ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

IMPROVING BIG DATA EFFICIENCY BY COMPRESSION AND TIME MANAGEMENT APPROACH ON CLOUD

[1] Leelambika.K.V, [2]R.Monica,[2]B.Monisha,[3] R.Divyalakshmi

[1]kvleelambika@gmail.com, [2]monica25594@gmail.com, [2]monisha6994@gmail.com,[3]r.divyalakshmi95@gmail.com

[1,2,3]Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College.

ABSTRACT:

Cloud computing has become a mainstream solution for data processing , storage and distribution. Cloud computing is the commodification of computing time and data storage by means of standardised technologies.
But moving a large amount of data in and out of the cloud presented a kind of challenge. Moving large amount of data can sometimes result in, inefficiency in data transmission, loss of data, bad quality , delay in time and taking more time not getting acknowledgement. Data compression approach not only reduce the Quantity of data but inturn reduces the time taken to execute a data. Since,small amount of data can be executed at a faster rate than the large amount of data . Data compression can reduce the disadvantage, occuring without a data compression.The works can be divided and the divided works are allocated with a specific time during the specific time the divided works can be worked out efficiently.For efficient data processing, scheduling approach is given for dividing the works into separate part executing at a specific time.The workloads can be divided in the form of cluster and the data contained in the cluster can be shared or managed at the same time thereby we are able to save time.The experiment results, shows that the data compression and the time management greatly increases the efficiency of the bigdata on the cloud than the normal techniques used in the cloud.

1.INTRODUCTION:

Since there is a fast development in the modern technology area, we are destined to save large amount of data and that is where the problem arises, the problem is that when we keep on storing data there is a problem of experiencing data explosion which causes problems in operating the data. Since the data is more , the execution of the data takes more time, that is processing big data becomes difficult. Big data is being generated by everything around us at all times. Every digital process and social media produces it and the Systems, sensors , mobile devices transmit it. Big data is arriving from multiple sources at an alarming velocity, volume and variety. To extract meaningful value from big data, we need optimal processing power, analytics capabilities and skills.Cloud computing is the practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer.In the simplest terms, cloud computing means storing and accessing data and programs over the Internet instead of your computer's hard drive. In cloud computing," we can access data or programs over the Internet, or at the very least, have that data synchronized with other information over the Web.

II. COMPRESSION APPROACH:

A.Compression approach by clustering:

With the help of the clustering also we are able to do the compression process , the clustering we use work with time series , the time series works with regression which helps in making the data transfer very easier in a cluster manner . The exchange of data among the clusters also works very well , during the data exchange the data exchange process which is going to make the data exchange is reduced , which is done by the

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

clustering. So whenever there is a reason for the occcurence or chance for the occurrence of the clustering which is taking place then the process which is actually making it to reduce the clustering by the interference process data exchanging.

B. Influence on Clustering Method:

The clustering algorithm we are going to create is based on the fact that it always check whether the two time series are similar , the similarity makes the clustering to execute by nodes set partitioning ,the participants of the data or linear arrangement what ever may be the result will be the reduced fro m the execution of the time .The results of the two time series can be a high level similarity , but the time series can be totally different.

C. Clustering by changing data:

In this technique we specify which is suitable for the clustering algorithm. Here however the data changes, Under extreme conditions it can make all the leaf node the head. Proposed clustering algorithm can be ineffective to provide .But using compensation method, the compensation by still saving the environment. The order of compression can be started and status is that they where able to complete the work, it can happen in a complex environment.

D. Order compression with Compression :.

Compressing multiple attributes with data information with the order, there are only 2posiblilities, they are

(1) transmit the data to the parent bode.
(2) Supress the data document into predefined error change

E. Data driven scheduling on the cloud:

In order to show or get the efficiency of the approach is easy

Two mapping strategies are

They are, with real network nodes while the other is the exchange edges and also with data exchange quality

V.DETAILED APPROACH OF THE COMPRESSION:

Order compression is the time where we learned about the relationship between the compressed

On processing with scheduling with edges is that the process during that we should able and should execute the edges repeatedly

After getting a long list of information from this company

VI CONCLUSION:

Thus we reduce the data size compared to the previous big data processing techniques on Cloud.To reduce the quantity and the processing time Wemade this paper which could reduce the quantityInturn it reduces processing time, Since less quantity takes less time to execute Already, reduce the data size compared to the previous big data processing techniques on Cloud data processing quality.

Cloud computing has become a viable, mainstream solution for data processing, storage and distribution, but moving large amounts of data in and out of the cloud presented an insurmountable challenge for organizations with terabytes of digital content.

Traditional WAN-based transport methods cannot move terabytes of data at the speed dictated by businesses; they use a fraction of available bandwidth and achieve transfer speeds that are unsuitable for such volumes, introducing unacceptable delays in moving data into, out

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

of, and within the cloud. Asper has solved the big data challenge Already, Cloud computing is the modification of computing time and data storage by means of standardized technologies

REFERENCES:

[1] G. Varaprasad and R. S. D. Wahidabanu, "Flexible routing algorithm
for vehicular area networks," in *Proc. IEEE Conf. Intell. Transp. Syst.*
Telecommun., Osaka, Japan, 2010, pp. 30–38.

[1] S. Tsuchiya, Y. Sakamoto, Y. Tsuchimoto, V. Lee, Big data processing in Cloud environments, Fujitsu Sci. Tech. J. 48 (2) (2012) 159–168.

[2] C. Ji, Y. Li, W. Qiu, U. Awada, K. Li, Big data processing in Cloud environments, in: 2012 International Symposium on Pervasive Systems, Algorithms and Networks, 2012, pp. 17–23.

[3] Big data: science in the petabyte era, Nature 455 (7209) (2008) 1.

[4] Balancing opportunity and risk in big data, a survey of enterprise priorities and strategies for harnessing big data, http://www.informatica.com/Images/1943_big-data-survey_wp_en_US.pdf, accessed on March 01, 2013.

[5] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, M. Zaharia, A view of cloud computing, Commun. ACM 53 (4) (2010) 50–58.

[6] R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, I. Brandic, Cloud computing and emerging it platforms: vision, hype, and reality for delivering computingas the 5th utility, Future Gener. Comput. Syst. 25 (6) (2009) 599–616.

[7] L. Wang, G. Von Laszewski, A. Younge, X. He, M. Kunze, J. Tao, C. Fu, Cloud computing: a perspective study, New Gener. Comput. 28 (2) (2010) 137–146.

[8] L. Wang, J. Zhan, W. Shi, Y. Liang, In cloud, can scientific communities benefit from the economies of scale?, IEEE Trans. Parallel Distrib. Syst. 23 (2) (2012) 296–303.

[9] X. Yang, L. Wang, G. Laszewski, Recent research advances in e-science, Clust. Comput. 12 (4) (2009) 353–356.

[10] S. Sakr, A. Liu, D. Batista, M. Alomari, A survey of large scale data management approaches in cloud environments, IEEE Commun. Surv. Tutor. 13 (3) (2011) 311–336.

[11] B. Li, E. Mazur, Y. Diao, A. McGregor, P. Shenoy, A platform for scalable one-pass analytics using MapReduce, in: Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD'11), 2011, pp. 985–996.

[12] J. Dean, S. Ghemawat, MapReduce: simplified data processing on large clusters, Commun. ACM 51 (1) (2008) 107–113.

[13] K.H. Lee, Y.J. Lee, H. Choi, Y.D. Chung, B. Moon, Parallel data processing with MapReduce: a survey, SIGMOD Rec. 40 (4) (2012) 11–20.

[14] Hadoop, http://hadoop.apache.org, accessed on March 01, 2013.

[15] X. Zhang, C. Liu, S. Nepal, J. Chen, An efficient quasi-identifier index based approach for privacy preservation over incremental data sets on Cloud, J. Comput. Syst. Sci. 79 (5) (2013) 542–555.

[16] X. Zhang, C. Liu, S. Nepal, W. Dou, J. Chen, Privacy-preserving layer over MapReduce on Cloud, in: The 2nd International Conference on Cloud and Green Computing (CGC 2012), Xiangtan, China, 2012, pp. 304–310.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

[17] X. Zhang, C. Liu, S. Nepal, S. Pandey, J. Chen, A privacy leakage upper-bound constraint based approach for cost-effective privacy preserving of inter-mediate datasets in Cloud, IEEE Trans. Parallel Distrib. Syst. 24 (6) (2013) 1192–1202.

[18] X. Zhang, T. Yang, C. Liu, J. Chen, A scalable two-phase top-down specialization approach for data anonymization using MapReduce on Cloud, IEEE

Trans. Parallel Distrib. Syst. 25 (2) (2013) 363–373.

[19] C. Yang, Z. Yang, K. Ren, C. Liu, Transmission reduction based on order compression of compound aggregate data over wireless sensor networks, in:
Proc. 6th International Conference on Pervasive Computing and Applications (ICPCA'11), Port Elizabeth, South Africa, 2011, pp. 335–342.

[20] S.H. Yoon, C. Shahabi, An experimental study of the effectiveness of clustered aggregation (CAG) leveraging spatial and temporal correlations in wireless sensor networks, ACM Trans. Sens. Netw. (2008) 1–36.

[21] P. Edara, A. Limaye, K. Ramamritham, Asynchronous in-network prediction: efficient aggregation in sensor networks, ACM Trans. Sens. Netw. 4 (4)
(2008), article 25.

[22] M.J. Handy, M. Haase, D. Timmermann, An low energy adaptive clustering hierarchy with deterministic cluster-head selection, in: Proc. 4th International
Workshop on Mobile and Wireless Communications Network (MWCN), 2002, pp. 368–372.

[23] A. Ail, A. Khelil, P. Szczytowski, N. Suri, An adaptive and composite spatio-temporal data compression approach for wireless sensor networks, in: Proc.
of ACM MSWiM'11, 2011, pp. 67–76.

[24] R. Kienzler, R. Bruggmann, A. Ranganathan, N. Tatbul, Stream as you go: the case for incremental data access and processing in the cloud, in: IEEE
ICDE International Workshop on Data Management in the Cloud (DMC'12), 2012.

[25] P. Bhatotia, A. Wieder, R. Rodrigues, U.A. Acar, R. Pasquin, Incoop: MapReduce for incremental computations, in: Proceedings of Proceedings of the 2nd ACM Symposium on Cloud Computing (SoCC'11), 2011, pp. 1–14.

[26] C. Olston, G. Chiou, L. Chitnis, F. Liu, Y. Han, M. Larsson, A. Neumann, V.B.N. Rao, V. Sankarasubramanian, S. Seth, C. Tian, T. ZiCornell, X. Wang,
Nova: continuous Pig/Hadoop workflows, in: Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD'11), 2011,
pp. 1081–1090.

[27] S. Lattanzi, B. Moseley, S. Suri, S. Vassilvitskii, Filtering: a method for solving graph problems in MapReduce, in: Proc. 23rd ACM Symposium on
Parallelism in Algorithms and Architectures (SPAA'11), San Jose, California, USA, 2011.

[28] J. Conhen, Graph twiddling in a MapReduce world, Comput. Sci. Eng. 11 (4) (2009) 29–41.

[29] K. Shim, MapReduce algorithms for big data analysis, Proc. VLDB Endow. 5 (12) (2012) 2016–2017.

[30] Big data beyond MapReduce: Google's big data papers, http://architects.dzone.com/articles/big-data-beyond-mapreduce, accessed on March 01, 2013.

[31] Real time big data processing with GridGain, http://www.gridgain.com/book/book.html, accessed on March 03, 2013.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
Available online at www.ijarbest.com

International Journal of Advanced Research in Biology Ecology Science and Technology (IJARBEST)
Vol. I, Special Issue I, August 2015 in association with VEL TECH HIGH TECH DR. RANGARAJAN DR. SAKUNTHALA ENGINEERING COLLEGE, CHENNAI
National Conference on Recent Technologies for Sustainable Development 2015 [RECHZIG'15] - 28th August 2015

[32] Managing and mining billion-node graphs, http://kdd2012.sigkdd.org/sites/images/summerschool/Haixun-Wang.pdf, accessed on March 05, 2013.