International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06, 2020
ISSN: 1475-7192

Updating of data through atomic ring Model in Big data

[1]P. Kirubanantham, [2]C. Jeeva,[3] B. Gracelin Sheena

ABSTRACT-- *Big data is the collection of huge data that is been transacted and analyzed in day to day basis. Data analysis is the huge task to be manipulated in time to time basis. Updating and prediction of data is the uphill task and it is a greatest challenge to provide a updated information to the user. Atomic ring based pattern will provide a differentiation between old data and updated data among the data to the user. Pruning process will remove the older data when the new data gets updated.*

Keywords--*Atomic ring structure, big data, data update****.***

I. INTRODUCTION

Big data is data centric technology which needs a procedure of updating of data at once while updating the data it has issues based on the nature of the data novice. As the data manipulation large in the Big Data it has number of challenges to be sorted out with solutions. Those challenges are discussed in this paper

Major challenges can be listed as

Data safety

Data seclusion

Data segregation

Data Streaming

- ***Data safety***

As the big data is the technology that grows day by day in enormous manner the security is one of major issue. In that as data flow increases in time basis providing a security in time concern is major issue and research are going on for sorting out a solution for it. [1]

- ***Data seclusion***

Data seclusion is the privacy the data owner who expect. Even his personal data is exposed to world when comes as data in the internet arena. Privacy is one of the issue that the technology is tending to find a better solution for it. [1]

[1]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, Tamilnadu, India, kirubanp@srmist.edu.in.

[2] Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, Tamilnadu, India, jeevac@srmist.edu.in.

[3] Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, Tamilnadu, India, gracelib@srmist.edu.in.

Received: 22 Sep 2019 | Revised: 13 Oct 2019 | Accepted: 15 Jan 2020

. Data Segregation

Data segregation is the form of data sorting in large manner for the big data .It's not an easy task as the tremendous data are pouring in second to second. Ordering the data based on time period is the major task. To be sorted as solution in the big data technology. This data segregation is taken as a work for this research proposal.

. Data Streaming

Data streaming is the one of the property of the big data in data handling technique. The streaming of the data should be handled with care, if there is any delay in streaming the quality of service will be considered a major issue. To rectify the problem researchers are sorting with certain number of challenges to be met out. The proposed work will try to get solution for the streaming based issues.

II. DATA STREAMING CHALLENGES IN BIG DATA

Time factor: The system of the big data is designed in which to handle hefty amount data in short period of time in effective manner. This is one concern In another concern the data should be accuracy and usable for the other process. Till now the time based data handling procedures are not effective .It lags in either in data synthesis or in time. Speed of data accumulation varies sector to sector and field to field. For this concern a methodology is proposed to handle the accuracy of the data and accumulation of the data in better manner. Processing time of the data is taking a long time for processing large data that is accumulating in very short period of time. The data is also changing in varying speed .It is a major task that a streaming algorithm faces .The proposed algorithm takes this challenge to solve this by segregating the data based on the time is generated .This keeps the data in different levels to manipulate and retrieve whenever it needs in required form.

Relaying on old streaming and clustering Techniques : This is an factor that can also be consider as an challenges to solve .The old streaming methods don't have proper ordering methods to sorting the data based on the time of data developed or generated .Clustering is not so accurate . The similar name based different ideology is considered as a similar character and grouped in a single common domain. This possess a challenges to solve. This issue is solved through new proposed algorithm in that .New algorithm separates the old data with new data based on the time data is made update and generated .By this a clarity is obtained among the data.

Scalability based hurdles: generally the big data want to handle huge amount of data, In the limited time period. Big data feels the heat at present to handle huge amount of data in a short limited period .This issue is somewhat of major task to meet with, with some hefty work. Because the networking of systems to make a reach among the people us and also huge. Categorizing of data based on time and time of data made to be update and changes based on new time is factor to sort the data, in which is the one of the major function of big data .The proposed system will try to separate the data based on time it is generated.

Security of the Data: Privacy of the data is one of the major concern that cannot be avoided while processing the data. Without the privacy of the data .The data can be tampered or it can be eavesdropped. The chance of masquerade attack has much possible in this point of big data, data manipulation .The most expected research proposal should meet out the security breaches challenges. The proposed work is tailor made to meet out data

streaming based concern. Work has the provisions of security of encrypting data, while the data is been moved from one level to another. While data is updated.

III. DATA SEGREGATION IN BIG DATA

Separating the data based on the data nature is segregating the data in this data segregation can be done based on many categories those are nature of the data, time of data update, type of data. This three categories of data are taken into consideration issues is been identifies and solutions are sorted out.

Data Segregation based on Nature of Data

Data segregation data are many types based on the sectors the data are identified in an organization itself there are different types of data can be identified and it can be sorted out. Sector wise data are varied .based on the field the data can be varies that data are to be separated based on their nature.

Data Segregation based on type of Data

It is general, the data segregation is made based on the type of the data, enormous segregation can made through type of data .Data are in the form of document image or in the form of media files. Segregation of data is happens to make the data handling in a simpler manner.

Data Segregation based on Time

Data segregation can also be done based on the time data got generated. Data newness has been considered as a one of the important feature in the big data environment for the data users. Newness of the data is always connected with information system success. An authorized agent for the data newness should be identified to prove the data quality from the particular person or agency.This kind of data segregation is most hectic to segregate, because this segregation of data changes time to time. Data update will be pouring in seconds. A model is required to monitor the data movement so a model is proposed based on Atomic ring structure type.

Generally time of the data is essential as the time period of data updating that takes place based on the time period of data change can be categorized as second, minute and hour based time data update. Based on this sequence time based data segregation can be sorted as.

(i)Second based data update

(ii)Minute based data update

(iii)Hour based data update.

(i)Second based data update

Some of the data update normally change time to time for example when it is considered some scientific results show difference in seconds and in microseconds too. In sports lie athletes the results are changes in time seconds. The measured index points show the time difference in seconds to seconds. That update change can be done through big data model through data segregation.

Received: 22 Sep 2019 | Revised: 13 Oct 2019 | Accepted: 15 Jan 2020

(ii)Minute Based data Update

Minute based data update is somewhat normally happens in most of the fields and its expected one also .In most of the sports events the minute based data change is expected. Other than sports event .Normally news update is made in the time span between 5, 15, 30, 45, minute.

This time change record is essential that to be update in proper sequence manner, that data also makes a considerable data change and value change. If time sequence changes.

(iii)Hour based data Update.

Hour based data update can expected while counting election votes .This makes considerable change in data number that to be considered as a proper data change time sequence, based on that also data update is considered . In normal news update happens in hour basis .This also considered for Time based Data Segregation for proper differentiation to identify which data is earliest and which is in old in the record is recorded .For analysis purpose in big data.

To classify the data according to time sequence a novel algorithm can be designed and compared with prominent algorithm based on the proposed algorithms techniques.

Different types of study is made on time based data .The time based data needs the accuracy and newness of the data .to transfer the data from one person to another old data lacks accuracy. An auditing tool should be considered to differentiate the newness of the data by comparing with the existing data. This technique should reveal the novelness of the data to the user.

In this proposed work the dimensions of the novelness of the data is measured in the basis of time of freshness. As we consider for any food product that is bought from the market , its freshness is evaluated as 1 day fresh product , 1 week old product or 1 month old produt .Likewise the freshness or novelness of the data is measured as 1 minute before data , 10 mins before data or 1 hr before data.

IV. DATA NOVELNESS

Data novelty is described here with some measuring aspects, technology and analysis .Some areas that affects the analysis and incorporating o data novelty. A overall study and summarization is analysed of concept workout feasibility.

In another aspect the novelness of the data is measured with certain measuring aspects or scales those aspects and scales are defined here. As currently data, obsoleteness of data, Novelness of the data, Timeliness of the data.

Currently Data; this aspect is time based, the actual time of the data revealed and the current nature of the data .The difference of the data period is measured here as n.

Obsoleteness of Data; this factor of measuring aspect is meant to measure the accuracy of the data. This may differ from source to source. This also is one of the important factor to be considered for measuring the nature of the data.

Novelness of the Data; This also comparison on basis of tuples level .The comparison is the level of various attributes of database. This will be an accurate mode of measuring the novel nature of a data in root level.

Deviations in the data is also identified in this way of measuring approach. Deviations are mentioned in percentage basis.

Timeliness of Data: This factor measures both the features like query time and nature of the data at the period of query. This somewhat different aspect, because mostly answers are expected for query within fraction second. But in data collection query may take its own time period to receive the answer. So the user have to compare the difference in the data at the time of the query. Arise and at the time answer received. So this difference in data is considered as a factor of measurement.

The data from the source is changeable in nature. Based in the different types of views and other supporting parameters the consistency of the message or the data is not trustworthy or it's not maintaining the accuracy as the user expects, In this scenario novelty of the data is definitely expected for the other related works.

Here another two factors can be considered as metrics for measuring the novlness of the data .They are Caching and Virtual systems .These two metrics are in opposite in its nature but measuring novelness of the data.

Caching: Here the data is considered as novel in nature and data novelness is considered as current data, so while refreshing the last data which was stored as last as updated data .This caching will update the last data which was finally stored before refreshment.

Virtual system: It is quite different than caching, here novelness of the data is not much important, but the time span is given a preference. The updated information for last one hour or a particular time period is taken into consideration mostly minimum of time span is taken ,it vary according to the field of domain it varies .This is taken into consideration and presented as new, till the time interval, then it is renewed.

The time span of novelness of the data is categorized in three different types based on their duration of novel character they are Constant data, fast changing data, duration based changing data.

Constant data; these data would not change mostly like date of birth, street name like wise, some rules numbers.

Fast changing data ; These data will changed with in particular short time duration minimum within seconds maximum with in a day .These data are been handled mostly with quick modulations daily. Maximum all the data are categorized in this data.

Duration based changing data; these data takes more than a day to change, weekly data, and prices of consumer commodities some stock exchange share.

V. SYSTEM BASED SUPPORT

The novelness of the data is sometimes based on the user the extracted information need to be properly presented to the user in accepting format It has some process like extraction ,integration and delivery in the form the user accepts. As the user access of the data is based on the query process .The answers to the queries of the user can be categorized into three different categories they are.

Cache answers, virtual answers, materialized answers. These are different form of answering to the queries of the user. To meet their query and to accept the answer.

Cache answers: cache answers are which is based on the multiple queries from the different users are categorized into a set of the questions in the name of frequently asked questions. Whatever the form of question.

It is made to be answered with certain set of answers. Which will not deviate with question that is raised by the user.

Virtual answers: Here whatever the query is the system will try to get answers from the sources in web or from different sources. If the system receives different types or different answers from the sources. The answering systems will correlate all the answers and reply with a single form of answer in user accepting format.

Materialized answer: In this form of answer are direct answering form for the queries of the users. As a search engine does. This is somewhat natural and actual answering methodology that answers are reliable .The answers are novel in nature.

Here virtual answers and cache answers to the user query does not possess the freshness of the data rather materialize data possess somewhat the new information the user requires as the user arises new form of queries often and answers are also delivered in the new form.

Synchronization of the Data

In this methodology the updated data and query is matched with in the synchronizing method as the query is a separate part and the data are in different configure. Both should be linked to match each other. As the research part the data will be always in the update as it is segregated based on its generating time. It is clearly clustered based on the time of the data is been developed in the each of the sectors those sectors are linked with the data that is required.

For the Big Data this mode of synchronization of the data is required on the basis of the time. Here there is no such grouping of data in the big data based on the time of data creation and replacing the old data with the new one. Here in this model an entirely new approach is introduced by that a clear view of the data is visualized by the user. Confusion based on the data is averted.

VI. ATOMIC RING STRUCTURE MODEL FOR DATA SEGREGATION

This model of data segregation is structuring the data in a form of atomic ring model in which the each and every data unit are placed in the form of ring structure.

Atomic Ring Structure Model

The tiny atomic nucleus is the center of an atom constituting positively charged particles protons and uncharged particles neutrons. On the other hand, the extra nucleus part is a much larger region which is composed of a cloud of negatively charged particles called an electron**.** Electrons revolve around the orbit or center of the nucleus. The attraction between the protons and electrons holds the structure of an atom together.

VII. APPLYING ATOMIC RING MODEL FOR DATA SEGREGATION

This model is based on the atomic model that was applied in the chemistry for the atomic model. This model is chemistry based model applied for data segregating from new data to old.

In normal atomic model there will various energy level rings around nucleus and proton that present in middle of the atom. Similarly consider the data source point at the centre of the data segregation point. Based on the data update the data will be moved from first level to next level, based on the data freshness. Recent update data will be in nearest ring of the data source. Old data will be moved from one level of data circle to another data circle. At

one point old data will be out of the data segregation ring. The old data that is in the last level of the atomic ring model will be moved to the atomic structural model which maintains the old data in same manner. The old datum will be moved likewise from this kind of structure to structure as a simple record to stored and retrieved back as a reference in future.[2] , [3] , [4]

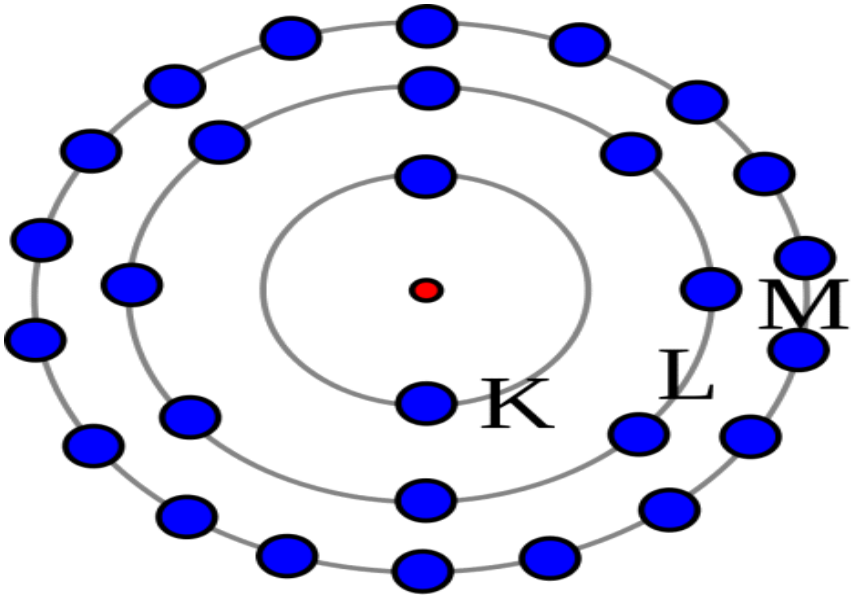

Figure 1: Atom structure.

In this atomic structure Fig: 1 it is noticed that center point is nucleus which is the main source of the atom .let it be considered as the data source. Then it is revolved by the number of rings in that there are number of electrons based on rings the number of the atoms will be increased. Likewise most recent update of data will be less comparing to the other data. As it is in the atom most recent data will be near to the data source that is in center. It is surrounded old data.

As the electrons which are in outside of the ring or out of the control of the nucleus will be losing the rotation similarly the old data will be out of the data consideration. It will be out of the data server. By that old data will be pruned from the data server.[3]

VIII. ALGORITHM FOR DATA PROPAGATION ONE LEVEL TO ANOTHER LEVEL

Step 1: Initially the new data N_d is generated at nucleus at the center point that is called as source point let it be server drive slot which maintains new data.

Step 2: A time interval N_t is provided to monitor the freshness of the data.

Step 3: Le the interval be (N_t -N_t- 1) ,it is based on the flow of data determined by the speed of data that emerging out in the time sequence.

Step 4: After the time sequence data which is in the time slot moves from one server drive slot to another.

Step 5:Let the server can be divided into n number of drives which can be considered as energy levels in the atom. (K,L,M,N,O,P) .

Step 6:Data in the Nucleus i.e source point is also a server drive where data is initially stored. After the given time interval the data is moving on KN_d after the time interval; $(N_t - N_t - 1)$ is consider as T_d the data is as $KN_d T_d$

Step 7: N_d is changeable based on flow of speed of data v ,so the speed of the can be considered as vT_d The propagation of data from one drive slot to the another drive slot is considered as on vT_d So $KN_d . vT_d = L . T_d$

Step 7: Similarly the data will propagate from one drive to another drive slot based on the time sequence and nature of data flow till the end of the drive slot i.e P energy level then the drive slot is renamed as PN_d . Similarly vT_d is applied to make the out of the server as it is old data as per the time concern.[2]

Step 8: In this point the data is totally transferred to another server which maintains the older data in the similar format.

Step 9: likewise data will be moving from one server to another as the electron moves from one energy level to another level. Similarly the old data will be moved from one drive to another.

Step 10: By this data can be stored in a different manner by which data is retrieved in easier manner.

By the proposed a new approach of algorithm model for data segregation challenges that generally occurs while data update occurs in large and fast in manner.

Security Provision for this Model: As per the security provision for this model is considered as a supplement part, that will enable security as an integrated work, as model itself provides security as itself. When the new comes from the source point. The old data will go to the next level that is Nucleus point next orbit level called K then K N ,like wise actual data is going to move .So attackers cant attack data at single point Whenever the data changes its position or it moves from one orbit to another when a new data is been originated from the source point automatically the data will be moving so attackers will not find a point to attack .

In another way of protecting the data, data encrypted through new block ciphering technique called Tigris block enciphering model .This model of block ciphering functions as other old ciphering models but differs in some areas like speed of operations that will highly support for the proposed model of data segregation

In the proposed block enciphering methodology the message block is 128 bits and key length varies 192 bits, and 256 in number the ciphering key can be realized as 4*4 state matrix in every state matrix has 16 positions bytes and 4 main function basically. The proposed system follows the functions that was maintained by other block chipper does .The main feature of this enciphering technique is it performs efficiently in security tasks in limited area resource handling. As our data in the orbit is limited at a time period this data can be considered for the proposed work. For high speed encryption tasks this block enciphering is recommended .As this block enciphers any data with high bandwidth and high speed. It can handle Big data security to encipher and decipher .As TIGRIS is tailor made to provide high volume and largely oppose the malicious attacks. As it provides complicated key generations ,it is very hard to break in .So Tigris block encipher will provide a large protection in all aspects [15].

IX. CONCLUSION

Data propagation from one level to another level based on the time condition is a new approach will levy the updated data confusions among the data users. Updated data is been generated at one point and old data will be

stored at another point. This will enhance the clear data segregation comparing with other models of data segregation based on time duration.

Future Enhancement

This atomic ring model is a new algorithm recommended for data segregation at this level .This algorithm approach can implemented for time based sequence operations. This will provide better results, while implementing this algorithm. Security short comes are discussed in this paper an idea to execute in better in future. If threat level is increases to this model.

REFERENCES

1. Data mining and Machine Learning techniques supporting Time-Based Separation concept
2. Deployment, Ivan De Visscher Wake Prediction Technologies(WaPT) Louvain-la-Neuve, Belgium ivan.devisscher@wapt.be Vincent Treve Airport division EUROCONTROLBrussels, Belgium vincent.treve@eurocontrol.int Guillaume Stempfel Eura Nova Marseille, France guillaume.stempfel@euranova.eu Frédéric Rooseleer Airport division EUROCONTROL Brussels, Belgium frederic.rooseleer@eurocontrol.int, 978-1-5386-4112-5/18/$31.00 ©2018 IEEE.
3. Build-Up Algorithm for Atomic Correspondence between Chemical Structures, Takeshi Kawabata†,‡ † Institute of Protein Science, Osaka University, Osaka 565-0871 Japan ‡ Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, NARA 630-0192 Japan ,Journal of Chemical Information and Modeling.
4. Algorithms for Solving Atomic Structures of Nanodimensional Clusters in Single Crystals Based on X-ray and Neutron Diffuse Scattering Data, N. M. Andrushevskiœ, B. M. Shchedrin*, and V. I. Simonov**, * Faculty of Computational Mathematics and Cybernetics, Moscow State University, ** Shubnikov Institute of Crystallography, Russian Academy of Sciences, Crystallography Reports, Vol. 49, No. 5, 2004, pp. 863–870. Translated from Kristallografiya, Vol. 49, No. 5, 2004, pp. 953–960.
5. Curved-line search algorithm for ab initio atomic structure relaxation, Zhanghui Chen,1,2 Jingbo Li,2 Shushen Li,2 and Lin-Wang Wang1,* PHYSICAL REVIEW B 96, 115141 (2017)
6. Verlak S and Heremans P 2007 Phys. Rev. B **75** 115127
7. Kalb W L, Haas S, Krellner C, Mathis T and Batlogg B 2010Phys. Rev. B **81** 155315
8. Martin M G and Siepmann J I 1998 J. Phys. Chem. B **102** 2569
9. Wick C D, Martin M G and Siepmann J I 2000 J. Phys.Chem. B **104** 8008
10. Vukmirovic N and Wang L-W 2008 J. Chem. Phys.**128** 121102
11. Canning A, Wang L, Williamson A and Zunger A 2000J. Comput. Phys. **160** 29
12. Schwoerer M and Wolf H C 2008 Organic Molecular Solids(Weinheim: Wiley-VCH)
13. Analyzing and Evaluating Data Freshness in Data Integration Systems Verónika Peralta* — Raúl Ruggia** — Mokrane Bouzeghoub*Laboratoire PRISM, Université de Versailles 45, avenue des Etats-Unis 78035, Versailles cedex, FRANCE, Instituto de Computación, Universidad de la República Julio Herrera y Reisig 565, 5to piso 11300, Montevideo, URUGUAY
14. Big data stream analysis: a systematic literature review Taiwo Kolajo1,2* , Olawande Daramola3 and Ayodele Adebiyi1,4, Kolajo et al. J Big Data (2019) 6:47 https://doi.org/10.1186/s40537-019-0210-7

15. https://searchdatamanagement.techtarget.com/feature/11-real-time-data-streaming-roadblocks-and-how-to-overcome-them
16. The New Block Cipher Design (Tigris Cipher), I. J. Computer Network and Information Security, 2015, 12, 10-18, Assist. Instructor Omar A. Dawood College of Education for Humanities Science, English Department, Anbar University and Ph.D. student at Computer, Prof. Abdul Monem S. Rahma Computer Science Department, University of Technology, Baghdad, Iraq. Assist. Prof Abdul Mohsen J. Abdul Hossen Computer Science Department, University of Technology Baghdad, Iraq.