DOR-SS: Unified Storage and Monitoring Solution for Diverse IOT Data

R.Priyambiga[1], Pallavi Kulkarni[2] , Uday B.Mane [3] ,Shivaprasad S.More[4]

[1,2,3,4] Department of Computer Science and Engineering, Sou. Sushila Danchand Ghodawat Charitable Trust's Sanjay Ghodawat Group of Institutions,Atigre and priyambiga.r@sginstitute.in[1], kulkarni.ps@sginstitute.in[2], mane.ub@sginstitute.in[3],more.ss@sginstitute.in[4]

ABSTRACT

Storage system plays a vital role in Internet of Things (IoT). Storage of diverse data generated by IoT and monitoring is a crucial task. To ease the task, we propose a Unified Storage and monitoring solution for heterogeneous IoT data. The proposed solution consists of three stages: In the first stage, the diverse data is collected from sensors during the monitoring and then with the help of Sensors Web Enablement (SWE) diverse data is transformed into a unified format. The transformed data undergoes preprocessing which includes different methods such as merging datasets, binning and de-duplication in second stage. Finally, the third stage stores the preprocessed data in a unified storage system known as Document Object Relational Storage System (DOR-SS) which is deployed in the cloud.

Key Words: Internet of Things, DOR-SS, Unified Storage, SWE

INTRODUCTION

With the exponential growth of IoT, Number of things connected to Internet is growing. Vast amount of heterogeneous data is being generated. This data can be utilized for analysis and prediction in various applications. Data can be of streaming, text, application, time series, geospatial, video & image, relational, social network, etc. The challenges associated are (1) Clean, unify and store the huge data. (2) Analyze the data for timely and effective actions. Relational Database Management Systems has strong mathematical base and are widely used but they are unable to scale as per today's storage requirements. Object/ NoSQL databases provide greater flexibility and horizontal scalability. This makes NoSQL database systems popular for IoT data storage [9]. MogoDB, PostgreSQL are mostly used to store IoT data[1].But there is lack of standards and storage as well as access techniques and queries adapted are application specific[2].

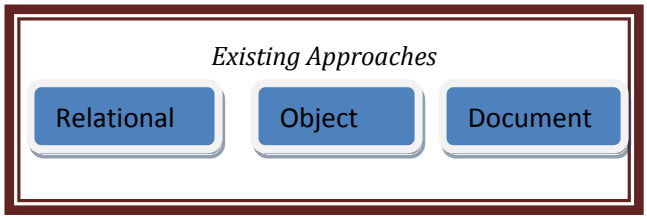

Figure 1: Existing approaches handle data separately by type

Figure 2: Our approach handles multiple data type in a unified entry

This paper discusses the architecture of IoT environment and analyses different possible approaches that can be adopted at each stages. It also proposes a design of new database management system which can store and access all three types of data efficiently such as Document, Object and Relational. In this paper, Section 2 describes literature review. IoT system architecture and methodology adopted are explained in Section 3. In Section 4, performance analysis experimental results are discussed. Our conclusion and future scope are summarized in Section 5

LITERATURE REVIEW

Storage and management of data generated by IoT is a crucial task. Data is most important entity in IoT. Data is collected from multiple sources, heterogeneous, huge in amount, multidimensional and there is Temporal-spatial correlation. Many researchers have proposed efficient data management techniques for diverse data. To manage

 DOI : 10.18231/2454-9150.2018.0238

cost and performance of storage M. Fazio et al propose cloud based half and half stockpiling engineering that couple Document and protest arranged systems, keeping in mind the end goal to streamline information stockpiling, questioning and recovery.Tingli LI et al provide storage solution for huge, heterogeneous data; propose the data sharing mechanism amongst various applications based on ontology[1] and also provide query syntax for storage and retrieval of data. Ontology based semantic representation and management enables cross application data sharing and analysis. Ming Tao et al developed ontology based semantic fusion model for smart home applications [8]. Designing queries for all the data management operations demands construction cost, data scale problems, application complexity. Ting Lu et al, build adaptation layer to connect the bottom-layer relational database and non-relational database clusters where in top-layer applications can perform database operations through unified SQL interfaces [3]. This interface supports majority of SQL query commands, which include add, delete, update and query for both table and data levels. There is a need of database management system which caters the need of all types of data generated by connected devices. This encouraged us to design the storage solution for diverse data.

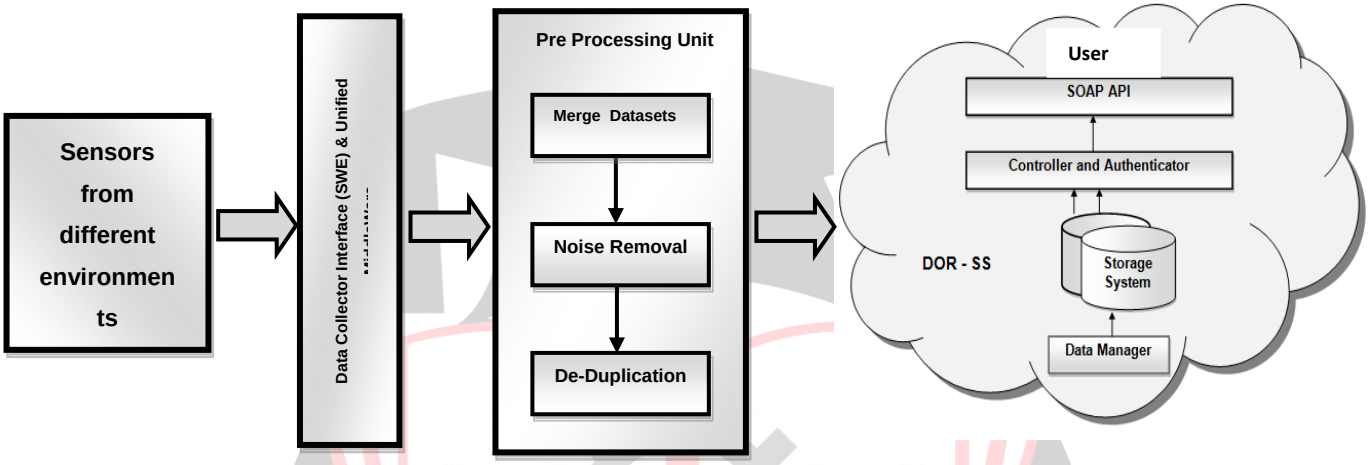

Fig 3 System Architecture

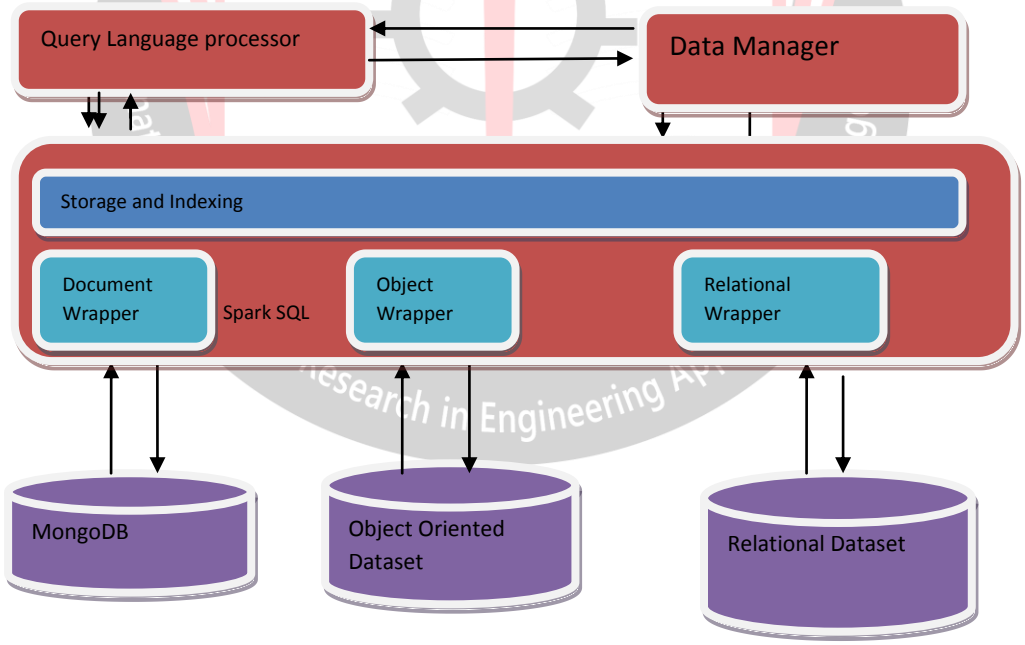

Fig 4 Unified Middleware

SYSTEM ARCHITECTURE

One of the best options for centralized IoT data processing—cloud database. NoSQL database is widely used in cloud databases. NoSQL database provide fast access to IoT data. With the help of NoSQL database spatial-temporal relationships of IoT data can be represented and stored efficiently.So,NoSQL database is the best option to deal with IoT data.The proposed architecture consists of three stages: In first stage, the diverse data is collected from sensors during the monitoring and then with the help of sensor web enablement diverse data is transformed into a unified format. The transformed data undergoes preprocessing which includes different methods such as merging datasets, binning and de-duplication in second stage.

 DOI : 10.18231/2454-9150.2018.0238

Merging Datasets

Statistical Analysis Software is used to merge datasets. After the first stage, collected IoT data is merged with the help of Merge Dataset. For example, if we are considering weather dataset and vehicle dataset the common attribute is timestamp.

MERGE Data-Set 1 Data-Set 2 by Common Variable

A one-to-one merge, appeared in the accompanying figure, joins perceptions in view of their situation in the informational indexes. You utilize the MERGE proclamation for one-to-one merging. [10]

Binning

Data binning or bucketing is a data preprocessing technique used to reduce the effects of minor observation errors. The original data values which fall in a given small interval, a bin, are replaced by a value representative of that interval, often the central value. It is a form of quantization. Statistical data binning is a way to group a number of more or less continuous values into a smaller number of "bins". For example, if you have data about a group of people, you might want to arrange their ages into a smaller number of age intervals. [1] It can also be used in multivariate statistics, binning in several dimensions at once.

De-Duplication:

With post-process de-duplication, new data is first stored on the storage device and then a process at a later time will analyze the data looking for duplication. The benefit is that there is no need to wait for the hash calculations and lookup to be completed before storing the data, thereby ensuring that store performance is not degraded. Implementations offering policy-based operation can give users the ability to defer optimization on "active" files, or to process files based on type and location. One potential drawback is that duplicate data may be unnecessarily stored for a short time, which can be problematic if the system is nearing full capacity. Finally, preprocessed data is stored in a unified storage system known as Document Object Relational Storage System (DOR-SS) which is deployed in the cloud.

CONCLUSION

Storage system plays a vital role in Internet of Things (IoT). We propose the unified storage system for storing three types of data in the cloud environment such as Document, Object and Relational data. In existing approach MangoDB, SQL and OODBMS is used separately to store their respective data. Our approach is very easy to analyze the data efficiently and performance also improved by maintaining all the data in one location rather than scattered in different location.

REFERENCES

[1] Tingli LI, Yang LIU, Ye TIAN, Shuo SHEN, Wei MAO, A Storage Solution for Massive IoT Data Based on NoSQL, IEEE International Conference on Green Computing and Communications, Conference on Internet of Things, and Conference on Cyber, Physical and Social Computing, 2012

[2] M. Shona and B. N. Arathi, ―A survey on the data management in IoT,International Journal of Scientific and Technical Advancements, Volume 2, Issue 1, pp. 261-264, 2016.

[3] Ting Lu, Jun Fang, Cong Liu, A Unified Storage and Query Optimization Framework for Sensor Data, 12th Web Information System and Application Conference, 2015

[4] M. Fazio, A. Celesti, A. Puliafito, M. Villari, Big Data Storage in the Cloud for Smart Environment Monitoring, The 6th International Conference on Ambient Systems, Networks and Technologies, Procedia Computer Science 52 (2015) 500 – 506

[5] Mohamed Omar Kebailia, Karim Foughalia, Karim FathAllah, Ali Frihida, TaharEzzeddinea, Christophe Claramuntc, Landsliding early warning prototype using MongoDB and Web of Things technologies, Geospatial Big Data Trends, Applications, and Challenges (GBD - TAC), Procedia Computer Science 98 (2016) 578 – 583

[6] Deepu S R,BhaskarR,Shylaja B S, Performance Comparison Of Deduplication Techniques For Storage In Cloud Computing Environment, Asian Journal of Computer Science And Information Technology 4 : 5 (2014) 42 - 46.

[7] JayavardhanaGubbi, RajkumarBuyya, SlavenMarusic, MarimuthuPalaniswami, Internet of Things (IoT): A vision, architectural elements, and future directions, Future Generation Computer Systems 29 (2013) 1645–1660

[8] Ming Tao, Kaoru Ota, Mianxiong Dong, Ontology-based data semantic management and application in IoT- and cloud-enabled smart homes, (2016) http://dx.doi.org/10.1016/j.future.2016.11.012

[9]SharvariRautmare, Dr. D. M. Bhalerao, SQL & NoSQL Database Study for Internet of Things, International Journal of Innovative Research in Science, Engineering and Technology Vol. 5, Issue 5, May 2016, 9069-9073

[10]http://support.sas.com/documentation/cdl/en/basess/58133/HTML/default/viewer.htm#a001293113.htm

 DOI : 10.18231/2454-9150.2018.0238