SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I1P105

Volume 12 Issue 1, 72-82, January 2025
© 2025 Seventh Sense Research Group®

Original Article

An Efficient Management Technique for Optimization of Compute, Storage, and Communication in Big Data Processing

Vijay Kumar Vasantham[1], D. Haritha[2]

[1,2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India.

[*]Corresponding Author : vijaykumarvasantham@gmail.com

Received: 09 November 2024 Revised: 15 December 2024 Accepted: 06 January 2025 Published: 30 January 2025

Abstract *- There are specific challenges like handling a variety of datasets, velocity, and largest, high-dimensional datasets to be handled efficiently in big data processing and optimization. In this regard, three crucial aspects were considered: computing, storage, and communication in big data optimization. To achieve accurate and timely data assessment and analysis, these three aspects are focused on the extraction of insights for the rest to be assured for future endeavours. The strategies used in the computing aspect enable the provision of fast data access, effective allocation of resources, and parallel processing. The approaches considered are distributed frameworks, in-memory systems, lazy evaluations, partitioning, and selective algorithms. The second aspect is the computation storage, which requires hardware, software tools, Orchestration tools, Analytics tools, and data management tools for faster access, efficient data movements, enhanced scalability, and effective security. The third aspect of big data optimization is communication, in which the approaches that were considered related to infrastructure, serialization, protocols, and notification/alerting systems to experience less overhead, proactive issue detection, and effective data exchange. This integrated aspect leads to efficient, informed decision-making based on insights and significant analysis. The customized framework for computing, storage, and communication services in big data optimization ensures better accuracy, scalability, processing efficiency, and cost-effectiveness against existing approaches.*

Keywords *- Big Data Optimization, Communication Services, Computation storage, Communication Management, Efficiency and Accuracy, Resource Allocation, Workload Distribution.*

1. Introduction

The exponential growth of data in recent years has necessitated the development of efficient processing techniques to handle large-scale datasets effectively. Traditional approaches often struggle to manage the computational and communication resources required for big data processing tasks, leading to inefficiencies and bottlenecks. The challenges of traditional approaches are to be minimized using novel optimization techniques over big data to enhance efficiency and better resource utilization. There is a demand for a novel optimization methodology that addresses these challenges by optimizing the computing and communication services, thereby improving overall performance.

The ever-growing volume, velocity, and variety of data sets present significant challenges in big data processing. Efficiently handling of these massive, high-dimensional datasets requires innovative optimization techniques. The three crucial aspects of big data optimization, computing, computation storage, and communication are considered. As a result, achieves accuracy and timely data analysis, paving the way for informed decision-making. For optimizing the computation aspect, the advantages guaranteed are fast data access, efficient resource allocation, effective privacy, and the power of parallel processing. Distributed frameworks, in-memory systems, lazy evaluations, partitioning, and selective algorithms are some key approaches considered for the computing aspect. The tools such as Computation storage Devices (CSDs), Solid state Devices (SSDs), Orchestration tools, Analytics tools, and data management tools are used in computation storage. Regarding communication, the advantages achieved by using suitable approaches such as optimizing infrastructure, serialization methods, protocols, and notification/alerting systems can significantly address advantages such as reducing communication overhead.

Additionally, proactive issue detection and efficient data exchange are explored. This framework is designed to deliver superior results compared to existing approaches by achieving better accuracy.

In the past two decades, big data processing has evolved from a struggle to manage massive datasets to a sophisticated ecosystem with computing, storage, and communication solutions.

This framework is designed to deliver superior results compared to existing approaches by achieving better accuracy and efficiency. Over the past two decades, big data processing has evolved from a struggle to manage massive datasets to a sophisticated ecosystem with advanced computing, storage, and communication solutions. This research aims to further this evolution by providing a comprehensive optimization framework addressing current gaps and challenges.

Existing research has explored various optimization techniques for big data processing. For example, a novel approach using memory mapping and a parallel message-passing interface has been proposed to handle big data with high performance and low cost. Another study highlights the use of convergent parallel algorithms and network analytics for large-scale optimization.

However, these approaches often focus on specific aspects of big data processing without providing an integrated solution. Our research addresses this gap by offering a unified framework that simultaneously optimizes computing, computation storage, and communication, leading to superior performance and resource utilization.

Table 1. Evolution of trends in Bid data optimization

Aspect	Early 2000s	Mid 2000s onwards	Recent Trends
Compute	Centralized processing	Distributed computing (Hadoop)	Cloud-based computing (AWS, Azure, GCP)
Storage	Limited storage capacity	Distributed File Systems (HDFS)	Object storage
Communi cation	Standard networking	High-performance networking	Software-Defined Networking (SDN)

From Table 1, computing is evolving from central computing to cloud-based processing, storage from small datasets to real-time objects, and communication from small networks to software-defined networks. Hence, a framework is needed that would assess three aspects of its environment in terms of computing, storage, and communication. Without these aspects, processing would not get any meaning and would not result in insights.

Table 2. Technologies that handle bulk data and their challenges

Technology over Big Data	Challenges	Remarks
Cloud Computing	Diverse vendor's comfort and efficiency	Efficient storage
Big Data and Cloud Computing	Readymade tools for efficient analysis	Efficient decision making
Internet of Things	Massive real-time data and unstructured	Tremendous data resulting

In the above scenarios, bulk data resulted from using IoT and computers for daily internet observation. To make intelligent decisions, data must be organized or processed. Later, cloud computing is required to store enormous amounts of data and must guarantee the security of the platform. Later, performance is a metric that would be a key factor for assessing these integrated technologies. To satisfy the requirements of performance in aspects of computing, computation storage, and communication, a customized approach efficiently guarantees accuracy and performance in these aspects. The significant methods of computing are demonstrated in Table 3, the significant methods for storage are demonstrated in Table 4, and specific approaches for communication are demonstrated in Table 5.

When designing a customized and efficient framework, an analysis should be performed on the existing approaches used in each aspect, such as computing, computation storage, and communication.

Table 3. Methods used in computing

Method	Purpose
Distributed frameworks	Guarantees fault tolerance, and scalability by distributing sub datasets of large datasets over a cluster for faster efficiency that involves parallel processing.
In-memory systems	Guarantees faster access using main memory RAM than taking time on disk accesses for storage and processing operations.
Lazy evaluations	Guarantees unnecessary computations and uses only required data from large datasets only when evaluating current tasks.
Partitioning	Processes complex tasks into manageable sub-tasks and reduces input-output operations for performance improvement.
Selective algorithm	Focuses customized or available algorithms that process only required portions, enhancing faster access.

Table 4. Methods used in computation storage

Method	Purpose
Hardware Tools	Should have physical infrastructure to store large datasets that support distributed and network capabilities.
Software tools	Should have frameworks or software that support data across multiple storage devices and parallel processing.
Orchestration tools	Should reduce human intervention and automate data processing between hardware and software.
Analytics tools	Should have machine learning and AI methods for extraction of insights, identify required patterns and enable decision making.
Data management tools	Should result from highly reliable data over data combined from various sources and cleanse and preprocess the uneven sources.

Table 5. Methods used in communication

Method	Purpose
Infrastructure	Should provide storage devices, distributed systems, and data transfer across the system to achieve scalability.
Serialization	Support data transfer easily when data is in the form of a stream of bytes over the networks and different systems.
Protocols	Defines rules and standards for transfer among different systems to achieve interoperability and scalability.
Notification/alerting systems	Should act proactively and monitor issues to minimize the error rate and maximize reliability.

Hence, our proposed model guarantees many significant benefits compared to the traditional method, as demonstrated in Table 6.

Table 6. Traditional versus Efficient, customized approach

Feature	**Traditional Approaches**	**Efficient Customized Approach**
Scalability	Limited	Highly scalable
Flexibility	Rigid	Greater flexibility
Real-time Processing	Batch-oriented	Real-time or near-real-time
Cost-Efficiency	Expensive	More cost-effective
Data Variety	Primarily structured	Handles diverse data types
Fault Tolerance	Vulnerable	Built-in fault tolerance
Integration	Challenging	Seamless integration
Advanced Analytics	Limited	Enables advanced analytics

Table 7. Customized Approach vs Traditional Approaches

Method	Accuracy	Performance	Cost-effectiveness
Customized Approach	Higher accuracy by tailored algorithms	Faster	The initial setup is high + cost-effective for the long run
Traditional Approaches	Lower accuracy if complex patterns involved	Slower	Initial setup is low + but the cost is expensive in the long run

In the consolidation of the customized method and traditional method from Table 7, accuracy, performance, and cost-effectiveness are the key metrics to be involved in the evaluation process.

2. Literature Review

In this, the focus is on how to optimize the big-data scenarios when there is a massive amount of unstructured data and not in a position to make informed decisions. A few studies are demonstrated for a further level of enhancement of our proposed system.

Chandrima Roy et al. (2018) demonstrate the drawbacks of conventional approaches, various optimization best practices, and tools like Hadoop over big data. It is implemented based on certain parameters such as process, memory, map reduction, data node, and name nodes. As per M. Wu et al. (2021), graph computations and the methodology of this model were compared against existing methods where the advantages of graph models were exploited. The analysis of the SFA algorithm is discussed towards graph modeling. The study of Hira Zahid et al. (2019) demonstrates various challenges and opportunities of big data in the field of telecommunications. Architectures like partial big-data implementation and full-fledged implementation of big-data over telecommunication are demonstrated. The proposed LambdaTEL makes practitioners aware of the practical implications for BDA applications. In the view of Agnieszka, Smalec (2021) demonstrates that the variety of communications that can be adapted in big data processing leads to large datasets by human contribution and the transfer of data from the usage of devices. There should be challenges to the multidimensional usage and management of communication in the hypervolume of media. It creates a path to initiate research in this communication.

Naghib A. et al. (2023) demonstrate complex environments like IoT where four mechanisms are used such as big data processing, architectures, significant features, and types of analytical tools. This study brings reliable and adaptable BDM for IoT environments. From Madhavi Vaidya et al. (2014), the discussion is initiated on issues involving large databases and possessing labor intensiveness. Three approaches have been proposed to overcome the the challenges of traditional approaches and are such as synchronization bulk methods, Parallel database systems, and map-reduce approach. In the aspect of Sandeep Dasari, Rajesh Kaluri (2023) discusses the 6-S approach over big data, which is formed daily from various sources. This framework helps make more intelligent decisions than any other approach based on enough available data and prevents privacy to the greatest extent possible. From Rahat Iqbal et al. (2021), the challenges raised from bulk-connected devices and daily routines are addressed and minimized by the proposed study. The method called hierarchical temporal, spatial machine, also titled a biologically inspired generative model, would follow physical, cyber architecture. This architecture provides a lot of benefits and shifts towards a modern paradigm.

In regard to Ahmed Hadi Ali, AL-Jumaili et al. (2023) demonstrate many issues when working with power processing systems by parallel conventional methods. To overcome these, a cloud-based infrastructure framework with real-time needs is proposed that provides effective monitoring and efficient processing in these power-enabled systems. In the view of Qingqing Chang et al. (2022), DSS is proposed to support effective decision-making and efficient intelligence computation. The challenges, such as inefficient decision-making due to the huge amounts of information on big data, have been overcome and addressed appropriately. Agnieszka Smalec et al. (2021) address the potential problems with daily results of information from the usage of the internet and connected devices. The hypermedia faces challenges like inefficient interpretation approaches that would be overcome by providing a comprehensive approach. This would be a multi-directional marketing and communication approach that allows efficient decision-making. In regard to [12], the demonstration of big data into chunks forms a cluster after processing, like removing duplication and later compressing them. Later, these are combined as a processed dataset and stored on the cloud rather than servers. Here, storage happens efficiently, and parallelism increases the performance. In the view of [13], demonstration of the new design of B+-tree with expansion and conversion into ring format would minimize the overheads. This increases the storage optimization compared to other approaches that result in efficient management of resources. In the aspect of [14], the drawbacks and bottlenecks in accessing the big data are overcome using the ETL concept, Big Query, and SQL formats. These tools enable the processing of big data in an optimized manner. From [15], it addresses many challenges like viewing through visual aids, different formatting, storage, security and rising issues such as hardware taking costly investment, time-consuming process to handle big data, etc. Mohan Naik Ramachandra (2022) demonstrates that TDES, which provides effective security over health data and the time taken to transmit encrypted and receive decrypted information, is less compared to the existing approach. The study of Ahmed Hadi Ali AL-Jumaili et al. (2023) demonstrates the challenges of big data in power management systems, results coming with new cloud computing with recommendations that address challenges over infrastructure and management. Neelay Jagani et al. (2021) address challenges over security, technologies, and applications when combined with both cloud computing and big data. The significant factor considered in this scenario is efficiency.

C. Stergiou et al. (2017) demonstrate algorithms in the combined technologies of big data and cloud computing to provide protection for privacy and enhance security. This contribution would analyze the repeated functionalities and provide security for them. Akansha Gautam et al. (2020) demonstrate the issues and challenges of big data and provide solutions when cloud computing is applied. Better analytics is applied to extract the information, make decisions, and refine the service.

Lingqi Xue (2021) demonstrates the algorithms for financial organizations that depend on wireless and network communication networks. The two proposed approaches of CoMP are IAN and IA variants, with measures such as spectrum, social utility, and re-transmission with computed accuracies and other metrics. Mahdavisharif, M. (2021) discusses big data, intrusion detection, LSTM, and SPARK framework to provide security against intrusions and unauthorized access. The designed method, called BDL-IDS, would produce better results than traditional methods.

2.1. Limitations of Solid State Drives (SSDs)

2.1.1. Higher Cost

SSDs are generally more expensive than traditional hard drives (HDDs), which can increase the overall cost of the system.

2.1.2. Limited Write Cycles

SSDs have a finite number of write and erase cycles, which can lead to wear and reduced lifespan under heavy write-intensive workloads.

2.1.3. Data Recovery Challenges

In the event of a failure, data recovery from SSDs can be more difficult compared to HDDs.

2.2. Limitations of In-Memory Systems

2.2.1. In-Memory Systems

Cost and Scalability: Storing large volumes of data in memory can be costly, as it requires significant investment in RAM. This can be a limitation for organizations with budget constraints or those needing to store vast amounts of data.

2.2.2. Data Persistence

In-memory systems can face challenges with data persistence, as data stored in RAM is volatile and can be lost in the event of a power failure.

2.2.3. Complexity

Implementing and managing in-memory systems can be complex, requiring specialized knowledge and expertise.

.

Table 8. Significant Studies over Big Data Environment

Theme	Study	Key Findings
Analysis and Processing of Bigdata	Chandrima Roy et al (2018)	The advantages of map reduction and Hadoop over big data are explored.
Opportunities, Challenges of Big Data	Hira Zahid et al (2019)	Telecommunications are evolving with respect to present challenges.
Big Data Applications	Naghib et al (2023)	Proposing a reliable and adaptable BDM framework using an IoT environment.
Frameworks and Technologies of Bigdata	C. Stergiou et al. (2017)	Combined cloud computing and big data issues on security and privacy.
Emerging Trends and Challenges	Agnieszka Smalec (2021)	Addressed the challenges of multidimensional data and hypermedia.

3. Proposed Methodology

When customization is applied, it benefits the big data system in a number of ways, such as balancing, which is possible over data processing, storage optimization, and communication. It supports scalability in terms of small datasets as well as growing data volumes. The resource allocation would be optimized by dynamically adjusting the computational power. It supports parallel processing and produces faster analysis for easy decision-making support. From Fig. 2, the modules such as DAP, SPFE, DPC, ARM, and CRC are used for performance and optimization.

The interaction between these modules is demonstrated in Figure 1. All five modules are defined through algorithms demonstrated in pseudo procedures PS1, PS2, PS3, and PS4. In this, DAP represents Data acquisition and preprocessing, where data is collected from various sources, and applying data cleaning as well as data transformation is to produce data ready for immediate analysis. The SPFE denotes Selective partitioning and feature engineering where partitioning is applied, then apply selective features are applied, and combine existing features are combined into new features for improving model performance.

3.1. Security Aspects in Big Data Environments

Big data environments are particularly vulnerable to various security threats due to the sheer volume and complexity of the data involved. The proposed framework addresses these threats through several key strategies:

3.1.1. Data Breaches and Unauthorized Access

Threat

Unauthorized access to sensitive data can lead to data breaches, compromising personal and financial information.

Solution

Implementing robust user access control mechanisms and encryption techniques ensures that only authorized personnel can access sensitive data. This includes multi-factor authentication and role-based access controls.

3.1.2. Data in Transit Vulnerabilities

Threat

Data transmitted across networks can be intercepted by malicious actors, leading to data leaks and breaches.

Solution

The framework employs advanced encryption protocols and secure communication channels to protect data in transit. Intrusion detection systems (IDS) and intrusion prevention systems (IPS) are used to monitor and secure network traffic.

3.1.3. Data Storage Security

Threat

Data storage systems, such as data lakes and warehouses, are prime targets for cyberattacks due to the valuable information they hold.

Solution

Utilizing encryption for data at rest, along with regular security audits and vulnerability assessments, helps protect stored data. Additionally, implementing access controls and monitoring tools ensures that any unauthorized access attempts are quickly detected and mitigated.

3.1.4. Insider Threats

Threat

Employees or insiders with access to sensitive data can intentionally or unintentionally cause data breaches.

Solution

The framework includes insider threat detection mechanisms, such as monitoring user behavior and implementing strict access controls. Regular training and awareness programs for employees also help mitigate the risk of insider threats.

3.1.5. Compliance and Regulatory Challenges

Threat

Non-compliance with data protection regulations can result in legal penalties and loss of reputation.

Solution

The framework ensures compliance with relevant data protection regulations, such as GDPR and CCPA, by incorporating data governance and compliance management tools. Regular audits and updates to security policies help maintain compliance.

The DPC denotes Distributed Processing and Caching where a distributed framework is applied for distributing across different machines over a cluster, frequently accessing data partitions must reside in RAM for immediate access, and then lazy evaluation is for processing only required data. The ARM denotes adaptive resource management, where uninterrupted processing is guaranteed using failure management, monitoring the resources, and then adding or removing them to achieve optimal performance. The CRC denotes communication and result consolidation, where data exchange between the processing units, partial results from different partitions are combined, and alerts are produced based on analysis results.

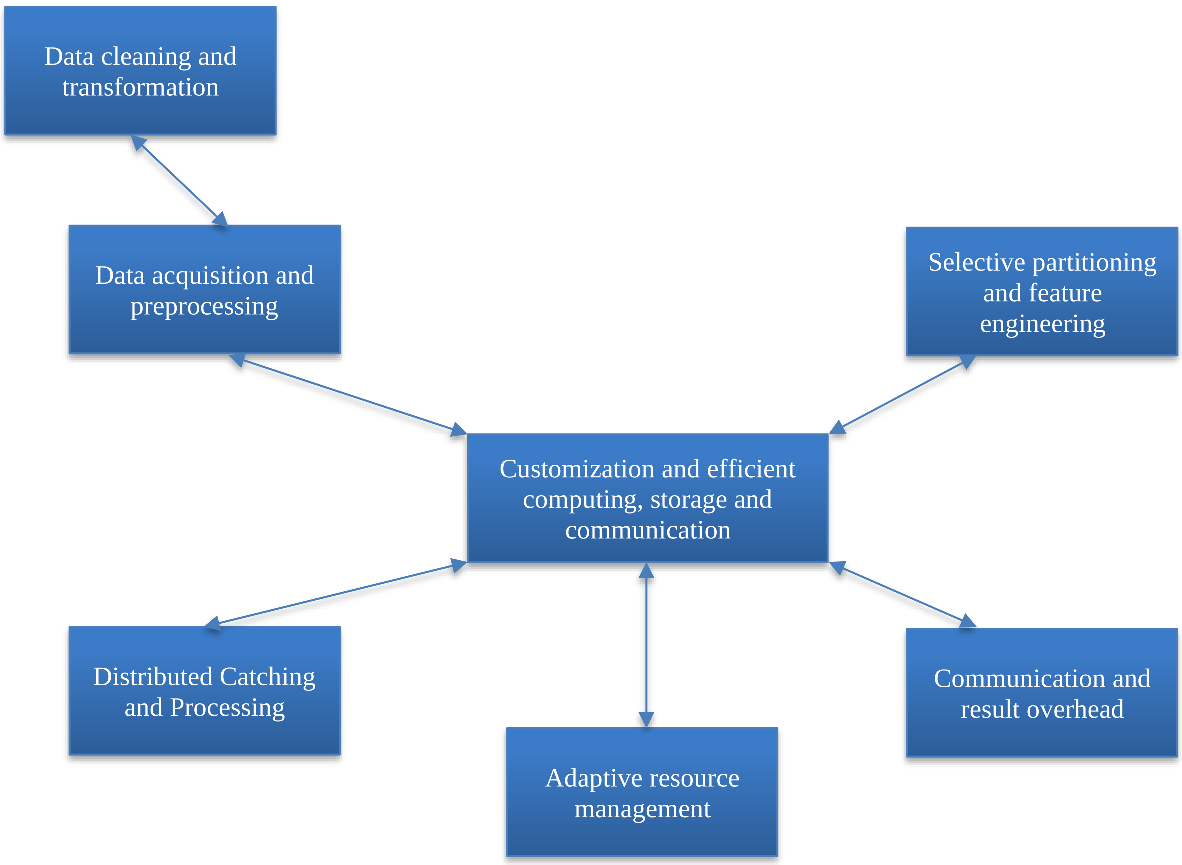

Fig. 1 Modules interaction in the proposed system

PS1: Pseudo_Procedure Computing_computation storage_communication ():
Input: Sources [], In_memory []
Output: Result [] []
Step 1: Identify the sources from which data is extracted, Call Computing_Bigdata ()
Step 2: Use caching for efficient data access and use distributed computing for parallel processing.
Call Computation_storage_Bigdata() for doing this task.
Step 3: For efficient communication between processing units, Call Communication_Bigdata ().
Step 4: Domain_specific_accuracy= Depends on (Missing_items, Inconsistencies, Errors)
Step 5: If (Domain_specific_accuracy==0):
Alert (Accuracy is 100%)
Else if (Domain_specific_accuracy<=5)
Alert (Accuracy is 95%)
Else
Alert (Accuracy is 90%)
Step 6: Domain_specific_Performance = time taken to process dataset
Step 7: Cost-effectiveness = Initial setup investment + cost taken for long-run maintenance.
Step 8: Visualize the Domain_specific_accuracy, Domain_specific_Performance, cost-effectiveness against existing methods

From PS1, the two metrics are evaluated, such as accuracy and performance.

Fig. 2 ER diagram for Customized computing, computation storage, and communication

The below algorithms PS1, PS2, PS3, and PS4 define the workflow of activities that enhance the performance over the big data and its optimization:

PS2: Pseudo_procedure Computing_Bigdata ():
Input: Sources []
Output: Normalized_dataset
Step 1: Execute DAP () for the following three steps
Step 1.1: Apply tools to collect data efficiently
Step 1.2: Set criteria as per context to eliminate irrelevant or duplicate data
Step 1.3: Apply scaling over existing data to prepare it for analysis
Step 2: Execute SPFE () for the following three steps
Step 2.1: Divide data into independent chunks for parallel processing
Step 2.2: Identify the most important attributes for efficient analysis
Step 2.3: Combine a few features into the new feature to improve performance.

In PS2, two techniques will be applied over the source normalization and enhance the quality.

PS3: Pseudo_procedure Computation_storage_Bigdata ():
Input: In_memory []
Output: Prepared dataset for analysis
Step 1: For efficient access, execute DPC () for the following three steps.
Step 1.1: Use software like Hadoop or Spark to process tasks from multiple machines in a cluster.
Step 1.2: Caching frequently access data so that faster access is guaranteed.
Step 1.3: Use lazy evaluation for processing only the data that is needed
Step 2: For scaling and failure management, execute ARM() for the following two steps.
Step 2.1: Apply scaling to add or remove machines to maintain optimal performance.
Step 2.2: Implement failure management where tasks are re-assigned to ensure uninterrupted processing.

From PS3, the two techniques are applied, and the dataset is produced with immediate analysis.

PS4: Pseudo_procedure Communication_Bigdata ():

Input: consolidated_dataset
Output: Decisions
Step1: For efficient communication, execute CRC () for the following three steps
Step 1.1: Apply protocols for data exchange between distributed tasks.
Step 1.2: Combine partial results from different partitions to obtain the final analysis.
Step 1.3: Activate notifications on results obtained for further decision-making.

The diagram below demonstrates the flow of activities that makes the user ready to make efficient decision-making for their further improvement.

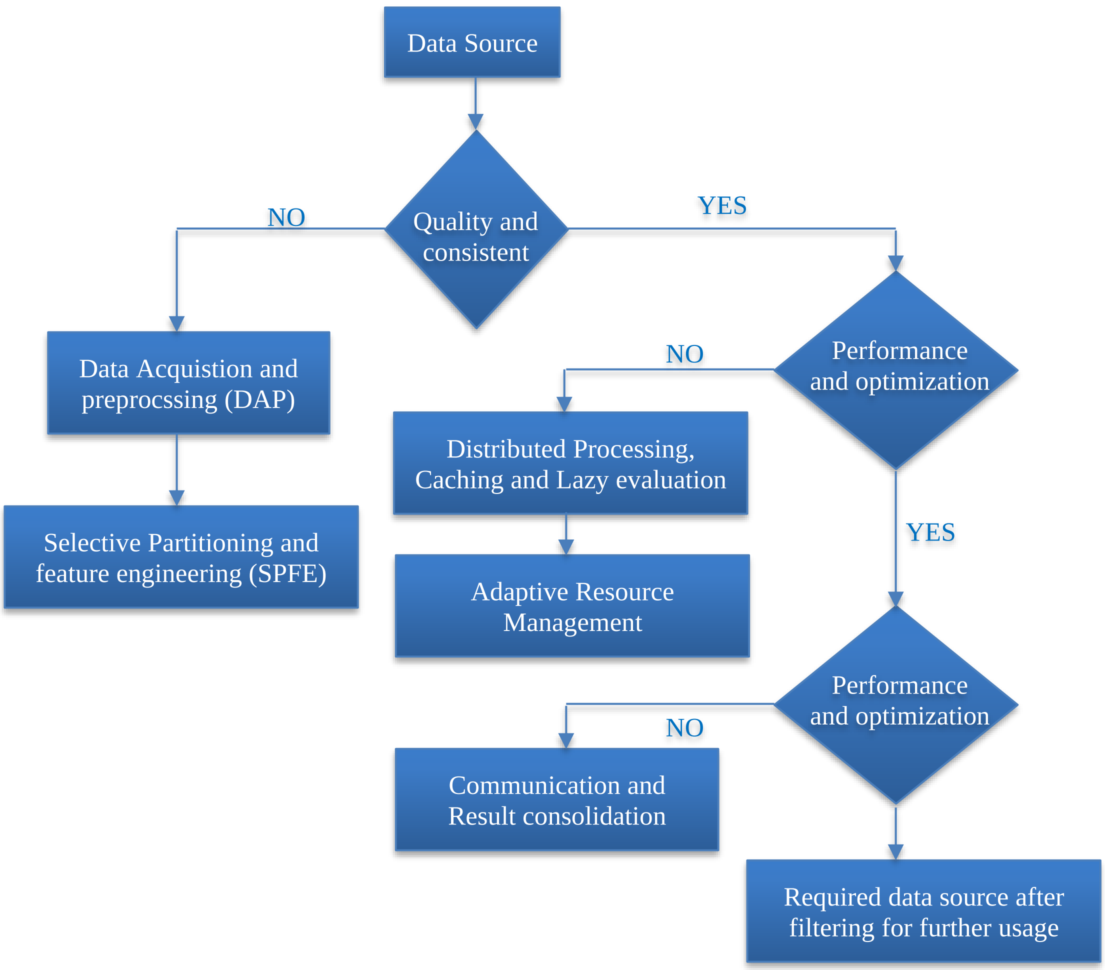

Fig. 3 Flow Diagram of an efficient, customized approach

4. Results

In this, accuracy and performance are considered as metrics to evaluate big data processing. The three aspects considered are computing, storage, and communication in big data processing. Tables 9, 10, 11, and 12 denote in order such strategies used in computing over big data, storage over big data, and communication over big data and their relation against specific metrics.

The computing strategies are measured against accuracy, performance, and cost-effectiveness in which varied specifies depends on the scenario, positive specifies better or outperform results, High means more expensive, and neutral specifies not effecting anything.

Table 9. Metrics impact during computing methods

Method	Accuracy	Performance	Cost-effectiveness
Distributed frameworks	Positive	Positive	Varied
In-memory systems	Positive	Positive	High
Lazy evaluations	Positive	Positive	Positive
Partitioning	Neutral	Positive	Positive
Selective algorithm	Positive	Positive	Positive

Table 10. Metrics impact during computation storage methods

Method	Efficiency	Scalability	Cost-effectiveness
Hardware Tools	Varied	Varied	Varied
Software tools	Positive	Positive	Varied
Orchestration tools	Positive	Positive	Positive
Analytics tools	Positive	Varied	Varied
Data management tools	Positive	Positive	Positive

Table 11. Metrics impact over communication methods

Method	Bandwidth	Latency	Reliability
Infrastructure	Varied	Varied	Positive
Serialization	Positive	Neutra	Neutral
Protocols	Varied	Varied	Positive
Notification/alerting systems	Varied	Varied	Positive

Tables 9, 10, 11, and 12 demonstrate the measures used over methods of computing, storage, and communication. Their impact is demonstrated in terms of fetching and changing nature. The computation of percentages over the above metrics for two methodologies such as customized and traditional, are demonstrated in Table 13 and is demonstrated in Figure 4. Over a 1-point scale, the evaluation is demonstrated.

Table 12. Metrics impact over communication storage methods

Method	Scalability	Efficiency
Infrastructure	Positive	Positive
Serialization	Neutral	Positive
Protocols	Positive	Positive
Notification/alerting systems	Positive	Varied

Table 13. Customized approach vs Traditional approaches

Method	Accuracy (%)	Performance (%)	Cost-effectiveness (%)
Customized Approach	0.99	0.99	0.95
Traditional Approaches	0.85	0.90	0.85

According to Table 13, the customized method guarantees performance and accuracy. For which determines another measure is the cost, which is moderate in the long run, although the initial setup takes more cost, whereas the traditional method consumes the initial setup, which is low, but maintenance is expensive in the long run. In Table 14, the evaluation is demonstrated over a 1-point scale.

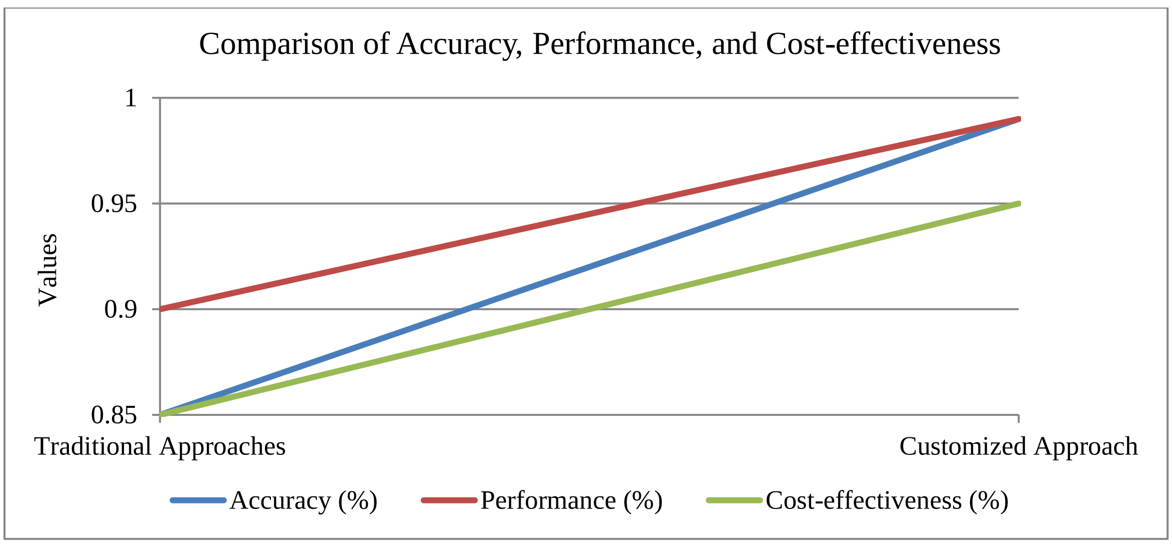

Fig. 4 Accuracy, Performance of customized approach against Traditional method

This research addresses the critical need for an integrated optimization framework in big data environments, focusing on computing, computation storage, and communication. Traditional approaches often fall short in managing the computational and communication resources required for efficient big data processing, leading to inefficiencies and bottlenecks. Our proposed framework offers a novel solution by integrating advanced techniques across these three crucial aspects. In the context of ever-growing data volumes, velocities, and varieties, this research contributes significantly to the field of big data processing. It offers a balanced view by acknowledging the limitations and constraints of the technologies used while highlighting the substantial benefits of the integrated optimization framework.

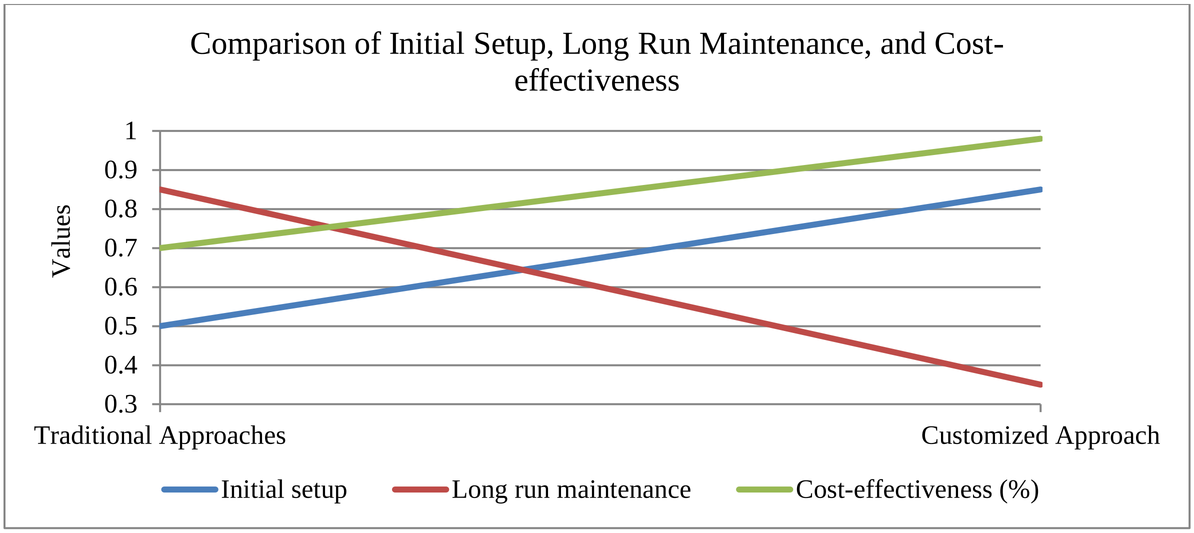

Fig. 5 Cost-effectiveness of efficient customized vs traditional

Table 14. Cost-effectiveness of efficient customized vs traditional

Method	Initial setup	Long run maintenance	Cost-effectiveness (%)
Customized Approach	0.85	0.35	0.98
Traditional Approaches	0.5	0.85	0.7

From Figure 5, it is clearly stated that an efficient, customized approach produces less maintenance compared to the traditional approach.

5. Conclusion

From a variety of aspects, big data is identified as the source of processing and guarantees the optimization of processing. There are five modules involved for efficient processing and optimization that are such as DAP, SPFE, DPC, ARM, and CRC. The purpose of these includes Data Acquisition and Preprocessing (DAP), where data is collected from various sources, data cleaning is applied, data transformation is done for quality datasets, and effective analysis is performed. Then, Selective Partitioning and Feature Engineering (SPFE), where partitioning is applied, then apply selective features, and combine existing features into new features to achieve performance. Then, Distributed Processing and Caching (DPC), where a distributed framework is applied across different machines over a cluster, In-memory computation is applied for immediate access of frequent items, apply lazy evaluation is for processing only required data. Then, Adaptive Resource Management (ARM), where uninterrupted processing is guaranteed using failure management, adding or removing resources in order to achieve optimal performance. Then, Communication and result consolidation, where data exchange between the processing units combines partial results from different partitions and produces alerts based on analysis results. This overall scenario would guarantee accuracy, performance, and cost-effectiveness using customized methods.

References

[1] Chandrima Roy, Siddharth Swarup Rautaray, and Manjusha Pandey, "Big Data Optimization Techniques: A Survey," *International Journal of Information Engineering and Electronic Business*, vol. 10, no. 4, pp. 41-48, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[2] Mengxuan Wu, Jingjing Jiang, and Lijuan Wang, "Research on the Optimization Algorithm of Big Data Computing System," *2021 International Wireless Communications and Mobile Computing (IWCMC)*, Harbin City, China, pp. 1783-1787, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] Hira Zahid et al., "Big Data Analytics in Telecommunications: Literature Review and Architecture Recommendations," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 1, pp. 18-38, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Agnieszka Smalec, "Big Data as a Tool Helpful in Communication Management," *Procedia Computer Science*, vol. 192, pp. 5156-5165, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[5] Arezou Naghib et al., "A Comprehensive and Systematic Literature Review on the Big Data Management Techniques in the Internet of Things," *Wireless Networks*, vol. 29, pp. 1085-1144, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Madhavi Vaidya, Shrinivas Deshpande, and Vilas Thakare, "Design and Analysis of Large Data Processing Techniques," *International Journal of Computer Applications*, vol. 100, no. 8, pp. 24-28, 2014. [Google Scholar] [Publisher Link]

[7] Sandeep Dasari, and Rajesh Kaluri, "Big Data Analytics, Processing Models, Taxonomy of Tools, V's, and Challenges: State-of-Art Review and Future Implications," *Wireless Communications and Mobile Computing*, vol. 2023, pp. 1-14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Rahat Iqbal et al., "Big Data analytics and Computational Intelligence for Cyber–Physical Systems: Recent Trends and State of the Art Applications," *Future Generation Computer Systems*, vol. 105, pp. 766-778, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Ahmed Hadi Ali AL-Jumaili et al., "Big Data Analytics Using Cloud Computing Based Frameworks for Power Management Systems: Status, Constraints, and Future Recommendations," *Sensors*, vol. 23, no. 6, pp. 1-37, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Qingqing Chang, Shah Nazir, and Xia Li, "Decision-Making and Computational Modeling of Big Data for Sustaining Influential Usage," *Scientific Programming*, vol. 2022, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Shubham Upadhyay et al., "Analytics and Storage of Big Data," *Proceedings of the International Semantic Intelligence Conference* (ISIC 2021), New Delhi, India, pp. 202-210, 2021. [Google Scholar] [Publisher Link]

[12] Cheng Luo, "Computer Data Storage and Management Platform Based on Big Data," *Journal of Physics: Conference Series*, vol. 2066, pp. 1-6, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Blend Berisha, Endrit Meziu, and Isak Shabani, "Big Data Analytics in Cloud Computing: An Overview," *Journal of Cloud Computing*, vol. 11, pp. 1-10, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] Amanpreet Kaur Sandhu, "Big Data with Cloud Computing: Discussions and Challenges," *Big Data Mining and Analytics*, vol. 5, no. 1, pp. 32-40, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Mohan Naik Ramachandra et al., "An Efficient and Secure Big Data Storage in Cloud Environment by Using Triple Data Encryption Standard," *Big Data and Cognitive Computing*, vol. 6, no. 4, pp. 1-20, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Neelay Jagani, and Parthil Jagani, "Big Data in Cloud Computing: A Literature Review," *International Journal of Engineering Applied Sciences and Technology*, vol. 5, no. 11, pp. 185-191, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[17] Christos Stergiou, and Kostas E. Psannis, "Algorithms for Big Data in Advanced Communication Systems and Cloud Computing," *2017 IEEE 19th Conference on Business Informatics (CBI)*, Thessaloniki, Greece, pp. 196-201, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[18] Akansha Gautam, and Indranath Chatterjee, "Big Data and Cloud Computing: A Critical Review," *International Journal of Operations Research and Information Systems*, vol. 11, no. 3, pp. 19-38, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[19] Lingqi Xue, "Financial Big Data Based on Internet of Things and Wireless Network Communication," *Wireless Communications and Mobile Computing*, vol. 2021, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] Mahzad Mahdavisharif, Shahram Jamali, and Reza Fotohi, "Big Data-Aware Intrusion Detection System in Communication Networks: A Deep Learning Approach," *Journal of Grid Computing*, vol. 19, 2021. [CrossRef] [Google Scholar] [Publisher Link]