23041/23042

‘The Comprehensive Big Data Paradigm’: Challenges, solutions and the future

Richard Fulton
Troy University

Diane Fulton
Clayton State University

Nate Hayes
Modal Corporation

Susan Kaplan
Modal Corporation

Abstract

The authors research the most pressing problems facing big data to date. As the use of big data technologies becomes mainstream, this paper examines key challenge categories (Velocity, Volume, Variety, Veracity, Value, Management/Culture, Privacy/Security/Governance and Resources) and delineates current issues within each. The ‘Big Data Ecosystem’ has evolved in the past three decades and the researchers summarize the current solutions in each of the challenge areas. Not only do decision makers need to know what is currently available in the ‘Big Data Ecosystem’, they also need to know the gaps in the system. Innovative solutions in the future must address these weaknesses. To that end, this article also focuses on future trends, predictions, research and technology innovations needed to catapult big data beyond 2023 and into the next decade. The ‘big picture’ challenges, solutions and the future of big data are meshed into “The Comprehensive Big Data Paradigm” while also embracing the concepts of integration, collaboration and sustainability necessary for managing the data and making decisions that have high value in organizations and society.

Keywords: Big data ecosystem, technology, future innovations, challenges, big data paradigm, sustainability

INTRODUCTION

The term ‘Big Data’ was initiated in the 1990s by John Massey (1998), popularized in 2008 and is extensively used today to describe “data that is so large, fast or complex that it’s difficult or impossible to process using traditional methods” (Thompson, 2023, online). Big data has greater “variety” (different formats, different sources), is arriving in increasing “volumes” (sheer amounts of data), with more “velocity” (speed) than ever before, which is accelerating the need to process the information to glean its “veracity” (accurate and meaningful data) and is changing the leader role into managing this useful data (intelligence) to address hard to tackle problems that add “value” to the firm or society (otherwise known as the 5Vs of data) (Oracle, Big data, 2023, online, Foote, 2018, online). Major examples of big data include the “Google search index, Facebook user profiles and the Amazon product list” (Lynch, 2019, online).

The goal of this research is to 1) provide an overview of the most pressing challenges associated with implementing big data; 2) examine current solutions to address these challenges; 3) elaborate on the future including directions, research, trends and solutions to the gaps in big data and 4) combine these elements into ‘The Comprehensive Big Data Paradigm’ integrating challenges, solutions and the future of big data.

MOST PRESSING BIG DATA CHALLENGES

In a comprehensive three-year study of 5,593 participants across industries called the ‘Big Data Maturity Scale’, researchers ranked the most important big data challenges: 1) skill set; 2) governance; 3) performance; 4) security and 5) management (At Scale Big Data Maturity Scale, 2018). Matillion (2022, online) believes that big data challenges are fivefold: 1) resources (finding the right data professionals; demand for specialized skills is outpacing supply); 2) value (must enable insights to be drawn from data); 3) veracity (ensuring data quality, data must be accurate and technology needs to cleanse out any errors or inconsistent data; all systems must be integrated so that data changes in one system automatically get changed in all systems for accurate analytics); 4) management (data should be fully accessible to tools decision makers use); and 5) privacy/security (control and ownership over the data and security of the data are key elements).

Harvey’s (2018) most pressing industry challenges comprise 5 categories: 1) veracity (data quality and ability to integrate different big data systems); 2) value (results or performance of the big data); 3) resources (need for talent and skills, cost in terms of time and money to hire or train professionals, costs of changing technology and infrastructure, big data analytics initiatives often run significantly over budget and require significant time resources); 4) management/culture (creating a data driven corporate culture to manage big data systems, changing technology, legacy system integration and decision making); 5) security/privacy/governance (complying with government regulations regarding data, handling cybersecurity risks, protecting sensitive data).

It’s important to understand that priorities may differ by industry as well. For example, the top three big data challenges identified in a meta-analysis of 28 research studies in healthcare organizations were: 1) veracity (data processing such as data structure, data standardization, storage and transfers), 2) the security, privacy and data governance area (protecting sensitive information, following regulation and industry protocols, and guards to prevent cybersecurity or fraud attacks) and 3) value (could help to improve quality of healthcare, population management and health, early detection of disease, decision making and reduce costs in healthcare) (Kruse, Goswamy, Ravel & Marawi, 2016).

In the agriculture sector, major big data challenges included: 1) veracity (need to process the data thoroughly and in a timely manner to get useful data and different entities such as governments, universities, research organizations and agri-businesses must cooperate to share data); 2) value (data that is analyzed and enhances decision making like the mobile device app for making pesticide spray decisions currently available); 3) privacy, security and governance (big agri-businesses and entities may not want to share data due to proprietary nature of data or data exists in silos or has not been anonymized) (Orts & Spigonardo, 2014); and 4) resources (automating data acquisition with processing based on open systems can allow data mining costs to remain low) (Maya Gopa & Chintala, 2020, 56-57).

In their review of the agricultural industry, the main big data difficulties identified were related to the 5V's (data integration complexities, inadequate infrastructure and insignificant data warehouse architecture) and resources (the lack of skilled personnel and sufficient resources) (Quafiq, Saadane & Chehri, 2022). Summary descriptions of the big data challenges associated with the categories of the 5V's, management/culture, privacy/security/governance and resources are indicated in Figure 1 (Appendix).

CURRENT SOLUTIONS TO BIG DATA CHALLENGES

The big data industry is huge and multi-factored. As challenges pop up in the industry, companies have been addressing them with innovative solutions such as small data, smart data, new storage solutions like the cloud, speed-oriented, scalable and sustainable solutions using new technologies and new infrastructures.

Solutions for Data Value, Variety, Velocity, Veracity and Volume (5V's)

Small Data

Data scientists are using corroborative analysis and contextual interpretation of big data implementing a "making big data small" strategy that enables the fast extracting of only relevant data from large sets of geosocial data. While contrary to many big data approaches – in which analysis is done on the entire dataset – much productive research can use smaller datasets – around the same size as census or survey data – within standard methodological frameworks and researchers do not need big data science skills (Poorthuis & Zook, 2017).

Smaller or compressed data is realized by using different approximation algorithms (Barron, 1993; Chen, Chen & Liu, 1995; Cybenko, 1989; Hornik, 1991; Hornick, Stinchcome & White, 1989).

Smart Data

A new solution to many big data issues is to create 'smart data' (using automated smart data sensors, refining, cleansing and sorting data by filters) can narrow down the data (decrease volume) to more meaningful data (better value since only useful information for solving the problem is presented). Value, velocity, and veracity (accuracy) should all increase with the decrease in volume (Foote, 2018).

One of the current solutions to control data flow by integrating data uses open-sourced extract-transform-load software such as Clover which is intuitive for the user, has a “lightweight footprint, flexibility and fast processing speed” (Bansal & Kagemann, 2015, 8).

Storage Solutions

Until fairly recently, companies performed big data analytics in their “on-premise computing environments, relying on clusters of servers for storage and compute power, managing the infrastructure with a layer of software on top, and running analytics with additional applications”. This era relied on Hadoop projects and a variety of technologies to manage and optimize these environments. Today, organizations move their data into “cloud data warehouses and data lakes for analytics” which “allows them to scale resources up or down as needed, gives them easy access to the latest cloud-native tools, and lessens the burden of managing the environment” (Matillion, 2022, online).

In the past, storing big data was very costly – but cheaper storage solutions using data lakes, Hadoop and the cloud have eased the burden. The cloud not only stores data, but it also compresses, processes and analyzes big data, affecting the data’s veracity and value (Latifian, 2022). Less costly alternatives to the big cloud storage companies such as Google, Dropbox, Salesforce, Amazon and Microsoft also exist such as Digital Oceans, Cloudways, Backblaze and Airtable (Gewirtz, 2018).

Scalability Solutions

The need for handling data has accelerated exponentially in the last decade. Scalability and costs are driving the solutions and new technology. Transistor scaling was the first technology solution. Then the industry moved to multi core – the recent Intel’s Core i9 series offers up to eight cores for general-purpose computing processors. The demand for high performance computing prompted the move from multi-core to many-core. The many-core system has been limited by the interconnection network, thermal and power constraints (Bai, 2018).

On-chip interconnection networks became the technological driving factor for many-core chips. As the number of cores rises, the interconnection network is constrained by communication management such as wire connection, link delay, channel contention and bandwidth. Bus-based and point-to-point interconnection networks worked well for small multi-core systems but are becoming infeasible for large size networks due to high channel contention and area costs. Many-core systems with high core counts moved to Networks-on-Chip (NoC) interconnects, solving both wire and power problems increasing scalability and flexibility (Bai, 2018).

In the short-term, “scalability is failing” and can only be achieved by “matching hardware to computational workload through heterogeneous computing” says Ron Black, CEO of Codasip (2021). For example, “RISC-V’s modularity combined with design automation provides a foundation for creating novel processors and domain-specific accelerators” (Black, 2021). In other words, as the needs for better solutions evolve, the more innovative technologies and revolutionary changes are required.

Speed data transfer

Technology solutions to speed up the transfer of data or improve velocity in near-real time include RFID tags, sensors and smart meters (Harvey, 2018). In situations which require speed (i.e. deciding when an aircraft carrier is under attack or when an underdeveloped country is failing and violence escalates) applications like HAMOC (Human Centric Analysis for Crisis and Conflict) combine AI-powered analytics to machine speed strategic analysis of the environment by providing access to large volumes of data, understanding what is important and enabling analysts to integrate this understanding into decision-making swiftly (Maxwell, 2021, online).

Sustainability solutions

'Dynamic Voltage and Frequency Scaling' (Semeraro, Magklis & Scott, 2002) and 'power gating' (Farrokhbakht, Kamali & Hessabi, 2017) are two current solutions to increase energy efficiency while maintaining performance. Chips reach their power and thermal budget capacity easily with multi-core chips. The solution of turning a part of the chip off (referred to as the dark silicon) (Esmaeilzadeh, Blem & Burger, 2011) makes it difficult to turn on all the processing elements to boost the system performance (Chandarana & Vijayalakshmi, 2014).

Many industries, such as telecom, health care, retail, pharmaceutical, financial services, and agriculture are generating large amounts of data and processing is being pushed to the limit to analyze in a reasonable amount of time. Two solutions to manage data volume with big data at rest are: 1) NoSQL systems for interactive data serving environments; and 2) systems for large scale analytics based on MapReduce paradigm, such as Hadoop. These work seamlessly in a distributed cloud-based environment. In contrast, one can use Hadoop-based systems to run long running decision support and analytical queries consuming and possibly producing bulk data. (Gupta, Gupta & Mohania, 2012)

For processing data in motion, solutions include creating algorithms in a data stream management system and integrative applications which can quickly process massive amounts of data from both systems (Gupta et al., 2012).

Management/Culture Solutions

Successful big data decision-making is enhanced by creating a data-driven culture based on several principles:1) commitment from the top; 2) start with small projects and make data accessible to everyone, 3) fix basic data access issues quickly; 3) make proofs of concept simple and robust; 4) offer specialized training where needed; 5) use analytics to help employees and customers; and 6) explain analytical choices (Waller, 2020).

In their model to overcome slow adoption and resistance to change, Enders, Martin, Sehgal & Schüritz (2020) recommend five strategies including starting small, transparent revenue model, great user experience, value-based approach and building trust.

Motivational leaders, workforce education, "learning capabilities"— "trial-and-error processes and reflective experimenting", "absorptive capacity"— "the ability to acquire external knowledge to problem solve"— as well as use of internal and external change agents in an organization lead to innovative change (Volberda, van den Bosch & Heij, 2013, 1). Giving incentives to employees who engage as change agents and involving employees across the organization in strategic planning and using data visualization tools like Tableau with access to all decision makers can increase big data change management success (Gebauer, Haldimann & Caroline, 2017, 517; Wielgus, 2023).

The leader's character influences the followers in any organization and people are more likely to change if the manager "leads by example and walks the talk regarding the change", "is persuasive; can convince people of the need for change", "drives execution and accountability" and "is an effective coach and mentor" (Hechanova, Caringal-Go & Magsaysay, 2018, 919).

An agile analytics framework – data science support across disciplines as a shared service, shared analytics network, a structured change program and partner for success— is recommended for big data change projects (Van Arnum, 2014).

Privacy/Security/Governance Solutions

To deal with cyberattacks and massive data breaches, businesses are implementing the General Data Protection Regulation (GDPR) (https://gdpr.eu/what-is-gdpr/) guidelines set by the European Union in 2018 by investing in processes, protocols and infrastructures which can protect data security and privacy and increase governance (Magimind, 2022). Companies are increasingly moving to cloud platforms for their big data analytics since these environments already have higher security protocols (Latifian, 2022).

Another common use for big data analytics — particularly in the financial services industry — is fraud detection. One of the big advantages of big data analytics systems that rely on machine learning is that they are excellent at detecting patterns and anomalies. These abilities give banks and credit card companies the ability to spot stolen credit cards or fraudulent purchases, often before the cardholder even knows that something is wrong (Harvey, 2018).

Resource Solutions

Low cost and sustainable solutions are needed to handle all the various forms and sources of data. As mentioned previously, cloud-based solutions are currently the most affordable solutions. Formal change management programs help increase decision making quality and lower employees' reluctance. These programs require resources such as time, commitment, people and budget to implement changes such as new technologies, projects and systems. Recruiting analytics talent within the company and retraining them can be a cost-effective method (Van Arnum, 2014).

Using strategic planning tools like cost/benefit analysis and SWOT analysis are effective low-cost ways to assess change projects in the big data ecosystem and determine what resources will be needed (Raeburn, 2022; Stobierski, 2019).

Current solutions in the 5V's, management/culture, privacy/security/governance and resource areas are indicated in Figure 2 (Appendix).

THE FUTURE OF BIG DATA

Value

A promising strategy to integrate diverse data formats and sources is contextualized and corroborated data analysis which increases data's veracity and value (O'Halloran et al., 2018; Pokhriyal & Jacques, 2017; Zhang et al., 2018). For example, researchers developed a "small" contextual analysis approach to big data to evaluate the effects of social distancing on focused subpopulations in U.S. society. They guided the selection and processing of big data to make big data small and structured using a difference in difference sampling model. More such proactive methodological innovations that confront the critical limitations of accessible big data are recommended for the future (Zhang, Xu & Zhao, 2022, 142).

Contextual data is gathered and analyzed without isolating the data from the broader scope of information. "Contextual data needs to become the new buzzword", because it is both powerful and effective and facilitates more competent data analysis and data value (Kh, 2018, online). Researchers can use "small data" analysis that selectively extracts the most relevant data from big datasets for use within traditional methodological frameworks, which turns big data into small, slow, simple and structured data (Shaozeng, Dafeng & Zhao, 2022).

Newer algorithms for approximating feed-forward networks and deep learning networks also will help to develop better tools for 'small data' and 'precision' analysis (Barron, 2018; Barron & Klusowski, 2018; Hassan, Natase & Goldstein, 2019; Klukowski & Barron, 2017, Mariusz & Marek, 2022).)

Veracity

Solutions that ensure proper data integration and orchestration in an integrated ecosystem, mean that any updates, cleansing or reshaping of data can automatically carry over across the entire data platform. That means integration of legacy systems and new system technologies will be key to future solutions (Mamillion, 2022, online).

Scientists also see the need to develop a dynamic edge computing method to establish a balance between centralized and distributed functionality (Abdallah, 2022).

According to McKinsey Company's recent survey, approximately 8% of firms are high performers and use a best practice of automating most data-related processes, both improving efficiency in AI development and expanding the number of applications they can develop which increases the data quality fed into AI algorithms (Chui, 2022).

Velocity

More real-time techniques are required in the future for synching with the ever-increasing volume and heterogeneity of big data (Anwar et al., 2021). Combining innovative AI, machine learning and deep learning tools to shorten the time used to analyze big data and increase the volume of data that an organization can process in a given time will be key to managing complex and rapidly changing problems (Maxwell, 2021, online).

Variety

Since the variety, velocity and volume of big data generated from multiple sources and different types of data can affect the veracity and value of data, deep learning capabilities (architecture based on deep neural networks) designed to handle both structured and unstructured data easily, make deep learning a future viable solution to data processing and analysis challenges (Elaraby, Elmogy & Barakat, 2016).

Handling structured, numeric data in traditional databases as well as unstructured text documents, emails, videos, audios, stock ticker data and financial transactions is an on-going challenge in the big data ecosystem. Integrating legacy systems into big data systems, modernizing them, incorporating data from new sources, specifically "live" sources, will continue to be a future technical challenge (Jha, Jha, O'Brian & Wells, 2014).

Bansal & Kagemann (2015, 42) proposed a new semantic extract-transform-load framework to integrate varied data from heterogeneous sources that provides "richer data integration". Future solutions on how to manage daily, seasonal and event-triggered peak data loads also need to be studied (Lynch, 2019).

Scalability

In the future, existing frameworks will have to address a wide range of issues related to energy consumption, scheduling, data distribution and access, communication and synchronization, to continue scalability of big data analysis applications (Talia, 2019). Current frameworks are expected to be constantly improved for coping with volume and value "challenges involved in processing, storing and analyzing big data, allowing the effective extraction of useful knowledge in several application domains" (Belcastro et al., 2022, 48).

Geoff Hinton (2017), the "godfather of neural networks", believes that the future of deep learning lies in "better architectures, better ways of making the learning work in very deep nets and better ways of getting neural networks to focus on the relevant parts of the input".

The future calls for a new dynamic edge computing method to establish a balance between centralized and distributed functionality (Abdallah, 2022). Data specialists should integrate "ethnographic methods and qualitative data" for more "localized baseline knowledge and more targeted decision-making" (Palinkas et al., 2020; Vindrola-Padros et al., 2020).

Organizations need to "compress the neural networks so they use less bandwidth, less energy and fewer human resources to get a better non-predictive answer with a feature identification component that relates directly to the problem and the solution" (Kaplan, S., 2022). New innovative technologies and approaches like ALiX (AI and machine learning system using modal arithmetic based on set theory) are needed to "reduce complexity of the neural network by eliminating least relevant nodes," be "scalable in performance and infrastructure" and provide a more sustainable solution for the future (http://www.modal.net/solutions.htm, 2023).

For example, in a collaborative precision oncology research project, Nate Hayes, CEO and Founder of Modal Technology Corporation, and Dr. Anthoula Lazaris, Scientific Director at MUHC-RI (McGill University's Research Institute) developed a novel method for biomarker identification and ranking from non-invasive blood liquid biopsy samples in the early diagnosis of metastatic liver cancer. The new ALiX tool successfully discovered a predictive biomarker profile and ranked biomarkers from least to most relevant, which is leading to development of a commercialized point of care lab kit (Lazaris & Hayes, N., 2022).

Future solutions to processing big data include data distributed processing or parallel processing—where data is divided into multiple parallel smaller jobs processed by different processors—which completes a job faster, reduces the time of execution and improves performance (Abdallah, 2022). Photonics and alternate materials such as carbon nanotubes have been suggested as revolutionary solutions to increase scalability in the future (Black, 2021).

Top big data performers deploy AI at scale and have a data architecture that is modular enough to accommodate new technologies and applications rapidly (Chui, 2022). They also follow certain advanced scaling practices, such as using standardized tool sets to create production-ready data pipelines and an end-to-end platform for AI-related data science, data engineering and application development that they've developed in-house (Chui, 2022).

Management/Culture

Changing to a data-driven culture is quite difficult and demanding. So far, approximately a third of the over 5000 firms surveyed were successful on this front (Harvey, 2018). An integrated approach will continue to be refined and additional tools to help companies track their successes made available such as the "Open Innovative Management Framework", "semantic mining" and "balanced scorecard" used to analyze CEO messages sent to employees (Kim & Cho, 2022, Na, Lee, Choi & Kim, 2020).

The challenges of organizational resistance and the inaccuracies and vagueness of the data warrant solutions that include employees throughout the firm in the entire data life cycle, assessing the reliability and accuracy of data and making important information accessible to decision makers throughout the firm to develop sustainable innovative business strategies (Anwar, Gill, Hussain & Imran, 2021).

For example, in their experiment in the healthcare industry, Phillips-Wren & McKniff (2020, 528) show that tools to increase 'interactive data visualization' throughout the organization, putting it into the "hands of primary decision makers" helped increase the acceptance of big data in operational decision making.

Big data is confronting various difficulties in managing the organization (Abdullah, 2022). Zhou, Chawla, Jin & Williams (2014, 62) suggest that bringing together groups of people from diverse perspectives, different geographical locations with different core research expertise, different affiliations and work experiences can increase the ability to innovate in big data.

There is a "unique opportunity for integrative system designs, such as bounded-error network synchronization and dynamic scheduling based on machine learning" that could implement "distributed stream processing engines" using a light version of MapReduce technology which will help data analysis become more robust and visual in the future (Lv, Song, Basanta-Val, Steed & Jo, 2017, 1896).

If enterprises don't address data quality issues, they may find that the insights generated by their analytics are worthless, so better models and tools are needed in the future to ensure data veracity and value (Lv et al., 2018).

Organizations need to find their own balance among four areas: (1) "their skills in programming capabilities and knowledge of languages; (2) hardware availability; (3) application domain and purposes; and (4) software community support" (Belacastro et al., 2022, 48). The most high performing big data users follow core practices that unlock value, such as linking their AI strategy to their business (Chui, 2022). The frameworks working with big data need to include the requirements of innovation and the client's needs (Abdallah, 2022).

Important capabilities for the future will need an ensemble algorithm for effectively classifying the storage and the management layer and an enabled export or import function that can increase system functionality (Abdallah, 2022). Most companies' big data analytic system is limited to the Apache Hadoop suite. Hence, the separation of the data storage layer and the management layer is a challenge still needing to be met (Abdallah, 2022).

Graph databases and graph techniques like Visual.ly and Power BI will become the "standard tool for analyzing complex data relationships" visually and will increase the "ability to derive insights in increasingly complex ways" which are needed to manage big data effectively (Oracle, What is graph database? 2023, online; Rock, 2023, online)

Privacy, Security and Governance

Future development of "integrated solutions" and future research that evaluates the effectiveness of solutions in the security, privacy and governance arena for "different organizational and user contexts" are needed along with a complete security framework that covers the entire lifecycle of big data— "data sources, the data itself, data storage and the results of output devices" (Anwar, Gill, Hussain & Imran, 2021, online).

In the future companies will need to address "security in a scalable way" says Ron Black, CEO of Codasip, and his company did this by acquiring Cerberus to "gain extensive experience in security design and implementation" … which will match "customers' needs for ensuring their products meet both their security and business requirements." (Black, 2022, online).

Due primarily to the newness of big data, no end-to-end empirical solution for the most mentioned challenges of privacy and security is reported (Anwar et al., 2021, online). In the future, solutions and their interdependencies need to be considered. For instance, "communication security, architecture security, user validation, confidentiality, privacy at social networks and recovery are attained if proper access control is in place" (Anwar et. al, 2021, online).

Resources

To address the lack of resources to perform big data analysis for smaller and medium-sized companies, new low-cost, scalable and sustainable technologies and software solutions are still needed—ones that are open sourced, simpler technologies, able to be streamed and closer to the source are currently being sought (Lynch, 2019). A newly developed AI "forward-forward algorithm" is nearly the same speed as existing algorithms, but provides a much "higher bandwidth way to share knowledge," is more energy efficient and sustainable (Hinton, 2022).

New models such as edge computing—a distributed computing paradigm that brings computation and data storage closer to the data sources—is expected to improve response time and save bandwidth (Ahmadvand & Goudarzi, 2019, 5760).

One such solution proposed is called BlueDBM, an appliance for big data analytics that uses flash storage, in-store processing and integrated networks for cost-effective analytics of large datasets, which is a "magnitude cheaper and less power hungry than a cloud-based system" (Jun, Liu, Lee, Hicks, Ankcorn, King, Xu, 2015, 12). More collaboration between profit firms, research institutions and public entities to defray the costs for smaller and medium organizations would also be beneficial as modeled in the agricultural arena (Ouafiq, Saadane & Chehri, 2022).

The shortage of 1.5 million highly trained data professionals and managers that McKinsey predicted for 2018 (Atkinson, 2013) continues today (Chui, 2022). More training programs are needed to increase the supply of IT professionals (at both the undergraduate and graduate levels) and should continue to focus on "database management systems and data warehousing, programming in Python and R, predictive analytic tools and techniques, AI, deep and machine learning tools and techniques" but add "data engineering skills" to meet the growing need for these skills in the future (Kesh, Rudramuniyaiah, & Ramanujan, 2022).

Big data high performers are more likely than other organizations to “engage nontechnical employees in creating AI applications by using emerging low-code or no-code programs, which allow companies to speed up the creation of AI applications” (Chui, 2022, online).

Small and medium sized companies could benefit from such simplification of coding in programs too and could either work directly with consultants who are willing to work from home on ‘ad hoc’ big data projects or team with modular technology companies who provide free extensive big data support. Such strategies which would require fewer resources and lower costs than hiring full-time data scientist professionals (Waller, 2020).

Sustainability

Leaders want to sustain high-performance gains while the chip is operated within the range of the power and thermal budget and need future solutions that balance these competing factors (Bai, 2018). Optimization of bandwidth scaling and area and power-efficient solutions in chip architecture are being proposed (Zhou & Kodi, 2013; Wang, Hu, Lee & Bagherzadeh, 2011).

Solutions to the problems of high-power dissipation are needed as current processors already require extremely high power budgets (Rezaei, Shao, Daneshtalb & Wu, 2016; Zhan, Xie & Sun, 2014). Greener enterprise solutions will need to be integrated with big data ecosystems to promote long term, sustainable and environmentally friendly strategies (Meiyou & Ye, 2022).

Agricultural researchers proposed a comprehensive architecture that includes big data technologies, IoT components and knowledge-based systems called, “Smart Farming Oriented Big-Data Architecture” to guarantee the system’s durability and data modeling to transform the business needs for smart farming into analytics. It relies on a “pre-defined big data architecture with an abstraction layer of the data lake that handles data quality, following a data migration strategy in order to ensure the data’s insights” (Ouafiq et al., 2022, 329).

This ‘Smart Big Data Architecture’ concept could be applied across industries. Big data demands future solutions and approaches including technologies, methodologies, directions, research and gaps that need attention a indicated in Figure 2 (Appendix).

‘THE COMPREHENSIVE BIG DATA PARADIGM’: CHALLENGES, SOLUTIONS AND THE FUTURE

‘The Comprehensive Big Data Paradigm’ was developed to bring all the disparate needs of the big data ecosystem together. This “big picture” framework provides an overview of the most pressing challenges faced in adopting big data, the current solutions that are now being implemented and what is needed in the future to catapult the big data ecosystem.

The paradigm is ‘comprehensive’ which means it is complete, covering the wide variety of aspects of big data. This research answers the call for “end to end solutions” that will solve big data ecosystem challenges and fill the “gaps” in current practices (Jha et al., 2017, 125-126).

Embedded in this comprehensive framework are ‘integrative elements’. A solution to cleanse data in the cloud should also mean data at an in-house legacy server is being cleaned. Such synchronization is important to gain high value data and insights (Jha et al., 2017,125). To be most resource effective, solutions should be integrative ones, not just fixing a problem in one part of the system (Bansal & Kagemann, 2015). These include integration of synchronized systems, applications, methods, techniques and architectures (Orenga-Rogla & Chalmeta, 2019).

'The Comprehensive Big Data Paradigm' also embraces 'collaboration' such as sharing corroborated and contextualized data, finding ways to work with public & private entities on big data joint and open-sourced projects, creating innovation across industries and disciplines and being diversity-oriented in all aspects of big data from getting talent to processing and analyzing data. Even within an organization, various groups such as privacy officers, data science experts and functional managers do not communicate and collaborate well sabotaging the organization's decision-making effectiveness (Hagen & Hess, 2021).

Included in the 'big picture' overview of data, is 'sustainability', encompassing new mindsets, looking for long term adaptable and scalable technologies and resource saving solutions— automated, standardized, smarter, smaller, greener, cheaper, faster and open-sourced (Fouhy, 2022; Shahzad, Maqbool, Rana, Mirza, Khan, Kim, et al., 2022) .

In their model, Ortega-Rojla & Calmeta (2019, 58) review the characteristics of data for each part of the big data ecosystem giving step by step actions to take in each phase of the big data life cycle. In this research, the comprehensive paradigm also addresses the 5V's or data characteristics but adds three areas of management/culture, privacy/security/governance and resources which impact the big data ecosystem (from generating data to data driven decision making) while looking at future directions of big data.

Jokonya's (2015) conceptual framework for big data outlines the potential benefits and costs of big data adoption in a strategic matrix categorizing companies into four boxes—low or high need for quality data and for new technologies. The paper concluded that "there is no one-size-fits-all to big data adoption in organizations" (Jokonya, 2015, 150).

While this research corroborates that each organization should consider how to match the big data ecosystem to their own strategic goals, resources and outcomes, the use of big data is impacting nearly all organizations, private and public sectors alike and is seen as a competitive tool that is essential today.

Only through revolutionary approaches and a new paradigm for big data challenges can organizations successfully meet their future needs for data and "answer questions that could never be answered before" (Jha et al., 2017, 120). For example, in the healthcare industry, current changes in using and integrating the big data ecosystem have brought substantial advancements "ranging from medical data management to drug discovery programs for complex human diseases including cancer and neurodegenerative disorders" (Dash, Shakyawar, Sharma & Kaushik, 2019, 23). This iterative aspect is captured in this paradigm as the 'impact cycle'—big data challenges lead to current solutions and then gaps lead to future solutions, which in turn bring about new challenges to face.

As a predictive system, big data has the potential to predict epidemics, produce early warnings of diseases and create novel biomarkers or interventions for "improved quality of life" (Dash et al., 2019, 24).'The Comprehensive Big Data Paradigm' emphasizes the challenges, solutions and the future of big data, as indicated in Figure 3 (Appendix).

REFERENCES

Abdalla, H. B. (2022). A brief survey on big data: Technologies, terminologies and data-intensive applications. *Journal of Big Data, 9*(1), 107. doi.org/10.1186/s40537-022-00659-3

Ahmadvand, H. & Goudarzi, M. (2019). Significance-aware approach to improve qor of big data processing in case of budget constraint. *Journal of Supercomputing, 75*(9), 5760–5781.

Anwar, M. J., Gill, A. Q., Hussain, F. K., & Imran, M. (2021). Secure big data ecosystem architecture: Challenges and solutions. *EURASIP Journal on Wireless Communications and Networking,* doi.org/10.1186/s13638-021-01996-2

At Scale 2018 Big Data Maturity Scale, https://www.atscale.com/press/industrys-first-multi-year-big-data-maturity-survey-reveals-surprising-state-enterprise-adoption/

Atkinson, D. (2013). McKinsey Global expects shortage of big data analysts. *Solutions Review*, *December 9.*

Bai, M. (2018). *Performance-driven hierarchical design and management of networks-on-chip in many-core system* (Order No. 13420526). ProQuest Central. (2207467734), University of Louisiana at Lafayette, Lafayette, Louisiana.

Bansal, S. & Kagermann, S. (2015). Integrating big data: A semantic extract-transform-load framework, *Computer 48*(3), 42-50.

Barron, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory, 39*(3), 930-945.

Barron, A. R. & Khusowski, J.M. (2018). Approximation and estimation for high-dimensional deep learning networks, *Working paper at Rudgers, September 19*, 1-41. *arXiv:1809.03090v2*

Belcastro, L., Cantini, R., Marozzo, F., Orsino, A., Domenico Talia & Trunfio, P. (2022). Programming big data analysis: principles and solutions. *Journal of Big Data* 9(4), 1-50.

Black, Ron. (2021). Scaling is failing. *RISC-V Summit North America, Keynote Speech*, San Francisco, CA, December 6-8, 2021. Retrieved December 21, 2022 from https://youtu.be/AEQKrWbcoFk

Black, Ron (2022). Codasip to boost RISC-V security through Cerberus acquisition. *M2 Presswire , November 10.*

Chandarana, P. & Vijayalakshmi, M. (2014). Big data analytics frameworks, *Proceedings of IEEE International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA), April 5,* 430–434.

Chen, T., Chen, H., & Liu, R. (1995). Approximation capability in C(R^n) by multilayer feedforward networks and related problems. *IEEE Transactions on Neural Networks*, *6*(1), 24-30.

Chui, M. (2022). Commentary on our insights on the state of AI in 2022 and a half decade in review. *McKinsey Report.* Retrieved November 16, 2022 from https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review#talent

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, *2*, 303-314.

Elaraby, N., Elmogy, M. & Barakat, S. (2016). Deep learning: Effective tool for big data analytics. *International Journal of Computer Science Engineering (IJCSE), 5,* 254-262.

Enders, T., Martin, D., Sehgal, G. & Schüritz, R. (2020). Igniting the Spark: Overcoming Organizational Change Resistance to Advance Innovation Adoption – The Case of Data-Driven Services. 217-230.. In Nóvoa, M. Dragoicea, M. & Kühl, N. (editors), *Proceedings of Exploring Service Science - 10th International Conference (IESS), Porto, Portugal, February 5-7, Springer, 377*, 217-230 .

Esmaeilzadeh, R.S., Blem, E. & Burger, D. (2011). Dark silicon and the end of multicore scaling, *IEEE Micro*, *32*(3), 122–134.

Farrokhbakht, H., Kamali, H.M., & Hessabi, S. (2017). Smart: A scalable mapping and routing technique for power-gating in noc routers," in *Proceedings of IEEE/ACM International Symposium on Networks-on-Chip (NOCS).October,* Seoul, South Korea: IEEE, 1–8.

Foote, K.D. (2018). Big data vs. smart data. *Dataversity*, *May 10*, https://www.dataversity.net/big-data-vs-smart-data/

Fouhy, J. (2022). How digital transformation—and a challenging environment—are building agility and resilience. *Deloitte Insights , April 29*.

Gebauer, H., Haldimann, M., & Caroline, J. S. (2017). A typology for management innovations: Analyzing the actual state of art and conducting case study research. *European Journal of Innovation Management, 20*(4), 514-533.

General Data Protection Regulation (GDPR). Retrieved on January 2, 2022 from https://gdpr.eu/what-is-gdpr/

Gewirtz, D. (2018). Best cloud services for small businesses. *CNET, July 10*.

Gupta, R., Gupta, H., & Mohania, M. (2012). Cloud computing and big data analytics: What is new from databases perspective? *Proceedings of international conference on big data analytics*, *December 24*, 42–61.

Hagen, J. & Hess, T. (2021). Collaboration for big data analytics: Investigating the (troubled) relationship between data science experts and functional managers. *Proceedings of the 54th Hawaii International Conference on System Sciences,* 254-264.

Harvey, C. (2018). Big data pros and cons. *Datamation, August 9*.

Hassan, U., Nastase, S. A., & Goldstein, A. (2020). Direct fit to nature: An evolutionary perspective on biological and artificial neural networks. *Neuron, 108*, 416-434.

Hechanova, M.R.M., Caringal-Go, J. F. & Magsaysay, J.F. (2018). Implicit change leadership, change management, and affective commitment to change. *Leadership & Organization Development Journal, 39*(7), 914-925.

Hinton, G. (2017) in Interview on *Gigaom: Four questions for: Geoff Hinton* (2017). Chatham: Newstex.

Hinton, G. (2022). The forward-forward algorithm: Some preliminary investigations. *Ithaca: Cornell University Library*, arXiv.org.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks, 4,* 251-257.

Hornik, K., Stinchcombe, M. & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks, 3*, 359-366.

Jha, S., Jha, M. O'Brien, L. & Wells, M. (2017). Supporting decision making with big data: Integrating legacy systems and data. *2017 4th Asia-Pacific World Congress on Computer Science and Engineering, IEEE, December 1*, Mana Island, Fiji, 120-128.

Jokonya, O. (2015). Towards a conceptual framework for big data adoption in organizations. In *Proceedings of the 2015 International Conference on Cloud Computing and Big Data (CCBD)* IEEE Computer Society, USA, 153–160.

Jun, S. W., Liu, M., Lee, S., Hicks, J., Ankcorn, J., King, M., & Xu, S. (2015). Bluedbm an appliance for big data analytics. In: Marr D, editor. *ACM/IEEE 42nd annual international symposium on computer architecture ISCA*. Manhattan: IEEE, 1–13.

Kaplan, S. (2022). "Future Directions of Big Data". Interview by Diane Fulton, *December 29, 2022.*

Kesh, S. Rudramuniyaiah, P. & Ramanujan, S. (2022). Bid data and analytics programs in the US: A pedagogical analysis. *Journal of Technology Research, 10,* 1-7.

Klusowski, J. M. & Barron, A. R. (2017). Minimax lower bounds for ridge combinations including neural networks. *Proceedings of IEEE International Symposium on Information Theory*, June, Aachen, Germany, 1377-1380.

Kruse, C. S., Goswamy, R., Raval, Y., & Marawi, S. (2016). Challenges and Opportunities of Big Data in Health Care: A Systematic Review. *JMIR medical informatics*, *4*(4), e38. https://doi.org/10.2196/medinform.5359

Kh, R.(2018). Contextual data collection is the future of modern decision making. *SAP Community, January 22*.

Kim, H.J. & Cho., K.T. (2022). Analysis of changes in innovative management of global insurers in the pre- and post-covid-19 eras. *Sustainability, 14*, 9976-9997.

Latifian, A. (2022). How does cloud computing help businesses to manage big data issues. *Kybernetes, 51*(6), 1917-1948.

Lazaris, A. & Hayes, N. (2022). "Riding the Third Wave of AI for Precision Oncology" in podcast interview led by Maria Palombini, *Re-Think Health*, *May 19, Season 3*, Episode 2.

Leshno, M., Lin, V.Y., Pinkus, A. & Schocken, S. (1993). Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks, 6*, 861-867.

Lynch, C.F. (2019). Is big data too big? *Worley Blog*, *September 10,* Retrieved September 9, 2022 from https://www.worleywarehousing.com/

Lv, Z., Song, H., Basanta-Val, P., Steed, A. & Jo, M. (2017). Next-Generation big data analytics: State of the art, challenges, and future research topics. *IEEE Transactions on Industrial Informatics, 13*(4), 1891-1899. doi: http://dx.doi/10.1109/TII.2017.2650204

Magnimind Academy. Blog. What are the advantages and disadvantages of big data. *Aug 10, 2022*.

Matillion, (2022), The big data analytics process in the cloud, *June 8,*

Maxwell, R. (2021). Machine Speed Strategic Analysis – Using AI to unlock information advantage. *Trilteral Research, Sociotech Insights, October 15*,

Maya Gopa, P.S., & Chintala, B. R. (2020). Big data challenges and opportunities in agriculture. *International Journal of Agricultural and Environmental Information Systems (IJAEIS), 11*(1), 48-66.

Meiyou, D. & Ye, Y. (2022). Establishment of big data evaluation model for green and sustainable development of enterprises, *Journal of King Saud University - Science, 34*(5), 102041, https://doi.org/10.1016/j.jksus.2022.102041.

Modal, Solutions (2023). Retrieved on January 20, 2023 from http://www.modal.net/solutions.htm.

Na, H.J.; Lee, K.C.; Choi, S.U.; Kim, S.T. (2020). Exploring CEO messages in sustainability management reports: Applying sentiment mining and sustainability balanced scorecard methods. *Sustainability, 12*, 590-608.

O'Halloran K.L., Tan S., Pham D.S., et al. (2018) A digital mixed methods research design: Integrating multimodal analysis with data mining and information visualization for big data analytics. *Journal of Mixed Methods Research, 12*(1), 11-30.

Oracle, What is Big Data? (2023). *Oracle Insights (OCI), January*, Retrieved January 14, 2023 from https://www.oracle.com/big-data/what-is-big-data/

Oracle, What is graph database? (2023) *Oracle Insights (OCI), January*.

Orenga-Rogla, S. & Chalmeta, R. (2019) Framework for implementing a big data ecosystem in organizations. *Communications of the Association for Computing Machinery (ACM), 62*(1), 58–65. https://doi.org/10.1145/3210752

Orts, E., & Spigonardo, J. (2014). *Sustainability in the age of big data*. Special Report. Initiative for Global Economic Leadership (IGEL)/Wharton. Philadelphia, PA: University of PA.

Ouafiq, E. M., Saadane, R., & Chehri, A. (2022). Data management and integration of low power consumption embedded devices IoT for transforming smart agriculture into actionable knowledge. *Agriculture, 12*(3), 329-345. doi.org/10.3390/agriculture12030329

Phillips-Wren, G., & McKniff, S. (2020). Overcoming Resistance to Big Data and Operational Changes Through Interactive Data Visualization. *Big data*, *8*(6), 528–539.

Mariusz, P. & Marek, B. (2022). Neural network training with limited precision and asymmetric exponent. *Journal of Big Data, 9*(1), 1-17.

Pokhriyal, N. & Jacques, D.C. (2017) Combining disparate data sources for improved poverty prediction and mapping. *Proceedings of the National Academy of Sciences of the United States of America* 114(46): E9783–E9792.

Poorthuis, A., & Zook M. (2017). Making big data small: Strategies to expand urban and geographical research using social media. *Journal of Urban Technology, 24*(4), 115–135.

Raeburn, A. (2022). SWOT analysis: What it is and how to use it, *Asana Resources, November 28*, https://asana.com/resources/swot-analysis

Rezaei, A., Zhao, D., Daneshtalab, M. & Wu, H. (2016). Shift sprinting: fine-grained temperature-aware Noc based mcsoc architecture in dark silicon age, in *Proceedings of IEEE Design Automation Conference (DAC)*, *June,* Austin, TX, USA: IEEE, 1–6.

Robertson, P. (1997). Integrating Legacy Systems with Modern Corporate Applications, *Communications of the ACM, 40*(5), 39-47.

Rock Content Editor, (2020). Big data visualization: what it is, techniques and best tools, *Rock Content, June 1* https://rockcontent.com/blog/big-data-visualization/

Semeraro, R., Magklis, G. &. Scott, M. (2002). Energy-efficient processor design using multiple clock domains with dynamic voltage and frequency scaling. *Proceedings of IEEE International, International Symposium on High-Performance Computer Architecture*, *February*, 29–40.

Shahzad, I., Maqbool, A., Rana, T., Mirza, A., Khan, W. Z., Kim, S. W., . . . Din, S. (2022). Blockchain-based green big data visualization: BGbV. *Complex & Intelligent Systems, 8*(5), 3707-3718.

Shaozeng, Z., Dafeng, X. & Zhao, B. (2022). “Small” analysis of big data: An evaluation of the effects of social distancing in the United States. *Methodological Innovations, 15*(2), 142-151.

Stobierski, T. (2019). How to do a cost-benefit analysis & why it’s important. *Business Insights Harvard Business School, September 5.*

Talia, D. (2019).A view of programming scalable data analysis: from clouds to exascale. *Journal of Cloud Computing, 8*(1), 1-16.

Thompson, W. (2023) Big data: Why it is and why it matters? *SAS Insights, January*

Van Arnum, P. (2014). Big Data: Five Mega Trends Shaping the Pharmaceutical Industry. *DECAT Value Chain Insights, June*,

Volberda, H.W., van den Bosch, F.A. & Heij, C.V. (2013). Management innovation: management as fertile ground for innovation. *European Management Review, 10*(1), 1-15.

Waller, D. (2020). 10 steps to creating a data-driven culture. *Harvard Business Review, February 20*, online.

Wang, C., Hu, W., Lee, S. & Bagherzadeh, N. (2011). Area and power-efficient innovative congestion-aware network-on-chip architecture, *Journal of Systems Architecture, 57*(1), 24-38.

Wielgus, D. (2023). Interview with Data Tech for Master Lock on February 11, 2023.

Zhan, J., Xie, Y. & G. Sun, G. (2014). Noc-sprinting: Interconnect for fine-grained sprinting in the dark silicon era," in *Proceedings of IEEE, Design Automation Conference (DAC). June*, San Francisco, CA, USA: IEEE, 1–6.

Zhang, S., Zhao, B., & Ventrella, J. (2018) Towards an archaeological-ethnographic approach to big data: Rethinking data veracity. *Ethnographic Praxis in Industry Conference Proceedings 2018*, 62–85.

Zhang, S. Xu, D. & Zhao, B., (2022) 'Small' analysis of Big Data: An evaluation of the effects of social distancing in the United States. *Methodological Innovations, 15*(2), 142-151.

Zhou, L. & Kodi, A. K. (2013). Probe: Prediction-based optical bandwidth scaling for energy-efficient nocs, in *Proceedings of IEEE/ACM International Symposium on Networks on Chip(NoCS), April,* Tempe, AZ, USA: IEEE, 1–8.

Zhou, Z. H., Chawla, N.V., Jin. Y. & Williams, G. J. (2014). Big data opportunities and challenges: Discussions from data analytics perspectives, *IEEE Computational Intelligence Magazine, 9*(4), 62-74.

APPENDIX

Figure 1. Overview of the Most Pressing Big Data Challenges By Category with Descriptions of Issues

Challenge Category	**Descriptions of Most Pressing Challenges**
Value	Meaningful data must bring insights and outcomes (i.e. performance, revenue, and strategic goals, be sustainable).
Variety	Data arrives in various forms (i.e. structured and unstructured) and from different sources needing to be meshed and communicated easily.
Velocity	Data comes in too fast; Handling of data needs to be timely (quick or real time); Seasonality, new products and cyclical issues impact data speed.
Veracity	Need high quality data—accurate, relevant and in analyzable format.
Volume	Too much data can be overwhelming; Storing and accessing data is complex; Processing massive data takes specialized technologies.
Management & Culture	Leaders need to align the big data ecosystem, strategies and outcomes; Creating a data-driven culture is difficult; Employees may be resistant to change; Balancing competing issues like performance and sustainability.
Privacy, Security & Governance	Data that is sensitive, personal or financial, is a target for cyberattacks; data needs to be handled carefully to ensure confidentiality and privacy; regulations are ever changing and firms needs to comply with laws and protect data from mishandling, breaches or fraud attempts.
Resources	Costs are increasing for new technologies, hardware, software, services and professionals; Difficult to find talent or train existing staff with new skills; Obsolete technologies; Integrating legacy systems with new systems is hard and costly; Time and money needed.

Figure 2. Challenges, Solutions and Future for Big Data

	Challenges	*Solutions*	*Future*
Value	▪ Meaningful data brings positive results	▪ Smart data ▪ Small data ▪ Cloud/warehouses/ data lakes	▪ Contextualized and corroborated data analysis to enhance value ▪ Small data analysis takes most relevant data giving added value
Variety	▪ Integrating data in various forms and from different sources	▪ Small data, corroborative analysis & contextual interpretation: structure from new algorithms ▪ Extract-transform-load software like Clover	▪ Deep neural network innovations to handle data variety: make data small ▪ Semantic ETL innovations in processing varied data ▪ Integrate legacy and big data systems into big data systems better
Velocity	▪ Data is fast ▪ Handling is quick/real time. ▪ Data speed varies.	▪ Smart data for fast or real time processing ▪ Smart sensors speed up data transfers ▪ Data stream algorithms ▪ Integrative applications	▪ More real-time techniques for synching big data ▪ Combine innovative AI, machine learning and deep learning tech/algorithms to shorten the time used to analyze and process big data
Veracity	▪ Need high quality data that is accurate, relevant and in a proper format to analyze.	▪ Smart sensors decrease errors ▪ Cloud tools, processing, data warehouses/lakes increase data accuracy ▪ AI-powered analytics with machine speed strategic analysis	▪ Integrate and articulate more diverse data sources and methods for more contextualized and corroborated data analysis ▪ Synchronized solutions to integrate the big data ecosystem ▪ Dynamic edge computing: centralized/distributed functions ▪ Automating data processes
Volume	▪ Too much data can be overwhelming ▪ Storing and accessing data is complex ▪ Processing massive data takes specialized technologies.	▪ Smart data sensors automate data collection ▪ "Small data" from big data. use smaller datasets ▪ Cloud data warehouses and data lakes scale ▪ Transistors to multi-core, to many core scaling ▪ Bus-based and point-to-point interconnection networks in multi-core ▪ Networks-on-Chip interconnects for wire and power in many core ▪ Heterogeneous computing matching hardware and workload ▪ NoSQL systems for interactive data serving ▪ Large scale analytics: MapReduce & Hadoop in cloud-based arena ▪ Hadoop systems to run long running decision support & queries ▪ Data stream algorithms/ integrative applications ▪ AI & machine learning	▪ New scalability frameworks for energy consumption, scheduling, data distribution, communication and synchronization of data ▪ Improve existing frameworks to cope with volume and value challenges to extract useful data through the entire ecosystem ▪ New dynamic edge computing method balances centralized and distributed functionality ▪ Integrate ethnographic/qualitative methods for richer insights ▪ Compress neural networks to use less bandwidth, energy and resources for decision quality ▪ Use new innovative, sustainable and scalable technologies like ALiX (AI and machine learning system based on set theory) ▪ Data distributed processing or parallel processing for faster job, better time and performance ▪ Photonics and alternate materials such as carbon nanotubes are revolutionary scalable solutions ▪ Scale AI with a modular approach to data architecture

Management & Culture	▪ Align the big data ecosystem with strategies and outcomes ▪ Creating a data-driven culture takes time and real commitment ▪ Employees may be resistant to change ▪ Balance performance with sustainability.	▪ Cloud data warehouses and data lakes lower the management burden ▪ AI-powered analytics & machine learning make decision-making faster. ▪ Dynamic voltage and frequency scaling and power gating balance energy and performance. ▪ Change management techniques can overcome employees' resistance ▪ Align big data system, strategies and outcomes ▪ Strategic analysis tools like SWOT or cost-benefit analysis ▪ Communicate with all company stakeholders ▪ Graphic/visual tools like Tableau to make decisions easier	▪ Low-cost, scalable & sustainable solutions. ▪ Edge computing brings computation and data storage closer to sources, improves response time and saves bandwidth. ▪ More cost effective and less power-hungry solutions like BlueDBM with flash storage, in-store processes ▪ Collaboration with various entities improves decisions. ▪ More training programs are needed to increase the supply of IT professionals and increase data science skills ▪ Use nontechnical employees to create big data approaches using low-code or no-code programs. ▪ Comprehensive models for smart big data architecture ▪ Data-driven culture- interactive visualization & wide data access: Visual.ly, PowerBI
Privacy, Security & Governance	▪ Sensitive or private data- cyberattacks ▪ Ensure data's confidential and private ▪ Comply with laws/ regs. ▪ Fix breaches mishandling, or fraud.	▪ Invest in processes, protocols, and infrastructures to protect data security and privacy and increase governance using GDPR guidelines ▪ Big data analytics/ma-chine learning detect fraud patterns and anomalies. ▪ Encryption, access control and anonymity.	▪ Integrated solutions for different contexts ▪ Security framework covering the entire big data lifecycle ▪ End-to-end empirical solutions ▪ Recognize interdependencies ▪ Partner/Acquire to gain better solutions to improving confidentiality, privacy, governance and stop mishandling, breaches or attempted fraud.
Resources	• Hardware, technologies, software, services & staff cost ▪ Hard to find talent or train existing staff ▪ Obsolete technologies ▪ Integrating legacy & new systems is hard & costly ▪ Resources for big data adoption.	▪ Cheaper to store in data lakes, Hadoop & clouds ▪ Use less costly alternatives to the big cloud storage like Cloudways. Backblaze or Digital Oceans. ▪ Cost/benefit or SWOT analysis to assess risks & get buy-in lower costs. ▪ AI-based dynamic OOP /tools to integrate systems and automate. ▪ Start with a small (less time/money) project for employees' confidence. ▪ Find people or train existing staff with new data skills most needed. ▪ Commit sufficient data budget and resources.	▪ Low-cost, scalable, greener and sustainable big data solutions ▪ New models: edge computing improves response time and saves bandwidth, forward-forward algorithms save energy. ▪ Cost effective analytics solutions use less power and money: Blue DBM uses flash storage, in-store processing, integrated networks. ▪ Greater collaboration among entities to defray costs: farming ▪ More training needed for new IT professionals/existing employees ▪ Training nontechnical low-code or no-code programs speeds up innovation ▪ Use ad hoc consultants on small data projects ▪ Partner with modular technology companies for free data support

Figure 3. Comprehensive Big Data Framework: Challenges, Solutions and the Future (An Impact Cycle)