Impacts of Big IoT on Data Analytics Platform

Sandeep Bhargava¹, Bright Keswani², Dinesh Goyal³ ¹Research Scholar, Suresh Gyan Vihar University, ²Professor, Suresh Gyan Vihar University, Department of CSE ³Professor, Department of CSE, Poornima Institute of Engineering & Technology

Abstract - IoT data is highly unstructured, which makes it difficult to analyze using traditional analysis and business intelligence tools designed to process structured data. IoT data comes from devices that frequently record noise processes (such as temperature, motion, or sound). There is often a large gap in the data in these devices. Damaged messages and erroneous readings must be cleared before they can be analyzed. This article discusses the impact of big IoT data on analytics and business benefits here.

Keywords - Big IoT Data Analytics, IoT Analytics

I. INTRODUCTION:

IoT-Internet of Things

The IoT is a dynamic and global network infrastructure, in which "Things"—subsystems and individual physical and virtual entities—are identifiable, autonomous, and self-configurable. "Things" are expected to communicate among themselves and interact with the environment by exchanging data generated by sensing, while reacting to events and triggering actions to control the physical world. The vision that the IoT should strive to achieve is to provide a standard platform for developing cooperative services and applications that harness the collective power of resources available through the individual "Things" and any subsystems designed to manage the aforementioned "Things". At the centre of these resources is the wealth of information that can be made available through the fusion of data that is produced in real-time as well as data stored in permanent repositories. This information can make the realization of innovative and unconventional applications and value-added services possible, and will provide an invaluable source for trend analysis and strategic opportunities. A comprehensive management framework of data that is generated and stored by the objects within IoT is thus needed to achieve this goal.

Massive amount of data which is generated from sensors at high speed of different variety is termed a Big IoT data. Sensors appeals across almost every sector of industry, the internet of things is going to trigger a massive influx of big data is termed as big IoT data. IoT (Internet of Things) and big data are closely intertwined and although they are not the same thing, it is very hard to talk about one without the other.

II. NECESSITY OF IOT AND BIG DATA IMPLEMENTATION:

IoT will enable big data, big data needs analytics, and analytics will improve processes for more IoT devices. IoT and big data can be used to improve various functions and operations in diverse sectors. Both have extended their capabilities to wide range of areas.

III. IMPACTS OF BIG IOT DATA ON DATA ANALYTICS PLATFORM

IoT is a network consisting of physical devices, which are also implanted with sensors, electronics, and software, thereby allowing these devices to exchange data. This ultimately allows better incorporation between real world physical entities and computer-operated systems.

IoT data is highly unstructured which makes it difficult to analyze with traditional analytics and business intelligence tools that are designed to process structured data. IoT data comes from devices that often record fairly noisy processes (such as temperature, motion, or sound). The data from these devices can frequently have significant gaps, corrupted messages, and false readings that must be cleaned up before analysis can occur.

Analytics derived from IoT data can help businesses solve problems before they happen, while also reducing the cost of operations and maintenance. "What is the business value of the data generated by the IoT? And what do we need to do to realize that value?"

. IoT data is complex, vast, and fast-moving. Aberdeen's survey of 68 IoT organizations revealed the areas where organizations struggle and hope to improve:

- The average IoT organization's total volume of data grew by 30% over the past year.
- 54% of IoT organizations reported that their current data analysis capabilities are insufficient.
- 50% of IoT organizations failed to improve time-to decision over the past year.

There are different types of data analytics that can be used and applied in the IoT investments to gain advantages. Some of these types have been listed and described below.

- Streaming Analytics: This form of data analytics is also referred as event stream
 processing and it analyzes huge in-motion data sets. Real-time data streams are analyzed
 in this process to detect urgent situations and immediate actions. IoT applications based
 on financial transactions, air fleet tracking, traffic analysis etc. can benefit from this
 method.
- Spatial Analytics: This is the data analytics method that is used to analyze geographic patterns to determine the spatial relationship between the physical objects. Locationbased IoT applications, such as smart parking applications can benefit from this form of data analytics.
- Time Series Analytics: As the name suggests, this form of data analytics is based upon the time-based data which is analyzed to reveal associated trends and patterns. IoT applications, such as weather forecasting applications and health monitoring systems can benefit from this form of data analytics method.
- Prescriptive Analysis: This form of data analytics is the combination of descriptive and predictive analysis. It is applied to understand the best steps of action that can be taken in a particular situation. Commercial IoT applications can make use of this form of data analytics to gain better conclusions.

IV. FACTORS THAT BIG DATA IS IMPACTED BY IOT ARE

Big Data Storage

At basis, the key necessities of big data storage are that it can handle very huge amounts of data and continuous balancing to keep up with expansion and that it can provide the

ISSN: 2395-0196

input/output operations per second (IOPS) necessary to deliver data to analytics tools. The data is of different form and format and thus, a datacenter for storing such data must be able to handle the load in changeable forms. Obviously IoT has a direct impact on the storage infrastructure of big data. Collection of IoT Big Data is a challenging task because filtering redundant data is mandatorily required. After Collection, the data has to transfer over a network to a data center and maintained. Many companies started to use Platform as a Service (PaaS) to handle their infrastructure based on IT. It helps in developing and running web applications. By this way, big data can be managed efficiently without the need of expanding their infrastructural facilities to some extent. IoT Big Data Storage is certainly a challenging task as the data grows in a faster rate than expected.

Data Security Issues

The IoT has given new security challenges that cannot be controlled by traditional security methods. Facing IoT security issues require a shift. For instance, how do you deal with a situation when the television and security camera at your home are fitted with unknown Wi-Fi access. Few security problems are 1. Secure computations in distributed environment 2. Secure data centers 3. Secure transactions 4. Secure filtering of redundant data 5. Scalable and secure data mining and analytics 6. Access control 7. Imposing real time security, etc., a multi-layered security system and proper network system will help avoid attacks and keep them from scattering to other parts of the network. An IoT system should follow rigorous network access control policies and then allowed to connect. Software-defined networking (SDN) technologies should be used for point-to-point and point-to multipoint encryption in combination with network identity and access policies.

• Data analytics:

Data analytics is the science of examining raw data with the idea of coming to conclusions about that information. Data analytics is used in many industries to allow them to make better business decisions and in the sciences to verify or disprove existing models or theories. IoT Big data analytics is very much needed to end up in an optimized

ISSN: 2395-0196

decision. Big data analytics will help you understand the business value it brings and how different industries are applying it to deal with their sole business necessities. According to the Gartner IT dictionary, Big Data is variety of information assets, high-volume, and high-velocity and, innovative forms of information processing for enhanced approach and decision making. Volume refers to the size of data. Data sources can be social media, sensor and machine-generated data, structured and unstructured networks, and much more. Enterprises are flooded with terabytes of big data. Variety refers to the number of forms of data. Big data deals with numbers, 3D data and log files, dates, strings, text, video, audio, click streams. Velocity refers to the speed of data processing. The rate at which data streams in from sources such as mobile devices, click streams, machine-to-machine processes is massive and continuously fast moving. Big data mining and analytics helps to reveal hidden patterns, unidentified correlations, and other business information.

BIG DATA PROCESSING & ANALYTICS PLATFORM

Apache Spark - The University of California, Berkeley's AMP Lab, developed Apache in 2009. Apache Spark is a fast large-scale data processing engine and executes applications in Hadoop clusters 100 times faster in memory and 10 times faster on disk. Spark is built on data science and its concept makes data science effortless. Spark is also popular for data pipelines and machine learning models development. Spark also includes a library – MLlib that provides a progressive set of machine algorithms for repetitive data science techniques like Classification, Regression, Collaborative Filtering, Clustering, etc.

AWS EMR - AWS EMR processing Engine: Amazon EMR provides a managed Hadoop framework that makes it easy, fast, and cost-effective to process vast amounts of data across dynamically scalable Amazon EC2 instances. Amazon EMR securely and reliably handles a broad set of big data use cases, including log analysis, web indexing, data transformations (ETL), machine learning, financial analysis, scientific simulation, and bioinformatics. Amazon EMR enables quickly and easily provisions as much capacity as you need, and automatically or manually add and remove capacity.

Hadoop MapReduce - Hadoop MapReduce is an open-source programming model for distributed computing. It simplifies the process of writing parallel distributed applications by handling all of the logic, while you provide the Map and Reduce functions. The Map function maps data to sets of key-value pairs called intermediate results. The Reduce function combines the intermediate results, applies additional algorithms, and produces the final output.

Azure Power BI Analytics Platform - Power BI is a business analytics service that delivers insights to enable fast, informed decisions.

AWS kensis- Amazon Kinesis makes it easy to collect, process, and analyze real-time, streaming data so you can get timely insights and react quickly to new information. Amazon Kinesis offers key capabilities to cost-effectively process streaming data at any scale, along with the flexibility to choose the tools that best suit the requirements of your application. With Amazon Kinesis, you can ingest real-time data such as video, audio, application logs, website click streams, and IoT telemetry data for machine learning, analytics, and other applications. Amazon Kinesis enables you to process and analyze data as it arrives and respond instantly instead of having to wait until all your data is collected before the processing can begin

Scala - It is one of the best known languages with one of the largest user bases. Since it was engineered to run on the JVM, anything that is written on Scala can run anywhere that Java runs. It is highly flexible and functional enough to play well with others. It is becoming a popular tool for anyone doing machine learning at large scales or building high-level algorithms.

Splunk - Splunk is a tool that analyzes and searches the machine-generated data. Splunk pulls all text-based log data and provides a simple way to search through it, a user can pull in all kind of data, and perform all sort of interesting statistical analysis on it, and present it in different formats.

Hive - Hive is an open source data warehouse and analytics package that runs on top of Hadoop. Hive is operated by Hive QL, a SQL-based language which allows users to structure, summarize, and query data. Hive QL goes beyond standard SQL, adding first-class support for map/reduce functions and complex extensible user-defined data types like JSON and Thrift. This capability Suresh Gyan Vihar University Journal of Engineering & Technology (*An International Bi- Annual Journal*)
Vol. 6, Issue 2, 2020, pp 31-38
ISSN: 2395-0196
allows processing of complex and unstructured data sources such as text documents and log files.
Hive allows user extensions via user-defined functions written in Java

Pig - Pig is an open source analytics package that runs on top of Hadoop. Pig is operated by Pig Latin, a SQL-like language which allows users to structure, summarize, and query data. As well as SQL-like operations, Pig Latin also adds first-class support for map/reduce functions and complex extensible user defined data types. This capability allows processing of complex and unstructured data sources such as text documents and log files. Pig allows user extensions via user-defined functions written in Java.

RapidMiner - RapidMiner is a powerful integrated data science platform developed by the same company that performs predictive analysis and other advanced analytics like data mining, text analytics, machine learning and visual analytics without any programming. RapidMiner can incorporate with any data source types, including Access, Excel, Microsoft SQL, Tera data, Oracle, Sybase, IBM DB2, Ingres, MySQL, IBM SPSS, and Dbase etc. The tool is very powerful that can generate analytics based on real-life data transformation settings, i.e. you can control the formats and data sets for predictive

V. THE LIMITATIONS

IoT data is largely sourced from sensors that are advancing in capability. These sensors gather information from their environment that the IoT connected device usually receives via cloud in the form of datasets. It's then up to the solutions provider how these datasets are translated and presented to the user – aka the data analysis. This means that, as hardware advances and devices are able to pick up on more attributes, the information available to the end-user also advances. However, as the IoT industry grows in popularity and becomes even more intertwined with daily life, it's important to bear in mind that there are still some potentially significant constraints. The limitations to the mutually beneficial relationship between excelling devices and information gathered are often dictated by roadblocks encountered during hardware development. Factors such as unforeseen costs and delays in production time are the main hindrance to IoT solutions that have the software aspect nailed down.

VI. CONCLUSION

Advanced data analytics are no longer a fancy add-on but an integral part of any IoT solution. They provide users with the knowledge necessary to make smarter business or personal decisions and can point out potential problem areas without requiring significant effort on the user's end. IoT is fueled by the power and capability of data. However, as much value as there is in pure quantitative data, there's more power in the way data is categorized and what insights a user can draw from it. Data analysis enables profitable decision making by consumers and, as the field of IoT technology expands in popularity, with it will grow the demand for advanced data analysis tools.

REFERENCES

- Anawar, M.R., Wang, S., Azam Zia, M., Jadoon, A.K., Akram, U. and Raza, S., 2018. Fog computing: An overview of big IoT data analytics. Wireless Communications and Mobile Computing, 2018.
- Sharma, R., Agarwal, P. and Mahapatra, R.P., 2020. Evolution in Big Data Analytics on Internet of Things: Applications and Future Plan. In Multimedia Big Data Computing for IoT Applications (pp. 453-477). Springer, Singapore.
- Bhargava, S., Goyal, D. and Keswani, B., 2019. Performance Comparison of Big Data Analytics Platforms. International Journal of Engineering, Applied and Management Sciences Paradigms (IJEAM), ISSN, pp.2320-6608.
- 4. "Big IoT Data Analytics: Literature Review, Opportunity and its Research Challenges", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.5, Issue 11, page no.451-455, November-2018, Available :http://www.jetir.org/papers/JETIR1811B68.pdf
- 5. Bhargava, Sandeep; Goyal, Dinesh; Keswani, Bright."Performance Assay of Big IoT Data Analytics Framework." International Journal of Recent Technology and Engineering (IJRTE), Volume 8, 4, page no:8593-8596(2019) or DOI: 10.35940/ijrte.D7383.118419