AOSIS

Human interaction with structure in designing big data architecture

CrossMark

Authors:
Wandisa Nyikana[1]
Tiko Iyamu[1]

Affiliations:
[1]Department of Information Technology, Faculty of Informatics and Design, Cape Peninsula University of Technology, Cape Town, South Africa

Corresponding author:
Tiko Iyamu,
iyamut@cput.ac.za

Background: There is an increasing interest in big data. However, challenges shape and affect the gathering, retrieval, use and management of big data in many organisations. Some of the challenges are linked to a lack of architecture that is specific to big data. Attempts have been made from both business and academic fronts, yet the challenges persist. The challenges are attributed to a lack of an understanding of the factors that influence the design of architecture for big data in an organisation.

Objectives: The study aims to propose big data architecture for enterprises.

Method: We employed the qualitative method, using document analysis to gather data. Activity theory (AT) was employed in the analysis of the data.

Results: From the analysis, governance, interactions, relationships and allocative were found to be the factors that influence the design of big data architecture. An interpretation was conducted following the inductive reasoning approach to gain a deeper insight of how the factors manifest themselves.

Conclusion: Big data architecture is proposed. The architecture is intended to address some of the challenges encountered in gathering, retrieving, using or managing big data in organisations.

Contribution: This study advances our understanding of the complex interplay of factors influencing the architecture of big data. Applying AT, the study fortifies our understanding of complex interactions between humans and big data including the architecture design.

Keywords: activity theory; architecture; big data; design; information technology governance; human interaction; information systems theory.

Introduction

Big data analysis enables organisations to use their current data to create knowledge that can be transformed to gain a competitive advantage (Hajli et al. 2021). Additionally, it helps to enhance the performance of an organisation, promote business innovation and increase sustainability (Zhang et al. 2022). According to Iyamu (2020), big data helps to uncover new patterns and make future predictions. Predictive analytics helps to create business value and provide decision support capabilities (Al-Sai & Abdullah 2019).

Organisations in various industries from the private to the public sector make use of big data for different activities and processes (Avci, Tekinerdogan & Athanasiadis 2020). For instance, governments use big data to forecast social and economic changes such as unemployment levels and to improve service delivery to citizens (Blazquez & Domenech 2018). In agriculture, it is used to determine the techniques that can be used to produce agricultural products (Prasetyo et al. 2019). This is done by evaluating data about the type of soil, the temperature and the biodiversity. Healthcare uses big data for diagnosing diseases and reducing healthcare costs (Manogaran, Thota & Lopez 2022) by reducing patient readmissions and preventing frequent emergency room (ER) visits (Pramanik, Pal & Mukhopadhyay 2022). In marketing, big data generated from transactions are often used to establish customer needs and buy patterns (Al-Sai & Abdullah 2019). In education, it improves educational effectiveness by promoting data-driven approaches to teaching and learning (Fischer et al. 2020). However, the gathering and use of big data by organisations is not by default; it entails approaches that are governed by standards and principles (Iyamu 2023).

Despite the wide range and reliance on the use of big data, many organisations are frequently or continually confronted with challenges, some of which are detrimental. One of the challenges is

the complex characteristics of big data, which include volume, variety, veracity and velocity (Nyikana & Iyamu 2023). According to Jeske and Calvard (2020), some of the challenges are in collecting, using and managing big data in an organisation. The challenges are of both technical and non-technical nature. From the technical perspective, the host infrastructure such as software, security and hardware are discussed (Zhang et al. 2022). The heterogeneous nature of big data makes the enabling infrastructure more critical and requires more attention such as the architectural guidelines. Jones (2019) argues that some of the challenges can be prohibitive. Therefore, the cost of collecting, analysing and storing big data should not be taken for granted.

Some of the challenges affect the value of big data in organisations. Nyikana and Iyamu (2022) state that the value of big data is influenced by its architecture. Iyamu (2022a) explains how architecture guides the selection and use of information technology (IT) solutions. Ferraris et al. (2019) argue that the value of data does not only depend on its accuracy and quality, but it also depends on other factors such as actors (humans) that are required to collect and analyse the data. Furthermore, the actors require governance in their use of processes for practices, which architecture provides.

The introduction section of the study indicates that some of the challenges encountered in gathering, retrieving, using or managing big data are because of a lack of architecture. Most importantly, the influencing factors are either unknown or the root is unclear in many organisations. Thus, the study's objective is to examine the determinants of the big data architecture design. This includes how the factors manifest themselves to influence the architecture, from both technical and non-technical perspectives.

The paper is structured into seven main sections. The first section introduces the phenomenon being studied. This is followed by a review of the literature in the second section. In the third section, activity theory (AT) that underpins the study is discussed. The methodology applied is discussed in the fourth section. The analysis and discussion are presented in the fifth and sixth sections, respectively. Finally, a conclusion is drawn in the last section.

Literature review

Big data architecture consists of layers that have different functions (Benhlima 2018). The functions and components of each layer are defined based on the organisational requirements and technological needs (Kalipe & Behera 2019). Also, the layers and components are influenced by the characteristics of big data. According to Wang et al. (2020), the layers include collection, storage, processing and visualisation of datasets. Governance is considered an important layer because of its cruciality, such as defining quality standards, security principles, compliance policies and determining access to data (Farooqi et al. 2019). Consequently, when big data architecture does not have a governance layer, the overall effectiveness and performance of the system can be affected. This results in

challenging implications such as a lack of compliance, a lack of standardised processes and security vulnerabilities.

The design of big data architecture that is based on organisational requirements generates value and improves competitiveness (Blazquez & Domenech 2018). Another perspective is that an organisation that deploys scalable and flexible big data architecture tend to be more responsive to rapid changes in its business environment, and has a positive impact on its overall performance. However, organisations struggle to design or find a big data architecture that is suitable for their environment (Ruiz et al. 2021). Avci et al. (2020) linked this challenge to fit or align with the application and business requirements of the organisations, which are the technical and non-technical factors that influence the design of a big data architecture.

Also, big data architectures can be difficult to implement because of various factors required to support the adoption of new architecture (Farooqi et al. 2019). Additionally, finding the right tools to utilise big data architecture is a challenge (Gökalp et al. 2019). Furthermore, the integration of the existing models with new models such as big data architecture is another challenge highlighted by Bansal et al. (2022) and Iyamu (2013). It is, therefore, essential to consider these challenging factors in designing an architecture for big data, which did not seem to be in existence at the time of this study (Tschoppe & Drews 2022; Saggi & Jain 2018).

Different big data architectures have been proposed and developed for various purposes. An example is the lambda and kappa architecture, which is concerned with processing real-time scalable big data (Barradas et al. 2022). There is also the Hadoop framework, which consists of a Hadoop distributed file system (HDFS), to store huge amounts of structured and unstructured data sets (Oussous et al. 2018). Another example is MapReduce, whose primary focus is to process data sets (Farooqi et al. 2019). Boumlik and Bahaj (2018) state that HDFS and MapReduce are used in parallel to process, store and retrieve large data volumes.

In attempts to address existing gaps, Wehn et al. (2021) propose a big data architecture that is concerned with integrating big data from different data sources' perspectives. Filaly et al. (2022) designed a big data architecture focussing on ensuring the security of the data from attacks and vulnerabilities. Even though there are different architectures that have been designed or developed, the challenges persist. This means none of the existing architecture seems to cover all the characteristics of big data. Hence, Mostefaoui et al. (2022) highlight the need for an architecture that covers all the characteristics and layers of big data. This is to have a holistic architecture that addresses the huge volume, rapidity of veracity, fluidity of variety, contextualisation of value and precedented velocity of datasets.
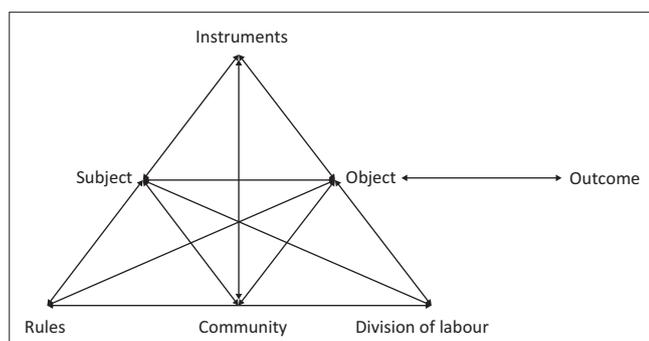
Theory underpinning the study

Because of the various activities involved in the design of architecture, some of which are highlighted earlier, the AT is

used to underpin the study. Activity theory is a socio-technical theory that has been adopted in information systems (IS) studies in the last three decades (Iyamu 2022b). The theory consists of six components which are subjects, objects, tools (instruments), rules, community and division of labour (Park et al. 2013). The primary concern of the theory is the development of social activities (Shaanika & Iyamu 2015). The theory focusses on understanding the interactions and relationships that occur as activities are performed by humans (Iyamu & Shaanika 2019). Hence, according to Dennehy and Conboy's (2017) explanation, AT is a framework that is used to understand complex human activities within a social system. Nehemia, Iyamu and Shaanika (2018) described AT as a theory of consciousness. The reason for that is because the activities performed are consciously planned.

As shown in Figure 1, the components are interconnected and interrelated, indicative of the arrows. The interconnections and relationship of the components help to understand the overall activities of the system (Nehemia et al. 2018). Also, as expressed by AT, activities are not static, they constantly evolve because of the changes in the environment (eds. Engeström, Lompscher & Rückriem 2016).

In AT, rules are control mechanisms, which can include policies, regulations and legislations that guide and govern how subjects perform their activities (Kelly 2018). The rules help to maintain order and control conflicts within an activity. Also, rules are used to manage the interactions between actors during the process of allocating tasks, roles and responsibilities among community members (Karanasios & Allen 2013). In AT, a subject is a human being or a collective of people involved in an activity, also referred to as an actor (Sannino & Engeström 2018; Shaanika & Iyamu 2015). The activities carried out by subjects are consciously planned. This is to ensure that the activities have a purpose and are not aimless (Dennehy & Conboy 2017). Nehemia-Maletzky et al. (2018) state that human consciousness is the basic principle of AT.

Community is defined by Iyamu and Shaanika (2019) as a collective of individuals in a social system working towards the same goal. Division of labour refers to the allocation of tasks and responsibilities among community members (Lioutas et al., 2019). The object component of the AT model

is the motive for carrying out an activity, and it can be tangible or intangible (Iyamu 2022b). According to Sannino and Engeström (2018), the object gives an activity a sense of direction and significance. These are critical in gaining an understanding of roles and responsibilities including interrelationships, in designing a big data architecture (Garoufallou & Gaitanou 2021).

The components of AT as described and discussed precedingly are usually norms in IT solutions; hence, the theory is increasingly connected with IS research. Subjects depend on tools to mediate with the object. Tools can be technical or psychological artefacts (Hasan & Kazlauskas 2014). Technical tools such as computers intend to manipulate physical objects, while psychological tools such as language are used by human beings to influence each other. These are fundamental elements and factors that can manipulate or influence an IT solution such as the design of architecture for big data. Er, Kay and Lawrence (2010) suggest that tools shape the way subjects interact with objects and influence the outcome of the activity. This is because tools have an enabling and constraining function in them, whereby they can transform or limit the object, depending on motive and how the actor employs them.

Methodology

Document analysis was employed in the data collection, primarily, because it enables the streamlining of materials through analysis of documents such as books, newspaper articles, peer-reviewed articles and organisational reports (Morgan 2022). The data were collected using a set of criteria that included area of specialisation, publication timeframe and credible sources. The areas of specialisation were big data and architectural design, which are the core aspects of the study. A period of 10 years was considered, to gain an understanding of the historical background and meanings associated with the concepts over time (Iyamu Nehemia-Maletzky & Shaanika 2016). The documents helped to give a comprehensive and holistic view of big data architecture. Thus, only a small sample of the most appropriate and relevant literature could be gathered (Glass, Ramesh & Vessey 2004).

Materials published in journal outlets, books, conference proceedings and the Internet between 2013 and 2023 were gathered. Academic databases such as Ebscohost, IEEE, AIS, and Emerald were used as sources for the collection of the data. This helps to ensure the credibility and reliability of the data (Nyikana & Iyamu 2023). The analysis of the documents was two-fold: academic papers (peer-reviewed) and non-academic papers (white papers and green papers). As shown in Table 1, a total of 201 papers were collected.

Source: Engeström, Y., Lompscher, J. & Rückriem, G. (eds.), 2016, *Putting activity theory to work: Contributions from developmental work research*, 13th edn., Lehmanns Media, Berlin

FIGURE 1: Activity model.

TABLE 1: Collected papers.

Paper	Peer-reviewed	Non-peer-reviewed	Total
Big Data	86	23	109
Big Data Architecture	39	10	49
Enterprise	33	10	43
Total	**158**	**43**	**201**

The data were coded, and a format was formulated for ease of referencing the data. An example of the format is as follows: BDTDoc01; Pg#: Ln#. This means the first of big data documents, the page number and line number. As shown in Table 1, the documents (used as data) are categorised into three groups, big data, enterprise and architecture, the core aspects of the study. The categories are further divided into two groups: peer-reviewed and non-peer-reviewed. Each of the groups is assigned codename, BDTDoc01 . . . BDTDocn[+1] for big data; ENTDoc01 . . . ENTDocn[+1] for the enterprise; and ARCDoc01 . . . ARCDocn[+1] for architecture.

Data analysis and discussion

Qualitative materials (data) were gathered, in achieving the objective of the study (Patel & Patel 2019). This was systematically conducted, to ensure that the most appropriate materials were referenced (Iyamu 2022a; Kothari 2020). The AT model, as shown in Figure 1, was employed by using the components to guide the analysis. This helps in achieving the objective of the study in three ways. Firstly, it assisted in gaining a better understanding of how big data are stored and governed in enterprises. Secondly, it fortifies the fathoming of insights and evidence provided in the materials examined, to gain and understand the factors that influence the design of big data architecture for enterprises. Thirdly, it helps to comprehend the relationships between architectural components (technical and non-technical factors) that suit big data in the context of enterprises.

Activity theory is employed as a lens to provide a frame for the use of the hermeneutic approach in the analysis. This helps to gain a fathomed and in-depth view in proposing the architectural design for big data. The analysis is presented following the six components, tools, subject, rules, community, division of labour, and object of AT.

Activity theory: Tools

In AT, tools refer to artefacts used in an activity to transform an object into an outcome (Sannino & Engeström 2018). Tools differ depending on the objective and the context of the study. Tools may include machines, instruments, signs, procedures and laws (Nehemia et al. 2018). There are different types of tools when it comes to big data architecture. This includes technical artefacts (software and hardware) and non-technical artefacts (language) that need to be considered when designing the architecture of big data. Scalable storage and processing play a role in the design of big data architecture. Tools can be used in isolation while others may be integrated with others. For instance, HDFS and MapReduce are used in parallel to process, store and retrieve large volumes of data (Boumlik & Bahaj 2018). Additionally, when designing an architecture for big data, some strategies and practices need to be followed (Oussous et al. 2018). It is stated as follows in some of the materials used for this study:

'This scheme was put about cloud computing, whose potential and benefits for storing huge amounts of data and performing powerful calculus are positioning it as a desirable technology to be included in the design of a Big Data architecture.' (BDTDoc01; Pg2: Ln11-14)

'Without appropriate organizational structures and governance frameworks in place, it is impossible to collect and analyze data across an enterprise and deliver insights to where they are most needed.' (BDTDoc51; Pg417: Ln25-27)

When incorrect tools such as storage and processing software are selected, challenges are sometimes encountered in the areas of performance, reliability, flexibility and limited features (Chen, Kazman & Haziyev 2016). Also, selecting an incorrect tool potentially results in an organisation not gaining value from big data technology (Nyikana & Iyamu 2022). Hence, Iyamu (2022a) argues that the selection of IT solutions should be guided by the architecture. Additionally, there are costs associated with tools, from purchase to support and maintenance perspectives. Thus, how and why tools are selected, used and managed becomes critical. This helps to avoid prohibitive circumstances that affect the effective and efficient use of the tools.

Activity theory: Subject

Subject refers to a human being or a collective of people involved in an activity (Shaanika & Iyamu 2015). There are different subjects involved in the activity of designing big data architecture. These subjects need to have the right skills and knowledge that are required to achieve the objective. Boumlik and Bahaj (2018) highlight that developers need to have language query skills to extract and present the correct data that is valuable to an organisation. The skills of the subjects need to align with their roles to perform the big data-related tasks assigned to them (Mohammad, Mcheick & Grant 2014). In one of the articles, it is revealed as follows:

'Organisations need to continuously plan and manage a trained workforce that can handle its Big Data technologies, and as such, this capability too can be considered a critical success factor for sustainable implementation of Big Data.' (BDTDoc14; Pg5: Ln6-8)

Organisations that do not have subjects with the right skills in their environment often struggle to sustain big data technologies such as big data architecture. This is primarily because the creation and management of the architecture require special or specific types of skills. In substituting the specific architecture skill with other types of skills, challenges arise. Lack of the appropriate skills sometimes results in data that cannot be converted into strategic resources, for the operationalisation of the organisational goals and objectives. In one of the studies, the implication of lack of skill is stated as follows:

'Many organisations have not been able to develop and implement architecture primarily because they do not have skilled personnel. What is even more challenging is the availability of the training facilities.' (ENTDoc30; Pg52: Ln15-17)

Activity theory: Rules

Rules refer to control mechanisms, which can include policies, regulations and legislations that guide and govern how subjects perform their activities, in AT (Kelly 2018). Additionally, the rules help to maintain order and control conflicts within an activity. The architecture of big data provides a layer that deals with the governance of the data throughout its lifecycle. The architecture of big data requires standards, laws and regulation controls to collect, use, share, store and disseminate data (Pratsri & Nilsook 2020). Following are extracts from some of the articles:

> 'The governance layer is in charge of applying policies and regulations to the whole data lifecycle, as well as managing the licenses related to the data sets.' (ARCDoc01; Pg108: Ln7-9)

> 'In other words, there are strong rules relating to data standardization and compliance within the infrastructure.' (BDTDoc52; Pg468: Ln40-42)

Data analysts and other specialists who are responsible for the governance and management of big data require certain protocols, processes or regulations for operationalisation, such as gaining access to the data (Yaseen & Obaid 2020). Also, rules are used to manage the interactions between actors during the process of allocating tasks, roles and responsibilities. Wang et al. (2020) suggest that organisational structures such as departments are also used to put a restriction on who accesses the resources.

Activity theory: Community

Community is defined by Iyamu and Shaanika (2019) as a collective of individuals in a social system working towards the same goal. The design of big data architecture involves a group of individuals such as data scientists, architects, analysts, software developers and business users (Chen et al. 2016). Forming teams of developers and data scientists helps to share strategies and ideas that help to increase the effectiveness and efficiency of the data (Kim et al. 2016). These individuals form communities based on their roles and skills which contribute towards achieving the desired goal. In addition, it becomes easy to achieve the goals when there is transparency and clear communication within the teams (Pau et al. 2022). Furthermore, communities are also responsible for making decisions regarding the appropriate big data architecture that aligns with the business goals:

> 'Thus, collaboration across teams and workstreams is critical when designing data architecture to help reveal as many areas for improvement or threats as possible.' (NPARCDoc04; Pg6: Ln26-29)

> 'Companies use big data to achieve value creation in collaboration with stakeholders, which is manifested in strengthening connection and interaction, synergistically improving operational performance, and reducing operating costs through platform integration.' (BDTDoc53; Pg 6: Ln 23-27)

When the organisation does not have teams with the right skills, they outsource the development of the architecture (Shaanika & Iyamu 2018). This helps to avoid operational costs and waste of resources that may arise from an architecture that is not designed properly (Pääkkönen & Pakkala 2020). Hence, the communities need to be guided by the rules when performing activities such as designing the architecture of big data (Dennehy & Conboy 2017). Some of the rules mentioned by Pau et al. (2022) are to integrate a security framework into the big data architecture, to ensure that data is secure and to avoid vendor lock-in as it prevents future integration of services to the architecture.

Activity theory: Division of labour

Division of labour refers to the allocation of tasks and responsibilities among community members (Lioutas et al. 2019). Division of labour helps to promote accountability of the actions taken by the individuals within a community. The design of big data architecture requires engagement from various stakeholders (actors) and an understanding of the functional and non-functional requirements of the big data architecture applications and their environment (Pau et al. 2022). The stakeholders perform distinct tasks that are interconnected to achieve a holistic design of big data architecture. For instance, the data scientist role requires programming and decision-making skills (Uden, Lu & Ting 2017). There were some corroborative views from the data, such as the following:

> 'On the technical side, data architects create data models themselves and supervise modelling work by others.' (NPARCDoc09; Pg15: Ln18-19)

> 'Big Data architectures offer remarkable solutions to complex data issues but do not cover the complete flow of information that is required.' (ARCDoc40; P1:20-21)

The tasks for the specialists are very specific and specialised. The developers are responsible for the integration of software applications and hardware components (Saggi & Jain 2018). Some of the tasks of software engineers are to understand how the business and decision-makers are going to use the data and therefore coordinate the analytical and storage tools needed (Chen et al. 2016). The use of the appropriate tools that align with the organisation's goals helps to deliver value for the business (Uden et al. 2017). Gökalp et al. (2019) state that specific tasks require specific knowledge and experience because of their complexity.

Activity theory: Object

The object component of the AT model is the motive of carrying out an activity, and it can be tangible or intangible (Iyamu 2022b). Big data architecture provides organisations with a competitive advantage, improves performance and generates value (Blazquez & Domenech 2018). A well-defined big data architecture drives innovation and provides useful insights to the organisation (Avci et al. 2020). This includes defining the interactions and relationships between the elements of the architecture (Mohammad et al. 2014). A detailed analysis of the characteristics of the existing data architecture is required to help decide on the new architecture (Kalipe & Behera 2019). Also, it helps to understand the

limitations of the existing data architecture to accommodate the new requirements (Uden et al. 2017):

'The right design of the big data architecture is a vital foundation for building an effective system to be used by the business on an everyday basis.' (ARCDoc38; Pg460: Ln30-32)

The design of big data architecture is a complex exercise that needs to be tailored based on the organisations' needs, drivers and available resources. Selecting an incorrect big data architecture can result in overlapping functionalities that can hinder the success of the organisation (Kalipe & Behera 2019). Iyamu (2023) suggests that for an organisation to successfully implement big data architecture, it needs skilled employees, to exert the power of a dynamic environment.

Findings and interpretation

In gaining a better understanding of the determinants of the big data architecture, which is the objective of the study, the focus is on two main areas that are revealed in the analysis. Firstly, it identifies the factors that influence the activities of big data in its use and management in the organisation. Secondly, it is to understand how the factors manifest themselves, as actors execute tasks. The relationships and interactions between the actors involved in the activities are examined. This is to mitigate risks in the activities involved in storing, accessing and managing big data in an organisation.

From the analysis presented in the preceding section, three factors, interactions, relationship and allocative are fundamental to the architecture of big data. As shown in Figure 2, the factors are interrelated and influence one another in the activities of big data, such as data gathering, retrieval, security, governance and use. The factors were interpreted following the subjective reasoning approach. This was done towards achieving the aim of the study, which is to understand the determinants in designing a big data architecture, purposely to enhance business continuity and improve the efficiency and effectiveness of operations and services in an organisation. The factors manifest into
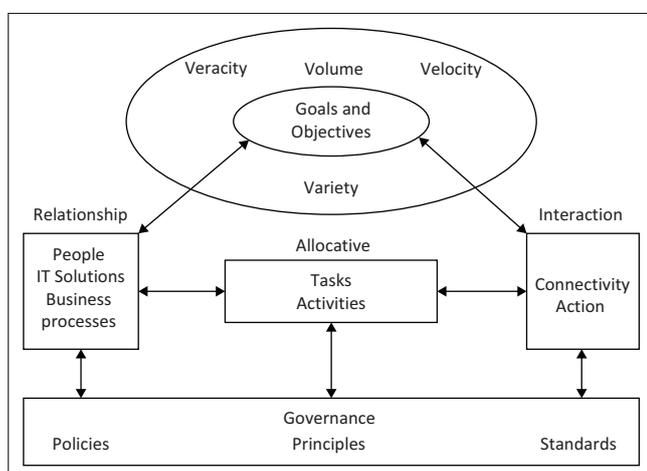
IT, information technology.
FIGURE 2: Big Data architecture.

attributes, as shown in the architecture. The attributes are from both technical and non-technical standpoints. The factors are discussed in the remainder of this section. The discussion should be read in conjunction with Figure 2 to gain a better understanding of the big data architecture.

The study aimed to design a big data architecture for enterprises, purposely to enhance business continuity and improve efficiency and effectiveness of operations and services in the use of big data in an organisation. The analysis revealed factors, which fundamentally influence the design of architecture for big data through their manifestations. The factors are relationships, interactions and allocative, and manifest through governance. It helps to gain a better understanding of the factors that can influence the design of big data architecture in an enterprise. The relationship is between humans such as business personnel, and IT architects and technology.

The relationship between the constituent entities (people, IT solutions and business processes) is guided by governance, which shapes the design of big data architecture in an organisation. Also, relationships influence the activities and interactions between the actors that enrol in defining and designing including the use of big data. Enacted by governance, humans interact with rules and IT solutions such as big data to transform business activities and objectives. During the process of designing big data architecture, the interactions between humans allow them to share requirements, ideas and knowledge, and allocate tasks. The allocation of tasks promotes the alignment of interests between business and IT architects.

The analysis revealed that governance can be used to define the standards, principles and policies within which events and activities are performed when developing and implementing architectures in organisations. The governance of activities helps to maintain uniformity, reduce complexity and enable flexibility in the activities such as how big data are generated, stored, governed and used in the enterprises. This helps to gain an understanding of the influencing factors, to leverage and utilise big data to improve efficiency and performance (Calic & Ghasemaghaei 2021). Ostensibly, the use of big data to enhance business continuity is influenced by theoretical and practical implications. Ghasemaghaei and Calic (2019) explain how business processes and IT solutions' relationship with big data contains significant theoretical and practical implications. We viewed the implications from the business and IT units. Operationalisation, innovation and integration were identified as significant implications towards improving organisational efficiency and performance.

From the operationalisation perspective, organisations need to develop an operational approach to support the architecture of big data. Batyashe and Iyamu (2020) emphasised and explained how operationalisation enables and supports an architectural design. In the innovation component, the IT units need to develop metrics that can be used to measure the value of big data architecture to the organisation. Babu et al. (2021) argued that big data

architecture has implications in defining, designing and implementing innovation in an organisation, to provide meaningful insight into constructiveness, efficiency and effectiveness. The business units need to understand how big data architecture can be used to reduce costs and promote business innovations. Integration ensures the unification of IT solutions and business artefacts, to reduce complexity, increase effectiveness and efficiency, promote seamlessness of processes and enable product interconnectivity. Thus, Wang, Kung and Byrd (2018) suggest that integration is a crucial aspect of big data architectural design.

Conclusion

The analysis revealed the factors that influence the design of big data architecture. The findings were interactions, relationships and allocative factors. The factors constitute technical and non-technical operations and processes involving big data. The architecture of big data was designed based on these factors. The combination of both technical and non-technical factors in the architecture makes it critical for business continuity. Also, theoretical and practical implications (operationalisation, innovation and integration) towards business continuity were identified using subjective reasoning.

The study contributes practically and theoretically. Practically, the designed architecture can be used to guide the development of governance (policies, standards and principles) in an enterprise. Based on the governance, big data can be better stored, its retrieval can be eased, and it can enhance usability and manageability. Better management reduces complexity and improves effectiveness and efficiency in the use of big data for service delivery. Additionally, in practice, the study reveals the factors that influence the design of big data architecture, which can be used to develop a research stream. The architecture enables enterprises to align with their evolving needs and challenges in dealing with the characteristics of big data. Also, the study highlights the architectural components that influence the use and management of big data in an environment. From an academic domain standpoint, the architectural design forms part of the enterprise architecture research stream. The theory was used to navigate existing materials, which makes the study a good methodological contribution. Thus, this contributes to advancing the application of the theory, in IT studies.

However, the big data architecture provided in this study has not been tested, which makes it theoretical. This creates an opportunity for validation through further research studies. Researchers and students focussing on an understanding of the architecture of big data can benefit from the study. Very importantly, it contributes to the body of knowledge in areas such as big data architecture and architectural design where literature is currently limited.

Acknowledgements

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article. The author, T.I., serves as an editorial board member of this journal. The peer review process for this submission was handled independently, and the author had no involvement in the editorial decision-making process for this manuscript. The authors have no other competing interests to declare.

Authors' contributions

W.N. problematised the topic. T.I. led the data collection and analysis processes. W.N. and T.I. both contributed in writing the article.

Ethical considerations

Ethical approval was obtained from the Faculty of Informatics and Design (FID) Ethics Committee, Cape Peninsula University of Technology with ethical clearance number 203168283/2023/24.

Funding information

Data availability

The data that support the findings of this study are available from the corresponding author, T.I. upon reasonable request.

Disclaimer

The views and opinions expressed in this article are those of the authors and are the product of professional research. The article does not necessarily reflect the official policy or position of any affiliated institution, funder, agency or that of the publisher. The authors are responsible for this article's results, findings and content.

References

Al-Sai, Z.A. & Abdullah, R., 2019, 'Big data impacts and challenges: A review', in K.M. Jaber (ed.), *2019 IEEE Jordan international joint conference on electrical engineering and information technology*, Jordan, 09–11 April, 2019, pp. 150–155.

Avci, C., Tekinerdogan, B. & Athanasiadis, I.N., 2020, 'Software architectures for big data: A systematic literature review', *Big Data Analytics* 5(1), 1–53. https://doi.org/10.1186/s41044-020-00045-1

Babu, M.M., Rahman, M., Alam, A. & Dey, B.L., 2021, 'Exploring big data-driven innovation in the manufacturing sector: Evidence from UK firms', *Annals of Operations Research* 333, 1–28.

Bansal, B., Jenipher, V.N., Jain, R., Dilip, R., Kumbhkar, M., Pramanik, S. et al., 2022, 'Big Data architecture for network security', in S. Pramanik, D. Samanta, M. Vinay & A. Guha (eds.), *Cyber security and network security*, pp. 233–267, Crivener Publishing LLC, New Jersey.

Barradas, A., Tejeda-Gil, A. & Cantón-Croda, R.M., 2022, 'Real-time big data architecture for processing cryptocurrency and social media data: A clustering approach based on k-means', *Algorithms* 15(5), 1–11. https://doi.org/10.3390/a15050140

Batyashe, N.R. & Iyamu, T., 2020, 'Operationalisation of the information technology strategy in an organisation', *Journal of Contemporary Management* 17(2), 198–224. https://doi.org/10.35683/jcm20018.71

Benhlima, L., 2018, 'Big data management for healthcare systems: Architecture, requirements, and implementation', *Advances in Bioinformatics* 2018, 1–11. https://doi.org/10.1155/2018/4059018

Blazquez, D. & Domenech, J., 2018, 'Big Data sources and methods for social and economic analyses', *Technological Forecasting and Social Change* 130(2018), 99–113. https://doi.org/10.1016/j.techfore.2017.07.027

Boumlik, A. & Bahaj, M., 2018, 'Big data and IoT: A prime opportunity for the banking industry', in M. Ezziyyani, M. Bahaj & F. Khoukhi (eds.), *Advanced information technology, services and systems: Proceedings of the International Conference on advanced information technology, services and systems*, 14–15 April, Tangier, Morocco, Springer International Publishing, pp. 396–407.

Calic, G. & Ghasemaghaei, M., 2021, 'Big data for social benefits: Innovation as a mediator of the relationship between big data and corporate social performance', *Journal of Business Research* 131, 391–401. https://doi.org/10.1016/j.jbusres.2020.11.003

Chen, H.M., Kazman, R. & Haziyev, S., 2016, 'Agile big data analytics for web-based systems: An architecture-centric approach', *IEEE Transactions on Big Data* 2(3), 234–248. https://doi.org/10.1109/TBDATA.2016.2564982

Dennehy, D. & Conboy, K., 2017, 'Going with the flow: An activity theory analysis of flow techniques in software development', *Journal of Systems and Software* 133, 160–173. https://doi.org/10.1016/j.jss.2016.10.003

Engeström, Y., Lompscher, J. & Rückriem, G. (eds.), 2016, *Putting activity theory to work: Contributions from developmental work research*, 13th edn., Lehmanns Media, Berlin.

Er, M., Kay, R. & Lawrence, E., 2010, 'Information systems and activity theory: A case study of doctors and mobile knowledge work', in S. Latifi (ed.), *2010 Seventh International Conference on Information Technology: New generations*, 12–14 April, 2010, Las Vegas, Nevada, pp. 603–607.

Farooqi, M.M., Shah, M.A., Wahid, A., Akhunzada, A., Khan, F., ul Amin, N. et al., 2019, 'Big data in healthcare: A survey', in F. Khan (ed.), *Applications of intelligent technologies in healthcare*, pp. 143–152, Springer, Switzerland AG.

Ferraris, A., Mazzoleni, A., Devalle, A. & Couturier, J., 2019, 'Big data analytics capabilities and knowledge management: Impact on firm performance', *Management Decision* 57(8), 1923–1936. https://doi.org/10.1108/MD-07-2018-0825

Filaly, Y., Berros, N., Badri, H., Mendil, F.E. & EL Idrissi, Y.E.B., 2023, 'Security of Hadoop Framework in Big Data', in Y. Farhaoui, A. Rocha, Z. Brahmia & B. Bhushab (eds.), *Artificial Intelligence and smart environment*, pp. 709–715, Springer International Publishing, Cham.

Fischer, C., Pardos, Z.A., Baker, R.S., Williams, J.J., Smyth, P., Yu, R. et al., 2020, 'Mining big data in education: Affordances and challenges', *Review of Research in Education* 44(1), 130–160. https://doi.org/10.3102/0091732X20903304

Garoufallou, E. & Gaitanou, P., 2021, 'Big data: Opportunities and challenges in libraries, a systematic literature review', *College & Research Libraries* 82(3), 410–435. https://doi.org/10.5860/crl.82.3.410

Ghasemaghaei, M. & Calic, G., 2019, 'Does big data enhance firm innovation competency? The mediating role of data-driven insights', *Journal of Business Research* 104, 69–84. https://doi.org/10.1016/j.jbusres.2019.07.006

Glass, R., Ramesh, V. & Vessey, I., 2004, 'An analysis of research in computing disciplines', *Communications of the ACM* 47(6), 89–94. https://doi.org/10.1145/990680.990686

Gökalp, M.O., Kayabay, K., Zaki, M., Koçyiğit, A., Eren, P.E. & Neely, A., 2019, 'Open source big data analytics architecture for businesses', in A. Varol, A. Yazici, C. Varol, G.S. Aygüneş & A. Çotur (eds.), *2019 1st International Informatics and Software Engineering Conference*, 6–7 November, 2019, Ankara, Turkey, pp. 1–6.

Hajli, N., Shirazi, F., Tajvidi, M. & Huda, N., 2021, 'Towards an understanding of privacy management architecture in big data: An experimental research', *British Journal of Management* 32(2), 548–565. https://doi.org/10.1111/1467-8551.12427

Hasan, H. & Kazlauskas, A., 2014, 'Activity theory: Who is doing what, why and how', in H. Hasan (ed.), *Being practical with theory: A window into business research*, pp. 9–14, THEORI, Wollongong.

Iyamu, T., 2013, 'Institutionalisation of the enterprise architecture: The actor-network perspective', in A. Tatnall (ed.), *Social and professional applications of actor-network theory for technology*, pp. 144–155, IGI Global, Hershey.

Iyamu, T., 2020, 'A framework for selecting analytics tools to improve healthcare big data usefulness in developing countries', *South African Journal of Information Management* 22(1), 1–9. https://doi.org/10.4102/sajim.v22i1.1117

Iyamu, T., 2022a, *Enterprise architecture for strategic management of modern IT solutions*, 1st edn., CRC Press, Boca Raton.

Iyamu, T., 2022b, *Applying theories for information systems research*, 1st edn., Routledge, New York, NY.

Iyamu, T., 2023, *Advancing Big Data analytics for healthcare service delivery*, Taylor & Francis, London.

Iyamu, T. & Shaanika, I., 2019, 'The use of activity theory to guide information systems research', *Education and Information Technologies* 24, 165–180. https://doi.org/10.1007/s10639-018-9764-9

Iyamu, T., Nehemia-Maletzky, M. & Shaanika, I., 2016, 'The overlapping nature of business analysis and business architecture: What we need to know', *Electronic Journal of Information Systems Evaluation* 19(3), 169–179.

Jeske, D. & Calvard, T., 2020, 'Big data: Lessons for employers and employees', *Employee Relations: The International Journal* 42(1), 248–261. https://doi.org/10.1108/ER-06-2018-0159

Jones, M., 2019, 'What we talk about when we talk about (big) data', *The Journal of Strategic Information Systems* 28(1), 3–16. https://doi.org/10.1016/j.jsis.2018.10.005

Kalipe, G.K. & Behera, R.K., 2019, 'Big Data architectures: A detailed and application oriented review', *International Journal of Innovative Technology and Exploring Engineering* 8(9), 2182–2190. https://doi.org/10.35940/ijitee.H7179.078919

Karanasios, S. & Allen, D., 2013, 'ICT for development in the context of the closure of Chernobyl nuclear power plant: An activity theory perspective', *Information Systems Journal* 23(4), 287–306. https://doi.org/10.1111/isj.12011

Kelly, P.R., 2018, 'An activity theory study of data, knowledge, and power in the design of an international development NGO impact evaluation', *Information Systems Journal* 28(3), 465–488. https://doi.org/10.1111/isj.12187

Kim, M., Zimmermann, T., DeLine, R. & Begel, A., 2016, 'The emerging role of data scientists on software development teams', in L. Dillon, W. Visser & L. Williams (eds.), *Proceedings of the 38th International Conference on Software Engineering*, 14–22 May, 2016, Austin, Texas, pp. 96–107.

Kothari, C.R., 2020, *Research methodology methods and techniques*, 2nd edn., New Age International (P) Ltd., Publishers, New Delhi.

Lioutas, E.D., Charatsari, C., La Rocca, G. & De Rosa, M., 2019, 'Key questions on the use of big data in farming: An activity theory approach', *NJAS-Wageningen Journal of Life Sciences* 90(1), 1–12. https://doi.org/10.1016/j.njas.2019.04.003

Manogaran, G., Thota, C. & Lopez, D., 2022, 'Human-computer interaction with big data analytics', in M. Khosrow-Pour, S. Clarke, M.E. Jennex & A. Anttiroik (eds.), *Research anthology on Big Data analytics, architectures, and applications*, pp. 1578–1596. IGI Global, Hershey.

Mohammad, A., Mcheick, H & Grant, E., 2014, 'Big data architecture evolution: 2014 and beyond', in M.S.M.A. Notare (ed.), *Proceedings of the fourth ACM International Symposium on Development and Analysis of intelligent vehicular networks and Applications*, 21–26 September, 2014, Montreal, Canada, pp. 139–144.

Morgan, H., 2022, 'Conducting a qualitative document analysis', *The Qualitative Report* 27(1), 64–77. https://doi.org/10.46743/2160-3715/2022.5044

Mostefaoui, A., Merzoug, M.A., Haroun, A., Nassar, A. & Dessables, F., 2022, 'Big data architecture for connected vehicles: Feedback and application examples from an automotive group', *Future Generation Computer Systems* 134, 374–387. https://doi.org/10.1016/j.future.2022.04.020

Nehemia-Maletzky, M., Iyamu, T. & Shaanika, I., 2018, 'The use of activity theory and actor network theory as lenses to underpin information systems studies', *Journal of Systems and Information Technology* 20(2), 191–206. https://doi.org/10.1108/JSIT-10-2017-0098

Nyikana, W. & Iyamu, T., 2022, 'A Guide for selecting big data analytics tools in an organisation', in T. Bui (ed.), *Proceedings of the 55th Hawaii International Conference on System Sciences*, 04–07 January, 2022, Maui, Hawaii, pp. 5451–5461.

Nyikana, W. & Iyamu, T., 2023, 'The taxonomical distinction between the concepts of small data and big data', in M.B. Nunes, P. Isaías & P. Powell (eds.), *Proceedings of the 16th International Association for Development of Information Society*, 11–13 March, 2023, Lisbon, pp. 138–146.

Oussous, A., Benjelloun, F.Z., Lahcen, A.A. & Belfkih, S., 2018, 'Big Data technologies: A survey', *Journal of King Saud University-Computer and Information Sciences* 30(4), 431–448. https://doi.org/10.1016/j.jksuci.2017.06.001

Pääkkönen, P. & Pakkala, D., 2020, 'Extending reference architecture of big data systems towards machine learning in edge computing environments', *Journal of Big Data* 7(1), 1–29. https://doi.org/10.1186/s40537-020-00303-y

Park, S., Cho, Y., Yoon, S.W. & Han, H., 2013, 'Comparing team learning approaches through the lens of activity theory', *European Journal of Training and Development* 37(9), 788–810. https://doi.org/10.1108/EJTD-04-2013-0048

Patel, M. & Patel, N., 2019, 'Exploring research methodology', *International Journal of Research and Review* 6(3), 48–55. https://doi.org/10.4324/9781351235105-3

Pau, M., Kapsalis, P., Pan, Z., Korbakis, G., Pellegrino, D. & Monti, A., 2022, 'MATRYCS – A Big Data architecture for advanced services in the building Domain', *Energies* 15(7), 2568. https://doi.org/10.3390/en15072568

Pramanik, P.K.D., Pal, S. & Mukhopadhyay, M., 2022, 'Healthcare big data: A comprehensive overview', in M. Khosrow-Pour, S. Clarke, M.E. Jennex & A. Anttiroik (eds.), *Research anthology on big data analytics, architectures, and applications*, pp. 119–147, IGI Global, Hershey.

Prasetyo, B., Aziz, F.S., Faqih, K., Primadi, W., Herdianto, R. & Febriantoro, W., 2019, 'A review: Evolution of big data in developing country', *Bulletin of Social Informatics Theory and Application* 3(1), 30–37. https://doi.org/10.31763/businta.v3i1.162

Pratsri, S. & Nilsook, P., 2020, 'Design on Big Data platform-based in Higher Education Institute', *Higher Education Studies* 10(4), 36–43. https://doi.org/10.5539/hes.v10n4p36

Ruiz, M.D., Gómez-Romero, J., Fernandez-Basso, C. & Martin-Bautista, M.J., 2021, 'Big data architecture for building energy management systems', *IEEE Transactions on Industrial Informatics* 18(9), 5738–5747. https://doi.org/10.1109/TII.2021.3130052

Saggi, M.K. & Jain, S., 2018, 'A survey towards an integration of big data analytics to big insights for value-creation', *Information Processing & Management* 54(5), 758–790. https://doi.org/10.1016/j.ipm.2018.01.010

Sannino, A. & Engeström, Y., 2018, 'Cultural-historical activity theory: Founding insights and new challenges', *Cultural-Historical Psychology* 14(3), 43–56. https://doi.org/10.17759/chp.2018140304

Shaanika, I. & Iyamu, T., 2015, 'Deployment of enterprise architecture in the Namibian government: The use of activity theory to examine the influencing factors', *The Electronic Journal of Information Systems in Developing Countries* 71(1), 1–21. https://doi.org/10.1002/j.1681-4835.2015.tb00515.x

Shaanika, I. & Iyamu, T., 2018, 'Developing the enterprise architecture for the Namibian government', *The Electronic Journal of Information Systems in Developing Countries* 84(3), 1–11. https://doi.org/10.1002/isd2.12028

Tschoppe, N. & Drews, P., 2022, 'Developing digitalization strategies for SMEs: A lightweight architecture-based method', in T. Bui (ed.), *Proceedings of the 55th Hawaii International Conference on System Sciences*, 04–07 January, 2022, Maui, Hawaii, pp. 1–10.

Uden, L., Lu, W. & Ting, I.H., 2017, 'Knowledge management in organizations', in L. Uden, P.W. Lu & I. Ting (eds.), *12th International conference in knowledge management in organisations*, 21–24 August, 2017, Beijing, pp. 460–469.

Wang, J., Yang, Y., Wang, T., Sherratt, R.S. & Zhang, J., 2020, 'Big data service architecture: A survey', *Journal of Internet Technology* 21(2), 393–405.

Wang, Y., Kung, L. & Byrd, T.A., 2018, 'Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations', *Technological Forecasting and Social Change* 126, 3–13. https://doi.org/10.1016/j.techfore.2015.12.019

Wehn, C., Yang, J., Gan, L. & Pan, Y., 2021, 'Big data-driven Internet of Things for credit evaluation and early warning in finance', *Future Generation Computer Systems* 124, 295–307. https://doi.org/10.1016/j.future.2021.06.003

Yaseen, H.K. & Obaid, A.M., 2020, 'Big data: Definition, architecture & applications', *JOIV: International Journal on Informatics Visualization* 4(1), 45–51. https://doi.org/10.30630/joiv.4.1.292

Zhang, Z., Shang, Y., Cheng, L. & Hu, A., 2022, 'Big data capability and sustainable competitive advantage: The mediating role of ambidextrous innovation strategy', *Sustainability* 14(14), 1–17. https://doi.org/10.3390/su14148249