EVALUATING THE CONVERGENCE OF HIGH-PERFORMANCE COMPUTING WITH BIG DATA, ARTIFICIAL INTELLIGENCE AND CLOUD COMPUTING TECHNOLOGIES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

YEŞİM DİLDAR KORKMAZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

JANUARY 2023

Approval of the thesis:

EVALUATING THE CONVERGENCE OF HIGH-PERFORMANCE COMPUTING WITH BIG DATA, ARTIFICIAL INTELLIGENCE AND CLOUD COMPUTING TECHNOLOGIES

Submitted by Yeşim Dildar Korkmaz in partial fulfillment of the requirements for the degree of **Master of Science in Information Systems Department, Middle East Technical University** by,

Prof. Dr. Banu Günel Kılıç
Dean, **Graduate School of Informatics** ____________________

Prof. Dr. Altan Koçyiğit
Head of Department, **Information Systems, METU** ____________________

Assoc. Prof. Dr. P. Erhan Eren
Supervisor, **Information Systems, METU** ____________________

Dr. Kerem Kayabay
Co-Supervisor, **TÜBİTAK** ____________________

Examining Committee Members:

Prof. Dr. Banu Günel Kılıç
Information Systems Dept., METU ____________________

Assoc. Prof. Dr. P. Erhan Eren
Information Systems Dept., METU ____________________

Assoc. Prof. Dr. Ebru Gökalp Aydın
Computer Engineering Dept., Hacettepe University ____________________

Date: 24.01.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Yeşim Dildar Korkmaz

Signature :

ABSTRACT

EVALUATING THE CONVERGENCE OF HIGH-PERFORMANCE COMPUTING WITH BIG DATA, ARTIFICIAL INTELLIGENCE AND CLOUD COMPUTING TECHNOLOGIES

Dildar Korkmaz, Yeşim

MSc., Department of Information Systems

Supervisor: Assoc. Prof. Dr. P. Erhan Eren

Co-Supervisor: Dr. Kerem Kayabay

January 2023, 69 pages

The advancements in High-Performance Computing (HPC), Big Data, Artificial Intelligence (AI), and Cloud Computing technologies have led to a convergence of these fields, resulting in the emergence of significant improvements for a wide range of fields. Identifying the state of development of technology convergence and forecasting promising technology convergence is critical for both academia and industry. That's why technology assessment and forecasting for HPC-Big Data-AI-Cloud Computing convergence is needed. The purpose of this thesis is to evaluate the convergence of HPC with Big Data, AI, and Cloud Computing technologies. In this thesis, a bibliometric analysis approach is conducted, including performance analysis and network analysis to identify the research trends and themes for the convergence of these technologies. The results of the analysis reveal a rapidly growing literature with a significant increase in research activities in this field in recent years. This study identifies key trends and patterns in the literature, including top published authors, most productive institutions, cited articles, and influential publications. In addition, research trends and thematic evolution analysis are carried out in this study. Existing studies that assess and forecast computational technologies do not consider the effect of convergence and do not apply bibliometric analysis in the field of HPC. This thesis provides valuable insights by identifying the bibliometric trends across the concept of technological convergence of HPC- Big Data-AI-Cloud Computing technologies.

Keywords: technology management, technology convergence, bibliometric analysis, high-performance computing, artificial intelligence

ÖZ

YÜKSEK BAŞARIMLI HESAPLAMA İLE BÜYÜK VERİ, YAPAY ZEKÂ VE BULUT HESAPLAMA TEKNOLOJİLERİNİN YAKINSAMASININ DEĞERLENDİRİLMESİ

Dildar Korkmaz, Yeşim
Yüksek Lisans, Bilişim Sistemleri Bölümü
Tez Yöneticisi: Doç. Dr. P. Erhan Eren
Tez Eş Danışmanı: Dr. Kerem Kayabay

January 2023, 69 sayfa

Yüksek Başarımlı Hesaplama (YBH), Büyük Veri, Yapay Zekâ ve Bulut Hesaplama teknolojilerindeki ilerlemeler, bu alanların yakınlaşmasına yol açarak, çok çeşitli alanlar için yeni ve önemli gelişmelerin ortaya çıkmasına neden olmuştur. Teknoloji yakınsamasının gelişim durumunu belirlemek ve gelecek vaat eden teknoloji yakınsamasını tahmin etmek hem akademi hem de endüstri için kritik öneme sahiptir. Bu nedenle, YBH-Büyük Veri-Yapay Zekâ-Bulut Hesaplama yakınsaması için teknoloji değerlendirmesi ve tahmini gereklidir. Bu tezin amacı, YBH' nin Büyük Veri, Yapay Zekâ ve Bulut Hesaplama teknolojileri ile yakınsamasını değerlendirmektir. Bu tezde, bu teknolojilerin yakınsaması için araştırma eğilimlerini ve temalarını belirlemek için performans analizi ve ağ analizi içeren bibliyometrik analiz yaklaşımı yürütülmektedir. Analiz sonuçları, son yıllarda bu alanda araştırma faaliyetlerinde önemli bir artış ile hızla büyüyen bir literatür olduğunu ortaya koymaktadır. Bu çalışma, en çok yayınlanan yazarlar, en üretken kurumlar, alıntılanan makaleler ve etkili yayınlar dahil olmak üzere literatürdeki temel eğilimleri ve kalıpları belirlemektedir. Ayrıca, bu alandaki araştırma eğilimleri ve tematik evrim analizi bu çalışmada yürütülmektedir. Hesaplama teknolojilerini değerlendiren ve tahmin eden mevcut çalışmalar, yakınsama etkisini dikkate almamakta ve YBH alanında bibliyometrik analiz uygulamamaktadır. Bu tez, YBH-Büyük Veri-Yapay Zekâ-Bulut Hesaplama teknolojilerinin teknolojik yakınsama kavramı genelindeki bibliyometrik eğilimlerini belirleyerek değerli bilgiler sağlamaktadır.

Anahtar Sözcükler: teknoloji yönetimi, teknoloji yakınsaması, bibliometrik analiz, yüksek başarımlı hesaplama, yapay zekâ

To my family

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor Assoc. Prof. Dr. P. Erhan Eren for his support and guidance throughout this research. His valuable knowledge, helpful advice, and insightful perspectives were instrumental in the completion of this thesis. I would like to thank my co-advisor, Dr. Kerem Kayabay for his support, suggestions, patience, and for sharing his knowledge during this research. I would also like to express my sincere gratitude to Prof. Dr. Mustafa İlhan Gökler. Without his encouraging attitude, I would not have been able to write this thesis.

Special thanks to my husband, Özgür Korkmaz, for his love and support. His endless encouragement is invaluable. Thank you for loving and believing in me.

Finally, I would like to thank my parents for their support throughout my life. I am particularly grateful to my sister and also my best friend, Yasemin Dildar, for her support and advice throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AWS	Amazon Web Services
HPC	High-Performance Computing
HPDA	High-Performance Data Analytics
ICT	Information and Communication Technology
IDC	International Data Corporation
IoT	Internet of Things
MPI	Message Passing Interface
RQ	Research Question
SLURM	The Simple Linux Utility for Resource Management
TORQUE	Terascale Open-source Resource and Queue Manager
WOS	Web of Science

CHAPTER 1

INTRODUCTION

1.1. Research Background

In today's interconnected and digital age, the volume of data produced by science and industry is increasing rapidly every year. According to the International Data Corporation (IDC), global data is projected to grow by 61% to reach 175 zettabytes by 2025 (Patrizio, 2018). It is essential to extract meaningful insights from these vast amounts of data for both scientific and business purposes. In the business world, for example, analyzing large data sets and gaining insights from them can help companies make better, more accurate, and efficient decisions in a timely manner and give them a competitive edge over their rivals.

As the volume of data continues to grow exponentially, the challenges associated with big data also increase. A key challenge is efficiently analyzing the data and arriving at a solution as quickly as possible. Emerging technologies like AI, big data, IoT, and cloud computing are having a major impact on business processes and operations (Alcácer & Cruz-Machado, 2019; Chun et al., 2019; Jagatheesaperumal et al., 2022; Nguyen et al., 2019). It's becoming increasingly crucial for organizations to adopt these technologies. Businesses turn to HPC capabilities to efficiently process, store, and analyze vast amounts of complex data on a large scale when faced with the challenge. The demanding workloads of big data are driving the convergence of these technologies (Amodei et al., 2016; Augonnet et al., 2009; Gorelick et al., 2017; Narasimhamurthy et al., 2019; Purawat et al., 2021). In addition to the growth in data volume, the increasing complexity of the AI techniques that organizations use to analyze data is also driving the convergence of these technologies (Huerta et al., 2020).

Technology convergence refers to the integrating and overlapping of different technological systems, resulting in new and enhanced products, services, and processes (W. S. Lee et al., 2015). Technology convergence is an active area of research and development in Information and Communication Technology (ICT), and it is expected to significantly impact future technological advancements (Jeong et al., 2015). It is crucial to identify patterns of technology convergence in order to address technology trends.

Combining HPC with technologies like AI, big data, and cloud computing helps solve complex computational problems. The convergence of HPC, big data, AI, and cloud computing can significantly impact various industries, including healthcare, finance, manufacturing, and transportation. HPC is an essential component of technology convergence, as it enables businesses to extract value from their data and make informed decisions in a rapidly changing technological landscape. HPC systems can be used to analyze and extract insights from big data, making it possible to process and analyze large datasets quickly. Big data and AI allow organizations to analyze and interpret vast amounts of data, enabling informed decision making and fostering innovation. AI and HPC frequently work together, as HPC systems are capable of providing the large amounts of data and computing power that AI applications require. In their paper, Yi and Loia (2019) provide HPC systems and applications for AI. HPC can be used to simulate complex phenomena, like molecular dynamic simulations for analyzing the physical movements of atoms and molecules (Thompson et al., 2022). Furthermore, cloud computing allows organizations to access scalable computing resources on demand, enabling them to quickly adapt to changing business needs (Buyya et al., 2009). HPC systems can be deployed in the cloud, making it possible for users to access HPC resources on a pay only for the number of resources they use.

1.2. Problem Statement

There has been previous research that has investigated the technology convergence in various computational technologies (Borgman & Rice, 1992; Chand Bhatt et al., 2021; Ejarque et al., 2022; Hussain et al., 2022; Kose & Sakata, 2019; Samadi, 2022; Singh et al., 2020; Sukumar et al., 2022). In addition, there have been multiple research studies focusing on forecasting computational technologies (Adamuthe & Thampi, 2019; Bildosola et al., 2017; Chand Bhatt et al., 2021; Lim et al., 2015; Shibata et al., 2018). However, several types of research are currently underway that are exploring the convergence of HPC, big data, AI, and cloud computing, and these projects have been proceeding independently of one another (Amodei et al., 2016; Georgiou et al., 2020; Gorelick et al., 2017; Madany et al., 2020; Nurmi et al., 2009; Thompson et al., 2022). To effectively understand the current state of technology convergence and anticipate promising technological trends, it is crucial to engage in technology assessment and forecasting for HPC-big data-AI-cloud computing convergence. This convergence has the potential to impact a wide range of fields greatly, and it is therefore crucial for both academia and industry to stay informed about the current state of development and forecast promising technological trends in these areas.

Currently, existing approaches in various computational domains need to effectively address the convergence of HPC with big data, AI, and cloud computing or utilize bibliometric analysis in the domain of HPC. In order to address this issue, a new approach is required that incorporates bibliometric analysis to assess the convergence of HPC, big data, AI, and cloud computing in technology forecasting. This approach will contribute to a better understanding of the interrelationships between these technologies, how they impact the field, and emerging trend topics.

1.3. Research Aim and Objectives

This thesis aims to present a comprehensive and current analysis of the convergence of HPC, big data, AI, and cloud computing to understand the underlying forces, impacts, and emerging trends in this field. The goal of identifying and synthesizing key themes and trends is to provide a clear and structured overview of this complex and rapidly evolving area. Objective of this thesis is to provide a structured and comprehensive overview of the convergence of HPC and AI, big data, and cloud computing. The following research questions (RQ) are the motivation for this thesis study:

RQ1. What are the driving forces for the convergence of HPC, big data, AI, and cloud computing?

RQ2. What are the effects of convergence on computing infrastructure, system architectures, and technologies?

RQ3. What are the various research areas and themes emerging from the convergence with regard to research areas?

1.4. Significance of the Study

This thesis's main contributions are: (1) Technological assessment of HPC, big data, AI, and cloud computing considering the driving forces, the effects of convergence on computing infrastructure technologies, the emerging research areas and themes. (2) The first study to apply bibliometric analysis in the field of HPC. Accordingly, (3) A bibliometric analysis of 3748 publications related to performance and network analysis to develop a new approach in the research field. (4) Identification of the most influential authors, institutions, countries, and most frequently cited publications in the field. (5) Identification and examination of the thematic map and thematic evolution through network analysis within the research field.

1.5. Thesis Structure

This thesis is organized into five distinct sections. Chapter 2 discusses the theoretical foundations of the research related to the convergence of HPC and AI, big data, and cloud computing technologies. In Chapter 3, the research approach and bibliometric methodology are provided. The procedure for bibliometric analysis, analysis of the results obtained, and evaluation are given in Chapter 4, with the research and future implications. Conclusion is provided in Chapter 5, which includes the contributions and future work.

CHAPTER 2

LITERATURE REVIEW

2.1. Background Information

This section presents a background information on the technologies involved in the study.

2.1.1 HPC

HPC is the collection of many separate servers (computers), named nodes, clustered together, and these clusters process distributed workloads using parallel processing (Hager & Wellein, 2010). HPC clusters are a collection of a number of compute servers that are working closely together. This is done because workloads that require high computational intensity and complex computation are provided at high speed and in parallel.

HPC involves the use of supercomputers and other specialized systems to perform a wide range of compute-intensive tasks, including scientific exploration, simulation, advanced engineering designs, and complex mathematical analysis. A lot of industries use HPC, such as manufacturing, energy, financial services, healthcare, media and entertainment, logistics and supply chain management, and weather forecasting, to gain valuable insight. With such a wide application area, designing an efficient infrastructure according to the needs requires having knowledge about the applications to be used.

With the increasing complexity of scientific and engineering problems, the need for processing power in the computing landscape has increased. Along with these developments, a new era for HPC, named Exascale, has been led (Fiore et al., 2018). Scientific and engineering problems in both simulation and data analysis have been exceeding petascale and are approaching exascale computing needs. Exascale computing refers to the ability to perform a quintillion (one billion billion) operations per second. It is significantly faster and more powerful than petascale computing, which is 1,000 times less powerful.

Application Areas of HPC

Scientific and engineering applications often require more computational power than standard computers can provide in order to effectively handle complex tasks involving large amounts of data. These tasks may require greater processing speed, large number of repeated calculation steps, memory capacity, or other resources to be completed efficiently. HPC provides a solution to address this demanding computing needs. HPC is used in a variety of fields to address complex computing tasks that require significant processing power and other resources. These tasks may include simulations, data analysis, and other types of computations that require significant computational resources. An example of how HPC is used is in the acceleration of simulations involving large amounts of numerical data, such as those used in computational fluid dynamics. These simulations can be time-consuming and require significant computational resources, making HPC an essential tool for efficiently completing such tasks. HPC reduces the amount of time required for computation and allows for more design iterations to be completed. This can improve the efficiency and effectiveness of the design process (Lange et al., 2021). Using HPC in molecular dynamics simulations can be useful because it allows for the rapid analysis of the movements of atoms and molecules (Chiu & Herbordt, 2010). It allows multiple computations to be carried out simultaneously. This can improve the speed and efficiency of the simulation process. Furthermore, graphical visualizations including image processing, requires great computational power due to a large amount of data and the need to generate results in a short time (Carnielli et al., 2021). Therefore, HPC has become important for such research areas. Another application of HPC is in the simulation of artificial neural networks (Strey, 2001). These networks can be complex and require significant computational resources to model and analyze. Another application area where HPC is used is self-driving vehicles. The self-decision of vehicles is achieved through complex machine-learning algorithms powered by HPC. HPC is also strong in weather and climate research. Prediction of climate catastrophes and extreme weather phenomena are better with the use of HPC technologies. HPC has applications in the computational finance area, such as development of complex computational financial models (Hong et al., 2010). In addition to their use in other fields, HPC systems can also be used to perform various computation-intensive tasks, such as climate modeling, financial modeling, drug discovery, genomics, oil and gas exploration, neuroscience, and psychology.

HPC Architecture

An HPC cluster architecture typically includes components that need to work together efficiently and scale together (Yang & Guo, 2005), these are login node, compute node, accelerator node, storage node, network switch, and application software.

1. Login Node

It is identified as a head node. It is one of the most important parts of a cluster and provides cluster management and scheduling services (Sterling et al., 2017). HPC login node authenticates the user and acts as an access point to the system. There are some basic tasks, such as data uploading, file management, software compiling, and job management. Clusters are generally composed of a single head node, but it may be risky for large-scale use. If too many users request resources, there may be a problem of exceeding the resource. For this reason, multiple head nodes may be a better application. This makes user access easier by providing multiple gateways to access the cluster.

2. Compute Node

The highly complex data-intensive tasks which are called HPC workloads are processed across multiple nodes in parallel (Sterling et al., 2017) . Each HPC workload requires different levels of CPU in order to complete tasks. Works can be dissected in smaller parts and then run. For parallel programming, different models can be used, such as Message Passing Interface (MPI) and OpenMP.

The compute node is responsible for executing workloads using the resources available locally, such as CPU, GPU, and FPGA. Clusters are groups of servers that are connected through a network and consist of hundreds or thousands of compute servers. A cluster contains at least 4 nodes or 16 cores. Each individual computer within the cluster is referred to as a node. These nodes work in parallel to deliver HPC. The resulting answers are communicated with the other calculations and stored in storage. HPC applications need powerful CPUs for high-speed tasks and powerful GPUs for graphics-intensive tasks. With the development of accelerators that are connected to the main CPU, applications got much faster. NVIDIA made major innovations in this area by making the GPU programmable as a General Purpose GPU.

3. Accelerator Node

The accelerator node provides CPU, memory, and network bandwidth capabilities (Sterling et al., 2017). These nodes may contain one or multiple accelerators. In smaller HPC clusters, each node may contain an accelerator. It adds performance to the cluster on some HPC mixed workloads.

4. Storage Node

Storage is an essential component for the faultless and effective running of HPC applications (Sterling et al., 2017). A significant amount of data is required to perform simulations like weather forecasting or protein structure alignment. Applications have to read and write huge amounts of data in parallel. Storage systems are important to provide this parallel I/O and scale together. During this process, high performance, low latency, and high-bandwidth storage infrastructure are needed. For efficient I/O, the orchestration of the servers needs to be simple for the system administrator or developer. Lustre File System, a kind of parallel and distributed file system, is commonly used in complex HPC storage environments.

Lustre File System

Lustre is a parallel file system that is open-source and specifically designed for use in HPC clusters. (Donovan et al., 2003). Lustre is composed of three primary components; these are metadata server, object storage server, and clients. Lustre file system separate data services and metadata services in order to allow multiple users to access different files simultaneously, which improves performance compared to traditional file systems. The architecture includes a series of I/O servers known as Object Storage Servers (OSS), Metadata Servers (MDS), Object Server Target (OST) and Metadata Targets (MDT).

5. Network Switch

The software runs on multiple nodes and accesses storage to read and write data, which requires high bandwidth and low latency for optimal performance. In order for the system to function properly, communication between all nodes and the storage system must be continuous (Andújar et al., 2022). Due to the high volume of data and the need for fast communication between nodes and the storage system, it is essential that the interconnection between these components be efficient. InfiniBand (J. Liu et al., 2003), Ethernet, and Omni-Path (Birrittella et al., 2016) are the most commonly used for interconnection. Ethernet, as a low-cost solution, is used to connect smaller clusters. InfiniBand and Omni-Path are preferred for larger workloads as they provide high throughput and low latency.

6. Application Software

An HPC cluster needs both basic infrastructure and software to control the infrastructure in order to run applications. Software is needed for the allocation of a certain number of servers, network bandwidth and storage (Sterling et al., 2017). Computing systems are growing larger and more complex with the convergence of the HPC, AI, and big data. Therefore, the software is essential for processing many simultaneous tasks, such as simulating physical processes, networking, executing large volumes of I/O traffic, and reading and writing data in parallel.

7. HPC Software Components

The compute nodes in the HPC are not accessible directly. It can run under the control of a batch system. A batch system, which is in charge of receiving and executing jobs, is made up of a job scheduler and a resource manager. Job scheduler like Moab (Alemzadeh, 2009) handles the queuing of jobs. The job waits in the queue until the requested resources are available. Resource managers like Torque (Staples, 2006) which is a component within the batch system identifies jobs, select resources, and decide the running time of the job. These two software modules work together. The batch jobs and resource usage are managed by working these components together.

MPI

MPI is an open library standard used to program parallel computers (Gropp et al., 1996). MPI Application Programming Interface is independent of programming language. It can be used in many programming languages like C, Fortran, Python. MPI provides writing practical, flexible, and efficient parallel programs. The last version of the standard is MPI-4.0. MPI provides portability (because it can be implemented for almost every distributed memory architecture) and standardization (being a generally accepted industry standard).

Another alternative for portable message passing is OpenMP (Dagum & Menon, 1998). It is an API for developing parallel programs on shared memory architectures. Unlike MPI which uses a distributed memory model, OpenMP uses a shared memory model.

a. Operating System

According to TOP500 list, that ranks the most powerful supercomputers in the world, Linux is the most used operating system, followed by CentOS and Cray Linux Environment respectively (*List Statistics | TOP500*, n.d.).

b. HPC Resource Managers

HPC scheduling, also named batch scheduling, is a software that is responsible for allocating jobs to resources according to site policies and resource availability. It runs on a cluster and decides who uses which machines at what time. With the growing processing power of supercomputers, the scheduling of these jobs has started to be more crucial. Traditional HPC job schedulers are simple. With the emergence of new hardware resources and the increase of compute-intensive and data-intensive workloads, more efficient schedulers are needed (Fan, 2021). Some well-known job schedulers are Slurm, Torque, and PBS.

- **Slurm:** Slurm, or the Simple Linux Utility for Resource Management, is a resource management system that is appropriate for use in both small and large Linux clusters (Yoo et al., 2003). Slurm is a cluster management and job scheduling system that is open source, fault-tolerant, and capable of handling a large number of nodes. Slurm is able to perform three main tasks. It assigns users to compute nodes, either exclusively or non-exclusively, and provides a framework for launching and monitoring jobs on allocated nodes. Additionally, it manages the queue of pending jobs. Slurm uses a centralized manager called slurmctld to monitor resources and manage the system. In the event of a failure, there is a backup manager to take over. Each computes server has a slurmd daemon that is responsible for launching and managing jobs and a slurmdbd daemon that collects accounting information for multiple Slurm-managed clusters and stores it in a single database. Slurm can be interacted with through a REST API using the REST API daemon (slurmrestd).

- **Torque:** Terascale Open-source Resource and Queue Manager (TORQUE) is an open-source resource manager that controls batch jobs and manages distributed compute nodes (Staples, 2006). It can integrate with Moab Workload Manager and Maui Cluster Scheduler to increase system utilization and optimize performance of the application. TORQUE is managed by Adaptive Computing and developed from the open-source Portable Batch System (OpenPBS) project. Due to the improvements, it provides over PBS, in the areas of scalability, reliability and functionality, it is highly preferred for both academic and commercial uses. Some common Torque commands include qsub (for submitting a job), qstat (for monitoring the status of a job), and qdel (for deleting a job). Commonly used Torque commands are qsub (submit job), qdel (deleting job), and qstat (monitor the status of the job).

- **PBS:** PBS is another important resource manager for HPC systems. It allocates the computational tasks according to available computing resources (Henderson, 1995). From big HPC workloads to small ones, PBS is designed to optimize efficiency, improve productivity and simplify administration for clusters. Three versions of PBS are OpenPBS Torque (a fork of OpenPBS that is maintained by Adaptive Computing Enterprises), PBS Pro (maintained by Altair Engineering).

c. Modules

Module systems in the supercomputer provide the use of different software in a controlled manner (Sochat & Scott, 2021). All these software packages need different settings that can negatively affect each other and cannot be used in parallel on the same system. All these software system settings are provided by the module system.

d. Compilers and libraries

The source code is written and compiled in mainly three classic HPC languages; these are C/C++ and Fortran. For the development of parallel applications, libraries such as MPI, OpenMPI (Dagum & Menon, 1998) and OpenCL (Stone et al., 2010) are used.

2.1.2 Big Data

When data is too large, fast, or complex to be processed with traditional data processing tools and techniques, it is referred to as big data (M. Chen et al., 2014). It can come from a variety of sources, such as IoT devices, social media, customer databases, and transaction processing systems. Big data provides valuable insight and enables the organization to make better decisions. On the other hand, it also has significant challenges in terms of storage, processing, and analysis. To effectively analyze and leverage big data, organizations use technologies such as machine learning and artificial intelligence, as well as specialized tools such as HPC systems and technologies (such as distributed computing frameworks, data lakes, and data warehouses). Big data has applications in many industries, for example healthcare, finance, manufacturing, and retail. Big data is a rapidly developing field that has great importance in the transformation of industries and in making more informed decisions.

HPDA

High-Performance Data Analytics (HPDA) involves tasks that use highly complex algorithms to discover patterns and trends in large datasets and require HPC resources to process them (Dube et al., 2021). The term HPDA was coined by IDC to represent the combination of the two terms, HPC and big data, as the lines between these two terms continue to blur (Grandinetti et al., 2018). With HPDA, it is possible to tackle today's most complex scientific and business challenges by finding patterns in large datasets and identifying trends. HPDA enables enterprises to gain more agile, productive, and competitive advantage, by managing computational and data intensive workflows. Over time, the development of HPDA solutions has led businesses to adopt HPC technology in order to gain a competitive advantage. The global market size for HPDA is estimated to be around 25.2 billion US dollars, and it is expected to reach 82 billion by 2022.

2.1.3 AI

According to Kaplan and Haenlein (2019), AI refers to a system's ability to accurately interpret external data, learn from that data, and use those learnings to adapt and achieve specific goals and tasks. AI technologies offer more efficient and accurate computing methods in a variety of fields, including computer science, engineering, psychology, and business. They are used in a range of applications, from personal assistants like Apple Siri and convenient retail experiences like Amazon Go, to more complex and potentially controversial areas like self-driving cars and autonomous weapons (Wang & Siau, 2019). Among other major technological trends in the last 10 years, the emergence of AI is one of the most important driving forces because it is an important innovation for both HPC suppliers and consumers (Feldman, 2020).

AI is a term that refers to a range of technologies, including machine learning and deep learning. Machine learning is an application of AI that learns from data and then applies what they have learned to make accurate predictions. Deep learning is considered the subfield of machine learning based on artificial neural networks concerned with algorithms inspired by the structure of the brain (Ongsulee, 2017).

The growth of AI has been driven by advances in machine learning techniques, which require a large amount of training data in order to improve the quality of predictions. Therefore, computers with higher computational power are needed. This increased computational power has revealed the need to distribute workloads to multiple machines, which is called Distributed Machine Learning (Verbraeken et al., 2021). For faster executions and large-scale training, datasets distributed machine learning platforms are available such as MapReduce (Gillick et al., 2006) and Apache Spark (Meng et al., 2015). Distributed deep learning as a sub-area of general distributed machine learning, is used when the model training process wanted to speed up using multiple GPUs. Sevilla et al. (2022) examine the three distinct eras, and three distinct trends of the compute, which is one of the main factors guiding the progress of machine learning. The article notes that before 2010, the computing power used for training increased in accordance with Moore's law. It is stated that with the emergence of deep learning in the early 2010s, the scaling of compute has accelerated and doubled every 6 months. TensorFlow (Abadi et al., 2016), PyTorch (Paszke et al., 2019), and Mxnet (T. Chen et al., 2015) are some of the libraries that offer distributed training. DDL frameworks such as Horovod and Tarantella develop approaches for distributed training of existing models. The goal is to make distributed training fast, easy, and portable. Both frameworks are open-source and implemented in Python and C++, and they support TensorFlow and CUDA.Veroneze Solórzano and Mello Schnorr (2022) present the comparison and evaluation of two distributed training frameworks.

2.1.4 Cloud Computing

According to the National Institute of Standards and Technology (NIST), cloud computing is a model that allows users to access a shared pool of configurable computing resources (such as networks, servers, storage, applications, and services) over the internet on an as-needed basis. These resources can be quickly and easily provisioned and released with minimal effort or interaction with the service provider. (Mell & Grance, n.d.). Cloud computing is the delivery of computing services over the internet with a pay-as-you-go model (Sunyaev, 2020). Without buying a physical data center or server, users can rent servers, storage, databases, networking, software, or analytics. Many different organizations are using the cloud for different use cases, such as file storage, backing up data, disaster recovery, software development, testing new projects, and big data analytics. There are three main types of cloud computing, these are Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).

Edge computing is a new way of processing data closer to where it is generated (Satyanarayanan, 2017). It enables faster data processing, improved response times, and increased bandwidth. According to McKinsey report published in 2022, the networks of the future will consist not only of traditional cloud data centers but also of computational resources located at edge nodes closer to end users (*McKinsey Technology Trends Outlook 2022 | McKinsey*, n.d.). This is intended to improve data latency and increase data autonomy. AI technology is frequently used in edge computing to improve performance by bringing computational resources closer to end users and enabling faster processing of data and better response times (Hao et al., 2021).

A prominent concept in cloud environments is containers. Cloud computing requires high portability. Containerization as a virtualization technology ensures this. Containers are packages of software that include code and all its dependencies, allowing the application to run quickly and reliably in any environment. Unlike Virtual Machines, which virtualize the entire machine down to the hardware layers, containers only virtualize software layers above the operating system level. (Xing & Zhan, 2012). Container orchestrators are used to managing large numbers of containers. There are many different container and orchestration technologies available for use with HPC applications. Docker, Singularity, and Shifter are the most widely known container solutions. Container orchestration tools are Docker Swarm, Kubernetes, Google Container Engine, Amazon ECS, Azure Container Services (Beltre et al., 2019).

2.1.5 Cloud-Native Technologies

Cloud-native is a term that means applications that are built to run in the cloud as opposed to on-prem. Cloud-native technologies, as defined by the Cloud Native Computing Foundation (CNCF), are designed to allow organizations to develop and operate scalable applications in various cloud environments, including public, private, and hybrid clouds (Kratzke & Quint, 2017). These technologies enable organizations to effectively operate in modern, dynamic environments. The terms "cloud-native" and "cloud-based" are often used interchangeably, but they refer to different types of applications. Cloud-based applications are designed to run on cloud platforms and take advantage of the benefits of the cloud, such as elasticity and scalability. Cloud-native applications, on the other hand, are specifically designed for the cloud and are built using cloud-native principles, such as microservices and containerization. Cloud-native applications are designed to run on public cloud infrastructure, such as Amazon Web Services (AWS), Microsoft Azure (Azure), or Google Cloud Platform (GCP). There are many cloud-native applications, such as Netflix movie streaming service, Facebook, Twitter, Google's cloud collaboration. Cloud-native supercomputing is a type of HPC that leverages the capabilities of cloud computing services, such as virtually unlimited resources and the ability to scale elastically on demand. For instance, Kubernetes, defined as a cloud-oriented operating system, is a growing community and platform developed for managing clusters of containerized applications and services (Zhou et al., 2021). It automates the deployment and management of cloud-native applications. It provides a framework for running distributed systems flexibly.

2.2. HPC, Big Data, AI, and Cloud Computing Convergence

For the last decade, innovations in technology have increased the demand for computing power. As the volume of data is growing, the need for processing power is increased in order to obtain a solution urgently. Therefore, in many industries, the use of HPDA is increasing. Handling large amounts of data in real-time applications with a fast response time can be challenging. This is a key factor that has led to the convergence of HPC and big data (Usman et al., 2018).

After the potential in HPC has been discovered, its use became widespread not only in scientific simulation and computing, but also in AI. Deep learning is growing fast and achieving tremendous results in many fields. Deep learning applications mainly require hardware platforms, because of the large number of neural network parameters. Therefore, the utilization of HPC systems leads to important results. HPC systems are suitable for distributed deep learning as a method because it provides parallel execution over many devices (GPUs/TPUs/servers) and reduce training time. For instance, deep neural networks use distributed training to avoid memory capacity limitations when training large models. HPC-enabled AI has been used to optimize supply chains, manufacturing, logistics, and healthcare to resolve any problem(Yi & Loia, 2019).

As the demand for HPDA increases, HPC applications have shifted towards data-driven. This has resulted in the convergence of HPDA, AI, and HPC. The convergence of large-scale data analytics, AI technologies, and HPC as a trend is due to the extreme computational power of HPC (Antoniu et al., 2021). Federated use of these technologies provides conveniences such as easier access, greater flexibility, lower response times, and more efficient use of essential resources. Besides, the convergence of these technologies provides incredible advances in many application areas. For example, the complex workflows that need data-intensive computing such as autonomous vehicles, smart buildings, weather forecasting, and smart agriculture require integrated usage of HPC, big data, AI, and cloud computing. In their book, Terzo and Martinovič, (2022) explain with examples of how different fields, such as agriculture, health, or mechanical engineering, can benefit from this convergence.

In a study by Georgiou et al. (2020) a new model for analyzing precision agriculture and livestock data using hybrid data analytics was proposed for the Cybele project, which is funded by the European Union. The new model combines HPC, big data, and cloud technologies to perform data analytics for precision agriculture on supercomputers. This approach allows for the efficient analysis of large amounts of data in order to support decision-making in the field of precision agriculture.

The LEXIS (Large-scale Execution for Industry & Society) project is another example of the convergence of HPC with cloud computing, and big data (Hachinger et al., 2022). Its goal is to utilize the power of scientific supercomputing to analyze large amounts of data in the industrial and enterprise sectors. As part of the project, a platform has been developed for processing data- and compute-intensive workflows, enabling organizations to handle large amounts of data and perform complex calculations.

Another study about convergence is the Expanse system (Strande et al., 2021) at the San Diego Supercomputer Center (SDSC), offering a platform for composable systems. It is a National Science Foundation (NSF)-funded supercomputer. Studies are conducted on the use of composable systems such as scientific computing, AI, and sensor data integration. Their aim is to make HPC more extensible and make future workloads more flexible by integrating cloud, AI, and HPC.

Madany et al. (2020) introduce an automated neuroscience reconstruction framework called NeuroKube that utilizes artificial intelligence (AI) and Kubernetes as a cloud-native computing engine to perform large-scale processing and labeling of neuroimages. NeuroKube is designed to help researchers and scientists more easily analyze and understand complex neuroimaging data sets. The study shows the effective use of Kubernetes for large-scale computing. Instead of using traditional data center-driven computers, the study leveraged the Pacific Research Platform (for high bandwidth, interconnected, optic fiber network) and CHASE-CI (for distributed node hardware) together.

TemPredict is a research project that uses a big data analytical platform to study and track personalized multimodal data related to COVID-19 (Purawat et al., 2021). The platform is designed to be used for AI research on COVID-19 physiology and includes features for collecting data and delivering results to clinical users. The developed infrastructure includes HPC and Kubernetes-managed GPU and CPU clusters for speed and efficiency of execution and accuracy of prediction of machine learning algorithms.

In another paper (Milojicic et al., 2021), a new vision for HPC is presented. Their main hypothesis is that future HPC requires diversification across hardware and software to ensure scaling performance and maintain energy efficiency. Increasing this heterogeneity affects the whole HPC stack, from infrastructure to delivery models. They discuss the convergence of data-intensive workloads and science. It is mentioned that combining HPC, analytics, and machine learning technologies and deploying them in distributed networks across various locations (such as the edge, supercomputers, and the cloud) is necessary in order to achieve convergence.

The authors of (Eicker et al., 2013) present the DEEP (Dynamical Exascale Entry Platform) project. It aims for implementing an energy-efficient system architecture that fits HPC and HPDA workloads. The project aims to develop new software and hardware technologies to enable exascale computing. The project involves creating a hardware platform that consists of two parts: a standard HPC cluster and a cluster of many-core processors called Booster. It aims to co-designing the software and programming environment of the upcoming European exascale systems.

2.2.1 Precision in HPC

A convergence of the computing application of HPC and AI is becoming more common. In computer science the precision of a numerical quantity is measured in bits, and computer performance is measured in floating point operations per second (FLOPS). Some standard precision formats are half (16-bit), single (32-bit), and double (64-bit). It can be important to use high floating-point precision for scientific computing and simulation tasks. Therefore, HPC simulations generally have been implemented in higher precision. On the other hand, in most cases AI applications do not need to be as precise as single or double precision. For example, data science and big data use mixed precision. For tasks such as image classification and speech recognition single or half-precision formats are required. Using lower precision has become more important because it allows for faster processing, as single precision can be processed at twice the speed of double precision (Abdelfattah et al., 2021). Using lower precision instead of double precision can increase the communication bandwidth and result in acceleration of the computing performance. Because of the trade-offs between accuracy and performance, computer architectures have been supporting mixed precision (Parasyris et al., 2020). Increasing precision may lead to a decrease in performance, and in other cases, it may require more memory. For example, the US Exascale Computing Project is attempting to create technology that takes advantage of computing power available in low precision and adapts the communication format to the specific needs of the application with mixed precision (Abdelfattah et al., 2020). Another example, Lang et al. (2021) study shows that reducing the precision of the Integrated Forecasting System (IFS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) forecasting model to single precision provides significant computational savings without reducing forecast accuracy. The world leading companies, such as NVIDIA, AMD and Intel have started designing accelerators according to the demand for high computing power in low-precision formats. Consequently, mixed precision is expected to become more common in scientific applications as HPC systems become increasingly heterogeneous (Parasyris et al., 2020).

2.2.2 Benchmarks for HPC and AI

Benchmarking for HPC is important for measuring the performance of parallel programs in order to make optimal use of software and hardware on HPC resources. Evaluating HPC systems and workloads and identifying their bottlenecks is difficult. High-Performance Linpack is one of the most important HPC benchmarks in the HPC community (Sterling et al., 2017). The TOP500 project (*List Statistics | TOP500*, n.d., p. 500), launched in 1993, ranks and detects the most powerful HPC systems in the world. The Linpack Benchmarks are used to measure the floating-point computing power of the systems. HPC systems are ranked by their ability to solve a set of linear equations. Another important benchmark is Graph500 which is focused on data-intensive workloads unlike floating-point-intensive computations as in High-Performance Linpack (Ang et al., 2010).

More complexity arises when deep learning workloads are run in HPC systems using multiple parallel and accelerator programming models, because DL workloads require heterogeneous computing platforms including CPU, GPU and various domain specific processors. So, convergence of deep learning workloads and HPC raise more challenges in HPC AI benchmarking. For this reason, different metrics are emerging for assessing resource capability and predicting performance quickly and accurately.

Jiang et al. (2019) have developed a benchmark suite called HPC AI500 for evaluating the performance of HPC systems running deep learning workloads. The benchmark suite includes a set of metrics for evaluating the accuracy, performance, power consumption, and cost of HPC-AI systems.

In another study, a comprehensive benchmark suite is defined, whose name is IO-500 (Kunkel et al., n.d.). The IO-500, which is similar to the TOP500 list, was created to monitor the performance increase of compute architectures over the years. Unlike other benchmarks, this benchmark provides documentation and sharing of best practices within the community.

With the increased importance of power in supercomputing, the energy efficiency of supercomputers is becoming important as a design constraint. To raise the awareness of greenness, the Green500 list is ranked among the TOP500 list of supercomputers in terms of energy efficiency. Different metrics, workloads, and methodologies are used for ranking.

2.2.3 Other Projects

There are also some other projects developed for the use of HPC resources and other e-infrastructures together. It is expected that HPC systems will eventually transform into e-infrastructure service providers that are based on various types of computing and storage resources (Haus et al., n.d.). This is because the data from scattered scientific instruments and sensors need to be processed, analyzed, and incorporated into simulations in order to gain scientific insights and make innovative predictions. Therefore, different initiatives are emerging. One of these initiatives is Fenix (Alam et al., 2020). Fenix aim is to establish suitable e-infrastructure governance, collaborating with HPC centers. The goal is to support a variety of science and engineering communities. This involves connecting and integrating various data repositories, scalable supercomputing systems, and private cloud instances. E-infrastructure has certain key characteristics, such as interactive computing services, flexible access to scalable compute resources and federated data infrastructure. Another initiative is Gaia-X (Braud et al., 2021). Its goal is to establish a sharable data ecosystem in a trustworthy environment. The infrastructure consists of network and interconnection providers, cloud services providers, edge locations and HPC. Mainly identified specific data spaces are industry 4.0, energy, health, finance, and smart living. It is foreseen that smart third-party services such as AI can also be added to the architecture. Another project is Sage (Narasimhamurthy et al., 2019). The goal of the Sage project is to build an Exascale system prototype by addressing the evolving overlap between big data analytics and HPC. Sage system includes tools for data-management, programming models and data analytics methods.

As previously mentioned in the literature review, there are several projects focusing on the convergence of HPC, big data, AI, and cloud computing that are progressing independently of one another. It is important for both academia and industry to understand the current state of development of technology convergence and predict promising technology trends (J. Kim & Lee, 2017). To provide a structured and concise view and guide researchers and practitioners, we should understand the past and predict the future. That's why we need technology assessment and forecasting for HPC-Big Data-AI-Cloud Computing convergence.

2.3. Studies Assessing and Forecasting Computational Technologies

Technology assessment and technology forecasting are methodologies that examine the early identification and evaluation of the ultimate effects of technological changes and attempt to understand and evaluate the probability of various future technological developments (Haleem et al., 2018). In the literature, various studies discuss the forecasting of computational technologies. The main focus of this section is to identify these researches.

Adamuthe and Thampi (2019) investigate the growth of six computational technologies, namely mainframes, minicomputers, cloud computing, cluster computing, grid computing, and autonomic computing. They use patent analysis to forecast technology trends by applying trend projection methods and growth curve methods. The life cycle of computational technologies is profiled on time series data of patent and research papers. The goal of their research is to find trend projections and suitable growth models for technology forecasting.

(Chand Bhatt et al., 2021) aim to understand how blockchain and Industry 4.0 technologies (IR 4.0) are converging and how they are being used together. It uses a bibliometric approach to analyze trends in the convergence of these technologies and forecast their potential future applications.

Lim et al. (2015) use technology forecasting to study the technological progress of supercomputer development in terms of both energy efficiency and performance, taking into account the trade-offs between power consumption, multi-core processors, and maximum performance. Technology forecasting is made using data envelopment analysis, which is data-driven forecasting technique to measure technological progress.

Bildosola et al. (2017) aim to provide a relevant profile and a technology roadmap for cloud computing technology based on technological forecasting methods. The first phase of the research is presented in this paper. The main purpose is to raise awareness of cloud technologies among enterprises. This study utilizes a new approach based on four categories of technological forecasting methods, which are statistical methods using bibliometrics and data mining, trend analysis, technological road mapping, and expertise.

In another research (Shibata et al., 2018), trend analysis of the technology development on FPGA (Field-Programmable Gate Array) is made using patent analysis with link mining method. Researchers use the US patent data filed by FPGA suppliers and employ patent classifiers for making the technology structure graph.

Forecasting of trends in computer technology at the national laboratories is proposed (Peskin, 1980). A group of scientists and computer professionals are organized to prepare an annually updated technology forecast. Groups charged with forecasting try to take as many objective readings as possible. Group members meet periodically over several months and the papers are critically reviewed by the entire group. Finally, group reaches a consensus and the final forecasting report is revealed.

2.4. Literature Summary and Knowledge Gaps

The literature review was presented in Chapter 2. Section 2.1 provided background information on the technologies involved in the research that is driving the confluence. The studies and projects about the convergence of HPC, big data, AI, and cloud computing were covered in Section 2.2. The studies that assess and forecast the computational technologies were reviewed and presented in Section 2.3.

Technological convergence can be defined as a phenomenon that creates new value, leads to new products and services, and triggers technological change with the increasing exchange of information between different technologies and industries (W. S. Lee et al., 2015). Therefore, it is very important to identify patterns of technology convergence and forecast new convergences in order to overcome future challenges (T. S. Kim & Sohn, 2020). Previous studies have attempted to identify technology convergence in a variety of fields. For example, the technology convergence in the electric vehicle domain has been studied (Feng et al., 2020). Research has been conducted on the technological convergence assessment of smart factory (Hussain et al., 2022). The convergence of technology among the various fields of robotics has been analyzed (Kose & Sakata, 2019). There are several methods that have been utilized in the literature to identify the convergence of various technologies. For instance, research using patent information on technology convergence has been conducted (Caviggioli, 2016; Gauch & Blind, 2015; J. Kim & Lee, 2017; Song et al., 2017). T. S. Kim and Sohn (2020) have proposed using semantic analysis and bibliometric-based indicators to predict the emergence of a converged technology. C. Lee et al. (2021) have forecasted convergence using link prediction. Zhu and Motohashi (2022) have suggested a new semantic method by demonstrating how a graph convolutional network model can be used to track technology convergence.

The convergence of HPC with big data, AI, and cloud computing has gained significant attention in recent years as a means to accelerate scientific research and industrial innovation. However, despite the various publications and projects related to convergence of these technologies (HPC-big data-AI-cloud computing) that have been investigated in the literature, there is a lack of a comprehensive study.

The following knowledge gaps were identified by reviewing the literature:

1. The literature lacks a comprehensive study of the driving forces behind the convergence of HPC, big data, AI, and cloud computing.

2. There is no comprehensive study that considers the effects of convergence on computing infrastructure, system architecture, and technologies.

3. There is an insufficient study that examines the effects of convergence on emerging research areas and themes.

4. This is the first study to utilize bibliometric analysis in the field of HPC.

CHAPTER 3

RESEARCH METHODOLOGY

3.1. Research Questions and Objectives

This study is an attempt to provide a comprehensive and up-to-date analysis of the convergence of HPC, big data, AI, and cloud computing, with a focus on understanding the driving forces, impacts, and emerging trends in this field. The goal is to provide a clear and structured overview of this complex and rapidly evolving field by identifying and synthesizing key themes and trends. This thesis aims to address the following three RQ:

RQ1. What are the driving forces for the convergence of HPC, big data, AI, and cloud computing?

RQ2. What are the effects of convergence on computing infrastructure, system architectures, and technologies?

RQ3. What are the various research areas and themes emerging from the convergence with regard to research areas?

3.2. Research Approach

Technology forecasting is a systematic approach to understanding the direction, speed, timing, characteristics, and effects of technological change (Porter et al., 1991). It is a tool used to evaluate the probability and significance of different future technological changes (Kang et al., 2013). Technology forecasting addresses the potential direction and impacts of technological change and the exploration of emerging technologies. Choosing the right technique is very important for the effective use of technology forecasting. Mishra et al. (2002) suggest a methodology to match the appropriate technology forecasting techniques for a technology. As mentioned in section 2.3, technology forecasting approaches exist in various computational domains, but they neither assess the convergence of HPC, big data, AI, and cloud computing nor apply bibliometric in HPC.

3.2.1. Bibliometric Analysis

One of the methods used for technology forecasting is bibliometric analysis. Bibliometric analysis is a kind of quantitative literature review to identify and quantify the relationship between authors, journals, institutes, publications, and geography and evaluate their influence on specific topics over a specific time period. Firstly, the concept of bibliometric analysis was introduced by (Pritchard, 1969) as a scientific method for understanding the evolution of a research field over time. Other key leaders have guided research over time (Kostoff, 1995; Porter & Detampel, 1995). However, today bibliometric analysis can be utilized to both comprehend past trends and forecast future trends (Morris et al., 2002). Bibliometric data can be collected from databases like Web of Science, Scopus, and PubMed, which are used for indexing and abstracting. Significant authors or publications, publications number, the average number of citations across authors or publications, and directions in the research domain can be identified. According to Donthu et al. (2021), a well-done bibliometric studies enable researchers to get a comprehensive overview, identify knowledge gaps, come up with new research ideas, and reveal intended contributions to the field. There are numerous studies in the literature that provide bibliometric analysis, which is about AI, deep learning, cloud computing, machine learning for big data analytics, Apache Hadoop, and parallel computing (El-Alfy & Mohammed, 2020; Gao & Ding, 2022; Garg et al., 2019; Knani et al., 2022; Li et al., 2020; Z. Liu et al., 2013; Penteado et al., 2021; Yu et al., 2018; Zhang & Lin, 2022). There are various studies providing convergence assessment for technologies, disciplines, and industries with bibliometric analysis (Borgman & Rice, 1992; Chand Bhatt et al., 2021; Y. Chen et al., 2022; Cui et al., 2022).

Likewise, this research aims to forecast the convergence between HPC, big data, AI, and cloud computing using bibliometric analysis. The reason for conducting the bibliometric analysis approach in this research is that bibliometric analysis is an effective and precise method for researching and analyzing large volumes of scientific data, mapping scientific knowledge, and identifying research trends and themes (Donthu et al., 2021). According to Donthu et al. (2021) if the scope of the review is broad and the dataset is too large to be reviewed manually, bibliometric analysis can be utilized. In this study, bibliometric analysis is used to reveal the driving force for convergence and the effects of convergence.

There are two main techniques in bibliometric analysis, these are performance analysis and science mapping (Donthu et al., 2021; Zupic & Čater, 2015). Performance analysis is applied to see the contributions of research components in a particular field. There are different metrics in performance analysis, such as publication-related metrics (total publications, number of contributing authors …etc.), citation-related metrics (total citations, average citations), and citation-and-publication-related metrics (collaboration index, collaboration coefficient, number of cited publications…etc.). Science mapping is the analysis of relationships and structural connections between research constituents. Science mapping includes different analysis techniques, these are citation analysis, bibliographic coupling, keyword co-occurrence analysis, and co-authorship analysis. In order to enrich the analysis techniques used in bibliometric studies, some enrichment techniques can be applied, such as network analysis. Different metrics in network analysis are network metrics (degree of centrality, betweenness centrality), clustering (exploratory factor analysis, hierarchical clustering, Louvain method) and visualization (Bibliometrix R, VOSviewer, Gephi).

The procedure that is used to select the most appropriate techniques for convergence assessment will be described in the next chapter.

CHAPTER 4

TECHNOLOGY CONVERGENCE ASSESSMENT OF HPC WITH BIG DATA, AI, AND CLOUD COMPUTING

4.1. Bibliometric Analysis Procedure

In this section, steps for conducting bibliometric analysis provided by Donthu et al. (2021) are presented. As described in the article, the stages of bibliometric analysis are examined in 4 steps. These 4 steps are as follows:

- **Step 1:** The Goals and Scope of the Bibliometric Study
- **Step 2:** Choosing the Methods for Bibliometric Analysis
- **Step 3:** Gathering the Data for Bibliometric Analysis
- **Step 4:** Running the Bibliometric Analysis and Reporting the Results

4.1.1. The Goals And Scope of the Bibliometric Study

The goal of this study is to address the research questions outlined in Section 3.1 and achieve a high level of research impact by working with large amounts of scientific data. Because the scope of the field in this study is large enough, it is considered suitable for bibliometric analysis (Donthu et al., 2021). The aim of bibliometric analysis is to retrospectively examine the performance and science of a research field. Bibliometric analysis was thought to be appropriate for this study, because it is essential to reveal the important research components (authors, countries, institutions, and journals) in the research field and the bibliometric structure between research constituents. Furthermore, bibliometric analysis, it is aimed to reveal the effects of convergence, driving forces for the convergence, and emerging themes through a comprehensive study.

4.1.2. Choosing the Methods for Bibliometric Analysis

In the second stage, the methods for bibliometric analysis were selected in order to achieve the goals and focus of the study. As recommended in the article (Donthu et al., 2021), firstly techniques were chosen and then bibliometric data were prepared according to selected technique. The bibliometric techniques used in this study and why these techniques were chosen are explained below.

Performance Analysis

This method was chosen to evaluate the influence of various research constituents (authors, institutions, countries, and journals) on the research field (Donthu et al., 2021). There are many measures for performance analysis. According to some metrics (publication related, citation related) critical reviews were made. In this study, analyzes such as the effectiveness of scientific actors (researchers, journals, institutions, country) in the study area and the most influential publications were made.

Network Analysis

- **Co-occurrence Network Analysis:** It is a visualization approach used in bibliometric analysis showing the relationship between groups of terms within a specific text by examining their connection (Ravikumar et al., 2015). In this study, co-occurrence network analysis was used to identify connections between subjects in the research area and to analyze the hotspots and trends in the research field.

- **Thematic Map:** In the thematic map method, which was first introduced by Callon et al. (1991), there is a coordinate plane consisting of 4 quadrants representing centrality (x-axis) and density (y-axis). Each quadrant represents a theme module, which are motor themes, basic and transversal themes, emerging or disappearing themes, and isolated/niche themes. Keywords in the research area take place in this plane by forming certain clusters according to their proximity. The stronger the centrality for a particular cluster, the more important that cluster is considered. Density shows the strength of the links that connect the words that make up the clusters. The stronger these connections are, the more the themes corresponding to the cluster form a coherent and integrated whole.

- **Thematic Evolution:** Thematic evolution is a technique used in bibliometric analysis to examine the evolution of a particular research topic over time, as well as to identify trends and changes in the focus of research within that topic. The emergence of new subtopics or the decline of older ones is identified, and a broader understanding of how a particular theme has evolved within the field over time is provided by thematic evolution.

4.1.3. Gathering the Data for Bibliometric Analysis

Web of Science (WoS) scientific citation indexing service is used in this research for searching and gathering the data. As recommended in the paper (Donthu et al., 2021), only one database is selected as minimizing the unnecessary action items can help minimize possible human errors. The main motivation for choosing WoS is its wide scope, quality of resources, and powerful research engine (Birkle et al., 2020),

Search Strategy

In order to determine the search strategy, the keywords were identified from the literature. A separate query series for HPC and other technologies (AI, big data, cloud computing) was created, and the results were combined. The query was made on the basis of title, author keywords, abstract, keyword plus, research area. The search results were limited to English-language publications that included journal articles (including early access), proceeding articles, and book chapters. Proceedings articles were included because it might be a starting point for new research. The reason to add book chapters is that books provide theoretical knowledge of the technology domain. After the search result, 3,748 publications were obtained from the WoS as shown in Table 1. Due to the limitation of exporting 1000 records at a time, the results were exported in four separate queries in BibTeX file format.

Table 1: Search strategy based on Chand Bhatt et al. (2021)

	Search String	Result	Combination
1	"high$performance comput*" OR HPC OR supercomput*	30,908	(#1 AND #2)
2	"artificial intelligence" OR "machine learning" OR "deep learning" OR "data science" OR "data mining" OR "data analytic*" OR "data management" OR "big$data" OR "data analys?s" OR "data process*" OR "data$intensive" OR "data govern*" OR "cloud comput*" OR "edge comput*" OR "fog comput*" OR "micro$service*" OR virtualiz* OR "virtual machine*" OR container* OR "cloud$native"	1,061,537	3,748

Refining Criteria: Languages: (English); Document types: (Article or Proceedings paper or Book chapter or Early access); Timespan: All years.

4.1.4. Running the Bibliometric Analysis and Reporting the Results

The fourth and final stage is reporting the bibliometric findings. Performance analysis and science mapping results are provided in Section 4.2.

4.2. Result Analysis

4.2.1. Results of the Performance Analysis

After the queries, four BibTex files were obtained from the search results. These files were manually merged to form a single file for performance analysis using Bibliometrix R-package.

According to search query, 3,748 documents from 1864 sources were obtained. Brief information is provided in Table 2.

Table 2: Main information about data

Description	Results
MAIN INFORMATION ABOUT DATA	
Timespan	1988:2023
Sources (Journals, Books, etc)	1864
Documents	3748
Annual Growth Rate %	3,19
Document Average Age	5,52
Average citations per doc	10,67
References	79302
DOCUMENT CONTENTS	
Keywords Plus (ID)	2178
Author's Keywords (DE)	7529
AUTHORS	
Authors	11532
Authors of single-authored docs	203
AUTHORS COLLABORATION	
Single-authored docs	220
Co-Authors per Doc	4,81
International co-authorships %	22,6
DOCUMENT TYPES	
article	1353
article; book chapter	37
article; data paper	2
article; early access	41
article; proceedings paper	90
editorial material; book chapter	1
proceedings paper	2221
proceedings paper; retracted publication	1
review; early access	2

Figure 1 illustrates the annual scientific production. Applications of HPC with other domains started to increase, and research publications followed an exponential curve after 2006. While there were just 15 publications in 2006, the number of publications continues to increase, with 56 in 2010, 146 in 2012, and 398 in 2018. It is seen that the most publications were made in 2019 with 447. Additionally, Figure 2 displays the distribution of the top 20 relevant scientific production of HPC and AI, big data, and cloud computing. Future Generation Computer Systems-The International Journal of Escience has the highest number of published articles among others, with 70 articles. Following, Concurrency and Computation: Practice and Experience is 2nd with 69 articles. Then, followed by Journal of Supercomputing and IEEE Transactions on Parallel and Distributed Systems respectively with 60 and 57.

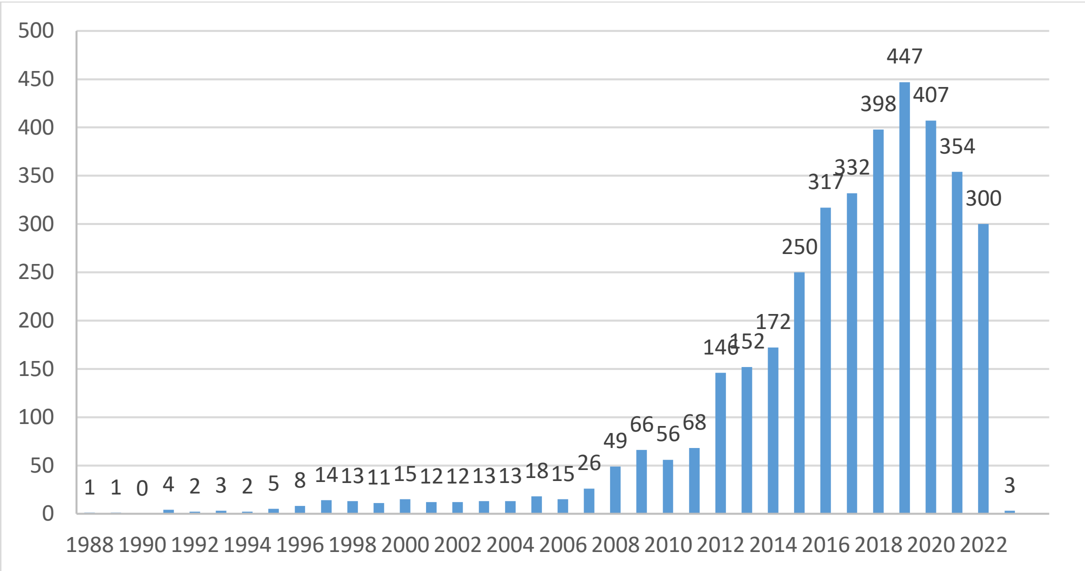

Figure 1: Annual scientific production

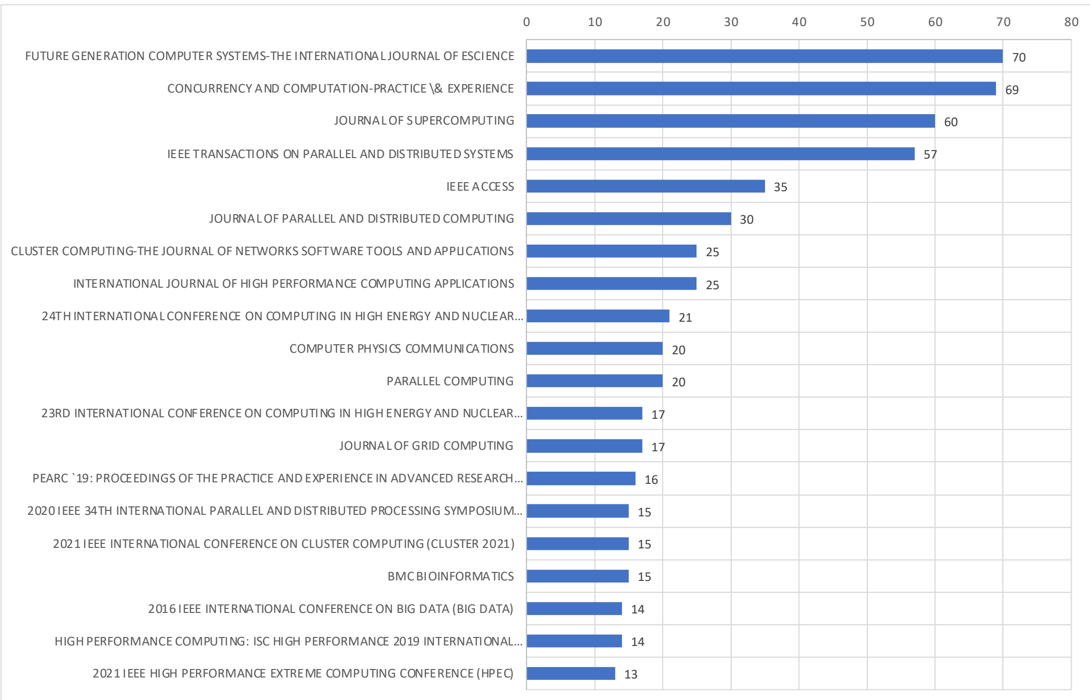

Figure 2: Top 20 relevant sources

Table 3 highlights the sources' growth dynamics. Sources from 2010 to the present were selected because there have not been many contributions in this area before. The top 10 sources have been determined that have made the highest contribution to the domain of HPC and AI, big data, and cloud computing convergence. The most significant growth is seen in Future Generation Computer Systems-The International Journal of Escience with % 19,5, followed by Concurrency and Computation: Practice and Experience with % 15,9. Among other prominent sources in the ranking based on the average of all years are IEEE Transactions on Parallel and Distributed Systems with %13,4 and Journal of Supercomputing with %11,4.

Table 3: Sources growth dynamics

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	Total	Perc(%)
FUTURE GENERATION COMPUTER SYSTEMS-THE INTERNATIONAL JOURNAL OF ESCIENCE	6	7	9	16	18	23	25	33	46	56	63	64	68	70	504	19,5
CONCURRENCY AND COMPUTATION-PRACTICE \& EXPERIENCE	3	4	5	7	12	17	21	26	33	45	49	58	66	66	412	15,9
JOURNAL OF SUPERCOMPUTING	3	3	4	7	8	10	16	18	22	25	32	43	52	52	295	11,4
IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS	1	3	5	7	10	13	19	25	29	35	41	45	57	57	347	13,4
IEEE ACCESS							1	3	6	15	21	26	35	35	142	5,5
JOURNAL OF PARALLEL AND DISTRIBUTED COMPUTING	3	4	5	8	8	10	10	13	16	19	22	25	30	30	203	7,8
CLUSTER COMPUTING-THE JOURNAL OF NETWORKS SOFTWARE TOOLS AND APPLICATIONS	1	1	1	2	4	5	6	9	10	14	18	20	23	23	137	5,3
INTERNATIONAL JOURNAL OF HIGH PERFORMANCE COMPUTING APPLICATIONS	3	4	6	7	7	7	8	10	12	14	16	22	24	24	164	6,3
24TH INTERNATIONAL CONFERENCE ON COMPUTING IN HIGH ENERGY AND NUCLEAR PHYSICS (CHEP 2019)											21	21	21	21	84	3,2
COMPUTER PHYSICS COMMUNICATIONS	3	3	4	6	6	7	9	10	11	11	14	15	20	20	139	5,4
PARALLEL COMPUTING	5	5	5	5	6	6	8	10	13	18	19	19	20	20	159	6,1

Table 4 displays a list of the top 20 authors, ranked by the number of publications they have produced. The most productive author is Panda, D.K., who has 28 papers. Lu, X. has 24 publications, followed by Chen, J., with 23 publications. All the top 20 authors have at least 19 publications.

Table 4: Top 20 authors ranked by the publication count

Rank	Authors	Articles
1	PANDA, DK	28
2	LU, X	24
3	CHEN, J	23
4	KEPNER, J	23
5	WANG , X	23
6	WANG, Y	23
7	ZHANG, J	23
8	CHEN, Y	22
9	SUBRAMONI, H	22
10	BYNA, S	21
11	FOSTER, I	21
12	JHA, S	21
13	LIU, Y	21
14	WANG, J	21
15	ZHANG, Y	21
16	LIU, X	20
17	REUTHER, A	20
18	ZHANG, Z	20
19	BESTOR, D	19
20	JONES, M	19

The most remarkable organizations according to the number of publications are presented in Figure 3. The top affiliation with 167 publications is University of Illinois. The top second affiliation is both Argonne National Laboratory and Oak Ridge National Laboratory, with 166 publications, and the top third is Lawrence Berkeley National Laboratory, with 145 publications. Furthermore, Ohio State University ranks 5th with 133 publications. All the top 20 organizations have at least 48 publications.

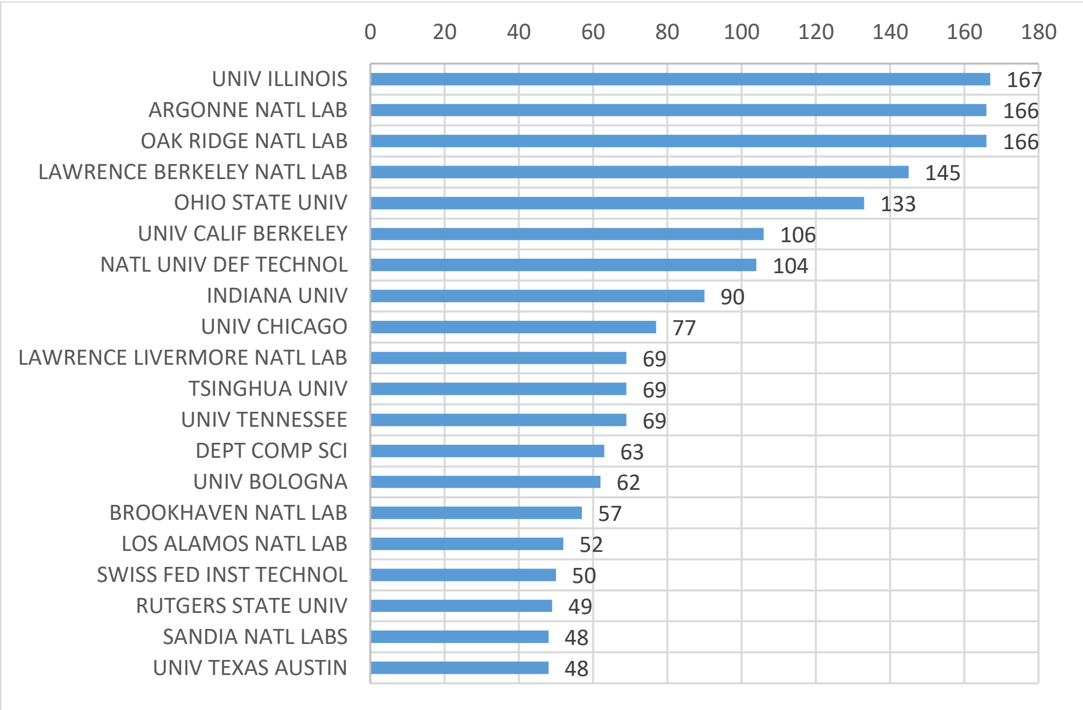

Figure 3: Top institutes based on publication count

Table 5 indicates the total number of citations and annual total citations for the 10 most cited publications globally. In order to identify important publications with a high impact on the research field, they were ranked according to the total number of citations and annual citations. The number of citations shows how popular the publications are in the research field. The most cited publication (Gorelick et al., 2017) was published in Remote Sensing of Environment with 4013 citations. This is followed by the publication (Buyya et al., 2009) in Future Generation Computer Systems ranking second with 2934 citations. Similarly, the study by Amodei et al. (2016) published at the International Conference on Machine Learning ranks 3rd with 1058 citations. The least cited document from 10 publications received 279 citations.

Table 5: Top 10 globally cited publications

Rank	Authors, Soruce	Paper	Total Citations	TC per Year
1	GORELICK N, 2017, REMOTE SENS ENVIRON	Google Earth Engine: Planetary-scale geospatial analysis for everyone	4013	668,83
2	BUYYA R, 2009, FUTUR GENER COMP SYST	Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th	2934	209,57
3	AMODEI D, 2016, INTERNATIONAL CONFERENCE ON MACHINE	Deep Speech 2 : End-to-End Speech Recognition in English and Mandarin	1058	151,14
4	CHEN Y, 2014, 2014 47TH ANNUAL IEEE/ACM INTERNATIONAL	DaDianNao: A Machine-Learning Supercomputer	765	85,00
5	AUGONNET C, 2011, CONCURR COMPUT -PRACT EXP	StarPU: a unified platform for task scheduling on heterogeneous multicore architectures	586	48,83
6	NURMI D, 2009, CCGRID: 2009 9TH IEEE INTERNATIONAL SYMPOSIUM	The Eucalyptus Open-Source Cloud-Computing System	533	38,07
7	THOMPSON AP, 2022, COMPUT PHYS COMMUN	LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and	493	493,00
8	KURTZER GM, 2017, PLOS ONE	Singularity: Scientific containers for mobility of compute	470	78,33
9	WOLSTENCROFT K, 2013, NUCLEIC ACIDS RES	The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in	379	37,90
10	KOHLHOFF KJ, 2014, NAT CHEM	Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways	279	31,00

Figure 4 shows the ranking of the top 20 countries based on the number of publications they have produced. The USA ranks first with 1444 in the number of publications, followed by China with 349. It is followed by Germany (194), Spain (154), Italy (136), United Kingdom (131), Japan (118), Russia (110), and Brazil (102), respectively. 5 countries with more than 50 but less than 100 publications were identified, these are India (96), France (85), Australia (80), Korea (76), and Switzerland (50). Furthermore, the collaboration indices of the countries are indicated in Figure 4 as intra-country (SCP) and inter-country (MCP) in different colors.

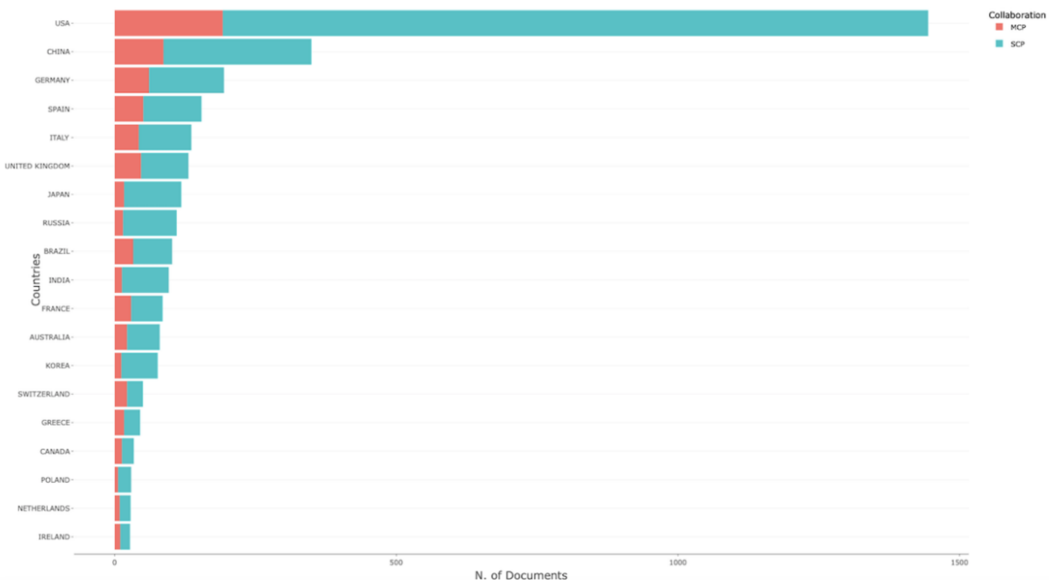

Figure 4: Top 20 countries based on the number of publications they have produced

Table 6 and Figure 5 present the keyword analysis results and word cloud related to the field. Table 6 reveals the most common terms used by authors in their articles. ‘High-performance computing’ is the term with the highest occurrences (2996), this points out its importance. The second most frequent term is ‘cloud computing’ with 1626 occurrences, then ‘machine learning’ comes in third position with 814 occurrences. Other important terms in the list are ‘deep learning’ (540), ‘virtual machines’ (501), ‘data analysis’ (487), ‘data processing’ (450), and ‘neural network’ (411). Finally, ‘energy consumption’ (341) and ‘data management’ (311) are relevant themes.

Table 6: Keyword analysis

Terms	**Frequency**
high-performance computing	2996
cloud computing	1626
machine learning	814
deep learning	540
virtual machines	501
data analysis	487
data processing	450
neural network	411
energy consumption	341
data management	311

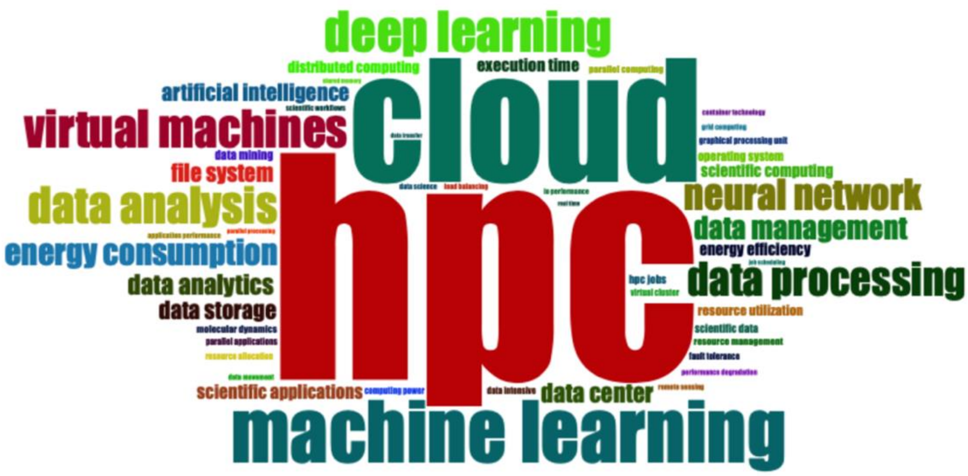

Figure 5: Word cloud

Analysis of Top 10 Globally Cited Publications

The 10 most cited publications globally were presented in Table 5. Here is a more detailed analysis of these publications.

The most cited publication is an application of a cloud-based platform with HPC resources (Gorelick et al., 2017). It presents Google Earth Engine, the cloud-based platform that brings together Google's computing capabilities to address a variety of important social issues such as climate monitoring, environmental protection, disease, food security and water management. As the demands for data and computation continue to increase to solve more complex problems, the convergence of these technologies is inevitable to tackle these challenges.

The 2nd most highly cited paper suggests an architecture for the allocation of resources within clouds (Buyya et al., 2009). They provide a vision for the development of a global Cloud exchange for trading services. Furthermore, the differences between Internet-based services workloads and HPC workloads are examined in this study. As stated in the article, with the significant progress in Information and Communications Technology (ICT) over the past 50 years, it is a widely held belief that computing will eventually be considered just as necessary as water, electricity, gas, and telephony. The top third cited paper is an application-based paper. The end-to-end deep learning approach is presented to improve speech recognition systems (Amodei et al., 2016). The key approach here is the usage of the HPC techniques to enable experiments more quickly. The article shows that by converging technologies, it may be possible to deal with increasing data more effectively and efficiently.

Rank #4 article is an application of multi-chip machine-learning architecture (Y. Chen et al., 2014). In the article, an architecture is proposed called ‘machine learning supercomputer’ to achieve high sustained machine-learning performance by using a multi-chip system.

Rank #5 is again an application-based article, which presents STARPU (Augonnet et al., 2009). It is a new runtime system that is efficient at utilizing multi-core architectures with different hardware to produce parallel tasks, and it makes it easy to create and optimize scheduling algorithms.

Rank #6 article is an application of EUCALYPTUS (a software framework for cloud computing that is open-source) that combines cloud computing systems with HPC (Nurmi et al., 2009). It provides a system that is easy to deploy existing resources, and it is designed to be modular and open-source, which allows for experimentation.

Rank #7 is an article that describes several aspects of the LAMMPS molecular dynamics package (including the usage of supercomputers in simulations) (Thompson et al., 2022). It is a simulation tool that can be used for material modeling and can run on any platform, from a single CPU core to the most powerful supercomputers with accelerators. It is flexible and adaptable. The use of molecular dynamics, which simulates the behavior of molecules, has become more popular due to improvements in hardware and the incorporation of machine learning potentials (Sharma & Jadhao, 2021).

A new open-source software called Singularity is offered in Rank #8 (Kurtzer et al., 2017). The aim is to provide mobility of computing (the ability to create and deploy reproducible environments) for both users and HPC centers. Rank #9 discusses the design and execution of workflows involving Web Services on the desktop, web, or in the cloud (Wolstencroft et al., 2013). Article Rank #10 is an application of Google's Exacycle cloud computing platform for simulations (Kohlhoff et al., 2014). With the developed cloud-based approach, it is possible to run molecular dynamic simulations more effectively and efficiently.

4.2.2. Results of the Network Analysis

Co-occurrence Network Analysis

It was performed to show the relationship between various themes and applications in the research area. In this study, co-occurrence network analysis was used to discover linkages among subjects in the research area in order to identify the driving force for the convergence and the effects of convergence on technologies. Figure 5 highlights the co-occurrence network analysis. The size of the nodes in the visualization reflects the frequency of the keywords, and the lines represent the relationship between the two keywords. After the analysis, three main clusters were identified, as shown in Figure 5. The main keywords per cluster are 'high-performance computing' and 'cloud computing' (red cluster), 'data analysis' (blue cluster), and 'machine learning' (green cluster). Explanations about three identified clusters are as follows:

Cluster 1: (Red cluster with 28 keywords). In this cluster, connection, among the main themes of HPC and cloud computing are illustrated in different nodes. This cluster shows the correlation between keywords related to HPC and cloud computing technologies, such as virtual machines, energy consumption, energy efficiency, resource utilization, fault tolerance, data center, distributed computing, grid computing, and so on.

Cluster 2: (Blue cluster with 17 keywords). This cluster primarily identified themes related to data analysis with keywords such as data processing, data management, data analytics, molecular dynamics, real-time, scientific workflows, remote sensing, data storage, data mining, and so on. It is interesting that the molecular dynamics theme is in the blue cluster. The reason for that might be the simulations in molecular dynamics for understanding and predicting the behavior of complex systems. It highly relies on data analysis and data processing in order to extract useful insights and make predictions about the behavior of atoms and molecules. One of the top-cited articles by (Thompson et al., 2022) confirms that the increasing popularity of molecular dynamics is due to the combination of hardware advancements and the use of machine learning potentials. These advances have enabled the development of more sophisticated and computationally intensive potentials, which have significantly improved the accuracy of material property predictions.

Cluster 3: (Green cluster with 5 keywords). This cluster shows various themes based on 'machine learning'. The green cluster illustrates connections among machine learning, deep learning, neural network, AI, and graphical processing units. It is observed that some applications of machine learning, deep learning, and AI are connected to HPC.

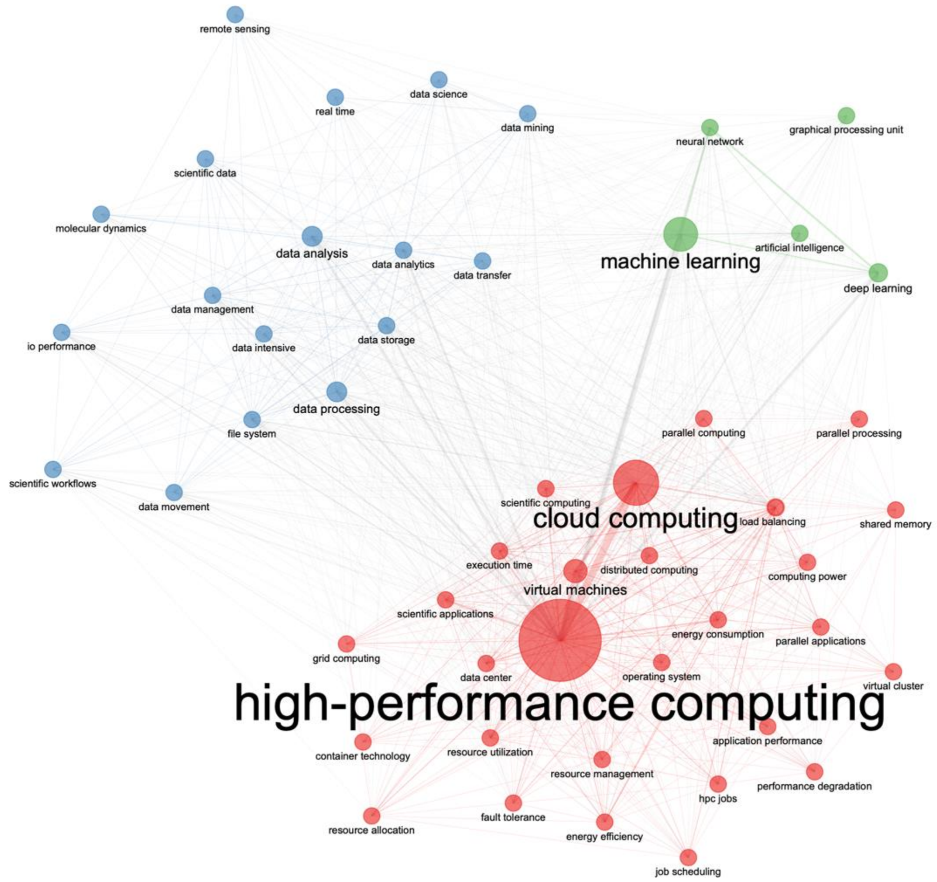

Figure 6: Conceptual structure based on co-occurrence network

Thematic Mapping

In this section, a thematic map was performed using the bibliometrix R-package to identify the research focus. The diagram as shown in Figure 7 was analyzed by considering the placement of themes, which were evaluated based on two measures; centrality and density.

The upper-right section of the thematic map, which contains themes with strong centrality and high density, represents the "motor themes". The most dominant cluster in this quadrant contains data-related keywords, such as ‘data analysis’, ‘data processing’, and ‘data management’. Other prominent keywords are ‘data analytics’, ‘data storage’, ‘scientific computing’, ‘distributed computing’, and ‘data mining’. These themes can be considered the most significant drivers of the convergence as a response to RQ1. Data analysis and data processing are important to efficiently manage and gain inside based on large volumes of data, which can help increase efficiency and improve business operations. Especially considering industry 4.0, to collect, analyze, and make use of large volumes of data from a variety of sources, like sensors, machines, and other devices in real-time, traditional methods may be insufficient (Jagatheesaperumal et al., 2022; Khan et al., 2017). HPC provides the processing power required to process this type of data.

The themes located in the bottom-right section of the thematic map are basic and transversal themes with high centrality but low density. This quadrant includes one cluster that overlaps the upper right quadrant. Major themes in this cluster are ‘HPC’, ‘data center’, and ‘energy consumption’. HPC can be considered as a basic and enabling theme for many other technologies, because it enables the processing and analyzing large amounts of data quickly and efficiently. Because it has a high level of computing performance, it can handle a variety of workloads, such as simulations, modeling, machine learning, and data analysis. Other keywords in this cluster are ‘execution time’, ‘energy efficiency’, ‘parallel computing’, ‘computing power’, ‘application performance’, and ‘graphical processing unit’. HPC requires a high amount of processing power involving powerful computers to solve complex problems. It is usually located in data centers that contain large numbers of computer servers and other equipment. One of the main challenges with HPC is high levels of energy consumption. The emerging concept here is ‘green computing’, which refers to the use of computing technologies in an environmentally friendly and energy-efficient manner (Gai et al., 2016). Green computing can be considered one of the most important driving forces for convergence (RQ1). Therefore, HPC requires the use of virtualization and cloud computing technologies to reduce the number of physical servers and other hardware components needed and to get cost-effective and energy-efficient solutions.

The lower-left quadrant contains two clusters as emerging themes. The most prominent cluster mainly contains ‘machine learning’, ‘deep learning’, and ‘neural network’. Other considerable keywords in this cluster are ‘artificial intelligence’, ‘random forest’, ‘computational fluid dynamics’, ‘distributed training’, and so on. Deep learning, as a subfield of machine learning, is an emerging theme in the field of machine learning. For instance, deep neural networks are used for complex tasks such as image and speech recognition (Amodei et al., 2016). It is particularly well-suited for processing large amounts of data. As more data becomes available and advances in computing power continue to increase, the neural network becomes important in solving complex problems and providing valuable insight. Deep learning is being utilized in a variety of applications, including image and speech recognition, natural language processing, and recommendation systems. Another cluster in this quadrant includes ‘hardware performance’, which is also included as a driving force (RQ1). It is an important factor in computing. The increasing complexity of applications such as machine learning and the need for fast and powerful hardware is constantly increasing. Therefore, the development of new and improved hardware technologies is an ongoing area of research and development in order to solve complex problems quickly and efficiently.

Themes located in the upper-left quadrant are the specific (niche) themes in the field. This quadrant includes one cluster that overlaps the upper right quadrant. ‘Cloud computing’, ‘virtual machines’, and ‘operating systems’ are the main keywords in this cluster. Other prominent keywords are ‘resource utilization’, ‘hpc jobs’, ‘resource management’, ‘resource allocation’, ‘load balancing’, ‘fault tolerance’, ‘grid computing’, ‘container technology’, and ’hpc cloud’. These keywords provide the answer to the RQ2 in the thesis. More information about this topic will be provided in Section 4.3.

Cloud computing is a constantly evolving field (Buyya et al., 2009; Gorelick et al., 2017; Kohlhoff et al., 2014; Nurmi et al., 2009). As for niche topics in cloud computing, there are many different possibilities depending on the industry and application. Some of these topics include the use of artificial intelligence and machine learning in the cloud, such as the adoption of hybrid and multi-cloud architectures, the security and privacy of cloud-based systems, serverless computing, containerization, edge and fog computing. All these developments are important for enabling real-time, low-latency applications and services to provide distributed computing with highly scalable, flexible, and responsive.

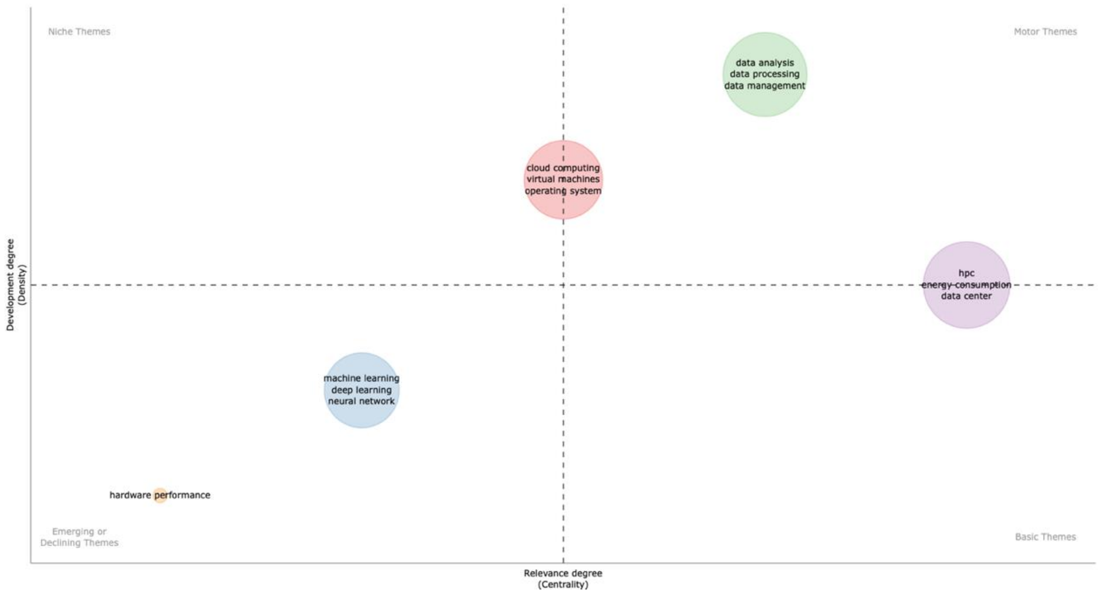

Figure 7: Thematic map

Further, the evolution of the themes with time is plotted in 3 intervals (1988-2016, 2017-2019, 2020-2023), as shown in Figure 8. It presents how identified clusters evolved over the three period of times. The period from 1988 to 2016 is found to revolve around the themes of ‘data processing’ and ‘high-performance computing’. Gradually, other themes evolved in 2017-2019; for instance, ‘embedded systems’, ‘hardware performance’, malware detection’, ‘deep learning’, ‘data management’, ‘cloud computing ’, and ‘data analysis’. The evolution of the themes is represented in Tables 7 and 8 with essential keywords. For instance, ‘high-performance computing’ evolved into three different fields in the first time slice; these are ‘cloud computing’, ‘data management’, and ‘data analysis’. While evolving, these themes led to significant other topics, such as ‘workflow management’, ‘neural network’, ‘data mining’, ‘virtual machines’, ‘data center’, ‘grid computing’, ‘resource allocation’, ‘energy consumption’ ,’benchmark suites’, and so on (as shown in Table 7). These are considered as driving force for the convergence of HPC with other technologies that can address RQ1 and RQ2. The period 2020–2023 has witnessed new topics (as shown in Table 8), such as ‘computational fluid dynamics’, ‘artificial intelligence’, ‘molecular dynamics’, ‘malware detection’, ‘anomaly detection’, ‘random forest’, ‘monte carlo’, ‘computer vision’, ‘performance degradation’. These research areas may be considered emerging themes from the convergence (answer of the RQ3).

Figure 9 depicts themes trends over time to answer RQ3 of the thesis. It indicates the various research areas and themes emerging over time. In the trend topics figure, the size of the circles indicates the frequency of the term and the length of the lines indicates how long it has been studied. When Figure 9 is examined, it is seen that the ‘intelligent computing’, ‘transfer learning’, ‘reinforcement learning’, ‘deep learning’, ‘neural network’, ‘artificial intelligence’ and ‘molecular dynamics’ has gained popularity in recent years. This is largely due to advancement in computational power and abundance of data that has become available. HPC is another important topic that has started to gain popularity due to its increasing use in fields such as data analytics, machine learning, and AI since about 2017.

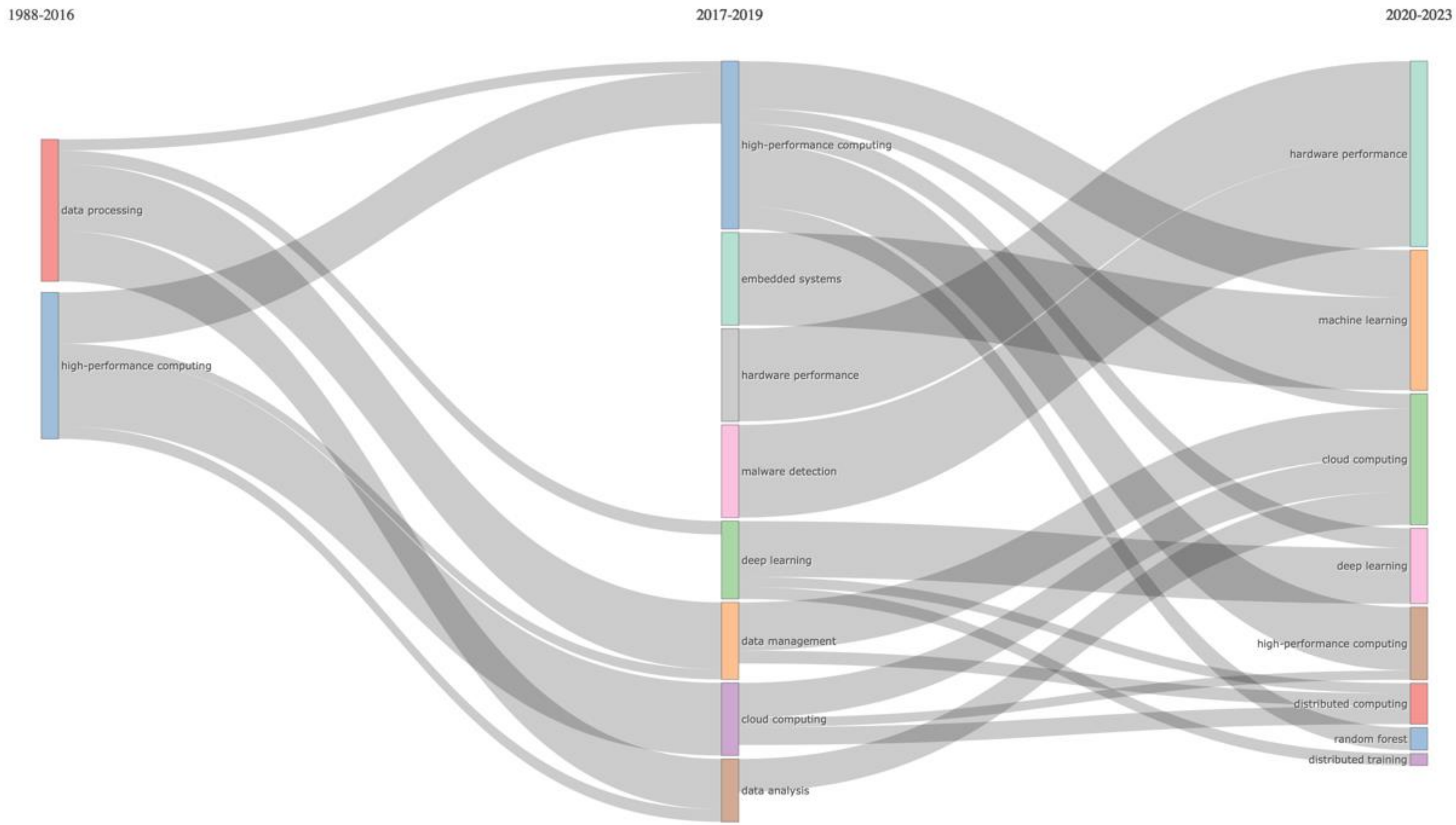

Figure 8: Thematic evolution

Table 7: Evolution of the themes with essential keywords

From	To	Words	Occurrences
data processing--1988-2016	cloud computing--2017-2019	scientific community;scientific research;central processing unit;intensive applications;execution times	14
data processing--1988-2016	data analysis--2017-2019	data processing;data analysis;data analytics;data intensive;data products;data transfer;large-scale data;map reduce;system architecture;big data analytics;data science;scientific simulations;data applications;data reduction;apache spark;data sources;raw data;storage devices	186
data processing--1988-2016	data management--2017-2019	data management;data storage;file system;scientific data;data access;data movement;data locality;io performance;parallel io;workflow management;access patterns;research data;data size	110
data processing--1988-2016	deep learning--2017-2019	neural network;performance results;deep learning	30
data processing--1988-2016	high-performance computing--2017-2019	machine learning;data mining;molecular dynamics;remote sensing;artificial intelligence;real time;image processing;peak performance;time series;monte carlo;hpc architectures	81
high-performance computing--1988-2016	cloud computing--2017-2019	cloud computing;virtual machines;data center;distributed computing;virtual cluster;execution time;resource management;fault tolerance;grid computing;resource allocation;resource utilization;hpc jobs;job scheduling;load balancing;scheduling algorithm;web	406
high-performance computing--1988-2016	data analysis--2017-2019	scientific applications;parallel processing;distributed systems;improve performance;job execution;data security	70
high-performance computing--1988-2016	data management--2017-2019	scientific workflows;scientific workflow;data structures;task scheduling	13
high-performance computing--1988-2016	deep learning--2017-2019	performance evaluation;network topology	21
high-performance computing--1988-2016	high-performance computing--2017-2019	high-performance computing;energy consumption;operating system;scientific computing;energy efficiency;parallel computing;parallel applications;computing power;shared memory;application performance;parallel programming;performance degradation;hpc clouds;performance	602

Table 8: Evolution of the themes with essential keywords

From	To	Words	Occurrences
cloud computing--2017-2019	cloud computing--2020-2023	cloud computing;virtual machines;resource utilization;hpc jobs;load balancing;resource management;resource allocation;container technology;programming models;fault tolerance;scientific research;central processing unit;container technologies;scheduling algorithm;scheduling algorithms;system software	252
cloud computing--2017-2019	deep learning--2020-2023	computational power;heterogeneous computing	17
cloud computing--2017-2019	distributed computing--2020-2023	distributed computing	35
cloud computing--2017-2019	high-performance computing--2020-2023	data center;execution time;edge computing	51
cloud computing--2017-2019	random forest--2020-2023	genetic algorithm	12
data analysis--2017-2019	cloud computing--2020-2023	data analysis;data processing;data analytics;scientific applications;data transfer;data applications;data intensive;data science;distributed systems;large-scale data;parallel processing;system architecture;high-performance data	118
data analysis--2017-2019	deep learning--2020-2023	cray xc	11
data analysis--2017-2019	high-performance computing--2020-2023	raw data	8
data management--2017-2019	cloud computing--2020-2023	data management;file system;data storage;scientific workflows;io performance;data movement;simulation data;workflow management;access patterns;parallel io;data structures;task scheduling	67
data management--2017-2019	distributed computing--2020-2023	scientific data	24
deep learning--2017-2019	cloud computing--2020-2023	performance results	15
deep learning--2017-2019	deep learning--2020-2023	deep learning;neural network;training time	125
deep learning--2017-2019	distributed computing--2020-2023	computer vision	12
deep learning--2017-2019	distributed training--2020-2023	distributed training;training process	13
deep learning--2017-2019	high-performance computing--2020-2023	performance evaluation	15
deep learning--2017-2019	machine learning--2020-2023	classification accuracy	10
embedded systems--2017-2019	machine learning--2020-2023	embedded systems	5
hardware performance--2017-2019	hardware performance--2020-2023	hardware performance	11
high-performance computing--2017-2019	cloud computing--2020-2023	scientific computing;computing power;memory access;parallel computing;application performance;message passing interface;compute intensive;exascale computing;linear algebra;performance improvements;distributed memory;performance improvement;memory bandwidth;communication patterns;parallel programming;hpc architectures;hardware resources;memory accesses;memory usage	45
high-performance computing--2017-2019	deep learning--2020-2023	artificial intelligence;molecular dynamics;remote sensing;computationally expensive;image processing;performance analysis;training data;computational fluid dynamics;computational cost	65
high-performance computing--2017-2019	distributed computing--2020-2023	graphical processing unit	38
high-performance computing--2017-2019	high-performance computing--2020-2023	high-performance computing;energy consumption;energy efficiency;data mining;parallel applications;operating system;performance degradation;real time;finite element;application developers	548
high-performance computing--2017-2019	machine learning--2020-2023	machine learning;anomaly detection;time series	219
high-performance computing--2017-2019	random forest--2020-2023	prediction model;random forest;monte carlo;prediction accuracy	10
malware detection--2017-2019	hardware performance--2020-2023	malware detection	6

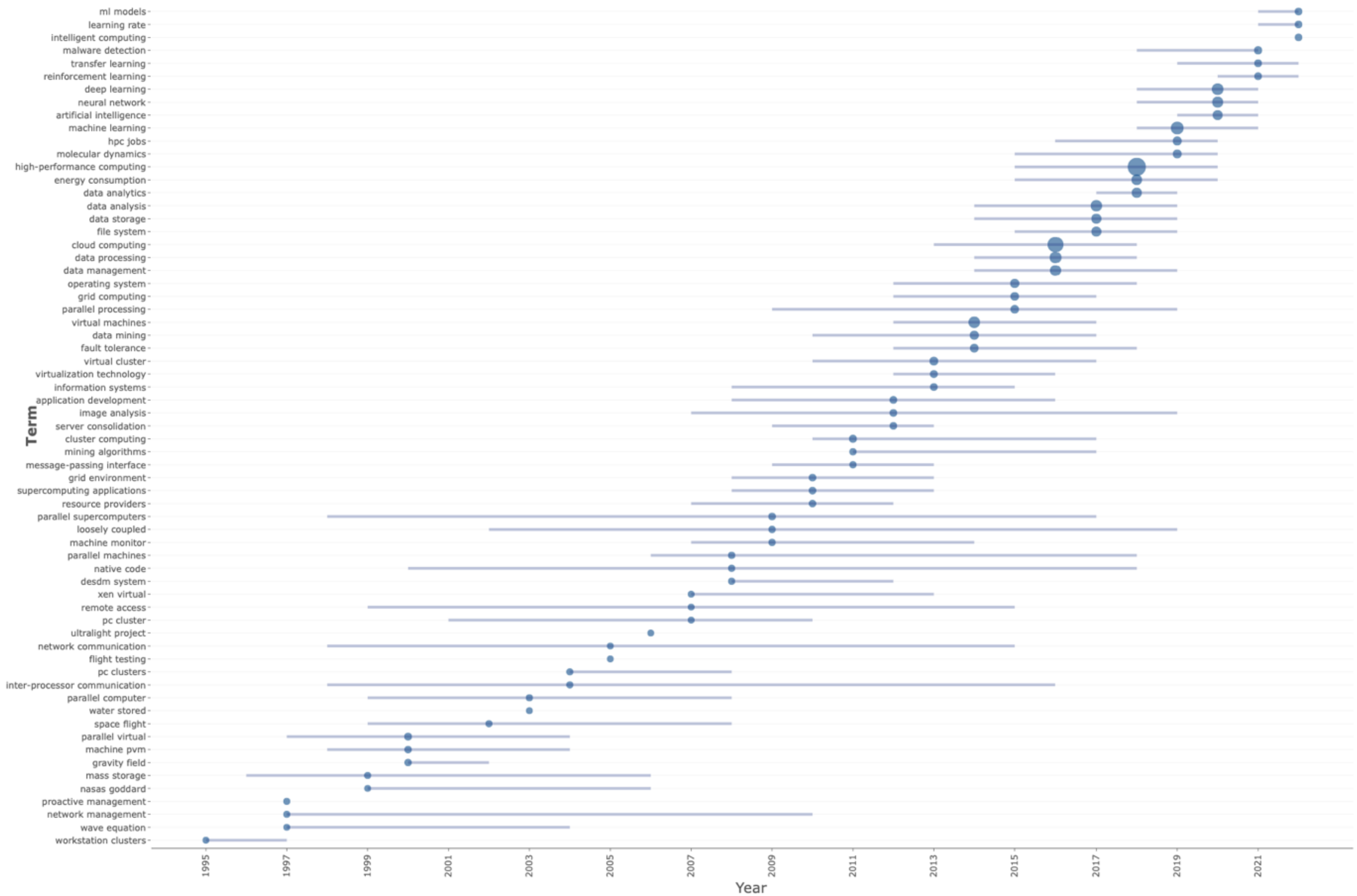

Figure 9: Trend Topic

4.3. Discussion of the Findings

Based on the results obtained from the bibliometric analysis, the answers to the research questions are summarized in this section. The research conducted has given us a comprehensive understanding of the ways in which HPC, AI, big data, and cloud computing technologies can be used together. There are several research and future implications that were identified regarding the potential future convergence of these technology areas.

The decline in publications as of 2019 depicted in the Annual Scientific Production presented in Section 4.2, Figure 1, is considered to be primarily due to the impact of the Covid-19 pandemic. It should be noted that this graph encompasses both articles and proceeding papers. Upon removal of proceeding papers from the graphic, as depicted in Figure 10, a linear increase in article numbers becomes apparent.

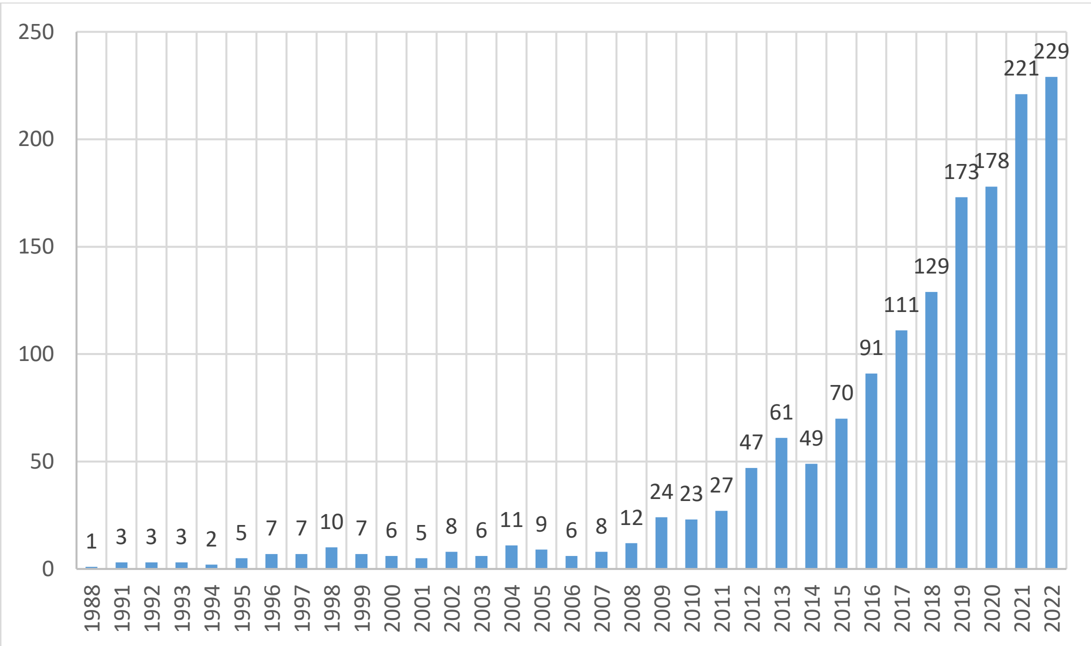

Figure 10: Article numbers (without proceeding papers)

Examining the situation within the European Union, it is evident, as depicted in the Figure 11, that a rise in publications has occurred. The top 10 organizations that finance these publications with the number of articles is given in Figure 12. The establishment of the European High-Performance Computing Joint Undertaking (EuroHPC JU) in 2018 and the calls made through the European Union Framework Programme may have played an important role in catalyzing advancements in this field. Important calls have been made and continue to be made in the field of HPC in the European Commission 7th Framework Program (between 2007 and 2013), Horizon 2020 (2014-2020) and Horizon Europe (2021-2027). For example, a call for proposals was initiated on 20 April 2022 under the title of National Competence Centers for High Performance Computing (DIGITAL-EUROHPC-JU-2022-NCC-01-01). On 28 December 2022 another call was launched, entitled Study for Hyper-connectivity for HPC Resources (EuroHPC/LUX/2022/OP/01). In addition, a call has announced on 26 January 2023 under the title EuroHPC traineeships in Hosting Entities, Centers of Excellence and Competence Centres, SMEs and Industry (DIGITAL-EUROHPC-JU-2022-TRAINING-03-01).

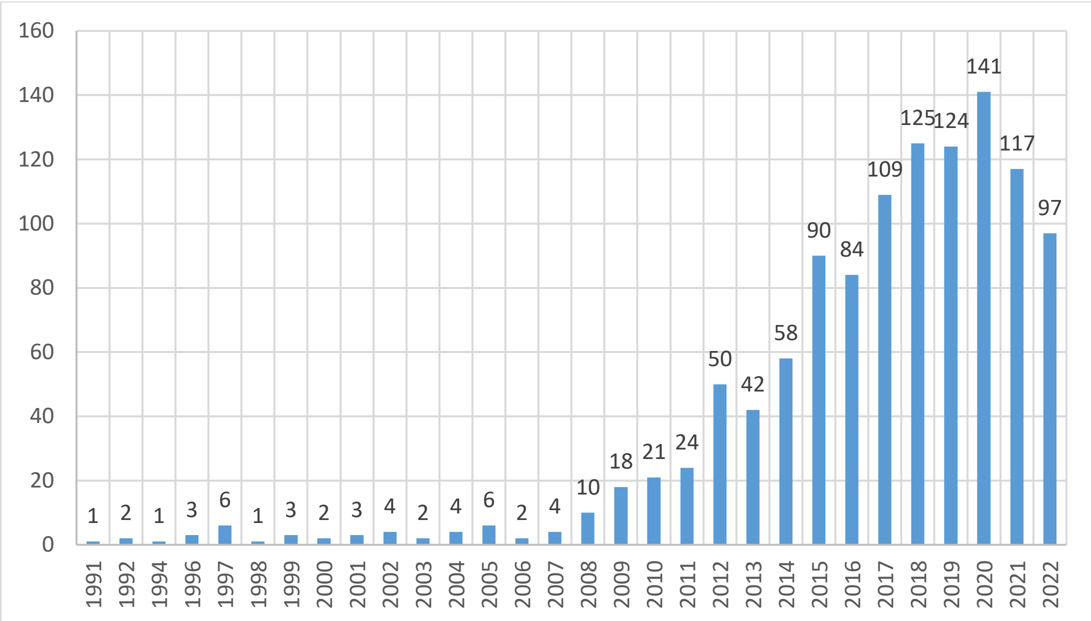

Figure 11: European Union publications

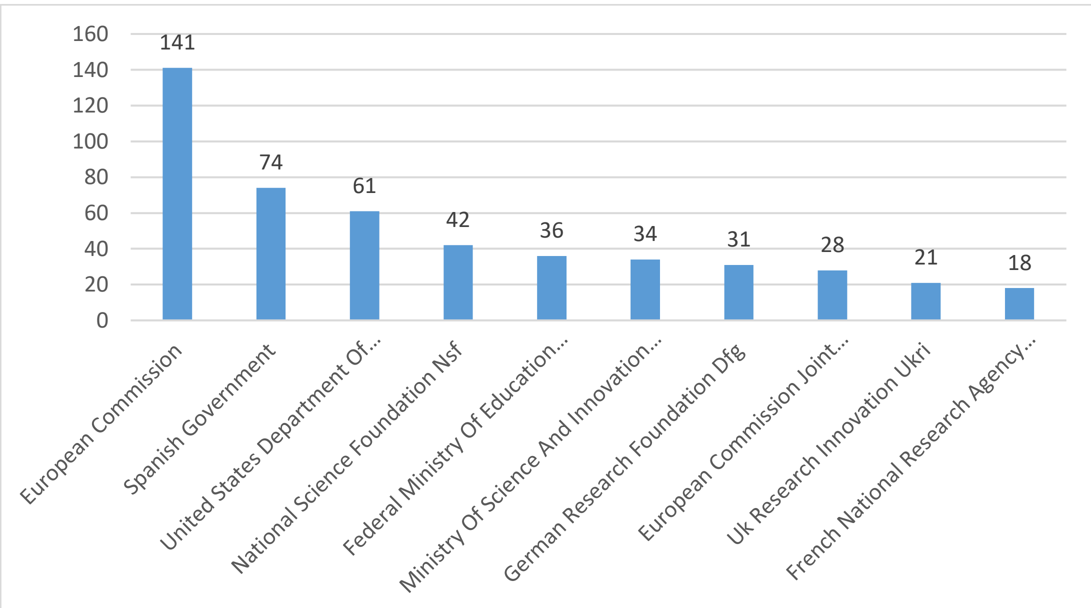

Figure 12: Top 10 funding agencies in EU with publication number

In this study, the driving forces for the convergence, which is RQ1, are determined. As determined in performance analysis and network analysis, the need for data analysis and data processing that emerged with the developments, especially in Industry 4.0 (Jagatheesaperumal et al., 2022), the increasing importance of energy efficiency issues, the advances in deep learning applications (Amodei et al., 2016), the growing complexity and volume of data being generated (Khan et al., 2017), and the need for fast and powerful computing (Y. Chen et al., 2014) are the main drivers of the convergence. Thematic evolution results indicate parallel results with thematic map.

Convergence has a significant impact on computing infrastructure, system architectures, and technologies (answer to RQ2). As determined in the thematic map and thematic evolution, some of the significant themes, such as 'workflow management', 'operating systems', 'virtual machines', 'data center', 'grid computing', 'resource allocation' , 'resource utilization', 'resource management', 'load balancing', 'fault tolerance', and 'container technology' lead us to answer of RQ2. The complexity of data-intensive and AI-driven applications emerging with convergence drives the need for more powerful and scalable computing infrastructure and the development of new HPC architectures. To accommodate the demands of these applications, system architectures must evolve to incorporate features such as distributed processing, parallelism, resource utilization, resource management, and resource allocation. This increases the need for more parallel processing, which can significantly improve performance. HPC architectures have changed to meet these demands by increasing the number of cores and processors. As mentioned in the top cited article Rank #5, HPC architectures will be not only multi-core, but also have heterogeneous technologies (Augonnet et al., 2009). Furthermore, new system architectures are being developed, such as those based on containerization (Abraham et al., 2020).

Various research areas and themes emerging from the convergence are analyzed (RQ3). These are shown in Figure 8 and Figure 9. The most prominent topics identified are 'computational fluid dynamics', 'molecular dynamics', 'malware detection', 'random forest', 'monte carlo', 'intelligent computing', 'transfer learning', and 'reinforcement learning'. Computational fluid dynamics is a branch of fluid mechanics that is often used for engineering and scientific applications. The use of HPC has become widespread in computational fluid dynamics for the analysis of complex flows and development of more accurate and efficient simulations (Lawson et al., 2012). The use of molecular dynamics has also become more popular due to improvements in computational power (Thompson et al., 2022). Advancement in machine learning has led to increased interest in malware detection as an important aspect of cybersecurity. Transfer learning, as an attractive option for many applications, is a machine learning technique that has the potential to reduce the amount of data and computation required to train a new model (T. Liu et al., 2019).

4.3.1. Research Implications

This thesis explored the convergence of HPC, big data, AI, and cloud computing technologies. The research contributes to the understanding of HPC and these other technologies by examining the convergence of these prominent technologies. It provided an in-depth understanding of how these technologies are being used together and identified potential future developments in this area. This study analyzed the convergence by implementing performance and network analysis.

Convergence represents a shift in the landscape of computational technologies and as such, reveals a significant rich area of study. This thesis is a study that tries to reveal the driving force of convergence, the changes in system architecture and technologies, and new research areas and themes revealed by convergence. Results of the research can be used in addition to demonstrating the latest developments in this field, and it can also be useful to plan future research projects and identify potential collaborations. The results of the study can be instrumental in enabling both academia and industry to identify areas of research and development that warrant further investigation and investment. These findings can facilitate the formulation of R&D strategies and priorities for organizations, as well as the identification of potential partners for collaboration. Results like publication patterns, citation rates, and identified institutions that emerge from the analysis can serve as a useful guide for academia in understanding the direction and impact of research. The findings of this study can be considered by organizations as they revise their strategic planning efforts. Overall, the results of this research can help to shape understanding of these technologies' convergence and their role in the larger technological landscape.

4.3.2. Future Implications

Previous research has established the trends and future of HPC as a standalone technology (Dongarra, 2004; Kindratenko & Trancoso, n.d.; Parnell et al., 2019; Strohmaier et al., 2005). This thesis explored the future of HPC in conjunction with big data, AI, and cloud computing technologies.

Various research fields, including agriculture (Georgiou et al., 2020), health (Purawat et al., 2021), industry (Hachinger et al., 2022), neuroscience (Madany et al., 2020), and various projects and platforms (Amodei et al., 2016; Buyya et al., 2009; Eicker et al., 2013; Gorelick et al., 2017; Milojicic et al., 2021; Narasimhamurthy et al., 2019; Strande et al., 2021; Thompson et al., 2022) (Gorelick et al., 2017) have been identified as a result of the literature review and bibliometric analysis of this study. This trend is projected to persist into the future, with a growth and diversification of these fields, initiatives, and platforms.

According to the result of the study, it is expected that the importance of data will increase in the future, and data-driven workflows will be used more in HPC-related research areas. In addition, the driving forces determined in the study, including energy consumption, the growing complexity, and volume of data, and the need for data analysis and data processing, may gradually increase their effect. In this case, HPC platforms, technologies, and research areas will reshape over time. It will lead to more dynamic, intelligent and powerful HPC platforms. Cloud-based computing platforms integrated with HPC techniques will likely become increasingly prevalent in the future. It is expected that cloud-native supercomputing will play a key role in the future. HPC platforms are being optimized for this approach and hybrid architectures that blend elements of both HPC and cloud architectures are being developed. These systems use a microservices architecture, which means that individual components are modular and self-contained, allowing for greater flexibility and agility in deploying and managing HPC workloads. The use of cloud-native technologies will make it easier to integrate new technologies and capabilities as they become available.

Another implication is that there will be a continued trend toward the integration of HPC systems within federated e-infrastructures (Alam et al., 2020). This involves bringing together HPC resources, cloud resources, and data infrastructure to better meet the needs of various research communities, provide more flexible access to a wider range of resources, and allow for access to a greater diversity of computing resources. Many initiatives towards the federation of e-infrastructure services such as Gaia-X and Fenix were mentioned in the literature review. It is anticipated that there will be growth in similar studies.

This convergence will likely drive a greater focus on the use of HPC in machine learning and deep learning approaches. There will likely be a need for research projects focused on developing new architectures that can accommodate this convergence. In the context of simulations such as molecular dynamics and computational fluid dynamics, efforts to accelerate complex calculations using HPC features such as parallelization, optimization, and specialized hardware will likely increase.

4.4. Evaluation of Bibliometric Analysis

4 steps were followed while performing bibliometric analysis. To evaluate the bibliometric analysis, the best practice guidelines provided by Donthu et al. (2021) were followed, as depicted in Table 9.

Table 9: Evaluation of Bibliometric Analysis based on Donthu et al. (2021)

No	Step	Evaluation Questions	Answers
Step 1	The Aims and Scope of the Bibliometric Study	1. What are the aims and scope of the study? 2. Is the scope of the study large enough to warrant the use of bibliometric analysis?	1. The aim of the study is to provide a comprehensive and up-to-date overview of the convergence of HPC with big data, AI, and cloud computing technologies, and to use bibliometric analysis to investigate and analyze the relationship between these technologies and how they are being influenced by each other. 2. After conducting numerous trials, keywords were attempted to select that would encompass the scope of the literature as much as possible. A separate query series for HPC (30,908 results) and other technologies (AI, big data, cloud computing) (1,061,537 results) was created, and the results were combined. The data obtained as a result of the queries was large enough to perform bibliometric analysis.
Step 2	Choosing the Techniques for Bibliometric Analysis	3. What bibliometric analysis techniques should be chosen to meet the aims and scope of the study?	3. Bibliometric analysis methods were chosen to meet the purpose and scope of this study and to answer research questions. Co-occurrence network analysis was chosen in order to identify the most important or influential concepts or ideas within a given field of research. Thematic map was used to identify the most important or influential themes within a given field of research and to provide a visual representation of the conceptual structure in different thematic areas. Thematic evolution was selected to understand how research topic has evolved over time, and to identify trends and shifts in research focus.
Step 3	Collecting the Data for Bibliometric Analysis	4. Do the search terms exemplify the scope of the study? 5. Is the coverage of the database adequate for the study? 6. Is the data free of errors such as duplicates and erroneous entries? 7. Does the final dataset fulfil the requirements of the bibliometric analysis techniques chosen for the study?	4. The search terms for this study were carefully selected to ensure that the scope of the investigation was adequately covered. A comprehensive search query was developed by including key words and phrases related to HPC as well as other technologies such as AI, big data, and cloud computing. This allowed us to identify relevant literature from a variety of sources and provided a comprehensive overview of the current state of HPC and related technologies. 5. The study followed the recommendations of Donthu et al. (2021) by only using one database in order to minimize potential human errors. According to Donthu et al. (2021), a dataset of 500 or more papers in a given research field is sufficient to justify the use of bibliometric analysis. The study conducted a bibliometric analysis of 3,748 documents, using a broad search query that included the title, author, keywords, abstract, keyword plus, and research area. 6. As recommended in the paper (Donthu et al., 2021), only one database is selected as minimizing the unnecessary action items can help minimize possible human errors. Furthermore, to ensure the data was accurate and free of errors such as duplicates and erroneous entries, the study iteratively updated and finalized a list of files that needed to be removed and synonyms by examining the results of various analyses. This process helped to purify the data. 7. In this study, bibliometric analysis was carried out with 3,748 documents (final dataset). As Donthu et al. (2021) says, if there are a significant number of papers, such as 500 or more (or even thousands) in a given research field, it can be considered large enough to justify the use of bibliometric analysis.
Step 4	Running the Bibliometric Analysis and Reporting the Findings	8. Can the bibliometric summary be easily understood by readers? 9. Does the writing align with the bibliometric summary presented? 10. Does the writing explain the peculiarities and implications of the bibliometric summary? 11. Does the writing align with the target outlet for publication?	8. The bibliometric summary was made easily understandable through the use of tables, figures, and explanations. 9. In Chapter 4, results of the performance analysis and network analysis were summarized in a step-by-step, orderly manner. Explanations were provided for all the resulting figures and tables, aligning with the bibliometric results. 10. The bibliometric analysis provides answers to the research questions posed in the study, using the results of the analysis as support. So the writing explains the peculiarities and implications of the bibliometric summary. 11. The bibliometric summary and accompanying comments, which are designed to answer the research questions and achieve the purpose of the thesis, are consistent and effective in meeting the aim of the thesis.

CHAPTER 5

CONCLUSION

5.1. Summary

The convergence of HPC with big data, AI, and cloud computing has the potential to greatly impact many fields. An examination of the graph depicting 3,748 publications by domain, as illustrated in Figure 13, reveals the presence of various research domains. It is evident that the highest number of publications is in the field of computer science, followed by engineering.

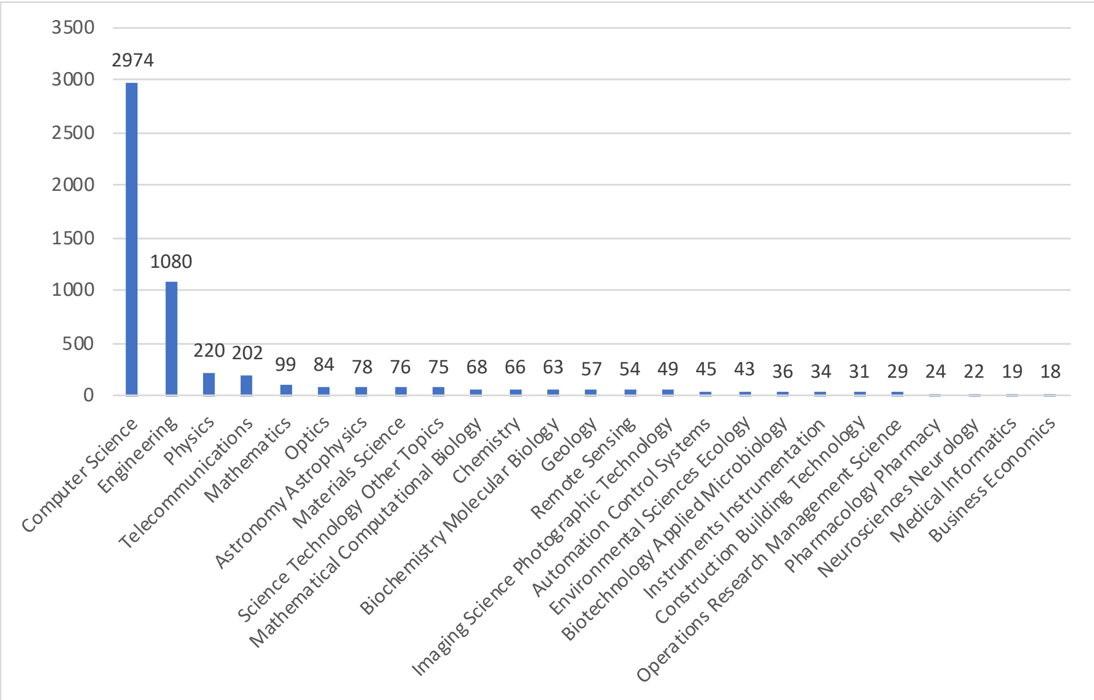

Figure 13: Research domain

It is important for both academia and industry to keep up to date on the latest developments and forecast promising technological trends in these areas. Currently, current approaches in different computational fields do not effectively address the convergence of HPC with big data, AI, and cloud computing or use bibliometric analysis in the field of HPC. It has been identified that there is a gap in the existing literature on this topic, and this study was initiated to address this gap. To address this issue, we need a new approach that uses bibliometric analysis to evaluate the convergence of HPC, big data, AI, and cloud computing in technology forecasting. This approach will improve our understanding of the interrelationships between these technologies, their overall impact on the field, and emerging trend topics.

The convergence of HPC with big data, AI, and cloud computing technologies has led to significant improvements in various fields. The results of the bibliometric analysis show that this convergence is driven by several factors, including the increasing demand for data analysis and processing, the rising importance of energy efficiency, the advancements in deep learning algorithms, the growing volume and complexity of data, and the need for fast and powerful computing capabilities.

The convergence of HPC with AI, big data, and cloud computing has had a significant impact on computing infrastructure, system architectures, and technologies. This convergence reveals the need for more powerful and scalable computing infrastructure that can support the requirements of these applications. In order to fulfill these requirements, the design of system architectures must progress to incorporate aspects such as distributed processing, parallelism, resource utilization, resource management, and resource allocation. HPC architectures are expected to become significantly multi-core and diverse to fulfill the growing demand for greater computational capacity (Augonnet et al., 2009). In addition, new system architectures based on containerization are being developed (Abraham et al., 2020). Hybrid architectures, which combine elements of both HPC and cloud architectures, are being developed to provide the computational power of HPC systems with the scalability and adaptability of cloud architectures (Zhou et al., 2021).

The research has also identified various research areas and themes emerging from the convergence, including computational fluid dynamics, molecular dynamics, malware detection, transfer learning, and reinforcement learning.

Ultimately, the use of HPC in conjunction with big data, AI, and cloud computing technologies is expected to continue to evolve and become increasingly integrated. These technologies will likely be used to solve complex problems and drive advancements in fields such as medicine, biology, physics, and finance, as well as in industries such as manufacturing, energy, and transportation (Terzo & Martinovič, 2022). There will also likely be a focus on improving the efficiency and performance of these technologies and reducing their environmental impact.

5.2. Contributions and Limitations

In this thesis, the main contributions are (1) Technological assessment of HPC, big data, AI, and cloud computing considering the driving forces, the effects of convergence on computing infrastructure technologies, the emerging research areas, and themes. (2) The first study to apply bibliometric analysis in the field of HPC. Accordingly, (3) A bibliometric analysis of 3748 publications related to performance and network analysis to develop a new approach in the research field. (4) Identification of the most influential authors, institutions, countries, and the most frequently cited publications in the field. (5) Identification and examination of the thematic map and thematic evolution through network analysis within the research field.

There are several limitations to this study. First, the choice of the database was limited to WoS, and future research may consider including other databases such as Scopus, PubMed, or a combination of these. Additionally, expanding the search criteria or considering different combinations of technologies may provide more in-depth insights into the convergence of these technologies. Finally, the bibliometric approach used in this study is only one way to analyze and forecast the trends in this field. Other approaches, such as systematic literature review, patent analysis, or machine learning techniques, may also be useful in identifying and predicting future developments in the convergence of these technologies.

5.3. Future Work

Technology assessment and forecasting are necessary for improving an organization's technology management capabilities and predicting potential technological shifts. By implementing these techniques, organizations can improve their technology management skills and make informed decisions about future technological changes.
Within the realm of academia, road-mapping studies have been utilized by organizations as a means of addressing data science challenges, attaining data analytics goals, and evaluating and forecasting emerging technologies. To achieve the most optimal results in these studies, experts recommend utilizing a combination of methods, such as bibliometric analysis and expert-based approaches (Daim et al., 2006, 2018).

For future work, a combination of bibliometric analysis and expert-based approaches can be used for technology management and road-mapping to strengthen the study's findings by providing a more comprehensive and robust understanding of the convergence of these technologies. Based on the results of the evaluation of the convergence of HPC with big data, AI, and cloud computing, it is projected that organizations will require updated strategic planning in this area. It can be utilized as a precursor to formulating strategic planning at various levels of data and technology (national, sectoral, organizational) while considering the market's driving forces. This can help organizations to make more informed decisions about technology roadmap and to better align their roadmaps according to current needs. By combining these two approaches, organizations can identify opportunities for innovation and growth and better understand the potential risks and challenges associated with this technology convergence.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., & Zhang, X. (2016). *TensorFlow: A system for large-scale machine learning*.

Abdelfattah, A., Anzt, H., Boman, E. G., Carson, E., Cojean, T., Dongarra, J., Fox, A., Gates, M., Higham, N. J., Li, X. S., Loe, J., Luszczek, P., Pranesh, S., Rajamanickam, S., Ribizel, T., Smith, B. F., Swirydowicz, K., Thomas, S., Tomov, S., … Yang, U. M. (2021). A survey of numerical linear algebra methods utilizing mixed-precision arithmetic. *International Journal of High Performance Computing Applications*, *35*(4), Article LLNL-JRNL-826451; NREL/JA-2C00-77220. https://doi.org/10.1177/10943420211003313

Abdelfattah, A., Anzt, H., Boman, E. G., Carson, E., Cojean, T., Dongarra, J., Gates, M., Grützmacher, T., Higham, N. J., Li, S., Lindquist, N., Liu, Y., Loe, J., Luszczek, P., Nayak, P., Pranesh, S., Rajamanickam, S., Ribizel, T., Smith, B., … Yang, U. M. (2020). *A Survey of Numerical Methods Utilizing Mixed Precision Arithmetic* (arXiv:2007.06674). arXiv. https://doi.org/10.48550/arXiv.2007.06674

Abraham, S., Paul, A. K., Khan, R. I. S., & Butt, A. R. (2020). On the Use of Containers in High Performance Computing Environments. *2020 IEEE 13th International Conference on Cloud Computing (CLOUD)*, 284–293. https://doi.org/10.1109/CLOUD49709.2020.00048

Adamuthe, A. C., & Thampi, G. T. (2019). Technology forecasting: A case study of computational technologies. *Technological Forecasting and Social Change*, *143*, 181–189. https://doi.org/10.1016/j.techfore.2019.03.002

Alam, S. R., Bartolome, J., Bassini, S., Carpene, M., Cestari, M., Combeau, F., Girona, S., Gorini, S., Fiameni, G., Hagemeier, B., Hater, T., Herten, A., Kiapidou, N., Klijn, W., Krause, D., Lafoucriere, J.-C., Leong, C., Leibovici, T., Lippert, T., … Wiber, G. (2020). Archival Data Repository Services to Enable HPC and Cloud Workflows in a Federated Research e-Infrastructure. *2020 IEEE/ACM International Workshop on Interoperability of Supercomputing and Cloud Technologies (SuperCompCloud)*, 39–44. https://doi.org/10.1109/SuperCompCloud51944.2020.00012

Alcácer, V., & Cruz-Machado, V. (2019). Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems. *Engineering Science and Technology, an International Journal*, *22*. https://doi.org/10.1016/j.jestch.2019.01.006

Alemzadeh, M. (2009). *A toolset to aid moab scheduler configuration* [Master thesis, University of Calgary]. https://doi.org/10.11575/PRISM/3054

Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., Casper, J., Catanzaro, B., Cheng, Q., Chen, G., Chen, J., Chen, J., Chen, Z., Chrzanowski, M., Coates, A., Diamos, G., Ding, K., Du, N., Elsen, E., … Zhu, Z. (2016). Deep speech 2: End-to-end speech recognition in English and mandarin. *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, 173–182.

Andújar, F. J., Coll, S., Alonso, M., Martínez, J.-M., López, P., Sánchez, J. L., & Alfaro, F. J. (2022). Energy efficient HPC network topologies with on/off links. *Future Generation Computer Systems*. https://doi.org/10.1016/j.future.2022.09.012

Ang, J., Barrett, B., Wheeler, K., & Murphy, R. (2010). *Introducing the graph 500*.

Antoniu, G., Valduriez, P., Hoppe, H.-C., & Krüger, J. (2021). *Towards Integrated Hardware/Software Ecosystems for the Edge-Cloud-HPC Continuum*.

Augonnet, C., Thibault, S., Namyst, R., & Wacrenier, P.-A. (2009). StarPU: A Unified Platform for Task Scheduling on Heterogeneous Multicore Architectures. In H. Sips, D. Epema, & H.-X. Lin (Eds.), *Euro-Par 2009 Parallel Processing* (pp. 863–874). Springer. https://doi.org/10.1007/978-3-642-03869-3_80

Beltre, A. M., Saha, P., Govindaraju, M., Younge, A., & Grant, R. E. (2019). Enabling HPC Workloads on Cloud Infrastructure Using Kubernetes Container Orchestration Mechanisms. *2019 IEEE/ACM International Workshop on Containers and New Orchestration Paradigms for Isolated Environments in HPC (CANOPIE-HPC)*, 11–20. https://doi.org/10.1109/CANOPIE-HPC49598.2019.00007

Bildosola, I., Rio-Bélver, R., Cilleruelo, E., & Gavilanes, J. (2017). Forecasting Cloud Computing: Producing a Technological Profile. In M. Amorim, C. Ferreira, M. Vieira Junior, & C. Prado (Eds.), *Engineering Systems and Networks* (pp. 23–31). Springer International Publishing. https://doi.org/10.1007/978-3-319-45748-2_3

Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, *1*(1), 363–376. https://doi.org/10.1162/qss_a_00018

Birrittella, M. S., Debbage, M., Huggahalli, R., Kunz, J., Lovett, T., Rimmer, T., Underwood, K. D., & Zak, R. C. (2016). Enabling Scalable High-Performance Systems with the Intel Omni-Path Architecture. *IEEE Micro*, *36*(4), 38–47. https://doi.org/10.1109/MM.2016.58

Borgman, C. L., & Rice, R. E. (1992). The convergence of information science and communication: A bibliometric analysis. *Journal of the American Society for Information Science*, *43*(6), 397–411. https://doi.org/10.1002/(SICI)1097-4571(199207)43:6<397::AID-ASI1>3.0.CO;2-M

Braud, A., Fromentoux, G., Radier, B., & Le Grand, O. (2021). The Road to European Digital Sovereignty with Gaia-X and IDSA. *IEEE Network*, *35*(2), 4–5. https://doi.org/10.1109/MNET.2021.9387709

Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. (2009). Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation Computer Systems-the International Journal of Escience*, *25*(6), 599–616. https://doi.org/10.1016/j.future.2008.12.001

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics*, *22*(1), 155–205. https://doi.org/10.1007/BF02019280

Carnielli, G., Arthur, R., Monteiro, A., França, R., & Iano, Y. (2021). *An Overview of High-Performance Computing Techniques Applied to Image Processing* (pp. 141–161). https://doi.org/10.1201/9781003201038-8

Caviggioli, F. (2016). Technology fusion: Identification and analysis of the drivers of technology convergence using patent data. *Technovation*, *55–56*, 22–32. https://doi.org/10.1016/j.technovation.2016.04.003

Chand Bhatt, P., Kumar, V., Lu, T.-C., & Daim, T. (2021). Technology convergence assessment: Case of blockchain within the IR 4.0 platform. *Technology in Society*, *67*, 101709. https://doi.org/10.1016/j.techsoc.2021.101709

Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, *19*(2), 171–209. https://doi.org/10.1007/s11036-013-0489-0

Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., Xiao, T., Xu, B., Zhang, C., & Zhang, Z. (2015). *MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems*.

Chen, Y., Luo, T., Liu, S., Zhang, S., He, L., Wang, J., Li, L., Chen, T., Xu, Z., Sun, N., & Temam, O. (2014). DaDianNao: A Machine-Learning Supercomputer. *2014 47th Annual IEEE/ACM International Symposium on Microarchitecture*, 609–622. https://doi.org/10.1109/MICRO.2014.58

Chen, Y., Shen, L., Zhang, X., & Chen, Y. (2022). Industry convergence and value innovation: A bibliometric analysis and systematic review. *Kybernetes, ahead-of-print*(ahead-of-print). https://doi.org/10.1108/K-01-2022-0097

Chiu, M., & Herbordt, M. (2010). Molecular Dynamics Simulations on High-Performance Reconfigurable Computing Systems. *ACM Transactions on Reconfigurable Technology and Systems*, *3*. https://doi.org/10.1145/1862648.1862653

Chun, K. W., Kim, H., & Lee, K. (2019). A Study on Research Trends of Technologies for Industry 4.0; 3D Printing, Artificial Intelligence, Big Data, Cloud Computing, and Internet of Things. In J. J. Park, V. Loia, K.-K. R. Choo, & G. Yi (Eds.), *Advanced Multimedia and Ubiquitous Engineering* (pp. 397–403). Springer. https://doi.org/10.1007/978-981-13-1328-8_51

Cui, W., Tang, J., Zhang, Z., & Dai, X. (2022). A bibliometric analysis on innovation convergence. *Library Hi Tech*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/LHT-12-2021-0430

Dagum, L., & Menon, R. (1998). OpenMP: An industry standard API for shared-memory programming. *IEEE Computational Science and Engineering*, *5*(1), 46–55. https://doi.org/10.1109/99.660313

Daim, T. U., Rueda, G., Martin, H., & Gerdsri, P. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting and Social Change*, *73*(8), 981–1012. https://doi.org/10.1016/j.techfore.2006.04.004

Daim, T. U., Yoon, B.-S., Lindenberg, J., Grizzi, R., Estep, J., & Oliver, T. (2018). Strategic roadmapping of robotics technologies for the power industry: A multicriteria technology assessment. *Technological Forecasting and Social Change*, *131*, 49–66. https://doi.org/10.1016/j.techfore.2017.06.006

Dongarra, J. (2004). Trends in High Performance Computing. *Computer Journal*, *47*. https://doi.org/10.1093/comjnl/47.4.399

Donovan, S., Symposium, L., Kleen, A., Hewlett-packard, M., Ibm, G., Hutton, A., Petersen, M., Source, W., & Schwan, P. (2003). *Lustre: Building a File System for 1,000-node Clusters*.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Dube, N., Roweth, D., Faraboschi, P., & Milojicic, D. (2021). Future of HPC: The Internet of Workflows. *IEEE Internet Computing*, *25*(5), 26–34. https://doi.org/10.1109/MIC.2021.3103236

Eicker, N., Lippert, T., Moschny, T., & Suarez, E. (2013). The DEEP Project—Pursuing Cluster-Computing in the Many-Core Era. *2013 42nd International Conference on Parallel Processing*, 885–892. https://doi.org/10.1109/ICPP.2013.105

Ejarque, J., Badia, R. M., Albertin, L., Aloisio, G., Baglione, E., Becerra, Y., Boschert, S., Berlin, J. R., D'Anca, A., Elia, D., Exertier, F., Fiore, S., Flich, J., Folch, A., Gibbons, S. J., Koldunov, N., Lordan, F., Lorito, S., Løvholt, F., … Volpe, M. (2022). Enabling dynamic and intelligent workflows for HPC, data analytics, and AI convergence. *Future Generation Computer Systems*, *134*, 414–429. https://doi.org/10.1016/j.future.2022.04.014

El-Alfy, E.-S. M., & Mohammed, S. A. (2020). A review of machine learning for big data analytics: Bibliometric approach. *Technology Analysis & Strategic Management*, *32*(8), 984–1005. https://doi.org/10.1080/09537325.2020.1732912

Fan, Y. (2021). *Job Scheduling in High Performance Computing* (arXiv:2109.09269). arXiv. http://arxiv.org/abs/2109.09269

Feldman, M. (2020, January 9). *HPC In 2020: AI Is No Longer An Experiment*. The Next Platform. https://www.nextplatform.com/2020/01/09/hpc-in-2020-ai-is-no-longer-an-experiment/

Feng, S., An, H., Li, H., Qi, Y., Wang, Z., Guan, Q., Li, Y., & Qi, Y. (2020). The technology convergence of electric vehicles: Exploring promising and potential technology convergence relationships and topics. *Journal of Cleaner Production*, *260*, 120992. https://doi.org/10.1016/j.jclepro.2020.120992

Fiore, S., Bakhouya, M., & Smari, W. W. (2018). On the road to exascale: Advances in High Performance Computing and Simulations—An overview and editorial. *Future Generation Computer Systems*, *82*, 450–458. https://doi.org/10.1016/j.future.2018.01.034

Gai, K., Qiu, M., Zhao, H., Tao, L., & Zong, Z. (2016). Dynamic energy-aware cloudlet-based mobile cloud computing model for green computing. *Journal of Network and Computer Applications*, *59*, 46–54. https://doi.org/10.1016/j.jnca.2015.05.016

Gao, H., & Ding, X. (2022). The research landscape on the artificial intelligence: A bibliometric analysis of recent 20 years. *Multimedia Tools and Applications*, *81*(9), 12973–13001. https://doi.org/10.1007/s11042-022-12208-4

Garg, D., Sidhu, J., & Rani, S. (2019). Emerging Trends in Cloud Computing Security: A Bibliometric Analyses. *IET Software*, *13*. https://doi.org/10.1049/iet-sen.2018.5222

Gauch, S., & Blind, K. (2015). Technological convergence and the absorptive capacity of standardisation. *Technological Forecasting and Social Change*, *91*, 236–249. https://doi.org/10.1016/j.techfore.2014.02.022

Georgiou, Y., Zhou, N., Zhong, L., Hoppe, D., Pospieszny, M., Papadopoulou, N., Nikas, K., Nikolos, O. L., Kranas, P., Karagiorgou, S., Pascolo, E., Mercier, M., & Velho, P. (2020). Converging HPC, Big Data and Cloud Technologies for Precision Agriculture Data Analytics on Supercomputers. In H. Jagode, H. Anzt, G. Juckeland, & H. Ltaief (Eds.), *High Performance Computing* (pp. 368–379). Springer International Publishing. https://doi.org/10.1007/978-3-030-59851-8_25

Gillick, D., Faria, A., & Denero, J. (2006). *MapReduce: Distributed Computing for Machine Learning*.

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. https://doi.org/10.1016/j.rse.2017.06.031

Grandinetti, L., Mirtaheri, S. L., & Shahbazian, R. (2018). *Big Data and HPC: Ecosystem and Convergence*. IOS Press.

Gropp, W., Lusk, E., Doss, N., & Skjellum, A. (1996). A high-performance, portable implementation of the MPI message passing interface standard. *Parallel Computing*, *22*(6), 789–828. https://doi.org/10.1016/0167-8191(96)00024-5

Hachinger, S., Golasowski, M., Martinovič, J., Hayek, M., García-Hernández, R. J., Slaninová, K., Levrier, M., Scionti, A., Donnat, F., Vitali, G., Magarielli, D., Goubier, T., Parodi, A., Parodi, A., Harsh, P., Dees, A., & Terzo, O. (2022). Leveraging High-Performance Computing and Cloud Computing with Unified Big-Data Workflows: The LEXIS Project. In E. Curry, S. Auer, A. J. Berre, A. Metzger, M. S. Perez, & S. Zillner (Eds.), *Technologies and Applications for Big Data Value* (pp. 159–180). Springer International Publishing. https://doi.org/10.1007/978-3-030-78307-5_8

Hager, G., & Wellein, G. (2010). *Introduction to High Performance Computing for Scientists and Engineers*. CRC Press.

Haleem, A., Mannan, B., Luthra, S., Kumar, S., & Khurana, S. (2018). Technology forecasting (TF) and technology assessment (TA) methodologies: A conceptual review. *Benchmarking: An International Journal*, *26*(1), 48–72. https://doi.org/10.1108/BIJ-04-2018-0090

Hao, T., Zhan, J., Hwang, K., Gao, W., & Wen, X. (2021). AI-oriented Workload Allocation for Cloud-Edge Computing. *2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid)*, 555–564. https://doi.org/10.1109/CCGrid51090.2021.00065

Haus, U.-U., Narasimhamurthy, S., Perez, M. S., Pleiter, D., & Wierse, A. (n.d.). *Federated HPC, Cloud and Data infrastructures*. 16.

Henderson, R. L. (1995). Job scheduling under the Portable Batch System. In D. G. Feitelson & L. Rudolph (Eds.), *Job Scheduling Strategies for Parallel Processing* (pp. 279–294). Springer. https://doi.org/10.1007/3-540-60153-8_34

Hong, L., Zhong-hua, L., & Xue-bin, C. (2010). The Applications and Trends of High Performance Computing in Finance. *2010 Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science*, 193–197. https://doi.org/10.1109/DCABES.2010.45

Huerta, E. A., Khan, A., Davis, E., Bushell, C., Gropp, W. D., Katz, D. S., Kindratenko, V., Koric, S., Kramer, W. T. C., McGinty, B., McHenry, K., & Saxton, A. (2020). Convergence of artificial intelligence and high performance computing on NSF-supported cyberinfrastructure. *Journal of Big Data*, *7*(1), 88. https://doi.org/10.1186/s40537-020-00361-2

Hussain, A., Jeon, J., & Rehman, M. (2022). Technological Convergence Assessment of the Smart Factory Using Patent Data and Network Analysis. *Sustainability*, *14*, 1668. https://doi.org/10.3390/su14031668

Jagatheesaperumal, S. K., Rahouti, M., Ahmad, K., Al-Fuqaha, A., & Guizani, M. (2022). The Duo of Artificial Intelligence and Big Data for Industry 4.0: Applications, Techniques, Challenges, and Future Research Directions. *Ieee Internet of Things Journal*, *9*(15), 12861–12885. https://doi.org/10.1109/JIOT.2021.3139827

Jeong, S., Kim, J.-C., & Choi, J. Y. (2015). Technology convergence: What developmental stage are we in? *Scientometrics*, *104*(3), 841–871. https://doi.org/10.1007/s11192-015-1606-6

Jiang, Z., Gao, W., Wang, L., Xiong, X., Zhang, Y., Wen, X., Luo, C., Ye, H., Lu, X., Zhang, Y., Feng, S., Li, K., Xu, W., & Zhan, J. (2019). HPC AI500: A Benchmark Suite for HPC AI Systems. In C. Zheng & J. Zhan (Eds.), *Benchmarking, Measuring, and Optimizing* (pp. 10–22). Springer International Publishing. https://doi.org/10.1007/978-3-030-32813-9_2

Kang, D., Jang, W., Lee, H., & No, H. J. (2013). *A Review on Technology Forecasting Methods and Their Application Area*. *7*(4), 5.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004

Khan, M., Wu, X., Xu, X., & Dou, W. (2017). Big data challenges and opportunities in the hype of Industry 4.0. *2017 IEEE International Conference on Communications (ICC)*, 1–6. https://doi.org/10.1109/ICC.2017.7996801

Kim, J., & Lee, S. (2017). Forecasting and identifying multi-technology convergence based on patent data: The case of IT and BT industries in 2020. *Scientometrics*, *111*(1), 47–65. https://doi.org/10.1007/s11192-017-2275-4

Kim, T. S., & Sohn, S. Y. (2020). Machine-learning-based deep semantic analysis approach for forecasting new technology convergence. *Technological Forecasting and Social Change*, *157*, 120095. https://doi.org/10.1016/j.techfore.2020.120095

Kindratenko, V., & Trancoso, P. (n.d.). *Trends in High-Performance Computing*. 4.

Knani, M., Echchakoui, S., & Ladhari, R. (2022). Artificial intelligence in tourism and hospitality: Bibliometric analysis and research agenda. *International Journal of Hospitality Management*, *107*, 103317. https://doi.org/10.1016/j.ijhm.2022.103317

Kohlhoff, K. J., Shukla, D., Lawrenz, M., Bowman, G. R., Konerding, D. E., Belov, D., Altman, R. B., & Pande, V. S. (2014). Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nature Chemistry*, *6*(1), Article 1. https://doi.org/10.1038/nchem.1821

Kose, T., & Sakata, I. (2019). Identifying technology convergence in the field of robotics research. *Technological Forecasting and Social Change*, *146*, 751–766. https://doi.org/10.1016/j.techfore.2018.09.005

Kostoff, R. N. (1995). Research requirements for research impact assessment. *Research Policy*, *24*(6), 869–882. https://doi.org/10.1016/0048-7333(94)00798-5

Kratzke, N., & Quint, P.-C. (2017). Understanding cloud-native applications after 10 years of cloud computing—A systematic mapping study. *Journal of Systems and Software*, *126*, 1–16. https://doi.org/10.1016/j.jss.2017.01.001

Kunkel, J. M., Bent, J., Lofstead, J., & Markomanolis, G. S. (n.d.). *Establishing the IO-500 Benchmark* (p. 1).

Kurtzer, G. M., Sochat, V., & Bauer, M. W. (2017). Singularity: Scientific containers for mobility of compute. *PLOS ONE*, *12*(5), e0177459. https://doi.org/10.1371/journal.pone.0177459

Lang, S. T. K., Dawson, A., Diamantakis, M., Dueben, P., Hatfield, S., Leutbecher, M., Palmer, T., Prates, F., Roberts, C. D., Sandu, I., & Wedi, N. (2021). More accuracy with less precision. *Quarterly Journal of the Royal Meteorological Society*, *147*(741), 4358–4370. https://doi.org/10.1002/qj.4181

Lange, C., Barthelmäs, P., Rosnitschek, T., Tremmel, S., & Rieg, F. (2021). Impact of HPC and Automated CFD Simulation Processes on Virtual Product Development—A Case Study. *Applied Sciences*, *11*, 6552. https://doi.org/10.3390/app11146552

Lawson, S. J., Woodgate, M., Steijl, R., & Barakos, G. N. (2012). High performance computing for challenging problems in computational fluid dynamics. *Progress in Aerospace Sciences*, *52*, 19–29. https://doi.org/10.1016/j.paerosci.2012.03.004

Lee, C., Hong, S., & Kim, J. (2021). Anticipating multi-technology convergence: A machine learning approach using patent information. *Scientometrics*, *126*(3), 1867–1896. https://doi.org/10.1007/s11192-020-03842-6

Lee, W. S., Han, E. J., & Sohn, S. Y. (2015). Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents. *Technological Forecasting and Social Change*, *100*, 317–329. https://doi.org/10.1016/j.techfore.2015.07.022

Li, Y., Xu, Z., Wang, X., & Wang, X. (2020). A bibliometric analysis on deep learning during 2007–2019. *International Journal of Machine Learning and Cybernetics*, *11*. https://doi.org/10.1007/s13042-020-01152-0

Lim, D.-J., Anderson, T. R., & Shott, T. (2015). Technological forecasting of supercomputer development: The March to Exascale computing. *Omega*, *51*, 128–135. https://doi.org/10.1016/j.omega.2014.09.009

List Statistics | TOP500. (n.d.). Retrieved September 5, 2022, from https://www.top500.org/statistics/list/

Liu, J., Wu, J., Kini, S. P., Wyckoff, P., & Panda, D. K. (2003). High performance RDMA-based MPI implementation over InfiniBand. *Proceedings of the 17th Annual International Conference on Supercomputing*, 295–304. https://doi.org/10.1145/782814.782855

Liu, T., Alibhai, S., Wang, J., Liu, Q., He, X., & Wu, C. (2019). Exploring Transfer Learning to Reduce Training Overhead of HPC Data in Machine Learning. *2019 IEEE International Conference on Networking, Architecture and Storage (NAS)*, 1–7. https://doi.org/10.1109/NAS.2019.8834723

Liu, Z., Liu, Y., Guo, Y., & Wang, H. (2013). Progress in global parallel computing research: A bibliometric approach. *Scientometrics*, *95*(3), 967–983. https://doi.org/10.1007/s11192-012-0927-y

Madany, M., Marcus, K., Peltier, S., Ellisman, M. H., & Altintas, I. (2020). NeuroKube: An Automated and Autoscaling Neuroimaging Reconstruction Framework using Cloud Native Computing and A.I. *2020 IEEE International*

Conference on Big Data (Big Data), 320–330. https://doi.org/10.1109/BigData50022.2020.9378053

McKinsey Technology Trends Outlook 2022 | McKinsey. (n.d.). Retrieved August 25, 2022, from https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/the-top-trends-in-tech

Mell, P., & Grance, T. (n.d.). *The NIST Definition of Cloud Computing*. 7.

Meng, X., Bradley, J., Yavuz, B., Sparks, E., Venkataraman, S., Liu, D., Freeman, J., Tsai, D. B., Amde, M., Owen, S., Xin, D., Xin, R., Franklin, M. J., Zadeh, R., Zaharia, M., & Talwalkar, A. (2015). *MLlib: Machine Learning in Apache Spark* (arXiv:1505.06807). arXiv. https://doi.org/10.48550/arXiv.1505.06807

Milojicic, D., Faraboschi, P., Dube, N., & Roweth, D. (2021). Future of HPC: Diversifying Heterogeneity. *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 276–281. https://doi.org/10.23919/DATE51398.2021.9474063

Mishra, S., Deshmukh, S. G., & Vrat, P. (2002). Matching of technological forecasting technique to a technology. *Technological Forecasting and Social Change*, *69*(1), 1–27. https://doi.org/10.1016/S0040-1625(01)00123-8

Morris, S., DeYong, C., Wu, Z., Salman, S., & Yemenu, D. (2002). diva: A visualization system for exploring document databases for technology forecasting. *Computers & Industrial Engineering*, *43*(4), 841–862. https://doi.org/10.1016/S0360-8352(02)00143-2

Narasimhamurthy, S., Danilov, N., Wu, S., Umanesan, G., Markidis, S., Rivas-Gomez, S., Peng, I. B., Laure, E., Pleiter, D., & Witt, S. de. (2019). SAGE: Percipient Storage for Exascale Data Centric Computing. *Parallel Computing*, *83*, 22–33. https://doi.org/10.1016/j.parco.2018.03.002

Nguyen, H.-D., TRAN, K. P., Zeng, X., Koehl, L., & Castagliola, P. (2019, July 1). *Industrial Internet of Things, Big Data, and Artificial Intelligence in the Smart Factory: A survey and perspective*.

Nurmi, D., Wolski, R., Grzegorczyk, C., Obertelli, G., Soman, S., Youseff, L., & Zagorodnov, D. (2009). The Eucalyptus Open-Source Cloud-Computing System. *2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid*, 124–131. https://doi.org/10.1109/CCGRID.2009.93

Ongsulee, P. (2017). Artificial intelligence, machine learning and deep learning. *2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)*, 1–6. https://doi.org/10.1109/ICTKE.2017.8259629

Parasyris, K., Laguna, I., Menon, H., Schordan, M., Osei-Kuffuor, D., Georgakoudis, G., Lam, M. O., & Vanderbruggen, T. (2020). HPC-MixPBench: An HPC

Benchmark Suite for Mixed-Precision Analysis. *2020 IEEE International Symposium on Workload Characterization (IISWC)*, 25–36. https://doi.org/10.1109/IISWC50251.2020.00012

Parnell, L. A., Demetriou, D. W., Kamath, V., & Zhang, E. Y. (2019). Trends in High Performance Computing: Exascale Systems and Facilities Beyond the First Wave. *2019 18th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, 167–176. https://doi.org/10.1109/ITHERM.2019.8757229

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (pp. 8026–8037). Curran Associates Inc.

Patrizio, A. (2018, December 3). *IDC: Expect 175 zettabytes of data worldwide by 2025*. Network World. https://www.networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html

Penteado, B. E., Fornazin, M., & Castro, L. (2021). The Evolution of Artificial Intelligence in Medical Informatics: A Bibliometric Analysis. *Progress in Artificial Intelligence: 20th EPIA Conference on Artificial Intelligence, EPIA 2021, Virtual Event, September 7–9, 2021, Proceedings*, 121–133. https://doi.org/10.1007/978-3-030-86230-5_10

Peskin, A. M. (1980). *Computer technology forecasting at the National Laboratories* (BNL-28153; CONF-801110-1). Brookhaven National Lab., Upton, NY (USA). https://www.osti.gov/biblio/5115058

Porter, A. L., & Detampel, M. J. (1995). Technology opportunities analysis. *Technological Forecasting and Social Change*, *49*(3), 237–255. https://doi.org/10.1016/0040-1625(95)00022-3

Porter, A. L., Roper, A. T., Mason, T. W., Rossini, F. A., & Banks, J. (1991). *Forecasting and Management of Technology*. John Wiley & Sons.

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation*, *25*, 348.

Purawat, S., Dasgupta, S., Song, J., Davis, S., Claypool, K. T., Chandra, S., Mason, A., Viswanath, V., Klein, A., Kasl, P., Wen, Y., Smarr, B., Gupta, A., & Altintas, I. (2021). TemPredict: A Big Data Analytical Platform for Scalable Exploration and Monitoring of Personalized Multimodal Data for COVID-19. *2021 IEEE International Conference on Big Data (Big Data)*, 4411–4420. https://doi.org/10.1109/BigData52589.2021.9671441

Ravikumar, S., Agrahari, A., & Singh, S. N. (2015). Mapping the intellectual structure of scientometrics: A co-word analysis of the journal Scientometrics (2005–2010). *Scientometrics*, *102*(1), 929–955. https://doi.org/10.1007/s11192-014-1402-8

Samadi, S. (2022). The convergence of AI, IoT, and big data for advancing flood analytics research. *Frontiers in Water*, *4*. https://www.frontiersin.org/articles/10.3389/frwa.2022.786040

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

Sevilla, J., Heim, L., Ho, A., Besiroglu, T., Hobbhahn, M., & Villalobos, P. (2022). *Compute Trends Across Three Eras of Machine Learning* (arXiv:2202.05924). arXiv. https://doi.org/10.48550/arXiv.2202.05924

Sharma, P., & Jadhao, V. (2021). *Molecular Dynamics Simulations on Cloud Computing and Machine Learning Platforms* (arXiv:2111.06466). arXiv. https://doi.org/10.48550/arXiv.2111.06466

Shibata, M., Ohtsuka, Y., Takahashi, M., & Okamoto, K. (2018). Advanced FPGA technology trend based on patent analysis with link mining. *2018 International Conference on Electronics Packaging and IMAPS All Asia Conference (ICEP-IAAC)*, 147–151. https://doi.org/10.23919/ICEP.2018.8374691

Singh, S., Sharma, P. K., Yoon, B., Shojafar, M., Cho, G. H., & Ra, I.-H. (2020). Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city. *Sustainable Cities and Society*, *63*, 102364. https://doi.org/10.1016/j.scs.2020.102364

Sochat, V., & Scott, A. (2021). Collaborative Container Modules with Singularity Registry HPC. *Journal of Open Source Software*, *6*(63), 3311. https://doi.org/10.21105/joss.03311

Song, C. H., Elvers, D., & Leker, J. (2017). Anticipation of converging technology areas—A refined approach for the identification of attractive fields of innovation. *Technological Forecasting and Social Change*, *116*, 98–115. https://doi.org/10.1016/j.techfore.2016.11.001

Staples, G. (2006). TORQUE resource manager. *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, 8-es. https://doi.org/10.1145/1188455.1188464

Sterling, T., Brodowicz, M., & Anderson, M. (2017). *High Performance Computing: Modern Systems and Practices*. Morgan Kaufmann.

Stone, J., Gohara, D., & Shi, G. (2010). OpenCL: A Parallel Programming Standard for Heterogeneous Computing Systems. *Computing in Science & Engineering, 12*, 66–72. https://doi.org/10.1109/MCSE.2010.69

Strande, S., Altintas, I., Cai, H., Cooper, T., Irving, C., Kandes, M., Majumdar, A., Mishin, D., Perez, I., Pfeiffer, W., Shantharam, M., Sinkovits, R. S., Sivagnanam, S., Tatineni, M., Thomas, M., Wolter, N., & Norman, M. (2021). Expanse: Computing without Boundaries. *Practice & Experience in Advanced Research Computing (PEARC)*. https://par.nsf.gov/biblio/10291594-expanse-computing-without-boundaries

Strey, A. (2001). High Performance Computing. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 6693–6697). Pergamon. https://doi.org/10.1016/B0-08-043076-7/00574-X

Strohmaier, E., Dongarra, J. J., Meuer, H. W., & Simon, H. D. (2005). Recent trends in the marketplace of high performance computing. *Parallel Computing, 31*(3), 261–273. https://doi.org/10.1016/j.parco.2005.02.001

Sukumar, S. R., Balma, J. A., Rickett, C. D., Maschhoff, K. J., Landman, J., Yates, C. R., Chittiboyina, A. G., Peterson, Y. K., Vose, A., Byler, K., Baudry, J., & Khan, I. A. (2022). The Convergence of HPC, AI and Big Data in Rapid-Response to the COVID-19 Pandemic. In J. Nichols, A. ‘Barney’ Maccabe, J. Nutaro, S. Pophale, P. Devineni, T. Ahearn, & B. Verastegui (Eds.), *Driving Scientific and Engineering Discoveries Through the Integration of Experiment, Big Data, and Modeling and Simulation* (pp. 157–172). Springer International Publishing. https://doi.org/10.1007/978-3-030-96498-6_9

Sunyaev, A. (2020). Cloud Computing. In A. Sunyaev (Ed.), *Internet Computing: Principles of Distributed Systems and Emerging Internet-Based Technologies* (pp. 195–236). Springer International Publishing. https://doi.org/10.1007/978-3-030-34957-8_7

Terzo, O., & Martinovič, J. (2022). *HPC, Big Data, and AI Convergence Towards Exascale: Challenge and Vision* (1st ed.). CRC Press. https://doi.org/10.1201/9781003176664

Thompson, A. P., Aktulga, H. M., Berger, R., Bolintineanu, D. S., Brown, W. M., Crozier, P. S., in ’t Veld, P. J., Kohlmeyer, A., Moore, S. G., Nguyen, T. D., Shan, R., Stevens, M. J., Tranchida, J., Trott, C., & Plimpton, S. J. (2022). LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications, 271*, 108171. https://doi.org/10.1016/j.cpc.2021.108171

Usman, S., Mehmood, R., & Katib, I. (2018). Big Data and HPC Convergence: The Cutting Edge and Outlook. In R. Mehmood, B. Bhaduri, I. Katib, & I. Chlamtac (Eds.), *Smart Societies, Infrastructure, Technologies and Applications* (pp. 11–

26). Springer International Publishing. https://doi.org/10.1007/978-3-319-94180-6_4

Verbraeken, J., Wolting, M., Katzy, J., Kloppenburg, J., Verbelen, T., & Rellermeyer, J. S. (2021). A Survey on Distributed Machine Learning. *ACM Computing Surveys*, *53*(2), 1–33. https://doi.org/10.1145/3377454

Veroneze Solórzano, A. L., & Mello Schnorr, L. (2022). *Understanding Distributed Deep Learning Performance by Correlating HPC and Machine Learning Measurements*. https://doi.org/10.1007/978-3-031-07312-0_14

Wang, W., & Siau, K. (2019). Artificial Intelligence, Machine Learning, Automation, Robotics, Future of Work and Future of Humanity: A Review and Research Agenda. *Journal of Database Management*, *30*, 61–79. https://doi.org/10.4018/JDM.2019010104

Wolstencroft, K., Haines, R., Fellows, D., Williams, A., Withers, D., Owen, S., Soiland-Reyes, S., Dunlop, I., Nenadic, A., Fisher, P., Bhagat, J., Belhajjame, K., Bacall, F., Hardisty, A., Nieva de la Hidalga, A., Balcazar Vargas, M. P., Sufi, S., & Goble, C. (2013). The Taverna workflow suite: Designing and executing workflows of Web Services on the desktop, web or in the cloud. *Nucleic Acids Research*, *41*(W1), W557–W561. https://doi.org/10.1093/nar/gkt328

Xing, Y., & Zhan, Y. (2012). Virtualization and Cloud Computing. In Y. Zhang (Ed.), *Future Wireless Networks and Information Systems* (pp. 305–312). Springer. https://doi.org/10.1007/978-3-642-27323-0_39

Yang, L. T., & Guo, M. (2005). *High-Performance Computing: Paradigm and Infrastructure*. John Wiley & Sons.

Yi, G., & Loia, V. (2019). High-performance computing systems and applications for AI. *The Journal of Supercomputing*, *75*(8), 4248–4251. https://doi.org/10.1007/s11227-019-02937-z

Yoo, A. B., Jette, M. A., & Grondona, M. (2003). SLURM: Simple Linux Utility for Resource Management. In D. Feitelson, L. Rudolph, & U. Schwiegelshohn (Eds.), *Job Scheduling Strategies for Parallel Processing* (pp. 44–60). Springer. https://doi.org/10.1007/10968987_3

Yu, J., Yang, Z., Zhu, S., Xu, B., Li, S., & Zhang, M. (2018). A Bibliometric Analysis of Cloud Computing Technology Research. In B. Xu (Ed.), *Proceedings of 2018 Ieee 3rd Advanced Information Technology, Electronic and Automation Control Conference (iaeac 2018)* (pp. 2353–2358). Ieee. https://www.webofscience.com/wos/woscc/full-record/WOS:000457576900466

Zhang, J., & Lin, M. (2022). A comprehensive bibliometric analysis of Apache Hadoop from 2008 to 2020. *International Journal of Intelligent Computing and Cybernetics*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/IJICC-01-2022-0004

Zhou, N., Georgiou, Y., Pospieszny, M., Zhong, L., Zhou, H., Niethammer, C., Pejak, B., Marko, O., & Hoppe, D. (2021). Container orchestration on HPC systems through Kubernetes. *Journal of Cloud Computing*, *10*(1), Article 1. https://doi.org/10.1186/s13677-021-00231-z

Zhu, C., & Motohashi, K. (2022). Identifying the technology convergence using patent text information: A graph convolutional networks (GCN)-based approach. *Technological Forecasting and Social Change*, *176*, 121477. https://doi.org/10.1016/j.techfore.2022.121477

Zupic, I., & Čater, T. (2015). Bibliometric Methods in Management and Organization. *Organizational Research Methods*, *18*(3), 429–472. https://doi.org/10.1177/1094428114562629