Peculiarities of creating a database for the IoT system of urban forest management in the city of St. Petersburg

O Kolbina¹, E Istomin¹, N Yagotinceva¹, T Safonova¹ and M Kalambet²

¹Department of applied information science, Russian state hydrometeorological university, Voronezhskaya Street 79, St. Petersburg 192007, Russian Federation
²Joint-stock company "Mashtab research institute", Kantemirovskaya Street 5, letter A, St. Petersburg 194100, Russian Federation

*Corresponding e-mail: olya_kolbina@mail.ru

Abstract. The article deals with the issues related to problems of creating databases for IoT-systems in general and in conjunction with geographic information systems of urban forest management. Highlighted are specific problems and ways of solving them on the basis of modern information technology. As one way to solve most of the problems faced by the developer, the paper reviews the use of open source database management system Tarantool with a non-blocking application server. The article also describes a tree structure for mapping devices, which are connected to the IoT system. It greatly simplifies their management and mapping. The results described in the article allow to judge the possibility of combining two systems such as a geographic information management system and the Internet of Things.

1. Introduction

As the standard of living is rising, the indicator of the "good life" has shifted from the concept of income to the concept of quality of life. One of the criteria in forming this concept is the possibility to be in harmony. The beneficial effects of nature on the ecology, physical and spiritual condition of human beings have been known for a long time.

A study conducted by the United Nations World Health Organization (UNWHO) found that city dwellers should have at least 50 square meters of urban green space and 300 square meters of suburban forests. Back in 2005, the urban area per inhabitant in St. Petersburg was 54.7 m² and 306.7 m² respectively. Since that time, massive urban planning and expansion has resulted in the clearing of nearby forests, leading to a decline in the quality of this indicator [1]. According to recent data, the area of forests from 1990 to 2020 has decreased by almost 2 times.

St. Petersburg occupies one of the first places on the green area in comparison with the countries of Western Europe. This cannot be said about the number of personnel who carry out activities for the protection, care, custody, reproduction and use of urban forests, as well as technical capacity.

At present, the management of urban forests is the responsibility of the executive body of state power of St. Petersburg, the Committee for the Urban Improvement of the Government of St. Petersburg. According to the statute of the committee, it is designed to ensure the implementation of state policy and state management in the field of landscaping gardening, forest park management; maintenance of roads and other objects of improvement; waste management on the territory of St. Petersburg [2]. The tasks of the city committee indicate the presence of numerous objects of the green areas, including urban forests, located in one management system. The management process is
complicated by the distribution of management elements over the city and region territory, forcing the use of geographic information technology to automate the management process, which are designed to bring together spatial data. However, technological advances and minimization of manpower in the subordinate units bring about the need for more extensive automation of the management process and the introduction of IoT-technologies.

2. Methods and Materials

2.1. Implementation of infocommunication systems in the management process using GIS

The introduction of information and communication systems in the management process using GIS should simplify access to information about the on the ground situation and reduce the time for managerial response. Each city district now may have special structural units engaged in [3]:

- measures to protect forest fires;
- measures to protect forests from illegal logging;
- measures to protect forests from pests and forest diseases;
- measures to protect wildlife objects and their habitats;
- biotechnical activities;
- reforestation activities;
- carding of forest plantations.

With the introduction of IoT technologies, information about the state of forest areas, parks and urban forests will be obtained automatically through all kinds of communication channels (figure 1).

![Figure 1. Scheme of obtaining GIS information using IoT technology.](image)

Every minute terabytes of information from various sensors and devices connected to the Internet are transmitted in space. This data is to be processed it and presented for analysis.

At the moment information technologies are already used to monitor the state of forests to prevent fires and combat illegal logging. Wide-angle cameras and heat sensors are installed. It remains to connect them into a unified monitoring system.

Any device for data transfer can serve to collect information, ranging from GIS-systems in forest districts to the internet posts of the forest parks and squares visitors.
To solve technical problems of construction and integration of geoinformation systems it is necessary to solve a number of problems of [3]:
- data collection, storage and analysis;
- handling big volumes of heterogeneous data;
- presenting the collected data on different GIS layers;
- visualization of the necessary data.

2.2. Features of creating databases for the geographic information management system with IoT elements

Databases (DB) have been and remain the heart of any information system, but creating a database for an IoT-supported system requires a special approach. One of them is to create a mix of all kinds of database technologies.

The peculiarity of databases of a geoinformation system that works with IoT technologies is many terabytes of different-format information coming in at great speed from tens of thousands of devices and sensors. This requires a database capable of reading and recording series of data in distributed time aspects, as well as supplementing applications with computational capabilities. Reliability is another criterion for the database, because one of the conditions of IoT-technology application is constant access to the data storage. Easy scalability and specificity in responding to requests leads to a debugged operation and prevent latency and obsolete data accumulation. [4]

There is no universal database for the Internet of Things system, because of its construction complexity and the volume and speed of incoming information.

IoT-system of urban forest management in Saint Petersburg should be a complete, scalable and fault-tolerant IoT-platform, that would support the following features:
1. Connecting simultaneously hundreds of thousands of devices.
2. Receiving millions of events per second.
3. Streaming near real-time processing.
4. Storing raw data for at least 5 years.
5. Tools for both streaming and historical data analytics.
6. Support for deployment across multiple data centers for maximum disaster tolerance.

Based on the presented indicators, we can formulate the main requirements for the database:
1. Use of ACID transaction for data resiliency when modifying the system.
2. Strict consistency to get the same responses from all nodes of the database.
3. Horizontal scaling on writes and reads for fast processing and near real-time processing.
4. Fault tolerance. Availability of multiple interconnected data centers.
5. Availability. Use of cloud technologies.
6. The amount of primary data storage over several years will be petabytes of data, so good storage compression is needed.
7. Performance. Providing quick access to raw data and analytical tools.
8. Application of well-known and proven SQL (Structured Query Language).

After a thorough analysis of these requirements it was concluded that none of the existing universal databases for the creation of this system is now suitable.

Database of the IoT (Internet of Things) contains a set of properties of several universal databases, so one can define the types of stored information and create a base for each of them [5].

The information stored in the urban forest management system can be divided into two types: metainformation and raw data from devices. Metainformation includes almost all data except for those that are transmitted by the end devices. Raw data can be readings from devices, service information from devices, data that have a structure of time series with the content of the value and timestamp in the message [6].

Based on the peculiarities of the data types the information can be selected under the characteristics of classical databases. Suitable should be any classical relational database with cluster support and asynchronous replication like PostgreSQL or MySQL. It because this data has a small volume and
is rarely modified. An important feature of working with this type of data is the need to build trees with certain requirements, which will combine the required number of user devices in a tree structure to facilitate their management and mapping (figure 2).

![Figure 2. Displaying devices as a tree structure](image)

Not every relational database is capable of managing large, arbitrary trees. Two databases can be used: a graph database to store the tree and a relational database to store the rest of the meta-information. But this approach has significant drawbacks:
- A need to add external software to regulate transactions between the two databases while ensuring consistency;
- Difficult to support and not very reliable design;
- Creating and maintaining two databases, one of which is needed only to support limited functionality.

The only true solution is to use a relational database and a graph database on the basis of Tarantool primitives set. This product contains all the necessary tools that will allow:
- Use spaces for storing data.
- Create full-featured ACID (Anonymity, Consistency, Isolation, Durability) transactions.
- Make asynchronous replication.
- Make relations on stored procedures.
- Make trees on the basis of stored procedures.

So one instance of Tarantool has a read request processing performance of up to 1000, with a tree traversal at the beginning of the process.

The other type of data is the raw data from devices represented in the form of a time series, which is characterized by a large volume of stored information and frequent recording. Therefore, the main criterion for the selection of the database for this type of data will be the horizontal scaling, availability and fault tolerance [7]. The availability of ready-made tools for analytical activities in the familiar to all SQL language will be a great advantage.

Since the data coming from the sensors have the form of time series it is advisable to use for their storage and processing time series databases. They have a special mechanism to work with data storage. Another prerequisite for the choice of the database is the condition of effective compression of the stored data, because their volume will only increase in the future. The most suitable for all requirements is a product by Yandex - ClickHouse. This is a columnar database management system (DBMS) for online analytical query processing (OLAP), which has all the necessary resources in its arsenal.

However, streaming analytics processes require another database containing the current state of Internet of Things devices and sensors, as well as many global variables. This information is
characterized by frequent recording of device state. Thus any database will be suitable, that is capable of horizontal scaling for reading and writing, supporting high availability and fault tolerance, any key value database.

The geographic information system of urban forest management in the city of St. Petersburg should fit well into the concept of a set of diverse databases (figure 3) [8].

Figure 3. Structural diagram of the urban forest management GIS data bank.

The structural diagram of the urban forest management GIS data bank in the city of St. Petersburg shows the interdependence of diverse databases, responsible for storing a particular type of information, with its characteristic features.

3. Conclusion

Preservation and effective management of large territories is unthinkable in the age of information progress without the use of new information technologies. Creation of geoinformation system of management of urban forests of St. Petersburg should accumulate the management process in a single information space, with the widespread introduction of sensors, tracking devices and neural networks. The implementation of such a project should begin with the definition of the initial information, methods and technologies of their use. The result of this study is to identify the types of data stored in the system, their structure and relationship through the use of different databases. Their choice is made on the basis of a thorough analysis of the original information.

References

[1] Petrov V N and Zubko A A 2005 [in Russian – Economic evaluation of forest recreation]. (St. Petersburg: Nauka) p 195
[2] Petrov V.N.2011 Urban forests: problems of organization and legal regulation [in Russian – LesPromInform] No 2 (76)
[3] Kolbina O, Yagotinceva N, Bogdanov P, Istomin E and Vagizov M 2020 The use of information and infocommunication technologies in the management of urban forests of St. Petersburg using geographic information systems IOP Conf. Series: Earth and Environmental Science (2020) All-Russian scientific-technical conference "Digital technologies in forest sector" 26 – 27 March 2020, Saint Petersburg, Russian Federation. vol 507 DOI:10.1088/1755-1315/507/1/012013
[4] Five technologies required to create IoT-applications [in Russian –Electronic resource]. Access
point: https://www.itweek.ru/iot/article/detail.php?ID=187126 (Accessed 15.03.21)

[5] Kalambet M, Yagotinceva N, Kolbina O, Yagotinceva T and Aksenova A 2020 Opportunities and prospects of IoT application in landscape architecture, design and information and communication technology // *IOP Conf. Series: Earth and Environmental Science* (2020) All-Russian scientific-technical conference "Digital technologies in forest sector" 26 – 27 March 2020, Saint Petersburg, Russian Federation. vol 507 DOI: 10.1088/1755-1315/507/1/012007

[6] Malikova D M, Slashchev E S, Istomin E P, Vagizov M R, Kolbina O E 2020 Method for solving problems of the theory restrictions of infocommunication systems using linear equations with many unknowns *IOP Conf. Series: Earth and Environmental Science* (2020) All-Russian scientific-technical conference "Digital technologies in forest sector" 26 – 27 March 2020, Saint Petersburg, Russian Federation vol 507 DOI:10.1088/1755-1315/507/1/012018

[7] Abramov V M, Istomin E P, Mikheev V L, Novikov V V and Palkin I I 2019 Model of geo-information support for decision-making while natural risk management *International Multidisciplinary Scientific GeoConference SGEM*. vol 19 No 2.1 pp 951-956

[8] Istomin E P, Kolbina O N, Stepanov S Y and Mirankov V A Database "Ice_Concentration". Database Registration Certificate RU 2014621110, 07.08.2014.