American Journal of Information Systems, 2024, Vol. 9, No. 1, 1-10
Available online at http://pubs.sciepub.com/ajis/9/1/1
Published by Science and Education Publishing
DOI:10.12691/ajis-9-1-1

Use of Conventional Business Intelligence (BI) Systems as the Future of Big Data Analysis

Molla Ehsanul Majid[1,*], Dora Marinova[2], Amzad Hossain[3], Muhammad E. H. Chowdhury[3], Farah Rummani[3]

[1]Computer Applications Department, Academic Bridge Program, Qatar Foundation, Doha, Qatar
[2]Department of Electrical Engineering, Qatar University, Doha, Qatar
[3]Environmental and Public Health Promotion, Shire of Serpentine-Jarrahdale, Western Australia
*Corresponding author: mollamajid@gmail.com

Received May 26, 2024; Revised June 28, 2024; Accepted July 04, 2024

Abstract Traditional Business Intelligence (BI) systems employ a combination of source systems, databases, data repositories, data warehouses, and analytical tools to gain insights into business operations and chart future organizational strategies. These BI systems typically rely on structured data extracted from the underlying source system databases. However, organizations are increasingly harnessing vast amounts of big data from diverse sources, which often include semi-structured and unstructured data. The BI systems currently in use were initially designed with structured organizational datasets in mind. As the volume of big data required for informed decision-making continues to grow, a pressing question arises: can the existing BI systems effectively analyse this diverse and expansive data landscape? This research seeks to assess the adaptability of current BI systems to analyse big data and presents potential strategies for addressing this evolving data landscape.

Keywords: *Conventional Business Intelligence, data analysis, structured data, big data*

Cite This Article: Molla Ehsanul Majid, Dora Marinova, Amzad Hossain, Muhammad E. H. Chowdhury, and Farah Rummani, "Use of Conventional Business Intelligence (BI) Systems as the Future of Big Data Analysis." *American Journal of Information Systems*, vol. 9, no. 1 (2024): 1-10. doi: 10.12691/ajis-9-1-1.

1. Introduction

A Business Intelligence (BI) system, typically comprising a suite of software tools, serves to seamlessly amalgamate data from various departments, locations, and sources within an organization, offering real-time visualizations of its performance [1,2]. Moreover, it furnishes the company with both short-term and long-term strategic insights. The adoption of a BI system can bestow a substantial competitive edge by expediting and enhancing decision-making processes, ultimately bolstering overall organizational performance [2,3]. Nonetheless, it's worth noting that the current implementation of BI systems within corporations and organizations often proceeds with limited regard for their capacity to effectively analyse big data. Big data, characterized by their complexity, often encompass semi-structured and unstructured elements, and are generated in vast quantities daily [2,4]. This extensive data landscape includes information stemming from system logs, emails, text messages, voice calls, video clips, as well as copious volumes of online data and data from IoT devices, among others. Therefore, this research endeavours to address a pivotal query: "Do conventional BI systems possess the capabilities to proficiently analyse the burgeoning expanse of big data?" The primary objective of this research is to scrutinize the aptitude of contemporary BI systems in the realm of big data analysis, with a specific focus on their handling of semi-structured and unstructured data components. To accomplish this goal and respond to the research question, this paper embarks on a comprehensive exploration. It commences by delving into the architecture of BI systems, while concurrently shedding light on the intricacies of big data. This is succeeded by a quantitative investigation designed to unveil the performance of current BI systems in their analysis of the expanding realm of big data, along with any associated limitations.

2. Literature Review

In this section, we will examine the framework of a typical BI system and explore the various categories of datasets that constitute the realm of 'big data'. Additionally, we will delve into any existing provisions within BI systems for the inclusion and management of these diverse datasets.

2.1. Business Intelligence (BI) System

A traditional Business Intelligence (BI) system typically comprises a suite of tools that operates in conjunction with a data warehouse. The data warehouse, in turn, serves as the repository for data extracted from various source system databases, including accounting, finance, HR, supply chain, and more, ultimately shaping the output of the BI system [1,5]. Therefore, to comprehend the inner workings of a BI system, it is imperative to grasp the architecture of the data warehouse.

A data warehouse fulfills the crucial role of storing an organization's historical data in an aggregated form. In contrast, relational database systems are employed to store and analyze real-time business transactions. These systems are designed to manage atomic data and provide concurrent user access for query retrieval. Consequently, they are often referred to as online transaction processing (OLTP) systems [6] Conversely, a data warehouse represents a specialized data repository primarily geared toward the storage of summarized and historical data for analytical purposes. It is not intended for day-to-day transactions or daily-level querying and is commonly categorized as an online analytical processing (OLAP) system. OLTP systems are optimized for quick, small-scale transactions with short response times, while OLAP systems handle intricate queries that yield extensive records, resulting in longer response times [7].

OLTP systems allow various online operations, such as reading, updating, inserting, and deleting records, often in real-time. On the other hand, OLAP systems are typically geared towards read and append-only operations, with offline and periodic updating options [7]. Moreover, the design of OLTP systems frequently employs an Entity-Relationship (E-R) modelling approach, facilitating the creation of highly normalized schemas. In contrast, OLAP systems are designed using a schema known as multidimensional modelling. In the context of multidimensional modelling, data is represented in the form of a cube, often referred to as a data cube or hypercube [8,9]. For instance, Figure 1-a, presented below, illustrates a data cube depicting information sourced from a database with entities encompassing customers, products, orders, employees, suppliers, and regions.

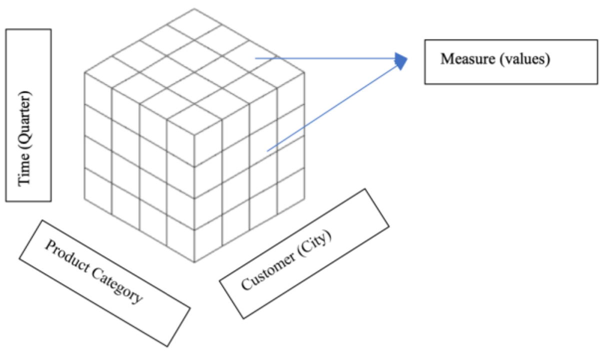

Figure 1. -a: A 3-dimensional data cube with sales data by product, time, and customer Region

A data cube serves as a representation of aggregated data, using dimensions and facts as its building blocks. To illustrate, in the context of a sales data cube, dimensions might encompass Product, Customer, and Time. It's important to note that each dimension encompasses a hierarchy, as depicted in Figure 2-A. Within this hierarchy, a higher-level structure is referred to as the parent, while a lower level is termed the child. As an example, within the Product dimension, the highest-level category might be labelled as "All Products." Subsequently, "All Products" can be further broken down into subcategories like Beverages, Seafood, Condiments, and more, and the Beverages category, in turn, can be subcategorized into items like Chai, Chang, Coke, and so forth. Each of these subcategories is referred to as a member of the hierarchy.

The purpose of a data cube is to exhibit a collection of these members within the dimensions. To clarify, this means that City could be considered a member within the Customers dimension, Quarter within the Time dimension, and Categories within the Product dimension. Conversely, facts within the context of a data cube are like small, specific slices of the larger cube, serving to display various measures. As an example, a measure like "Quantity" provides information about the number of products sold within a specific quarter, category, and the customer's city [8,9].

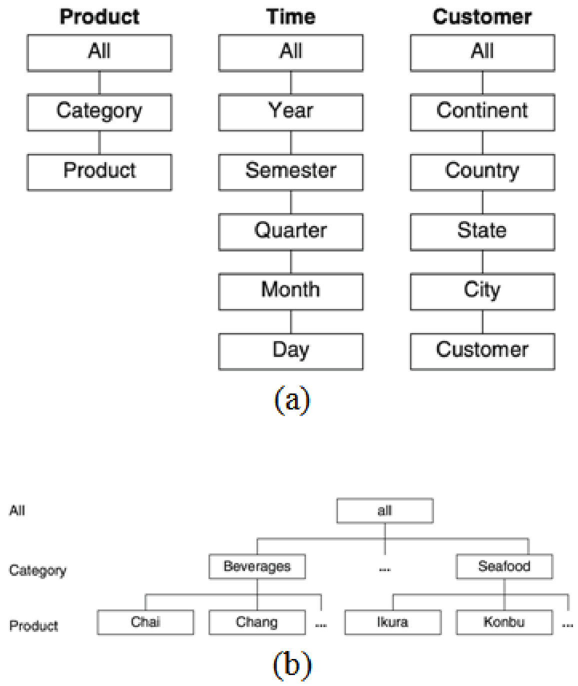

Figure 2. Hierarchy of dimensions (a) and Members of hierarchy (b). Adapted from [9]

Measures within a data cube can be categorized as additive, semi-additive, or non-additive, as outlined by Horner et al. [10]. Additive measures, the most common type, permit data summarization along all dimensions. For instance, in Figure 1-A, Sales Quantity serves as an additive measure, allowing summation across all hierarchical movements within the Product, Time, and Customer dimensions. Semi-additive measures, on the other hand, enable data summarization along some dimensions. For example, Inventory Quantity is considered a semi-additive measure since adding it along the Time dimension (e.g., total inventory in Q1+Q2) would lack meaningful context. Instead, it can be averaged along the Time dimension (e.g., average inventory quantity in Q1+Q2+Q3, etc.) and added along the Product and Customer dimensions. Non-additive measures, like unit cost or item price, do not support summation along any dimension. In the case of semi-additive or non-additive measures, other functions such as count, maximum, minimum, median, etc., may be more appropriate [10].

The cube depicted in Figure 1-A can be employed to analyse different hierarchy levels, such as Chai (Product) -

Year (Time) - Country (Customer), or Seafood (Product) - Month (Time) – State (Customer), through OLAP operations. These OLAP operations encompass a range of actions, including drill-down, roll-up, sort, pivot, slice, dice, and drill-across, allowing for comprehensive analysis across all hierarchy levels.

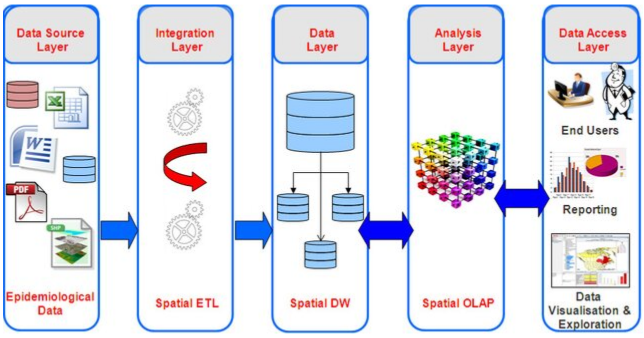

Figure 3. Data Warehousing architecture (Source: [23])

In a typical data warehouse architecture (Figure 3), there are several tiers. The back-end tier includes a data staging area, known as the operational data store, where data from various sources and databases is staged for ETL (extract, transform, load) operations [11]. The extraction operation gathers data from various sources, the transformation process converts data into standard warehouse formats, and loading loads the transformed data into the warehouse. The data warehousing tier encompasses the data warehouse itself, along with a metadata repository and data marts. While the data warehouse contains summarized transactional data from the ETL process, the metadata repository stores both business and technical metadata. Business metadata includes information on business policies, IT policies, and rules, while technical metadata includes details regarding data warehousing, data storage in computer systems, and data retrieval methods. Data marts, essentially smaller-scale data warehouses, house information related to individual departments or functional areas. They can pull data from the data warehouse or directly from data sources [9,11]. The OLAP tier facilitates the presentation of warehouse data to users in a multidimensional format, as explained in previous sections.

The representation of multidimensional data warehouses in the relational model encompasses several well-known schemas: the star schema, snowflake schema, starflake schema, and constellation schema [12]. In a star schema, there is a central fact table connected directly to multiple dimension tables. The primary keys of these dimension tables collectively form the key for the fact table, ensuring referential integrity. However, data redundancy can be a concern in star schemas as the dimensions are not normalized [12].

To address the redundancy issue, the snowflake schema seeks to normalize the dimensions by splitting them into separate dimension tables, such as product category, department, city, and state. These tables are then linked to their parent tables through referential integrity constraints. A normalized schema not only mitigates redundancy but also facilitates additional queries on the data cube, enabling operations like 'total sales by city' or 'total sales by category.' In both cases, the fact tables provide aggregated information on total sales by product (category, department), promotion, store (city, state), and time (month, quarter, year) dimensions.

A star flake schema combines some normalized dimension tables with denormalized ones, offering a compromise between the two approaches. In contrast, a constellation schema employs multiple fact tables, and dimension tables may be shared among these fact tables. This schema includes both normalized and denormalized tables, which allows for more extensive OLAP operations but can introduce data redundancy issues due to denormalized dimensions [12]. Relational data warehouse schemas are generally derived from the ER model of the underlying database system [9]. The process typically begins with a conceptual schema, which is then transformed into a logical schema before generating the physical schema model. Once the physical data warehousing model is created, data is extracted from various source systems, including databases, flat files, and other sources, using a set of processes collectively known as extraction, transformation, and loading (ETL) [13,14]. ETL involves three key steps: data extraction from source systems to a staging area, data cleansing and transformation, and placement into an operational data store (ODS), from which it is mapped and loaded into the data warehouse. It's worth noting that the specific practices for these ETL operations can vary depending on the platforms used [14]. For a visual representation of the ETL process within the data warehousing example presented in Figure 3, please refer to Figure 4.

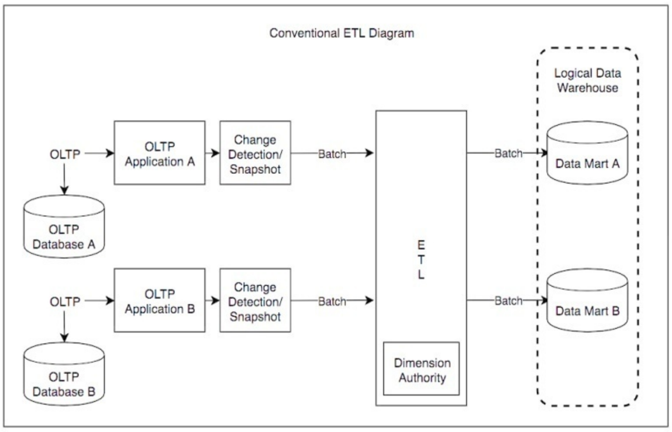

Figure 4. Conceptual view of ETL process (Source: [14,17])

2.2. Big Data

In addition to transactional systems, organizations generate data from a diverse array of systems and sources, with much of this data being semi-structured or unstructured in nature. Sources of such data encompass radio-frequency identification (RFID) readers, software and system logs, voice recordings, sensors, text messages, internet search logs, computer-operated machine logs, and mobile phones [13,15,16]. Furthermore, the proliferation of the Internet of Things (IoT), a network of interconnected devices and systems, is progressively contributing to the burgeoning volume of data due to the growing reliance on this global network [18]. The exponential growth in data from these diverse sources has reached a point where traditional information systems are ill-equipped to store and analyse them. Moreover, the nature and sheer variety of these data types transcend the capacity of existing information systems to manage them within conventional rows and columns [19]. The complexity of utilizing this data demands intelligent and expert systems, setting them apart from traditional organizational data and designating them as "big data" [15] [20]. The scientific discipline focused on acquiring and effectively utilizing big data is known as data science. International Data Corporation (IDC) predicts that the volume of big data may exceed 150 zettabytes by 2025 [21]. The analysis of such vast data necessitates significant computational power with parallel processing, which involves running computer programs simultaneously on multiple computers to meet the extensive computational requirements.

Now, let's delve into the technologies that underpin the storage and analysis of big data and understand their workings. The big data landscape boasts several key players, with Hadoop being one of the most prominent. Originally conceived as the Nutch project by Doug Cutting and Mike Cafarella, Hadoop is a distributed data storage and processing infrastructure project. It was later released as an open-source project in 2008 by Yahoo after Cutting joined the company and brought the technology with him. Presently, Hadoop is managed by the Apache Software Foundation, a nonprofit global community of software developers. Hadoop excels at storing vast volumes of data from diverse sources such as social media, sensors, data logs, and the Internet of Things (IoT). Thanks to its distributed architecture, wherein multiple computers can function as nodes working in parallel and contributing their combined computing power, Hadoop enables rapid data processing. Even if one of these nodes experiences a failure, the system automatically reroutes the work to other nodes, ensuring uninterrupted data processing and making the system fault tolerant. In contrast to relational databases or data warehouses, Hadoop allows data to be stored in a data lake, a vast reservoir of data in its original, unprocessed form. Although data lakes are a subject of security concern for IT and many organizations, ongoing efforts are aimed at securing them, as they are crucial for storing big data [22]. The low cost, owing to its open-source nature, and scalability have solidified Hadoop's position as a major platform in the realm of big data. An illustrative application of Hadoop is in web-based recommendation systems, where it swiftly analyses large volumes of data in the background to provide customer recommendations based on their purchasing behaviour and demographics [22].

3. Research Methodology

This research employs a quantitative approach to address the research question. Data collection was carried out through a survey that involved the participation of 29 respondents hailing from 9 distinct organizations. It was ensured that these organizations were of medium to large size, each employing a minimum of 100 personnel, and utilized various Business Intelligence (BI) systems for analysing their organizational data. This specific size range was chosen due to the expectation that organizations of this magnitude would possess practical experience in handling larger volumes of data. Respondents who actively worked with BI systems, holding roles such as BI analysts, consultants, or managers, were selected for data collection.

The selection of these organizations was facilitated through internet search engines, and contact was established using the provided contact information. All the chosen organizations were based in the Riyadh region of Saudi Arabia. This location was preferred to ensure that the investigator could personally meet with the participants, making data collection a more convenient and effective process. The organizations represented various sectors, including manufacturing (4 organizations), service (3 organizations), and finance (2 organizations). The following figure displays a flow diagram of the processes adopted in data collection and analysis.

The participants were presented with a series of questions related to their organization's type and size. They were inquired about the use of any Business Intelligence (BI) system for data analysis within their organizations. If they confirmed the utilization of BI systems, they were further queried about the specific type of BI systems they employed, distinguishing between isolated BI tools and integrated BI systems. Additionally, participants were asked to characterize the type of data their BI systems processed, whether it was structured, semi-structured, or unstructured, and to specify the volume of data they typically handled. Satisfaction levels regarding the capability of their BI systems to manage various data types were also gauged. In cases where respondents expressed dissatisfaction, they were encouraged to provide insights into the reasons for their discontent.

The questionnaire used for data collection is included in Appendix A of this paper. Subsequently, the collected responses were transferred to a spreadsheet for further analysis.

4. Results and Analysis

The survey responses and results are presented in Figure 5. The header row corresponds to the survey questionnaire items from 1 to 7 (Q1 to Q7). Q1 specifies the types of organizations, as per question 1. The responses to question 2 (Q2) have been categorized into three levels, represented by 1 to 3, signifying small, medium, and large organizations with employee sizes of 100-500, 500-1000, and over 1000, respectively.

Question 3 (Q3) addresses the utilization of BI systems within the organizations for data analysis. The responses to question 4 (Q4) have been quantified as 1 to 3 to denote the type of BI systems used within the corresponding organizations: 1 for integrated BI systems, 2 for multiple isolated BI tools, and 3 for both integrated and isolated BI solutions.

Question 5 (Q5) characterizes the data types employed by the BI systems in these organizations and is represented as types a to d: a) Mostly structured with some semi and/or unstructured, b) Mostly structured with moderate volumes of semi-structured and unstructured, c) Structured with a large volume of semi-structured and unstructured, d) Mostly semi-structured and unstructured.

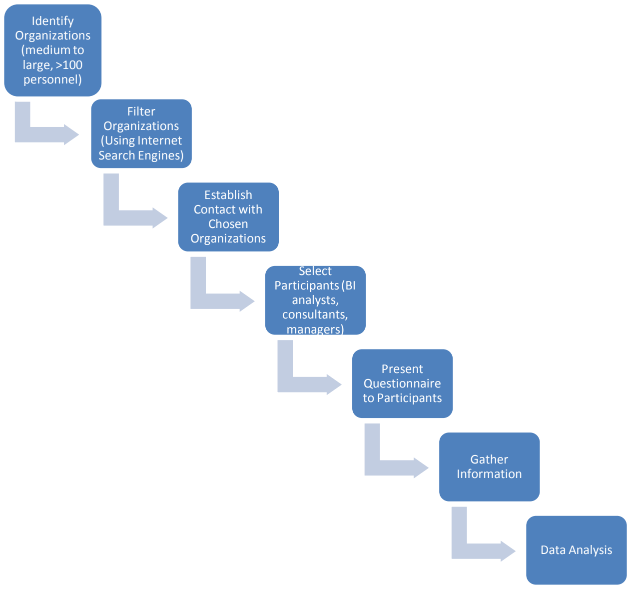

Figure 5. Survey and Data Collection method

Responses to question 6 (Q6) are quantified on a scale of 1 to 5, with 1 indicating "not at all satisfied" and 5 representing "extremely satisfied." This scale is used to gauge the overall satisfaction levels of BI users regarding the speed and output of data processing within their respective organizations.

Finally, answers to question 7 (Q7), which pertain to the reasons for user dissatisfaction with conventional BI systems, are represented through options a to e: a) BI system has no major issues, b) Facing difficulty in handling large data volumes, c) Having difficulty in managing semi-structured and unstructured data, d) Struggling with the modelling and mapping of semi-structured and unstructured data, e) Other reasons. While some of these quantifications were used for data analysis (e.g., Q2 and Q6), others served as a means of data entry (e.g., Q4, Q5, and Q7).

The survey results reveal that 41% of the respondents are affiliated with manufacturing organizations, while the service sector and financial organizations account for 34% and 24% of the respondents, respectively. Although these three categories are not exactly equal in proportion, their representation does not significantly vary, indicating a reasonable mix of sectors.

In terms of organization size, 7 organizations fall into the small category (with 100-500 employees), 13 are categorized as medium-sized (with 500-1000 employees), and the remaining 9 belong to the large category (with over 1000 employees). The median value of 2.07 reflects the average medium size of the chosen organizations.

Regarding the use of BI systems for organizational data analysis, 90% of the respondents mentioned that their organization uses a BI system, with 10% responding partially. This high percentage is expected, given that the organizations were purposefully chosen based on their utilization of a BI system. For the same reason, all the respondents indicated that their organizations use either an integrated BI system (48%), multiple isolated BI tools (6%), or both (24%).

In terms of data types employed by the BI systems, 71% of small organizations predominantly use structured data with some semi and unstructured data, while the remaining 29% use structured data with a moderate volume of semi and unstructured data. Among medium-sized organizations, 23% primarily use structured data with some semi and unstructured data, 69% use structured data with a moderate volume of semi and unstructured data, and the remaining 8% use structured data alongside a large volume of semi and unstructured data. Among large organizations, 22% use structured data with a moderate volume of semi and unstructured data, while 78% use structured data with a large volume of semi and unstructured data.

Although all respondents mentioned the use of BI systems in organizational data analysis, their satisfaction with processing speed and output varies significantly. As indicated in Figure 6 (Q6), the average satisfaction level among all respondents is 2.24 out of 5. This satisfaction level is 3.43 among respondents working in small organizations, 2.38 in mid-sized organizations, and 1.11 in large organizations. Figure 7 illustrates the relationship between organization size and satisfaction regarding the data processing performance of BI systems, showing a decline in satisfaction as organization size and data volume increase.

In Figure 7, the blue line illustrates the organization's size in terms of employee count, while the orange line represents user satisfaction with the performance of their BI analytics. Notably, in small organizations (ranging from 1 to 7 on the category axis), users tend to express higher satisfaction with their BI systems. In contrast, in medium-sized organizations (ranging from 8 to 20), users report a lower satisfaction level. Large organizations (ranging from 21 to 29) exhibit the lowest satisfaction levels in terms of BI analytical performance.

Figure 8 illustrates the data types utilized by different organizations. It is evident that the volume of semi-structured and unstructured data types increases as the organization's size grows.

Regarding the reasons for dissatisfaction with the analytical performance of BI systems, as displayed in Figure 9, option 'a' was selected by 17% of the organizations, while options 'b', 'c', 'd', and 'e' were chosen by 21%, 31%, 31%, and 0%, respectively. Notably, all those who selected reason 'a' belong to small organizations (100%).

In the case of reason 'b', 83% are from medium-sized organizations, and 17% are from large organizations. For reason 'c', 11% are from small organizations, 33% are from medium-sized organizations, and 56% are from large organizations. When selecting reason 'd', only 11% are from small organizations, whereas 56% are from medium-sized organizations, and 33% are from large organizations. In summary, 71% of the small organizations chose option 'a,' 38% of the medium-sized organizations selected reason 'b,' and the remaining 42% chose either 'c' or 'd.' Meanwhile, 89% of the large organizations opted for either 'c' or 'd.'

Finally, when comparing data types with reasons for dissatisfaction, the results indicate that 63% of the organizations using data type 'a' (mostly structured with some semi and unstructured) opted for reason 'a' (the BI system does not have any major issues), and the remaining 37% chose reason 'b' (facing difficulty with large data volume). Among organizations using data type 'b' (structured with moderate volume of semi and unstructured), 23% selected reason 'b,' while the remaining 77% opted for reasons 'c' or 'd' (difficulty with handling, modeling, and mapping of semi- and unstructured data). Finally, 100% of the organizations using data type 'c' (structured with a large volume of semi and unstructured) chose either reason 'c' or 'd.'

Q1 (Org type)	Q2 (Org size)	Q3 (Use of BI System)	Q4 (Type of BI system)	Q5 (data type)	Q6 (Satisfaction level)	Q7 (Resaon for dissatisfaction)
Finance	1	yes	3	a	3	a
Finance	2	yes	3	a	3	b
Finance	2	yes	2	b	2	c
Finance	2	yes	2	b	2	d
Finance	2	yes	3	b	3	d
Finance	2	yes	2	c	2	c
Finance	3	yes	1	c	1	c
Manufacturing	1	yes	2	a	4	a
Manufacturing	1	yes	3	a	3	a
Manufacturing	2	yes	1	a	2	b
Manufacturing	2	yes	2	b	2	c
Manufacturing	2	yes	1	b	2	b
Manufacturing	2	yes	1	b	3	b
Manufacturing	2	yes	2	b	2	d
Manufacturing	3	yes	3	b	1	c
Manufacturing	3	yes	1	c	1	c
Manufacturing	3	Partially	1	c	1	d
Manufacturing	3	Partially	1	c	2	d
Manufacturing	3	yes	1	c	1	c
Service	1	yes	3	a	4	a
Service	1	yes	3	a	3	a
Service	1	yes	2	b	4	c
Service	1	yes	1	b	3	d
Service	2	yes	1	a	3	b
Service	2	yes	2	b	2	d
Service	2	yes	1	b	3	d
Service	3	yes	1	b	1	b
Service	3	Partially	1	c	1	c
Service	3	yes	1	c	1	d
Percentages	Median Value	Percentages	Percentages	Percentage	Median Value	Percentages
Manufaturing	2.07	yes	Integrated BI	small org (data type)	2.24	a=17%
41%	Small	90%	48%	a=71%,b=29%,c=0%		sm=100%
Service	24%	No	Multiple tools	Medium org	Small org	b=21%
34%	Medium	0%	28%	a=23%,b=69%,c=8%	3.43	med=83% lg=17%
Finance	45%	Partially	Both	Large org	Medium org	c=31%
24%	Large	10%	24%	a=0%,b=22%,c=78%	2.38	sm=11% med=33% lg=56%
	31%				Large org	d=31%
					1.11	sm=11% med=56% lg=33%

Figure 6. Survey responses and results

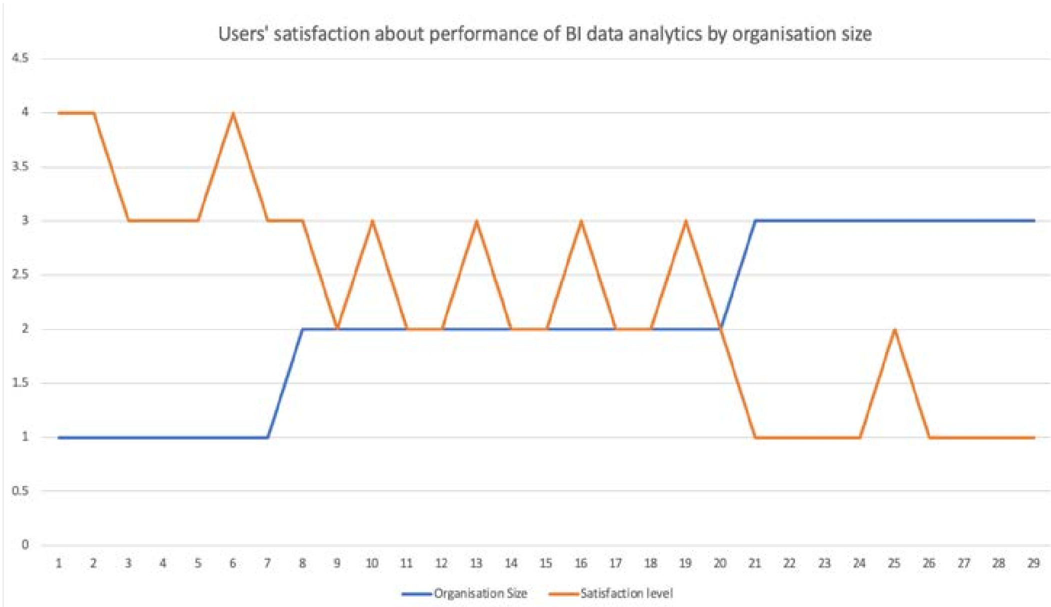

Figure 7. BI analytical performance as data volume increases

Figure 8. Types of data analysed by various sizes or organisations (Mostly structured with some semi and/or unstructured (a), Mostly structured with moderate volume of semi and unstructured (b), Structured with a large volume of semi-structured and unstructured (c), Mostly semi-structured and unstructured (d))

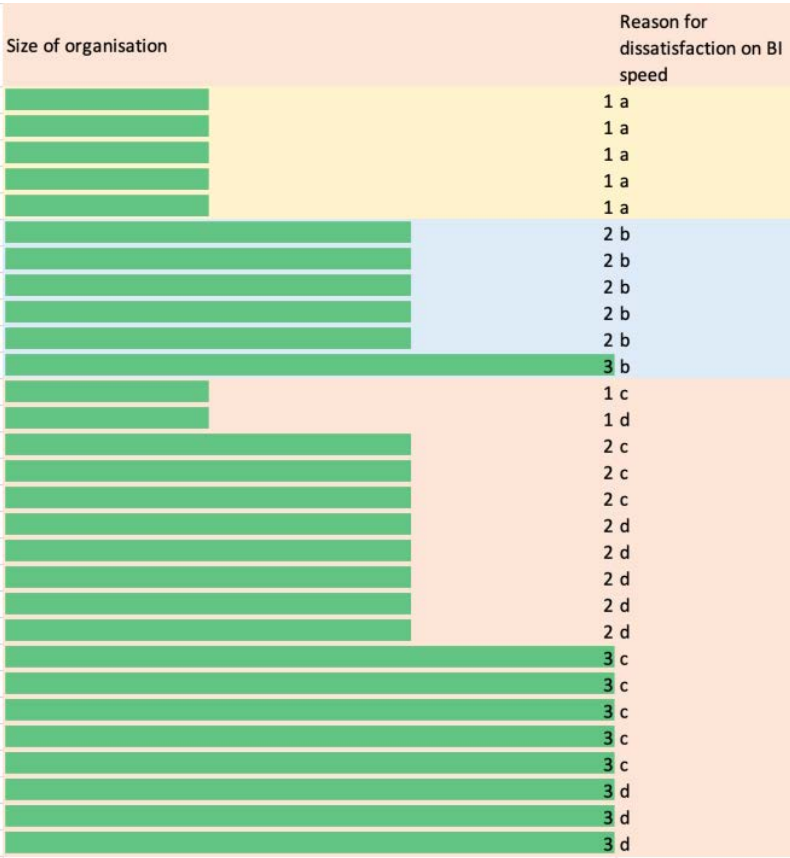

Figure 9. Reason for dissatisfaction over BI analytical performance by size of organisation (a-BI system does not have any major issues, b-Facing difficulty in handling large data volume, c-Having difficulty in handling semi-structured and unstructured data, d-Difficulty in modelling and mapping of semi-structured and unstructured data, e-Other reasons)

5. Discussion and Possible Remedy

The results presented in the previous section provide a clear indication of the relationship between the size of an organization and the level of satisfaction regarding the analytical performance of their BI systems. As organizations expand in size, so does their number of employees, and in tandem with this growth, the volume of data used by the organization also increases. Consequently, it can be inferred that the escalation in data volume contributes to a decrease in users' satisfaction with the processing speed and overall performance of the BI system. This observation raises an essential question about the capability of BI systems to handle big data, particularly those with semi-structured and unstructured elements.

The findings reveal that smaller organizations predominantly rely on structured data with a small volume of semi and/or unstructured data. In contrast, medium-sized organizations work with structured data that incorporates a moderate volume of semi and unstructured information, and large organizations deal with structured data alongside a substantial volume of semi and unstructured data. This trend of increasing data complexity aligns with the pattern of user dissatisfaction with BI system performance.

Moreover, the reasons behind dissatisfaction further shed light on the issue. Smaller organizations tend to opt for options 'a' or 'b', implying that, given their reliance on structured data, they believe their BI systems don't face significant issues aside from handling large volumes of structured data. In contrast, medium and large organizations, with their substantial volumes of big data, express concerns primarily related to the mapping and analysis of semi-structured and unstructured data (options 'c' and 'd'). Given the higher dissatisfaction levels observed in medium and large organizations, it is apparent from this study that current BI systems struggle when dealing with big data that includes semi-structured and unstructured components.

Currently, organizations typically employ separate systems to analyse structured data and big data. This division stems from the structured nature of organizational data, while big data exhibits complexity in its semi-structured and unstructured forms, necessitating columnar databases or data lakes for storage. Such data types possess intricate characteristics that require more extensive refinement to make them compatible with relational systems. To address the challenges associated with big data analysis, it is imperative to revamp and redesign current BI systems during the implementation phases. This transformation should enable the proper mapping and loading of semi-structured and unstructured data into data warehouses. The solution might involve the introduction of parallel data lakes and non-relational (NoSQL) databases to host, refine, and convert semi-structured and unstructured data into columnar data models, which can then be mapped onto data cubes. Once this data transformation process is complete and the data is loaded into a data warehouse, it can be readily analysed using BI tools or systems without significant obstacles. This suggests that further research is needed to develop a viable model or prototype that can address these challenges effectively.

6. Limitations

This research has several limitations, which should be taken into consideration in future investigations. Firstly, all the participants in this study were drawn from organizations located within the same geographic region. This geographical restriction may raise concerns about the generalizability of the research findings to a global context. However, it is important to note that the research primarily focuses on the volume of data handled by BI systems, which is a concern irrespective of geographic location. Consequently, similar research conducted in different locations should yield comparable results, although further research is needed to confirm this.

Additionally, the scope of the survey questionnaire may be considered limited, particularly regarding the options provided in question 6. The available options might have influenced users' perceptions of the reasons for their dissatisfaction with current BI systems. Nonetheless, the questionnaire did include an 'other' option, allowing respondents to express any additional reasons they deemed relevant. The fact that none of the respondents availed of the 'other' option suggests that the available answer choices appropriately represented users' satisfaction levels with organizational BI systems.

Another limitation pertains to the assumption that organizational size, as measured by the number of employees, is directly proportional to data volume. For instance, it was assumed that a medium-sized organization with 500-1000 employees generates significantly more data than a small-sized organization with 100-500 employees. However, survey results reveal that, as employee numbers increase, the type of data utilized by an organization change. Smaller organizations predominantly use mostly structured data, while medium and larger organizations employ varying volumes of semi-structured and unstructured data. Therefore, the assumption that data volume correlates directly with employee numbers has been corroborated by the findings.

Furthermore, this study operates on the assumption that hardware does not significantly contribute to the slowdown of BI analytical performance. This assumption may raise questions about the validity of the research. However, it is essential to note that all participants were BI power users, including BI consultants, analysts, and managers. Furthermore, these users had the opportunity to indicate hardware-related issues as one of the reasons for their dissatisfaction (option 'e' - other reasons), and yet no respondents cited hardware as a concern. As a result, any adverse effects stemming from hardware can be reasonably discounted.

7. Conclusion

Organizations rely on Business Intelligence (BI) systems to analyse data pertinent to their operational needs, extracting valuable insights that guide them toward informed decision-making. However, this research reveals a critical gap in the current landscape of BI systems: they are ill-prepared to cope with the burgeoning volume of big data, particularly semi-structured and unstructured datasets. To explore this issue, this study engaged twenty-nine BI consultants, analysts, and managers across organizations of various sizes. The analysis of survey data indicates that users dealing with expanding datasets, particularly those of a semi-structured and unstructured nature, exhibit higher levels of dissatisfaction with the analytical performance of conventional BI systems.

Furthermore, users have pinpointed the challenges related to modelling, mapping, and handling of semi-structured and unstructured datasets as the principal underlying causes of their dissatisfaction. In light of these findings, this research recommends adapting and reconfiguring the handling of big data within current BI systems during the implementation phase. This entails considering the integration of non-relational databases to ensure these systems remain effective tools for data analysis in the era of big data.

Appendix A: Questionnaire

Select the answer that best suits your circumstances.

1. What is the category of your organisation?

Ans: Manufacturing ( ), Service ( ), Finance ( )

2. What is the employee size of your organisation?

Ans: 100-500 (1), 500-1000 (2), >1000 (3)

3. Does your organisation utilize a Business Intelligence (BI) system in organisational data processing?

Ans: Yes ( ), No ( ), Partially ( )

4. What type of BI system does your organisation use?

Ans: An integrated BI system (1), Multiple Isolated BI tools (2), Both (3)

5. What sort of data does your organisation process with the help of the BI system?

Ans: Mostly structured with some semi-structured and/or unstructured (a), Mostly structured with moderate volume of semi-structured and unstructured (b), Structured with a large volume of semi-structured and unstructured (c), Mostly semi-structured and unstructured (d)

6. Overall, how much are you satisfied with the analytical performance of the BI system?

Ans: 1 (Not satisfied at all), 2, 3, 4, 5 (Extremely satisfied)

7. What, do you believe, are the reasons for your dissatisfaction (if any) with the analytical performance of the BI system in your organisation?

Ans: BI system does not have any major issues (a), Facing difficulty in handling large data volume (b), Having difficulty in handling semi-structured and unstructured data (c), Difficulty in modelling and mapping of semi-structured and unstructured data (d), Other reasons (e)

The questionnaire and collected answered were paper based. After collection, data were transferred to Ms Excel spreadsheet for analytical purposes.

References

[1] Yiu, L.M.D., Yeung, A.C.L. and Jong, A.P.L. (2020), "Business intelligence systems and operational capability: an empirical analysis of high-tech sectors", Industrial Management & Data Systems, Vol. 120 No. 6, pp. 1195-1215.

[2] Zhang, L., Qi, Z., Fan Meng, (2022) A Review on the Construction of Business Intelligence System Based on Unstructured Image Data, Procedia Computer Science, Volume 199, 2022, Pages 392-398, ISSN 1877-0509.

[3] Calitz, A., Bosire, S., & Cullen, M. (2018). The role of business intelligence in sustainability reporting for south african higher education institutions. International Journal of Sustainability in Higher Education, 19(7), 1185-1203.

[4] Lin, Yang & Jun, Zhou & Hongyan, Mei & Zhongwei, Zhang & Zhanfang, Feng. (2018). A Method of Extracting The Semi-structured Data Implication Rules. Procedia Computer Science. 131. 706-716. 10.1016/j.procs.2018.04.315.

[5] Arshad, M., Brohi, M.N., Soomro, T.R., Ghazal, T.M., Alzoubi, H.M., Alshurideh, M. (2023). NoSQL: Future of BigData Analytics Characteristics and Comparison with RDBMS. In: Alshurideh, M., Al Kurdi , B.H., Masa'deh, R., Alzoubi , H.M., Salloum, S. (eds) The Effect of Information Technology on Business and Marketing Intelligence Systems. Studies in Computational Intelligence, vol 1056. Springer, Cham.

[6] Laudon, Kenneth C., Laudon, Jane P. (2014), Management Information Systems: Managing the Digital Firm, Thirteenth Edition, Pearson Education Limited, England.

[7] Chung, J. Z. A. , Chang , M. M. L. and Phang, S. K. (2023) "Recent Advances and Benchmarking of NewSQL for OLTP and OLAP in the Big Data Age," 2023 IEEE 13th Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 2023, pp. 168-174.

[8] Francia, M., Marcel, P., Peralta, V. et al. Enhancing Cubes with Models to Describe Multidimensional Data. Inf Syst Front 24, 31–48 (2022).

[9] Vaisman, A and Zimanyi, E, (2014) Data Warehouse Systems, Data-Centric Systems and Applications.

[10] Horner, J., Song, II., Chen, P. (2004), An analysis of additivity in OLAP systems, Conference: DOLAP 2004, ACM Seventh International Workshop on Data Warehousing and OLAP, Washington, DC, USA, November 12-13, 2004, Proceedings.

[11] Jameel, K., Adil, A., & Bahjat, M. . (2022). Analyses the Performance of Data Warehouse Architecture Types. Journal of Soft Computing and Data Mining, 3(1), 45-57. https:// publisher.uthm.edu.my/ojs/index.php/jscdm/article/view/9561.

[12] Kavishwar, P., Pande, S. R. (2023), comparison of data models in data warehouse, international journal of researches in biosciences, agriculture and technology, IJRBAT, Issue (XI) Vol (II) May 2023 : 222-230, e-ISSN 2347 – 517X.

[13] Majid, Molla E. (2018) "Role of ICT in promoting sustainable consumption and production patterns - a guideline in the context of Bangladesh," *Journal of Environmental Sustainability*: Vol. 6: Iss. 1, Article 1. Available at: https:// scholarworks.rit.edu/ jes/vol6/iss1/1.

[14] Qaiser, A., Farooq, M., Mustafa, M., Abrar, N., (2023), Comparative Analysis of ETL Tools in Big DataAnalytics, Pakistan Journal of Engineering and Technology, PakJETISSN (p): 2664-2042, ISSN (e): 2664-2050Volume: 6, Number: 1, Pages: 7-12, Year: 2023.

[15] Favaretto M, De Clercq E, Schneble CO, Elger BS (2020) What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. PLoS ONE 15(2): e0228987.

[16] Segaran, Toby; Hammerbacher, Jeff (2009). Beautiful Data: The Stories Behind Elegant Data Solutions. O'Reilly Media. p. 257. ISBN 978-0-596-15711-1.

[17] Ralph., Kimball (2004). The data warehouse ETL toolkit : practical techniques for extracting, cleaning, conforming, and delivering data. Caserta, Joe, 1965-. Indianapolis, IN: Wiley. ISBN 978-0764579233. OCLC 57301227.

[18] Wang, Z. Jenipher (March 2017). "Big Data Driven Smart Transportation: the Underlying Story of IoT Transformed Mobility". URL: http://www.wiomax.com/big-data-driven-smart-transportation-the-underlying-big-story-of-smart-iot-transformed-mobility/ accessed Feb 2023.

[19] MarketLine (March 2015), Big Data: A road map for business intelligence, MarketLine case study, Reference code ML00019-004, www.marketline.com.

[20] Oesterreich T. D., Anton E.,Teuteberg F., (2022) What translates big data into business value? A meta-analysis of the impacts of business analytics on firm performance, Information & Management, Volume 59, Issue 6, 2022, 103685, ISSN 0378-7206.

[21] Reinsel, David; Gantz, John; Rydning, John (2017). "Data Age 2025: The Evolution of Data to Life-Critical" (PDF). seagate.com. Framingham, MA, US: International Data Corporation. Retrieved Feb 2023.

[22] SAS (2020), Hadoop: What is it and why it matters, SAS Institute Inc., Published online, URL: https:// www.sas.com/ en_us/ insights/ big-data/hadoop.html Accessed: Feb 2023.

[23] Cembalo, A., Ferrucci, A., Pisano, F., Pigliasco, G., (2013), A Spatial Data Warehousing Application in Epidemiology Domain, Conference: 15th International Conference on Data Warehousing and Knowledge Discovery - DaWaK 2013 At: Prague, Czech Republic Volume: L. Bellatreche and M.K. Mohania (Eds.): DaWaK 2013, LNCS 8057, pp. 97–109, 2013.© Springer-Verlag Berlin Heidelberg 2013.