Data in Industrial Manufacturing:A Case Study of General Electric's Predix™ Platform. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(05). https://doi.org/10.55041/ijcope.v2i5.275

https://doi.org/10.55041/ijcope.v2i5.275

Abstract

The exponential growth of industrial sensor data has compelled large-scale manufacturers to fundamentally rethink their data infrastructure strategies. This paper examines how General Electric (GE) — one of the world's largest diversified industrial corporations — confronted the challenge of managing, processing, and deriving value from machine-generated data at unprecedented scale. Through a structured case-study methodology, this paper analyses GE's development and deployment of the Predix™ Industrial Internet of Things (IIoT) platform, including its integration with Amazon Web Services cloud infrastructure. The paper maps each of GE's Big Data initiatives — predictive maintenance, real-time equipment monitoring, aviation fuel optimisation, wind farm management, and digital twin technology — against established Big Data frameworks, including the three Vs (Volume, Velocity, Variety), Hadoop/MapReduce distributed processing, NoSQL storage architectures, and the descriptive, predictive, and prescriptive analytics hierarchy. Findings indicate that GE's $1 billion investment in data infrastructure generated substantial operational benefits: issue resolution times decreased from weeks to days, wind farm energy output improved by two to five per cent, and individual customers avoided multi-million-dollar unplanned outages. The study also identifies the key implementation challenges GE encountered — including scalability, latency, data security, talent acquisition, and system integration — and documents the solutions applied. The paper concludes that Big Data transformation in heavy industry requires not only technological investment but also organisational commitment, talent strategy, and a willingness to reimagine the product from hardware to data-driven services.

Keywords: Big Data, Industrial Internet of Things, Predix, General Electric, predictive maintenance, Hadoop, MapReduce, cloud computing, NoSQL, digital twin, descriptive analytics, prescriptive analytics, manufacturing analytics

1. Introduction

The digital transformation of industrial manufacturing has accelerated dramatically in the twenty-first century, driven primarily by the proliferation of sensor-enabled machinery and the emergence of cloud-based distributed computing platforms. Organisations that once relied on periodic inspection schedules and reactive maintenance regimes are increasingly adopting data-driven approaches that permit continuous real-time monitoring, failure prediction, and prescriptive operational guidance (Davenport & Dyché, 2013). Yet the implementation of such capabilities is non-trivial: the sheer volume, velocity, and variety of industrial sensor data routinely exceed the capacity of conventional relational database management systems, necessitating fundamentally different approaches to data architecture (Manyika et al., 2011).

General Electric (GE), founded by Thomas Edison in 1892, occupies a particularly instructive position in this landscape. As a manufacturer of jet engines, gas turbines, wind turbines, MRI machines, and power generation equipment, GE's industrial machines collectively produce more than thirty per cent of the world's electricity and are deployed in over one hundred and fifty countries (GE, 2016). The sensor networks embedded within this equipment generate approximately 500,000 data records per second and twenty billion sensor tags per day — a volume that quickly overwhelmed GE's existing on-premises data infrastructure and demanded a transformative architectural response (AWS, 2016).

GE's response was the development of Predix™ — a proprietary Industrial Internet of Things (IIoT) platform representing a $1 billion investment between 2012 and 2016. Predix integrates real-time data streaming, cloud-based distributed processing, machine learning, and prescriptive analytics to support use cases ranging from predictive equipment maintenance to aviation fuel optimisation and wind farm management. The Predix case is significant not merely because of its scale, but because it offers a comprehensive empirical illustration of virtually every major Big Data concept applied within a single industrial context.

This paper is structured as follows. Section 2 provides a review of the relevant Big Data literature. Section 3 describes the methodology. Section 4 presents the company background. Section 5 analyses the Big Data problem GE faced. Section 6 examines the Predix architecture and technology stack. Section 7 details GE's key use cases. Section 8 maps the theoretical Big Data concepts applied. Section 9 presents the results and business impact. Section 10 addresses implementation challenges and solutions. Section 11 provides a discussion, and Section 12 offers a conclusion.

2. Literature Review

2.1 The Big Data Paradigm

The term 'Big Data' entered mainstream academic and practitioner discourse around 2011, catalysed by McKinsey Global Institute's landmark report estimating that data could generate $300 billion in annual value in healthcare alone and $600 billion in retail (Manyika et al., 2011). While definitions vary, the most widely cited conceptual framework characterises Big Data according to three fundamental dimensions — Volume, Velocity, and Variety — collectively known as the three Vs (Laney, 2001). Volume refers to the quantity of data generated; Velocity to the rate at which data arrives and must be processed; and Variety to the heterogeneous formats in which data exists, including structured relational records, unstructured text and images, and semi-structured log files (Zikopoulos et al., 2012).
Subsequent scholarship has extended this framework to include additional dimensions — Veracity (data quality and reliability), Value (actionable insight derivable from data), and Variability (inconsistency in data meaning over time) — reflecting the growing recognition that raw data accumulation is insufficient without systematic governance and analytical transformation (Beyer & Laney, 2012). These extensions are particularly relevant in industrial contexts, where sensor calibration errors, network latency, and device heterogeneity can introduce significant data quality challenges.

2.2 Distributed Processing: Hadoop and MapReduce

The computational challenge of processing Big Data at scale was fundamentally addressed by Google's publication of the MapReduce programming model in 2004 (Dean & Ghemawat, 2004) and the Google File System paper (Ghemawat et al., 2003), which inspired the open-source Apache Hadoop framework. Hadoop enables the distributed processing of large datasets across clusters of commodity hardware through a two-phase computation model: the Map phase transforms input data into key-value pairs, and the Reduce phase aggregates these pairs into summary outputs (White, 2015). The Hadoop Distributed File System (HDFS) provides fault-tolerant distributed storage by replicating data blocks across multiple nodes.

In industrial applications, MapReduce has been applied to sensor log aggregation, anomaly detection across large fleets of equipment, and the batch processing of historical operational data to train predictive maintenance models (Lee et al., 2014). Cloud-based implementations such as Amazon Elastic MapReduce (EMR) have further democratised access to distributed processing by eliminating the need for organisations to maintain physical Hadoop clusters.

2.3 NoSQL Database Systems

The variety dimension of Big Data — encompassing unstructured and semi-structured data types — challenged the rigid schema requirements of traditional relational database management systems. NoSQL (Not Only SQL) databases emerged as a response, providing flexible document, key-value, column-family, and graph-based storage models capable of accommodating diverse data formats at scale (Cattell, 2011). NoSQL systems generally trade the strong consistency guarantees of relational databases for horizontal scalability and schema flexibility — trade-offs that are frequently acceptable in industrial IoT contexts where write throughput and schema evolution are prioritised over transactional integrity (Sadalage & Fowler, 2012).

2.4 Industrial IoT and Predictive Maintenance

The Industrial Internet of Things (IIoT) refers to the application of IoT sensor networks, connectivity, and data analytics within industrial and manufacturing contexts. Lee et al. (2014) identify predictive maintenance as the primary value-generating application of IIoT, estimating that unplanned equipment downtime costs manufacturers approximately $50 billion annually in the United States alone. Predictive maintenance leverages continuous sensor telemetry and machine learning models to detect early warning indicators of equipment degradation, enabling maintenance interventions to be scheduled before failure occurs rather than in response to it.

Digital twin technology — the creation of a real-time virtual replica of a physical asset, continuously synchronised with operational sensor data — has emerged as a key enabling concept for advanced predictive maintenance and lifecycle management. Grieves and Vickers (2017) conceptualise the digital twin as a three-component model comprising the physical entity, its virtual counterpart, and the data connection linking them — a framework directly applicable to GE's implementation.

2.5 Analytics Maturity: Descriptive, Predictive, and Prescriptive

The analytics maturity model identifies three ascending tiers of analytical capability (Davenport & Harris, 2007). Descriptive analytics characterises what has happened or is currently happening through dashboards and reports. Predictive analytics forecasts what is likely to happen, using statistical models and machine learning. Prescriptive analytics recommends specific actions to achieve desired outcomes, integrating prediction with optimisation and decision logic. Industrial Big Data initiatives like GE's Predix represent the full spectrum of this maturity model — descriptive dashboards for equipment health, predictive models for failure probability, and prescriptive alerts and recommendations for maintenance scheduling and operational adjustment.

3. Methodology

This study employs a single-organisation case study design, following Yin's (2018) framework for explanatory case research. The case study method is appropriate when examining a contemporary phenomenon within its real-world context, particularly where contextual conditions are central to understanding the phenomenon — conditions that apply directly to GE's Big Data implementation, which cannot be separated from the specific engineering, commercial, and organisational context of large-scale industrial manufacturing.

Data collection drew on multiple secondary sources including GE's official technical documentation and public announcements, Amazon Web Services published case studies, peer-reviewed academic literature, and specialist technology journalism. Triangulation across these sources was employed to enhance construct validity. The analysis follows a pattern-matching approach (Yin, 2018), mapping empirical observations from the GE case against theoretical

Big Data frameworks established in the literature review. Limitations of the study include reliance on publicly available information, the absence of primary interview data, and potential retrospective rationalisation in published corporate accounts.

4. Company Background

General Electric was founded by Thomas Alva Edison in 1892 and has since grown into one of the world's most diversified industrial conglomerates. Its core product lines include jet engines for commercial and military aviation, gas turbines and generators for power generation, wind turbines for renewable energy, and MRI and imaging equipment for healthcare. GE serves customers in over one hundred and fifty countries, and its power generation equipment collectively produces more than thirty per cent of the world's electricity.

GE's industrial scale creates a distinctive data challenge. Each piece of equipment — from a single wind turbine blade to a commercial jet engine — is embedded with hundreds or thousands of sensors monitoring parameters including temperature, pressure, vibration, rotational speed, fuel flow, and electrical output. Aggregated across GE's global installed base, this sensor network generates approximately 500,000 data records per second and twenty billion sensor tags per day — data volumes that dwarf the capacity of conventional on-premises data infrastructure and demand a purpose-built Big Data architecture (AWS, 2016; GE, 2016).

5. The Big Data Problem GE Faced

Prior to its Big Data transformation, GE's approach to machine data was fundamentally reactive. Equipment failures were detected after they occurred — a paradigm that imposed enormous costs in the form of unplanned production downtime, emergency repair mobilisation, and reputational damage to customers whose operations depended on GE's machinery. The company's traditional IT infrastructure was simply unable to accommodate the three Vs of the data its machines generated.

5.1 Volume

Sensor data volumes grew from 8 gigabytes to over 200 terabytes — a twenty-five-thousand-fold increase that rendered on-premises storage and processing architectures economically and operationally unviable. Conventional relational databases could not ingest, index, or query data at this scale without prohibitive hardware investment and unacceptable query latency.

5.2 Velocity

Sensor readings arrive at a rate of 500,000 records per second across thousands of machines distributed globally. This velocity demands streaming data ingestion architectures capable of processing data in near-real-time — a requirement fundamentally incompatible with traditional batch-oriented extract, transform, and load (ETL) processes.

5.3 Variety

GE's sensor network generates not only structured numeric telemetry but also unstructured data types including log files, thermal images, audio signals from vibration sensors, and maintenance notes in natural language. The simultaneous management of these heterogeneous data types requires storage and processing architectures that relational databases — optimised for tabular structured data — cannot efficiently accommodate.

5.4 Reactivity

Perhaps most commercially damaging was GE's inability to detect equipment failures before they occurred. Without predictive capability, GE and its customers could only respond to failures after the fact — incurring the full cost of unplanned outages, emergency maintenance, and associated production losses. In power generation, a single forced outage at a major plant can cost customers millions of dollars in lost generation revenue.

6. The Predix™ Platform: Architecture and Technology Stack

GE's strategic response to these challenges was the development of Predix™ — a proprietary Industrial Internet of Things platform designed to ingest, store, process, and analyse sensor data from industrial machines at global scale. GE invested $1 billion in Predix and its supporting analytics infrastructure between 2012 and 2016 (GE, 2016; MIT Sloan Management Review, 2015).

6.1 IoT Sensor Layer

Thousands of sensors installed on turbines, jet engines, and power plant components form the data collection layer of the Predix architecture. These sensors continuously stream real-time telemetry — including temperature, pressure, vibration, rotational speed, and electrical output — into the platform's ingestion tier.

6.2 Real-Time Data Streaming: Amazon Kinesis

Predix leverages Amazon Kinesis Data Streams to ingest and stream 500,000 data records per second from 900 GE Power customer sites worldwide (AWS, 2016). Kinesis provides a managed, highly available streaming data service capable of handling the sustained high-velocity data ingestion that GE's sensor network demands — a capability that batch-oriented data pipelines cannot replicate.

6.3 Distributed Processing: Amazon EMR and Hadoop MapReduce

Sensor data ingested via Kinesis is processed using Amazon Elastic MapReduce (EMR) — a cloud-based managed Hadoop service. EMR executes MapReduce jobs that aggregate, transform, and analyse the sensor data at scale, enabling GE to process terabytes of machine data without maintaining physical Hadoop infrastructure. This cloud-based distributed processing approach provided the elastic scalability that on-premises systems could not deliver.

6.4 Distributed Storage: Amazon S3

Amazon Simple Storage Service (S3) serves as the distributed archival and historical storage layer of the Predix architecture — functionally equivalent to a cloud-hosted Hadoop Distributed File System (HDFS). Historical sensor data and analytical outputs are stored in S3, enabling the training of machine learning models on large historical datasets and supporting retrospective analytical queries.

6.5 NoSQL Data Stores

Unstructured machine data — including log files, images, and audio signals — is stored in document-based NoSQL databases that provide the schema flexibility required to accommodate the variety of data types generated by GE's heterogeneous sensor network. NoSQL storage also supports the high write throughput required by continuously streaming sensor data without the locking and schema constraints of relational database systems.

6.6 The Predix Application Layer

The Predix application layer executes more than one million data processing operations per day and generates automatic real-time alerts when equipment anomalies are detected. Machine learning models embedded within Predix continuously analyse incoming sensor telemetry against historical performance baselines, identifying deviations that may signal impending equipment failure and triggering prescriptive maintenance alerts.

7. Key Use Cases

7.1 Predictive Maintenance

Predictive maintenance is the flagship application of Big Data at GE. Rather than waiting for machines to fail or adhering to fixed time-based maintenance schedules, Predix continuously analyses sensor readings to detect early warning patterns that precede equipment failures. Machine learning models identify statistical anomalies in vibration, temperature, pressure, and other parameters that correlate with impending component degradation. When anomalies are detected, customers receive automatic alerts enabling them to schedule maintenance during planned downtime — eliminating the cost of unplanned emergency outages. At least one GE Power customer avoided a forced outage worth millions of dollars through Predix-generated predictive alerts (AWS, 2016).

The analytics applied span both predictive (forecasting failure probability) and prescriptive tiers (recommending the specific maintenance action and optimal timing), representing the highest levels of the analytics maturity model (Davenport & Harris, 2007).

7.2 Real-Time Equipment Monitoring

GE Power monitors 900 customer sites globally, streaming half a million data records per second through Kinesis into the Predix platform. Customers gain live visibility into the operational health of their assets through real-time dashboards, enabling faster and better-informed operational decisions. Before Predix, issue identification and resolution required weeks or months; after Predix deployment, the same cycle was reduced to days (AWS, 2016). This application exemplifies descriptive analytics — providing a continuous, accurate representation of what is currently happening across all monitored assets.

7.3 Aviation Fuel Efficiency

Airlines spend approximately $200 billion per year on aviation fuel, making fuel efficiency a primary commercial and environmental concern for the aviation industry (GE, 2016). GE's software analyses flight data — including engine telemetry, weather conditions, payload weight, and flight path variables — to recommend optimal fuel management strategies to pilots in real time. A two per cent improvement in fuel efficiency across the global aviation industry would represent $4 billion in annual savings. This application is prescriptive analytics in its clearest form: integrating multiple data streams to recommend the optimal course of action.

7.4 Wind Farm Optimisation

Wind turbines positioned at the front of a wind farm create turbulence that propagates downstream, causing aerodynamic loading and vibration in turbines behind them. Sustained vibration accelerates component wear and risks stress fractures in turbine blades — failures that are extremely costly to remediate at scale. GE's sensors detect real-time wind patterns and blade vibration, and the Predix platform automatically adjusts blade pitch angles to reduce stress loads and maximise energy capture. The result is a two to five per cent improvement in overall wind farm energy output (GE, 2016). This use case combines predictive detection of vibration risk with prescriptive automatic actuation — a closed-loop application of Big Data analytics.

7.5 Digital Twin Technology

GE creates a virtual replica — a Digital Twin — of every physical machine it manufactures. The digital twin is continuously updated with real sensor data and historical performance records, enabling engineers to simulate failure scenarios, test maintenance strategies, and accurately forecast the remaining useful life of individual components — all without touching or disrupting the actual machine. Grieves and Vickers (2017) describe this as a foundational technology for lifecycle management in industrial manufacturing. GE's implementation of digital twins at scale represents one of the most sophisticated applications of IIoT and Big Data in manufacturing.

8. Big Data Concepts Applied at GE

The GE case provides a comprehensive empirical illustration of the major Big Data concepts established in the academic and practitioner literature. Table 1 summarises the mapping between theoretical concepts and their specific application in the GE context.

The three Vs are directly instantiated: Volume manifests in the 200 TB of sensor data generated across GE's global operations; Velocity in the 500,000 records streamed per second via Amazon Kinesis; and Variety in the combination of structured telemetry, unstructured log files, thermal imagery, and audio signals (Laney, 2001; Zikopoulos et al., 2012). The Hadoop/MapReduce distributed processing framework — implemented through Amazon EMR — provides the computational engine for processing data at this scale, while Amazon S3 serves as the distributed storage layer functionally equivalent to HDFS (Dean & Ghemawat, 2004; White, 2015). NoSQL databases address the variety dimension by providing schema-flexible storage for unstructured data types (Cattell, 2011). All three tiers of the analytics maturity model are represented: descriptive analytics in real-time dashboards, predictive analytics in failure forecasting models, and prescriptive analytics in automated maintenance alerts and pilot recommendations (Davenport & Harris, 2007). The digital twin implementation exemplifies the integration of IoT sensor networks with advanced analytical simulation (Grieves & Vickers, 2017).

9. Results and Business Impact

GE's Big Data transformation generated substantial and measurable operational and commercial benefits. At the infrastructure level, Predix now processes 500,000 data records per second, ingests twenty billion sensor tags per day, and executes over one million data processing operations per day across GE Power's 900 monitored customer sites worldwide (AWS, 2016). These figures represent a capability that was entirely beyond the reach of GE's pre-Predix on-premises infrastructure.

At the operational level, the most significant impact was the reduction in equipment issue resolution time from weeks or months to days — a transformation with direct commercial implications for GE's customers, whose production operations depend on the continuous availability of GE's equipment. At least one major GE Power customer avoided a forced outage that would have cost millions of dollars through a Predix- generated predictive maintenance alert. Wind farm customers achieved two to five per cent improvements in energy output through automated blade optimisation — improvements that compound significantly at the scale of large wind farms. Aviation customers gained access to fuel optimisation recommendations with the potential to reduce industry-wide fuel costs by $4 billion annually if a two per cent efficiency gain is achieved across the global fleet.

From a strategic perspective, GE's Big Data investment enabled a fundamental shift in its business model — from a hardware manufacturer selling capital equipment to a data-driven services provider whose value proposition is centred on continuous operational optimisation and outcome guarantees. This transition from product to service is consistent with the servitisation literature (Vandermerwe & Rada, 1988) and represents a significant competitive differentiator in markets where hardware specifications are increasingly commoditised.

10. Implementation Challenges and Solutions

10.1 Scalability

On-premises data infrastructure was unable to keep pace with data volumes growing from 8 GB to 200 TB. GE's solution was to migrate from on-premises infrastructure to a cloud-based architecture centred on Amazon Web Services, providing elastic scalability that could accommodate continued data growth without proportionate capital investment. The Predix IIoT platform was designed from the outset for cloud-native elastic scaling (AWS, 2016).

10.2 Latency

GE's customers required real-time alerts and operational guidance — a requirement incompatible with batch processing architectures that introduce hours of latency. GE addressed this challenge by adopting real-time streaming analytics tools, including Apache Spark for in-memory computation, and edge computing architectures that process data at or near the machine before transmission to the cloud — reducing latency to levels compatible with real-time decision-making.

10.3 Data Security

The migration of sensitive industrial operational data to cloud infrastructure raised significant security concerns for GE and its customers. GE addressed these concerns through strong encryption of data in transit and at rest, rigorous identity and access management controls, and compliance with relevant industry security standards. Cloud provider security frameworks, including those provided by AWS, were leveraged to enhance security posture without requiring GE to develop bespoke security infrastructure.

10.4 Talent Acquisition

Data scientists and industrial data engineers remain scarce globally. GE recognised that organisational capability was as critical as technology in enabling its Big Data transformation, and responded by establishing a dedicated analytics centre in San Ramon, California, staffed by over 300 data science and engineering specialists. This investment in human capital was essential to translating the Predix platform's technical capabilities into commercially valuable analytical products (MIT Sloan Management Review, 2015).

10.5 System Integration

Connecting thousands of heterogeneous sensors and legacy operational systems — manufactured over decades and operating across diverse communication protocols — into a single unified platform was a significant engineering challenge. GE addressed this through the use of the Predix IIoT platform as an integration layer, supplemented by application programming interfaces (APIs) and middleware connectors that translated legacy data formats and protocols into the standardised schemas required by the Predix analytics pipeline.

11. Discussion

The GE case study offers several broader insights for scholars and practitioners engaged with Big Data strategy in industrial contexts. First, the case demonstrates that the three Vs framework, while analytically useful, understates the organisational complexity of large-scale Big Data implementation. GE's most significant challenges — talent acquisition, legacy system integration, and organisational change management — are not captured by the three Vs, underscoring the relevance of extended frameworks that incorporate dimensions of Veracity and Value (Beyer & Laney, 2012).

Second, the case illustrates the critical enabling role of cloud infrastructure in making industrial Big Data economically viable. The elastic scalability, managed services, and consumption-based pricing of AWS enabled GE to build a data processing capability that would have been prohibitively expensive to replicate on-premises — a finding consistent with the emerging literature on cloud-enabled digital transformation (Bharadwaj et al., 2013). Cloud adoption is not merely a technical architecture decision but a strategic enabler of new business models.

Third, the case provides compelling empirical support for the analytics maturity model (Davenport & Harris, 2007). GE's progression from descriptive monitoring dashboards to predictive failure models to prescriptive automated alerts reflects the maturity progression identified in the literature — and suggests that organisations should plan their Big Data investments as a staged capability-building journey rather than a single transformation event.

Finally, the GE case challenges a common assumption that Big Data transformation is primarily relevant to technology-native organisations such as internet platforms and financial services firms. GE demonstrates that traditional heavy manufacturing — at its heaviest and most capital-intensive — can be fundamentally reimagined through intelligent data collection, processing, and analysis. The barriers to Big Data adoption in manufacturing are more organisational and strategic than technological: the willingness to invest at scale, the ability to attract and retain analytical talent, and the courage to reimagine the core value proposition from hardware to data-driven service outcomes.

12. Conclusion

This paper has examined General Electric's Big Data transformation through a structured case study of the Predix™ IIoT platform. The analysis demonstrates that GE's $1 billion investment in Big Data infrastructure — spanning cloud-based streaming ingestion, distributed MapReduce processing, NoSQL storage, machine learning, and prescriptive analytics — generated substantial operational benefits: issue resolution times reduced from weeks to days, wind farm energy output improved by two to five per cent, and predictive maintenance alerts enabled customers to avoid multi-million-dollar unplanned outages. Strategically, Predix enabled GE to transition from a hardware manufacturer to a data-driven industrial services provider — a transformation with enduring competitive significance.

The case illustrates, comprehensively, how every major Big Data concept — the three Vs, Hadoop/MapReduce distributed processing, NoSQL storage, distributed file systems, and all three tiers of the analytics maturity model — applies in a real industrial setting at global scale. GE's journey also demonstrates that the most significant barriers to Big Data adoption in manufacturing are rarely technological; they are organisational: the readiness to change, the capacity to acquire and develop analytical talent, and the strategic will to invest at the scale the transformation demands.

Future research should examine the sustainability of GE's data-driven business model in the face of evolving cloud infrastructure costs, increasing competitive pressure from technology-native industrial analytics providers, and the accelerating commoditisation of machine learning tooling. Longitudinal studies tracking the financial returns on GE's Big Data investment would also contribute valuable evidence to the emerging literature on digital transformation value creation in industrial manufacturing.

References

Amazon Web Services. (2016). GE Power uses AWS to unlock the power of industrial IoT data. Amazon Web Services Case Studies. https://aws.amazon.com/solutions/case-studies/ge-power/

Beyer, M. A., & Laney, D. (2012). The importance of 'big data': A definition. Gartner. https://www.gartner.com/en/documents/2057415

Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. (2013). Digital business strategy: Toward a next generation of insights. MIS Quarterly, 37(2), 471–482. https://doi.org/10.25300/MISQ/2013/37:2.3

Cattell, R. (2011). Scalable SQL and NoSQL data stores. ACM SIGMOD Record, 39(4), 12–27. https://doi.org/10.1145/1978542.1978544

Davenport, T. H., & Dyché, J. (2013). Big data in big companies. International Institute for Analytics. https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/big-data-in-big-companies-106461.pdf

Davenport, T. H., & Harris, J. G. (2007). Competing on analytics: The new science of winning. Harvard Business School Press.

Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified data processing on large clusters. Proceedings of the 6th Symposium on Operating System Design and Implementation (OSDI '04), 137–150. https://dl.acm.org/doi/10.1145/1327452.1327492

Digital Vidya. (2016). How General Electric built big data software analytics for the industrial internet. https://www.digitalvidya.com/blog/general-electric-ge-built-big-data-software-analytics-for-industrial-internet/

General Electric. (2016). GE big data & analytics. GE Official Website. https://www.ge.com/taxonomy/term/1664

Ghemawat, S., Gobioff, H., & Leung, S.-T. (2003). The Google file system. Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP '03), 29–43. https://doi.org/10.1145/945445.945450

Grieves, M., & Vickers, J. (2017). Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In F.-J. Kahlen, S. Flumerfelt, & A. Alves (Eds.), Transdisciplinary perspectives on complex systems (pp. 85–113). Springer. https://doi.org/10.1007/978-3-319-38756-7_4

Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. META Group Research Note. Gartner.

Lee, J., Bagheri, B., & Kao, H.-A. (2014). A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manufacturing Letters, 3, 18–23. https://doi.org/10.1016/j.mfglet.2014.12.001

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation

MIT Sloan Management Review. (2015). GE's big bet on data and analytics. MIT Sloan Management Review Case Study. https://sloanreview.mit.edu/case-study/ge-big-bet-on-data-and-analytics/

Sadalage, P. J., & Fowler, M. (2012). NoSQL distilled: A brief guide to the emerging world of polyglot persistence. Addison-Wesley.

Vandermerwe, S., & Rada, J. (1988). Servitization of business: Adding value by adding services. European Management Journal, 6(4), 314–324. https://doi.org/10.1016/0263-2373(88)90033-3

White, T. (2015). Hadoop: The definitive guide (4th ed.). O'Reilly Media.
Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). SAGE Publications.Zikopoulos, P., Eaton, C., deRoos, D., Deutsch, T., & Lapis, G. (2012). Understanding big data: Analytics for enterprise class Hadoop and streaming data. McGraw-Hill.