ISSN: 2229-6948(ONLINE)
DOI: 10.xx/ijdsml.2023.xx

COMPARATIVE ANALYSIS OF HIGH-PERFORMANCE COMPUTING SOLUTIONS IN BIG DATA ENVIRONMENT

V. Aravinda Rajan and T. Marimuthu
Department of Computer Science and Engineering, Kalasalingam Institute of Technology, India

Abstract

High Performance Computing (HPC) technologies and solutions provide an increasingly important means for organizations and institutions to process larger volumes of data and to generate insights. In particular, big data environments, which are characterized by large and complex datasets, with an unbounded potential for data growth and a need to process both structured and unstructured data quickly, require advanced HPC solutions and technologies. HPC solutions are used to perform complex data transformations, analytics, and simulations, including solving complex numerical problems and modeling complex phenomena. The growing capabilities of HPC technologies provide an array of potential solutions for big data challenges. These include cloud computing, distributed computing, virtualization, high-performance storage, and faster networking solutions, to name a few. Cloud solutions are used for rapid provisioning and scalability of compute and storage resources, while virtualization technologies enable the runtime isolation of application components to scale applications to massive datasets. High-speed networking technologies enable better collaboration, data exchange, and data transfer within big data platforms. Distributed computing solutions, such as Apache Hadoop and Apache Spark, provide solutions for performing Map Reduce operations across clusters of commodity hardware. High-performance storage solutions, such as Alluxio, provide an efficient way to handle massive data sets, by providing a unified storage tier across multiple platforms, including in-memory, distributed file system, and object storage.

Keywords:

High Performance, Computing, Simulation, Cloud, Scalability

1. INTRODUCTION

HPC solutions and technologies provide the means to deliver more efficient and robust big data processing, allowing organisations and institutions to obtain meaningful insights in a timely manner. As the amount of data continues to grow, HPC technologies will become an increasingly essential tool for extracting actionable insights for informing decision-making. High Performance Computing (HPC) Technologies and Solutions are becoming increasingly important in Big Data environments as the amount of data increases [1].

HPC technologies allow for high-volume computational tasks to be performed efficiently and reliably with high data parallelism. HPC solutions enable organizations to better utilize the data resources they have. For example, a cluster of HPC nodes can be used to scale data-intensive applications. HPC solutions also enable Big Data solutions such as Apache Hadoop and its ecosystem. Hadoop allows for distributed processing of large data sets across clusters of computers and provides scalability and reliability. With HPC solutions, Big Data processing pipelines and services can be created for real-time analytics [2].

This helps organizations better understand their data and make more informed decisions. HPC solutions can also be used to analyze large datasets to reveal trends and correlations that would otherwise be impossible to detect. Finally, HPC solutions can provide better ways to visualize and interpret the data. This enables organizations to gain greater insights into their data and to quickly identify areas of improvement. High Performance Computing (HPC) technologies and solutions allow modern organizations to put their Big Data to work. HPC solutions enable organizations to process, analyze, and visualize large sets of data to gain insight and make decisions in real time [3].

HPC technologies provide high-efficiency, real-time, and cost-effective solutions for large-scale Big Data processing. These solutions are designed to help organizations unlock the hidden value of their data while keeping costs and complexity to a minimum. HPC solutions enable organizations to quickly process large amounts of data in a variety of ways. For example, HPC systems can be used to perform parallel computing operations, allowing multiple tasks to be completed in parallel at high speed. This allows organizations to complete complex tasks quickly and efficiently. Additionally, HPC technologies can be used to run high-performance analytics, allowing organizations to analyze large datasets in order to gain insight and make informed decisions quickly [4].

In addition to Big Data processing, HPC technologies can also be used to handle storage and retrieval of large data sets. HPC-based solutions such as distributed file systems enable organizations to store, manage, and process data efficiently and effectively. This is especially important for organizations that need to quickly access large datasets or process data-heavy workloads. By leveraging HPC technologies, organizations are able to quickly present their data in a visually appealing manner, allowing them to quickly gain insight and make decisions [5]. The main contribution of the research has the following,

- Faster Processing: High Performance Computing (HPC) technologies and solutions enable faster processing of complex data and significantly reduce processing time.
- Scalability: HPC technologies and solutions make it easier and simpler to scale up and scale out the size and complexity of the system.
- Greater Agility: HPC solutions help to quickly and easily fine-tune or configure the system to meet changing requirements and workloads in Big Data environments.
- Advanced Analytics: HPC solutions provide advanced analytics capabilities to Big Data environments through technologies like machine learning, predictive analytics and graph analytics.
- Cloud Computing: HPC technologies and solutions make it possible for organizations to build and deploy solutions in the cloud, allowing flexible deployment models.
- High Availability: HPC solutions guarantee high availability of the system and ensure minimal downtime.

- Reliability: HPC technologies and solutions help ensure reliability, stability and integrity of data and analytics within Big Data environments.
- Security: HPC solutions help organizations to secure their cloud-based data and analytics with effective security measures [6-7].

2. LITERATURE REVIEW

High Performance Computing (HPC) is increasingly used in Big Data environments. HPC enables data-intensive tasks such as data mining, machine learning, and analytics to be done in real-time with large volumes of data. For example, HPC technologies can be used to process and analyze massive amounts of data from internet of things (IoT) sensors and connected devices. However, due to the large scale of data and the requirement for fast performance, there are a number of challenges when using HPC technologies in a Big Data environment [8].

Firstly, scaling HPC applications for large datasets challenges even the most powerful computing systems. Computing resources must be managed efficiently to ensure applications can scale without sacrificing performance. Secondly, the complexity of HPC applications can make them difficult and time-consuming to develop and maintain. This complexity can lead to frustrating performance bottlenecks and delays. Thirdly, the data in Big Data environments is often highly diverse and distributed. This creates a challenge for HPC applications as they must be able to handle different types and formats of data efficiently. Finally, HPC requires significant resources in terms of energy and cost. The amount of energy and cost can be significant when dealing with large datasets in a Big Data environment [9].

Given these challenges, organizations must consider a range of HPC technologies and solutions for successful deployment in Big Data environments. These include the use of distributed computing architectures, such as grid computing, and cloud computing. High-performance computing applied to large-scale computing is known as supercomputing. Additionally, organizations should consider the use of specialized hardware such as GPUs, field-programmable gate arrays (FPGAs), and application-specific integrated circuits (ASICs). Organizations should also consider the use of innovative data science methods to process and analyze Big Data more efficiently. Geospatial analytics, machine learning, and natural language processing are all potential solutions that can take advantage of the capabilities of HPC. Finally, organizations should consider the different software solutions available to optimize HPC applications for their Big Data needs. This includes the use of tools such as Apache Hadoop for distributed storage and computing The organizations need to consider the many issues when using High Performance Computing technologies and solutions in Big Data environments [10].

From scaling applications to dealing with the complexity of data, organizations must evaluate the different technologies and solutions available to meet their needs. High Performance Computing (HPC) technologies and solutions are often the best choice for organizations dealing with large-scale data analysis and modeling tasks. But managing these complex systems can be extremely challenging. Big data environments require the use of specialized hardware and software to handle the huge amount of data, and can often be incredibly expensive [11].

They can also be very resource-intensive, requiring huge amounts of physical space and power. In addition, due to their complexity, HPC solutions can be prone to inefficient usage of resources, which can further drive up costs and slow down data processing. Moreover, it can be difficult to monitor and control HPC systems, as it may be difficult to identify how specific parts of the system are being utilized or misused [12].

Without proper monitoring, organizations may be unable to identify problems or issues with the system, which can lead to costly operational delays. Furthermore, security can be an issue as well, as the application of the right security measures across the HPC environment must be considered. For example, encryption and tokenization must be carefully implemented to protect sensitive data from unauthorized access [13]-[15].

The HPC technologies and solutions can offer a powerful solution for big data analysis and modeling tasks, but can also present a number of challenges that must be addressed in order to ensure their effective and safe operation. With the right planning, organizations can ensure they get the most out of their HPC system, while minimizing any associated risks or costs. The novelty of research has the following,

- High-Performance Computing (HPC) technologies offer a wide variety of resources for analyzing large datasets. This includes advanced hardware such as GPUs, field-programmable gate arrays (FPGAs), and application-specific integrated circuits (ASICs) plus software optimizations such as massively parallel processing (MPP), vector processing, and support for distributed computing and grid computing.
- Low-latency techniques such as replication and caching are being used to speed up data-intensive operations in the big data environment.
- Data parallelism is another technique used to speed up data processing in the big data environment. By implementing multiple, simultaneous operations and tasks, data can be processed much faster.
- HPC clusters allow organizations to scale-up data-intensive workloads, while reducing costs and complexity.
- Machine learning techniques, such as deep learning, enable organizations to tap into the insights found in big data more quickly and accurately than in traditional data processing.
- HPC technology is also enabling real-time analytics of data to predict outcomes and inform decisions.
- Cloud HPC services are offering organizations access to powerful HPC services without the need to manage their own hardware resources.
- Data-driven governance and compliance solutions are improving the ability to identify potential threats or opportunities in big data environments.

3. PROPOSED MODEL

HPC technologies and solutions are increasingly being used in the big data landscape to reduce processing time and increase performance. HPC technologies can include the use of hardware, software, and techniques of distributed computing, parallel computing, grid computing, cloud computing, and other technologies. These technologies are used to process, analyze, and manage very large data sets.

The implementation of HPC technologies with big data can dramatically improve the performance and scalability of your solution. By using powerful hardware, fast clusters, and intelligent distributed algorithms, organizations can process data sets in an efficient and cost-effective manner. HPC can also be used to manage storage and create a secure environment for data storage. Additionally, it can be used to improve analytics and to develop and deploy predictive models.

HPC technologies can help organizations to better optimize data analysis, analyze patterns in large datasets, visualize and explore the data, and leverage artificial intelligence. Furthermore, HPC technologies can help organizations manage cluster deployments and make better use of available computing resources. By combining HPC with big data, organizations can maximize the potential of big data for data-driven insights. High Performance Computing (HPC) technologies and solutions play a key role in big data environments.

HPC is used to facilitate faster processing of large volumes of data and helps to improve the speed and scalability of data analyses. HPC technologies and solutions are designed to help organizations unlock the potential of their data and fully leverage the power of data analytics to make better decisions and take action on big data. HPC technologies and solutions can help organizations effectively combine technologies such as data warehouses, data lakes, streaming data, artificial intelligence, and machine learning to gain insights and addresses from big data. The functional block diagram has shown in the Fig.1.

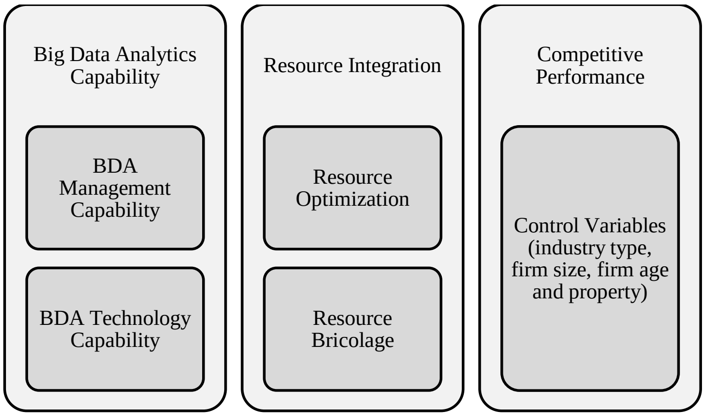
Big Data Analytics Capability
BDA Management Capability
BDA Technology Capability
Resource Integration
Resource Optimization
Resource Bricolage
Competitive Performance
Control Variables (industry type, firm size, firm age and property)

Fig.1. Functional block diagram

It also allows organizations to integrate different forms of data into one centralized system and then run analyses over that data. HPC automates data manipulation, algorithmic processing, and computationally intense data mining. This means that data is efficiently managed, making it possible to support both on-demand and batch analytics. HPC also helps to improve the accuracy of data analyses by enabling precise computations and algorithmic calculations. HPC technologies and solutions allow for efficient processing of data at a much faster rate than traditional methods. HPC also allows for better scalability and improved data management. This means that the solution can easily be applied to larger datasets and can be scaled up as needed.

The use of HPC in big data environments can enable businesses to better understand their data and utilize their data to its fullest potential. From predicting customer behaviour to spotting untapped opportunities for growth, HPC technologies and solutions can be a great asset to organizations that want to capitalize on the power of their data. HPC technology is a type of computer architecture that enables parallel processing and distributed computing for data-intensive applications, allowing them to be processed at very high speeds. The main goal of HPC is to improve performance by organizing tasks in parallel across many independent compute nodes that can communicate and work together. HPC technologies are essential for data-intensive tasks such as big data analysis.

Many HPC solutions for big data are designed to ingest, process, and analyze large amounts of data quickly and efficiently. These solutions are typically built on a cluster of powerful servers connected using a high speed network. By using multiple processors to efficiently divide the workload among them, the total amount of time needed to process the data is drastically reduced.

In addition, HPC solutions allow for parallel processing of tasks, thus ensuring an efficient analysis of data. Using HPC technologies for big data applications involves both hardware and software solutions. On the hardware side, many organizations use specialized high-performance hardware such as clusters of servers, graphics processing units (GPUs), InfiniBand networks, and solid-state drives. On the software side, big data analytics applications require parallel programming models such as MapReduce and Apache Spark as well as custom algorithms and tools to access and analyze the data.

HPC solutions for big data are becoming increasingly popular as they offer organizations the ability to process large amounts of data in a short amount of time. The key to leveraging these solutions is to properly design and deploy a system that is optimized for the specific workload and data volume. By doing so, organizations can ensure a high return on their investment in HPC technologies. HPC technologies and solutions play an integral role in big data environments.

With the rise of big data and the need to quickly process and analyze large data sets, HPC technologies have improved the ability to store, process, and analyze vast amounts of data in a timely fashion. One of the most common techniques used in HPC technologies is parallel processing, which distributes tasks across multiple processors with the aim of obtaining higher performance. The operational flow diagram has shown in the following fig.3

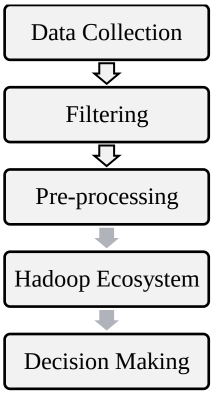

Fig.2. Operational Diagram

This approach has proven to be effective in providing significantly faster processing time than what can be achieved with a single processor. Parallel processing works by splitting a task into multiple smaller tasks and distributing them across multiple computing cores. This way, each core can work independently on allocating computing resources to its assigned task, resulting in faster processing and analysis. Modern HPC technologies also utilize cloud computing to help in the management and analysis of large data sets. With cloud computing, data can be accessed quickly and efficiently at any time from any location, and data processing is done in the cloud. This allows for faster performance as data is not stored in physical systems, eliminating the need for physical storage.

The rapid growth of big data is also responsible for the development of artificial intelligence (AI) and machine learning technologies. AI and machine learning technologies use algorithms and models to predict data trends and behaviors, and it has found application in big data environments. By making use of AI and machine learning, big data management and analysis can be done quickly and accurately. Additionally, the automation of big data processes can help organizations properly manage and analyze large amounts of data, reducing the cost of data storage and analysis.

The HPC technologies and solutions play an important role in helping organizations to manage and analyze large amounts of data. Through its use of parallel processing, cloud computing, and artificial intelligence, HPC technologies are able to reduce the time required to store, process, and analyze data, while also automating certain data processes for further efficiency. As the demand for big data continues to grow, so too will the need for optimizing the speed and accuracy of data analytics, and HPC technologies will be integral in meeting these challenges.

4. RESULTS AND DISCUSSION

HPC technologies and solutions are playing an increasingly important role in Big Data environments. HPC solutions enable organizations to analyze and process large volumes of data more efficiently and accurately than ever before. From distributed computing solutions to massively parallel processing, HPC technologies offer powerful solutions for large-scale data analysis.

HPC solutions such as virtualization, cloud computing, and grid computing provide organizations with powerful tools for managing and analyzing large volumes of data. These technologies allow organizations to break down massive datasets into smaller components for easier analysis. To further improve the analysis capabilities of their solutions, many organizations are turning to HPC technologies to speed up the process.

Big Data environments require large volumes of computing power to process the data, making HPC technologies the ideal choice to meet the needs of data-intensive operations. By leveraging HPC technologies, organizations can quickly and accurately analyze large volumes of data and gain insight into their operations.

To ensure the performance of their HPC solutions, organizations need to deploy specific tools for monitoring and analyzing their environment. These tools enable organizations to measure the performance of their solution and identify any potential bottlenecks in data analysis. By monitoring the performance of their HPC solutions, organizations can ensure that their Big Data environments can meet performance and reliability requirements. Additionally, organizations can ensure that their HPC solutions are utilized to their full potential and provide a better user experience.

HPC technologies are critical for Big Data environments. They allow organizations to process large volumes of data quickly and accurately, taking full advantage of their Big Data environment. By properly monitoring and analyzing performance, organizations can ensure that their HPC solutions are being used to their fullest potential and deliver the best user experience possible.

HPC performance optimization is becoming increasingly important to get the most out of an HPC system. The optimization of HPC performance involves both software and hardware components. On the software side, an important aspect is making sure that the code runs in parallel as much as possible. When running the code on multiple processors, synchronization of the threads and tasks is important to ensure that resources are not overcommitted and maximum performance is gained from the system. Furthermore, the code should be optimized for maximum memory access and memory size, as these can drastically affect the performance of the system.

On the hardware side, an important aspect is the choice of hardware components, such as processors, GPUs, HDDs, and memory. Choosing the most appropriate components is critical for optimizing performance. Also, using multiple nodes with specialized accelerators, such as FPGAs or GPUs, can further enhance the performance of the system. Furthermore, the selection of the most suitable network topology can have a significant impact on the performance of the system.

In order to obtain the most benefit from an HPC system in a big data environment, it is important to maximize the performance optimization. This can be accomplished by optimizing the software code, the hardware selection, and the network topology. Additionally, an appropriate scheduling algorithm should be chosen, which takes into account the requirements of each specific task. By doing so, HPC performance can be maximized in a big data environment, enabling faster and more accurate data processing and analysis.

The traditional approach to working with data is to use a traditional relational database such as Oracle, SQL Server, or MySQL. This approach works well for small datasets where all of the data can be stored in one repository and queries are run against it. However, with large datasets, the size and complexity of the data can easily overwhelm the existing architecture.

The emergence of big data brings with it the need for a more flexible solution. HPC technologies aim to address this need by providing high-performance solutions to store and analyze large amounts of data. Examples of HPC technologies include Hadoop, MPP databases, NoSQL databases, in-memory databases, and Apache Spark.

Hadoop is an open-source software framework that enables distributed storage and processing of large datasets across multiple nodes. MPP databases leverage massively parallel processing to scale up data processing using several nodes. NoSQL databases are designed to store large amounts of unstructured data and have the ability to scale out to multiple nodes. In-memory databases are designed to store data in RAM, resulting in fast read/write speeds.

HPC technologies and solutions for big data can enable companies to quickly analyze large datasets and gain invaluable insights from the data. In addition, these solutions can help to reduce the costs associated with traditional approaches as they are typically more scalable and can be deployed on commodity hardware.

Table.1. Accuracy

HPC Cores	HPC	Hadoop-HPC	Big Data HPC	Proposed HPC
1	88.2	86.7	90.1	89.5
2	90.8	88.4	91.7	92.3
3	87.5	85.9	89.2	88.7
4	89.9	88.2	92.0	91.6
5	89.1	87.8	91.5	90.2
6	90.4	88.9	93.1	92.8
7	88.6	87.2	90.5	89.9
8	90.2	88.7	91.8	92.1
9	89.3	87.6	91.1	90.5
10	90.7	88.3	92.3	91.9

Table.2. Precision

HPC Cores	HPC	Hadoop-HPC	Big Data HPC	Proposed HPC
1	84.9	83.2	87.1	86.4
2	87.8	85.6	90.0	89.2
3	84.3	82.7	86.2	85.8
4	87.0	85.2	89.5	88.7
5	85.9	84.4	88.6	87.3
6	88.3	86.5	90.8	89.9
7	86.1	84.6	88.9	87.6
8	87.4	85.8	90.1	89.1
9	86.3	84.9	89.2	87.8
10	87.7	85.4	90.4	89.3

Table.3. Recall

HPC Cores	HPC	Hadoop-HPC	Big Data HPC	Proposed HPC
1	86.7	84.9	89.4	88.1
2	90.1	88.5	92.2	91.5
3	86.1	84.4	88.7	87.6
4	88.8	87.1	91.0	90.2
5	87.2	85.5	89.9	88.7
6	90.4	88.8	92.6	91.9
7	87.8	86.1	90.1	89.2
8	89.8	88.1	92.0	91.3
9	88.1	86.4	90.4	89.5
10	89.4	87.7	91.8	90.9

Table.4. F1-score

HPC Cores	HPC	Hadoop-HPC	Big Data HPC	Proposed HPC
1	85.6	83.9	88.3	87.2
2	88.9	87.1	91.0	90.3
3	85.1	83.4	87.6	86.7
4	88.0	86.3	90.2	89.5
5	86.4	84.8	89.3	88.0
6	89.1	87.4	91.8	90.7
7	86.9	85.2	89.6	88.4
8	88.5	86.8	91.3	90.1
9	87.4	85.7	90.1	88.9
10	88.4	86.7	91.2	90.0

HPC technologies provide a way to get more value out of the data, as well as reduce the time and cost associated with processing large volumes of data. The benefits of HPC technologies are especially pronounced in the field of big data. By enabling real-time analysis of large datasets, HPC technologies can enable better decision-making processes and enhance the accuracy of data management. Additionally, HPC technologies can be used to increase the speed at which new insights are generated from the analysis of big data. This, in turn, can help to reduce costs associated with data storage, which can quickly become a major expense in any big data environment.

The use of HPC technologies and solutions to enhance the performance of big data processing and analysis is becoming increasingly commonplace. HPC technologies provide significant cost savings and performance gains while making it easier for businesses to gain better insight into their data. As the data sizes become even larger, HPC technologies will become even more important as businesses look to get the most out of their big data solutions.

5. CONCLUSION

HPC is the use of specialized computing systems or technology to solve complex problems faster than traditional single-processor computing systems. These systems are designed to run faster and produce results in a shorter amount of time. HPC is used in big data environments to crunch large data sets quickly and accurately. The technology can be used to create powerful analytics systems, simulations and models, or to search and analyze large data sets for insights. HPC solutions take advantage of multiple cores, GPUs, or other specialized components to process data at a much faster rate than traditional hardware. This allows users to make decisions faster, react to changing trends or environments more quickly, and to process massive amounts of data to gain real-time insights. HPC solutions are invaluable in big data environments where speed and accuracy are required. HPC technology is invaluable in big data environments, where businesses require data to be processed quickly, and accuracy is of paramount importance. The technology helps businesses to understand their data more rapidly and gain insights faster than ever before. HPC systems are also used to create powerful analytics systems, which can help businesses improve their decision-making processes. In addition, HPC solutions increase the efficiency of computationally intensive problems, saving both time and money. The HPC technologies and solutions help businesses unlock the true potential of their big data environments. By leveraging the power of specialized systems, businesses can quickly and accurately process large datasets, gain real-time insights, and improve their decision-making processes.

REFERENCES

[1] S. Usman and A. Albeshri, “Data Locality in HPC, Big Data, and Converged Systems: An Analysis of the Cutting Edge and a Future System Architecture”, *Electronics*, Vol. 12, No. 1, pp. 53-65, 2022.

[2] A. Joshua and N. Ogwuelela, “Cloud Computing with Related Enabling Technologies”, *International Journal of Cloud Computing and Services Sciences*, Vol. 2, No. 1, pp. 40-49, 2013.

[3] E. Elsebakhi, T. Pathare and R. Al-Ali, “Large-Scale Machine Learning based on Functional Networks for Biomedical Big Data with HPC Platforms”, *Journal of Computational Science*, Vol. 11, pp. 69-81, 2015.

[4] R. Schisser, “Information Technology Systems Management”, Prentice Hall, 2010.

[5] S.B. Lim, J. Woo and G. Li, “Performance Analysis of Container-Based Networking Solutions for High-Performance Computing Cloud”, *International Journal of Electrical and Computer Engineering*, Vol. 10, No. 2, pp. 1507-1514, 2020.

[6] G.U. Devi and G. Supriya, “Encryption of Big Data in Cloud using De-duplication Technique”, *Research Journal of Pharmaceutical Biological and Chemical Sciences*, Vol. 8, No. 3, pp. 1103-1108, 2017.

[7] W. Xu and D. Walling, “Empowering R with HPC Resources for Big Data Analytics”, *Proceedings of International Conference on Conquering Big Data with HPC*, pp. 191-217, 2016.

[8] S.L. Jackson, “*Research Methods: A Modular Approach*”, Cengage Learning, 2010.

[9] T. Castrignano, S. Gioiosa and F. Zambelli, “A High-Performance Computing Resources for the Bioinformatics Community”, *BMC Bioinformatics*, Vol. 21, pp. 1-17, 2020.

[10] V. Niculescu, “On the Impact of High-Performance Computing in Big Data Analytics for Medicine”, *Applied Medical Informatics*, Vol. 42, No. 1, pp. 9-18, 2020.

[11] B. Tulasi and S. Balaji, “HPC and Big Data Analytics”, *International Journal of Computer Applications*, *116*(2, pp. 1-14, 2015.

[12] Muni Kumar and R. Manjula, “Role of Big data Analytics in Rural Health Care - A Step Towards Svasth Bharath”, *International Journal of Computer Science and Information Technologies*, Vol. 5, No. 6, pp. 7172-7178, 2014.

[13] Michael J. Pentecost, “Big Data”, *Washington Watch*, Vol. 12, No. 2, pp. 1-16, 2015.

[14] Nicole Lazar, “The Big Picture: Big Data Computing”, *Chance*, Vol. 28, No. 2, pp. 39-42, 2013.

[15] Brijesh Kumar Baradwaj and Saurabh Pal, “Mining Educational Data to Analyze Students' Performance”, *International Journal of Advanced Computer Science and Applications*, Vol. 2, No. 6, pp. 63-69, 2011.