# ijRASET

# INTERNATIONAL JOURNAL
# FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

www.ijraset.com

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

# Big IoT Data Analytics by Combining Batch and Stream Processing Technologies

Sneha Kharole[1], Ms. Nisha Balani[2], Ms. Parul Jha[3]

[1,2,3]*Computer Science & Engineering, Jhulelal Institute of Technology RTMNU, India*

*Abstract- The Internet of things is a current major developing technology, which is a network of everyday physical objects that enhances the quality of lifestyle. Application of the internet of things encounters dealing with huge amount of data. One of the directions of big data is this huge amount of data with respect to the internet of things. As the name implies, big data refers to the data that cannot be analyzed by a traditional data processing software. The key challenge of this phenomenon is to use a proper way to analyse, which can provide useful features from the data absorbed by the perception layer of the internet of things in order to provide feedback to end users, which helps them in better decision making and improves the performance of the corresponding internet of things network. Analysis of big data in the internet of things is obviously a hard task. Data storages are distributed and there should be parallel data processing. Transmission of the data across the network can slow down because of the massive amount of data. In this regard, this paper focuses on how to analyze the massive and heterogeneous data of the internet of things in a proper way. At first, the internet of things and the big data are discussed separately with architectures, applications, challenges etc. Since these two technologies are interrelated, data analysis in the internet of things is discussed with various methodologies and challenges. Finally, the study discusses a proper framework that can analyze the big data in the internet of things (IOT) in an efficient way.*
*Keywords: Data storages, heterogeneous, big data, IOT.*

## I. INTRODUCTION

The Internet of things (IoT) is one of the most famous research challenges in the current world. The IoT implies a network of physical objects that are being used in day to day life. They are connected through the internet and allowed to collect and exchange data. Objects are connected by wires or in a wireless way. When they transfer data between each other, a unique address is needed for each object in order to differentiate them on the internet. Internet protocol (IP), which is an open protocol, provides unique addresses for internet connected devices. Internet is the fourth version of internet protocol, addresses.

Because of the rapid development of the internet, this quantity has been getting limited. When the IoT came on to the stage, it has obviously become insufficient. As a result of that, protocol version 6 (IPv6) was introduced. It is a 128 bits address and theoretically addresses. The basic IoT architecture consists several layers. Bottom layer of the IoT architecture which is called Perception layer consists of components such as Sensors, RFID (Radio Frequency Identification) tags, bar code labels, GPS (Global Positioning System) devices, cameras etc. Data collection (sensing) is done on this layer. Then, there is the network layer which collects data from the lower layer and sends them to the internet. Middleware layer is the next layer which serves the management and storage of data. After that, there is the application layer. Its purpose is the final presentation of data. Finally, it has the business layer, which forms meaningful services by the data from the application layer.It is obvious that IoT sticks with a massive amount of data, which can be called big data. Big data refers to a large amount of structured or unstructured data that cannot be processed by a traditional application software. To gain values from data, the big data should be analyzed. Data scientists have defined four characteristics of the big data. They are volume (amount of data), variety (types of data), velocity (speed at which data are collected, stored, analyzed and visualized) and veracity (completeness and accuracy of data).

### A. Batch Processing Method

Batch layer is responsible for deploying this method. It keeps the master copy of a dataset and pre- computes batch views on that dataset. Normally, this output is stored in a read only data base and updates replace the existing batch views. Queries run on those batch views instead of master dataset.

### B. Steam Proceeding Method

This process is accomplished by the speed layer. It manages the low latency requests through real time data processing, which generates real time views. Those views are updated within a very short time period. It is responsible for filling gaps which are generated because of the batch layer. There is another layer which is called the serving layer, and it is responsible for indexing and exposing the views. It responds to low latency ad hoc queries.

## II. RESEARCH ANALYSIS

The key challenge of this phenomenon is to use a proper way to analyse, which can provide useful features from the data absorbed by the perception layer of the internet of things in order to provide feedback to end users, which helps them in better decision making and improves the performance of the corresponding internet of things network. Analysis of big data in the internet of things is obviously a hard task. Trying to convert day to day systems into the

IoT is a major consideration today. Researches are being conducted to achieve this macro goal, which helps to distribute the practice of the internet of things. As previously mentioned, analysis of the big data is one of the major challenges on the way to achieve this goal. MongoDB stores data and ways for devices to behave in the form of three types of patterns which are accessed by the storm for real time processing. It is clear that, the IoT is expanding over various industries. For instance, it is becoming a part of the healthcare sector today. Therefore, a lot of studies are being conducted over the IoT integration with e-health. A literature survey with main challenges (i.e. data management, privacy, data mining, security, chaos of data) on this topic is done by [5]. The IoT can also be adapted to the manufacturing market, which can be a great advantage as described in [6]. It is a competitive advantage. Authors entered this goal by considering sensor system and mobile devices. As they discuss, embedded sensors in machines connect to cloud solutions. The predictive analysis can create an advantage. The study introduces a monitoring tool which is formed in the wireless sensor network. The monitoring tool consists of data acquisition device. It provides status of the machine, power consumption and failures. These devices are organized in star topology. Data which come from these devices are collected by a central gateway, which coordinates data transmission. The data packetizing process is also done by it. Then, the data are uploaded to the cloud for further processing. The proposed architecture supports a human operator to upload information which are aggregated in the cloud.

A novel architecture for this imagined technology, the IoT and data analysis of it. As a whole, this scenario can be classified into to two major layers. First one is the bottom layer, which includes the IoT system. The other one is the top layer which consists of the cloud. architecture directly as shown in the figure 2. It causes to reduce the complexity and to increase the security. The agent layer consists of different kinds of agents

The user can configure this scenario according to his/her application. For example, the user may have a re-computational batch layer, which employs full re-computation on the master data set once a month. Then, there may be an incremental batch layer which operates once in every four hours. It only executes on the data which are not represented in the re-computational batch layer. Then the speed layer compensates data which are not represented in both batch layers. As a result of that, requirements for the layers are reduced. Another benefit is the gaining of both incremental and re-computation advantages. The bottom layer consists of different kinds of components. As the IoT objects in which objects are connected via communication protocols such as WiFi, Bluetooth and Zigbee are lying in the lower level. Most of these every-day objects do not include an internet connection ability. Objects which belong to a particular environment are connected to a particular environmental access point. They manage their objects while providing routing function. It uses less resources since it only executes new data rather than all the master data. On the other hand, the recovery from mistakes is ensured by the re-computational batch layer. In our system, data coming from agents are distributed using Hadoop distributed file system (HDFS) in the batch layer cluster. MapReduce is used for batch computation and MongoDB is used to store immutable master dataset. Thus, MapReduce executes the master dataset which is stored in the MongoDB in terms of map and reduce functions in a distributed and robust manner. As the batch layer cluster is connected to the serving layer. The serving layer uses as the database in our system. The serving layer is distributed in a cluster, and it has the ability to respond to ad-hoc queries in a low latency manner with regard to the indexing mechanism.

### III.  PROPOSED SYSTEM

Batch layer agent is responsible for forwarding data coming from all environmental access points to the batch layer (for master dataset).

Speed layer agent is responsible for forwarding data coming from all environmental access points to the speed layer.

Security agent

We have explored the feasibility of bringing multi-antenna benefits to single RF-chain wireless devices. Our findings are synthesized in a practical cross-layer design,that uses antenna index to carry extra bits and adaptive antenna hopping to ensure robustness/efficiency of communication. Decoding components are simple and built from existing WiFi/ZigBee modules. By integrating antenna hopping with the inherent modulation structures of such practical wireless systems, It is able to achieve multiple folds of capacity gain – even higher than existing theoretical prediction . Thus, it represents a viable and effective means of realizing multi-antenna networking between energy-constrained wireless devices.
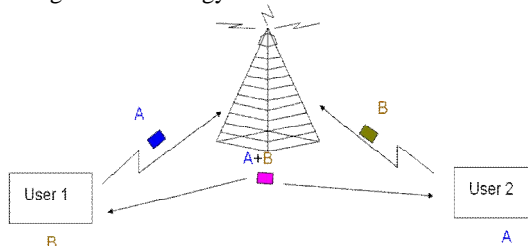


**Fig. 1. IoT system and cloud layer**

Ensures the safety of data. Employs the encryption and decryption and other security algorithms when necessary. Backup agent – Responsible for keeping backups and providing data when the master dataset fails. The user can configure this scenario according to his/her application. For example, the user may have a re computational batch layer, which employs full re computation on the master data set once a month. Then, there may be an incremental batch layer which operates once in every four hours. It only executes on the data which are not represented in the re-computational batch layer. Then the speed layer compensates data which are not represented in both batch layers. As a result of that, requirements for the layers are reduced. Another benefit is the gaining of both incremental and re-computation advantages. The bottom layer consists of different kinds of components. As the IoT objects in which objects are connected via communication protocols such as WiFi, Bluetooth and Zigbee are lying in the lower level. Most of these every-day objects do not include an internet connection ability. Objects which belong to a particular environment are connected to a particular environmental access point. They manage their objects while providing routing function. Its purpose is the final presentation of data. Finally, it has the business layer, which forms meaningful services by the data from the application layer. It is obvious that IoT sticks with a massive amount of data, which can be called big data. Big data refers to a large amount of structured or unstructured data that cannot be processed by a traditional application software.
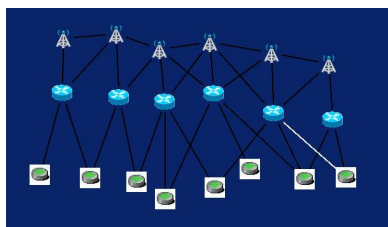


**Fig 2. Data packet flow**

The number of agents and their features can be adjustable according to the application.

*A. Batch Layer Agent*

Responsible for forwarding data coming from all environmental access points to the batch layer (for master dataset).

*B. Speed Layer Agent*

Responsible for forwarding data coming from all environmental access points to the speed layer.

*C. Security Agent*

Ensures the safety of data. Employs the encryption and decryption and other security algorithms when necessary.

*D. Backup Agent*

Responsible for keeping backups and providing data when the master dataset fails.

## IV. PERFORMANCE

*A. Latency*

Indexing mechanism reduces the latency in serving layer and speed layer. Speed layer is updated with low latency since it supports incremental algorithm.

*B. Energy And Resource Consumption*

Minimizes the power consumption of day to day objects in lower level in IoT model. Incremental sub layer of the batch layer uses less resources.

*C. Scalability*

Serving layer and speed layer are in distributed manner. Therefore, they achieve the scalability.

*D. Reliability*

Serving layer and speed layer are in distributed manner, which helps to replicate the data. Therefore, the fault tolerance can be achieved. On the other hand, recovery from mistakes are ensured by re-computational batch layer.

## V. CONCLUSION

We find approaches of big data analysis in the IoT applications. When the IoT applications increase by interconnecting devices through the whole world, the complexity, cost and power consumption are increased. We propose a new way to build the data flow between IoT and big data analysis system, which supports to reduce them. We discussed not only the connection between IoT and cloud but also the analysis mechanism which uses the System architecture that consists of popular technologies (i.e. Apache MongoDB, ElephantDB). Scalability, low latency and fault tolerance are some of the major factors that are considerable with respect to this scenario in the proposed architecture. They cause to increase performance of the big data analysis in IoT applications.

## VI. FUTURE SCOPE

Framework that can analyze the big data in the internet of things in an efficient way. hence to enhance the security model.

## REFERENCES

[1] T.T. Mulani and S.V.Pingle (March 2016). "Internet of things." International research journal of multidisciplinary studies & sppp's [online], Vol. 2, Special Issue 1, ISSN: 2454-8499.

[2] S. Chandrakanth, K.Venkatesh, J.U. Mahesh, K.V.Naganjaneyulu, "Internet of Things," International Journal of Innovations & Advancement in Computer Science, Vol. 3, Issue 8, ISSN 2347 – 8616, October 2014.

[3] D. Evans, "Internet of things: How the next evolution of the internet is changing everything," Cisco internet business solutions group, 2011.

[4] M. Villari, A. Celesti, M. Fazio, A. Puliafito, "AllJoyn Lambda: an Architecture for the Management of Smart Environments in IoT," International conference on IEEE, pp 9-14, November 2014.

[5] C. Ifrim, A.M. Pintilie, E. Apostol, C. Dobre, F. Pop (2017), "The art of advanced healthcare applications in big data and IoT systems," In advances in mobile cloud computing and big data in the 5G era[online], C.X. Mavromoustakis et al. (eds.), Springer International Publishing Switzerland, pp 133-149, 2017. Available: https://cs.pub.ro

[6] D. Mourtzis, E. Vlachou, N. Milas (2016). "Industrial big data as a result of IoT adoption in manufacturing," 5th CIRP Global web conference research and innovation for future production[online], Vol. 55, pp 290-295, Available: http://www.sciencedirect.com

[7] Y. Sun, H. Song, A.J. Jara, R. Bie, "Internet of things and big data analytics for smart and connected communities," IEEE access, vol. 14, August 2015.

[8] P. Goel, D. Grag, "The internet of things: A main source of big data analytics," Computer engineering and intelligent systems, vol. 8, ISSN 2222-1719, pp 12-16, 2017

[9] H. Cai, B. Xu, L. Jiang, A. V. Vasilakos, "IoT- Based big data storage systems in cloud computing: perspectives and challenges," IEEE internet of things journal, vol. 4, pp 75-87, February 2017.

[10] A. Bera, A. Kundu, N.R.D. Sarkar, D. Mou, "Experimental analysis on big data in IoT-based architecture," Proceedings of the international conference on data engineering and communication technology, Springer Singapore, pp 1-9, 2017.

[11] Y. Simmhan, S. Perera, "Big data analytics platforms for real-time applications in IoT," Big data analytics, Springer India, pp 115-135,2016.

[12] X. Liu, N. Iftikhar, X. Xie, "Survey of real-time processing systems for big data," Proceedings of the 18th International database engineering & applications symposium, ACM, pp 356-361, July 2014.

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)