

# Implementation of 5G Oriented IoT Big Data Analysis

Asmeeta Rohit, Prof. Y.S. Thakur

*Department of Electronics and Communication Engineering, UEC, Ujjain*

Date of Submission: 28-03-2023

Date of Acceptance: 07-04-2023

**ABSTRACT-** The application degree and application scope of 5G Internet of Things technology and big data analysis technology are becoming wider and wider, bringing opportunities for the development of traditional enterprises and providing technological innovation support for the development of new enterprises. Based on 5G Internet of Things technology and big data technology, this paper designs and studies an intelligent agricultural monitoring platform. We collect crop growth data and monitor crop growth status through this platform to study the 5G-oriented IoT big data analysis method system. This paper studies the data collection and storage issues involved in the huge agricultural IoT data environment. This article analyzes the specific sources of agricultural big data, the specific methods of data collection, and the methods of various database storage technologies. Combining wireless sensor network technology, large-source data processing technology, and distributed data storage technology, a method is proposed to solve the problem of rural Internet data collection and storage in the big data environment. In the experimental part of this article, experiments are carried out on the key parts of the IOT-HSQM system model that may limit storage or query performance. Experimental results show that this article compares TSBPS and direct writing methods. The maximum write speed increased by 79%, and the average write speed increased by 42%. The IOT-HSQM system model can meet the requirements of compiling and query performance and statistical analysis.

## I. INTRODUCTION-

From the emergence of 5G-oriented Internet of Things technology to today's Internet of Things technology and big data analysis technology having entered almost every field of society, 5G-oriented Internet of Things technology is developing rapidly and will become more and more important in the future. Similarly, the Internet

of Things technology has also affected the agricultural sector, leading to major changes in production methods in many sectors.

In foreign countries, many scholars have conducted research on big data analysis methods and have achieved good results. The architecture combines the powerful functions of big data and emerging cloud computing technology and includes three key technologies, namely, multidimensional index based on distributed grid files, automatic SQL to conversion tool and hybrid data storage model, and support data update operation. Based on big data technology, analyzed the advanced operation management, maintenance schedule, and network planning and asset management of the distribution network and studied the auxiliary decision-making platform of the distribution network.

The importance of big data to the Internet of Things, an open IoT ecosystem framework with advanced communication and analysis technologies, and the integration of big data and the Internet of Things. The resulting new applications explored the big data technology framework in the Internet of Things.

## II. RESEARCH ON 5G-ORIENTED IOT BIG DATA ANALYSIS METHOD SYSTEM-

### 2.1 Data Collection and Design Based on 5G IoT Operating Platform-

- **Big Data Collection Based on Wireless Sensor Network-** The development and application of 5G-oriented Internet of Things technology in the agricultural field has slowly begun to spread, and the concept of 5G-oriented agricultural Internet of Things has been introduced. Based on the ZigBee protocol, sensor data will be collected on a computer and then can be stored in a database and displayed on the interface in real time. The data transmitted from the serial port to the computer is received through programming

and stored in the HBase database. HBase table database can store this modeling data in the table format we designed.

- **Data Migration-** Data migration mainly involves two aspects of migration: one is table conversion, and the other is application tools, such as sqoop and kettle tools. Sqoop supports a variety of relational database migration. There are several conversion methods to convert traditional arrays into HBase tables, such as basic conversion, segment conversion, and embedded conversion. The basic conversion is to add the functions of a traditional table to the column clusters in the table. Here, we mainly use sqoop tool for data migration.
- **Agricultural Internet of Things System-** The agricultural Internet of Things technology uses various sensors and other detection equipment to obtain and collect field information in various fields of agriculture, forestry, and such as cultivated land, horticultural parks, and aquaculture, and stabilize the continuous transmission of this data through a reliable transmission channel. Operators can make corresponding adjustments through the IoT agricultural system to maximize agricultural production profits. It can be seen that the agricultural Internet of Things is an Internet of Things application system that covers all aspects of agriculture, such as seedling selection, planting, fertilization, harvesting, production, processing, and circulation. Platform business-level technologies are applied in a multilayer structure from low to high, including the following five levels:
- **Transmission layer:** it consists of two parts: wireless communication (long-distance and short-distance) and wired communication (industrial Ethernet ring network). The backbone network is mainly wired communication, and the industrial Ethernet ring network is mainly used in the industrial field to ensure network transmission. Equipment level communication technologies include ZigBee, rf433, and RS232, which can provide reliable, convenient, and fast data transmission and exchange.
- **Business layer:** all types of information used in the growth factor and quality traceability monitoring system, including basic environmental data, business management information, geographic information, and video information. The quality inspection and quality traceability database of IoT

applications needs to be extended and improved by providing basic services such as business intelligence (BI), collaboration technology, and workflow.

- **Application layer:** it is divided into multiple entrances according to roles, namely, regulators, producers, suppliers, channel providers, and consumers.
- **User layer:** it uses multiple terminal methods based on Android, PC, IOS, and other multi-operating system query support architectures, including various information security methods, data storage, access and transmission security, through security systems and technologies to ensure safe access to information in the event of malicious attacks, to protect data security and data integrity. The main key technologies are information security technology and system security technology in network integration and heterogeneous system environments.

**2.2 Designing Interpersonal Data Query-** The fast statistical analysis model can ensure effective statistical analysis of the loaded data, but before statistical analysis, the query and loading of the data depend on the query speed of the storage system. Therefore, it is particularly important to increase the speed of receiving sensory data from the storage system.

**2.2.1. Establishing an Index for Intermediate Data Stored on HBase-** Druid's built-in index supports fast placement in many cases and will not become an obstacle to fast statistical analysis. Since HBase only supports fast row key scanning, for the intermediary perception of HBase storage, the correspondence between the unanswered key field and the array key of the common question must be determined. Similar to the HBase index processing method at the micro data level, the index is stored on the index server. The index server uses HBase and Redis to store index data. HBase stores all index data, and Redis temporarily stores index data. Since Redis is a cache based on fast access to memory, frequently accessed index data is stored in Redis. When running a data query, first get the results from Redis, which facilitates quick placement and saves query time. When searching index data, if the corresponding data is not in the Redis cache, it will search HBase and asynchronously replace the index cache on the index server.

**2.2.2 Query Design Priority-** When designing a query, follow the following principles:

- First query the cache, reduce the cache, and then query the mesoaware data from the data storage system.
- Query data in the mesoknowledge storage system. For the same query requirements, if Druid and HBase are used at the same time, Druid queries the data first.
- If the statistical analysis is to receive results from Druid and is based on a fast statistical analysis model, first obtain the results from Druid.
- When the query results of different rough and detailed data in the mediation perception data are available, the coarse-grained meso perception data will be queried first.
- When searching for data or performing analysis and calculations from a media perception system, you are performing tasks that do not interact with each other in parallel.

### 2.3 Building and Extracting Big Data-

**2.3.1. Data Mining Process-** The rapid development of ICT and computer technology has made data resources an increasingly important factor of production. The connection method of the Internet of Everything has led to an explosive growth in data volume, leading to changes in social driving forces. Data is the basis of information and knowledge, and the rapid growth of data has also led to the explosive growth of information and knowledge in human society. Data mining is the basic technology of knowledge discovery. Data mining refers to the process of discovering new relationships, trends, and patterns hidden in the data by processing and analyzing large amounts of data. Data mining is to extract information and discover knowledge from a large amount of data under unknown conditions. Data mining expands the application of data from basic data indexing and query to advanced applications of decision-making knowledge and analysis and prediction, which reflects the connotation of deep mining of data

**2.3.2. Data Preprocessing-** As the first and most important step of the entire data mining process, high-quality data is an important prerequisite for successful data mining. First of all, the goal of data mining must be clearly defined, and the business direction of the data mining application must be clearly defined. In other words, determine what important knowledge and information can be discovered from a large amount of data Once the target data is determined, the data must be processed (such as noise removal and data conversion) to make it high-quality data to be

mined. The following data mining algorithms are processed on this database.

**2.3.3. Data Mining-** According to the different characteristics of the data and the requirements of the user or the actual operating system, select specific algorithms for data mining, such as prediction, sorting, grouping, finding sequence patterns, finding related rules, or anomaly analysis. This article focuses on analyzing specific data mining algorithms.

**2.3.4. Data Analysis Screen-** If the data mining model is redundant or missing, or does not meet the needs of the user, it should be deleted and the previous steps of the process should be repeated. If the data mining model can meet the verification criteria and user needs, the results should be analyzed and expressed in natural language. The model or relationship obtained from data mining may be redundant or irrelevant to the mining goal, so it should be eliminated at this time. Even the results obtained do not meet the needs of users or do not meet the requirements of mining. For now, it is necessary to reselect data sources and algorithms and set new parameter values for re-mining. The result of mining is ultimately useful knowledge for users, so it must be visualized, and the mining results must be obtained after completion, induction, and evaluation.

**2.4 Data Collection and Agreement-** In the Internet of Things, most data collection is done through cutting-edge devices, which are often referred to as cutting-edge devices because they are usually located at the edge of the network. In other words, they exist in one place, such as a person or an element interacting with the network. The original data is received by these devices and can be retransmitted to the network. The Internet of Things is driven by economic value, which can significantly improve social and economic benefits and save social production costs.

**2.4.1. Data Collection Agreement-** In order to create an appropriate channel to receive data from the latest devices, although there is no consistent protocol to support data collection, similar mechanisms have emerged in specific areas. Of course, in order to accept the protocol between devices communicating with each other, a cloud capable of centralized analysis is needed. Some companies follow a specific path, while others are influenced by open source software.

**2.4.2. Data Integration and Management-**Data retention strategies are also important. This is because storing these huge high-resolution data is neither feasible nor necessary in many cases. Strategies include data collection and sampling and corresponding storage of source data, especially when the data has a fairly high level of redundancy. These data structures often provide various analysis tools and libraries.

**2.4.3. Big Data Technology Framework-** Hadoop execution analysis requires low latency requirements, including real-time model development, iterative, or interactive analysis, which is one of the main limitations of distributed settings. In this case, especially when multiple data (such as many machine learning algorithms) must be transferred, the Hadoop framework can be very expensive for basic HDFS communication. For example, the Spark framework was developed to solve Hadoop problems and has now evolved into its own ecosystem.

**2.4.4. Machine Learning and Data Mining-** From the organization or application level, a bunch of IoT terminals are needed. Stack elements should include M2M-level data collection, data processing functions, Internet data sharing, insight and device delivery functions (potential functions), and decision-makers (automatic or manual). Many groups and companies define the nature and structure of this IoT stack, which can provide the core framework, platform, related services, and solution development for front-end applications and computer and back-end support.

### III. EXPERIMENTAL RESEARCH ON 5G-ORIENTED IOT BIG DATA ANALYSIS METHOD SYSTEM-

**3.1. System Hardware Path Test-**The intelligent agricultural system based on the Internet of Things contains several functional modules. In order to ensure the normal connection of the intelligent agricultural system, a communication program is designed to test whether each unit can be successfully connected. The communication program mainly includes the following test points: gateway management, controller management, sensor management, and connection status display. In the system change settings, configure the IP address and port number of the client and server to ensure that communication can be connected.

**3.2. Comparative Experiment of TSBPS Writing Method-** Aiming at the experiment of the data

writing throughput of the micro sensing data layer, different amounts of data are extracted from the historical data of the sensor for experimental writing. When writing, HDFS uses the IOT-HSQM directory structure and file naming method. Compare the average data writing speed and maximum data writing speed of direct data storage in HDFS, as well as the caching method, and then group them according to space-time blocks. In order to ensure the accuracy of the test results, most of the original perception data will be temporarily stored in the memory, and when writing, the display time of the data will be modified to the current time in advance. The experiment should use the maximum writing speed and the maximum average speed of the two writing methods, where the maximum writing speed refers to the maximum instantaneous writing speed that can be achieved during the writing process, and the average maximum writing speed refers to maintaining stability. Under this premise, the maximum write speed that can be maintained without exceeding the Redis cache threshold is the standard here.

In the experiment, 12 simulated writing clients were deployed on 6 computers. Socket Server is used to monitor the actual status of Redis and uniformly control the upload speed and data volume of each simulated data terminal

**3.3. Storm System Performance Experiment and Data Collection-** In the continuous data generation process in the data source system, too much data may appear. At present, the computing power of the Storm system will be reduced, and the entire system may crash in severe cases. In order to test whether the Kafka message queue in the system can temporarily store the data cache when the data increases, multithreading technology is used to control its performance. Each thread produces the same amount of data, which can be verified by comparing the number of threads and computing performance.

### IV. EXPERIMENTAL RESEARCH AND ANALYSIS OF 5G-ORIENTED IOT BIG DATA ANALYSIS METHOD SYSTEM-

**4.1. Application Analysis of K-means Clustering Algorithm in Plant Growth Condition Clustering-**

We put the experimental data through 15 iterations and put the results of each iteration into the out folder, and check the files to get the final cluster centers and clustering results. After five sets of test data, almost each requires 15 iterations to get the final category. After five sets of test data,

the similarity in the same category is higher and more compact. Run the K-means Map Reduce program, after each iteration, a new cluster center is generated, and then iterated, and finally all the data

is divided into 4 clusters: Finally, output the results of our iteration in the out folder, and finally the iteration is completed and divided into 4 clusters. Table 1 shows the centers of 4 clusters.

Table 1 Algorithm execution cluster center point results.				
	Cluster1	Cluster 2	Cluster 3	Cluster 4
Ambient temperature	26	15	23	28
Air humidity	58	42	46	53
Carbon dioxide concentration	326	418	276	341
Light time	9	11	7	6
Soil moisture	25	26	28	30

As shown in Table 1 from the results of the clustering, the optimal state of the cluster shows that the suitable temperature for tomatoes is 24–26 degrees during the day. When the temperature and humidity are high, the light is between 30,000 and 40,000 lux. The relative humidity is 65–85%, the relative air humidity is 45%–65%, the pH is 6–7, and the concentration of carbon dioxide in the air is 300 micro liters/liter. Some variables close to the center are clustered into this category, and then plants grow well under this category of conditions. We can adjust the conditions of different greenhouses according to these conditions to improve plant growth. From this, we can know that such an approximate range is better for our plant growth. In other conditions, we try to gather those conditions to the variable range of the conditions, and adopt heating, increasing carbon dioxide concentration, and increasing light intensity.

It can be seen that the speed-up ratio between the cluster and the time of a single machine will be more obvious as the number of computing nodes increases. However, when the amount of data is small, the time it takes to initialize the Map Reduce calculation cannot be ignored, and the calculation time consumes more than the time of a single machine. When the amount of data is small, the extra time comes from the overhead of the Hadoop library. Before starting any map or reduce tasks, this library needs to do a lot of checks, but once started, Hadoop’s mapper and reducer will run at full speed. A single machine is more suitable for calculation of small data volume, as shown in Figure 1.

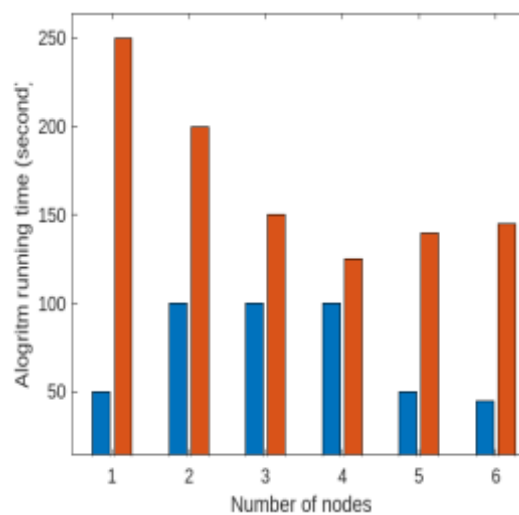


Figure 1

Comparison of the time it takes to run K-means clustering on different samples.

In this figure we can use MATLAB-

```
y=[50 250; 100 200; 100 150; 100 125; 50 140; 45 145];
bar(y)
xlabel('Number of nodes');
ylabel('Alogritm running time (second)');
```

The result of the calculation is shown in Figure 1. It can be seen that the cluster takes more time than the single machine, and as the number of nodes increases, the time change will not be too

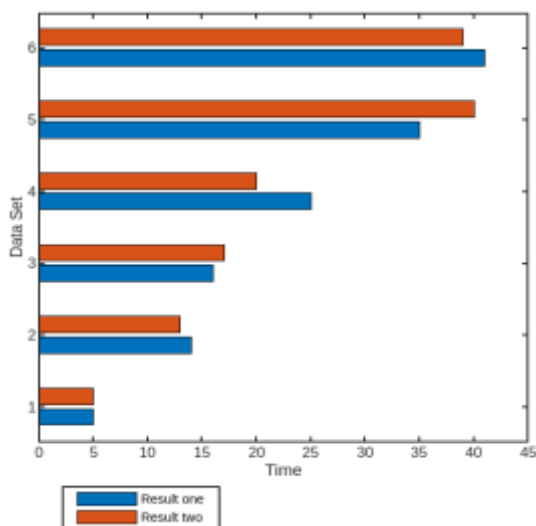
large. The data in this article has still more clusters than single machines, relatively speaking, it takes more time than the first group, and the cluster time has decreased with the increase in the number of nodes.

**4.2. Storm System Performance Test and Analysis-** In the process of continuously generating data by the data source system, excessive data volume may occur. At this time, the computing power of the storm system will decline, and the entire system may crash in severe cases.

Each thread produces the same amount of data, which is verified by comparing the number of threads and the computational efficiency. Part of the data is shown in Figure 2. The method of repeatedly sending data is used to achieve the requirement of large data volume. The storm system accumulates the abnormal values in the sent data to test storm's performance system.

It can be seen from Figure 2 that, with the increase in the amount of calculated data, the calculation time curve of storm does not show a geometric change, but an approximate straight line. Therefore, it can be concluded that the increase in the amount of calculation data does not weaken storm's computing power

For the read performance test, when 10,000, 100,000, and 1 million records have been written in the data table, test the query of 10, 100, and 1,000 records from the data table, respectively; the required system response time is shown in Figure 3.



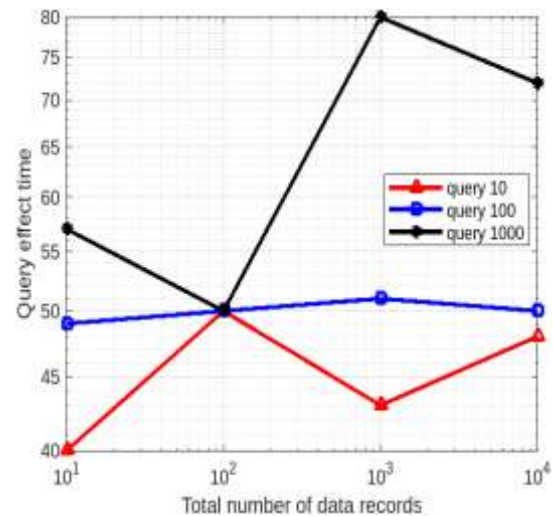
**Figure 2**

**The relationship between the amount of data calculated by the storm system and the calculation time.**

In this figure we can use MATLAB-

```
y=[5 5; 14 12; 16 17; 25 20; 35 40; 42 39 ];
barh(y)
xlabel('Time');
ylabel('Data set');
```

As shown in Figure 3 it can be seen from the test results that the response time of the data query is generally maintained at the millisecond level. When querying in the same data table that stores the data records, the system time-consuming will increase with the increase of the query data. When the number of records in the data storage is different, but the number of records in the query is the same, the system response time will increase as the total number of storage tables increases.



**Figure 3**

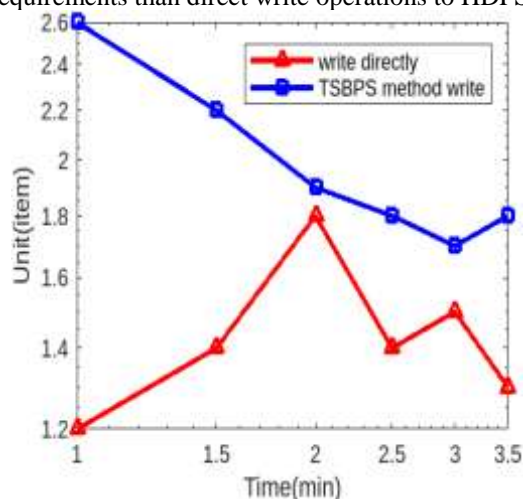
**Data query performance test result graph.**

In this figure we can use MATLAB-

```
x=[10 100 1000 10000]
y1=[40 50 43 48]
y2=[49 50 51 50]
y3=[57 50 80 72]
loglog(x,y1,'-^r','Linewidth',2)
hold on
loglog(x,y2,'-sb','Linewidth',2)
hold on
loglog(x,y3,'-k*','Linewidth',2)
hold off
xlabel('Total number of data records')
ylabel('Query effect time')
legend('query 10','query 100','query 1000')
grid on
```

**4.3. Experimental Analysis of the Writing Method of the Micro sensing Data Layer-** The microware data layer is based on the TSBPS method of writing, and the maximum instantaneous speed depends on the data compression situation and the service capacity of the Redis cache. The maximum average speed is the data speed that can be received when the amount of data cached by Redis is basically stable. It is restricted by data compression and the speed of Redis writing to HDFS. In theory, the speed of writing to HDFS from Redis can be close to the maximum of HDFS and since the data written is compressed data, compared to the original data, the writing speed is further improved according to the compression ratio. The experimental results are shown in Figure 4.

As can be seen from Figure 4, the TSBPS method is compared with direct writing, the maximum writing speed has increased by 79%, and the average writing speed has increased by 42%. The writing speed of TSBPS method is much higher than that of direct writing. Therefore, the method of caching and then clustering writing in the model can better meet the high throughput requirements than direct write operations to HDFS.



**Figure 4**  
**Write speed comparison chart.**

```
In this figure we can use MATLAB-
x=[0 1 1.5 2 2.5 3 3.5]
y1=[1 1.2 1.4 1.8 1.4 1.5 1.3]
y2=[2.7 2.6 2.2 1.9 1.8 1.7 1.8]
loglog(x,y1,'-r','Linewidth',2)
hold on
loglog(x,y2,'-sb','Linewidth',2)
hold off
xlabel('Time(min)')
ylabel('Unit(item)')
```

legend('write directly','TSBPS method write')

## V. CONCLUSION-

This paper designs an agricultural data processing model represented by sensor data in the agricultural field. First, determine the characteristics and types of agricultural information, and then sub modules to realize the various functions of the station, and involve the collection, transmission, storage, and processing of the entire agricultural field data. The data interface unit and data loading unit are based on wireless sensor networks. The data storage unit based on the distributed file system and the table structure for storing data, sensor data, and file management module are data query modules, user access, and user management modules that apply web query. Based on the research and analysis in the field of agricultural Internet of Things, this paper conducts design research on data collection and storage in the field of Internet of Things under the big data environment. In order to collect data from the sensor network in the agricultural Internet of Things, the real-time data processing system storm big and HBase data storage were analyzed and researched. Then, various new technical problems appeared. When developing sensor networks, more reliable, stable, and efficient structures can be explored, and data fusion technology of nodes in the network can be used to reduce the energy consumption of data transmission. In the big data processing system, further research and exploration are needed to achieve load balancing and system processing efficiency, and the application of the system in data collection and data mining needs to be further improved. The research on offline data processing also needs further research. This paper proposes a spatiotemporal block processing TSBPS to store the first detection data. The method uses spatiotemporal preblocking, data compression, and caching to significantly improve the recording speed of near real-time storage and micro detection data

## REFERENCES-

- [1]. Z. Lv and A. K. Singh, "Big data analysis of internet of things system," ACM Transactions on Internet Technology (TOIT), 2020.
- [2]. S. S. Gill, I. Chana, and R. Buyya, "IoT based agriculture as a cloud and big data service," Journal of Organizational and End User Computing, vol. 29, no. 4, pp. 1-23, 2017.
- [3]. N. Saxena, A. Roy, B. J. R. Sahu, and H. Kim, "Efficient IoT gateway over 5G

- wireless: a new design with prototype and implementation results,” IEEE Communications Magazine, vol. 55, no. 2, pp. 97–105, 2017.
- [4]. S. Rajeswari, K. Suthendran, and K. Rajakumar, “A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics,” International Journal of Pure and Applied Mathematics, vol. 118, no. 8, pp. 365–369, 2018.
- [5]. M. Hussain Iqbal, T. R. Soomro, and T. Rahim Soomro, “Big data analysis: Apache storm perspective,” International Journal of Computer Trends and Technology, vol. 19, no. 1, pp. 9–14, 2015.
- [6]. a.Ahad, M. Tahir, K.-L.A. Yau, 5G-based smart healthcare network: Architecture, taxonomy, challenges and future research directions. IEEE Access 7, 100747–100762 (2019).
- [7]. Mistry, S. Tanwar, S. Tyagi, N. Kumar, Blockchain for 5G-enabled IoT for industrial automation: A systematic review, solutions, and challenges. Mech. Syst. Signal Process. 135, 106382 (2020).
- [8]. S. Rajeswari, K. Suthendran, and K. Rajakumar, “A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics,” International Journal of Pure and Applied Mathematics, vol. 118, no. 8, pp. 365–369, 2018.
- [9]. Z. Lv, R. Lou, J. Li, A. K. Singh, and H. Song, “Big data analytics for 6G-enabled massive internet of things,” IEEE Internet of Things Journal, vol. 8, no. 7, 2021.
- [10]. B. Tang, Z. Chen, G. Hefferman et al., “Incorporating intelligence in fog computing for big data analysis in smart cities,” IEEE Transactions on Industrial Informatics, vol. 13, no. 5, pp. 2140–2150, 2017.