iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: VII Month of publication: July 2023

DOI: https://doi.org/10.22214/ijraset.2023.55108

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

An In-Memory Data-Cube Aware Distributed Data Discovery Across Clouds for Remote Sensing Big Data

Chetan Jane[1], Vinay S Navale[2]
B.E in Computer Science & Engineering, Dept of CSE, Bellari Institute of Technology and Management

Abstract: The recent proliferation of high-resolution Earth observation data, along with the advent of open-data initiatives, has led to the availability of petabyte-scale Earth observation datasets accessible for free. However, this abundance of big data presents significant challenges for regional to global spatio-temporal analysis. The traditional approach of "download-preprocess-store-analyze" introduces excessive data downloading overhead and computational barriers, hindering efficient analysis. The Earth observation data cube (EODC) paradigm offers a solution by revolutionizing the storage and management of spatio-temporal RS data. Yet, the presence of multiple EODC solutions has resulted in "information silos," making the sharing and joint use of RS data across EODCs difficult.

To tackle these challenges, we propose a novel method of in-memory distributed data cube autodiscovery and retrieval across clouds. Our approach involves constructing a distributed in-memory data orchestration system that shields the heterogeneity of EODC storage solutions, effectively overcoming the "information silos" problem. Additionally, we introduce a "larger-sites-first" and spatio-temporal aware RS data discovery strategy that automatically identifies and retrieves data across clouds based on specific requirements. Leveraging the data cube paradigm, our article presents a "quality-first" data filtering strategy, enabling users to extract high-quality data relevant to their target spatio-temporal range from the vast amount of available data. This method ensures efficient data cube joint retrieval and utilization across cloud platforms.Through comparative experiments, we have demonstrated the effectiveness and efficiency of our proposed approach. By addressing the complexities of RS data management and promoting seamless data sharing across cloud-based EODCs, our method opens up new possibilities for collaborative spatio-temporal analysis, empowering researchers and organizations to make informed decisions and gain valuable insights from the vast Earth observation data resources.

Keywords: Clouds, data cube, data discovery, data integration, distributed computing, GEE, remote sensing (RS) big data, in-memory distributed file system.

I. INTRODUCTION

The advancement of high-resolution Earth observation technology and open-data initiatives has ushered in an era of unprecedented access to petabyte-scale Earth observation data. This wealth of data offers immense potential for regional to global spatio-temporal analysis, enabling researchers to gain valuable insights into our planet's dynamic environment. However, with such an enormous influx of big data, the challenges of managing, processing, and analyzing it have become apparent. Cumbersome cycles of "download-preprocess-store-analyze" have led to excessive data downloading overhead and significant computational barriers. Moreover, the traditional acquisition-oriented 2-D file-based structure is ill-suited for spatio-temporal analysis, making efficient utilization of remote sensing (RS) data an arduous task. In response to these challenges, the Earth observation data cube (EODC) paradigm has emerged, revolutionizing the storage, management, and analysis of spatio-temporal RS data to some extent. However, the proliferation of different EODC solutions has resulted in isolated "information silos," hindering the sharing and joint use of RS data across EODCs.

To overcome these obstacles, this article proposes an innovative method of in-memory distributed data cube autodiscovery and retrieval across clouds. By constructing a distributed in-memory data orchestration system, we aim to overcome the heterogeneity of EODC storage solutions and break down the barriers posed by "information silos." Additionally, we introduce a novel "larger-sites-first" and spatio-temporal aware RS data discovery strategy that automatically identifies and retrieves data across clouds based on specific user requirements. Leveraging the data cube paradigm, we present a "quality-first" data filtering strategy to efficiently extract high-quality data relevant to the target spatio-temporal range from the vast data repositories. Through comparative experiments, we demonstrate the effectiveness and efficiency of our proposed method, paving the way for seamless collaborative spatio-temporal analysis and empowering researchers to harness the vast Earth observation data for informed decision-making and valuable insights into our planet's ecosystems.

The proposed method represents a significant step towards tackling the challenges associated with big Earth observation data, facilitating regional to global spatio-temporal analysis with enhanced efficiency and ease of use. By leveraging in-memory data cube technology and distributed data orchestration across cloud platforms, researchers can overcome computational barriers and optimize data retrieval processes. The elimination of "information silos" fosters a collaborative environment, enabling the seamless sharing and joint utilization of remote sensing data across diverse EODC solutions. Moreover, the "larger-sites-first" and spatio-temporal aware RS data discovery strategy automates the process of data identification, streamlining the selection of relevant datasets for analysis. The "quality-first" data filtering strategy further enhances the analysis process by enabling users to pinpoint high-quality data within the vast repositories, optimizing data cube joint retrieval and usage. Through the successful implementation of our method and demonstrated effectiveness in comparative experiments, we anticipate that this research will pave the way for groundbreaking advancements in spatio-temporal RS data analysis, driving informed decision-making and addressing pressing global challenges with greater precision and insight.

II. OBJECTIVES

1) *Efficient Data Management:* We strive to design an efficient and scalable data management system that can handle the exponential growth of RSBD. By leveraging distributed cloud resources, we intend to create a flexible infrastructure that accommodates the ever-expanding remote sensing datasets.
2) *In-Memory Data-Cube Implementation:* Our framework emphasizes the implementation of in-memory data-cube technology. By employing this approach, we aim to accelerate data processing and reduce the query response time, thereby empowering researchers to gain valuable insights rapidly.
3) *Distributed Data Discovery:* Facilitating distributed data discovery across multiple cloud platforms will enable seamless collaboration among researchers and organizations. This collaborative environment will foster knowledge exchange and encourage innovative approaches to remote sensing data analysis.
4) *Scalability and Performance:* The proposed framework seeks to provide a highly scalable solution that can adapt to changing data requirements and increased user demands. Moreover, we will extensively evaluate the performance of our system to ensure it meets the rigorous demands of processing RSBD.

III. LIMITATIONS

In-Memory Data-Cube Aware Distributed Data Discovery Across Clouds for Remote Sensing Big Data," while offering innovative solutions for handling Remote Sensing Big Data (RSBD), has several limitations. Firstly, implementing an in-memory data-cube and distributed data discovery across clouds can be resource-intensive, leading to higher operational costs. Secondly, data ingestion overhead and the need for preprocessing RSBD may affect real-time data discovery capabilities. Thirdly, managing distributed data introduces challenges in maintaining data consistency and synchronization, leading to potential data integrity issues. Moreover, network latency may impact real-time data discovery, and security and privacy concerns arise due to the distributed nature of the framework. Additionally, compliance with data governance policies and the complexity of handling data heterogeneity from diverse sources may affect data quality and analysis. Despite its potential, addressing these limitations is crucial for successful implementation and effective utilization of the framework for RSBD management and analysis.

IV. LITERATURE SURVEY

In Giuliani, Camara, Killough, and Minchin (2019) [1] delves into the domain of Earth observation open science and its potential for enhancing reproducible science through the use of data cubes. The paper explores the concept of data cubes in the context of Earth observation data, highlighting their significance in enabling efficient and reproducible spatio-temporal analysis. By employing data cubes, researchers can store, manage, and analyze large-scale Earth observation datasets with greater ease and flexibility. The survey identifies the role of data cubes in overcoming the challenges posed by big data deluge, tedious data processing cycles, and the need for increased computational capabilities. Additionally, the study addresses the importance of open science principles, promoting data sharing and collaboration among the scientific community. By presenting a comprehensive analysis of the existing literature, the paper emphasizes the potential of data cubes to revolutionize Earth observation research, fostering greater reproducibility and advancing scientific knowledge in this critical field.
In [2] provides a comprehensive overview of platforms designed for the management and analysis of big Earth observation data. The paper explores various existing platforms and technologies that cater to the challenges posed by the vast amount of Earth observation data available today. Through an extensive analysis, the survey highlights the functionalities and capabilities of these platforms in terms of data management, processing, and analysis.

It discusses the significance of efficient data handling and the integration of advanced analytics tools to extract valuable insights from the massive datasets. Furthermore, the study assesses the performance and scalability of the platforms in dealing with the growing volume of Earth observation data. By presenting a comparative analysis of different platforms, the paper serves as a valuable resource for researchers and practitioners seeking to navigate the landscape of big Earth observation data management and analysis, ultimately fostering advancements in the field of remote sensing and geospatial research.

In [3] Hargreaves and Watmough (2021) explore the application of satellite Earth observation for supporting sustainable rural development. The paper focuses on the utilization of satellite-based technologies and geospatial data to address the challenges and opportunities in rural areas. By examining a range of case studies and research findings, the survey highlights the potential of Earth observation data in monitoring and managing various aspects of rural development, such as land use, agricultural practices, natural resource management, and environmental conservation. It emphasizes the importance of integrating remote sensing techniques into sustainable rural development strategies, enabling evidence-based decision-making and promoting informed policies for socio-economic and environmental well-being. The survey contributes to the growing body of knowledge on the significant role of satellite Earth observation in fostering sustainable development in rural regions, offering valuable insights to policymakers, researchers, and stakeholders in the field of geoinformatics and applied Earth observation.

[4] Investigates the application of Earth system data cubes in unraveling global multivariate dynamics. The paper delves into the concept of Earth system data cubes, which offer a multidimensional approach to analyzing complex Earth system processes. By integrating diverse Earth observation datasets, the survey explores how data cubes facilitate the representation, management, and analysis of spatio-temporal variables at a global scale. Through a comprehensive review of research findings and case studies, the study highlights the significance of data cubes in understanding the intricate interactions between various Earth system components, such as climate, vegetation, and land use. The survey contributes to the advancement of Earth system science, offering valuable insights into the potential of data cubes to support informed decision-making, environmental monitoring, and sustainable management of global dynamics.

The challenges posed [5] by a Big Data Earth. The paper explores the implications of the vast and diverse Earth observation datasets available today and their impact on various scientific domains. By examining the complexities associated with handling and analyzing big Earth data, the survey sheds light on the computational, storage, and processing barriers faced by researchers and organizations. Moreover, it delves into the need for advanced data management and analysis techniques to harness the full potential of these datasets. The study emphasizes the importance of developing scalable and efficient solutions to cope with the ever-increasing volume of Earth observation data. By providing insights into the challenges and potential solutions, this survey contributes to the ongoing efforts to navigate the era of big Earth data and its transformative impact on Earth science and related disciplines.

In [6] focuses on the utilization of Big Data for remote sensing, particularly in the domains of visualization, analysis, and interpretation. The book explores the transformative impact of Big Data on remote sensing technologies, providing valuable insights into the integration of advanced visualization techniques for processing and presenting large-scale Earth observation datasets. Through a comprehensive analysis, the survey delves into the challenges and opportunities of analyzing and interpreting Big Data from remote sensing platforms, emphasizing the significance of data-driven decision-making in various applications such as environmental monitoring, disaster management, and urban planning. By presenting case studies and research findings, the survey contributes to the understanding of how Big Data is reshaping the landscape of remote sensing, opening up new avenues for innovative research and applications in this critical field.

In [7] investigate user needs and requirements for future Landsat missions. The paper delves into the perspectives and expectations of stakeholders and end-users in the field of remote sensing, focusing on the critical aspects that should be considered in designing and planning future Landsat missions. Through a comprehensive analysis of user feedback and case studies, the survey highlights the evolving demands for Earth observation data and the role of Landsat missions in addressing these needs. It examines the application domains where Landsat data plays a vital role, such as agriculture, forestry, land cover mapping, and environmental monitoring. By identifying the gaps and opportunities for improvement, the study offers valuable insights to satellite mission planners and policymakers, ensuring that future Landsat missions are tailored to meet the evolving user requirements and contribute effectively to global environmental monitoring and sustainable development efforts.

Sudmanns et al. [8] investigate the paradigm of Big Earth data and its potential to disrupt Earth observation data management and analysis. The paper explores the transformative changes brought about by the exponential growth of Earth observation data and its implications for data management and analysis techniques. Through a comprehensive analysis, the survey sheds light on the challenges and opportunities presented by the vast and diverse datasets, focusing on the need for scalable and efficient solutions to

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

handle big Earth data. It examines the role of advanced technologies, such as data cubes and cloud computing, in enabling effective data storage, processing, and analysis. By presenting case studies and research findings, the study contributes to the understanding of the impact of Big Earth data on Earth sciences, remote sensing, and related disciplines, ultimately fostering innovative approaches to harnessing the potential of Earth observation data for informed decision-making and sustainable development efforts.

In [9] examines the concept of Satellite Analysis Ready Data (ARD) and its significance in supporting the United Nations' Sustainable Development Goals (SDGs). The paper explores how ARD plays a crucial role in facilitating the effective monitoring and implementation of SDGs by providing readily accessible and pre-processed satellite data. Through an in-depth analysis, the survey highlights the potential of ARD in addressing the data challenges posed by the SDGs, including data availability, processing complexity, and interoperability. It emphasizes the importance of ARD in enabling evidence-based decision-making, policy formulation, and progress tracking towards achieving sustainable development objectives. By presenting real-world applications and case studies, the study contributes to the understanding of how ARD can act as a valuable tool for policymakers, researchers, and stakeholders in the realm of Earth observation applications and global policy frameworks, ultimately driving positive changes in the pursuit of sustainable development worldwide.Moser et al. [10] focus on the significance of mountain lakes as crucial indicators of global environmental change. The paper delves into the role of these lakes as sensitive ecosystems, vulnerable to the impacts of climate change, human activities, and other environmental stressors. Through a comprehensive analysis, the survey highlights the importance of monitoring and studying mountain lakes as valuable sentinels of environmental shifts on a global scale. It explores the diverse methods and technologies employed to assess lake characteristics, water quality, and ecological changes. By presenting case studies and research findings, the study contributes to the understanding of how mountain lakes act as essential indicators for assessing the broader implications of environmental change and its potential consequences on freshwater resources, biodiversity, and ecosystem dynamics. The survey serves as a valuable resource for researchers and policymakers in the field of global environmental change, guiding efforts to protect and manage mountain lake ecosystems in the face of a changing climate and evolving human impacts.

V. DESIGN AND IMPLEMENT

A. Main Solution

This article introduces an innovative approach to address the growing demand for analyzing and processing large-scale multisource remote sensing (RS) big data. The proposed solution, as depicted in Fig. 1, comprises several key components. Firstly, a distributed in-memory data orchestration system is established across clouds, allowing virtual mounting of multisource RS datasets. This offers users a unified in-memory access view and operation interface, promoting seamless sharing and joint utilization of RS data across clouds. Secondly, a distributed data discovery architecture, leveraging STAC standards, facilitates data cube discovery and indexing across clouds. The article proposes the LSA data discovery strategy, which efficiently identifies target RS data from multiple cloud platforms, enhancing the spatio-temporal aware data discovery process. Lastly, data cube retrieval across clouds is implemented using QF data filtering, enabling the selection of high-quality data from various cloud platforms that cover a specific spatio-temporal range. Additionally, a data cube retrieval interface across clouds is provided to support large-scale RS analysis and applications.

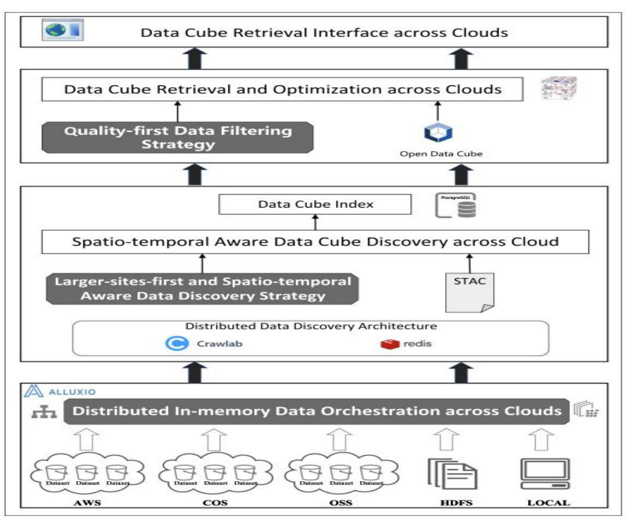


Fig. 1. Main solution.

B. In-Memory Data Orchestration Across Clouds With Alluxio

To address the issue of "information silo" among cloud platforms of different Earth Observation Data Cube (EODC) solutions, we implement a distributed in-memory data orchestration across clouds using Alluxio. Alluxio, previously known as Tachyon, is a memory-centric distributed virtual storage system that bridges the gap between data-driven applications and various storage systems, such as Amazon S3, Google Cloud Storage, HDFS, and more. In this article, we deploy Alluxio in highly available mode via a Zookeeper cluster to construct the in-memory data orchestration across clouds, as depicted in Fig. 2. By virtually mounting RS datasets from diverse clouds, including AWS S3, COS, OSS, HDFS, and LOCAL, we address the heterogeneity of underlying storage systems. This virtual data mounting mode provides users with a unified access view of all mounted datasets under the same namespace, organized as a directory tree (Fig. 3). This approach ensures data sovereignty for data providers while enabling users to access and analyze data on demand directly from different clouds. The cloud platforms or data providers maintain and update the original RS image data, and when users initiate data access requests, the master node locates the original data and instructs workers to read and cache the data from remote clouds to local in-memory (MEM) and hard disk drive (HDD) storage. This way, users can access and analyze data from various clouds without the need for local storage of the entire dataset, thus avoiding the high local storage capacity requirement for large-scale datasets.

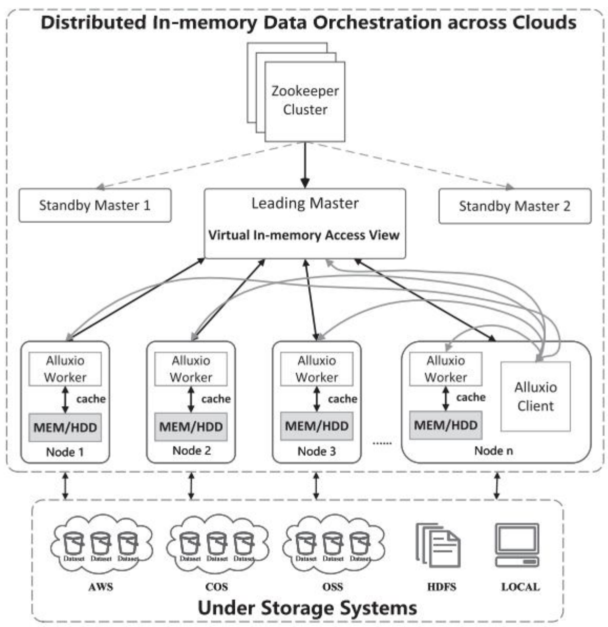


Fig. 2. Architecture of distributed in-memory data orchestration across clouds.

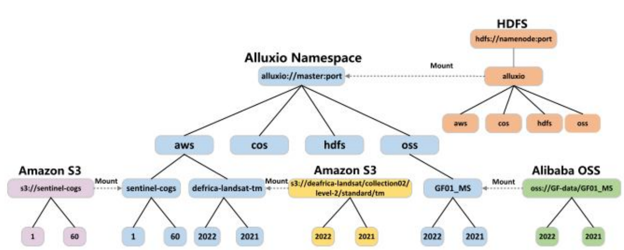


Fig. 3. Data in-memory access view across clouds

The data orchestration approach proposed in this article offers a powerful solution by virtually mounting RS data from multiple cloud platforms and local file systems, providing users with a unified and global virtual view of the data. Moreover, the system intelligently caches frequently accessed data, particularly from remote locations, which significantly improves memory-level data I/O throughput for subsequent data discovery, data cube retrieval, and data access across clouds.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

This effective data orchestration solution successfully tackles the challenges of sharing and accessing data across various Earth Observation Data Cube (EODC) solutions, enabling efficient and seamless access to multisource RS data across cloud platforms and data cube platforms. By facilitating streamlined data access and eliminating "information silos," researchers can benefit from enhanced collaboration, analysis, and decision-making in the field of remote sensing and geospatial research.

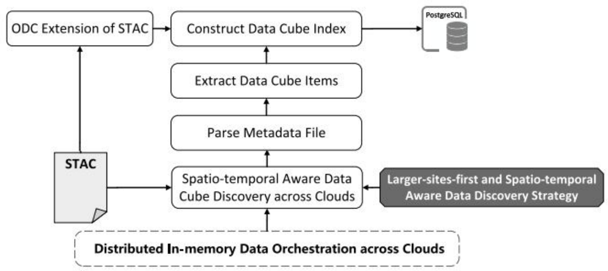
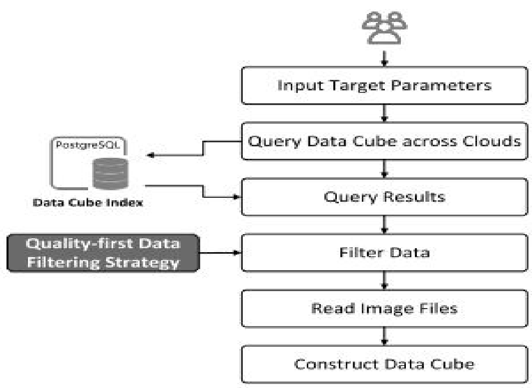


Fig. 4. Process of spatio-temporal aware data cube discovery across clouds.

C. Process of data cube Retrieval and Optimization Across Clouds



The process of data cube retrieval and optimization across clouds involves several key steps to efficiently access and analyze remote sensing big data. Here is an overview of the process:

1) *Data Cube Discovery:* The first step is to discover the relevant data cubes across different cloud platforms. Leveraging a distributed data discovery architecture and utilizing STAC standards, the system identifies data cubes that match specific user requirements, such as spatio-temporal extent and data variables.
2) *In-Memory Data Orchestration:* Once the data cubes are identified, the framework establishes a distributed in-memory data orchestration system using technologies like Alluxio. This system allows for the virtual mounting of the identified data cubes from various cloud platforms, providing users with a unified in-memory access view and operation interface.
3) *Data Cube Retrieval:* With the data cubes virtually mounted across clouds, users can initiate data cube retrieval requests directly from their analysis environment. The master node coordinates the data retrieval process by locating the original data from the remote clouds and instructing workers to read and cache the data into local in-memory and hard disk drive storage.
4) *Quality-First (QF) Data Filtering:* To optimize the data retrieval process, a QF data filtering strategy is employed. This strategy involves filtering out high-quality data that covers the target spatio-temporal range from the multiple cloud platforms, minimizing the amount of unnecessary data transfer and improving data retrieval efficiency.
5) *Transparent Caching:* The distributed in-memory data orchestration system transparently caches frequently accessed data, particularly from remote locations. This caching mechanism enables memory-level data I/O throughput for subsequent data discovery, data cube retrieval, and data access, further enhancing the performance of the framework.
6) *Network Speed and Stability Considerations:* The performance of data discovery and retrieval may be impacted by network speed and stability, especially when sourcing data from multiple public cloud storage platforms. Future work may involve addressing network-related challenges to ensure efficient data retrieval even in diverse cloud environments.

7) *Optimization Strategies:* The framework may employ various optimization strategies to improve the overall performance and efficiency of data cube retrieval across clouds. This may include load balancing mechanisms, adaptive caching policies, and auto-scaling strategies to handle varying workloads and cloud resource management effectively.

VI. CONCLUSION

The exponential growth of global archived Earth Observation (EO) data, amounting to exabytes, has prompted the innovative paradigm of Earth Observation Data Cube (EODC), revolutionizing traditional EO data acquisition, storage, processing, and sharing. However, the proliferation of diverse data cube solutions has resulted in limited interoperability among existing infrastructures, impeding data sharing and collaborative use. To address these challenges, we propose an in-memory distributed data cube autodiscovery and retrieval method for Remote Sensing (RS) big data across clouds. Our approach involves constructing a distributed in-memory data orchestration system, offering users a unified access view and operation interface across different cloud platforms. The proposed LSA data discovery strategy and QF data filtering strategy facilitate efficient data discovery and retrieval across clouds, overcoming data cube joint use challenges. However, network speed and stability may impact data discovery performance due to data sourced from multiple public cloud storage. To enhance performance, future work will consider spatio-temporal RS data characteristics and design access pattern-aware cache prefetching and replacement strategies. Overall, our method enables users to focus on data analysis and processing, eliminating tedious data acquisition tasks. By harnessing the full potential of spatio-temporal RS data, our approach facilitates large-scale scientific research on global environmental change and sustainable development, contributing to informed decision-making and advancements in the field of remote sensing and environmental sciences.

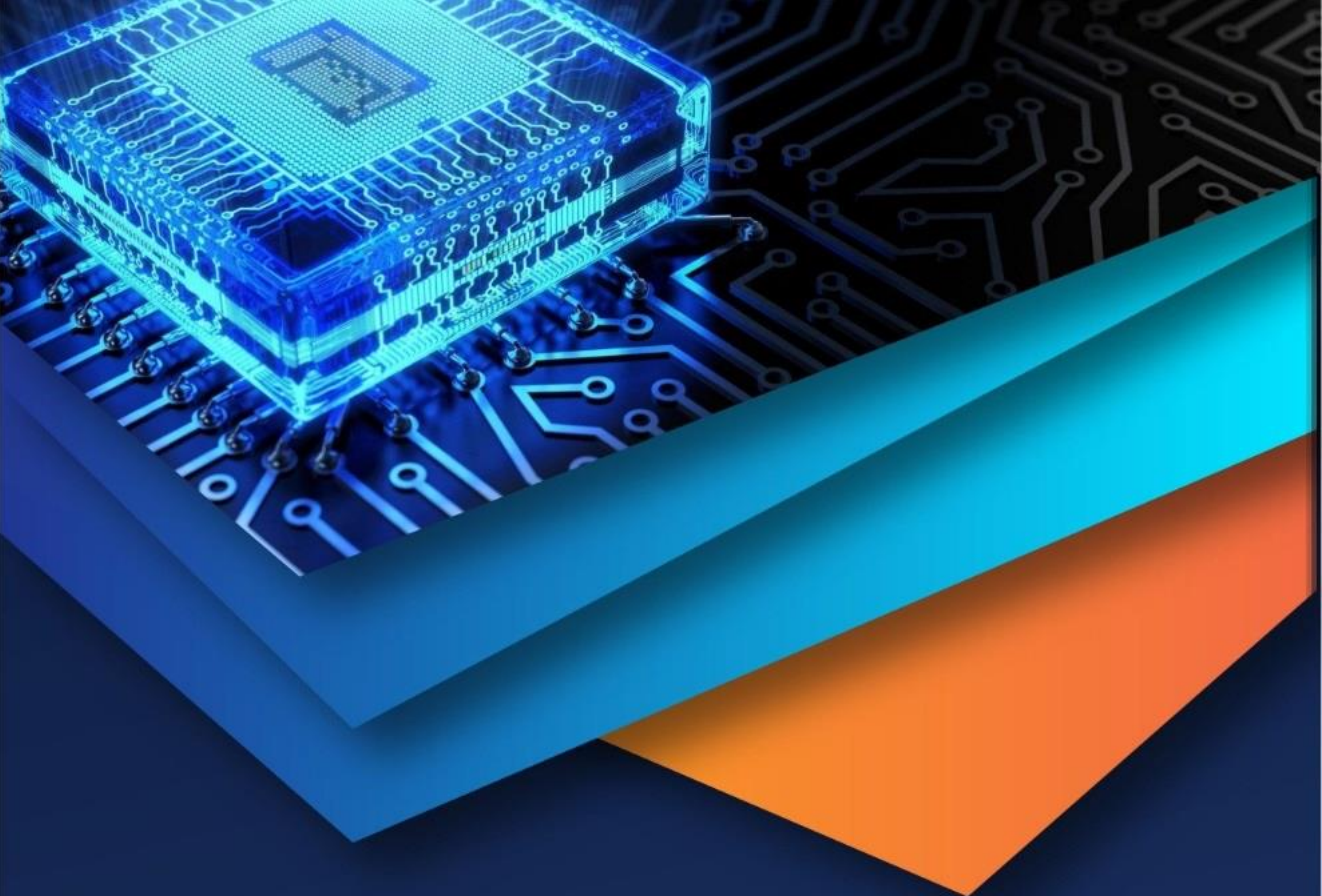
VII. FUTURE WORK

As part of future work, this innovative framework of "In-Memory Data-Cube Aware Distributed Data Discovery Across Clouds for Remote Sensing Big Data" can be further enhanced and expanded in several ways. Firstly, efforts can be focused on optimizing the performance and scalability of the distributed data orchestration system to handle even larger volumes of remote sensing big data efficiently. Secondly, exploring the integration of advanced data cube functionalities, such as on-the-fly data aggregation and real-time updates, can provide researchers with more powerful and flexible data analysis capabilities. Thirdly, achieving semantic interoperability by developing standardized data models and ontologies for EODC metadata can enable seamless collaboration and data sharing among different cloud platforms. Additionally, ensuring data privacy and security through robust encryption techniques and access controls is crucial for handling sensitive remote sensing data. Lastly, conducting real-world case studies and collaborations with research institutions can validate the framework's practicality and performance in operational remote sensing applications, providing valuable insights for further refinements and improvements.

REFERENCES

[1] G. Giuliani, G. Camara, B. Killough, and S. Minchin, “Earth observation open science: Enhancing reproducible science using data cubes,” vol. 4, no. 4, 2019, Art. no. 147.
[2] V. C. 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, 2020, Art. no. 1253.
[3] P. K. Hargreaves and G. R. Watmough, “Satellite Earth observation to support sustainable rural development,” Int. J. Appl. Earth Observ. Geoinf.,vol. 103, 2021, Art. no. 102466.
[4] M. D. Mahecha et al., “Earth system data cubes unravel global multivariate dynamics,” Earth Syst. Dyn. Discuss., vol. 11, no. 1, pp. 201–234, 2020.
[5] G. Boulton, “The challenges of a Big Data Earth,” Big Earth Data, vol. 2,no. 1, pp. 1–7, 2018.
[6] N. Dey, C. Bhatt, and A. S. Ashour, Big Data for Remote Sensing:Visualization, Analysis and Interpretation. Cham, Switzerland: Springer, 2018.
[7] Z. Wu et al., “User needs for future Landsat missions,” Remote Sens.Environ., vol. 231, 2019, Art. no. 111214.
[8] M. Sudmanns et al., “Big Earth data: Disruptive changes in Earth observation data management and analysis?,” Int. J. Digit. Earth, vol. 13, no. 7, pp. 832–850, 2020.
[9] B. D. Killough, “Satellite analysis ready data for the sustainable development goals,” Earth Observation Applications and Global Policy Frameworks. Hoboken, NJ, USA: Wiley, 2022, pp. 133–143.
[10] K. A. Moser et al., “Mountain lakes: Eyes on global environmental change,” Glob. Planet. Change, vol. 178, pp. 77–95, 2019.
[11] P. Liu, J. Li, L. Wang, and G. He, “Remote sensing data fusion with generative adversarial networks: State-of-the-art methods and future research directions,” IEEE Geosci. Remote Sens. Mag., vol. 10, no. 2, pp. 295–328, Jun. 2022.
[12] L. Mu, L. Wang, Y. Wang, X. Chen, and W. Han, “Urban land use and land cover change prediction via self-adaptive cellular based deep learning with multisourced data,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 12, no. 12, pp. 5233–5247, Dec. 2019.
[13] W. Han et al., “Sample generation based on a supervised Wasserstein generative adversarial network for high-resolution remote-sensing scene classification,” Inf. Sci., vol. 539, pp. 177–194, 2020.
[14] L. Zhang, P. Liu, L. Zhao, G. Wang, W. Zhang, and J. Liu, “Air quality predictions with a semi-supervised bidirectional LSTM neural network,” Atmospheric Pollut. Res., vol. 12, no. 1, pp. 328–339, 2021.
[15] Y. Wang, L. Wang, X. Chen, and D. Liang, “Offshore petroleum leaking source detection method from remote sensing data via deep reinforcement learning with knowledge transfer,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 15, pp. 5826–5840, Jul. 2022.

[16] G. Giuliani et al., “Live monitoring of Earth surface (limes): A framework for monitoring environmental changes from Earth observations,” Remote Sens. Environ., vol. 202, pp. 222–233, 2017.
[17] A. Shelestov, M. Lavreniuk, N. Kussul, A. Novikov, and S. Skakun, “Large scale crop classification using Google Earth engine platform,” in Proc. IEEE Int. Geosci. Remote Sens. Symp., 2017, pp. 3696–3699.
[18] N. Sánchez, Á. González-Zamora, J. Martínez-Fernández, M. Piles, and M. Pablos, “Integrated remote sensing approach to global agricultural drought monitoring,”Agricultural Forest Meteorol., vol. 259, pp. 141–153, 2018.
[19] P. Potapov et al., “Mapping global forest canopy height through integration of Gedi and Landsat data,” Remote Sens. Environ., vol. 253, 2021, Art. no. 112165.
[20] S. Nativi, P. Mazzetti, and M. Craglia, “A view-based model of data-cube to support Big Earth Data systems interoperability,” Big Earth Data, vol. 1, no. 1/2, pp. 75–99, 2017.
[21] Y. Zhang and J. Cheng, “Spatio-temporal analysis of urban heat island using multisource remote sensing data: A case study in Hangzhou, China,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 12, no. 9, pp. 3317–3326, Sep. 2019.
[22] F. Zellweger, P. De Frenne, J. Lenoir, D. Rocchini, and D. Coomes, “Advances in microclimate ecology arising from remote sensing,” Trends Ecol. Evol., vol. 34, no. 4, pp. 327–341, 2019.
[23] J. Li, Y. Pei, S. Zhao, R. Xiao, X. Sang, and C. Zhang, “A review of remote sensing for environmental monitoring in China,” Remote Sens., vol. 12, no. 7, 2020, Art. no. 1130.
[24] C. A. Lee, S. D. Gasster, A. Plaza, C.-I. Chang, and B. Huang, “Recent developments in high performance computing for remote sensing: A review,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 4, no. 3, pp. 508–527, Sep. 2011.
[25] Y. Ma, L. Wang, A. Y. Zomaya, D. Chen, and R. Ranjan, “Task-tree based large-scale Mosaicking for massive remote sensed imageries with dynamic DAG scheduling,” IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 8, pp. 2126–2137, Aug. 2014.
[26] A. Lewis et al., “Rapid, high-resolution detection of environmental change over continental scales from satellite data–the Earth observation data cube,” Int. J. Digit. Earth, vol. 9, no. 1, pp. 106–111, 2016.
[27] S. Kopp, P. Becker, A. Doshi, D. J. Wright, K. Zhang, and H. Xu, “Achieving the full vision of Earth observation data cubes,” Data, vol. 4, no. 3, 2019, Art. no. 94.
[28] G. Giuliani, B. Chatenoux, T. Piller, F. Moser, and P. Lacroix, “Data cube on demand (DCOD): Generating an Earth observation data cube anywhere in the world,” Int. J. Appl. Earth Observ. Geoinf., vol. 87, 2020, Art. no. 102035.
[29] P. Baumann, “The datacube manifesto,” 2018. Accessed: Jan. 1, 2023. [Online]. Available: http://www.earthserver.eu/tech/datacube-manifesto
[30] P. Baumann, D. Misev, V. Merticariu, and B. P. Huu, “Datacubes: Towards space/time analysis-ready data,” in Service-Oriented Mapping. Berlin, Germany: Springer, 2019, pp. 269–299.
[31] G. Giuliani et al., “Building an Earth observations data cube: Lessons learned from the Swiss data cube (SDC) on generating analysis ready data (ARD),” Big Earth Data, vol. 1, no. 1/2, pp. 100–117, 2017.
[32] A. Lewis et al., “The Australian geoscience data cube–foundations and lessons learned,” Remote Sens. Environ., vol. 202, pp. 276–292, 2017.
[33] P. Merodio Gómez, A. Ramírez Santiago, O. J. Juárez Carrillo, and F. J. Jiménez Nava, “The potential contribution of Earth observation data cubes for the production of information for sustainable development in emerging countries,” Geomatics Environ. Eng., vol. 16, no. 3, pp. 131–155, 2022.
[34] C. Ariza-Porras et al., “CDCOL: A geoscience data cube that meets colombian needs,” in Proc. Colombian Conf. Comput., 2017, pp. 87–99.
[35] P. Baumann et al., “Big data analytics for earth sciences: The Earthserver approach,” Int. J. Digit. Earth, vol. 9, no. 1, pp. 3–29, 2016.
[36] G. Camara et al., “The e-sensing architecture for big Earth observation data analysis,” in Proc. Conf. Big Data From Space, 2017, pp. 28–30.
[37] M. Stonebraker, P. Brown, D. Zhang, and J. Becla, “SCIDB: A database management system for applications with complex analytics,” Comput. Sci. Eng., vol. 15, no. 3, pp. 54–62, 2013.
[38] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, “Google Earth engine: Planetary-scale geospatial analysis for everyone,” Remote Sens. Environ., vol. 202, pp. 18–27, 2017.
[39] Amazon, “EAWS-Earth on Amazon web services,” 2023. Accessed: Jan. 10, 2023. [Online]. Available: https://aws.amazon.com/cn/earth
[40] P. Baumann et al., “Fostering cross-disciplinary Earth science through datacube analytics,” in Earth Observation Open Science and Innovation. Cham, Switzerland: Springer, 2018, pp. 91–119.
[41] G. Giuliani, J. Masó, P. Mazzetti, S. Nativi, and A. Zabala, “Paving the way to increased interoperability of Earth observations data cubes,” Data, vol. 4, no. 3, 2019, Art. no. 113.
[42] L. Wang, Y. Ma, A. Y. Zomaya, R. Ranjan, and D. Chen, “A parallel file system with application-aware data layout policies for massive remote sensing image processing in digital Earth,” IEEE Trans. Parallel Distrib. Syst., vol. 26, no. 6, pp. 1497–1508, Jun. 2015
[43] L. Wang, Y. Ma, J. Yan, V. Chang, and A. Y. Zomaya, “pipsCloud: High performance cloud computing for remote sensing Big Data management and processing,” Future Gener. Comput. Syst., vol. 78, pp. 353–368, 2018.
[44] Z. Sun, F. Chen, M. Chi, and Y. Zhu, “A spark-based Big Data platform for massive remote sensing data processing,” in Proc. Int. Conf. Data Sci., 2015, pp. 120–126.
[45] L. Li, W. Jing, and N. Wang, “An improved distributed storage model of remote sensing images based on the HDFs and pyramid structure,” Int. J. Comput. Appl. Technol., vol. 59, no. 2, pp. 142–151, 2019.
[46] R. D. Price, M. D. King, J. T. Dalton, K. S. Pedelty, P. E. Ardanuy, and M. K. Hobish, “Earth science data for all: EoS and the EoS data and information system,” Photogrammetric Eng. Remote Sens., vol. 60, no. 3, pp. 277–285, 1994.
[47] ESA, “Copernicus open access hub,” 2023. Accessed: Jan. 16, 2023. [Online]. Available: https://scihub.copernicus.eu/
[48] FENGYUN Satellite Data Center, “FY satellite remote sensing data service web portal,” 2023. Accessed: Jan. 16, 2023. [Online]. Available: http://satellite.nsmc.org.cn/portalsite
[49] Y. Shao, L. Di, Y. Bai, H. Wang, and C. Yang, “Federated catalogue for discovering Earth observation data,” Photogrammetrie-FernerkundungGeoinf., vol. 2013, no. 1, pp. 43–52, 2013.
[50] M. Appel and E. Pebesma, “On-demand processing of data cubes from satellite image collections with the gdalcubes library,” Data, vol. 4, no. 3, 2019, Art. no. 92



10.22214/IJRASET

INDEX COPERNICUS

45.98

ISRA JIF

IMPACT FACTOR: 7.129

SJIFactor

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY