8-10 June 2015- Istanbul, Turkey
Proceedings of SOCIOINT15- 2nd International Conference on Education, Social Sciences and Humanities

573

A STUDY OF CLOUD COMPUTING BASED MODEL TO FACILITATE COLLABORATIVE RESEARCH

Rahul Wargad[1*] and Manimala Puri[2]

[1]Rearch Scholar, Savitri Bai Phule Pune University, India rahulbw@gmail.com
[2] Prof. Dr., JSPM, Savitri Bai Phule Pune University, India manimalap@yahoo.com
*Corresponding author

Abstract

This paper examines how Cloud computing technology can facilitate collaborative research amongst the researchers belonging to different universities and countries. Research over has evolved from simple, controlled research to very complex having multiple dimensions. With the fast development in areas, it is becoming increasingly difficult for individual researchers to conduct innovative research and trailblaze on their own. This paper is conceptual in nature wherein qualitative method has been used to substantiate the significant issues in collaborative research. An attempt is made to explore the possible solutions to facilitate collaborative research with the help of Cloud computing technology in order to make certain vital observations to lay down conclusion. For hundreds of years hunam beings have been conducting research independently or in various universities or labs. However in recent years due to the advent of advanced communication technologies the world has became a global village. Reseachers in various parts of world are involved in research in various areas with or without much experience. Cloud computing in Collaborative research can facilitate (a) Combined efforts of many good researchers for excellent outcomes, (b) Creativity, (c) Sharing experience of all researchers, (d) Get fundings from multiple organizations from different countries, (e) Sharing of techninques, (f) Cost effective use of common computing resources.

The findings of this paper contribute to literature on Cloud computing technology for collaborative research. In addition, the research has focussed on the benefits of collaborative research which adds value and is key aspect of the novelty of this research.

Keywords: Cloud computing , Collaborative research, Cloud computing technology in research.

1. INTRODUCTION

Collaboration at different levels, at inter-institutional and international collaboration's exited for a long time however it had lot of limitations. Cloud computing technology can built strong bonds collaborations.

Global high capacity bandwidth connectivity, powerful computing resources and digitalization is changing the landscape of collaboration in research. Collaboration at global scale are vital for various common problems like climate change, poverty and common curiosity in the field of astronomy and astrophysics and common requirements in areas of biomolecular simulation, & drug design, and big data analysis for global business needs.. Cloud computing gives mandate for collaborative research and quality higher education for all, where in content and resources like computing, storage, infrastructure, platform and softwares can be shared. As a disruptive technology cloud computing can transform the very nature of how education and research use information technology.

8-10 June 2015- Istanbul, Turkey
Proceedings of SOCIOINT15- 2nd International Conference on Education, Social Sciences and Humanities

574

Cloud computing has many benefits like speeding up scientific discovery, doing collaborative research. Cloud computing has ability to provide resources that meet the requirements of a much wider range of researchers.

Cloud computing has been envisioned as the technology for researchers to collaborate and think beyond the boundaries of our planet. Today researchers are tied up with the boundaries of lab, research institute, university or country. Cloud can break these boundries and make researchers collaborate and work globally for common objectives of mankind. Research will get a new identity of planet earth research by some one from our universe or other universe from our or any other milky way.

With cloud technology numerous initiatives can be taken with the aim of developing collaboration among individual researchers working in same domain by bringing them together under one umbrella of either larger centres of excellence irrespective of their locations, and bringing together researchers for interdisciplinary research.

2. NEED FOR COLLABORATION IN RESEARCH

The term "Research Collaboration" is used to describe the relationship between individuals but also relationship between organizations and between individuals with organization. Collaboration can be defined as "Social process whereby human beings pool their experience, knowledge and social skills with the objective of producing new knowledge, including knowledge as embedded in technology". (Bozeman, 2014, p.2). Collaboration usually involves two or more researchers within an institution or in different institutions, working either in the same field or in different scientific fields or sectors of the economy

A number of factors have contributed to a trend towards increased collaboration. The growing complexity and cost of research, especially in disciplines requiring specialized instrumentation or facilities, tend to take collaboration an imperative. Similarly, interdisciplinary and multidisciplinary research has assumed greater significance, as new ideas are explored at the intersection of previously distinct fields and as public expectations about the outcome of research look to more holistic and global approaches to research challenges. Finally, issues such as climate change or global pandemic health challenges, which by their nature transcend national borders, demand international collaboration.

Collaboration facilitates the sharing of meaning and completion of tasks with respect to a mutually shared super ordinate goal, and which takes place in a particular social or work setting (Sonnenwald, 2000,p.461). Interdisciplinary collaboration occurs when people with different educational and research backgrounds bring complementary skills to bear on a problem or task. Interdisciplinary collaboration is often needed to address complex problems, the solutions to which require knowledge and resources from more than one discipline. Clark (Clark, 1995) identified over 8530 disciplines and fields of knowledge. During interdisciplinary scientific collaboration, scientists exchange tools and methods, utilize diverse paradigms and cross-fertilize disciplinary concepts (Klein, 1994, p.7).

In the areas of High-energy physics, human-genome project, social impact of computing, telemedicine etc. the questions that researchers ask are becoming more and more complex. Collaborative efforts are required to investigate to answer these complex questions. Physicists throughout the world pool resources to solve the problems related to high energy physics, nuclear physics , bio molecular simulation, astronomy and astrophysics and work together to develop models, methods and software tools.

Normally a significant amount of efforts are is required to carry out research. An attempt to pool individual talent larger group collaborations helps in speeding up of research and gaining more credibility. In one of the reputed journal "S*cience"* a single drug discovery study may include experiments conducted by synthetic chemists, x-ray crystallographers, cell biologists, and behavioural pharmacologists.

Researchers at different institutions may work separately and yet collaborate on a particular project. The need could be because the researchers will be working on different aspects of the same project, exchanging data, compiling data for the entire project, and then doing joint data analysis, reporting, and publication. Clinical trials are a particular type of collaboration that occurs in the biomedical field. They involve the testing of experimental treatments, medications, or devices on human participants. Frequently, several sites and many researchers will be involved. Another requirement of academicians could be to collaborate with industry. With collaboration researcher, can get access for hardware and software resources, sharing of information such that both researchers make intellectual contributions to the project.

Some important factors facilitating of collaborative research include, funding from multiple organizations, different skill sets to work on various dimensions of research, connectivity with other research institutions, technology and sharing of resources...

8-10 June 2015- Istanbul, Turkey
Proceedings of SOCIOINT15- 2ⁿᵈ International Conference on Education, Social Sciences and Humanities

575

3. CLOUD COMPUTING TECHNOLOGY

Cloud computing can be either a centralized or a distributed computing system. Clouds can be built with physical or virtualized resources over large data centres that are centralized or distributed. Some authors consider cloud computing to be a form of utility computing or service computing (Kai Hwang, 2012, p.31-80).

3.1 The Cloud service models

The three cloud service models are:

3.1.1   Infrastructure as a Service (IaaS)
This model puts together infrastructures demanded by users—namely servers, storage, networks, and the data centre fabric. The user can deploy his applications and run on multiple virtual machines The user does not manage or control the underlying cloud infrastructure, but can specify when to request and release the needed resources.

3.1.2   Platform as a Service (PaaS)
This model enables the user to deploy user-built applications onto a virtualized cloud platform. PaaS includes middleware, databases, development tools. The platform includes both hardware and software integrated with specific programming interfaces. The provider supplies the API and software tools (e.g., Java, Python, Web 2.0, .NET). The user is freed from managing the cloud infrastructure.

3.1.3 Software as a Service (SaaS)
This refers to browser-initiated application software over thousands of paid cloud customers. The SaaS model applies to business processes, industry applications, consumer relationship management (CRM), enterprise resources planning (ERP), human resources (HR), and collaborative applications. On the customer side, there is no upfront investment in servers or software licensing. On the provider side, costs are rather low, compared with conventional hosting of user applications.

3.2 Transparency in Programming Environments

Cloud computing provides a transparent computing infrastructure for future computing platforms. The user data, applications, OS, and hardware are separated into four levels.
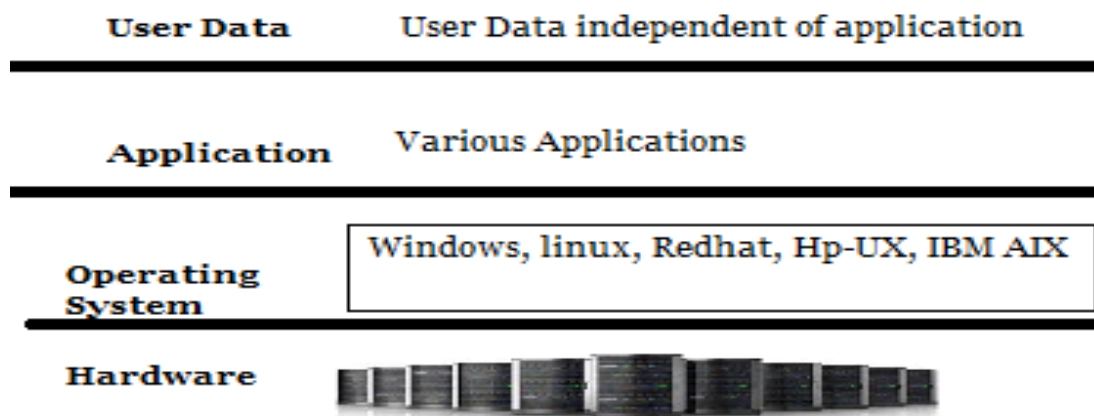

Fig. 1 A transparent computing environment that separates the user data, application, OS, and hardware in time and space – an ideal model of cloud computing for research.

Data is owned by users, independent of the applications. The OS provides clear interfaces, standard programming interfaces, or system calls to application programmers. In future cloud infrastructure, the hardware will be separated by standard interfaces from the OS. Thus, users will be able to choose from different OS on top of the hardware devices they prefer to use. To separate user data from specific application programs, users can enable cloud applications as SaaS. Thus, users can switch among different services. The data will not be bound to specific applications. The figure above shows how transparency can be achieved with cloud:

8-10 June 2015- Istanbul, Turkey
Proceedings of SOCIOINT15- 2ⁿᵈ International Conference on Education, Social Sciences and Humanities

576

3.3 Scalability of Cloud Resources

The researcher needs a system that can achieve scalable performance. Procuring peak load hardware in research labs is not cost-effective. Cloud computing supports static or dynamic scaling of resources. The following parameter of cloud support scalability of computing resources. (Kai Hwang, 2012, pp.50-80).

3.3.1 Size scalability

This refers to achieving higher performance or more functionality by increasing the machine size. The word "size" refers to adding processors, cache, memory, storage, or I/O channels. The most obvious way to determine size scalability is to simply count the number of processors installed. Not all parallel computer or distributed architectures are equally size scalable. For example, the IBM Blue Gene/L system scaled up to 65,000 processors.

3.3.2 Software scalability

This refers to upgrades in the OS or compilers, adding mathematical and engineering libraries, porting new application software, and installing more user-friendly programming environments. Some software upgrades may not work with large system configurations. Testing and fine-tuning of new software on larger systems is a nontrivial job.

3.3.3 Application scalability

This refers to matching problem size scalability with machine size scalability. Problem size affects the size of the data set or the workload increase. Instead of increasing machine size, users can enlarge the problem size to enhance system efficiency or cost-effectiveness.

3.3.4 Technology scalability

In case of a standalone research lab, while scaling a system design with new technology one has to consider three aspects: time, space, and heterogeneity. Time refers to generation scalability. When changing to new-generation processors, one must consider the impact to the motherboard, power supply, packaging and cooling, and so forth. Based on past experience, most systems upgrade their commodity processors every three to five years. Space is related to packaging and energy concerns. Technology scalability demands harmony and portability among suppliers. Heterogeneity refers to the use of hardware components or software packages from different vendors. Heterogeneity may limit the scalability. By completely relying on cloud facility researcher will not have to worry about all these concerns and focus on his core research activity.

4. HOW THE OPERATIONAL LAYERS OF CLOUD MODEL CAN FACILITATE COLLABORATIVE RESEARCH?

In order to understand the requirements of research in various domains the cloud model is proposed with five operational layers: (1) The application layer, (2) The Middle ware layer, (3) the resource layer and (4) Network fabric inside cloud data centre, (5). External Connectivity (from research organisations/ researchers to cloud resources).

4.1 Application Layer

Examples of software applications in some important areas of research have been given below. These software's if available on cloud can be accessed by researchers irrespective of their location.

Research Domain	Astronomy	Simulation / Analysis
Software Application's	1.Astronomical Image Processing System (AIPS) 2. Astropy in Python 3.Starlink - leading edge interactive data reduction and analysis tools 4. Flexible Image Transport System 5. ECLIPSE: a suite of IR datacube processing software. 6. SKYCAT: Graphical access to catalogs and images. (Dr. Warren Brown)	1. GROMOS is an acronym of the GROningen Molecular Simulation computer program package. 2. *CHEM* code is used by NASA for launch vehicle, propulsion, and missile systems analysis. 3. LAS X – The integrated imaging platform for advanced life science research 4. The Leica Application Suite (LAS) for visualization, archiving, analysis and detailed documentation of work at the microscope.

8-10 June 2015- Istanbul, Turkey
Proceedings of SOCIOINT15- 2nd International Conference on Education, Social Sciences and Humanities
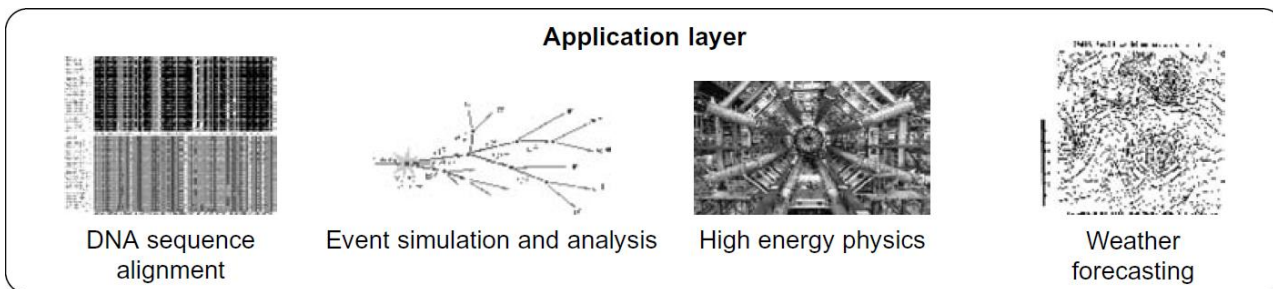
577

Fig. 2 Application Layer – different software for different areas of research on cloud.

4.2   Middleware layer

The middleware layer acts as a bridge between the application layer and the resource layer. This layer provides resource broker, communication service, task analyzer, task scheduler, security access, reliability control, and information service capabilities. It is also responsible for applying energy-efficient techniques, particularly in task scheduling. Until recently, scheduling was aimed at minimizing make span, that is, the execution time of a set of tasks.
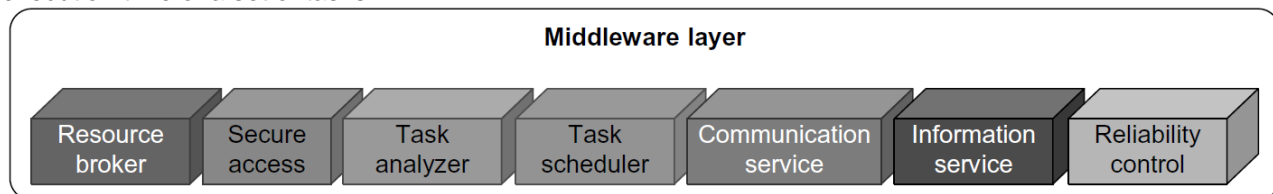

Fig. 3 Middle ware layer – Task scheduling for different research threads / processes running on cloud.

4.3   Resource Layer

The resource layer consists of a wide range of resources including computing nodes and storage units. This layer generally interacts with hardware devices and the operating system; therefore, it is responsible for controlling all distributed resources in distributed computing systems
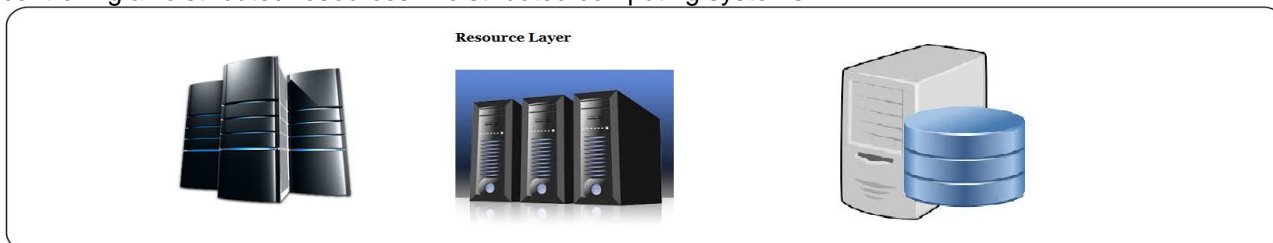

Fig. 4 Resource layer – the processing and storage capabilities on cloud.

4.4   Network Fabric

Routing and transferring packets and enabling network services to the resource layer are the main responsibility of the network layer in cloud computing systems. This layer is basically the physical connectivity within the cloud data centre connecting the servers, storage etc. And connectivity between different cloud data centres.

4.5   External Connectivity to research organizations / researchers

Connectivity from research labs / researchers to cloud computing resources is very important. Requirement of bandwidth may vary from research project to project. Security is also a concern. However it can be seen that almost of the countries have created high bandwidth connectivity to connect research organizations and universities within country. Some examples of these connectivity's is as below:

8-10 June 2015- Istanbul, Turkey
Proceedings of SOCIOINT15- 2nd International Conference on Education, Social Sciences and Humanities

578

SR NO	Country	NREN	Core Bandwidth
1	India	ERNET, NKN	2.5/10 Gbps
2	Australia	AARNET	20Gbps
3	France	RENATAR	60Gbps
4	Germany	DFN	100Gbps
5	Thailand	ThaiREN; Uninet	51Gbps
6	Japan	NII	20Gbps/80Gbps
7	UK	Janet	40Gbps
8	United States	Internet2; NLR; CalREN	100Gbps
9	Ireland	HEAnet	30Gbps
10	Turkey	ULAKBIM	10 Gbit/s

Table 1 National Research and Educational networks (TERENA Compendium, 2013, p.95)

5. CASE STUDY OF CLOUD MODEL FOR COLLABORATIVE RESEARCH IN THE FIELD OF BIOMOLECULAR SIMULATION

5.1 Cloud Computing requirements for Biomolecular Simulation

To build, design, cloud computing framework for creating or modifying high performance applications for the cloud, we establish a set of rules to follow, which are presented below:

Rule 1: Scalability. The cloud computing framework must allow applications to scale in size.

Rule 2: Resource Adaptability. The framework must dynamically harness resources as they become available and allow applications to utilize these resources to progress in their execution. (*Example hypervisor in case of Microsoft Azure)*

Rule 3: Fault tolerance. The framework must provide robust fault-tolerance and allow the application to continue execution even in the presence of hardware failures, communication failures, site failures, and execution errors and failures. The framework must dynamically rerun failed tasks or seamlessly migrate them to other sites in the event of such failures.

Rule 4: Portability. The framework must be able to deploy and run the application on different cloud computing platforms with minimal effort and intervention from the user.

Rule 5: Platform independence. It must be independent of any platform, operating system, and hardware characteristics from the application.

Rule 6: Ease of effort. The framework must allow users to migrate their applications with minimal effort to the cloud.

5.2 Cloud computing framework suggested Collaborative research in the field of Bio molecular Simulation

Cloud computing frame work for one of the important area of research i.e. Bio molecular simulation has been discussed here. This is a research area wherein huge amount of computing and storage resources are required. However the resources are required only during the period when the simulation process runs. Thus the resource can remain ideal when not in use. Hence the resource can be shared with multiple labs across the globe. Since more labs can participate, in order to avoid overload on processing capabilities, the requests can be queued and monitored by master management software and then allot the same to the free computing resources i.e. workers to process the data or execute codes or algorithms. The results of simulation can be stored and can be shared with collaborating institutes to avoid repetitive work and move ahead fast for new inventions / discoveries.

The above can be achieved by using this Work Queue framework which is based on the master-worker paradigm, where multiple worker processes can receive and execute workloads sent by the master. The master coordinates the execution of a given application by assigning and scheduling work units to each of the workers. Figure illustrates the master-worker architecture of Work Queue. The arrows describe the communications between the master and worker. The communications occur at the following times: (a) transfer of input including application executables, binaries, input files etc., from master to workers, (b) communication of the task execution commands and their arguments by master to its workers, and (c) transfer of output including output files and logs from workers to master.(C. Hoffa, 2008, pp. 640 –645).
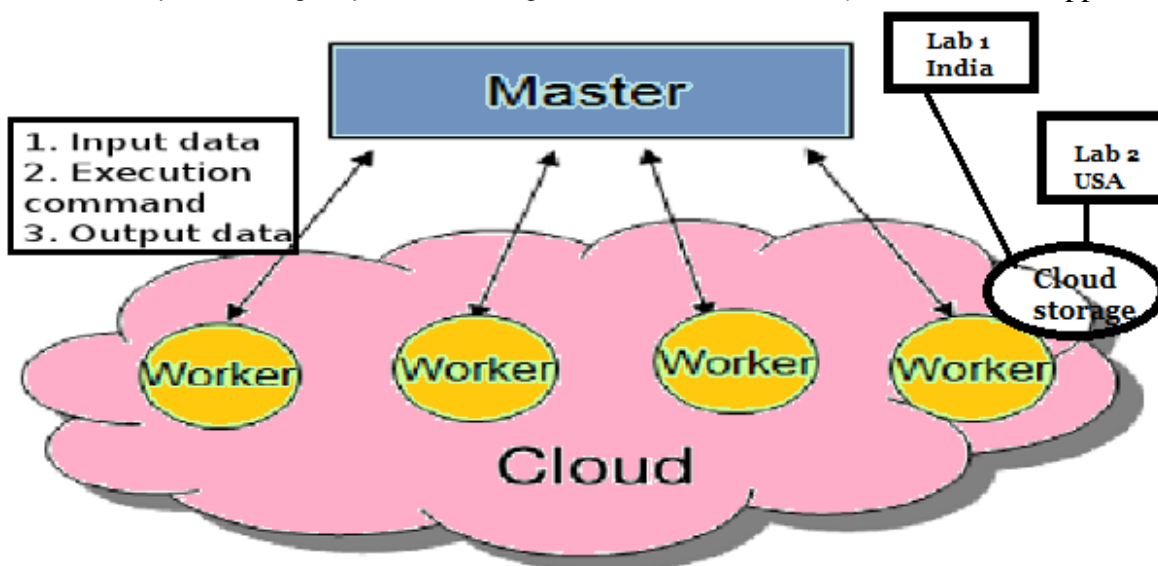

Fig. 5 Cloud based Work Queue framework.

The workers are deployed as executables on the cloud platform and they are invoked and run as jobs on these platforms through their respective job submission interfaces. The workers can be compiled, installed, and run on any Portable Operating System Interface (POSIX) compliant environment. This implies that the worker can virtually be deployed and run on any operating environment including Microsoft Windows based environments. (Andrew R. Leach , 2001)

The master script implemented by the user is often relatively simple. This is because the master script only contains the input file specifications, the executables required for task execution, the output file specifications, and the task execution command and arguments.

6. CONCLUSION

The paper has discussed need and importance of collaboration. In cloud environment how a transparent computing environment separates the user data, application, OS, and hardware in time and space wherein number of researchers can work simultaneously without sharing data or collaborate to share data or results. The paper has discussed the operational layer of cloud infrastructure for collaborative research.

Finally cloud based work queue framework for bio molecular simulation has been discussed wherein researchers from various labs in world can collaborate and work together.

Cloud computing , where computing, memory, storage facilities are provided on demand from shared data centres will enables effective research collaboration to blossom. Instead of procuring a cluster of computers or struggle to find space at the lab, researchers can outsource their computing storage needs to remote facilities in the cloud and make this data accessible to colleagues. Thus facilitating collaboration.

The collaboration can be in following ways, but not limited to sharing the computing infrastructure, funding to create the common cloud infrastructure, development of cloud model for specific high computing research requirements, sharing of information and data, development of software's for common research objectives etc.

Thus it is possible to create global research platform for researchers across the globe to researchers to collaborate and answer to the complex questions for betterment of humanity.

7. ACKNOWLEDGEMENT

Sciences, Pune for guidance during this research work.

REFERENCE LIST

Bozeman, Barry, Boardman, Craig (2014). Research Collaboration and Team Science - A State-of-the-Art Review and Agenda.

Klein, J. (1994). Finding interdisciplinary knowledge and information. In J. Klein & W. Doty (Eds.), Interdisciplinary Studies Today.

Clark, B. (1995). Places of inquiry: Research and advanced education in modern universities. Berkeley, CA University of California Press.

Sonnenwald, D.H. & Pierce, L. (2000). Information behaviour in dynamic group work contexts: Interwoven situational awareness, dense social networks, and contested collaboration in command and control. Information Processing & Management.

Kai Hwang , Geoffrey C. Fox, Jack J. Dongarra (2012) Distributed and Cloud Computing: From Parallel Processing to the Internet of Things.

Dr. Warren Brown The SAO Telescope Data Center (http://tdc-www.harvard.edu/index.html).

Andrew R. Leach (2001) Molecular modelling : principles and applications / ISBN 0-582-38210-6 (pbk) 2 ed. Harlow : Prentice Hall,

C. Hoffa, G. Mehta, T. Freeman, E. Deelman, K. Keahey, B. Berriman, and J. Good (2008), "On the use of cloud computing for scientific workflows," in eScience, 2008. eScience '08. IEEE Fourth International Conference.

TERENA Compendium (2013); Partnership on Measuring ICT for Development WSIS Targets Questionnaire.