Data Modernization - Blending IOT and Cloud Big Data Analytics

Sreenivasa Rao Veeranki

Senior Technical Architect, Genpact India Pvt Ltd &

Department of Computer Science and Engineering, School of Engg. & Tech., Maharshi university of Information Technology, Lucknow, INDIA

Abstract

The Internet is constantly evolving and changing. The Internet of Things (IoT) can be seen as the future of Internet applications that include machine learning (M2M). Actionable intelligence can be derived by combining big data and real-time analytics with the Internet of Things. Big Data and IoT can be seen as two sides of the same coin. With the connection between Big Data and objects on the Internet, the benefits of IoT can easily be reaped. IoT applications are proliferating in various fields such as healthcare, retail management and disaster management. Despite the benefits of synthesizing big data analytics and the Internet of Things, there are inherent complexities and challenges that need to be addressed. This article focuses on the combination of big data analytics and the Internet of Things.

Keywords: Internet of Things (IoT), Big Data, Real time analytics, Machine to machine learning (M2M), Data intensive,Cloud .

INTRODUCTION:

Internet applications are now closely associated to human life. Ranging from e-mails to e-learning, many applications are in vogue which has simplified otherwise complex processes. Technology is being refined to integrate itself seamlessly into the routines of human world. The digital space has provided the platform to implement various applications which can facilitate andease the complicated processes in various domains.

This has resulted in deployment of various types of sensors, computing devices in huge number. Devices which are interconnected share and create data being termed as "Internet of Things (IoT)" [2]. Instead of most data on the Internet being produced and consumed by people (text, audio, video), more and more information would be produced and consumed by machines, communicating between themselves to improve the quality of our

live. It has been seen that IoT has augmented people in various fields like healthcare, natural disaster management [5]. The integration of IoT into the health care system could prove to be incredibly beneficial for both an individual and a society [7]. A chip could be implemented into each individual, allowing for hospitals to monitor the vital signs of the patient. By tracking their vital signs, it could help indicate whether serious assessment is necessary.

IoT can also function as a tool that can conserve energy within households. Home appliances communicating and operating would lead to efficient energy usage. IOT allows for the communication between devices, commonly referred to as Machine-to-Machine (M2M) communication. With this being possible, physical devices would be able to communicate to people letting them know their condition and where it is located. Devices such as trucks or ships allow for the maximum capacity to be filled by communication amongst devices and then relaying that information to a person to capitalize on the data supplied. Another advantage of IoT is the ability to track individual consumers and targeting these consumers based on the information supplied by the devices. In a way, it provides a more "personalized" system that could potentially increase business sales and increases their demographic. Utilities due to optimal utilization of IoT are many. The various advantages of IoT has been visualized in fig 1. IoT comprises of not only the computing devices but also humans who can sense, communicate, and compute. Thus, along with the advantages IoT comes with it inherent complexities and challenges. The major concerns related with IoT are complexity of the system, space, size, security and privacy. Due to huge number of interconnections, there is a great possibility of increase in complexity of the system. The Internet of objects would encode 50 to 100 trillion objects and would be able to follow the movement of those objects [9]. The size of IoT would be a major concern. Direct collection of sensitive personal information, such as precise geo location, financial account numbers, or health information may create privacy risk. The data intensive nature of IoT can be channeled with Big Data as a part of the solution to the challenges faced by IoT.

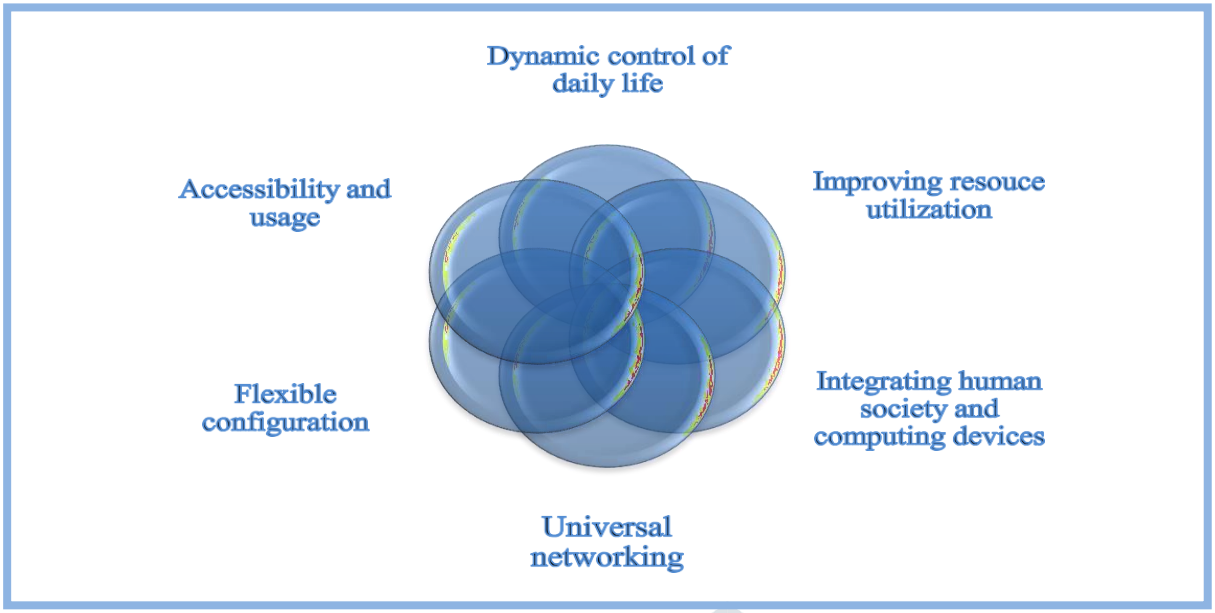


Fig 1 Utilities of IoT

The deployment of various devices is also contributing to the deluge of data or now more predominantly known as 'Big Data'. IoT are one of the major sources for Big Data. With the count of interconnected devices increasing the data associated with them is mounting to a humongous one.IoT intersects with Big Data and it is evident that the two trends would fit one another. IoT and Big Data are connected by the definition "billions of internet-connected 'things' ". The size of the digital universe and the number ofinterconnected 'things' is being amplified. This is not being complimented by the actionable data [9] . The Big Data analytics would provide a platform to enhance and obtain actionable data for the humongous data being collected. This paper deals with the relationship between IoT and Big Data and its significance.

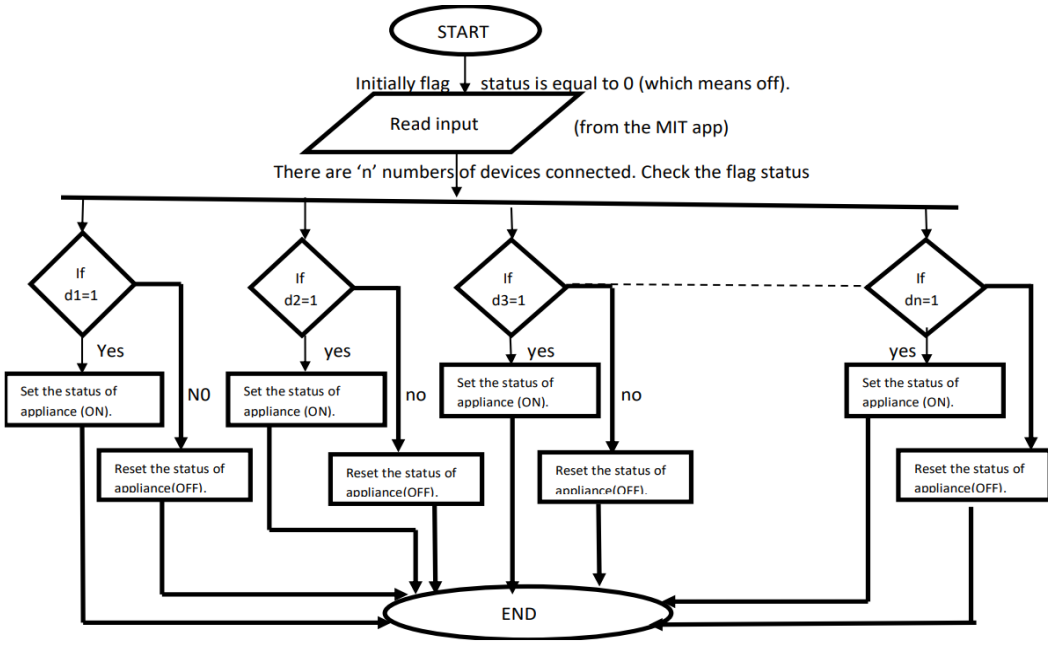
The Internet of Things is the most trending technology today, alongside wearables and robotics. The concept is simple: for example, Devices in your home (or wherever they are) have the capability to communicate with each other via the internet. This technology usually uses sensors to pass data to the internet. Imagine a sensor installed in your garden that uploads data like temperature, humidity, and soil purity to the internet, and this data is visible to you from anywhere in the world. Or imagine home automation systems you can use to control appliances in your home like lights, door locks, and air conditioning through a web interface or smartphone application. A lot of technologies are being developed around this concept, such as independent lightweight IoT networks and protocols for passing data. The example is to provide a smart control for home appliances using a Raspberry pi board with Wi-Fi being remotely controlled by any smart phone. As technology is advancing so houses are also getting smarter. Presently, Conventional wall switches located at different

parts of the house makes it difficult for the user to go near them to operate. Even more it becomes more difficult for the elderly or physically handicapped people to do so. Remote controlled systems provide a most modern solution with smart phones. In order to achieve this, a Wi-Fi module is interfaced to the Raspberry pi board at the receiver end while on the transmitter end, GUI application on the cell phone sends ON/OFF commands to the receiver where loads are connected. By touching the specified Button on the GUI, the loads can be turned ON/OFF remotely through this technology.

This paper is designed for IOT Cloud Big data Analytics. In this paper, I would like to demonstrate my solution of web application and Raspberry Pi code on Azure cloud with home automation system. Our home automation systems comprise in the form of individual control devices, 9 distributed control systems, and centrally controlled systems Individual control devices, the simplest of the home automation devices, are programmable devices that can be set to user preferences like thermostats, lighting controls, and occupancy sensors, etc. Distributed and centrally controlled systems, on the other hand, are home automation systems that can communicate and control the individual devices remotely. The difference between the distributed and central systems is that of the control whereby the former has no central controller and latter has a controller like a mobile or computer to control all the devices centrally. The downside of a centrally controlled system is the failure when the controller fails. Commercial home automation systems are expensive due to custom equipment, components, and installation cost. An alternative approach is to utilize open-source platforms and IoT sensors to do customized home automation to meet the user's own needs. There are two major problems with this approach: IP-based devices' cost and user's lack of expertise to operate the open source solutions for home automation. IP-based devices are expensive and home automation requires the installation of several such sensors that are out of reach of low-income houses. This paper presents an approach that can provide a low-cost home automation solution for such homes. The demonstration of my example is on ordinary home appliances and makes use of Relay Circuit Pack with Raspberry pi to control these appliances. The term "ordinary electrical appliances" refers to those appliances that are readily available in the market. Such devices are without any special IP connector or IoT circuit. These appliances cost much less than the special IP-based smart devices like Phillips hue bulb. The system allows the user to check and change the status of electronic home appliances and the working state of sensors. Appliances include common lighting, heating, ventilation, and air-conditioning. It provides the online status of various appliance that enables the user to control and set the appliance's usage accordingly to save the energy. The control is provided through an Android built smartphone application that is connected to the Azure cloud database. The local network provides control over the appliances when the internet is not available. An Azure cloud database is used for data logging that restores the appliances to a previous state if the system fails. The smartphone app lets the user to custom build the house on the smartphone analogous to the

actual house and place each appliance that needs to be controlled.

High level Algorithms Implementation Flow



Data Intensive Iot

Data is usually perceived in modular or transactional form, like a sale, data on a product. But IoT will be creating streams of data similar to social networking. The concept of IoT can be visualized in a smart warehouse where data is being stored regarding the opening of the door of warehouse like duration, temperature, time and date, frequency per hour, per day, per week so on. This is a continuous stream of data which is being captured by various sensors deployed. Similarly in a use case of smart home issues like roof damage, water and gas leakage, power consumption can be effectively handled with the help of the sensors and computing devices which would be streaming data at a very regular interval of time. These scenarios indicate huge growth of data in implementation of IoT. Understanding, analyzing the data to produce actionable data would be an essential exercise. Big Data and IoT infuse at this junction.

Iot And Big Data Analytics

Data though collected by the devices need to be filtered to make it relevant and useful. The redundancy in the data being collected is predominant due to the sheer nature ofthe framework of IoT. The data is continuous hence the extraction of valuable information is not simple. This requires a good mechanism of protocols and software to ensure that the data is secured and also significant. [17] Data is generally collected by the sensor devices in which these devices collect and transmit data to a centralized server. Similarly the data is distributed back to the devices also. These activities require performance efficiency of the network to be optimum. IoT involves a number of heterogeneous networks like wireless Sensor Networks (WSN), Wireless mesh networks, Wireless LAN. These networks would help in transmission of data and also involve various types of quality issues ranging from performance to energy efficiency.

Big Data and IoT are complimentary to each other and are two dimensions of a perception. Managing the data and extracting information from it is a very vital task associated with IoT. An appropriate analytical platform is required to enable to derive knowledge from IoT data. IoT devices generate continuous streams of data in a scalable way. It is essential to handle the high volume of stream data and exploit the data. In a normal scenario Big Data, the data might not be stream data, but the actions are. While in IoT data, it is continuous flow. Applying real time analytics is the need in IoT environment. The advantages of IoT can be seen only when real time analytics is applied on the data stored. Real time Big Data analytics and IoT equates to value creation which is depicted in fig 2.

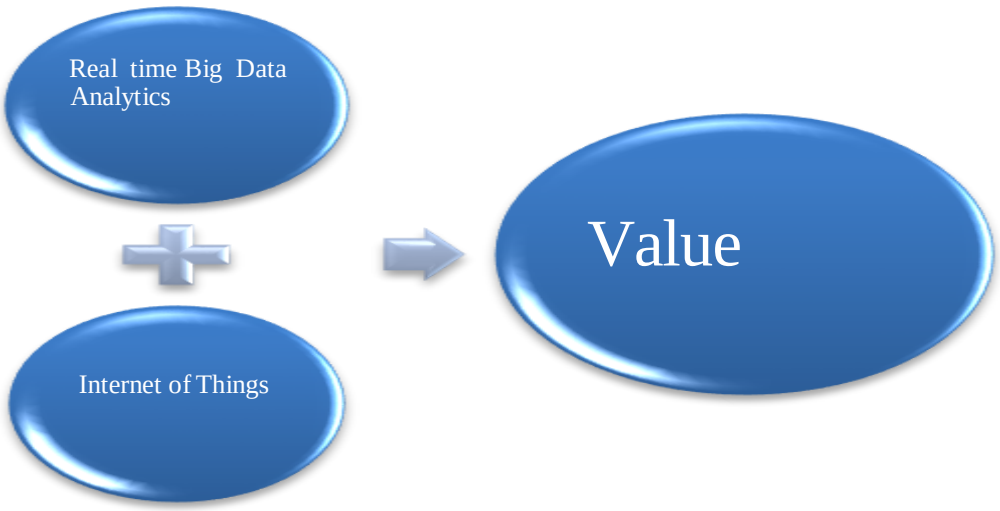


Fig2. Value Creation

Though real time analytics provides a greater avenue to generate actionable data implementing to the fullest is a huge task. Implementing real-time analytics in an IoT environment is challenging due to following reasons

- The large number of IOT devices and volume ofdata generated
- The need for processing and analyzing data at lowlatencies
- The need for specialized visualization andreporting
- Non-standardized stack techniques and solutions

Advanced analytics on IoT data would help in analyzing the data collected over a large period of time, and in turn gain a better insight into systems and their behavior. To create models to forecast future outcomes and to optimize the same [18]. Collect information to estimate factors that would not be directly measured by sensors, by determining the relationship between different system parameters, and their impact on each other.

TECHNOLOGY AND TECHNIQUES IN BIG DATA

Technology and techniques in big data are large, varied fast growing extracting relevant information. Technologies and analytical techniques employed, attacking a big data project being analyzing and researching. Several software technology products are available. Hadoop is key technology used to handle big data. Apache Hadoop is an open-source framework that deals with distributed computing of large data sets across clusters of computers using simple programming models. Its software library is a framework Hadoop is designed to scale up from single servers to thousands of machines. Major advantages that Hadoop offers are we can use inexpensive hardware. Hadoop distributed file system provides high-throughput access to application data stores large amounts of data.HBase is a scalable and distributed database supports structured data storage for large tables provides fortransactional kind of capabilities by allowing updates inserts deletions etc; HBase allows for random check. Pig a high-level data flow language and execution framework parallel computations. Apache Pig is a scripting language, Map reduce transformations including summarizing. Hive a data warehouse infrastructure that provides data summarization software tool used for managing and analyzing large datasets. Hive QL is SQL like language. SQL is traditional languages. Sqoop software tool designed to transfer bulk data. Sqoop is used to import data from external data bases into HDFS or HBASE. Zookeepers a high-performance co-ordination service for distributed applications. It is a centralized service

used for maintainingconfiguration information named registry. Avro is a data serializations system. Cassandra a scalable data base with no points of failure. Apache Cassandra is a high availability, capability with multiple servers. Tez a generalized data flow programming framework, Tez is being adopted by Hive, Pig, and other framework in the Hadoop eco-system. Apache spark is fast data analytics. Flume is a reliable distributed service for efficiently collecting aggregating and moving large amount of Big Data. Hadoop framework's base characteristics provide deep information on the various components perceptive.

Big Data & Cloud Computing and Analytics with Artificial Intelligence

As we can see, there are infinite possibilities when we combine Big Data and Cloud Computing! If we simply had Big Data alone, we would have huge data sets that have a huge amount of potential value just sitting there. Using our computers to analyze them would be either impossible or impractical due to the amount of time it would take.

However, Cloud Computing allows us to use state-of-the-art infrastructure and only pay for the time and power that we use! Cloud application development is also fueled by Big Data. Without Big Data, there would be far fewer cloud-based applications, since there wouldn't be any real necessity for them. Remember, Big Data is often collected by cloud-based applications, as well!

In short, Cloud Computing services largely exist because of Big Data. Likewise, the only reason that we collect Big Data is because we have services that are capable of taking it in and deciphering it, often in a matter of seconds. The two are a perfect match, since neither would exist without the other!

The Cloud Big data analytics is the use of processes and technologies, including AI and machine learning, to combine and analyze massive datasets with the goal of identifying patterns and developing actionable insights. This helps you make faster, better, data-driven decisions, data management, pattern management, context management, decision management, action management, goal management, and risk management. that can increase efficiency, revenue, and profits. AI can identify data types, find possible connections among datasets, and recognize knowledge using natural language processing. It can be used to automate and accelerate data preparation tasks, including the generation of data models, and assist in data exploration. It can learn common human error patterns, detecting and resolving potential flaws in information. And it can learn by watching how

the user interacts with an analytics program, surfacing unexpected insights from massive datasets fast. AI can also learn subtle differences in meaning, or context-specific nuances, in order to help users better understand numeric data sources. And it can alert users to anomalies or unexpected patterns in data, actively monitoring events and identifying potential threats from system logs or social networking data, for example.

Challenges

IoT can change the way of working of Internet but also has challenges which need to be addressed. The following are few key challenges:

- Naming and Identity Management: Identification of the huge number of devices connected is to be donein a unique and dynamic way.
- Interoperability and Standardization: Standardizationof the devices to have interoperability among the devices is major requirement.
- Information Privacy: The data being captured can bevital and privacy of it has to be considered.
- Objects safety and security: The devices security andsafety is a concern as there can be physical damage done to distributed devices.
- Data confidentiality and encryption: Data being transmitted needs to be encrypted to ensure the data is not misused.
- Spectrum: The spectrum on which the devices would transmit information needs to be efficient one.
- Green IoT: The energy consumption by the devices would be large if efforts to minimize the consumption is not done.

CONCLUSION

This paper focuses on the fusion of IoT and Big Data and the role of real time analytics in IoT. IoT is a emerging technology which can drive a new wave of application of analytics into the regular routines of humans. The scalability of IoT would lead to smarter applications in various domain ranging from health care to smart and secure homes. The actionable intelligence obtained by application of real time analytics on the data or "Big Data" of IoT is one of the main benefits of IoT. To tap in the advantages of IoT, Big Data analytics is needed.

Finally, it's important to note that both Big Data and Cloud Computing play a huge role in our digital society. The two linked together allow people with great ideas. They also allow established businesses to utilize data that they collect but previously had no way of analyzing.

More modern components of cloud infrastructure's typical "Software as a Service" model such as artificial intelligence also enable businesses to get insights based on the Big Data they've collected. With a well-planned system, businesses can take advantage of all of this for a nominal fee, leaving competitors who refuse to use these new technologies in the dust.

REFERENCES

[1] J. Zheng, D. Simplot-Ryl, C. Bisdikian, and H. Mouftah, "The Internet of Things," in IEEE Communications Magazine, Volume:49 , Issue: 11, pp:30-31, 2011.

[2] Y. Huang and G. Li, "Descriptive Models for Internet of Things," in IEEE International Conference on Intelligent Control and Information Processing (ICICIP), August 2010

[3] T. Fan and Y. Chen, "A Scheme of Data Management in the Internet of Things," in 2nd IEEE International Conference on Network Infrastructure and Digital Content, Sept. 2010.

[4] Y. Huang and G. Li, "A Semantic Analysis for Internet of Things," in International Conference on Intelligent Computation Technology and Automation (ICICTA), May 2010.

[5] Q. Zhou and J. Zhang, "Research Prospect of Internet of Things Geography," in 19th International Conference on Geoinformatics, June 2011.

[6] J. Li, Z. Huang, and X. Wang, "Countermeasure Research about Developing Internet of Things Economy," in International Conference on E -Business and E -Government (ICEE), May 2011.

[7] Y. Yu, J. Wang, and G. Zhou, "The Exploration in the Education of Professionals in Applied Internet of Things Engineering," in 4th International Conference on Distance Learning and Education (ICDLE), October 2010.

[8] L. Coetzee and J. Eksteen, "The Internet of Things: Promise for the Future? An Introduction," in IST-Africa Conference Proceedings, CSIR, Pretoria, South Africa, May 2011.

[9] L. Tan and N. Wang, "Future Internet: The Internet of Things," in 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), August 2010.

[10] G. Gang, L. Zeyong, and J. Jun, "Internet of Things Security Analysis," in International Conference on Internet Technology and Applications (iTAP), August 2011.

[11] M. Wu, T. Lu, F. Ling, J. Sun, and H. Du, "Research on the Architecture of Internet of Things," in 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), Aug. 2010.

[12] Z. Hu, "The research of several key question of Internet of Things," in International Conference on Intelligence Science and Information Engineering (ISIE), August 2011.

[13] http://bigthink.com/think-tank/the-internet-of-things-meets-big- data

[14] http://data-informed.com/the-impact-of-internet-of-things-on-big- data/

[15] TECHNOLOGY ANALYSIS FOR INTERNET OF THINGS USING BIG DATA LEARNING, Sunghae Jun International Journal of Research in Engineering and Technology Vol 3,Issue 12,2014

[16] Internet of Things in Industries: A Survey, Li Da Xu, Senior Member, IEEE, Wu He, and Shancang Li, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, VOL. 10, NO. 4, NOVEMBER 2014

[17] Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions, Jayavardhana Gubbi,a Rajkumar Buyya,b, Slaven Marusic,aMarimuthu Palaniswami, Future Generation Computer Systems 29 (2013) 1645–1660,Elsevier