

Development of a geographic information system for data collection and analysis based on microservice architecture

Alexander A. Dontsov¹, Igor A. Sutorikhin^{1,2}

¹*Institute for Water and Environmental Problems SB RAS, Barnaul, Russia*

²*Federal Research Center for Information and Computational Technologies, Novosibirsk, Russia*

Abstract

The paper discusses the use of microservice architecture in the development of geographic information systems (GIS) for collecting, processing and analyzing data. As a rule, microservice architecture is used to build applications in information systems related to solving business problems, and is not widespread in the development of geographic information systems in the scientific field. However, its application is now becoming increasingly important. Decomposition of the software implementation and GIS infrastructure associated with computations and data processing into components in the form of microservices has a number of advantages, such as: increased fault tolerance, increased flexibility, reduced maintenance effort, simplified scaling, and others. The first results of the application of the microservice approach in the development of a geoinformation system for the collection and processing of hydrological and hydrobiological data on the state of water bodies are shown. The architecture, main components, and features of the information infrastructure are shown.

Keywords

GIS, microservices, satellite data, geoportal, cloud technologies, information systems, lake, reservoir, measuring complexes.

1. Introduction

Monitoring the parameters of various natural objects is an urgent task of nature management. Currently, there is a growing need to provide data on the state of natural objects to a wide range of organizations and individuals, from government agencies to public organizations. Currently, to solve the problems of collecting, processing and providing data, web GIS systems are widely used in the form of geoportals, which allow automating data processing processes and organizing access to the results of calculations. The use of microservice architecture in the development of such web GIS makes it possible to achieve reuse of GIS components and their independent operation [1]. This approach is that the information system is implemented as a set of small services, each of which is executed as a separate process and communicates with others using interaction mechanisms, as a rule, technologies are used for this: REST, gRPC, RabbitMQ, Apache Kafka [2].


Our previous works [3, 4] show the results of the development of a GIS for the collection and

SDM-2021: All-Russian conference, August 24–27, 2021, Novosibirsk, Russia

 alexdontsov@yandex.ru (A. A. Dontsov)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

processing of hydrological and hydrobiological data on the state of water bodies; at present, a new version of it is being developed, the main difference of which is the decomposition of the information system into microservices.

Microservice architecture has several advantages such as.

1. Services are simple and easy to maintain.
2. Services are deployed independently of each other.
3. Services scale independently of each other.
4. There is an opportunity to conduct technological experiments and it is relatively easy to introduce new technologies.
5. Relatively high fault tolerance compared to monolithic architecture.

However, the microservice architecture also has disadvantages such as.

1. Difficulty at the initial stage of development and creation of infrastructure. Distributed systems are more difficult to develop, since you need to provide for the independence of one microservice from failure in another component.
2. Development of distributed systems imposes additional costs on the exchange of data between microservices: you need to choose the right communication protocols between the components so that the interaction is as efficient as possible [5, 6].

Monitoring of water bodies, such as lakes and reservoirs, is a topical area of environmental management. Inland water bodies play a very important role in natural and anthropogenic processes.

When developing a GIS designed for collecting and processing data on the state of water bodies, it is necessary to take into account the fact that complete and comprehensive information can be obtained by integrating different measurement methods, such as:

1. Satellite monitoring.
2. Data of ground measuring complexes.
3. Data from expeditionary research and field observations.

These are three unrelated sources of information that require different approaches, algorithms and technological solutions to work with them. In this case, it is relevant to use a microservice architecture, which allows the implementation of system components in the form of independent software modules.

2. Geographic information system implementation

2.1. Description of GIS infrastructure

Consider the developed GIS, it can be divided into components related to data collection, computational processing and analysis modules, data storage and provision systems. Docker technology is used to deploy and manage application infrastructure [7]. This technology implies application containerization, that is, applications (GIS components) run in independent software containers. Their interaction at the infrastructure level is defined using Docker Compose. It is

a tool included with Docker and is intended for solving tasks related to project deployment. Docker Compose uses YAML files to store container group configuration. Containerization is a lightweight type of virtualization and resource isolation at the OS kernel level. It makes it possible to run applications with the minimum required libraries in a standardized environment. Due to the fact that each container is an isolated environment, it can be viewed as a small service under the control of a programmer. Any container can be customized and updated without affecting other containers, while providing complete isolation and standardization. Since containers are configured through special files, they can be versioned in GIT. This approach, in general, is called IaC – Infrastructure as a Code [7].

2.2. General description of GIS

Figure 1 shows the technological flows of data transfer from sources to the user: data is received through software interfaces. In the most technologically difficult verification of satellite data, since it is necessary to make sure that the archives have been completely downloaded to the GIS server, then they need to be unpacked and saved. In the case of the results of measurements of automated complexes and field observations, the files are checked for integrity and format. Satellite data atmospheric correction and thematic processing. A thematic term means the use of a set of algorithms according to the task at hand, for example, the selection of a water surface. The highlighting of the water surface is carried out using spectral water indices, which enhance the contrast between water and other objects.

The results obtained in the form of vector polygons in GeoJSON or Shapefile format are written to the database. After saving the results of calculations and importing data, they are available to users in the form of maps, files, graphs and tables. The JavaScript library Leaflet [8] is used to generate maps; map generation on the server side is implemented using the MapServer software utility [9].

As you can see from the diagram, work with each of the data sources can be organized as an independent process. When developing the GIS architecture, this technological scheme was used to divide the information system into microservices. Decomposing systems into microservices is a complex task, and there are several techniques for separating services from a monolithic system. In industrial information systems, there is still no single approach to decomposition of systems, and in each case it is necessary to choose a technique based on the peculiarities

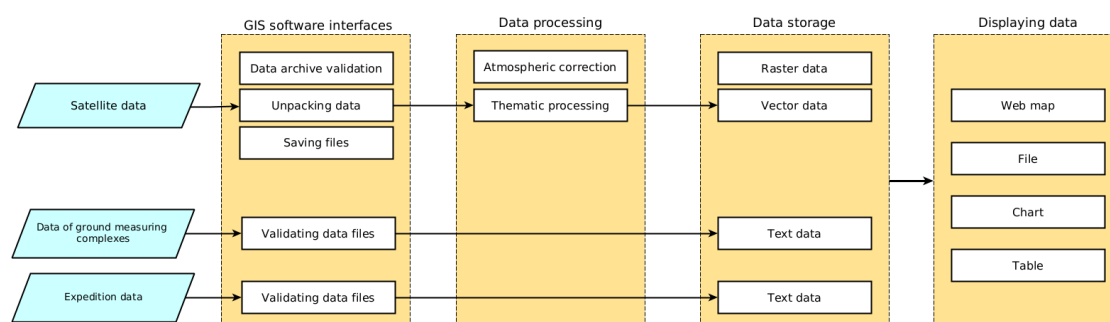


Figure 1: Process data transfer descendants in GIS.

of the subject area, the connectivity of subsystems and other parameters [10, 11]. In this case, an approach was used in which the sources of information were identified, the technological stages of data processing, based on this, a list of microservices was determined that should be present in the GIS.

GIS consists of the following main components (microservices).

1. Service for downloading satellite data from open archives ESA (European Space Agency) and USGS (United States Geological Survey).
2. Service of work with data of ground measuring complexes.
3. Service for working with data from expeditionary operations.
4. Service of atmospheric correction of satellite data.
5. Service of thematic processing of satellite data.
6. Service for converting files (from raster to vector formats).
7. Data cataloging service.

All services use a common database, where, along with the processed information, various system settings are stored, for example, the schedule of computing tasks, information about system users and their rights. Postgresql with the Postgis extension for storing geo objects is used as a DBMS [12].

In addition to services, GIS contains interfaces for interaction with other software systems (Figure 2). For example, desktop packages QGIS, GRASS, etc., as well as a web interface and an administration panel.

An important part of GIS is the management of computing processes, management of computing processes is carried out using the administrator panel (web interface), then the data

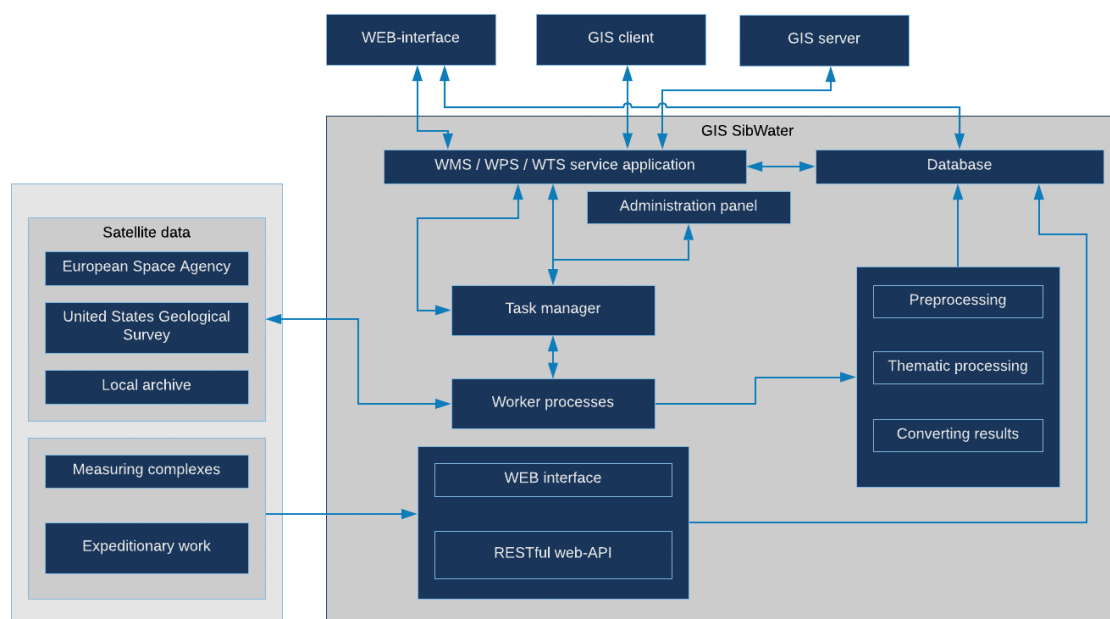


Figure 2: Block diagram of GIS.

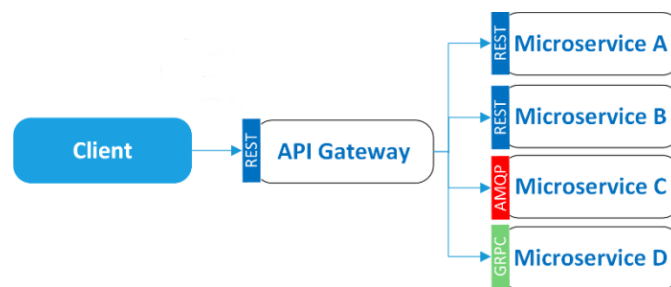


Figure 3: API Gateway pattern.

received from the user is processed and transmitted to the task manager to form tasks. Then they are written to the queue — for this, the RabbitMQ technology is used, the computational process is started. At the end of the work, the results are written to the database.

When developing the system architecture, the API Gateway pattern was used. This pattern is based on the use of a gateway that sits between the client application and microservices, providing a single entry point for the client. The use of this pattern reduces the number of calls, ensures client independence from the protocols used in services: REST, AMQP, gRPC, and provides centralized management of end-to-end functionality. Figure 3 shows the structure of the API Gateway pattern.

2.3. Integration of GIS with ground measuring complexes

The use of local automated systems for monitoring the parameters of natural objects is a promising area of research. Such measurements are carried out in order to monitor various natural processes, as a rule, with the further transfer of the measurement results to a wide range of interested parties.

Monitoring of inland water bodies is part of the monitoring of the natural environment as a whole. According to modern international approaches, monitoring of any component of the natural environment (including water bodies) should include a set of standardized observations and methods of processing, analyzing and transmitting the results of these observations to consumers [13].

The work uses the data of the measuring complex, which is designed to carry out systematic complex measurements of the parameters of water bodies. The measuring complex is located in the Altai Territory on Lake Krasilovskoye, on the shore of which the educational and scientific station of the Altai State University operates. The measuring complex allows, in an autonomous mode, with a period of 15 minutes, to receive information about 4 meteorological parameters of the atmosphere at heights of 2 and 4 meters, incident and reflected solar radiation, levels of lake and ground waters, water temperature from surface to bottom (depth 7 meters), and also the temperature of the soil from the surface to a depth of 3 meters.

The measuring complex consists of three autonomous units specially prepared for installation in the water area of the lake on a raft, on the bottom near the water's edge and permanently on the shore. APIK is equipped with a GSM modem for data transmission and a logging module (storage of measured data for subsequent download). To ensure the integration of GIS with

the measuring complex, a RESTfull-API was developed, which is based on the Django REST framework (DRF) extension. API data is transmitted in JSON format and, after validation using the Django form functionality, is written to the GIS database. The results of expeditionary work can also be added to the GIS via API or web interface with an add and import form.

In developing the RESTful GIS API and the data transfer format, the recommendations for building the REST API developed by the Open API Initiative were used. The OpenAPI specification comes with a set of guidelines for developing REST APIs. It provides a number of interoperability benefits, but requires additional design attention to comply with the specification. OpenAPI recommends that you start by creating a contract, not an implementation. This means that when developing an API, a contract (interface) must first be developed and only after that the program code for its implementation must be developed.

As noted earlier, a separate microservice was developed to work with measuring complexes; it has the following functionality: receiving data in JSON format or in the form of a CSV file, validating and writing to a database.

2.4. GIS work with satellite data

Determination of the parameters of water bodies based on satellite imagery data is of particular interest, since satellite data simultaneously cover a vast territory and reflect the current forms and areas of water bodies. Due to this, satellite imagery materials are becoming more and more popular [14].

Earth remote sensing data and geoinformation technologies allow solving many important tasks, including such as:

1. Inventory of reservoirs and other water bodies.
2. Regular monitoring of the condition of dams and other water protection and hydraulic structures. Assessment of the ecological state of water bodies, including the identification of areas of water bodies contaminated as a result of emergency discharges and spills of hazardous substances, identification of sources of pollution. Study of channel processes and mapping of the bottom microrelief in shallow water.
3. Forecasting and operational monitoring of floods, modeling the processes of inundation of the territory as a result of floods.
4. Determination of biological productivity of reservoirs, identification of aquatic biological resources, solution of fish farming problems.
5. Determination of the area of the water area of water bodies.

However, when working with satellite data, it is necessary to take into account a number of features, which are presented below:

1. Inland water bodies, as a rule, have a relatively small area, therefore, medium and high resolution data are suitable for effectively determining their spatial characteristics.
2. Dependence on weather conditions and time of day for satellite vehicles with measuring equipment operating in the optical range.
3. Inland water bodies are much less studied using Earth remote sensing systems than relatively large seas and oceans.

4. The efficiency of satellite data processing largely depends on the choice of optimal processing algorithms, technological solutions, well-developed techniques, information support.

Based on the above features, the work uses data from the Sentinel-2 and Landsat-8 spacecraft, which are available in the open archives of satellite information ESA (European Space Agency) and USGS (United States Geological Survey). To solve the problem of obtaining satellite data in an automated mode, a special software module was developed. It is based on a network connection to the interfaces of the satellite data archive servers. In the process of connection, a session of transmission of a request for data search is established, consisting of the coordinates of the required data area, the date of the survey and the type of satellite. The server, in response to the request, transmits a list of available data satisfying the request. After that, the data is downloaded in the form of an archive to the GIS server, then, according to the scheme in Figure 1, the files are processed.

3. Summary and conclusions

The presented geoinformation system allows to implement the processes of processing diverse information about the state of water bodies for solving fundamental and applied problems of hydrology and hydrobiology. However, it can be used in other tasks related to the collection and processing of spatial data. When developing a GIS, it is necessary to take into account the features of data sources, their processing stages and storage features. Microservice architecture allows you to organize a flexible system for collecting, processing and storing data on the state of natural objects. The source code of the developed GIS is open and available at <https://github.com/alexdontsov/sibwater>.

References

- [1] Wang Y., Han W., Nian Z. Design of satellite ground management system based on microservices // *Proceedings of the 2020 3rd International Conference on Computer Science and Software Engineering (CSSE 2020)*. N.Y.: Association for Computing Machinery, 2020. P. 119–123. DOI:10.1145/3403746.3403915.
- [2] Microservice architecture. Available at: <https://microservices.io/patterns/microservices.html> (accessed June 9, 2021).
- [3] Dontsov A.A., Sutorikhin I.A. Specialized geoinformation system for automated monitoring of rivers and reservoirs // *Computational Technologies*. 2017. Vol. 22. No. 5. P. 39–46.
- [4] Dontsov A.A., Sutorihin I.A., Frolenkov I.M. Geographi information system for bloom monitoring inland water bodies // *Limnology and Freshwater Biology*. 2020. No. 4(SI:7VBC). P. 914–915.
- [5] Villamizar M. Infrastructure cost comparison of running web applications in the cloud using AWS lambda and monolithic and microservice architectures // *16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)*. IEEE, 2016. P. 179–182.

- [6] Namiot D., Sneps-Snepp M. On micro-services architecture // *International Journal of Open Information Technologies*. 2014. Vol. 2. No. 9. P. 24–27.
- [7] Docker. Available at: <https://www.docker.com> (accessed June 9, 2021).
- [8] Leaflet. Available at: <https://leafletjs.com> (accessed June 9, 2021).
- [9] MapServer. Available at: <https://mapserver.org> (accessed June 9, 2021).
- [10] Balalaie A., Heydarnoori A., Jamshidi P. Microservices migration patterns // *Technical Report TR-SUTCEASE-2015-01*. Automated Software Engineering Group, Sharif University of Technology, Tehran, Iran, 2015.
- [11] Amaral M., Polo J., Carrera D., Steinder M. Performance evaluation of microservices architectures using containers // *14th IEEE International Symposium on Network Computing and Applications (NCA)*. IEEE, 2015. P. 27–34.
- [12] PostGIS. Available at: <https://postgis.net> (accessed June 9, 2021).
- [13] Dumitru A. et al. Approaches to monitoring and evaluation strategy development // *Evaluating the Impact of Nature-Based Solutions. A Handbook for Practitioners*. 2021.
- [14] Frazier P.S. et al. Water body detection and delineation with Landsat TM data // *Photogrammetric Engineering and Remote Sensing*. 2000. Vol. 66. No. 12. P. 1461–1468.