iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** III **Month of publication:** March 2022

DOI: https://doi.org/10.22214/ijraset.2022.41038

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

Review on Integrating Geospatial Big Datasets

Ms. Dhanya Anto[1], Ms. Mariya C.S[2], Dr. Sonia Sunny[3]
[1, 2, 3]*Department of Compute Science, Prajyoti Niketan College,Pudukkad ,Thrissur, Kerala*

Abstract: ***Big Data and Local Systems (GIS) are both expanding technologies has influenced many areas over the past 10 years and will continue to improve and help resolve a global crisis problems, such as the effects of climate change or a global epidemic. More GIS applications it works with the continued growth of large geospatial data sources to drive precise and informed decisions. Geospatial Big Data Integration is designed to achieve compatibility with different geospatial data sets without the availability of space. A large number of large geospatial data sources seek to operate effectively the integration of data storage and management of such data, which will be used for geospatial data analysis and vision. For example, risk management data sets are related to health care and the environment heterogeneous and distinct. Finding an integrated view of large geospatial databases is also difficult it is challenging, especially when we consider the problems associated with the health epidemic and natural disasters. Therefore, before we try to predict and mitigate the processes that take place in these domains, we should see that big data integration is very important in combining data sets. We explore and chat issues involved in compiling large geospatial data sets in this study. We then split the big data collection it processes it in three phases, namely, data storage, data conversion and integration methods. In addition, several research challenges focus on large geospatial data, large global data, data retention, data conversion and linked data are presented. Lastly, open up research issues and emerging styles requires in-depth research into the near future highlighted in this study.***
Keywords: Big Data, Geospatial

I. INTRODUCTION

Geographic information systems (GISs) work with the largest number of big data sources with various properties. GIS can import, export, store, manage, analyse, process and process visualize the location data displayed and play an important role in compiling and analysing a large number of geospatial data [1]. Most of the information is passed on such as geospatial data, which is collected using technology, such as global applications (GPS), radio frequency directions, and local volunteer details as well community-based social networks. Geospatial data used in applications related to land use, environmental management, health, tourism, marketing and many more. However, most of this data is only available on its own [2].

It incorporates geospatial data from different data sources is widely used and important in health care- and environmental-related applications for new information and make informed decisions [3], [4]. A growing number of geospatial data sources are stored in information contain sensor data, local volunteer information and location-based data. Even ordinary people can create something new or update existing geospatial data in a timely manner advanced web technology and location-based devices as well services. Combining geospatial appropriate with non-geospatial appropriate data can help provide relevant and timely information for the wrong people because different details will do is included in many sources.

The integration of data obtained from different sources is known as data aggregation [5]. The basic premise of the integration program is the same idea of one approach viewer access to data stored and accessed from a variety of sources data sources with a universal system for arbitrator or data shed (DW) [6]. Data aggregation is often fragmented in six stages, namely, manual assembling, provision shared user interface for users, integration implementation, use of middleware interface, development for associated data access, and for the construction of shared data storage [7]. Modern decision-making systems rely heavily on DWs, a concept first introduced in the 1980s [8]. Combined data from various sources are stored and made accessible to multidimensional formats in DWs for targeted research to help users improve their knowledge about business.

The extract-transform-load (ETL) process is widely used to collect a lot of DW data from company operating details and includes releases and converting data into a pre-mass format uploaded to DW as cubes for subsequent analysis with reporting and online processing tools (OLAP). Faster loading methods are often used for performance ETL always during the end of DW inactivity [9]. Although it collects general data from various databases systems are complex, many solutions have been proposed data integration from different relationship systems. However, the absence of homosexuality in data systems shows that different data models, such as relationships or different unrelated data models, are used [7]. Typical methods in data access, discovery and integration are reversed with linked data [10].

©IJRASET: All Rights are Reserved

Geospatial data can be shared fairly and access to local data infrastructure (SDI) based on linked data structures, such as standard data model, standard method of data access and data acquisition depending on the links. Semantic interactions between variations web applications and services can be accessed by addressing the issue of semantic heterogeneity on the basis of ontologies [6].

Problems associated with obtaining data from too many sources are a challenge. The main difference between Standard data integration (BDI) related to the amount of data sources, the structural homogeneity, changes environment and various data sources depending on their quality, such as punctuality, accuracy and integration [11].

In addition, Large geospatial data aggregation is designed to achieve the compatibility of a different geospatial database without looking at it Localization [12].

Data is converted from range formats, speculation or reference systems, followed by adjustment according to the data model provided [13]. This process makes it easier with integrated geospatial to analysed, processed and visualized. This is available from integration of data collected from various sources with various approaches to integrated viewing.

Significantly, many applications from transport planning to disaster management are based on BDI data access and analysis [14]. More lessons on file literature reviewed large geospatial data. For example, Eldawy and Mokbel [15] introduced a period of geospatial magnitude data. Li et al. [2] reviewed major geospatial data methods as well major challenges. Yao and Li [16] introduced a major geospatial vector data management.

However, these studies ignored the complete integration of large geospatial data. So, this article reviews the texts with general information and greater geospatial integration and aims to improve similar taxonomy that can help researchers understand the field and make progress complete geospatial BDI methods. Moreover, we believe that this review will be helpful to novice researchers appearing different sectors and domains. Extensive review of existence methods of geospatial and comprehensive data integration, BDI research in geospatial areas in the past

Five years were presented in this study. Extensive review examines existing definitions and signs of geospatial and big data integration. Relationship between big data, BDI and geospatial data are also discussed. Existing BDI subjects are divided into the following categories: (i) data retention, (ii) data modification, and (iii) integration methods. In addition, several research challenges are open research issues, as well as geospatial data trends are discussed.

The whole paper is organized as follows: Definitions and Large geospatial data features are presented in Phase II. BDI segmentation revealed in Section III. A summary of current research challenges as well open research issues, particularly geospatial BDI-related problems, are presented in sections IV and V, respectively. Trends in future development are presented in section VI. Finally, the end of the lesson on discussion and the future the work is presented in section VII.

II. DEFINITION AND AND CHARACTERISTICS OF GEOSPATIAL BIG DATA

According to Liu et al. [17], common symptoms of big data volume, variability and velocity, collectively known as the 3V model. IBM [18] is considered authenticity is an important factor in accurate interpretation big data. Volume refers to different data sources, including sensor and social networks, producing a large amount of daily data beyond the traditional processing capacity for information.

Variety pertains to the various structures of generated data, including structured and unstructured data. Examples include asset plans, blog posts, tweets and mobile applications. Tria et al. [19] note that speed has arrived since the need to quickly transfer data between sources to stay competitive.

Geospatial data includes position (e.g. to create geometry or linking) and lexical (e.g. construction words) details [2]. Various methods are used to capture geospatial data, including GPS, satellite imagery, social media, location-based services and remote sensors for high sensitivity, which will then be entered into the GIS database, which saves geo reference-related data that specifies location details related to the connection between data points, non-geospatial (attribute) features and other issues. Located locally data sets are important because they not only determine the location characteristics of any given data point but also reflect the time and type of event. In addition, this data can be represented by a type of raster or vector data. Lastly uses geometric shapes, such as dots and lines in a vector model to represent geospatial objects and has benefits of accuracy, low volume and high quality there considered in relation to the raster data model [1], [16]. Geospatial data has long been regarded as big data. Managing geospatial data is very important because the file a large amount of data is available in geo direction, as evidenced by the questionable claim that ‘80% of the data is geographically ’’ [2]. Nowadays, large geospatial data is important in many fields such as data analysis and acquisition. Many community applications, such as environmental Transformation and disaster risk management, may benefit from large geospatial data [20].

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

III. CLASSIFICATION OF BIG DATA INTEGRATION

With the exception of systematic, less structured details as well informal data is also continuously used by organizations to improve their decision-making business analysis performance [21]. As a result, data from various sources must be accessed, analysed and shared. Collaboration and integrating the functions of different information systems is challenging due to the volume, diversity and large speed structures data and lack of systemic similarity brightness data. The need for data integration by organizations as well businesses have led to the emergence of a new research field called BDI [11], determined by the key elements of big data. Therefore, BDI plays a major role in various combinations data sets. In this review, we introduce the BDI segmentation. This division will help geospatial data players find a complete understanding of how to combine differences geospatial or non-geospatial sources have also been beneficial details of high value.

A. Data Warehousing

DW was created within pre-existing conditions depending on the traditional servers used to use the connection for information. Although the term was used in the 1980s, the data storage became an independent research topic in the topic of in the late 1990s. Significantly, DW has a strong connection with related topics, such as data perception and data consolidation [23]. DW was established mainly due to the need for maintenance and query historical data and is a way to extract. The dispersed data is stored in a variety of information systems and is compiled into a single component and integrated storage system. Significant changes in the use of similarities structures for the required features of big data requirements are considered. Santoso and Yulia [24] note that various building methods, such as shareddisk, shared memory, shared-shared and shared methods can be accepted. DW is usually divided into five types, i.e., old, modern, equal integration techniques, local DWs, and data cubes.

Among the major categories of data, variations and volumes are present commonly used in almost all subjects. Velocity too authenticity is ignored in most studies due to lack of complete solutions due to the nature of big data.

1) *Classcal Data Warehouses:* Businesses often use a variety of IT systems to cover different jobs and places of business, as a customer relationship management, accounting and staff. DWs are usually made when managers need integrated and adapted information acquisition strategy [25]. Classical DW is built from a logical data collection that provides an overview of the organization. The details, however, are different from any other operation of information to allow the consolidation of shares of many different data regardless of the system or system obtained from. Centralization creates a one-department entry policy for managers to access and evaluate information strategic decision-making. However, the traditional DW is able to manage large amounts of data with the arrival of big data. Santoso and Yulia [24] noted that data processing should be done in the same way. Descriptive data is this built on an old DW, which can't handle it with formal or informal details. Therefore, modern DWs can be an option for large amounts of data.
2) *Modern Data Warehouses:* Modern DWs are designed to deal with emerging data sources, including sensor and social networks, generate a lot of data every day. Not just a structured query language (NoSQL) included for advanced processing volumes of organized, slow-moving and unstructured data volumes. Bicevska and Oditis [25] have shown that NoSQL technology is ready to process high-volume data from various settings. Akid and Ayed [26] expanded NoSQL technology and we proposed a completely new construction of modern DW to build NoSQL-focused document a database that stores and analyses large social data. However, both [25] and [26] ignored the ETL process under NoSQL DW, thereby reducing the performance of these DWs also restrict the transfer of required data.
3) *Parallel-Processing Architectures:* Big data is made on multiple devices, e.g. Internet of Things (IoT) and local-based social networks. Similar building structures (e.g. MapReduce) [27] can solve old DW issues, which cannot be discussed single node solutions due to the large load required in data analysis. Big data analysis in distributed areas actually made with MapReduce-based Hadoop structures [28]. Supporting the most effective queries geospatial data values are becoming increasingly important in various application domains. Answer questions quickly a key condition for geospatial applications involving large ones data values. This situation requires an unstable suspension that is capable of questioning large geospatial data. For this purpose, a spatial data warehousing (SDW) system Hadoop-GIS demonstrating distribution and efficiency was proposed in [29] as an option to use a large geospatial questions on Hadoop. In addition, GeoSpark [30], a computer memory system, was developed for use. Large geospatial data and support for various types of geospatial data, indicators and function by root enhancement of Apache Spark; hence, GeoSpark extends the concept of distributed data sets are powerful. Compared to Hadoop-based systems, GeoSpark has been strongly shown that improve performance time performance [15].

©IJRASET: All Rights are Reserved

4) *Spatial Data Warehouses:* SDW is used by geospatial decision support systems (DSSs) to query and analyze information related to information data [13], [31]. According to Baazaoui-Zghal [32], SDW has become increasingly needed in many areas, such as health care and environmental disaster risk management. SDW it is generally considered a viable option when it comes to large amounts of data. The decision-makers became to better understand the urgent need to successfully keep ever-increasing values both in order unstructured data. Li et al. [33] suggested the use of NoSQL data as the largest geospatial data repository at a time using a traditional geospatial database as an application server. Baazaoui-Zghal [32] suggested the use of ontology an incomprehensible SDW for content search and recommendations in the collection of uncertain data on many levels of a layer of information during decision-making, summary searches and recommendations remain the same. Background of relevant and interesting information from SDWs can be complex. Therefore, recommendation programs are intended to help users in their roaming through large data sets and acquisitions of relevant information based on their analytical objectives and your interests.
5) *Data Cubes:* Comprehensive listing is being developed [34], data cubes are among the newest and most efficient methods Big Earth Data (BED) storage and analysis. Based on these data cubes, the data from the combined data saves the location, for a while and various sizes are reduced, cut or removed very quickly [35]. Without providing the amount of development, data cubes also stimulate the application of open geospatial concert (OGC) standards in contact with geospatial data. There are several important efforts that have been launched in recent times for major solutions data problems faced by various scientific groups depending on use of data cubes. For example, EarthServer4 ensures consistency between BED analysis and the integrated list products [36], while global viewing data (EO) may be better organized and analyzed on the basis of data frameworks and tools integrated with the analytical framework for Open Data Cube [37]. In addition, a continuous encounter and edge analysis of data fragments and information details (e.g. standards and usage) are facilitated by Research Data Alliance [35]. However, data usage cubes are a problem because they want to be kept in the novel too processes paradigms to ensure the same speed of queries each size [34].

B. Data Transformation

The ETL process can be used to compile and upload data from completely different sources; this process starts with the creation of an integrated storage database [14]. Jobs involved in the ETL process is data acquisition from multiple sources there is only one source (domain), to process data according to levels of storage integrity (transformation), and data overcrowding in the form of new records (loaded) [38]. Therefore, the ETL process is important for DW architecture [39] because it loads all available data into DW before the start of user queries [9]. ETL the process is divided into three types, namely, traditional ETL, extract-load-transform (ELT) and spatial ETL (SETL).

1) *Traditional ETL:* Traditional ETL is called an aspiring process because it is just beginning by uploading complete data sources to an integrated database of data [14]. Data is poorly managed ETL determined mainly due to (1) maximum loading time and bandwidth, and (2) possible data authorizations surrounded when all data source content is uploaded to a file data aggregation; however, this passing will increase risk of data breach by allowing data access to users [14]. There are many suggested ways to resolve these issues discharge [9], [40]. The lazy ETL includes an aspiring combination as well only metadata uploads have real data entry [14]. Traditional ETL methods are a graphical user interface (GUI) to simplify the process of transferring data from the source of the targeted program and the ETL program with manually coded, providing high flexibility. Lots of open sources too ETL marketing tools, such as Talend, Informatica and IBM, are available. Iswas et al. [34] suggested an ETL approach supports near-real-time ETL processes through additional uploads. However, this model cannot support The GUI also requires users to work with a few lines of code. Sreemathy et al. [42] used Talend Open Studio as a GUI ETL tool to streamline data integration process and analysis process. GUI ETL tools are very useful because they provide users has many features that are ready to use and can be made with different fields due to the repeated ETL process.
2) *ELT:* A large number of health care [3], [4] and disaster management data [48], [49] are performed in a wide variety of settings. ETL system cannot manage file for order terabyte and petabyte data because it works often supported by only one machine known as ETL server. However, a complex state of big data is possible it may be treated by matching or distributing the data cluster processing. In this regard, the DSS community suggested that the ETL process should be split in half many cluster nodes to solve this problem. Therefore, ETL can perform better due to the same management of source data segregation by the various components of the ETL process [50]. Another option in such cases is the use of Hadoop, which includes ELT or 'extraction' of data from sources, 'uploading' data to HBase data, as well as 'modification' and data integration in a directed way to the nest [51].

3) *Spatial ETL:* Integrating GIS technology with data storage capabilities makes it possible to obtain SDWs. ETL extended to SETL in [38] for this purpose. Meanwhile, an additional 'Talend local extension' plug-in was developed [40] by using Talend Open Studio Data Combination to create a SETL process and enable complete data integration of geospatial features between information. Domain, conversion and upload of data format types made of many elements of the Talend. However, Shape files with geometric data are not read using the the standard Talend version. Jo and Lee [52] D_ETL as delayed ELT steps using the same data the processing process to deal with a large amount of IoT geospatial data. Test results showed better D_ETL data adjustment performance compared to traditional ETL and ELT when working in large quantities of data. However, D_ETL has shown malicious operation in simple data analysis and demonstrated the inability to support geospatial data queries. SpatialHadoop [47], GeoSpark [30] and other geospatial programs are proposed to address related issues big data processing and data analysis. However, a method that supports most of those programs involves the introduction of geospatial species or activities data in current big data systems, thus not providing direct use.

C. *Integration Methods*

Domain requirements play a key role in choosing the right method or process for data integration. The division of integration methods into two main types of Data models with semantic integration are proposed dealing with a wide range of different BDI methods as well strategies.

1) *Database Data Models*

a) *RDBMS:* The widely used relationship data model is the basis for systems relating to database management (RDBMS). This technology is applied to older DWs and legacy systems [53]. Without facilitating maintenance and operation for small and systematic data, RDBMS is also popular GIS data storage technology. PostgreSQL and related PostGIS geospatial extension shows benefits, such as as effective functions in vector and raster models [54]. Wide a range of models, schemas or formats used in a variety of ways organizations. The heterogeneity of the data can be (1) syntactic or (2) plan (55). Syntactic heterogeneity is caused by use of various data systems, such as relationships or object-oriented details and geometric presentations (e.g. raster presentations or vector). Scheme heterogeneity it caused by the use of different data models to represent the same real objects. RDBMS verifies the availability of information is consistent and robust in the context of data management. However, some problems persist in the end, access and data storage, management of less structured and unstructured data, as well as implementation horizontal thickness [56].

b) *NoSQL:* RDBMS achieves maximum capacity and is insufficient in the management of large data [57]. Therefore, NoSQL database data is used with such problems. Currently, large amounts of unplanned data are available on GIS. Previous studies have been tested NoSQL database for storing and query geospatial data. Vathy-Fogarassy and Hugyák [7] have proposed a data integration framework that supports different GIS data models. This framework allows data retrieval from RDBMS and NoSQL databases simultaneously, integrate different data sources, and monitor causal users. NoSQL uses a variety of methods and workflows without requiring system technology to users. Zhang et al. [58] suggested a way to store large geospatial data based on an unrelated MongoDB database. MongoDB and Python writing languages have benefited high performance compared to traditional information data. Rainho and Bernardino [59] proposed web GIS using NoSQL data to increase efficiency for finding big GIS data from the web. Web GIS welcomes MongoDB as a NoSQL database also provides geospatial data in GeoJSON format with web service. A lesson the results showed that storing and retrieving random geospatial data using MongoDB yields better results compared to RDBMS data details. In addition, MongoDB provides competent local operators and allows for the maintenance of various data structures. NoSQL systems restore the gap caused by storing large data and maintain performance by supporting multiple data models, such as column stores, key value stores, info details and graph information details [60]. NoSQL database information usually has it many of the issues are true because they fail to use data integrity or consolidation strategies as opposed to related information details, viz apply strict data policies to ensure simultaneous delivery of the same data to all users. Khalfi et al. [61] introduced another large data center process addressing energy issues and ensuring the consistency of geospatial data in a text-based framework with the NoSQL framework through workflow that authorizes certain consistency. GeoJSON used schema as a logical example of enhancing both integration and NoSQL data and semantic support issues fixed storage.

2) Semantic Integration

a) Linked Data: Semantic web allows universal access to the connected web data using the linked data paradigm. Details are provided with relevant formats and links to additional web data sources for continuous growth in geospatial capacity data connection file such geospatial data can be accessed via the web in government-run SDIs and volunteer, scientific activities and the nature of companies, and are influenced by factors, such as law, market complaints and social media [62]. Data can be edited, distributed, received, accessed and integrated in new ways with connected data. Geospatial resources in SDIs can be properly identified and shared exploiting the power of connected data, as usual data model, standard data access method and Data link-based identification. Normal OGC storage as well other geospatial web services that demonstrate interaction are designed to help implement SDI [10]. Therefore, user technology is not required to retrieve, access once share information on the semantic web with SDI. Connected data can improve the performance of the whole thing process, please create universal data sharing infrastructure depending on the publication of the data in the device description framework (RDF) as an example of standard data, and then use links to link different data. Purss et al. [64] proposed the standard OGC standard in the form of a smart earth grid system to provide a coherent integration area as well visualize vector geometry and geastery based raster data sources in the same way as information on the use of computer graphics to the computer screen pixels. Composition of streaming connected widgets. The database is expanded to [65] to facilitate creation. Mashups platform plug-in connected in real time. The basic principle of this construction is representation all data processing operations and provisioning of user assistance in creating semantic data stream as well end user mashups and processing visuals data flows to get information in real time. However, these levels are difficult to apply due to geospatial data Joining geometry statistics and user requirement technology.

b) Ontologies: Ontologies transmit information as a formal character of the targeted domain. Ontologies are widely used in data integration systems [66] and are important in data semantics because they clearly define the domain in a comprehensive way mechanical method. Ontologies facilitate interaction between various web applications and services in order to note the problem of semantic inhomogeneity [6]. An integrated set of unique and unique data sources can inquired by end users by category of called programs access to ontology-based data (OBDA) to reduce the need for IT assistance [67]. Ontology map and hybrid ontology can provide a description of the system (structure) as well semantic interactions, respectively [68]. Several studies have talked about BDI ontology. For example, Abbes and Gargouri [66] proposed a web of ontology language backed by NoSQL database, MongoDB and modular ontologies, with sources equal to big data. It produces local ontologies and creates the universe ontology by local construction. Storage and processing of unique data from more than one source in their original form allowed big data configuration. Nadal et al. [67] has developed a systematic supported ontology is RDF in the form of BDI ontology assisting modeling and integration to improve data from the plural there is only one provider. A composite heterogeneous collection too. Different data sources may be requested by end users via OBDA reducing the need for IT assistance. However, these studies are not able to analyze a large amount of data actually time. In addition, the minimum data status of NoSQL databases considered an obstacle to the ontology integration process, and schema maintenance in response to domain requirements will contribute to the achievement of ontology.

IV. RESEARCH CHALLENGES

The study of large geospatial data is in the development stage despite many organizations applying for it GIS. Volume, variety, speed and accuracy are an important factor which define big data and emphasizes the issues of homosexuality. However, the need for data integration by users has led in the emergence of a new research field called BDI, i.e. limited by key features of big data. This section identifies challenges in the following areas of focus: major geospatial data, big earth data, data storage, data conversion, data models, linked data and ontologies.

A. Geospatial Big Data

Geospatial information is available to public and private organizations. The main geospatial data management field is underdeveloped [16]. Therefore, developing technical and theoretical approaches and providing solutions to critical issues is necessary to understand the core of GIS theory. Currently, state-of-the-art techniques, including automation, are used to collect large geospatial data, especially sensory data, thereby creating novels and threats. The retrenchment rate also occurs due to high variability of formats, data providers, resources and usage. Widths of Barriers are also associated with continuous data recovery, data synchronization methods and various operations that high-level users need. As a result, the same Geospatial data can be obtained several times and the final products may vary despite facing similar locations [69].

Previous studies focused on textual data integration [70]. At the same time, large picture scrolls, audio and sensor data are available despite these types of data is rarely combined with text data on a shared information base. Therefore, the required database is determined components are essential to facilitate use, sharing and geospatial data consolidation and defining order levels in a geospatial data setting. Moreover, creation of responsive, accessible and easy-to-use web sites, viz users can rely on decision-making support, which is also important.

B. Big Earth Data

In view of the development of large geospatial data with the use of technology, the BED field is very attractive attention. The GIS domain has been continually filled with the building of theory and power strategies in line with the latest advances in cuttingedge computing technology such as NoSQL knowledge as well cloud computing [16]. BED is the bottom line of big data relating to EO data such as satellite details, weather data and human activity data. It is primarily intended with regard to simplifying world-based assessment and understanding communication through analysis and understanding available data [71]. There must be a variety of skills and technologies combined to perform a complex analysis process BED, which brings great difficulty especially with about how to keep them, analyze them and visualize them. In addition, it is necessary to create temporary-space data a model that successfully supports cloud configuration. Related to the non-relationship details and the file system submitted are currently the main data storage strategies. Given this circumstances, the difficulty in the final system can be reduced by using various storage methods data.

C. Data Warehousing

Relying on traditional data processing techniques not enough for the age of big data. Critical GIS barrier is related to the integration of large databases. NoSQL has been removed DW can provide novel features for data analysis, e.g. it did not happen under the old DW programs. Also, an important promise is associated with redistribution data matching methods and processes. New techniques may include heterogeneous geospatial as well non-geospatial data in the integrated SDW context urgently needed. To build DWs are actually distributed by and it is important because the data is reproduced slowly again shared between organizations. Novel making DW with NewSQL related data. NewSQL and NoSQL databases have similar structures, such as distribution structures and large associated processing. In addition, the NewSQL database is an effective management method big data. Although the structure of DW was tested by researchers, the DW trial was not evaluated in comparison [63]. Data quality testing is used to verify data sources in advance upload to target DW. Similarly, strategies have always been available It is proposed that a single-dimensional data verification be implemented methods ignore data changes or losses from ETL process.

D. Data Transformation

The ETL process can collect data from variance sources. However, the ETL should be re-evaluated to address a complex state of big data. Traditional ETL systems generally they operate on the same machine as the ETL server and cannot managing large data sets. Therefore, new methods were developed. It is capable of cloud computing, MepReduce and NoSQL data models and improvements and improvements to the Performance of current ETL methods is required. Remarkably, The ETL frameworks suggested in the books are not compatible with the automatic installation of people in this process. Therefore, human-induced ETL processes are manual, repetitive, and time-consuming. Distributed ETL processes can promote integrated data sharing and reduce duplication data uploading and integration efforts that accumulate high bandwidth. However, there is still room for research on this topic. Additional clues should be subject to testing physical structures for construction and testing the effort required to design an ETL section in relation to a specific domain [19]. To check for discharge problems, modify and upload geospatial data in the context of more geometric presentations are also required [13]. Currently, ETL processes cannot successfully integrate real-time data usage with historical data, such as data stored in data sources.

E. Database Data Models

Identify related information from of extensive information stored in various databases management systems (DBMSs) are becoming increasingly popular is challenging due to the high diversity in data technology and size [72]. Building an effective last resort is a thing it is necessary to read and write large geospatial data [73] because standard RDBMS is limited to maintenance and analysis of big data. Large amount of data and RDBMS final performance and informal geospatial questions data are major GIS issues. These problems are important because the required queries and real time require active access at large geospatial data rates across the Internet [74].

Future investigations should address the same issue schema comparison process and implementation of methods that use the use of a large amount of RAM memory than computers can provide right now. In addition, data dictionary and domain additional metadata database factors not considered [75]. Big data has it has been instrumental in the recent management of random NoSQL data. However, geospatial data details are no longer ignored.

F. Linked Data

Integration of heterogeneous data from multiple sources resources can be successfully accessed through the data linked in the context of big data, including initial modification of differential geospatial data is used current geospatial linked data in the integrated RDF can be read by machines [76]. However, the processing of Web content is usually intended for human users and not equipment. In addition, to achieve the main objectives of the connection data, i.e., linking and integration can be challenging and it is expensive [77]. Various organizations and authorities use different models, schema or formats. Offering of easy-to-use GUI integration tools that can facilitate the sharing, interpretation and re-use of information [10]. However, the accuracy and completeness of the data is still low.

G. Ontologies

A key problem in BDI is the automatic construction of an ontology model and the identification of potentially indirect semantics data sources. Personalization of ontologies is fast and error-free. Moreover, to keep and regenerating ontologies is a daunting task. Allowing users to obtain a combined concept in a powerful heterogeneous the data source group is complex and is referred to in literature as a data diversity challenge [67]. Therefore, the efficacy of ontologies is a critical research topic in yet. Further research could develop novel models of extracting ontologies from other commonly used data sources, as NoSQL databases. Easy-to-see recognition by a major goal even though it is linked to geospatial resources in the RDF formats can be processed by computer programs. GUI Tools for making clear semantic meanings are also appropriate provision. Excessive exposure to technical information is limited of existing SPARQL implementation [10]. So, it works user interface is required for queries than SPARQL questions.

V. OPEN RESEARCH ISSUES

One's direct understanding is quickly surpassed by the scale and the complex nature of big data. Therefore, machine support it is necessary in semantic analysis, editing and translation to find the number of such great techniques as well data on diversity [78]. Many studies cannot fund it large geospatial data.

Many traditional integration options are inaccessible due to their failure to address the problems of BDI was created by their volume, speed, variability and accuracy. Automation-based integration methods are needed to replace existing manual methods. Few of these problems discussed in the following paragraphs.

A. Data Sources

Continuous proliferation of data sources and volumes, this becomes difficult due to existing data in multiple domains, integrate multiple geospatial datasets into a single dataset [79]. Diversity of data sources in government agencies, private organization, geospatial resolution, estimation and Storage formats lead to significant challenges in geospatial BDI. Various geospatial principles and methods can be used to integrate traditional data as well as address geospatial large data [2]. Given the wide range of data providers, technical Tools and views are insufficient to obtain geospatial Data integration from multiple sources. Idea of these is also institutional, social, legal and policy specifications.

B. Semantic Integration

Personality differences occur when the real world is the same the object is interpreted differently by different categories or user groups [80]. Semantic separation can also take the formation of heterogeneity, in which words are different given the same thing of the real world or a different real world items have the same name. Geospatial data sharing is challenging due to its unique character and results in Data replication problems [55]. Although manual comparisons methods can be used for data integration, the process involved requires considerable effort in the case of large databases [68]. Significant progress has been made five years ago, but to deal with the problems caused different thinking and translation of geospatial data, exchanging information between different domains and integrating multilingual data still needs to continue investigation [81].

C. Data Quality

Data quality can be interpreted depending on the purpose of its use. Organizations, companies and users are responsible to set their quality requirements. According to [82],

Quality has several meanings; for example, 'quality is the level at which a set of natural features complements needs; suitability for use; compliance'. In addition, data quality consists of the following dimensions: data consistency, data duplication, data completeness, data cost and data accuracy [17]. Investigators users of large geospatial data must understand how the data provider's behavior affects large data quality, and check big data quality and possible errors, such as precise positioning, logical consistency, and so on Factors related to data accuracy, prior to data analysis. Different data sources can be used to improve reliability of these findings. Visualization and management of NoSQL data quality are fundamental challenges. Inside big data uncertainty, such as human data sets, impedes their development and implementation [2], [83].

D. Storage Geospatial Big Data

Data storage plays an important role in processing and analyzing a large number of structured, slow-moving structures random data [16]. The maximum number of details is

Paired performance issues with RDBMS for storage as well random data, while geospatial data queries are important challenges associated with GIS. Given that works well Access to geospatial data for Internet data is essential for making demands and real-time, several studies have recently examined the use of cloud paradigm in solving these problems [45]. It is driven by the great power of computing and storage infrastructure, many studies use NoSQL DBMS, such as MongoDB and HBase. As a result, an integrated metadata format with successful data integration. The framework is urgently needed. In addition, studies should be focus on in-depth learning algorithms, especially the use of semantic matching and conversion of integrated remote control format hearing metadata [84].

E. Processing Geospatial Big Data

Processing large data obtained from GIS has become increasingly difficult due to the volume of data [57]. Understanding Geospatial data is important because of its positive effect in a range of consecutive domains and applications. The inability to wait until complete data is available it is a critical problem of geospatial algorithms in real time big data processing. As a result, the distribution and comparison of geospatial algorithms is important. Identifying effective strategies for quickly processing user requests, such as answers in less than a second, it is a major obstacle. Similarly, additional novel performance that promotes the development of a geospatial data framework is necessary. Scientists currently works with effective strategies for quickly processing user requests, such as answers in less than a second, it is a major obstacle. Similarly, additional novel performance that promotes the development of a geospatial data framework is necessary.

F. Big Earth Data Analytics

Life time analysis for preparation, analysis, mining, and visualizing enough volumes of different types of space data constitutes BED analytics. In this process, understanding of the world system can be improved and the challenges posed by global and regional transformation. The level can be better assessed by obtaining a range of relevant data, including patterns, causes, and information [85]. Although there have been many attempts to make it create an integrated model for the analysis of large data [71], the analysis process remains a difficult task that requires that a combination of various skills and technologies. Performing such meta-analyzes is done more frequently, It is hard to imagine a huge amount of data shortages structure. Moreover, the distribution and similarity are also and problems, as with big data in general. So, inside of this age of big data, geospatial statistics need to adapt structures that can use existing information as far as possible, the effective measurement of the volume of the data, shows compliance with the different models of the model, and provides users with options to view and visualize data jointly [86].

G. Big Spatiotemporal Data Analytics

Temporary large-scale data analysis is required to investigate and apply appropriate techniques, frameworks, and solutions for large data generated by space and time stamps [91]. The checking procedures and your texts are both natural and social events are greatly enhanced by to feel the new technology as it happens in following the point of COVID-19 problems [4]. However, space-term analysis he has not yet matured, there are still many problems to come overcome, including pattern types that can be found from time series data and practical strategies as well high efficiency.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Development of new real-time strategies event availability can help resolve such issues. In addition, it is important to achieve improved data integration with the identification of an inclusive event and a wide range with we used a rapid spatio-temporal spread data sources. This can have far-reaching effects about how events are scientifically understood, but also of operational procedures that support decision-making associated with events [92].

VI. EMERGING BIG GEOSPATIAL DATA TRENDS

In the current context of large geospatial and EO data, growth over the years reflects its emergence from the norm areas to additional areas of application such as health disaster risk management, natural disaster risk reduction and predicting, dealing with human activities and driving themselves cars. To address these needs, emerging technologies are available was designed and developed. This section points to a major geospatial data styles in the following areas of focus: large geospatial cloud computing, large geospatial data in context of artificial intelligence (AI) and machine learning (ML), smart geospatial data acquisition and geospatial data content understanding.

A. Big Geospatial Data Cloud Computing

Cloud computing can facilitate the distribution of computer resources by ensuring that they are used as efficiently as possible in relation to CPU, RAM, network and storage [93]. The field of Computer Sciences currently makes a paradigm change in the direction of cloud computing [94], which has been useful in many operating systems and greatly improved storage and computer cost-effective [95]. Given the open access to greatness geospatial data value, appropriate storage, processing, transfers, and the analysis of such data brings difficulties with standard SDI. This created new creation cloud-based technologies, such as GeoRocket, which is among the first targeted cloud-based technology especially in the management of geospatial data [94]. Other technology that works in the cloud and helps geospatial data sets will be scientifically analyzed and visualized scale is Google Earth Engine and EO System Data Availability, Processing and Global Monitoring Analysis [37]. However, there is one important problem related to cloud computing platforms, namely, vendor lock Data migration can be hampered by the fact that management and processing functions are not the same cloud platforms [71]. In addition, maintenance and performance of insufficient geospatial data within remote cloud servers can be it depends on the delay and the use of force for the sake of the country some form of geospatial data [96].

B. Big Geospatial Data In The Context Of Ai And Ml

Like computer power, learning and application algorithms conditions have become more complex and varied; the use of AI in various fields has intensified. Geospatial scientific data is of great benefit to AI, especially when used alongside large-scale data analysis [97]. In addition, BSD analytics can be refined with new graphics methods, such as descriptive AI and ML rendering [98]. With the provision of various domains where there is a large geospatial data relevant (e.g. tracking infection, climate change simulation, disaster risk management, etc.), the study focused on providing geospatial extensions to current ML solutions or build completely new solutions to make them more efficient analysis and intelligence of existing applications [99]. However, further research is needed to determine which ones Geospatial applications have a major impact as well combine geospatial and parallelization techniques in this period of big data [100]. The lack of homogeneity of big data causes some difficulty in researching the most advanced construction techniques that will be used with great success [101].

C. Smart Geospatial Data Discovery

The production of big data every day brings great difficulties the field of Earth Sciences with respect to geospatial data availability and accessibility [102]. Basically, an application of connected data and direct data acquisition is affected by lack of semantic homogeneity of geospatial data [103]. In this context, the construction of geospatial data sites is proposed as a solution for generating large geospatial data easily accessible [104].

Jiang et al. [102] has developed an intelligent web-based geospatial data acquisition program, where the metadata behavior of the mine is used as well use data compatibility. Besides, in the context of sensible urban and environmental institutions built, there are many and semantics analysis can benefit from the integration of semantics architectural data processing (BIM) and GIS, which can and provide opportunities for information access and informed decision making [105]. In addition, creating graphs of information from multiple sources can be it is conceptually connected in a space-temporary manner, thus providing researchers with reliable and efficient services [106].

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

D. Geospatial Data Content Understanding

Combining unique data also leads to better data representation and understanding. General research data can be exploitation for research purposes to a very large extent with a combination of geospatial statistics and big data strategies. For example, highlighting how income inequality enters and life was related. Haithcoat et al. [107] included large geospatial statistics and large general research data. Another important use of large geospatial data in controlling self-driving cars as a new smart border transportation, using the ability of these types of cars to feel the environment and operate less with or without human intervention [108]. Moreover, it is important insight into travel behavior, traffic flow, and the environment can be found in the knowledge of large geospatial details. The programs and tools of the novel are required to successfully scan existing data archives, if further provision expansion of EO data [101]. However, the accuracy of the data is reasonable considered to gain an inclusive understanding such data [109].

VII. CONCLUSION

We provide a review of geospatial BDI methods and infrastructure. Classification of BDI methods, including data retention, data modification and integration methods, are proposed in this study. A large number of subjects related to data storage and ETL tools are re-evaluated it is summarized, and their relaxation and limitations are highlighted. Many studies focus on the limitations of so-called structured data while ignoring random data, especially large geospatial data. Our study clearly shows that complete geospatial BDI methods require further investigation. Finally, there are a number of persistent challenges, research issues as well; Trends are discussed on the basis of current big data period. This review covers data sources, semantic integration, data quality, processing and storage of geospatial data; as well the development of these research ideas will benefit the field.

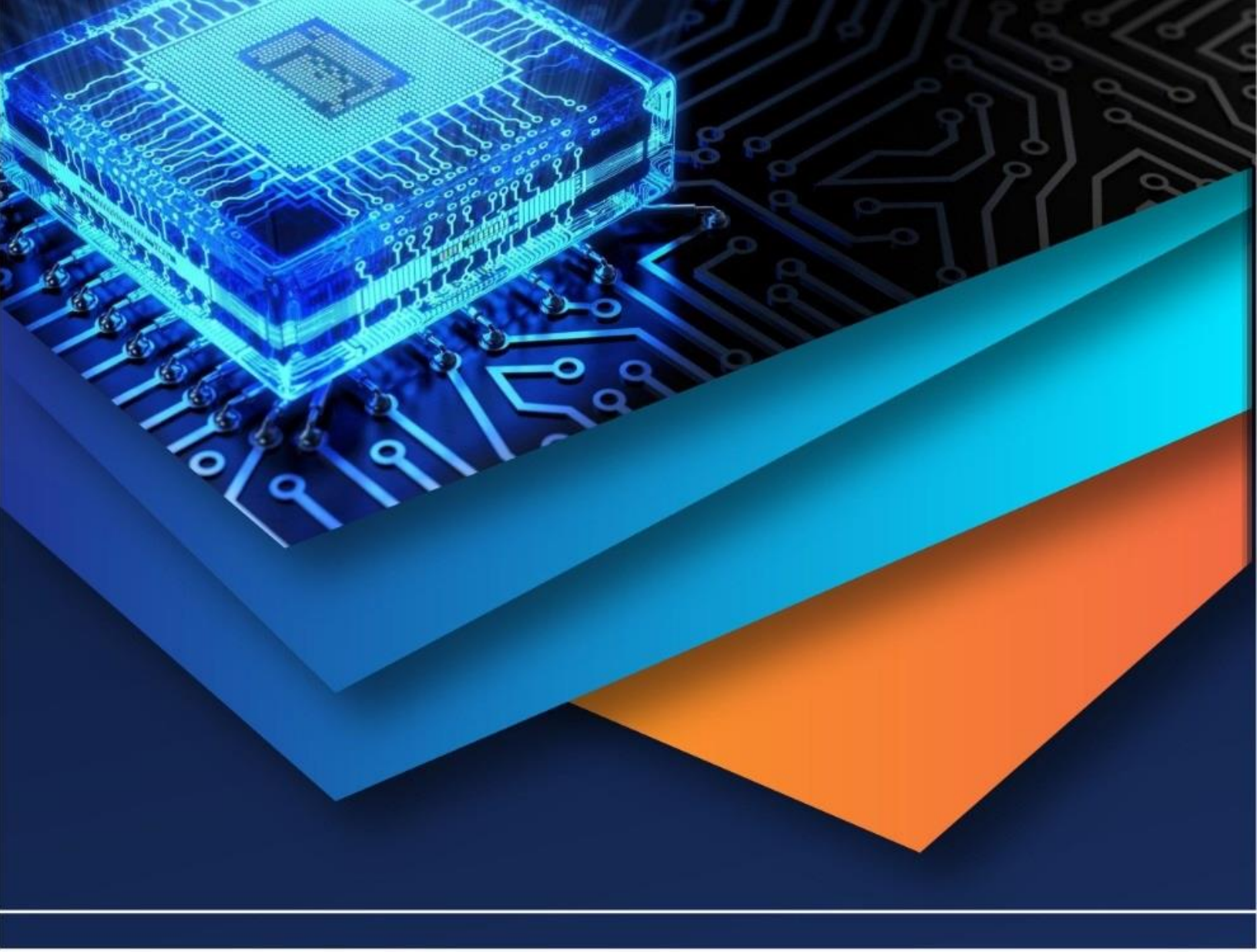
REFERENCES

[1] Goyal, C. Sharma, and N. Joshi, ``An integrated approach of GIS and spatial data mining in big data," Int. J. Comput. Appl., vol. 169, no. 11, pp. 8887_8975, 2017.
[2] S. Li, S. Dragicevic, F. A. Castro, M. Sester, S. Winter, A. Coltekin, C. Pettit, B. Jiang, J. Haworth, A. Stein, and T. Cheng, ``Geospatial big data handling theory and methods: A review and research challenges," ISPRS J. Photogramm. Remote Sens., vol. 115, pp. 119_133, May 2016.
[3] J. P. Mcglothlin, A. Madugula, and I. Stojic, ``The virtual enterprise data warehouse for healthcare," in Proc. 10th Int. Joint Conf. Biomed. Eng. Syst. Technol., vol. 5, pp. 469_476, Feb. 2017.
[4] C. Zhou et al., ``COVID-19: Challenges to GIS with big data," Geography Sustainability, vol. 1, no. 1, pp. 77_87, Mar. 2020.
[5] M. Lenzerini, ``Data integration: A theoretical perspective," in Proc. 21st ACM SIGMOD-SIGACT-SIGART Symp. Princ. Database Syst. (PODS), 2002, pp. 233_246.
[6] O. El Hajjamy, L. Alaoui, and M. Bahaj, ``Semantic integration of heterogeneous classical data sources in ontological data warehouse," in Proc. Int. Conf. Learn. Optim. Algorithms, Theory Appl., May 2018, pp. 1_8.
[7] Á. Vathy-Fogarassy and T. Hugyák, ``Uniform data access platform for SQL and NoSQL database systems," Inf. Syst., vol. 69, pp. 93_105, Sep. 2017.
[8] M. Golfarelli and S. Rizzi, ``From star schemas to big data: 20C years of data warehouse research," in A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years. Cham, Switzerland: Springer, 2018, pp. 93_107.
[9] L. Baldacci, M. Golfarelli, S. Graziani, and S. Rizzi, ``QETL: An approach to on-demand ETL from non-owned data sources," Data Knowl. Eng., vol. 112, pp. 17_37, Nov. 2017.
[10] P. Yue, X. Guo, M. Zhang, L. Jiang, and X. Zhai, ``Linked data and SDI: The case onWeb geoprocessingwork_ows," ISPRS J. Photogramm. Remote Sens., vol. 114, pp. 245_257, Apr. 2016.
[11] X. L. Dong and D. Srivastava, ``Big data integration," in Proc. IEEE 29th Int. Conf. Data Eng., Apr. 2013, pp. 1245_1248.
[12] R. Flowerdew, ``Spatial data integration," Geogr. Inf. Syst., vol. 1, pp. 375_387, 1991.
[13] M. Ponjavic, A. Karabegovic, E. Ferhatbegovic, and I. Besic, ``Spatial data integration in heterogeneous information systems' environment," in Proc. 42nd Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO), May 2019, pp. 1559_1564.
[14] P. Kathiravelu, A. Sharma, H. Galhardas, P. Van Roy, and L. Veiga, ``Ondemand big data integration," Distrib. Parallel Databases, vol. 37, no. 2, pp. 273_295, Jun. 2019.
[15] A. Eldawy and M. F. Mokbel, ``The era of big spatial data: A survey," Found. Trends Databases, vol. 6, nos. 3_4, pp. 163_273, 2016.
[16] X. Yao and G. Li, ``Big spatial vector data management: A review," Big Earth Data, vol. 2, no. 1, pp. 108_129, Jan. 2018.
[17] J. Liu, J. Li,W. Li, and J.Wu, ``Rethinking big data: A review on the data quality and usage issues," ISPRS J. Photogramm. Remote Sens., vol. 115, pp. 134_142, May 2016.
[18] (2013). The Four V's of Big Data. Accessed: Jan. 7, 2020. [Online]. Available: http://www.ibmbigdatahub.com/infographic/four-vs-big-data
[19] F. D. Tria, E. Lefons, and F. Tangorra, ``Evaluation of data warehouse design methodologies in the context of big data," in Proc. Int. Conf. Big Data Anal. Knowl. (DaWaK), vol. 10440, Aug. 2017, pp. 3_18.
[20] J.-G. Lee and M. Kang, ``Geospatial big data: Challenges and opportunities," Big Data Res., vol. 2, no. 2, pp. 74_81, Jun. 2015.
[21] I. A. Ajah and H. F. Nweke, ``Big data and business analytics: Trends, platforms, success factors and applications," Big Data Cognit. Comput., vol. 3, no. 2, p. 32, Jun. 2019.
[22] N. J. van Eck and L. Waltman, ``Software survey: VOSviewer, a computer program for bibliometric mapping," Scientometrics, vol. 84, no. 2, pp. 523_538, Aug. 2010.

[23] R. Venkatraman and S. Venkatraman, ``Big data infrastructure, data visualisation and challenges," in Proc. 3rd Int. Conf. Big Data Internet Things (BDIOT), 2019, pp. 13_17.
[24] L. W. Santoso, ``Data warehouse with big data technology for higher education," Procedia Comput. Sci., vol. 124, pp. 93_99, Dec. 2017.
[25] Z. Bicevska and I. Oditis, ``Towards NoSQL-based data warehouse solutions," Procedia Comput. Sci., vol. 104, pp. 104_111, Jan. 2017.
[26] H. Akid and M. Ben Ayed, ``Towards NoSQL graph data warehouse for big social data analysis," in Proc. Int. Conf. Intell. Syst. Design Appl., 2017, pp. 965_973.
[27] Z. Han, F. Qin, C. Cui, Y. Liu, L. Wang, and P. Fu, ``Mr4Soil: A MapReduce-based framework integrated with GIS for soil erosion modelling," ISPRS Int. J. Geo-Inf., vol. 8, no. 3, p. 103, Feb. 2019.
[28] D. Rammer, S. L. Pallickara, and S. Pallickara, ``ATLAS: A distributed _le system for spatiotemporal data," in Proc. 12th IEEE/ACM Int. Conf. Utility Cloud Comput., Dec. 2019, pp. 11_20
[29] S. Wang, Y. Zhong, and E. Wang, ``An integrated GIS platform architecture for spatiotemporal big data," Future Gener. Comput. Syst., vol. 94, pp. 160_172, May 2019.
[30] J. Yu, J.Wu, and M. Sarwat, ``GeoSpark: A cluster computing framework for processing large-scale spatial data," in Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst., Nov. 2015, pp. 4_7.
[31] H. Haroun, A. R. Ghomari, M. Lahlouh, and A. Mehdi, ``Towards a spatial data warehouse for occupational health risk management," in Proc. 1st Int. Conf. Innov. Res. Appl. Sci., Eng. Technol. (IRASET), Apr. 2020, pp. 1_6.
[32] H. Baazaoui-Zghal, ``Fuzzy ontology-based spatial data warehouse for context-aware search and recommendation," in Proc. 11th Int. Joint Conf. Softw. Technol., 2016, pp. 161_166.
[33] Q. Li, S. J. Yang, H. J. Huang, and Y. H. Zhou, ``Geo-spatial big data storage based on NoSQL database," Geomatics Inf. Sci. Wuhan Univ, vol. 42, no. 2, pp. 163_169, 2017.
[34] P. Baumann, D. Misev, V. Merticariu, and B. P. Huu, ``Datacubes: Towards space/time analysis-ready data," in Service-Oriented Mapping. Cham, Switzerland: Springer, 2019, pp. 269_299.
[35] M. Sudmanns, D. Tiede, S. Lang, H. Bergstedt, G. Trost, H. Augustin, A. Baraldi, and T. Blaschke, ``Big Earth data: Disruptive changes in Earth observation data management and analysis?" Int. J. Digit. Earth, vol. 13, no. 7, pp. 832_850, Jul. 2020.
[36] G. A. Pagani and L. Trani, ``Data cube and cloud resources as platform for seamless geospatial computation," in Proc. 15th ACM Int. Conf. Comput. Frontiers, May 2018, pp. 293_298.
[37] V. C. F. Gomes, G. R. Queiroz, and K. R. Ferreira, ``An overview of platforms for big Earth observation data management and analysis," Remote Sens., vol. 12, no. 8, pp. 1_25, 2020.
[38] U. Drescek, M. K. Fras, J. Tekavec, and A. Lisec, ``Spatial ETL for 3D building modelling based on unmanned aerial vehicle data in semi-urban areas," Remote Sens., vol. 12, no. 12, p. 1972, Jun. 2020.
[39] S. Laraichi, A. Hammani, and A. Bouignane, ``Data integration as the key to building a decision support system for groundwater management: Case of Saiss aquifers, Morocco," Groundwater Sustain. Develop., vols. 2_3, pp. 7_15, Aug. 2016.
[40] M. Lupa, W. Sarlej, and K. Adamek, ``Harmonization of datasets in the frame of spatial data infrastructure using ETL tools: A case study of BDOT500 and BDOT10k databases," in Proc. Baltic Geodetic Congr. (BGC Geomatics), Jun. 2018, pp. 217_220.
[41] N. Biswas, A. Sarkar, and K. C. Mondal, ``Ef_cient incremental loading in ETL processing for real-time data integration," Innov. Syst. Softw. Eng., vol. 16, no. 1, pp. 53_61, Mar. 2020.
[42] J. Sreemathy, I. J. V, S. Nisha, C. P. I, and G. P. R. M., ``Data integration in ETL using TALEND," in Proc. 6th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS), Mar. 2020, pp. 1444_1448.
[43] H. Moulai and H. Drias, ``From data warehouse to information warehouse: Application to social media," in Proc. Int. Conf. Learn. Optim. Algorithms, Theory Appl., 2018, pp. 1_6.
[44] M. Mazzei and S. Di Guida, ``Spatial data warehouse and spatial OLAP in indoor/outdoor cultural environments," in Proc. Int. Conf. Comput. Sci. Appl. Cham, Switzerland: Springer, May 2018, pp. 233_250.
[45] S. Bimonte, M. Zaamoune, and P. Beaune, ``Conceptual design and implementation of spatial data warehouses integrating regular grids of points," Int. J. Digit. Earth, vol. 10, no. 9, pp. 901_922, Sep. 2017.
[46] M. Barkhordari and M. Niamanesh, ``Atrak: A MapReduce-based data warehouse for big data," J. Supercomput., vol. 73, no. 10, pp. 4596_4610, Oct. 2017.
[47] A. Eldawy and M. F. Mokbel, ``SpatialHadoop:AMapReduce framework for spatial data," in Proc. IEEE 31st Int. Conf. Data Eng., Apr. 2015, pp. 1352_1363.
[48] A. G. Rumson, S. H. Hallett, and T. R. Brewer, ``Coastal risk adaptation: The potential role of accessible geospatial big data," Mar. Policy, vol. 83, pp. 100_110, Sep. 2017.
[49] V. Bhanumurthy, K. R. M. Rao, G. J. Sankar, and P. V. Nagamani, ``Spatial data integration for disaster/emergency management: An Indian experience," Spatial Inf. Res., vol. 25, no. 2, pp. 303_314, Apr. 2017.
[50] M. Bala, O. Boussaid, and Z. Alimazighi, ``A _ne-grained distribution approach for ETL processes in big data environments," Data Knowl. Eng., vol. 111, pp. 114_136, Sep. 2017.
[51] H.-K. Lin, J. A. Harding, and C.-I. Chen, ``A hyperconnected manufacturing collaboration system using the semantic Web and Hadoop ecosystem system," Procedia CIRP, vol. 52, pp. 18_23, Jan. 2016.
[52] J. Jo and K.-W. Lee, ``MapReduce-based D_ELT framework to address the challenges of geospatial big data," ISPRS Int. J. Geo-Inf., vol. 8, no. 11, p. 475, Oct. 2019.
[53] H. Dhayne, R. Haque, R. Kilany, and Y. Taher, ``In search of big medical data integration solutions_A comprehensive survey," IEEE Access, vol. 7, pp. 91265_91290, 2019.
[54] D. Guo and E. Onstein, ``State-of-the-art geospatial information processing in NoSQL databases," ISPRS Int. J. Geo-Inf., vol. 9, no. 5, p. 331, May 2020.
[55] F. Yu, D. A. McMeekin, L. Arnold, and G. West, ``Semantic Web technologies automate geospatial data con_ation: Con_ating points of interest data for emergency response services," in Proc. 14th Int. Conf. Location Based Services (LBS). Cham, Switzerland: Springer, 2018, pp. 111_131.

[56] F. Gao, P. Yue, Z. Wu, and M. Zhang, ``Geospatial data storage based on HBase and MapReduce," in Proc. 6th Int. Conf. Agro-Geoinformat., Aug. 2017, pp. 1_4.
[57] D. R. D. Almeida, C. D. S. Baptista, F. G. D. Andrade, and A. Soares, ``A survey on big data for trajectory analytics," ISPRS Int. J. Geo-Inf., vol. 9, no. 2, pp. 1_24, 2020.
[58] X. Zhang, W. Song, and L. Liu, ``An implementation approach to store GIS spatial data on NoSQL database," in Proc. 22nd Int. Conf. Geoinfor- mat., Jun. 2014, pp. 4_8.
[59] F. D. C. Rainho and J. Bernardino, ``Web GIS: A new system to store spatial data using GeoJSON in MongoDB," in Proc. 13th Iberian Conf. Inf. Syst. Technol. (CISTI), Jun. 2018, pp. 1_6.
[60] M. R. Ahmed, M. R. Ahmed, M. A. Khatun, M. A. Ali, and K. Sundaraj, ``A literature review on NoSQL database for big data processing," Int. J. Eng. Technol., vol. 7, no. 2, pp. 902_906, 2018.
[61] B. Khal_, C. De Runz, S. Faiz, and A. Herman, ``A new methodology for storing consistent fuzzy geospatial data in big data environment," IEEE Trans. Big Data, Jul. 2017.
[62] S. Wiemann and L. Bernard, ``Spatial data fusion in spatial data infrastructures using linked data," Int. J. Geogr. Inf. Sci., vol. 30, no. 4, pp. 613_636, 2016.
[63] H. Homayouni, S. Ghosh, and I. Ray, ``An approach for testing the extracttransform- load process in data warehouse systems," in Proc. 22nd Int. Database Eng. Appl. Symp. (IDEAS), 2018, pp. 236_245.
[64] M. B. J. Purss, R. Gibb, F. Samavati, P. Peterson, and J. Ben, ``The OGC discrete global grid system core standard: A framework for rapid geospatial integration," in Proc. Int. Geosci. Remote Sens. Symp., Nov. 2016, pp. 3610_3613.
[65] A. M. Tjoa, P.Wetz, E. Kiesling, T.-D. Trinh, and B.-L. Do, ``Integrating streaming data into semantic mashups," Procedia Comput. Sci., vol. 72, pp. 1_4, 2015.
[66] H. Abbes and F. Gargouri, ``Big data integration: A MongoDB database and modular ontologies based approach," Procedia Comput. Sci., vol. 96, pp. 446_455, Jan. 2016.
[67] S. Nadal, O. Romero, A. Abelló, P. Vassiliadis, and S. Vansummeren, ``An integration-oriented ontology to govern evolution in big data ecosystems," Inf. Syst., vol. 79, pp. 3_19, Jan. 2019.
[68] C. Prudhomme, T. Homburg, J.-J. Ponciano, F. Boochs, C. Cruz, and A.-M. Roxin, ``Interpretation and automatic integration of geospatial data into the semantic Web: Towards a process of automatic geospatial data interpretation, classi_cation and integration using semantic technologies," Computing, vol. 102, no. 2, pp. 365_391, Feb. 2020.
[69] Z. Li, ``Geospatial big data handling with high performance computing: Current approaches and future directions," Jul. 2019, arXiv:1907.12182. [Online]. Available: http://arxiv.org/abs/1907.1218210618
[70] X. L. Dong and T. Rekatsinas, ``Data integration and machine learning: A natural synergy," in Proc. Int. Conf. Manage. Data, May 2018, pp. 1645_1650.
[71] P. Merritt, H. Bi, B. Davis, C. Windmill, and Y. Xue, ``Big Earth data: A comprehensive analysis of visualization analytics issues," Big Earth Data, vol. 2, no. 4, pp. 321_350, Oct. 2018.
[72] D. G. D. Reis, M. Ladeira, M. Holanda, and M. de Carvalho Victorino, ``Large database schema matching using data mining techniques," in Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW), Nov. 2018, pp. 523_530.
[73] L. Zhang, Q. Li, Y. Li, and Y. Cai, ``A distributed storage model for healthcare big data designed on HBase," in Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2018, pp. 4101_4105.
[74] L. van den Brink, P. Barnaghi, J. Tandy, G. Atemezing, R. Atkinson, B. Cochrane, Y. Fathy, R. García Castro, A. Haller, A. Harth, K. Janowicz, A. Kolozali, B. van Leeuwen, M. Lefrançois, J. Lieberman, A. Perego, D. Le-Phuoc, B. Roberts, K. Taylor, and R. Troncy, ``Best practices for publishing, retrieving, and using spatial data on theWeb," Semantic Web, vol. 10, no. 1, pp. 95_114, Dec. 2018.
[75] A. Holemans, J.-P. Kasprzyk, and J.-P. Donnay, ``Coupling an unstructured NoSQL database with a geographic information system," in Proc. 10th Int. Conf. Adv. Geogr. Inf. Syst. Appl. Serv. Coupling, 2018, pp. 23_28.
[76] S. Athanasiou, G. Giannopoulos, D. Graux, N. Karagiannakis, J. Lehmann, A. C. Ngomo, K. Patroumpas, M. A. Sherif, and D. Skoutas, ``Big POI data integration with linked data technologies," Adv. Database Technol.-EDBT, vol. 2019, pp. 477_488, Mar. 2019.
[77] M. Mountantonakis and Y. Tzitzikas, ``Large-scale semantic integration of linked data: A survey," ACM Comput. Surv., vol. 52, no. 5, pp. 1_40, Oct. 2019.
[78] D. J. Kim, J. Hebeler, V. Yoon, and F. Davis, ``Exploring determinants of semantic Web technology adoption from IT professionals' perspective: Industry competition, organization innovativeness, and data management capability," Comput. Hum. Behav., vol. 86, pp. 18_33, Sep. 2018.
[79] H. Abbes and F. Gargouri, ``MongoDB-based modular ontology building for big data integration," J. Data Semantics, vol. 7, no. 1, pp. 1_27, Mar. 2018.
[80] L. Ding, G. Xiao, D. Calvanese, and L. Meng, ``Consistency assessment for open geodata integration: An ontology-based approach," Geoinfor- matica, pp. 1_26, Dec. 2019.
[81] M. Kokla and E. Guilbert, ``A review of geospatial semantic information modeling and elicitation approaches," ISPRS Int. J. Geo-Inf., vol. 9, no. 3, p. 31, 2020.
[82] L. Cai and Y. Zhu, ``The challenges of data quality and data quality assessment in the big data era," Data Sci. J., vol. 14, p. 2, May 2015.
[83] A. A. Frozza and R. D. S. Mello, ``JS4Geo: A canonical JSON schema for geographic data suitable to NoSQL databases," GeoInformatica, vol. 24, no. 4, pp. 987_1019, Oct. 2020.
[84] J. Fan, J. Yan, Y. Ma, and L.Wang, ``Big data integration in remote sensing across a distributed metadata-based spatial infrastructure," Remote Sens., vol. 10, no. 1, pp. 1_20, 2018.
[85] C. Yang, M. Yu, Y. Li, F. Hu, Y. Jiang, Q. Liu, D. Sha, M. Xu, and J. Gu, ``Big Earth data analytics: A survey," Big Earth Data, vol. 3, no. 2, pp. 83_107, Apr. 2019.
[86] C. Robertson, C. Chaudhuri, M. Hojati, and S. A. Roberts, ``An integrated environmental analytics system (IDEAS) based on a DGGS," ISPRS J. Photogramm. Remote Sens., vol. 162, pp. 214_228, Apr. 2020.
[87] K. Graff, C. Lissak, Y. Thiery, O. Maquaire, S. Costa, M. Medjkane, and B. Laignel, ``Characterization of elements at risk in the multirisk coastal context and at different spatial scales: Multi-database integration (Normandy, France)," Appl. Geography, vol. 111, Oct. 2019, Art. no. 102076.
[88] K. Bin and T. Rahim, ``Spatiotemporal applications of big data," Int. J. Comput. Appl., vol. 181, no. 21, pp. 5_10, Oct. 2018.

[89] R. P. D. Nath, K. Hose, T. B. Pedersen, and O. Romero, ``SETL: A programmable semantic extract-transform-load framework for semantic data warehouses," Inf. Syst., vol. 68, pp. 17_43, Aug. 2017.
[90] W. Li, M. Song, B. Zhou, K. Cao, and S. Gao, ``Performance improvement techniques for geospatial Web services in a cyberinfrastructure environment_A case study with a disaster management portal," Com- put., Environ. Urban Syst., vol. 54, pp. 314_325, Nov. 2015.
[91] C. Yang, K. Clarke, S. Shekhar, and C. V. Tao, ``Big spatiotemporal data analytics: A research and innovation frontier," Int. J. Geographical Inf. Sci., vol. 34, no. 6, pp. 1075_1088, Jun. 2020
[92] M. Yu et al., ``Spatiotemporal event detection: A review," Int. J. Digit. Earth, pp. 1_27, Mar. 2020.
[93] Y. Li, M. Yu, M. Xu, J. Jhang, D. Sha, Q. Liu, and C. Yang, ``Big data and cloud computing," Manual Digit. Earth, pp. 325_355, Nov. 2019.
[94] M. Krämer, ``GeoRocket: A scalable and cloud-based data store for big geospatial _les," SoftwareX, vol. 11, Jan. 2020, Art. no. 100409
[95] S. Bebortta, S. K. Das, M. Kandpal, R. K. Barik, and H. Dubey, ``Geospatial serverless computing: Architectures, tools and future directions," ISPRS Int. J. Geo-Inf., vol. 9, no. 5, pp. 1_26, 2020.
[96] J. Das, A. Mukherjee, S. K. Ghosh, and R. Buyya, ``Spatio-fog: A green and timeliness-oriented fog computing model for geospatial query resolution," Simul. Model. Pract. Theory, vol. 100, Apr. 2020, Art. no. 102043.
[97] D. Li, Z. Shao, and R. Zhang, ``Advances of geo-spatial intelligence at LIESMARS," Geo-spatial Inf. Sci., vol. 23, no. 1, pp. 40_51, Jan. 2020.
[98] B. Huang and J.Wang, ``Big spatial data for urban and environmental sustainability," Geo-Spatial Inf. Sci., vol. 23, no. 2, pp. 125_140, Apr. 2020.
[99] I. Sabek and M. F. Mokbel, ``Machine learning meets big spatial data," in Proc. IEEE 36th Int. Conf. Data Eng. (ICDE), Apr. 2020, pp. 1782_1785.
[100] Z. Li, W. Tang, Q. Huang, E. Shook, and Q. Guan, ``Introduction to big data computing for geospatial applications," ISPRS Int. J. Geo-Inf., vol. 9, no. 8, p. 487, Aug. 2020.
[101] K. Alonso, D. Espinoza-Molina, and M. Datcu, ``Multilayer architecture for heterogeneous geospatial data analytics: Querying and understanding EO archives," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 10, no. 3, pp. 791_806, Mar. 2017.
[102] Y. Jiang, Y. Li, C. Yang, F. Hu, E. Armstrong, T. Huang, D. Moroni, L. McGibbney, F. Greguska, and C. Finch, ``A smart Web-based geospatial data discovery system with oceanographic data as an example," ISPRS Int. J. Geo-Inf., vol. 7, no. 2, p. 62, Feb. 2018.
[103] K. Sun, Y. Zhu, P. Pan, K. Luo, D. Wang, and Z. Hou, ``Morphologyontology of geospatial data and its application in data discovery," in Proc. 23rd Int. Conf. Geoinformat., Jun. 2015, pp. 1_6.
[104] Y. Jiang, Y. Li, C. Yang, F. Hu, E. M. Armstrong, T. Huang, D. Moroni, L. J. McGibbney, and C. J. Finch, ``Towards intelligent geospatial data discovery: A machine learning framework for search ranking," Int. J. Digit. Earth, vol. 11, no. 9, pp. 956_971, Sep. 2018.
[105] M. Breunig, P. E. Bradley, M. Jahn, P. Kuper, N. Mazroob, N. Rösch, M. Al-Doori, E. Stefanakis, and M. Jadidi, ``Geospatial data management research: Progress and future directions," ISPRS Int. J. Geo-Inf., vol. 9, no. 2, p. 95, Feb. 2020.
[106] Y. Zhu, ``Geospatial semantics, ontology and knowledge graphs for big Earth data," Big Earth Data, vol. 3, no. 3, pp. 187_190, Jul. 2019.
[107] T. L. Haithcoat, E. E. Avery, K. A. Bowers, R. D. Hammer, and C.-R. Shyu, ``Income inequality and health: Expanding our understanding of state-level effects by using a geospatial big data approach," Social Sci. Comput. Rev., pp. 1_19, Sep. 2019.
[108] S. Shang, J. Shen, J.-R. Wen, and P. Kalnis, ``Deep understanding of big geospatial data for self-driving cars," Neurocomputing, pp. 1_2, 2020.
[109] S. Zhang, B. Zhao, Y. Tian, and S. Chen, ``Stand with# StandingRock: Envisioning an epistemological shift in understanding geospatial big data in the `post-truth' era," Ann. Amer. Assoc. Geogr., pp. 1_21, Aug. 2020.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)