*International Journal of Internet, Broadcasting and Communication Vol.12 No.4 218-225 (2020)*
*http://dx.doi.org/10.7236/IJIBC.2020.12.4.218*

IJIBC 20-4-27

# Design and Development of Big Data Platform based on IoT-based Children's Play Pattern Analysis

Seon-Jin Jung*

**Part-time Instructor, College of liberal and convergence, Honam University, Korea*
*icare79@naver.com*

***Abstract***

*The purpose of this paper is to establish an IoT-based big data platform that can check the space and form analysis in various play cultures of children. Therefore, to this end, in order to understand the healthy play culture of children, we are going to build a big data platform that allows IoT and smart devices to work together to collect data.*

*Therefore, the goal of this study is to develop a big data platform linked to IoT first in order to collect data related to observation of children's mobile movements. Using the developed big data platform, children's play culture can be checked anywhere through observation and intuitive UI design, quick information can be automatically collected and real-time feedback, data collected through repeaters can be aggregated and analyzed, and systematic database can be utilized in the form of big data.*

***Keywords:*** *IoT, Big-data, Children's play culture, Hadoop, HDFS.*

## 1. Introduction

Recently, various big data collection has become possible in real time due to the development of information and communication and wide-area networks, and ways to utilize them in each industry are being discussed. Recently, demographic data using public data, the number of businesses by region/industry and average sales information are included, and the Seoul Metropolitan Government provides services using public data such as the Woori Village Shop Commercial Area Analysis Service [1]. It is also developed and applied as a technology to prevent accidents in advance by collecting, analyzing, and predicting information on location, operation status, etc. by attaching sensors to facilities, equipment, etc. to various workplaces by applying IoT and big data technologies to prevent accidents, such as work personnel and facilities [2]. These big data used in a variety of industries can be defined as data with the Volume, Velocity, and Variety attributes [3]. Recent advances in technology and data have added meaning of value and accuracy to Value, Veracity, defined as 5V [4].

Typical analysis methods for big data include Word Cloud, Topic Modeling, and Semantic Network Analysis [5]. The frequency analysis used in Word Cloud is the simplest and most commonly used of text-based analyses, and can give a brief overview of what is emphasized in the literature. Topic Modeling is a technique for extracting and analyzing keywords by topic and topic, and the setting of the number of topics is

Manuscript Received: October. 12, 2020 / Revised: October. 18, 2020 / Accepted: October. 23, 2020
Corresponding Author: icare79@naver.com
Tel: +82-62-952-1318, Fax: +82-62-952-1319
Part-time Instructor, College of liberal and convergence, Honam University, Korea

determined by the researcher's judgment, which ultimately translates the classification according to the topic modeling. Semantic Network Analysis is an analysis method that applies social network theory and analyzes network composed of connections between objects in a structured form. Connectivity centroid refers to the direct influence of a node in a network and identifies how many other nodes it connects to. Through the interaction relationship of the language structure that constitutes text, it is a method of analyzing and explaining meaning [6].

For big data analysis, this paper aims to establish a big data platform in which IoT and smart devices can work together to collect data. Therefore, the final objectives of this paper are to utilize the developed big data platform to check children's play culture from place to place through observation and intuitive UI design, to enable quick information collection and real-time feedback, to aggregate and analyze data collected through repeaters, and to utilize structured databases in the form of big data.

## 2. Associated Research

### 2.1 Big Data Research

Recently, a large amount of data used in various industries has been used to digitize and store, and big data has been processed. The size and variety of such big data are also becoming more advanced and expanding in the wake of the rapid development of storage and memory semiconductor technologies. In addition, collection, storage, processing and analysis of big data in the medical, bio and energy sectors, as well as handling multimedia big data, which was previously easily accessed and utilized, such as images, video and audio, are becoming increasingly possible [7].

Big data processing is a process that supports processes of data processing or data analysis to find useful information and meaningful knowledge in big data. There are various ways to process and analyze big data. There are Hadoop, MapReduce, Hbase, and non-relational databases. Hadoop is an open source software framework that supports large-scale data processing analysis on distributed systems, and is also the result of Google's implementation of MapReduce as an open source. H-base is an open source distributed nonrelational database developed by referring to Google's Big Table.

MapReduce was implemented in Hadoop, a software framework proposed by Google to handle large data sets on distributed systems. Nonrelational databases are databases that do not store data in tables and have a contrasting concept with relational databases [8].

Big data requires a preprocessing process to utilize various data collected in various environments. Refining consistent data by removing unnecessary data through pre-processing process. The process of pre-processing big data is distinguished as shown in Figure 1.

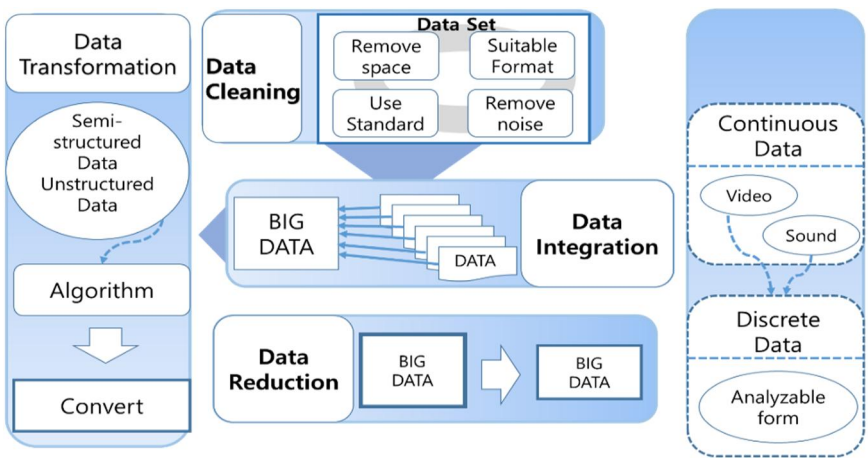


**Figure 1. A typical big data the pre-treatment process**

### 2.2 Hadoop Data Distributed File System

Among big data processing systems, Hadoop is a Java-based open-source framework that can distribute large amounts of data, and the distributed file system (HDFS) is drawing attention as a computing technology that can use inexpensive computers as if they were one. Figure 2 shows a distributed file system structures [9].

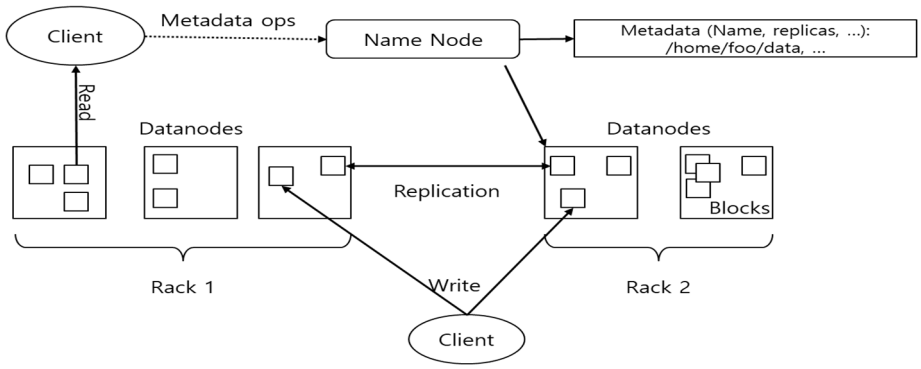


**Figure 2. Hadoop distributed file system architecture**

HDFS architecture consists largely of a namenode that manages the names, locations, and other metadata of files, and a datanode that is responsible for processing (reading, writing) actual data. The process of reading a file from HDFS can first request a file from the client to namenode, and then signal the client to the datanode location where the file resides, and send or read data while the client communicates directly with datanode. The role of namenode in Figure 2 here is to retain metadata about files, directories, back up files that supplement metadata's persistence, determine failures through datanode Monitoring, manage blocks to store data blocks in new datanode or in case of datanode failure, and checkpoint operations using editslog and memory fsimages. datanode's role is to block the file, which is the actual data, and the block size of the file is fixed at 64/128MB, and sends a heart beat signal to namenode every 3 seconds.

### 2.3 Existing Research

The integrated heat demand management system based on the big data platform was proposed to address a number of chronic problems in management, including energy loss, deterioration in heating quality, and failure of facilities, through a collective energy supply method that produces large amounts of heat energy at a given regional level and supplies it to thermal user facilities such as multi-family housing, public and business buildings [9], and Figure 3 [7].

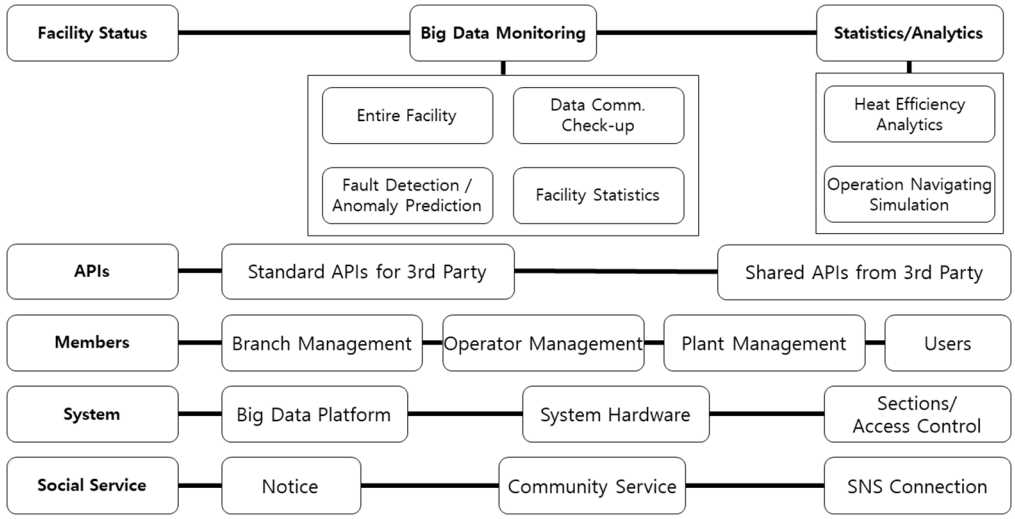


**Figure 3. Design of heat demand integrated management system**

## 3. Design of the proposed IoT-based Big Data System

Figure 4 is a structure designed to build an IoT-based big data platform that can check the space and shape analysis of various play cultures of children proposed in this paper.

It is designed to observe children's movements using IoT, to automatically collect, aggregate and analyze related data, and to utilize a structured database in the form of big data.

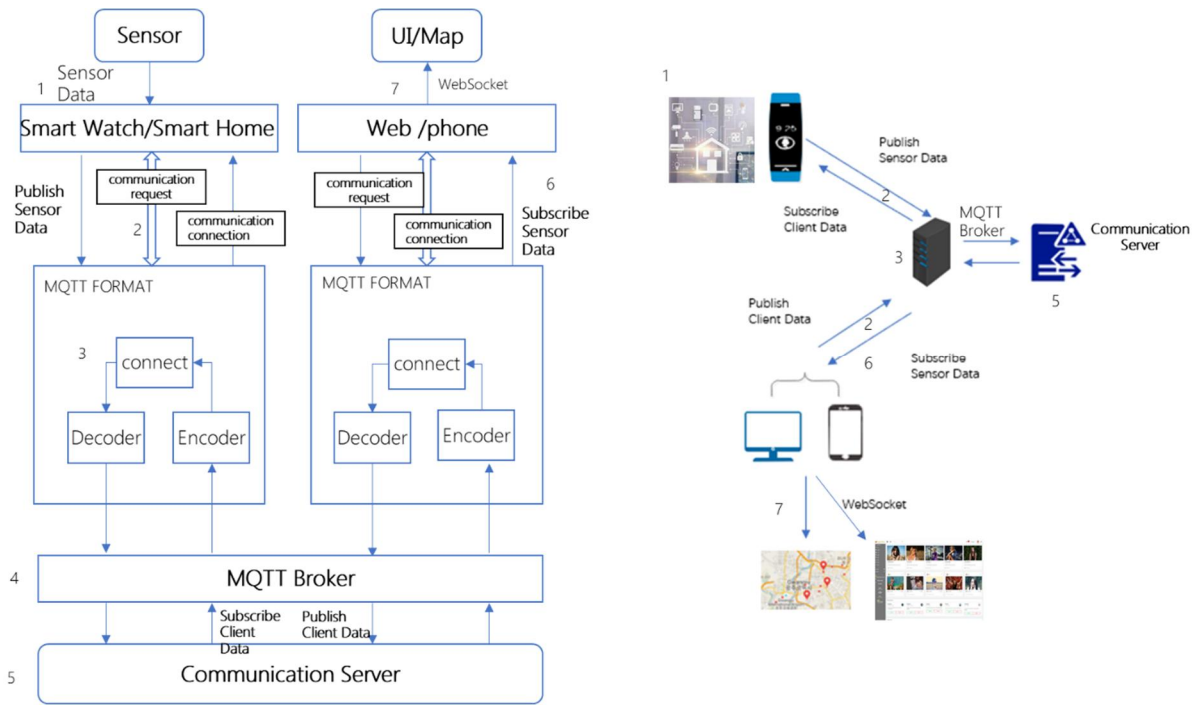


**Figure 4. Structural drawing of IoT-based big data system**

The MQTT used here is designed with a broker, Publisher, and Subscriber structure, rather than a client-

server structure, such as HTTP, TCP, etc.

①Sensor Data: Smart Watch/Smart Home records user behavior through IoT sensors

②Commination request: Smart Watch/Smart Home applies for connection to MQTT Server

③MQTT FORMAT: Data obtained from smart devices is converted to MQTT format or posted to MQTT Broker.

④MQTT Broker Server: Checked the transmitted data and converted to json format

⑤Communication Server: Data sent by MQTT Broker is acquired and processed and the results are sent back to MQT Broker in json format.

⑥Subscriber Sensor: The MQTT server sends data from the communication server to the Web/Smart Phone

⑦Web/ Smart Phone is displayed on a map or UI via WebSocket

Figure 5 shows the following. The main server of the big data system deployed utilized one MQTT server, one intermediate communication server, and consisted of DB Server, Admin Server, Log Server, and Bag Data Server. Typically, a distributed-processed server system consists of a master node that commands and controls other nodes to handle a given task, and a slave node that processes data by performing the control signals and commands and commands. It is designed to extract data from Offline (2-4) and Online (5-6).

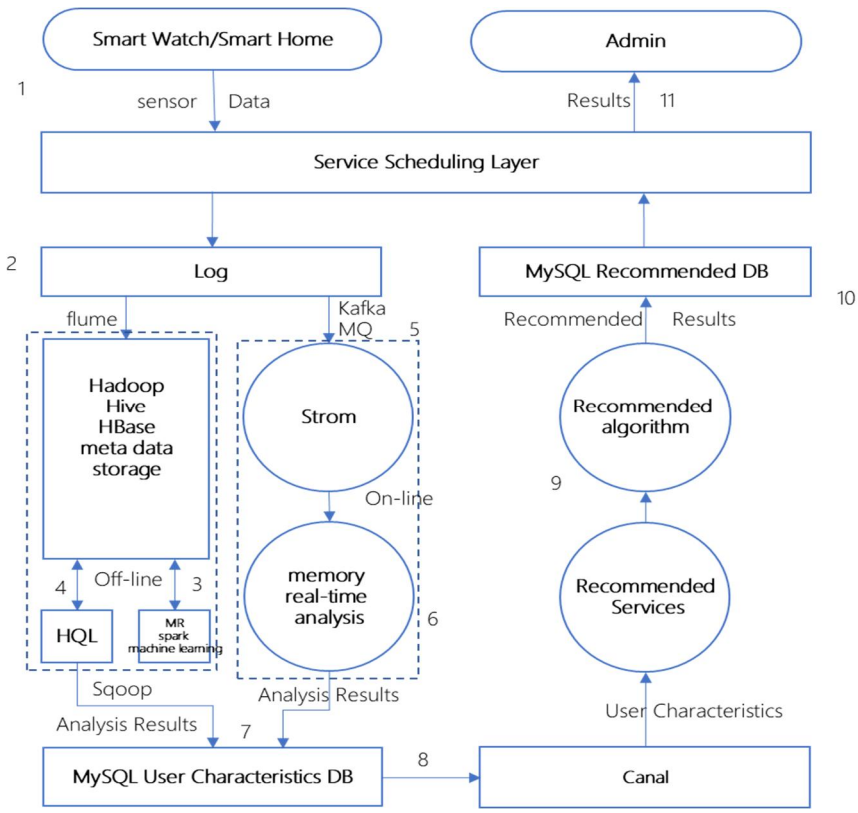


**Figure 5. Design of big data architecture**

①Smart Watch/Smart Home : Data from sensors is sent to the task scheduling layer via RocektMQ

②Log: Receives data from Log Server via flume technology from hadoop

③MR(MapReduce)/spark machine learning: Extracted data is analyzed using technologies such as MR speak spark machine learning (a distributed programming model that divides processing into Map and Reduce steps to process large data)

④HQL(Hibernate Query Language): Store the analysis results of ③ as HQL to prepare for re-

⑤Kafka MQ: Log data and real-time data sent to Strom via Kafka for quick analysis of real-time data

⑥memory real-time analysis: Transfers real-time data from the user to extract user characteristics(gender, age,..) such as Spark/ Strom/Flink, etc.

⑦MySQL User Characteristics DB: Store user features in MySQL

⑧Canal: Send user features stored in MySQL to Recommended Services via Canal

⑨Recommended algorithm: Recommend data analyzed through extracted user characteristics results

⑩MySQL Recommended DB: ⑨Recommended data is saved as MySQL

⑪Admin: Communicate final data to the administrator

Figure 6 shows the structure of the communication server in this paper. The communication server used here sends data to each server via Rocket MQ.

The Feign Client helps you create Web Service clients more easily, and when you create an interface and attach Annotation, code complexity is reduced because it can be used without details. Rocket MQ is a solution specialized in large capacity real-time log processing, and delivering data safely without loss is the main focus of processing data with stable architecture and fast performance in the message system.

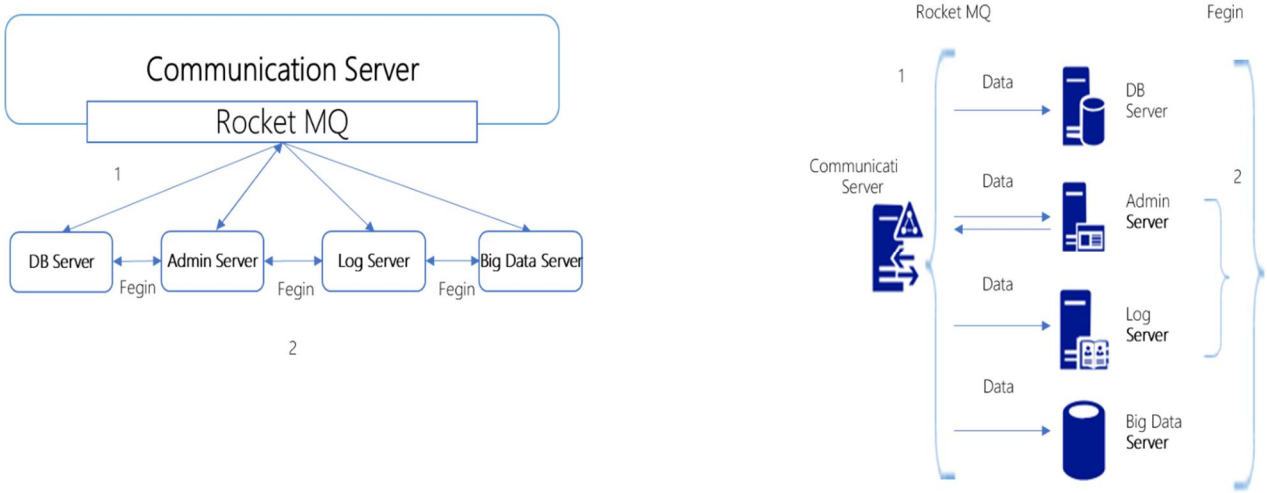


**Figure 6. Communication server structure diagram**

## 4. Development of UI to Check the Location of Children

Figure 7 is the UI of a big data system using IoT. The function of the UI enables interactive communication using IoT and mobile applications on the map, and the moving position of children is displayed as coordinates.

The moving section of children is stored as data in the big data system. The developed UI can be checked by dividing the location and movement of children by hourly, daily, weekly, and monthly. It was also developed to be able to check on mobile. A telephone connection is possible between a teacher in charge of field trips and a student, and it is designed to easily track the location of a child who has fallen behind. All these data can be collected and managed.

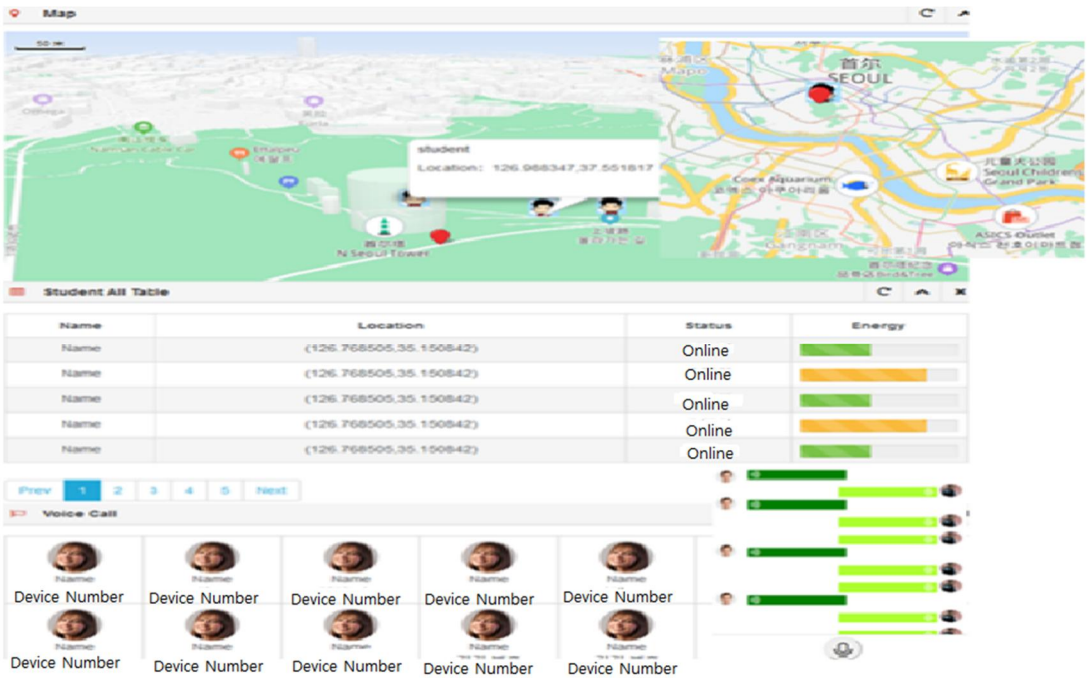


**Figure 7. UI of the developed big data system**

## 5. Conclusion

In this paper, a customized big data platform was designed based on IoT-based child play patterns analysis. The proposed big data platform wants to use big data analysis to see how the space and behavior patterns in children's play culture appear. These findings help teachers understand the characteristics of children immersed in play, and can also be used as basic data for further research on how play behavior patterns can affect the development of children.

## References

[1] Seoul Metro City, https://golmok.seoul.go.kr/main.do
[2] W. L. Kang, H. G. Kim, and Y, J. Lee, "Reducing IO Cost in OLAP Query Processing with Map Reduce," IEICE Trans. Information. & System, Vol. E98-D, No. 2, pp. 444-447, Feb. 2015.
[3] Hyeon Gyu Kim. “Developing a Big Data Analysis Platform for Small and Medium-Sized Enterprises,” Journal of The Korea Society of Computer and Information, 25(8), pp.65-72, 2020.
DOI: https://doi.org/10.9708/jksci.2020.25.08.065

[4] Shin Dong-jin, Hwang Seung-yeon, Lee Don-hee, Moon Jin-yong, Kim Jung-jun, "Analysis of Professional Baseball Data based on Big Data," Korean Internet Broadcasting and Telecommunications Association, Vol.20, No. 4, pp.177-185, 2020. DOI: 10.7236/JIIBC.2020.20.4.177
[5] Mun-Ki Kim, Hee-jung Choi, Sang-Lin Han. "Research Trends in Distribution Study by using Big-Data Analysis," Journal of channel and retailing, 25(3), pp.85-103, 2020.
[6] Jeongmyeong Lee, Hyesun Kim and Junho Choi. "A Study on User Experience Factors of Display-Type ArtificialIntelligence Speakers through Semantic Network Analysis: Focusing on Online Review Analysis of the Amazon Echo," The Journal of the Convergence on Culture Technology (JCCT), Vol. 5, No. 3, pp.9-23, 2019. DOI: https://dx.doi.org/10.17703/JCCT.2019.5.3.
[7] Mingoo Song. "Development of Big Data System for Energy Big Data." KIISE Transactions on Computing Practices 24(1), 24-32, 2018.
[8] Seo-ha Jeong, Chang-hyun Kim, In-Jae Choi, Yeong-Jin Kim, Jin-suk Bang. "A Study on the Web Service-Based Platform for Pregnant Women Based on Big Data," The Korea Institute of Information and Communication Engineering, 24(1), 486-489, 2020.
[9] M. Song, "Development of Heat Demand Management System for District Heating based on Big Data Platform," Communications of the Korean Institute of Information Scientists and Engineers, pp. 31-33, 2017.
[10] https://cyberx.tistory.com/141