I J C T A, 9(26) 2016, pp. 257-265

Implementation of Bellman-Ford Algorithm for Vehicular Ad Hoc Networks using Big Data Techniques

V. Kalpana Mary[1] and M. Sughasiny[2]

ABSTRACT

Big Data technology is getting to be omnipresent and portraying key consideration of researchers in all territories. VANET is a special form of MANET that utilizations vehicles as hubs in a network. By applying Big Data advances to Vehicular Adhoc Network (VANET), one can increase helpful knowledge from a huge measure of operational data, to enhance movement administration procedures, for example, arranging, designing and operations. VANETs get to extensive data amid the continuous operations. In this paper we outline qualities to Big Data characteristics expressed in writing. Further, we assess the execution of Bellman–Ford calculation utilized for steering as a part of vehicular systems on Hadoop Map Reduce standalone disseminated framework and also on multi-node cluster with 2, 3, 4 and 5 hubs separately. The outcomes got affirm that expanding the quantity of hubs in Hadoop system, handling time for the calculation is extraordinarily diminished.

Keywords: Big Data; Hadoop; Map Reduce; VANETs; Bellman–Ford Algorithm; GSR; Shortest Path; Distributed computing.

1. INTRODUCTION

The exponential development regarding unpredictability and limit of data in most recent couple of years has prompted outstanding exploration in the field of Big Data technology. The wording of Big Data is getting to be universal these days. Big Data advances are still in the beginning phase of improvement. Big Data today, is the thing that the web was there in 1993. The data on the web is developing and the web may get greater and greater, yet not very many of us truly perceived what this "Huge" implied. Today, we trust that we aren't notwithstanding rubbing the surface of the Big Data opportunity. Huge Data is the most smoking catchphrase in current figuring environment. Worldwide Data Corporation (IDC) [1] characterizes Big Data as: "Large Data advances portray new era of technologies and frameworks intended to monetarily remove esteem from huge volumes of a wide assortment of data, by empowering high-speed catch, revelation and/or investigation".

Communication frameworks in Big Data technology are acknowledged by gathering and coordinating data from heterogeneous administrations. In heterogeneous systems administration environment, further examinations are expected to make technology more users centric. Research here incorporates data stockroom items with extensive scale conveyed data preparing stages, for example, Hadoop, and data scientific advancements, for example, machine learning and data mining.

Big Data is as an eminent technology for both professionals and specialists, which emerge critical difficulties for the educated community, industry, and different associations. There is a need of novel systems to oversee and investigate Big Data to make esteem that build the exactness of forecasts, enhance the

1 M.Phil. Research Scholar, Department of Computer Science, Srimad Andavan Arts & Science College, Trichy-05, Tamil Nadu, India.

2, Research Supervisor, Research Scholar, Department of Computer Science, Srimad Andavan Arts & Science College, Trichy-05, Tamil Nadu, India.

administration and security and empower educated basic leadership. Prior meanings of Big Data concentrated just on the organized data, yet now specialists have understood that the greater part of the data dwells in semi-organized or even in un frameworkd configuration (otherwise called multi-organized), basically as media, sensor data, XML archives, and long range interpersonal communication data to give some examples.

Vehicular Ad-hoc Network (VANET) an uncommon type of Mobile Ad-hoc Network (MANET) is a technology that has its roots in movement designing and has taken tremendous consideration lately. VANET abuses vehicles as hubs in a system and includes vehicle to vehicle (V2V) and vehicle to roadside base (V2I) remote communication [2]. Each sharing vehicle is transformed into a switch or remote hub that can interface and turn into a part of the system in the scope of 100 m to 300 m around. System is dropped out if there should be an occurrence of vehicles drop out of the extent. Any vehicles can participate in the system, on the off chance that it comes in the reach to shape a VANET. “Altered foundation has a place with the legislature or private system administrators or service suppliers” [3].

Propelling patterns in impromptu systems are conceived to prompt an accumulation of remote technology as a sort of Wi-Fi, Satellite, Cellular, Dedicated Short Range Communication (DSRC) [4] and WiMAX. VANET can be seen as the segment of the Intelligent Transportation System (ITS) [5].

Fig 1 demonstrates the run of the mill VANET Architecture, including vehicles and street side units (RSUs) that exchange for the most part wellbeing messages. It gives an ideal opportunity to respond forever imperiling occasions to the drivers. The fundamental objective of VANET is giving wellbeing and solace to individuals and products [6]. For this a unique electronic hardware is being set inside every vehicle to give specially appointed system network.

Rest of the paper is composed as takes after: Section II portrays the qualities of VANET and contends VANET as a Big Data issue. Big Data Analytics is talked about in segment III. Steering in VANETs utilizing Hadoop as trial study is portrayed in segment IV lastly area V finishes up the paper.

2. VANET AS A BIG DATA PROBLEM

With organizations now alternation to make computerized representations of existing data and obtaining everything that is most recent, data development rates in the course of the most recent couple of years have been high. Numerous commercial ventures fall under the shade of new data creation and digitization of existing data, and subsequently are getting to be suitable hotspots for Big Data assets. Indeed, nowadays about every one of the segments have begun receiving advanced representation of data, bringing about big data volume. Big Data has been extending quickly into the transportation field *i.e.* VANETS. The greater parts of the misgivings to VANETs are of worries in MANETs, yet their details fluctuate. Fairly walking at random way, vehicles grade to advance predetermined. The affiliations can be delineated all the more precisely among roadside units. The vehicles are delimited in their extent of movement, obliged to take after a cleared way.

This paper focuses on VANETs which get to expansive data progressively to plan and administration of sheltered and productive activity stream in transportation frameworks. “Big Data” is described at first by Doug Laney [1], a META Cluster and now Gartner expert as 3 V’s: Volume, Variety and Velocity in 2011. In 2012 IBM concluded two more V’s as Value and Veracity [7] along these lines making 5 V’s of Big Data [8]. These 5 V’s are currently recorded as: Volume, Variety, Velocity, Value and Veracity. The writing review by different writers [9], incorporate some exceptional qualities of VANETs. In this paper we have attempted to guide some of these VANET qualities to ascribe of Big Data to legitimize that VANET issues can be dealt with as Big Data issue and can be unraveled utilizing procedures of Big Data. Different VANET qualities we have mapped to the traits Volume, Variety, Velocity, Value and Veracity of Big are as given beneath:

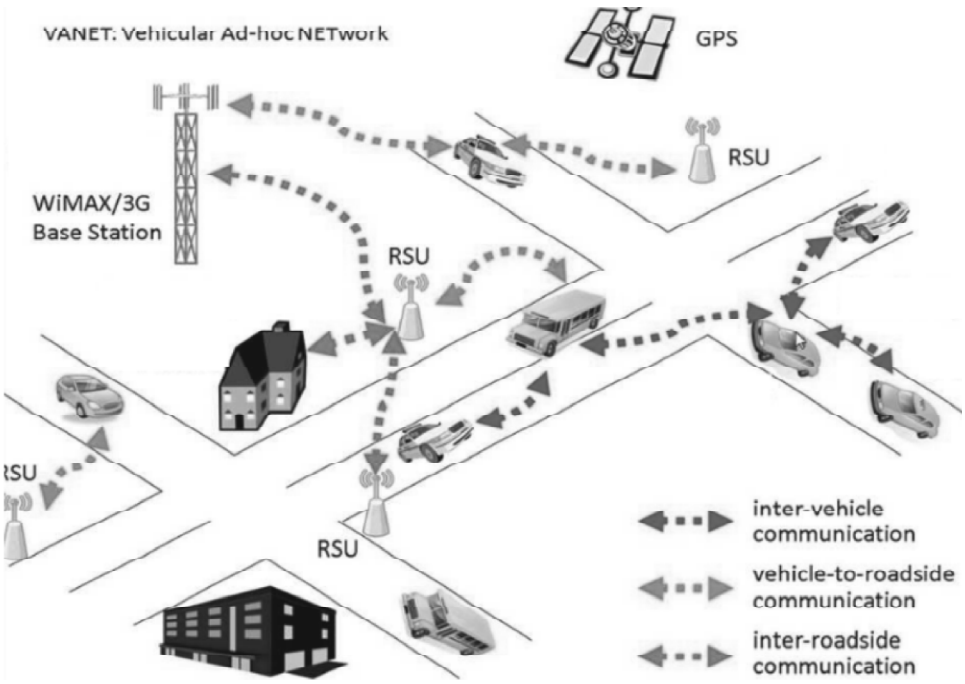

Figure 1: Architecture of Vehicular Ad Hoc Networks

Constant Data

VANET Data is ongoing and naturally redesigned after standard interims, put away in databases bringing about vast tables required for steering choices [10].

Variable system thickness

It relies on upon the exceptionally variable vehicular thickness. These Vehicles are outfitted with different GPS empowered sensor gadgets that are delivering distinctive type of data.

Highly Dynamic topology and Mobility modeling

In VANETs, the hubs allude to the vehicles that are versatile and exceedingly dynamic in nature [11]. It results in quick topology changes because of rapid that outcomes in continuous fracture as far as little compelling width and system thickness [9], [12].

Large Scale network and high computational ability

Many GPS empowered gadgets expand the computational limit of hubs in extensive scale systems [9], [13]. This requires the definite estimation of data for anticipating the steering choice.

Anonymous addressee and potential support from Infra-framework

Identification of vehicles in a specific zone support from the accessible base is the fundamental necessity of the most applications in VANETs. Subsequently there is a necessity from the neighboring hubs legitimate as they would like to think [10].

3. BIG DATA ANALYTICS TECHNIQUES FOR VANET

The conventional social databases can't prepare this vast volume of information. One needs new methods to handle extensive volume of information. Hadoop is an open stage to process huge measure of information

[14]. At the point when an organization needs to store gigantic measure of information they can embrace two alternatives: either get a major machine with additional CPU, RAM, circle space and so forth or counsel some database sellers for a greater arrangement. The two alternatives have their own issues. There is a farthest point on obtaining a major machine. Besides scaling is a testing undertaking. The second alternative derives scaling on a level plane. Arrangement gave by this alternative is not modest and requires huge venture and concentrated abilities. Today,

> *"the information is being produced not just by the clients and different applications and/or gadgets but at the same time is machine created and such information is exponentially driving the adjustment in the Big Data space. Managing huge datasets in the request of terabytes or even petabytes is extremely testing" [15].*

In Modern registering environment, a far reaching information preparing motor utilized for Big Data is Hadoop, an execution of the Map Reduce framework.

Hadoop give solid, shared and distributional capacity through Hadoop Distributed File System (HDFS) and appropriated figuring abilities through Map Reduce, a programming model.

A. Hadoop

Hadoop is an open-source system that procedure huge measure of information for creating and executing different appropriated applications. It indicates the strategy that appropriates over a swarm of machines in a cluster and that pushes examination code to hubs nearest to the information being dissected. Hadoop gives disseminated capacity called Hadoop dispersed document framework (HDFS) and appropriated processing through a programming model called Map Reduce [16], [17]. Hadoop is not an option for databases and information product houses. It gives a system to Big Data preparing. The handling of organized information is much less demanding in social databases

B. HDFS (Hadoop Distributed File System)

A HDFS cluster is working in expert laborer design with two sorts of hub: a NameNode (the expert) and various DataNodes (specialists). Document framework namespace is overseen by the NameNode. It additionally states the document framework tree and the metadata of the considerable number of records and catalogs in the tree. The workhorses of the record framework are DataNodes. They gather and recuperate squares when they are passed on to (by customers or the Name Node), and they pass on back to the NameNode occasionally with arrangements of obstructs that they are gathering.

The replication of information pieces is chosen by Name Node. The default piece size is 64MB in a common HDFS with a replication component of 3 (neighborhood rack holds second duplicate and remote rack holds third copy).The replication variable can be changed by client wish. HDFS have default replication arrangement such that each Data Node contains at most one reproduction of any piece and every rack contain at most two copies of the same square, the length of there are adequate racks on the cluster. Every piece has adequate number of reproductions is guaranteed by the NameNode. It will ceaselessly check whether the piece is over-reproduced or under-duplicated. In the event that it is over duplicated, NameNode will choose an imitation to evacuate. On the off chance that it is under-imitated, NameNode will make another copy in any hub or in the same hub. The imitation evacuation or copy creation is done taking into account the replication strategy.

C. Map Reduce

Map Reduce is a straightly adaptable programming model presented by Google that makes it simple to prepare in parallel hugely expansive information on extensive number of PCs. MapReduce works

fundamentally through two capacities, Map capacity and reduce capacity. Map capacity takes an arrangement of key/worth information and maps it into zero or more arrangement of key/quality. The diminish work then takes every exceptional key and gathering its related qualities into a one of a kind key/esteem set [18]. Map Reduce gives convenience, adaptability, and failover properties. Map work: The guide capacity is compressed as takes after [18],

"Guide, composed by the client, takes an info match and creates an arrangement of middle of the road key/esteem sets. The MapReduce library gathers together all transitional qualities connected with the same middle of the road key (I) and passes them to the diminish function."

We can further say, Map (k1, v1) → list (k2, v2)

The information keys and values have diverse spaces when contrasted with the areas to which the moderate keys and values have a place.

Reduce function: The diminish capacity can be outlined as takes after [19],

"The lessen capacity, additionally composed by the client, acknowledges a moderate key (I) and an arrangement of qualities for that key. It consolidates together these qualities to frame a potentially littler arrangement of qualities. Commonly only zero or one yield quality is delivered per lessen conjuring. The middle qualities are supplied to the client's lessen capacity by means of an iterator. This permits us to handle arrangements of qualities that are too expansive to fit in memory."

We can further say, Reduce (k2, list (v2)) → (k2, v3)

The halfway keys and values have the same area as the yield keys and values. Be that as it may, the first information keys and values have diverse keys and values when contrasted with the last yield keys and values.

D. Execution Overview of Hadoop

The Map Reduce library first parts the information records of User system into M bits of ordinarily 16MB to 64MB for each piece. At that point, numerous duplicates of the project on a cluster of machines are started. Every one of the capacities is controlled by the expert. Whatever remains of the hubs that are doled out work by the expert are laborers. There are M map errands and R diminishes undertakings to dole out. The unmoving specialists are picked by the expert and appoints everyone a guide errand or a diminish assignment. The dividing capacity then apportioned them into R locales. The yield of the Map is put away in the neighborhood plate transiently which is evacuated when the Reduce capacity is finished. This makeshift storage of Map yield helps in adaptation to internal failure and recuperation. That is the reason, when a Reduce capacity comes up short, the framework need not begin the Map procedure once more. The Map

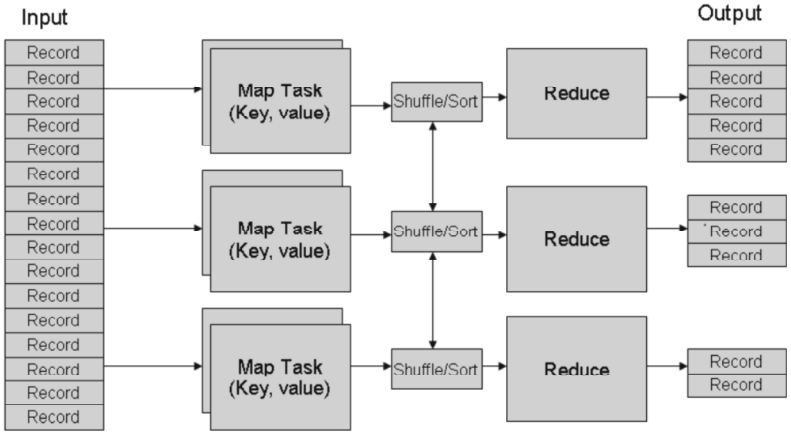

Figure 2: Architecture of Map Reduce

yield is then passed back to the expert on the neighborhood plate. The expert is at risk for diverting these areas to the lessen specialists. In the wake of getting a notice about these areas from the expert, the diminish laborer read the supported information from the nearby plates of the guide specialists through remote strategy calls. The expert fortifies the client program after the fulfillment of all guidance and reduces undertakings. The ordinary Map Reduce system is appeared in Fig 2.

4. ROUTING IN VANET USING HADOOP

The objective of VANET is to set up a productive course between hubs *i.e.* vehicles furthermore to alter this effectively to vary topology of moving vehicles. Hence, Routing has dependably been a noteworthy test in VANETs because of portability and successive changes in the system topology. To accomplish this a steering convention is required that can coordinate the path through which the information can be traded between two communication elements and joins the system of a course foundation, sending choice information and activity in keeping up the course and/or recouping from the directing disappointment. The Bellman–Ford calculation is a calculation that registers most limited ways from a solitary source vertex to the greater part of alternate vertices in a weighted digraph. It is slower than Dijkstra's calculation for the same issue, yet more flexible, as it is equipped for taking care of charts in which a portion of the edge weights are negative numbers. [20], [21].

> *"For a given source hub in the chart, the calculation finds the way with most minimal expense (*i.e. *the most brief way) between that hub and each other hub. For instance, if the hubs of the chart speak to urban communities and edge way costs speak to driving separations between sets of urban areas associated by an immediate street, Bellman–Ford's calculation can be utilized to locate the most brief course between one city and whatever other city" [21].*

A noteworthy suspicion for the diagram look calculation is to have a solitary source that is taking care of most brief way issue for a chart with nonnegative edge way costs and creating a most brief way tree. Different suspicions incorporate the accessibility of adequate number of courses, communication joins, nearness of two way radio, framework, for example, GPS, presence of Road side units (RSU) at proper positions for engendering of sign. The fundamental suspicion is that VANET system topology can be seen as a coordinated chart with vehicles speaking to hubs and most limited way speaking to joins in the diagram. The Bellman–Ford's most brief way calculation is a sort of Geographic Source Routing (GSR) calculation in light of position for VANETs [22].

> *"GSR expects the guide of a road map in city situations to know the city topology. GSR consolidates geographic directing and topological information from road maps. The sender decides the intersections that must be navigated by the parcel utilizing the Bellman–Ford's most limited way calculation and after that forward the bundle in a position-based style between the intersections" [23].*

This calculation can be productively coded utilizing Big Data investigation. In this paper, the execution of Bellman–Ford calculation in Vehicular Ad-hoc Networks is assessed on Big Data system *i.e.* Hadoop stage open environment running in standalone hub and also on multi-hub cluster with 2, 3, 4 and 5 hubs individually. The outcomes are additionally thought about utilizing Net Beans IDE 7.4 environment running on a Windows stage. The outcomes as far as preparing time are arranged by expanding the quantity of hubs in the framework. The calculation was keep running for a chart comprising of 100 hubs at first Then we expanded the quantity of hubs with a stage of 100, *i.e.* 200, 300, 400 and 500 in each consequent emphasis.

5. IMPLEMENTATION RESULTS AND DISCUSSIONS

The Bellman–Ford's algorithm was executed in JAVA for single standalone dispersed Hadoop map-reduce framework and assessed for the given number of hubs then further assessed on multi-node completely appropriated Hadoop cluster environment with 2, 3, 4 and 5 hubs individually. In Map Reduce programming,

the task is performed by dividing the whole program into smaller program and executing them independently. Numerous information examination calculations are included in parallelization of Map Reduce strategy for ease-ofuse and viability.

"An imperative normal for Hadoop is the apportioning of information and calculation across thousands of hubs, and the execution of use calculations in parallel near their information. A Hadoop cluster scales computation limit, storage limit and I/O bandwidth by simply adding commodity servers" [19].

It can utilize the existing product of hardware and in this manner the expense of setting Hadoop cluster is generally low. For running a Map Reduce program a little Hadoop cluster is required that may incorporate a solitary expert hub and various worker hubs.

"The expert hub comprises of a JobTracker, TaskTracker, NameNode and DataNode. A slave or specialist hub goes about as both a DataNode and TaskTracker, however it is conceivable to have information just laborer hubs and process just laborer hubs" [19].

The stream of steps required in MapReduce programming model are as per the following [19]:

(i) Cycle over the info

(ii) Calculation of key/quality sets from every bit of info

(iii) Gathering of every single transitional worth by key

(iv) Cycle over the subsequent gatherings

(v) Reduce of every gathering'

The calculation gets an arrangement of information key/esteem matches, and yields an arrangement of yield key/esteem sets. Two capacities, guide and lessen, are utilized to express the calculation. In this execution, we took beginning hub as the key and separation it focuses to as the worth in Mapper part and current hub as the key and least of all the conceivable separations to this point as the quality in Reducer part for key/esteem pair.

Results are organized in Table 1 which demonstrates the time taken (in Seconds) for a vocation in various cases with expanded number of hubs. It was watched that in Net Beans IDE 7.4 with 100 hubs, the employment takes 1.1 seconds to finish, while in Hadoop standalone hub it takes 1.2 seconds to finish the same assignment. Further Hadoop appropriated framework with 2, 3, 4, 5 hubs it takes 1.2 seconds, 1.1 second, 1.0 second and 0.8 seconds separately to execute the employment.

It is construed from the outcomes that when the quantity of hubs in chart builds, the preparing time of the occupation in Hadoop environment tends to diminish with the expanded number of cluster hubs because of conveyed handling of the information on various hubs and the distinction in handling time increments with existing Java situation as the quantity of hubs in Hadoop environment increment.

Table 1
Time Taken (in Seconds) by Bellman–Ford's Algorithm

	Time Taken (in Seconds) by Map Reduce Hadoop Distributed System with number of nodes				
Number of Nodes in the system	*Single Node (hub)*	*2 nodes*	*3 nodes*	*4 nodes*	*5 nodes*
100	1.2	1.2	1.1	1	0.8
200	1.8	1.7	1.6	1.4	1.1
300	2.3	2.1	1.8	1.7	1.5
400	2.9	2.6	2.3	2.1	1.9
500	3.1	2.9	2.6	2.5	2.2

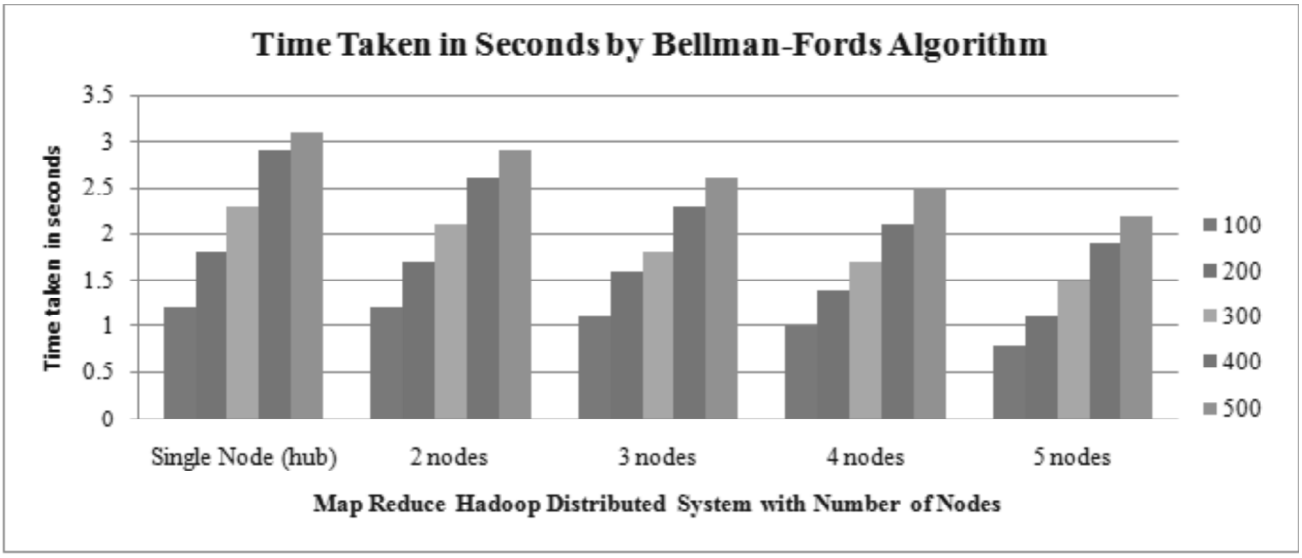

Figure 3: Graphical Representation of Time Taken (in Seconds) by Bellman–Ford's Algorithm

6. CONCLUSIONS

Big Data is receiving the consideration of investigators in all regions as a lot of information is being made by various day-today applications nowadays. Volume, Variety, Velocity, Veracity and Value are the characteristics associated with Big Data in composing. In this paper, the issue of VANET is discussed and the mapping of some VANET qualities to Big Data attributes is done to substantiate that VANET issue can be considered as a Big Data issue and distinctive techniques for Big Data can be associated with VANET information remembering the deciding objective to upgrade development administration shapes In this paper, Bellman–Ford's count used as a piece of Vehicular Ad-hoc Networks for controlling is taken as a relevant examination to be comprehended using Big Data propels. The execution of Bellman–Ford count on Big Data framework using standalone passed on Hadoop Map Reduce environment and what's more Hadoop scattered environment with 2, 3, 4 and 5 centers is moreover assessed. The results procured from the trials were differentiated and eventual outcomes of routine NetBeans IDE environment with respect to increase in the amount of center points in the framework. It was found that extending the amount of gathering center points using Hadoop framework, taking care of time for the guiding count is fundamentally reduced.

REFERENCES

[1] Gantz, J., & Reinsel D, "The 2011 Digital Universe Study: Extracting Value from Chaos", 2011.

[2] Danda B Rawat, Gongjun Yan, Bhed Bista, "Trust on the Security of Wireless Vehicular Ad Hoc Networking", *Ad Hoc and Sensor Wireless Networks*, 283-305, January 2015.

[3] Consortium, C.-t.-C. C. Retrieved from http:/www.carto- car.org, April 2014.

[4] Home, D. S.. Retrieved from http://www.leearmstrong.com/DSRC/DSRCHomeset.htm, 2014.

[5] Chirstos Bouras, Vaggelis Kapoulas and Enea Tsanai, "Performance Evaluation of Routing Mechanisms for VANETs in Urban Areas", *Springer*, 133-135, 2015.

[6] Dhyey Patel, Mohammed Faisal, Priyanka, Sidharth, M. Mani Roja, "Overview of Routing Protocols in VANET", *International Journal of Computer Applications*, 4-7, February 2016.

[7] Demchenko, Y., Worring, M., Los, W., & Laat, C. d. (2013, September 16). 2013-09-16-rda-bdaf.pdf. Retrieved December 11, 2013, from http://www.delaat.net: http://www.delaat.net/~cees/posters/2013-09-16- rda-bdaf.pdf.

[8] Taimoor Abbas, Katrin Sjoberg, Johan Karedal and Fredrik Tufvesson, "A Measurement based Shadow Fading Model for Vehicle-to-Vehicle Networks Simulations", *International Journal of Antennas and Propagation, Hindawi Publishing Corporation*, 2015.

[9] Wenshuang Liang, Zhuorong Li, Hongyang Zhang, Shenling Wang and Rongfang Bie, "Vehicular Ad Hoc Networks: Architectures, Research Issues, Methodologies, Challenges and Trends", *International Journal of Distributed Sensor Networks*, 2015

[10] Shaima M. Shakeel, Osama M.H. Rehman, Mohamed Ould-Khaoua, Hadj Bourdoucen, "Open Soure Software Support for Field Experiments of Vehicular Ad Hoc Networks", *Free and Open Soure Software Conference*, Muscat, 69-72, February 2015.

[11] Kamran Zaidi and Muttukrishnan Rajarajan, "Vehicular Internet: Security & Privacy Challenges and Opportunities", *Future Internet*, 257-275, 2015.

[12] Iftikhar Ahmad, Uzma Ashraf and Abdul Ghafoor, "A Comparative QoS Survey of Mobile ad hoc Network Routing Protocols", *Journal of the Chinese Institute of Engineers*, 2016.

[13] Amir Gandomi, Murtaza Haider, "Beyond the Hype: Big Data Concepts, Methods and Analytics", *International Journal of Information Management*, Volume 35, Issue **2**, 137-144, April 2015.

[14] Barské, A. (2013). *Big Data Whitepaper*,. Erdogan: Arzu Barské - Erdogan 2013 © version 2.0.

[15] Warden, P. (2012). "Big Data Glossary". Sebastopol: O'Reilly Media, Inc.

[16] Hai Wang, Zeshui Xu, Hamido Fujita, Shousheng Liu, "Towards Felicitious Decision Making: An Overview on Challenges and Trends of Big Data", *Information Science*, Volume 367-368, 747-765, November 2016.

[17] Brijendra Singh, Rohit Miri, "A Review Paper on Parallel Association Rules Mining Algorithm in Data Mining and MapReduce Framework", *Data Mining and knowledge Engineering*, 2016.

[18] Lämmel, R. (2008), "Google's MapReduce programming model—Revisited". *Science of computer programming*, **70(1)**, 1-30.

[19] Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. "Introduction to Algorithms". *MIT Press and McGraw-Hill*, 2001.

[20] D.Jinil Persis and T. Paul Robert, "Ant Based Multi Objective Routing Optimization in Mobile Ad Hoc Networks", *Indian Journal of Science and Technology*, 875-888, 2015.

[21] Shilpy Agarwal, R.S.Raw, Neeraj Tyagi and A.K. Misra, "Fuzzy Logic based Greedy Routing (FLGR) in Multi Hop Vehicular Ad Hoc Networks", *Indian Journal of Science and Technology*, 2015.

[22] B, S. S., G, N. K., Rani, U. H., K, D. C., George, G., & Murali, A. (2012). "GPS Based Shortest Path for Ambulances using VANETs". 2012 *International Conference on Wireless Networks (ICWN 2012)*. Singapore: IPCSIT vol. 49, IACSIT Press, 190-196.

[23] Ryu, M. W., & Cho, K. H. (2012). "A Survey of Greedy Routing Protocols for Vehicular Ad Hoc Networks". *Smart Computing Review*, vol. 2, no. **2**, 125-137.