



African Virtual University

Applied Computer Science: ITI 4104

# GRID CLOUD COMPUTING

---

Jackson G. Kabira

---

# Foreword

The African Virtual University (AVU) is proud to participate in increasing access to education in African countries through the production of quality learning materials. We are also proud to contribute to global knowledge as our Open Educational Resources are mostly accessed from outside the African continent.

This module was developed as part of a diploma and degree program in Applied Computer Science, in collaboration with 18 African partner institutions from 16 countries. A total of 156 modules were developed or translated to ensure availability in English, French and Portuguese. These modules have also been made available as open education resources (OER) on [oer.avu.org](http://oer.avu.org).

On behalf of the African Virtual University and our patron, our partner institutions, the African Development Bank, I invite you to use this module in your institution, for your own education, to share it as widely as possible and to participate actively in the AVU communities of practice of your interest. We are committed to be on the frontline of developing and sharing Open Educational Resources.

The African Virtual University (AVU) is a Pan African Intergovernmental Organization established by charter with the mandate of significantly increasing access to quality higher education and training through the innovative use of information communication technologies. A Charter, establishing the AVU as an Intergovernmental Organization, has been signed so far by nineteen (19) African Governments - Kenya, Senegal, Mauritania, Mali, Cote d'Ivoire, Tanzania, Mozambique, Democratic Republic of Congo, Benin, Ghana, Republic of Guinea, Burkina Faso, Niger, South Sudan, Sudan, The Gambia, Guinea-Bissau, Ethiopia and Cape Verde.

The following institutions participated in the Applied Computer Science Program: (1) Université d'Abomey Calavi in Benin; (2) Université de Ougadougou in Burkina Faso; (3) Université Lumière de Bujumbura in Burundi; (4) Université de Douala in Cameroon; (5) Université de Nouakchott in Mauritania; (6) Université Gaston Berger in Senegal; (7) Université des Sciences, des Techniques et Technologies de Bamako in Mali (8) Ghana Institute of Management and Public Administration; (9) Kwame Nkrumah University of Science and Technology in Ghana; (10) Kenyatta University in Kenya; (11) Egerton University in Kenya; (12) Addis Ababa University in Ethiopia (13) University of Rwanda; (14) University of Dar es Salaam in Tanzania; (15) Université Abdou Moumouni de Niamey in Niger; (16) Université Cheikh Anta Diop in Senegal; (17) Universidade Pedagógica in Mozambique; and (18) The University of the Gambia in The Gambia.

Bakary Diallo

The Rector

African Virtual University

# Production Credits

## **Author**

Jackson Kabira

## **Peer Reviewer**

Dessaegn Mequanint

## **AVU - Academic Coordination**

Dr. Marilena Cabral

## **Overall Coordinator Applied Computer Science Program**

Prof Tim Mwololo Waema

## **Module Coordinator**

Robert Oboko

## **Instructional Designers**

Elizabeth Mbasu

Diana Tuel

Benta Ochola

## **Media Team**

Sidney McGregor

Michal Abigael Koyier

Barry Savala

Mercy Tabi Ojwang

Edwin Kiprono

Josiah Mutsogu

Kelvin Muriithi

Kefa Murimi

Victor Oluoch Otieno

Gerisson Mulongo

---

# Copyright Notice

This document is published under the conditions of the Creative Commons

[http://en.wikipedia.org/wiki/Creative\\_Commons](http://en.wikipedia.org/wiki/Creative_Commons)

Attribution <http://creativecommons.org/licenses/by/2.5/>



Module Template is copyright African Virtual University licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. CC-BY, SA

## Supported By



AVU Multinational Project II funded by the African Development Bank.

---

# Table of Contents

<b>Foreword</b>	<b>2</b>
<b>Production Credits</b>	<b>3</b>
<b>Copyright Notice</b>	<b>4</b>
<b>Supported By</b>	<b>4</b>
<b>Course Overview</b>	<b>7</b>
<b>Welcome to Grid &amp; Cloud Computing.</b>	<b>7</b>
<b>Units</b>	<b>8</b>
Unit 0: Pre-Assessment	8
Unit 1: Overview of grid and cloud computing.	8
Unit 2: Cloud Applications	8
Unit 3: Grid & Cloud computing security	9
Unit 4: Grid Projects and Applications	9
<b>Readings and Other Resources</b>	<b>10</b>
Unit 1	10
Unit 2	10
Unit 3	11
Unit 4	11
<b>Unit 0: Pre-assessment.</b>	<b>11</b>
Unit Objectives	11
Unit Introduction	12
<b>Unit Assessment</b>	<b>13</b>
<b>Unit 1. Overview of Grid and Cloud Computing.</b>	<b>17</b>
<b>Unit Introduction.</b>	<b>17</b>
<b>Learning activities</b>	<b>19</b>
Activity 1:- Cloud computing.	19
Activity 2 Grid computing.	21
Activity 3 Cloud Services	24
Commercial cloud providers	28

Conclusion . . . . .	29
Unit assessment . . . . .	29
<b>Unit 2. Cloud Applications</b>	<b>33</b>
Unit Introduction. . . . .	33
Learning activities . . . . .	34
Activity 1:- Cloud applications Paradigms	34
Activity 2: Cloud applications architectural styles	36
Activity 3:- Cloud computing Limitations	42
Assessment . . . . .	44
<b>Unit 3. Security in grid and cloud Computing</b>	<b>48</b>
Unit Introduction. . . . .	48
Unit Objectives	48
Learning activities . . . . .	49
Activity 1:- Security in cloud computing	49
Activity 2: Security in Grid Computing	50
Conclusion . . . . .	54
Assessment . . . . .	57
<b>Unit 4 : Grid Projects and applications</b>	<b>58</b>
Unit Introduction . . . . .	58
Unit Objectives	58
Learning activities . . . . .	59
Activity 1:- DataGrid	59
Activity 2:- BioGrid (Biological General repository for Interaction Datasets)	59
Activity 3 European Grid Infrastructure	60
Activity 4 Other Large scale grids	61
Activity 5 Grid Applications	61
Unit Assessment . . . . .	63
Unit Readings and Other Resources . . . . .	64
Module Assessment . . . . .	64

# Course Overview

## Welcome to Grid & Cloud Computing.

Cloud computing is a means which allows you access applications and other information that reside at a location other than your computer or server which typically may be hundreds or thousands of miles away. One advantage of cloud computing is that another company hosts the application(s) which implies that they are responsible for the cost of hardware and software and therefore you as the end user pays less for the services.

The learner should realize that grid computing is often confused with cloud computing. However, a grid computing network harnesses the unused processing cycles of all computers in the grid pool to solve problems that may be too intensive for any one stand alone computer.

### Prerequisites/Required Knowledge

You should be able to access the Internet through any of the browsing tools such as Mozilla FireFox or Internet Explorer. You will also find references to materials which can be obtained from selected e-books on the Internet.

### Materials

The materials needed to complete this course will be:-

1. Bahga Arshdeep and Vijay Madiseti, Cloud Computing: A hands-on Approach, Arshdeep
2. Bahga & Vijay Madiseti Publishing, 2014.
3. Erl Thomas, Zaigham Mahmood and Ricardo Puttini, Cloud Computing: Concepts, Technology and Architecture, Prentice Hall, 2013.
4. Magoules Frederic, Jie Pan et al, Introduction to Grid Computing, CRC Press, Taylor & Francis group Boca Raton Florida, 2009.
5. Velte Toby and Anthony Velte, Cloud Computing: A Practical Approach, McGrawHill, 2009.
6. Information Resources Management Association, Grid and Cloud Computing: Concepts, Methodologies, Tools and Applications, Vol 1
7. Wilkinson Barry, Grid Computing: Techniques and Applications, Chapman & Hall/CRC, 2010

### Number of Hours

120 hours

### **Course/module rationale**

#### Course Goals

Upon completion of this course the learner should be able to:

- Articulate main concepts and key technologies behind grid & cloud computing.
- Evaluate various grid and cloud computing architectures.
- State the benefits of cloud computing model.
- Analyze various grid and cloud computing solutions.

### **Learning outcomes**

- The learner will distinguish grid from cloud computing.
- The learner will select an appropriate cloud computing model.
- The learner will state the benefits and limitations of cloud computing.
- The learner will evaluate when to use or not to use the cloud.

## **Units**

### Unit 0: Pre-Assessment

This part of the module is to remind you of some of the concepts that you need to have covered and are assumed in the module.

### Unit 1: Overview of grid and cloud computing.

Cloud and grid computing are now among the leading emerging technologies in the world of computing today. Major corporations such as Microsoft, Amazon, IBM, HP and Salesforce.com are now involved in providing cutting-edge and innovative solutions for small and big businesses via the Cloud.

### Unit 2: Cloud Applications

The efforts in support of large scale distributed computing have encountered major difficulties over a long period of time such as users having difficulties locating systems to run their applications. The future success of cloud computing rests on the ability of companies promoting utility computing to convince a large segment of user population on the merits using cloud applications.

**Unit 3: Grid & Cloud computing security**

Security is the main issue when it comes to grid and cloud computing. Since a third party stores the data in the cloud, its unlikely the user will know what is going on. Similarly with the grid, many entities are interconnected together in form of a hybrid network and coupled with lack of a single point of control, security can easily be compromised.

**Unit 4: Grid Projects and Applications**

Grid computing has lead to the proliferation of many grid projects geared towards the solution of many complex issues such as astronomy and engineering. These and many others attempt to harness the unused power of computers to make the computations faster and efficient. Grid computing is an interesting field where more research in still ongoing and it seems to have a great future.

**Assessment**

For each unit covered, you will conclude with an activity and a CAT (Continuous Assessment Test)

1	Activity-based assessment	30%
2	Assignments/Laboratory work	20%
3	Final exam	50%
TOTAL		100%

**Schedule**

Unit	Activities	Estimated time
Unit 0	Pre-assessment	[10 hours]
Unit 1	Overview of grid and cloud computing	[20 hours]
Unit 2	Cloud applications	[30 hours]
Unit 3	Grid and Cloud security	[30 hours]
Unit 4	Grid Projects and Applications	[30 hours]

### Readings and Other Resources

The readings and other resources in this course are:

#### Unit 0

Required readings and other resources:

- <https://www.siteground.com/tutorials/cloud/>
- <https://www.aerofs.com/blog/the-history-and-development-of-cloud-computing/>

Optional readings and other resources:

- <http://www.business2community.com/cloud-computing/brief-history-cloud-computing-0858476>
- <http://www.salesforce.com/uk/socialsuccess/cloud-computing/the-complete-history-of-cloud-computing.jsp>

#### Unit 1

Required readings and other resources:

- <http://aws.amazon.com>
- <http://code.google.com/appengine/>

Optional readings and other resources:

- USA Information resources management association, Grid and Cloud Computing: Concepts, Methodologies, Tools and Applications, 2012.

#### Unit 2

Required readings and other resources:

- Reese George, Cloud Application Architectures: Building Applications and Infrastructure in the Cloud, Oreily, 2009.
- Wilder Bill, Cloud Architecture Patterns: Using Microsoft Azure, Oreily, 2012

Optional readings and other resources:

- Amies Alex and harm Sluiman, Developing and Hosting Applications on the Cloud, IBM Press, 2012
- Fehling Christoph and Frank Leymann, Cloud Computing Patterns: Fundamentals to Design, Build and Manage Cloud Applications. Springer books, 2014.

### Unit 3

Required readings and other resources:

- <http://www.isaca.org/TICOcloud>
- <http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf>
- Wilkinson Barry, Grid Computing: Techniques and Applications, Chapman & Hall/ CRC, 2010

Optional readings and other resources:

- L. Wang, R. Ranjan, J. Chen, and B. Benatallah, Cloud Computing: Methodology, Systems, and Applications, CRC Press, Boca Raton, FL,USA, October 2011

### Unit 4

Required readings and other resources:

- [www.opensciencegrid.org](http://www.opensciencegrid.org)
- [www.teragrid.org](http://www.teragrid.org)

Optional readings and other resources:

- [www.naregi.org](http://www.naregi.org)
- [www.thebiogrid.org](http://www.thebiogrid.org)

## **Unit 0: Pre-assessment**

### Unit Objectives

At the end of this unit, you should be able to:

- Examine the history of grid and cloud computing
- Differentiate cluster from grid computing
- Explain basic grid and cloud terminologies.

### KEY TERMS

**Parallel processing** - It is the processing of program instructions by dividing them among multiple processors with the objective of running a program in less time since in earliest computers only one program ran at a time.

**ARPANET** - Advanced Research Agency Project Network. It is the network that became the basis of the Internet and a test bed for many networking technologies linking universities and research centers.

**Protocol** - Its is an agreed upon format or language for transmitting data between two network communicating devices eg TCP/IP or HTTP

**Ubiquitous/pervasive computing** - The idea that technology is moving beyond the personal computer to everyday devices since computing devices are becoming more smaller and more powerful.

**VLSI** - Very Large Scale Integration. The process of placing thousands of electronic components on a single chip

### Unit Introduction

Although cloud computing might be a term that you've only heard about in recent years, the concept behind it has been around for quite some time. This concept was preceded by a file-sharing model which involved sharing of computer data in a network with various levels of access privilege. This allowed a number of people to use the same file by being able to read, write or modify it. It had been a feature of mainframe computer systems for many years. With the advent of the Internet, a file transfer system called File Transfer Protocol (FTP) became widely used to access files shared among users with a password "anonymous" used to gain access. FTP was mostly used to upload files to a web server and these files could then be downloaded by users from any particular place in the world as long as they had Internet access. Later, a centralized computing model, consisting of super computers located behind walls of an internal data center was conceived. These supercomputers were quite expensive typically costing millions of dollars and were primarily used in intensive computational tasks such as quantum mechanics, weather forecasting, climate research and oil exploration. The 1980s brought the growing demand for increasingly more powerful and less expensive microprocessors, laptops and personal computers which were relatively low in cost.

Grid and cloud computing came into play in the early 1990s as the Internet exploded exponentially moving away from centralized, client-server models to Internet based computing. A client-server model is a distributed communication framework via the Internet consisting of a client (service requester) and a server (service provider). The server manages the processes and stores all data while a client requests specific data or processes which are relayed to it by the server. This differs from the peer to peer model where there is no dedicated server but both nodes are equal in status as both are client or server. The notion behind grid computing was to make computer power as easy to access as an electric power grid. Grid computing provided people from different organizations the opportunity to work together to reach a common goal. Cloud computing allows people to rent computing services such as Internet access which cuts back on cost and makes computing more affordable for small and medium sized businesses.

Grid and cloud computing use distributed interconnected computers to collectively achieve higher performance computing and share resources for better workplace productivity.. The history of grid and cloud computing dates back to the 1960s with the development of packet switched networks the most ground breaking being the department of defence funded ARPANET network which became functional in 1969 with four interconnected nodes, University of Utah, University of California (Santa Barbara), University of Los Angeles and Stanford research institute. Transmission Control Protocol Internet Protocol (TCP/IP) was conceived in 1974 provided a protocol for reliable communication including the development of Ethernet which became the principle way of connecting computers on a local area network. The Internet began to be commercially viable in 1990s as the World Wide Web with the introduction of the browser and Hyper Text Markup Language. Computer systems started as single processor systems where it soon became evident that increased speeds could potentially be achieved by having more than one processor in one computer system leading to the adoption of the term parallel processing. Individual computer systems were then interconnected together in a bid to boost computing performance. The common denominator between grid and cloud computing lies in the use of the Internet to share resources. Grid computing focuses more on collaborative and shared resources while cloud computing is geared towards placing resources for users to pay while shared on a common platform.

### **Unit Assessment**

Check you understanding!

Formative assessment

- 1) What is the correct definition of cloud computing?
  - a) A large pool of usable and accessible virtualized resources.
  - b) A network of globally interconnected client computers
  - c) Service architecture based on thin clients
  - d) Service offered by service provider, not limited by service level agreements

2. Where did the Internet begin?
  - a) Cluster of cooperating universities in the US
  - b) Computer hobbyists
  - c) The CIA
  - d) US department of defense.
- 3) What is an important benefit of grid & cloud computing?
  - a) Highly protected data
  - b) Independent from the Internet
  - c) reduced cost
  - d) Small bandwidth
- 4) Cloud computing is also known as on-demand computing.
  - a) True
  - b) False
- 5) Which of the following is NOT a disadvantage of cloud computing.
  - a) privacy
  - b) security
  - c) reliability
  - d) costs
  - e) Loss of control
- 6) Cloud computing usually charges users a flat rate of usage.
  - a) True
  - b) False
- 7) Grid computing:-
  - a) run an application on one server
  - b) is ideal for transaction processing systems
  - c) divides the processing of an application among servers
  - d) does not require any special software
  - e) is difficult to scale up and down

8) Server virtualization uses software based partitions to create multiple virtual servers on a single physical server.

- a) True
- b) False

9) Which of the following is NOT a benefit of employing server virtualization?

- a) cost savings in space
- b) cost savings in energy
- c) cost savings in maintenance
- d) cost savings in training
- e) cost savings in hardware

10. Which of the following statements concerning cloud computing is NOT true?

- a) User access computers in the cloud usually via the Internet
- b) The cloud includes computers software and a network
- c) A cloud can be private or public
- d) A cloud can be server farm
- e) Data can be stored on the users computer

**Answers**

- 1. A
- 2. D
- 3. C
- 4. B
- 5. D
- 6. B
- 7. C
- 8. A
- 9. D
- 10. E

### Feedback

You are requested to provide this feedback on all the four units of this course. Answer the questions objectively as this will go a long way in assisting in the future development and improvements of materials presented in each of the units.

- Does this assessment adequately test the materials contained in this unit?
- Do you think the questions are clearly stated?
- Do the questions relate to the unit objectives?
- Suggest any improvements on the assessment styles.

### Unit Readings and Other Resources

The readings in this unit are to be found at the course-level section "Readings and Other Resources".

- <https://www.siteground.com/tutorials/cloud/>
- <https://www.aerofs.com/blog/the-history-and-development-of-cloud-computing/>
- <https://web.stanford.edu/class/ee204/Publications/Amazon-EE353-2008-1.pdf>

# Unit 1. Overview of Grid and Cloud Computing.

## Unit Introduction

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, storage, applications and services) that can be rapidly provisioned and release with minimal management effort or service provider interaction<sup>1</sup>. The word cloud is used to represent the Internet where various services such servers, storage devices and a myriad of applications are delivered to an organization's computing devices e.g desktops and workstations via the Internet. The basic idea behind cloud computing is to apply high performance computing power associated with the military and other research facilities to perform trillions of computations in applications such as finance, assets management and even computer based games.

Grid computing on the other hand applies the resources of numerous computers to work to work on a problem simultaneously. In most instances, this is done to address a scientific problem such as the Search for Extraterrestrial Intelligence whose goal is to detect intelligent life outside earth. It uses radio telescopes to listen for narrow bandwidth radio signals from space. Such signals are not known to occur naturally so a detection would provide evidence of extraterrestrial technology.

People globally contribute the unused power of their computers to the SETI project and you too could participate and learn more on what this scientific expedition is all about! You may also have heard about the Berkeley Open Infrastructure for Network Computing (BOINC) which conducts protein-folding experiments in order to create better and more durable rice crops to mitigate against world wide hunger. You probably didn't know you can elect to dedicate a part of your idle CPU processing power to help in this noble project!

## Unit Objectives

Upon completion of this unit you will be able to:

- Differentiate grid and cloud computing.
- Categorize various cloud services.
- Evaluate various cloud delivery models.
- Outline security threats in grid and cloud infrastructure.

### KEY TERMS

**Virtualization:** It is the creation a virtual version of a devise or resource such as server, storage device or operating system where the framework divides the resource into more than one execution environments.

**Paravirtualization:** It allows multiple operating systems to run on a single hardware device at the same time more efficiently using system resources like processors and memory. It runs better than the full virtualization model because in a fully virtualized deployment, all elements must be emulated.

**Full virtualization:** Is a technique in which a complete installation of one machine is run on another. The result is a system in which all software running on the server is within a virtual machine.

hardware virtulization:

**PDA:** (Personal Digital assistant) It is a hand held device that combines computing, telephone, Internet and networking features such as browsing and personal organizer.

**SLA:** (Service Level Agreement) An agreement between a business providing a service and a customer stating the minimum quality of service, availability etc expected by the user and the charges levied on those services.e.g The centers SLA states that 80% of calls will be answered within 30 seconds.

**VO:** (Virtual Organization) - Dynamic collection of multiple organizations providing coordinated resource sharing.

**Floating point operations:** Computers equipped with floating point unit perform certain types of applications much faster than computers that lack one.

**OS: (Operating System)** - A set of programs that control the way a computer system works such as how its memory is used and how different programs work together.

**Data grid:** Is a grid for managing and sharing a large amount of distributed data.

**API:** Application Programming Interface -

**Metadata:** data about data

**Indexing:** Is the process of organizing content in some order to allow efficient search and retrieval operations to identify sections of interest within the content.

**Multitenancy:** The aspect of allowing a resource to be shared by many users.

**SSL:** Secure Sockets Layer

**HTML:** Hyper Text Mark up Language

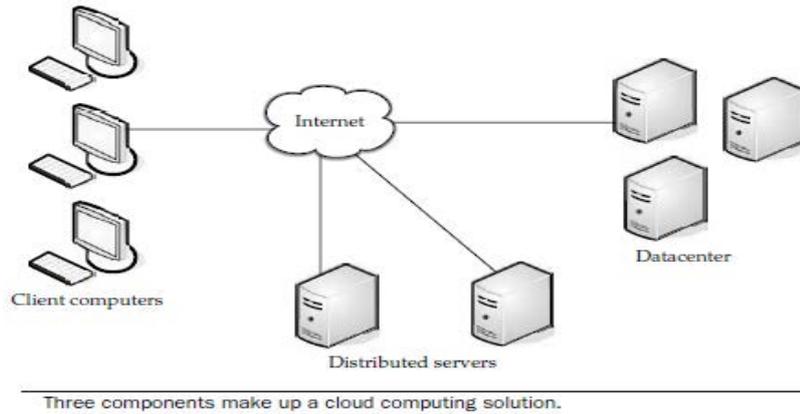
### Learning activities

#### Activity 1:- Cloud computing.

##### **Introduction**

A cloud computing model is made up of several components i.e clients, datacenter and distributed servers. Clients are devices you use to interact and manage data on the cloud and these include tablets, mobile phones, desktops, workstations, PDA's and laptops. These clients can further be grouped into Mobile clients (Smart phones such as iPhone and Blackberries), thin clients which do not have any processing power of their own but only display information they get from the servers, and thick clients which are desktop computers that use web browsers such as Internet Explorer or Mozilla FireFox to access the cloud. The applications that you subscribe in the cloud for your usage are housed in a datacenter as a collection of servers. This can be a physical server though logically can be instances of multiple servers which is often referred to as virtualization. The distributed servers are in most cases geographically dispersed so as to mitigate against natural and man-made disasters. The distributed nature of these servers ensures that if the cloud were to require additional software or hardware, this can be provisioned at any site that is part of the cloud.

Therefore, having a third party host the applications means you need not pay for power nor the personnel to maintain them. Another added benefit is the convenience accorded telecommuters who can simply log on wherever they are and use their applications on the go. Others are improved performance (applications can be scaled up or down), reduced costs (no costs for initial capital expenditure),outsourced management (your organization only pays for operational costs) and improved reliability through SLA. Some shortcomings with cloud computing include Internet outage, cloud failure and security of your mission critical data.



Source: Cloud computing: Techniques and applications

### Characteristics of cloud computing

- Network Access - Cloud services can be accessed via the Internet which through use of heterogeneous platforms such as smartphones, desktops, workstations, tablets and laptops.
- On-demand service - Users can provision cloud services without the intervention of the service provider since the process is fully automated.
- Pooled resources - Multiple users can access the cloud resources that reside on the same physical server. This is made possible by the various forms of virtualizations such as full virtualization, para-virtualization and hardware virtualization.
- Fast elasticity - Cloud resources can be scaled up or down depending on the users request. example you can scale vertically which involves changing the computing capacity assigned to the server while keeping the number of servers constant. Horizontal scaling involves providing additional server resources.
- Metered service - Cloud computing resources are billed on per usage basis based on metrics such as CPU cycles consumed, network I/O requests and storage capacity used.

### Activity details -Cloud computing. (Group Activity) Assessment

a) Other than the examples given in the activity, list three other examples of each these clients below:-.

- i) Mobile clients
- ii) Thin clients
- iii) Thick clients

b) With the aid of a diagram, illustrate the concept of virtualization.

c) How does cloud virtualization differ from OS virtualization and virtual memory creation?

### **Assessment Submission**

This assessment is on the just concluded activity. Do do it in groups of three and submit your work within three using the class email address to the instructor whose email that will be provided in due course.

### **Conclusion**

In this activity, you have learnt about the components of a cloud computing model and its fundamental characteristics. The key defining feature of the cloud is the aspect of virtualization which refers to the seamless integration of geographically distributed and heterogeneous systems to enable you use the services provided by the grid in a transparent manner.

### **Activity 2 Grid computing.**

#### **Introduction**

Grid computing has emerged as an important field analogous to high performance computing. You will note that contrary to other systems where the focus is to achieve greater performance measured in terms of the number of floating point operations the system can perform per minute, grid performance is measured in terms of the amount of work they are able to deliver over a period of time. Grids are actually not considered as an revolutionary technology but have evolved from existing technologies such as distributed computing, Internet, cryptography, web services and virtualization. As you can see, none of these technologies is entirely new but have existed for quite some time. Grid technology has taken features from these to develop systems that provide specific tasks such as SETI. Grid computing involves use of software which can be divided and then send pieces of the program to hundreds of thousands of computers as a form of public collaboration. Sun Microsystems provides Grid engine software which allows engineers of some companies to pool the computer cycles of up to eighty workstations per given time.

Grid computing offers several benefits such as :-

- Solves problems requiring massive amounts of computing power eg SETI and Globus.
- Several computers can be shared on a peer to peer basis.
- Cost affective method with a certain number of computer resources.
- Enable formation of virtual organizations. It enables sharing of resources among VO such as data, special hardware, processing and information dissemination about other resources within the grid. In addition, members of a grid can be part of other VOs and grid can be used to delimit security policies for members thus enabling prioritization.

- Fault tolerance. Grids make provision for automatic resubmission of jobs to other available nodes when failure is detected. You need to note that where data grids are used, they increase file transfer speed and several copies of the data can be created in geographically dispersed locations. Should you require data for any purpose, it can be accessed from the nearest machine hosting the data. Moreover, if it is known prior that a particular machine will access the data more frequently than other machines, the data will be hosted close to that machine.
- Load balancing. This feature enables the grid to evenly distribute jobs to the available resources and if a machine becomes overloaded, the scheduling algorithm can reschedule the tasks to other underutilized systems.
- Parallel processing. You may reckon that some tasks such as mathematical modelling and 3D imaging and animation can be broken down into independent sub-tasks and the results combined to arrive at the desired output. However, one constraint is that these sub-tasks need to operate on same set of data structures. A locking mechanism similar to concurrency control in databases or semaphores in OS needs to be implemented so the data structure does not become inconsistent.
- Quality of Service: You may require the services of the grid for e.g. a real time application and therefore a more strict QoS level than other students. Therefore it is important that a grid scheduler gives your job a higher priority than other jobs and hence provide the needed QoS to your real time application. This is implemented by reserving grid resources for certain jobs but once this is done, the grid scheduler can report its status to a resource management module in the grid. The resource is then freed for other jobs.

You may be asking now, how do grid and cloud computing differ? Just realize that grid computing divides a large project among multiple computers while cloud computing allows multiple applications to run at the same time among the various platforms on the cloud.

### **Grid Architecture**

Grid architecture is mainly concerned with aspects that are important in the design and implementation of the grid system. It is a layered model where the top layer consists of grid applications and APIs from a user's point of view. The second layer is the middleware which consists of software and packages for grid implementation such as glite and Globus Toolkit. The third layer consists of resources for the grid such as storage and processing capabilities. The fourth layer is composed of network components like routers, bridges, switches and protocols.

Applications	Application Programming Interfaces (API)
Middleware	Glite, Globus toolkit
Resources	Storage, processing
Network components	Routers, bridges, switches

The four layered Grid computing model

### **Security**

You should know that security forms an important aspect of grid computing. The security features include sign-on, authentication and authorization. Sign on implies that a user can log-in in using his security credentials and access the grid services. Authentication is concerned with the user providing proof in order to confirm his identity. Finally, authorization is a process that will check and assign a user privileges based on e.g a guest may be allowed to perform basic tasks while a registered user is able to perform more advanced tasks.

### **Resource Management**

This includes submission of a job remotely, checking its progress report and getting the output upon completion. Once a job is submitted, the available resources are obtained through a directory service from where resources for a particular job are selected to run the job. This decision is made by the grid scheduler based on the priority of the job as indicated in the SLA. If an application requires sequential execution where results of the job are required by another job, then the scheduler schedules these jobs in a sequential manner.

### **Data Management**

This implies that a grid scheduler should ideally assign a job close to the data in lieu of transferring vast amounts of data over the network a process that can result in significant performance overheads. Other aspects of data management include data security, replication, migration, metadata, indexing and caching.

### **Information Discovery and Monitoring**

You have learnt that a grid scheduler queries the directory service for available resources and puts constraints such as finding resources relevant and best suited for the job. For example, if a job requires fast CPU's for its execution, the scheduler should select only those machines fast enough for the timely completion of the job. The information discovery service can access these resources through a well defined interface of web services or the scheduler can query it for the list of available resources.

### Activity details - Grid computing Assessment

- a) A distributed system provides features such as fault tolerance, sharing of resources and parallel processing. How then is a grid different?
- b) Think about real systems that have implemented grid computing in your country. List three features of each that make them qualify to be grid computing systems.
- c) Grid security is concerned with sign-on, authentication and authorization. Explain how these tasks can be automated through computer biometrics such as fingerprints, retina scanning etc.
- d) What problems exist when data is moved through great network distances to the processing location? Suggest ways in which the grid can solve this problem.

### Assessment Submission

This assessment is on the just concluded activity. Do it individually and submit your work within two days using the class email address to the instructor email that will be provided in due course.

### Conclusion

In this activity you have learnt that grid computing and high performance computing are related only that grid computing places more emphasis on the amount of work achievable in a given period of time. Additionally, grid computing is seen to have evolved from existing technologies such as the Internet and distributed computing. Grid computing offers a wide array of benefits as discussed in the learning activity such as parallel processing and load balancing among others.

### Activity 3 Cloud Services

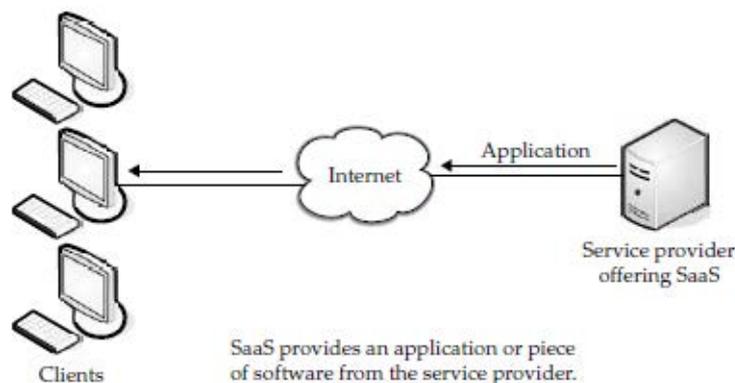
#### Introduction

You will recall that the word cloud is used to represent the Internet where various services such servers, storage devices and a myriad of applications are delivered to an organization's computing devices e.g desktops and workstations via the Internet. The term services in cloud computing refers to reusable and fine grained components across an internet network. This concept is widely referred to "as a service" which includes features such as low entry barriers hence available to even small businesses, scalability, heterogeneity and multitenancy.

### Software as a Service (SaaS)

This provides a complete software application or the user interface to the actual application which is accessible via the Internet. Since this is hosted off-site, the user does not have to incur charges to maintain or support the application. It is the job of the cloud service provider to manage the underlying cloud infrastructure such as servers, network, storage, application software and operating systems. In this scenario, the user is unaware of underlying cloud architecture as access is via a thin client such as a web browser. SaaS applications are platform independent eg they can run on a variety of operating systems such as Windows, Linux, apple Macintosh etc and can equally be accessed from various client devices such as smart phones, tablets and workstations

SaaS services can be billed as a once upfront fee or it can be on per application usage basis. Many softwares lend themselves well to the SaaS but typically, those that perform simple tasks without much interaction with other systems are the most ideal. Such applications include, CRM, accounting, web content management, video conferencing and web analytics.



Source: Cloud computing; Concepts, technology and architecture

### SaaS benefits

- Staff - Outsourcing the SaaS applications reduces need for many IT staff.
- Customization - SaaS applications can be customized to an organization as opposed to earlier applications which needed a lot of coding.
- Marketing - Since the application is hosted on the web, its marketing is open to wider audience.
- Security - SaaS uses SSL a secure application which has been around for long and widely trusted by many organizations world wide.
- Learning - Many employees are familiar with Internet operations and therefore the learning curve for using applications in SaaS is much shorter.
- Bandwidth - Many organizations now experience better bandwidths due to proliferation of fiber and other high speed technologies and hence applications can be accessed with low latencies.

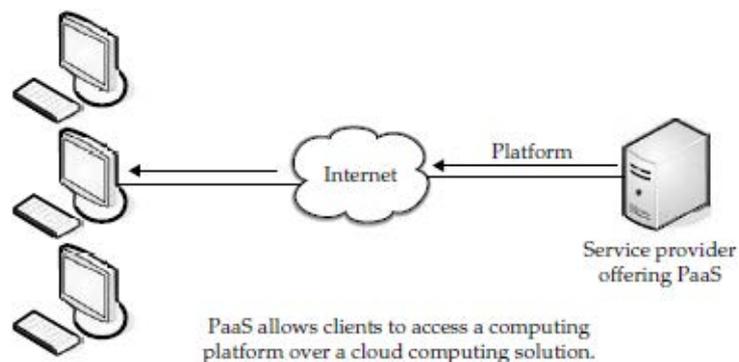
### Limitations

- Organizations with very specific computing needs may be unable to find a suitable application on SaaS for their use. They may therefore find that they need to buy the software and install it in their local servers and workstations.
- Switching costs - Some vendors may charge their clients heavy fees to port an application to another SaaS vendor or the client may find the applications are incompatible.
- SaaS hardware and software may be very expensive and these costs are passed on to the customer by companies hosting SaaS services.
- Unavailability of open source applications.

### Platform as a Service (PaaS)

This provides users with the capability to develop and deploy software in the cloud such as application design, development, testing, deployment and hosting. PaaS provides the development tools to accomplish this such as application programming interfaces, and software libraries based on HTML or javascript. PaaS also provides other web development interfaces such as Simple Object Access Protocol (SOAP) and Representational State Transfer (REST), that makes it possible to develop multiple web services which are sometimes referred to as mashups.

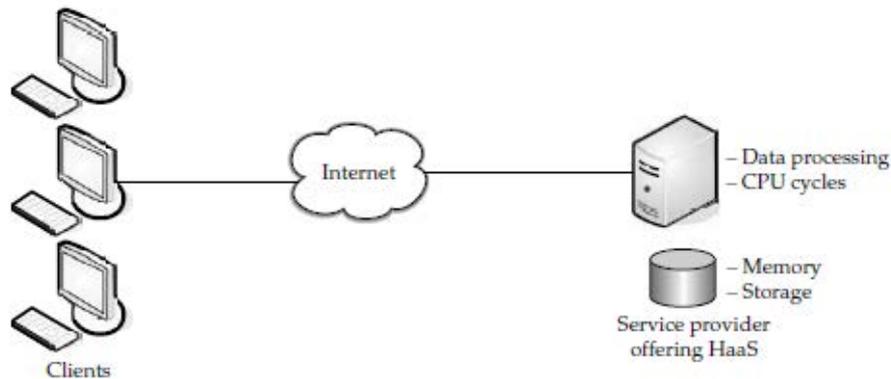
You should note that one drawback of PaaS is lack of interoperability and portability among various cloud vendors. Should you create an application with one vendor and decide to switch to another, it may be costly or not possible. In addition, should the cloud vendor close shop, your applications and data will be lost.



Source: Cloud computing; A hands on approach

### Infrastructure as a Service (IaaS)

You should realize that whereas SaaS and PaaS are about provisioning of applications to clients, IaaS or hardware as a Service (HaaS) offers hardware so you or your organization can put onto it whatever you want. IaaS allows you to rent resources such as server space, memory, storage, CPU cycle and networking equipment (firewalls, routers etc) which can dynamically be scaled up or down depending on the users resource requirements. These resources are also provided as virtual machine instances and virtual storage and are billed on a pay per use basis.



Source: Cloud computing; A Practical Approach Cloud deployment models

### Public cloud

You can access cloud services in this model since they are available to the general public and the services are shared among individuals, organizations and even government. They are ideal for users who would want to use cloud infrastructure for development and testing of applications without large investments in IT infrastructure.

### Private cloud

This type of a cloud is for use by a single organization and can be provided in house by a third party provider and the management can equally be in-house or a third party. They are ideal where security is paramount and the organization desires tight control over the data.

### Community cloud

In the community cloud model, the services are shared by organizations having the same information requirements such as applications and data and the costs are shared amongst them.

### Hybrid cloud

The hybrid model combines the services of private and public clouds though individual clouds retain their unique identities. Since they are intertwined by technology, data and applications are easily portable across the clouds.

This type of cloud is great for organizations desiring the security of a private cloud and the savings accruing from hosting applications in the public cloud.

### Commercial cloud providers

#### **Amazon**

It was among the first to offer cloud services to the public. Its services include:-

- Elastic Compute Cloud (EC2) - Offers CPU cycles and virtual machines.
- Simple Storage Services (S3) - you can store up to 5GB in the virtual storage service.
- Simple Queue Service (SQS) - Machines are able to talk to each other using this API message passing service.

These services can be accessed via the command line interface and virtual machines are Linux based.

#### **Google**

Google has an offering of spreadsheets and other online documents and developers can build their online software using the Google App Engine.

#### **Microsoft**

You will learn that Microsoft cloud solution lies in an operating system called Windows Azure which lets organizations run Windows applications and store files using Microsoft datacenters. Azure Services Platform has services allowing developers to manage work flows, synchronize data and build software programs on Microsoft's online computing platform.

- Windows Azure - Storage, computation and networking.
- Microsoft.NET - Implementation of .NET Framework concepts eg workflow.
- Microsoft SharePoint Services and Microsoft Dynamics CRM Services - Collaboration, business content and solutions development.
- Live Services - Synchronize documents, photos, files, storage across workstations, phones and Internet sites.

### **Activity details - Cloud Services Assessment**

- a) List any four software or hardware services that can be offered in SaaS, PaaS and IaaS from your own local perspective.
- b) What are the benefits of IaaS and PaaS? What are the limitations?
- c) Use the Internet to find out companies offering private, hybrid, community and public clouds.

What are their similarities and differences?

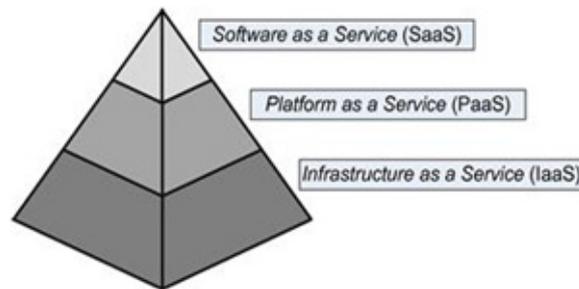
d) What is the SPI model?

### Conclusion

You may now want to ask yourself the question, what does cloud computing actually do? Well look at the applications running on your laptop, servers and phones. You realize that they are either on the cloud or the cloud has the potential to bring them to you! So in short, cloud computing lets you share applications and store data on the cloud without incurring any upfront charges. The most popular methods of data manipulation on the cloud are storage and databases.

### UNIT SUMMARY

In this unit, you have learnt what grid and cloud computing involve and the strengths and weakness of each. The various cloud computing services and models have been discussed including the architecture that forms a grid computing network.



Source: <https://www.safaribooksonline.com>

This is a pyramid model of cloud computing services. The infrastructure provides the basic resources; the platform adds an environment to facilitate the use of these resources, while software allows direct access to services.

### Unit assessment

Check your understanding!

Formative assessments

a) What is the difference between grid and cloud computing?

- b) PaaS and IaaS are the two of the three main categories of cloud computing. Which is the third category?
- c) Where does "cloud" in cloud computing come from?
- d) In a \_\_\_\_\_ cloud an organization might use a public cloud service such as Amazon Elastic Compute Cloud (EC2) for general computing but store data within its own data center.
- e) Payment for cloud computing services is based on this model. What is it?
- f) \_\_\_\_\_ an important characteristic of cloud computing is the ability of hardware and software to continue to function well when it is changed in size or volume to meet a user need.
- g) Is the private cloud or public cloud the standard cloud computing model?
- h) What is the benefit of cloud based anti-virus protection over standard anti-virus programs?
- i) According to industry buzz these small but low power laptop computers known as \_\_\_\_\_ may lead to mainstream acceptance of cloud computing.
- j) Which area of research led to initial development of grid technology?
- k) What is the acronym of the international grid that supports LHC activities?
- l) Where does the term "grid" come from?
- m) Analysis of the huge quantities of data recorded at Large Hadron Collider (LHC) is performed using a distributed computing system called a grid. This links computing resources located at partner institutes (grid sites) from around the world. Approximately how many sites are there in the LHC grid?
- n) Approximately what quantity of data is recorded annually by the LHC experiments?
- o) What are the common set of technologies used to create a manageable cloud computing environment?
- p) What is a common trait of cloud architectures?
- q) Which billing model allows companies to have a predetermined and recurring costs for services used in a cloud environment?
- r) What type of hypervisor provides the highest efficiency and performance?

### Answers:

- a) Cloud computing is the general term for the delivery of hosted services over the Internet.  
Grid computing is harnessing power of unused computers such as CPU cycles to boost the processing power for intensive jobs such as batch processing.
- b) Software as a Service (SaaS)
- c) Internet is represented as a cloud.
- d) hybrid cloud
- e) Utility computing
- f) Scalability
- g) Public cloud
- h) CloudAV - It is a program that combines multiple antivirus applications and scans user files over a network of servers. It was developed at the University of Michigan.
- i) Netbook - It is a small low power notebook computer that has less processing power than a full sized laptop but is still suitable for word processing, running a web browser and connecting wirelessly to the Internet. It has a slimmed down OS, keyboard and screen.
- j) Particle physics
- k) Worldwide LHC Computing Grid (WLCG)
- l) Analogy with the electricity grid which is easy to use and provides power on demand 24/7.
- m) 150 sites spread over 35 countries and 6 continents.
- n) 11 petabytes or 11,000,000,000,000,000 bytes.
- o) A hypervisor, a hypervisor manager, a self service portal, automated provisioning, and monitoring.
- p) While in operation the application automatically scales up or down based on resource needs
- q) Subscription
- r) firmware based

### Unit Readings and Other Resources

- Magoules Frederic, Jie Pan et al, Introduction to Grid Computing, CRC Presss, 2009.
- Velte Anthony, Toby J.Velte and Robert Elsenpeter, Cloud Computing: A Practical Approach, 2010.
- [http://www.gvsu.edu/e-hr/cloud-computing-1.htm?gclid=CP3Zg-LM0sQCFcsBwwod\\_2MAvg](http://www.gvsu.edu/e-hr/cloud-computing-1.htm?gclid=CP3Zg-LM0sQCFcsBwwod_2MAvg)

# Unit 2. Cloud Applications

## Unit Introduction

In the previous unit, you learnt that cloud computing makes computer infrastructure such as hard disk, development platform, database or complete solutions available "on-need basis". Grid computing on the other hand, harnesses e.g. the unused processing cycles of all computers in the grid pool to solve problems that may be too intensive for any one stand alone computer. In this unit, we analyze the different types of cloud applications such as processing pipelines, batch processing and web applications and for each, practical examples are briefly discussed.

## Unit Objectives

Upon completion of this unit you should be able to:

- Evaluate categories of cloud application paradigms.
- Examine cloud computing architectural styles.
- Outline the limitations of cloud computing.

### KEY TERMS

**MPEG**- Moving Picture Experts Group. It refers to the family of digital video compression standards and file formats developed by the group. It is a working group of the ISO.

**AVI** - Audio Video Interleave. It is the file format for Microsoft video for windows standard.

**PDF** - Portable Document Format. Is a file system developed by Adobe systems.

**TCP** - Transmission Control Protocol. Its the suite of communication protocols used to connect hosts to the Internet.

**RPC** - Remote Procedure Call. It is a type of protocol that allows a program on one computer to execute a program on a server computer.

**XML** - Extensible Markup Language. Its a specification that allows designers to create their own customized tags enabling the definition, transmission, validation and interpretation of data between applications and organizations.

**DOD** - Denial of service. A type of attack on a network that is designed to bring a network to its knees by flooding it with useless traffic.

**Indexing** - A collection of information stored in a computer in alphabetical order.

**OCR** - Optical Character Recognition. The process by which an electronic device recognizes printed or written letters and numbers.

**OMG** - Is an international, open membership, not for profit technology standards consortium founded in 1989. OMG standards are driven by vendors, end users, government agencies and academic institutions.

**IDL** - Interface Definition Language

### Learning activities

#### Activity 1:- Cloud applications Paradigms

##### **Processing pipelines**

These compute and data intensive applications represent a big chunk of applications that are currently running on the cloud. I will discuss them here below:-

- i) Indexing. Large databases created by search engines are indexed by the use of the processing pipeline model.
- ii) Image processing. This is where you can store your image on the cloud eg www.flickr.com and you later do image conversion such as enlargement, compression and encryption of the said image.
- iii) Video encoding. This is where you can convert one video format to another such as MPEG to AVI.
- iv) Data mining. This is the examination of large amounts of information stored in a computer in order to look for patterns or changes e.g. buying patterns or new trends in computing.
- v) Document processing. Here, you can convert large documents written in WORD to PDF or you can use Optical Character Recognition to generate digital images of documents for emailing to a third party.

### **Batch Processing**

These are mainly enterprise applications that are characterized by deadlines of which non compliance results in dire economic consequences. Examples of these include the following:-

- Daily,weekly,quarterly or yearly reports for organizations e.g. in engineering.
- Inventories for enterprise organizations
- Billing and payroll processing
- Automated testing of software and hardware systems
- Software development

### **Web access**

As you may be aware, some web sites have temporary or periodic presence e.g. conferences, workshops or even a country's revenue authority's deadline for submission of income taxes. These and other web sites used for promotional activities are inactive at night but auto-scale during the day. It therefore makes sense to store data in a cloud close to where these applications will be often used so as to lower transmission and processing costs.

### **Activity Details - Cloud applications Paradigms Assessment**

- a) Search the web for reports of cloud system failures in pipeline, batch and web access applications and discuss the causes of each incident.

### **Assessment Submission**

This assessment is on the just concluded activity. This is an individual assignment which should be four pages long with 1.5 line spacing. Submit it within a week using the class email to the instructor email that will be provided in due course.

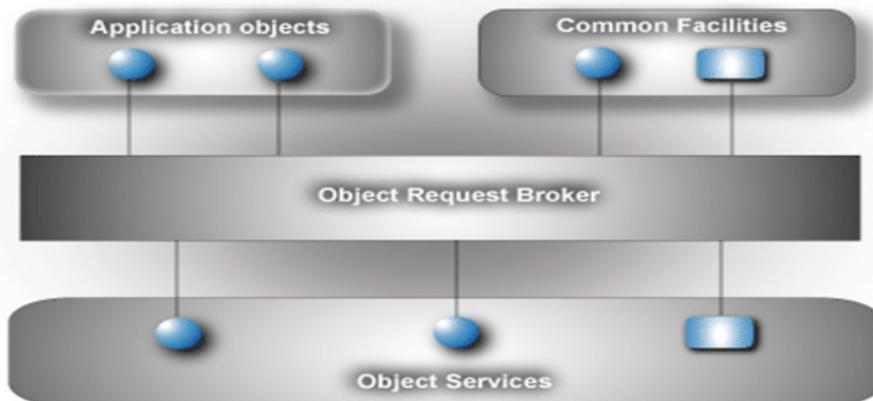
### **Conclusion**

In this activity, you have learnt cloud computing paradigms such as processing pipelines, batch processing and web access services. Processing pipelines represent the vast majority of the applications while batch processing is concerned with periodic reports such as billing, payroll and inventories. Web access is mainly for activities that are short term or seasonal in nature such as promotions or other periodic events such as tax compliance which have a limited duration.

### Activity 2: Cloud applications architectural styles

Cloud applications are generally based on the client/server model where clients communicate with stateless servers. A Stateless server does not require a client to establish a connection before hand to the server. Stateful servers incur considerable overhead because they maintain the state of all the connections and therefore recovering from a failure has considerable overheads. A stateless server is scalable and simpler as the client does not have to be concerned with the state of the server. When a client receives a response from the server that indicates that the server is up and running and if it does not receive a request it means it should resend later. A web server is stateless and it responds to HTTP requests without maintaining the history of past connections with the client. The browser which is the client is equally stateless as it sends requests and has to wait for a response. HTTP makes use of TCP which is a connection oriented protocol. However, this exposes the web server to DOS attacks where malicious clients attempt a TCP connection forcing the server to allocate CPU time for the bogus connection. The servers and clients that run on the cloud communicate using RPC which use stubs to convert the parameters in a RPC call with the stub marshaling the data structures and serialization.

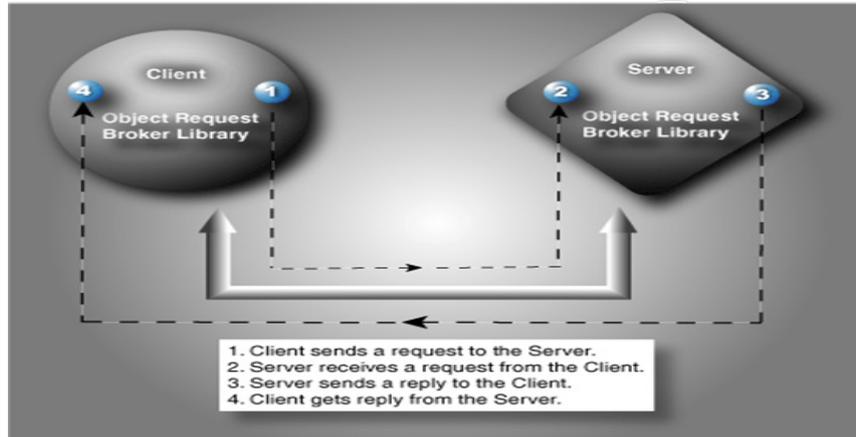
- Common Object Request Broker Architecture (CORBA) - Is an open vendor-independent architecture and infrastructure from OMG that computer applications use to work together over networks. It was developed to allow applications developed in various programming languages in a networked environment and running on different architecture and software to work harmoniously. Because of the easy way that CORBA integrates machines from so many vendors with sizes ranging from mainframes, desktops, handheld and embedded systems, it is the middleware of choice for large, medium and small size enterprises. One of its most important as well as most frequent uses is in servers that must handle large numbers of clients at high bit rates with of high reliability. It works behind the scenes in computer rooms of many of the worlds largest websites, ones that you probably use everyday with specifications for scalability and fault tolerance inbuilt in the systems. The CORBA IDL allows the development of language and location independent interfaces to distributed objects and application components can communicate with one another no matter where they are located.



Primary components in a CORBA implementation

Source: <http://www.ois.com/Products/what-is-corba.html>

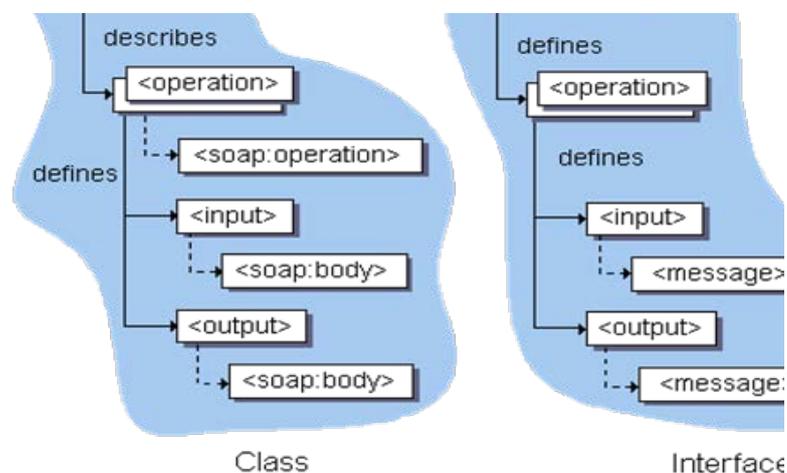
The Object Request Broker determines the location of the target object sends a request to that object and returns any response back to the caller. Through this technology, developers can take advantage of features such as inheritance, encapsulation, polymorphism and runtime dynamic binding. These features allow applications to be changed modified and reused with minimal changes to the parent interface.



Client sending a request to the server through Object Request Broker

Source: <http://www.ois.com/Products/what-is-corba.html>

- Web Services Description Language (WSDL) - It describes communication between end points in a network application. The definition of the elements include services, collection of end points of communication, types, containers for data type definitions, port types, operations descriptions of actions supported by a service, port type, operations supported by endpoints, bindings, protocols and data formats supported by a particular port type and port, an end point as a combination of a binding and network address.



Elements present in a Web Services Description language.

Source: [http://download.oracle.com/otn\\_hosted\\_doc/jdeveloper/1012/web\\_services/ws\\_wsdstructure.html](http://download.oracle.com/otn_hosted_doc/jdeveloper/1012/web_services/ws_wsdstructure.html)

```
<?xml version = '1.0' encoding = 'UTF-8'?>

<!--Generated by the Oracle JDeveloper 10g Web Services WSDL Generator-->

<!--Date Created: Tue Aug 17 10:44:49 BST 2004-->

<definitions

    name="MyJavaClass1WS"

    targetNamespace="http://mypackage/JavaClass1.wsdl"

    xmlns="http://schemas.xmlsoap.org/wsdl/"

    xmlns:xsd="http://www.w3.org/2001/XMLSchema"

    xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/"

    xmlns:tns="http://mypackage/JavaClass1.wsdl"

    xmlns:ns1="http://mypackage/IMyJavaClass1WS.xsd">

    <types>

        <schema

            targetNamespace="http://mypackage/IMyJavaClass1WS.xsd"

            xmlns="http://www.w3.org/2001/XMLSchema"

            xmlns:SOAP-ENC="http://schemas.xmlsoap.org/soap/encoding/" />

    </types>

    <message name="getDate0Request" />

    <message name="getDate0Response">

        <part name="return" type="xsd:string" />

    </message>

    <portType name="JavaClass1PortType">

        <operation name="getDate">

            <input name="getDate0Request" message="tns:getDate0Request" />

            <output name="getDate0Response" message="tns:getDate0Response" />

        </operation>

    </portType>

    <binding name="JavaClass1Binding" type="tns:JavaClass1PortType">

        <soap:binding style="rpc" transport="http://schemas.xmlsoap.org/soap/http" />

        <operation name="getDate">
```

```
<soap:operation soapAction="" style="rpc"/>

<input name="getDate0Request">
    <soap:body use="encoded" namespace="MyJavaClass1WS"
encodingStyle="http://schemas.xmlsoap.org/soap/encoding/" />
</input>

<output name="getDate0Response">
    <soap:body use="encoded" namespace="MyJavaClass1WS"
encodingStyle="http://schemas.xmlsoap.org/soap/encoding/" />
</output>
</operation>
</binding>
<service name="MyJavaClass1WS">
    <port name="JavaClass1Port" binding="tns:JavaClass1Binding">
        <soap:address location="http://UKP16211:8888/
Application1-Project-context-root/MyJavaClass1WS" />
    </port>
</service>
</definitions>
```

An example of a WSDL for a simple web service that returns current date and time.

Source: [http://download.oracle.com/otn\\_hosted\\_doc/jdeveloper/1012/web\\_services/ws\\_wsdstructure.html](http://download.oracle.com/otn_hosted_doc/jdeveloper/1012/web_services/ws_wsdstructure.html)

- Simple Object Access Protocol (SOAP) - It was developed in 1998 as an application protocol for web based applications and the message format is build on XML. It is basically a messaging protocol for exchanging information among computers. The following briefly describe the general characteristics of SOAP:-
  - i) It is communication protocol designed to communicate over the Internet.
  - ii) It provides data transport for web services
  - iii) Can be used for broadcasting a message
  - iv) It is platform and language independent
  - v) It is the XML way of defining what information is sent and how
  - vi) It can be used to exchange complete documents o call a remote procedure
  - vii) It enables client applications to easily connect to remote services and invoke remote methods.

A SOAP message is an ordinary XML document containing the following elements:-

Envelope - Defines the start and end of message. It is a mandatory element.

Header - Contains any optional attributes of the message used in the processing of the message either at the intermediary point or at the end point. It is an optional element.

Body - Contains the XML data comprising the message being sent. It is a mandatory element.

Fault - An optional element that provides information about errors that occur while processing the message.

```
<?xml version="1.0"?>
```

```
<SOAP-ENV:Envelope xmlns:SOAP-ENV="http://www.w3.org/2001/12/soap-envelope" SOAP-ENV:encodingStyle="http://www.w3.org/2001/12/soap-encoding">
```

```
  <SOAP-ENV:Header>
```

```
</SOAP-ENV:Header>
```

```
  <SOAP-ENV:Body>
```

```
    <SOAP-ENV:Fault>
```

```
  </SOAP-ENV:Fault>
```

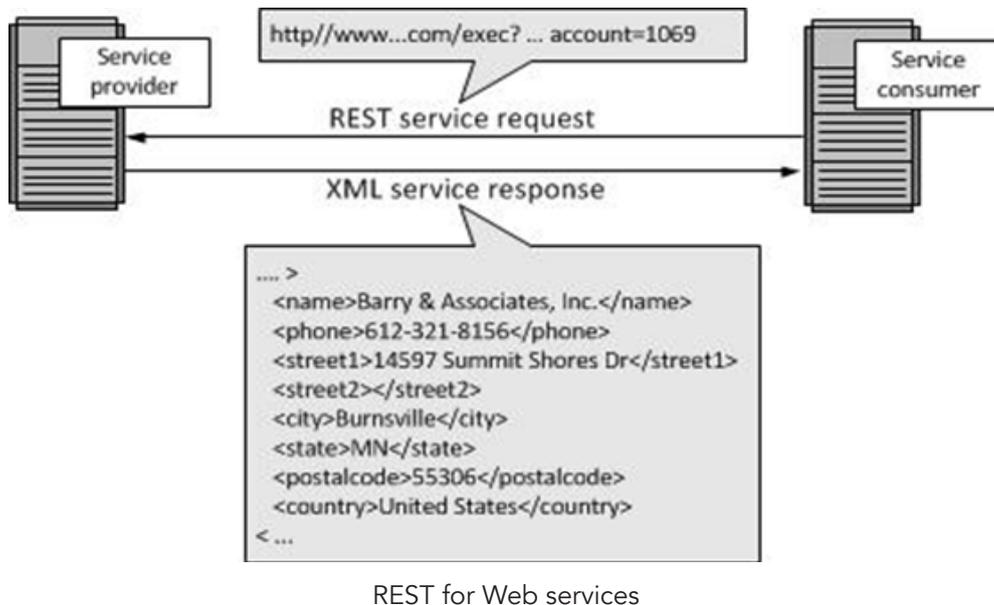
```
</SOAP-ENV:Body>
```

```
</SOAP_ENV:Envelope>
```

The general structure of a SOAP message.

Source: [http://www.tutorialspoint.com/soap/soap\\_message\\_structure.htm](http://www.tutorialspoint.com/soap/soap_message_structure.htm)

Representational State Transfer (REST) - It is a software architecture for distributed hypermedia systems supporting client communication via stateless servers which is language and platform independent. It is generally viewed as a style of software architecture as opposed to a set of standards. Hence, such applications are often referred to as RESTful or RESTstyle applications or architectures. It has proved to be a popular choice for implementing web services for example books suggested at Amazon.com are dynamically generated in part using REST architecture.



Source: [http://www.service-architecture.com/articles/web-services/representational\\_state\\_transfer\\_rest.html](http://www.service-architecture.com/articles/web-services/representational_state_transfer_rest.html)

You need to note that cloud applications need completion of several independent tasks which is a complex activity known as a workflow.

### Activity Details - Cloud applications architectural styles Assessment

Workflow models are abstractions revealing the most important properties of the entities participating in a workflow management system. Task is the central concept in workflow modeling and is a unit of work to be performed on the cloud. State the meaning of the following attributes of a task:-

- Name
- Description
- Actions
- Preconditions
- Post-conditions
- Attributes

- Exemptions
- Composite task
- Routing task
- Fork routing task

### **Assessment Submission**

This assessment is on the just concluded activity. This individual assignment consists of brief descriptions of the terms with appropriate examples. Submit the work within three days using the class email to the instructor whose email will be availed in due course.

### **Conclusion**

In this activity you have learnt that cloud applications are based on the client/server via a stateful or stateless server model. Remote procedure calls are the primary means of communication with technologies such as CORBA, SOAP and REST among others being the intermediaries facilitating this RPC call architecture.

### **Activity 3:- Cloud computing Limitations**

#### **Hardware dependencies**

If you have an application requiring some very specific hardware, the cloud may be inappropriate for you since its unlikely that the cloud provider will have your precise hardware requirements.

#### **Server Control**

Should you require complete control over the server such as memory, CPU and other interfaces, a cloud solution may not work for you since these are the things entrusted to the cloud provider.

#### **Cost**

You may notice that ultimately, the cost of cloud subscription over time may equal to investing in your own computing infrastructure. So, before going the cloud way, factor in the total cost of ownership if you invested in your own facilities vis a vis leasing hosted facilities from the cloud.

#### **Applications integration**

It may not be wise to have some non sensitive applications on the cloud and other sensitive applications locally since chances are that the sensitive data will eventually find its way on the cloud. In equal measure, if you are running a high-speed application in house that's relying on cloud data, that application's speed will depend on how fast the cloud is leading to reliability issues.

### **Latency**

Your data in the cloud is often located on servers dispersed in several geographical areas meaning that there is a delay between making a request on the server and receiving the reply on your client. For time sensitive data e.g. tele-medicine applications, a delay of even a few seconds can make a big difference.

### **Legal issues.**

Different countries have different sets of laws governing the Internet eg Canadian government has a law that forbids government workers from accessing network services operating with U.S borders. Therefore, if you are in Canada and wanted to post data on an American cloud it would not suffice. The U.S government also allows the FBI access to data even on the cloud without a warrant or the consent of the owner whereas that would be against the law in some countries. Posting patients health care and customers financial data on the cloud can also attract a torrent of legal suits in some countries.

### **Activity Details - Limitations of cloud computing Assessment**

Apart from the legal issues discussed above on cloud computing, analyze the moral, social and ethical issues pertaining to cloud computing with specific examples.

### **Assessment Submission**

This assessment is on the just concluded activity. This individual assignment requires you to analyze the moral, social and ethical issues raised by use of cloud computing technologies. You are required to submit a typed, 1.5 line spaced five page report in form of hard copy to the instructor within five days.

### **Conclusion**

In this activity, you have learnt about various shortcomings pertaining to cloud computing such as latency, costs and others such as lack of control over the server that hosts your applications. Cloud computing is an emerging technology and as with any new technology, its bound to present developers and users with new challenges as it matures and gains a critical mass.

## **UNIT SUMMARY**

In this unit, you have learnt about cloud computing applications such as processing, batch and web access. Cloud computing architectural styles include the Common Object Request Broker Architecture and Simple Object Access Protocol among others. You have also seen that cloud computing as with any other technology has its own unique limitations such as cost, integration of hardware and software including various underlying moral, social and legal issues.

### Unit Readings and Other Resources

- Reese George, Cloud Application Architectures: Building Applications and Infrastructure in the Cloud, Oreily, 2009.
- Wilder Bill, Cloud Architecture Patterns: Using Microsoft Azure, Oreily, 2012
- [https://www.exin.com/assets/exin/exams/2017/samples/english\\_sample\\_exam\\_exin\\_cloudf\\_201304.pdf](https://www.exin.com/assets/exin/exams/2017/samples/english_sample_exam_exin_cloudf_201304.pdf)
- [http://www.ics.uci.edu/~fielding/pubs/dissertation/rest\\_arch\\_style.htm](http://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm)

### Assessment

Check your understanding!

#### Formative assessment

- 1) Why did virtualization boost the emergence of cloud computing?
  - a) A virtual machine is more secure than a physical machine
  - b) Virtualization has made it easier and cheaper to share resources between users.
  - c) Virtualization machines have greater performance than their physical counter parts.
  - d) Virtualization leads to better network utilization.
- 2) What is an example of a single purpose environment?
  - a) any application on any server
  - b) Interface to large computer.
  - c) Interface to large storage.
  - d) Mainframe.
- 3) What is an important concern for a customer in a multi-tenant environment?
  - a) availability
  - b) network bandwidth
  - c) network latency
  - d) security
- 4) What can be done to make maximum use of the interoperability principle of cloud computing?
  - a) Employ multiple system integrators to built your private cloud.
  - b) Only use cloud providers located in Europe.
  - c) Use hardware and software of a single vendor.

- d) Use standard protocols.
- 5) What is not a valid reason for a customer asking a cloud provider where their servers are located?
- a) Geographical location may tell something about network latency.
  - b) The geographical location may tell something about legislation.
  - c) The number of sites tells you something about disaster recovery possibilities.
  - d) When a server breaks down, the customer wants to send a technician to fix the problem soonest.
- 6) What is an important requirement for applications to be accessible in the cloud?
- a) The application should be compatible with the browser of the users computer.
  - b) The application should use the same programming language as the clients.
  - c) The user should know on which server the application is located.
  - d) The users identity should be known by the application.
- 7) Which model allows a customer to choose more layers in the computing architecture?
- a) IaaS
  - b) PaaS
  - c) SaaS
  - d) There is no difference between the service models.
- 8) How does cloud computing change the relationship between provider and customer?
- a) Increased focus on SLA
  - b) Less compliance to standards.
  - c) Less focus on SLA
  - d) More focus on training.
- 9) How can confidentiality of information be achieved?
- a) Ensuring enough resources to make information available for all users.
  - b) Preventing unauthorized access
  - c) Regularly backing up the information.
  - d) Restricting access to information.

10. Which of the following is NOT a mitigating measure against data loss?
- a) Audits
  - b) Authentication & authorization.
  - c) Encryption.
  - d) Storage Area Network.
11. What is location information used for in the cloud?
- a) To determine the geographic location of the user of an application.
  - b) To determine the identity of a user of an application.
  - c) To determine whether a user of an application is online.
  - d) To determine who has accessed a document stored in the cloud.
12. What is a benefit of cloud computing for IT staff?
- a) Higher payment of IT staff involved in cloud computing.
  - b) Less interruption of work caused by users asking for support.
  - c) Less knowledge needed. Cloud computing does not require special skills.
  - d) Lower stress levels: Less worry about normal daily activities like making backups.
13. What is the benefit of storage availability in the cloud?
- a) Additional storage does not require budget for new large storage devices.
  - b) Storage in the cloud has a higher availability than storage devices in the LAN.
  - c) Storage in the cloud has shorter access times than storage in the LAN.
  - d) Storage in the cloud is easier to protect against intrusion.
14. What is a Virtual Private Network (VPN)?
- a) Secured connection for remote access to a LAN.
  - b) Secured private cloud for a single user.
  - c) Virtual network device for private purpose.
  - d) An operating system for private network systems.
15. What is an important benefit of cloud?
- a) Highly protected area.

b) Independence from the Internet.

c) Reduced cost.

d) Small bandwidth.

**Answers:**

1. B

2. D

3. D

4. D

5. D

6. A

7. A

8. A

9. D

10. D

11. A

12. D

13. A

14. A

15. C

# Unit 3. Security in grid and cloud Computing

## Unit Introduction.

Computer security has always been an issue since the advent of computing when computers were all stand alones and threats could only have been posed from within most likely by malicious insiders. The threat came with the interconnection of computers to form large networks which were able to communicate over large geographical boundaries. This meant that various types of malware can easily migrate from one network to another, cross international boundaries infecting systems as they propagate. Cloud and grid computing security takes on a new urgency as society increasingly becomes dependent on the information infrastructure whereby even a nations mission critical applications can be attacked by hackers exploiting flaws in computer systems security.

## Unit Objectives

- Describe threats inherent in cloud computing.
- Distinguish between different levels of grid security
- Evaluate cloud and grid security approaches

### KEY TERMS

**Phishing** - An attempt to trick someone into giving information over the Internet or email that would allow someone else to take money from them.

**SQL injection** - A form of attack on a database driven web site in which the attacker executes unauthorized SQL commands by taking advantage of insecure code on a system connected to the Internet by passing the firewall.

**Cross-site scripting -(XSS)** - Is a security breach that takes advantage of dynamically generated web pages where a web application is sent with a script that activates when it is read by an unsuspecting user or application that has not protected itself against cross-site scripting.

**DDOS (Distributed Denial of Service Attack)** - Is an attack where multiple compromised systems are used to target a single system causing a denial of service attack.

**VO** - Virtual Organization. It distinguishes something that is merely conceptual from something that has physical reality e.g. virtual memory refers to an imaginary set of locations where you can store data.

### Learning activities

#### Activity 1:- Security in cloud computing

You realize that security is a major concern for cloud computing users as the cloud is a target rich environment for malicious individuals and criminal entities. You already know that cloud computing is an entirely new paradigm to computing based on emerging technologies and therefore new methods to attack the computing assets are evolving every day. One consequence of rapid strides in information technology is that standards, regulations and laws governing cloud computing are yet to be devised let alone adopted. Coupled with this, different legal systems across different countries governing activities such as e-business have not made this situation any better.

#### **Cloud Security Risks**

##### a) Traditional Threats

These are experienced from time to time by any system that is connected to the Internet but in this case, the problem is cloud specific. This impact is made worse by the large amount of data resources available in the cloud and the large user community that can potentially be impacted. There is also the problem of assigning responsibility between the cloud provider and the users plus identifying the root cause of the problem. Many of the traditional threats emanate from the user and the burden therefore is on the user to protect their infrastructure which connects to the cloud through technologies such as firewall, authentication and authorization mechanisms.. It is also important for an organization to assign users distinct levels of privilege depending on their roles plus harmonizing the security policies of the organization with those on the cloud. Cloud computing threats are mainly seen thorough distributed denial of service attacks, cross-site scripting, phishing and SQL injection. Identifying the path followed by an attacker in the cloud is also much more difficult due to the multitenancy nature of the sites.

##### b) Cloud services availability

You notice that there are many catastrophic events can negatively affect the cloud services such as system failures, power outages and even outright sabotage. Should such situations lend themselves, data lock-in many prevent an organization from accessing its data of which it depends to function. When this happens, users cannot be guaranteed that the applications on the cloud will yield correct results due to the resulting instability of the system a situation referred to a s phase transition phenomena.

##### c) Third party control

In a cloud environment, you have limited control and the provider may in some instances subcontract some services to a third party whose level of trust is in doubt. In some cases, this has resulted in loss of data due to substandard storage mechanisms and this makes auditing a very difficult proposition in cloud computing scenario.

### d) Data loss and leakage

If the only copy of data you have is on the cloud, it can be permanently lost if file replication mechanism fails which is in most cases followed by the storage media failure. Should the failure result in leakage of sensitive data, the loss occasioned would be unreparable with severe and far reaching consequences.

### e) Service hijacking

Your user authorization credentials such as user-names and passwords can be revealed through repeated guesses or even brute force attacks. Unknown risk profile is the general term used to refer to the underestimation of cloud security risks or the ignorance resulting from being unaware of other security risks associated with cloud computing.

## **Activity Details - Security in cloud computing Assessment**

a) How secure is cloud computing?

b) Is cloud computing mostly for larger or smaller organizations?

c) Is cloud computing for governments?

d) Is there a difference in PaaS, SaaS and IaaS as far as security is concerned?

e) Do you think cloud computing is appropriate for governments?

f) What is cloud security alliance and what does it do?

## **Assessment Submission**

This assessment is on the just concluded activity. You are supposed to answer each of the above questions in a paragraph format. This is an individual assignment and work should be submitted within a week to the instructor via an email to be provided.

## **Activity 2: Security in Grid Computing**

Grid security computing is challenging as a result of virtual organizations (VO) which are geographically dispersed organizations created to share resources and services within themselves. Their cross organizational nature renders the problem of implementing security if grid entities a herculean one. You may also want to appreciate that grids do not have a central point of control which implies that each service provider has to make their own risk impact assessment prior to interacting with the others.

### **Authentication**

This involves establishment of user identity, process or a resource which normally is validated using a username and a password. Kerberos is a network authentication process for client/server applications that uses symmetric key cryptography. Authentication in grids is via Public Key Infrastructure (PKI) which describes a security system that identifies entities through the use of X.509 certificates. Highly trusted organizations which are known as certifying authorities (CA) are responsible for issuing those identity certificates in which various VO have agreed on their usage terms.

### **Authorization**

It is the second step of the trust establishment in grid organization and involves validation of privileges assigned to a user or process to access a given resource in the grid. You can only be authorized to access a grid resource only after a successful authentication has been performed. This is mainly left to resource providers to grant or deny access based on the membership to the VO.

Globus Toolkit Gridmap file was one of the first authorization methods applied in grid computing. The gridmap file contains a list of global names of the grid users and local account names to where they are mapped. Access control is done by the host operating system in conjunction with prevailing local security policies which involves quite some work in maintaining a current version of the gridmap file. The Community Authorization Service (CAS) allocates resources to those who need them in the VO based on what the resource owners defined access rights. A CAS server acts as a trusted intermediary between the resources and the users in the VO.

### **Confidentiality**

You need to look at confidentiality as a way to hide sensitive data from people who have no rights to see it and of course, grids contain databases holding such data such as medical and financial information. Such data needs to be protected due to privacy laws in various countries or intellectual property rights. Cryptography is the most common approach to data confidentiality and it involves transforming data into unreadable format called cipher text and only those who possess a secret key can decipher the message into plain text for it to be read.

### **Current Grid and Cloud Security technologies**

All these security technologies are based on open standards and form a key part of grid security.

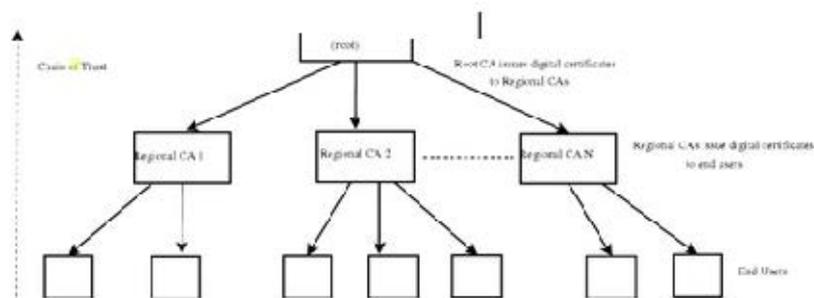
#### a) Public key Infrastructure (PKI)

It provides a way for secure communication in an insecure public network using public and private keys. It incorporates a trusted third party called a certifying authority (CA) which issues a digital certificate to individuals and organizations. It conforms to X.509 system where a distinct name of the user of the certificate is tied with its public key by a certifying authority. The private key of the certificate is kept securely with owner of the certificate while the digital certificate containing the public key is available for use by the public. A piece of data signed the private key can be decrypted using the public key and vice-versa.

Certifying authorities are responsible for publishing the Certificate Revocation List (CRL) which contains the serial numbers of those certificates that have become invalid due to expiration of validity period or due to some form of fraudulent activities. These CRLs are also signed by the CA that issues them and then published on their websites to preempt generation of false CRLs.

A comprehensive PKI must implement the following items:-

- Public key certificates
- A Certificate repository
- Certificate revocation
- Backup and recovery
- Support for non repudiation of digital signatures
- Automatic update of key pairs and certificates
- Management of key histories
- Support for cross certification
- Client side software interacting with all of the above in a secure, consistent and trustworthy manner.



Hierarchical structure of PKI

Source: Introduction to Grid computing, Frederic Magoules et al CRC Press 2009

PKI uses a hierarchical structure for establishing a chain of trust where at the lowest level are the end users who are issued with the digital certificate. At the next level are CAs who are authorized to issue certificates on a regional level. There is no fixed specification for the size of a region which can be as small as an organization or as big as a country. Each of the CAs has a digital certificate which are in turn signed by another CA which is at a higher level in the hierarchy. At the uppermost level are those CAs that issue certificate for small CAs. Please note that there can be more than one CA at the top of the hierarchy. These CAs are in the business of issuing digital certificates and are trusted by everyone. Supposing a user obtains the digital certificate of an individual. The user examines whether the certificate has been signed by a trusted CA and if the user does not trust the CA who has signed the certificate he or she may request the digital certificate of the CA which is turn signed by another CA at an upper layer in the hierarchy.

This can continue until the user finds that the certificate has been signed by a CA which he or she trusts. This chain of trust is important because a user may trust very few recognized CAs.

### b) Kerberos

It is a network authentication protocol developed by the Massachusetts Institute of Technology. It is a distributed authentication protocol that provides manual authentication to the client and server using symmetric key cryptography which means that the same key is used for both encryption and decryption of the message.

It provides a symmetric-key cryptography for client and servers which implies that the same key is used for both encryption and decryption of the message. The following are key terms as used in Kerberos network authentication:-

- Principal - The entity whose identity is being verified
- Verifier - Is the entity that verifies the identity of the principal
- Kerberos Realm - A realm is an administrative domain in which Kerberos function. These usually consist of the Internet Domain Name of the organization.
- Security Principals - These are entities recognized by the Kerberos realm which may either be a user or a process running on behalf of the user.

Key components of Kerberos authentication system

1. Key Distribution Center - KDC is a trusted third party which runs a service on a physically secure

system. It maintains a database of account information of all security principals in its docket.

KDC has two entities which are Authentication Server (AS) and Ticket Granting Server (TGS).

2. Authentication Server (AS) - Its purpose is to issue the Ticket Granting Ticket (TGT) which is used by the client to authenticate to the Ticket Granting Server (TGS). TGT is issued by the AS to the client and a normal ticket issued by the TGS to the client.

3. Ticket Granting Server (TGS) - It is a trusted third party that uses short term keys known as TGT to provide tickets to clients that want to authenticate to the server. All security principals in Kerberos share a TGT with TGS.

4. Ticket - It contains information which may be in plain and encrypted texts needed for clients authentication to the server.

### c) Grid Security Infrastructure (GSI).

It defines the complete architecture for implementation of security in grid computing as is part of Globus Toolkit. Some of its key highlights include:-

- i. Single sign-one. A proxy certificate is created the first time a user authenticates and can be used repeatedly for the period defined in the certificate.
- ii. Privileges delegation. Suppose you have three entities X,Y and Z. Y trusts X and Z trusts X. Suppose X wants Y to perform a task that requires access to a resource on Z. But as Z does not trust Y it cannot allow Y to access its resource. To solve this, X can delegate a part of its privileges to Y by issuing a proxy certificate which along with its private key is referred to as proxy credential. This proxy credential provides Y restricted access to those resources on Z that are needed to complete X job. This delegation of privileges is what is called credential delegation.
- iii. Inter-domain security support. A good grid security solution must provide support for interaction among different entities located in various security domains. However, a proxy agent on a local system should provide access to on behalf of a remote client based on local security policies.
- iv. Secure communication among the grid computing entities should in place.
- v. Authentication & Authorization. A good grid security infrastructure should be provide for secure and scalable authorization and authentication mechanisms.

### **Activity Details - Security in grid and cloud computing Assessment**

- a) How can a grid computing network be made more secure?
- b) Compare and contrast PKI and Kerberos grid security strategies.
- c) How do cloud and grid security approaches differ? Are similar?

### **Assessment Submission**

This assessment is on the just concluded activity. This individual assignment requires you to answer the questions in short essay type answers and submit them to your instructor within a week in an email to be provided.

### **Conclusion**

You have learnt that cloud and grid computing security threats include the traditional ones such as phishing, SQL injection and distributed denial of service attack among others. Others include unavailability of cloud services which can potentially be caused by power outages, sabotage and system failures. The sub contraction of services to third parties is also a potential problem for cloud services and may lead to data loss. These threats can be mitigated by using authentication, authorization, confidentiality, PKI, Kerberos and Grid Security Infrastructure.

### Lab exercise

#### **Description**

Amazon Web Services elastic compute cloud can launch instances of windows or Linux virtual machines. The cloud computing lab will provide students with access to virtual machines running various operating systems and applications. These virtual machines are grouped into several categories such as storage, content delivery, databases, networking, Internet of things, Enterprise applications and management tools among others. What is required of the students is a computer with Internet access which therefore eases the financial burden from them of acquiring the needed hardware and software.

#### **Objectives**

The major goal of this exercise is to create a lab that that accomplishes the following:-

- a. utilize a cloud platform that is remotely accessible
- b. available to students 24/7
- c. minimal financial cost to students
- d. scalable
- e. Variety of heterogeneous applications

Resource required

- a. Internet access
- b. PC/laptop

#### **Time**

1 hour or less (depending on your Internet connectivity)

#### **Submission requirements**

At step ix, you will receive an RSA private key file. Send this RSA private key file as an attachment to your instructor as evidence that you have followed the above instructions.

#### **Assessment criteria**

This lab exercise will contribute 10% of your overall marks for this course.

#### **Reference**

<https://aws.amazon.com>

Create and launch a Virtual Machine at Amazon Web Services

- i. Sign up for a free account on Amazon Web Services (<http://aws.amazon.com/>) and subscribe to EC2 service. You will be asked to provide a credit card number after which a verification code will be printed on the web page and which you enter on your mobile phone. This lab will be within the free tier category. (Read what the other tiers entail)
- ii. Go to EC2 console and create a virtual server which is referred to as an Amazon EC2 instance.
- iii. You can now choose one of the many ready to use Amazon Machine Image (AMI) e.g. Red hat linux, Ubuntu server, Microsoft Windows depending on the software running on your PC.
- iv. In the "choose an instance type" - choose under family the default general purpose since it is free tier eligible
- v. In the next step "configure instance details" - leave the choices at their default levels.
- vi. In the "add storage" step it shows a virtual HD volume type /dev/sda1 of 30GB - This can be edited but leave them for now.
- vii. This brings you to a page on "adding tags to your instance". Use your full name and your student registration number for the value.
- viii. This brings us to a page on Create a new Security group (Configure Firewall). Leave the rules at their default. Continue.
- ix. Go to the next step "review instance launch" where a dialog box will ask you to "select an existing key pair or create a new key pair" - Create and download. You will receive a file with a .pem extension. This is the RSA private key file that you should keep on your local machine. Send the RSA private key file as an attachment to your instructor as evidence that you have followed the above instructions.
- x. You now be in "launch status" - Save the AMI you created and launch in later EC2 sessions.

### UNIT SUMMARY

In this unit, we have explored the security vulnerabilities that apply to both cloud and grid computing. It has been noted that these being fairly new technologies, the threats they are exposed to are legion. Coupled with the fact that there is no common point of control e.g. in grid computing makes the problem even more catastrophic should a security breach occur. The geographical dispersion of resources in both cloud and grid computing is another area of security concern and you should be aware of these threats in order to mitigate against the risks involved.

### Assessment

Check your understanding!

#### Formative assessment

- a) What is a stand alone computer?
- b) Why is it important for an organization to assign users different levels of privilege in cloud computing?
- c) Explain what is meant by data lock-in?
- d) What is service hijacking?
- e) Why is difficult to enforce security in grid networks?
- f) What is an example of a client server application?
- g) What is the purpose of a gridmap file?
- h) Name three current grid security technologies.

### Answers

- a) No processing power of its own
- b) Its based on a need basis eg guest, administrator, network analyst etc
- c) Its where cloud services are overwhelmed by system outage such as power failure, system breakdown, sabotage etc.
- d) Theft of user credentials such as usernames and passwords.
- e) No central point of control
- f) User and cloud/grid applications
- g) List of global names of grid users and local account names
- h) PKI, Kerberos and GSI

# Unit 4 : Grid Projects and applications

## Unit Introduction

You will note that the proliferation of many worldwide grid projects has directly contributed to the development of the grid. In many of these projects, many dedicated networks have been set up and these serve as the main basic building blocks of grid research. In the last decade or so, grid computing has become a very dynamic discipline where many scientific fields are seeking solutions in grid computing such as earth observation, astronomy, e-business, engineering and high energy physics.

## Unit Objectives

- Compare and contrast various grid projects
- Discuss real life applications of grid computing
- Examine how grid applications can be used in the local context

### KEY TERMS

**Data grid** - It is a distributed infrastructure integrating intensive computational capability and large scale databases.

**Metadata** - It describe the content and structure of data, information content represented by the file and the circumstances under which the data was obtained.

**Genome** - The complete set of genetic material of a human, animal, plant or other living thing.

**CERN(European Organization for Nuclear Research)** - Is a European research organization established in 1954 that operates the largest particle physics laboratory in the world.

**EMBL(European Molecular Biology Lab)** - Is a molecular biology research institution created in 1974 operating from Germany, UK, France, and Italy to train scientists in development of new instruments, methods and technology.

**Particle accelerator** - It speeds up and increases the energy of a beam of particles by generating electric fields that accelerate the particles and magnetic fields that steer and focus them.

**Biodynamic** - Is a method of organic farming that proponents describe as holistic understanding of the agricultural processes.

### Learning activities

#### Activity 1:- DataGrid

This is a European project that targets building the next generation of computing infrastructure providing intensive computational ability in shared large scale databases. It consists of Replica Management called Reptor, which is a a middleware that enables management of files on the grid. Reptor is a virtual single access point which the enables the user to access the replica management system providing transparent access to the underlying grid infrastructure through a SOAP interface. Reptor can be configured as a distributed service providing file transfer, file registration and file access. Access to requests coming from one VO are scheduled through well defined priorities so the available resources are made of in an optimal manner. Reptor offers consistency services in order for files to be consistent with the replicas such as:-

- Lifetime management - These replicas are created for temporary use and are deleted once their predefined lifetimes have expired.
- Update propagation - This is where once you make changes to a file, the said change should be propagated to all the replicas.
- Inconsistency detection - Should a system fail or crash, this feature should find the inconsistency caused by such.

#### **Activity Details - DataGrid Assessment**

Explain the functions of the following as pertains the DataGrid:-

a) Replica Location Service (RLS)

b) Replica Optimization Service (ROS)

c) Mass storage management in relation to Storage Resource Broker (SRB) and Sequential Access

to data via Metadata (SAM).

#### **Assessment Submission**

This assessment is on the just concluded activity. Answer each of the above questions individually in a short answer format and submit the work within a three days to the instructor via an email to be provided.

#### Activity 2:- BioGrid (Biological General repository for Interaction Datasets)

It is motivated by 500 databases which are composed of a collection of individual databases regarding genome information on the Internet. Different formats are used for storing data for instance the same protein could be described in a different way in two databases.

To overcome this, a database conversion system has been developed to transform heterogeneous database format into an XML standard format using conversion rules. You can then access the database services through XML based SOAP protocol. An automatic update system is also being developed to update new sequences of data that are added into the system everyday including a data comparison mechanism to reduce data redundancy.

### **Activity Details - BioGrid Assessment**

Log on to <http://www.biogrid.org/> and write a three page report on the activities of the biogrid including the latest research developments. This report should include who funds the biogrid projects and the different search mechanisms available on the site.

### **Assessment Submission**

This assessment is on the just concluded activity. This should be a three page report double spaced and should be submitted within a week to the instructor email to be provided.

### **Activity 3 European Grid Infrastructure**

The EGI is a publicly funded e-infrastructure put together to give scientists access to more than 530,000 logical CPU's, 200 PB of disk capacity and 300 PB of tape storage to drive research and innovation in Europe. It is a federation of resource infrastructure providers working together to provide leading edge computing services needed by European researchers. The infrastructure provides high throughput computing and cloud compute and storage capabilities. Resources are provided by about 350 resource centers distributed across 56 countries in Europe, Asia-Pacific region, Canada and Latin America. The National Grid Initiatives (NGI) are organizations set up by individual countries to manage the computing resources they provide to the EGI. They also provide the country's single point of contact for governments, research communities and resource centers as regards ICT services for e-science. The NGI's are the main stakeholders together with CERN and EMBL.

### **Activity Details - EGI Assessment**

Log on to <http://www.egi.eu/> and analyse the services provided and solutions. Click on the case studies section and under engineering and technology, read the case study on "predicting the risk of dam failure".

### **Assessment Submission**

This assessment is on the just concluded activity. Write a detailed report covering five pages on the services and solutions provided by the European Grid Infrastructure. On the case study write a one- page report on how grid technology can mitigate the effects of a dam spillage in the event of a toxic spill.

### Activity 4 Other Large scale grids

TeraGrid is one of the most significant high speed networks constructed for grid computing and funded by the National science Foundation (NSF) in 2001 to connect five supercomputer centers which are:-

- Argonne National Laboratory (Chicago)
- Pittsburgh super computing center
- California Institute of Technology (Los Angeles)
- San Diego super computer center
- National Center for Super Computing Applications (NCSA)

TeraGrid provides open access for scientific research with users making requests for resource allocation from a big range of extremely powerful computer systems. Open Science Grid (OSG) is also another NSF funded initiative with a very large number of participants with interests in particle and nuclear physics, astrophysics, bio-informatics and gravitational-wave science. The South Eastern Research Association (SURA) is a collaborative venture between universities that provides a shared grid computing platform but is not focused on any application specific domain. It can take in new members working on new and specific projects but they need to know the required software and hardware.

### Activity 5 Grid Applications

a) Physical Sciences.

The Large Hadron Collider (LHC) is the worlds largest and most powerful particle accelerator which started in 2008 and is the latest addition to CERN's accelerator complex. It consists of a 27 km ring of superconducting magnets with a number of accelerating structures to boost the energy of the particles along the way. Inside the accelerator, the two high energy particle beams travel at close to the speed of light before they are made to collide. The beams travel in opposite directions in separate beam pipes and are guided around the accelerator ring by a strong magnetic field maintained by superconducting electromagnets.

b) Astronomy

There is a large amount of astronomy based data on the Internet in which research is being conducted by among others D-Grid, DutchGrid, Chinese OSG, ChinaGrid and AmericanSDG. Astronomy is a dynamic filed of research as human beings have always been curious about the outer space since time immemorial.

c) Biomedical

This is an area of research that is aimed at pharmaceutical development and biodynamic elucidation where experimental and measuring equipment is directly connected to computer resources to rapidly obtain and process data.

### d) Earth Observation

Earth observation satellites return gigabytes of data daily which can be used to analyze the ozone profiles, forecast floods or detect oil. The grid allows data to be shared between different countries and establish a testbed specialized for ozone data.

### Conclusion

In this activity, you have learnt about some active grid projects such as the DataGrid, BioGrid and the European Grid infrastructure. Common to these projects is the aspect of sharing information in form of large databases spanning several countries. It is important you realize that grid infrastructures due to the massive scale and the quantity of resources involved are in most cases government funded.

### Activity Details - Grid Applications

In this activity, you are required to do some research on BelnGrid and list at least seven business applications that can be solved using grid technology.

### Assessment

This assessment is on the just concluded activity. You are required to list seven business applications that can be supported by way of grid technology. This is a one page assignment where you are expected to list the application solved by grid technology accompanied by a brief explanation.

## UNIT SUMMARY

In this unit, you have learnt about the various grid projects such as the data grid which is a large European project geared towards building the next generation computing infrastructure. The BioGrid is a biological database that strives to provide a comprehensive resource of protein-protein and genetic interactions of all major model organism species and the European Grid Infrastructure for driving research in Europe. A number of grid solutions in physical sciences such as astronomy, biomedical and earth observation applications have also been briefly discussed. The largest grid computing project in the world which is found in the US is known as the TeraGrid and it connects a geographically distributed group of super computers in five major cities.

### Unit Assessment

Check your understanding!

- a) What is the D-grid?
- b) What is name of the Chinese based grid computing platform?
- c) What is metadata?
- d) Name three benefits of hosting big scientific applications on the grid.
- e) What is the hot-spot of grid technology research?
- f) List three job scheduling ongoing research projects.
- g) Which Japanese grid has carried out nanoscience research?

### Answers.

- a) It is a grid devoted to the processing of and access to large amounts of scientific data.
- b) SDG
- c) It refers to the description of data within data in databases.
- d) Storage, distributed computations,, sharing
- e) Job scheduling - It deals with the interaction of services through an extensible and integrated resource management.
- f) CrossGrid, GridPP,CNGrid
- g) NAREGI

### MODULE SUMMARY

This module has covered introduction to Grid and Cloud computing. The module has covered overview of grid and cloud computing in which the concepts of grid, cloud computing and cloud services were introduced; cloud applications in which an attempt was made to cover various aspects of cloud applications such as cloud computing paradigms, cloud applications architectural styles and cloud computing limitations; the module also covered security in grid and cloud computing including security in cloud computing and security in grid computing; and finally the module covered grid projects and applications with a focus on data grid, biogrid, the European Grid infrastructure and other large scale grids as well as grid applications. The module specifically brought out clearly the advantages of cloud computing as well as the differences in concept and architecture of grid computing and cloud computing.

### Module Assessment

Attempt all the questions:

1. Discuss the major trends in computing that have led to the emergence of Cluster computing
2. Describe the design issues and the architecture of Cluster computing systems.
3. What is a Single System Image (SSI) ? Describe different SSI services that cluster middleware need to support.
4. Discuss SSI architecture of implementing at Operation System and Tool levels with a suitable example.
5. What are the key distinctions between Cluster and Grid computing?
6. Discuss two commercial applications of Clusters and Grids.
7. Discuss in detail the architecture of Grid Computing systems.
8. Discuss the design issues of Grid Resource management systems.
9. Discuss the architecture of a Grid Resource Broker with a suitable example.
10. Discuss Deadline-and-Budget Constrained Time and Cost optimization scheduling algorithms for Grid Computing.
11. List the methods to compare computing cluster performance.
12. List the most powerful clusters in the World.
13. Explain how monitoring of the computer cluster can be done
14. Describe the bases for cloud computing (demands, technical possibilities).
15. Explain with examples the meaning of IaaS.
16. Explain with examples the meaning of SaaS.
17. Explain with examples the meaning of PaaS.
18. How can one achieve reliability/security in computer cluster.

### Unit Readings and Other Resources

The readings in this unit are to be found at course level readings and other resources







**The African Virtual University  
Headquarters**

Cape Office Park

Ring Road Kilimani

PO Box 25405-00603

Nairobi, Kenya

Tel: +254 20 25283333

[contact@avu.org](mailto:contact@avu.org)

[oer@avu.org](mailto:oer@avu.org)

**The African Virtual University Regional  
Office in Dakar**

Université Virtuelle Africaine

Bureau Régional de l'Afrique de l'Ouest

Sicap Liberté VI Extension

Villa No.8 VDN

B.P. 50609 Dakar, Sénégal

Tel: +221 338670324

[bureauregional@avu.org](mailto:bureauregional@avu.org)