2023 IEEE 3rd International Conference on Data Science and Computer Application (ICDSCA) | 979-8-3503-4154-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICDSCA59871.2023.10393383

Research and design of a parking Internet of Things device management platform based on Hadoop big data technology

Chongwen Wu
School of Information Engineering
Wuhan University of Technology
Wuhan, China
wcw0122005830125@whut.edu.cn

***Abstract*—This paper focuses on the management of parking IoT devices and introduces the practice of using Hadoop and NB-IoT technologies for data processing and intelligent management. The paper describes the basic infrastructure and components of the Hadoop platform, as well as methods for implementing big data processing and analysis, while delving into the application scenarios and advantages of NB-IoT technology in IoT device management. Based on this, the paper constructs the architecture and functional modules of the IoT device management platform, including data acquisition, storage, analysis, visualization, etc., and proposes application solutions such as intelligent warning, real-time monitoring, and decision support. The research results have high reference value for the research and practice in the field of parking IoT device management.**

Keywords—Hadoop, IoT device, NB-IoT technology, Management platform

I. INTRODUCTION

Parking IoT is an important component of the urban transportation system, and its operational efficiency and service quality are directly related to the smooth flow of urban traffic and the convenience of citizens' travel. However, faced with the acceleration of urbanization and the increase of car ownership, parking IoTs have encountered a series of problems such as space constraints, resource waste, and management chaos. In order to address these challenges, this paper designs and implements an Internet of Things device management platform suitable for parking IoT scenarios based on Hadoop big data technology. This platform can real-time monitor the status of parking IoT devices, such as the occupancy of parking spaces, the working condition of devices, and collect massive parking data through sensors and data acquisition devices. These data are stored, processed and analyzed to provide real-time parking IoT data analysis, parking space management, parking flow prediction and other functions. Through this platform, parking IoT managers can better grasp the operation of the parking IoT, improve the efficiency of parking resource utilization, realize optimized scheduling of parking flow, and improve parking service and user experience.

II. KEY TECHNOLOGY INTRODUCTION

A. Hadoop Big Data Platform Architecture

Hadoop is an open-source distributed system infrastructure, mainly used for storing and processing massive unstructured data. Its core architecture consists of four components: HDFS, MapReduce, YARN, and HBase. HDFS is a distributed file system that divides data into fixed-size data blocks and stores multiple copies on multiple nodes to achieve high reliability, high throughput, and fault tolerance data storage. MapReduce is a distributed computing framework that decomposes large-scale data processing tasks into two stages: Map and Reduce. The Map stage assigns the input data to multiple nodes for parallel processing and generates intermediate results; the Reduce stage merges the intermediate results and outputs the final results. YARN is a resource management platform responsible for managing computing resources in the cluster and allocating and scheduling resources for users' applications. YARN separates the resource management and job control functions in MapReduce, which are implemented separately by ResourceManager and ApplicationMaster to improve cluster scalability and flexibility. HBase is a distributed database developed based on Google's BigTable principle for storing large-scale structured data. As a column-oriented NoSQL database built on HDFS, HBase supports the need for high-speed reading and writing of large amounts of data. [3][4]

In addition to these four core components, Hadoop also has many other auxiliary components, forming a complete big data ecosystem. Hive is a data warehouse infrastructure built on Hadoop, providing HQL, a SQL-like query language, for easy querying and analysis of data in Hadoop. Pig is a high-level data flow language and execution framework that supports parallel computing and abstracts MapReduce programming. PigLatin script language is used to write data processing programs, which are then converted to MapReduce tasks running on Hadoop. Zookeeper is a software that provides distributed coordination services for configuration maintenance, domain name service, distributed synchronization, and group services, providing necessary support for distributed application programs. Mahout is a scalable machine learning and data mining library that provides machine learning algorithms based on MapReduce, such as clustering, classification, and recommendation.

These components together form the Hadoop ecosystem, providing powerful tools and frameworks for processing big data. By leveraging Hadoop's distributed processing and big data storage capabilities, enterprises and research institutions are able to better manage, process, and analyze massive unstructured data, obtaining valuable information and insights. The architecture diagram of Hadoop is shown in Figure 1.

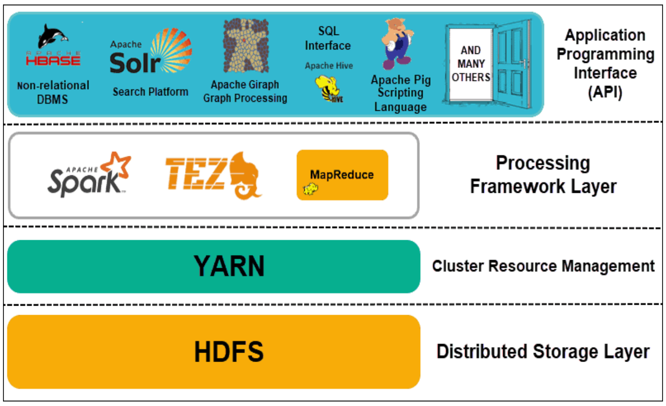

Fig. 1. Architecture diagram of Hadoop

B. Internet of Things Communication Technology

NB-IoT (Narrowband Internet of Things) is a low-power wide-area network (LPWAN) communication technology designed specifically for IoT device connections. It adopts narrowband modulation and multiple access technologies and communicates through existing mobile network infrastructure (GSM, UMTS, or LTE) to provide a broad coverage, long battery life, and low-cost connectivity solution. The NB-IoT chip is shown in Figure 2.[1]

Fig. 2. NB-IoT chip

The following are some important features and advantages of NB-IoT technology:

Low power consumption: NB-IoT devices are designed for low power consumption and can be powered by regular batteries for several years. This makes NB-IoT particularly suitable for applications that require long-term operation without frequent battery replacement, such as remote monitoring and management of parking IoT equipment.

Wide coverage range: NB-IoT has the characteristics of broad coverage and can achieve long-distance communication indoors and outdoors. Compared with traditional mobile networks, NB-IoT can penetrate walls and obstacles, providing better indoor coverage. High connection density: NB-IoT supports a large number of devices connected simultaneously, which can meet the connection needs of large-scale devices in parking IoTs. Through the NB-IoT network, sensors and equipment in the parking IoT can be interconnected and provide real-time data and monitoring services. [2]

Low cost: Since NB-IoT is built on existing mobile network infrastructure, it can share existing network resources and devices, reducing deployment and maintenance costs. In addition, the cost of NB-IoT modules and chips is relatively low, making NB-IoT an economical and affordable connection choice.

Security: NB-IoT provides end-to-end data transmission encryption and authentication mechanisms to ensure data security and privacy protection. This is crucial for protecting sensitive data and user information in parking IoTs.

NB-IoT technology has a wide range of application prospects in the field of parking IoTs, such as parking space monitoring, vehicle positioning, parking navigation, payment systems, etc. Through NB-IoT technology, parking IoT managers can obtain real-time information on the status and use of parking spaces, optimize the utilization of parking resources, and provide more convenient and intelligent parking services. At the same time, NB-IoT also supports remote management and control of parking IoT equipment, achieving functions such as device monitoring, fault diagnosis, and maintenance. The advantages of NB-IoT are shown in Figure 3.

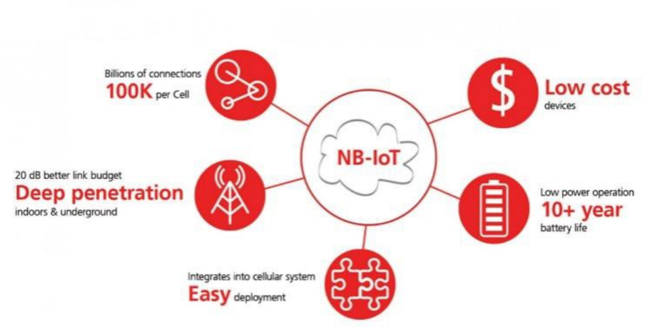

Fig. 3. Advantages of NB-IoT

C. Data Storage and Management Technology

In the Hadoop big data platform architecture, the bottom layer provides data storage and management services through the Hadoop Distributed File System (HDFS). The design goal of HDFS is to achieve high throughput data access on inexpensive hardware, while also having fault tolerance and high availability.

In the Hadoop big data platform architecture, the bottom layer provides data storage and management services through the Hadoop Distributed File System (HDFS). The design goal of HDFS is to achieve high throughput data access on inexpensive hardware, while also having fault tolerance and high availability.

In terms of data access and read/write, when the client writes data to HDFS, it communicates with NameNode and informs it of the file information to be written. NameNode returns a list of DataNodes containing storage data blocks. Then, the client divides the data into segments equal to the size of the data block and sends each segment to the corresponding DataNode. After receiving the data, DataNode writes it to the local disk and sends a confirmation to the client and NameNode. When reading data, the client requests the DataNode list containing the data block from NameNode and reads the data directly from DataNode.

The running process of HDFS file management and storage is shown in Figure 4.

HDFS has fault tolerance and recovery capabilities. Even if a node or data block fails, the system can still work normally. When a data block replica is unavailable, HDFS automatically selects another replica to provide data to ensure data reliability and availability. If a node fails, HDFS copies the data blocks on that node to other healthy nodes to achieve data recovery and rebalancing.[5]

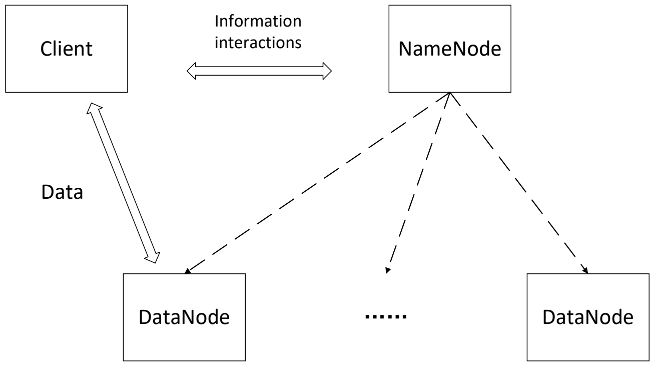

Fig. 4. HDFS Management and Storage Process

In summary, as one of the core components of Hadoop, HDFS provides a stable and efficient solution for storing and managing large-scale data. By dividing data into data blocks and storing multiple replicas in distributed clusters, HDFS achieves high throughput, fault tolerance, and high availability. At the same time, the design of HDFS also fully considers the efficiency and security of data access. By using NameNode as the master node to manage the namespace and metadata of the file system, all file read and write operations by clients are effectively authorized and coordinated, and the fault tolerance and recovery mechanisms safeguard the reliability and availability of data.

In addition, there are also some industry-leading solutions based on HDFS, such as HBase, which is suitable for fast read/write and random access to large-scale structured data. HBase is based on the storage and management technology provided by HDFS, and achieves high reliability, scalability, and deep integration with Hadoop, enabling real-time querying and high-speed writing of massive data. Integrating HDFS and HBase can build more complete and efficient big data processing and analysis solutions, better cope with increasingly complex data challenges.[6][7]

III. Equipment Management Platform Design

A. Platform Function and Architecture Design

The main functions of the IoT equipment management platform for parking IoTs include real-time monitoring and management, data acquisition and storage, data analysis and mining, parking space management and optimization, user services and experience enhancement, as well as reporting and visualization. The implementation of these functions can improve the utilization efficiency of parking resources, optimize parking flow scheduling, provide better parking services and user experience, and provide decision support and data references for parking IoT managers.

Specifically, in terms of real-time monitoring and management, the platform can monitor the status of IoT devices in the parking IoT in real-time, including the occupancy status of parking spaces and the working status of devices. By monitoring device data, managers can understand the operational status of the parking IoT in a timely manner and take necessary measures to adjust and optimize it. In terms of data acquisition and storage, the platform can collect a large amount of parking IoT data, such as parking space occupancy, vehicle access records, and device working status. After pre-processing and conversion, these data are stored in appropriate data storage systems, such as Hadoop distributed file systems or databases, for subsequent data analysis and applications.

In terms of data analysis and mining functions, the platform uses big data analysis technology to conduct in-depth analysis and mining of parking IoT data to extract valuable information and insights. For example, analyzing the utilization rate of parking spaces, predicting the change trend of parking flow, and discovering parking behavior patterns. These analysis results can provide decision-making basis for reasonable planning of parking resources and traffic flow scheduling.

In terms of parking space management and optimization functions, based on analysis results and real-time data, the platform can carry out intelligent management and optimization of parking spaces. It can help managers understand the use of parking spaces, including the number and location of empty parking spaces, as well as predicting future parking needs. Through intelligent parking space allocation and navigation, the platform can improve the utilization efficiency of parking resources, reduce the time for vehicles to find parking spaces, and alleviate traffic congestion.

In terms of user services and experience enhancement, the platform can provide a user-friendly interface and functions to facilitate users to select, reserve, and navigate parking IoTs. Users can obtain real-time parking space information, parking IoT capacity, and prices through the platform, which can improve their parking experience and satisfaction.

In terms of reporting and visualization, the platform can generate reports and visualizations to visually present the operational status and analysis results of the parking IoT. Through dashboards, charts, and reports, managers can quickly understand the real-time status, utilization, and performance indicators of the parking IoT, and make effective management decisions.

The functional architecture diagram of the IoT equipment management platform for parking IoTs is shown in Figure 5.

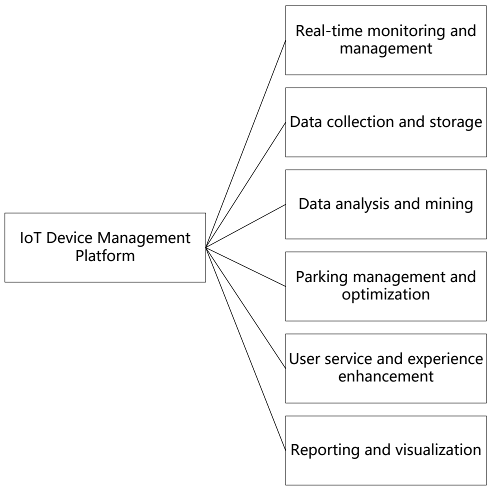

Fig. 5. Functional Architecture of IoT Equipment Management Platform

B. Data Analysis and Resource Scheduling

The designed IoT equipment management platform for parking IoTs can achieve data analysis based on Hadoop big data platform architecture. The basic workflow includes the following steps:

1) Data Collection and Import: IoT devices in the parking IoT, such as sensors and cameras, collect vehicle access information, parking duration, and other data. These data are

transmitted to the HDFS of the Hadoop cluster through the IoT gateway or other communication devices.

2) Data Preprocessing: Using tools and libraries in the Hadoop ecosystem, such as Pig and Hive, to clean, deduplicate, handle outliers, and transform data. Incomplete or invalid data can be identified and processed to ensure data accuracy and consistency.

3) Data Analysis and Mining: Based on the MapReduce programming model of Hadoop, custom MapReduce programs are developed to perform analysis and mining tasks for parking IoT data. In the Map phase, input data is split into multiple data blocks and processed in parallel on different nodes in the cluster. Each Map task on each node transforms input data into a series of key-value pairs. In the Shuffle phase, the output of Map tasks is sorted by key, and the values with the same key are grouped together. Then, these grouped data are transmitted to the node where the Reduce task is located. In the Reduce phase, the Reduce task processes the received data according to the key and generates the final results. The MapReduce computing model is shown in Figure 6.

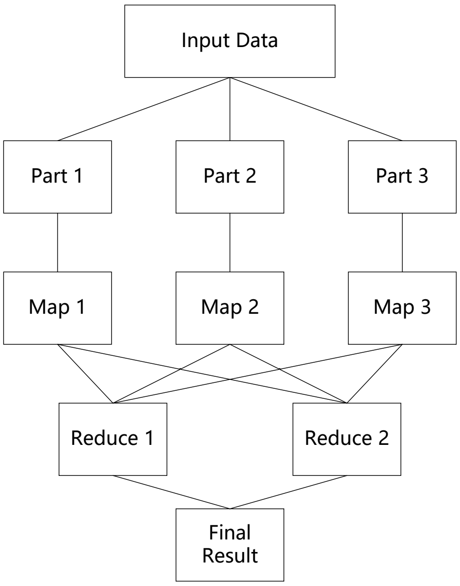

Fig. 6. MapReduce Computing Model

4) Data Application: After completing data mining and analysis, data can be displayed using visualization tools such as charts and reports, providing an intuitive display of the parking space usage, IoT device operation, and traffic flow distribution information and analysis summary.

The overall architecture of data analysis is shown in Figure 7.

In terms of resource scheduling in the parking IoT management platform, YARN modules are mainly used. The overall operation of this module mainly includes three parts:

1) Application submission: The application is submitted to the ResourceManager through YARN. The application consists of one or more tasks, and each task is managed by an ApplicationMaster.

2) Resource allocation: The ResourceManager allocates resources to the application by dividing the computing resources into containers and assigning them to ApplicationMaster.

3) Task scheduling: The ApplicationMaster is responsible for coordinating the execution of tasks. It communicates with the ResourceManager to request resources and assigns them to tasks. The ApplicationMaster also monitors the execution of tasks and handles failed tasks.

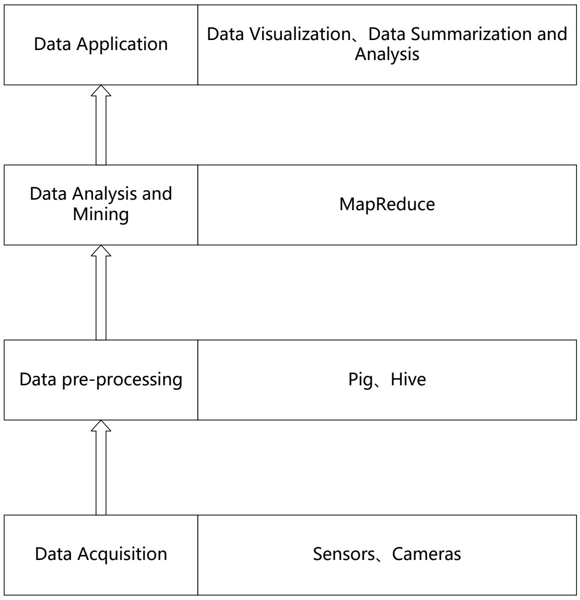

Fig. 7. Overall Architecture of Data Analysis

The resource scheduling process is shown in Figure 8.

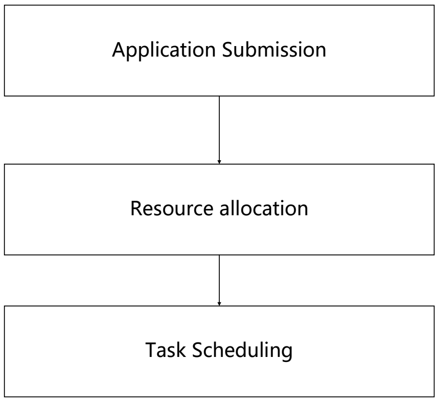

Fig. 8. Resource Scheduling Process

Through the workflow of YARN, the Hadoop cluster can effectively manage resources and schedule tasks. YARN provides a flexible resource allocation mechanism, allowing applications to dynamically obtain the required computing resources and implement parallel and distributed task execution in the cluster. This architecture allows the Hadoop cluster to handle large-scale data analysis and computing workloads.

C. Detailed Scheme for Data Application

The data application of the IoT equipment management platform for parking IoTs mainly consists of three parts: real-time data visualization, intelligent alarms, and data analysis.

In the real-time data visualization part, real-time parking IoT data, such as vehicle access information and parking duration, are analyzed and summarized. Combined with forms such as charts, dashboards, and maps, the usage of the parking IoT is displayed in real-time. This can intuitively show the distribution of traffic flow, parking space occupancy rate, and specific occupancy situation. In addition, dynamic 3D modeling can be achieved by using 3D modeling technology and real-time parking data, to monitor the operational status of the parking IoT in real-time and promptly detect abnormal situations such as congestion and low availability of parking spaces.

In the intelligent alarm part, the parking IoT is monitored based on real-time data, and corresponding alarm thresholds are set. When the parking status indicators (such as parking occupancy rate and queue length) or other analog quantities (such as temperature, humidity, and smoke concentration) exceed the set threshold, the system will issue intelligent alarms and remind relevant personnel to take corresponding measures.

The data analysis part mainly includes two sub-modules: historical data analysis and decision support. By analyzing historical data, the usage patterns, peak periods, and seasonal changes of the parking IoT can be explored. At the same time, machine learning algorithms can be used to analyze and predict the demand and trend of the parking IoT based on historical data. The predicted results can be used for decision-making such as optimizing resource allocation and vehicle flow scheduling to improve the utilization rate and service quality of the parking IoT. In terms of decision support, based on the results of data analysis and visualization display, decision support suggestions and insights are provided. Parking IoT managers can develop more reasonable parking strategies, adjust fee standards and improve operational processes based on real-time and historical data.

The framework diagram of the detailed data application scheme is shown in Figure 9.

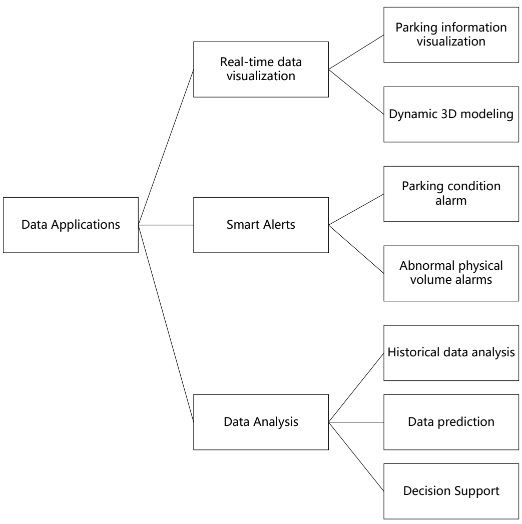

Fig. 9. Framework Diagram of Detailed Data Application Scheme

IV. Conclusion

This paper has investigated the scheme of using Hadoop platform and NB-IoT technology to achieve parking IoT devices management, and constructed a corresponding data processing and intelligent management platform, including data acquisition, processing, storage, analysis, visualization and display, etc. On this platform, functions such as intelligent warning, real-time monitoring, and decision support have been realized, which solves some practical problems in parking IoT device management and improves management efficiency and quality. By using Hadoop platform and NB-IoT technology, the parking IoT device management platform has the ability to handle big data and multi-source heterogeneous data. Meanwhile, the advantages of NB-IoT technology such as low power consumption, wide coverage, and high reliability are also beneficial to the comprehensive upgrade and optimization of parking IoT device management.

References

[1] Zhao J, Zhang J, Feng Y, et al. The study and application of the IOT technology in agriculture[C]//2010 3rd international conference on computer science and information technology. IEEE, 2010, 2: 462-465.

[2] Beyene Y D, Jantti R, Tirkkonen O, et al. NB-IoT technology overview and experience from cloud-RAN implementation[J]. IEEE wireless communications, 2017, 24(3): 26-32.

[3] Cheng Y, Zhao X, Wu J, et al. Research on the smart medical system based on NB-IoT technology[J]. Mobile Information Systems, 2021, 2021: 1-10.

[4] Babar M, Arif F, Jan M A, et al. Urban data management system: Towards Big Data analytics for Internet of Things based smart urban environment using customized Hadoop[J]. Future Generation Computer Systems, 2019, 96: 398-409.

[5] Zhang X, Wang Y. Research on intelligent medical big data system based on Hadoop and blockchain[J]. EURASIP Journal on Wireless Communications and Networking, 2021, 2021(1): 1-21.

[6] Bobade V B. Survey paper on big data and Hadoop[J]. Int. Res. J. Eng. Technol, 2016, 3(1): 861-863.

[7] Braten A E, Kraemer F A, Palma D. Autonomous IoT device management systems: structured review and generalized cognitive model[J]. IEEE Internet of Things Journal, 2020, 8(6): 4275-4290.