ASIAN JOURNAL OF INFORMATION TECHNOLOGY
MEDWELL PUBLICATIONS

A Paper on Comparison Between Big Data and IoT

P. Sunanda
G. Pulla Reddy Engineering College (Autonomous), GPREC, Kurnool, India

Key words: Big Data, internet of things, characteristics, tools, applications, challenges

Corresponding Author:
P. Sunanda
G. Pulla Reddy Engineering College (Autonomous), GPREC, Kurnool, India

Page No.: 102-107
Volume: 19, Issue 5, 2020
ISSN: 1682-3915
Asian Journal of Information Technology

Abstract: Big Data and IoT (Internet of Things) are the emerging technologies in today's computer world. Big Data is nothing but a huge amount of data that can't be analysed and processed easily. Whereas the Internet of Things (IoT) is nothing but a system of interrelated computing devices which can be either mechanical or digital machines that provide the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction. These two terms are being used by everyone without knowing their difference. This study will help you in comparing Big Data and IoT starting from their definition to applications.

INTRODUCTION

In the current digital era, according to massive progress and development of the internet and online world, cloud storages, social networks and other websites are producing big volume of data. Although, this massive volume of data can be really useful for people and corporations, it can be problematic in storing this data using traditional databases like Oracle. So, we move towards the new technology called "Big Data". In the same way along with big data there is a new technology called IoT (Internet of Things) that mainly deals with how the things are interconnected and communicated with one another without any human interaction.

Definition

Big data: Big data can be defined as a massive volume of either structured or unstructured data or the combination of both that can't be processed easily by traditional database and software techniques (Fig. 1).

Or big data can be defined as a collection of large and complex data sets that on-hand database management tools or traditional data processing applications can't process.

Fig. 1: Sample figure of big data

Examples on big data: Data from different websites like ecommerce sites (Flipkart, Snapdeal) and social media sites (Facebook and twitter) is the combination of structured and unstructured and huge in size that can't be processed easily. So, this data is considered a best example for big data.

IoT: The Internet of Things (IoT) is a system of interrelated computing devices, objects, mechanical and digital machines, animals or people that have the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction by using their unique identifiers[1] (Fig. 2).

Fig. 2: Sample figure of IoT

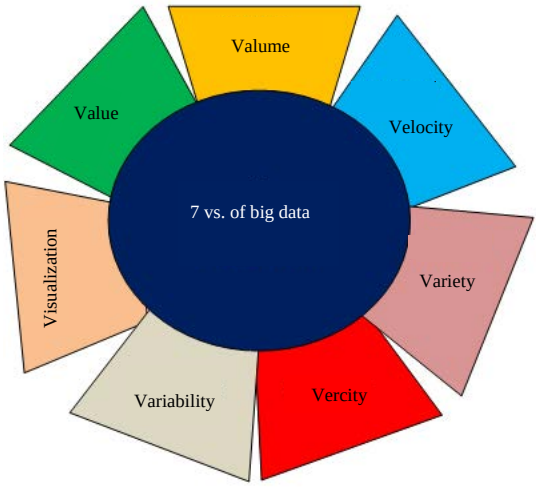

Fig. 3: 7V' of Big Data

Examples on IoT: The main example of internet of things is smart city which consist of Smart surveillance, safer and automated transportation, Bigbelly smart waste and recycling system (helps with smart trash picking, avoid overflows and generate notifications making waste management), smarter energy management systems like city sense is a smart and wireless outdoor lighting control system (helps in saving electricity by intuitively adjusting brightness of streets lights based on presence of automobiles and pedestrians (Fig. 3).

Characteristics

Characteristics of big data

Volume: In terms of volume big data can be characterized as a collection of large amount of data like some Gigabytes to Zettabytes (ZB) or even Yottabytes (YB) of data.

Velocity: It refers to the speed at which data is being generated, produced, created or refreshed. It mainly deals with how much time the huge amount of data takes to pass through network of systems. Variety big data is a collection of different types of data like structured, semi-structured and unstructured data. This implies that it can be defined as variety of data.

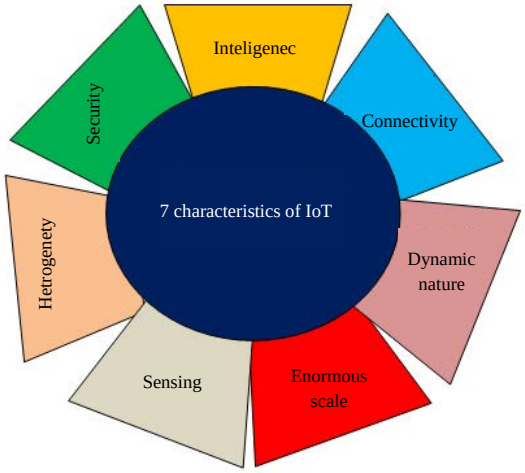

Fig. 4: Characteristics of IoT

Veracity: The power of big data lies in analyzing the data. It has many analytics tools to analyze data carefully and perfectly to maintain veracity.

Variability: The variability of big data lies in multitude of data dimensions resulting from multiple disparate data types, sources and inconsistent speed at which data is loaded into the database.

Visualization: Big data visualization mainly deals representing huge volume of data to the user for his easy understanding. Representation of data is done using some charts and graphs.

Value: Last but not least the most important of all, is value. The other characteristics of big data are meaningless if business value can't be obtained from the data. By better understanding the customers, targeting them accordingly, optimizing processes and improving machine or business performance substantial value can be found in big data.

Characteristics of IoT

Intelligence: The combination of algorithms and computation, software and hardware made IoT smart. IoT enhanced its capabilities with it's ambient intelligence which facilitate the things to respond in an intelligent way to a particular situation and supports them in carrying out specific tasks (Fig. 4).

Connectivity: As IoT is a system of interrelated computing devices that provide the ability to transfer data over network, connectivity places an important role in connecting these devices together for effective communication among them.

Fig. 5: Sample figure for IoT with Big Data

Dynamic nature: The primary activity of internet of things is to collect data from its environment; this is achieved with dynamic changes that take place around devices. The states of these devices change dynamically. In addition, to the state of the device, the number of devices also changes dynamically with a person, place and time (Fig. 5).

Enormous scale: In addition to the state of the device, the number of devices also changes dynamically with a person, place and time which results in production of huge amount of data. But the management of data generated from these devices and their interpretation for application purposes becomes more critical.

Sensing: IoT without sensors won't be possible because they help to detect or measure any change in the environment to generate data that can report on their status or even interact with the environment. Sensing technologies provide the means to create capabilities that reflect a true awareness of the physical world and the people in it. The sensing information is simply the analogue input from the physical world but it can provide the rich understanding of complex world.

Heterogeneity: Heterogeneity in internet of things is one of the key characteristics. Devices in IoT are based on different hardware platforms and networks and can interact with other devices or service platforms through different networks. IoT architecture should support direct network connectivity between heterogeneous networks. The key design requirements for heterogeneous things and their environments in IoT are scalabilities, modularity, extensibility and interoperability.

Security: IoT devices are naturally vulnerable to security threats. There is a high level of transparency and privacy issues with IoT. It is important to secure the endpoints, the networks and the data that is transferred across all of means creating a security paradigm. There are a wide variety of technologies that are associated with internet of things that facilitate in its successful functioning[2].

Tools

Big data tools

Splice machine: This is a SQL-on-Hadoop real-time database tool for developers that help to scale out on commodity hardware by using standard SQL like commands.

Google charts: This tool developed by Google to visualize data of websites in the form of charts and trees that helps for easy sorting, modifying and filtering of data.

Cambridge semantics: This is an analysing tool. It has developed mainly to collect, integrate and combine data from multiple sources and customized dashboards.

MongoDB: Is an open-source documental database that comes with full index support to have a precise control over the final results. It consists of some document-based queries and GridFS for storing files.

Pentaho: Is an open and embeddable platform comes with extensive analytics capabilities with data mining and predictive analysis. It mainly joins data integration and business analytics for visualising, analysing and blending Big Data[3].

IOT tools

Swarm: New services to the devices in IOT network can be easily added using Swarm, an IoT development platform. Swarm Dashboards offer real-time, visual access to device features by serving as central, device-specific home pages. Real-time Analytics and reporting, Integration with CRM/ERP products, historical activity logs, compliance testing are the key features of swarm.

Axeda: It is a platform used to transform machine data into valuable insights, build and run applications and integrate machine data with other applications and systems to optimize business processes by providing a comprehensive cloud-based platform for managing connected products and machines and implementing IoT and M2M applications.

IoBridge: Whether the need to connect a single product or more than one million products, ioBridge speeds your time to market and lowers your cost per connected product. To internet-able their devices, gather product usage data, perform remote device maintenance and gain real-time, actionable insights to drive decision-making >50,000 users in 40 countries are already relying on ioBridge.

Zatar: A new cloud-based infrastructure that automatically detects the devices and connects them to the

internet. It is an open platform for managing entire device network. By connecting all devices to the same platform and facilitating M2M communication, all connected devices and their respective users can share data and collaborate seamlessly.

Thingworx: ThingWorx facilitates rapid, streamlined creation of end-to-end smart applications for agriculture, cities, water, building and telematics. Thing Worx reduces the time, cost and risks of building M2M and IoT applications. Users can build comprehensive mobile interfaces with zero coding, take advantage of thing work composer for application modelling as well as real-time dashboards and collaborative work spaces all with the scalability to support millions of devices[4].

APPLICATIONS

Big data applications

Transportation: In transportation industry of all government, private and individual sectors Big data have applications like: In government sector big data is used for traffic control and route planning. In private sectors it is used for revenue management and technological enhancements. Individuals use big data for route planning, for saving fuel and time for travel arrangements in tourism etc.

Health care: The healthcare sector has access to huge amounts of data. But this data is inadequate or unusable if it is not processed properly. Additionally, the healthcare databases that hold health-related information have made it difficult to link data that can show patterns useful in the medical field. This data can be analyzed easily using big data tools. For example, Google Maps have been used to create visual data for faster identification and efficient analysis of healthcare information from free public health data.

Manufacturing and natural resources: In the natural resources industry, big data allows predictive modelling to support decision making that has been utilized to ingest and integrate large amounts of data from geospatial data, graphical data, text and temporal data. Areas of interest where this has been used include; seismic interpretation and reservoir characterization. Big data has also been used in solving today's manufacturing challenges and gaining competitive advantage among other benefits.

Insurance: Big data has been used in the insurance industry to provide customer insights for transparent and simpler products by analyzing and predicting customer behaviour through data derived from social media, GPS-enabled devices and CCTV footage. The big data also allows for better customer retention from insurance companies.

Retail and whole sale trade: From traditional brick and mortar retailers and wholesalers to current day e-Commerce traders, the industry has gathered a lot of data over time. This data, derived from customer loyalty cards, POS scanners, RFID, etc. is not being used enough to improve customer experiences on the whole. In retail industry big data analytics is used for optimized staffing through data from shopping patterns, local events and so on reduced fraud and timely analysis of inventory[5].

APPLICATIONS OF IoT

Airline: With the help of equipment tracking app an airline's engineers can maintain the equipment easily by having the live view of locations of equipments. This IoT application increases the efficiency of engineers by not only generating significant cost savings and process improvements but also by impacting the customer experience through more reliable on-time flights.

Pharmaceutical: The temperature of medications is a key in meeting quality standards and regulations. A medication temperature monitoring app uses sensors as a way to ensure the best possible delivery of medical supplies and monitors every shipment in order to ensure that proper temperature is maintained and to remind patients to take their medication on time.

Manufacturing: A lighting manufacturer for the horticultural industry built a Smart App that leverages IoT sensors and predictive analytics to perform predictive maintenance and optimize lighting, power consumption and plant photosynthesis. The IoT application transformed their business from a lighting systems manufacturer to a greenhouse optimization as-a-service.

Insurance: An insurance company offers policyholders discounts for wearing internet-connected Fitbit wristbands. The fitness tracking service is part of the insurer's Vitality program aimed at integrating wellness benefits with life insurance. Through this IoT application, this insurer is creating smart life insurance products and rewarding customers for their positive actions.

Business services: A facility services company are using their multi-device IoT application to enable support personnel to receive alerts about service issues and take immediate action. By aggregating data from thousands of sensors in places like coffee machines, soap dispensers, paper towel dispensers and mouse traps rather than manual checks, the application has significantly cut costs and improved service levels.

CHALLENGES

Big data challenges

Meeting the need for speed: In today's hyper competitive business environment, companies not only have to find and analyze the relevant data they need, they must find it quickly. Visualization helps organizations perform analyses and make decisions much more rapidly but the challenge is accessing the sheer volumes of data at high speed. This challenge only grows as the degree of granularity increases. One possible solution is hardware. Some vendors are using increased memory and powerful parallel processing to crunch large volumes of data extremely quickly. Another method is putting data in-memory but using a grid computing approach where many machines are used to solve a problem. Both approaches allow organizations to explore huge data volumes and gain business insights in near-real time.

Understanding the data: It takes a lot of understanding to get data in the right shape. For example, if the data comes from social media content, there is a need to know who the user is in a general sense such as a customer using a particular set of products. One solution to this challenge is to have the proper domain expertise in place. Make sure the people analyzing the data have a deep understanding of where the data comes from what audience will be consuming the data and how that audience will interpret the information.

Addressing data quality: Even if one can find and analyze data quickly and put it in the proper context for the audience that will be consume huge amount of time to visualize it properly. This is a challenge with any data analysis but when considering the volumes of information involved in big data projects, it becomes even more pronounced. Again, data visualization will only prove to be a valuable tool if the data quality is assured. To address this issue, companies need to have a data governance or information management process in place to ensure the data is clean. It's always best to have a pro-active method to address data quality issues, so that, problems won't arise later.

Displaying meaningful results: Plotting points on a graph for analysis becomes difficult when dealing with extremely large amounts of information or a variety of categories of information. For example, imagine having 10 billion rows of retail data that you're trying to compare. The user trying to view 10 billion plots on the screen will have a hard time seeing, so many data points. One way to resolve this is to cluster data into a higher-level view where smaller groups of data become visible. By grouping the data together or "binning," one can more effectively visualize the data and display meaningful results.

Dealing with outliers: The graphical representations of data made possible by visualization which can communicate trends and outliers much faster than tables containing numbers and text. Users can easily spot issues that need attention simply by glancing at a chart. Outliers typically represent about 1-5% of data but when working with massive amounts of data, viewing 1-5% of the data is rather difficult. How to represent those points without getting into plotting issues? Possible solutions are to remove the outliers from the data (and therefore from the chart) or to create a separate chart for the outliers. One can also bind the results for both viewing the distribution of data and seeing the outliers.

IoT challenges

Security: There are many entry points for malware in IoT as it connects more devices together. Less expensive devices that are in the network of IoT are subject to security issues. More layers of software, integration middleware, APIs, machine-to-machine communication, etc. create more complexity and new security risks. Handling these security issues in IoT network is a difficult task users facing even now.

Trust and privacy: As it is a combination of different devices like sensors, cctv camera and different people from different places, the data produced by them will be so sensitive that there is need to maintain privacy in data and people should have trust while communicating with devices in IoT system. So, it is a great challenge for IoT to get trust from users by maintaining privacy.

Complexity: Confusion and integration issues. As IoT system is a combination of multiple platforms, numerous protocols and large numbers of APIs it facing challenges like integration and testing.

Evolving architectures, protocol wars and competing standards: With so many (propositions) devices involved with the IoT, many companies are using them for their use. As legacy companies seek to protect their proprietary systems advantages and open systems proponents try to set new standards these propositions are bound to be ongoing turf wars. Based on different requirements determined by device class, power requirements, capabilities and uses there may evolve multiple standards. This presents opportunities for platform vendors and open source advocates to contribute and influence future standards.

Concrete use cases and compelling value propositions: Lack of clear use cases will slow down adoption of the IoT. In order to get rid of this IoT mainly require well-grounded, customer-oriented communications and

messaging. When buyers are looking for a "whole solution" or complete value-added service the detailed explanations of a specific device or technical details of a component will compel value for whole solution. So, IoT providers have to explain the key benefits of their services and have to compel value proportions of their use cases and services[6].

CONCLUSION

Though big data and IoT differ in many ways, they provide many benefits when combined with one another. Let's see the combination of both IoT and Big Data i.e., IoT with big data.

IoT with big data: Big Data and the internet of things: a perfect match made in Heaven IoT a disruptive technology requires new infrastructures including hardware, software applications as well as an operating system. Enterprises using IoT have the need to deal with the influx of data that starts flowing in and by every minute as the data grows they have to analyze it in real-time. It's here where the big data comes in. As IoT devices that produce a continuous stream of information, big data analytics tools are capable of handling this masses of data transmitted from them. But just to differentiate the two, it can be said that the IoT delivers the information from which big data analytics can draw the information to create the insights required of it. However, the IoT brings data on a different scale and from different devices; the big data analytics tools should rapidly process the data providing accurate and fast extraction in order to accommodate its needs.

Another example for IoT with Big Data is let's consider a connected car. The connected car is actually a combination of multiple use cases of IoT data. In this connected car the data comes from different aspects like the data of scheduled maintenance, recalls and safe operations, the data related to insurance premiums, data about the environment collected by sensors, the driver, and the car itself to assist with route optimization and autonomous driving and so on. Everything is a tangled web of information create an unprecedented deluge of data never before possible. In order to collect, store and analyze this data it is impossible by traditional databases so, here the use of big data tools helps in better understanding and storage of data.

REFERENCES

01. Gillis, A., L. Rosencrance, S. Shea and I. Wigmore, 2005. Internet of Things (IoT). TechTarget Marketing Company, Newton, Massachusetts. https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT.
2. Elena-Lenz, C., 2015. Internet of things: Six key characteristics. LinkedIn Corporation Company, Sunnyvale, California.https://www.inkedin.com/pulse/internet-things-six-key-characteristics-carlos-elena-lenz.
03. Nunns, J., 2015. 10 of the most popular big data tools for developers. Computer Business Review, England.
04. Anonymous, 2020. Enterprise cloud: The flexible infrastructure as a service solution for your future-proof cloud computing applications. IONOS by 1&1 Internet Company, Europe. https://www.ionos.com/enterprise-cloud
05. Anonymous, 2009. Top 10 big data applications across industries. Simplilearn Solutions, San Francisco, California.
06. Kocher, C., 2014. The internet of things: Challenges and opportunities. Sand Hill Group, Brandon. http://sandhill.com/article/the-internet-of-things-challenges-and-opportunities.