International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 7
July 2017

A New Framework for Effective Processing of WSN Application Generated Big Data

Mamidi Kiran Kumar[1], Dr.V. Kamakshi Prasad[2]
[1,2]*Dept of CSE, JNTUHCEH, JNTU Hyderabad, Telangana, India.*

ABSTRACT
In real word automation applications needs Wireless Sensor Networks (WSNs) for many controlling or observation activities. In Zig-bee based WSN, end node sensors collect the data continuously from its' surroundings or objects and transfer to the sink node directly or with the help of anchor nodes (routers). Data will be accumulated from long periods at the sink node. WSN generated data has the characteristics of volume, velocity, variety, variability, and veracity. This kind of data characteristics is similar to big data. Localization is an important activity in the WSN applications. Localization can give location information for the applications to provide effective services. WSN generated big data characteristics make it difficult to process using traditional database or software techniques. To process such kind of big data, Text Analytics, Audio/Video Analytics, Social Media Analytics and Predictive Analytics are some of the traditional techniques. The traditional technologies have their own limitations like Hardware: Hardware performance and capacity, Data accessing: Size of data accessing, Data computation: Data distribution and processing. In this paper, a novel framework was proposed to overcome the limitations of traditional techniques.

Keywords: WSN= Wireless Sensor Network, GPU=Graphics Processing Unit, MPP= Massive Parallel Processing, NoSQL= Not only SQL, vCPU=Virtual CPU

1. Introduction
Nowadays, in this digital world, everything is digitalized like automated homes, automatic ticket vending machines, area monitoring, disaster management, online bookings, etc. Wireless sensor networks play a major role in the automation of anything [1]. A huge amount of data will be generated from wireless sensor networks [2], social network sites, and smart home appliances for longer terms, various organizations. Such data have to be accumulated in the data centers and, are to be considered as big data. In WSN applications, data may be stored at Sink node and which will be stored for longer period. In the case of flash flood prone areas, the Emergency team will give sensors enabled smart bands (Multiple sensors embedded on it) to the people. People may wear the smart band [3], and that will be connected to a sink node using the Zig - bee network is given in Figure-1. Smart band continuously collects data from the human body like body temperature, heartbeat, acceleration, etc. and transfer the collected data to the sink node. Data at the sink node will be huge in volume, continuous in velocity, various sensor data in variety, latest data in variables. In WSN applications, data at sink node have the characteristics of big data. If people are more in numbers, then WSN generated data will be stored in multiple data centers in case of multiple sinks. WSN generated big data should be processed effectively and, post processing information will be used for monitoring the people, to know their current health status and their location information. The processed data not only used for monitoring, but also useful for applications like localization of smart bands which will helpful for rescue team [4], efficient packet routing, secure data communication, etc. There are increasing hopes that analysis and processing of WSN generated big data, which will provide insights into the localization of smart bands and, that may be valuable to both rescue teams as well as people who are in danger during the disaster. Since thousands of terabytes (TBs) or hundreds of petabytes (PBs) of data will be accumulated at sink node and it should be processed properly to make full use of it in the form of reports.

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 7
July 2017

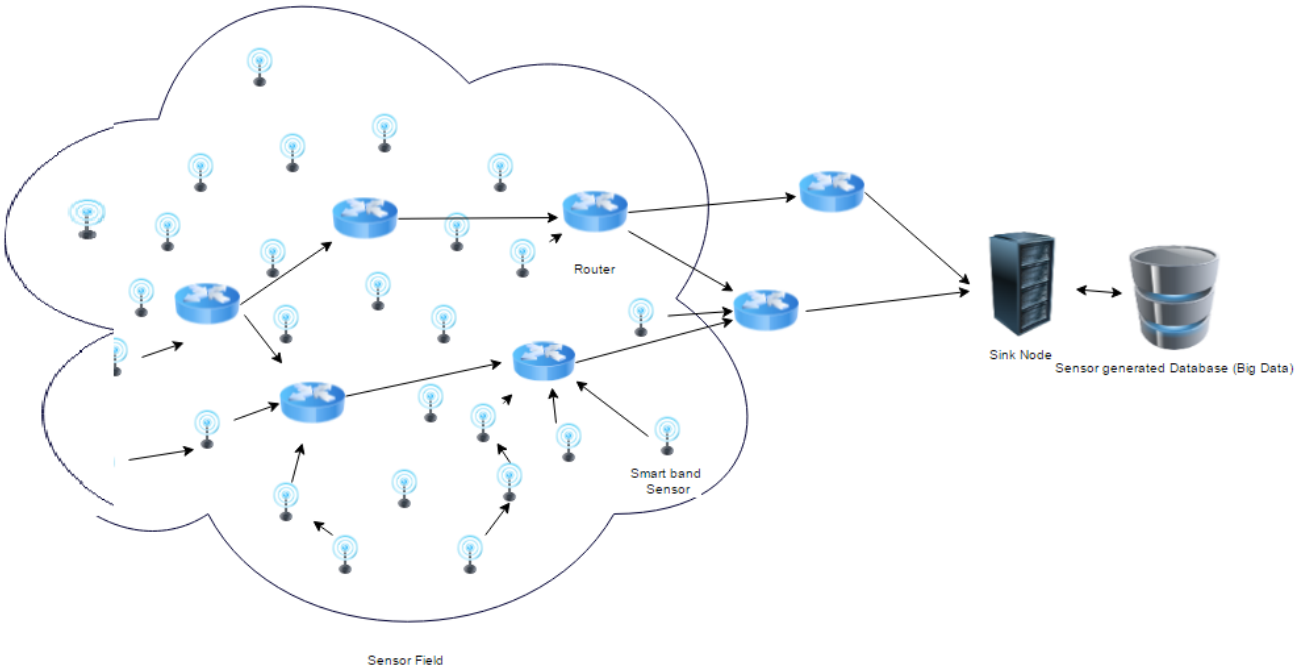

Figure-1: Zigbee Protocol based WSN

There is a need for a new type of technologies for processing of big data [5] and, those technologies should be different from the traditional technologies. In this paper, proposed a new framework, which consists the following technologies like improved hardware performance and capacity, reduced size of data for parallel processing, distributed computing techniques [6] for effective and efficient data processing. The proposed framework can give ultimate solutions for processing of WSN generated big data and provide useful reports than the existing techniques. Reports may be useful to improve the accuracy of target nodes in the WSN localization process.

2. The Problem Statement

Proposed "***A Novel Framework for Effective Processing of WSN Generated Big Data***" which consists of technologies like improved hardware performance and capacity, reduced size of data accessed in parallel processing, distributing the data and parallel processing techniques for effective and efficient data processing.

3. Methodology

The design of the proposed research problem is based on the limitations of the traditional data processing technologies such as the limited processing capacity of the hardware, the capacity of data access mechanisms and distributed data computation techniques. In this research work, solutions for limitations of the existing or traditional data processing techniques are discussed. The proposed framework is represented in Fig 2.

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 7
July 2017

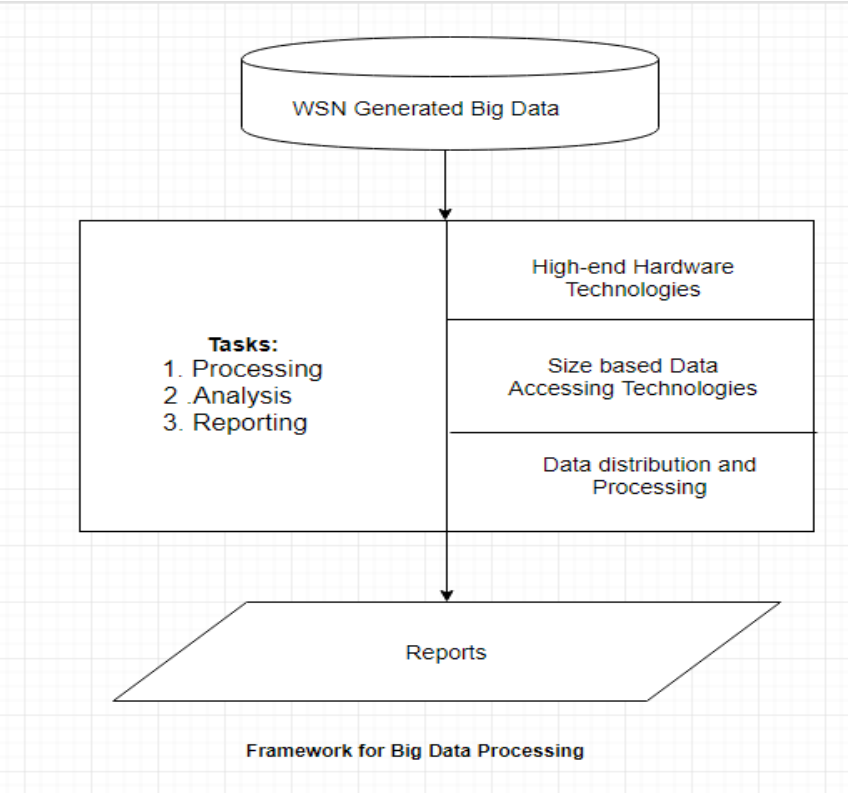

Figure-2: proposed framework architecture

3.1 Problem: Hardware Performance and Capacity

Hardware performance is very poor in the processing of big data [7] like WSN applications generated data. Limited processing and storage memory resources are available to process and a store huge amount of data generated by continuous monitoring applications of WSN. For these challenges, improved hardware performance and capacity techniques have been introduced.

Solution: ***Improved Hardware Performance and Capacity,*** this technique will give directions to hardware developers to use a number of and faster CPU's cores (requires parallel operations to take advantage of multi-core CPU's), increase disk capacity, and data transfer throughput, increased network throughput for processing of WSN generated big data.

***Example*:**

3.1.1 GPU (Graphics Processing Unit): Which has low cost and massively parallel processing power [8]; Due to these characteristics, it is getting importance in HPC (High-Performance Computing). It is physically [9] small to fit inside a desktop machine and is massively parallel as a small scale supercomputer, capable of handling up to thousands of threads in parallel [10] which makes it the maximum exponent of parallel computing.

3.1.2 MPP (Massively Parallel Processing): MPP extends the capabilities of RDBMS-based [11] data warehouses and can store, process petabytes of structured data. MPP spreads data over a number of independent servers, or nodes, in a manner transparent to those using the WSN application generated the database. With modern multi-core CPUs, MPP techniques can be configured to treat each core as a node and run tasks in parallel on a single server. By distributing data across nodes and running database operations across those nodes in parallel, MPP techniques are able to provide fast performance even when handling very large data stores.

3.2 Problem: Size of data accessing

To process the big data, some data must be accessed into RAM. In the case of big data, logical connections will be more among segments of data while processing. Logical links may create data inconsistency problems during its report generation.

Solution: ***Reduced the size of data accessed***

Data comparison and data structures of the user application design, limit the amount of data required for queries.

Example:

3.2.1 No SQL or Bitmaps

Not only SQL **(NoSQL)** databases [12] overcomes the constraints of a traditional RDBMS to deliver higher performance and scalability by introducing a column-oriented database in which the data is stored by columns, and, when possible, turned into bitmaps or compressed in other ways to reduce the amount of data stored. Compressing columns reduced the data to be stored; the combination of compressed data and retrieving only the requested columns speeds query performance by reducing the amount of I/O required and increasing the amount of query data that can reside in memory. NoSQL databases [13] can extend the capabilities of Hadoop clusters by providing low-latency object retrieval functionality. WSN generated big data should be stored in multiple data centers using NoSQL databases. Examples of NoSQL databases include Key – Value systems, columnar systems, Document Databases, Graph Database systems.

3.3 Problem: Data Distribution and Processing

In general, WSN generated data will be distributed and processed in a traditional manner. If the data is huge in size, the traditional methods will not useful for data distribution and processing in a systematic manner.

Solution: ***Distributing the data and Parallel Processing***

Putting data into more disks is to parallelize disk i/o, put slices of data on separate computing nodes that can work on these smaller slices in parallel, use, massively distributed architectures with emphasis on fault tolerance and performance monitoring with high-throughput networks to improve data transfer between nodes.

Example:

3.3.1 Hadoop is a distributed, scalable, Java-based file system [14] that provides a storage layer for large volumes of unstructured data. Hadoop is comprised of Hadoop Distributed File System (HDFS) and the Map Reduce model. WSN application data can be processed with this tool.

3.3.2 Map Reduce is a software framework [15] that simplifies the development and execution of highly parallel applications. The Map function divides a query into multiple parts and processes data [16] at the node level. The Reduce function aggregates Map results to determine the answer to the query. With its distributed, parallel processing capabilities, a Hadoop cluster can rapidly ingest, store, and process petabytes of poly-structured data [17]. Hadoop software coordinates local storage and computation across tens, hundreds, or even thousands of data centers (servers). Each server stores and processes a subset of the data. Since applications execute in parallel, performance and capacity scale with each server added to the cluster.

3.3.4 Virtual Processors and virtual cluster machines

3.3.4.1 Virtual Processors: A virtual CPU (vCPU) also known as a virtual processor [18]. vCPU is a physical central processing unit (CPU) that is assigned to a virtual machine (VM). Virtual machines are allocated one vCPU each. Parallel Virtual Machine (PVM) is a program that enables distributed computing among networked [19] computers on different platforms so that they can perform as a single, large unit for the computer-intensive application. PVM has been used as an educational tool to teach parallel programming.

3.3.4.2 Virtual Cluster Machines: Clustering is an effective technique for ensuring high availability. It's even more effective, flexible and cost-efficient when combined with virtualization technology. Virtual clusters are built with virtual machines (VMs) installed on distributed servers from one or more physical cluster. The

VMs in a virtual cluster are interconnected logically by a virtual network across several physical networks. Virtual clusters are formed with physical machines or a VM hosted by multiple physical clusters [20]. The purpose of using VMs is to consolidate multiple functionalities on the same server [21]. This will greatly enhance server utilization and application flexibility. We can also distribute the template VM to several physical hosts in the cluster to customize VMs.

4. Conclusion and Future work

WSN application generated big data will be processed effectively by the proposed framework, which supports multiple technologies at a time. This framework consists of various advanced big data processing techniques which will solutions to the limitations of the existing techniques. With this framework, utilization of the basic hardware resources in the organization is improved, the speed of data processing higher than the existing techniques and advanced distributed computing introduced systematically. Finally, Processing of data at the sink node in Zig-bee based networks will more useful to perform localization process effectively and efficiently in WSN.

Acknowledgements:

The author would like to thank the Ministry of Human Resource Development (MHRD) for providing financial support for this work under the research scholar's grant.

References:

Case Study:

[1] Home automation using WSN, A case study:
http://asterix.ist.utl.pt/rmr/WSN/CaseStudies/ WSN_Home_Enviroment

Journal Paper:

[2]. Azhar Mahmood, Ke Shi, Shaheen Khatoon, and Mi Xiao, "Data Mining Techniques for Wireless Sensor Net works: A Survey", International Journal of Distributed Sensor Networks Volume 2013, Article ID 406316, 24 pages

[3]. ELKISSANI Kaoutar , Settat,Morroco Pr.Moughit Mohammed r Settat,Morroco Pr.Nasserdine Bouchaib, "Zigbee Routing Opnet Simulation for a Wireless Sensors Network " , (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 12, 2014

[4]. F. Dominici, A. Defina, and F. Dovis, "An Augmented GPS/EGNOS Localization System for Alpine Rescue Teams Based on a VHF Communication Infrastructure", 0-7803-9454-2/06/$20.00/©2006 IEEE, pg 531

[5]. Arindam Hazra Corneliu Holban, "Performance and Capacity Implications for Big Data: redpaper", International Technical Support Organization Performance and Capacity Implications for Big Data January 2014.

[6]. C.L. Philip Chen, Chun-Yang Zhang, "Data-intensive applications, challenges, techniques and technologies: A sur vey on Big Data", Information Sciences 275 (2014) 314–347

[7]. Josepha , J.Jakkulin Joshi, "Data Mining with Parallel Processing Technique for Complexity Reduction and Charac terization of Big Data ", Global research of advance journals, Vol-1, Issue-1 PP. 69-80 ISSN: 2394-5788

[8]. Crist ´obal A. Navarro, Nancy Hitschfeld-Kahler and Luis Mateu, "A Survey on Parallel Computing and its Applica tions in Data-Parallel Problems Using GPU Architectures", Commun. Comput. Phys.Vol. 15, No. 2, pp. 285-329 February 2014.

[9]. *Crist´obal,A.Navarro,NancyHitschfeld-Kahler and LuisMateu,* A Survey on Parallel Computing and its Applica tions in Data-Parallel Problems Using GPU Architectures Commun. Comput. Phys.Vol. 15, No. 2, pp. 285-329, Feb ruary 2014.

[10]. John D. Owens, Mike Houston, David Luebke, Simon Green, "GPU Computing", Proceedings of the IEEE, Vol. 96, No. 5, May 2008

White Paper:

[11]. Moty Fania, Parviz Peiravi, Ajay Chandramouly,Predictive Analytics and Interactive Queries on Big Data, *White Paper:* Intel Reference Architecture Big Data Analytics.

[12]. Oracle, "Hadoop and NoSQL Technologies and the Oracle Database", An Oracle White Paper February 2011

[13]. Oracle, "Hadoop and NoSQL Technologies and the Oracle Database", An Oracle White Paper February 2011

Journal Paper:

[14]. Pooja S.Honnutagi, "The Hadoop distributed file system", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014, 6238-6243

University Notes:

[15]. Google, "Introduction to Parallel Programming and MapReduce", https://courses.cs.washington.edu/courses/cse490h/07wi/readings/IntroductionToParallelProgrammingAndMapRedu_ce.pdf.

Journal Paper:

[16]. Madhavi Vaidya, "Parallel Processing of cluster by Map Reduce ", International Journal of Distributed and Parallel Systems (IJDPS) Vol.3, No.1, January 2012

White Paper:

[17]. Intel, "Extract, Transform, and Load Big Data with Apache Hadoop", White Paper Big Data Analytics

[18]. VmWare, "The CPU Scheduler in VMware vSphere® 5.1", TECHNICAL WHITE PAPER / 5

University Notes:

[19]. Ekaterina Elts, "Comparative analysis of PVM and MPI for the development of physical applications on parallel clusters", Faculty of Physics, Saint-Petersburg State University, Russia JASS 2004, St. Petersburg

Journal Paper:

[20]. Joshua Eckroth, "Teaching Big Data with a Virtual Cluster", SIGCSE '16, March 02-05, 2016, Memphis, TN, USA 2016 ACM. ISBN 978-1-4503-3685-7/16/03

[21]. Mr. Nilesh C. Thakkar, Mr. Nitesh M. Sureja, High Performance Computing: A Survey. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 1, Issue 9, November 2012.