AFRICA UNIVERSITY
(A United Methodist Related Institution)

THE ESSENCE OF HIGH-PERFORMANCE COMPUTING IN ZIMBABWE'S NATIONAL SPATIAL DATA INFRASTRUCTURE ANALYSIS

BY

MUNYARADZI MWARIANESU MUCHAPONDWA

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF SCIENCE HONOURS IN COMPUTER SCIENCE IN THE COLLEGE OF BUSINESS, PEACE, LEADERSHIP AND GOVERNANCE

2023

Abstract

The study looked at the essence of high-performance computing in national spatial data analysis, towards inclusive national development in Zimbabwe. This qualitative research explored the perceptions of the increasing availability of national spatial data, looking at the challenges being faced in the handling and analysis of this big data, hence the growing need for more sophisticated tools and techniques to analyse this data efficiently and accurately. Spatial data has become ubiquitous, for example, Global Positioning System data, with increasingly sheer sizes in recent years. The applications of big spatial data span a wide spectrum of interests including tracking infectious disease, climate change simulation, and natural disaster detection, among others. High-performance computing (HPC) has emerged as a critical tool in addressing this need of efficiently collecting, managing, storing, and analysing geospatial data streams as well as handling the "Vs" (volume, variety, velocity, veracity, and value) of big data, since the massive datasets must be analysed in the context of space and time. This dissertation, therefore, explores the role of High-Performance Computing in national spatial data analysis, examining its benefits and challenges for this purpose. Drawing on a review of the literature, case studies, questionnaires, and semi-structured interviews with experts in the field, the dissertation presents an overview of High-Performance Computing architecture and software, as well as a framework for evaluating its efficacy in national spatial data analysis. The main findings from the research suggest that HPC can significantly improve the speed and accuracy of national spatial data processing, enabling users to better analyse large and complex datasets. However, there are some challenges such as cost, data management, and the need for specialized skills and expertise that were taken note of and needed to be addressed. The inquiries about these points appearing, concerns and challenges of the appropriation of High-Performance Computing innovation in spatial data analysis, were taken off from a case study about Zimbabwe National Geospatial and Space Agency (ZINGSA) in utilizing the graphic explanatory strategy. Eighteen surveys were recalled out of the twenty surveys that were conveyed and analysed. The investigate suggested that ZINGSA can embrace High-Performance Computing innovation in its operations, on the off chance that it is curiously on the side of ICT human assets through training, scientific missions, and advancements. In expansion, curiously on the side of security through putting the non-critical application and information within the cloud, or through making hybrid HPC which comprises of cloud-based HPC for non-basic applications like email and physical HPC data centre for basic and touchy applications and information. Without a question, the best administration has a crucial part in the appropriation of this innovation in its operations even through its decisions and facilities. Overall, this dissertation highlights the importance of High-Performance Computing in Zimbabwe's national spatial data analysis and offers practical recommendations for its effective use.

Keywords: Spatial Data, Big Data, High-Performance Computing, Zimbabwe National Geospatial and Space Agency (ZINGSA), Zimbabwe

Declaration Page

I declare that this dissertation is my original work except where sources have been cited and acknowledged. The work has never been submitted, nor will it ever be submitted to another university for the award of a degree.

______________________________	______________________________
Student's Full Name	Student's Signature (Date)
______________________________	______________________________
Main Supervisor Full Name	Main Supervisor's Signature (Date)

Copyright

Acknowledgements

My greatest appreciation goes to my supervisor Mrs L. Tembani - Fundisi, from the College of Business, Peace, Leadership, Peace and Governance for her great effort in making sure that my research work was a success. She equipped me with the knowledge and skills I needed to make it in such a competitive environment. Without her distinguished relentless direction and guidance, this research could not have been possible.

In addition to this, I am particularly indebted to the Zimbabwe National Geospatial and Space Agency family for taking me in and giving me more than I expected. Their patience, and cooperation in my learning process have been extraordinary. I want to extend my gratitude to my friends and classmates who stood and believed in me from the first day, giving me all the encouragement towards the attainment of this professional qualification despite the challenges imposed on us like the Covid 19 during the period of study.

Above all, I'm grateful to the Lord Almighty for blessing me with the opportunity and intellect to grow in the field of Information Technology. It is only through Him that the l has come this far in the academic journey, for everything begins and ends with Him.

Dedication

My life has been a collage of random luck, accidents, consistent self-doubt, mistakes, opportunities, chance meetings and the odd bout of fortitude and hard work. To all the queens who stood with me and took place of my late mother, I love you all, and to all the father figures who stood in for my late father, I salute you all. To the Muchapondwa clan, thank you for raising a warrior.

This dissertation is dedicated to the poor child growing up in the lowly and dusty streets of any ghetto, without hope, without opportunity, without love, you too matter, you too are capable and you can change the world.

List of Acronyms and Abbreviations

AU – Africa University

CPU – Central Processing Unit

DGSS – Discrete Global Grid System

ERP – Enterprise Resource Planning

FGPA – Field Programmable Gate Array

FLOPS – Floating Point Operations Per Second

GB – Gigabytes

GIS – Geographic Information System

GNSS – Global Navigation Satellite Systems

GPS – Global Positioning System

GPU – Graphics Processing Unit

HPC – High-Performance Computing

ICT – Information and Communication Technology

IT – Information Technology

LAN – Local Area Network

MIPS – Million Instructions Per Second

MPI – Message Passing Interface

PC – Personal Computer

TB – Terabyte

UAV – Unmanned Aerial Vehicle

USGS – United States Geological Survey

ZCHPC – Zimbabwe Centre for High-Performance Computing

ZINGSA – Zimbabwe National Geospatial and Space Agency

Table of Contents

CHAPTER 1: INTRODUCTION

1.1. Introduction

As it has been the norm of the day that the Zimbabwean Government is striving to centre its national development on inclusivity such that all sectors are equally involved in the developmental process. Issues around geospatial and earth observation intelligence have become a discourse that entails that we can’t efficiently implement spatial and space development as in Zimbabwe. Withholding such, national development has been thriving to encompass Global Navigation Satellite Systems, aerial mapping, Real Time Kinetic (RTK) surveys, topographic map creation, ground control point establishment, cadastre surveying, and examinations.

It is equally important to note that satellites, ubiquity sensors, and other devices will be installed and they will produce larger quantities and more diverse types of geospatial data. Over time, with the increasing amount of spatial data being collected, the use of that spatial data will be more likely to often be limited by accessibility, human capacity, and resources. The use of spatial data in particular will be limited by, among other factors, software needed to work with geospatial data, and the computational resources often required to prepare large and complex datasets for analysis. Overcoming these technological barriers is a significant focus of the geo-intelligence organizations including the Zimbabwe Government and its partners, who monitor and evaluate the progress of national development efforts through The Ministry of Higher and Tertiary Education, Innovation Science and Development.

Therefore, this study is sought to look at the essence of using high-performance computers for National Spatial Data Analysis as compared to traditional methods and the use of ordinary computers and the relative benefits to the surveyors and the related stakeholders as much as data capturing, manipulation, distribution and analysis of big spatial can be made easier. While using computers in these spatial and geographic information systems applications, new technologies, such as multicore high-performance processing computers, as well as reliable and robust broadband connectivity, will help bring up a long-term solution that makes it practically possible to implement geographical information systems, geo-engineering, earth observation, and monitoring, weather forecasting, space science, space exploration and satellite communications in line with the national development strategy.

1.2. Background of the Study

Spatial data analysis was first attempted dating back to the 1960s in Canada using traditional descriptive or inferential statistics for numerical comparison or by displaying the outputs with the spatial features associated with the spatial data sets. The earliest main objective for the subject of GIS applications was to systematize national natural resources but, the analysis was however unable to provide insight into the spatial attributes and they could not interpret the spatial relationships among the given spatial data sets.

It is difficult to imagine the world, a country, or a super industry that does not benefit from geo-referencing directly or indirectly. Spatial data analysis allows for determining if the selected area is well suited for a particular business activity, detecting changes, identifying tendencies, assessing risks, outlining possible outcomes, and preventing losses. This, therefore led to the proliferation in amounts of generated data as it has propelled the rise of high-performance computing solutions to efficiently analyse and extract useful insights from such data. Meanwhile, spatial data has become ubiquitous, GPS data, with increasingly sheer sizes in recent years. The applications of big spatial data span a wide spectrum of interests including tracking infectious disease, climate change simulation, and drug addiction, among others.

This study sought to comprehensively review the essence of work in the intersection of high-performance computing and big spatial data. The advancement of technology will radically benefit this unique and advanced research, innovation, and advisory sector through the use of different software applications as there had been an unprecedented explosion in the amounts of generated and collected data, for example, space telescopes can generate roughly up to 150 GB weekly spatial data, and satellite earth images archive has more than 500 TB. To efficiently process such tremendous amounts of spatial data, researchers and developers need to take a stride toward the use of HPCs. Up-to-date tools and spatial analysis techniques will undoubtedly provide more accurate analytics that will assist design, promoting, coordinating, conduct research and development initiatives that promote efficient advances in Geospatial Sciences and Earth Observations, Space Engineering, Space Science, Aeronautical Engineering, Mechatronics, Satellite Communication Systems, Global Navigation Satellite Systems (GNSS), Land Positioning Systems, Unmanned Aerial Vehicles (UAV) and Launch of Satellites to efficiently harness the development and application of Geospatial and Space Technologies for the advancement of Zimbabwe’s modernization and industrialization agenda, there is a strong need for the adaption to the use of high-performance computers (HPC), contributing more to Zimbabwe’s Government's national development strategy.

1.3. Statement of the Problem

While there is no consensus on the definition of big data (Ward and Barker, 2013; De Mauro et al., 2015), the widely used definition is: “datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse” (Manyika et al., 2011, p.1). Geospatial big data refers to a specific type of big data that contains location information, and most data today are inherently spatial, collected with ubiquitous location-aware sensors such as satellites, GPS, and environmental observations. Geospatial big data offers great opportunities for advancing scientific discoveries across a broad range of fields, including climate science, disaster management, public health, precision agriculture, and smart cities. However, what matters the most is the ability to efficiently and promptly extract meaningful information from spatial data, an aspect reflected in the widely used big data definition provided above. Efficiently extracting such meaningful information and patterns is challenging due to big data’s 5 -Vs characteristics which are volume, velocity, variety, veracity, and value (Zikopoulos and Eaton, 2011; Zikopoulos et al., 2012; Gudivada et al., 2015), and geospatial data’s intrinsic feature of space and time.

Volume refers to the large amounts of data being generated. Velocity indicates the high speed of data streams and the accumulation that will gradually occur over time. Variety refers to the high heterogeneity of data, such as different data sources, formats, and types. Veracity refers to the uncertainty and poor quality of data, including low accuracy, bias, and misinformation. For geospatial big data, these four Vs must be handled in the context of dynamic space and time to extract the ‘value’ from big data, which creates further challenges.

High-performance computing (HPC) provides an essential solution to geospatial big data challenges by allowing fast processing of massive data collections in parallel. Handling geospatial big data with HPC can help us make quick and better decisions in time-sensitive situations, such as emergency response (Bhangale et al., 2016). It also helps us to solve larger problems, such as high-resolution global forest cover change mapping in reasonable timeframes

(Hansen et al., 2013) and to achieve interactive analysis and visualization of big data (Yin et al., 2017).

Therefore, this study seeks to identify the benefits and challenges of adopting High-Performance Computing and services technology in Zimbabwe's national spatial data infrastructure analysis.

1.4. Research Objectives

The study has the following specific objectives:

1. To determine the extent to which the Zimbabwe National Geospatial and Space Agency is aware of high-performance computing.
2. To determine the extent to which Spatial Data Analysis and High-performance computing synergize.
3. To establish the importance of high-performance computing in national spatial data analysis.
4. To identify the main factors that influence ZINGSA to adopt high-performance computing.
5. To assess the promising future of high-performance computing in spatial data analysis.

1.5. Research Questions

1. To what extent is ZINGSA aware of high-performance computing?
2. To what extent do Spatial Data Analysis and High-performance computing synergize and relate?
3. What is the importance of high-performance computing in national spatial data analysis?
4. What are the main factors that are more likely to influence ZINGSA to adopt high-performance computing?
5. Where can high-performance computing take national spatial data analysis in the future?

1.6. Assumptions/ Hypothesis

1. Zimbabwe National Geospatial and Space Agency will adopt High-Performance Computing and services to increase efficiency.

2. Stakeholders and respondents within the organization would be able to comply and hence provide honest data which will assist in the results of the study.
3. High-Performance Computing has a positive impact when it comes to reducing big data processing time.
4. The Respondents are acquainted with the knowledge of High-Performance Computing.

1.7. Significance of the study

This dissertation attempts to fill this gap in the nexus between the concerns of national spatial data analysis and high-performance computing in Zimbabwe. This is an attempt to add perspectives on the essence of high-performance computing. In exploring these issues, the research will seek to lay the groundwork for the understanding of big data handling as much as storage and management are concerned, spatial indexing, tasks scheduling, query processing, and handling, and workflow-based systems. This research may also help to make the geo-intelligence professional practice more accountable to spatial data consumers and users.

1.8. Delimitation of the Study

While traditional coordinate systems (such as the system based on latitude and longitude) have been successful as a frame of reference, a relatively new framework called the discrete global grid system (DGGS) is believed to work better in managing and processing the heterogeneous geospatial big data associated with the curved surface of the Earth (Sabeur et al., 2019). DGGS represents “the Earth as hierarchical sequences of equal area tessellations on the surface of the Earth, each with global coverage and with progressively finer spatial resolution” (OGC, 2017). It aims to provide a unified, globally consistent reference framework to integrate heterogeneous spatial data—such as raster, vector, and point cloud—with different spatiotemporal scales and resolutions. The design of DGGS makes it natively suitable for parallel processing with HPC, as the data that it stores and manages has already been decomposed into discrete subdomains. However, this research will remain based on traditional reference frameworks. Future research is needed to investigate spatiotemporal indexes, parallel algorithms, and big data computing platforms in the context of DGGS and HPC.

CHAPTER 2: REVIEW OF RELATED LITERATURE

2.1. Introduction

The introduction of High-Performance Computing in recent years has piqued the interest of a wide range of people. Organizations, institutions, and end users are all examples of beneficiary users. This is due to the adoption of a new economic model for High Performing computing that promises to transform the Information Communication Technology (ICT) department. The fashion model promises a shift away from an organization that is forced to invest heavily in limited ICT resources.

According to Zikopoulos and Eaton, 2011; Zikopoulos et al., 2012; Gudivada et al., 2015., efficient extraction of meaningful information and patterns is challenging due to big data's 5-V characteristics—volume, velocity, variety, veracity, value, and geospatial data's intrinsic feature of space and time. For geospatial big data, these four Vs must be handled in the context of dynamic space and time to extract the 'value' from big data, which creates further challenges.

High-performance computing (HPC) provides an essential solution to geospatial big data challenges by allowing fast processing of massive data collections in parallel. Handing geospatial big data with HPC can help us make quick and better decisions in time-sensitive situations, such as emergency response (Bhangale et al., 2016). It also helps us to solve larger problems, such as high-resolution global forest cover change mapping in reasonable timeframes (Hansen et al., 2013) and to achieve interactive analysis and visualization of big data (Yin et al., 2017).

Therefore, this section explores how High-Performance Computing is used to handle geospatial big data. It first summarizes four typical sources of geospatial big data. It then describes the four key components, including data storage and management, spatial indexing, domain decomposition, and task scheduling. It also defines High-Performance Computing, a brief history of High-Performance Computing, its benefits to the organization, underlying technologies, service and delivery models offered by High-Performance Computing, characteristics, factors to be considered in HPC adoption, where and where not to adopt HPC, the economics of HPC and the recommendations for adoption.

2.2. Theoretical Framework

Huge quantities of data are being generated across a broad range of domains, including banking, marketing, health, telecommunications, homeland security, computer networks, e-commerce, and scientific observations and simulations. These data are called *big data*, which generally refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse. Geospatial big data is also a specific type of big data that contains location information. Most data in this modern day are inherently spatial, collected with ubiquitous location-aware sensors such as satellites, global positioning systems, and environmental observations. This type of big data offers great opportunities for advancing other industries and sectors such as climate science, disaster management, public health, precision agriculture, and smart cities. However, the ability to efficiently and promptly extract meaningful information from spatial data has been limited due to its demand for efficient computational power.

HPC can be used to handle geospatial big data and solve geospatial big data challenges by allowing the fast processing of massive datasets in parallel. Handling geospatial big data with HPC can help the responsible authorities to make quick and better decisions in time-sensitive situations, such as emergency response. It also helps the responsible to solve larger problems, such as high-resolution global forest cover, changing the mapping in reasonable timeframes and achieving interactive analysis and visualization of big data.

Four typical sources of geospatial big data are summarized below.

Geospatial and earth observations

Geospatial and earth observation systems generate massive volumes of disparate, dynamic, and geographically distributed data with in-situ and remote sensors. Remote sensing, with its increasingly higher spatial, temporal, and spectral resolutions, is one primary approach for collecting earth observation data on a national scale. The Landsat archive, for example, exceeded one petabyte and contained over 5.5 million images several years ago (Wulder et al., 2016; Camara et al., 2016). When efficiently operating, the national surveying team can bring in more than a terabyte of spatial data to its archives every day, which a general computer cannot quickly handle. In Zimbabwe, the wide use of drone-based remote sensing has opened another channel for big Earth observation data collection.

- *Geoscience model simulations*

  The rapid advancement of computing power can allow Zimbabwe to model and simulate earth phenomena with increasingly higher spatiotemporal resolution and greater spatiotemporal coverage, producing huge amounts of simulated national geospatial data. A typical example can be a climate model simulation. According to the Intergovernmental Panel on Climate Change (IPCC), they produced ten petabytes of simulated climate data from their Fifth Assessment Report (AR5) alone, and the next IPCC report is estimated to have produced hundreds of petabytes (Yang et al., 2017; Schnase et al., 2017). (Murphy et al., 2014) suggest that calibrating the geoscience models also produces large amounts of geospatial data, since a model often must be run many times to sweep different parameters. A general-use computer will struggle to process this.

- *Internet of Things*

  The term *Internet of Things* (IoT) was first coined by Kevin Ashton in 1999 in the context of using radio frequency identification (RFID) for supply chain management (Ashton, 2009). The IoT connects "things" to the internet and allows them to communicate and interact with one another, making them form a vast network of connected things. The "things" include devices and objects such as sensors, cell phones, vehicles, appliances, medical devices, and so on. These things, in the modern day, have been coupled with ubiquitous location-based sensors, which are generating massive amounts of geospatial data. IoT continuously generates unstructured or semi-structured geospatial data streams across the globe, which are more dynamic, heterogeneous, and noisy and will need robust computing power to be able to be processed in real-time.

- *Volunteered geographic information*

  Volunteered geographic information (VGI) refers to the creation and dissemination of geospatial data from the public, in which citizens are regarded as sensors moving "freely" over the surface of the earth (Goodchild, 2017). The whole VGI is enabled by the internet, GPS, and smart technologies, and as a result, massive amounts of location-based data are being generated and disseminated by billions of citizen sensors inhabiting the world. Through geotagging (location sharing), for example, social media platforms such as Twitter, Facebook, Instagram, and Tinder provide environments for digital interactions among millions of people in the virtual space while leaving “digital footprints” in the physical space. For example, about 500 million tweets are sent per day according to Internet Live Stats (2019); assuming the estimated 1% geotagging rate (Marciniec, 2017), five million tweets are geotagged daily.

This study, therefore, relies on this framework model of big data to make the case that high performance can solve the challenges associated with big data and contribute more to the national development processes. The challenges of handling big data had led to a shift in focus away from digital spatial data collection toward a more traditionally focused way of collecting spatial data. According to this research, the adoption of high-performance computing can be a crucial stepping stone toward the next researched big data handling models nationally through ZINGSA.

Key Components of Geospatial Big Data Handling with HPC

1. **Data storage and management**

   Data storage and management is essential for any data manipulation system, and it is especially challenging when handling any type of big data, particularly spatial data. Massive volumes of data require large and reliable data storage. Traditional storage and general-purpose computers cannot efficiently handle data at the petabyte scale (Robinson, 2012). Second, the fast velocity of the data requires storage with the flexibility to scale up or out to handle the ever-increasing storage demands (Katal et al., 2013).

2. **Spatial indexing**

   Using HPC, many processing units must concurrently retrieve different pieces of data to perform various data processing and spatial analysis in parallel, for example, clipping, road network analysis, and remote sensing image classification. Spatial indexing is used to quickly locate and access the needed data, such as some specific image tiles for raster data or specific geometries for vector data, from a massive dataset; to quickly locate and retrieve a specific unit from terabytes of spatial data at the byte level. Since the performance of the spatial index determines the efficiency of concurrent spatial data visits (Zhao et al., 2016), it directly impacts the performance of parallel data processing.

3. **Domain decomposition**

This technique devises a divide-and-conquer approach, by first dividing a big problem into concurrent small problems and then processing the small problems in parallel using multiple processing units. This procedure is called *decomposition*. Based on the problem to be solved, the decomposition will take one of three forms: domain decomposition, function decomposition, or both. Domain decomposition treats the data to be processed as the problem and decomposes the big problem into many small datasets. Parallel operations are then performed on the decomposed data and then focus on the computation, dividing the big computation problem (e.g., a climate simulation model) into small ones (e.g., ocean model, atmospheric model).

4. **Task Scheduling**

   Task scheduling refers to distributing subtasks to concurrent computing units (e.g., CPU cores or computers) to be processed in parallel. Task scheduling is essential in HPC because the time spent to finish subtasks has a direct impact on parallelization performance. Determining an effective task schedule depends on the HPC programming paradigms and platforms, the problems to be parallelized, and the underlying computing resources. Regardless, two significant aspects must be considered to design efficient task scheduling approaches for geospatial big data processing: load balancing and data locality.

2.3. Relevance of Theoretical Framework

Spatial big data is relevant and valid to modern-day activities. The literature highlighted practical examples that are currently being faced and ongoing in the geo-intelligence world. However, the literature did not to a greater extent highlight how the growing need for greater computational needs may be managed and addressed. It also did not address the geospatial data processing platforms that can be employed.

2.4. Definitions

High-Performance Computing

The most normative definition of High-Performance Computing is that issued by the United States Geological Survey (USGS) which states that High-Performance Computing most generally refers to the practice of aggregating computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation in solving large problems in science, engineering, or business.

Geospatial Data

IBM defines geospatial data as information that describes objects, events, or, other features with a location on or near the surface of the earth. Geospatial data typically combines location information (usually coordinates on the earth) and attribute information (the characteristics of the object, event, or, phenomena concerned) with temporal information (the time or life span at which the location and attributes exist). The location provided may be static in the short term (for example, the location of a piece of equipment, an earthquake event, and children living in poverty) or dynamic (for example, a moving vehicle or pedestrian, the spread of an infectious disease).

The Internet and High-Performance Computing

The Internet is defined as a network of networks, which provides software/hardware infrastructure to establish and maintain connectivity of computers around the world**.** High-Performance Computing, is a technology that delivers a variety of resources over a network. Therefore, in establishing a connection and operating on High-performance putting to deliver services, one needs Internet access, and services can be offered within enterprises through LANs but in reality, they cannot operate nationally, regionally, or globally without the Internet.

HPC Models

- Embedded Software.
- Supercomputer.
- Cluster Computing.
- Embedded Systems.
- Graphics Processing Unit.
- Message Passing Interface.
- Grid Computing.
- Cloud Computing.
- Quantum Computing.
- Workload.

The Embedded HPC software is the response to the growing appetite for better computational power in embedded applications. This kind of application requires the CPU and field-programmable gate array (FPGA) to be on the same board and ideally in the same chip.

One of the best-known types of HPC solutions is the supercomputer. A supercomputer. A supercomputer is a computer with a high level of performance as compared to a general-purpose computer as it contains thousands of compute nodes that work together to complete one or more tasks. This is called parallel processing. It's similar to having thousands of general-purpose computers networked together, combining computing power to complete tasks faster. The performance of a supercomputer is commonly measured in floating-point operations per second (FLOPS) instead of million instructions per second (MIPS).

In HPC, clustering refers to the use of multiple computers, typically PCs or UNIX workstations, multiple storage devices, and redundant interconnections, to form what appears to users as a single highly available system. Cluster computing can be used for load balancing as well as for high availability.

An embedded system is a microprocessor- or microcontroller-based system of hardware and software designed to perform dedicated functions within a larger mechanical or electrical system, either as an independent system or as a part of a large system. At the core is an integrated circuit designed to carry out computation for real-time operations.

A graphics processing unit (GPU) is a specialized electronic circuit designed to manipulate and alter memory to accelerate the creation of images in a frame buffer intended for output to a display device. GPUs are efficient at manipulating computer graphics and image processing. Their parallel structure makes them more efficient than general-purpose central processing units (CPUs) for algorithms that process large blocks of data in parallel.

The message-passing interface (MPI) is a standardized means of exchanging messages between multiple computers running a parallel program across distributed memory. In parallel computing, multiple computers – or even multiple processor cores within the same computer – are called nodes. Each node in the parallel arrangement typically works on a portion of the overall computing problem. The challenge then is to synchronize the actions of each parallel node, exchange data between nodes, and provide command and control over the entire parallel cluster. The message-passing interface defines a standard suite of functions for these tasks. The term message passing itself typically refers to the sending of a message to an object, parallel process, subroutine, function, or thread, which is then used to start another process.

An HPC workload is a highly complex, data-intensive task that is spread across compute resources, which each runs part of the task in parallel. An HPC system can run millions of scenarios at once, using terabytes (TB) of data at a time, which helps organizations get insights faster.

Grid Computing; This model is similar to cluster computing but operates on a larger scale. It involves multiple clusters of computers connected over the internet to form a computational grid. Grid computing is suitable for complex simulations and data analysis that require a large distributed infrastructure.

The Cloud Computing model provides on-demand access to computing resources, including processing power, storage, and software, over the Internet. Cloud computing allows users to quickly scale up or down their computing resources according to changing needs. It is suitable for users who require occasional access to HPC resources without the high cost of building and maintaining a dedicated infrastructure.

The Quantum Computing model is still in its early stages but has the potential to revolutionize computing power by using quantum bits (qubits) to perform computations. While traditional computers operate on binary states of 0 and 1, quantum computing works with states that may exist in both states simultaneously, multiplying the computational power exponentially.

Chapter 3: Research Methodology

3.1. Introduction

This chapter outlines the research design, the study's population, sample and sampling techniques, the data collection procedures, and the data analysis.

Research Design

The purpose of this research is to identify the concerns and challenges of the adoption of High-Performance Computing in the National Spatial Data Analysis for Zimbabwe thus a descriptive-analytical approach was used in this research as it will give an in-depth. This type of research will require quantitative research methods. Therefore, questionnaires will be used to collect the data from participants however depth interviews have to be used for data collection. The Questionnaires are given to the people at the Zimbabwe National Geospatial and Space Agency and interviews will be carried out with the management of ZINGSA.

3.2. Population, sampling procedure, and Sample size

Population

The population sample includes the ZINGSA family and this study was conducted at ZINGSA thus the Harare Head Office and Mazowe Ground Station. Also, the study was conducted at ZCHPC which is found in a nearby complex to the ZINGSA complex. The possible population will include almost all the employees at ZINGSA with general desktop computers thus those who can access their emails thus approximately 50 employees.

Sampling procedure

The researcher is going to use Stratified Random. According to Haynes, 2019, this is whereby the population is broken down into relatively homogenous categories in stratified samples, named layers, and examples of strata may be sex/gender, age groups, educational levels, or management levels. Selection is conducted randomly in the individual strata and this ensures the possibility of selection of each variable in the population (Hayes A. , 2019).

Sample size

The researcher will use the stratified random sampling method, by using questionnaires and conducting interviews with the employees at ZINGSA and ZCHPC. The stratified sampling will be carried out from the strata thus the selected individuals will get the chance to choose between taking a questionnaire and being interviewed. The researcher will research a sample size population of ZINGSA. This technique for sampling strata is used to distinguish between employees that prefer working on traditional data processing and analysis and those that support the adoption of HPC.

3.3. Data Collection Instruments

A data instrument is a tool used to collect, measure, and analyse that is related to the topic. Questionnaires, surveys, tests interviews are examples of data collection tools. So, in this research, the researcher will use interviews, questionnaires, and the analysis of documents to collect the necessary data. The researcher instruments will allow the researcher to collect the relevant data which answered the various research questions.

3.4. Analysis of Data

Data will be presented qualitatively and quantitatively. Research findings will be presented descriptively inner thus through the narration of findings. The quantitative analysis will include pie charts, graphs, and diagrams which will be used to give a quick visual impression but qualitative explanations will accompany every quantitative presentation**.**

3.5. Confidentiality/ Privacy

The employees who will have participated will not be entitled to give out their identities as a way of keeping the information private and highly confidential.

Investigator Experience/qualifications

The researcher was recently on an 8-month attachment period with ZINGSA.

3.6. Ethical Considerations

Fouka in 2011, described ethical considerations as the basic principles that must be followed by an individual. They make sure that no human rights are violated and that the research being conducted has no hidden agenda.

In my study, the following ethical principles will be adhered to:

i. The researcher will first seek permission to undergo the study at ZINGSA and ZCHPC and will adhere to the rules.

ii. No interviewee will be threatened or coerced into cooperating with the investigation. They will have the right to choose whether or not to submit the report.

iii. The respondents' anonymity and confidentiality, as well as their personal information, will be preserved in the study. For this purpose, no reference will be made to a single respondent, whether by name, code, or identification, and various tools were in place to protect the respondent's identity.

iv. Only people who will give their consent and understanding of what they will be doing will take part in the study. Respondents will be given sufficient information about the study before being asked to answer some questions.

v. The researcher treated all respondents with the same regard and will be complimentary about the sample.

Summary

This proposal is a descriptive review of the adoption of high-performance computing at the Zimbabwe National Geospatial and Space Agency. It is an extensive survey that will use quantitative analysis to develop knowledge. The research will discuss, among other topics, research paradigms, the architecture, research processes, data collection tools, data analysis, description and presentation procedures, and research ethics.

CHAPTER 4: DATA PRESENTATION, ANALYSIS AND INTERPRETATION

4.1. Introduction

This chapter is focused on the comprehensive presentation, analysis and interpretation of the data gleaned from the research conducted with the help of the tools for data gathering described in the previous chapter and a detailed analysis of the research findings concerning the objectives. Understanding how to make sense of the raw data collected in the previous chapter is the main contribution of this section. The study's objectives informed the design of online questionnaires and semi-structured interviews with study participants, who were asked about their perspectives on the essence of adopting high-performance computing in national spatial data analysis. In this way, the study serves as a bridge between the research findings and the dissertation's literature review analysing the various factors of high-performance computing technology, and the benefits and consequences of this technology. This chapter further discusses the demographic statistics of the participants.

4.2. The response rate of questionnaires

The researcher sent the google forms-based questionnaire to 20 randomly selected participants via email, as stated in the previous chapter. Of these 20 questionnaires, 18 were completed successfully, providing a 90% response rate to continue with the rest of the study. The researcher concluded that some of them were involved in other duties and they followed up on non-represent ants. No response was given to the two questionnaires.

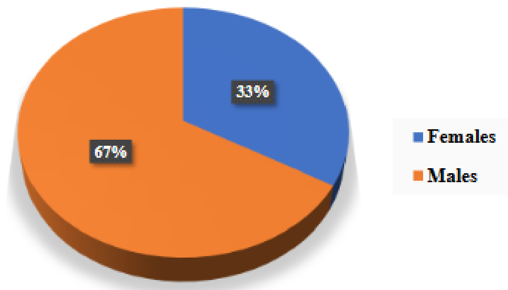
4.3. Data Presentation and Analysis

4.3.1. Gender

The majority of the respondents were men at 66.67 % and females at 33.33% as summarized in the figure. Figure 4.2.1 shows explicitly that the populations utilizing and largely involved around the information and technology and in the GIS, departments were mostly men responsible for all activities and utilization of the technology tools. The questionnaire was answered whether the ZINGSA employees had little or more knowledge of HPC technology.

GENDER OF RESPONDENTS

Gender	Number of Respondents	Percentage
Females	6	33.3333333
Males	12	66.6666667

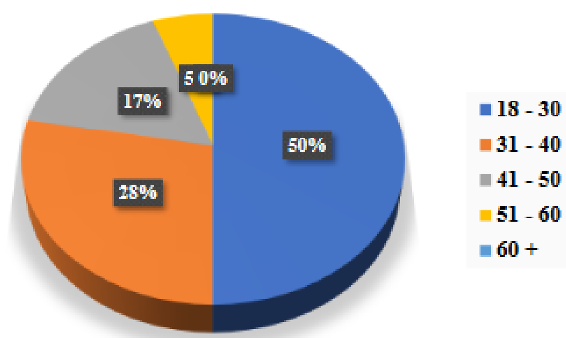


4.3.2. Age

Most people were between 18 and 30 years of age. As shown below, this age group comprised 50 % of the population. The statistics show that the age group 18–30 had about 50% and has the bulk of respondents from the chart below, followed by the age group 31–40, with 27.78%. The age group 41–50 had about 16.67%. The 51 to 60 age group had 5.56% and the 60 and above community was the last with no representatives.

AGE OF RESPONDENTS

Age Range	Number of Respondents	Percentage
18 - 30	9	50
31 - 40	5	27.7777778
41 - 50	3	16.6666667
51 - 60	1	5.55555556
60 +	0	0

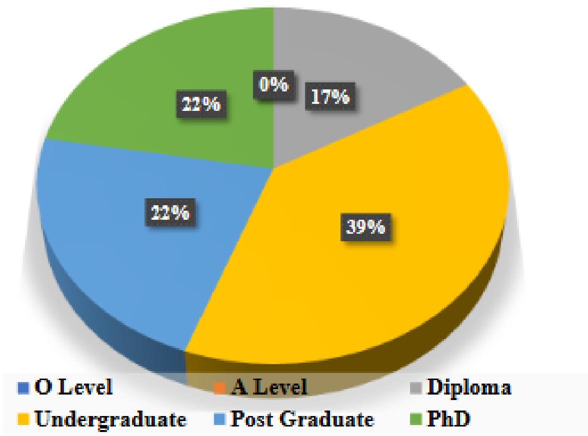


4.3.3. Educational Level

The statistics show that most of the population are undergraduate holders at (39%), followed by postgraduates at (22%), on an equilibrium with Doctor of Philosophy (PhD) at 22 % also. The least population were diploma holders with a representation of 17%.

LEVEL OF EDUCATION OF RESPONDENTS

Level Of Education	Number of Respondents	Percentage
O Level	0	0
A Level	0	0
Diploma	3	16.6666667
Undergraduate	7	38.8888889
Post Graduate	4	22.2222222
PhD	4	22.2222222



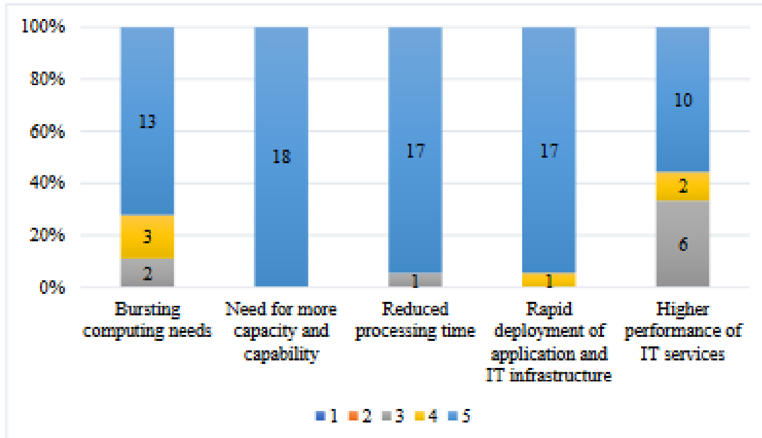
4.3.4. Knowledge of HPC Technology

From the study, almost all of the respondents in the company have a certain knowledge of cloud computing thus 99.9% probability of HPC awareness in the company. The whole proportion of people who have knowledge of high-performance computing at ZINGSA is due to their exposure to the technological world, and the sister company they share the complex with, just a stone's throw, which is the Zimbabwe Centre for High-Performance Computing.

4.3.5. Factors Driving the adoption of Cloud Computing

FACTORS DRIVING THE ADOPTION OF HPC

Factors driving the adoption of HPC	1	2	3	4	5
Bursting computing needs			2	3	13
Need for more capacity and capability					18
Reduced processing time			1		17
Rapid deployment of application and IT infrastructure				1	17
Higher performance of IT services			6	2	10



Presented above are the factors driving the adoption of high-performance computing at the Zimbabwe National Geospatial and Space Agency, (ZINGSA) and upon analysing the data from the respondents at ZINGSA, it was revealed that 72.22% of the respondents strongly agreed that one of the major factors driving the adoption of this technology is the bursting of computer needs. 16.67% of the respondents agreed to the same driving factor towards the adoption of High-Performance Computing, whilst 11.11% of the respondents were neutral and none of the respondents disagreed that bursting computer needs is a factor that drives the organization towards the adoption of cloud computing. Now, based on the above findings, it is obvious that most participants 72.22% agreed that the bursting computing needs are one of the factors driving the adoption of cloud computing at ZINGSA.

Another factor is the need for more capacity and capability for the organisation as it is still in its growth phase. The data from the respondents on this factor showed that all the respondents strongly agreed to this factor, as ZINGSA is still just at the beginning of its trajectory and there are a lot more expectations and goals to be achieved, giving it a 100 per cent of the respondents strongly agreeing to this factor.

Moving on to the other factor that is likely to drive ZINGSA to adopt high-performance computing. 94.44% of the respondents strongly agreed that reduced processing time is also another factor driving the adoption of this technology, whereas 5.56%, only one of the 18 respondents was not sure if reduced processing time is a driving factor for high-performance computing, thus they are on the neutral side. Therefore, it is noteworthy to mention that the majority of the respondents were sure and agreed to the fact that reduced processing speed is a factor that will drive the adoption of high-performance computing at the Zimbabwe National Geospatial and Space Agency.

Another interesting factor driving the adoption of high-performance computing is the rapid deployment of applications and ICT services and infrastructure which resulted in 94.44% strongly agreeing to this, 5.56% of the respondents agreed and none were neutral or had to disagree. Hence, from the above findings, it is obvious that the majority of the respondents 94.44% agreed that one of the factors driving the adoption of high-performance computing at ZINGSA is the rapid deployment of applications and ICT services as well as infrastructure.

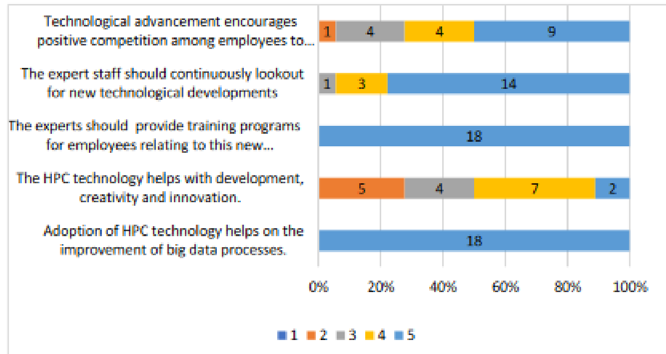
The last proposition on the factor in question was the higher performance of general-purpose IT services. For this proposed factor, nearly half of the respondents strongly agreed to this, pegged at 55.56% of the respondents. 11% of the respondents fairly agreed with the proposition that higher performance of the IT services can drive ZINGSA to adopt high-performance computing. 33.33% took a neutral position regarding this proposition and none of the respondents disagreed with this proposition.

We cannot talk of any computing technology without networking and internet accessibility, it was revealed that all of the respondents strongly agreed that on the network accessibility, the internet quality is very high. they all agreed to the same proposition

4.3.6. Promising Active HPC Evaluation

PROMISING ACTIVE HPC EVALUATION

	1	2	3	4	5
Adoption of HPC technology helps on the improvement of big data processes.					18
The HPC technology helps with development, creativity and innovation.		5	4	7	2
The experts should provide training programs for employees relating to this new technologies					18
The expert staff should continuously lookout for new technological developments			1	3	14



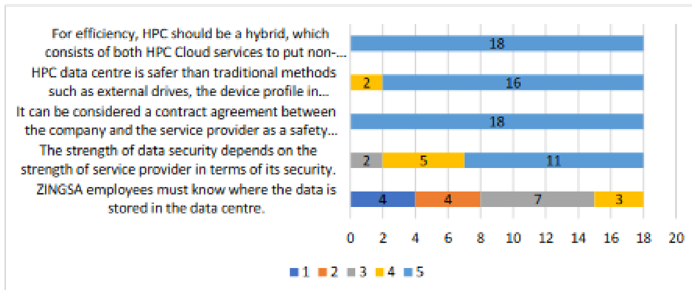
From the above illustration, all the respondents 100% strongly agree that the adoption of high-performance computing technology helps with the development of big data processes in the future and none of the respondents agree that high-performance technology helps with big data processes and hence it’s a benefit to consider for ZINGSA as an organization, and Zimbabwe as a country. Another promising active HPC evaluation for the future is that it helps with development, creativity and innovation. 11.11% strongly agreed with this evaluation whilst 38.89% mildly agree on the same proposition. 22.22% of the respondents maintained a neutral response whereas 27.78% disagreed with the proposition that HPC helps employees with development, creativity and innovation. All in all, a higher percentage of the respondents agreed with this evaluation.

The other proposition was that the ICT team and HPC experts should continue to look for newer HPC technology developments. 77.78% also strongly agree with the same. Of the respondents, 16.67% just agreed to the proposition and 5.56% took a neutral position on the proposition. Therefore, the majority seems to agree on the continuous lookout for new technological developments as much as HPC is concerned. 100% of the respondents strongly agreed that training and advancement should be offered to employees, and there were no neutral or negative standings to this proposition. By analysing the findings, the majority of the respondents confirmed that there is a brighter proposition that lies in favour of ZINGSA in the future, for the organisation, and the nation.

After analysing all the findings on the promising active high-performance evaluation, a greater percentage seems to be agreeing with the proposition that the promising future of high-performance computing in spatial data analysis is bright, only a few are neutral or seem to disagree on these factors.

4.3.7. Security

SECURITY	1	2	3	4	5
ZINGSA employees must know where the data is stored in the data centre.	4	4	7	3	
The strength of data security depends on the strength of service provider in terms of its security.			2	5	11
It can be considered a contract agreement between the company and the service provider as a safety and reliability of the data.					18
HPC data centre is safer than traditional methods such as external drives, the device profile in putting their reports and researches.				2	16
For efficiency, HPC should be a hybrid, which consists of both HPC Cloud services to put non-sensitive and public applications and also from the Private Data Centre to maintain the confidentiality and security of data.					18



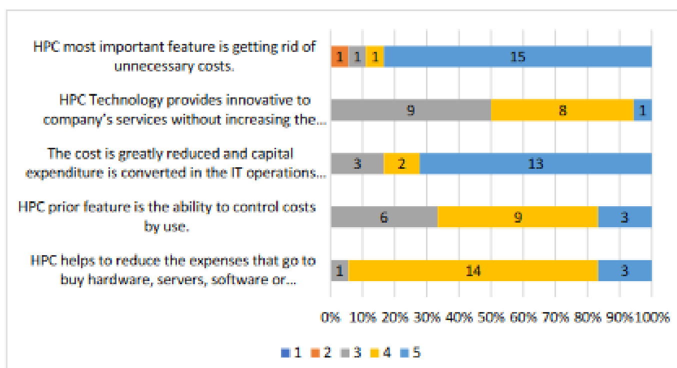
The most important feature in any organisation is the security of data. Therefore, the responses form the survey showed that none of the respondents strongly agree that ZINGSA employees must know where the data is stored in the HPC data centre, 16.67% agree to this same proposition whilst 38.89% remained neutral thus they are not sure of what they want. But the minority of the responding population want to know where the organisation's data is being stored in the HPC data centre. Another factor which was included is the strength of data security which they say depends on the service provider. The majority constituting 61.11% strongly agree with this factor. Of the respondents, 27.78% agree with this whilst 11.11% remained neutral on this factor.

The whole 100% of the respondents strongly agree that it can be considered a contract agreement between ZINGSA and the HPC service provider for the safety and reliability of the data and they also agreed that for efficiency, the HPC system should be a hybrid thing, which consists of HPC Cloud services and therefore, an on-site HPC data centre, and there were no neutralities or negativities.

The whole 100% of the respondents all agreed that an HPC data centre is safer than traditional methods such as external drives of which the greater percentage, 88.89% strongly agreed and 11.11% of the respondents agreed.

4.3.8. Efficiency through High-Performance Computing

EFFICIENCY IMPROVEMENT THROUGH HPC ADOPTION	1	2	3	4	5
HPC helps to reduce the expenses that go to buy hardware, servers, software or maintenance.			1	14	3
HPC prior feature is the ability to control costs by use.			6	9	3
The cost is greatly reduced and capital expenditure is converted in the IT operations to ongoing expenses.			3	2	13
HPC Technology provides innovative to company's services without increasing the cost or the price of the service.			9	8	1
HPC most important feature is getting rid of unnecessary costs.		1	1	1	15



From the data collected, it is shown that 16.67% strongly agree that high-performance computing helps to reduce expenses to regularly go buy hardware at ZINGSA, and 77.78% somewhat agree that high-performance computing helps to reduce expenses to regularly go buy hardware at ZINGSA. 5.56% were neutral on the fact that high-performance computing helps to reduce expenses to regularly go buy hardware whilst none disagreed.

The gathered data indicate that 66.67% of the sample was for high-performance computing helps to control expenses by use. 16.67% of the sample strongly agreed to the proposition that HPC's prior feature can control cost by use, and the other 50% agreed, with 16.67% taking a neutral stand on this.

The other efficiency contributor through high-performance computing adoption that was indicated by the respondents is the reduction of cost and capital expenditure, getting it converted in the IT operations to ongoing expenses. A greater proportion of the population strongly agreed with this fact, with a total percentage of 72.22% of the respondents, with 11.11% also agreeing on this, and just a third of the sample were neutral and 0% disagree on this.

Another factor includes the innovation that high-performance computing brings without increasing cost appeared to be one of the leading HPC efficiency factors with 50% strongly agreeing while the cost and the other 50% were neutral on this factor.

High-performance computing's most important feature is getting rid of unnecessary costs and promoting efficiency considering ZINGSA would benefit from this proposition with 83.33% strongly agreeing, 5.56% also agreeing to this while 5.56% were neutral and 5.56% disagreed that high-performance computing is a most important feature is getting rid of unnecessary costs within the organization.

The responses obtained by the researcher showed that to a larger extent, the respondents agree that high-performance computing most important feature in improving efficiency and getting rid of unnecessary costs.

4.4. Chapter Summary

This chapter presented and visualised the results of the research, examined them and reviewed them in association with the literature review. The beginning of the chapter showed that participants' rate of response was obtained using questionnaires and interviews. The data visualization and analytics were done after and were organised following the research objectives. The next chapter looks at the summary, conclusions, and recommendations of the study.

Chapter 5: Summary, Conclusions and Recommendations

5.1. Introduction

This chapter focuses on ending the study, the conclusions accompanied by recommendations on the conclusion. The chapter also access whether the key factors of the objectives of the study were achieved and the areas of further study are highlighted in this chapter as well.

5.2. Summary of Findings

The study aimed to investigate the factors driving the adoption of high-performance computing at the Zimbabwe National Geospatial and Space Agency. It focused on establishing whether the employees at this organization knew this technology and the benefits and challenges associated with this technology. The relevant literature review looked at many writers' views on the key subject areas within the research topic. Chapter 3 further explained the methodology of the study detailing the research instruments employed for the study and Chapter 4 moved on to present the research findings and gave an analysis of the findings. This final chapter will therefore provide summaries of the four preceding chapters, conclusions, suggestions for future research and recommendations.

5.3. Discussion

From the overview of the Zimbabwe National Geospatial and Space Agency, certainly, high-performance computing (HPC) plays a central role in the analysis of Zimbabwe's national spatial data infrastructure (NSDI) due to the large amounts of data involved and the complexity of the analysis. NSDI is a collection of technologies and policies that enable the management, sharing, and dissemination of geospatial data, such as maps, satellite images, and sensor data. These data sets can be large and complex, requiring sophisticated analysis techniques to extract meaningful information.

HPC provides the computational power needed to analyse such large and complex data sets quickly and accurately. For example, HPC clusters can be used to process large amounts of satellite imagery to detect changes in land use over time. Similarly, HPC can be used to analyse large amounts of sensor data to detect patterns and anomalies.

The researcher observed that HPC in SDI analysis is the ability to perform simulations and modelling to test different scenarios. For example, HPC can be used to simulate the effects of climate change on sea levels or to model the spread of a disease outbreak.

Furthermore, the researcher observed that HPC can be used to integrate and analyse data from multiple sources. For example, it can be used to combine satellite imagery with ground-based sensor data to create a more complete picture of environmental conditions.

In conclusion, HPC plays an essential role in SDI analysis by providing the computational power needed to process large and complex geospatial data sets. With the growing importance of SDI in decision-making processes, the use of HPC will continue to evolve and expand, opening up new opportunities for data-driven insights and analysis.

More specifically, the loss of stakeholders' confidence in high-performance computing stems from a lack of knowledge.

The researcher observed that the perceived costliness of HPC technologies can discourage some entities from adopting HPC.

Existing skills and knowledge of ICT professionals in organizations to implement and manage

The researcher also observed that the expanse of ZINGSA's Internet of Things (IoT) is a green light for the adoption of HPC and due to this, it can result in high performance of the organization and efficiency of the organization.

The researcher observed that the adoption of high-performance computing at ZINGSA will greatly assist the ICT team or department in reducing its expenses on the company and focus more on advancing to the latest technology.

5.4. Conclusions

This study has presented the factors that drive the adoption of high-performance computing at the Zimbabwe National Geospatial and Space Agency and this study has defined high-performance computing in terms of its functions and presented its benefits in national spatial data analysis and highlighted its challenges in the organization. The results showed that the employees and various stakeholders at ZINGSA are more than ready to adopt cloud computing due to its various advantages and therefore we will conclude this study by referring to the research questions that were posed at the beginning of the research.

5.4.1. To what extent is ZINGSA aware of high-performance Computing?

Based on the findings, it revealed that to a greater extent, ZINGSA is very much aware of high-performance computing, its functions together with benefits. As indicated, the majority of the participants responded yes to the knowledge of cloud computing and they also managed to identify the main HPC service providers and the different classifications. They also managed to identify the key factors which will drive the company towards the adoption of this technology which includes the reduction of unnecessary costs thereby giving a highly competitive advantage to the organization therefore through higher performance within the organization.

However, there are a few who were not aware of this technology and this can be addressed by the zealous ICT team by carrying out activities such as training and campaigns of high-performance computing awareness among the employees.

5.4.2. To the extent does Spatial Data Analysis and High-performance computing synergize and relate

Big data is characterised by "Vs" which are volume, variety, velocity, veracity, and value, and massive datasets are to be dealt with in the context of space and time. High-performance computing (HPC) involves the use of advanced computer systems and software to solve complex problems that require significant computational power. These systems typically utilize clusters of interconnected computers, specialized processors, and multi-threaded software to perform computations at high speeds. Spatial data is then complex and it requires advanced computing solutions and significant power, so HPC is an answer and a solution to the storage, processing and analysis of big spatial data hence the synergy.

5.4.3. The Importance of high-performance Computing in the national spatial data analysis

High-performance computing is very important in the NSDI as it helps in solving quite a range of problems such as acquiring and maintaining expensive hardware and software resources that we use in our daily operations to streamline these resources most efficiently for our benefit and the benefit of the society at large. HPC is also important in national spatial data analysis because it addresses all the concerns that are connected with big data handling.

5.4.4. The main factors that are more likely to influence ZINGSA to adopt high-performance computing

1. Bursting computer needs and need for more capacity and capability - The amount of spatial data being generated by the GIS Internet of Things and machines is growing at an unprecedented rate. HPC provides a way to process and analyse these large spatial datasets quickly and efficiently.
2. The Complexity of Problems - Many of the challenges faced by researchers and engineers today are highly complex and they require significant computational resources to solve. HPC provides the necessary computing power to tackle these problems such as climatic simulations. HPC enables researchers to create more accurate models by simulating real-world scenarios with greater precision
3. Reduced Processing Time - HPC systems are designed to perform parallel processing, which allows multiple computations to be carried out simultaneously. This greatly speeds up the processing time for complex spatial data analysis and simulations.
4. Cost-Effective - HPC offers a cost-effective alternative to building large-scale physical prototypes or conducting costly experiments.

5.4.5. Can high-performance computing take national spatial data analysis in the future

The future of high-performance computing (HPC) looks very promising as the demand for computational power and advanced technology increases in national spatial data analysis with real-time analysis, cloud-based HPC solutions, Artificial intelligence and machine learning solutions and high-resolution imaging. Technological advancement of HPC will encourage innovation and positive competition among employees to motivate themselves and thus serve the general interest of the company and nation

5.4.6. What are the benefits and barriers of high-performance Computing?

High-performance computing has been demonstrated to offer colossal benefits to organizations handling big data. Unravelling these benefits is broadly classified into three categories that incorporate efficiency, flexibility, and strategic value.

Efficiency

To start with, efficiency in national spatial data analysis is significantly achieved by the speed of processing large-scale spatial data. Traditional methods of processing large spatial data involve sequential and serial processing, which can be slow and time-consuming.

Overall, the use of HPC in national spatial data analysis has the potential to significantly improve the efficiency, accuracy, and speed of processing large-scale spatial data as data can be processed in parallel, on multiple processors or computing cores, enabling much shorter processing times. This is particularly important for spatial data applications that require the analysis of large datasets, such as climate modelling, predicting the spread of diseases, urban planning, and natural resource management.

Flexibility

The flexibility of HPC in national spatial data analysis makes it a valuable tool that can be used for processing and analysing an extensive range of spatial data at different scales, which makes it ideal for researchers, industries, and government entities. handle a wide range of complex spatial data analyses with varying computational requirements. The flexible nature of HPC enables it to handle very complex analysis tasks such as simulation modelling, spatial network analysis, and geostatistics let alone simple tasks like mapping and geocoding. This flexibility enables users to take advantage of the parallel processing capability of the HPC systems to complete tasks much more quickly and accurately than with traditional methods.

Strategic Edge

Lastly, this technology allows any entity to gain a strategic edge in their niche in the sense that there will be Increased productivity, as the entities will be able to focus their efforts more on their core business operations other than on how to get the computational work done. The strategic edge of employing High-Performance Computing (HPC) in National Spatial Data Analysis is primarily through its ability to handle large and complex spatial datasets with enormous speed and accuracy, providing more substantial and reliable insights that could be leveraged for decision-making processes, which are otherwise too difficult to be processed using traditional computing methods.

5.4.7. **What are the factors that will influence Zimbabwe National Geospatial and Space Agency to adopt cloud computing**?

The factors that will influence Zimbabwe National Geospatial and Space Agency to adopt high-performance computing are the benefits that come with this technology which include the higher performance of IT services, Increased Processing Speed, Improved Accuracy, Reduced Operational Costs, Enhanced Data Visualization and Improved Scalability. In conclusion, its benefits are the factors that will influence this organization to adopt cloud computing.

5.5. Recommendations for ZINGSA

Due to HPC's various advantages, the researcher recommends ZINGSA adopt cloud computing to address the problem highlighted at the beginning of this research. Also, the top management and the ICT team should be on the continuous lookout for advanced technology use. And lastly, the ICT team should develop plans which are flexible enough to accommodate any changes required by the adoption of high-performance computing

5.6. Implications

1. Speed and Efficiency: HPC can handle large amounts of data in a short time, enabling quick analysis and decision-making. This allows national spatial data analysts to process large amounts of data to create accurate insights, models, and simulations. With HPC, scientists and researchers can perform complex simulations and analyses in less time, which can speed up the discovery process. HPC can assist in predicting natural phenomena and developing new materials and products.
2. Accuracy and Precision: HPC allows for accuracy and precision in data analysis, which is essential for national spatial data analysts. By providing the ability to perform complex calculations, researchers can produce high-quality models and simulations, enabling better decision-making. HPC allows businesses to analyse large datasets quickly and efficiently, making it easier to identify patterns and trends. This can help organizations make better and faster decisions based on insights from big data.
3. Data Sharing: HPC provides a platform for sharing data, allowing researchers to collaborate across organizations and geographical boundaries. This promotes sharing of

knowledge and intellectual property, which can lead to discoveries and advances in national spatial data analysis.

4. Predictive Analytics: HPC makes it possible to perform predictive analytics in real-time, making it easier to predict and respond to disasters or other events that can influence national spatial data. This can help organizations to make better and faster decisions and take preventive measures.
5. Optimization of Resource Utilization: HPC helps in optimizing the utilization of resources, allowing researchers to focus on other important aspects of their jobs. By offering a scalable way of conducting research, research time is optimized, and national spatial data analysts can handle more complex projects. With HPC, scientists and researchers can perform complex simulations and analyses in less time, which can speed up the discovery process. HPC can assist in treating diseases, predicting natural phenomena, and developing new materials and products.
6. High-Fidelity Simulations: HPC enables complex simulations with unprecedented accuracy, which helps designers and engineers optimize products, architecture, and systems.
7. Enhances customer services: Stakeholders may presently get information that will help at any time and from any area. A self-service web-based site can be hosted progressively commonplace on the HPC servers. Client issues may be settled rapidly and successfully as a result. This improves general client fulfilment by lessening ticket times and enlivening request responsiveness.

5.7. Recommendations

Based on results reached by the researcher during the research study and analysis, it revealed that there are concerns and challenges in the adoption of High-Performance Computing technology at ZINGSA. However, several recommendations are suggested in light of the preceding research for ZINGSA to adopt High-Performance Computing technology in national spatial data analysis, which is neatly outlined below;

- The Zimbabwean Space Agency should adopt High-Performance Computing technology in its operations, which is an attractive technological and efficient option for the organization.

- The top management and ICT team should be informed of the ongoing developments of High-Performance Computing technology and the importance of its use.
- The ICT team should develop plans which are flexible enough to accommodate any changes required by the adoption of high-performance computing.
- The management team should support the ICT department by adopting new technology like High-Performance Computing technology to enhance its competitive advantage. The top management must have plans to get rid of obstacles that hinder the use of any new technology at ZINGSA.
- The management and responsible ministries should provide the support and the needed requirements to adopt High-Performance Computing technology at ZINGSA.
- ZINGSA should provide training programs, relating to the new technology which is High-Performance Computing.
- The organization can create a hybrid HPC infrastructure, which consists of an on-site data centre to put in sensitive and private applications and from the cloud to maintain the backups, confidentiality and security of data.

5.8. Suggestions for Further Research

Since this study primarily surveyed a space agency in Zimbabwe, future research should also inquire about other industries dealing with big data about the efficiency they are aiming at achieving and their inclusion in national progress. The researcher felt that there is rare research about High-Performance Computing technology and this is because HPC is a new topic in the ICT field thus the door is open for more academic research about this technology. The researcher suggested the following topics which may provide good research ideas for the future:

- A study to measure the effects high-performance Computing technology has on manufacturing industries
- To conduct a study to measure the effects of High-Performance Computing technology on Enterprise Resource Planning (ERP)
- A study to create a prototype of using Hogh Performance Computing technology in e-government.
- To conduct a study about the monitoring and tracking of all manufacturing activities through using High-Performance Computing.

5.9. Chapter Summary

The chapter concluded the study by giving a summary of the entire research and discussed the findings that were made from the research and in the end proposed recommendations which should be addressed and enforced at ZINGSA.

Appendices

Appendix 1: Questionnaire to Zimbabwe Centre for High-Performance Computing (ZCHPC)

My name is Munyaradzi Mwarianesu Muchapondwa and I am studying Bachelor of Honours in Computer Science at Africa University (AU). In the partial fulfilment of my study programme, I am required to carry out a study in a business field. So, I am kindly asking for your participation in my research by completing this questionnaire. **In this questionnaire**, **I would like to know your knowledge towards the adoption of high-performance computing in national spatial data analysis. I am kindly asking that you choose what you prefer most.**

The responses given will be combined as a whole and analysed.

Your participation in the study will be greatly appreciated.

Thank you very much for your time and assistance.

SECTION A: DEMOGRAPHIC INFORMATION

This section requires you to put an "X" in the appropriate box,

A1	**Gender**	**Male**		**Female**	

A2	18-31	32-42	43-53	54-64	65+

A3	**Level of education**	PhD	Postgraduate	Undergraduate	Diploma	O level	No education

SECTION B: AWARENESS OF CLOUD COMPUTING

B1. Any knowledge about high-performance computing?

YES		NO	

B2. Any knowledge about big data?

YES		NO	

B2	**Who are the main HPC users in Zimbabwe?**	Geospatial, space and earth sciences	Engineering	Life Sciences	Heritage and Philosophy

B3	**Which computing classifications are you familiar with? (Select all applicable)**	Networking and security	System Administration	Parallel programming	Cluster computing

B4	**Which operating system(s) are you using? (Select all applicable)**	CentOS	FreeBSD	Linux	Windows	CNK

SECTION C:

To the best of your knowledge, please rate how you strongly agree or disagree on the factors that will drive your organization into adopting high-performance computing (Single response only per statement). Please tick the applicable using Likert's 5-point scale as indicated below

Key: 1 strongly agree, 2 disagree, 3 neutral, 4 agree and 5 strongly agree.

Factors driving the adoption of HPC	**1**	**2**	**3**	**4**	**5**
Bursting computing needs					
Need for more capacity and capability					
Reduced processing time					
Rapid deployment of application and IT infrastructure					
The higher performance of IT services					

PROMISING ACTIVE HPC EVALUATION	**1**	**2**	**3**	**4**	**5**
The adoption of HPC technology helps in the improvement of big data processes.					
The HPC technology helps with development, creativity and innovation.					
The experts should provide training programs for employees relating to these new technologies					
The expert staff should continuously lookout for new technological developments					
Technological advancement encourages positive					

competition among employees to motivate themselves and thus serve the general interest of the company and nation					

NETWORK ACCESSIBILITY (NA)	**Yes**	**No**
Does your network service provider support 4G?		
Internet quality is very high.		
The network service provider can handle large volumes of users and the signal strength is not disturbed.		
The network is easily accessible in remote areas.		
Signal strength is always excellent.		

SECTION D:

To the best of your knowledge, please rate how you strongly agree or disagree with the following tables. (Single response only per statement). Please tick the applicable using Likert's 5-point scale as indicated below

Key: 1 strongly disagree, 2 disagree, 3 neutral, 4 agree and 5 strongly agree.

SECURITY(S)	**1**	**2**	**3**	**4**	**5**
ZINGSA employees must know where the data is stored in the data centre.					
The strength of data security depends on the strength of the service provider in terms of its security.					
It can be considered a contractual agreement					

between the company and the service provider for the safety and reliability of the data.					
HPC data centre is safer than traditional methods such as external drives, and the device profile in putting their reports and research.					
For efficiency, HPC should be a hybrid, which consists of both HPC Cloud services to put non-sensitive and public applications and also from the Private Data Centre to maintain the confidentiality and security of data.					

EFFICIENCY IMPROVEMENT THROUGH HPC ADOPTION	**1**	**2**	**3**	**4**	**5**
HPC helps to reduce the expenses that go to buy hardware, servers, software or maintenance.					
HPC's prior feature is the ability to control costs by use.					
The cost is greatly reduced and capital expenditure is converted in the IT operations to ongoing expenses.					
HPC Technology provides innovation to the company's services without increasing the cost or the price of the service.					
HPC's most important feature is getting rid of unnecessary costs.					

Thank you for your assistance. It is greatly appreciated.

Appendix 2: Interview for Zimbabwe National Geospatial and Space Agency (ZINGSA)

INTERVIEW QUESTIONS GUIDE……………

1. May you please care to share your educational background?
2. Do you feel your education prepared you to work as a GIS specialist?
3. What interests you about GIS?
4. Why are geo-statistics important?
5. What skills do you think are most valuable for a GIS specialist?
6. Is there a difference between geo-coding and geo-referencing?
7. How does GPS support GIS?
8. Which GIS applications have you used in your previous roles?
9. How is spatial data analytics set to benefit the nation?
10. According to your understanding, what is big data?
11. Do you have any experience in big data handling, particularly spatial data?
12. If you did, what are some of the challenges that come with big data handling?
13. Are you aware of the 5 Vs of big data?
14. Can spatial data be classified as big data?
15. Why is HPC so popular in big data analytics?
16. What are HPC and its main components?
17. What are some common hardware components used for High-Performance Computing?
18. What are the key steps that are followed in deploying data into the HPC platform?
19. What kind of operating system is suitable for high-performance computing clusters?
20. What types of applications can benefit from using HPC?
21. What are GPUs and how are they different from CPUs?
22. Can you describe your understanding of distributed processing? How is it different from multicore processing?
23. Looking at the current NSDI, how does it work? What could you improve?

References

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

Abramova, V., & Bernardino, J. (2013, July). NoSQL databases: MongoDB vs cassandra. In *Proceedings of the international C* conference on computer science and software engineering* (pp. 14-22). ACM

Aji, A., Wang, F., Vo, H., Lee, R., Liu, Q., Zhang, X., & Saltz, J. (2013). Hadoop gis: a high performance spatial data warehousing system over mapreduce. Proceedings of the VLDB Endowment, 6(11), 1009-1020.

Alarabi, L., Mokbel, M. F., & Musleh, M. (2018). St-hadoop: A mapreduce framework for spatio-temporal data. GeoInformatica, 22(4), 785-813.

Altintas, I., Berkley, C., Jaeger, E., Jones, M., Ludascher, B., & Mock, S. (2004, June). Kepler: an extensible system for design and execution of scientific workflows. In Proceedings. 16th International Conference on Scientific and Statistical Database Management, 2004. (pp. 423-424). IEEE.

Armstrong, M P. 2000. Geography and computational science. Annals of the Association of American Geographers, 90(1): 146–56.

Asadi, R., & Regan, A. (2019). A Spatial-Temporal Decomposition Based Deep Neural Network for Time Series Forecasting. arXiv preprint arXiv:1902.00636.

Ashton, K. (2009). That ‘internet of things’ thing. RFID journal, 22(7), 97114.

Athanasis, N., Themistocleous, M., Kalabokidis, K., & Chatzitheodorou, C. (2018, October). Big Data Analysis in UAV Surveillance for Wildfire Prevention and Management. In European, Mediterranean, and Middle Eastern Conference on Information Systems (pp. 47-58). Springer, Cham.

Audebert, N., Le Saux, B., & Lefèvre, S. (2018). Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. ISPRS Journal of Photogrammetry and Remote Sensing, 140, 20-32.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12), 2481-2495.

Baumann, P., Dehmel, A., Furtado, P., Ritsch, R., & Widmann, N. (1998). The multidimensional database system RasDaMan. Acm Sigmod Record, 27(2), 575-577.

Baumann, P., Dehmel, A., Furtado, P., Ritsch, R., & Widmann, N. (1999, September). Spatio-temporal retrieval with RasDaMan. In VLDB (pp. 746-749).

Baumann, P., Mazzetti, P., Ungar, J., Barbera, R., Barboni, D., Beccati, A., ... & Campalani, P. (2016). Big data analytics for earth sciences: the EarthServer approach. International Journal of Digital Earth, 9(1), 3-29.

Baumann, P., Misev, D., Merticariu, V., Huu, B. P., Bell, B., Kuo, K. S., & Bayesics, L. L. C. (2018). Array Databases: Concepts, Standards, Implementations. Research Data Alliance (RDA) Working Group Report.

Bhangale, U. M., Kurte, K. R., Durbha, S. S., King, R. L., & Younan, N. H. (2016, July). Big data processing using hpc for remote sensing disaster data. In 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 5894-5897). IEEE.

Blumenfeld J. (2019). Getting Petabytes to People: How EOSDIS Facilitates Earth Observing Data Discovery and Use. Available at: https://earthdata.nasa.gov/getting-petabytes-to-people-how-theeosdis-facilitates-earth-observing-data-discovery-and-use , last accessed on May 1, 2019

Bouziane, H. L., Pérez, C., & Priol, T. (2008, August). A software component model with spatial and temporal compositions for grid infrastructures. In European Conference on Parallel Processing (pp. 698-708). Springer, Berlin, Heidelberg.

Callahan, S. P., Freire, J., Santos, E., Scheidegger, C. E., Silva, C. T., & Vo, H. T. (2006, June). VisTrails: visualization meets data management. In Proceedings of the 2006 ACM SIGMOD international conference on Management of data(pp. 745-747). ACM.

Camara, G., Assis, L. F., Ribeiro, G., Ferreira, K. R., Llapa, E., & Vinhas, L. (2016, October). Big earth observation data analytics: Matching requirements to system architectures. In Proceedings of the 5th ACM SIGSPATIAL international workshop on analytics for big geospatial data (pp. 1-6). ACM.

Chang, F., Dean, J., Ghemawat, S., Hsieh, W. C., Wallach, D. A., Burrows, M., ... & Gruber, R. E. (2008). Bigtable: A distributed storage system for structured data. *ACM Transactions on Computer Systems (TOCS)*, *26*(2), 4.

Chen, Q., Wang, L., & Shang, Z. (2008, December). MRGIS: A MapReduce-Enabled high performance workflow system for GIS. In 2008 IEEE Fourth International Conference on eScience (pp. 646-651). IEEE.

Chen, X. W., & Lin, X. (2014). Big data deep learning: challenges and perspectives. IEEE access, 2, 514-525.

Cheng, G., Zhou, P., & Han, J. (2016). Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 54(12), 7405-7415.

Clarke, K. C. (2003). Geocomputation's future at the extremes: high performance computing and nanoclients. Parallel Computing, 29(10), 1281-1295.

Cudré-Mauroux, P., Kimura, H., Lim, K. T., Rogers, J., Simakov, R., Soroush, E., ... & DeWitt, D. (2009). A demonstration of SciDB: a science-oriented DBMS. Proceedings of the VLDB Endowment, 2(2), 1534-1537.

De Mauro, A., Greco, M., & Grimaldi, M. (2015, February). What is big data? A consensual definition and a review of key research topics. In AIP conference proceedings (Vol. 1644, No. 1, pp. 97-104). AIP.

Dean, J. (2016). Large-scale deep learning for building intelligent computer systems.

Dean, J., & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. Communications of the ACM, 51(1), 107-113.

Desjardins, M. R., Hohl, A., Griffith, A., & Delmelle, E. (2018). A space– time parallel framework for fine-scale visualization of pollen levels across the Eastern United States. Cartography and Geographic Information Science, 1-13.

Ding, Y. and Densham, P., 1996. Spatial strategies for parallel spatial modelling. International Journal of Geographical Information Systems, 10 (6), 669–698. doi:10.1080/02693799608902104

Duffy, D., Spear, C., Bowen, M., Thompson, J., Hu, F., Yang, C., & Pierce, D. (2016, December). Emerging Cyber Infrastructure for
NASA’s Large-Scale Climate Data Analytics. In AGU Fall Meeting Abstracts.

Eldawy, A., & Mokbel, M. F. (2015, April). Spatialhadoop: A mapreduce framework for spatial data. In *2015 IEEE 31st international conference on Data Engineering* (pp. 1352-1363). IEEE.

Eldawy, A., & Mokbel, M. F. (2015, April). Spatialhadoop: A mapreduce framework for spatial data. In 2015 IEEE 31st international conference on Data Engineering (pp. 1352-1363). IEEE.

Eldawy, A., Mokbel, M. F., Alharthi, S., Alzaidy, A., Tarek, K., & Ghani, S. (2015, April). Shahed: A mapreduce-based system for querying and visualizing spatio-temporal satellite data. In 2015 IEEE 31st International Conference on Data Engineering(pp. 1585-1596).
IEEE.

Engélinus, J., & Badard, T. (2018). Elcano: A Geospatial Big Data Processing System based on SparkSQL. In GISTAM (pp. 119128).

Esri. (2013). GIS Tools for Hadoop. https://github.com/Esri/gis-tools-forhadoop, last accessed on April 25, 2019

Fahmy, M. M., Elghandour, I., & Nagi, M. (2016, December). CoS-HDFS: co-locating geo-distributed spatial data in hadoop distributed file system. In Proceedings of the 3rd

IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (pp. 123-132). ACM.

Gabriel, E., Fagg, G. E., Bosilca, G., Angskun, T., Dongarra, J. J., Squyres, J. M., ... & Castain, R. H. (2004, September). Open MPI: Goals, concept, and design of a next generation MPI implementation. In European Parallel Virtual Machine/Message Passing Interface Users' Group Meeting(pp. 97-104). Springer, Berlin, Heidelberg.

Gong J., Wu H., Zhang T., Gui Z, Li Z., You L., Shen S., (2012). Geospatial Service Web: towards integrated cyberinfrastructure for GIScience. Geo-spatial Information Science, 15(2):73-84.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. GeoJournal, 69(4), 211-221

Google. (2019), Google BigQuery GIS, https://cloud.google.com/bigquery/docs/gis-intro, last accessed on April 25, 2019

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18-27.

Gropp, W., Lusk, E., Doss, N., & Skjellum, A. (1996). A highperformance, portable implementation of the MPI message passing interface standard. Parallel computing, 22(6), 789-828.

Guan, Q. (2009). pRPL: an open-source general-purpose parallel Raster Processing programming Library. Sigspatial Special, 1(1), 57-62.

Guan, Q., Zhang, T., & Clarke, K. C. (2006, December). GeoComputation in the grid computing age. In International Symposium on Web and Wireless Geographical Information Systems (pp. 237-246). Springer, Berlin, Heidelberg.

Gudivada, V. N., Baeza-Yates, R., & Raghavan, V. V. (2015). Big data: Promises and problems. Computer, (3), 20-23.

Gui, Z., Yu, M., Yang, C., Jiang, Y., Chen, S., Xia, J., Huang, Q., Liu, K., Li, Z., Hassan, M.A. and Jin, B., (2016). Developing Subdomain

Allocation Algorithms Based on Spatial and Communicational Constraints to Accelerate Dust Storm Simulation. *PloS one*, 11(4), p.e0152250.

Guo, Z., Fox, G., & Zhou, M. (2012, May). Investigation of data locality in mapreduce. In Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012) (pp. 419-426). IEEE Computer Society.

Guttman, A. (1984). R-trees: a dynamic index structure for spatial searching (Vol. 14, No. 2, pp. 47-57). ACM.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A. A., Tyukavina, A., ... & Kommareddy, A. (2013). High-resolution global maps of 21st-century forest cover change. science, 342(6160), 850-853.

He, Z., Wu, C., Liu, G., Zheng, Z., & Tian, Y. (2015). Decomposition tree:
a spatio-temporal indexing method for movement big data. Cluster Computing, 18(4), 1481-1492.

Hohl, A., D Griffith, A., Eppes, M. C., & Delmelle, E. (2018).
Computationally Enabled 4D Visualizations Facilitate the
Detection of Rock Fracture Patterns from Acoustic Emissions.
Rock Mechanics and Rock Engineering, 51, 2733-2746.

Hohl, A., Delmelle, E. M., & Tang, W. (2015). SPATIOTEMPORAL DOMAIN DECOMPOSITION FOR MASSIVE PARALLEL
COMPUTATION OF SPACE-TIME KERNEL DENSITY. ISPRS
Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 2(4).

Hu, F., Xia, G. S., Hu, J., & Zhang, L. (2015). Transferring deep convolutional neural networks for the scene classification of highresolution remote sensing imagery. Remote Sensing, 7(11), 1468014707.

Hu, F., Yang, C., Schnase, J. L., Duffy, D. Q., Xu, M., Bowen, M. K., ... & Song, W. (2018). ClimateSpark: An in-memory distributed computing framework for big climate data analytics. Computers & geosciences, 115, 154-166.

Huang, Z., Chen, Y., Wan, L., & Peng, X. (2017). GeoSpark SQL: An effective framework enabling spatial queries on spark. ISPRS International Journal of Geo-Information, 6(9), 285.

Hughes, J. N., Annex, A., Eichelberger, C. N., Fox, A., Hulbert, A., & Ronquest, M. (2015, May). Geomesa: a distributed architecture for spatio-temporal fusion. In Geospatial Informatics, Fusion, and Motion Video Analytics V (Vol. 9473, p. 94730F). International Society for Optics and Photonics.

Hull, D., Wolstencroft, K., Stevens, R., Goble, C., Pocock, M. R., Li, P., & Oinn, T. (2006). Taverna: a tool for building and running workflows of services. Nucleic acids research, 34(suppl_2), W729-W732.

Internet Live Stats, (2019), available at https://www.internetlivestats.com/twitter-statistics/, last accessed on May 3, 2019

Jaeger, E., Altintas, I., Zhang, J., Ludäscher, B., Pennington, D., & Michener, W. (2005, June). A Scientific Workflow Approach to Distributed Geospatial Data Processing using Web Services. In SSDBM (Vol. 3, No. 42, pp. 87-90).

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., ... & Darrell, T. (2014, November). Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia (pp. 675-678). ACM.

Katal, A., Wazid, M., & Goudar, R. H. (2013, August). Big data: issues, challenges, tools and good practices. In *2013 Sixth international conference on contemporary computing (IC3)* (pp. 404-409). IEEE.

Kini, A., & Emanuele, R. (2014). Geotrellis: Adding geospatial capabilities to spark. Spark Summit.

Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. IEEE Geoscience and Remote Sensing Letters, 14(5), 778-782.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436.

Lee, K., & Kim, K. (2018, July). Geo-Based Image Analysis System Supporting OGC-WPS Standard on Open PaaS Cloud Platform. In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium (pp. 5262-5265). IEEE.

Li Z., Hodgson M., Li W. (2018) A general-purpose framework for largescale Lidar data processing, International Journal of Digital Earth, 11(1), 26-47

Li Z., Huang Q., Jiang Y., Hu F. (2019) , SOVAS: A Scalable Online Visual Analytic System for Big Climate Data
Analysis, International Journal of Geographic Information
Science, DOI: 10.1080/13658816.2019.1605073

Li Z., Yang C., Yu M., Liu K., Sun M.(2015) Enabling Big Geoscience Data Analytics with a Cloud-based, MapReduce-enabled and Service-oriented Workflow Framework, *PloS one*, 10(3), e0116781.

Li Z., Yang, C., Huang, Q., Liu K., Sun, M., Xia, J., (2017a). Building Model as a Service for Supporting Geosciences,*Computers, Environment and Urban Systems*. 61(B), 141-152

Li Z., Huang Q., Carbone G., Hu F. (2017b) A High Performance Query Analytical Framework for Supporting Data-intensive Climate
Studies, Computers, Environment and Urban Systems, 62(3), 210221

Li, J., Jiang, Y., Yang, C., Huang, Q., & Rice, M. (2013). Visualizing 3D/4D environmental data using many-core graphics processing units (GPUs) and multi-core central processing units (CPUs). Computers & Geosciences, 59, 78-89.

Li, Z. (2015). Optimizing geospatial cyberinfrastructure to improve the computing capability for climate studies (Doctoral dissertation), available at:
http://ebot.gmu.edu/bitstream/handle/1920/9630/Li_gmu_0883E_
10873.pdf?sequence=1&isAllowed=y

Li, Z., Yang, C. P., Wu, H., Li, W., & Miao, L. (2011). An optimized framework for seamlessly integrating OGC Web Services to support geospatial sciences. International Journal of Geographical Information Science, 25(4), 595-613.

Li, Z., Hu, F., Schnase, J. L., Duffy, D. Q., Lee, T., Bowen, M. K., & Yang, C. (2017c). A spatiotemporal indexing approach for efficient processing of big array-based climate data with MapReduce. International Journal of Geographical Information Science, 31(1), 17-35.

Li, Z., Yang, C., Liu, K., Hu, F., & Jin, B. (2016). Automatic Scaling Hadoop in the Cloud for Efficient Process of Big Geospatial Data. *ISPRS International Journal of Geo-Information*, 5(10), 173.

Ling, F., & Foody, G. M. (2019). Super-resolution land cover mapping by deep learning. Remote Sensing Letters, 10(6), 598-606.

Ma, L.; Li, M.; Ma, X.; Cheng, L.; Du, P.; Liu, Y. A review of supervised object-based land-cover image classification. ISPRS J. Photogramm. Remote Sens. 2017, 130, 277–293.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity, Big Data: The Next Frontier for Innovation, Competition & Productivity, pp. 1-143, Available online at https://bigdatawg.nist.gov/pdf/MGI_big_data_full_report.pdf

Marciniec, M. (2017), Observing World Tweeting Tendencies in Realtime, available at https://codete.com/blog/observing-worldtweeting-tendencies-in-real-time-part-2, last accessed on May 3, 2019

Minsky, M. (1961). Steps toward artificial intelligence. Proceedings of the IRE, 49(1), 8-30.

Murphy JM, Sexton DM, Barnett DN, Jones GS, Webb MJ, et al. (2004) Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430: 768-772

Nvidia, C. U. D. A. (2011). Nvidia cuda c programming guide. *Nvidia Corporation*, *120*(18), 8.

OGC (2017), OGC announces a new standard that improves the way information is referenced to the earth

O'Leary, D. E. (2013). BIG DATA', THE 'INTERNET OF THINGS'AND THE 'INTERNET OF SIGNS. Intelligent Systems in Accounting, Finance and Management, 20(1), 53-65.

Ooi, B. C. (1987). Spatial kd-tree: A data structure for geographic database. In Datenbanksysteme in Büro, Technik und Wissenschaft (pp. 247-258). Springer, Berlin, Heidelberg.

Ooi, B. C., Tan, K. L., Wang, S., Wang, W., Cai, Q., Chen, G., ... & Xie, Z. (2015, October). SINGA: A distributed deep learning platform. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 685-688). ACM.

Planthaber, G., Stonebraker, M., & Frew, J. (2012, November). EarthDB: scalable analysis of MODIS data using SciDB. In Proceedings of the 1st ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data (pp. 11-19). ACM.

Ramsey, P. (2005). Postgis manual. Refractions Research Inc, 17.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., ... & Bloom, S. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. Journal of climate, 24(14), 3624-3648.

Robinson , 2012, The Storage and Transfer Challenges of Big Data, http://sloanreview.mit.edu/article/the-storage-and-transferchallenges-of-big-data/ (Last accessed Nov. 25, 2015)

Sabeur, Z, Gibb, R., Purss, M., 2019. Discrete Global Grid Systems SWG, Available at http://www.opengeospatial.org/projects/groups/dggsswg. Last accessed on March 13, 2019

Samet, H. (1984). The quadtree and related hierarchical data structures. ACM Computing Surveys (CSUR), 16(2), 187-260.

Shook, E., Hodgson, M. E., Wang, S., Behzad, B., Soltani, K., Hiscox, A., & Ajayakumar, J. (2016). Parallel cartographic modeling: a methodology for parallelizing spatial data processing. International Journal of Geographical Information Science, 30(12), 2355-2376.

Shvachko, K., Kuang, H., Radia, S., & Chansler, R. (2010, May). The hadoop distributed file system. In *MSST* (Vol. 10, pp. 1-10).

Stonebraker, M. (1986). The case for shared nothing. *IEEE Database Eng. Bull.*, *9*(1), 4-9.

Tan, X., Di, L., Deng, M., Fu, J., Shao, G., Gao, M., ... & Jin, B. (2015). Building an elastic parallel OGC web processing service on a

cloud-based cluster: A case study of remote sensing data processing service. Sustainability, 7(10), 14245-14258.

Tang, W., Feng, W., & Jia, M. (2015). Massively parallel spatial point pattern analysis: Ripley's K function accelerated using graphics processing units. International Journal of Geographical Information Science, 29(3), 412-439.

Taylor, Ian, Ian Wang, Matthew Shields, and Shalil Majithia. "Distributed computing with Triana on the Grid." Concurrency and Computation: Practice and Experience 17, no. 9 (2005): 11971214.

Taylor, R. C. (2010, December). An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics. In BMC bioinformatics (Vol. 11, No. 12, p. S1). BioMed Central.

Thain, D., Tannenbaum, T., & Livny, M. (2005). Distributed computing in practice: the Condor experience. Concurrency and computation: practice and experience, 17(2-4), 323-356.

Tschauner, H., & Salinas, V. S. (2006, April). Stratigraphic modeling and 3D spatial analysis using photogrammetry and octree spatial decomposition. In Digital Discovery. Exploring New Frontiers in Human Heritage. CAA2006. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 34th Conference, Fargo, United States(pp. 257-270).

Unat, D., Dubey, A., Hoefler, T., Shalf, J., Abraham, M., Bianco, M., ... & Fuerlinger, K. (2017). Trends in data locality abstractions for HPC systems. IEEE Transactions on Parallel and Distributed Systems, 28(10), 3007-3020.

VoPham, T., Hart, J. E., Laden, F., & Chiang, Y. Y. (2018). Emerging trends in geospatial artificial intelligence (geoAI): potential applications for environmental epidemiology. Environmental Health, 17(1), 40.

Vora, M. N. (2011, December). Hadoop-HBase for large-scale data. In *Proceedings of 2011 International Conference on Computer Science and Network Technology* (Vol. 1, pp. 601-605). IEEE.

Wang, F., Aji, A., Liu, Q., & Saltz, J. (2011). Hadoop-GIS: A high performance spatial query system for analytical medical imaging with MapReduce. Center for Comprehensive Informatics, Technical Report. Available at: http://www3. cs. stonybrook. edu/~ fuswang/papers/CCI-TR-2011–3. pdf (access 21 September 2015)(online).

Wang, L., Chen, B., & Liu, Y. (2013, June). Distributed storage and index of vector spatial data based on HBase. In *2013 21st international conference on geoinformatics* (pp. 1-5). IEEE.

Wang, L., Chen, B., & Liu, Y. (2013, June). Distributed storage and index of vector spatial data based on HBase. In 2013 21st international conference on geoinformatics (pp. 1-5). IEEE.

Wang, S. (2008, November). GISolve toolkit: advancing GIS through cyberinfrastructure. In Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems (p. 83). ACM.

Wang, S. (2010). A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. Annals of the Association of American Geographers, 100(3), 535-557.

Wang, S., & Armstrong, M. P. (2003). A quadtree approach to domain decomposition for spatial interpolation in grid computing environments. Parallel Computing, 29(10), 1481-1504.

Wang, S., Anselin, L., Bhaduri, B., Crosby, C., Goodchild, M. F., Liu, Y., & Nyerges, T. L. (2013). CyberGIS software: a synthetic review and integration roadmap. International Journal of Geographical Information Science, 27(11), 2122-2145.

Ward and Barker (2013). School of Computer Science, University of St Andrews, Undefined By Data: A Survey of Big Data Definitions UK

Whitby, M. A., Fecher, R., & Bennight, C. (2017, August). Geowave: Utilizing distributed key-value stores for multidimensional data. In *International Symposium on Spatial and Temporal Databases* (pp. 105-122). Springer, Cham.

Widlund, O. B. (2009). Accomodating irregular subdomains in domain decomposition theory. In Domain decomposition methods in science and engineering XVIII (pp. 87-98). Springer, Berlin, Heidelberg.

Wu, D., Liu, S., Zhang, L., Terpenny, J., Gao, R. X., Kurfess, T., & Guzzo, J. A. (2017). A fog computing-based framework for process monitoring and prognosis in cyber-manufacturing. Journal of Manufacturing Systems, 43, 25-34.

Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., ... & Roy, D. P. (2016). The global Landsat archive: Status, consolidation, and direction. Remote Sensing of Environment, 185, 271-283.

Xia, J., Yang, C., Gui, Z., Liu, K., & Li, Z. (2014). Optimizing an index with spatiotemporal patterns to support GEOSS
Clearinghouse. International Journal of Geographical Information Science, 28(7), 1459-1481.

Xie, D., Li, F., Yao, B., Li, G., Zhou, L., & Guo, M. (2016, June). Simba: Efficient in-memory spatial analytics. In Proceedings of the 2016 International Conference on Management of Data (pp. 10711085). ACM.

Yang, C., Huang, Q., Li, Z., Liu K., Hu, F. (2017). Big Data and cloud computing: innovation opportunities and challenges, *International Journal of Digital Earth*, 10(1),1-41.

Yin, D., Liu, Y., Padmanabhan, A., Terstriep, J., Rush, J., & Wang, S. (2017, July). A CyberGIS-Jupyter framework for geospatial analytics at scale. In Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact (p. 18). ACM.

Yin, J., Foran, A., & Wang, J. (2013, October). Dl-mpi: Enabling data locality computation for mpi-based data-intensive applications. In *2013 IEEE International Conference on Big Data* (pp. 506511). IEEE.

Yoon, G., & Lee, K. (2015). WPS-based Satellite Image Processing onWeb Framework and Cloud Computing Environment. Korean Journal of Remote Sensing, 31(6), 561-570.

Yu, J., Wu, J., & Sarwat, M. (2015, November). Geospark: A cluster computing framework for processing large-scale spatial data. In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems(p. 70). ACM.

Yue, P., Gong, J., & Di, L. (2010). Augmenting geospatial data provenance through metadata tracking in geospatial service chaining. Computers & Geosciences, 36(3), 270-281.

Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., ... & Ghodsi, A. (2016). Apache spark: a unified engine for big data processing. *Communications of the ACM*, *59*(11), 56-65.

Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., ... & Ghodsi, A. (2016). Apache spark: a unified engine for big data processing. Communications of the ACM, 59(11), 56-65.

Zhang, C., Di, L., Sun, Z., Eugene, G. Y., Hu, L., Lin, L., ... & Rahman, M. S. (2017, August). Integrating OGC Web Processing Service with cloud computing environment for Earth Observation data.

In 2017 6th International Conference on Agro-Geoinformatics (pp. 1-4). IEEE.

Zhang, C.; Sargent, I.; Pan, X.; Li, H.; Gardiner, A.; Hare, J.; Atkinson, P.M. An object-based convolutional neural network (OCNN) for urban land use classification. Remote Sens. Environ. 2018, 216, 57–70.

Zhang, J., Pennington, D. D., & Michener, W. K. (2006, May). Automatic transformation from geospatial conceptual workflow to executable workflow using GRASS GIS command line modules in Kepler.

In International Conference on Computational Science (pp. 912919). Springer, Berlin, Heidelberg.

Zhang, X., Song, W., & Liu, L. (2014, June). An implementation approach to store GIS spatial data on NoSQL database. In *2014 22nd international conference on geoinformatics* (pp. 1-5). IEEE.

Zhao, L., Chen, L., Ranjan, R., Choo, K. K. R., & He, J. (2016). Geographical information system parallelization for spatial big data processing: a review. Cluster Computing, 19(1), 139-152.

Zheng, M., Tang, W., Lan, Y., Zhao, X., Jia, M., Allan, C., & Trettin, C. (2018). Parallel Generation of Very High Resolution Digital
Elevation Models: High-Performance Computing for Big Spatial Data Analysis. In Big Data in Engineering Applications (pp. 2139). Springer, Singapore.

Zikopoulos, P., & Eaton, C. (2011). Understanding big data: Analytics for enterprise class hadoop and streaming data. McGraw-Hill Osborne Media.

Zikopoulos, P., Parasuraman, K., Deutsch, T., Giles, J., & Corrigan, D.
(2012). Harness the power of big data The IBM big data platform. McGraw Hill Professional.