iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4628

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Healthcare Data Analytics Using GPU

Barkat Hafeez[1], Abu Rehan[2], Mohammad Atif[3], Umiya Mushtaq[4]
[1, 2, 3]*Department of Electronics & Communication Engineering, Al-Falah University, India*

Abstract: Due to recent technological advancements in the field of wearable sensors, we see sudden surge in the amount of data related to the human beings health. Such data include body temperature, blood pressure, oxygen content, heart beat rate etc. Moreover, there is tremendous growth in the social network such as Facebook and WhatsApp. The growth of Smartphone users have also increased by tense of folds. It is expected that this growth will further keep on moving and we will have more data as compared to what we have just few years back. Data related to patient's health can be of great help to the doctors and other Healthcare professionals. It can help them analyze the disease faster at an early stage and also take important decisions regarding a particular patient or regarding some epidemic. In this paper we present our initial ideas of using a GPU for clustering different health groups, namely, healthy, unhealthy and marginally healthy people. We do this clustering using K-means Algorithm and as part of our ongoing work we plan to implement this K-means algorithm on a GPU. Our proposed GPU implementation is expected to be much faster than traditional CPU based implementation. We also highlight various advantages of using a GPU instead of a CPU based solution and also instead of using much specialized solutions such as using ASICS and FPGAs.

Keywords: health monitoring; GPU; ECG; epidemics; datamining

I. INTRODUCTION

When we have real time requirements in a health care problem, often such systems which can respond in real-time are costly. So one of the goals of this research is to find a solution that can be used in real-time and is cost effective also; especially when we have access to huge amount of patients' data [1]. Today we have huge amount of data in general and related to patient's health in particular, thanks to advancement in wearable health care technologies. This data can be used to get various insights about patient's health [2]. An application of a GPU in ANN has been discussed in [4].

We plan to explore commodity graphics cards, which have Graphical Processing Unit (GPU), for this purpose. Since the GPU has huge computational power, it motivates us to use it in our research. We use our parallel version of K-means algorithm to prove our point.

K-means clustering algorithm is one of the most popularly and widely used clustering algorithms in the fields of data mining and image processing. NP complete problem for which the time of computation is un-predictable. We go for data clustering when we have to search interested data in a huge amount of data set. In data mining, the clustering algorithm plays an important role. When the data grows to more extant, the time to produce results also increases. To get faster results in our work GPGPU computing technology has been implemented. This technique uses GPU for normal general computation purpose along with CPU. Since GPU has more computing units compared to CPU, the computation speed is increased and results are obtained at faster rate [6-7]. In this work we apply K means on patient's simulated data such as BP, Temp, Oxygen content etc.

The NVIDIA GPU is a highly parallel architecture with thousands of processing cores. Each core can work in parallel resulting in several order of speed up as compared to normal CPU based serial execution. We first present an overview of NVIDIA GPUs. Then in subsequent sections we discuss algorithm and some interim results. Finally, we present conclusion based on our research done so far.

II. MASSIVELY PARALLEL ARCHITECTURE OF GPU

GPU or a Graphical Processing Unit has lot of processing power. This power comes from a huge number of scalar processors inside a GPU. These processors work in parallel and the computational job is offloaded to all these processors by the job scheduler. The processors inside a GPU will process all the data simultaneously. As a result we expect that we will get performance benefits as compared to a CPU only solution, even though this approach has certain limitations as discussed in [5]. It may be noted that we cannot have a GPU-only system solution and the reason is because there is an operating system and the operating system will run on a CPU, so a CPU should always be there. So in this paper our comparison would be based on a CPU only solution and a CPU-GPU solution. CPU-GPU solution can also be termed as a hybrid system.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

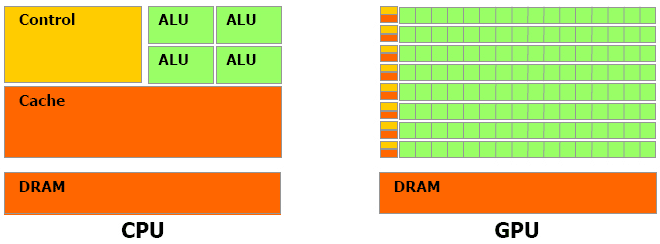


Fig 1: CPU Vs. GPU

Fig shows the layout of the silicon on the chip in case of CPU and GPU. It can be seen that most of the silicon area in GPU is devoted for ALU, shown as small green boxes, while in CPU most of this area is devoted for control and cache. We can see that a GPU provides an economical supercomputing platform for masses! Though there are certain problems such as cache coherence problem. But various solutions have been proposed to solve it[5].

One of the most important conditions for using a CPU GPU system is that there should be a matching between the algorithm that we are putting on the GPU and the GPU architecture. If there is a huge mismatch between these two components we may not get the desired performance and in such a case a CPU based system would be better than a GPU-CPU based hybrid system.

III. GOAL OF ON-GOING RESEARCH

Our health monitoring system will be designed to exploit such features in GPU along with our algorithm to provide real-time or near real time response for health care problems.

IV. OUR APPROACH

A. In this work we trying to understand how commodity GPU based data mining algorithms (specifically K-means algorithm) can help classify various healthcare data in different groups faster than traditional CPU based systems.

B. It involves discovering various hidden patterns in patient's data.

C. K-means algorithm is a widely used data mining algorithm and we hope it can help gain useful insights to decision makers and health professionals.

D. When we have real time requirements in a health care problem, often such systems which can respond in real-time are costly.

E. So one of the goals of this research is to find a solution that can be used in real-time and is cost effective also; especially when we have access to huge amount of patients' data.

F. Today we have huge amount of data in general and related to patient's health in particular, thanks to advancement in wearable health care technologies. This data can be used to get various insights about patient's health.

G. We plan to explore commodity graphics cards, which have Graphical Processing Unit (GPU), for this purpose. Since the GPU has huge computational power, it motivates us to use it in our research.

H. We use our parallel version of K-means algorithm to prove our point.

V. WHAT IS A K-MEAN ALGOROTHM ?

A. The Following Points Have Been Noted About This Algorithm [7-8]

1) Unsupervised learning algorithms that solve the well known clustering problem.

2) The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result.

3) The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycenter of the clusters resulting from the previous step.

4) After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated.

As an example following distribution of patient's health related data is to be clustered among healthy and unhealthy category. We can also have a marginal situation and thus can have three clusters as our goal.

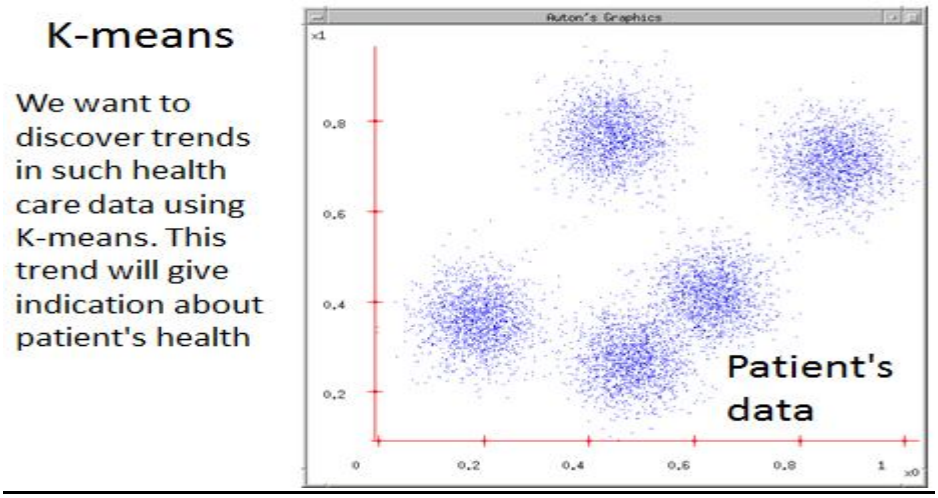


Fig 2: Distribution of patient's related data

The above stated steps of K-means algorithm will give rise to desired no. of clusters:

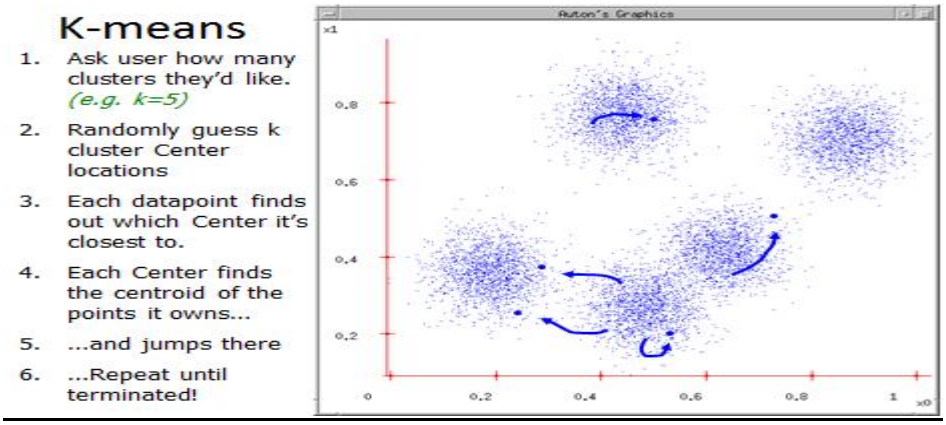


Fig 3: Clustering on patient's related data

VI. INTERIM RESULTS

To understand the computational power of a GPU, we did large no. of experiments (with and without shared memory of GPU). The results are shown below :

TABLE I
COMPUTATION RESULTS ON GPU

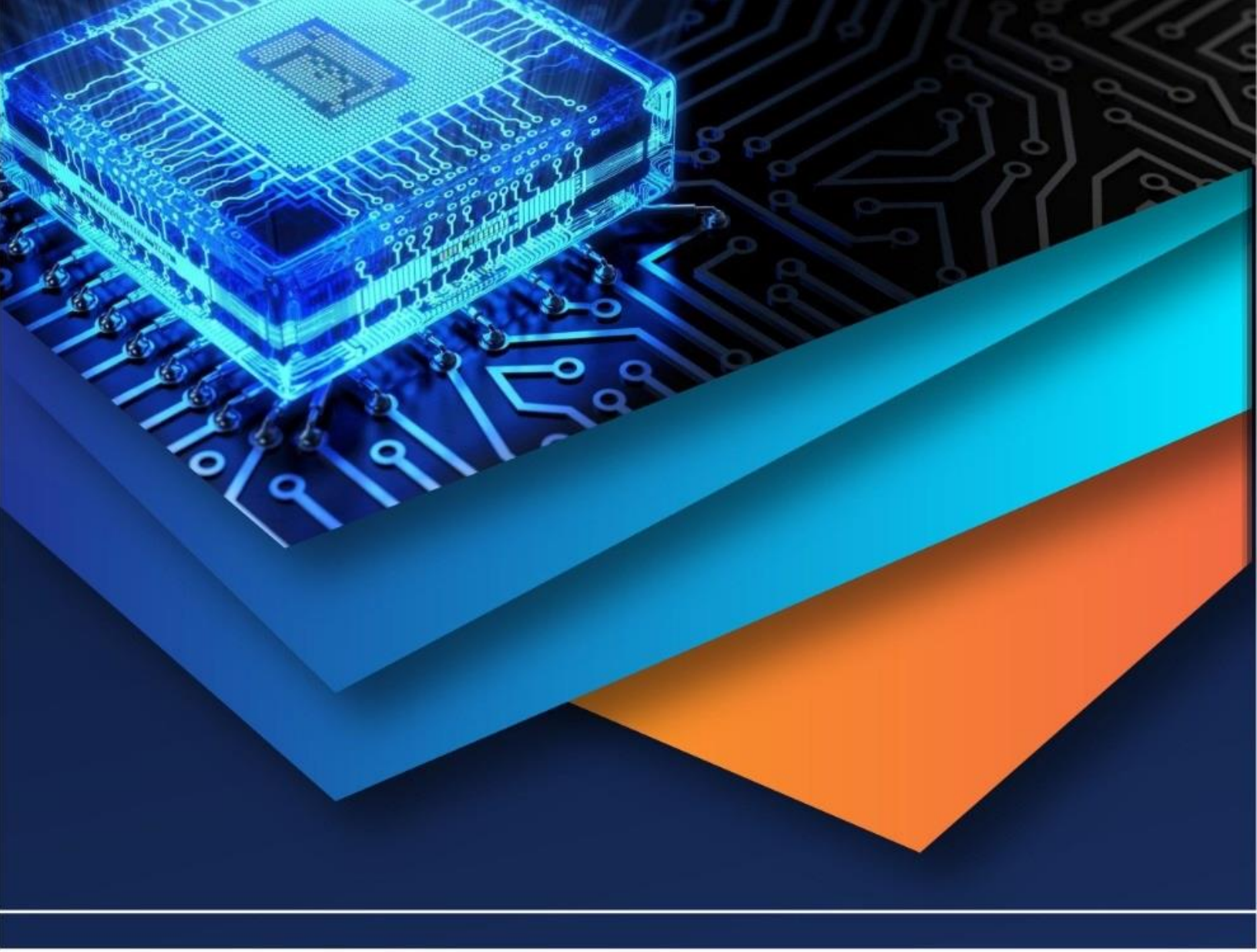
Problem	Grid Configuration	Time Taken Without Shared Memory(ms)	Time Taken with Shared Memory(ms)	Speed-Up
Addition of numbers	1,1,1	0.0101		
Addition of two Vectors	1,1,1	0.0106		
Addition of two matrices	1,1,1	0.0113	0.01136	1
Square of First 500 natural numbers	1,1,1	0.0104		
16*16 Matrix Multiplication	1,1,1	0.0255	0.0168	1.52
128*128 Matrix Multiplication	8,8,1	11.523	0.8630	13.35
1024*1024 Matrix Multiplication	64,64,1	4074.699	401.508	10.148

VII. CONCLUSION

In this paper we presented some of the ideas and results obtained related to our on-going research on an economical and real-time health care analytics system. Healthcare analytics which is economical and real-time can be of great help to various government and non-government bodies, medical practitioners, decision makers etc. It can help to understand and find useful trend from already available healthcare databases. Through different experiments with common and time consuming operations we observed that a GPU based solution can outperform a CPU based solution by more than 10 folds in terms of speed up. We used shared memory of a GPU as a cashing mechanism. Our future work will try to exploit more features of a GPU and thus we expect more performance gain while maintaining accuracy of results.

REFERENCES

[1] Koh H. C., Tan G.Data mining applications in healthcareJournal of Healthcare Information Management, 2005.
[2] Shyamal Patel, Hyung Park, Paolo Bonato, Leighton Chan and Mary Rodgers, A review of wearable sensors and systems with application in rehabilitation, Journal of NeuroEngineering and Rehabilitation,2012.
[3] Jiewen Zheng, Zhengbo Zhang,Taihu Wu, Guang Zhang, Emerging Wearable Medical Devices towards Personalized Healthcare , Conference: Proceedings of the 8th International Conference on Body Area Networks, September 2013.
[4] S Tanweer , A Mobin, A Alam, "Environmental Noise Classification using LDA, QDA and ANN Methods", Indian Journal of Science and Technology, Vol 9(33) , September 2016 ,ISSN (Print) : 0974-6846 ISSN (Online) : 0974-5645
[5] Inderpreet Singh, Arrvindh Shriraman, Wilson WL Fung, Mike O'Connor, Tor M Aamodt, Cache coherence for GPU architectures, IEEE 19th International Symposium on High Performance Computer Architecture (HPCA2013), 2013.
[6] Ramzi A. Haraty, Mohamad Dimishkieh,Mehedi Masud, An Enhanced k-Means Clustering Algorithm for Pattern Discovery in Healthcare Data, International Journal of Distributed Sensor Networks vol. 11,June 2015.
[7] Analysis of K-Means and K-Medoids Algorithm For Big Data, Procedia Computer Science, Volume 78, 2016, Pages 507-512 in 1st International Conference on Information Security & Privacy ,2015.
[8] Bottou L.,Bengio Y.Tesauro and G., Touretzky D.Convergence properties of the k-means algorithms *Advances in Neural Information Processing Systems 7*New York, NY, USAMIT,1995.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY