SELF-ORGANIZED MAPPING BASED MAP-REDUCE TECHNIQUE IN BIG DATA ANALYTICS: A NEURAL NETWORK APPROACH

Vijaya Kumar B. P. and Gazala H. Tidagundi
M S Ramaiah Institute of Technology, Industrial and Systems Engineering Department, Bangalore, India
E-Mail: hod_is@msrit.edu

ABSTRACT

Data mining technology has become a boon for all the engineering fields in today's world. "With the great outburst of data comes a greater responsibility (i.e., organizing and managing the data). Large amount of data is fragmented into smaller chunks of data and ran over hundreds or thousands of servers in parallel ways to extract useful data in big data analytics. With the boom in technology and the huge growth in the data, clustering has become a crucial part in identifying the similarities based on many parameters in a given data set that influences in decision making. To reinforce this clustering, we propose to use machine learning methods to influence over data redundancy and grouping i.e. Self Organized Mapping (SOM). Solving such problem involves in tackling the issues like clustering, visualization, abstraction and de-duplication. Here the work involves in the usage of Artificial Neural Network (ANN) for Self Organized Mapping that has a unique features, like construction of maps, self-organization to form different clusters dynamically to support the volume, variety and variance of big data. In this paper, a novel technique is proposed that supports Hadoop in classifying or clustering of data using Self Organized Mapping. MapReduce has always been used in combination with Hadoop, but in this implementation MapReduce is used external to Hadoop job. Using MapReduce externally is an advantage as Logical Block Address (LBA) and Hash Tag are obtained easily. From the results we found that this implementation has increased speed, data is structured and redundancy is achieved with improved efficiency.

Keywords: data mining, kohonen self-organizing maps, MapReduce, self-organizing map, unsupervised learning.

1. INTRODUCTION

The sudden surge of data because of the boom in the use of internet by the various financial institutions, Medical institutions, social networking sites, Universities, Scientific domains, business, including e-commerce, etc from the past two decades has given way to a newer trend called "Big Data". Big Data is the sheer increase in the volume of the variable and unstructured data. This data is of great potential for the above stated institutions. Hence we need a technique to analyze and process this heterogeneous data in a distributed manner involving hardware and software systems that can serve the purpose of the institutions. The amount of data generated is colossal, in this digital world. This data consists of various YouTube videos, images, uploads by mobile phone, various banking transactions through ATMs, surveillance footage, etc.

a) **Big data sources:** Data is generated at an enormous rate in various fields, some of them are listed below:

A. Medical field, a large amount data is generated through record keeping, regulatory requirements, etc. most of the data is stored in the form of hard copy but as the world is moving towards digitization, each and every record has to be digitized. The data is generated through ECG, MRI, X-ray, bio-metric data, EMR, etc.

B. Transportation field, the data here comes from ticketing services that has been made online, the services that are provided through the websites, Passenger information, transport information, GPS enabled locations that help to trace and navigate the geographical locations. It also helps to provide the traffic details in a certain place.

C. Politics, recent advancement in the technology has allowed the electoral department to go online. Nowadays the campaigning is being carried out through social networking sites, television ads and even web page display. This has led the opponents to campaign against each other in a digitized manner thereby creating huge amounts of data.

D. Agriculture, in this field mobile payment for agricultural products, input purchases and subsidies is generating large amount of data. The use of mobiles extensively is giving a pattern that could help the government to identify farmers or the regions which are in distress.

E. Sports, the data in this field is flooding rigorously; some of the instances are sensors that are placed in the race cars to analyze their speed and other aspects. Fans can interact with their dream players or teams through facebook, twitter and other social networking sites. News about the players and the teams also generates a large amount of data.

F. Media, every song or video that is listened to or watched and every click of a page generate data. Prediction of shows or movies is also one of the reasons for generating data. The Weather channel is also generating massive amount of data by selling the weather news and other insights as news services.

One of the most important challenges for researchers and programmers in this digital world are to categorize data into structured and unstructured data. The amount of unstructured data and the speed by which it is generated is difficult for the existing computer infrastructure to handle the Big Data. Hence Big Data Analytics has emerged. Big Data analytics is the area where new techniques can be operated on big data sets. It is a technique of examining and processing large sets of data to find the required information and also to uncover the hidden patterns within it for a better decision making. Example: To determine the genes in the DNA that would be responsible for certain disease, predict the election results, predict the places where earthquakes can happen, etc. To handle this type of data which is voluminous and variable we have proposed a novel technique involving self organizing maps using Kohonen ANN that provides support for Map reducing technique that has been implemented at application level.

The rest of paper is discussed as follows. Some of related works and research outcomes towards MapReduce, Big Data, and Self Organizing Maps are briefly explained in Section 2. The existing system and analysis for Self Organizing Maps, a Kohonen Neural Network and functionality of MapReduce technique is discussed in Section 3. In Section 4 we have described a proposed model and architecture that shows how Self Organized Maps is integrated with MapReduce. Section 5 includes the implementation details of the proposed technique and the performance evaluation along with the results is given in Section 6. Concluding remarks and future scope for improvements are discussed in Section 7.

2. RELATED WORKS

Many researchers have put forth their views by publishing papers on Big Data and SOM. The speed and the volume with which the data is generated in today's world make it hard for traditional methods to analyze and organize this data in a structured way. So the author in [1] suggest various Big Data Challenges and solution and catering to the problems in hand through Map Reduce framework over Hadoop Distributed File System (HDFS).It also explains various Big data opportunities and detail architecture of Map Reduce. In [2] the author describes the analysis of distribution of data as well as scheduling of the tasks for execution and effective communication. He describes Map Reduce has a programming model for data intensive applications and has also proposed a model for scheduling the divisible loads. Use of Kohonen Self Organizing Map-Neural Networks (KSOM-NN) to study about cluster formation and simulations on Wireless Sensor Networks are carried on to determine the performance with respect to given application parameters and requirements in [3]. Simulations are carried on a number of parameters of sensor networks to understand the chaotic environment with respect to parameters of WSNs.Electronic Health Records (EHRs) have come up in today's world because digitization has become very necessary. Hence the author in [4], uses the layer-wise framework to access the EHRs and data storage is done using MongoDB, a NOSQL open source that stores structured data. He also explains about the data sharing technique that is being carried out with the help of EHRs and data mining components. A kernel method is proposed by the author in [5], this algorithm is based on the energy functions. He has also suggested ways to determine the parameters initialization and has given a detail explanation about overcoming description of clustering centers. In this paper the author concludes that KSOM is better compared to SOM in terms of performance. A technique called Growing Hierarchal Self Organizing Map (GHSOM) on Intrusion Detection System (IDS) traces to detect if there are any signature attacks based on topological distances between clustering is used in paper [6]. This approach has helped Analysts and System Administrator to detect if there are any suspicious attacks and to provide a mechanism to recover for security threats. This approach can be used for the internet security mechanism to detect if there are attacks and prevent any future attacks based on the historical data. Self Organizing Maps has gained its popularity because of ability to preserve the topology in projection. However this topology is not perfectly gained due to static structure of SOM. In order to solve this author in [7] has put forth a novel architecture called MIGSOM (Multilevel Interior Growing Self-Organizing Maps) which organizes itself over a time. As the addition of number of nodes increases, the map can have three-Dimensional structure with multi-levels oriented maps. The experiment also demonstrates how MIGSOM is better than SOM and Growing grid in terms of topology. In [8] the author uses a neural-network-based Multicast routing algorithm for constructing a reliable multicast tree that connects the participants of a multicast group. KSOM is used for clustering and simulations are carried on. A multicast distribution tree (MDT) for multicast routing in mobile networks is proposed. He has shown the ability that the neural networks for solving the computational problems.

Mobile Multimedia networks that handle communication should be efficient and reliable, in order to attain this the author in [9] has used two algorithms called Hopfield Neural Network (HNN) and Kohonen Neural Network (KNN) for the construction of efficient multiple MDTs. This paper is an extension of the previous paper [8]. It is used for on demand multi-cast applications. It is evident that for the less number of Multi-cast groups low number of multi-cast distribution tree is significant. Online Social Networks has bloomed in a wide range of application in today's modern world. This has caused a huge development of data in order to handle this author in [10] has used MapReduce technique to manage and predict the data. Experiments are carried on Obama's Twitter network on the 2012 U.S. presidential election. The results showed the increase in the performance was 90% when the network size was increased. Clustering methods are used in varied areas by researchers; among neural network approach Self Organizing Map hold the highest popularity, using this in [11] the author has come up with the findings of centroids in homogenous data sets. There are two steps used, firstly the learning phase is carried on where parameters reinitialized and secondly the algorithm is run to obtain clustering of map. This is repeated till certain iterations are received. In [12] the author describes about spatial Big data and its role in emerging Wireless Networks Applications. This spatial data is in turn fed to Hadoop frameworks of Map reduce to provide highly parallel implementation of data.

3. SYSTEM ANALYSIS AND DESIGN

a) Self organizing maps:

Our brain is a creation that has fascinated number of scientists time and again. It receives the commands, processes them, analyses and then sends the output as an action. This mechanism of the brain of processing, analyzing and then sending output as an action gave way for a new concept called “Neural Networks”. Neural Networks is an amazing concept, it is totally different from other computer programs because each computer program has its own output or is unique, but in the case of Neural Network each node of the network works independently. It doesn’t depend on the other nodes to carry out its functions, which makes the neural networks fast and efficient. Prior to the network becoming useful it should learn about the information that is present. There are 2 types of learning they are, Supervised learning in which the output is already known, to train the network for the given problem. It compares both input and output to know if there is some error. In Unsupervised Learning, the output is not known, the network must learn from its own to discover patterns from the input data. Hence no human interaction is needed here. This is very important to solve complex problem. One such kind of learning is Self Organizing Maps.

Kohonen Self Organizing Maps (K-SOM) are a kind of neural networks; they were devised by Tuevo Kohonen in 1982. Self Organizing Maps are named so because there is no inspection required. They learn on their own through unsupervised competitive learning. Maps because they assign themselves according to the weights that has been given to the input nodes or data. Another unique aspect of SOM is termed as Vector Quantization, which is a data compression technique which helps in representing multi-dimensional data into lower dimensional space (1-D or 2-D). This helps in the visualization, as humans are apprehended to low dimensional data.KSOM consists of two layers an output competitive layer and an input layer. Every input neuron is in connection with every output neuron there by forming a 2 dimensional grid, the Figure-1 mentioned below shows 4 inputs and 12 outputs [10].

Learning algorithm for clustering
STEP 1: Start
STEP 2: Read the file.
STEP 3: Split the file and extract the words from the file.
STEP 4: Remove the unnecessary words
STEP 5: Compare the each words with the trained data
STEP 6: If the word is in the trained data by Euclidian distance get the identical word and calculate the weightage of the word using equation (1).
STEP 7: Based on the total weightage of the words ,(using equation 2) related to the clusters, categorizing the file in to the respective clusters and store the words in the respective cluster database as well as store the file in the respective cluster folder.
STEP 8: Stop

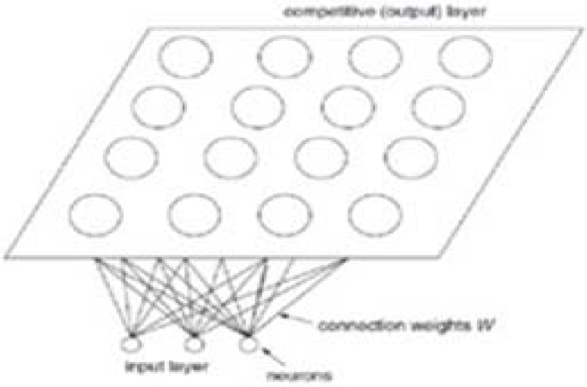


Figure-1. A typical KSOM-NN model with 4 input and 20 output neurons [10].

$$D_j=\sqrt{\sum_{j=1}^{|V|}(I_j - W_{ji})^2} \tag{1}$$

$$W_j^{new}=W_j^{old} + \alpha(I - W_j^{old}) \tag{2}$$

b) MapReduce

Hadoop MapReduce provides a reliable and a fault tolerant way of distributing huge variety of data in to large clusters in a parallel manner for processing. A MapReduce job splits the input data-set into independent chunks which are processed in a parallel manner by the map tasks.

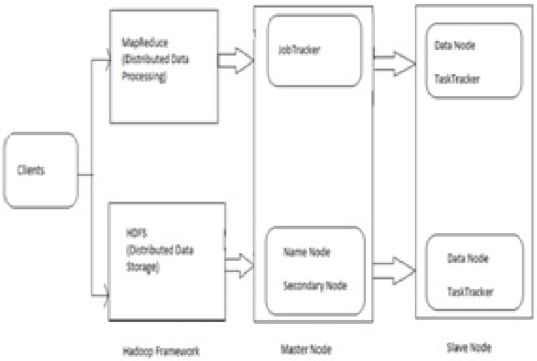


Figure-2. Hadoop configuration with MapReduce and Hadoop.

The outputs of the maps, are then fed as input to the reduce tasks by framework. Both the input and the output of the job are stored in a file-system. The Monitoring task, Scheduling tasks and re-execution of the failed tasks are taken care by framework. The configuration includes storage node Hadoop Distributed File System (HDFS) and computational node MapReduce to include in the same node. This organization provides framework to schedule huge aggregate of data which is already residing at the node in to large clusters in an efficient manner. The framework includes a sole master node called Job Tracker which accounts to Monitor and schedule all the jobs residing on the slave node and also to re-execute the failed tasks. This slave node is called TaskTacker which works as per the instructions from Master Node. The Hadoop Configuration with MapReduce and HDFS is shown in Figure-2.

The MapReduce framework works entirely on <key, value> pairs that is –(input) <k1, v1> ->map -><k2, v2> ->unify -><k2, v2> ->reduce -><k3, v3> (output)

4. ARCHITECTURE

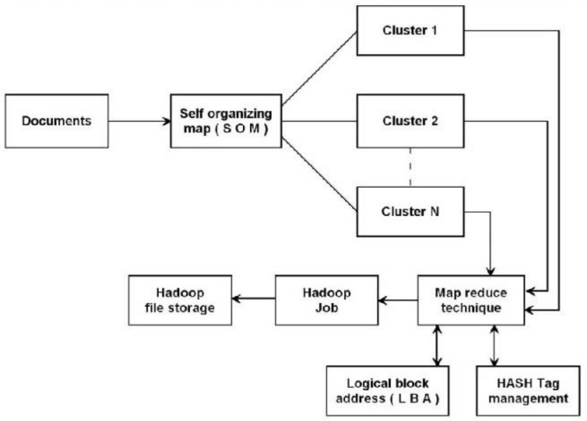


Figure-3. Adapting self-organizing maps to map reduce.

The existing system consists of Self Organizing map and Hadoop MapReduce which were treated as two different embodiments, but as the creativity has emerged far beyond the reach and hence has made the possibility to bring together both these system. Technology has given us the way to use MapReduce technique externally to Hadoop systems.

A simple block diagram of the architecture that is implemented is shown in Figure-3. The three major components of this architecture are evident, they are Self Organizing Maps (SOM) which is unsupervised learning and uses trained data. Followed by this we have MapReduce technique which is called explicit to Hadoop, this provides us Logical Block Address and Hash Code that gives way for de-duplication process. MapReduce technique saves the memory space by overcoming storage of duplicate and redundant data. Last and the most important aspect is storage of blocks in the Hadoop system. Blocks are stored in the Hadoop as it provides efficient way of storing huge amount of data in a secured manner. The original file can be merged and retrieved back based on the logical block addresses assigned.

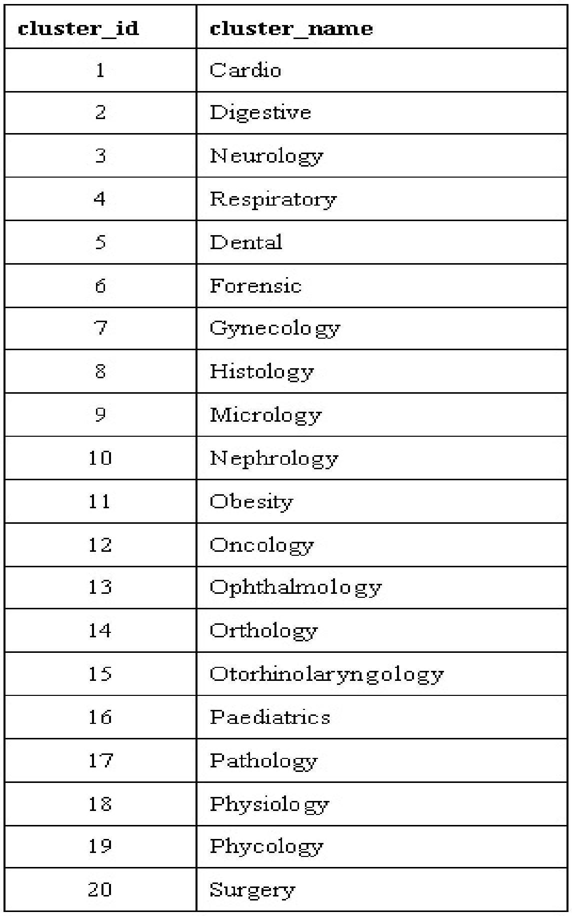
5. IMPLEMENTATION

One of the most important aspects of SOM is that any number of output neurons can be used. The proposed architecture consists of Self Organizing map in combinations with Map reduce technique as shown in Figure-3. Twenty clusters that are obtained from medical background such as, Dental, Neurology, Gynecology, Cardiology etc conforms to the output neurons. The neurons are assigned with cluster_id's as shown in the Table-1. These clusters contain data related to the above backgrounds which also conforms to the output neurons. Self Organizing is named so, because they learn on their own through trained data. Henceforth the above mentioned clusters serve as trained data. When a particular document is given as an input to this machine learning technique (SOM) the code ignores articles, punctuations etc and fetches only the important keywords which are related to trained data. This happens with the help of Euclidean distance i.e. every input neuron is compared with the trained data based on collection frequency. The frequency here is the keywords which are related to trained data and accordingly the weight updating takes place which complies with the increase in the instances of the keyword. This in turn conforms to the weight updating mechanism which activates the particular output neuron and according to the winning neuron, the cluster_id's are returned. Based on these cluster_id's the respective documents are stored in the cluster folders.

The next crucial part is mapping and reducing technique within our application and outside Hadoop. Here the cluster documents are extracted and split into number of blocks. This generates a unique Hash code for each and every block. The hash code here is 128-bit or 32 byte hexadecimal value as shown in Table-2. While mapping, Hash code is of extreme significance to check duplicate content or redundant data. This serves as a verification to check whether the block is present. If the block is present it increases the instances, based on MapReduce technique, else it's uploaded to Hadoop Distributed Files System (HDFS). Logical Block Address (LBA) is fetched in either of the cases. Thus MapReduce technique is carried out separately within the SOM and outside Hadoop which is an added advantage.

Finally, Hadoop is used to store these chunks of data in the form of blocks. To download these blocks, we need a Logical Block Address (LBA) which is fetched from cluster table. These blocks are merged and original file is returned. Figure-4 shows flow chart for execution.

Table-1. Cluster_id and cluster_name.



cluster_id	cluster_name
1	Cardio
2	Digestive
3	Neurology
4	Respiratory
5	Dental
6	Forensic
7	Gynecology
8	Histology
9	Micrology
10	Nephrology
11	Obesity
12	Oncology
13	Ophthalmology
14	Orthology
15	Otorhinolaryngology
16	Paediatrics
17	Pathology
18	Physiology
19	Phycology
20	Surgery

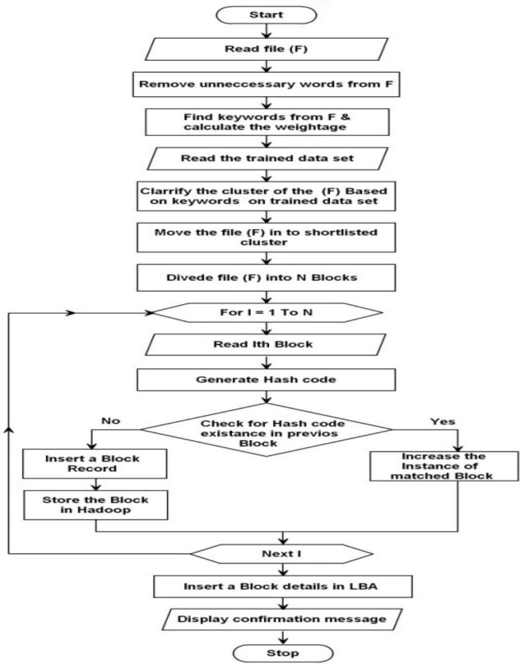
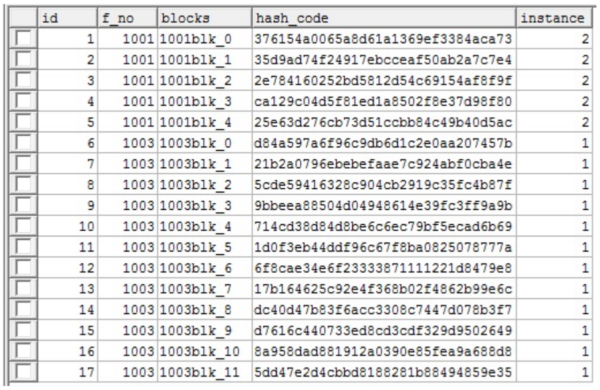

Start
Read file (F)
Remove unneccessary words from F
Find keywords from F & calculate the weightage
Read the trained data set
Clarrify the cluster of the (F) Based on keywords on trained data set
Move the file (F) in to shortlisted cluster
Divede file (F) into N Blocks
For I = 1 To N
Read Ith Block
Generate Hash code
Check for Hash code existance in previos Block
No
Yes
Insert a Block Record
Store the Block in Hadoop
Increase the Instance of matched Block
Next I
Insert a Block details in LBA
Display confirmation message
Stop

Figure-4. Flow chart of project execution.

6. RESULTS

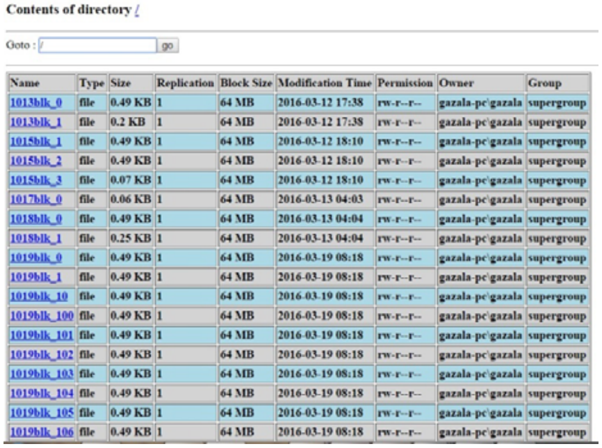
The implementations are tested for different scenarios of data set. We have separately tested for SOM and Hadoop architecture and outputs are discussed. As a part of execution, simulations are carried out using tomcat as Database server and SQL log is used to maintain all the details of data tables. Data table gives us hash codes and more importantly the LBA by which original data is retrieved in its form. The below Table-2 shows the output of the same.

Table-2. Hash code and instances are generated.



	id	f_no	blocks	hash_code	instance
□	1	1001	1001blk_0	376154a0065a8d61a1369ef3384aca73	2
□	2	1001	1001blk_1	35d9ad74f24917ebcceaf50ab2a7c7e4	2
□	3	1001	1001blk_2	2e784160252bd5812d54c69154af8f9f	2
□	4	1001	1001blk_3	ca129c04d5f81ed1a8502f8e37d98f80	2
□	5	1001	1001blk_4	25e63d276cb73d51ccbb84c49b40d5ac	2
□	6	1003	1003blk_0	d84a597a6f96c9db6d1c2e0aa207457b	1
□	7	1003	1003blk_1	21b2a0796ebebefaae7c924abf0cba4e	1
□	8	1003	1003blk_2	5cde59416328c904cb2919c35fc4b87f	1
□	9	1003	1003blk_3	9bbeea88504d04948614e39fc3ff9a9b	1
□	10	1003	1003blk_4	714cd38d84d8be6c6ec79bf5ecad6b69	1
□	11	1003	1003blk_5	1d0f3eb44ddf96c67f8ba0825078777a	1
□	12	1003	1003blk_6	6f8cae34e6f23333871111221d8479e8	1
□	13	1003	1003blk_7	17b164625c92e4f368b02f4862b99e6c	1
□	14	1003	1003blk_8	dc40d47b83f6acc3308c7447d078b3f7	1
□	15	1003	1003blk_9	d7616c440733ed8cd3cdf329d9502649	1
□	16	1003	1003blk_10	8a958dad881912a0390e85fea9a688d8	1
□	17	1003	1003blk_11	5dd47e2d4cbbd8188281b88494859e35	1

Table-3. Hadoop for storage of documents in to blocks.



Contents of directory /

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
1013blk_0	file	0.49 KB	1	64 MB	2016-03-12 17:38	rw-r--r--	gazala-pc\gazala	supergroup
1013blk_1	file	0.2 KB	1	64 MB	2016-03-12 17:38	rw-r--r--	gazala-pc\gazala	supergroup
1015blk_1	file	0.49 KB	1	64 MB	2016-03-12 18:10	rw-r--r--	gazala-pc\gazala	supergroup
1015blk_2	file	0.49 KB	1	64 MB	2016-03-12 18:10	rw-r--r--	gazala-pc\gazala	supergroup
1015blk_3	file	0.07 KB	1	64 MB	2016-03-12 18:10	rw-r--r--	gazala-pc\gazala	supergroup
1017blk_0	file	0.06 KB	1	64 MB	2016-03-13 04:03	rw-r--r--	gazala-pc\gazala	supergroup
1018blk_0	file	0.49 KB	1	64 MB	2016-03-13 04:04	rw-r--r--	gazala-pc\gazala	supergroup
1018blk_1	file	0.25 KB	1	64 MB	2016-03-13 04:04	rw-r--r--	gazala-pc\gazala	supergroup
1019blk_0	file	0.49 KB	1	64 MB	2016-03-19 08:18	rw-r--r--	gazala-pc\gazala	supergroup
1019blk_1	file	0.49 KB	1	64 MB	2016-03-19 08:18	rw-r--r--	gazala-pc\gazala	supergroup
1019blk_10	file	0.49 KB	1	64 MB	2016-03-19 08:18	rw-r--r--	gazala-pc\gazala	supergroup
1019blk_100	file	0.49 KB	1	64 MB	2016-03-19 08:18	rw-r--r--	gazala-pc\gazala	supergroup
1019blk_101	file	0.49 KB	1	64 MB	2016-03-19 08:18	rw-r--r--	gazala-pc\gazala	supergroup
1019blk_102	file	0.49 KB	1	64 MB	2016-03-19 08:18	rw-r--r--	gazala-pc\gazala	supergroup
1019blk_103	file	0.49 KB	1	64 MB	2016-03-19 08:18	rw-r--r--	gazala-pc\gazala	supergroup
1019blk_104	file	0.49 KB	1	64 MB	2016-03-19 08:18	rw-r--r--	gazala-pc\gazala	supergroup
1019blk_105	file	0.49 KB	1	64 MB	2016-03-19 08:18	rw-r--r--	gazala-pc\gazala	supergroup
1019blk_106	file	0.49 KB	1	64 MB	2016-03-19 08:18	rw-r--r--	gazala-pc\gazala	supergroup

Once the Mapping and Reducing technique is carried out internally within SOM, Hadoop comes into picture. This stores data in the form of blocks as shown in Table-3. Thus the original file is retrieved.

7. CONCLUSION AND FUTURE WORK

We use machine learning algorithm i.e. SOM technique to cluster the data and solve the problem which involves tasks like clustering, visualization, abstraction and de-duplication. Data from YouTube videos, images, uploads by mobile phone, various banking transactions through ATMs, surveillance footage; etc has given rise to a Data Driven World where data is flooded in an enormous rate. There is a need to be able to collect this data, analyze and visualize and process in a parallel manner in a distributed way. This project aims in giving the individuals the real feel of the Map Reduce technique, which is superficial to the Hadoop. With the help of this technique it is possible to store large amount of efficient and structured data like forensic, medical sports political etc.This data is stored in the form of blocks with unique LBA assigned to each of them. With the help of this LBA it is possible to retrieve the original file. As MapReduce is always used with Hadoop and Cloud Computing applications, this context helps MapReduce to be used external to the Hadoop and within the application using SOM, irrespective of infrastructure overhead of Hadoop Systems. As a part of future work a Hierarchical clustering models using KSOM can provide more abstraction for handling 5V's of Big Data and this can be cost effective. Data classification at different level can be achieved and thus purifies the heterogeneous and unstructured data. The technique of using ANN that has robust and distributed decision making can be leveraged, across Hadoop architecture and in any Big Data Analytic tools.

REFERENCES

[1] Jaseena, K. U., and Julie M. David. Issues, Challenges, and Solutions: Big Data Mining. Academy & Industry Research Collaboration Center 4 (2014).

[2] Gu, Tao, *et al*. Scheduling Divisible Loads from Multiple Input Sources in MapReduce. Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on. IEEE, 2013.

[3] Veena, K. N., and B. P. Vijaya Kumar. Dynamic clustering and analysis for sensor networks. Information, Communications & Signal Processing, 2007 6th International Conference on. IEEE, 2007.

[4] Wassan, Jyotsna Talreja. Modelling Stack Framework for Accessing Electronic Health Records with Big Data Needs. International Journal of Computer Applications 106.1 (2014).

[5] Chen, Ning, Hongyi Zhang, and Jiexin Pu. A novel kernel self-organizing map algorithm for clustering. Mechatronics and Automation, 2009. ICMA 2009. International Conference on. IEEE, 2009.

[6] Huang, Shin-Ying, Yennun Huang, and Neeraj Suri. Event Pattern Discovery on IDS Traces of Cloud Services. Big Data and Cloud Computing (BdCloud), 2014 IEEE Fourth International Conference on. IEEE, 2014.

[7] Ayadi, Thouraya, Tarek M. Hamdani, and Adel M. Alimi. A new data topology matching technique with multilevel interior growing self-organizing maps. Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on. IEEE, 2010.

[8] Veena, K. N., and BP Vijaya Kumar. Dynamic clustering for Wireless Sensor Networks: a neuro-fuzzy technique approach. Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on. IEEE, 2010.

[9] Vijaya Kumar B. P., and Dilip Kumar S. M .Neural Networks Based Efficient Multiple Multicast Routing for Mobile Networks. at International Journal of Information and Electronics Engineering, Vol. 4, No. 2, March 2014

[10] De C, Gatti, *et al*. Handling big data on agent-based modeling of Online Social Networks with MapReduce. Simulation Conference (WSC), 2014 Winter. IEEE, 2014.

[11] Harchli, Fidae, Es-Safi Abdelatif, and Ettaouil Mohamed. Novel method to optimize the architecture of Kohonen's topological maps and clustering.Logistics and Operations Management (GOL), 2014 International Conference on. IEEE, 2014.

[12] Jardak, Christine, Petri Mähönen, and Janne Riihijärvi. Spatial big data and wireless networks: experiences, applications, and research challenges.Network, IEEE 28.4 (2014): 26-31.