A SURVEY ON BIG DATA AND ITS CHALLENGES RELATED TO IOT

Surajit Mohanty
Asst. Prof., Computer Science & Engineering,
DRIEMS, Cuttack, BPUT, India

Sameer Kumar Das
Asst. Prof., Computer Science & Engineering,
GATE, BPUT, Berhampur, India

Shekharesh Barik
Asst. Prof., Computer Science & Engineering,
DRIEMS, Cuttack, BPUT, India

Mamata Mayee Rout
Research scholar, Computer Science & Engineering,
DRIEMS, BPUT, Cuttack, India

Abstract- Now a day’s Big Data is an emerging technology that has an ability to change the world. It solves the velocity-volume-variety problem. While dealing with unstructured and structured data, data science plays an important role for data cleansing, preparation, and analysis. Apart from this Big Data refers to homogenous volumes of data that cannot be processed effectively with the traditional applications. Big Data is something that can be used to analyse insights which can lead to better decision and strategic business moves of an organisation. For this it requires proper data analytic tools which are able to examine raw data with the purpose of drawing conclusions about that information.

The Internet of Things will comprise many millions of systems, in which some of them belongs to big data. The aggregated information from these systems represent, really big data systems. The problems arising from so many devices, data and processing coming together are likened to an international market, with similar challenges. These big data problems in the IoT are reviewed from a reliability engineering perspective.

Keywords: Big Data, Big Data Challenges, Volume of data, Velocity of data, Iot

I. INTRODUCTION

Big Data is a new term in today technology era. In a wide range of application areas, data is being collected at an unprecedented scale. Decisions that previously were based on guesswork, or following some traditional way, can now be replaced by using data-driven mathematical models. Such Big Data analysis now drives nearly every aspect of society, including mobile services, retail, manufacturing, financial services, life sciences, and physical sciences. Big data has existed long before the IoT burst out into the scene to perform analytics. Information is defined as big data when it demonstrates the 4 V’s: volume, variety, velocity, and veracity.[1]

Data analysis is considerably more challenging than simply locating, identifying, understanding, and citing data. For effective large-scale analysis all of this has to happen in a completely automated manner. This requires differences in data structure and semantics to be expressed in forms that are computer understandable, and then robotically resolvable. The Internet of Things (IoT) is a collective noun for any system consisting of sensors, actuators, computational elements and other devices communicating locally and across the Internet which create large no of unstructured data in a large volume. The aggregation
of data from a large number of IoT ecosystems, can lead to large data sets for analytic purposes. Furthermore, IoT applications could connect (deliberately or accidentally) to one or more big data systems outside the ecosystem, thus creating an aggregate of big data system orders of magnitude larger than any of the constituents. In this sense, every IoT system, even a small, local IoT ecosystem, is a potential big data system.

II. PROBLEM ASSOCIATE WITH BIG DATA
A problem with current Big Data analysis is the lack of coordination between database systems, which host the data and provide SQL querying, with analytics packages that perform various forms of non-SQL processing, such as data mining and statistical analyses. Today’s analysts are impeded by a tedious process of exporting data from the database, performing a non-SQL process and bringing the data back. There are two basic challenge associated collection and processing along with some reliability challenges with concern to IoT.[8]

a) The first step is data acquisition from data sources, such as sensor networks, can produce staggering amounts of raw data. Much of this data is of no interest, and it can be filtered and compressed by orders of magnitude. One challenge is to define these filters in such a way that they do not discard useful information.

b) The second big challenge is to automatically generate the right metadata to describe what data is recorded and how it is recorded and measured. This metadata is likely to be crucial to downstream analysis.

Creating value from Big Data is a multistep process: Acquisition, information extraction and cleaning, data integration, modelling and analysis, and interpretation and deployment. Many discussions of Big Data focus on only one or two steps, ignoring the rest.[4] There are numerous reliability challenges to deploying practical, large scale IoT systems. These challenges include, communications problems (e.g. lost signals, noise), fault-tolerance (e.g. sensor failure) and securing the network. But let’s focus on the reliability issues specific to big data IoT systems.[10] In particular, with respect to the data there are three fundamental challenges:
1. Authentication
2. Security
3. Uncertainty

The significant increase in connected devices that’s due to happen at the hands of the Internet of Things will, in turn, lead to an exponential increase in the data that an enterprise is required to manage. Here’s where IoT intersects wonderfully with big data and where it becomes evident that the two trends fit one another like a glove.

SOLUTIONS APPROACH FOR BIG DATA CHALLENGES
A. Heterogeneity
If data is not in natural language or in heterogeneity format then it may not provide valuable depth, because machine algorithms are suited for homogeneous data. So, for data analysis data must be structured carefully [1]. Data can be structured by using Map Reduce techniques, because all keys generated by the Map Reduce must fit into main memory [7].

B. Inconsistency and Incompleteness
Since big data comprises of lot of information coming from different sources, so error and missing values is a challenge for managing them. These errors and incompleteness must be managed in data analysis [2]. However, there is also challenges include with data analysis such as efficient representation, access, and analysis of unstructured or semi structured data in the further researches. To remove noise and correct
inconsistencies, different types of data preprocessing techniques can be applied such as data cleaning, data integration, data transformation and date reduction [4].

C. SCALE

Size of big data is an important challenge, although there are many researches to handle this issue such as handle big data with processor speed but some time increasing volume of data is faster than processor speed [1]. In Big Data applications, the state-of-the-art techniques and technologies cannot ideally solve the real problems, especially for real-time analysis. As Big Data requires a more storage and medium, if Hard Disk Drives (HDDs) are used for such purpose, then HDDs is slower than data processing engines, this challenges can be handled by using Solid State Drives (SSDs) and Pulse-code modulation (PCM) technologies [5]. To store such data there is also memory issues. Faster growth of data and memory issues can be solved by using grid computing approach.

D. VELOCITY OF DATA

In Big Data, data is generated with very high speed, and this speedy data requires processing in timely manner. This problem in learning algorithm can be solved by using online learning approaches [9].

E. TIMELINESS

Timeliness says How data can be filtered at real time for storage purpose [3]. This issue can be handled through Index structure of traffic management system. If data is not analysed quickly and there is not proper framework for users, then data for decision making will not be fruitful.

F. VISUALIZATION

Due to large and high dimension of Big Data, visualization of data is very difficult. There are some tools for visualization but mostly have poor performance and response time. New framework for visualization is highly necessary [6]. There is no communication path between data points, so companies cannot aggregate and manage the data across the enterprise.

G. BIG DATA FUNCTIONS

Leveraging Big Data often means working across functions like IT, engineering, finance and procurement. These issues require a path from scratch for collaboration over functions and businesses [5].

H. SECURITY

Internet security platforms like Zscaler offer IoT devices protection against security breaches with a cloud based solution. You can route the traffic through the platform and set policies for the devices so they won’t communicate with unnecessary servers.

Data protection is a major issue related to security. If it is solved from all dimensions then companies could take full advantage of their data [5].

CONCLUSION

Two terms that have been discussed that is big data and The Internet of Things (IoT) are not same but It’s hard to talk about one without the other, and both are closely intertwined. The Internet of Things and big data share a closely knitted future. There is no doubt the two fields will create new opportunities and solutions that will have a long and lasting impact. Our traditional RDBMS are not able to store large volume of data. Due to three big advantage (Cost reduction, faster, better decision making, new products and services) of Big Data, it is impossible to ignore the survival of Big Data in current IT Market. These advantages can be achieved after resolving the issues related to different challenges. Here we studied some of the challenges such as inconsistency and incompleteness, displaying meaningful results, scale, timeliness, and security while some of them have not been paid proper attention such as velocity and
visualization. Besides, remaining challenges such as finding talent for Big Data, IT architecture for Big Data are yet to be under research. It’s not just that big data and the IoT help each other, they also greatly impact each other. The more the Internet of Things grows, the more demands are placed on businesses’ big data capabilities.

REFERENCES


[10] M. arcinjedyk, making big data, small, using distributed systems for processing, analysing and managing large huge data sets, software professional’s network, cheshire data systems ltd.