An Efficient Divide and Conquer Approach for Big Data Analytics in Machine to Machine Communication

Mukesh Nair, Prof. Amruta Chadawar, Ashitesh Bhosle

Mukesh Nair, Department of MCA, Lokmanya Tilak College of Engineering, KoparKhairne, University of Mumbai, India.

Prof. Amruta Chadawar, Department of MCA, Lokmanya Tilak College of Engineering, KoparKhairne, University of Mumbai, India.

Ashitesh Bhosle, Department of MCA, Lokmanya Tilak College of Engineering, KoparKhairne, University of Mumbai, India.

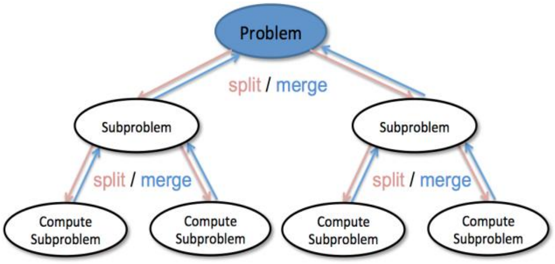
***Abstract*— Many Machine-to-Machine communications relies on the physical objects like satellites, sensors etc interconnected with each other, creating mesh of machines producing massive volume of data about large geographical areas. Thus, the Machine-to-Machine is an ideal example of Big Data. On the contrary, the Machine-to-Machine platforms that handle Big Data might perform poorly or not according to the goals of their operator in terms of the cost, database utilization, data quality, processing and computational efficiency, analysis etc. Therefore, to address the aforementioned needs, we propose a new effective, memory and processing efficient system architecture for Big Data in M2M, which, unlike other previous proposals, does not require whole set of data to be processed (including raw data sets), and to be kept in the main memory. Our designed system architecture exploits divide-and-conquer approach and data block-wise vertical representation of the data- base follows a particular petitionary strategy, which formalizes the problem of feature extraction applications. The architecture goes from physical objects to the processing servers, where Big Data set is first transformed into a several data blocks that can be quickly processed, then it classifies and reorganizes these data blocks from the same source. In addition, the data blocks are aggregated in a sequential manner based on a machine ID, and equally partitions the data using fusion algorithm. Finally, the results are stored in a server that helps the users in making decision.**

***Index Terms*— Machine-To-Machine Communications, Logarithmic, Speech Signals, Data Hiding Technique, Frequency Masking.**

I. INTRODUCTION

Well to speak about big data, it is one of the central and influential research challenges for the analyst as well as researchers in of our generation. The archetype relies on the acquisition and aggregation of the massive volume of data to support innovation in the upcoming years. The groundwork of Big Data exploitation is to empower the existence data sets to extract new information, helps in enrichment of business values chains like stock markets, shares and many more. According to the IDC group, the quantity of world data will be 44 times bigger in the next few years (such as 0.8–35 zettabytes). Therefore, in this context, the machine-to-machine archetype relies on the world of interconnected object , which can be used for acquisition, aggregation and analyzing the data depending on context.

While businesses across industries recognize the imperative of big data, there are many challenges that face the research in this field. The most prevalent are skill set shortages, cultural barriers, processes and structures, and technology maturity levels.



The proposed divide-and-conquer data analytical architecture for Big Data in M2M has several advantages, such as, at data acquisition stage, the data is concatenated to form a Big Data block that helps the system to combine the same data type, the fusion domain helps in enhancing the efficiency of D&CPU by dividing the data into smaller data blocks. Each block is then sent to a single server for further processing, which helps in increasing the processing efficiency, and finally, users can use the desired results for comparison purpose.

II. WORKS RELATED TO THIS APPROACH

Big Data and its analysis are at the verge of modern science and business, where author highlights the identity of number of sources on Big Data such as online transactions, emails, audios, videos, search queries, health records, social networking interactions, images, click-streams, logs, posts, search queries, health records, social networking interactions, mobile phones and applications, scientific equipment, and sensors. The concept of Big Data is stimulating a broad range of curiosity in the industrial sector. A massive volume of data is generated by numerous machines deployed in the supply lines of utility providers, which are constantly monitoring the production quality, safety, maintenance, and so forth. The electronic sensors that are frequently monitoring the mechanical and atmospheric conditions are high-quality example of sensors generating a bulk of Big Data. Furthermore, sensors are used for healthcare sectors are huge source of information for Big Data presented in [11]. However gathering the sensors' data from numerous sensors in an energy efficient method remains, beyond expertise's of the report.

III. ORGANIZATIONAL ADVANTAGES

Much like with most disruptive embarking on big data utilization projects will accure a number of organizational benefits.Improve decision making by lowering the cost of better quality information analysis.

Improve business performance by disseminating information more effectively across the organization.

Improve collaboration by developing a common, enterprise-wide business intelligence, integrating views on identified business opportunities.

Generate and pretest value propositions utilizing advanced and discovery analytics.

Fraud can be detected the moment it happens and proper measures can be taken to limit the damage. The financial world is very attractive for criminals. With a real-time safeguard system, attempts to hack into your organization are notified instantly. Your IT security department can take immediately appropriate action.

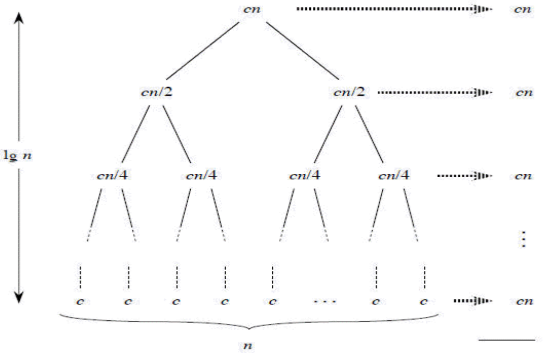
IV. CHALLENGES OF REAL TIME BIG DATA ANALYSIS

It requires special computer power: The standard version of Hadoop is, at the moment, not yet suitable for real-time analysis. New tools need to be bought and used. There are however quite some tools available to do the job and Hadoop will be able to process data in real-time in the future.

Using real-time insights requires a different way of working within your organization: if your organization normally only receives insights once a week, which is very common in a lot of organizations, receiving these insights every second will require a different approach and way of working. Insights require action and instead of acting on a weekly basis this action is now in real-time required. This will have an effect on the culture. The objective should be to make your organization an information-centric organization.

V. PROPOSED DIVIDE AND CONQUER ALGORITHM

On the basis of analysis findings, we propose continuous natured feature extraction algorithm using traditional divide-and- conquer mechanisms, such as, merge sort. We consider River as a continuous natured feature presented in satellite EO products and provide detection algorithm for extracting rivers from Satellite products. REPTree is a fast classification mechanism using tree structure, however, the technique cannot be applied directly to detect continuous natured feature extraction as River. We used REPTree for small sub-block classification with other ED and sta- tistical measures such as mean, S.D, differences, etc., to find continuous features in EO products. In this section, we present the algorithm details including its parameters, functions, flow charts, and pseudo-code.



The above diagram is an example based on merge sort. The flow of the diagram is as follows.

At each recursive step the input is split into two parts, then the conquer part takes O(n), so each level of the tree costs O(n), the tricky part might be how is it possible that the number of recursive levels (tree height) is logn. That is more or less simple. So at each step we divide the input in 2 parts of n/2 elements each, and repeat recursively, until we have some constant size input. So at the first level we divide n/2, on the next n/4, then n/8, until we reach a constant size input that will be a leaf of the tree, and the last recursive step.

So at the i-th recursive step we divide n/2^i, so lets find the value for i at the last step. We need that n/2^i = O(1), this is achieved when 2^i = cn, for some constant c, so we take the base 2 logarithm from both sides and get that i = clogn. So the last recursive step will be the clogn-th step, and thus the tree has clogn height.

Thus the total cost of MergeSort will be cn for each of the clogn recursive (tree) levels, which gives the O(nlogn) complexity.

The algorithm works as follows :

```
Divide (image_matrix M)
{
    If (size (M) o¼ B size ∂ _ )
    { Set_Rivers ¼ Analyze (M);
        Return Set_Rivers;
    } //end of if
    If (Width_MoHeight_M)
    { // divide m into two parts vertically
        B1¼ M [0- Width_M /2][ Height_M]; //first half of M
        B2¼M [Width_M /2- Width_M][ Height_M];    //2nd half of M
    }
    Else
    {
        B1¼ M [width_M] [0-Height_M/2]; //Uper half of M
        B2¼M [width_M] [ Height_M/2- Height_M]; // Lower half of M
    } // end of if else
    //Recursion and division
    Set_Rivers1¼Divide (B1);
    Set_Rivers2¼Divide (B2);
    Conquer (Set_Rivers1, Set_Rivers2); //combining blocks and results of blocks.
}//end of Divide
Analyze (Image_Matrix_Block B)
{
```

```
Calculate $\bar{X}$_B, S.D_B;
If (S.D_B o Min RB SD ∂ ) // Block does not have any river.
{
    Set_RiverDataClass_Set¼ Φ return Set_RiverDataClass_Set;
}// end of if
Set_RiverDataClass_Set¼Rivers_in_BlocK(B, X ¯B);
For each (RiverDataClass RDC: Set_RiverDataClass_Set)
{
    If (( X X RDC B_¯ −¯ _ o Mean diff ∂ _ ) ‖ (NP_RDC o NP RDC ∂ _ )
    Remove RDC from Set_RiverDataClass_Set;
}
ReturnSet_RiverDataClass_Set; // Return set of rivers detected.
}
Rivers_in_Block (Matrix block B, Double $\bar{X}$_B)
{
    Define Set_R¼ Φ;
    Devide B in to sub blocks S_B of size 10 x 10;
    For each (S_B) Do
    {
        Calculate $\bar{X}$_SB, SD_SB, |$\bar{X}$_B-$\bar{X}$_SB|;
        If(REPTree($\bar{X}$_SB, SD_SB, |X ¯B-X ¯_SB|)¼¼river)
        { Set_R¼ Set_R U S_B; }
    }
    ∀ SBi, SBj ∈ Set_R where (i≠j), if (ED (SBi, SBj) o¼2) the merger SBi, SBj return
    Set_R
}
Conquer (Set of Rivers Set_Rivers1, Set of Rivers Set_Rivers2);
{
    If (Set_Rivers1¼¼Φ) then return Set_Rivers2;
    If (Set_Rivers2¼¼Φ) then return Set_Rivers1; //if either set is empty then no need to combine.
    For each (River R1: Set_Rivers1)
    For each (River R2: Set_Rivers2)
    {
        If(ED (R1, R2) o Ed Rivers ∂ _ ) then R1 þ R2;
        //Combine R1 and R2, remove individual entries of R1 and R2
    }//end of for each loop
    Return (Set_Rivers1 U Set_Rivers2); Combine RiverSet1 and Set_Rivers2
}
```

The proposed algorithm implementation is based on simple Java programming as well as Hadoop. The algorithms are executed on ASAR and MERIS products for correctness and processing time measurements. It detects four rivers from Product 1, 2 rivers from Product2, 2 rivers from Product3, 3 rivers from Product4, and 2 rivers from Product 5. The detection mechanism could be improved depending upon the satellite image quality and its image taking height. The implementation of the proposed algorithm using Map Reduce divide-and-conquer mechanism is more efficient than simple java iteration implementation due to its divide and conquers nature.

VI. CONCLUSION

In this paper, we proposed an architecture Big Data in M2M that uses a divide-and-conquer mechanism for analysis purposes. The proposed system architecture is capable arranging data block in a sequential manner by using machine ID. In order to achieve the computation efficiency, the data fusion domain is used to partition the data block. These data blocks can be equally distributed among various servers that follow the in divide and conquer mechanism. These units implement and design algorithms for each level of the architecture depending on the required analysis. The proposed system architecture is a generic model that is used for any Big Data analysis. The advantage of the proposed system is to extract the features from Big Data depending upon the user requirements. So, we are planning to extend the proposed architecture to make it compatible for efficient and real-time Big Data analysis for all applications like networking, satellite, indexing etc. Furthermore, we are also planning to use the proposed divide-and-conquer architecture for performing a complex real time analysis for observatory data that can help in decision making based on various crucial factors.

REFERENCES

[1] C. Cecchinel, M. Jimenez, S. Mosser, M.Riveill, An architecture to support the collection of Big Data in the Internetof things, in:Proceedings of the 2014 IEEE World Congress on Services (SERVICES), 2014 , pp.442–449, doi:10.1109/ SERVICES.2014.83.

[2] S. Lindsey and C. Raghavendra, PEGASIS: power-efficient gathering in sensor information systems, in: Proceedings of the IEEE Aerosp. Conf., 2002, pp. 1125–1130.

[3] L.I., Xiaoquan, Fujiang, Zhang, Yongliang, Wang, Research on big data archi- tecture, key technologies, and its measures, in: Proceedings of the IEEE 11th International Conference on Dependable, Autonomic and Secure Computing,.

[4] Samuel Marchal, Xiuyan Jiang, Radu State, Thomas Engel, A big data archi- tecture for large scale security monitoring, in: Proceedings of the IEEE Inter- national Congress on Big Data, 2014.

[5] Yi Xiaomeng, Fangming Liu, Jiangchuan Liu, Hai Jin, Building a network highway for big data: architecture and challenges, Network, IEEE 28 (4) (2014) 5–13

[6] Tzu-Chuan Juan, Shih-En Wei, Hung-Yun Hsieh, Data-centric clustering for data gathering in machine-to-machine wireless networks, in: Proceedings of the 2013 IEEE International Conference on Communications Workshops (ICC), 2013, pp. 89–94, doi: 10.1109/ICCW.2013.6649207.

[7] A. Zaslavsky C. Perera D. Georgakopoulos Sensing as a Service and Big Data, arXiv preprint arXiv: 1301.0159 2013.

[8] S. Mosser, F. Fleurey, B. Morin, F. Chauvel, A. Solberg, and I. Goutier, SENSAPP as a Reference Platform to Support Cloud Experiments: From the Internet of Things tothe Internetof Services, in: Managementof resources and services in Cloud and Sky computing (MICAS), workshop, Timisoara: IEEE, Sep. 2012