Cloud Computing using Internet of Thinks (IoT) in Big Data Storage

K. Thamizhmaran
Assistant Professor, Department of Electronics and Communication Engg., Government College of Engineering, Bodinayakkanur, Theni (deputed from Annamalai University), Tamilndu, India

Corresponding Author
E-Mail Id: tamil10_happy@rediff.com

ABSTRACT
Applications for the Internet of Things (IoT) have become a significant area of study for researchers and engineers alike, particularly data analysts. This is because of the volume and influence of data-related issues that need to be resolved in high-level corporate organizations, particularly in the cloud. It offers a functional framework that delineates the domains of IoT big data collecting, management, processing, and mining. Additionally, a number of related technical modules are defined and characterized based on their salient features and capacities. IoT-related research is being processed at the moment. Additionally, the potential and difficulties related to IoT big data research are noted. Studying important IoT application publications and research topics based on relevant scholarly and commercial publications are also aided by it. It raises issues, some common ones being provided inside the suggested framework for IoT-related research.

Key words: *IoT applications, big data, cloud computing, data management*

INTRODUCTION

Large-scale networked sensors have created large volumes of data, making the acquisition, integration, storing, processing, and utilization of this data a critical and pressing issue for businesses looking to meet their objectives. Consequently, managing vast amounts of heterogeneous data in widely dispersed settings presents a problem for researchers and engineers alike, particularly in cloud platforms. The features of IoT data in cloud platforms can be summed up as follows: Large-scale dynamic data, high heterogeneity data from multiple sources, low-level data with poor semantics, and inaccurate data. The perception layer, network layer, and application layer make up a typical Internet of Things application structure. Because it is made up of middle wares and business models, the application layer in cloud computing is very important for IoT-based storage systems.

Much work has been finished to empower compelling and smart information handling and examination in application layer in light of distributed computing. Radio Frequency Identification (RFID), the Wireless Sensor Network (WSN), and various Smart Things are components of the front-end layer. In view of the handling system of IoT application, a structure of IoT-based information capacity frameworks in distributed computing is given.

The system comprises of a few modules, which are information securing and incorporation, information capacity, information the executives, information handling, information mining and application improvement module. Information Securing and Joining Module, Information Stockpiling Module, Information The executives Module, Information Handling Module, Information Mining Module.[1-5]

METHODS AND CHALLENGES

Information Procurement and Coordination Module[6-9]

The purpose of the data acquisition and integration module is to collect data from a variety of sensors, including RFID, ZigBee, GPS, temperature sensors, and others. When designers of Internet of Things applications need to integrate massive amounts of structured, semi-structured, and unstructured data, heterogeneous information poses a significant challenge. According to the viewpoint of information handling in figuring, we can characterize the three primary techniques during the time spent information procurement and coordination: information portrayal models, multi-source information combination, and information transmission and correspondence. Information Portrayal Models: Information Portrayal models are utilized for IoT information procurement and mix in a general sense. There exist a few unique kinds of sensor gadgets, like messages, occasions, pictures, recordings, status information, and so forth. Information portrayal models ought to be planned in view of various application purposes with an adaptable and normal configuration.

Data Storage Module

In IoT applications, the monstrous information from sensors consume huge extra room. Meanwhile, because that different roles and tenants require different service and security levels, data should be isolated for various requirements. Therefore, how to share and isolate these data in cloud platform are the main challenges in IoT data storage.

Data Storage Types

In the aspect of data storage in cloud platform, existing works can be classified into several types as: RDBMS, NOSQL DBMS, DBMS based on HDFS, Main-Memory DBMS and Graph DBMS.

RDBMS: Many structured data storage platforms are based on RDBMS. The relationships between massive amounts of data are always essential for a multi-tenant data storage system, even though they are generated rapidly and in a variety of ways [16]. Then different traditional relational data with virtual relational data is combined in a single schema, but exports a unified data access view to act as a multi-tenant database for different tenants. There is an approach of RDBMS called Ultrawrap, which encodes a logical representation of each RDBMS as an RDF graph, and uses SPARQL queries to get the data on the existing relational stored views.

NoSQL DBMS: Unlike RDBMS, NoSQL DBMS s- tores and manages unstructured data in a key-value model. The NoSQL DB is free in schema structure. It can provide some properties, such as horizontal scalability, distributed storage, dynamically schema, etc. On the other hand, NoSQL DB is not good at keeping Atomicity, Consistency, Isolation, and Durability of data. Besides, it cannot support well for some crowd sensing applications is presented to combine different kinds of protocols for different types of data. Three sorts of conventions like HL7 for clinical information, BACnet for building checking, and Perception and Estimation model for ecological information, are coordinated into one information model to oversee sensors and activities. Attempt to work in a diverse network; an original methodology is proposed to screen a commonplace plastic industry climate with WSN, Administrations and Google Devices. It utilizes miniature infusion to utilize a heterogeneous organization of remote sensor hubs and these sensor hubs send climate information, for example, material capacity conditions, encompassing lab temperature and stickiness. In IoT assets are coordinated as an original programmed asset type on the business cycle layer in order to oblige more changes in future venture conditions. The research solves the problem of the integration of the IoT paradigm and its devices coming with native software components as resources. Thing Broke aims to integrate IoT objects with different characteristics, protocols, interfaces and constraints while maintaining the simplicity and flexibility required for a variety of applications. Thing Broker provides a uniform access interface to different IoT objects. Using a single abstraction to represent IoT objects with their own configurable attributes, Thing Broker involves all sorts of objects, from physical sensors to high-level services.

Data Transmission and Communication[10-13]

Once data acquisition from different sensors is finished, data transmission and communication should be carried out so as to transmit data to the back-end or to communicate with other sensors for business purpose. There are many protocols for data transmission and communication, such as UDP and TCP. Developers often use User Data Protocol (UDP) to transmit multimedia data due to its real-time characteristic. But when network congestion and channel noise occur, packet losses happen easily by UDP protocol. To address this issue, a brand-new real-time multimedia transmission protocol over UDP known as Control over UDP (CoUDP) [13] has been developed. The performance of CoUDP protocol is better than UDP and TCP because it adds rate control and fast retransmission mechanism over UDP application and gives up redundant feedback like TCP. In a layered fault management scheme with fault managing program control and separate layer functions is designed to ensure the reliability of end-to-end transmission for IoT applications. This proposed plot suits well to the IoT necessities. They implement integrated evaluation and prediction of the possible fault through the use of fuzzy cognitive maps, which may offer a solution to the issue that current relative algorithms are not suitable for complex conditions. In a proximity-based authentication approach is proposed to utilize the wireless communication interface.

Data Storage Module

In IoT applications, the monstrous information from sensors consumes huge extra room. In the meantime, data ought to be isolated for various requirements because distinct roles and tenants necessitate distinct service and security levels. Thusly, how to share and separate this information in cloud stage are the primary difficulties in IoT information capacity.

Data Storage Types

In the aspect of data storage in cloud platform, existing works can be classified into several types as: RDBMS, NOSQL DBMS, DBMS based on HDFS, Main-Memory DBMS and Graph DBMS.

RDBMS: Many organized information capacity stages depend on RDBMS. The relationships between massive amounts of data are always essential for a multi-tenant data storage system, even though they are generated rapidly and in a variety of ways [16]. Then different traditional relational data with virtual relational data is combined in a single schema, but exports a unified data access view to act as a multi-tenant database for different tenants. There is an approach of RDBMS called Ultrawrap, which encodes a logical representation of each RDBMS as an RDF graph, and uses SPARQL queries to get the data on the existing relational stored views.

NoSQL DBMS: Unlike RDBMS, NoSQL DBMS s- tores and manages unstructured data in a key-value model. The NoSQL DB is free in schema structure. It can provide some properties, such as horizontal scalability, distributed storage, dynamically schema, etc. On the other hand, NoSQL DB is not good at keeping Atomicity, Consistency, Isolation, and Durability of data. Besides, it cannot support well for some distributed queries. Besides, on the aspect of database access, some work has been done on the potential integration of an ontology-based date access approach in NoSQL stores. It is a new data management style that exploits the semantic information represented in on topologies when searching the IoT data stores. In this way, some connected work has attempted to coordinate the element of circulated document framework to the IoT information stores.

Data Storage Module

In IoT applications, the monstrous information from sensors consume huge extra room. In the meantime, data ought to be isolated for various requirements because distinct roles and tenants necessitate distinct service and security levels. Thusly, how to share and separate these information in cloud stage are the primary difficulties in IoT information capacity.

Data Storage Types

In the aspect of data storage in cloud platform, existing works can be classified into several types as: RDBMS, NOSQL DBMS, DBMS based on HDFS, Main-Memory DBMS Graph DBMS.

RDBMS: Many organized information capacity stages depend on RDBMS. The relationships between massive amounts of data are always essential for a multi-tenant data storage system, even though they are generated rapidly and in a variety of ways [16]. Then different traditional relational data with virtual relational data is combined in a single schema, but exports a unified data access view to act as a multitenant database for different tenants. There is an approach of RDBMS called Ultrawrap, which encodes a logical representation of each RDBMS as an RDF graph, and uses SPARQL queries to get the data on the existing relational stored views.

NoSQL DBMS: Unlike RDBMS, NoSQL DBMS s- tores and manages unstructured data in a key-value model. The NoSQL DB is free in schema structure. It can provide some properties, such as horizontal scalability, distributed storage, dynamically schema, etc. On the other hand, NoSQL DB is not good at keeping Atomicity, Consistency, Isolation, and Durability of data. Besides, it cannot support well for some distributed queries. Besides, on the aspect of database access, some work has been done on the potential integration of an ontology-based date access approach in NoSQL stores. It is a new data management style that exploits the semantic information represented in on topologies when searching the IoT data stores.

DBMS integrated with HDFS

HDFS can likewise be stretched out to an exceptional conveyed document vault, which processes enormous unstructured records productively. In the IoT data stream, many of the data are generated in the XML format, making it difficult to deal with these small XML files with huge volumes. One methodology is to advance putting away and getting to monstrous little XML documents in HDFS. Little XML documents are converted into a bigger record to lessen the metadata at Name Hub, hence related component could be utilized to further develop the information store execution with the assistance of another focal ordering administration finding framework in light of Hadoop HBase information store the presentation of administration finding is expanded.

Main-Memory DBMS
For large-scale IoT applications, IO stream processing needs to be fast. Creator in executes a huge scope RFID application in the fundamental memory data set framework H2. Plus, it likewise gives a multi-faceted hash-based file configuration outline work and accomplishes a beat execution assessment. Specialists investigate the actual information base construction that acknowledges fast data set movement, which is a significant piece of the cloud information capacity and propose a few recuperation ways to deal with the transitory principal memory data sets.[15]

Graph DBMS
A database called Graph DBMS stores and represents data using graph structures with nodes, edges, and properties. A data set that utilizations chart structures for semantic questions with hubs, edges and properties to address and store information. With Chart DBMS, the relationship among sensor information can be overseen effectively. Gives a superior presentation Diagram DBMS the executives situation, supporting productive control of huge charts that comprises of enormous scope hubs and edges.[16]

RDF-based data storage
RDF is a semi-organized information model for web data assets the executives. RDF Outline (RDFS) gives a cosmology indicating language to gathering the assets into ideas and recognizing the relationship among these ideas. Cloud-based RDF Information the executives give a principled order of existing work on RDF information the board. Meaning to help various information kinds of IoT sensors, various information types ought to be consolidated in order to acknowledge powerful information stockpiling.[17]

Table 1: *Comparison between Data Storage Types.*

Features	RDBMS	NOSQL DBMS	DBMS integrated with HDFS	Main-Memory DBMS	Graph DBMS
Support for ACID	Not well	Yes	Yes	Common	Common
Support for semi-structured data and unstructured data	Not well	Yes	Yes	Yes	Use graph structures with nodes, edges
Support for structured Data	Yes	Not well	Yes	Yes	Use graph structures with nodes, edges
Support for scalability	Not well	Yes	Yes	Not well	Yes
Support for massive and distributed processing	Not well	Yes, but not flexible	Yes	Yes, but not flexible	Yes

Table 2: *Comparison between Data Indexing Methods.*

Features	Bitmap Index	Complex Data Structure Index	Inverted Index
Data structure	Bitmap	Tree, graph or others	Mapping from content to location
Suitable data characteristics	Values of a variable repeat frequently	Values of a variable repeat frequently	New key values monotonically increase, such as sequence numbers

Suitable scene	Analytical process, such as OLAP	Transaction process, such as OLTP	Large scale process, such as search engines
Performance	Less efficient	Efficient	More efficient
Cost	Less space	More space	Most space and increased processing
current usage status	Common	Common	Rising trend

Table 3: *Comparison between Data Mining Methods.*

Data Mining Dimensions	Goals	Shortcomings or Future Work
Parallel programming	Classification	Need more experiments in different environments
	Clustering	Need to verify the results about the efficiency and convergence
	Association Rules	Need to optimize the algorithm for smaller data sets or nodes
	Prediction	Only focus on one specific area
Mobile Computing	Frequent patterns, belongs to association rules	Lack of comparison experiments
	Data stream classification	Have some outliers to fix
	Survey on clustering problem	Focus on one specific area
Graph Mining	BSP-based Parallel Graph Mining (BPGM) based on BSP computing model	Limits on the scale of graph data
	cloud-based Spider Mine (c-Spider Mine) based on cloud computing	Examine more real big data sets and introduce more data mining algorithms
	Graph mining on Map Reduce	Improvements on efficiency

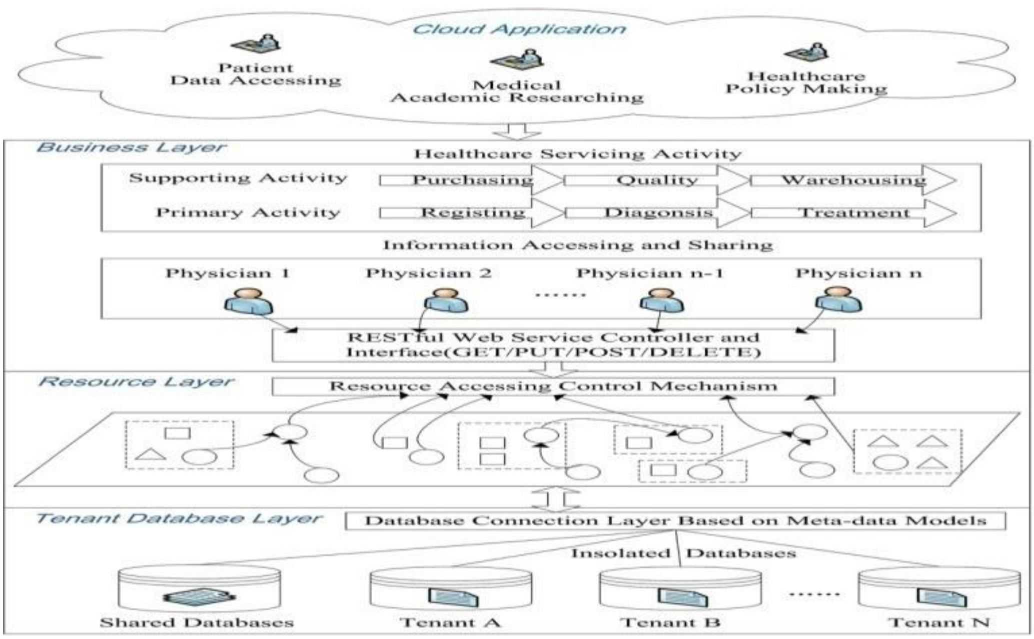

Fig. 1: *An Ontology-based CoT Platform.*

Table 4: *Weaknesses in map reduce and solving technology for IOT big data disposing.*

IoT Data features	Disposing Requirement	Weakness in Map Reduce	Improved Techniques
Distributed multi-source high heterogeneity data Huge scale dynamic data	Data access	Selective access to data	Indexing, Data Layout, International Data Placement
	Data communication	High communication cost	Partitioning, Colocation
	Process n-way operation from multiple sources	Lack of support for n-way operations	Additional MR phase, Redistribution of keys, Record duplication
	Work allocation	Load balancing	Pre-processing sampling, repartitioning, batching
	Real-time data processing	Lack of interactive or real-time processing	Streaming pipelining, in-memory processing, pre-computation
	Data stream disposing	Redundant and wasteful processing	Result sharing, Incremental Processing, Batch Processing of Queries, Result Materialization
Low-level with weak semantics data	Interactive query in data analysis tasks	Sequence execution lack of interaction	Looping, Caching, pipelining, Recursion, incremental processing
	Query Optimization	lacks of management and future reuse of results	Processing optimizations, parameter tuning, plan refinement, operator reordering, code analysis, dataflow optimization
Inaccuracy data	Exploratory queries	Lack of quick retrieval of approximate result	Reasoning based on formal expression such as ontology
	rank-aware processing such as top-k queries	Lack of early termination	Sorting, Sampling

CONCLUSION

This paper is an ideal exploration which outlines the current and potential IoT enormous information stockpiling frameworks in distributed computing and simultaneously reviews the condition of-craftsmanship in writing from the perspective on information handling process. The Internet of Things storage system makes it possible to follow important information about things as they travel through cloud platforms. It shows huge incentive for IoT applications by giving exact information on the ongoing IoT information handling, which brings about higher accessibility and adaptable asset arrangement. Information capacity framework supporting IoT gadgets can be used to further develop the whole information handling productivity and proposition enormous upper hand to the IoT applications. Contextual business scene, semantic annotation, and multi-device cooperation *et al.* demonstrated that semantic relationships among IoT data will increase global intelligent and inter-operational capabilities. IoT Information stockpiling frameworks will empower undertaking to obtain such ability.

REFERENCES

1. Ma, M., Wang, P., & Chu, C. H. (2013, August). Data management for internet of things: Challenges, approaches and opportunities. In *2013 IEEE International conference on green computing and communications and IEEE Internet of Things and IEEE cyber, physical and social computing* (pp. 1144-1151). IEEE.
2. Aazam, M., Khan, I., Alsaffar, A. A., & Huh, E. N. (2014, January). Cloud of Things: Integrating Internet of Things and cloud computing and the issues involved. In *Proceedings of 2014 11th International Bhurban Conference on Applied Sciences & Technology (IBCAST) Islamabad, Pakistan, 14th-18th January,*

2014 (pp. 414-419). IEEE.
3. Thamizhmaran, K. (2020). EE-APTSP: Evolution Node Life Time for WSN. *i-Manager's Journal on Wireless Communication Networks*, *8*(4), 27.
4. Thamizhmaran, K. (2020). Issues in wearable electronics devices for wireless sensor network. *I-Manager's Journal on Communication Engineering & Systems*, *9*(1).
5. Thamizhmaran, K., Lakshmi, S. J., & Hemalatha, M. (2018). Performance Comparison of ABR using EPKCH in MANET. *i-Manager's Journal on Information Technology*, *7*(2), 16.
6. Kayalvizhi, A., Thamizhmaran, K., & Ramprabhu, G. (2017). Implement PAPR Reduction in OFDM System Using SAS DCT with Companding. *International Journal of Advanced Research in Computer Science*, *8*(3).
7. Thamizhmaran, K. (1993). EEQRP-Energy Efficient Quality Routing Protocol for Wireless Sensor Networks. *Journal of Signal Processing*, *3*(1).
8. Thamizhmaran, K. (2018). Comparison of AODV and M-AODV IN MANET. *IOSR Journal of Engineering*, *8*(5), 11-14.
9. Thamizhmaran, K., & Girishkumar, G. (2019). ACO of energy efficient clustering routing method. *Journal of gujarat research society*, *21*(7), 66-71.
10. Thamizhmaran, K., & Akshayadevi, A. (2015). Survey on Cluster Head with Cluster Head Selection Techniques and algorithm for MANET. *Journal of Applied Sciences Research*, *11*(23), 43-46.
11. Arivazhagan, A., & Thamizhmaran, K. (2015). Performance comparison of proactive and reactive routing protocol under wormhole for MANET. *International journal of applied engineering research*, *10*(46), 32193-32198.
12. Prabu, K., & Thamizhmaran, K. (2016). Cluster Head Selection Techniques and Algorithm for Mobile Ad-hoc Networks (MANETS). *Advance Research in Computer science and software Engg*, *6*(7), 169-173.
13. Thamizhmaran, K. (2019). IOT-Fundamentals and Applications. *International Journal of Advance Research and Review*, *4*(3), 10-13.
14. Thamizhmaran, K. (2019). Security Attacks in Wireless Sensor Networks–A Study. *i-manager’s Journal on Information Technology*, *9*(1), 35-43.
15. Thamizhmaran, K. (2020). IOT supported security considerations for network. *WSEAS Transactions on Communications*, *19*, 113-123.
16. Dao, M. S., Pongpaichet, S., Jalali, L., Kim, K., Jain, R., & Zettsu, K. (2014, April). A real-time complex event discovery platform for cyber-physical-social systems. In *Proceedings of International Conference on Multimedia Retrieval* (pp. 201-208).