Publication of Little Lion Scientific R&D, Islamabad PAKISTAN
Journal of Theoretical and Applied Information Technology
15th April 2011. Vol. 26 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

THE DIRECTS IMPACT TO PRE-FILTERING PROCESS TO WEATHER DATASET

[1]WA'EL JUM'AH AL-ZYADAT, [2]RODZIAH BINTI ATAN, [3]HAMIDAH IBRAHIM, [4]MASRAH AZRIFAH AZMI MURAD

[1]PhD Student, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor

[2]Dr. Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, MALAYSIA.

[3]Asstt Prof. Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, MALAYSIA.

[4]Dr. Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, MALAYSIA.

E-mail: wael_abade@yahoo.com , rodziah@fsktm.upm.edu.my , hamidah@fsktm.upm.edu.my , masrah@fsktm.upm.edu.my

ABSTRACT

Monitoring through of data near real-time (dynamic data), in this paper, we address the issue of efficiently monitoring the satisfaction of sequential process to receive on the way to data storage. The research question then we is answer is how is it possible to monitor and efficiently fulfill the requirement of data storage by receiving a streaming of weather data? We propose a monitoring structural Algebra representing. We adapt the concept of data similarity in acclimated data from environment constraint a weather data from Malaysia. This enables contribute and construction of filtering data to doing the data meaningful and relevance. These performance enhancing techniques have general applicability.

Keywords: *Algorithm, Monitoring Data, Storage Data, Structural Data, Pre-Filter Data.*

1. INTRODUCTION

the monitoring task below [1-2] [3] data pre-processing takes more than half of the total time spent by solving the data mining problem. There are a number of different tools and methods used for preprocessing; data pre-processing is high complicated task [4] as well a transformation, or conditioning, of data designed to make modeling easier and more robust describes any type of processing performed on raw data to prepare it for another processing procedure. Commonly used as a preliminary data mining practice.

Data pre-filter transforms the data into a format that will be more easily and effectively processed for the purpose of the user for example, in a network. There are a number of different tools and methods used for pre-processing, including: sampling, which selects a representative subset from a large population of data; transformation, which manipulates raw data to produce a single input; demising, which removes noise from data; normalization, which organizes data for more efficient access; and feature extraction, which pulls out specified data that is significant in some particular context.

The objective using pre-filter toward for acquisition data and find out the information (quality data) within keep the regular relation for one set as a unit furthermore pre-filter technique in general appear like pre-process data in begin process but in core its

Publication of Little Lion Scientific R&D, Islamabad PAKISTAN
Journal of Theoretical and Applied Information Technology
15th April 2011. Vol. 26 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

different path as pre-process main functionality aim the feature extraction. The challenges for pre-filter involve generation for monitoring data when finished collected data and gathering implies it request long term process to direct effect from time complexity as recursive technique, on other hand; for when any modifying from context data (Insert, Update) request to re-monitoring happen to exchange the structure data and location.

usually the data when aggregate are generally containing incompletely, inconsistent, duplicate, and noisy, which proves the importance of this stage, we can use the data well and clear and explore the information and support the filter in a positive way [5].

This paper will propose method to monitoring raw data (primary data) when collected and aggregated data from environment, unfortunately the monitoring data start when captured collected data which mean provide the static data work additional does not have ability to dynamic data as high modifying data or irregular data; the role for monitoring requesting minimum semi-structural data to simulate the data and know the volume of data as vector or matrix base on domain the data .on other hand ,about processing and categorization to data recommended to be un-ambiguity such (Sequential process, Batch process, distributed process, parallel process) .

Due the Queue-Buffer (QB) structure functionality classifier under queue method plus involve indexing to sort data dependability for time to benefits reduce duplicate data and omitted the data entry problems are include missing value noisy data as well flexible to handle gap data; absolutely distortion smooth action whatever the time complexity suffer to companion .

2. RELATED WORK

Monitoring of arbitrary threshold functions over distributed data stream in an efficient manner [3, 6] as well the network fields appear for transfer stream data such End-to End monitoring [7-8] the main challenges of monitoring method call delay process additional losses at the sending buffer as shown, for instance [9-10]; reduce the probability of failure [11] , the structure of data important strategy provide fast and accuracy for monitoring method as well short description of data set as one unit [12-13]; a monitoring system minimize unnecessary computation furthermore ability to discard data that have become too old to be meaningful [14], SPMON incorporate probabilistic models of uncertainty in constraint monitoring to generalize the notion of data similarity to cover data objects.

3. METHOD

The conceptual work of monitoring method covering the structural data or semi-structural to utilizing ability data contaminated as well modification function for cloud index is investigated as the change of time and space in this study. Furthermore prerequisite for monitoring method based on (1) could represent the data collected as mathematical model to optimization the schema model and straightforward process It.; (2) indexing data are major variants to supporting sort data and usually in this issue using the time parameters to introduce the data location and references.

In initial present data by algebra will see at two dimensional elucidate how to represent the data usage algebraic and each cell can determinate by location by pair values are ***(I,J),*** based on location of data become suitable to determinate the redundant and missing data which a result become more accuracy ; the first equation to merge data its :

$$Y=\sum (Xi,j) \quad \text{——} \quad (1)$$

I

X1,1	X1,2	X1,3	X1,4
X2,1	X2,2	X2,3	X2,4
X3,1	X3,2	X3,3	X3,4
X4,1	X4,2	X4,3	X4,4
X5,1	X5,2	X5,3	X5,4
X6,1	X6,2	X6,3	X6,4
X7,1	X7,2	X7,3	X7,4

J

Figure1 represented data

The scan record is initial operation to make realize the data is finished transfer and gathered all; as well as pass conditional otherwise need it return previous process , about a scan record start from i =1 and j= {1,2,...,length of row}, when j =length will jump to next record i =2 and keep j counter as previous. In one record ***i*** status is static but j call dynamic based on length.

It has two main types of monitor data:

Publication of Little Lion Scientific R&D, Islamabad PAKISTAN
Journal of Theoretical and Applied Information Technology
15th April 2011. Vol. 26 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1- Horizontal: is familiar with query and search data easier and faster because individual work, weakness not support for pre-process data because care for cells.
2- Vertical: is familiar with comfortable with pre-process module and easy adopt with other method as shown; the advantage more dynamically and ability agree to update.

Hence, for each one using based on kind of data and structural if the relational between data close up prefer use vertical as query, but the relation between data ambiguity use horizontal

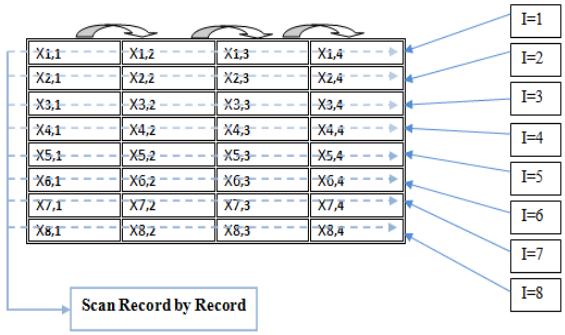


Figure2 check record by record

Due the element ***i*** in this method will as primary in addition static to clear how represented collected and aggregated data from environmental and one main issue provide determinate the location data in storage to become more fetch data from storage during saving considered one of the most important forms of intellectual property as well the ***j*** variables is dynamic variants to sorting data receive it such queue/queue method as show the below figure.

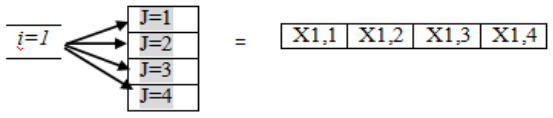


Figure 3sorting data

The equation proof a sorting it important to storage and save contain; the ***i*** variants aim numbers of records and can use multi-function as Primary key and numbers of records .on other hand, The sorting record consider the frequently path to be high accuracy which mean turned out worth attributes in dataset.

$$(J1, J2, J3, J4, J5, J6, J7, J8) * \begin{pmatrix} I=1 \\ I=2 \\ I=3 \\ I=4 \\ I=5 \\ I=6 \\ I=7 \\ I=8 \end{pmatrix} \neq \begin{pmatrix} I=1 \\ I=2 \\ I=3 \\ I=4 \\ I=5 \\ I=6 \\ I=7 \\ I=8 \end{pmatrix} * (J1, J2, \ldots, J8)$$

Figure 4condition equation

4. APPROACH

This will persuaded the missing of the data and therefore the time to make declaration of location and what suitable area should be ignored [15] , rarely to collected data from environment without error or symbol (miss understandability) for this reason shall be an approach propose utility data within low time complexity [16-18] , the weakness about monitoring algorithm s the time complexity its high and make delay from structural of dataset.

The main elements for monitoring algorithms are conditional ***If and Only If*** to determinate the location unwanted data and fetch to omit the value second elements ***loop*** as recursive method aim to restructure the datasets.

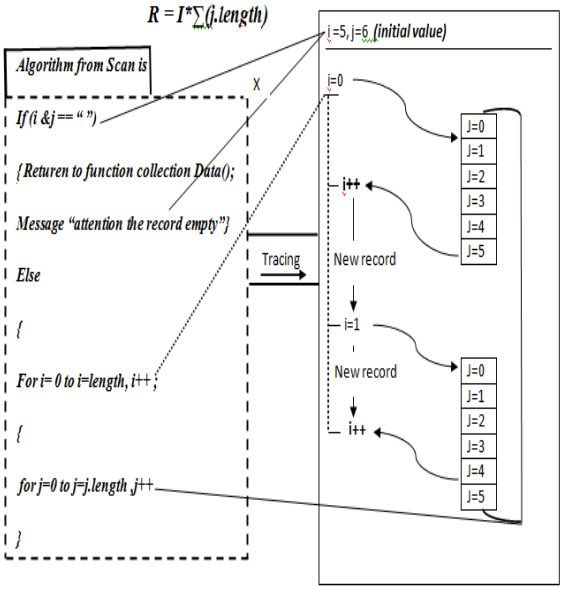


Figure 5 Algorithm

Initially process for algorithm check collection data if carry data or empty to determine the path and through data processing, which means if the content is empty will alert appears that the process of

Publication of Little Lion Scientific R&D, Islamabad PAKISTAN
Journal of Theoretical and Applied Information Technology
15th April 2011. Vol. 26 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

receiving data failed to repeat the process of the reception of other facial explained if it holds data transferred to the stage of data processing for alleged quality of the data smoothly. The type of data come is raw data, furthermore The process of examining the data the next depends on the length of the variable ***(i)*** to know the length of the data and here is connected in time in all alone move the variable ***(j)*** the words of Serial data within one unit, and here be clarified to us that the variable ***(i)*** is linked to the event and the variable ***(j)*** is linked in the value of the data contained in.

Due tracing about algorithm appear how to ejected the exception cause as you show the conditional ***if (i&j == "")*** directly rollback to function ***collection_data()*** to attend recollection data unfortunately about this action delay the time additional make sometimes damage to data transfer but newly has one way to solve it using downlink/uplink popular using in wireless network ; the ***(i &j)*** will content values are will starting create structural dataset and sorting data based on length data as well transfer time , for this issue the ***i*** value is clear to find initial data to end to jump to second values but ***j*** values on sorting inside ***i*** value which mean the ***j*** is high modifying after sort data to become at least semi-structural data can start cleaning and remove unwanted data will show the figure5.

```
If exist (i,j) ⟶ ( symbol || empty || null)
{
Get (i ) static value;
For (j=length to j=0 ,j--)
Delete (i, j.length)
}
```

Figure5 Cleaning data

For this function start after forward from traffic forward [6] to start clean content the dataset toward meta-information , Beginning to get meta-information will be determined by the location of data such as unwanted, missing and empty, which influenced the quality of data , furthermore here concept functional same pointer get rid of unwanted data and safest way is to change the title with the deletion of content, which proves the deletion, and the nucleus of the work is the comparison in the table consists of the type of data that do not wish to found in database. In observation about delete function for any record the pointer for which record want delete is static value from ***(i)*** moreover the variants action to sequence remove the content is ***(j)*** a bear in support of smoothly transport between location in dataset Flow data with measurement of the amount of data flowing depending on the length and quantity of the variable which shows in the case deletion using the length of the variable to delete the data unwanted and that makes controlling in the amount of data within the variable itself, but the risk that possible I could get is the intersection of the data in one unit and to avoid this risk as possible add variable function respectively and install the data in a fixed location, a time-use. Here will be used normalization until the inventory data in the field of filtration, which makes it easier to track this process through the facts of nature that cannot be ignored and that specially in the field of weather,

Which we will need rule to support the filtration results in an accurate and logical to predicated.

5. EVALUATE

The experiment method is fast estimate time start from collected data until pre-filter also within normalization when compare with recursive method as well return in this method is break of f and limits power.

The steps of method beginning receive data from environment to pre-filter during the process appear the ambiguity data Important features of performance with speed and accuracy of the results showed the importance (norm) to support the accuracy of the data link with the rules of direct data quality weather data and a risk that can happen is to ignore the time than we would have an overlap in the data and the low quality.

Publication of Little Lion Scientific R&D, Islamabad PAKISTAN
Journal of Theoretical and Applied Information Technology
15th April 2011. Vol. 26 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

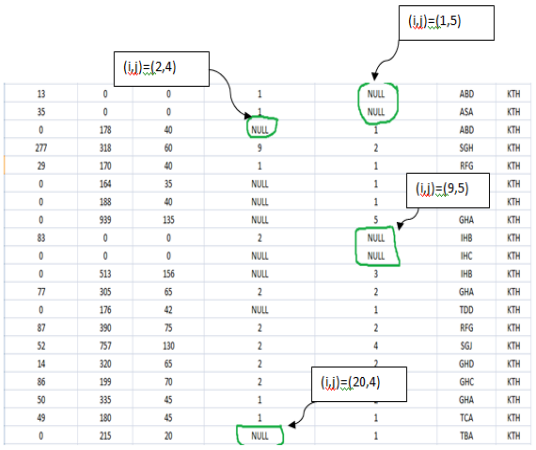
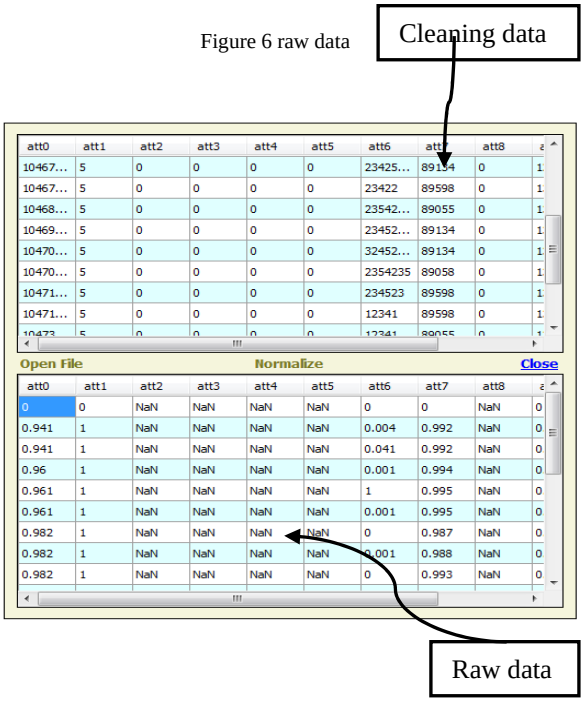


Figure 6 raw data



6. CONCLUSION

Monitoring method is important for various filtering data tasks and application. Recently, the importance of the completeness and homogeneity as evaluation criteria for such as recursive and loop method; the presented data using algebra is high flexibility and support to determinate location intended for unwanted data. Due the measurement proof monitoring method done well shall appear in filter application and direct effect storage data.

Future work enhancement the capacity for monitoring method using three dimensional variables are apply in presented data without using vector structure.

ACKNOWLEDGMENT

This paper determination dataset scope Malaysia weather climate from 2008 and 2009 the data contract from Malaysia Meteorological Department (MMD).

This research was supported in part “A Dynamic Environment Safety Precaution System Using Formal Specification” and special coordination funds for Science Funds, Malaysia.

REFERENCES:

[1] N. Zhang and W. F. Lu, "An Efficient Data Preprocessing Method for Mining Customer Survey Data," in *Industrial Informatics, 2007 5th IEEE International Conference on*, 2007, pp. 573-578.

[2] Available: http://www.pcmag.com/encyclopedia_term/0,2542,t=filter&i=43200,00.asp

[3] I. Sharfman, *et al.*, "A geometric approach to monitoring threshold functions over distributed data streams," presented at the Proceedings of the 2006 ACM SIGMOD international conference on Management of data, Chicago, IL, USA, 2006.

[4] P. Miksovsky, *et al.*, "Data pre-processing support for data mining," in *Systems, Man and Cybernetics, 2002 IEEE International Conference on*, 2002, p. 4 pp. vol.5.

[5] B. E. a. T.-Y. L. Fu Zhao, Ph.D., "A Data Preprocessing Framework for Supporting Probability-Learning in Dynamic Decision Modeling in Medicine," *NCBI*, p. 5, 2000.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[6] I. Sharfman, *et al.*, "A geometric approach to monitoring threshold functions over distributed data streams," *ACM Trans. Database Syst.*, vol. 32, p. 23, 2007.

[7] T. Lindh, *et al.*, "A performance monitoring method for wireless sensor networks," presented at the Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments, Athens, Greece, 2008.

[8] G. Allen, "Building a Dynamic Data Driven Application System for Hurricane Forecasting," in *Computational Science – ICCS 2007*. vol. 4487, Y. Shi, *et al.*, Eds., ed: Springer Berlin / Heidelberg, 2007, pp. 1034-1041.

[9] F. Cacace and L. Vollero, "A delay monitoring method for up-link flows in IEEE 802.11e EDCA networks," presented at the Proceedings of the 2nd International Conference on Simulation Tools and Techniques, Rome, Italy, 2009.

[10] W. Honguk and A. K. Mok, "Real-Time Monitoring of Uncertain Data Streams Using Probabilistic Similarity," in *Real-Time Systems Symposium, 2007. RTSS 2007. 28th IEEE International*, 2007, pp. 288-300.

[11] L. Xiaoling, *et al.*, "PSMM: A Plug-In Based Software Monitoring Method," in *Information Engineering and Computer Science, 2009. ICIECS 2009. International Conference on*, 2009, pp. 1-5.

[12] J. Aguilar-Saborit, *et al.*, "Dynamic adaptive data structures for monitoring data streams," *Data & Knowledge Engineering*, vol. 66, pp. 92-115, 2008.

[13] H. Takeshita and N. Henmi, "A novel data format free bit-by-bit quasi-error monitoring method for optical transport network," in *Optical Fiber Communication Conference, 1999, and the International Conference on Integrated Optics and Optical Fiber Communication. OFC/IOOC '99. Technical Digest*, 1999, pp. 149-151 vol.4.

[14] T. A. Russ, "Use of data abstraction methods to simplify monitoring," *Artificial Intelligence in Medicine*, vol. 7, pp. 497-514, 1995.

[15] R. B. A. WA'EL JUM'AH AL-ZYADAT, "AN APPROACH OF DYNAMIC FILTER FOR WEATHER
ENVIRONMENT DEPENDENT ON TIME," *JATIT*, vol. 18, p. 6, 2010.

[16] F. Li, *et al.*, "Model-based monitoring and fault diagnosis of fossil power plant process units using Group Method of Data Handling," *ISA Transactions*, vol. 48, pp. 213-219, 2009.

[17] J. F. MacGregor, *et al.*, "Data-based latent variable methods for process analysis, monitoring and control," *Computers & Chemical Engineering*, vol. 29, pp. 1217-1223, 2005.

[18] H. Reuter, *et al.*, "Information system for monitoring environmental impacts of genetically modified organisms," *Environmental Science and Pollution Research*, vol. 17, pp. 1479-1490, 2010.