

# Big Data Optimization Techniques: A Survey

**Chandrima Roy**

Kiit University, Bhubaneswar, India  
Email: chandrima.roy.1914@gmail.com

**Siddharth Swarup Rautaray and Manjusha Pandey**

Kiit University, Bhubaneswar, India  
Email: { siddharthfcs, manjushafcs } @kiit.ac.in

Received: 13 November 2017; Accepted: 19 January 2018; Published: 08 July 2018

**Abstract**—As the world is getting digitized the speed in which the amount of data is over owing from different sources in different format, it is not possible for the traditional system to compute and analysis this kind of big data for which big data tool like Hadoop is used which is an open source software. It stores and computes data in a distributed environment. In the last few years developing Big Data Applications has become increasingly important. In fact many organizations are depending upon knowledge extracted from huge amount of data. However traditional data technique shows a reduced performance, accuracy, slow responsiveness and lack of scalability. To solve the complicated Big Data problem, lots of work has been carried out. As a result various types of technologies have been developed. As the world is getting digitized the speed in which the amount of data is over owing from different sources in different format, it is not possible for the traditional system to compute and analysis this kind of big data for which big data tool like Hadoop is used which is an open source software. This research work is a survey about the survey of recent optimization technologies and their applications developed for Big Data. It aims to help to choose the right collaboration of various Big Data technologies according to requirements.

**Index Terms**—Big Data, Hadoop, Optimization, Scalability.

## I. INTRODUCTION

Data is one of the most important and vital aspect of different activities in today's world. Therefore vast amount of data is generated in each and every second. A quick development of data in current time in different domains requisite an intellectual data analysis tool that would be useful to fulfill the requirement to analysis a vast quantity of data.

### A. Big Data

The data which are beyond the storage space of the server and beyond to the processing power is called Big Data. It is not manageable by traditional [4] RDBMS [16] or conventional statistical tools. Big data Increases the

storage capacities as well as the processing power. Market prediction & forecasting can be done so easily using big data. Big Data is going to change the world completely. Figure 1 demonstrates the several characteristics of Big Data, known as the 9V's of Big Data.

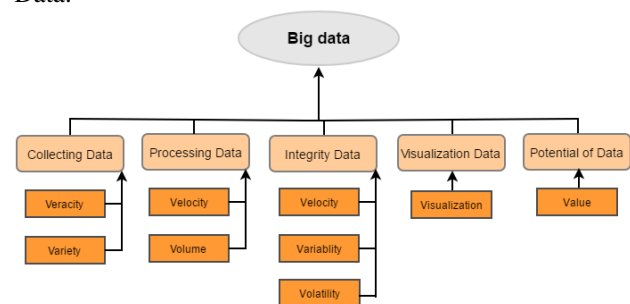


Fig.1. Big Data with 9V's Characteristics

Volume is the amount of data generated every day. The sources may be internal or external. Velocity is how fast data is processed, batch and streaming data. Variety deals with a wide range of data types and sources of data. This is generally studied under three categories: Structured data, Semi-structured data, and unstructured [18] data. Veracity refers to the uncertainty of data, i.e., whether they obtained data is correct or consistent. Out of the huge amount of data that is generated in almost every process, only the data that is correct and consistent can be used for future analysis. Value is the identification of data which is valuable than transformed and analyzed.

### B. Issues in Big Data

Big data issues are real challenges on which action should be taken as soon as possible. Without handling these challenges, it will lead to failure and unpleasant result.

**Privacy and Security:** One of the important challenges of big data it includes legal, technical as well as computational [8] significance. Some time with large data-sets personal information of the people is also added which conclude new facts about that personal information of the people are also added. Which conclude new facts about that person and it may be possible the fact about that person is secretive and should not be shared by data owner or that person.

*Accessing and Sharing of Information:* If the data in the enterprises are used to create precise decision it should be complete accurate and in timely manner, which increases the complexity of governance process and data management.

*Analytical Challenges:* One of the key job of big data is to do analysis on huge data. These data are structured, semi-structured and unstructured and advanced skills is needed to process these kind of data, more over decision is made upon the data which is obtained that is decision making.

*Scalability:* Big data scalability issue leads to cloud computing. There is a high end resources sharing involved in it which is expensive and also challenges to efficiently run various jobs so that the goal of each work load cost effectively. System failure is also dealt efficiently while working on large data clusters. These combined factors make it difficult to write program, even machine learning complex tasks. Changes are being made in technologies used, solid state drive is used in place of hard drive and phase change technologies are not performing as good as sequential and random transfer. Thus it's a big question that what kind of storage devices should be used to store data.

*Quality of Data:* Collecting and storing huge amount of data are costly; if more data is stored for predictive [12] analysis and decision making in business then it will yield better results. Big data has interest in having quality rather than unused data so that better conclusion and result can be drawn. This further arise the question on whether the data is relevant or not, can the stored data be accurate enough to produce right result, how much data would be enough in making decision.

*Heterogeneous Data:* In recent days data there is increase in unstructured data from social media, to record meetings, emails, fax transfer, black box data and more. Working with unstructured data is costly and difficult to store. Converting unstructured data into structured is not easy. Structured [18] data is highly organized and well managed data on the other hand unstructured data is unorganized and raw.

*Optimization:* Finding the alternative way to achieve the highest performance in cost effective manner and under given limitation by mostly utilizing the desired factors. In comparison, maximization is the process of attaining highest or maximum performance deprived of cost or expenditure. Practice of optimization is restricted when there is shortage of information and lack of time to estimate what information is obtainable.

### C. Motivation

The world is changing rapidly with time it is growing towards the hi-tech era where technology is playing a great role. As the world is rapidly pacing into hi-tech era one thing has also grown tremendously within a few years that is data. Today data has rose from terabytes to zettabytes of data. As data is growing tremendously in a rapid velocity and different varieties and structure, efficient tools are needed to organize these huge data known as big data. Some of the tools are already playing

vital roles in efficiently working with big data, one such tool is hadoop. As hadoop is a new technology there are still some areas in hadoop which can be better optimized to increase the efficiency and throughput of hadoop ecosystem.

### D. Big data Applications

Some of the major applications of big data have been summed up in the following section and also represented diagrammatically in Figure 2.

*Big data in Banking:* Big Data has delivered opportunities to bank to understand the scenario due to maintain the re-active nature of the data for convey significance to users along with the confidentiality and safety of sensitive information.

*Big data in Finance sector:* Financial amenities have embraced big data analytics [14] to notify improved investment decisions with steady revenues. Big data is being used in a collective number of applications, such as employee observing and surveillance, developing algorithms to estimate the track of financial markets.

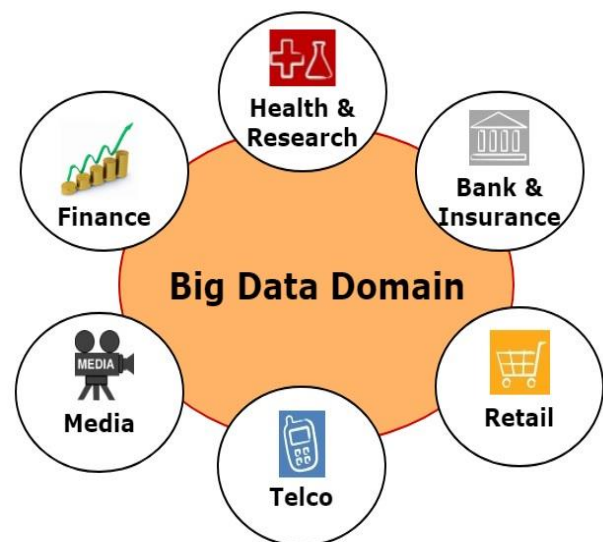


Fig.2. Application of big data

*Big data in Telecom:* Telecom companies require a proper searching and analysis [15] of data to get deeper understanding into customer behavior, their service usage preferences, patterns and real-time interests. Here is where Big Data comes in.

*Big data in retail sector:* Merchants use Big Data to provide customer custom-made shopping involvements. Exploring how a customer came to make a purchase, or the road to purchase, is one way big data techiest making a spot in marketing. 66% of vendors have made monetary improvements in customer relationship organization through big data.

*Big data in HealthCare:* The data coming from electronic medical care (EMR) system includes both unstructured and structured [17] data from pathology reports, physician notes etc. These data are analyzed with the target to improving patient health care, reducing expenses and improve patient's health.

II. HADOOP ECOSYSTEM

NameNode control the HDFS. It does not hold any actual file data; rather it holds all file system metadata for the cluster. Such as name of the file and its path, what blocks makeup a file and where those blocks are located in the cluster. It holds the health of each and every data node. Backup node or secondary name node connects to the name node (occasionally) and take a copy metadata store into Name Node’s memory. If the NameNode dies, the files preserve by secondary NameNode used to recover the NameNode. DataNode sends heartbeats to the NameNode every 3 seconds to inform the NameNode that it is still alive. Every 10<sup>th</sup> heart beat is a block report, DataNode tells the NameNode about all the blocks contains within it.

Node Manager takes care of the computational nodes in the Hadoop cluster. It determines the health of the executing nodes. It keeps a track of the resources available and sends the track report to Resource manager. Journal Node is used to keep the state synchronized between the NameNode and Backup Node or standby NameNode. When any modification is performed by the NameNode, it informs the journal node about the modification. The secondary node is able to read the update from the Journal node.

A. Scalability

Scalability is a characteristic of a system or function that defines its ability to manage and accomplish under an improved or increasing load. A system that scales well will be capable to keep or even increase its level of performance or productivity when verified by bigger operational demands. Figure 3 demonstrate the Vertical and Horizontal Scaling.

Vertical Scalability:

It is the process of increasing the processing power and capacity of the hardware or software by adding several resources to make it work more efficiently and faster.

Pros	Cons
All data is in a particular device. No need to manage numerous entities.	The problem is the cost efficiency. A powerful machine having huge number of CPU and higher RAM capacity is expensive than a set of small size entities.
Reduced software costs.	Limited scope of upgradeability in the future.
Implementation isn’t difficult.	The overall cost of implementing is really expensive.

Horizontal Scalability:

It is the capability to link several individuals so that they perform as a solo logical entity.

Pros	Cons
All is in smaller pieces so program can process them very fast with parallel job distribution among all entities.	The problem is about managing those instances and the complex distributed architecture.
Easy to upgrade	The licensing fees are more
Much cheaper compared to scaling-up	The licensing fees are more

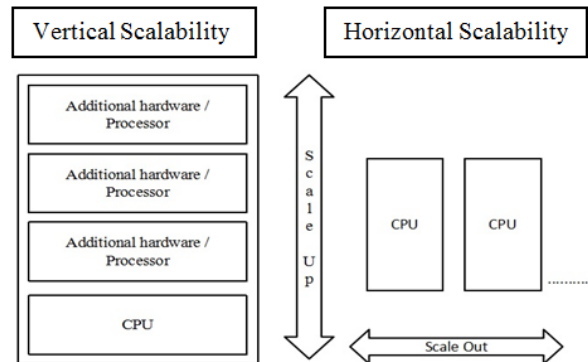


Fig.3. Vertical and Horizontal Scaling

B. Processing capabilities

SQL:

Structured Query Language is used to be in communication with a database. "Select", "Insert", "Update", "Delete", "Create", and "Drop" are the standard SQL commands which are used to accomplish transactions with a database.

No SQL (Not Only SQL):

These are non-relational, open source, distributed databases. Structured, unstructured and semi-structured data can be deal with the help of No SQL. It has scale out architecture instead of the monolithic architecture of relational databases. Insertion of data in the NO SQL database, do not need a predefined schema. This supports faster development, easy code integration and require less database administration.

C. Optimization

Optimization is the act of design and developing systems in such a way that it can take greatest advantage of the available resources represented in Figure 4. Optimization of applications [23] can be done to take advantage from the huge amount of memory space present on a specific computer, or the hardware speed, or the processor being used. Optimization is to done to attain the finest strategy relative to a set of selected constraints which include maximizing factors such as efficiency, productivity, reliability, longevity, strength, and utilization.

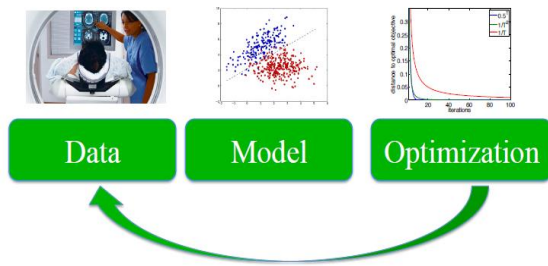


Fig.4. Optimization process

While converged infrastructures in general provide easier and faster scalability than traditional architectures [6], the optimum environment for Big Data analytics is a system that allows you to scale computing power separately from storage. The idea is to be able to scale storage and compute nodes separately. Aim for as much memory and as little cost and power, as possible, he adds,

since companies doing analytics will scale out to hundreds or even thousands of nodes.

D. Mapreduce

Hadoop accomplishes its operations with the help of the MapReduce model, which comprises two functions namely, a mapper and a reducer [22].

The Mapper function is responsible for mapping the computational subtasks to different nodes and also responsible for task distribution, load balancing and managing the failure recovery. The Reducer function takes the responsibility of reducing the responses from the compute nodes to a single result. The reduce component has the responsibility to aggregate all the elements together after the completion of the distributed computation. The overall function of the map-reduce process is explained in Figure 5.

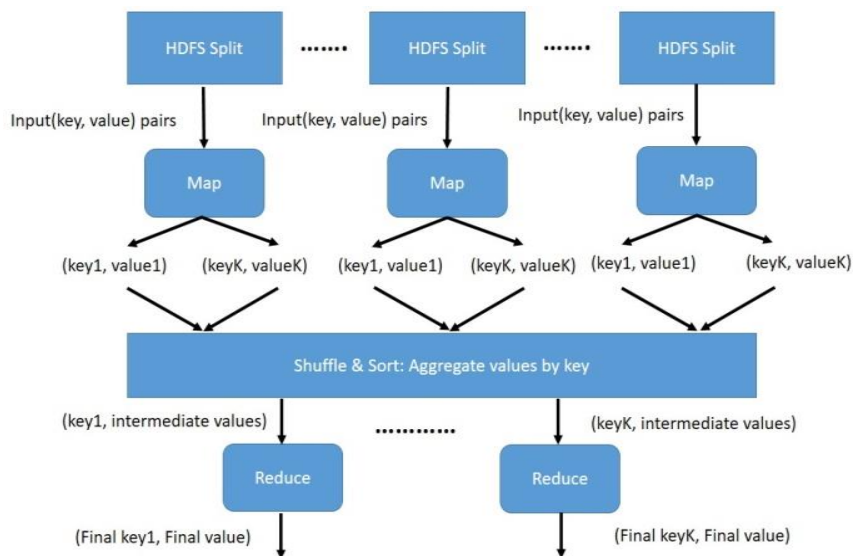


Fig.5. Overview of the Map-Reduce process

III. RELATED WORKS

The below Table [1] present a brief idea about some optimization process and big data analytics. It also discusses about the characteristics of various optimization techniques and how these techniques are helpful for improving the decision making process in big data domains.

The above Table 1 describes the objective of the paper like some paper have worked on the optimization area of big data tools, the methods used to do optimization such as some techniques which have been defined in this paper such as Flex allocation scheduler, Recursive Chunk Division (RCD) for Optimal Pipelining, Optimal

Parallelism-Concurrency-Pipelining (PCP), Merging of several small files and prefetching for structurally-related small files [21] and grouping of files and prefetching for logically-related small files and the application domain in which these have been implemented.

The above Table 2 describes the survey paper and their respective technique and algorithm used to achieve the objective and the application in which it has been performed.

The Presented survey paper has analyzed the previously explained research work according to the emphasis given on the enhancement techniques provided on it, represented in the Table 3.



Table 1. Optimization Implemented

Author	Objective	Technique/ Algorithm	Application
Joel Wolf et.al[6]	FLEX: A Slot Apportionment Scheduling Optimizer for MapReduce Jobs	Flex allocation scheduler	Job Scheduling
Bo Dong et.al[7]	Determine an optimized method for loading and retrieving small files on cloud storage	Merging of several small files and prefetching for structurally-related small files and grouping of files and prefetching for logically-related small files	Small File distribution
Esma Yildirim et al.[8]	Application-Level Optimization through Pipelining, Parallelism and Concurrency.	Recursive Chunk Division (RCD) for Optimal Pipelining, Optimal Parallelism-Concurrency-Pipelining (PCP)	Map reduce optimization
Maumita Bhattacharya et.al[9]	An evolutionary algorithm is used to enhance the ability to cope up with the difficulties of great dimensionality and scatteredness of data.	The recommended prototype uses informed genetic operators to present diversity by increasing the range of search procedure at the cost of redundant less promising associates of the population.	Ensuring constructive population diversity
Kostas Kolomvatsos et.al [10]	Effective Time Optimized Arrangement for Advanced Analytics in Big Data.	Two consecutive decision building prototypes are used to conduct the entering partial results. The first prototype is centered on a finite horizon time-optimized model and the second one is built on an infinite horizon optimally scheduled model.	Improve the performance of querying big data clusters.

Table 2. Optimization Survey

Author	Objective	Key fields
Marisiddanagouda. M et al.[7]	Improve the performance of hadoop mapreduce, overcome the limitations of the Map-Reduce framework, decrease the performance degradation because of interaction and heavy dependency across different MapReduce phases	Optimization methods, MapReduce Performance.
Dilpreet Singh et al.[13]	Evaluates the advantages and drawbacks of different hardware platforms available for big data analytics.	Graphics processing units, scalability, k-means clustering, real-time processing, big data platforms
Shivaraj B. G. et al. [11]	Handle Multiple Jobs in Hadoop Cluster. Resources handled by MapReduce schedulers assigning resources in the form of MapReduceTasks.	HDFS, MapReduce, Schedulers optimization
Sunith Bandaru et al.[1]	Knowledge detection in multi-objective optimization using data mining techniques present real-world optimization difficulties where numerous objectives are optimized at the same time with respect to several variables.	Multi-objective optimization, Descriptive statistics, Visual data mining process, Machine learning, Knowledge-driven optimization
Hao Zhang et al.[18]	Design ideologies for in-memory data organization, handling, and practical procedures for planning and implementing efficient and high-performance in-memory systems.	Primary memory, DRAM, relational databases, distributed databases, query processing

Table 3. Enhancement Techniques used in Optimization Process

References	Process Capability Enhancement	Memory Management Enhancement	Map Reduction Enhancement	Data Node Enhancement	Name Node Enhancement
1	✗	✗	✓	✗	✗
2	✓	✗	✗	✗	✗
3	✓	✗	✗	✗	✗
4	✓	✗	✗	✗	✗
5	✓	✗	✗	✗	✗
6	✗	✗	✓	✗	✗
7	✗	✓	✗	✗	✗
8	✓	✗	✗	✗	✗
9	✗	✗	✗	✓	✗
10	✗	✗	✓	✗	✗

IV. ANALYSIS

A. Process Capability Enhancement

Application-Level Optimization: As per the research on Big Data Application-Level Optimization [17] Transfers over Pipelining, Parallelism and Concurrency, process capability enhancement has been done to improve

performance. Application-level transfer alteration parameters just like pipelining, parallelism and concurrency are very needful appliances for come across data transmission bottlenecks for scientific cloud applications. These parameters can optimize to achieve supreme transmission rate. The research proposes the optimization algorithms via these guidelines and representations can offer a regular growth to the

maximum throughput on inter-cloud [20] and intra cloud transfers.

*Evolutionary optimization:* The following research on Evolutionary optimization emphasis on process capability enhancement. The proposed model introduces diversity by using informed genetic operators. The research also introduces an algorithm to deal with the high

dimensionality problem. The POPULATION\_EA [2] algorithm introduce in this study mainly aim to manage advanced dimensional problem field. The POPULATION\_EA prototype is capable in handling high dimensional optimization problems including complex multi model solution space.

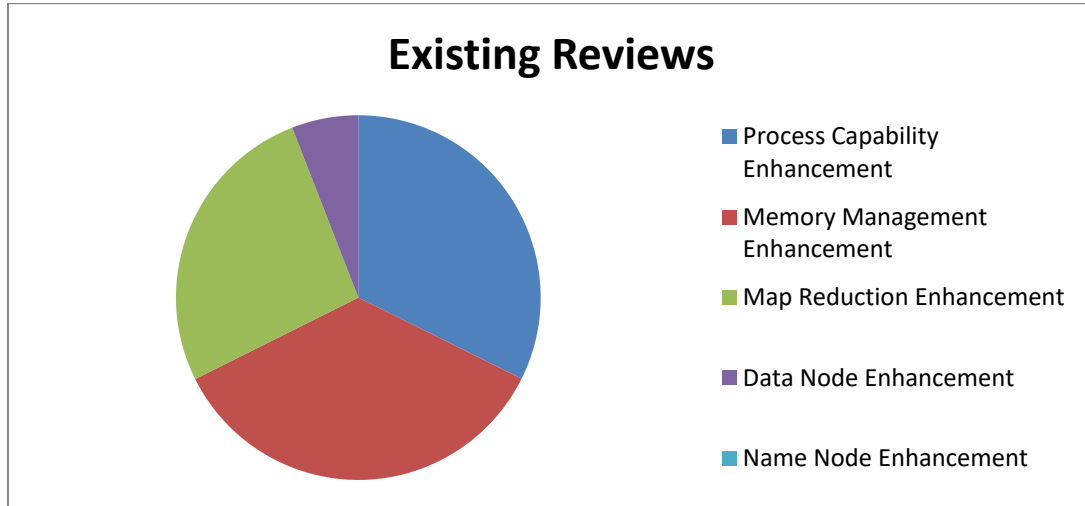


Fig.6. Statistics of Survey on Various Enhancement Techniques of Big Data

*Optimized Job Allocation Scheme:* Flex, a flexible and intelligent distribution system for MapReduce loads. It is flexible in the logic that it can enhance any of standard scheduling metrics, such as average response time, stretch, deadline-based penalty functions [19], and even slas. It is intellectual in the sense that it can attain performance near to the theoretical Optimal.

#### B. Memory Management Enhancement

*In-Memory Big Data Optimization:* This survey on In-Memory Big Data Management and Processing illustrate Memory Management Enhancement. The research delivers a complete analysis of essential technology in memory management [18] and exploration of associated works. This survey dedicated on the design values for in-memory data management and handling, and practical methods for planning and executing effective and operational in-memory systems.

#### C. Map Reduction Enhancement

*Platform optimization according to Big Data:* The choice to select a specific platform for a specific application typically depends on some factors: data size, speed or throughput optimization and model development. There are numerous big data platforms [13] present with various features and selecting the exact platform needs in-depth information about the skills of all these platforms.

#### D. Data Node Enhancement

*Knowledge discovery in multi-objective optimization:* The survey on data mining techniques for knowledge discovery in multi-objective optimization signifies data node enhancement. Maximum multi-objective optimizers

[1] start with a population of arbitrarily produced results or entities, which endure deviation and assortment iteratively, until a definite number of generations or performance target or function estimations is touched.

#### E. Name Node Enhancement

*Small files optimization:* Bo Dong et.al done a research on storing and fetching small file where the author design an enhanced method to increase the storage and fetching proficiencies of small files on HDFS. Merging of several small files and prefetching for structurally-related small files and grouping of files and prefetching for logically-related small files have been proposed in this research.

## V. CONCLUSION

In this survey paper various optimization techniques has been presented using big data tools. It is also esteemed the existing optimization technologies, mechanism and techniques of big data framework. The individuality of this paper is that this paper gives a summary of several methods and highlights most of the substantial outcomes of existing research which is presented concisely in above tables. The paper also highlights most of the major research challenges and issues related with the existing skills. Figure 6 demonstrate that most importance has been given to process capability enhancement by the researcher. As we can perceive from the Table, the most of the research solved Hadoop problems that are independent of data. Most of the examined methods deal with utilization, fairness, and starvation difficulties. This survey will be helpful for the further improvement and development of

Big Data Analytics. In the next phase name node enhancement will be done and ensured that the computational time will be optimized and reduced storage overhead will be attained.

#### ACKNOWLEDGMENT

I express my profound gratitude to the Dean of School Of Computer Engineering, KIIT University Dr. Samaresh Mishra for allowing me to proceed with the report and also for giving me full freedom to access the lab facilities. My heartfelt thanks to Dr. Manjusha Pandey & Dr. Siddharth Swarup Rautaray for taking time and helping me through my work. They have been a constant source of encouragement without which the work might not have been completed on time. I am very grateful for their support. Their ideas and thoughts have been of great importance.

#### REFERENCES

- [1] Bandaru, S., Ng, A. H., and Deb, K. Data mining methods for knowledge discovery in multi-objective optimization: Part a-survey. *Expert Systems with Applications* 70 (2017), 139-159.
- [2] Bhattacharya, M., Islam, R., and Abawajy, J. Evolutionary optimization: a big data perspective. *Journal of network and computer applications* 59 (2016), 416-426.
- [3] Dong, B., Zheng, Q., Tian, F., Chao, K.-M., Ma, R., and Anane, R. An optimized approach for storing and accessing small files on cloud storage. *Journal of Network and Computer Applications* 35, 6 (2012), 1847-1862.
- [4] Gu, R., Yang, X., Yan, J., Sun, Y., Wang, B., Yuan, C., and Huang, Y. Shadoop: Improving mapreduce performance by optimizing job execution mechanism in hadoop clusters. *Journal of parallel and distributed computing* 74, 3 (2014), 2166-2179.
- [5] Hua, X., Wu, H., Li, Z., and Ren, S. Enhancing throughput of the hadoop distributed file system for interaction-intensive tasks. *Journal of Parallel and Distributed Computing* 74, 8 (2014), 2770-2779.
- [6] Kolomvatos, K., Anagnostopoulos, C., and Hadjiefthymiades, S. An efficient time optimized scheme for progressive analytics in big data. *Big Data Research* 2, 4 (2015), 155-165.
- [7] Mr. Marisiddanagouda. M, M. R. M. Survey on performance of hadoop map-reduce optimization methods. *International Journal of Recent Research in Mathematics Computer Science and Information Technology* 2 (2015), 114-121.
- [8] Nagina, D., and Dhingra, S. Scheduling algorithms in big data: A survey. *Int.J. Eng. Comput. Sci* 5, 8 (2016).
- [9] Nghiem, P. P., and Figueira, S. M. Towards efficient resource provisioning in mapreduce. *Journal of Parallel and Distributed Computing* 95 (2016), 29-41.
- [10] Rumi, G., Colella, C., and Ardagna, D. Optimization techniques within the hadoop eco-system: A survey. In *Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2014 16th International Symposium on* (2014), *IEEE*, pp. 437-444.
- [11] Shivaraj B. G., N. N. Survey on schedulers optimization to handle multiple jobs in hadoop cluster. *International Journal of Science and Research* 4 (2013), 1179-1184.
- [12] Shu-Jun, P., Xi-Min, Z., Da-Ming, H., Shu-Hui, L., and Yuan-Xu, Z. Optimization and research of hadoop platform based on fifo scheduler. In *MeasuringTechnology and Mechatronics Automation (ICMTMA), 2015 Seventh International Conference on* (2015), *IEEE*, pp. 727-730.
- [13] Singh, D., and Reddy, C. K. A survey on platforms for big data analytics. *Journal of Big Data* 2, 1 (2015), 8.
- [14] Tamboli, S., and Patel, S. S. A survey on innovative approach for improvement in efficiency of caching technique for big data application. In *Pervasive computing (ICPC), 2015 International Conference on* (2015), *IEEE*, pp. 1-6.
- [15] Ur Rehman, M. H., Liew, C. S., Abbas, A., Jayaraman, P. P., Wah, T. Y., and Khan, S. U. Big data reduction methods: a survey. *Data Science and Engineering* 1, 4 (2016), 265-284.
- [16] Wolf, J., Rajan, D., Hildrum, K., Khandekar, R., Kumar, V., Parekh, S., Wu, K.-L., et al. Flex: A slot allocation scheduling optimizer for mapreduce workloads. In *Proceedings of the ACM/IFIP/USENIX 11th International Conference on Middleware* (2010), Springer-Verlag, pp. 1-20.
- [17] Yildirim, E., Arslan, E., Kim, J., and Kosar, T. Application-level optimization of big data transfers through pipelining, parallelism and concurrency. *IEEE Transactions on Cloud Computing* 4, 1 (2016), 63-75.
- [18] Zhang, H., Chen, G., Ooi, B. C., Tan, K.-L., and Zhang, M. In-memory big data management and processing: A survey. *IEEE Transactions on Knowledge and Data Engineering* 27, 7 (2015), 1920-1948.
- [19] Jena, Bibhudutta, et al. "Name node performance enlarging by aggregator based HADOOP framework." *I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2017 International Conference on.* *IEEE*, 2017.
- [20] Yadav, Kusum, Manjusha Pandey, and Siddharth Swarup Rautaray. "Feedback analysis using big data tools." *ICT in Business Industry & Government (ICTBIG), International Conference on.* *IEEE*, 2016.
- [21] Jena, Bibhudutta, et al. "A Survey Work on Optimization Techniques Utilizing Map Reduce Framework in Hadoop Cluster." *International Journal of Intelligent Systems and Applications* 9.4 (2017): 61.
- [22] Chakraborty, Sabyasachi, et al. "A Proposal for High Availability of HDFS Architecture based on Threshold Limit and Saturation Limit of the Namenode." (2017).
- [23] Kanaujia, Pradeep Kumar M., Manjusha Pandey, and Siddharth Swarup Rautaray. "Real time financial analysis using big data technologies." *I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2017 International Conference on.* *IEEE*, 2017.

#### Authors' Profiles



**Chandrima Roy** was born on the 10th of March 1994 in Kolkata. She is continuing her studies in School of Computer Engineering as a M.Tech postgraduate at KIIT University, Bhubaneswar. Her research areas include Data Analytics, Big Data and Data Mining. Her paper "A Proposal for Optimization of Horizontal Scaling in Big Data Environment" has been accepted in "ICDIS 2017". She can be reached at chandrima.roy.1914@gmail.com



**Siddharth Swarup Rautaray**, PhD (Computer Science), Member IEEE is Professor at the School of Computer Engineering, KIIT University, Bhubaneswar. He has more than a decade of teaching and research experience. Dr Rautaray has published numbers of Research Papers in peer- reviewed International Journals and Conferences. His areas of interest are Image Processing, Data analytics, Human Computer Interaction. He can be reached at [siddharthfcs@kiit.c.in](mailto:siddharthfcs@kiit.c.in).



**Manjusha Pandey**, PhD (Computer Science), Member IEEE is Professor at the School of Computer Engineering, KIIT University, Bhubaneswar. She has more than a decade of teaching and research experience. Dr Pandey has published numbers of Research Papers in journals and conferences. Her areas of interest WSN, Data analytics. She can be reached at [manjushafcs@kiit.ac.in](mailto:manjushafcs@kiit.ac.in)

**How to cite this paper:** Chandrima Roy, Siddharth Swarup Rautaray, Manjusha Pandey," Big Data Optimization Techniques: A Survey", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.10, No.4, pp. 41-48, 2018. DOI: 10.5815/ijieeb.2018.04.06