High Performance Predictive Analytics in IoT

Sandeep Singh Rajpoot, Dhanraj Verma

Abstract*: Internet of things (IoT) is a quick-moving gathering of web associated sensors implanted in a wide-extending assortment of physical articles. While things can be any physical item (energize or lifeless) on the planet, to which you could associate or implant a sensor. Sensors can take countless potential estimations. Sensors produce gigantic measures of new, organized, unstructured, ongoing information, and structures enormous information. IoT information is exceptionally huge and confused, which can give genuine-time setting and supposition data about genuine articles or nature. Among the different challenges that the present IoT is facing, the three prime areas of concern are, need of efficient framework to receive IoT data, a need of a new scalable parallel indexing technique for efficiently storing IoT data and securing IoT generated data at all the stages i.e. from the edge devices to the cloud. A new efficient framework is introduced, which can retrieve meaningful information from these IoT devices and efficiently index it. For processing such enormous real time data generated from IoT devices, new techniques are introducing which are scalable and secure. The research proposes a general IoT network architecture. It describes the interconnectivity among the different things such as sensors, receivers and cloud. The proposed architecture efficiently receives real time data from all the sensors. The prime focus is on the elimination of the existing issues in IoT. Along with this, the provision has to make for standard future proofing against these new proposed schemes.

***Index Terms*: Internet of Things (IoT), cloud data, sensor data, indexing technique, interconnectivity, gigantic data**

I. INTRODUCTION

Industrial revolution took place before 200 years ago, to increase the productivity, which leads to accelerate the economy. Where mechanical power replaces human and animal muscle power. Internet revolution started in 1950, which depends on networking, computing and communication in order to increase the productivity and boosted global economy much better than industrial revolution. The next revolution that will drive a new thread to increase the productivity is Internet of Things (IoT). IoT is the amalgamation of smart devices, smart systems and intelligent decision-making capabilities. With the help of IoT, all things across the physical world can be connected and communicate with each other. All things can have their unique identities, so that they can be uniquely identify and interact with other things in the network of interconnected things. The word "Internet of Thing" is devised by Kevin Ashton. However, from past decade, IoT shelters almost all

the applications i.e. home automation, smart healthcare, utilities, smart transport, etc. Key enabling technologies for IoT includes RFID technology, wireless sensor network and smart detection technology, wireless technology, Wi-Fi, Bluetooth, internet technology, cloud computing, intelligent computing (Artificial Intelligence), etc.

Internet of Things (IoT) is a major wave in computing, which were used to connect traditional end-user devices. These "things" typically provide data, act on the environment and/or encompass points of control. For example, lighting can be adjusted based on data from an occupancy sensor and time of day, and chillers can be adjusted based on temperature sensors. As the cost of sensors and network connectivity becomes less expensive there is an increased interest in applications.

Sensors can be built small enough to be embedded into many physical objects and wireless communication technologies have improved to provide data connectivity that allows data to be produced frequently. These sensors will massively increase the amount of data available for analysis. Developing applications that make use of this data is challenging since application developers have to deal with heterogeneous devices and the underlying network for accessing the sensor data. After capturing the raw sensor data, the application has to transform the data to a proper format and apply analytics to data in order to extract valuable information.

IoT provides massive opportunities but also poses data management challenges. One of these challenges is that many IoT applications require that queries are long-running over data streams where the data is continuously generated. There are some variations of database management systems such as pipelinedb, EP-SPARQL, and Nile devoted to addressing this issue.

II. LITERATURE REVIEW

Exhaustive literature survey has been done, because it is very important task to know about the basics of IoT. It will also provide current research conditions and the possibilities of future research in that area. A survey has been done in related area.

Brighty et al(2015) The Internet of Things (IoT) is the following pattern of advancement that possibilities to enhance and upgrade our every day life in view of keen sensors and shrewd articles cooperating.

Manuscript published on November 30, 2019.
* Correspondence Author
Sandeep Singh Rajpoot*, Computer Application, Dr. APJ Abdul Kalam University, Indore, India, Email: sandeepraj413@gmail.com
Dr. Dhanraj Verma, Computer Application, Dr. APJ Abdul Kalam University, Indore, India, Email: Dhanrajmtech@gmail.com

The essential vision of the IoT was to in a general sense change the method for working together, creating more noteworthy skills, rousing further client organizes and displaying new plans of action. IoT incorporates gadgets, for example, radio recurrence distinguishing pieces of proof (RFID), sensors, and actuators, and in addition different instruments and savvy apparatuses that are turning into a basic piece of the Internet. It is critical to accumulate precise crude information capability; however more huge is to review and mine the crude information to digest more profitable data, for example, connections among things and administrations to give web of things or Internet of services. Devices can be associated with the Internet utilizing novel IP addresses in Internet Protocol (IP) network, subsequently concurring them to be perused, controlled, and overseen at all around and unfailingly. Security is an indispensable perspective for IoT arrangements. Because of the hearty assaulting capacity, speed, straightforward execution and extra highlights, differential blame investigation has end up being a critical strategy to measure the security in the Internet of Things.

Vyas et al. (2015) With calculation, availability, and information stockpiling winding up further developed and all-inclusive there has been a blast of IoT based application arrangements in broadened spaces from human services to open wellbeing, from sequential construction system planning to assembling and different other mechanical areas. IoT can be characterized as a system of physical objects, devices that contain installed innovation (like clever sensors, controllers and so forth.) which can convey, sense, or cooperate with inner or outside frameworks. As such, when articles can detect and convey, it changes how and where choices are made, and who influences them and as needs be operations to can be conveyed out, In this overview a cognizant exertion has been advanced to examine the cutting edge associated with IoT and its different differentiated building applications. Latest patterns and evolution, challenges, and future extent of this innovation have been touched upon for better understanding. Various IoT based applications have been investigated and conceivable approach for upgrading the utilization of this innovation have been talked about in this paper. Future headings and proposals for adequately and effectively enhancing the IoT based application zones have been touched upon. This paper will give a superior knowledge to any individual who wishes to convey out research in the field of IoT. In this paper they have endeavoured to give an all-encompassing viewpoint on IoT and IoT based applications, application zones, explore challenges in IoT, patterns and future conceivable outcomes in IoT.

Resul Das and Gurkan Tuna(2015) Machine to Machine (M2M) can be portrayed as advancements which permit both wired and remote frameworks to speak with different gadgets of the same ability.M2M conveys a few advantages to industry and business, since it can be utilized as a part of an extensive variety of uses for observing and control purposes. It is normal that M2M advancements when joined with PDAs will end up noticeably necessary components in savvy homes. In like manner, in this investigation, an example utilization of M2M advances is introduced. In the exhibited application, utilizing temperature information gave by sensors, the savvy aeration and cooling system consequently changes itself. In spite of the fact that the displayed application is only a basic case of how M2M can be utilized, it has the capability of influencing all regions enhancing our everyday life.

Deepika et al.(2016) exhibit a novel plan strategy for limiting the line at the power charging counters and to confine the utilization of power consequently, if the bill isn't paid. They are likewise proposing a framework that will decrease the loss of energy and income because of energy burglaries and other illicit exercises. The work framework embraces an absolutely new idea of "Prepaid Electricity". The IoT based idea is utilized so we can ceaselessly screen the utilization of energy (in watts) and on the off chance that it achieves the base sum, it would naturally caution the customer to revive through web.

KAA it is used for developing smart IoT applications. It enables exchange of data amongst all the connected devices and IoT cloud services too. Major services provided by this platform are fetching device information, device provisioning, and configuration of all the devices. With the help of SDK it provides security to the data, reliability, interoperability, and management of data. KAA SDK works on very small memory. It supports various databases such as MongoDB NoSQL, Hadoop and Cassandra. Pros: Applications supports Big Data base and NoSQL. Cons: Support of hardware model is very less.

Carriots it facilitates speedily building IoT applications, by drastically reducing time, expenses and difficulties. It facilitates device management remotely and controlling device remotely, action logging, activating custom made alarms, and exporting data. It supports RESTful API and update end user about the current status of the devices through sending Email, Twitter and Short Message Service. It supports data storage in Big Data and NoSQL. Pros: application which required triggering are mostly supported. Cons: Not user friendly.

SensorCloud it is used to procure, picturize, display, and investigate the received data from IoT devices. Basically, it is an outstanding platform which leverages powerful cloud computing facilities. It facilitates developers to perform complicated mathematical task on the captures data. Along with this it also provides Graphical User Interface (GUI) to process and analyse the data stored on cloud. It supports RESTful API for uploading data to the cloud. Pros: Huge devices accompanied in a network. Cons: difficult to support open source hardware.

III. ARCHITECTURE OF INTERNET OF THINGS

There is no single standard design for IoT exists, which is concurred ordinarily. The three-layer design characterizes the primary thought of the Internet of Things. It consists of three layers i.e. first one is the perception; second one is the network and third is the application layer, as described below:

The Perception Layer:

The first foremost layer works like human facial skin and five sense organs. Mainly this layer deals with the identifying things, collecting data from the things. It includes RFID tags, QR code, and various types of sensors, camera, terminals, wireless sensor network, 2-D bar code labels and readers. Main function of this layer is to identify the objects uniquely and collect the information.

The Network Layer:

The network layer works like the neural network and it is called as brain of loT. Mainly this layer deals with the transmitting and processing information received from perception layer. It consists of Internet network, information processing centre and intelligent information processing centre, network management centre etc. Main function of this layer is to process and forwards the data received from perception layer.

The Application Layer

The layer is liable for conveying application-explicit administrations to the end-client. It manages different applications where the IoT can be deployed, such as smart agriculture, smart homes, smart cars, and smart healthcare.

IV. APPLICATIONS OF INTERNET OF THINGS

The IoT presence is in everywhere because of its potential usefulness in almost all the possible domains. Out of this very small fraction of this services are currently available to the respective end users. Still there exist large number of domain, where new application usefulness is likely to come up and enhance the life of human. IoT applications are very vast so to accommodate under one diagram, therefor it must be categorized in to various domain such as described below:

Smart transport and logistics

Assisted Driving, Mobile Ticketing, Environment monitoring, Augmented maps.

Smart Health

Tracking, Identification Authentication, Data Collection, Sensing.

Smart environment

Home office monitoring, Industrial Plants, Smart Museum & Gym.

Personal and social

Social Networking, Historical Queries, Losses, Theft.

Futuristic

Self Driving Cars, City Information Model, Enhanced Game Rooms.

Predictive analytics information is time arrangement information gotten from the readings from a huge number of sensors spilled at high speed at real time. Sensor readings mirror the condition of the machine execution at that time. So the time arrangement information must be changed into a timestamped populace, furthermore, new inferred highlights must be characterized to discover what changes in the sensor readings over a timeframe give greatest data about the likelihood of failure of a machine. The information administrator segment in Predictive analytics empowers the client to make ultra-wide explanatory informational collections for a time-stamped populace, what's more, consequently concentrate a large number of inferred highlights that can be utilized for making strong models.

V. PROBLEM STATEMENT & RESEARCH CHALLENGES

This work focuses on the development of a data stream systems that have several properties. Firstly, IoT applications often require real-time processing of high-volume stream data. The processing logic should satisfy the requirements of IoT applications in various domains. Secondly, IoT scenarios require a decoupling of data consumers and data producers. We assume that data streams from different sensors are available to multiple applications. Essentially it should be possible for a sensor's data to be shared. Any application should be able to connect or disconnect to any desired stream data at any time. Furthermore, it should be possible to share the data such that the data is only sent once and ideally if two applications require the same subset of a data stream then the subset only has to be generated once. Finally, IoT will be pervasive in the near future. IoT application developers should be able to specify the data required through an easy to understand interface.

The primary contribution of this work is that a programming abstraction is provided through the use of an SQL-like syntax to be used by application programmers. An architecture of a platform is proposed that manages the data flows for application programmers that allows for sharing of a single data stream from a sensor.

Research challenges

Among the different challenges that the present IoT is facing, the two prime areas of concern are, efficiently processing IoT generated data and providing security to the IoT generated data. In the ongoing research, everybody is focusing on collecting and processing non-spatial IoT generated data somehow or somewhere, the issues regarding processing and securing IoT generated spatial data are side-lined. The main motivations behind choosing this topic are:

One of the requirements for the data processing in the future is the ability to digest petabytes of data.

Designing an efficient scalable framework for collecting and can able digest petabytes of IoT generated geo spatial data not only for the present internet but also for the future internet too. Need efficient and optimized indexing technique for IoT generated data. Analysing IoT generated data in real time for particular events or patterns of interest. With the introduction of new technologies, new architectures and over dependence of these technologies on the internet, create the situation for the security threats, which needs to be overcome. The Internet of Things (IoT) demands tailor-made security solutions.

VI. PROPOSED SYSTEM DESIGN

The framework is designed to be available via the Platform as a Service (PaaS) Model. The platform is licensed on a subscription basis and is hosted on a server. Sensor data can be accessed by a client IoT application using a TCP connection. Various kinds of IoT applications can be developed by leveraging this platform.

The platform provides two main services to its users: executes basic analytics on sensor data based on the user query and dispatches processed sensor data to applications who subscribe to this information. The platform gathers data generated by data sources and sends data to IoT applications. This platform assumes that there is a queue for each data stream coming from a data source. The Stream Manager maintains these queues.

Data Sources

The Data Sources refers to sensors attached to smart devices, which are used to sense the surrounding environment, periodically produce sensor data and send sensed data to the cloud platform. In general, the devices or things can be categorized.

Query Parser

This component is used for receiving and parsing query messages issued by user applications. Only after parsing the query message successfully, can the query can be executed by the query engine.

Stream Manager

This component maintains a list of queues, where each queue represents a data stream where the queue contains data from this stream. A data stream can be used by more than one query, and therefore we have multiple queues associated with this data stream with each belonging to a query. Since a query may need data from multiple streams, we associate multiple queues with each query.

Data Dispatcher

This component is aware of the information of all the active queries, including the streams that each query is interested in. Using this information, the sensor data can be dispatched to corresponding queues at the Stream Manager, so that each query has access to its requested stream data from the queues at the Stream Manager.

Query Processing Task

Each Query Processing Task is started for handling one query. As all data streams required by this query is handled by the Stream Manager, the task is able to continuously receive data from the corresponding queues. Upon receiving the data, the task will apply analysis to the streaming data including aggregate, calculate average and filter on the stream data according to user needs, and generate final query results.

Task Manager

This component is used to manage a group of Query Processing Tasks. There is one task per user query. The tasks can be started or removed based on user requirements.

Query Result Publisher

This component is responsible for publishing the query results on a query topic, the query topic is the query string. It allows users who subscribe to that query topic to receive the query results.

VII. RESULTS & DISCUSSION

Sensors are an extension between the physical world and the web. They will assume an always expanding part in pretty much every field comprehensible, and driving the "Internet of Things".

Potential Uses of Sensor Data

Sensors can be utilized to screen machines, framework, and environment, for example, ventilation hardware, spans, vitality meters, plane motors, temperature, lowliness, and so on. Utilization of this information is for prescient support, to repair or supplant the things before they break.

Types of Maintenance:

Corrective Maintenance is essentially settling things after they endure a breakdown and can likewise be called Reactive upkeep.

Preventive Maintenance is about supplanting or recharging consumables at planned interims.

Predictive Maintenance or Condition-construct upkeep centres in light of identifying failure before they happen. Prescient Maintenance joins investigations of the framework at foreordained interims to decide framework condition. Contingent upon the result of a ceaseless assessment, either a preventive or no support movement is performed. The computerized time arrangement in conjunction with an automated relapse calculation helps in issues identified with anticipating, for example, the remaining time before the machine or its parts break down at the end of the day, the time to next maintenance. The computerized grouping part empowers the client to discover hearty quantities of groups in the information and can be utilized for issues like gathering machines carrying on in a comparative mold and consequent irregularities in the sensor information. Variation of data generated by same sensor in just in a second can be easily seen in the graph.

VIII. CONCLUSION

The research addressed challenges in the development of a data stream system that supports real-time analytics but also allows for the decoupling of sensors from applications in a way that allows for sharing of sub stream computation. Integrating the data collection and analysis functions within a common platform is important for IoT application developers. Developers do not have to deal with heterogeneous and complex smart devices and the underlying network protocols to access sensor data. Users can specify the sub streams and analysis on sub streams without being concerned about efforts to transform the data to the desired format and apply analytics to data in order to extract valuable information. Instead, the platform provides an interface to access sensor data and carry out data analysis in the cloud on users' request. We reduce their development effort by not having these tasks done by each of these applications. From the perspective of overall IoT applications ecosystem, the design also has its advantage: the pattern has eased the overall communication to a large extent. In a direct communication model, one user application should talk to multitudes of sensors. On the other hand, one sensor should maintain communication with several users. With our platform, only one connection is needed for each sensor multiple user applications. The IoT requires custom-made security solutions. Today, there exist number of proposals to meet and tackle the attacks from outsider 's context.

REFERENCES

1. Ahmed A. and J. F. Naughton (2004). Static Optimization of Conjunctive Queries with Sliding Windows over Infinite Streams, ACM SIGMOD int. conf. on Management of data, Paris, pp. 419–430, France.

2. Albert, B. (2013). Mining big data in real time, Informatica (Slovenia), Vol. 37, No. 1, pp. 15–20.
3. Amini, T., Y. Wah and H. Saboohi (2014). On density-based data streams clustering algorithms: a survey, J. Comput. Sci. Technol., Vol. 29, No. 1, pp. 116–141.
4. Arun, K., K. Vamshikrishna, V. Kaladhar and G. V. Prabhakara Rao (2012) context aware scheduler for Hadoop, Int. Conf. on Advances in Computing, Communications and Informatics, Chennai, pp. 2-61, India.
5. Baskiyar, S. and R. Aamid Kunal (2010). Energy aware DAG scheduling on heterogeneous systems, j. Cluster Comput., Vol. 13, No. 4, pp. 373–383.
6. Benkhelifa, E., M. Abdel-Maguid, S. Ewenike and D. Heatley (2014). The Internet of Things: The eco-system for sustainable growth. IEEE/ACS 11th Int. Conf. on Computer Systems and Applications, Doha, pp. 836-842, Qatar.
7. Bhandarkar, M. (2010). MapReduce programming with apache Hadoop, IEEE Int. Sym. on Parallel and Distributed Processing, Oak Ridge, pp. 1-5, USA.
8. Charu, C., K. Aggarwal, J. Han, J. Wang and Philip S. Yu (2006). A framework for on demand classification of evolving data streams, IEEE Trans. Knowl. Data Eng., Vol. 18, No. 5, pp.577–589.
9. Cugola G. and A. Margara (2012). Processing flows of information: from data stream to complex event processing, ACM Computing Surveys, Vol. 44, No.3, pp. 53-62.
10. Daoud, M.I. and N. Kharma (2011). A hybrid heuristic-genetic algorithm for task scheduling in heterogeneous processor networks, J. Parall. Distr. Comput., Vol. 71, No. 11, pp. 1518–1531.
11. Deepak, P., N. Surya, R. Rajiv and C. Jinjun (2016). A dynamic prime number based efficient security mechanism for big sensing data streams, J. of Computer and System Sciences, Vol. 83, No. 1, pp. 22-42.
12. Demirkan, D. (2013). Leveraging the Capabilities of Service-oriented Decision Support Systems: Putting Analytics and Big Data in Cloud, Int. J. Decision Support Systems, Vol. 55, No. 1, pp. 412–421.
13. Li, J., Bao, Z. and Z. Li (2015), Modeling Demand Response Capability by Internet Data Centers Processing Batch Computing Jobs, IEEE Trans. on Smart Grid, Vol. 6, No. 2, pp. 737–747.
14. Liu, X., N. Iftikhar and X. Xie (2014). Survey of Real-Time Processing Systems for Big Data, 18th Int. Database Engineering and Applications Symposium, New York, pp. 356–361, USA.
15. Pedro, D. and Geoff Hulten (2000). Mining high-speed data streams. ACM int. conf. on Knowledge discovery and data mining, Massachusetts, pp. 71–80, USA.

AUTHORS PROFILE

Sandeep Singh Rajpoot is Bachelor of Computer Application from Dr. Hari Singh Gour University Sagar 2011, & Master of Computer Applications from RGPV, Bhopal. & presently a Ph.D. research scholar from Computer Application from Dr. APJ Abdul Kalam University, Indore, India.

Dr. Dhanraj Verma, Professor in Department of Computer Science & Engineering, College of Engineering, Dr. APJ Abdul Kalam University, Indore, India. He is a member of IEEE & IEEE Computer Society Since 2012, Life member of the CSI since 2012, He has published more than 24 Research paper National / international journals including. His main research work focuses on Network Security, Cloud Security and Privacy, Big Data Analytics, Data Mining, IoT and Computational Intelligence based education. He has 18 years of teaching experience and 6 years of Research Experience.