International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 11 | Nov -2017 www.irjet.net p-ISSN: 2395-0072

Data Management for Internet of Things: A Survey and Discussion

Bhavana Bachhav[1], Parikshit N. Mahalle[2]

[1]Department of computer Engineering, Smt Kashibai College of Engineering, Pune, SPPU (India)
[2] Professor and Head, Department of computer Engineering, Smt Kashibai College of Engineering, Pune, SPPU (India)

---***---

Abstract - *The Internet of Things (IoT) concepts are applied to a number of applications ranging from home automation to industrial IoT, where connecting various physical things from anywhere in network. The techniques to manage and utilize the massive volume of data produced by these objects are yet to mature. Traditional database management solutions are not enough in satisfying the sophisticated application needs of an IoT network that has a truly global-scale. Current solutions for IoT data management shows partial aspects of the IoT environment with special focus on sensor networks. In this paper, we survey the data management lifecycle of the IoT. We finally propose a data management framework for IoT that takes into consideration.*

***Key Words*: Internet of Things, Data management, huge data, Real time application, Framework**

1. INTRODUCTION

The internet of things (IoT) is a network that connects various types of objects to the internet through different kinds of information perception devices so that all the physical objects are able to exchange information with each other. Data is one of the most valuable aspects of the IoT. The term IoT has different meaning for different people - IoT includes sensors, objects, smart devices, services etc that can interact with user and among themselves [4]. One of the objectives of the Internet of Things (IoT) research and development is to enable real world objects to be connected to the Web, so data generated by those objects can be discovered, collected, processed, shared and utilized to create intelligent and useful applications and services in many domains such as smart cities, environment monitoring, health and energy.

IoT data has distinctive characteristics that make traditional relational-based database management an obsolete solution. A massive volume of heterogeneous, streaming and geographically dispersed real time data will be created by million diverse devices periodically sending observations about monitored phenomena or reporting the occurrence of abnormal events[1][3].

From the data processing point of view, one of the challenges in managing the IoT data is how to deal with the large number of heterogeneous sensing sources in a particular application domain. If we take the smart city applications, the real world data made available to the city applications is not only from sensor networks installed by city authorities at fixed locations, but also from mobile sources such as buses and taxis equipped with environment monitoring sensors and participatory sensing from citizens' smart phones. Traditional data management systems handle the storage, retrieval, and update of elementary data items, records and files. In the context of IoT, data management systems must summarize data online while providing storage, logging, and auditing facilities for offline analysis. This expands the concept of data management from offline storage, query processing, and transaction management operations into online-offline communication/storage dual operations. We first define the data lifecycle within the context of IoT and then outline the energy consumption profile for each of the phases in order to have a better understanding of IoT data management.

2. MOTIVATION

IoT is a network that connects objects to the internet through various kinds of information perception devices so that ordinary physical objects are able to exchange information with each other. Data is one of the most important aspects of IoT. It is collected from various kinds of sensors in IoT environment. In most of the IoT application, large number of sensors and data receivers sends information to server. The server gathers the information that later on becomes the huge amount in short time. IoT application faces the challenge of real time managing/extracting client useful information from whole data stored on server. As this entire situation occurs in IoT environment, there is huge need of data management. The right data management strategy allows production processes to be optimized more precisely, errors to be avoided and cost to be minimized. Thus data management plays a central role. A good data management system enables it to provide the right data to the right time at right speed, no matter which sources they are from and where they are located. It covers the entire data life cycle, from data collection, storage, classification and prioritization to achieving or deletion. Data management also enables data migration as well as aspect of data security and integrity.

3. LITERATURE SURVEY

1] "Large data management in IoT application" focuses that IOT applications can face the challenge of real time managing or displaying or extracting client useful information from the whole data stored on servers. Especially in critical situations, client's database query can take too long. A distinct layer of data processing is used to "cache" fields based on selected or most frequent database queries.

2] In "Data management for Internet of things: Green direction", the life cycle of data within the Internet of Things and survey the current research in the data management field for the Internet of Things has discussed. The discussion will focus on the research which is related to the optimization of communication overhead and storage mechanisms as they have the most significant impact on energy consumption.

3]"Data management for internet of things: design primitives and solutions" focuses on the survey of the data management solutions that are proposed for IoT or subsystems of the IoT has been done. The distinctive design primitives are highlighted. Finally, a data management framework for IoT is proposed that takes into consideration the discussed design elements and acts as a seed to a comprehensive IoT data management solution.

4] " Enabling Query of Frequently Updated Data from Mobile Sensing Sources", Wei Wang[4] focuses on two problems: (i) how to design a common, structured sensing layer for the heterogeneous, mobile data sources and, (ii)how to query FUTS (Frequently Updated, Time stamped and Structured) data from these sources.

5]"Efficient Storage of Multi-Sensor Object-Tracking Data" proposed the first read/write-optimized solution for storing multi-sensor object-tracking data on HDFS. The results suggest the efficiency of the proposal with respect to disk-write throughput, memory-write throughput, search performance, and sensor clustering.

6]"A Unified storage and query optimization framework for sensor data" proposed that Traditional data storage and query approaches cannot handle large amount of sensor data properly. To deal with such limitations, a unified storage and query optimization framework, named *DeCloud-RealBase*, is proposed towards the management of large volumes of sensor data.

7]"A storage solution for massive IoT data based on NoSQL" proposed a storage management solution called IOTMDB based on NoSQL as current storage solutions are not performing well support storing massive and heterogeneous data collected by IoT devices. Some evaluations are also carried out instead of considering only about expressing and organizing of IoT data.

8]"IoT data management methods and optimization algorithms for mobile publish/subscribe services in cloud environment" focuses on design principles for data management methods in IoT and optimization algorithms by way of publish/subscribe middleware and linked data which spread over mobile network for producing a coherent IoT ecosystem.

9]"Data management in ambient assisted living platforms approaching IoT : a case study" analyzes the issues related to data management starting from a review of state of art for drawing a general approaches. In this paper , investigation has been done on data handling and management issues from the adoption of IoT paradigm in ambient assisted living platform.

10]In "When things matter: A data centric view of the Internet of Things", main techniques in IoT from data centric point of view, which includes data stream processing , data storage models , complex event processing and searching has been discussed.

4. GAP ANALYSIS

Paper no	Data format	Storage	architecture	Processing speed	Server response
1	Sensor data ,RFID	Cloud based	Centralized	High	Good
2	Time series data	Local	Decentralized	High	Good
3	Sensor data	Local	centralized	Low	Good
4	Historical data	Cloud based	Centralized	Medium	moderate
5	Sensor data	Cloud based	Centralized	Low	Moderate
6	Time series data	Local	Decentralized	High	Good
7	Sensor data	Local	Centralized	High	Moderate
8	Sensor data	Local	Decentralized	Medium	Good
9	Historical data	Cloud based	Centralized	High	Good
10	Sensor data	Cloud based	Centralized	Low	Good

Table- 1: Gap Analysis

It is useful to classify data of the IoT into a number of categories. Some data is discrete and some continuous, some automatically generated and some an input by humans. For gap analysis, data format, storage, architecture, processing speed and server response are the points which are taken into consideration We have categorized the data into the following areas: RFID, address/unique identifiers, descriptive data, positional and environmental data, sensor data, historical data. Radio Frequency Identification refers to identification and tracking using radio waves and is becoming a common place technology. In paper [1],[3],[5],[7],[8],[10] ,sensor data is used where in [2] and [6] time series data and in [4],[9] historical data used. Processing speed is high in [1][2][6][7][9] where medium and low in [4],[8] and [3],[5],[10] respectively.

5. IOT DATA LIFECYCLE AND DATA MANAGEMENT

The lifecycle of data within an IoT system is illustrated in Fig 1. Querying and analysis are the end points that initiate and consume data production [3]. We divide an IoT data management system based on the data lifecycle into an online frontend that interacts directly with the interconnected IoT objects and sensors, and an offline backend that handles the mass storage and in-depth analysis of IoT data. The data management frontend is communication-intensive; involving the propagation of query requests and results to and from sensors and smart objects. The backend is storage-intensive; involving the mass storage of produced data for later processing and analysis and more in-depth queries [8][11][12]. Although the storage elements reside on the back end, they interact with the front end on a frequent basis via continuous updates and are thus referred to as online [5][6].

Data-intensive systems rely on *querying* as the core process to access and retrieve data. In an IoT context, queries can be issued either to request real-time data to be collected for temporal monitoring purposes or to retrieve a certain view of the data stored within the system. Data *production* involves sensing, collecting and sending data by the Things within the IoT framework and reporting this data to interested parties periodically, pushing up the network to aggregation points and subsequently to database servers, or triggered by queries that pull the data from sensors and smart objects. *Fusion* techniques deploy summarization and merging operations in real-time to compress the volume of data to be stored and transmitted. In *delivery*, the objects within the IoT may store data for a certain time interval or report it to governing components. Wired or wireless broadband communications may be used from there to transfer data to permanent data stores. Data may need to be *pre-processed* to handle missing data, remove redundancies and integrate data from different sources into a unified schema before being committed to storage. Storage phase handles the efficient storage and organization of data as well as the continuous updates of data. *Archiving* refers to the offline long-term storage of data that is not immediately needed for the system's ongoing operations. The core of centralized storage is deployment of storage structures that adapt to various data types and frequency of data capture. Processing phase involves the ongoing retrieval and analysis operations performed on stored and archived data in order to gain insights into historical data and predict future trends, or to detect abnormalities in the data that may trigger further investigation or action. Task-specific pre-processing may be needed to filter and clean data before meaningful operations take place. This is all about the various phases in the data lifecycle in IoT.

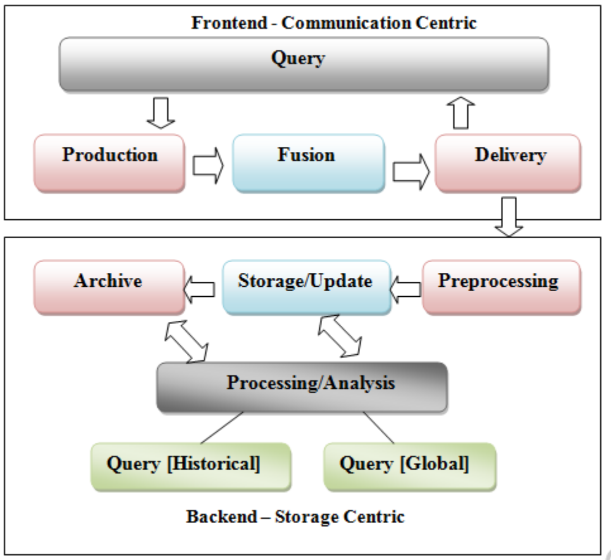

Fig - 1: IoT data lifecycle and data management

6. DATA MANAGEMENT FRAMEWORK FOR IOT

The proposed IoT data management framework consists of six layers, two of which include sub-layers and complementary or twin layers. The framework layers relate closely to the phases of the IoT data lifecycle, as shown in Figure 2 [3][9]. The “Things” Layer encompasses IoT sensors and smart objects (data production objects), as well as modules for in-network processing and data collection/real-time aggregation (processing, aggregation). The Communication Layer provides support for transmission of requests, queries, data, and results (collection and delivery). The Data layers respectively handle the discovery and cataloguing of data sources and the storage and indexing of collected data (data storage/archival). The Data Layer also handles data and query processing for local, autonomous data repository sites (filtering, preprocessing, processing) [7][10][12].

The Federation Layer provides the abstraction and integration of data repositories that is necessary for global query/analysis requests, using metadata stored in the Data Sources layer to support real-time integration of sources as well as location-centric requests (preprocessing, integration, fusion). The Query Layer handles the details of query processing and optimization in cooperation with the Federation Layer as well as the complementary Transactions Layer (processing, delivery).

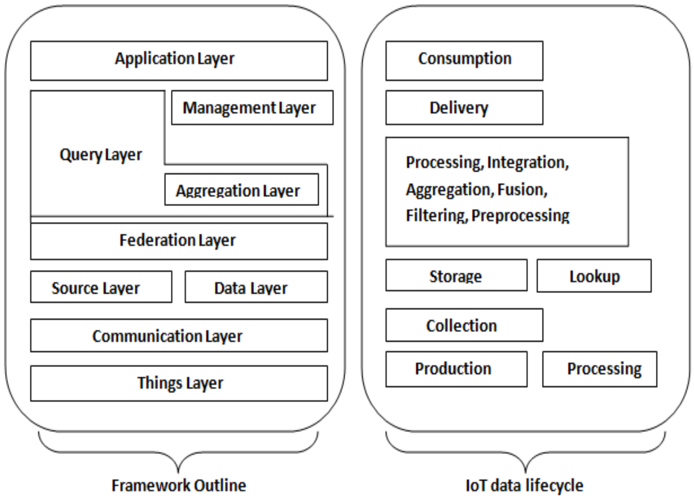

Fig – 2: IoT data management Framework

7. MATHEMATICAL MODEL

Input = Sensory Data {**sd**}

Sd = {s1, s2, s3 ...sn}

Output = Managed / Evaluated Information
Mathematical model using Moore machine

(Q, Σ, O, δ, X, q0)

Q = {1) Unique identification of data sources,

2) Generate clear unified format for data,

3) Identify communication network with enough speed,

4) Identify data processing capacity of server,

5) Identify storage capacity of server,

6) Aggregate or fuse the data,

7) Inference generation and decisions as per application }

Σ = {Sensory data –sd , Sources}

O = { 1) Assume output symbol # if data resources are ok,

2) Assume output symbol $ if unified information is generated,

3) Assume output symbol @ if data is well managed and evaluated as per requirement,

4) Assume output symbol & for accurate decision upon data analysis}

Present State	Next State	Output
Q1	Q2	#
Q2	Q3	$
Q3	Q4	#
Q4	Q5	#
Q5	Q6	#
Q6	Q7	@
Q7	Q7	&

Fig – 3: State transition diagram

8. CONCLUSIONS

In this paper, we discussed the IoT data management lifecycle, in which data goes through various stages of processing. Also we outlined the framework which highlights the need for two-way, cross-layered design approach that can address both real-time and archival query, analysis, and service needs.

Future work involves mapping the details of the proposed framework more closely to the reference model in the IoT-A, in-depth investigation and development of a data management solution that builds upon the proposed framework, and adding considerations of data security and privacy into the framework design in compliance with the considerations that need to be addressed in the IoT dynamic and heterogeneous environment. Industry wide global standards, unified communication protocols, and highly enhanced security aspects and middleware problems are left for future work.

REFERENCES

[1] Catalin Constantin CERBULESCU, Claudia Monica “Large data management in IoT application”, ,2016 IEEE,978-1-4673-8606-7

[2] Abu-Elkheir, M., “Data management for the Internet of Things: Green directions”, Globecom Workshops (GC Wkshps), 2012 IEEE, Anaheim,CA, pp 386 – 390, 3-7 Dec. 2012.

[3] Yuchao Zhou, Suparna De ; Wei Wang ; Klaus Moessner, “Enabling Query of Frequently Updated Data from Mobile Sensing Sources”, Inst. for Commun. Syst., Univ. of Surrey, Guildford, UK, Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on, 19-21 Dec. 2014, Page(s): 946 – 952, ISBN: 978-1-4799-7980-6

[4] Mervat Abu-Elkheir, Mohammad Hayajneh and Najah Abu Ali, "Data Management for the Internet of Things: Design Primitives and Solution",Sensors 2013, 13, 15582-15612; doi:10.3390/s131115582 ,14 November 2013 ,ISSN 1424-8220

[5] Xingjun Hao, X. Hao, Peiquan Jin ; Lihua Yue, "Efficient Storage of Multi-Sensor Object-Tracking Data", School of Computer Science and Technology, University of Science and Technology of China, Hefei, IEEE Transactions on Parallel and Distributed Systems (Volume:PP ,Issue: 99), Page(s): 1-5 ISSN : 1045-9219

[6] Ting Lu, Jun Fang, Cong Liu, "A Unified Storage and Query Optimization Framework for Sensor Data", 2015 12th Web Information System and Application Conference (WISA), 11-13 Sept. 2015, 229 -234, ISBN: 978-1-4673-9371-3

[7] Cooper, J.; James, Challenges for database management in the internet of things. IETE Tech.Rev. 2009, 26, 320–329.

[8] https://www.sap.com/india/trends/internetof-things.html.

[9] https://en.wikipedia.org/wiki/Internet of things.

[10] Y. Qin, Q. Z. Sheng, N. J. G. Falkner, S. Dustdar, H. Wang, and A.V. Vasilakos, "When Things Matter: A Data-Centric View of the Internet of Things",in CoRR, 2014.

[11] L. Chen, M. Tseng, and X. Lian, "Development of foundation models for Internet of Things," Frontiers of Computer Science in China, vol. 4, no.3, pp. 376-385, September 2015.

[12] L. Chen, M. Tseng, and X. Lian, "Development of foundation models for Internet of Things," Frontiers of Computer Science in China, vol. 4, no.3, pp. 376-385, September 2015.

[13] K. Aberer, M. Hauswirth, and A. Salehi, "A middleware for fast and flexible sensor network deployment," in Proc. 32nd International Conference on Very Large Databases, 2006