International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org **Original Research Paper**

Analysis of Exploratory Data For The Interactive Visualization Gap

Md Shahid Ahmad[1], Ravi Kumar[2], Md. Talib Ahmad[3], Niraj Kumar[4]

Submitted: 06/02/2024 **Revised:** 14/03/2024 **Accepted:** 20/03/2024

Abstract: Complexity is one of the defining features of data scaling. When it comes to big data and data integration, heterogeneous data is a major factor. Both are necessary, but enormous amounts of data processing and storage make it hard to see and understand big databases. Data extraction in a way that the human brain can understand is a major challenge in this data-driven era of exponential data growth. This study summarizes and offers a description of heterogeneous distributed storage, data visualization, and the difficulties associated with it, drawing on a variety of approaches from prior studies. In addition, we compare the findings of the examined research works and talk about the major change happening in the field of virtual reality huge data visualization.

Keywords: Big Data, Heterogeneous, Visualization, Distribute Data Value

1. INTRODUCTION

Computers, social media, mobile devices, and more are producing vast amounts of data, and data analytics and visualization are embracing this new era of Big Data (Mustafa et al., 2020; Obaid et al., 2020; Zebari et al., 2020). Given the massive amounts of data processed and stored, visualizing and understanding large-scale databases is crucial yet challenging (Dino et al., 2020; Mahmood et al., 2021; Zhu et al., 2015). Over the last two years, 90% of the world's technology has been produced (Dragland, 2013; Zebari et al., 2018; Zeebaree et al., 2020), according to Science Daily, demonstrating an astonishing speed of data creation. To survive this onslaught, data processing philosophies, methods, and methodologies will need to undergo a radical shift (Alzakholi et al., 2020; Caldarola et al., 2014; Zeebaree et al., 2019). In order to effectively identify this data explosion and disseminate ground-breaking technological solutions that can handle this enormous quantity of data, the term "Big Data" was created (Franks, 2012; Jader et al., 2019; Zebari et al., 2019). Actually, the term "Big Data" has been trending upwards in Google Trends since 2011 (Saeed et al., 2020a; Weinberg et al., 2013; Zeebaree et al., 2020), as shown in previous studies. From different perspectives, big data may be defined in different ways when dealing with large data collections. Technically speaking, "Collection of data that are not capable of recording, store, handle and review traditional computing resources in database" is what Big Data shows (Dino and Abdulrazzaq, 2019; Manyika et al., 2011; Zeebaree et al., 2020). According to Dino & Abdulrazzaq (2020), Weinberg et al. (2013), and Zeebaree et al. (2020), marketers see big data more as an internal issue and a matter of decision-making than as a technological obstacle. Data that extends beyond the capabilities of computer systems and applications may also fall under this category.

often used in its user-controlled recording and processing over an acceptable time period" (Caldarola and Rinaldi, 2017; Haji et al.,2020; Saeed et al., 2020b). Last but not least, from the perspective of the user, Big Data should beseen as cutting-edge, intricate computing technologies that augment the existing ones (Osanaiye et al., 2019; Zeebaree et al.,2017; Zeebaree et al.,2019).

Nowadays, many different things generate a lot of data: social media, traffic sensors, satellite images, audio transmission, financial institutions, stock markets, etc. Data management structures like connection database servers are presented by the 3Vs, which stand for "Volume," "Speed," and "Velocity." These servers are capable of handling multiple relationship records, but they aren't very adaptable when it comes to handling unstructured or semi-structured data (Shukur et al.,2020; Zeebaree et al., 2020). Hence, it's critical to develop new innovation that can gather data from many sources, like social media, stock market, multi-sensor data, etc. Data Ingestion, Storage, Processing, and Analysis, and Data Insight are the overarching categories of operations covered (Chawla et al., 2018).

1. Assistant professor, Rashtrakavi Ramdhari Singh Dinkar College of Engineering, Begusarai. md.shahid.md@gmail.com

2. Assistant professor, Rashtrakavi Ramdhari Singh Dinkar College of Engineering, Begusarai. ravikmr.04@gmail.com

3. Assistant professor, Katihar engineering College, Katihar. talib@keck.ac.in

4. Assistant professor, Bakhtiyarpur College of Engineering, Bakhtiyarpur. saytoniraj@gmail.com

What follows is a framework for the remainder of the paper: Section 2 provides context for Big Data Visualization and the visualization process, explains their approaches, and then describes the difficulties of Big Data Visualization. The third section provides a literature review on data visualization using big data analysis. In the fourth section, we will talk about and compare the methods and outcomes of the two connected studies. The article is finally concluded in section five.

2. BACKGROUND THEORY

2.1 Visualization and big data and characteristics

Visualization seems to be a picture or graphic display of data. Data visualization has to be interpreted formally to evaluate and extract more in-depth perspectives from big data. Visualization of data helps to pull multiple data points together, understand data relations discuss problems in real-time, and determine more easily where to concentrate analysis(Abdullah et al., 2020; Haji et al., 2020; Khalifa et al., 2019). It allows data scientists to find secret data patterns and how they are stored. Business analysts may also use techniques of data visualization to define areas that require change or enhancement, concentrate on variables that affect consumer behavior, and forecast revenue volumes (Ali et al., 2016; Dino et al., 2020).

2.2 Big Data Visualization Process

As seen in figure 1, the method of visualization consists of the following steps (Chawla et al., 2018; Chen and Zhang, 2014):

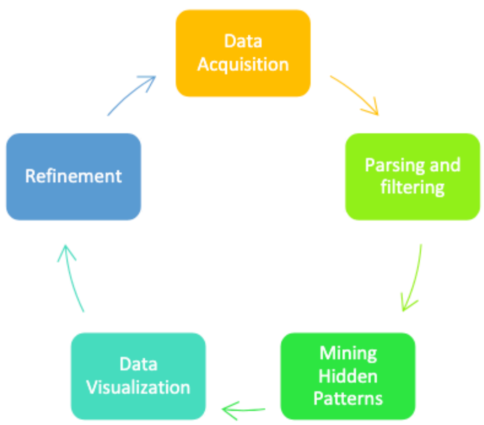

Figure1: Big Data Visualization Process

The visualization process begins with gathering data from various sources. Parsing the data in a structured manner is necessary since it may include unstructured or semi-structured information gathered from diverse sources (Zeebaree, 2020). The next step is to remove any irrelevant data as not all of it may be required for display. Afterwards, helpful patterns are deduced and shown using diagrams and charts. The user's basic comprehension of hidden information is revealed by the visualization of helpful patterns through charts and graphs (Chawla et al., 2018; Sallow et al., 2020; Zebari et al., 2017).

3. BIG DATA VISUALIZATION METHODS

The display of large data has been approached from several angles. Data size, data diversity, and data dynamics are the three criteria used to rank these approaches. Various approaches to data visualization include:

3.1. Tree map

Hierarchical data may be seen as a series of nested rectangles using this approach. Using a tiling method, the parent rectangle is divided into sub-rectangles. As a rule, a trained approach is used. The number assigned to a category is defined by the rectangular area. Therefore, only tree maps are subject to the restriction of negative and zero values. There is an imbalance in the hierarchy due to the increased number of pixels (Khalifa et al., 2019; Tennekes and Jonge, 2011).

3.2. Circle Packing

Another way to draw tree maps, this one employs circles to stand in for the different levels of a hierarchy. The number of a kind is determined by the circular area. The tree map is only one example of how it employs several colors in distinct groupings. Compared to the tree map, this method wastes a lot of space. (Peters et al., 2013; Mohammed et al., 2017).

3.3. Parallel Coordinates

This approach is a way to display large datasets. You can view the forest and the tree in parallel coordinates, and you may map the data components individually through numerous sizes. Consistent findings are obtained by drawing line trends. To see the exact output of specific data items, person lines may be highlighted. However, over plotting may be caused by a multitude of data items. Data categorical in nature does not use this technique. According to Johansson et al. (2008).

3.4 Stream Graph

Value displacement along a distinct central chronology may be shown using this approach. It shows that data from several categories has improved over time. Every stream form's size is proportional to the values of each category in a stream graph. Perfect for displaying a large dataset (Byron and Wattenberg, 2008). Data visualization tools can swiftly make sense of a mountain of data. With the right data visualization tool, people could see outliers, hidden patterns, or groupings they were unaware of. With these tools, you may also explore data sets that are in a constant state of flux. In table 1 we can see the key characteristics of large data visualization applications (Alzakholi et al., 2020; Byron and Wattenberg, 2008). One more is Space Titans 2.0, which facilitates comprehensive exploration of the Solar System. Using the advantages of current VR's enhanced spatial awareness, we aimed to get a fresh perspective on the way our environment seems. From the vantage point of Large Data Visualization, the skill necessary to scan an information branch in search of specific meaning or knowledge is hindered by multidimensional structures (Olshannikova et al., 2015; Zeebaree et al., 2019). The integration of virtual items with real-world scene perception is another area of interest for researchers. In addition to potentially slowing down the device, this mapping might also distort the actual picture. The development of a suitable structural framework to fortify the connection was motivated by the fact that even real and virtual distances are different. Effective research in physics, paleontology, MRI, and type interpretation is also required (Chawla et al., 2018).

Table 1: Big Data Tool Characteristics

Tools	Applications	Characteristics
Tableau	Market intelligence tool for the visual collecting of data used by academics and government agencies	Supports Google Big Data Query API, can handle massive volumes of data, can filter many sets at once, lets users create and share interactive dashboards showing patterns and variations, has built-in R support, and more.
Plotly	Python, R, MATLAB, Perl, J, Arduino, and the Restate graphics libraries are available online for use with static tools, graphing, and analysis.	A brand-new, freely-available agile framework for researching and analyzing data.
SAS Visual Analytics	Data visualization program; report, dashboard, and analysis delivery system	Data that may not be immediately obvious may be analyzed with the help of this comprehensive research tool.
Microsoft Power BI	Creating interactive dashboards, graphs, and visualizations with natural language queries	Power dash boards provide a 360-degree perspective for corporate users to keep all of their critical metrics in one spot, updated in near real-time, and accessible on any device.
D3.js	Making use of the widely used SVG, CSS, and HTML5	Visualization package for collaborative, immersive web browsers built with JavaScript

4. BIG DATA VISUALIZATION CHALLENGES

The quantity, diversity, and velocity of data make big data visualization challenging. Dealing with massive amounts of data and effectively presenting the useful and practical results of data visualization and analysis is the main challenge when working with big data. It is necessary to develop a new system that analyzes data in a manner that allows policymakers to rapidly and visibly understand it via the use of maps and graphs. Large data sets are too much for traditional visualization tools to handle. With the presenting tool, we can achieve the smallest conceivable display lag. This kind of data visualization often requires parallel processing. The most important part of visualizing vast amounts of data is finding interesting patterns. For pattern mining to work, the data measurements need to be precise. If we choose too many dimensions, our visualization will be too complicated and miss out on some interesting patterns; picking too few will make the view too complicated and unusable for consumers. For instance, considering the resolution of typical display (1.3 million pixels), over plotting, overlapping, and the user's perceptual and cognitive capabilities are all potential outcomes of visioning any points of data (Keahey, 2013). According to Ali et al. (2016), the majority of current visualization technologies have low efficiency when it comes to scalability, accessibility, and reaction time.

5. LITERATURE REVIEW

In order to speed up comprehensive visualization performance and enhance I/O speed, Zhu et al. (2015) suggested Visualization via the Heterogeneous Distributed Storage Infrastructure (VH-DSI). As an alternative to the traditional parallel file system, their suggested distributed file system type better supports visualization applications. In order to distribute compute tasks to data nodes taking data proximity and cluster heterogeneity into account, the authors also suggested a new scheduling method in VH-DSI, HetetShe. A distributed file system's POSIX-IO may be found in VH-DSI as well. In order to achieve better performance in both reaction time reduction and visualization acceleration, the testing results demonstrated the significance of the suggested VH-DSI solution and HeteSchi algorithm for visualization applications. The new Scalable Uniform Storage (SUORA) processing technique for heterogeneous devices was introduced by Ali et al. (2016). It addresses optimally adaptive and randomized numbers. Data is distributed evenly via a tiered and hybrid storage cluster using their suggested technique, which is a bogus random algorithm. It classifies gadgets into distinct categories, then assigns them to specific subcategories within each category. In addition, the writers generated a series of deterministic and pseudo-random numbers to map data between segments and devices. To improve read throughput performance and maintain load balance regarding bucket threshold and data hotness, data movement is also used. Successful adaptive data distribution for diverse storage systems and data centres was shown by the evaluation performance results of the SUORA algorithm.

To better distribute enhanced data in a heterogeneous object-based storage system and to make greater use of the abilities of heterogeneous computers, Zhou et al. (2016) suggested the HiCH method. Based on the results of the Sheepdog evaluation, HiCH sorts heterogeneous devices into distinct buckets and uses distinct consistent hashing rings for each bucket. It partitions data into distinct hashing rings based on hotness, data access, access duration, and behaviours. Findings shown that the HiCH algorithm has the potential to improve storage system efficiency and use heterogeneous storage devices more effectively. Using sys-stat, Kaneko et al. (2016) compared the read/write data throughput among conventional data placement methods and examined a storage system set up according to the established guideline. The rule of thumb is that all servers should have an equal amount of data that clients may access. While expanding the number of access sources, the findings revealed that the total data throughput rate increased using the proposed technique. Technical representation of heterogeneous CPUs using the Legion runtime framework was suggested by Yu and Yu (2016). We have covered the fundamentals of scientific visualization, which might include several processes with different data requirements. Multiple processes may be executed simultaneously on present and future supercomputers with the aid of this method, which optimizes storage partition programming, data visualization, and data transportation for heterogeneous distributed-memory systems. To improve the efficacy and utility of data analytics, Fiaz et al. (2016) laid forth methods for Big Data as well as Data Visualization. The authors noted that most businesses lack the necessary expertise to do the necessary data analysis, and that any techniques utilized to handle Big Data are difficult. Data visualization technologies make this task easier and provide better management and interpretation of data.

A technique that does not need information leaking for data translation was developed by Malik et al. (2016). This encompasses both data and metadata in a way that prevents them from increasing complexity while yet maintaining their inherent link. In order to make metadata more applicable to certain types of data sources, it is enhanced. The process of converting text into RDF inside a relational database is shown by an example. These techniques could be useful for covering all the bases in the data model for audio, video, images, and text. Using data value, Loorak et al. (2017) introduced the heterogeneous storage architecture. Various data values were used to dynamically choose the convenience shop. A more efficient system is achieved by using the SSD strategy, which prioritizes high-level data values, and the HHD approach, which prioritizes low-level data values. Hadoop, a distributed file system, was also suggested as the foundation for research and evaluation. The data-value-based heterogeneous storage architecture was introduced by Li et al. (2017). Various data values were used to dynamically choose the convenience shop. A more efficient system is achieved by using the SSD strategy, which prioritizes high-level data values, and the HHD approach, which prioritizes low-level data values. Hadoop, a distributed file system, was also suggested as the foundation for research and evaluation. Data analysis methods for large data and heterogeneous data were presented by Wang (2017). These methods include large Data techniques, machine learning (ML), and some traditional data mining (DM) approaches. A synopsis of extensive understanding and its capabilities in Big Data analytics is provided. It lays forth the advantages of HPC, Deep Learning, Big Data Analytics, and Heterogeneous Computing Integration. Discussed as well are challenges associated with handling large datasets and diverse data types, as well as research into big data. For better visualization, Liu et al. (2017) suggested a new approach. Based on the initial visual structure, the graphic framework is dynamically updated to meet user needs. Furthermore, the connections between entities may evolve in real time in response to data changes. Additionally, rather of using SQL to query the database, they employ the improvement technique. The procedure proposed in this article is not data set specific, and any data set may make use of it. Researchers observed that even when consumers don't understand the data set structure, they can still visually examine and enjoy the data by investigating the data interaction. Zhi (2017) proposed a distributed optimization storage model using hash distribution as a result of research into cloud computing's inherent data processing features. There is a 12.2% improvement in throughput and a 9.8% reduction in response time when using the distributed optimization storage model as opposed to the sequential storage distribution method. Around 8Mb/s is the haphazard writing speed. Evidence from the simulation's points to cloud computing as the foundation of the hash distribution-based distributed storage device model architecture. There were three major advancements that Iturbe et al. (2017) announced: We first take a look at and examine several Large Data Anomaly Detection Systems (ADSs) that can be useful for INs. To categorize existing ADSs according to IN, a new taxonomy was created. (3) The potential for continued growth of INs was the subject of a discussion on open and accessible data ADSs. The subject of anomaly detection in industrial networks using big data is still in its early stages, but it is showing promise as a place to focus research efforts on these unanswered questions. A systematic tool was suggested by Kammer et al. (2018) to facilitate the evaluation and construction of ML-based clustering algorithms. This tool takes use of a number of visualization elements, including histograms, semantic zoom, and glyphs. With the use of clustering and classification, machine learning (ML) enables data discovery, online commerce, and adaptive learning environments to construct structures. Interactive Big Data Landscapes were conceptualized as a consequence of the study's findings.

In their 2018 study, Mahfoud et al. suggested a platform for immersive visualization utilizing Microsoft HoloLens to examine diverse sensor data. Their approach delves into the fundamentals of how a viewer might conceptualize dynamic data and unearth hidden parallels in mixed reality. It also presents event recognition algorithms that can detect suspicious data automatically. Interactive mixed-reality analytics capabilities, shown in the demonstration framework, enable analysts to monitor and understand time series data regardless of their physical location, freeing them from the constraints of traditional computer settings. To accomplish efficient data aggregation for diverse storage systems, Zhou et al. (2019) used a contemporary PRS data duplication strategy. The PRS groups adapt to data access patterns by creating objects and distributing functioning copies to computers with different architectures. PRS takes storage system efficiency and capacity into account while using a pseudo-random method to improve the design of parallels. The experimental findings proved that PRS is an excellent method for replicating in systems with different components. A comprehensive HBase-based data storage strategy for remote sensing photos was suggested by Liang and Zhou (2019). This method stores large data remote sensing images using a columnar open-source database called HBase. To construct the tile pyramid for the remote sensing picture, it uses tile pyramid technology in conjunction with MapReduce, a parallel processing engine. Lastly, the data blocks containing the remote sensing images are placed in the distributed database HBase. The storage issue of massive data picture remote sensing may be efficiently addressed by this technology, which boasts high processing quality, scalability, and dependability.

In order to gather all the necessary information for further research, Mehmood et al. (2019) suggested using big data analysis technologies. The data may be processed, retrieved, integrated, and viewed and studied further using this interface. This method is CUTLER's first foray into integrating data points from four pilot regions. A framework for typing graph behaviour in heterogeneous networks and higher-order spectral clustering was presented by Carranza et al. (2020). A higher-order structure may be built using the proposed method by constructing clusters that maintain connectedness from typed graph allows. Previous research on higher-order clustering of spectrum has been extended by this method. Technically, the authors demonstrated a number of important outcomes, such as a Cheeger-like inequality for typed-graph let activity, which suggests that the method approaches its optimum boundaries. Theoretical results provide a unified theoretical grounding for researching spectral approaches of a higher order and considerably simplify prior work. Three main scientific applications—clustering, compression, and relation estimation—quantitatively demonstrate the method's effectiveness. To cut down on computation time and provide a new approach to calculation allocation, Woolsey et al. (2020) used Maximum Distance Separable (MDS) storage assignment for heterogeneous Coded Elastic Computing (CEC) networks. Suggest a new way of looking at combinatorics optimization and a "filling problem," and then break it down into its component optimization problems in order to identify the optimal computational load and solutions.

Following an examination of relevant literature on the topic of big data visualization. In order to acquire a substantial amount of comprehensive data, the researchers employed a wide variety of techniques and methodologies. The satisfaction of such algorithms is different. This is why related studies that use VR for big data visualization overcame obstacles and offered solutions, paving the way for the observation and analysis of varied and complicated data structures.

6. DISCUSSION

Numerous research investigations have focused on their relevance, as is obvious from the literature reviews mentioned above. The results of this study demonstrated that researchers approached the problem of Big Data visualization from several angles. It makes it possible to use clusters of a programming paradigm to spread processing of big quantities utilizing basic datasets. It is scalable, cost-effective, reliable, and has several important properties including high availability and fault tolerance. Table 2 displays the statistical evaluation of these investigations.

Table 2: COMPARISON OF THE PROPOSED APPROACHES BY PREVIOUS RESEARCHES

Author(s)	**Algorithm**	**Objectives**	**Significant Results**
Zhu et al. (2015)	HeterSche	Using a Diverse Distributed Storage Architecture to Accelerate Data Visualization	cuts reaction time in half, if not in half again
Ali et al. (2016)	(SUORA)	Algorithms for heterogeneous storage with standard and adaptable application providers	Highly effective adaptive data distribution for heterogeneous storage systems and data centers.
Zhou et al. (2016)	HiCH	Storage of Heterogeneous Objects with Hierarchical Consistent Hashing	Raise the capacity of storage systems and expand access to heterogeneous computing.
Kaneko et al. (2016)	sysstat and compare read or write data throughput	Instructions for storing data in distributed storage systems with different types of hardware	guideline maximise the flow of aggregated data while increasing the capacity of all available access points
Yu and Yu (2016)	Legion runtime system	Evidence-Based Scientific Visualization on Multi-Processor Systems with Legion	proven the method of data distribution, scalable execution, and simple use with hybrid data partitioning
Fiaz et al. (2016)	Techniques for Big Data and Data Visualization together	Making Big Data More Valuable and Adaptable with Data Visualization	allow for more effective data interpretation and control.
Malik et al. (2016)	Relationship between data and metadata	Improved data transformation for mixed sources	helpful for a data model that encompasses a broad variety of media types (video, audio, images, and text).
Loorak et al. (2017)	Heterogeneous Embedded Data Attributes (HEDA)	Delving into the potential by combining diverse data characteristics	fresh approaches to well-known visualization methods

Li et al. (2017)	data value and hedoop distributed file system	data-value-based distributed heterogeneous storage	bring the system up to speed
Wang (2017)	------------	Processing heterogeneous and large datasets	----------------
Liu et al. (2017)	------------	Different data support schemes and a novel approach to data representation	strategy for investigating the data interaction may be significantly simplified
Zhi (2017)	hash distribution	Study of a Cloud-Based Storage Model for Distributed Data Optimization	The cloud is the foundation of the hash distribution-based storage system template.
Iturbe et al.	-------------	Detecting Heterogeneous Anomalies	Exciting potential areas for the future of
(2017)		Infrastructure for Business Networks:	studies of the dynamic production networks
Kammer et al. (2018)	Clustering algorithm for machine learning	Enhancing the Visualization of Clustering Algorithms for Machine Learning in Big Data Landscapes	Created the notion of interactive Big Data Landscapes.
Mahfoud et al. (2018)	Microsoft HoloLens	Interactive visual representation of diverse on-site decision-making for anomaly detection	interactive mixed-reality analytics capabilities, freeing analysts from traditional computer settings and enabling them to monitor and understand time series data regardless of their physical location.
Zhou et al. (2019)	pseudo random algorithm	Storage of Heterogeneous Objects with a Pattern-Directed Replication Scheme	very efficient replication method for diverse inventory management systems
Liang and Zhou (2019)	HBase and parallel processing system MapReduce	Studies on HBase-Remote Sensing Image-Based Distributed Big Data Storage	The storage issue of massive data picture remote sensing is efficiently addressed, and the system is dependable, scalable, and has high processing quality.
Mehmood et al. (2019)	CUTLER	Deploying large data lakes to integrate diverse data sets	This strategy in the CUTLER project is the first of its kind to combine several data points from four test fields.
Carranza et al. (2020)	clustering, compression, and relation estimation	Advanced Grouping in Multi-Heterogeneous Systems	Making previous research easier Hence, a theoretical framework was born.
Woolsey et al. (2020)	novel calculation allocation	Elastic Coded Computing on Speed and Speed Store with Heterogeneity	fresh approach to solving combinatorics optimization problems and Get to the bottom of it by

			breaking it down into an optimization strategy.

7. CONCLUSION

Issues with data integration and big data processes are exacerbated by heterogeneous data. Both are necessary, but enormous amounts of data processing and storage make it hard to see and understand big databases. In this article, we take a look back at the literature on data visualization using big data analysis. It also compares the outcomes based on the algorithms and methodologies used. This is why related studies that use VR for big data visualization overcame obstacles and offered solutions, paving the way for the observation and analysis of varied and complicated data structures.

8. REFERENCES

1. Abdullah, P. Y., Zeebaree, S. R., Jacksi, K., & Zeabri, R. R. (2020). AN HRM SYSTEM FOR SMALL AND MEDIUM
2. ENTERPRISES (SME) S BASED ON CLOUD COMPUTING TECHNOLOGY. International Journal of Research- GRANTHAALAYAH, 8(8), 56–64.
3. Ali, S. M., Gupta, N., Nayak, G. K., & Lenka, R. K. (2016). Big data visualization: Tools and challenges. 656–660.
4. Alzakholi, O., Haji, L., Shukur, H., Zebari, R., Abas, S., & Sadeeq, M. (2020). Comparison Among Cloud Technologies and Cloud Performance. Journal of Applied Science and Technology Trends, 1(2), 40–47. https://doi.org/10.38094/jastt1219
5. Byron, L., & Wattenberg, M. (2008). Stacked graphs–geometry & aesthetics. IEEE Transactions on Visualization and Computer Graphics, 14(6), 1245–1252.
6. Caldarola, E. G., & Rinaldi, A. M. (2017). Big Data Visualization Tools: A Survey. Research Gate.
7. Caldarola, E. G., Sacco, M., & Terkaj, W. (2014). Big data: The current wave front of the tsunami. Applied Computer Science, 10(4), Article 4.
8. Carranza, A. G., Rossi, R. A., Rao, A., & Koh, E. (2020). Higher-order clustering in complex heterogeneous networks.
9. 25–35.
10. Chawla, G., Bamal, S., & Khatana, R. (2018). Big Data Analytics for Data Visualization: Review of Techniques'.
11. International Journal of Computer Applications, 182(21), 37–40.
12. Chen, C. P., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, 275, 314–347.
13. Dino, H., Abdulrazzaq, M. B., Zeebaree, S. R., Sallow, A. B., Zebari, R. R., Shukur, H. M., & Haji, L. M. (2020). Facial Expression Recognition based on Hybrid Feature Extraction Techniques with Different Classifiers. TEST Engineering & Management, 83, 22319–22329.
14. Dino, Hivi I., & Abdulrazzaq, M. B. (2020). A Comparison of Four Classification Algorithms for Facial Expression Recognition. Polytechnic Journal, 10(1), 74–80.
15. Dino, Hivi Ismat, & Abdulrazzaq, M. B. (2019). Facial Expression Classification Based on SVM, KNN and MLP Classifiers. 2019 International Conference on Advanced Science and Engineering (ICOASE), 70–75.
16. Dino, Hivi Ismat, Zeebaree, S. R., Ahmad, O. M., Shukur, H. M., Zebari, R. R., & Haji, L. M. (2020). Impact of Load Sharing on Performance of Distributed Systems Computations. International Journal of Multidisciplinary Research and Publications (IJMRAP), 3(1), 30–37.
17. Dino, Hivi Ismat, Zeebaree, S. R., Salih, A. A., Zebari, R. R., Ageed, Z. S., Shukur, H. M., Haji, L. M., & Hasan, S. S. (2020). Impact of Process Execution and Physical Memory-Spaces on OS Performance. Technology Reports of Kansai University, 62(5), 2391–2401.
18. Dragland, A. (2013). Big Data–for better or worse. SINTEF. No. 22 May 2013. Web. 27 Oct.
19. Fiaz, A. S., Asha, N., Sumathi, D., & Navaz, A. S. (2016). Data visualization: Enhancing big data more adaptable and valuable. International Journal of Applied Engineering Research, 11(4), 2801–2804.
20. Franks, B. (2012). Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics (Vol. 49). John Wiley & Sons.
21. Haji, L. M., Zeebaree, S. R., Ahmed, O. M., Sallow, A. B., Jacksi, K., & Zeabri, R. R. (2020). Dynamic Resource Allocation for Distributed Systems and Cloud Computing. TEST Engineering & Management, 83(May/June 2020), 22417–22426.
22. Haji, L., Zebari, R. R., R. M. Zeebaree, S., Abduallah, W. M., M. Shukur, H., & Ahmed, O. (21, May). GPUs Impact on Parallel Shared Memory Systems Performance. International Journal of Psychosocial Rehabilitation, 24(08), 8030–8038. https://doi.org/10.37200/IJPR/V2418/PR280814
23. Iturbe, M., Garitano, I., Zurutuza, U., & Uribeetxeberria, R. (2017). Towards large-scale, heterogeneous anomaly detection systems in industrial networks: A survey of current trends. Security and Communication Networks, 2017.
24. Jader, O. H., Zeebaree, S. R., & Zebari, R. R. (2019). A State of Art Survey For Web Server Performance Measurement And Load Balancing Mechanisms. INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, 8(12), 535–543.
25. Johansson, J., Forsell, C., Lind, M., & Cooper, M. (2008). Perceiving patterns in parallel coordinates: Determining thresholds for identification of

relationships. Information Visualization, 7(2), 152–162.
26. Kammer, D., Keck, M., Gründer, T., & Groh, R. (2018). Big data landscapes: Improving the visualization of machine learning-based clustering algorithms. 1–3.
27. Kaneko, S., Nakamura, T., Kamei, H., & Muraoka, H. (2016). A guideline for data placement in heterogeneous distributed storage systems. 942–945.
28. Keahey, T. A. (2013). Using visualization to understand big data. IBM Business Analytics Advanced Visualisation, 16.
29. Khalifa, I. A., Zeebaree, S. R., Ataş, M., & Khalifa, F. M. (2019). Image Steganalysis in Frequency Domain Using Co- Occurrence Matrix and Bpnn. Science Journal of University of Zakho, 7(1), 27–32.
30. Li, H., Li, H., Wen, Z., Mo, J., & Wu, J. (2017). Distributed heterogeneous storage based on data value. 264–271. Liang, C., & Zhou, L. (2019). Research on Distributed Storage of Big Data Based on HBase Remote Sensing Image. 1,
31. 2628–2632.
32. Liu, Q., Guo, X., Fan, H., & Zhu, H. (2017). A novel data visualization approach and scheme for supporting heterogeneous data. 1259–1263.
33. Loorak, M., Perin, C., Collins, C., & Carpendale, S. (2017). Exploring the Possibilities of Embedding Heterogeneous Data Attributes in Familiar Visualizations. IEEE Transactions on Visualization and Computer Graphics, 23(1), 581–590.
34. Mahfoud, E., Wegba, K., Li, Y., Han, H., & Lu, A. (2018). Immersive visualization for abnormal detection in heterogeneous data for on-site decision making. Proceedings of the 51st Hawaii International Conference on System Sciences.
35. Mahmood, M. R., Abdulrazzaq, M. B., Zeebaree, S. R., Ibrahim, A. K., Zebari, R. R., & Dino, H. I. (2021). Classification techniques' performance evaluation for facial expression recognition. Indonesian Journal of Electrical Engineering and Computer Science, 21(2), 176–1184.
36. Malik, K. R., Ahmad, T., Farhan, M., Aslam, M., Jabbar, S., Khalid, S., & Kim, M. (2016). Big-data: Transformation from heterogeneous data to semantically-enriched simplified data. Multimedia Tools and Applications, 75(20), 12727–12747.
37. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.
38. Mehmood, H., Gilman, E., Cortes, M., Kostakos, P., Byrne, A., Valta, K., Tekes, S., & Riekki, J. (2019). Implementing big data lake for heterogeneous data sources. 37–44.
39. Mohammad, O. F., Rahim, M. S. M., Zeebaree, S. R. M., & Ahmed, F. Y. (2017). A survey and analysis of the image encryption methods. International Journal of Applied Engineering Research, 12(23), 13265–13280.
40. Mustafa, O. M., Haji, D., Ahmed, O. M., & Haji, L. M. (2020). Big Data: Management, Technologies, Visualization, Techniques, and Privacy. Technology Reports of Kansai University, 62(05).
41. Obaid, K. B., Zeebaree, S. R., & Ahmed, O. M. (2020). Deep Learning Models Based on Image Classification: A Review. International Journal of Science and Business, 4(11), 75–81.
42. Olshannikova, E., Ometov, A., Koucheryavy, Y., & Olsson, T. (2015). Visualizing Big Data with augmented and virtual reality: Challenges and research agenda. Journal of Big Data, 2(1), 22.
43. Osanaiye, B. S., Ahmad, A. R., Mostafa, S. A., Mohammed, M. A., Mahdin, H., Subhi, R., Zeebaree, D. A. I., & Obaid, O.
44. (2019). Network Data Analyser and Support Vector Machine for Network Intrusion Detection of Attack Type. REVISTA AUS, 26(1), 91–104.
45. Saeed, R. H., Dino, H. I., Haji, L. M., Hamed, D. M., Shukur, H. M., & Jader, O. H. (2020). Impact of IoT Frameworks on Healthcare and Medical Systems Performance. International Journal of Science and Business, 5(1), 115–126.
46. Sallow, A. B., Zeebaree, S. R., Zebari, R. R., Mahmood, M. R., Abdulrazzaq, M. B., & Sadeeq, M. A. (2020). Vaccine Tracker/SMS Reminder System: Design and Implementation. International Journal of Multidisciplinary Research and Publications, 3(2), 57-63
47. Shukur, H., Zeebaree, S., Zebari, R., Ahmed, O., Haji, L., & Abdulqader, D. (2020). Cache Coherence Protocols in Distributed Systems. Journal of Applied Science and Technology Trends, 1(3), 92–97.
48. Shukur, H., Zeebaree, S., Zebari, R., Zeebaree, D., Ahmed, O., & Salih, A. (2020). Cloud Computing Virtualization of Resources Allocation for Distributed Systems. Journal of Applied Science and Technology Trends, 1(3), 98–105.
49. Tedesco, J., Dudko, R., Sharma, A., Farivar, R., & Campbell, R. (2013). Theius: A streaming visualization suite for hadoop clusters. 177–182.
50. Tennekes, M., & de Jonge, E. (2011). Top-down Data Analysis with Treemaps. 236–241.
51. Wang, L. (2017). Heterogeneous data and big data analytics. Automatic Control and Information Sciences, 3(1), 8– 15.
52. Weinberg, B. D., Davis, L., & Berger, P. D. (2013). Perspectives on big data. Journal of Marketing Analytics, 1(4), 187–201.
53. Woolsey, N., Chen, R.-R., & Ji, M. (2020). Coded Elastic Computing on Machines with Heterogeneous Storage and Computation Speed. ArXiv Preprint ArXiv:2008.05141.
54. Yu, L., & Yu, H. (2016). A study of scientific visualization on heterogeneous processors using Legion. 107–108. Zebari, D., Haron, H., & Zeebaree, S. R. (2017). Security Issues in DNA Based on Data Hiding: A Review.
55. International Journal of Applied Engineering Research ISSN 0973-4562, 12(24), 15363–15377.
56. Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., & Saeed, J. (2020). A Comprehensive Review of Dimensionality Reduction Techniques for Feature

Selection and Feature Extraction. Journal of Applied Science and Technology Trends, 1(2), 56–70. https://doi.org/10.38094/jastt1224
57. Zebari, R. R., Zeebaree, S. R., & Jacksi, K. (2018). Impact Analysis of HTTP and SYN Flood DDoS Attacks on Apache
58. 2 and IIS 10.0 Web Servers. 2018 International Conference on Advanced Science and Engineering (ICOASE), 156–161.
59. Zebari, R., Zeebaree, S., Jacksi, K., & Shukur, H. (2019). E-Business Requirements for Flexibility and Implementation Enterprise System: A Review. International Journal of Scientific & Technology Research, 8, 655–660.
60. Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zeebaree, S. R. (2017). Combination of K-means clustering with Genetic Algorithm: A review. International Journal of Applied Engineering Research, 12(24), 14238– 14245.
61. Zeebaree, S. R. (2020). DES encryption and decryption algorithm implementation based on FPGA. Indonesian Journal of Electrical Engineering and Computer Science, 18(2), 774–781.
62. Zeebaree, S. R., Jacksi, K., & Zebari, R. R. (2020). Impact analysis of SYN flood DDoS attack on HAProxy and NLB cluster-based web servers. Indonesian Journal of Electrical Engineering and Computer Science, 19(1), 510–517.
63. Zeebaree, S. R., Sallow, A. B., Hussan, B. K., & Ali, S. M. (2019). Design and Simulation of High-Speed Parallel/Sequential Simplified DES Code Breaking Based on FPGA. 2019 International Conference on Advanced Science and Engineering (ICOASE), 76–81.
64. Zeebaree, S. R., Shukur, H. M., & Hussan, B. K. (2019). Human resource management systems for enterprise organizations: A review. Periodicals of Engineering and Natural Sciences, 7(2), 660–669.
65. Zeebaree, S. R., Zebari, R. R., & Jacksi, K. (2020). Performance analysis of IIS10.0 and Apache2 Cluster-based Web Servers under SYN DDoS Attack. TEST Engineering & Management, 83(March-April 2020), 5854–5863.
66. Zeebaree, S. R., Zebari, R. R., Jacksi, K., & Hasan, D. A. (2019). Security Approaches For Integrated Enterprise Systems Performance: A Review. INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, 8(12).
67. Zeebaree, Subhi R. M., Haji, L. M., Rashid, I., Zebari, R. R., Ahmed, O. M., Jacksi, K., & Shukur, H. M. (2020). Multicomputer Multicore System Influence on Maximum Multi-Processes Execution Time. TEST Engineering & Management, 83(May/June), 14921–14931.
68. Zeebaree, Subhi R M, M. Shukur, H., M. Haji, L., Zebari, R. R., Jacksi, K., & M.Abas, S. (2020). Characteristics and Analysis of Hadoop Distributed Systems. Technology Reports of Kansai University, 62(4), 1555–1564.
69. Zeebaree, SUBHI R.M., Salim, B. wasfi, R. Zebari, R., Shukur, H. M., Abdulraheem, A. S., Abdulla, A. I., & Mohammed, S. M. (2020). Enterprise Resource Planning Systems and Challenges. Technology Reports of Kansai University, 62(4), 1885–1894.
70. Zhi, C. (2017). Research of Distributed Data Optimization Storage Model in the Cloud Computing Environment.
71. 193–198.
72. Zhou, J., Chen, Y., Xie, W., Dai, D., He, S., & Wang, W. (2019). PRS: A Pattern-Directed Replication Scheme for Heterogeneous Object-Based Storage. IEEE Transactions on Computers, 69(4), 591–605.
73. Zhou, J., Xie, W., Gu, Q., & Chen, Y. (2016). Hierarchical consistent hashing for heterogeneous object-based storage.1597–1604.
74. Zhu, Y., Juniarto, S., Shi, H., & Wang, J. (2015). VH-DSI: speeding up data visualization via a heterogeneous distribu*ted storage infrastructure*. 658–665