Enabling the Big Earth Observation Data via Cloud Computing and DGGS: Opportunities and Challenges

Xiaochuang Yao 1,* Guoqing Li 2, Junshi Xia 3, Jin Ben 4, Qianqian Cao 2, Long Zhao 2, Yue Ma 2, Lianchong Zhang 2 and Dehai Zhu 1

1 College of Land Science and Technology, China Agricultural University, Beijing 100081, China; zhudehai@cau.edu.cn
2 Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China; ligq@radi.ac.cn (G.L); caoqq@radi.ac.cn (Q.C); zhaolong@aircas.ac.cn (L.Z.); mayue@radi.ac.cn (Y.M); zhangle@radi.ac.cn (L.Z.)
3 Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan; junshi.xia@riken.jp
4 Institute of Surveying and Mapping, Information Engineering University, Zhengzhou 450052, China; benj@lreis.ac.cn
* Correspondence: yxc@cau.edu.cn; Tel.: +86-010-62-737-554

Received: 3 December 2019; Accepted: 19 December 2019; Published: 23 December 2019

Abstract: In the era of big data, the explosive growth of Earth observation data and the rapid advancement in cloud computing technology make the global-oriented spatiotemporal data simulation possible. These dual developments also provide advantageous conditions for discrete global grid systems (DGGS). DGGS are designed to portray real-world phenomena by providing a spatiotemporal unified framework on a standard discrete geospatial data structure and theoretical support to address the challenges from big data storage, processing, and analysis to visualization and data sharing. In this paper, the trinity of big Earth observation data (BEOD), cloud computing, and DGGS is proposed, and based on this trinity theory, we explore the opportunities and challenges to handle BEOD from two aspects, namely, information technology and unified data framework. Our focus is on how cloud computing and DGGS can provide an excellent solution to enable big Earth observation data. Firstly, we describe the current status and data characteristics of Earth observation data, which indicate the arrival of the era of big data in the Earth observation domain. Subsequently, we review the cloud computing technology and DGGS framework, especially the works and contributions made in the field of BEOD, including spatial cloud computing, mainstream big data platform, DGGS standards, data models, and applications. From the aforementioned views of the general introduction, the research opportunities and challenges are enumerated and discussed, including EO data management, data fusion, and grid encoding, which are concerned with analysis models and processing performance of big Earth observation data with discrete global grid systems in the cloud environment.

Keywords: big Earth observation data; cloud computing; discrete global grid systems

1. Introduction

With the rapid development of Earth observation (EO) technology and continuous launch of remote sensing satellites, the resolution of Earth observation data is getting higher and higher, and the data quantity and variety are also increasing, which indicate that EO data is gradually stepping into the era of big data [1]. According to statistical data from the Committee on Earth Observation Satellites
(CEOS), over 500 EO satellites have been launched in the last half-century, and more than 150 satellites will be launched in the next 12 years [2,3]. China has launched more than 60 satellites since 1970 for comprehensive observation of the Earth’s systems, including HuanJing (HJ), FengYun (FY), China-brazil earth resource satellite (CBERS), and GaoFen (GF) series. From the European Space Agency (ESA), in terms of Sentinel family, a total of 50,964,670 products had been downloaded by users since the start of data access operations, with a total data volume of 41.35 PB [4]. The big Earth observation data (BEOD) has gradually promoted the development of global industries, research institutions, and application sectors, which has had a profound impact on the study of the Earth system [5,6], contributing to human activities, environmental monitoring, and climate changes, and also provided abundant data resource for the construction of digital Earth [7–9].

BEOD also poses challenges to uses in terms of problem complexity, automatic analysis, and processing efficiency [10–13]. Fortunately, the rapid advancement of cloud computing technology in recent years provides strong computing power, especially for the efficiency of big geospatial data management and processing, which makes it possible to perform complex simulations on a global scale [14]. Cloud computing is seen as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction [15]. By providing at least three types of services, namely, infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS), cloud computing has completely changed the traditional information system architecture. Spatial cloud computing (SCC) [16] with a layer of data as a service (DaaS) was proposed to handle the endemic problems from spatial data model. The proliferation of cloud-based applications illustrates the enormous opportunities that cloud computing presents, and it is also a revolution for the geographic information science [17]. For example, some open source cloud systems, such as Hadoop and Spark, have promoted the development of spatial data distributed storage and parallel spatial algorithms [18,19].

Whether the accumulation of data resource or the improvement of computing power is of great significance for Earth science research in the era of big data remains to be seen. If there is no standard spatiotemporal unified framework, EO data integration and fusion on-demand will be a great challenge, and cloud computing will never play its biggest advantage. To better solve the problem of global data processing, in this paper, we integrate the unified spatiotemporal unified framework, discrete global grid systems (DGGS), with BEOD and cloud computing into a closed-loop solution. Over the past two decades, DGGS has had a relatively perfect mathematical theoretical model and basic functions [20]. In 2014, the Open Geospatial Consortium set up the DGGS Standards Working Group and issued a relevant international standard. In this standard [21], DGGS is defined as an Earth reference system (ERS) that uses a hierarchical tessellation of cells to partition and address the globe. Based on the DGGS infrastructure, which is better than regular grids from map projections [22,23], the system will enable the integrated analysis of very large, multisource, multiresolution, multidimensional Earth observation data [24].

As shown in Figure 1, the trinity of BEOD, cloud computing, and DGGS can give a more perfect solution, and in this trinity, BEOD is the data resource, cloud computing is the computing power, and DGGS is the unified framework. In this paper, we analyze and discuss these three aspects. In such a case, the Earth observation data should not only consider the unified data structure but also combine the characteristics of cloud computing model to design and implement more favorable processing algorithms to maximize the mining and utilization of data values.

The rest of this paper is organized as follows: Section 2 summarizes the development and main characteristics of BEOD. Sections 3 and 4 discuss the technology frontier from cloud computing and DGGS, respectively, and review related work on BEOD. Section 5 highlights the opportunities and challenges for the trinity solution of BEOD, cloud computing, and DGGS. The conclusions of this paper are provided in Section 6.
2. Big Earth Observation Data (BEOD)

2.1. Big Data and BEOD

In recent years, big data has become a hot topic as a proper term, mainly due to the rapid development of the Internet, cloud computing, mobile, and Internet of Things (IoT). The trinity of understanding big data will include those who own the data, those who can process and analyze the data, and those who utilize the data [25]. In an age when everything is a sensor, the global-oriented datasets have shown explosive growth potential. In terms of datasets, according to the white paper by the International Data Corporation (IDC), as shown in Figure 2, the global datasphere will grow from 33 ZB in 2018 to 175 ZB by 2025, and now, more than 5 billion consumers interact with data every day—by 2025, that number will be 6 billion or 75% of the world’s population [26]. The amount of data to be processed is too large and grows too fast, and business requirements and competitive pressures put forward higher requirements for the effectiveness of real-time (near real-time) data processing, and traditional conventional technical methods cannot cope.

![Annual Size of the GlobalDatasphere](image)

Figure 2. Annual size of the global datasphere [26].

In the field of space science, at present, there are more than 1700 global orbiting satellites and about one-third of these satellites are Earth observation satellites [27]. In general, one image data is about a few hundred megabytes, and the data covering China reach terabytes. Therefore, Earth observation data becomes a natural “testing ground” for big data. As shown in Table 1, since 1998, China has launched FY, ZiYuan (ZY), HJ, and GF series of EO satellites [28], which generate a lot of remote sensing data that are applied to support the development of the national economy. With increasing spatial resolutions of sensors and shorter revisiting times, a call for ‘bring the user to the data and not the data to the user’ has started in the EO community [29]. Now, more and more EO data that are made freely available by space agencies come from various archives [30]. From digital Earth to big Earth data, all EO data play an important role and will provide a new vision and methodology to Earth sciences [7].
Table 1. Time series of Earth observation satellites launched by China.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FY-1A</td>
<td></td>
<td></td>
<td></td>
<td>FY-1B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FY-2A</td>
<td></td>
<td></td>
<td></td>
<td>FY-1B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FY-3A</td>
<td></td>
<td></td>
<td></td>
<td>ZY-1 01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FY-1C</td>
<td>FY-2B</td>
<td></td>
<td></td>
<td>ZY-2 02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FY-1D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ZY-1 02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FY-2C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FY-3B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td>Beijing-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HY-2A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HY-1B</td>
<td>SH-1 03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HY-2A</td>
<td>JY-1 02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>JY-1 02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>JY-1 02</td>
<td>HY-2A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>JY-1 02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>JY-1 02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>JY-1 02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>JY-1 02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>JY-1 02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>JY-1 02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>JY-1 02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>JY-1 02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.2. The Characteristics of BEOD

Earth observation data are images of the objective world, reflecting specific characteristics in a time and space interval in the objective world [31]. Therefore, EO data should have at least three dimensions for characterizing space, time, and observed characteristics [32]. From the big data perspective, the characteristics of BEOD have all the features of big data [33], mainly characterized by “3Vs”, i.e., volume, variety, and velocity [34,35], and are more and more obvious. Here, we enumerate the following six aspects: large volume, great variety, multiple resolutions, time series, global scale, and data intensive [10,36,37], as shown in Figure 3.

Figure 3. The characteristics of BEOD.

- Large volume: One remote sensing image amounts to 1 GB, which has led to an explosion in the volume of data as a whole. For example, only one of NASA archives holds 7.5 PB of data with nearly 7000 unique datasets, which only contain in-domain EO data [25].
- Great variety: Due to the difference of sensors, the format of EO data obtained is also varied, such as Hierarchical Data Format (HDF), network Common Data Format (NetCDF), and GeoTIff. Meanwhile, the data structure is also different, which will be a complicated process in the fusion analysis of multisource data [38].
- Multiple resolutions: With the improvement of remote sensing technology, data resolution is also getting better. Most of the network available data are also in the meter level, such as SPOT-5 (2.5 m), IKONOS (1 m), and QuickBird (0.61 m).
- Time series: Remote sensing satellites dynamically monitor the changes according to old and new data, which is unmatched by artificial field measurement and aerial photogrammetry. For example, land satellites 4 and 5 can cover the Earth every 16 days, and the National Oceanic and Atmospheric Administration (NOAA) weather satellite can receive two images per day.
- Global scale: Remote sensing can detect large-scale areas from the air and even the space in a short period, and obtain valuable remote sensing data. These data expand people’s visual space, for example, a terrestrial satellite image, covers an area of more than 30,000 square kilometers.
- Data intensive: According to the statistics of the EO data processing link, the rate of data preprocessing and information extraction is much lower than the rate of data acquisition and transmission. Speed and efficiency remain weak points for EO data [27,39].

3. Cloud Computing

3.1. Cloud Computing and Spatial Cloud Computing

Cloud computing is seen as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction [15]. Cloud computing technology has begun to penetrate all walks of life, especially in the field of data storage. Increasingly,
consumers are fine with lower storage capacity on endpoint devices in favor of using the cloud. By 2020, we believe that more bytes will be stored in the public cloud than in consumer devices (Figure 4), and there will be more data stored in the public cloud than in traditional data centers by 2021 [26]. Meanwhile, a variety of commercial cloud development companies, such as Google, Amazon, and Ali Cloud, have provided versatile services, including infrastructure and scientific computing services.

![Where is the data stored?](image)

Figure 4. Where is the data stored? [26].

For spatial data, cloud computing is unable to make the most of itself directly as it is designed ignoring characteristics of spatial dataset essentially [40]. Hence, spatial cloud computing (SCC) [16] is proposed. The initiative aims to solve the geospatial problems of four intensiveness issues, including data, computing, concurrent, and spatiotemporal intensities. After several years of development, cloud computing technologies or platforms for Earth observation data are also being rapidly developed, for example, Google Earth Engine [6,41] and Esri Geospatial Cloud. And there are also some open-source cloud computing solutions for geosciences [42].

3.2. Cloud Computing for BEOD

Big data management in EO is not just an information technology (IT) issue; the responsibilities go beyond handling large data volumes, providing high-throughput processing capacity, and making available large network bandwidths for swift data access [43]. Cloud computing has emerged as a new paradigm to provide computing as a utility service, including IaaS (infrastructure as a service), PaaS (platform as a service), and SaaS (software as a service) [44]. In terms of efficiency of EO data processing, Figure 5 shows that cloud computing provides some services for big Earth observation data (BEOD), including spatial data infrastructure (SDI), EO data resource, algorithm or model library, processing and computation, systems and applications [11,45,46]. For example, the parallel mosaic and interpretation algorithms based on cloud computing have advantages in efficiency [47,48].
The most straightforward pattern is to provide spatial data infrastructures, such as AWS, Google Cloud, and Aliyun, which have a large number of clusters, machines, and servers that can provide infrastructure level services for users on-demand. The second one is to provide big Earth observation data resources for users, which is called the EO data cloud. Data cloud is the most mature and basic cloud service mode at present. For example, with the cloud-based discovery and access solutions, the Global Earth Observation System of Systems (GEOSS) has built a flexible framework for global and multidisciplinary data sharing in EO realm, and now is evolving from a data infrastructure to an information system [49]. The third and fourth types are to provide an algorithm or model library and processing and computing power. These two parts are relatively more professional, so a few research teams or commercial companies can provide these services. For systems and applications, they are the fastest-growing patterns [50]. For example, SpatialHadoop [51], which is an open-source system, has formed a complete ecological service form geospatial computing to applications. pipsCloud [52], a cloud-enabled high performance computing (HPC) platform for large-scale remote sensing applications, provides Hilbert-R+ tree and RS workflow processing across data centers. The applications are more extensive, involving environmental change [53], urban facilities [54], land cover [55], precision agriculture [33,56], and disaster warning [57] among others.

Although systems based on cloud computing have been shown to make great progress in all aspects for big Earth observation data, challenges still remain regarding the progressive incorporation of the concept of spatial thinking into cloud computing [10,58]. Literature for big data in remote sensing mainly focuses on the voluminous issue and considers it as a data-intensive computing problem [25].

4. Discrete Global Grid Systems (DGGS)

4.1. DGGS Standards and Models

Discrete global grid systems (DGGS) are presented and studied from the 1980s [59]. At present, there are several standards and models accepted by academia about DGGS. In 1994, Goodchild formulated an early version of the general evaluation criteria, serving as comparison standards for different global grids. Kimerling supplemented and improved the standards in 1999. In 2014, OGC (Open Geospatial Consortium) established the standards committee, DGGS Standards Working Group, which is finalizing its work based on inputs and reviews from experts around the world with experience using multiple DGGS. In this standard [60], a formal definition is proposed for DGGS as
spatial reference systems that use a hierarchical tessellation of cells to partition and address entire the globe, which will be a framework for the next era in big Earth data [61]. Discrete global grid systems are characterized by the properties of their cell structure, geo-encoding, quantization strategy, and associated mathematical functions.”

As shown in Figure 6, there are only five Platonic solids [62,63], including the tetrahedron, cube (hexahedron), octahedron, dodecahedron, and icosahedron, which can be used as female parents to construct DGGS. In general, DGGS solutions should be constructed by specifying five substantially independent design choices [59]: a base polyhedron, a fixed polyhedron orientation, a hierarchical spatial partitioning method, transformation, and a method for assigning point representations to grid cells. After a long period of DGGS development, remarkable achievements have been made in subdivision modeling, encoding computation, and grid quality assessment [20,64,65]. In the applications, as shown in Table 2, some DGGS models have been developed and used in academia and business. For example, PYXIS WorldView [60] is a web-based DGGS platform, which adopted ISEA3H (icosahedral Snyder equal area aperture 3 hexagonal grid system) to enable complex spatial queries and analysis on-demand.

![Figure 6. (a) Tetrahedron, (b) cube, (c) octahedron, (d) dodecahedron, and (e) icosahedron.](image)

### Table 2. Several existing DGGS solutions.

<table>
<thead>
<tr>
<th>DGGS</th>
<th>Base Polyhedron</th>
<th>Partitioning Method</th>
<th>Providers</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCENZ-Grid [66]</td>
<td>Cube</td>
<td>Square and triangle</td>
<td>Landcare Research, Lincoln, New Zealand</td>
</tr>
<tr>
<td>Open DataCube [67]</td>
<td>Cube</td>
<td>Square</td>
<td>Geoscience Australia, Canberra, Australia</td>
</tr>
<tr>
<td>DGGRID [68]</td>
<td>Icosahedron</td>
<td>Hexagon</td>
<td>Southern Oregon University, Ashland, Ore, USA</td>
</tr>
<tr>
<td>PYXIS WorldView [60]</td>
<td>Icosahedron</td>
<td>Hexagon</td>
<td>PYXIS Innovation, Kingston, ON, Canada</td>
</tr>
<tr>
<td>HLQT [69]</td>
<td>Icosahedron</td>
<td>Hexagon</td>
<td>Institute of Surveying and Mapping, Information Engineering</td>
</tr>
<tr>
<td>Uber H3</td>
<td>Icosahedron</td>
<td>Hexagon</td>
<td>University, Zhengzhou, Henan, China</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Uber Technologies Inc., San Francisco, CA, USA</td>
</tr>
</tbody>
</table>

### 4.2. DGGS for BEOD

Because of the natural advantages of DGGS, it has been applied to management, analysis, and visualization of spatial data [68–70]. The indexing method based on grid encoding from DGGS is the most common application mode for big EO data [71,72]. Besides, simulation-visualization based on DGGS [70] was carried out for BEOD, and a multiresolution digital Earth model has been developed for managing and processing spatial datasets [73]. DGGS is also integrated to process multidimensional EO data, such as point clouds [74].

It is worth mentioning that, as a relatively complete grid-based ecology, Open Data Cube is increasingly being used to systematically manage and process Earth observation data [63], such as the Australian Geoscience Data Cube [67], Colombian Data Cube [75], Swiss Data Cube [76], China Data Cube [77], and Armenian Data Cube [78]. Based on these data cubes, EO data management, analysis,
and application have been carried out to monitor and detect environmental change and surface water [53].

5. Opportunities and Challenges

The era of big data brings not only rich data resources, but also profound changes to all walks of life [17]. Both the replacement of information technology (cloud computing) and the improvement of basic theory (DGGS) will face various opportunities and challenges in processing and handling BEOD [24,79]. The life cycle of big data processing includes management, access, mining analytics, simulation, and forecasting [44]. Based on the process of EO data governance, this paper enumerates several opportunities and challenges in the following section.

5.1. Opportunities

5.1.1. EO Data Organization and Management

As a global unified space–time framework, the most significant advantage of DGGS is the ability to organize and store a wide variety of Earth observation data according to a unified rule. The goal of data organization and management is to be able to quickly retrieve and query any content in the dataset [14]. DGGS itself has natural advantages, such as the use of grid coding can not only quickly achieve spatial positioning, but also can easily find child nodes and parent nodes [71,80,81]. Besides, the multilevel of DGGS can be highly consistent with the multiscale of EO data, so that the data at different scales (medium, high, and low resolutions) can be uniformly organized and managed, which provide a good data foundation for multisource data fusion analysis in the later stage. In terms of EO data processing performance, the existing literature [41,82–85] is sufficient to demonstrate the advantages of cloud computing.

5.1.2. Fusion Analysis of Global- or Regional-Scale EO Data

The grid division of spatial data in the Cartesian coordinate system can be completed by employing projection [86]. Although it can satisfy people’s direct cognition, when facing large-scale or global data and applications, the data obtained by projection will have serious deformation problems, especially in polar or high latitudes [87]. The global discrete grid system is a spherical mathematical model, which is equivalent at any point on the Earth and is more suitable for the analysis and visualization of global datasets [60,88]. DGGS is a spatiotemporal framework, not only conducive to fusion analysis of remote sensing data with different resolutions and multiple spectrums, but also able to achieve seamless integration with other geographic information system (GIS) data, which will improve the accuracy of data analysis and data utilization [89].

5.1.3. Integration with Cloud Computing Technologies

Cloud computing provides computing technologies for the potential solution of transformation of big data’s four Vs into the fifth V (value) [44]. The relevance and regionalization of spatial data hinder the optimal use of cloud computing technologies in GIS [40]. Based on the DGGS framework, the Earth is divided into multiple, continuous cells, and these discrete cells have the same geometry and the same area. More importantly, they are independent of each other, which makes the discretized spatial data perfectly integrated with distributed storage and parallel computing mechanism in the process of spatial operation and analysis. The integration with cloud computing technologies is the way to discover the value of big Earth observation data.
5.2. Challenges

5.2.1. DGGS Grid Coding in Cloud Environment

In terms of grid retrieval efficiency and performance, on the one hand, it needs to rely on the advantages of grid coding [90] and on the other hand, it needs to improve the query efficiency of data with the help of cloud computing [40]. Grid coding has been the hotspot of DGGS basic theory research [71,72,91]. According to the coding principle, coding schemes are roughly divided into three categories [20,92]: hierarchical coding operation, filling curve coding operation, and integer coordinate coding operation. These coding schemes place more emphasis on spatial proximity, which contradicts distributed or parallel mechanisms. It is not easy to design and implement grid coding methods compatible with distributed storage and parallel computing.

5.2.2. Spatiotemporal DGGS Framework for EO Data

At present, more research is focused on two-dimensional data [74]. EO data apparently have the characteristics of time dimension [1,9], so the research on spatiotemporal DGGS framework is also an important direction. Near-real-time or real-time analyses of Earth observation data are also imminent or asked in applications [11,93]. Time has the same problem as space partition and space scale [94,95]. Comparatively, the time dimension is more challenging to deal with. On the one hand, time itself is recorded and expressed in various ways. On the other hand, the calculation of time is more complicated and unintuitive. We believe there will be better solutions in the future.

5.2.3. Data Interface with Modeling via Cloud Computing and DGGS

Hundreds of conventional analysis models have been developed for EO data analysis, processing, and application. The construction of the model knowledge library can improve the value of information sharing and reuse [96]. The parallel implementation of existing EO data models is not easy [97]. Therefore, to solve this problem, we can consider adopting a way of encapsulating each model as a service for people to access in parallel. Deep integration with the cloud computing environment still faces some challenges. Another thorny issue is the implementation of the model within the DGGS framework. Since the spherical coordinate system is adopted, the data analysis algorithm based on the plane coordinate system may not be perfectly executed. At present, simple operations, such as distance measurement and area measurement, are used. The improvement of the traditional model or algorithm may be determined according to the actual situation because of the difference between the two reference coordinates, which has particular challenges.

6. Conclusions

This paper mainly focuses on big Earth observation data and proposes a trinity solution consisting of BEOD, cloud computing, and DGGS, which separately provide a data resource, computing power, and a unified framework. We review the current situations and achievements and discuss the opportunities and challenges that enable big Earth observation data via cloud computing and DGGS. From the perspective of data resources, Earth observation data has entered the era of big data, which not only faces the opportunities of abundant datasets, but also faces the challenges of data value mining. From the perspective of information technology, cloud computing technology has been able to provide a good solution for the mining of big data values. Due to its inherent defects, spatial thinking needs to be added. From the perspective of basic theory, the global discrete grid system provides a unified space–time framework. Although some research achievements have been made, the integration with data resources and information technology still faces some challenges. Based on this trinity theory, the organization and management of spatiotemporal data, as well as fusion analysis, can be realized through a unified framework, and the values of big Earth observation data (BEOD) can be fully discovered.
Future work will focus on the design, implementation, and global applications of the presented trinity solution. First, DGGS models for BEOD will be designed to cover remote sensing datasets with different resolutions. Second, cloud computing technology and advanced systems and platforms will be implemented to meet functional and performance requirements. Third, some global application cases will be carried out by the solution proposed in this paper.

Author Contributions: X.Y., G.L., and J.B. proposed the concept and designed the research frame; Q.C., L.Z. (Long, Zhao), Y.M., and L.Z. (Lianchong, Zhang) provided some materials; X.Y. wrote the original draft; J.B. and J.X. reviewed and edited the manuscript; D.Z. provided the project. All the authors contributed equally to the revisions of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Strategic Priority Research Program of Chinese Academy of Sciences (grant number XDA19020103).

Acknowledgments: We would like to thank the anonymous reviewers and editors for commenting on this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References
2. Guo, H. Big earth data: A new frontier in earth and information sciences. Big Earth Data 2017, 1, 4–20. [CrossRef]
12. Waldner, F.; Duveiller, G.; Dufourny, P. Local adjustments of image spatial resolution to optimize large-area mapping in the era of big data. Int. J. Appl. Earth Obs. 2018, 73, 374–385. [CrossRef]


24. Yao, X.; Li, G. Big spatial vector data management: A review. Big Earth Data 2018, 2, 108–129. [CrossRef]


34. Laney, D. 3-d data management: Controlling data volume, velocity, and variety. META Group Res. Note 2001, 6, 1.


64. Tao, M.; Wen, Y.; GuangXiong, P.; Su, L.; Zhangang, W. The study on error analysis of discretization area in discrete global grid system. *Procedia Environ. Sci.* 2011, 10, 1122–1128. [CrossRef]


76. Giuliani, G.; Chatenoux, B.; De Bono, A.; Rodila, D.; Richard, J.-P.; Allenbach, K.; Dao, H.; Peduzzi, P. Building an earth observations data cube: Lessons learned from the swiss data cube (sdc) on generating analysis ready data (ard). *Big Earth Data* **2017**, *1*, 100–117. [CrossRef]


92. Amiri, A.; Samavati, F.; Peterson, P. Categorization and conversions for indexing methods of discrete global grid systems. ISPRS Int. Geo.-Inf. 2015, 4, 320–336. [CrossRef]


