Original Research Paper Computer Science

Quantum Computing Approach for Big Data

Siddhartha Sankar Biswas Department of Computer Science & Engineering School of Engineering Sciences and Technology Jamia Hamdard University Hamdard Nagar, New Delhi – 62, INDIA.

Parul Agarwal Department of Computer Science & Engineering School of Engineering Sciences and Technology Jamia Hamdard University Hamdard Nagar, New Delhi – 62, INDIA.

ABSTRACT Quantum computing presently at baby stage, even inferior to that. It is expected that by the year 2025, a first version of quantum computer may be developed by the computer scientists. In this paper the author explains that quantum computer may easily deal with big data without using Hadoop or ADS of r-atrain. In Hadoop or in ADS, the space complexity is resolved using distributed system with a centralized master link. These require a number of machines to constitute the distributed system. But this problem of space complexity of today's scenario for handling big data will be no more any issue in quantum computers. A huge amount of parallelism in processing the big data can be possible using quantum computers.

KEYWORDS : quantum, qubit, qubyte, q-register, machine state, big data, train, atrain, ADS, Hadoop.

INTRODUCTION

Quantum computing is the latest expectation to the world scientists and engineers to deal with big data. The genuine dependence on an extraordinary type of computing like quantum computing is being studied by several authors (e.g. Devitt, Munro & Nemoto, 2011; Kanamori, Yoo, Pan, & Sheldon, 2006; Kauffmann, 2012; Manay, 1998; Needham, 2013; Nielsen & Chuang, 2000; Pandey & Ramesh, 2015;Prantosh, 2015). As on today there is no existence of any physical machine called by quantum computer. Even the existing supercomputers are becoming inefficient in processing the big data in many cases. It is because of the reason that superco mputers can solve one or two Vs out of 5Vs of big data. Consequently the quantum computing is becoming very popular as a modern set of tool to the computer scientists, researchers, engineers, programmers to develop and enhance computation capabilities much better than the classical computing capabilities. Quantum computers are expected to become a new approach to solving this issue. The quantum approach to big data could make impossibly complex problems solvable, however time will say when the first ever quantum computer will be developed physically by the computer scientists and engineers.

In this research paper the authors discuss a review analysis of the latest concept about the potential benefits of quantum computing while dealing with big data.

CHALLENGES OF BIG DATA

Big-Data is everywhere and at every moment in and around our daily life frame. It includes data sets with sizes beyond the ability of available software tools to capture, curate, manage, and process the data within a reasonable amount of time. Big data sizes are exponentially expanding in 5Vs, even in many cases to large number of petabytes of data in a single data set. The 5Vs are:

Volume : Large amount of data.
Variety: Data today comes in varieties of formats.
Velocity: data is being produced at high velocity and the data must be processed at high velocity.
Veracity: the quality or habit of retaining the truth. Veracity of a data or information can be challenged by infinite number of false data from variety of fraudulent.
Value: the value of data at any cost must be retained.
In the area of computer science, to represent the big sizes we use the **terms like :** Gigabyte (roughly 109 bytes), Terabyte(roughly 1012 bytes), Petabyte (roughly 1015 bytes), Exabyte (roughly 1018 bytes), Zettabyte (roughly 1021 bytes), Yottabyte (roughly 1024 bytes), etc. For details about big data one could study the literatures (Berman, 2013; Chen, Chaiang & Storey, 2014; Feinleib, 2013; Needham, 2013; Pandey & Ramesh, 2015; Schönberger & Cukier, 2013).

To handle 5 Vs, the basic requirements are:

Massively parallel processing (ex. r-atrain data structure (Biswas, 2015a) which is exclusively for big data, MapReduce framework), Distributed computing platform (e.g., Apache Hadoop, ADS distributed system (Biswas, 2015a), Increase of storage capacities, Increase of processing power , managing the big data using cloud (Agarwal & Biswas, 2017), etc.

Hadoop is an open source software which exploits the concept of "divide and conquer" approach to store the data and follows the client-server architecture. Hadoop makes use of Hadoop distributed file system (HDFS) for maintaining and storing the data.
In (Biswas, 2015a) a special type of distributed system called by 'Atrain Distributed System' (ADS) is proposed which is very suitable for processing big data of any amount of 5Vs using the heterogeneous data structures 'atrain' which is exclusively for big data or the homogeneous data structure 'train' (Biswas, 2015a; Biswas, 2011; Biswas, 2012; Biswas, 2013b; Biswas, 2014a; Biswas, 2014b). The ADS is scalable upto any extent as desired in any 5Vs. Two new type of network topologies are defined for ADS called by 'multi-horse cart' topology and 'cycle' topology which can support increasing volume of big data. Where r-atrain and r-train data structures are introduced exclusively for the processing of big data, the data structures 'heterogeneous data structure MA' and 'homogeneous data structure MT' are introduced for the processing of big data including temporal big data too, however in the form of n-dimensional hypermatrix or hyperlatrix(Alam, 2013; Biswas, 2015a; Biswas, 2013a). In the next section we propose the machine like Quantum Supercomputer to handle big data of any 5Vs.

QUANTUM SUPERCOMPUTER FOR BIG DATA

Classical Supercomputers can give velocity, but added with quantum memory for storage management can make it "Quantum Supercomputer" to deal with big data of all the 5Vs. Because the potential of quantum computing alone in a machine may not be sufficient to handle big data today. The notion of "Quantum Supercomputer" is analogous to the classical supercomputer, but the component computers/processors are just to be replaced by quantum computers.

The beautiful phenomenon of superposition in quantum computing

In quantum computers the basic storage unit is qubit whereas in the conventional computers the basic storage unit is bit. If we have n bits, then assuming the superposition of them we will get 2n qubits. A qubyte is a single unit composed of a sequence of 8 qubits.

In the classical computer one byte stores single number between 0 and 256 at a time whereas in a quantum computer one qubyte can store 256 numbers at the same time, which is the key point to speed up the memory utilization and computation power.

For the sake of presentation, consider all possible combinations of a 3 bit data system with 8 possible states 000, 001, 010, 011, 100, 101, 110 and 111. One 3 bit classic computer can have at the most one of these 8 possible states at one time. In order to go for all eight of them the computer would have to repeat each operation separately eight times. But a 3 bit quantum computer is able to analyse all of these possibilities simultaneously in one operation .This is possible due to the beautiful phenomenon of superposition in quantum computing. For example, the 3-bit can contain one data (say 110) at any instant of time as shown in Table 1, whereas the 3-qubit can contain all the eight data at any instant of time:-

3-Qubit	3-Bit
000	-
001	-
010	-
011	-
100	-
101	-
110	110
111	-

Table 1: The 3- Qubit and 3-Bit data

How Quantum Computers can deal with Big Data of all 5Vs

A machine with n-qubits can be in superposition of 2^n states at the same time. Thus an n-qubit computer could analyse 2n parallel states in a single operation; in comparison a n-bit classical computer can only analyse one state. To achieve the same solution as the Quantum computer classic computer has to repeat this operation 2n times. This powerful potential can be easily exploited to deal with big data.

Thus quantum computer can tackle problem on scale beyond any classical computer, and it is clear that in this way quantum computing can be utilized in a huge scale for analysing the problem of rapidly growing big data.

CONCLUSION

In this research paper we have discussed the possible application of quantum computing to deal with big data which are rapidly growing in 5Vs.

Although the quantum computation is presently at its baby stage, at its initial stage, but the future computing for big data analytics will be very much dependent upon quantum computing only. However, we have proposed a notion of Quantum Supercomputer to cover all the 5Vs smoothly. The concept of Quantum Supercomputers is analogous to the concept of the existing classical supercomputers, but the component computers in them are just to be replaced by quantum computers. Use of quantum computing can enable the world to deal with big data of daily life very efficiently, viz. to predict statistical inferences of large population of even NR-Statistics (Biswas, 2014c; Biswas, 2016) decision making, atmospheric forecasting, large scale pattern matching, large scale web data mining, etc. to list a few only out of many.

REFERENCES

1. Agarwal, P., & Biswas, S.S.(2017). "Big Data on Cloud: A Review". International Journal of Advanced Research in Computer Science, Vol. 8(2), ISSN No: 0976-5697.
2. Alam, B. (2013). Matrix Multiplication using r-Train Data Structure, 2013 in AASRI Conference on Parallel and Distributed Computing Systems: Vol.5. AASRI Procedia. (pp 189-193). Elsevier
3. Berman, J.J. (2013). Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information. U.S.A., Morgan Kaufmann (Elsevier) Publisher.
4. Biswas, R. (2015a). "Atrain Distributed System" (ADS): An Infinitely Scalable Architecture for Processing Big Data of Any 4Vs in Acharjya, D. P., Dehuri, S., Sanyal, S.,(ed.) Computational Intelligence for Big Data Analysis Frontier Advances and Applications:, Springer International Publishing. Switzerland. Part-1, pp 1-53.
5. Biswas, R. (2011). r-Train (Train): A New Flexible Dynamic Data Structure. INFORMATION : An International Journal (Japan), Vol.14(4), pp 1231-1246.
6. Biswas, R. (2012). Heterogeneous Data Structure "r-Atrain". INFORMATION : An International Journal (Japan), Vol.15 (2). pp 879-902.
7. Biswas, R. (2013b). Heterogeneous Data Structure "r-Atrain", Chapter-12 in Tripathy, B. K., Acharjya, D. P.,(Ed.), "Global Trends in Knowledge Representation and Computational Intelligence" IGI Global, USA.
8. Biswas, R. (2014a). Processing of Heterogeneous Big Data in an Atrain Distributed System (ADS) Using the Heterogeneous Data Structure r-Atrain. International Journal of Computing and Optimization. 1(1), 17-45.
9. Biswas, R. (2014b). Data Structures for Big Data. Int. Journal of Computing and Optimization, Vol.1. pp 73 – 93.
10. Biswas, R. (2014c). Introducing Soft Statistical Measures. The Journal of Fuzzy Mathematics. 22(4). pp 819-851.
11. Biswas, R., (2013a). Theory of Solid Matrices & Solid Latrices, Introducing New Data Structures MA, MT : for Big Data. International Journal of Algebra, 7(16), 767–789.
12. Biswas, R. (2016). Introducing 'NR-Statistics': A New Direction in "Statistics". Chapter - 23 in John, S.,J.,(Ed.),"Generalized and Hybrid Set Structures and Applications for Soft Computing", IGI Global, USA.
13. Chen, H., Chaiang, R.H.L. & Storey, V.C. (2012), Business intelligence and analytics: From big data to big impact, MIS Quarterly. Special issue: Business intelligence research, 36(4).
14. Devitt, J.S. , Munro, J.W. & Nemoto, K. (2011). High performance quantum computing. Special issue : Quantum information technology , No. 8, pp.49-55.
15. Feinleib, D. (2013). "Big Data Demystified: How Big Data Is Changing The Way We Live, Love And Learn", The Big Data Group Publisher, LLC San Francisco, USA.
16. Kanamori, Y., Yoo, S.M., Pan, W.D. & Sheldon, F.T. (2006). A short survey on quantum computer. International Journal of Computers and Applications, 28(3).
17. Kauffmann, S. K. (2012). Computer Hardware of the Future: Will the Classical-wave Simulated "Long Qubyte" Trump the True-quantum Qubit? Prespacetime Journal, 3(6) pp. 538-541.
18. Manay, K. (1998). Quantum computers could be a billion times faster than Pentium III. USA Today. Retrieved on December 1st, 2002 from: http:// www.amd1.com /quantum_computers.html.
19. Needham, J.(2013). Disruptive Possibilities: How Big Data Changes Everything. O'reill y Publisher, Cambridge.
20. Nielsen, M.A. & Chuang, I.L. (2000). Quantum Computation and Quantum Information. Cambridge University Press.
21. Pandey, A. & Ramesh, V. (2015). Quantum computing for big data analysis. Indian Journal of Science, 14(43), pp 98-104.
22. Prantosh, P.K. (2015). Quantum Information Science: Emerging basic science focused information science domain. Abhinav-National Monthly Refereed Journal of Research in Science and Technology, 2(9), ISSN 2277-1174.
23. Schönberger , M. V., & Cukier, K. (2013). BIG DATA : A Revolution That Will Transform How We Live, Work, and Think, Eamon Dolan/Houghton Mifflin Harcourt: Publisher.