Original Article
ISSN (Online): 3048-9245

ShodhAI: Journal of Artificial Intelligence
July-December 2024 1(1), 131–149

GEOSPATIAL BIG DATA PROCESSING USING THE HIGH-PERFORMANCE COMPUTING TECHNOLOGY

Thomas. U. Omali [1], Ibrahim Garba [2]

[1] National Biotechnology Development Agency (NABDA), Nigeria
[2] Department of Mathematics, College of Education, Kazaure Jigawa, Nigeria

Received 01 August 2024
Accepted 03 September 2024
Published 09 October 2024

Corresponding Author
Thomas. U. Omali, t.omali@yahoo.com

DOI 10.29121/ShodhAI.v1.i1.2024.13

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

A significant portion of the data used for different purposes includes location information. As a result, there is an increasing effort to find a general framework by utilizing crucial technological drivers. Similarly, as geospatial data becomes more accessible; the possibilities for changing scientific discoveries and methods in modern society are endless. Naturally, many applications depend on the accurate and efficient geospatial data processing. However, using the traditional method to process and compute geospatial data frequently presents several difficulties. The explanation for this is the large volume, heterogeneity, and distributed nature of these data. Consequently, it is imperative to implement contemporary techniques that can manage the computational and analytical demands of this massive amount of geospatial data. Many of these systems—particularly those related to High-Performance Computing (HPC) technology—have been described in the literature. Therefore, the goal of this article is to review the application of HPC technology to big data processing in geospatial applications. The review's findings show that predictive data science tools like parallel and grid computing are available in geographic big data applications, enabling quick and effective processing that will help create a sustainable ecosystem. Also, the efficient management of large data sets requires storage, visualization, analytics, and analysis. Of course, this review demonstrated how recent advancements in computing have impacted geospatial data handling.

Keywords: Analytics, Big Data, Computation, Geospatial Database, Spatial Data

1. INTRODUCTION

Many industries, including banking, healthcare, telecommunications, and homeland security, create massive amounts of data, or "big data." The quick generation of huge volumes of geospatial data (geospatial big data) from both in-situ and remote sensors is another important effect of the current revolution in sensor and computing technologies Bill et al. (2022). According to Lee and Kang (2015), there is a prediction that the amount of personal location data will increase by 20% annually, with location-aware information accounting for a significant portion of the 2.5 quintillion bytes of data generated daily. There has been application of geospatial big data in various areas of endeavors like public health, business analysis, natural hazard prediction and mitigation, sustainable development, and climate change. Every day, a growing amount of data is produced by computers, smartphones, sensors, and even people. Numerous applications of geospatial big data have been documented in the literature, including urban planning (see Huang et al. (2021)), mobility analysis, and social network analysis (e.g., Dong et al. (2021); Huang et al. (2021)). Others are, managing of event, modelling of building occupancy (see Versichele et al. (2014)), studying of travel orientation, and modelling of urban functions (see Hu et al. (2018), Wei and Yao (2021). As a result, it is essential to effectively and efficiently extract information from these massive geospatial data sets and produce new knowledge. Computational techniques have demonstrated their effectiveness in achieving the intended result in this case. There are many research opportunities with computational methods and digital data Watts (2014), and it is likely that these tools will lead to new discoveries O'Sullivan and Manson (2015).

Geospatial information computing is the computational task necessary to make geospatial data meaningful to users. A few of the crucial elements that are included in this are storage of data, management of data, processing of data, analysis of data, and mining of data (Liu et al. (2016); Baralis et al. (2017); Hu et al. (2018). Unfortunately, trying to solve certain problems makes geospatial information computing a challenging task. For example, the amount of global geospatial data, measured in petabytes (pb), exceeds the computational capacity of desktop-era analytical tools and traditional computing technologies. The speed at which thousands of geotagged tweets are collected every minute and terabytes (tb) of satellite data are acquired each day affects the capacity of traditional computing and data storage techniques. Furthermore, gathering geospatial data is typically accomplished through a variety of methods (such as social media, remote sensing, mapping, surveying, location-based data, and Internet of Things [see Yao and Li (2018)]. Various data models, such as raster and vector [refer to Li et al. (2017)] can also be used to abstract geospatial data. According to Chen et al. (2015), they also have varying spatial and temporal resolutions and are encoded using a variety of data formats, including geodatabases. Because of these varied qualities, tools for data processing and spatial analysis tasks require interoperability and standards. Global geospatial data are also frequently gathered by dispersed sensors and kept on servers. Regrettably, it is difficult to move data from one place (like a local server) to another (like the cloud) for processing due to high volume, high velocity, and the need to make real-time judgment Yang et al. (2013). In response to the necessity of addressing the above issues however, many processing and computing tools have evolved. High-performance computing technology, for instance, has proven effective in solving problems related to large-volume processing and geospatial information processing.

For numerous applications, many studies indicate that HPC has been utilized for solving geospatial issues (e.g., Hegeman et al. (2014); Pektürk and Ünal (2018); Yang et al. (2019). Initially, the complex nature of computations inherent in geospatial analyses was a driving force behind efforts to improve performance. Non-trivial examples of these issues remain unsolvable today, requiring a significant amount of memory and processing time. The solutions produced by other spatial analysis techniques also need a huge processing.

Although studies on geospatial big data has been carried out by many researchers within the academic and industrial sectors, more studies that will capture the most recent state-of-the-art methodology geospatial big data processing using high-performance computing technology is required as demonstrated by this review. The study is organized in seven sections. In section 2, we presented an overview of geospatial big data. In section 3, we described the principles of High-performance computing. Section 4 dealt with geospatial database management systems based on HPC. In section 5, the classification of computational systems was discussed. Section 6 has to do with the paradigm shift in geospatial big data computing. Section 7 concluded the research work.

2. GEOSPATIAL BIG DATA

Big data has been defined differently from industrial, academic, and technological standpoint (Chen et al. (2014); De Mauro et al. (2015)). However it is generally understood to be datasets larger than what can be handled by the typical modern data management tools (Batty (2013)). Big Spatial Data (BSD) fits the above characteristics and gives rise to specialized systems, techniques, and algorithms. Even before the big data era officially began, there was a wave of people using BSD. Geospatial big data facilitates real-time assistance, cost savings through increased efficiency, and the analysis of spatial relationships. Global elevation, remote sensing, and sensor data from the Internet of Things (IoT) are just a few sources of spatial big data. Other examples are land use, social media, public transportation, navigation, ontological, heterogeneous data via online services, and climate (Gaigalas (2019); Wu. (2019)).

Big data that includes location information is referred to as geospatial big data. Location information is crucial in the big data era (Huang et al. (2018)), since most data are spatial by nature. Lee and Kang (2015) believe that through the geospatial big data application, there are numerous opportunities for scientific advancement in many domains, such as precision agriculture, public health, climate science, disaster management, and smart cities. However, the ability to quickly and effectively extract useful information from big data is more important than the data itself. But, there are challenges in extracting important information and configurations due to the intrinsic space and time features of geospatial data Gudivada et al. (2015)]

2.1. MAIN SOURCES OF GEOSPATIAL BIG DATA

2.1.1. EARTH OBSERVATION

The statistical data from the Committee on Earth Observation Satellites (CEOS) suggest that over 500 Earth observation (EO) satellites have been launched in the last 50 years, and over 150 satellites are anticipated (Guo (2017)). The swift advancement of EO technology and the ongoing deployment of remote sensing satellites have contributed to the rise in EO data resolution, quantity, and variety. This suggests that EO data has transitioned into big data (Xia et al. (2018)). Of course, it is feasible to produce enormous amounts of diverse, dynamic, and widely distributed geospatial data using EO systems. Remote sensing has been one of the main ways to gather Earth observation data globally. For instance, the Landsat archive held more than 5.5 million images (Wulder et al. (2016), which are larger than one petabyte [Cervone et al. (2016)). Also, more than nine petabytes of data were being managed by EOSDIS as of 2014, and roughly 6.4 tb of data is being added daily by the system to its records. Furthermore, big Earth observation data collection now has an additional avenue through the application of drone-based remote sensing (Athanasis et al. (2018)).

2.1.2. GEOSCIENCE MODELLING

Through the quick advancement of computing power, the Earth occurrences can now be replicated with ever-higher spatiotemporal characteristics, producing vast amounts of simulated geospatial data (Blais and Esche (2014)). Common examples are the climate modelling by the Intergovernmental Panel on Climate Change (IPCC). The IPCC Fifth Assessment Report (AR5) alone produced simulated climate data amounting to 10pb, and hundreds of petabytes are expected to be created for the upcoming IPCC report (Yang et al. (2017)). Furthermore, we know that to sweep different parameters requires that a model frequently needs to be run multiple times. Thus, the process of standardizing the geoscience models generates huge volumes of geospatial data in addition to simulations. For instance, calibrating Model E (a NASA climate model) produced 3 terabytes of climate data from 300 models that were run in a single test (Li et al. (2015)).

2.1.3. INTERNET OF THINGS

The IoT is rapidly developing, and becoming a vital tool in almost every industry (Kumar et al. (2019)). Its origins can be traced to Kevin Ashton, who first used it in 1999 when he discussed the utilization of radio frequency identification (RFID) in supply chain management. The Internet of Things includes everything that has access to a network, including sensors that can provide recommendations on where to put pesticides or fertilizer locally [i.e., agro application (Maschi et al. (2018); Andreazi et al. (2021)]. It creates a massive network of interrelated things by connecting "things" to the Internet and allowing them to interact and communicate.

Various formats (e.g., discrete & streaming data, images, and social media) can deliver data on this network (De Azevedo et al. (2022)), as such, sensors can be used, with or without humans. Connecting the network to the Internet allows the virtual and physical worlds to communicate, and decisions can be made without human intervention. With the IoT, unstructured or semi-structured geospatial data streams are constantly produced globally. But, these data are more heterogeneous & noisy than structured multi-dimensional geospatial data generated by Earth Observations and model simulations.

2.1.4. VOLUNTEER GEOGRAPHICAL INFORMATION SYSTEM

Volunteered geographic information (VGI) refers to the production and sharing of geographic data from the general public. According to Haklay et al. (2014), it is crowdsourced geographic information delivered by many contributors. Also, Zook and Breen (2017). define VGI as the spatial subcategory of user-generated content that emerged during the Web 2.0 era; which was associated with the growth of GPS and smartphone technologies, blogs, social media, and wikis.

With the above-mentioned technologies, billions of citizen sensors around the globe are producing and sharing vast amounts of location-based data. For example, social media sites like Facebook, Instagram, Twitter, and Facebook use location sharing, or geotagging, to create virtual spaces where millions of people can connect digitally. Of course, 500 million tweets are sent daily (Internet Live Stats, 2019), and 5 million tweets are geotagged every day, based on the estimated 1per cent rate (Marciniec (2017)). In general, social media provides an abundance of resources for researching people's experiences in the outdoors and comprehending online conservation discussions or debates (Di Minin et al. (2015)).

3. PRINCIPLE OF HIGH-PERFORMANCE COMPUTING

HPC is a sophisticated system for processing massive amounts of data and resolving computing- and data-intensive issues (Niculescu (2020)), which was invented in the 1960s. HPC technology combines parallel programming and system administration (such as network and security expertise). Even though supercomputing is now considered a subset of HPC, the latter emerged after the former. But supercomputing has recently given way to the grid in HPC.

The Graphic Processing Units (GPUs) and Central Processing Units (CPUs) are the primary components that power the HPC. CPUs carry out serial processing, in which a single task is normally handled by a single CPU at a time. However, Ji et al. (2017) stated that parallel processing is performed using the GPUs. In HPC, "parallel architecture" describes the simultaneous execution of multiple processes. In this case, computation is separated into many parallelizable subtasks or decomposition. Once the processing is finished, the final output is typically combined. For processing spatial data on a large scale, some popular HPC platforms are listed in Table 1. The platforms in table 1 can be broadly categorized based on how much parallelism the hardware can support. For example, parallelism on CPUs is aided by MPI, UPC, and OpenMP. Numerous HPC applications can be realized more easily using the HPC platforms. These include fog computing Steffenel (2018), cloud computing (Mauch et al. (2013)), and the developing edge computing (Shi et al. (2016); Cao et al. (2020)). In general, modern computational science and scientific research are linked to HPC. As a result, it has been primarily used in numerous areas of operations. Furthermore, geospatial information processing can benefit greatly from its computational capability.

Table 1

Table 1 HPC Platforms for Large-Scale Processing of Spatial Data

HPC platforms	Description
Message Passing Interface (MPI)	They work with highly parallel computing architectures.
Open Multi-Processing (OpenMP)	An API for C/C++ and Fortran that facilitates multi-platform shared memory parallel programming.
Unified Parallel Computing (UPC)	By extending the C programming language, UPC allows programmers to work with a single shared, partitioned address space. Though the variables contained in this address space contains are only physically owned by one processor, they can be read and written by any processor.
General-purpose computing on Graphics Processing Units (GPGPU)	This employs GPUs to carry out computations that are handled by CPUs. It is possible for a GPU to outperform many CPUs in calculation speed if the computational operation is divided into manageable subtasks because it has multiple cores for basic tasks
Apache Hadoop, and Apache Spark	The open-source software package Apache Hadoop is built on the MapReduce programming prototype, with the capacity to automatically manage failures in hardware that are taken for granted. The MapReduce model has limitations that need a dataflow structure in linear format to read and write data to and from the disk. In response, Apache Spark was created. Distributed shared memory is used by Apache Spark in place of a hard drive disk.

4. GEOSPATIAL DATABASE MANAGEMENT AND DATA PROCESSING ON HPC

The emergence of geospatial big data brings new applications and issues (Yue and Jiang (2014)). Effective storage, management, and querying of geospatial data has become a research focus, and these are issues that need to be addressed (Schmid et al. (2015); Liu et al. (2016); Baralis et al. (2017); Hu et al. (2018). Before any spatial analysis can begin, a geospatial database must be designed and developed. The first thing to do in this case is to identify and define the database's content (database design). Next, the growing collection of publicly available spatial datasets from multiple sources may be used to create the geospatial database. The primary sources are social media (Tsou (2015); Cervone et al. (2016)); remote sensing (Mulyono and Fanany (2015); Chi et al. (2016); surveying and mapping (Lu et al. (2017)); location-based (Liu et al. (2015); Zhuang et al. (2017)); and Internet of Things (Ding et al. (2014); Alelaiwi (2017)).

Big data attributes have generally grown from the original "3Vs (Volume, Velocity, Variety)" to the more recent "4Vs (+ Veracity)" and "5Vs (+ Value)" (Li and Li (2014)). Therefore, processing contemporary geospatial data requires sophisticated computational tools. For example, scalable algorithms are necessary for real-time data processing, and large, inexpensive, and dependable storage is needed for massive volumes of data. Geospatial big data processing and analyses often involves many floating-point calculations, such as changing coordinate reference systems, transforming geometry, and assessing spatial relationships. To speed up these calculations, frameworks and systems built on MapReduce and Spark, such as SpatialHadoop and GeoSpark (Yu et al. (2015)) emerged.

The development and improvement of HPC tools, such as cloud computing, are having a major impact on the possibility of utilizing high-volume or high-velocity geographic data acquisition in more applications. In particular, the first organized systematic platforms for handling remotely sensed big data have improved the remote sensing method (Wang et al. (2018)). Additionally, big data analytics software can be easily implemented on distributed, parallel computing platforms thanks to big data platforms like Hadoop (Lu et al. (2017)). The ability to handle geospatial big data with HPC is required for making timely and improved decisions in time-sensitive circumstances, such as emergency response (Bhangale et al. (2016)). Larger issues can also be solved with it, like mapping and change detection of forest at global scale within acceptable timeframes (Yin et al. (2017)).

5. CLASSIFICATION OF COMPUTATIONAL SYSTEMS

5.1. SINGLE-CORE SEQUENTIAL ALGORITHMS

On a computer, a single-core sequential algorithm executes serially. It consists of many actions that convert an input into an output, such as computations, loops, and decisions. A typical single-core sequential algorithm is shown in Figure 1. In this case, accessing the Level 1 cache naturally takes only a few clock cycles, whereas accessing other levels inevitably takes more cycles.

Figure 1

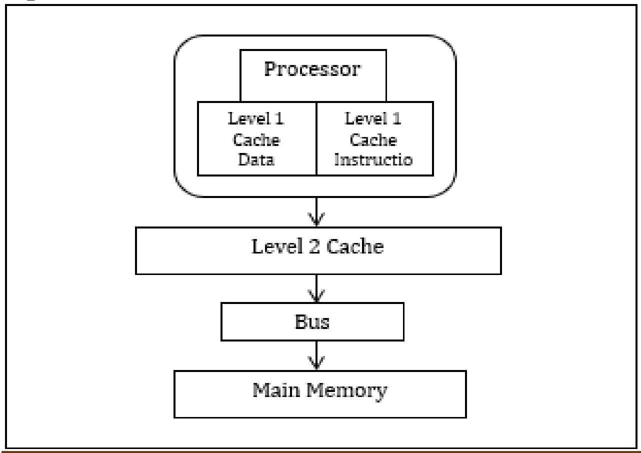

Figure 1 Single-Core Sequential Algorithm

The conventional argument for sequential algorithms is their worst-case performance; however, this can be deceptive (Roughgarden (2019)). Consequently, the database community has created a highly valued collection of benchmark datasets to support such Werner Parallel Processing Techniques for High-Performance Big Geodata.

5.2. PARALLEL ALGORITHMS

Early research on using HPC for geospatial analyses concentrated on uniprocessors with comparatively little parallelism. On the other hand, later work used pipelining and more processors that operate in parallel. Additionally, features of the data and algorithm typically require careful consideration for the algorithmic design to improve the performance of a parallel algorithm for handling geospatial data (Guan et al. (2014); Li et al. (2018)). By and large, numerous studies on parallel computing and the adaptation of current computing frameworks have been carried out for geospatial data preprocessing, parallel algorithm design, simulation modelling, and data analysis, (e.g., Zhao et al. (2019); Kang et al. (2019); Safanelli et al. (2020). The subsequent subsections provide two key parallel algorithm categories.

5.2.1. SHARED MEMORY PARALLEL ALGORITHMS

Figure 2

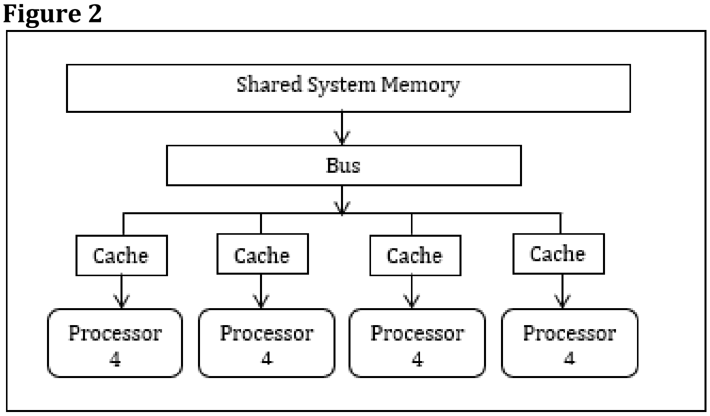

Figure 2 A Simplified View of Four-Processor Shared Memory Design

The usual setup for Parallel Algorithms is a multi-core computer with a single main memory space as shown in Figure 2. Modern multi-core CPU-based workstations that share the same main memory among all cores are notable examples of such systems (Schmidt et al., 2018). Despite operating in parallel, it can carry out a specific subset of operations in an atomic fashion. This suggests that concurrent activities cannot stop the CPU (Sterling et al. (2018). Programs with shared memory have the advantage of having a straightforward, consistent joint state due to global variables. However, their limited scalability is a major drawback (Lee (2014).

6. DISTRIBUTED MEMORY PARALLEL PROCESSING ALGORITHM

A system architecture in which separate, dispersed constituents cooperate to finish an operation without frequently utilizing joint resources is known as a distributed memory. Stated differently, it is a computer system with multiple processors, each of which has a private memory (Pardo et al. (2021). Here, a collection of independent PCs is used, which gives the impression to users that it is a single, cohesive system. When remote data is needed, computational tasks must communicate (via explicit messages) with remote processors to transfer the necessary data. However, computational tasks can function effectively with local data. Supercomputers with thousands of computing nodes typically use this kind of parallel computing. To coordinate their work, the computers—also referred to as nodes—speak with one another via a network. The underlying principles of the communication are typically rather ambiguous; for instance, there is no assurance that a message will be received at all, or even precisely once, nor is there any guarantee on when this will happen. In general, coordinating such systems is difficult and broadly classified into two: either adding a central management component (which makes sense), such as it is often undertaken in cloud computing algorithms (like Hadoop's Node Manager) and HPC algorithms (like MPI rank zero), or introducing a set of guidelines to be adhered by all distributed parts for creating a reliable combined outcome.

7. PARADIGM SHIFT IN GEOSPATIAL BIG DATA COMPUTING

7.1. CLOUD COMPUTING

There is rapid advancement in cloud computing technology, which has culminated in the possibility to execute global-scale multifarious simulations. This is especially true for the efficient management and processing of big geospatial data (Li and Huang (2017). Cloud computing, according to Sugumaran and Armstrong (2017) is a general distributed model that makes network-based configurable computer services, like storage. It provides easy, on-demand, and widespread access to a shared pool of reconfigurable means of computing that can be quickly released with little involvement from providers of service or management. A typical Geospatial Cloud is provided by the Environmental Systems Research Institute (see Figure 3).

Figure 3

Figure 3 The Esri Geospatial Cloud. Source: Environmental Systems Research Institute (ESRI, 2019).

The broader range of technologies and products that Esri offers is embodied in the Esri Geospatial Cloud. Because Esri Geospatial Cloud is designed to scale easily, users can query its billions of records to ask sophisticated questions and perform location analytics. Generally, the proliferation of cloud-based applications highlights the enormous potential that cloud computing offers and represents a revolution in GIScience (Li and Li (2017)). For example, the development of distributed storage for spatial data and parallel spatial algorithms has been aided by certain open-source cloud systems like Spark and Hadoop (Yao et al. (2018); Yao et al. (2018a)).

Large companies (like Google and Amazon) that offer enticing, customizable hardware and software configurations are more likely to make cloud computing services available to the general public. Numerous geospatial problem domains have demonstrated the effectiveness of cloud computing (Hegeman et al. (2014)). Summarily, cloud computing offers vital assistance in processing big data to address the 4Vs and obtain value improved research, operations, and decision support across many geospatial domains (Yang et al. (2016)). Though cloud computing has many advantages, it also has some disadvantages such as latency. Since communication can only happen at the speed of light, it takes place much more slowly (Satyanarayanan (2017)). The rise of Internet of Things-connected electronic devices generates data annually in zettabytes (1021 bytes), which makes bandwidth a key problem (Shi and Dustdar (2016). However, the emergence of fog and edge computing has gained popularity as concepts. Of course, decentralized processing (in fog and edge) reduces the need for bulk data transfers and boosts overall computational performance between distinct tools and the cloud.

7.2. FOG COMPUTING WITH HPC

Data processing at a cloud server is the initial cloud-based GIS model (Barik et al., 2016). A very large time for processing and high internet bandwidth is required for this kind of system. By providing the computation overhead close to the client edge, fog computing solves the issue of lengthy processing times. The greatest enhancement potential in cloud GIS architecture comes from fog computing, which lowers latency and boosts throughput. As a computing paradigm falling between conventional cloud or data centres and smart end devices (Iorga et al. (2018)), it was first used by Cisco in 2012 (Dastjerdi et al. (2016)). In this sense, it is complementary to cloud computing in that it allows users to decentralize data centre resources, improving user experience and quality of service (Sareen, Gupta, and Sood, 2017). However, the processing of various services based on fog computing framework isn't limited to cloud data centres (Monteiro et al. (2016); Sareen et al. (2017); Verma et al. (2017).

With fog computing, the amount of cloud storage required for geospatial big data is typically reduced. Furthermore, a decrease in the needed transmission power leads to an enhanced general efficacy. In the study conducted by Barik et al. (2016), geospatial data was processed at the edge using a Fog computing device. The traditional IoT architecture usually uploads data generated by IoT devices (also called edge devices) directly to the cloud with slight processing because the processing power of edge devices are limited. In fog computing, a mid-computing level is created comprising of a group of fog nodes in between the edge devices and the cloud. An obvious plus of this system is that due to the fact that fog nodes are closer to the edge devices and with their lower network latency, it is possible to rapidly transfer data to them for processing and filtering in real-time. Subsequently, the data is transferrable (after filtering) to the cloud for data mining and analysis using conventional HPC, AI, or Hadoop-like systems. Furthermore, IoT generates geospatial big data because many edge devices use location-based sensors. Thus, real-time processing of geospatial data is important to fog computing.

7.3. DISCRETE GLOBAL REFERENCE FRAMEWORK WITH HPC

Heterogeneity has long been a limiting factor in traditional geospatial data handling techniques. Some examples of the various phases in which heterogeneity manifests itself are in the method by which data are collected, data models and formats, and spatiotemporal resolutions. Also, heterogeneity is produced by geospatial big data because location-based sensors are widely used to collect data from a variety of industries. Combination and integration of geospatial big data with HPC becomes a serious issue when there is much heterogeneity. Owing in part to the absence of a referencing framework capable of effective data storage, data integration and data management for integration of data and parallel processing, most HPC schemes and researches today handle a particular geospatial data kind with particular parallel algorithms.

As a reference frame, traditional coordinate systems like latitude and longitude have proven to be useful. However, a relatively new framework called Discrete Global Grid System (DGGS) is more efficient for managing and processing heterogeneous geospatial big data connected with the Earth's curved surface (Sabeur et al. (2019). Simply, the DGGS divides and addresses the world using a hierarchical tessellation of cells.

7.4. GEOSPATIAL ARTIFICIAL INTELLIGENCE WITH HPC

In computer science, artificial intelligence (AI) is concerned with the use of computer systems to simulate human intelligence in a problem solving environment. It encompasses various fields and subfields. Deep learning— a subfield of machine learning in AI has considerably advanced lately (LeCun et al, (2015). The integration of geospatial and AI technologies results in the emergence of geospatial artificial intelligence (GeoAI). Deep learning and other AI technologies are used by geospatial artificial intelligence (GeoAI) for extracting valuable information from geospatial big data (VoPham et al. (2018) as demonstrated by many studies in literature. A few noteworthy examples are land cover mapping (Kussul et al. (2017); Ling and Foody (2019)) or and remote sensing image classification (Hu et al. (2015)), and object detection (Cheng et al. (2016)). GeoAI presents a promising answer to issues associated with geospatial big data. Similarly, geospatial big data is crucial for training GeoAI's sophisticated deep neural networks (DNNs) and has recently sparked breakthroughs in deep learning.

Tech giants like Google, Microsoft, and IBM have shown a keen interest in creating large-scale platforms for artificial intelligence upon which massive computing clusters runs. However, most recent GeoAI study is carried out on workstations or single-node computers, based on comparatively lesser data volumes to train the model. Of course, developing high-performance, scalable GeoAI structures and platforms that fully utilize geospatial big data to build larger and more sophisticated simulations requires more researches. This can be accomplished by combining HPC technologies with general-purpose big data platforms in Hadoop to handle geospatial big data, as in the case of TensorFlow (Abadi et al. (2016)) and Apache SINGA, with deep learning platforms.

8. CONCLUSION

An enormous amount of geospatial data is produced at a very rapid pace, known as geospatial big data. The traditional computational approach that uses hardware, software, and database technologies for data acquisition, storage, manipulation, analysis, management, and presentation is insufficient for handling these kinds of data. However, handling geospatial big data has become possible by using the HPC technology. Of course, geospatial big data analytics now has a better method due to the application of HPC.

The application of HPC is becoming more important in solving problems related to geospatial big data. However, HPC is confronted with both fresh prospects and problems from geospatial big data. Of course, the geospatial data science has apparently changed in respond to the integration of geospatial big data, AI, cloud computing, fog computing, and big data. In conclusion, HPC will remain essential in this new era because it is sufficient for solving complex problems more quickly.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ..., Kudlur, M. (2016). Tensorflow: A system for large-scale machine learning. In 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16, 265-283.

Alarabi, L., and Mokbel, M. F. (2017). A Demonstration of ST-Hadoop: A Mapreduce Framework for Big Spatio-Temporal Data. Proceedings of the VLDB Endowment, 10(12), 1961-1964. https://doi.org/10.14778/3137765.3137819

Alelaiwi, A. (2017). A Collaborative Resource Management for Big Iot Data Processing in Cloud Cluster Computing. The Journal of Networks Software Tools and Applications, 20(2), 1791-1799. https://doi.org/10.1007/s10586-017-0839-y

Andreazi, G. T., Estrella, J. C., Bruschi, S. M., Immich, R., Guidoni, D., Alves Pereira Júnior, L., and Meneguette, R. I. (2021). MoHRiPA-An Architecture for Hybrid Resources Management of Private Cloud Environments. Sensors, 21, 6857.

Athanasis, N., Themistocleous, M., Kalabokidis, K., and Chatzitheodorou, C. (2018). Big Data Analysis in UAV Surveillance for Wildfire Prevention and Management. in European, Mediterranean, and Middle Eastern Conference on Information Systems, 47-58, Springer, Cham. https://doi.org/10.1007/978-3-030-11395-7_5

Audebert, N., Le Saux, B., and Lefèvre, S. (2018). Beyond RGB: Very High-Resolution Urban Remote Sensing with Multimodal Deep Networks. ISPRS Journal of Photogrammetry and Remote Sensing, 140, 20-32. https://doi.org/10.1016/j.isprsjprs.2017.11.011

Baralis, E., Valle, A. D., Garza, P., Rossi, C., and Scullino, F. (2017). SQL Versus NoSQL Databases for Geospatial Applications. In Proceedings of the 2017 IEEE International Conference on Big Data, IEEE, Boston, MA, USA, 11-14 December 2017, 3388-3397. https://doi.org/10.1109/BigData.2017.8258324

Barik, R. K., Das, P. K., and Lenka, R. K. (2016). Development and Implementation of SOA-Based SDI Model For Tourism Information Infrastructure Management Web Services. In Proceedings of the 2016 6th International Conference of Cloud System and Big Data Engineering (Confluence), 748-753. Academic Press. https://doi.org/10.1109/CONFLUENCE.2016.7723633

Barik, R. K., Dubey, H., Samaddar, A. B., Gupta, R. D., and Ray, P. K. (2016). FogGIS: Fog Computing for Geospatial Big Data Analytics. 3rd IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics Engineering, December 09-11, 2016, Indian Institute of Technology (Banaras Hindu University) Varanasi, India. https://doi.org/10.1109/UPCON.2016.7894725

Batty, M. (2013). Big Data, Smart Cities and City Planning. Dialogues in Human Geography. 3, 274-279. https://doi.org/10.1177/2043820613513390

Bill, R., Blankenbach, J., Breunig, M. et al. (2022). Geospatial Information research: State of the Art, Case Studies and Future Perspectives. PFG 90, 349-389. https://doi.org/10.1007/s41064-022-00217-9

Blais, R., and Esche, H. (2014). Geomatics and the New Cyber-Infrastructure, Geomatica, 62(1), 431- 443.

Bhangale, U. M., Kurte, K. R., Durbha, S. S., King, R. L., and Younan, N. H. (2016). Big Data Processing Using HPC for Remote Sensing Disaster Data. In 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 5894-5897. IEEE. https://doi.org/10.1109/IGARSS.2016.7730540

Biljecki, F., and Ito, K. (2021). Street View Imagery in Urban Analytics and GIS: A Review. Landscape and Urban Planning, 215, 104217. https://doi.org/10.1016/j.landurbplan.2021.104217

Cao, K., Liu, Y., Meng, G., and Sun, Q. (2020). An Overview of Edge Computing Research. IEEE Access, 8: 85714-85728. https://doi.org/10.1109/ACCESS.2020.2991734

Cervone, G., Sava, E., Huang, Q., Schnebele, E., Harrison, J., and Waters, N. (2016). Using Twitter for Tasking Remote-Sensing Data Collection and Damage Assessment: 2013 Boulder Flood Case Study. International Journal of Remote Sensing, 37(1), 100-124. https://doi.org/10.1080/01431161.2015.1117684

Chen, X. W., and Lin, X. (2014). Big Data Deep Learning: Challenges and Perspectives. IEEE Access, 2, 514-525. https://doi.org/10.1109/ACCESS.2014.2325029

Chen, M., Mao, S., Liu, Y. (2014). Big Data: A Survey. Mobile Networks and Applications, 19(2), 171-209.

Chen, B., Yuan, H., Li, Q., Shaw, S., Lam, W. H. K., and Chen, X. (2015). The Spatiotemporal Data Model for Network Time Geographic Analysis in the Era of Big Data. International Journal of Geographical Information Science, 30(6), 1041-1071. https://doi.org/10.1080/13658816.2015.1104317

Cheng, G., Zhou, P., and Han, J. (2016). Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing, 54(12), 7405-7415. https://doi.org/10.1109/TGRS.2016.2601622

Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., and Zhu, Y. (2016). Big Data for Remote Sensing: Challenges and Opportunities. Proceedings of the IEEE, 104(11), 2207-2219. https://doi.org/10.1109/JPROC.2016.2598228

Dastjerdi, A. V., Gupta, H., Calheiros, R. N., Ghosh, S. K., and Buyya, R. (2016). Fog Computing: Principles, Architectures, and Applications. arXiv preprintarXiv:1601.02752. https://doi.org/10.1016/B978-0-12-805395-9.00004-6

De Azevedo, L. J. d. M., Estrella, J. C., Delbem, A. C. B., Meneguette, R. I., Reiff-Marganiec, S. and de Andrade, S. C. (2022). Analysis of Spatially Distributed Data in the Internet of Things in the Environmental Context. Sensors, 22, 1693. https://doi.org/10.3390/s22051693

De Mauro, A., Greco, M., and Grimaldi, M. (2015, February). What is Big Data? A Consensual Definition and a Review of Key Research Topics. In AIP Conference Proceedings, 1644(1), 97-104. https://doi.org/10.1063/1.4907823

Di Minin, E., Fink, C., Hiippala, T., and Tenkanen, H. (2015). A Framework for Investigating Illegal Wildlife Trade on Social Media with Machine Learning. Conservation Biology, 33(1), 210. https://doi.org/10.1111/cobi.13104

Ding, Z., Chen, Z., and Yang, Q. (2014). IoT-SVKSearch: A real-Time Multimodal Search Engine Mechanism for the Internet of Things. International Journal of Communication Systems, 27(6), 871-897. https://doi.org/10.1002/dac.2647

Dong, W., Wang, S., and Liu, Y. (2021). Mapping Relationships Between Mobile Phone Call Activity and Regional Function Using a Self-Organizing Map. Computers, Environment and Urban Systems, 87, 101624. https://doi.org/10.1016/j.compenvurbsys.2021.101624

Environmental Systems Research Institute. (2019). The Emergence of a Geospatial Cloud and the Continuing Evolution of Cloud Computing. https://www.esri.com/about/newsroom/arcnews/the-emergence-of-a-geospatial-cloud/?srsltid=AfmBOopcNilqwmIjwnpqcLjIUxc2UNiPcZAZz6xTtNQ2X

Gaigalas (2019). Advanced Cyber Infrastructure to Enable Search of Big Climate Datasets in THREDDS. ISPRS Int. J. Geo Inf., 8, 494. https://doi.org/10.3390/ijgi8110494

Guan, Q., Zeng,W., Gong, J., and Yun, S. (2014). PRPL 2.0: Improving the Parallel Raster Processing Library. Trans. GIS, 18, 25-52. https://doi.org/10.1111/tgis.12109

Gudivada, V. N., Baeza-Yates, R., and Raghavan, V. V. (2015). Big data: Promises and Problems. Computer, (3), 20-23. https://doi.org/10.1109/MC.2015.62

Guo, H. (2017). Big Earth Data: A New Frontier in Earth and Information Sciences. Big Earth Data, 1, 4-20.

Haklay, M., Antoniou, V., Basiouka, S., Soden, R., and Mooney, P. (2014). Crowdsourced Geographic Information use in Government: Global Facility for Disaster Reduction & Recovery (GFDRR), World Bank: London, UK, 76. [Google Scholar]

Hegeman, J. W, Sardeshmukh, V. B., Sugumaran, R., and Armstrong, M. P. (2014). Distributed LiDAR Data Processing in a High-Memory Cloud-Computing Environment. Annals of GIS 20(4), 255-264. https://doi.org/10.1080/19475683.2014.923046

Huang, H. S., Gartner, G., Krisp, J. M., Raubal, M. and Van de Weghe, N. (2018) Location-Based Services: Ongoing Evolution and Research Agenda. Journal of Location Based Services, 12, 63-93. https://doi.org/10.1080/17489725.2018.1508763

Huang, F., Wen, C., Luo, H., Cheng, M., Wang, C., and Li, J. (2016). Local Quality Assessment of Point Clouds for Indoor Mobile Mapping. Neurocomputing, 196, 59-69. https://doi.org/10.1016/j.neucom.2016.02.033

Huang, H., Yao, X. A., Krisp, J. M., and Jiang, B. (2021). Analytics of Location-Based Big Data for Smart Cities: Opportunities, Challenges, and Future diqrections. Computers, Environment and Urban Systems, 90, 101712. https://doi.org/10.1016/j.compenvurbsys.2021.101712

Hu, F., Xia, G. S., Hu, J., and Zhang, L. (2015). Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing imagery. Remote Sensing, 7(11), 14680-14707. https://doi.org/10.3390/rs71114988

Hu, F., Xu, M. C., Yang, J. C., Liang, Y. S., Cui, K. J., Little, M. M., Lynnes, C. S., Duy, D. Q., and Yang, C. W. (2018). Evaluating the Open Source Data Containers for Handling Big Geospatial Raster Data. ISPRS Int. J. Geo Inf., 7, 144. https://doi.org/10.3390/ijgi7040144

Internet Live Stats, (2019), Available at https://www.internetlivestats.com/twitter-statistics

Iorga, M., Feldman, L., Barton, R., Martin, M. J., Goren, N., and Mahmoudi, C. (2018) Fog Computing Conceptual Model. NIST Special Publication 500-325. Gaithersburg, MD: NIST. https://doi.org/10.6028/NIST.SP.500-325

Ji, C., Xiong, Z., Fang, C., Lv, H., and Zhang, K. (2017). A GPU-Based Parallel Clustering Method For Electric Power Big Data', Proc. - 2017 4th Int. Conf. Inf. Sci. Control Eng. ICISCE 2017, 29-33. https://doi.org/10.1109/ICISCE.2017.16

Kang, J., Fang, L., Li, S., and Wang, X. (2019). Parallel Cellular Automata Markov Model for Land Use Change Prediction Over MapReduce Framework. ISPRS Int. J. Geo-Inf. 2019, 8, 454. https://doi.org/10.3390/ijgi8100454

Kumar, S., Tiwari, P. and Zymbler, M. (2019). Internet of Things is a Revolutionary Approach for Future Technology Enhancement: A Review. J Big Data 6, 111. https://doi.org/10.1186/s40537-019-0268-2

Kussul, N., Lavreniuk, M., Skakun, S., and Shelestov, A. (2017). Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. IEEE Geoscience and Remote Sensing Letters, 14(5), 778-782. https://doi.org/10.1109/LGRS.2017.2681128

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature, 521(7553), 436. https://doi.org/10.1038/nature14539

Lee, J.-G., and Kang, M. (2015). Geospatial Big Data: Challenges and Opportunities. Big Data Res., 2(2), 74-81. https://doi.org/10.1016/j.bdr.2015.01.003

Lee, G. (2014). Switch Fabric Technology, In Lee, G. (Ed) Cloud Networking. Morgan Kaufmann, 37-64. https://doi.org/10.1016/B978-0-12-800728-0.00003-5

Li, C., Wu, Z. and Yin, J. (2017). Research on Oracle-Based Integrative Storage and Management of Spatial Data. International Conference, 2017, 16-22. https://doi.org/10.1145/3077584.3077587

Li, G., and Huang, Z. (2017). Data Infrastructure for Remote Sensing Big Data: Integration, Management and on-Demand Service. Jisuanji Yanjiu Yu Fazhan/Comput. Res. Dev., 54, 267-283.

Li, Q., and Li, D. (2014). Big data GIS. Wuhan Daxue Xuebao (Xinxi Kexue Ban)/Geomatics and Information Science of Wuhan University, 39(6), 641-646. https://doi.org/10.13203/j.whugis20140150

Li, Z., Hodgson, M. E., and Li, W. (2018). A General-Purpose Framework for Parallel Processing Of Large-Scale Lidar Data. Int. J. Digit. Earth, 11, 26-47. https://doi.org/10.1080/17538947.2016.1269842

Li, Z., Yang, C., Yu, M., Liu, K., and Sun M. (2015). Enabling Big Geoscience Data Analytics with a Cloud-Based, MapReduce-Enabled and Service-Oriented Workflow Framework. PloS one, 10(3), e0116781. https://doi.org/10.1371/journal.pone.0116781

Ling, F., and Foody, G. M. (2019). Super-Resolution Land Cover Mapping by Deep Learning. Remote Sensing Letters, 10(6), 598-606. https://doi.org/10.1080/2150704X.2019.1587196

Liu, Z., Guo, H., and Wang, C. (2016). Considerations on Geospatial Big Data. IOP Conf. Ser., 46, 012058. https://doi.org/10.1088/1755-1315/46/1/012058

Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., … Shi, L. (2015). Social Sensing: A new Approach to Understanding our Socioeconomic Environments. Annals of the Association of American Geographers, 105(3), 512-530. https://doi.org/10.1080/00045608.2015.1018773

Lu, P, Chen, G., Ooi, B. C., Vo, H. T., and Wu, S. (2014). ScalaGiST: Scalable Generalized Search Trees for MapReduce Systems [innovative systems paper]. Proceedings of the VLDB Endowment, 7(14), 1797-1808. https://doi.org/10.14778/2733085.2733087

Lu, G., Yuan, L., and Yu, Z. (2017). Surveying and Mapping Geographical Information from the Perspective of Geography. Cehui Xuebao/Acta Geodaetica et Cartographica Sinica, 46(10), 1549-1556. https://doi.org/10.11947/j.AGCS.2017.20170338

Lee, J.-G., and Kang, M. (2015). Geospatial Big Data: Challenges and Opportunities. Big Data Res., 2, 74-81. https://doi.org/10.1016/j.bdr.2015.01.003

Ma, L., Li, M., Ma, X., Cheng, L., Du, P., and Liu, Y. (2017). A Review of Supervised Object-Based Land-Cover Image Classification. ISPRS J. Photogramm. Remote Sens., 130, 277-293. https://doi.org/10.1016/j.isprsjprs.2017.06.001

Marciniec, M. (2017). Observing World Tweeting Tendencies In Real-Time. https://codete.com/blog/observing-world-tweeting-tendencies-in-real-time-part-2

Maschi, L .F. C., Pinto, A. S. R., Meneguette, R. I., and Baldassin, A. (2018). Data Summarization in the Node by Parameters (DSNP): Local Data Fusion in an IoT Environment. Sensors, 18, 799. https://doi.org/10.3390/s18030799

Mauch, V., Kunze, M., and Hillenbrand, M. (2013). High-performance cloud computing. Future Generation Computer Systems 29(6), 1408-1416. https://doi.org/10.1016/j.future.2012.03.011

Monteiro, A., Dubey, H., Mahler, L., Yang, Q., and Mankodiya, K. (2016). FIT: A Fog Computing Device for Speech Tele-Treatments. 2016 IEEE International Conference on Smart Computing (SMARTCOMP), St. Louis, MO, 2016, 1-3. https://doi.org/10.1109/SMARTCOMP.2016.7501692

Mulyono, S., and Fanany, M. I. (2015). Remote sensing Big Data Utilization for Paddy Growth Stages Detection. Paper presented at the IEEE International Conference on Aerospace Electronics and Remote Sensing Technology, Bali, Indonesia, December 3-5. https://doi.org/10.1109/ICARES.2015.7429821

Niculescu, V. (2020). On the Impact of High-Performance Computing in Big Data Analytics for Medicine. Applied Medical Informatics. 42. 9-18.

O'Sullivan, D., and Manson, S. M. (2015). Do Physicists have Geography Envy? and What can Geographers Learn from It? Annals of the Association of American Geographers, 105, 704-722. https://doi.org/10.1080/00045608.2015.1039105

Pardo, D., Matuszyk, P. J., Puzyrev, V., Torres-Verdín, C., Nam, M. J., and Calo, V. M. (2021). Parallel Implementation. Modelling of Resistivity and Acoustic Borehole Logging Measurements using Finite Element Methods. Elsevier. https://doi.org/10.1016/C2019-0-02722-7

Pektürk, M.K., and Ünal, M. (2018). Performance-aware High-Performance Computing For Remote Sensing Big Data Analytics. Open Access Peer-Reviewed Chapter, Intechopen. https://doi.org/10.5772/intechopen.75934

Roughgarden, T. (2019). Beyond Worst-Case Analysis. Commun. ACM, 62, 88-96.

Sabeur, Z, Gibb, R., Purss, M., (2019). Discrete Global Grid Systems SWG. http://www.opengeospatial.org/projects/groups/dggsswg

Safanelli, J. L., Poppiel, R. R., Ruiz, L. F. C., Bonfatti, B. R., Mello, F. A. d. O., Rizzo, R., and Demattê, J. A. M. (2020). Terrain Analysis in Google Earth Engine: A Method Adapted For High-Performance Global-Scale Analysis. ISPRS Int. J. Geo-Inf., 9, 400. https://doi.org/10.3390/ijgi9060400

Sareen, S., Gupta, S. K., Sood, S. K. (2017). An Intelligent and Secure System for Predicting and Preventing Zika Virus Outbreaks Using Fog Computing. Enterprise Information Systems, 1-21. https://doi.org/10.1080/17517575.2016.1277558

Sareen, S., Sood, S. K., Gupta, S. K. (2017). Secure Internet of Things-based cloud Framework to Control the Zika Virus Outbreak. International Journal of Technology Assessment in Health Care, 1-8. https://doi.org/10.1017/S0266462317000113

Satyanarayanan, M. (2017). The Emergence of Edge Computing. IEEE Computer 50(1), 30-39. https://doi.org/10.1109/MC.2017.9

Schmidt, B., González-Domínguez, J., Hundt, C., and Schlarb, M. (2018). Introduction, In Schmidt, B., González-Domínguez, J., Hundt, C., and Schlarb, M. (Eds): Parallel Programming, Morgan Kaufmann, 1-19, https://doi.org/10.1016/B978-0-12-849890-3.00001-0

Schmid, S., Galicz, E., and Reinhardt, W. (2015). Performance Investigation of Selected SQL and NoSQL Databases. In Proceedings of the AGILE 2015, Lisbon, Portugal, 9-12 2015. https://doi.org/10.1109/MILTECHS.2015.7153736

Shi, W., Cao, J., and Zhang, Q. (2016). Edge Computing: Vision and Challenges. IEEE Internet Things J., 3(5), 637-646. https://doi.org/10.1109/JIOT.2016.2579198

Shi W, Dustdar S (2016). The Promise of Edge Computing. IEEE Computer 49(5), 78-81. https://doi.org/10.1109/MC.2016.145

Steffenel, L. A. (2018). Improving the Performance of Fog Computing Through the Use of Data Locality: 2018 30th International Symposium on computer Architecture and High-Performance Computing (SBAV-PAD), Lyon, France, 2018, 217-224. https://doi.org/10.1109/CAHPC.2018.8645879

Sterling, T., Anderson, M., and Brodowicz, M. (2018). HPC Architecture 1: Systems and Technologies. In Sterling, T., Anderson, M., and Brodowicz, M. (Eds)., High Performance Computing, Morgan Kaufmann, 2018, 43-82, https://doi.org/10.1016/B978-0-12-420158-3.00002-2.

Sugumaran, R., and Armstrong, M.P. (2017). Cloud Computing. The International Encyclopedia of Geography: People, the Earth, Environment, and Technology. New York, NY: John Wiley https://doi.org/10.1002/9781118786352.wbieg1017

Tsou, M. -H. (2015). Research Challenges and Opportunities in Mapping Social Media and Big Data. Cartography and Geographic Information Science, 42(Suppl 1), 70-74. https://doi.org/10.1080/15230406.2015.1059251

Verma, P., Sood, S. K., Kalra, S. (2017). Cloud-centric IoT-Based Student Healthcare Monitoring Framework. Journal of Ambient Intelligence and Humanized Computing, 1-17. https://doi.org/10.1007/s12652-017-0520-6

Versichele, M., de Groote, L., Claeys Bouuaert, M., Neutens, T., Moerman, I., and Van de Weghe, N. (2014). Pattern Mining in Tourist Attraction Visits Through Association Rule Learning on Bluetooth Tracking Data: A Case Study of Ghent, Belgium. Tourism Management, 44, 67-81. https://doi.org/10.1016/j.tourman.2014.02.009

VoPham, T., Hart, J. E., Laden, F., and Chiang, Y. Y. (2018). Emerging Trends in Geospatial Artificial Intelligence (geoAI): Potential Applications for Environmental Epidemiology. Environmental Health, 17(1), 40. https://doi.org/10.1186/s12940-018-0386-x

Wang, L., Ma, Y., Yan, J., Chang, V., and Zomaya, A. Y. (2018). Pips Cloud: High-Performance Cloud Computing for Remote Sensing Big Data Management and Processing. Future Generation Computer Systems, 78, 353-368. https://doi.org/10.1016/j.future.2016.06.009

Watts, D. J. (2014). Common Sense and Sociological Explanations. American Journal of Sociology, 120, 313-351. https://doi.org/10.1086/678271

Wei, X., Yao, X. A., (2021). Constructing and Analyzing Spatial-Social Networks from location-Based Social Media Data. 48(3), 258-274. https://doi.org/10.1080/15230406.2021.1891974

Wu. A. (2019). Novel Method of Missing Road Generation in City Blocks Based on Big Mobile Navigation Trajectory Data. ISPRS Int. J. Geo-Inf., 8, 142. https://doi.org/10.3390/ijgi8030142

Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., ..., Roy, D. P. (2016). The Global Landsat archive: Status, Consolidation, and Direction. Remote Sensing of Environment, 185, 271-283. https://doi.org/10.1016/j.rse.2015.11.032

Xia, J., Yang, C., and Li, Q. (2018). Building a Spatiotemporal Index for Earth Observation Big Data. Int. J. Appl. Earth Obs., 73, 245-252. https://doi.org/10.1016/j.jag.2018.04.012

Yang, C., Q. Huang, Z. Gui, Z. Li, C. Xu, Y. Jiang, and J. Li. (2013). Cloud Computing Research for Geosciences. In Yang, C., Huang, Q., Li, Z., Xu, C., and K. Liu, (eds.) Spatial Cloud Computing: A Practical Approach. 295-310. Boca Raton, FL: CRC Press/Taylor & Francis. https://doi.org/10.1201/b16106-23

Yang, C., Huang, Q., Li, Z., Liu, K., and Hu, F. (2017). Big Data and Cloud Computing: Innovation Opportunities and Challenges. International Journal of Digital Earth, 10(1), 13-53. https://doi.org/10.1080/17538947.2016.1239771

Yang, C., Yu, M., Hu, F., Jiang, Y., Li, Y., Yang, C., et al. (2016). Utilizing cloud computing to Address Big Geospatial Data Challenges. Computers, Environment and Urban Systems, http://dx.doi.org/10.1016/j.compenvurbsys.2016.10.010

Yang, C., Yu, M., Hu, F., Jiang, Y., and Li, Y. (2017). Utilizing Cloud Computing To Address Big Geospatial Data Challenges. Computers, Environment and Urban Systems, 61, 120-128. https://doi.org/10.1016/j.compenvurbsys.2016.10.010

Yang, J., Paudel, A., and Puri, S. (2019). Spatial Data Decomposition and Load Balancing on HPC Platforms. (Computer Science Faculty Research and Publications. 40. https://epublications.marquette.edu/comp_fac/40Geospatial

Yao, X., and Li, G. (2018). Big Spatial Vector Data Management: A Review. Big Earth Data, 2(1), 108-129. https://doi.org/10.1080/20964471.2018.1432115

Yao, X., Mokbel, M., Ye, S., Li, G., Alarabi, L., Eldawy, A., Zhao, Z., Zhao, L., and Zhu, D. (2018a). Landqv2: A MapReduce-Based System for Processing Arable Land Quality Big Data. ISPRS Int. Geo.-Inf., 7, 271. https://doi.org/10.3390/ijgi7070271

Yao, X., Yang, J., Li, L., Ye, S., Yun, W., and Zhu, D. (2018). A Parallel Algorithm for Partitioning Massive Spatial Vector Data in a Cloud Environment. Wuhan Daxue Xuebao (Xinxi Kexue Ban)/Geomat. Inf. Sci. Wuhan Univ. 2018, 43, 1092-1097.

Yin, D., Liu, Y., Padmanabhan, A., Terstriep, J., Rush, J., and Wang, S. (2017). A CyberGIS-Jupyter Framework for Geospatial Analytics at Scale. In Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact, 18. https://doi.org/10.1145/3093338.3093378

Yu, J., Wu, J., and Sarwat, M. (2015). Geospark: A Cluster Computing Framework for Processing Large-Scale Spatial Data. In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, Bellevue, WA, USA, 3-6 November 2015, 70. https://doi.org/10.1145/2820783.2820860

Yue, P., Jiang, L. C. (2014). BigGIS: How Big Data Can Shape Next-Generation GIS. In Proceedings of the Third International Conference on Agro-Geoinformatics, Beijing, China, 11-14 August 2014, 413-418. https://doi.org/10.1109/Agro-Geoinformatics.2014.6910649

Zhao, K., Jin, B., Fan, H., Song, W., Zhou, S., and Jiang, Y. (2019). High-Performance Overlay Analysis of Massive Geographic Polygons that Considers Shape Complexity in a Cloud Environment. ISPRS Int. J. Geo-Inf. 2019, 8, 290. https://doi.org/10.3390/ijgi8070290

Zhuang, Y., Fong, S., Yuan, M., Sung, Y., Cho, K., and Wong, R. K. (2017). Location-Based Big Data Analytics for Guessing the Next Foursquare Check-ins. The Journal of Super computing, 73(7), 3112-3127. https://doi.org/10.1007/s11227-016-1925-2

Zook, M., and Breen, J. (2017). Volunteered Geographic Information. In: Shekhar, S., Xiong, H., Zhou, X. (eds) Encyclopedia of GIS. Springer, Cham. https://doi.org/10.1007/978-3-319-17885-1_1656