Analysis of Data Stream Processing At Edge Layer for Internet of Things

Dr. Pasumponpandian,
Professor, Computer Science Engineering,
KGiSL Institute of Technology, Coimbatore, India.
Email id: pasumponpandian32@gmail.com

Abstract: The progress of internet of things at a rapid pace and simultaneous development of the technologies and the processing capabilities has paved way for the development of decentralized systems that are relying on cloud services. Though the decentralized systems are founded on cloud complexities still prevail in transferring all the information's that are been sensed through the IOT devices to the cloud. This because of the huge streams of information's gathered by certain applications and the expectation to have a timely response, incurring minimized delay, computing energy and enhanced reliability. So this kind of decentralization has led to the development of middle layer between the cloud and the IOT, and was termed as the Edge layer, meaning bringing down the service of the cloud to the user edge. The paper puts forth the analysis of the data stream processing in the edge layer taking in the complexities involved in the computing the data streams of IOT in an edge layer and puts forth the real time analytics in the edge layer to examine the data streams of the internet of things offering a data- driven insight for parking system in the smart cities.

Keywords: Smart Cities, Internet of Things, Cloud Services, Edge Layer Computation, Real Time Analytics

1. Introduction

The emergence of the internet of things has subsidized the rapid and the constant progress of the devices utilized along with the ubiquitous instruments in the personal, business and as well as in the public spaces, producing, computing and consuming the data. Some of the major examples of the Internet of things are the developing smart cities, the industries such as medical care, large scale manufacturing factories and automotive industries etc. Managing all these areas with the assistance of internet of things involves huge set of devices employed in monitoring the changes and the prevailing circumstances, causing a very big data to be generated. Mostly the data is generated are usually in the form of sequential data sets, for instance in the form of data streams. This necessitates the need for the software solutions that are able to compute the data in streams instead of processing in the data in the batches

As the Internet of things and its associated smart systems are intrinsically unpredictable, the count of devices that are entering and exiting the systems and the volumes of data produced by the each system are

26

rapidly changing. More over the big data generated from each IOT component, holds knowledge that would be helpful for the future development of a concern organization or a city. All the insights produced might not possess equal value some data might be of higher value and require to be processed immediately as it would lose its value if processed later , for e.g. intimations of theft in houses , health care intimations of the serious condition patients from the rural areas. Such data requires faster processing providing insights within milliseconds form the time the data was generated. So the batch processing become incompatible and the stream processing for the data was initiated.

The stream processing is a big data technology that is used to probe constant data streams and identify the state of affairs, at a higher pace; within a small duration from the time data was triggered usually the insight identification period ranges from few milliseconds to minutes. The following statement explains the need for the stream processing. The primary reason why stream processing is preferred over batch processing is the batch processing stores the gathered data and then computes it later. Unlike the batch processing stream processing computes data naturally detecting the patterns and invigilating the outcomes, providing a manifold side by side focus and views at the data from the multiple streams simultaneously, The computation in the stream processing is simultaneously performed with less hardware facilities than the batch processing, and further the stream processing handles the large fire horse style data and reserves only the bits that are useful.

Initially the stream processing to provide insights from data was handled by the cloud for the IOT devices, but this further incurred more time to process as each time the information's where sent to a remote cloud and processed and received back resulting in heightened latencies. So the processing the information's at the edge would reduce the latency, communication overhead and the cost of processing at a considerable rate. The figure.1 below shows the operations of batch processing.

The paper analyses the stream processing in the edge layer taking in the complexities involved in the computing the data streams of IoT in an edge layer and put forth the real time analytics in the edge layer to examine the data streams of the internet of things offering a data- driven insight for parking system in the smart cities. The paper is arranged with the related works in the.2 presenting the need for the edge layer stream processing and the processing strategies available. The analysis with the edge layer based data streaming in section 3. The real time analytics in the edge layer to examine the data streams of the internet of things offering a data- driven insight for parking system in the smart cities in section 4, analyses results in section.5 and conclusion in section.6
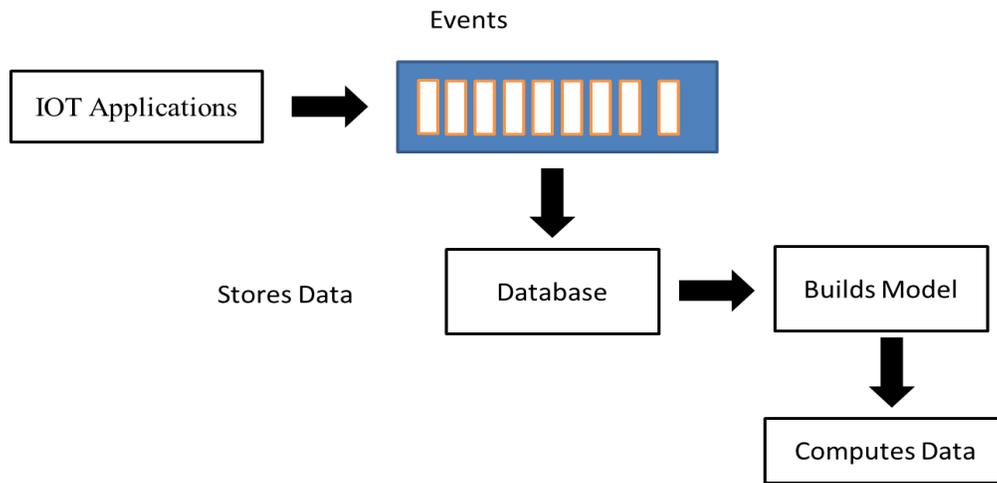
I-SMAC

Figure.1 Batch Processing

2. Literature Survey

Kolozali, et al [1] "is a strategy for performing a real time internet of things data stream annotation and processing utilizing the knowledge based approach to support the dynamic integration through the into the web , in order to improve the sustainability and the operational efficiency". Lomotey et al [2] "the author proffers a data streaming process by employing wearable architecture for performing the data streaming process in internet of things delivering the traceability of data routes form the source of origination to the health information system. The method in the paper utilizes the petri nets models that are enhanced to assist the generation of the route for the transparent data, tracking and detecting, making itself compatible for the IOT data from health care's transparency, connectivity, traceability, scalability and dis-connectivity".

De Francisci et al [3] "the tutorial gives the introduction to the mining of big IOT data streams and presents the data stream learners for classification, regression, clustering and frequent pattern mining followed with the further stages dealing with the scalability issues inherent in IOT applications and explains the strategies to mine data streams on the distributed engines such as spark, Flink, Storm and Samza."

Yasumoto et al [4] the paper presents "the survey on the emerging technologies towards the real time utilization of IOT data streams in terms of networking, processing, and proffers a new infrastructure for the IOT data streams called information flow of things that process analyses and analyses massive IOT data streams in the real –time based on the distributed processing among the IOT devices". Puschmann et al [5]

I-SMAC

the method adapts the adaptive clustering strategy to determine the various clusters that could be found distributed data stream , once the clusters are selected the online clustering method is selected to cluster the incoming streams of data.

Yang, et al [6], the paper is about the utilization of the Fog in processing the data stream instead of the cloud architecture in order to bring down the latencies, resilience and the cost. Hesse et al [7], "the paper is about the "conceptual survey of the data stream processing systems". Hochreiner et al [8], present a "VISP: An ecosystem for elastic data stream processing for the internet of things." Manate et al [9] " the paper handles the problem through the architecture, handling storage, analysis and processing of large amounts of data and can scale accordingly providing a corner stone for the internet of things"

de Assuncao et al [10] the paper is the " survey on the data stream processing engines and mechanism exploiting the resource elasticity features of cloud computing in stream processing and discusses the challenges in the solutions proposed in the literature to address the same". Stephen, et al [11] the author proffers "STYX a new programming abstraction and managed runtime system that ensures confidentially of the IOT applications while leveraging the public cloud for simultaneous query processing. The fundamental concepts are intelligently utilizes the partial homomorphic encryption to perform as many computational intensive operations as possible in the untrusted cloud."

Duraipandian, et al [12] presents the "Cloud based internet of things for smart connected objects." in Bashar, et al [13] proposes the "Intelligent development of big data analytics for manufacturing industry in cloud computing." , the author Bestak, et al [14] in his paper "big data analytics for smart cloud-fog based applications." Sivaganesan, D et al [15] "Design and development AI-enabled edge computing for intelligent-IoT applications" Bhalaji, N et al [16] "Delay diminished efficient task scheduling and allocation for heterogeneous cloud environment." Abul et al [17] "Secure and cost efficient implementation of the mobile computing using offloading technique."

3. Analyses of Data Stream Processing in the Edge Layer

The section presenting the analyses of the data stream processing in edge layer initially begins with the study on the steps and the challenges faced while shifting the data stream processing from the cloud to the edge devices, The edge computing layer brings the services of the cloud to its user edge, it is equipped with the computing devices of smaller size to large sized computing devices such as the edge servers, mini clouds, fog nodes, gateways etc. and handles the data processing at a very low latency eluding the necessity to transmit the information's gathered from the IOT devices to the cloud enabling a decision to be made

I-SMAC

within a short period of time for the applications that are critical with the higher user mobility. More over enhances the security of the system by improving the trust levels of the Internet of things devices.

The edge computing was actually developed to lighten the processing burden caused to the cloud using the internet of things. This done by offloading the data processing to the device in the edge closer to the location initially the data was triggered from. The explanation below provides the general phases in stream processing the data in the edge layer. The architecture below shows the flow process involved in the steam processing of the information's obtained through the IOT devices in the edge layer.
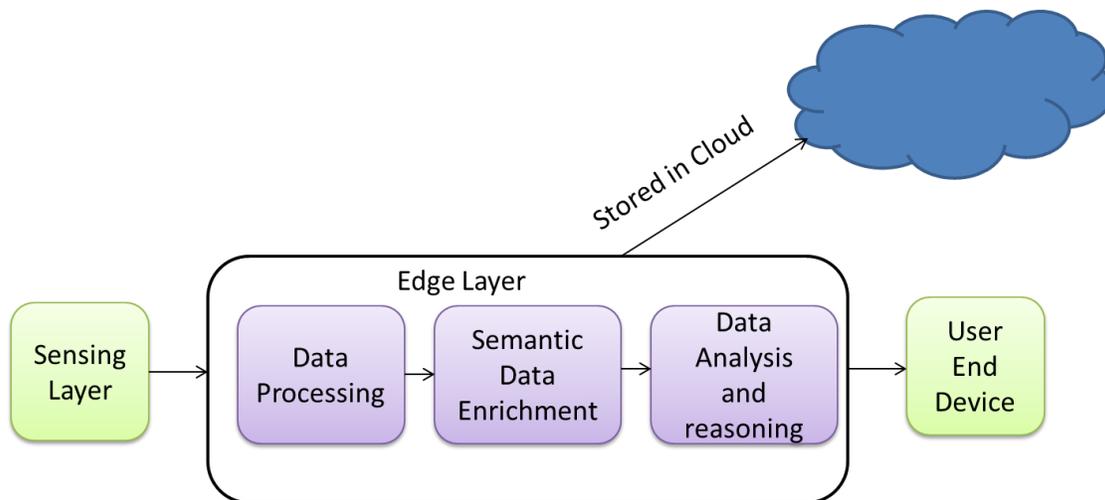

Figure.2 General System Framework

The process is segregated into four phases as (i) monitoring phase- minute notable changes sensed through the IOT devices, (ii) Edge layer computation , - where the data are preprocessed and the followed by the semantic data enrichment, (iii) does the analysis and the reasoning (iv) conveys the information to user application layer.

Monitoring Phase: This is very lowest layer in the internet of things paradigm, as it holds all the devices required in sensing the changes and conveys, the information to the edge layer, this layer takes the most important part in deciding the throughput characteristics of the system such that a rate of data generation is fixed. To over the complexities in the processing the generated data and ensure it has been properly processed it is important to take into consideration the rate of data generation during the evaluation.

I-SMAC

Edge layer Computation: this phase is further divided into two phases as data preprocessing and semantic data enrichment, the data processing eludes the redundant data and removes noises found in the original data and sends it to next layer where the semantic enrichment is done. The semantic enrichment " add values to the giving meaning to it , allows more efficient reasoning and mining , links the content of different IOT devices and allows a complex event processing. The utilization of the resource description framework enables the data to appear in the machine readable form making the reasoning and the rule based engine to automatically run on it and the use of metadata facilitates mining, links the data of different data streams, making possible the correlation allowing a complex event processing. The correlation plays a vital role in processing the IOT data stream as without correlation the events would appear independent of each other and at different times, so correlation allows easy computation of events with the higher complexities". Further the edge layer identifies the event, or the patterns, filers the errors and minimizes the data stream rate for the measurements this causes one to reduce the unnecessary amount of data.

Analysis and Reasoning: This process enables the real time and linked processing enhancing the reasoning latency for the large data set by offering a distributed reasoning. As the internet of things would cause more challenges in reasoning, as the reasoning are liable of occurring in any phase of the delivery process.

4. Real Time Analytics of Parking System

The real time analytics for the parking systems examining the data streams of the internet of things to avail data- driven insights are developed using the raspberry pi 3 B+ integration, an Ubuntu mate operating system, and further utilizes the IBM Node-Red Server to keep not on the outcomes in the desktop as well as the mobile view.

The parking system enables the drivers to identify the available parking spaces, by integrating the smart parking in to the city frame work minimizing the parking search time, emission as well as the fuel consumption along with the traffic congestion. The application holds the data streams generated in the real time, whenever a driver requests for the parking slot, immediately the information's are processed in the edge layer, the information's are gathered. The raspberry pi 3 B+ supported by the MQTT protocol message queuing telemetry transport to execute the edge processing and transferring of data to the server through the MQTT broker. The simple file transfer protocol was used in the proposed method to transfer the information to the raspberry pi 3 B+, processed using the free open source distributed real time processing system that has a reliably unbounded streams of data. The working resembles the Hadoop framework for the batch processing (apache storm). The processed information with the semantic enrichment is analyzed employing the C-SPARQL for the primary streaming and utilizes the reasoner to deduce the knowledge from the data and the ontology constructed over the rules that are predefined. The

I-SMAC

proposed method utilizes the commonly used reasoning and the inference engines such as the Racer Pro, pellet, Fact++, Rif4J, Hermit and Jena to support the various rules, ontologies and the OWL. The Flow diagram in figure .3 shows the orchestration of data streaming of the parking system in the edge layer.

Further the proposed method employs software's such as Active MQ to aid the message broker's by enabling them to have a data pipelines for the data streams of IOT using mechanism of the publish subscribe offering a heightened throughput and minimized latency and also employs the JSON parser to parse the data, Scikit-mulitflow to stream ML library, Python to deal with the structured data's that are flowing in identifying the events and the patterns, rethink DB to constantly process the queries coming in to have the outcomes real time using the ReQL query language. Superset and the Grafana are engaged to gain insights and the monitor the incoming data of the IOT devices. The system was finally orchestrated with the Apache Ambari and the Zookeeper and installed with the Wazuh to elude the security threats
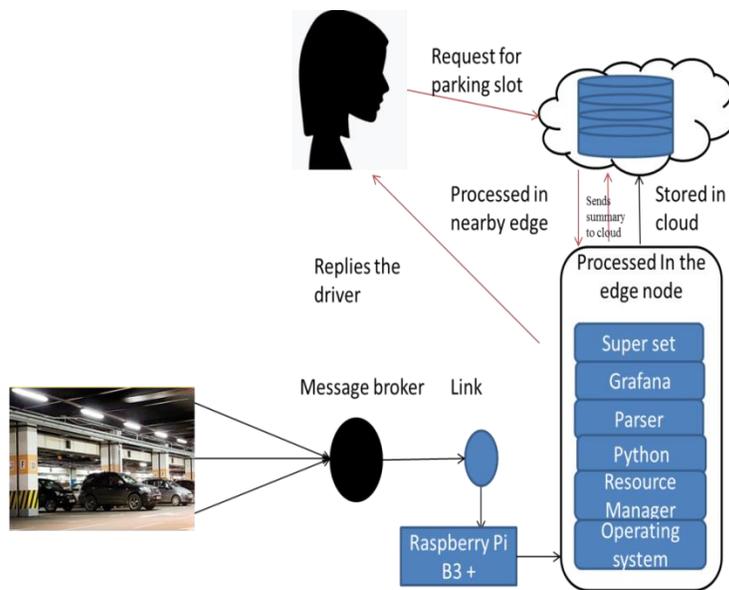
Proposed Parking System Data Streaming in Edge

5. Performance Analysis

The proposed method that is deployed over the edge layer saves the particulars of the parking slots every five seconds, and the data streams $\{D_1, D_2, \ldots \ldots D_n\}$ obtained are sent to the edge layer at proximal, The data fields of the each data received in the data stream is depicted in the table.1 below.

Data Fields	Attribute	Description
Particulars of the Parking Area	Name of the spot	Name of the parking slot
	Latitude- Identity	Latitude description
	Longitude-Identity	Longitude description
The Parking Particulars	Vehicle ID	Number allotted to vehicle
	Entry time	The time the vehicle entered the slot
	Total length	The duration of parking
	Spot ID	The identification for the each slot

Table.1 The Data Field Description

The data obtained are preprocessed enriched with the semantic data, filtering the unwanted data , the data on the edge layer is assigned with the labels and computed to determine the entry timing , the total parking time and the exit time and the represents the status of the parking slot. The copy of status is sent to the cloud for storage.

The proposed architecture was evaluated monitoring the data streams latencies; the latency was determined by gathering samples every 5 minutes and registering the arrival time slots of the datasets at the edge. The figure.4 below shows the latency observed while data streaming in the edge and is compared with the usual latencies that occur in the cloud.
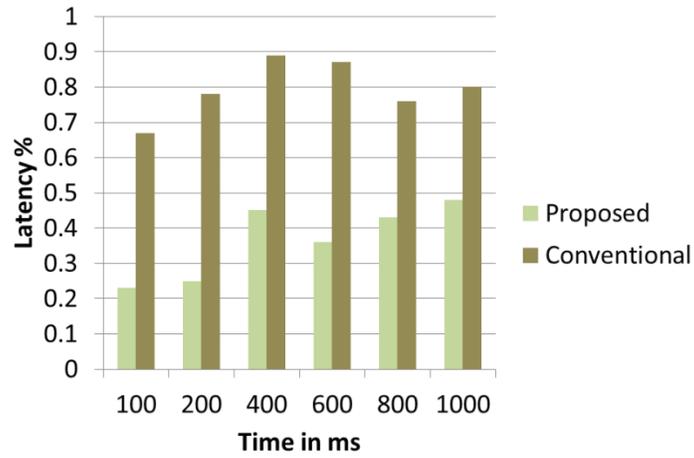
Figure.4 Latency Observed

The latencies observed shows that the edge layer data streaming enables to have an improved latency compared to the processing done in the cloud layer. Thefigure.5below shows the memory consumption over view of the proposed system for different datasets observed over time.
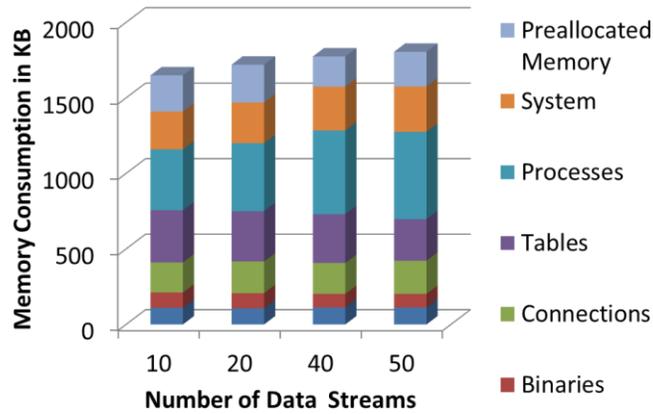

Figure.5 Memory Consumption

34

I-SMAC

It was noted that the total amount of memory used was less than 2000 KB including the memory that was allotted for the binaries, process, systems and the connection tables. The figure.6presents the data streaming rate execution time.
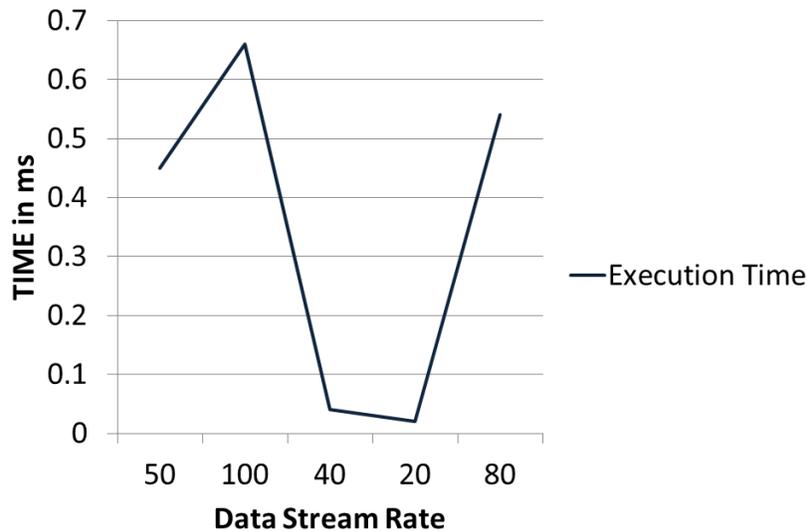

Figure.6 Data Rate Execution Time

The results shows that though there is an increase in the data stream rate, the execution time increases the continuous increase of the data rate crashes the system after 4-5 execution cycles due to the enormous workload and the thermal problems.

6. Conclusion

The paper discusses the system overview of the general edge layer data streaming and proposes the parking system for the smart city by proliferating the edge layer data streaming for the parking slot availability, to have an immediate and latency reduced responses that are processed within few microseconds. The analyses of the proposed system with the semantic data augmentation showed that the proposed system has an improved latency, memory consumption and execution time and also reduces the processing burden of the prevailing cloud paradigm. In future the paper

I-SMAC

is planned to extend propelling of the edge computing using the 5G technologies to attain sustainability and enhance the quality.

References

[1]   Kolozali, Sefki, Maria Bermudez-Edo, Daniel Puschmann, Frieder Ganz, and Payam Barnaghi. "A knowledge-based approach for real-time iot data stream annotation and processing." In *2014 IEEE International Conference on Internet of Things (iThings), and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom)*, pp. 215-222. IEEE, 2014.

[2]   Lomotey, Richard K., Joseph Pry, and Sumanth Sriramoju. "Wearable IoT data stream traceability in a distributed health information system." *Pervasive and Mobile Computing* 40 (2017): 692-707.

[3]   De Francisci Morales, Gianmarco, Albert Bifet, Latifur Khan, Joao Gama, and Wei Fan. "Iot big data stream mining." In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 2119-2120. 2016.

[4]   Yasumoto, Keiichi, Hirozumi Yamaguchi, and Hiroshi Shigeno. "Survey of real-time processing technologies of iot data streams." *Journal of Information Processing* 24, no. 2 (2016): 195-202.

[5]   Puschmann, Daniel, Payam Barnaghi, and Rahim Tafazolli. "Adaptive clustering for dynamic IoT data streams." *IEEE Internet of Things Journal* 4, no. 1 (2016): 64-74.

[6]   Yang, Shusen. "IoT stream processing and analytics in the fog." *IEEE Communications Magazine* 55, no. 8 (2017): 21-27.

[7]   Hesse, Guenter, and Martin Lorenz. "Conceptual survey on data stream processing systems." In *2015 IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS)*, pp. 797-802. IEEE, 2015.

[8]   Hochreiner, Christoph, Michael Vogler, Philipp Waibel, and Schahram Dustdar. "VISP: An ecosystem for elastic data stream processing for the internet of things." In *2016 IEEE 20th International Enterprise Distributed Object Computing Conference (EDOC)*, pp. 1-11. IEEE, 2016.

[9]   Manate, Bogdan, Victor Ion Munteanu, and Teodor-Florin Fortis. "Towards a scalable multi-agent architecture for managing iot data." In *2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, pp. 270-275. IEEE, 2013.

[10]   de Assuncao, Marcos Dias, Alexandre da Silva Veith, and Rajkumar Buyya. "Distributed data stream processing and edge computing: A survey on resource elasticity and future directions." *Journal of Network and Computer Applications* 103 (2018): 1-17.

I-SMAC

[11]   Stephen, Julian James, Savvas Savvides, Vinaitheerthan Sundaram, Masoud Saeida Ardekani, and Patrick Eugster. "STYX: stream processing with trustworthy cloud-based execution." In *Proceedings of the Seventh ACM Symposium on Cloud Computing*, pp. 348-360. 2016.

[12]   Duraipandian, M., and Mr R. Vinothkanna (2019). Cloud based internet of things for smart connected objects. Journal of ISMAC, 1(02), 111-119.

[13]   Bashar, Abul. "Intelligent development of big data analytics for manufacturing industry in cloud computing." *Journal: Journal of Ubiquitous Computing and Communication Technologies September* 2019, no. 01 (2019): 13-22.

[14]   Bestak, Robert, and S. Smys. "Big data analytics for smart cloud-fog based applications." *Journal of trends in Computer Science and Smart technology (TCSST)* 1, no. 02 (2019): 74-83.

[15]   Sivaganesan, D. "Design and development AI-enabled edge computing for intelligent-iot applications." *Journal of trends in Computer Science and Smart technology (TCSST)* 1, no. 02 (2019): 84-94.

[16]   Bhalaji, N. "Delay diminished efficient task scheduling and allocation for heterogeneous cloud environment." *Journal of trends in Computer Science and Smart technology (TCSST)* 1, no. 01 (2019): 51-62.

[17]   Bashar, Abul. "Secure and cost efficient implementation of the mobile computing using offloading technique." *Journal of Information Technology* 1, no. 01 (2019): 48-57.

I-SMAC