International Journal of Engineering & Technology, 7 (2.12) (2018) 184-188

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

a\A technique on novel based marching ants colonies clusters for operational big data sets

Konda Sreenu [1]*, Dr. Boddu Raja Srinivasa Reddy [2]

[1] Assistant Professor in Computer Science and Engineering Department, Sir C R Reddy College of Engineering, Eluru, Andhra Prdesh, India.
[2] Professor in Computer Science and Engineering Department, Ramachandra College of Engineering, Eluru, Andhra Pradesh, India.
**Corresponding author E-mail: sreenukcupid@gmail.com*

Abstract

Computer plays a key role in everywhere world. Data is growing along with the usage of computers. In everyday life we use computer for various purpose and store bulk of information. One or other way we want to retrieve data from the storage system. Retrieving bulk of data information is not a simple thing or it is magic show. Every user wants data in different forms like reports or output information. For doing all this exercises we require one process. Process is nothing but marching ants colonies. Data related databases and tables are collected, trivial data is selected from huge tables and databases, apply aggregate functions on data and output information or reports related to data. Paper focus on how efficiently we can use software for some extent on solving business related problems. Paper may not solve century year's data but we can achieve something. When century years data it is better to go for data mining approach because it accumulates large time to solve such big problems

.

Keywords: *Marching Ants; Database; Tables; Data Mining*

1. Introduction

Every day lot of data is added to the databases because we depend on them. As humans we cannot remember days, months and years long data. Where as a machine can remember years long data. If it appears to happen exist in database. There are two methods in using year's long data. One is big data analytics and one is operational big data. Getting reports on year's long data is not possible within seconds. It will go for minutes. As years are growing the time also increase. The requirement will be like that we have to compare past ten years annual sales. We can make system ready for annual sales prior. But there will be ad hoc requirements, where we have to mine year's long data for that purpose. It also depends on the organization management, how their perception of requirement is there.

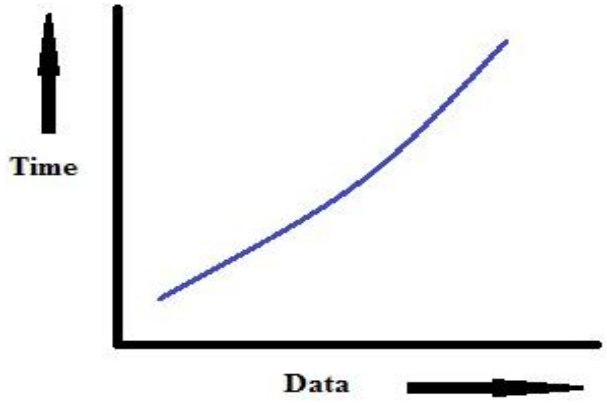


Fig. 1: Increase of Time.

But obtaining data report within minutes is possible. Preparing that data for operational is quite a challenge because our system or machines have some limitations on data type memories. We cannot accumulate years long data because of memory limitations in the machine. C like language has 1 or 2 bytes of size depending on compiler. SQL has 32 bits of size for integer. When the result of the data is more than those bytes, then it will give unrelated data information because of non-accommodation of real result. Based on hardware of the system also it depends. This paper focuses on using microcomputers. In use of word called microcomputers, that means there will be some limitations on data size. Due to this experience, we have to compress the data to make compactable with the memory.
In past days the usage of computer is less and nobody is aware of computers. After merging the electronic gadgets like mobile, tabs, computer systems the usage of memory has rapidly grown to more usage. Now the usage is of GBs. In future it may carry onto TBs. if you scatter all disks they will fit to an airport. It may rise to one medium size town or city. Not only saving, retrieving data is also a task. When we refer to large data, they are said to as "Big Data". There are two methods related to Big Data, one of them are big data analytics and other one is operational big data. This paper focuses on the operational big data. For big data analytics we need only business person and management person or group. Operational big data process can retrieve data with minutes or hours. Not as big data analytics. Both can be applied for year's long data. As years grow the time also increases as shown in figure 1. Both the data should be cleaned for mining or operational type. After recognizing big data, a new era started. Whereas most of the countries like USA, centered on big data by spending millions of amounts over it. Their main target is to retrieve data very fast, storage of data, security related to data and other aspects. Every business in the world is dependent upon data. They want to see data in different forms like

helping them in improving their customers, products, sales, profits and spread their new business operations.
Operational database management systems referred to as OLTP (Online Transaction processing databases), are used to update data in real-time. These types of databases allow users to do more than simply view archived data. Figure 2 for year's long data growth. When years grow data of organization also will grow. For growth there will be lot of reasons. An organization cannot be as same it started initially, it will spread it wings around and around. These are the reasons for growth of customers, products, sales, profits and other things.

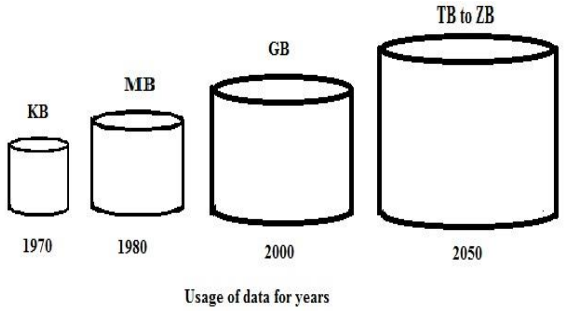


Fig. 2: Growth of Data for Years.

2. Literature review

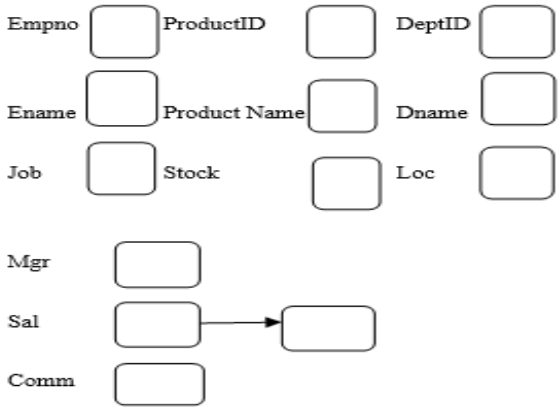
Konda Sreenu, ET. Al. introduced Ant Colony Cluster, a new model, but no implementation about ant colony clusters. Model focus on compression of the data sets for execution. Paper focus on methods like columnar databases, run length environment technique and ant colony clusters. No implementation related to the model.
O. A. Mohamed Jafar, et. al. introduced Ant-based clustering is a clustering technique aims at the unsupervised classification of patterns in different groups. These are so many algorithms developed for solving numerical and combinational optimization problems. Among them are swaron-based algorithms. It is a conventional clustering technique. It focus on the behavior of ants.
Wei Gao, introduced clustering analysis that are used in many disciplines and applications. The ant-colony clustering algorithm is a swaron-intelligent method used in various cluster problems. That inspired by the behavior of ant colonies. Author focus on the model of ant colony algorithm, but no implementation about ant colony cluster algorithm.
Ling chin, et. al. introduced an artificial ants sleeping model (ASM) and adaptive Artificial Ants Clustering Algorithm (AYC) are presented to solve clustering problems. They are simulation based solution but author does not focus implementation and no identification of experimental results.
Bao-Jiang Zhoo, introduced ant colony clustering algorithm for optimistically executing large data sets. This algorithm was implemented and tested on several real data sets. Time is the main drawback of the paper.
Urszula Boryczka, introduced bio-inspired techniques for clustering algorithms received attention for performance & stability to make mature tools for data mining. Author introduced algorithms and procedures to the new approach, but no focus on implementation.
Xiaoyong Liu, et. al. introduced clustering algorithm to solve the unsupervised clustering problem. Ant colony optimization (ACO) is used to solve combinational optimization problem which is based on stochastic best solution kept – EsaCC. This model cannot compute with present large data sets.
Jinbiao Wang, et. al. introduced ant-based clustering algorithm, Ant colony clustering have different characteristics like clustering, high clustering accuracy, irregular cluster shapes and so on. All these topics are related to clustering algorithm. No implementation details.
Hong Jiang, et. al. introduced basic model of ant colony clustering algorithm. Purposed IACC an improved ant colony clustering shows better performance than LF method and it should be improved.
Dong Liyan, et. al. introduced micro-clustering, traditional ant cluster algorithm is not accurate. Ant Colony clustering algorithm is based on swarm intelligence. This algorithm proposed a new policy of picking and dropping objects. Author proves that ant colony clustering is better algorithm based on swarm intelligence than traditional colony algorithm in terms of efficiency. It is based on dividing the similar objects into one cluster.

3. Existing system

Memory plays a key role in today's world. Memory size is growing every day. Retrieving and storing data is also becoming a task. Large companies have more records related to their employees, sales, billing and so on. Today's world is dependent upon computer systems. They want their search or results very fast in a finger tips. To get fast results on data sets, we have to focus on data mining. In data mining, data is divided into two parts, non-numeric and numeric data. Non-numeric data occupies more memory and almost worst case of search because there will no repeated data. All data will occupy memory. Numeric data we can save memory by applying compressing techniques. Data should be collected and then it must be cleaned, make schemas preparation for collected data, compress the data into ant colonies. One after the other, place data for computing. Expose the result in a report or graph format. Final result will contain report or a graph.
We design and implement a pattern to solve business user requirements without data mining expert and business analyst. The process of mining will take time to display results. But business user wants result very fast. Data is collected from different repositories and other sources. Collected data cleaning process will carry on. Data cleaning means removing unwanted non-key attributes, records with null values. After cleaning records are made to be ant colony cluster. Each cluster can be of 100 data records size or chunk with fixed size. If data block is more than 100 records, then it is added to next block. Data in blocks can be compressed. Data with numeric can be compressed but alphabets cannot be compressed. There are so many algorithms like Bitmap Index, Look-and-say Sequences, Run-length limited, LZW or run-length encoding for compression. It will be is easy to accumulate large data. There should be large memory enough to hold data collected from various sources. Results are made to display for user. Data can be displayed in reports form, picture form, text form or any other way which can be understood by business user.

2.1. Algorithm

Step 1:
Collect data from various database sources.
Step 2:
Data cleaning process. Remove unwanted non-key attributes, columns with null values.
Step 3:
There are three tables in a database. Table – 1 Employee, Table Two Product and Table.
Three Dept.



Step 4:

1) Arrange the data in columnar database manner.
2) Columnar database have data blocks for 100 record umn.
3) Normally table is given in the following manner.

X	Y	Z
1	Ravi	34
2	Smith	40

1	Ravi	34
2	Smith	40

Step 5:

Apply aggregate functions or any other calculation methods over the data.

2.2. Block diagram

The existing method is called Ant Colony Cluster Method as shown in below figure 1 (Fig. 1). The data blocks are made to available like ants row. Data blocks can be expanded to their right side. It can grow to maximum as much memory is available. After data blocks are prepared, based on business user requirements we can use aggregate functions like SUM, COUNT, AVG, MAX, MIN.

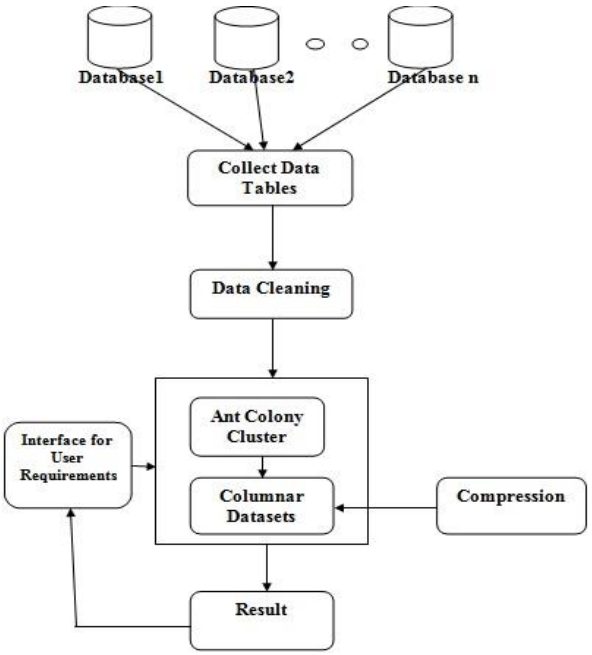


Fig. 1: Iv. Proposed System.

More data is added to the database(s) or repositories every day. An example for addition of data is youtube or social medias or personal life videos and images. Databases are increasing day by day for years. Retrieving year's long data is difficult. In retrieving them there are two types. One is trivial and second one is non-trivial. Trivial is that retrieving business related data for future planning and knowing current status of the business, which should be systematic and controlled manner. The systematic process should be like that as shown in figure 2.

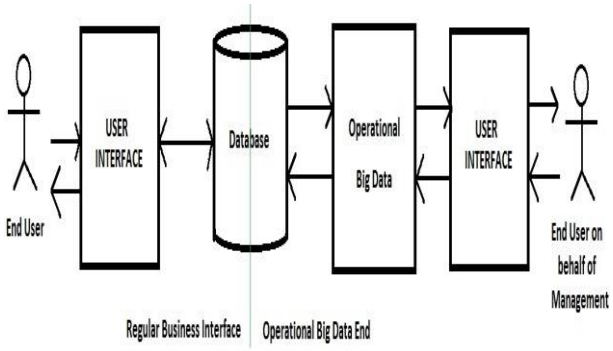


Fig. 2: Architecture for Operational Big Data.

In the above figure 2, the system will have input as usual to daily life. The data is feed to the database regularly. Every day data is added to the database, hence data will increases daily. Data which is stored in the databases will be years long data. Whereas other half of the system is retrieving data from the stored databases with-in fractions of minutes. Because retrieving of years long data is not possible with-in seconds. For retrieval purpose we have choose trivial data from the database. That means we want to calculate months sales. We needed not focus on non-trivial data like person name and other character oriented data. Operational big Data end is used to work on stored data for reports generations for planning and assumptions of future.

Regular business interface is related to software interfaces to end users to feed data for storage as shown in figure 3.

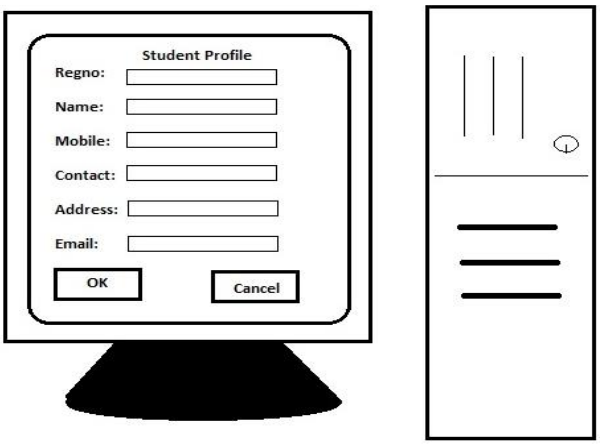


Fig.3: End User System.

Operational Big Data is operated on storage databases. There are steps in operational big data. The systematic way of approach for operational data is shown in below figure 4.

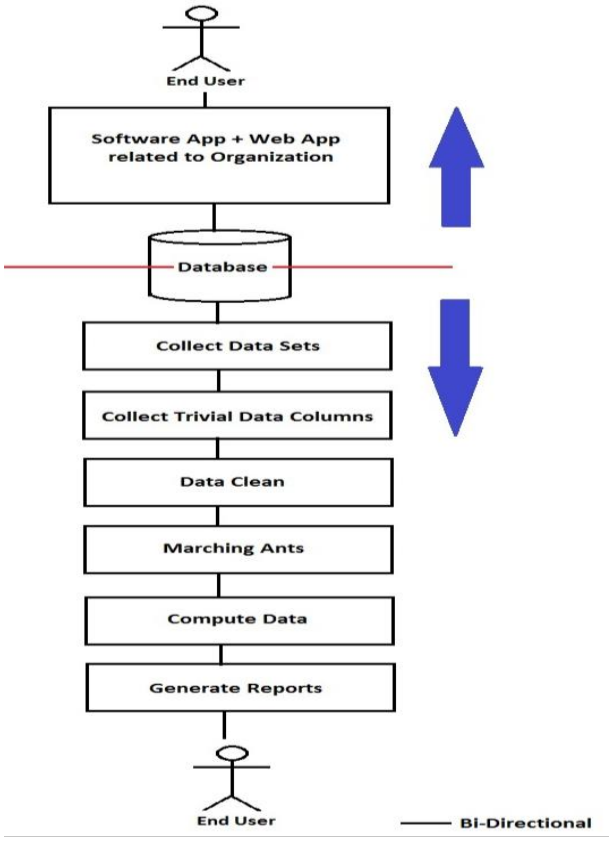


Fig. 4: Operational Big Data.

Data will be feed by end users of the organization and it will be stored at repositories of the organizations. From repositories we will collect the tables. Cleaning of the data will take place. In cleaning process we will select the data that can be computed. The data may be like a long railway track. In the next process we will reduce the marching ants to small shrinking marching ants. We can apply aggregate functions on the data. When we cumulative data from each row it will be overflow because of the memory limitations occupied by the data type. End user from the management side may use output of the computed data in form of reports or just as a result.

3.1. Proposed algorithm

Step 1: Collect the tables from varies repositories.

$Collect\ tables\ CT\ t1 = r1, r2, ..rn, t2 = r1, r2, ..rm$

Step 2: Get the trivial data from the tables.

$Data\ Cleaning\ t1 = r1, r3, .., t2 = r2, r4$

Step 3: Reduce the data in rows by applying aggregate functions like sum, avg, max, min, count. Arrange the required data in form of marching ants.

r1=d1, d2, d3, d4

r1=D1=SUM({d1,d2,..}),D2=SUM({d6,d7,..}),
Dn=SUM({da,db,dc,..} (3)

Step 4: Usage of aggregate functions depends on user, management and there requirements. My example below gives the monthly sales of the business.

$Jan = D1, Feb = D2, Mar = D3,$

Step 5: Generating report or output result depends on the user and management.
Generate Report or Output Result (5)

3.2. Implementation

Software should be like that it has range the data in computational way. Example mostly the operational big data focus on the sales, highest sales, lowest sales, monthly sales, monthly outgoing products, annual sales and soon. It is better to arrange the data in easy way. It has to collect all data from database and make primary arrangement of the data. For implementation purpose java is good.

3.3. Pseudo code of the program

```
DATA INTERFACE
|       |INPUT DATA
|       |       |FORM1
|       |       |FORM2
|       |       |FORM3
|       |       |….FORMn
|       |RETRIEVE DATA
|       |       |SELECT DATA FORM
|       |UPDATE DATA
|       |       |UPDATE DATA FORM
|       |DELETE DATA
|       |       |DELETE DATA FORM
OPERATIONAL BIG DATA
|       |COLLECT DATA SETS
|       |COLLECT TRIVIAL DATA
|       |APPLY AGGREGATE FUNCTIONS ON LARGE DATA
|       |       |PREPARE MARCHING ANTS DATA
|       |USE DATA TO GENERATE REPORTS OR OUTPUT
```

4. Conclusion

We are rich in data but poor in maintaining the large data. When data is growing the computation of large data is also a task. For business purpose the management requires some information from the large database. Large database may contain databases and tables. We need not require all data from the tables. We can categorize data into trivial and non-trivial. Mostly trivial data will be numeric type and non-trivial will be character type. After collecting trivial data apply aggregate functions on it. According to big data we can large databases and tables. Make trivial data in the form of marching ants' manner. Collect computed data and generate reports or outputs. Paper focus on how efficiently we can use software for some extent on solving business related problems. Paper may not also century year's data but we can achieve something. When century years data it is better to go for data mining approach because it accumulates large time to solve such big problems.

References

[1] "Ant Colony Clusters for Fast Execution of Large Data sets", Konda Sreenu, Dr. B Raja Srininivasa Reddy, Unpublished.
[2] "Ant-based Clustering Algorithms: A Brief Survey", O. A. Mohamed Jafar and R. Siva Kumar, ISSN: 1793-8201, IJCTE, Vol. 2, No. 5, October, 2010.
[3] "Improved Ant Colony Clustering Algorithm and its Performance Study", Wei Gao, Computational Intelligence and Neuroscience, Volume 2016, Article ID: 4835932, 14 pages, http://dx.doi.org/10.1155/2016/4835932.
[4] "An Adaptive Ant Colony Clustering Algorithm", Ling Chen, Xiao-Hua XU, Yi-Xin Chen, Proceeding of the third international conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004, IEEE.
[5] "An Ant Colony Clustering Algorithm", Bao-Jiang Zhao, ISSN (Print): 2160-133X, (Electronic), ISSN: 2160-1348, 29 Oct' 2007, IEEE Xplore.
[6] "Ant Clustering Algorithm", Urszula Boryczka, ISBN: 978-83-60434-44-4, Pages 377-386, Intelligent Information System, 2008.
[7] "An effective Clustering Algorithm with Ant Colony", Xiaoyong Liu, Hui Fu, Journal of Computers, Vol. 5, no. 4, April' 2010.

[8] "An Ant Colony Clustering Algorithm Improved from ATTA", Jinbiao Wang, Ailing Tu, Hongwei Huang, 1875-3892, Published by Elsevier BV. Selection and/or peer-review under responsibility.
[9] "An improved ant colony clustering algorithm", Hong Jiang, Qing-song Yu, Yu Gong, ISSN (Print): 1948-2914, (Electronic): 1948-2922, IEEE Xplore: 18th Nov' 2010.
[10] "Ant Colony Clustering Algorithm Based on Swarm Intelligence", ISBN (Electronic): 978-1-4799-2809-5, CD-ROM ISBN: 978-1-4799-2810-1, IEEE Xplore: 24 July' 2014.
[11] "Approach for Developing Business Statistics using Data Web usage mining", G VS CH S L V Prasad, Malapati Sri Rama Lakshmi Reddy, Kuntam Babu Rao, Chodagam Suresh Kumar, ISSN (Print): 2319-2526, Volume – 1, Issue – 2, 2012.
[12] "Approach for Developing Business Statistics using weblog mining", Konda Sreenu, Dr. B. Raja Srinivasa Reddy, ISSN (Online): 2347-2812, Volume – 2, Issue – 11, 2014.
[13] "Providing Security and efficiency in Attribute based Data sharing with novel technique", Gudipudi Ravi Chaitanya Kumar, Prathipati Ratna Kumar, ISSN (Online): 2347-2820, Volume – 3, Issue – 12, 2015.
[14] "Preprocessing a Unsupervised Approach for Web Usage mining", Addanki Ramya, Konda Sreenu, P Ratna Kumar, ISSN: 2252-8784, Volume – 1, No. 2, December 2012.
[15] "Data Method and Mining Techniques for better Business Organization", Adabala Hanumantha Rao, Prathipati Ratna Kumar, ISSN: 2347-2820, Volume – 3, Issue -12, 2015.
[16] "A Freight Stabilization Model Based on Cloud Segregating for the shared cloud", S Venkata Siva Satish, P Ratna Kumar, ISSN (Online): 2278-5841, ISSN (Print): 2320-5156, IJRCCT, Volume – 3, Issue – 11, November – 2014.
[17] "De-Duplication of Citation Data by Genetic Programming Approach", Seetalam Divya Manusha, Valiveti Karthik, Prathipati Ratna Kumar, IJRAET, ISSN(Online): 2347-2812, Volume – 1, Issue – 3, 2013.
[18] "Slicing: A New Approach to privacy preserving High dimensional Data Publishing", IJSETR, ISSN (Online): 2319-8885, Volume – 2, Issue – 15, November – 2013.
[19] "Fast Searching Technique on Sequential Large Databases in an organization", Samparthi V Suresh Kumar, Matcha Ganesh B abu, J S V Gopal Krishna, Suryadevra Mohan Babu Chowdary, Konda Sreenu, B Homer Benny, Prathipati Ratna Kumar.