© 2019 IJSRSET | Volume 6 | Issue 2 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology
DOI : 10.32628/IJSRSET196315

IoT Big Data Analytics – Architecture Design and Use Cases

Devinderpal Yadav
MS in Software Engineering, BITS Pilani, India

ABSTRACT

As sensors are adopted in almost all the fields of life, the Internet of Things (IoT) is triggering a massive influx of data. It is crucial to have efficient and scalable methods to process this data to gain valuable insight and take timely actions. The future vision of internet is to connect everything, such as connecting things like transportation networks, communication networks, etc. All these data which will be generated by all this connected millions of devices will not be useful without analytic power. Numerous big data, IoT, and analytics solutions have enabled people to obtain valuable insight into large data generated by IoT devices. However, these solutions are still in their infancy, and the domain lacks a comprehensive survey. This paper investigates the state-of-the-art research efforts directed toward big IoT data analytics. The relationship between big data analytics and IoT is explained. Moreover, this paper adds value by proposing a new architecture for big IoT data analytics. This paper will discuss about the possible architecture of the IoT Big data analytics and also will discuss some industry use-cases.

Keywords: Internet of Things (IoT), Big Data Analytics, Distributed Computing, Architecture

I. INTRODUCTION

As we move towards future, all the things are being automated and connected. The explosion of embedded and connected smart devices, systems, and technologies in our lives has created an opportunity to connect every 'thing' to the internet. From smartphone to the self-driving car, the journey has been quite futuristic. What's next? Are we heading towards future or this is it? The growth of data produced via IoT has played a major role on the big data landscape. Big data can be categorized according to three aspects: (a) volume, (b) variety, and (c) velocity. These categories were first introduced by Gartner to describe the elements of big data challenges. Immense opportunities are presented by the capability to analyse and utilize huge amounts of IoT data, including applications in smart cities, smart transport and grid systems, energy smart meters, and remote patient healthcare monitoring devices.

According to the Cisco report, the number of objects connected to the Internet has exceeded the number of human beings in the world. These Internet-connected objects, which include PCs, smartphones, tablets, WiFi- enabled sensors, wearable devices, and household appliances, form the IoT as shown in Fig. 1. Reports show that the number of Internet-connected devices is expected to increase twofold from 22.9 billion in 2016 to 50 billion by 2020 as shown in Fig. 2.

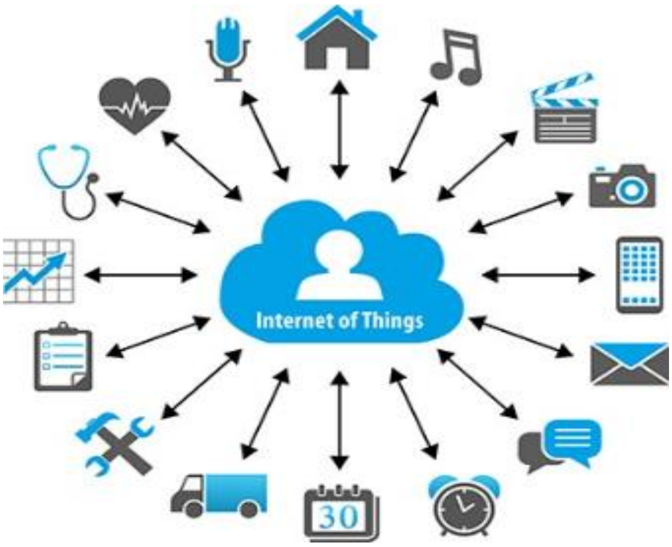

Figure 1: Big Data Sources in IoT

Most IoT applications do not only focus on monitoring discrete events but also on mining the information collected by IoT objects. Most data collection tools in the IoT environment are sensor-fitted devices that require custom protocols, such as message queue telemetry transport (MQTT) and data distribution service (DDS). Given that sensors are used in nearly all industries, the IoT is ex- pected to produce a huge amount of data. The data generated from IoT devices can be used in finding ppotential research trends and investigating the impact of certain events or decisions.

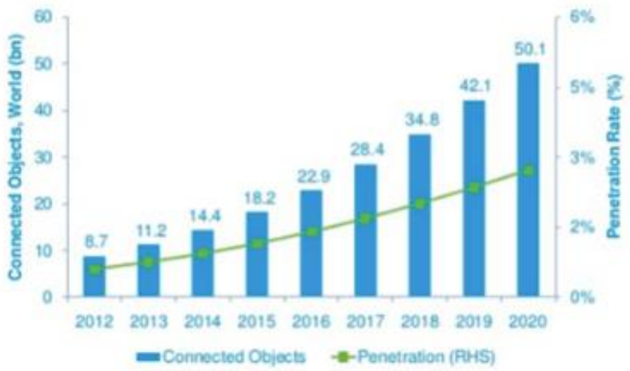

Figure 2: Number of connected devices

These data are processed using various analytic tools. Fig. 3 illustrates the process of data collection, monitoring, and data analytics.

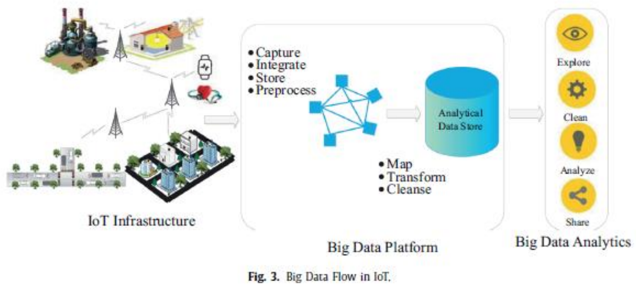

Figure 3: Data flow in IoT

Although IoT has created a huge number of opportunities, collecting large amount of data alone is not sufficient. To gain benefits from the Big Data and IoT platform enterprise must collect, manage and analyse a massive volume of sensor data in a scalable and cost effective way. Data integration and analytics allow organizations to revolutionize their business process. Specifically, these enterprises can use data analytics tools to transform a huge volume of sensor-collected data into valuable insights

Big data analytics is a rapidly expanding research area spanning the fields of computer science, information management, and has become a ubiquitous term in understanding and solving complex problems in different disciplinary fields such as engineering, applied mathematics, medicine, computational biology, healthcare, social networks, finance, business, government, education, transportation and telecommunications. The utility of big data is found largely in the area of Internet of Things (IoT). Big data is used to build IoT architectures which include things-centric, data-centric, service-centric architecture, cloud-based IoT. Technologies enabling IoT include sensors, radio frequency identification, low power and energy harvesting, sensor networks and IoT services mainly include semantic service management, security and privacy-preserving protocols, design examples of smart services. To effectively synthesize big data and communicate among devices using IoT, machine learning techniques are employed. Machine learning extracts meaning from big data using various techniques which include regression analysis, clustering, bayesian methods, decision trees and random forests, support vector machines, reinforcement learning, ensemble learning and deep learning.

II. OVERVIEW OF IoT

The IoT is significant because an object that can represent itself digitally becomes something greater than the object by itself. No longer does the object just relate to the process; it now connects to surrounding objects and database data, permitting "big data" analytics and insights. In particular, "things" might communicate autonomously with other things and other devices, such as sensors in manufacturing environments or an activity tracker with a smartphone. IoT has evolved from the convergence of wireless technologies, micro-electromechanical systems, microservices and the internet. This convergence has torn down the walls between operational technology and information technology, allowing unstructured machine-generated data to be analysed for insights that will drive improvements. Consumer IoT took another revolutionary path, either by becoming connected – for example, speed sensors on a bike – or being newly invented. In other instances, such as in healthcare, things have been there but not widely used, such as patient health status.

IoT – Key Features:

The most important features of IoT include artificial intelligence, connectivity, sensors, active engagement, and small device use. A brief review of these features is given below –

1. **AI:** IoT essentially makes virtually anything "smart", meaning it enhances every aspect of life with the power of data collection, artificial intelligence algorithms, and networks. This can mean something as simple as enhancing your refrigerator and cabinets to detect when milk and your favorite cereal run low, and to then place an order with your preferred grocer.
2. **Connectivity:** New enabling technologies for networking, and specifically IoT networking, mean networks are no longer exclusively tied to major providers. Networks can exist on a much smaller and cheaper scale while still being practical. IoT creates these small networks between its system devices
3. **Sensors:** IoT loses its distinction without sensors. They act as defining instruments which transform IoT from a standard passive network of devices into an active system capable of real-world integration.
4. **Active Engagement:** Much of today's interaction with connected technology happens through passive engagement. IoT introduces a new paradigm for active content, product, or service engagement.
5. **Small Devices:** Devices, as predicted, have become smaller, cheaper, and more powerful over time. IoT exploits purpose-built small devices to deliver its precision, scalability, and versatility.

The IoT system architecture with various devices in the network is shown in the fig. 4.

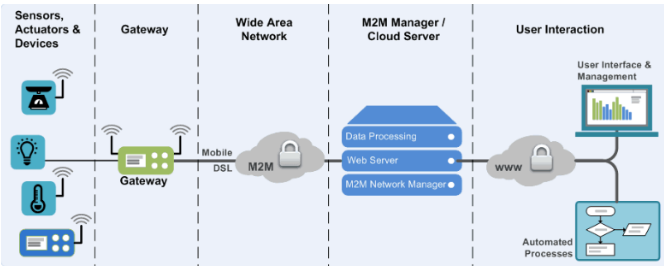

Figure 4 : IoT System Architecture

III. OVERVIEW OF BIG DATA

As one of the most "hyped" terms in the market today, there is no consensus as to how to define big data. The term is often used synonymously with related concept such as Business Intelligence (BI) and data mining. It is true that all three terms is about analyzing data and in many cases advanced analytics. But big data concept is different from the two others when data volumes, number of transactions and the number of data sources are so big and complex that they require special methods and technologies in order to draw insight out of data (for instance, traditional data warehouse solutions may fall short when dealing with big data).

This also forms the basis for the most used definition of big data, the three V: Volume, Velocity and Variety as shown in Fig. 5.

- **Volume:** Large amounts of data, from datasets with sizes of terabytes to zettabyte.
- **Velocity:** Large amounts of data from transactions with high refresh rate resulting in data streams coming at great speed and the time to act on the basis of these data streams will often be very short. There is a shift from batch processing to real time streaming.
- **Variety:** Data come from different data sources. For the first, data can come from both internal and external data source. More importantly, data can come in various format such as transaction and log data from various applications, structured data as database table , semi-structured data such as XML data, unstructured data such as text, images, video streams, audio statement, and more. There is a shift from sole structured data to increasingly more unstructured data or the combination of the two.

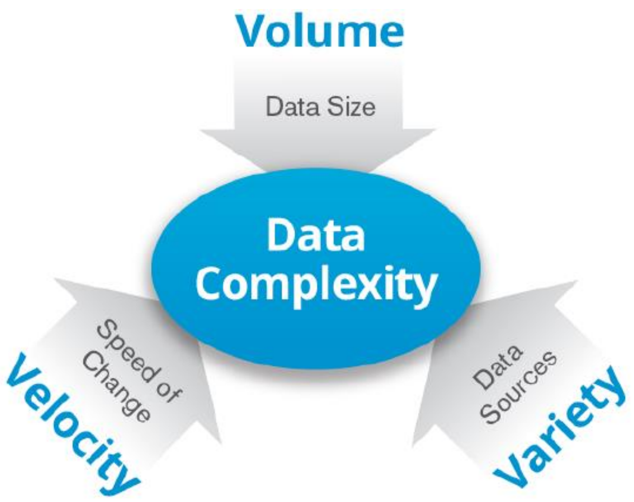

Figure 5: The three V of Big Data

Some of the key features of Big Data Analytics are as below:

1. **Data Processing:** Data processing features involve the collection and organization of raw data to produce meaning. Data modeling takes complex data sets and displays them in a visual diagram or chart. This makes it digestible and easy to interpret for users trying to utilize that data to make decisions. Data mining allows users to extract and analyze data from different perspectives and summarize it into actionable insights. It is especially useful on large unstructured data sets collected over a period of time.
2. **Predictive Applications:** Referred to as the "final frontier of analytic capabilities," prescriptive analytics entails the application of mathematical and computational sciences and suggests decision options to take advantage of the results of descriptive and predictive analytics. The first stage of business analytics is descriptive analytics, which still accounts for the majority of all business analytics today.[4] Descriptive analytics looks at past performance and understands that performance by mining historical data to look for the reasons behind past success or failure.
3. **Scalability:** With the exponential increases in the volume of data being produced and processed, many companies' databases are being overwhelmed with the deluge of data they are facing. To manage, store and process this overflow of data, a technique called "data scaling" has become necessary for many organizations dealing with exploding datasets. A scalable data platform accommodates rapid changes in the growth of data, either in traffic or volume. These platforms utilize added hardware or software to increase output and storage of data. When a company has a scalable data platform, it also is prepared for the potential of growth in its data needs.

4. **Simple Integration:** Another interesting feature to take into account. Simple integrations make it way easier to share results with other developers and data scientists. As a result, data analytics tools must support easy integration with existing enterprise and cloud applications and data warehouses.
5. **Security:** This has to be one priorities for companies investing in big data analytics tools, especially those based on the cloud. Consider tools that comprise of an extensive bunch of security provisions. For this, you can even seek help from a big data service provider.

IV. IoT BIG DATA ARCHITECTURE

The IoT Big Data Analytics architecture is positioned with a three-tier architecture pattern comprising edge, platform and enterprise tiers. Fig.6 gives a high level diagram and fig.7 gives a detailed architecture with relationships between different layers of communications.

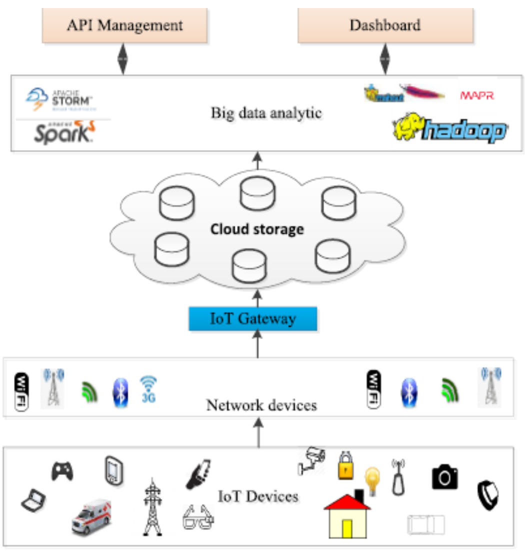

Figure 6: High Level Architecture of IoT and Big-Data Analytics

The edge-tier includes Proximity Networks and Public Networks where data is collected from devices and transmitted to devices. Data flows through the IoT gateway or optionally directly from/to the device then through edge services into the cloud provider via IoT transformation and connectivity.

The Platform tier is the provider cloud, which receives, processes and analyzes data flows from the edge tier and provides API Management and Visualization. It provides the capability to initiate control commands from the enterprise network to the public network as well.

The Enterprise tier is represented by the Enterprise Network comprised of Enterprise Data, Enterprise User Directory, and Enterprise Applications. The data flow to and from the enterprise network takes place via a Transformation and Connectivity component. The data collected from structured and non-structured data sources.

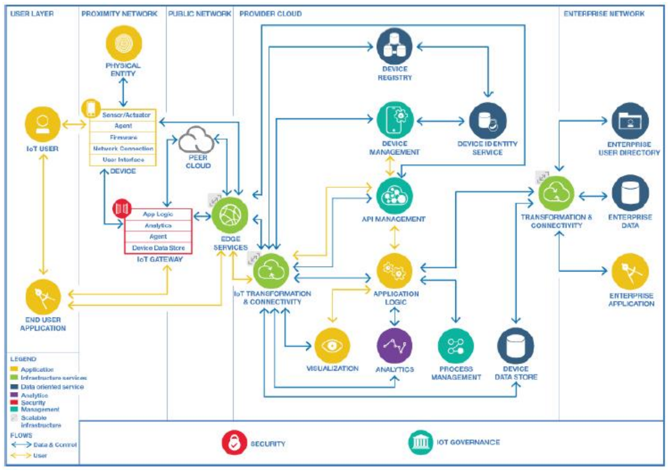

Figure 7: Detailed Architecture for IoT Big Data Analytics

The various components of the detailed architecture are described below:

1) User Layer: User layer contains IoT users and their end user applications. IoT User is person or alternatively an automated system that makes use of one or more end user applications to achieve some goal. The IoT User is one of the main beneficiaries of the IoT solution. End User Application is domain specific or device specific application. The IoT user may use end user applications that run on smart phones, tablets, PCs or alternatively on specialized IoT devices including control panels.

2) Proximity Network: It contains the physical entities that are at the heart of the IoT system, along with the devices that interact with the physical entities and connect them to the IoT system. Physical Entity is the real-world object that is of interest – it is subject to sensor measurements or to actuator behavior. It is the "thing" in the Internet of Things. This architecture distinguishes between the physical entities and the IT devices that sense them or act on them. For example, the thing can be the ocean and the device observing is it a water temperature thermometer. Another example is a depot shipping parcels: the parcels are the physical entities and there are devices with sensors capable of observing and identifying each parcel – e.g. via RFID tags or via Barcode readers. It is clear that the RFID Tag reader is one thing and the parcels are something completely different - the identity of the parcel is the physical entity here. Device contains sensor(s) and/or actuator(s) plus a network connection that enables interaction with the wider IoT system. There are cases where the device is also the physical entity being monitored by the sensors – such as an accelerometer inside a smart phone. Main components of the devices are:

a). ***Sensor/Actuator*** - senses and acts on physical entities. A sensor is a component that senses or measures certain characteristics of the real world and converts them into a digital representation. An actuator is a component that accepts a digital command to act on a physical entity in some way.

b). Agent - provides remote management capabilities for the device, supporting a device management protocol that can be used by the Device Management service or IoT management system.

c). ***Firmware*** - software that provides control, monitoring and data manipulation of engineered products and systems. The firmware contained in devices such as consumer electronics provides the low-level control program for the devices.

d). ***Network Connection*** - provides the connection from the device to the IoT system. This is often a local network that connects the device with an IoT gateway – low power and low range in many cases to reduce the power demands on the device. However, there are cases where the network connection is direct to the public network and no IoT gateway is required.

3) IoT Gateway: It acts as a means for connecting one or more devices to the public network (typically the Internet). It is commonly the case that devices have limited network connectivity – they may not be able to connect directly to the Internet. The IoT Gateway contains the following components:

App Logic - provides domain specific or IoT solution specific logic that runs on the IoT Gateway. For IoT systems that have Actuators which act on physical entities, a significant capability of the app logic is the provision of control logic which makes decisions on how the actuators should operate, given input from sensors and data of other kinds, either held locally or held centrally.

Analytics - provides Analytics capability locally rather than in the provider cloud.

Agent - allows management of the IoT Gateway itself and can also enable management of the attached devices by providing a connection to the provider cloud layer's Device Management.

Device Data Store - stores data locally. Devices may generate a large amount of data in real time it may need to be stored locally rather than being transmitted to a central location. Data in the device data store can be used by the application logic and analytics capability in the IoT Gateway.

4) Public Network: contains the wide area networks (typically the internet), peer cloud systems, the edge services.

Peer Cloud - a 3rd party cloud system that provides services to bring data and capabilities to the IoT platform. Peer clouds for IoT may contribute to the data in the IoT system and may also provide some of the capabilities defined in this IoT architecture.

Edge Services - services needed to allow data to flow safely from the internet into the provider cloud and into the enterprise.

5) Provider Cloud: This provides core IoT applications and associated services including storage of device data; analytics; process management for the IoT system; create visualizations of data. Also hosts components for device management including a device registry. Provider Cloud elements include:

• IoT Transformation and Connectivity
• Application Logic
• Visualization
• Analytics
• Process Management
• Device Data Store
• API Management
• Device Management
• Device Registry
• Device Identity Service
• Transformation and Connectivity

6) Enterprise Network: Enterprise Network host a number of business specific enterprise applications that deliver critical business solutions along with supporting elements including enterprise data. Typically, enterprise applications have sources of data that are extracted and integrated with services provided by the cloud provider. Analysis is performed in the cloud computing environment, with output consumed by the enterprise applications.

Enterprise Data: Enterprise Data includes metadata about the data as well as systems of record for enterprise applications. Enterprise data may flow directly to data integration or the data repositories providing a feedback loop in the analytical system for IoT

Enterprise User Dictionary - stores user information to support authentication, authorization, or profile data. The security services and edge services use this to control access to the enterprise network, enterprise services, or enterprise specific cloud provider services.

Enterprise Applications - Enterprise applications consume cloud provider data and analytics to produce results that address business goals and objectives.

V. IoT BIG DATA ANALYTICS – USE CASES

Big data technologies can offer data storage and processing services in an IoT environment, while data analytics allow business people to make better decisions. IoT applications are the major sources of big data. This section explains the role of big data and analytics in different IoT applications, including smart grids, smart healthcare, smart transportation, and smart inventory systems and many others.

1. ***Smart Transportation:*** Finding valuable information has become a key concern in this modern age of technologies where vehicles are connected to the Internet and generate large amounts of data. Data analytics can help transport management authorities to find out the history of road mishaps (e.g., under what circumstances did the accident oc- cur and at what speed were the drivers driving during the mishap), minimize the number of road accidents, determine the time when the traffic load reaches its peak, and prepare an optimal route plan that can help minimize traffic congestion. The analytics of smart transport data can indirectly optimize shipment movements, improve road safety, and enhance end-to- end user experience in terms of delivery time.
2. ***Smart healthcare***: Over the past few years, voluminous amounts of data have been created in the healthcare sector. However, such rapid increase in data production has created challenges in extracting valuable in- formation from big healthcare datasets that can help predict epi- demics and find cures for various diseases. Data analytics can help healthcare specialists analyze a large amount of patient data and learn the history of a disease (in the case of family doctors). Insurance companies may also use data analytics

when making policies. Healthcare specialists may also detect serious illnesses at their early stages and subsequently prevent the loss of life.

3. ***Smart grid:*** Smart grids rapidly generate data, and finding useful information from these data has become imperative. In a smart grid environment, large amounts of data are collected from various sources, such as the power utilization habits of users, pharos measurement data for situational awareness, and energy consumption data measured by widespread smart meters, to name a few. Proper analytics can help decision maker's measure the appropriate level of electricity supply that they must provide to their customers. Analytics may also help business people predict electricity demands in the near future. The strategic objectives of specific organizations can also be met through proper analytics (e.g., pricing plans that are consistent with supply, demand, and production models).
4. ***Smart Metering:*** Smart metering is one of the IoT application use cases that generates a large amount of data from different sources, such as smart grids, tank levels, and water flows, and silos stock calculation, in which processing takes a long time even on a dedicated and powerful machine. A smart meter is a device that electronically records consumption of electric energy data between the meter and the control system. Collecting and analyzing smart meter data in IoT environment assist the decision maker in predicting electricity consumption. Furthermore, the analytics of a smart meter can also be used to forecast demands to prevent crises and satisfy strategic objectives through specific pricing plans. Thus, utility companies must be capable of high-volume data management and advanced analytics designed to transform data into actionable insights.
5. ***Smart Agriculture:*** Smart agriculture is a beneficial use case in big IoT data analytics. Sensors are the actors in the smart agriculture use case. They are installed in fields to obtain data on moisture level of soil, trunk diameter of plants, microclimate condition, and humidity level, as well as to forecast weather. Sensors transmit obtained data using network and communication devices. These data pass through an IoT gateway and the Internet to reach the analytics layer. The analytics layer processes the data obtained from the sensor network to issue commands. Automatic climate control according to harvesting requirements, timely and controlled irrigation, and humidity control for fungus prevention are examples of actions performed based on big data analytics recommendations.
6. ***Smart Cities:*** Big data collected from smart cities offer new opportunities in which efficiency gains can be achieved through an appropriate analytics platform/infrastructure to analyze big IoT data. Various devices connect to the Internet in a smart environment and share information. Moreover, the cost of storing data has been reduced dramatically after the invention of cloud computing technology. Analysis capabilities have made huge leaps. Thus, the role of big data in a smart city can potentially transform every sector of the economy of a nation. Hadoop with YARN resource manager has offered recent advancement in big data technology to support and handle numerous workloads, real-time processing, and streaming data ingestion.
7. ***Smart Inventory System:*** Finding useful information from large amounts of inventory systems data can help business owners generate more profit. The analytics of inventory-systems-generated datasets can help one ac- quire knowledge about market trends. Product recommendations can be generated after analyzing seasonal variations. The analytics of inventory data can also help detect fraudulent cases. Analytics may aid advertisers in strategically placing their advertisements. Predictive analytics can help people make valuable decisions and understand further their customers and products. Data analytics can also

help companies identify their potential risks and opportunities.

VI. CONCLUSION

IoT is one of the biggest sources of big data, which are rendered useless without analytics power. IoT interacts with big data when voluminous amounts of data are needed to be processed, trans- formed, and analyzed in high frequency. This work specifically focus on the big data context. First, we investigate the recent literature on big data processing and analytics solutions for IoT. Second, we identify the numerous requirements for big data and analytics in IoT. Third, we taxonomies the literature. Fourth, we determine the various opportunities that are brought about by big data. Fifth, we highlight the role of data analytics in IoT applications. Sixth, we present the open research challenges that must be addressed in the future. Seventh, we conclude that the existing big data solutions in the IoT paradigm are still in their infancy and the challenges associated with them must be solved in the future.

The growth rate of data production has increased drastically over the past years with the proliferation of smart and sensor devices. The interaction between IoT and big data is currently at a stage where processing, transforming, and analyzing large amounts of data at a high frequency are necessary. We conducted this survey in the context of big IoT data analytics. First, we explored recent analytics solutions. The relationship between big data analytics and IoT was also discussed. Moreover, we proposed an architecture for big IoT data analytics. Furthermore, big data analytics types, methods, and technologies for big data mining were presented. Some credible use cases were also provided. In addition, we explored the domain by discussing various opportunities brought about by data analytics in the IoT paradigm. Several open research challenges were discussed as future research directions. Finally, we concluded that existing big IoT data analytics solutions remained in their early stages of development. In the future, real-time analytics solution that can provide quick insights will be required.

VII. REFERENCES

[1]. Mohsen Marjani, Fariza Nasaruddin, Abdullah Gani, Ahmad Karim, Ibrahim Abaker Targio Hashem, Aisha Siddiqa, Ibrar Yaqoob,"Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges".

[2]. Hongming Cai, Boyi Xu, Lihong Jiang, Athanasios V. Vasilakos,"IoT-based Big Data Storage Systems in Cloud Computing: Perspectives and Challenges".

[3]. Victoria Moreno-Cano, Fernando Terroso-Saenz, Antonio F. Skarmeta-Gómez,"Big Data for IoT Services in Smart Cities", Department of Information and Communications Engineering University of Murcia, Murcia, 30100 Spain

[4]. Tiainen, P., New opportunities in electrical engineering as result of the emergence of the Internet of Things. 2016.

[5]. S. Villarroya, D.M. Casas, M. Vilar, J. R.R. Viqueira, J.A. Taboada, J.M. Cotos,“Heterogeneous sensor data integration for crowdsensing applications”, In Proceedings of the 18th International Database Engineering & Applications Symposium. ACM, 2014.

[6]. Wikibooks: https://en.wikibooks.org/wiki/I_Dream_of_IoT/Chapter_5_:_IoT_and_Big_Data

Cite this article as :

Devinderpal Yadav, "IoT Big Data Analytics - Architecture Design and Use Cases", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 6 Issue 2, pp. 630-638, March-April 2019.
Journal URL : http://ijsrset.com/IJSRSET196315