Towards Reducing Energy Consumption in Big Data Networks using Fused Linear Programming

Khaled Alfawaz*

King Abdul-Aziz University, Jeddah KSA; kalfawaz@kau.edu.sa

Abstract

Background/Objectives: Big data is a relative concept of institutions. Some organizations may see that the data dealt with is very large while it is not worth anything to another organization. The essential target of this study is to investigate the potential effect of big data challenges and provide solution for research issues to improve energy of networks. Methods/Statistical Analysis: This is based on the third dimension of large data that measures the volume, diversity, variability and complexity of data (Velocity), and data processing speed and processing performance (Velocity). This article gives a stage to investigate enormous information at various stages and it provides the solution to minimize to clean the chunks of the database before pre-processing. In this study a new method, Fused Linear Programming is introduced. We used 18 of previous researches as a reference. Findings: Studies show that companies using large data have achieved 20% growth. Massive data can also allow businesses to analyze millions of tweets on Twitter, for example, to make a decision about a product, on the opinions and comments of the renegade! It also can be used in the health sector to identify and predict diseases, link characteristics and discover their relationship to diseases and drugs. This can be measured by reference to government decisions that have begun to link their decisions and orientations based on the analysis of their accumulated data. This study introduced Fused Linear Programming method to find the veracity of big data impact in energy efficient big data network in bypass Internet protocol over Wavelength Division Multiplexing core networks. Here we present a processing node, which can be integrated with Internet Service Provider data centres to host Internet Protocol and Wavelength Division Multiplexing nodes. The optimized energy saving us up to 56% when there no backup and 43% in the backup node. Improvements/Applications: A tremendous archive of terabytes of stored information is produced every day from current data frameworks and advanced advances, for example, Internet of Things and distributed computing. Analysis of these huge data requires a great deal of endeavours at different levels to retrieve knowledge for decision-making. Therefore, big data analysis keeps vital role in the current research and development. Based on the above facts. This research point to the possibility of measuring social interactions among individuals within educational environments to solve problems and collaborative skills, allowing for analysis that is more direct and review of performances related to standard search tools.

Keywords: Big Data, Linear Program, Network, BN, ICT, SN, SPN

1. Introduction

Many applications like telecommunication, video conferencing, online business, and their effect on human developments, have stamped Information Communication Technology as a domain amicable segment, be that as it may, there is a drawback of ICT. Because of the ICT accessibility, all over the place and anyplace in our day by day life, the energy expected to keep up and work the system considered as a basic issue related with data transmission development. Another viewpoint, consumption of the energy in the devices, computer and network hardware are turning into a huge piece of the worldwide vitality utilization because of the system development.

For instance, few years before, the Internet data transmission has expanded by roughly 50 to 100 times,
and appropriately, the system control utilization has expanded all the while. The lot of research concentrate energy consumption to increase the life time of the network.

A significant research in Information and Communication Sector is to reduce the energy consumption as the device to connect the internet increases around 40% and generate the huge number of data.

There are major challenges of data growth as a three-dimensional element to describe the massive data in the 3V model. These dimensions are (1) Volume; (2) Velocity (3) Diversity Variety.

In\textsuperscript{5} updated its definition to read: “Big data are large, high-speed, and / or high-volume information assets that require new forms of processing to enhance decision-making, deep understanding and process improvement”.

Data Storage location and power utilization is in the scope of 110-140 GWh every year, which estimated by the energy required adequacy (PUE) list, and the computer centre and air cooling machines consumes up to a large portion of the aggregate power devoted by the server farm\textsuperscript{2,4}. The measure of information made between the time of the sunrise civilisation and 2003 is evaluated to be five Exabyte. Right now, a similar measure of data is made each two days\textsuperscript{5,10}. This enormous increment in producing information and the gigantic information created is alluded to Big Data.

Data Science Researchers characterize the big data into 4 fundamental classifications: Initially the volume of data, type and source of data, manages the space at which data from various source and veracity. The above factors convey numerous difficulties that have suggestions on the power utilization of networks which conveying the traffic of big data.

Big data provides a competitive advantage for enterprises if they are well utilized and analysed because they provide a deeper understanding of their customers and requirements\textsuperscript{11}. This helps to make decisions within the organization more effectively based on the information extracted from customer databases and thus increase efficiency and profit and reduce waste. The results of the search for their products over the Internet by 10-15% while in a report to McKinsey - a leading company in the field of business consulting - that the US health sector if it uses the techniques of analysis of large data efficiently and efficiently it had produced more than 300 million US dollars, an annual surplus of the health budget due to two-thirds of spending cuts costs by 8%. According to a previous Gartner survey, 64% of companies and organizations have invested in adopting new technologies to deal with large data in 2013. The huge human resources project (the study of the entire genetic material of humans), which contains about 25 thousand genes, which in turn contains over 3 billion pairs of chemical bases, is not limited to the huge data on enterprises and commercial projects, but extends to many areas including energy.

2. Literature Review

There have been many statements and concepts about the idea of large data or giant “big data”, which became a full knowledge of his specialties and many aspects. This field has grown rapidly with the spread of the Internet revolution and social networks, and has become a target for many large companies, because of its importance, both in the current period and in the future. Scientists believe that this area will lead the world into a huge revolution, perhaps changing the future of the world. In order not to delve into complex scientific and academic concepts, we can explain the idea to a simple example, according to Digit Digital. Imagine if you are a company manager and have a database of your employees, skills, experience and levels, HR can certainly analyse this information to give you a report, a general picture of competencies, strengths and weaknesses, etc. But imagine if you had a larger database with a large number of companies? Imagine if you have a database of all the information (simple and large) about the population of a particular country, you will surely have a scary treasure of information through which you can draw some ideas and facts, as well as predict future things. This is exactly the idea of huge data, billions and billions of information, which is currently measured by “Exabyte” which equals = “1024 petabytes” which equals = “1 million terabytes Terabytes!!” Do you imagine the size of these data and the size of storage media that accommodate this amount of data?\textsuperscript{12}

No definition of large data. According to a definition, all the data are created and generated by the digital device only, tools and platforms supported by the Internet in day to day life. In all the time, millions of people around the world are using 7to 8 billion phones for different mode of communication. Different transaction like money
transfer, purchase, surfing, updates in social media and etc. All these activities have a digital impact, and this digital information constitutes the bulk of the large data. Since 2012, more than 1.2 gigabytes of data per year, 2110 bytes, is being produced, or enough to fill 80 billion 16 GB iPhones (which may span more than 100 Earth cycles). But the volume of these data is growing rapidly. Thus, size, speed, and diversity are the three “characteristics” that characterize large information called data, often adding the value that can be derived from it as the fourth characteristic.

Big data comes with various types. For example, a tiny piece of “deaf” data - figures - which are known as “structural” as it constitutes groups of variables which can described easily, and easily organized for systematic analysis. For example, records of phone calls collected by mobile operators and telephone call records are data descriptors (data) that record subscribers’ use of their mobile phones with the identity code, and the call location of the telephone tower which specified call path for every caller the recipient, at a minimum - and the time and duration of the call. Major telecoms companies keep records of phone calls on a daily basis. A second type of massive data includes video files, documents, blogging, and also content used in social media. It is very difficult to analyse as those data are “unstructured”. They different from the data of “bread crumbs” subject to the views of authors, and may draw a deceptive image because they are not objective. For example, you might write a post in which your region mentions a specific product, but your credit card statement - based on your actual purchases - may reveal a different truth. It combines a third type of large data remotely with sensors, and reflects the behaviour of a human. This type of devices may “smart meters” fixed at home to calculate the power consumption, and also satellite images can capture physical data. But some view the vast data world wider, it inclusive records in administration work, rate or data related to the weather.

The vast majority of the large data is machine-readable, produced by humans and centred at the same time. These data were not available before the era of Facebook, or the era of massive use of mobile phones, resulting from major technological and community changes. Massive data is that their sources are digital, and are collected from the table in databases even for basic purpose. That is, they were not collected with a view to drawing conclusions from them. This also makes the exploitation of large data difficult. Therefore, the massive description of the data may be a misnomer and misleading, because size is not its distinguishing feature. For example, a file size that contains an Excel spread sheet for telephone call details records may not be large, but the entire “World Bank Development Indicators database” occupies large file, remaining results of the file controlled operations, also surveys conducted by Official. The difference in the first place is the quality of the data, and the method of its generation. For some, the term “massive data does not revolve around data” is related to massive data analyses, which generally refer to improve the power of computing and analytical capabilities, like artificial intelligence algorithm.

Also possible to find the records of phone calls that record location and time that helps to know who they are, even when they have no personal information. According to this, four reference points were sufficient to identify specific person from a complete data set at 95% accuracy.

The risk in the decision making depends on the biased data. Though the people involving policy making are believe that "data do not lie", these types of threats can be very worrying. One of the major challenges in the massive data is that the people who produce them do so of their own volition, through their activity. This is technically a “biased choice”, meaning that the analysis of these large data is likely to get the result entirely differ when compare with traditional survey.

Another risk in flawed analysis is that it lacks “internal authority”. For example, a small sharp in the volume of phone call records in a particular place may be explained by past historical events of the user, such as the announcement of a looming conflict. The reason may actually be somewhat different, such as a signal tower of the mobile phone in the particular area. Another one risk data analysis of large data tends to focused on correlation and prediction at particular cause, diagnosis without the policies will lose all relevance. A good example is “keeping security on a predictive basis”. Around the year 2010, authorities of law and enforcement of the United States and the United Kingdom have been processing data to assess the likelihood of increased crime in some areas, which is expected to increase based on historical patterns. The police are sending their troops accordingly, and in most cases this crime has been reduced. However, unless the increase in crime rates is known, preventative policies cannot be developed that address root causes or contributing factors.
Another serious risk has not received the attention it deserves - the possibility that massive data will create a “new digital divide” that may contribute to widening rather than narrowing income and power gaps around the world. The “how to use data” is one of the key related challenges. All discussions on the “data revolution” assume that “data are important” and that bad data fall under the responsibility of developing bad policies. But history has shown that the lack of data or information has played only a marginal role - throughout history - in making decisions that lead to bad policies and thus to produce bad results. At the same time, the blind “arithmetic” future may undermine the same processes, which are designed to ensure that the way in which data are converted into decisions is subject to democratic control.

Another impact on the future of large data is how it overlaps with the “open” data movement - and its underlying social motivations - and develops alongside them. The term “open data” refers to easily accessible, machine-readable and accessible data for free or at a negligible cost, with minimal restrictions on their use, conversion and distribution. In the near future, it is envisaged that large data movements and open data will become key pillars of a larger “data revolution”. And that the star of both will emerge against the background of increased public demand for greater openness, rapid decision-making, transparency, and accountability for data and public actions. In addition, its political significance is clear. We should therefore aspire to a massive “real” data revolution in which the impact of data can be exploited in many vital areas, which can lead to the interests of humanity.

3. Problem Formulation

Big data has been the focus and study of mathematics for the last 100 years in some form. As a classic example, meteorology is where we need huge amounts of numbers that must be compressed to generate realistic ritual predictions. Similarly, the large data sets generate different type of climate models, data involved in geophysics, and also astronomy. In any case, data sets on this type of issues - although large - are well-ordered and understandable. This is the result of being coming from physical processes well understood by scientists. The real challenges lie in understanding and dealing with large data in biosciences, social sciences, and especially those based on human activity. Such data are often distorted, incomplete, unreliable, complex, and narrative rather than the same. Physical data are not.

Our major interest lies on how Big data be can photographed? How do we clean it and understand it? How do we experiment with the systems it generated, and how can control this type of systems? The technological and mathematical challenges beyond these questions vary, because these are most important, the sheer volume of data makes the process of automation imperative. This automation is based on mathematical algorithms. The questions we might ask about massive data include:

- How do we classify the importance of information in large networks found in Internet browsers such as Google? How do we define consumer activities, loyalty and even their feelings, and how do we conduct personal propositions? How do we simulate the uncertainties in patient health trends? How do we accomplish and deal with the real-time health monitoring process in 5G.
- How do we use smart data in power supply? I think any reasonable person would agree that numerous future improvements in Modern mathematics will either lead us to recreations by vast data based applications, or will rely upon the need to see extensive information of data. A significant number of the current numerical procedures are presently down to earth applications in our comprehension of vast information, for instance a noteworthy case of system hypothesis (organize hypothesis).

The main challenge with the Big Data researchers is veracity as it has to identify the difference between dirty data and actual required data. To keep away the dirty data out of the data centre databases is essential. To remove unwanted duplications and errors from database is enhance the quality of database. When managing numerous enormous data sources, the requirement for data purging winds up noteworthy, the sources contain grimy data because of overlap, duplications or conflict materials. So, it is a vital to clean the data with the goal which prepared for huge information examination. Thus, giving simple access to exact, reliable and combined information of various information shapes is required.15.
Network theory describes - as the name suggests - objects known as nodes that are connected to each other through what are known as edges. These nodes can be computers, or spider webs. The edges are connections between computers, or links between websites. The nodes can also be human beings, and links are friends on Facebook or Twitter. Or can be mobile phone groups, and the links in this case are conversations or simply neighbour proximity that may lead to interference. Network theory explains the nature of networks and allows us to search for links between individual points of data sets and can describe the movement of information about the network. In fact, the management of the mobile network (which is actually used in data loading) is also very important and a continuous application of growth to the field of graph colouring: finding ways to colour the edges or nodes in the network according to specific limitations such as the possession of contiguous nodes Different.

For example, these colours may represent the frequencies associated with mobile transmitters that must be selected so as to minimize interference, and therefore should be different for adjacent transmitters. It was not until late to look at the colouring of the drawings as coming back to the field of pure mathematics. Other examples that lead to massive data include the operation of organizational networks such as management networks, crime gangs, even voting behaviour in the European Song Contest, as well as technological networks such as power networks and circuits, information networks made of protein-Transport networks such as airlines, food logistics, underground and above-ground train systems, and environmental networks, such as food chains, diseases and infection mechanisms\[16, 17\].

Network theory can address many large data questions. When you deal with very large networks it will not always be easy to identify clusters - groups of nodes that are strongly associated internally - or divide data into groups that share common characteristics. Such information is very important in data mining and pattern recognition. This is particularly related to the retail sector, which is concerned with consumer behaviour and events, but it can also be associated with the creation of voting patterns in the European Song Contest, and the theory of networks provides the algorithms needed to determine aggregations and to divide data. Such analysis helps solve another important issue that can be encountered in many applications: linking data that depends on different levels of space and time. An example is the weather forecasting process, in which some data may come from Earth-orbiting satellites and transmit data up to several megabytes per second.

Another piece of data may result from individuals in isolated earth stations providing a few measurements each day. Some of these data may also be historical, such as records of sea captains, which include the last 100 years. These three data sets are useful and must be linked together in a clear and concise manner. How do you connect the network or spread the connections to the same importance? As well as the shortest paths within the network. These questions are essential for effective Internet, in addition to the interpretation of logistics data; understand the rapid communication, also marketing. Network theory is also essential in the search for nodes affecting giant networks. A strong communications contract - whether representing people, web sites, or airports - is critical to network cohesion, because deleting them will significantly affect the overall availability of communications. Thus, such information can be used to break terrorist organizations, stop the spread of epidemics, or maintain air traffic when the area is affected by bad weather.

The proposed method Fused Linear Programming (FLP). The data sources from multiple location and multiple applications. The raw data cleaned and transform it to data warehouse and extract transform load to clean data. Then it is applied analytical tool such as Hadoop or Map Reduce. Figure 1 shows Framework Architecture of Big Data. The cleansing the data this research introduced Fused Linear Programming. It optimizes the data base or storage location of cleaned the chunk of big data prior to processing as there is the limitation in processing node in the network. For the future reference and protection of the data chunks, the proposed method automatically takes the copy of cleansed chunk. So that it minimizes the energy consumption. When the network node consists as 1 to n, the optimization to find the location from 1 to n in Data Centre. The system accessed one backup of data centre for cleansed data and it named as backup node.

From the network all the data transmitted from various sources, a Source Processing Node (SPN) performed the cleaning process. It generated chunks with as very small volumes. After cleaning of chunks, it
processes in the energy efficient big data network. The processing node creates different impact of limitations in storage on the network to store cleaned chunks. A purified Chunk reproduction is a reinforcement Chunk made whether the first Chunk is eliminated or lost. This reinforcement Chunk is ideally put away in Backup Node. This Backup Node could be either Source Processing Node or Intermediate Processing Node. Nonetheless, choosing the area of the Backup Node in a similar area of one of the two areas upgraded Data Centre’s isn’t concentrated in this work. Quantity of utilized Backup Nodes can be chosen by the strength level wanted for huge information unique Chunks.

5. Evaluation of FLP

SN is the Source Node, SPN is the Source Processing Node, BNa is the Backup node when BN =1 otherwise it is 0. $BCH_{sd}$ is the flow of the traffic of the Chunk between the node s,d and i&j.

The router consumed power, it has been calculated as

$$PRPORT = \sum_{i\in N} PR \cdot (AB_i + AR_i + ACH_i + AI_i) + PR \cdot \sum_{j\in N \cap m} (C_{ij})$$

The above equation calculated the router port power during the data traffic. Again the energy consumption of all the switches and router are calculated as follows:

$$PSR = 2 \cdot \sum_{s,d \in N} BCH_{sd} \cdot (RS \cdot SEB \cdot RR \cdot REB)$$

$$+ \sum_{p \in N} \sum_{d \in N} CHT_{pd} \cdot (RS \cdot SEB + RR \cdot REB)$$

$$+ \sum_{p \in N} \sum_{d \in N} (CHT_{pd} + INT_{pd}) \cdot (RS \cdot SEB + RR \cdot REB)$$

$$+ \sum_{p \in N} \sum_{d \in N} INF_{pd} \cdot (RS \cdot SEB + RR \cdot REB)$$

The equation computes the power utilization of Routers and switches available in the interior in the Source Processing Nodes, Intermediate Processing Nodes and Data Centres, and the additional inside switches then switches control utilization in the source processing node and backup nodes coming about because of sending reinforcement Chunks between them. A homogeneous system and hardware, the aggregate power utilization in source processing nodes because of reinforcement Chunk movement is equivalent to the power utilization of the backup node getting that activity, henceforth the factor of two in condition. The energy utilization of the capacity for the backup node is shown: -

$$BNSTORAGE = \sum_{d \in N} SCH_d \cdot RSG \cdot PSB$$

Objective to minimize

$$PUN = \sum_{m \in N} \sum_{n \in N} PTR \cdot W_{mn} + \sum_{m \in N} \sum_{n \in N} PRG \cdot W_{mn} \cdot RG_{mn} + \sum_{m \in N} \sum_{n \in N} PE \cdot A_{mn} \cdot F_{mn} + \sum_{i \in N} EO_i$$

$$+ PU (\sum_{p \in N} BNW_p + 2 \sum_{s \in N} BCH_{sd} \cdot (RS \cdot SEB + RR \cdot REB)$$

$$+ \sum_{p \in N} \sum_{d \in N} CHT_{pd} \cdot (RS \cdot SEB + RR \cdot REB)$$

$$+ \sum_{p \in N} \sum_{d \in N} (CHT_{pd} + INT_{pd}) \cdot (RS \cdot SEB + RR \cdot REB)$$

$$+ \sum_{p \in N} \sum_{d \in N} INF_{pd} \cdot (RS \cdot SEB + RR \cdot REB) + \sum_{p \in N} SCH_{pd} \cdot RSG \cdot PSB)$$

From the above equation it minimizes the power consumption with the following constraints.

The backup node for chunks constraint

$$\sum_{d \in N} BCH_{sd} = \sum_{v \in N \cap CHV} CHV_{st}$$

When in consist of many backup node the constraints are: -

The constraint calculates the stored chunks backup.
6. Veracity Results using Fusing Linear Programming

There are many tools and techniques used to analyse large data such as: Hadoop, MapReduce, GridGain, HPCC, Storm, Cassandra, but Hadoop is one of the most popular tools, and Hadoop is an open source software or platform written in Java for storing and processing large data. Such as storing large data on multiple devices and then distributing the process on these devices to speed up the processing result. The analysis of these types of large data in education can be used to provide a variety of opportunities and options to improve student learning and personalize the student’s path to content mastery through adaptive learning or competency-based learning, resulting in better learning as a result of faster diagnosis and more in-depth learning or learning needs during the learning process, including assessment of skills such as structured thinking, collaboration, problem solving in a deep context, an authentic assessment of the field and subject matter of knowledge, Financial students and institutions, and use of existing environments and complex information in decision-making and identify policies.

These data can also provide modern and effective tools to measure students’ performance of learning tasks. These types of tasks can be measured by increasing the relevance and accuracy of results on how students learn, and can also help design learning environments tailored to the specific needs of students. A clear analysis of individual and collective responses to a range of educational issues. According to the fusing linear programming, we have implemented in Matlab tool to compare with energy efficient big data networks. The comparison of conservation method which not cleaned the chunks and passes to data centres.

### 6.1 Veracity with Sufficiently Extensive Capacity Limit

As in past areas, we looked at the energy efficient big data network approach to the established methodology where the Chunks not rinsed and specifically sent it to Data Centres. This implies the crude activity volumes in the mentioned networks are littler contrasted with the traditional methodology which purifying and handling occur inside the Data Centres as it were. Every methodology, we assessed two methods of activity. According to the main mode, there is backup node in the system and in the second mode backup node is not utilized. In this way, we thought about the two methodologies Energy Efficient Big Data Network versus Classical Big Data Network against one another for every mode.

For the energy efficient big data network, the purified volume of the Chunk change in an irregular and consistent dispersion between 15Gb and 225 Gb. Then again, a bigger volume extend for the traditional methodology is accepted between 15Gb and 335 Gb. In all the cases, the source processing nodes stockpiling is sufficiently extensive to keep the purified and cleaned data. The proposed research utilized the info esteems to look at the impact of veracity on system control utilization. Figure 2 demonstrates network power consumption for the classical big data network and energy efficient big data network with and without backup node for veracity. This framework execution yields imperative contrasts in the system control sparing within the two modes. When consider an example, the most extreme power sparing is 46% in the reinforcement mode when it is 59% for the no reinforcement mode at $\beta = 51$. Normal power sparing in the reinforcement mode is 40% and 50% in the no-reinforcement mode. Explanation behind the lower control reserve funds in reinforcement mode is because of nearness of the additional reinforcement movement between source processing nodes and backup node that expands the system control utilization and lessens the system control investment funds. Then again, there is no reinforcement movement in the no-reinforcement situation, however just CHT shows up in the system, or, in other words source procession nodes to intermediate processing nodes or from source processing nodes to data controls, in this way limiting the power utilization.
Figure 3 shows the process nodes and data centres storage with various values of β for veracity estimations of β for energy efficient big data network approach. This demonstrates that node 7 is chosen as the backup node at all estimations of β. It is because of vital area of node 7, which has the base number of jumps to every single other hub. What's more, the data control areas are chosen at node 5 and 14 for all estimations of β. Up to β = 35, the data control stockpiling use stays unaltering due to the first Chunks prepared either locally in the source processing nodes or transitionally in the Intermediate Processing Nodes. At β = 50 the backup chunks overwhelm the system contrasted with the Chunk Big Data Traffic and Information Big Data Traffic. At the phase where 35 < β ≤ 50, the data controls begin to get an impressive number of unique Chunks in light of the fact that most PN assets are used. In like manner, CHT increments impressively notwithstanding the current BCH, accordingly yielding a general increment in system control utilization as talked about before. At the point when 50 < β ≤ 60, all PN handling assets are drained, along these lines, the expansion in the data control stockpiling use is noteworthy as all additional Chunks are sent to data control for putting away and preparing. Thus, the consolidated activity is presently at its most extreme incentive at β = 60.

7. Conclusion and Future Work

The proposed work has introduced Fused Linear Programming method to find the veracity of big data impact in energy efficient big data network in bypass Internet protocol over Wavelength Division Multiplexing core networks. Here we present a processing node which can be integrated with Internet Service Provider data centres to host Internet Protocol and Wavelength Division Multiplexing nodes. Data Centre has small version of processing node and it contains limited capacity in the available space, it builds the processing node within the data centre. The cleaned data storage locations optimised and the backup node of the location optimised to store a copy of cleaned chunks for later usage. The optimized energy saving us up to 56% when there no backup and also 43% in the backup node.

As a future work, we can point to the possibility of measuring social interactions among individuals within educational environments to solve problems and collaborative skills, allowing for more direct analysis and review of performances related to standard search tools. We also point out that researchers can benefit from the analysis of large data by collecting accurate data on individual and group student work, which provides more details about the learning paths and actions taken to reach them. In addition, large evaluations provide information on the development of these experiences, such as recording the number of times the student searches among the pages of a set of sites related to the content of the textbook. The analysis of large data also helps researchers know how to create data, identify the process that originally produced the data, how it propagates, and help learners and professionals in learning how to build modern and
effective learning models to ensure the best quality in speed. Productivity, and predict future outcomes such as cycle taking patterns.

8. Acknowledgement

This paper has been presented and discussed during the International Conference on Communication, Management and Information Technology- ICCMIT 2019, Vienna, Austria.

9. References

1. Dong X, El-Gorashi T, Elmighani JM. IP over WDM networks employing renewable energy source. Journal of Lightwave Technology. 2011; 29(1):3–14. https://doi.org/10.1109/JLT.2010.2086434
2. Zikopoulos P, Eaton C. Understanding big data: Analytics for enterprise class hadoop and streaming data. McGraw-Hill Osborne Media; 2011.
3. Wirtz T, Ge R. Improving mapreduce energy efficiency for computation intensive workloads. International Green Computing Conference; 2011. p. 1–8. https://doi.org/10.1109/IGCC.2011.6008564
4. Ruj S, Pal A. Analyzing cascading failures in smart grids under random and targeted attacks. IEEE Proceeding 28th International Conference of Advanced Information and Network Application; 2014. p. 226–33. https://doi.org/10.1109/AINA.2014.32
5. Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems. 2013; 29(7):1645–60. https://doi.org/10.1016/j.future.2013.01.010
6. Ramaswami R, Sivaranjan K, Sasaki G. Optical networks: A practical perspective. Morgan Kaufmann; 2009. p. 1–928. https://doi.org/10.1016/B978-0-12-374092-2.50009-6
7. Al-Fares M, Loukissas A, Vahdat A. A scalable, commodity data center network architecture. ACM SIGCOMM Computer Communication Review; 2008. p. 63–74. https://doi.org/10.1145/1402946.1402967
8. Greenberg, Hamilton J, Maltz DA, Patel P. The cost of a cloud: Research problems in data center networks. ACM SIGCOMM computer communication review. 2008; 39:68–73. https://doi.org/10.1145/1496091.1496103
9. Ishizuka Y, Chen W, Paik I. Workflow transformation for real-time big data processing. 2016 IEEE International Congress on Big Data; 2016. p. 315–18. https://doi.org/10.1109/BigData.2016.47
10. McAfee A, Brynjolfsson E, Davenport TH, Patil DJ, Barton D. Big data: The management revolution. Harvard Business Review. 2012; 90(10):60–8.
11. Simmhan Y, Aman S, Kumbhare A, Liu R, Stevens S, Zhou Q, Prasanna V. Cloud-based software platform for big data analytics in smart grids. Computing in Science and Engineering. 2017; 15(4):38–47. https://doi.org/10.1109/MCSE.2013.39
12. Aiello M, Pagani GA. The smart grid’s data generating potentials. Computer Science and Information Systems; 2014. p. 9–16. https://doi.org/10.15439/2014F509
13. Sancho-Asensio A, Navarro J, Arrieta-Salinas I, Armend’ariz-T’engo JE, Jimenez-Ruano V, Zaballos A, Golobardes E. Improving data partition schemes in smart grids via clustering data streams. Expert Systems with Applications. 2018; 41(13):5832–42. https://doi.org/10.1016/j.eswa.2014.03.035
14. Mori H, Kurata E. An efficient kernel machine technique for short-term load forecasting under smart grid environment. IEEE Power and Energy Society General Meeting; 2012. p. 1–4. https://doi.org/10.1109/PESGM.2012.6345687. PMCID:PMC3293050
15. Yoo PD, Muhaidat S, Taha K, Bentahar J, Shami A. Intelligent consensus modeling for prolinecis-trans isomerization prediction. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2014; 11(1):26–32. https://doi.org/10.1109/TCBB.2013.132. PMid:26355504
16. Liang H, Choi BJ, Abd Rabou A, Zhuang W, Shen X. Decentralized economic dispatch in microgrids via heterogeneous wireless networks. IEEE Journal on Selected Areas in Communications. 2012; 30(6):1061–74. https://doi.org/10.1109/JSAC.2012.120705
17. Gaber MM, Zaslavsky A, Krishnaswamy S. Mining data streams: A review. SIGMOD Record. 2005; 34(2):18–26. https://doi.org/10.1145/1083784.1083789
18. Stojanova D, Ceci M, Appice A, Dzeroski S. Network regression with predictive clustering trees. Data Mining and Knowledge Discovery. 2012; 25(2):378–413. https://doi.org/10.1007/s10618-012-0278-6