ARPN Journal of Engineering and Applied Sciences

CAEMON: CLOUD ACCESS EXECUTION AND MONITORING FOR BIG DATA ANALYTICS OF SENSOR SYSTEM

R.S. Ponmagal[1, 2], M.P. Chitra[1, 2], P. Dineshkumar[1, 3], V.N.Rajavarman[1, 3] and G. Dhineshkumar[1]
[1]Department of CSE, Dr. M. G.R. Educational and Research Institute, University Chennai, India
[2]Department of ECE, Panimalar Institute of Technology, Chennai, India
[3]Department of CSE, Kodaikanal Institute of Technology, Kodaikanal, India
E-Mail: ponmagal.rs@drmgrdu.ac.in

ABSTRACT

This paper aims to implement an intelligent architectural system to analyze and access the sensor data using Big Data analytics. As cloud resources enable the Wireless Sensor Networks to store and analyze their vast amount of data, Sensor Cloud is designed using Service Oriented Sensor Architecture. Sensor Cloud acts as an enabler for big sensor Data analytics. In the current application these three become the compelling combination. It is proposed to use the Hadoop Distributed File Systems (HDFS) concept to store the streaming sensor data on to sensor cloud for further analysis using MapReduce technique. This paper describes a public sensor cloud delivery model through cloud data analytics for sensor services. The proposed architecture acts as a Cloud Access Execution and Monitoring environment for sensor systems and is able to respond to the requested sensor client applications with greater intelligence.

Keywords: sensor, cloud, big data, HDFS, MapReduce.

INTRODUCTION

We are living in an information era where data is proliferated from Institutions, Individuals and Machines at a very high rate. (Vaikkunth Mugunthan, 2014). Almost all these of these data are diverse in nature. They take different forms such as unstructured and semi structured data. This data is categorised as "Big Data" (Puneet Singh Duggal, *et al*., 2013) due to its sheer Volume, Variety, Velocity and Veracity. It is difficult for the currently available computing infrastructure to handle such Big Data. Since the data generated by heterogeneous wireless sensor networks would fall in Big Data, this research work is carried out to propose an intelligent architectural system to manage this data efficiently by integrating this big sensor Data with Cloud. The data is generated by the connected devices such as Personal Computers (servers or Gateways) and sensors. A wide variety of sensors to sense different environment parameters are collectively arranged to build a distributed sensor network (Weiss *et al*., 2011) topology. The gathered data can be made accessible (Lombriser *et al*., 2011) to other nodes, including a specialized one called sink through a variety of means. The sensor networks are comprised of assorted sensor devices supporting to a large range of applications. Interoperability is essential for such heterogeneous sensor devices. Sensor networks could not operate as standalone networks for such applications. An efficient and flexible method of accessing the data produced by these sensor networks are mandatory.

The proposed work provides the solution by integrating the sensor networks with Cloud through the Service Oriented Architecture model. Service Oriented Architecture (Laurel Reitman *et al*., 2007) is an architectural paradigm that may be used to build infrastructures enabling those with requirements called service consumers and those with potential to create and offer services called service providers to cooperate by means of services across incongruent domains of technology. The sensor nodes are considered as service providers and applications requiring the sensor services are sensor clients. The service oriented architecture for sensor network has been designed in such a way that for the explicit obligation of the sensor client, the services of the sensors are invoked through a registry. The enormous quantity of data delivered from the sensor network requires a massive storage and computing infrastructure to process and analyse the sensor data. Hence, the Service Oriented Sensor Network Architecture SOSA is extended to cloud architecture through Integration controller, in which this sensor services is deployed into a public cloud.
Big data analytics are a set of advanced technologies to work with large volumes (Intel, 2014) of heterogeneous data. Hence it is chosen as the technology to analyze the sensor data which are coming from heterogeneous sensor networks, such as temperature sensor networks, pressure sensor networks, humidity sensor networks, vision sensor networks, etc., The data formats from different sensor networks are physical quantities such as degree centigrade, degree Kelvin, psi, kg/m^2, etc., which are categorized as unstructured data. These data are converted into a unified data format i.e., XML, which is categorized as semi structured data. This xml conversion is necessary to change the sensor data into the web service message. This is essential to integrate WSN with SOA and then extend to the cloud. XML is used to carry the sensor data but at the same time, it is the protocol language of the Internet which can communicate with systems on any platform, and any architecture.

In this work three converging technologies are proposed to be used, hence XML is preferred for building the proposed intelligent architectural system. Due to the specific nature of Big Data in WSN, it is stored in distributed file system architectures. Hadoop and HDFS by Apache (Abouzeid A *et al*., 2009) are widely used for storing and managing the Big Data. HDFS is a file system designed for storing very large files with streaming data access patterns (TomWhite, 2009), and hence it is chosen to handle the data in the sensor system. In this paper, it is proposed to use sensor log files stored as HDFS in the cloud and map reduction process is executed for parallel analysis of very large sensor data. Hive available with Hadoop as the query command tool is used to query the sensor data. Presented cloud architecture takes care of all the complexities of the system such efficient storage using MapReduce and authentication of the sensor service providers through Integration Controller. Since the cloud acts as the backbone for the presented architecture, the sensor client application can easily acquire the sensor data with simple authentication. The proposed system is named as CAEMON, because it is a long running background process that answers requests for sensor services, any time and anywhere to the authenticated clients. Thus it is an intelligent architecture which takes care of all the complexities and keeps the end user simple. They can just provide a user name and password and access the required sensor data.

STATE OF THE ART

The sensor information can be transmitted to the requesting client as (Flavia Coimbra. *et al*., 2005) SOAP messages, which is used to access the sensed information with application independent protocol. A method to access the sensor information using structured data (Nissanka B. Priyantha. *et al*., 2008) and WSDL descriptions are proposed. A number of well-known service discovery protocols are reviewed (Åke Östmark *et al*., 2006) in the context of networked nodes. Integrating wireless sensor networks with other communication networks is an intricate task. A reason is the absence of standardized data exchange format that is supported in all participating networks. XML has evolved to be the communication language of the distributed network. Further, it offers the standard data exchange format between heterogeneous networks and systems. The XML based data representation model (Anbalagan. *et al*., 2010) is presented to model the power system reliability. An XML template objects are introduced making XML usage applicable within sensor networks and different optimized ways (Nils Hoeller *et al*., 2008) of using XML is specified. A cloud storage platform (Weimin Zheng *et al*., 2009) for pervasive computing environments such as wireless sensor networks is explained. The cloud provides scalable processing power and several kinds of (Werner Kurschl, *et al*., 2009) connectable services. In order to keep the end users' loyalty, the cloud should cope with the service failure (Liu, R *et al*., 2011). There is a need for a powerful and scalable high-performance computing and massive storage infrastructure for real-time processing and storing of the WSN (Atif Alamri, *et al*., 2013) data as well as analysis (online and offline) of the processed information under context using inherently complex models to extract events of interest.

The recent trend of digitising all the data has resulted a lot of large and real time data across a broad range of industries. Much of the unstructured, semi structured and hybrid data such as streaming, Geo-spatial and sensor generated (Pethuru Raj *et al*., 2014) data do not fit neatly into traditional structured relational data warehouse models. Many IT companies attempt to manage big data challenges using a NoSQL ("not only SQL") database, such as Cassandra or HBase, and may employ a distributed computing system (Jinbao Zhu, *et al*., 2012) such as Hadoop. Hadoop is an open source software framework which stores large unstructured data (HDFS) and processes (MapReduce) the unstructured data and allows the distributed processing of large scale data sets (Michael Minelli, *et al*., 2014). Large scale data processing frameworks like MapReduce (Dean J, Ghemawat S, *et al*., 2008) have been integrated with cloud to provide a powerful computation capability for applications. HDFS is the file system of Hadoop used for storage.

A project is developed using a web application to be made available as a software as a service (SaaS) for sensor data analytics (Kumaraswamy Krishnakumar *et al*., 2011) and visualisation. There are some limitations in terms of Hadoop's interfaces and performance (Abouzeid A, *et al*., 2009). A scalable two phase TDS approach, to ananymise a large scale data sets using the MapReduce framework (Xuyun Zhang, *et al*., 2014) on the cloud is proposed. A noteworthy quantity of research and profitable activity has focused on integrating MapReduce and structured database technologies (Ferrera P, 2013). With fully connected multi-layer network, the data-level parallelism (Kunlei, et al., 2014) is performed, concerning the communication cost. A Multiple Query Optimization framework, SharedHive (Tansel Dokeroglu, *et al*., 2014) is proposed using MapReduce, to improve the overall performance of Hadoop Hive.

PROPOSED SYSTEM

It is proposed to converge three promising technologies to achieve an intelligent architecture, which enables the sensor system client to access and analyse the sensor data. The critical industrial process parameters such as temperature, pressure and level are monitored at the control stations. The physical quantity measured with analog or digital instruments are converted into electrical quantity such as voltage or current. This is required that these parameters have to be transmitted to a control station for monitoring which is located in a remote place. The strategy adopted for monitoring, analysis and control through the said procedure is applicable within the coverage area of a particular industry only. Recent trends in VLSI technology to manufacture micro sensors, have led to the deployment of low cost and low power sensing devices to sense the process parameters with computational capabilities to a wide range. Sensors used in paper plant (José A. et al., 2006) monitoring system are considered.

Figure-1 shows the overall architecture with heterogeneous Wireless Sensor Networks, Cloud integration of Wireless Sensor Networks and arrives at a relevant data by analyzing the vast amount of sensor data after mapping it as a distributed file system through Big Data analytics using the Hadoop framework in the cloud.

WSN1 and WSN2 represent heterogeneous Wireless Sensor Networks. WSN1 could be temperature sensors, wirelessly networked together. WSN 2 could be pressure sensors, wirelessly networked together. Since the Wireless Sensor Networks produce physical quantities such as temperature, which is an unstructured data? This is converted into semi structured data such as XML, which is essential for the proposed web services based deployment in the Cloud. With appropriate APIs they are mapped onto the cloud servers. The Cloud acts as a backbone system which integrates Wireless Sensor Networks and Big-Data environment to analyse the sensor data as given in the Figure-1. Finally the sensor client applications can access their required sensor data for further analysis using Hive, a query command tool. This is possible only after the client is authenticated. It is also possible for the clients to access the sensor data through mobile, leading to any time anywhere, a ubiquitous computing environment. Figure-2 shows the layered architecture of presented system.

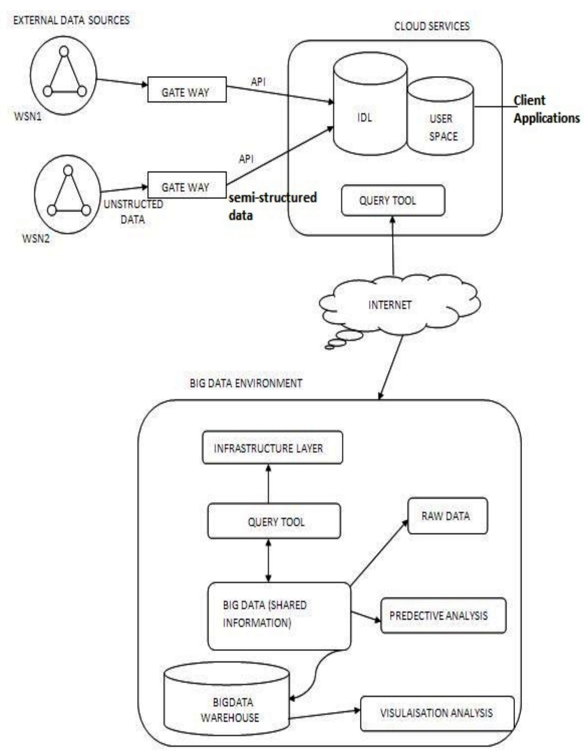

Figure-1. Big sensor Cloud data analytics architecture.

Service oriented sensor network architecture (SOSA)

The SOA is an approach to develop a distributed system such as sensor systems. Web services are a technology that is well suited for implementing service oriented architecture and hence it is chosen to implement the SOSA. The developed sensor services are used to support interoperability among the heterogeneous sensor networks and the server interaction over a network. This interoperability is achieved through a set of XML-based standards, such as the Sensor Service Description

Language (SSDL), Sensor System Registry, and Simple Object Access Protocol (SOAP) messages.

Figure-3 depicts the Service Oriented Sensor Architecture for distributed networked industrial sensors. The sensor services are deployed in the server as service description. To facilitate aggregation of services into applications, an eb-xml-registry is used. The sensor services are published into the eb-xml registry using sensor service description language. The lists of services are discovered and invoked by the sensor client applications.

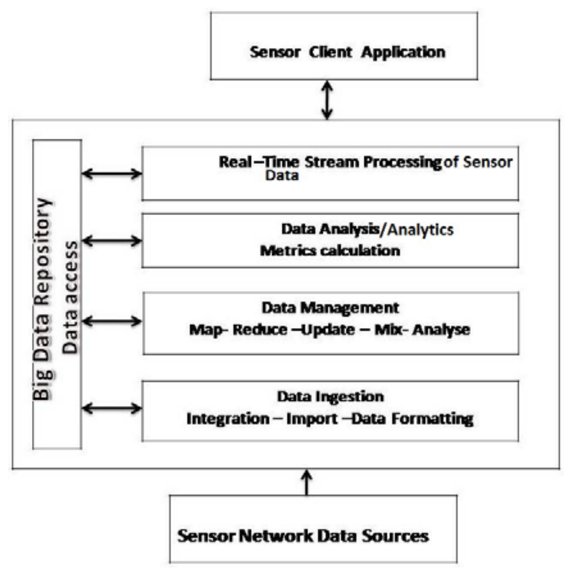

Figure-2. Layered architecture for big-data analytics of sensor system.

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

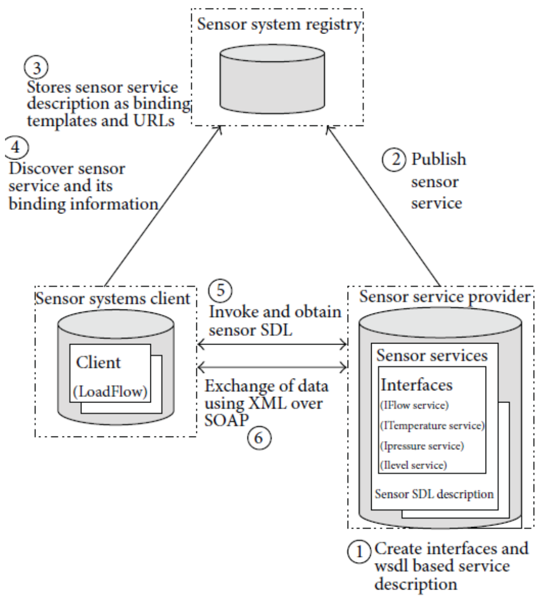

Figure-3. Service oriented sensor network architecture.

Sensor Cloud architecture

An environment is necessary to administer the massive amount of data from sensor networks. Hence, the designed SOSA is extended to Cloud architecture, where heterogeneous sensor networks are connected with the cloud. This connectivity is established through an intelligent component called Integration controller. IC plays the role of bridging the two technologies, SOSA and Cloud.

The preferred communication method from the client`s computer to the cloud computing environment is based on web services, which is implemented in SOSA. In this architecture the Integration Controller is used to upload the sensed data from diverse sensor networks to the public Cloud designed. A bridge program is written to read the Sensed data and convert into XML form of data and store it on a Web Server available in Cloud.

Integration Controller offers the authentication support to the sensor service providers and recommends privileged services to the sensor clients and hence it becomes intelligent component. Figure-4 shows the architecture of SOSA extended to Cloud. The proposed architecture facilitates the sensor clients to easily access, process and search large amounts of sensor data from different sensor networks. It chains complete sensor data lifecycle from data collection to the backend storage system. Sensor data access has thus moved from a loosely managed system to a well managed cloud. A complete data management system for sensor data is designed.

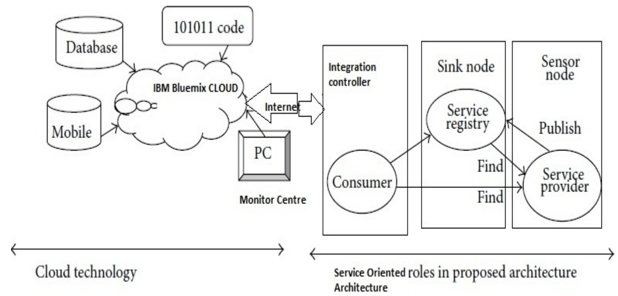

Figure-4. SOSA extended to Cloud architecture.

Big data analytics for sensor services

Pervasive and Ubiquitous connectivity of the sensors as intelligent systems have initiated a set up of new storehouse of valuable information from the sensor networks. Applying big data analytics to data from sensors becomes mandatory. This section describes how cloud and big data technologies are converging to offer an efficient delivery model for cloud-based big sensor data analytics. Sensor cloud is used to store sensor data from diverse sensor networks and to derive meaningful analytics that respond to sensor client application needs. Big data environments have to be supported with clusters of servers and tools to process the large volumes and high velocity assorted sensor data. Clouds are deployed on pools of server without any versioning problems, storage infrastructure, and also networking resources. Cloud computing offers an effective way to support big data technologies and the advanced analytics applications. Hence, it is proposed to consider big data analytics technique which are processed on the cloud is implemented to arrive at a value for sensor services. Since the proposed sensors in the sensor system are continuously providing data for monitoring, these streaming huge amounts of data have to be stored for future statistical report preparations and for further analysis.

Predictive analytics of sensor data enable industries to move to a future-oriented view of what is ahead and offers industries some of the most exciting opportunities for deriving value of big sensor data. Real-time data such as streaming sensor data from sensor networks provides the opportunity for fast, accurate, and flexible predictive analytics that quickly adapt to changing environmental conditions in the industries. The faster the data are analyzed, the timelier the results and the greater will be its predictive value.

The key capabilities of the proposed big sensor data analytics in sensor Cloud are: Capturing and extracting semi structured data (xml) from trusted sensor sources. Managing and controlling data in compliance with specific industry requirements (paper plant). Performing data integration, and analysis to deliver the accurate sensor data to the right location (client application) and at the right time.

IMPLEMENTATION

Temperature sensor as service deployment

The sensed data such as temperature is measured through WSN-EDU2110CB Wireless Sensor Network Educational Kit operating at 2.4 GHz with Data Acquisition boards with temperature (MDA100) sensor and PC Interface Boards (MIB520). The J2EE 1.4 platform provides comprehensive support (Qusay H. Mahmoud, 2004) for web services through the JAX-RPC 1.1 API. Here it is chosen to implement as the sensor service provider. The interface and implementation files for temperature, pressure and flow are coded. Configuration files are written to specify the XML namespace and target namespace. These files are compiled to generate SSDL which contains possible inputs and server's address for client reference and mapping file which contains port number and service endpoint location for server reference. War files are generated by the deployment tool from the services written, and are deployed in the server.

Sensor system registry

To facilitate aggregation of services into applications, sensor system registry is used. To publish the services, the eb-xml registry is used. The SSDL files for sensor services with the appropriate service bindings enabled the services to be registered in the repository. Figure-5 shows the list of services available at the registry.

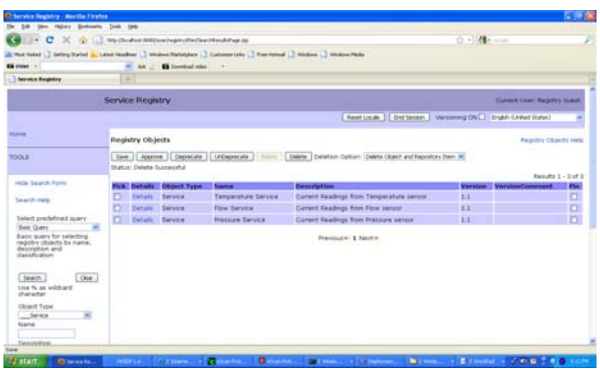

Figure-5. Sensor system registry.

Sensor as a service in Cloud

Integration Controller will upload the sensed data to the Cloud server. Figure-6 shows the sample XML code of sensed data deployed in the Cloud.

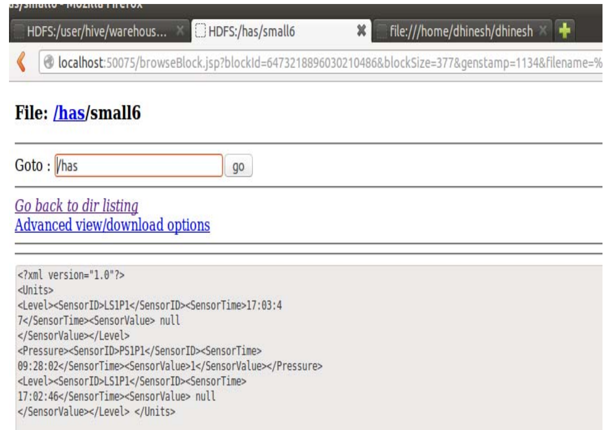

Figure-6. Sensor file (xml) deployed in the Cloud.

Sensor data analytics using MapReduce HDFS

The sensor data in XML format are stored in the cloud server. This is mapped on to the Hadoop file system in the cloud to store the sensor files. Hadoop (Apache, 2013) clusters are built based on the Open Stack cloud platform to facilitate such large-scale data processing. Hadoop is the open source cloud computing platform which supports MapReduce programming framework and mass data storage with good fault tolerance. Here it is used to store and process the sensor data. MapReduce is a popular distributed implementation model proposed by Google (S. Ghemawat, 2003) which is inspired by the map and reduces operations. Map reduce provides a programming paradigm for performing (Kuneli Zhang, et al., 2014) distributed computation.

Hadoop is written in Java and all Hadoop file system interactions are mediated through the Java API. Each component in hadoop is configured using an XML file. Standalone mode is suitable for running MapReduce programs (Tom White, 2009) during development. The Java abstract class org.apache.hadoop.fs.file System represents a file system in Hadoop. Hence Java is chosen as the best choice to implement the HDFS for XML sensor files in the Cloud. Figure-7 shows the sensor log file mapped in the cloud, named as large.

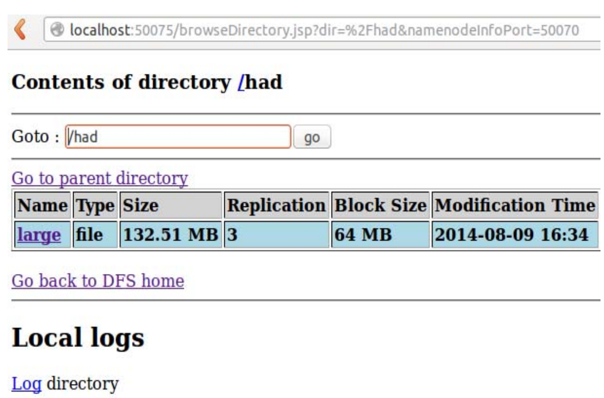

Figure-7. Large sensor file mapped as HDFS in Cloud.

VOL. 10, NO. 2, FEBRUARY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

This large log file is reduced into three files using Map reduce classes, methods and objects as given in Figure-8. These three files can be stored in different three servers in the cloud. These files can be processed concurrently to achieve one final required output. Hence the processing time can be reduced. The sensor output after successful MapReduce technique may also be made available to the mobile clients after appropriate authentication.

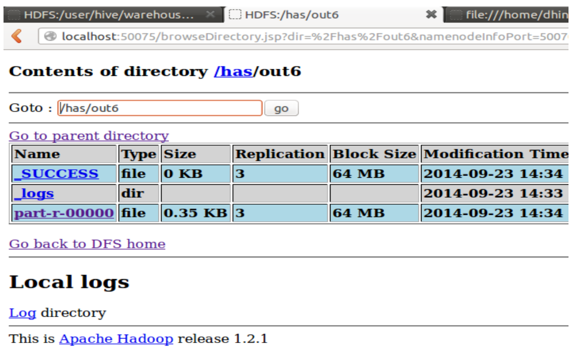

Figure-8. Large sensor file is MapReduced to three blocks.

The sensor data can be obtained from the sensor cloud using Hive query command tool available with Hadoop. A query command is executed to collect and interpret the sensor data using Hive. Figure-9 shows the analytics of the given sensor file using MapReduce technique.

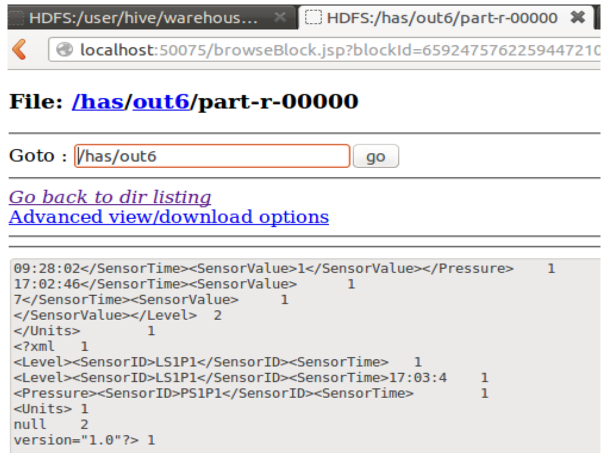

Figure-9. Sensor data analytics.

Figure-10 shows the hive query command executed to retreive the output for the pressure sensor. From the Figure-10, it is understood that it takes only 13.98 seconds to query the pressure sensor value using the Hive query command, which proves the quick response time of the proposed system.

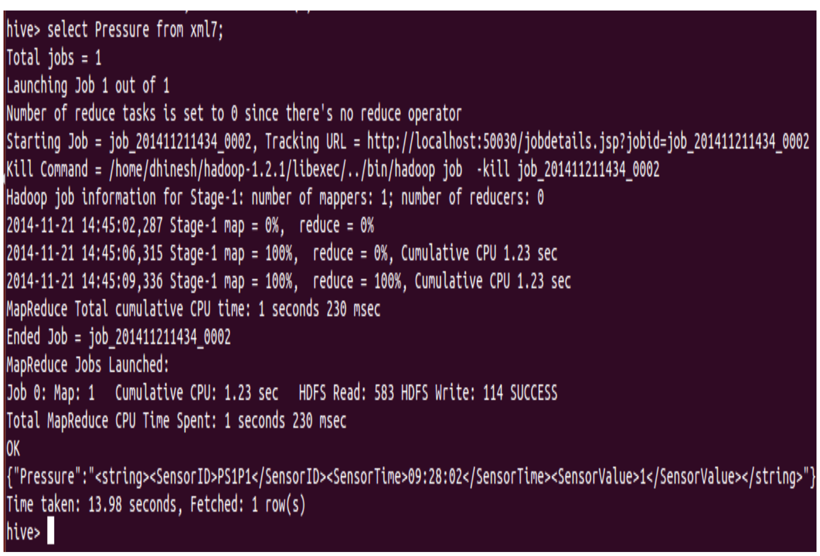

Figure-10. Pressure sensor value through MapReduce.

RESULT ANALYSIS

In this section, the effectiveness and efficiency of the proposed system are analysed. The experiments are recorded by running the system set up for 10 times. The average of the found result is represented. The average response time and execution time of sensor data at the cloud is anlysed.

The average response time to access the converted web service message of sensor data is compared with simple SOA and with Cloud. From the Figure-11, as the number of sensor nodes increases, the average execution time is also increasing proportionally, in both the cases. The noteworthy point is that the average response times is always higher with Cloud based access,

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

as the propagation time, queing time and transmit time for the data is to be taken into account. The proposed system consumes 25 percent more response time on average than the stand-alone SOA based Web service model. This considered as overhead incurred when the sensor data is processed in cloud servers. The performance overhead is introduced when the system is made to scale.

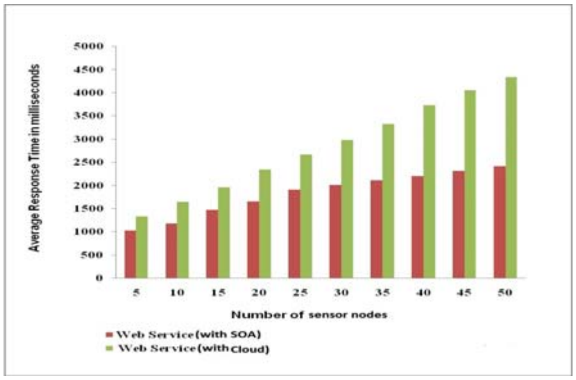

Figure-11. Average reponse time w.r.t no. of sensor nodes.

Figure-12 shows the execution time in milliseconds with respect to data size from 5GB to 15 GB sent from sensor nodes as web service messages. Form the Figure-12, it is noticed that the execution time goes up as the number of sensor node increases. It is important to notice that, the execution time decreases for 15GB of data. It shows the efficiency of the system, i.e. for processing large volumes of data, the proposed system suits well.

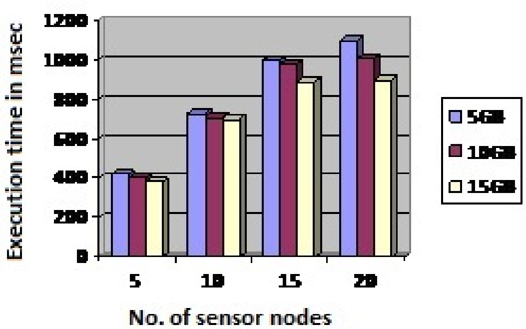

Figure-12. Change of execution time w.r.t data size.

The execution time is measured to be in seconds to run different partitions during MapReduce technique. Figure-13 shows that the execution time is converging, for 5GB, 15GB and 20 GB data with 8 partitions. The execution time is around 700 seconds for all the data sizes. This shows the parallelization of MapReduce technique brings benefit to the presented system.

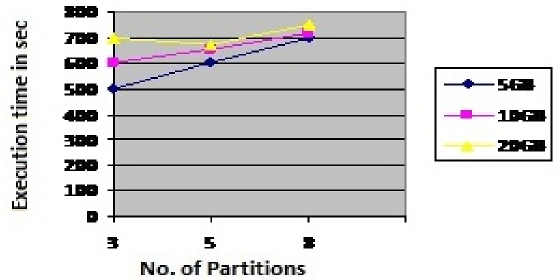

Figure-13. Change of execution time w.r.t no.of partitions.

CONCLUSIONS

An intelligent architectural system to analyze the sensor data with converging technologies such as Wireless Sensor Networks, Cloud computing and Big Data analytics is implemented and the result analysis are investigated.The combination of wireless sensor networks, with their large collected sensor data, with a cloud computing infrastructure makes it attractive in terms of integration of sensor network platforms from different vendors, scalability of data storage, scalability of processing power for different kinds of analysis, worldwide access to the processing and storage, and be able to share the results of sensor data analytics more easily. To start with service oriented architecture is used to build the sensor system and the necessary abstraction was implemented though web services and xml technologies. Hadoop Distributed File Systems (HDFS) is implemented to store the streaming sensor data on to sensor cloud for further analysis using MapReduce technique.

This paper described a public sensor cloud delivery model through cloud data analytics for sensor services. The proposed architecture acts as a Cloud Access Execution and Monitoring environment for sensor systems. The proposed system proves the performance improvement in terms of less execution time for partitioning the sensor data. The proposed intelligence is achieved through the Integration controller which does the authentication of sensor service providers and providing preferred services to the sensor clients in a quick manner. Further the CAEMON runs constantly in the background and handles all the complexities, and the system is able to react to the requested sensor client applications with simple security credentials such as user name and password. This feature also provides intelligence to the proposed system.

As a future work the streaming sensor data analytics could be implemented using Jubatus. Jubatus, a distributed processing framework with fault tolerance, can be used to handle huge streaming data which are coming on the fly. Using this tool, sensor data analytics can be implemented by correlating, consolidating, and contextualizing even more diverse sensor data sources for long periods of time and in real time basis.

REFERENCES

Vaikkunth Mugunthan. 2014. Big Data and Predictive Healthcare. International Journal of Scientific and Engineering Research. 5(8): 168-172.

Puneet Singh Duggal, Sanchita Paul. 2013. Big Data Analysis: Challenges and Solutions. International Conference on Cloud, Big Data and Trust 2013. pp. 269-276.

B. Weiss, H.L. Truong, W. Schott, T. Scherer, C. Lombriser and P. Chevillat. 2011. Wireless Sensor Network for Continuously Monitoring Temperatures in Data Centers, IBM Research Report, RZ 3807.

C. Lombriser, U. Hunkeler and H.L. Truong. 2011. Centrally controlled clustered wireless sensor networks, IBM Research Report, RZ 3811.

http://www.adobe.com/enterprise/pdfs/Services_Oriented_Architecture_from_Adobe.pdf Technical white pape.r

Big-Data in the Cloud: Converging Technologies, Intel IT centre, September 2014.

Anbalagan, Malarvizhi, Ponmagal. 2010. An XML data representation model for power system reliability assessment. 5th International Conference on Industrial and Information Systems, ICIIS. pp. 458-462

Nils Hoeller, Christoph Reinke, Jana Neumann, Daniel Boeckmann, Volker Linnemann. 2008. Efficient XML usage within wireless sensor networks. International Conference on Wireless Internet WICON08.

Werner Kurschl, Wolfgang Beer. 2009. Combining cloud computing and wireless sensor networks. International Conference on Information Integration and Web-based Applications and Services, pp. 512-518.

Abouzeid A, Bajda-Pawlikowski K, Abadi D, Silberschatz A, Rasin A. 2009. HadoopDB: an architectural hybrid of MapReduce and DBMS technologies for analytical workloads. Proc VLDB. 2(1): 922-933.

F. C. Delicato, P. F. Pires, L. Pirmez and L. F. R. da Costa Carmo. 2005. A service approach for architecting application independent wireless sensor networks. Cluster Computing, Elsevier B.V. 8(2-3): 211-221.

Weimin Z., Pengzhi X., Xiaomeng H. and Nuo W. 2009. Design a cloud storage platform for pervasive computing Environments. Cluster Computing. 13(2): 111-112.

Åke Ö., Jens E., Lindgren P., VanHalteren A. and Lianne M. 2006. An Infrastructure for Service Oriented Sensor Networks. Journal of Computers. 1(5): 20-29.

R. Liu and I. J. Wassell. 2011. Opportunities and challenges of wireless sensor networks using cloud services. In Proceedings of the Workshop on Internet of Things and Service Platforms (IoTSP '11), Tokyo, Japan.

Atif Alamri, Wasai Shadab Ansari, Mohammad Mehedi Hassan, M. Shamim Hossain, Abdulhameed Alelaiwi and M. Anwar Hossain. 2013. A Survey on Sensor-Cloud: Architecture, Applications, and Approaches. International Journal of Distributed Sensor Networks, Vol. 2013, Article ID 917923, 18 pages http://dx.doi.org/10.1155/2013/917923.

Pethuru Raj, Ganesh chandra Deka. 2014. Handbook of Research on Cloud Infrastructures for Big Data Analytics. IGI global book series.

Kumaraswamy Krishnakumar, Ling wan. 2011. A Framework for IoT Sensor Data Analytics and Visualisation in Cloud Computing Environments Distributed Computing Project COMP90019, University of Melbourne.

Michael Minelli,Michele Chambers, ambiga Dhiraj. 2014. Big data Big Analytics, Wiley CIO Series.

Jinbao Zhu,Allen Wang, "Data Modeling for Big Data", 2012.

Tansel Dokeroglu, Serkan Ozal, Murat Bayir, Muhammet Cinar, Ahmet Cosar. 2014. Improving the performance of Hadoop Hive by sharing scan and computation tasks. Journal of Cloud Computing. 3: 12 doi:10.1186/s13677-014-0012-6.

José A. Gutiérrez, David B. Durocher, Bin Lu, Ronald G. Harley, Thomas G. Habetler. 2006. Applying Wireless Sensor Networks in Industrial Plant Energy Evaluation and Planning Systems. IEEE IAS Pulp And Paper Industry Conference in Appleton.

Dean J, Ghemawat S. 2008. MapReduce: simplified data processing on large clusters. Commun ACM. 51(1): 107-113.

Ferrera P, De Prado I, Palacios E, Fernandez-Marquez JL, Serugendo GDM. 2013. Tuple MapReduceand Pangool: an associated implementation. Knowl Inf Syst 1-27. doi:10.1007/s10115-013-0705-z.

S. Ghemawat, H. Gobioff and S. T. Leung. 2003. The Google File Sys-tem. The 19th ACM Symposium on Operating Systems Principles. pp. 29-43.

Kunlei Zhang and Xue-Wen Chen. 2014. Large scale deep belief Nets with MapReduce. IEEE Access, DOI 10.1109. pp. 395-403.

ARPN Journal of Engineering and Applied Sciences

Tom White. 2009. Hadoop: The Definitive guide. O'Reilly Publications.

Apache. 2013. Hadoop. http://hadoop.apache.org.

Xuyun Zhang, Laurence T. Yang. 2014. A scalable Two-Phase Top-Down Specialisation Approach for data Anonymization using MapReduce on Cloud. IEEE Transactions on Parallel and Distributed Systems. 25(2): 363-373.

Qusay H. Mahmoud. 2004. Developing Web Services with Java 2 Platform, Enterprise Edition (J2EE) 1.4 Platform. Oracle Technology Network Article.

Jonathan R Owens, Jon Lentz, Brian Femiano. 2013. Hadoop real world solutions cook book. Packt Publishing Ltd. Livery Place35 Livery Street Birmingham B3 2PB, UK. ISBN 978-1-84951-912-0

Themis Palpanas. 2012. Real-Time Data Analytics in Sensor Networks. University of Trento.

Charu C. Aggarwal. Managing and Mining Sensor Data. IBM T. J. Watson Research Center, Yorktown Heights, NY, USA.

http://jubat.us/en/.