

Review Article

Development of Algorithm of Smart Geographic Area

Yusifov S.I.¹, Ragimova N.A.¹, Abdullayev V.H.^{1*}, Khalilov M.E.¹¹Department of Computer Engineering, Azerbaijan State Oil and Industry University*Corresponding author: abdulvugar@mail.ru

Received: 7 January 2021; Accepted: 1 March 2021; Published: 17 March 2021

Abstract: The rapid development of information technologies accelerates the approximations of Industry 4.0, which is why sectors of the economy and science must adapt to these changes. Global changes in geography have led to the emergence of a new scientific discipline called geoinformatics. It then provides insights into the Smart Geographic Area, its structure and the main components. To do this, there used methods for communicating the main components (IoT, IoE), for analyzing data (Big Data, Hadoop), for managing processes (CPs), for storing data (Cloud Computing, Fog Computing. As a result of the study, there was developed a Smart Geographic Area algorithm based on the MapReduce paradigm.

Keywords: Industry 4.0, Geography, Geoinformatics, Smart Geographic Area, Geographic Information System, Hadoop.

Introduction

There happen global processes of digitalization and informatization of all spheres of economics and science on the eve of the Industry 4.0. They based on modern developments in Information Technology: Internet of Things, Big Data, Cyber Physical Systems, Cloud Computing, Foggy Computing etc. [1]

One area that is undergoing global change is geography. Under the influence of these changes, there is an intellectualization of geographical areas. The mines, the farms, the kitchen gardens, the plants, the factories, the settlements all these objects are on a identified geographical area. In order to maintain the normal processing of these facilities, it is necessary to take into account and monitor the geographical indicators of the relevant geographical area. [2, 3]

The main problem of this research is to compile a smart geographic area algorithm based on modern methods of collecting, processing, communicating and storing geographical data. The main purpose of this paper is to define the structure of a Smart Geographical Area. For this, it need to determine the main components, methods of collection and processing of data, how components interact themselves.

Methods of Research

There used the following concepts to carry out this study: Internet of Things, Big Data, Cyber-Physical Systems, Cloud Computing.

The Internet of Things (IoT) combines devices into a network and allow them to collect, analyze, process and transfer data to other objects through software, applications or technical devices. In fact, it is a network of networks in which people can communicate with devices, and devices can communicate with each other, respond to changes in the environment and decisions making without participation of human. IoT devices function independently, although people can configure them or guarantee access to data. IoT systems work in real time and usually consist of a network of smart devices and a cloud platform. [4]

Big Data is structured and unstructured data that cannot processed by traditional data processing methods. The main characteristics of big data are volume, velocity, and variety. They also include veracity, value, viability, variability, and visualization. From the above characteristics, you can formulate the following basic principles for performing big data manipulations:

- Horizontal scalability is the main principle of big data analysis. This means that with more data, you need to increase the number of computing nodes without losing performance.
- Fault tolerance. It is possible that the number of computing nodes will grow, which means that the likelihood of their failure is growing. Therefore, big data tools must be prepared for such situations and able to respond appropriately.
- Data locality. It is necessary to store and process data in the same physical server, otherwise the cost of transferring data between servers can be enormous. [5]

The essence of **Cyber-Physical Systems** is that they connect physical production processes or digital processes (for example, transmission and distribution controls) that require the practical implementation of continuous control in real time. [6]

Cloud computing is a model that assumes that all servers, networks, applications, and other elements are available to the IT service and end users over the Internet. Cloud is not a place, but a way to manage IT resources, replacing local machines and private data centers with virtual infrastructure. [7]

Research Results and Discussion

Smart Geographic Area

Changes in geography mainly reflected in Geoinformatics. Geoinformatics is a technological science that combines science of Earth and computer science. [8] The main object of research in this scientific field is a digital representation of the geographical area. The development of Geoinformatics contributes to the creation of an intellectually developed geographical area. The assumed structure of the Smart Geographic Area (SGA) is shown in Figure 1.

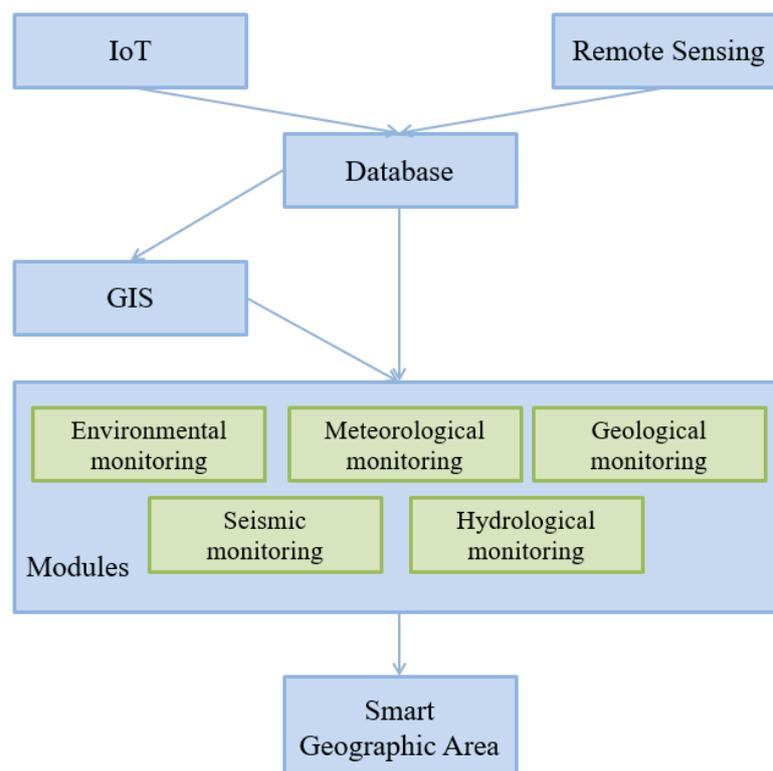


Figure. 1. Structure of the SGA.

To maintain correct operation of any system, it is necessary to have appropriate data. The necessary data for SGA can be obtained using field devices, which can be presented as remote sensing of the Earth and Internet of Everything.

Remote Sensing of the Earth is the receiving of information about objects without entering into physical contact with them. The principal advantage of remote sensing of the Earth is the rapid receiving of the necessary geographical data. The most common method is the application of spacecraft. Remote

sensing of the Earth is a method that allows you to obtain geographical data on a particular geographical area. [9]

The idea of **the Internet of Everything** based on the Internet of Things. Things occupy the main in the Internet of Things, while the main elements in the Internet of Everything, except things, are people, data and processes. The concept of IoE is aimed at connecting all devices to a single network. If RSE is a method, then IoE is the devices using which there are obtained the necessary geographical data. [10]

Once data is received from smart devices, all data is placed in **databases**. These databases are located in special data centers. To simplify and accelerate the processing of these centers, there should create an identified number of small centers for processing of primary data. After primary processing, the data is sent to the main data center, where the main processing will be carried out on them. There can be used cloud computing and fog computing concepts to implement this scheme.

The two concepts complement each other. There talked about cloud computing above. Note that the cloud is centralized memory. **Fog computing** offers that some work of data redirect to local computers. In other words, the main data center is the "cloud," and the collection of small data centers is the "fog." [11]

The main component of SGA is a **Geographic Information System (GIS)**. GIS is a system that enables the acquisition, storage, processing, analysis of spatial data and relevant metadata to obtain information and knowledge about the geographical area necessary for the decision-making process. [12] To create a GIS, there come spatial data and its metadata from the cloud.

Most of SGA is occupied by monitoring modules, which for each module collect and process the corresponding data. Each module generates a geographic map with corresponding parameters. These modules perform the following monitoring: ecological monitoring, seismological monitoring, meteorological monitoring, hydrological monitoring, geological monitoring and other types of monitoring.

Ecological monitoring is a comprehensive system for observing, evaluating and predicting changes in natural environments, natural resources, plant and animal life, which allows you to distinguish changes in their state and processes taking place in them under the influence of anthropogenic activity. From the very beginning, two points of view appeared in the interpretation of monitoring. Many foreign researchers have proposed a system of continuous observations of one or more components of the environment for a given purpose and according to a specially developed program. Another view suggested that monitoring should be understood only as a system of observations that made it possible to distinguish private changes in the state of the biosphere that occurred only under the influence of anthropogenic activities (i.e., monitoring anthropogenic changes in the natural environment). The monitoring process involves the sequential implementation of two tasks. The first task is to ensure a constant assessment of the "comfort" of human habitat and biological objects (plants, animals, microorganisms), as well as an assessment of the state and functional integrity of ecosystems. The second task is to create conditions for determining corrective actions in cases where targets for ecological quality assessment criteria are not achieved. [13]

Seismic monitoring refers to technologies for reducing the risk of natural hazards. It is based on the organization of a network of continuous long-term observations in the studied territory. In the modern interpretation, monitoring includes not only registration, but also further operational processing and interpretation of seismological data with access to forecast estimates. [14]

Meteorological monitoring is a set of measures for the collection, processing, storage of meteorological data (temperature and humidity of air, atmospheric pressure, wind direction and speed, type and amount of precipitation, etc.). The purpose of this monitoring is to determine pre-emptive meteorological conditions that can lead to economic, financial and human losses. [15]

Hydrological monitoring refers to a system of continuous (current) and integrated monitoring of the state of water resources, control and accounting of quantitative and qualitative characteristics over time, mutually agreed effects and changes in consumer properties, as well as a system for forecasting conservation and development in different modes of use. Hydrological monitoring of water bodies includes surface waters of land, seas, water management systems and structures (including reservoirs).

The object of ecological monitoring is to assess its quality and level of pollution as a prerequisite for making scientifically sound decisions on the effectiveness of ecological measures. [16]

Geological monitoring is carried out to study the soil. The results of monitoring can be a decisive factor for the development of land for agricultural purposes, for the construction of various construction facilities, for the determination of profitable deposits of minerals and deposits. This all contributes to the growth of the economy of the corresponding geographical area. [17]

After the collection and processing stages of geographical data, it is possible to explore the region from a geographical point of view. Under the concept of "geographical region" can be a whole country or region (Caucasus, Siberia, Apache, Alps, Sahara, etc.). In this case, it makes sense to divide the studied geographical area into smaller geographical units (for example, district). Then it is necessary to group the obtained maps into districts for SGA.

SGA and Hadoop

The data collection process will accumulate a huge amount of geodata. To ensure efficient and continuous operation, it is necessary to apply distributed tools for large geodata processing. One of the most common distributed tools is Hadoop, based on the MapReduce model.

MapReduce is a model of distributed computing developed by Google, used to perform reliably parallel computing of big data on large clusters up to several terabytes in size. The main advantage of MapReduce technology is the ease of scalability of data processing on several clusters. Each of these clusters can consist of several computers, if the number of computers changes, you should simply change the configuration. The MapReduce paradigm consists of "sending a computer to where the data is located," that is, processing big data do on the same cluster where it is stored. The MapReduce paradigm, like any other technology in addition to the advantages, has disadvantage. Usually, disadvantages mean in view of the situation when this model is undesirable to use. Below is a list of these cases:

- Real-time processing.
- It is not always very easy to implement everything and everything in the form of the MapReduce application.
- When processing requires many data to shuffle across the network.
- When to process streaming data. MapReduce is best suited for batch processing of huge amounts of data.
- When you can get the desired result using a standalone system. Obviously, setting up and managing an autonomous system is less painful than a distributed system.
- When there are OLTP needs. MapReduce is not suitable for a large number of short online transactions.[18]

The most common tool for big data processing is Apache Hadoop, implemented based on MapReduce. Apache Hadoop is a set of libraries and utilities that allows you to develop and execute distributed programs running on clusters that are capable of consisting of a large number of nodes (up to several thousand).

Given the large amount of data, it makes sense to use Hadoop to implement SGA (Fig. 2). The figure below shows the relationship of MapReduce steps to SGA components. As can be seen, the inputs of the Map step are data collected from field devices in real time. This data contains information on the physical processes required for the above-described monitoring. In this case, the Map phase has two tasks: creating a geographical database and compiling a GIS. The use of MapReduce allows designing databases that will contain current data and this data will update in the present time. The design of such a DB will make it possible to draw up maps of any area the GIS is based on. Accordingly, the outputs of this phase are the geographic DB and the GIS that is based on these data.

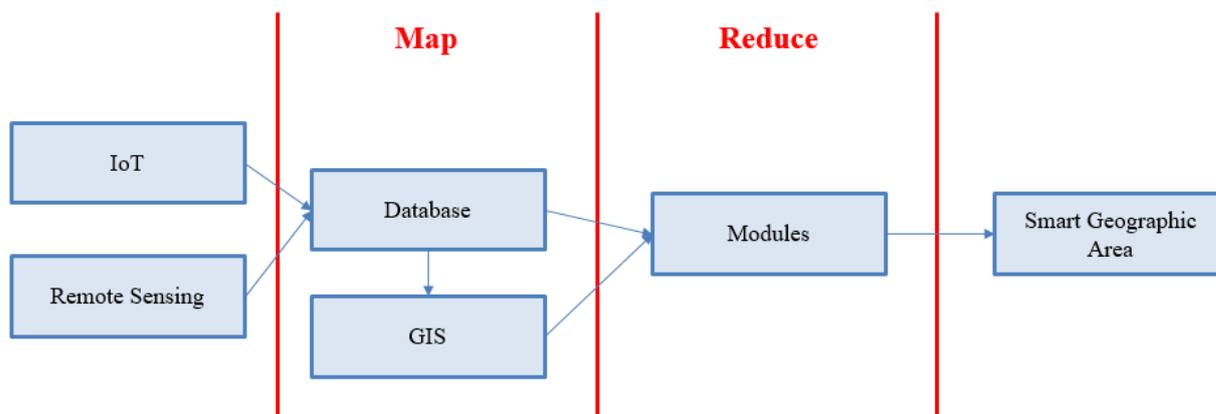


Figure. 2. MapReduce for Smart Geographical Area

MapReduce are received a pair values of GIS and corresponding data for each module at the input of the Reduce stage. During this stage, the main processing of geodata carry out based on which geographical phenomena monitor. This means that ecological data will use for environmental monitoring; seismological data will be used for seismological monitoring etc. After the main processing of geographical data, each monitoring module receives its own map (a meteorological map will obtain for meteorological monitoring). As is known at the output of the Reduce step, a key and a corresponding list of values obtain. In this case, the key is the name of a region, and the list of values is the collection of previously obtained maps for this region.

Conclusions

1. The study resulted in the structure of the SGA, including its components, implementation technologies and the relationship of these components.
2. Information technologies were also considered, which could lead to the further development of geography and make it more relevant on the eve of the fourth industrial revolution.
3. There considered using Hadoop to implement an SGA. The reason for using Hadoop is scalability and distributed processing of geodata.

References

- [1] Dorleon, G., Gervais, S., Industry 4.0-What we need to know. https://www.researchgate.net/publication/329388831_Industry_40-What_we_need_to_know, (2017).
- [2] Schabus S., Scholz J., Geographic Information Science and Technology as Key Approach to unveil the Potential of Industry 4.0 –How Location and Time Can Support Smart Manufacturing. ICINCO-2015 463-470, (2015).
- [3] Kiss E., Nedelka E., Geographical approach of Industry 4.0 based on information and communication technologies at Hungarian enterprises in connection with industrial space., Hungarian Geographical Bulletin 69(2) (2020) 99-117.
- [4] Subramanian, B., Nathani, K., Rathnasamy, S., IoT Technology, Applications and Challenges: A Contemporary Survey, Wireless Personal Communications 108 (2019) 363-388. Doi: 10.1007/s11277-019-06407-w.
- [5] Moorthy, J., Lahiri, R., Biswas, N., Sanyal, D., Ranjan, J., Nanath, K., Ghosh, P., Big Data: Prospects and Challenges, Vikalpa 40(1) (2015) 74-96. Doi: 10.1177/0256090915575450.
- [6] Jamaludin, J., Rohani, J. M., Cyber-Physical System (CPS): State of the Art. 2018 International conference on Computing, Electronic and Electrical Engineering (ICE Cube).

-
- https://www.researchgate.net/publication/330389785_Cyber-Physical_System_CPS_State_of_the_Art, (2018).
- [7] Birje, M., Challagidat, P., Goudar, R. H., Tapale, M. T., Cloud computing review: Concepts, technology, challenges and security, *International Journal of Cloud Computing* 6(1) (2017). Doi: 10.1504/IJCC.2017.083905.
- [8] Sinha, A. K., Malik, Z., Rezgui, A., Zimmerman, H., *Geoinformatics: Transforming Data to Knowledge for Geosciences*, *GSA Today* 20(12) (2010) 4-10. Doi: 10.1130/GSATG85A.1.
- [9] What is Remote Sensing? <https://earthdata.nasa.gov/learn/remote-sensing>.
- [10] Miraz, M. H., Ali, M., Excell, P., Picking, R. (2015). A review on Internet of Things (IoT), Internet of Everything (IoE) and Internet of Nano Things (IoNT), *Internet Technologies and Applications (ITA)* (2015).
- [11] Al-Doghman, F., Chaczko, Z., Rakhi, A., Klempous, R., A review on Fog Computing technology, 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC). DOI: 10.1109/SMC.2016.7844455.
- [12] Shrestha, H. L., GIS and Remote Sensing Applications for Resource Mapping. *The Initiation* 1 (2010) 56-61. DOI: 10.3126/init.v1i0.2681.
- [13] Jing Z., Jia Z., Xiangyang Du, Kang Hou, Minjuan Qiao, an overview of ecological monitoring based on geographic information system (GIS) and remote sensing (RS) technology in China, *IOP Conference Series Earth and Environmental Science* 94(1) (2017). DOI: 10.1088/1755-1315/94/1/012056.
- [14] Dzeboev, B., Krasnoperov, R., On the monitoring of seismic activity using the algorithms of discrete mathematical analysis. *Russian Journal of Earth Sciences* 18(3) (2018) 1-9. DOI: 10.2205/2018ES000623.
- [15] Ukhurebor, K. E., Abiodun, I. C., Azi, S. O., Otete, I., Obogai, L. E., A Cost-Effective Weather Monitoring Device, *Archives of Current Research International* 7(4) (2017) 1-9. DOI: 10.9734/ACRI/2017/32885.
- [16] Hund, S., Johnson, M., Keddie, T., Developing a Hydrologic Monitoring Network in Data-Scarce Regions Using Open-Source Arduino Dataloggers, *Agricultural & Environmental Letters* 1(1) (2016). DOI: 10.2134/ael2016.02.0011.
- [17] Chen Guo-ping, Zhao Jun-san, Yuan Lei, Ke Zun-jie, Gu Miao, Wang Tao., Implementation of a Geological Disaster Monitoring and Early Warning System Based on Multi-source Spatial Data: A Case Study of Deqin County, Yunnan Province. <https://nhess.copernicus.org/preprints/nhess-2017-191/nhess-2017-191.pdf>
- [18] What are the disadvantages of MapReduce?, <https://stackoverflow.com/questions/18585839/what-are-the-disadvantages-of-mapreduce>