International Journal of Database Theory and Application
Vol.8, No.5 (2015), pp.99-108
http://dx.doi.org/10.14257/ijdta.2015.8.5.08

Design and Implementation of Data Platform Based on Internet of Things Technology

De-Qiang Liu

Wei Fang Vocational College
sdwfinfo@126.com

Abstract

In recent years, the rapid development of Internet of Things has received wide attention of the social and academic circles. However, if there is no unified standard to store and process the huge data, the systems are still highly independent and interconnection is difficult to be realized. This paper researches the design and implementation of data platform based on Internet of Things technology. We firstly analyze the data sources and features to understand the platform requirements. Then we propose the data platform scheme with the function and performance requirements considered. It focuses on the resource identification and addressing, resource description and management, data storage, processing and analysis problem. With the data platform, the resources in Internet of Things system are managed in a unified way, which improves the system openness, access and transmission capability thus makes the system more flexible and open. However, the current design scheme can be improved in performance and safety in the future research.

Keywords: *Internet of Things, data platform, resource management, data processing*

1. Introduction

The concept of IoT (Internet of Things) is put forward by Kevin Ashton in 1999 [1]. At that time, IoT is considered as follows: each item stores the unique EPF (Electronic Product Code) of electronic labels, which collects the item information with RFID (Radio Frequency Identification) technology, and shares the EPC information with Internet. In 2005, ITU (International Telecommunication Union) releases the technical report named "Internet of Things" [2]. It extends the definition and coverage of IoT, and points out that IoT is a new dimension of information and communication technology.

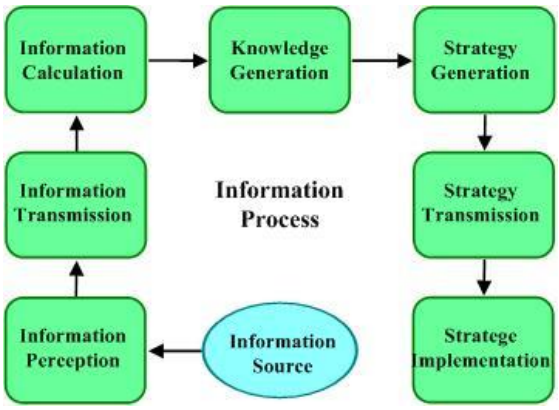

Figure 1. A Complete Information Process

As technology evolves, people have a deeper and broader understanding on information technology and IoT. A complete information process is shown in Figure 1,

ISSN: 2005-4270 IJDTA

including information perception, transmission, calculation, knowledge generation, strategy generation, transmission and implementation [3]. IoT system obeys the information process, in which knowledge generation and strategy generation represents for the intelligence development. From this perspective, IoT is a kind of network which connects any item into Internet with information sensing equipment for information exchanging and communication, thus realizing intelligent identification, location, tracking, monitoring and management. It can be considered as an extension and expansion of the Internet.

The core of traditional information network is human semantics, however, large amounts of devices with sensing, control and communication abilities in short distances exist in IoT system. These devices extends the system ability of obtaining information automatically, allowing the information contributed by the objects exist in IoT system as well as the human semantic information. At the same time, devices can be controlled with controllers in IoT system. Therefore, interactions between human and the real world are realized, and the digital ecosystem of human world and physical world is formed through IoT system [4].

The openness and simplicity of Web technology allow all users contribute information and applications to the Web. It provides service abilities for all the users, making them both service providers and consumers. The mechanism of separating the network builder and service provider is the main reason for the rapid development of Internet, and also makes Web technology the natural choice of IoT system. Therefore, this paper designs the data platform of IoT system with Web technology, which focuses on the resource identification, addressing and description, data compression, storage and cache technologies.

2. Key Technologies of Data Platform

2.1. Platform Middleware Technology

Middleware can achieve the resources abstraction and other functions in IoT system. It provides services for the upper applications in the form of API, which shields the underlying hardware environment, communication protocol and heterogeneity.

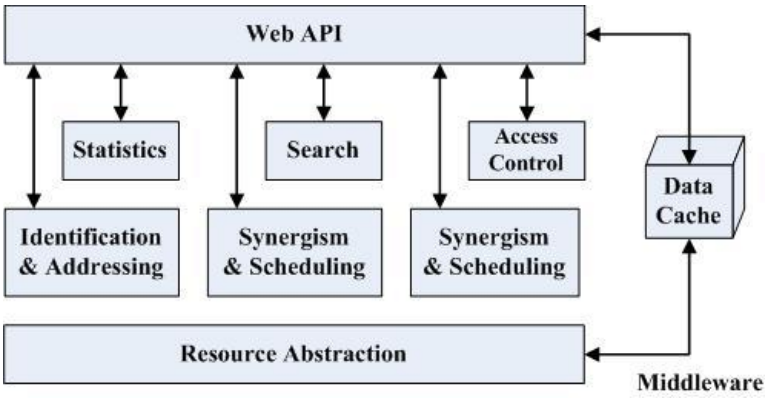

Figure 2. The Basic Functions of Middleware

Figure 2 shows the basic functions of middleware, in which the most core function is the resources abstraction and development. It also provides resources identification and addressing, synergism and scheduling, synergism and scheduling, statistics, search and access control functions. Data cache effectively improves data access speed.

2.2. Resource Identification and Addressing

ETSI (European Telecommunications Standards Institute) defines "resource" as the entity can be addressed independently, which includes both data and ability [5]. Resource identification of IoT satisfies four requirements: uniqueness, practicability, scalability and compatibility [6].

Resource identification of IoT can be divided into three types: physical identification, communication identification and application identification. The physical identification mainly identifies the hardware entities, such as EPC, IMEI, UID and NID. The communication identification identifies the logical entities, such as IPv4, IPv6 and Session ID. The application identification identifies the services and information resources, such as URL.

2.3. Resource Description Model

Resource description refers to the attribute of resources. ETSI defines "attribute" as the metadata which provides relevant resource identification features [5]. At present, resource description framework mainly contains RDF, WSDL and EEML.

- RDF (Resource Description Framework) is the W3C standard for describing network resources, such as the webpage title, author, content and copyright information. RDF provides model and grammar for data. It is written in XML, which can be read and understood by computer [7].
- WSDL (Web Service Description Language) is a kind of XML language describing Web services. Also, it provides detailed interface instructions for users [8].
- EEML (Extended Environments Markup Language) describes the sensor data collected from different entities in structured format. It supports the interaction and interpretation of EEML language between software packages, which provides a protocol for communications between physical and virtual entities [9].

2.4. Data Compression Technology

The purpose of data compression is to reduce the data storage space and transmission time [10]. From the perspective of compression distortion, data compression can be divided into lossless and loss compression.

The lossless compression includes Huffman coding, arithmetic coding and prediction. These traditional technologies mainly compress information on the grammatical level. However, a lot of redundant semantics information exists in IoT system. Feng and Yan propose a kind of sensor data storage and access method with the packet compression [11]. It reduces the data storage space by eliminating the redundant information, while ensuring the random access feature, which inspires us a lot.

2.5. Data Storage Technology

Relational and non-relational database are two main data storage technologies. IoT system should design appropriate storage structure considering its own features.

- Relational database is a kind of database based on the relational model, which processes data with mathematical methods. Relational model is composed of relational data structure, relational operation set and relational integrity constraints. SQL (Structured Query Language) is a kind of relational database [12].
- The traditional relational database has appeared to be inadequate in the face of dynamic web2.0 with large scale and high concurrency. NoSQL (Not Only

SQL) is generated to solve this challenge, especially for large data application problems.

2.6. Cache Technology

Cache is mainly to improve the data access speed [13]. As traffic brings bottleneck between the server and application clients, cache helps reduce interactions between them and greatly improves the program performance.

Cache strategies are based on time, access frequency or time distance distribution strategy. It improves the hit rate of cache content by designing good algorithms for data block, prediction, sequential pre-fetching and cache replacement.

3. Data Features of Internet of Things System

3.1. Data Sources of IoT System

Sensor network is one of the main data sources of IoT system. It is composed of multiple wired or wireless sensor nodes with low power consumption and small volume, providing information collection and communication functions among entities in a local area network. There are a large number of sensor devices in IoT system, such as temperature and humidity sensors, light sensors, carbon dioxide sensors, wind sensors, and so on.

IoT technology is originated from RFID, which is another important data source of IoT system. Besides IoT system, RFID has also been widely used in many other fields, such as access control system, logistics system, library system, and so on.

The control function of the underlying devices is required for IoT system. The new smart devices can be controlled by Wi-Fi or other protocols, however, the traditional devices can only be controlled through some simple devices, such as electrical relay.

Another important function of IoT system is monitoring, of which video monitoring is an effective method. In recent years, with the rapid development of image processing and transmission technology, video monitoring technology has also made great progress. The video camera devices transmit the video, audio or image to the video monitoring platform, providing intuitive, accurate, timely and rich information for IoT administrators.

In summary, data resources of IoT system include sensor data, RFID data, relay value, video, audio and picture files. Besides, other resources management and statistical information also exist in IoT system. So compatibility and scalability should be considered emphatically for data platform design of IoT system.

3.2. Data Classification of IoT System

The effective classification of data helps us design appropriate data structure and storage method of IoT system data platform. According to their different functions, the data of IoT system can be divided into four categories:

- Numerical values of resources, including the sensor values, relay status and RFID data.
- Multimedia data, including videos, audio and pictures.
- Metadata of resources, describing the resources attributes.
- Management and statistical data, including the gateway upload data and user access data.

3.3. Data Features of IoT System

Mass data: gateway and devices sends data to the data platform periodically or in a certain mode. The platform also generates new data according to the received information. It leads to the rapid expansion of the data scale of the platform.

Heterogeneity: as introduced in Section 3.1, IoT system includes sensor data, RFID data, relay value, video, audio and picture files. Different data sources and data types directly bring heterogeneity for IoT system.

Space-time correlation: data of IoT system reflects the changes in the real world with space and time attributes. Data sent to the platform are collected by devices located in different places at different time, so they are correlated with both space and time.

Timeliness: An important application of IoT technology is real-time monitoring, which reflects the real-time changes of the physical world. It brings two technical challenges for the data platform: one is how to choose a good caching algorithm to guarantee the data query performance; the other is how to process the fast growing log data.

4. Data Platform Requirement Research and Analysis

4.1. Resource Identification and Addressing Requirement

Resource identification and addressing is the basis of resource access, discovery and management. There are four requirements for resource identification in IoT system:

- Global uniqueness: the function of resource registration center is the allocation of resource identifications in IoT system. It requires each resource identification to be unique, or it cannot be registered in the system.
- Unified management: the unified distribution and management system should be applied to identify the resource identification.
- Scalability: as the increasing number and various types of the resources, the identification system should be with good scalability to ensure the resource identifications are available with the network evolution.
- Compatibility: there are already many identification systems and methods at the moment, so the IoT identification should be compatible to the current systems while satisfying the system requirements.

4.2. Resource Description Requirement

In IoT system, all the devices, data information and services can be abstracted as resources. However, there are great differences among resource description grammers and semantics due to the heterogeneity, which brings a great challenge to the effective organization and management of devices. IoT platform defines the unified resource description framework as the basis of resource discovery, search and management for more efficient and more intelligent polymerization of the resources.

4.3. Resource Management Requirement

Resources are the basic unit of the IoT system. The heterogeneity of gateways, devices and resources bring great challenges for resource management [14]. Resource description is the basis of the resource management. A good resource description model can shield the details of the underlying devices, thus the gateways, devices and resources can be effectively managed.

The consistency is an important problem in resource management process. It includes two aspects: one is the consistency of the gateway attributes information; the other is the gateway attributes and platform attributes information.

4.4. Data Storage Requirement

The various types of IoT system data brings great challenges for data storage [15], while the rapid growing data amount also has a great impact on the system search

and data management performance. The requirements for data storage include appropriate storage format, database and table structure.

4.5. Data Processing Requirement

Data repetition is abundant in IoT system, which is caused by two main reasons: high sampling frequency and invalid precision. The data change frequency is far less than the sampling frequency, and the accuracy of collected data is far more than users required, thus a large amount of duplicate data is collected. The main requirement for data processing module is to remove the duplicated data of IoT system [16].

4.6. Data Analysis Requirement

Data analysis refers to the statistics and analysis on the type, amount, value and usage of the platform resources [17]. Its requirements come from the commercial future of IoT system. The upper layer of IoT system is the application layer, which provides services for users with the underlying and devices and resources. Understanding user requirements to establish the billing model for IoT system is the inexorable trend.

5. Data Platform Design and Implementation

5.1. Function Framework of Data Platform

In order to meet the requirement of flexibility, scalability and high availability, we design the IoT data platform as the hierarchical structure. It can be divided into four layers: database layer, data access layer, business transact layer and user service layer, as shown in Figure 3.

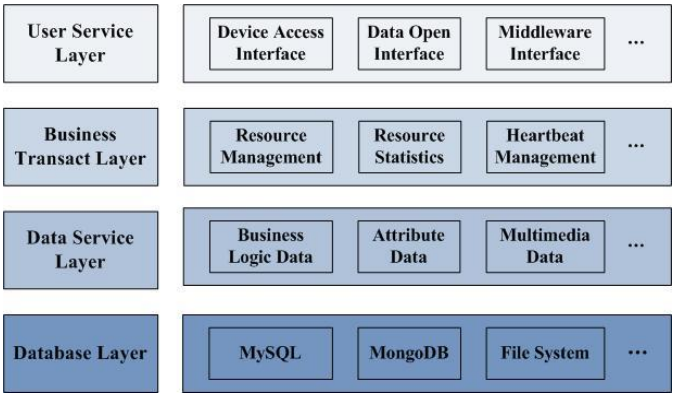

Figure 3. The function Framework of IoT Data Platform

- Database layer mainly stores the business data, gateway metadata and statistical data implemented by different database systems and file systems. In order to ensure the consistency of data access, we define the unified data access interfaces between the database layer and upper layers, which are encapsulated in the database layer. It improves the scalability of IoT system, which provides the basic data support for the upper business functions.
- Data service layer is composed of various data access objects corresponding to the different data types. The object of abstract resource access is to shield the physical details of database layer, which provides universal access method for the business transact layer. The upper layers ignore the complex data operations, thus realizing the decoupling of the database layer and the upper layers.

- Business transact layer obtains data by calling the methods in data service layer. The business transact layer includes four modules: gateway resource management module, gateway heartbeat management module, data processing module and cache module.
- User service layer is the presentation of the platform. It provides API (Application Programming Interface) for users, mainly including device access interface, data open interface and middleware interface.

5.2. Resource Identification and Addressing Design

The IoT system can be divided into gateway layer, device layer and resource layer from top to bottom. The gateway connects several devices, and these devices provide resources. So we apply a layered architecture for resource addressing.

Accordingly, the resources can be described as follows:

```
Platform Address / Gateway ID / Device ID / Resource ID
```

The above description is completely defined by users. It is with good flexibility and scalability, but is too generalized, thus goes against resource search and discovery. We define a set of common attributes, covering most of the attributes of gateways, devices and resources. Also the specific attribute descriptions are added on the basis of it. The improved description is designed as follows:

```
<MW_MSG_Attribute xmlns:xsi="http://www.w3.org/2015/XMLSchema-instance" xmlns:xsd
="http://www.w3.org/2015/XMLSchema">
      <MWID>01</MWID>
      <AttrEditions>
         <Attribute Language="en" Tag="Gateway" Name=" Attribute ">...</Attribute>
         <Attribute Language="en" Tag="IoT, Smart Home" Name="Attribute">...</Attribute>
      </AttrEditions>
      <RegTime>2015-04-10T10:00:00.995+08:00</RegTime>
</MW_MSG_Attribute>
```

5.3. Gateway Resource Management Module Design

The gateway management module includes gateway, devices and resources registration, information updating and information deleting.

The gateway firstly sends the registration information to the platform, and then the platform generates the gateway data and returns its ID. After successfully registered, it reports the gateway, devices and resource attributes information to the platform. The gateway manages the devices through updating their attributes information, including devices and resources registration, information updating and deleting. The gateway can be canceled when it ends services in the system.

5.4. Data Processing Module Design

As introduced above, the IoT data can be divided into four categories: numerical values of resources, multimedia data, metadata of resources, management and statistical data. We apply MySQL to store the resources data and statistical data, MongoDB to store the gateway metadata, the file system to store the multimedia data and store their path in MySQL.

The management and statistical data can be further divided into the upload data management and the resource access management. The upload data mainly includes the parameters uploaded by the gateway, such as the size, types and upload time. The resource access data mainly describes the resources accessed by the platform.

5.5. Gateway Heartbeat Management Module Design

The heartbeat management module manages the heartbeat information of the gateway to determine if the gateway is online [18]. After the gateway is successfully registered, its ID is recorded in the heartbeat management module. The gateway periodically sends its heartbeat message to the platform though the heartbeat interface. The heartbeat management module updates the latest time. If it does not receive the heartbeat information within five minutes, the gateway is considered to be offline.

5.6. Cache Module Design

Cache can be realized by several means, in which the simplest way is to store the cache data by queue or circular queue. This way is simple, but need to move large numbers of data objects when the cache is full or cache hits.

Another popular way is to store the cache data with linked list. Compared with the queue, this method is more flexible and adding or replacing data can be realized only by adjusting the pointer. So we apply the linked list as the storage structure of cache.

6. Conclusions

The IoT data platform provides data storage and resources management functions, which plays an important role in today's widely popular IoT system. This paper proposes a data platform scheme based on IoT technology. It mainly researches the sources and features of IoT data, resource identification and addressing, resource description and management, data storage, processing and analysis, and the platform design and implementation.

The platform scheme is proposed according to the features of IoT data, so it can meet the requirements of mass data, heterogeneity, space-time correlation and timeliness. However, there still lie the following problems in the platform:

1. Performance problem: the current storage method can hardly satisfy the requirements of huge data amount and high query efficiency. The distributed storage will be the main technology in the future.
2. Lack of effective security mechanism: the data platform proposed in this paper is totally open, and there's no effective security mechanism to ensure the data privacy. It is not conductive for the development of IoT commercialization.

The final direction of IoT is intelligence. It is still far from achieving that, but machine learning, artificial intelligence and data mining has made a lot of achievements. At present, the context aware based on IoT is also in the study. The future of IoT should combine these research results with its own characteristics to promote the new network technology to bring greater changes to people's life.

Acknowledgments

These should be brief and placed at the end of the text before the references.

References

[1] Ashton K., “That ‘Internet of Things’ Thing”, RFiD Journal, vol. 22, **(2009)**, pp. 97-114.
[2] “Strategy ITU, Unit P. ITU Internet Reports 2005: The internet of things”, Geneva: International Telecommunication Union (ITU), **(2005)**.
[3] Soos S., Kampis G. and Gulyás L., “Large-scale temporal analysis of computer and information science”, European Physical Journal Special Topics, vol. 222, no. 6, **(2013)**, pp. 1441-1465.
[4] Atzori L., Iera A. and Morabito G., “The internet of things: A survey”, Computer Networks, **(2010)**, pp. 2787 -2805.
[5] ETSI T S. 102 690: “Machine-to-Machine communications (M2M)”, Functional architecture.

[6] DG G., WC B., LT H., "The protein identification resource (PIR)", Nucleic Acids Research, vol. 14, no. 1, **(1986)**, pp. 11-15.
[7] Swick R. R., "Consortium WWW. Resource Description Framework (RDF) model and syntax specification", Recommendation, **(1999)**.
[8] Christensen E., Curbera F. and Meredith G., "WSDL: Web Service Description Language" Asg Deliverable, **(2001)**.
[9] Zhang M., Lu Y. and Gong C., "Energy-Efficient Maximum Lifetime Algorithm in Wireless Sensor Networks", International Conference on Intelligent Computation Technology & Automation-volume. IEEE, **(2008)**, pp. 931 - 934.
[10] Jain A. K., "Image data compression: A review", Proceedings of the IEEE, vol. 69, no. 3, **(1981)**, pp. 349 - 389.
[11] Tao F., Lin Y., Yao G. and Xiangqun C., "A Compression-Based Storage and Access Method for Sensor Data", Computer Science, vol. 10, **(2011)**, pp. 190-195.
[12] Burns R. K. and Koh Y., "A Comparison of Multilevel Structured Query Language (SQL) Implementations", Computer Security Applications Conference. IEEE Computer Society, vol. 192, **(1996)**.
[13] Fan L., Cao P. and Almeida J., "Summary cache: a scalable wide-area Web cache sharing protocol", IEEE/ACM Transactions on Networking, vol. 8, no. 3, **(2000)**, pp. 281 - 293.
[14] Fan T. and Chen Y., "A scheme of data management in the Internet of Things", Network Infrastructure and Digital Content, 2010 2nd IEEE International Conference on. IEEE, **(2010)**, pp. 110-114.
[15] Cosm, "Internet of Things Platform Connecting Devices and Apps for Real-Time Control and Data Storage", cosm.com, **(2012)**.
[16] Jiang L., Xu L. D. and Cai H., "An IoT-Oriented Data Storage Framework in Cloud Computing Platform", IEEE Transactions on Industrial Informatics, vol. 10, no. 2, **(2014)**, pp. 1443-1451.
[17] Lingjuan L., "Data Management and Intelligent Processing in IoT", Zte Technology Journal, **(2011)**.
[18] Zhi K. H., Duo C. H. and Wei H. G., "Remote monitoring system of rectifier based on improved heartbeat mechanism", Journal of Computer Applications, vol. 28, no. 2, **(2008)**, pp. 363-366.

Author

De-Qiang Liu, graduated from Shandong University at Weihai computer technology professional in 1995, and obtained a master's degree in computer science in Shandong University of science and technology in 2008, now from the things of internet technology in teaching and research work, professional and technical posts as associate professor.