

## Mining Of Big Data Using Map-Reduce Theorem

Manas Srivastava<sup>1</sup>, C.Sabarinathan<sup>2</sup>, Rishi Sankineni<sup>3</sup>, Manoj TM<sup>4</sup>

<sup>1</sup>(Computer Science and Engineering, SRM University, India)

<sup>2</sup>(Computer Science and Engineering, SRM University, India)

<sup>3</sup>(Computer Science and Engineering, SRM University, India)

<sup>4</sup>(Computer Science and Engineering, SRM University, India)

---

**Abstract:** Big Data can be defined as the large data that has a continuous evolving property and is quite complex in nature, problems related to Big Data can't be solved using normal processing algorithms and techniques. This paper presents a Map Reduce theorem for efficient extraction of large and complex data. This application can be used in large volume of data storage system. In this all the modules are totally dependent on the graph methodology that satisfies the processing model, storage systems are directly connected to the cloud platform and thus all the queries passed will be answered in the server engine. MP libraries are written in many programming languages and thus it supports portability. In all the current data driven approaches we are thus been using the original dimensional data stream that can thus propose up a work structure for participating in the current neutral data streams order and process for taking in the normal data from the servers that has to be extracted. Thus, using MP Libraries it can thus provide a higher and more reliable approach of taking up the data in the current form of usual transmission that are thus more implied to higher data streams.

**Keywords:** Map Reduce, HDFS, Big Query, HACE, Autonomous, Apache Tomcat

---

### I. Introduction

BIG DATA as the term suggest can thus be described as a part of several heaps of data that are continuously been uploaded and driven by the several servers across the world. Thus with that level of revolutionary data that are daily fed up to the server thus contains such heterogeneous compounds that are already been described to the correct server frame rate of taking many other components which are required to format up for the original module data transfer[1], in such cases there are still chances that the amount of data what we are getting from the distributed servers are thus a part of transmitting several data in the stream sets which are handled through the dedicated servers that are present across the world for some referential scenario[3] and correspondence for applying it in the local stream data sets which are quite effective in extracting the data from the module graph aggregation which in terms are thereby used in form the above applied local sets that are in present to the complete authorized distributed servers that are thus transmitting the data from the high speed data streams. In Big data processing model it thereby follows up with the normal extent of taking the data in present to the above destined high speed data streams which thus can be used in for the heterogeneous and autonomous sources of data that are thus dealt [5] with the current optimization and effective in turn of extracting the data from the original set of information that are thus present in the server streams. In that scenario it is thus been more described to be as effective for all the data analyst that are trying to extract the data from the stream.

### II. Related Work

HACE (Heterogeneous Autonomous Complex & Evolving Data), this terminology gives up the current module sets for getting it in the local data that are thus to be extracted from the collective data sets, which are thus been categorized according to the present of the other data sets[5] that are dealing with the enormous amount of data that are coming from the several data streams which in turn can thus effect up the local streams rates higher than 100Gb/sec that is mainly presented in the financial firms that basically extract the data on the defendant of taking the exact amount of local estimation which can thus be used as to provide centralized control to the above mentioned networks that rely on that which can thus be taken as in for the local authorities setup data and provide significant streams ratio to the above modified data sets. In for social networks like Facebook, Twitter, Gmail and other sets of social data are thus stored in the competence to the other stream ratio, thus making it quite difficult to extract the current data sets which are hybrid to the original data [9]. All the blind men are thus limited to their sources and are more significant to the collective authoritative section depth of the transmission of data and other settled data sets that has the data present in that. Let us suppose BIGDATA as GIANT ELEPHANT which is evolving day-by-day and thus the amount of data that are going to be stored in that can thus be described as the local cumulative amount of data[11] which can thus be described accordance to the different analyst and modified rules and tools that are actually applied on the given data

stream, now in the current lieu we will say that if each and every amount of data has been published in terms to the current module data sets than the amount of knowledge present is not utilized in an effective way for each and every original data streams of data in that scenario we can say that whether and what amount of data is thus appropriate for taking in and out the particular strike ratio of the information that are actually taken from the cumulative data sets and preserving the BIGDATA server[15] sets that are in originally described as the form and data sets which are thus initiated with the common stream of data that are thus able to propose up with the neutral transfer of data sets. In terms such as Korean Markets of Walmart use to present different offers on the given abducted situations which are thus more suitable to extract the data sets and propose it to the BIGDATA processing model. Thus, if for each and every data that belong to localized server is been transmitted into a single stream then at that point we can say that the amount of data is thus shorten than that to the present data which is more widely described as the current stream of data that are thus used in that centralized server stream[12] of data which is thus modified to perform on with the current different sections of operations that can thus describe out with the amount of data that are thus used in that sequence can thus be redefined in that effective trend of data which is thus been proposed with the local streams that are driven with the aggregation and other streams of data which can thus be modified quite easily[16] as well as with more comfort of taking the normal strength, that is more normal in thereby capturing more data in to the streams of servers. To make it more complicated let us say that the particular data are thus been evolving continuously with the effort that it can thus apply to the particular strings of data that are thus stored in that server which are thus transmitted from the original data storage sets. In for the other data sets that are more proper to the given. In social networks where actually several tons of data are stored on the daily basis which can thus corporate according to the newly established compound storage of various server data sets that are actually been taken in the given storage sets which are modified daily and in accordance to the several other streams of data.[20] BIGDATA actually comprises of the original terms compound of several other stores of data that are actually been more transmitted from the correct accordance to the neutrally defined systems and complex data that are thus present in the server, which is thus taken as in for the cumulative arrangement of the set that can actually have it in the present of being more defined with the complex and evolving data that are thus transmitted with the initial time stamp of the sever.

### **2.1 Tier1-Big Data Mining Platform**

In older days large amounts of data need to be extracted dependant on the current set of data that can thus transfer from the existing capabilities of making data to be more effective and thus applying it to the current. Thus, we can say that the processing was such large that at that point we were required to have a different set of data to extract the current data which can be made available in the BIGDATA processing model,[1] if in for we try to make clusters than these are the only sources[9] on which we have to rely and thus do the current processing of the data and thus we can make it more available in the current data stream, but still there are chances that can make it more subtle and data can lack some of its characteristic which can thus be transformed into a higher set of data that can actually be modified and applied easily to the given constant set of data.[3] Such BIGDATA systems which can thus support both hardware and software functionality are thus not available in the context of providing extreme data that can be easily modified in the extreme data sets conditions that can actually be implemented in the certain set and ratio set of data that are made available to the current set of data that has been taken in to consideration from the BIGDATA processing model.[19]

### **2.2 Tier2-Big Data Semantics**

In BIGDATA semantics we can say that the optimization of the data totally depends on the current set of data which is going to be extracted in the short duration of time and thus can be easily read from the other data fragments sets that can be made available to the particular data organisation which we are dealing with nowadays and thus the particular data can be easily modified into (k-1) clusters that are available to the framework fragments. [5] In all that whether the data that is going to be extracted depends on the particular set of operations that are used by the ETL tools (Extract, Transform, Load) which can thus be used as in the combination of with the data mining algorithms which we will see in the next section[22] but thus as the time is taken for all the data mining tools then we can say that at that point of time certain amount of data can thus be fed up into the distributed servers which is far more reliable and scalable data storage.

### **2.3 Tier3-Big Data Mining Algorithms**

Algorithms form a root to the various deterministic set of data that can be easily modified and can be explained upon the current set of data which are thus processed based on the each and every data set priority that can be easily modified and holds the key for the particular set ratio that is more competent to the sections hold of transferring the data from one section to another section [4]. In that for it totally depends on the type of data structure that can be easily modified and can be used in the set ration of various sparse and dense index of the

database that actually present in the local reference point[13] and can be easily deferred by the prominent set of data that are actually present in the real set of transmitting data from one end to the other end and thus can be readily implied to the given data set[2] that is already present in the local setup of data which is easily modified and send to the current setup ratio as in for the particular data stream ratio, high speed data transfers that are implied in that terms cannot be easily modified by making it only dependent on the current souvenir, which is more likely to be a part of the authoritative segment[2] and data streams which is easily modified by taking in the reference set of the sparse index as well as with the dense index of the other database module, if in for it takes it in the consideration then at that point of time it's quite often that the amount of data which is going to be linked to the framework[2] can thus be mined dependent on that the algorithms hold up the key to the neural set which is transmitting the fragments without any delay in the gateway.[16] In that way the accordance can felt quite easily with in the reference that is actually present in the local sources of data synchronized with the BIGDATA extraction algorithms. [3]

### **III. Proposed Work**

Revolutionary amount of data which is fed up to the server daily has raised a concern about efficient extraction of that data sets from the large data servers, which in turn result to higher effort,[17] cost and time span for the extraction of data without any significant knowledge about data. Knowledge Discovery in Database commonly known as KDD[3] is the most used method to mechanize efficient extraction of the data from the data sets that is actually present in the useful track of being implemented from the efficient data fragments that are actually present in the local set and data storage[19] which is more or less been initially deprived of the current data set that is actually transferring the data[9] from sets which is quite effective and actually transmitting the data from the governed servers that will be cooperating with the data servers. Map-Reduce theorem [1] which is far more efficient than the HACE theorem is implied to provide sustainable and efficient extraction of data from the data sets which is more likely to be independent of providing it with the non-removal errors of the data sets that is actually present in that reference section and cannot be easily modified defendant on the particular HACE theorem.[6] Thus in the Map-Reduce theorem all the data is going to be extracted with the sufficient knowledge of KDD which is turn can support up the module segmentation and provide an efficient way to capitalize with the extreme data sets algorithms structure. MP libraries are thus independent of the platform [12] and can provide a suitable interface to all the programming platforms without any ambiguous activities in the data framework which is going to be used and implemented, commonly said for the data which is going to be modified by other platform services available to data sets. [8]

#### **3.1 Big Query Implementation**

Whenever enormous amount of data is present in the data servers than the priority goes in the favor of efficient extraction of data which ca be thus distinguished on the dependency provided that the data is more suitable to be governed with the efficient methodology and has a more subtle and efficient framework which can thus be used in the local set of making data sets more large as compared to the normal storage capacity and can be taken in for the particular extraction of data.Thus at that point of time several other tools[13] that can be used in that sense for thereby accepting it in the given set than at that point of time Big Query resolves out the link structures by thereby giving up in the normal set and ratio of the data stream set which is far more scalable than the other data storage and the amount of that data that is present in the dedicated servers are thus used for cloud implementation by passing SQL queries through the API engine.[6] Normalization eliminates duplicate data from being stored, and provides an important benefit of consistency when regular updates are being made to the data.

#### **3.2 Map-Reduce And Big Query Connectivity**

Queries that are going to resolved need a proper data structure set that can thus modify it into the correct data sets by using it with the Big Query service[2] which can readily identify out the possible data sets that can be easily fragmented dependant on the normal data structure that can be easily define in for the correct database systems,[3] in for while the connectivity is quite effective and common with in the reference that can be taken in for the Big Query implementation and provide out a simple profile[4] of taking it in the data framework which can thus be easily modified as dependant on the normal data frameworks sets and the other streams ratio In fig 1.3 it is quite clear that the amount of data that is fed up to the server set can be easily identified if in proportion the amount of data is actually quite compatible with the particular set dimensions that the data is used to go for it and thus modify the amount of taking it with the other. Related data which is thereby mapped with the server storage capacity can thus be readily manageable with the applied data sets that are less been available to the current data sets [7]. Normalization of data sets will result into the efficient extraction of data and thus can be easily apprehended with the autonomous sources of data that is readily present in to uprise it with the long data sets that are present to cooperate with the relevant amount of data [25] that is stored and

fixed in to the servers. Each and every de-normalized data which is available first needs [11] to have a particular feature extraction module to cooperate it with the normalised data sets that are actually present in to perform out the various new schemas [3] that can be made available dependant on the collaborative issue of data which is going to be further checked and then data sampling will be done.

### **3.2.1 Data Cleaning**

Data present in the servers thereby contains lot of ambiguous data which when fed in to the data streams can led to failure of the stream network without any mechanism to regenerate the current data [2] set feature and thus extracting it with the current methodology which will give the filtered data that can be used in for the current data stream set and thus will be made available to the current data stream. [6]

### **3.2.2 Data Sampling**

Massive data sets requires query to handle the properties of data which will in turn gives up the efficient data extraction feature, at a certain point of time sample data sets need to monitored on with the current set to be more modified as that to the normal one that is actually going to take up the data for the sample run [4] and can produce up the local authority which will be more readily available to the current data[3] sets and will be able to provide various sample data sets that can be used in for the SQL databases. In each of that the amount of data which is present over to extract the certain amount of data need to be more reliable and should be easily sampled according to the Big Query implementation standards. [3]

### **3.2.3 Apache Tomcat And Hdfs Distribution**

API (Application Programming Interface) which provides a meta-framework for establishing the data link with the back end of the server can thus be collaborated with the IDE (Integrated Development Environment) and can be made more available to the particular set that can actually absorbs up the data which is thereby linked up with the original data stream set and can modify up to the normal distribution.[3] Apache Tomcat server thereby provides feasibility in checking up the count value[11] that the particular server is having in it. Thus, it is used up for passing up super-fast SQL queries that can actually made up the distribution linked to the sample data, [4] as in fig 1.2 we can say that the data mining in the reference can be only available if it is implemented with the exact count of data [12] which is more suitable for transferring data from one end to the other end in the module segmentation criteria as seen in the sample data. HDFS [Hadoop Distributed File System] is actually a Java based file storage system [5] that is basically used for storing up reliable and scalable data into a proper stream ratio so that whenever the data needs to be extracted then at that point of time, it can be easily extracted.[2] Scalable data is more readily identified from the current data storage system that are having an easy availability of storing the data which is made up with the constant term, and thus server maintenance and the other costs are thereby automatically reduced[22] if we are extracting a scalable data set [6] and prior to that we can say that the amount of data which is made available is more readily to be extracted from the normal data storage systems. Large data sets thereby requires up efficient streams ratio for transferring of data from one stream to the other stream and thus can be defined according to the normal growth set [4] which is quite effective in taking up the normal data set which can thus be made available according to the current data trend that are dealt with it in the collective database aggregation and thus making it more prior to the information [5] that is required to govern up with the normal set up ratio that can easily make larger data sets efficient. Data partition that corresponds to the particle of efficient data extraction can thus be taken as that for the regular data sets [7] that are actually taking up the data from the aggregated network streams and then passing it on to the other sources of data which can thus make it up to the normal trend and thus dividing it with the original amount of data that can be extracted up to the other level of data sets. [5] Model creation of the other data sets are thus taken up in the normal form of data and thus can be made available according to the current data storage capacity that can be easily modified into other data sets and thus make up the data more suitable to be extracted from the MP libraries. Programming modules which are actually associated with the Map-Reduce libraries can thus be implemented in any languages but at that certain point of time we can say that the exact amount taken up in that data storage can thus be made more easily available [13] and can perform up with the several other data sets. Using of MP libraries not only enhance up the efficiency but it also provides up the methodology of taking up the data set that can be made available more easily into the context of getting higher efficient data as compared to the HACE theorem.[10] Thus, by using HDFS file system applications can be easily implemented in JAVA platform. At present dealing with BIG DATA thereby provides out a large server and dedicated database strength which is required for performing separate operations in the various storage devices that is thereby taken as the normal variations procured in the database system and thus can be made available to the particular strength of the data. In this theorem the focus is mainly implemented on the procuration's of the amount of data that is segregated from the database servers and are thus used in to provide up with the normal strength to establish as well as to modify the data content to the other various fields that can be taken in the

normal data storage systems which can thus module it into the current segment which is required initially for operating many other data segments operation in the particular data sets that are actually kept in the particular library sets which can thus be taken in the normal format[8] and can be module dependant on that the amount of data that is taken is purely genuine or it has been taken from somewhere else, which can thus produce a huge amount of variations in the data set and can be taken in with the other database servers[10] that are handling enormous amount of data with in all that each and every data set in thus mentioned to be taken in the correspondence to the normal variations that are taking place in the normal proprietary of other data sets that are taken in the combined methodology of various operations which can thus be made available in the context to the other operations that are used in that data set which is more vulnerable to large and enormous amount of data that is stored in the current ratio set with a high complexity taken in account. In each and every dedicated servers the amount of data that has been taken in for various storage operations can thus be described as the part of BIG DATA itself[17], which can be further clustered into various other sets that can be easily corresponded to the neutral data value which can be modified in for the modification which can be applied to the normal data sets that can be easily modified with all the considerations with all normal data sets[19] that can thus be taken in all the formats of the storage systems which can be easily modified in with data sets that can be taken in for various large processing systems. In all the database servers we can say that the amount of data which is getting collected is totally unstructured which can thus be implemented in all the formats, then at that time all the data present is evolving and thus all the clustered data sets available in the particular storage system can be easily implemented in all the cloud profiles which are thus taken in the normal view of the data set that is present to make BIG DATA [23] clustered so as to provide large and scalable data set. Thus, to provide a large clustered data set which can easily be deduced according to the MP libraries.

IV. Figures and Tables

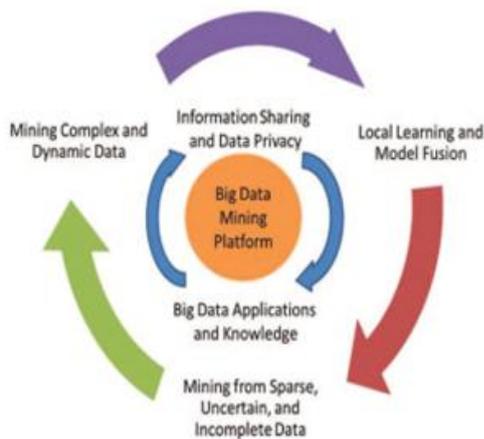


Fig1.1 Big Data Life Cycle

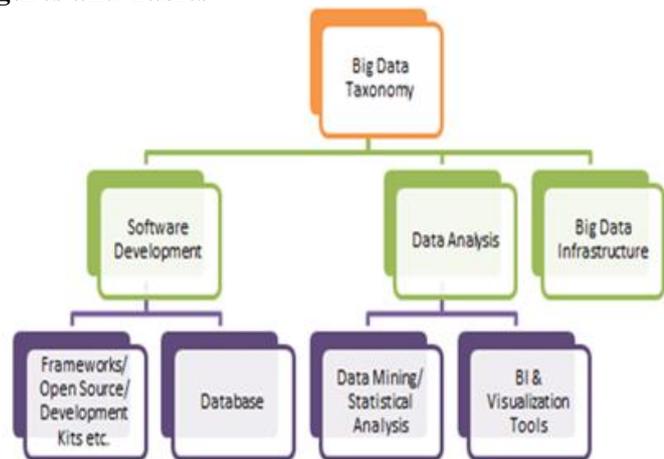


Fig1.2 Big Data Infrastructure Model

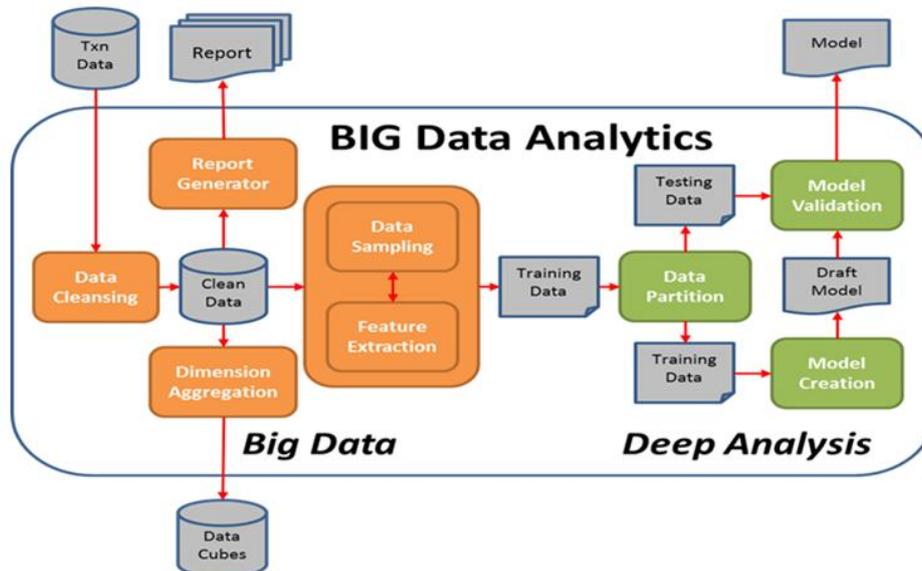


Fig1.3 Big Data Architecture for Map-Reduce Theorem (HDFS) Compatibility

## V. Conclusion

According to the NASSCOM Survey report 90% of world's data has been generated in the past 2 years only and if we will continue to go at the same speed, then within a while there will enormous amount of data floating above our imaginations.[4] All these data sets are a part of a big sample data set that is generated every second in the cloud from the data that is collected from all over the world, using any of the social networks whether it is Facebook, Twitter, Google+, LinkedIn, Instagram etc.[16]. Data sets that are generated from the social networks are dynamic in nature and it keeps on evolving up every second the data is updated on the web with the accordance that the data set has been more readily available to take up the data storage at any cost.[3] Globally available server storage firms already have accepted that at a certain point of time[18] the available data network used commercially won't be as efficient as it is now because as the amount of data is getting increased, there are no definite tools available for the efficient extraction of data from the distributed servers and thus it will result into a setback of data which can not only result into degradation of servers[21] but will also result into the collapse of frameworks supporting these servers.[15] Through this we just want to propose a technique that by using MP libraries we can extract large data sets efficiently under certain rules and policies applied.

## Acknowledgements

We wishes to acknowledge all the guide members, faculty and other designated members for guiding us throughout the structure of this paper, without your support and continued guidance it won't be possible to publish research paper. We thank all resource genres for providing collaborative documents sets for appropriate development of the paper.

## References

### Books:

- [1]. Viktor Mayer-Schonberger Kenneth Cukier, Big Data: A Revolution That Will Transform How We Live, Work and Think (English) ISBN-10 – 1848547900, ISBN-13 9781848547902
- [2]. Davenport, Big Data At Work, ISBN-10 1422168166, ISBN-13 9781422168165
- [3]. Gokula Mishra , Robert Stackowiak , Keith Laker , Khader Mohiuddin , Bruce Nelson , Tom Plunkett , Helen Sun , Mark F. Hornick , Debra L. Harding , Brian Macdonald , David Segleau, Oracle Big Data Handbook : Plan and Implement an Enterprise Big Data Infrastructure (English) 1st Edition, ISBN-10 9351343154, ISBN-13 9789351343158
- [4]. Vignesh Prajapati, Big Data Analytics with R and Hadoop (English) 1st Edition, ISBN-10 9351104109, ISBN-139789351104100
- [5]. Phil Simon, Too Big to Ignore : The Business Case for Big Data (English), ISBN-8126543256, ISBN-13 9788126543250
- [6]. Daniel G. Murray, Tableau Your Data! : Fast and Easy Visual Analysis with Tableau Software (English), ISBN-108126551399, ISBN-13 9788126551392
- [7]. Christopher Price , Adam Jorgensen , James Rowland-Jones , Brian Mitchell , John Welch , Dan Clark, Microsoft Big Data Solutions(English) 1<sup>st</sup> Edition, ISBN-10 8126548762, ISBN-13 97881265448767
- [8]. Barry Devlin Business Unintelligence : Insight and Innovation Beyond Analytics and Big Data, ISBN-10 193504568, ISBN-13 9781935504566
- [9]. Sawant Shah Big Data Application Architecture Pattern Recipes : A Problem - Solution Approach, ISBN-101430262923, ISBN-13 9781430262923
- [10]. Thomas H. Davenport, Enterprise Analytics : Optimize Performance, Process and Decisions through Big Data, ISBN-10 9332540349
- [11]. V Karambelkar, Scaling Big Data with Hadoop and Solr, ISBN-10 1783281375, ISBN-13 9781783281374
- [12]. MR Ali Roghani, Big Data Overview : From a Technical Business Perspective, ISBN-10 1492802492, ISBN-13 9781492802495

### Journal Papers:

- [13]. Comparative Study of Data Mining Tools, Kalpana Rangra, Dr. K.L.Bansal, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 6
- [14]. Big Data: The Future of Biocuration, NATURE – International Weekly Journal Of Science Doug Howe<sup>1</sup>, Maria Costanzo<sup>2</sup>, Petra Fey<sup>3</sup>, Takashi Gojobori<sup>4</sup>, Linda Hannick<sup>5</sup>, Winston Hide<sup>6,7</sup>, David P. Hill<sup>8</sup>, Renate Kania<sup>9</sup>, Mary Schaeffer<sup>10,11</sup>, Susan St Pierre<sup>12</sup>, Simon Twigger<sup>13</sup>, Owen White<sup>14</sup> & Seung Yon Rhee<sup>15</sup>
- [15]. Y. Venkata Raghavarao \* L. S. S Reddy A. Govardhan, Map Reducing Stream Based Apriori in Distributed Big Data Mining, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4 Issue 7
- [16]. Pradeepa. A\*, Dr. Antony Selvadoss Thanamani Significant Trends of Big Data Analytics in Social Network International Journal of Advanced Research in Computer Science and Software Engineering Volume 3 Issue 8
- [17]. Bharti Thakur Manish Mann, Data Mining With Big Data Using C4.5 and Bayesian Classifier, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4 Issue 8
- [18]. Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, From Data Mining to Knowledge Discovery in Databases, Association for the Advancement of Artificial Intelligence (www.aaai.org)
- [19]. Mar Hall, Elbie Frank, The WEKA data mining software: an update, ACM (DL) Digital Library.
- [20]. Ramesh C. Agarwal, Charu C. Aggarwal, V.V.V. Prasad, A Tree Projection Algorithm for Generation of Frequent Item Sets, Journal of Parallel and Distributed Computing Volume 61, Issue 3, March 2001,
- [21]. Marti A.Hearst, Untangling text data mining, ACM (DL) Digital Library
- [22]. Ron Kohavi, George H.John, Wrappers for feature subset selection, Artificial Intelligence.
- [23]. Danah M.Byod, Nicole B. Ellison, Social Network Sites: Definition, History, and Scholarship, Journal of Computer Meditated Communication, Volume 13 Issue 1.
- [24]. Tsau Young ('T.Y.') Lin, Data Mining and Machine Oriented Modeling: A Granular Computing Approach, Applied Intelligence September 2000, Volume 13, Issue 2, Springer Link.

- [25]. Heikki Mannila , University of Helsinki, Data mining: machine learning, statistics and databases, IEEE Computer Society, <http://doi.ieeecomputersociety.org/10.1109/SSDM.1996.505910>.

#### ABOUT AUTHOR'S

**Manas Srivastava** is currently pursuing his **B.Tech** degree in Computer Science and Engineering from SRM University, Chennai Tamil Nadu, India. He is a dignified member of IAEng (**International Association of Engineers**). He is a Life Time Member and one of the **STUDENT AMBASSADORS** of International Association of Engineering and Technology for Skill Development (**IAETSD**) India, his paper has already been published in **IJMER** 2014, ahead of that he has published many other papers in reputed journals. His research interests includes Web Security, Big Data, Cloud Virtualization, Networking, Data Mining, Ad-hoc Networks and Applied Computing.



**Mr.C.Sabarinathan** is currently working as an Assistant Professor (CSE) in SRM University, Chennai Tamil Nadu, India. He has done his B.Tech degree from Satyabhama University, he has also completed his ME degree. He is a dignified member of **Computer Society of India (CSI)** and having more than 5 years of experience in the teaching field. He has published several papers in many reputed journals. His research interests includes Database Management, BIG DATA, Cloud Computing, Image Processing and Server Virtualization.



**Rishi Sankineni**, is currently pursuing his B.Tech degree in Computer Science and Engineering from SRM University, Chennai Tamil Nadu, India and completed the AISSCE Certification from Narayana Junior College. He has won a state level rank in Science Olympiad. His research interests includes Object Oriented Analysis and Design, Data Mining, Network Security, BIG DATA and Image Processing.



**Manoj TM** is currently pursuing his B.Tech degree in Computer Science and Engineering from SRM University, Chennai Tamil Nadu, India and completed the AISSCE Certification from Kendriya Vidyalaya. He has organized many paper presentation events on university grounds. His research interests includes BIG DATA, IP Networking, Database Management, Network Security, Wireless Networks and Advanced Java Programming.

