Edge Hill University

Cloud Computing

Topic 8A, Session 1

Aim & Objectives

- This topic aims to introduce the development and management of applications deploying in a computing cloud, and explore the frameworks existing to support the management of both data and applications.

- Learn the concepts of distributed computing
- Introduce computing virtualisation
- Explore the cutting-edge techniques of cloud computing
- Understand how cloud application management works
- Introduce Apache Hadoop
- Explore the security issues associated with cloud computing

2

Introduction to Distributed Systems

- A **Distributed System** is a collection of autonomous computers interconnected by a computer network with distributed system software to form an integrated computing facility.

"A distributed system is one in which the failure of a computer you didn't even know existed can render your own computer unusable"

--Leslie Lamport (the winner of the 2013 Turing Award)

https://www.igibutorial.com/distributed-systems/challenges-for-a-distributed-system

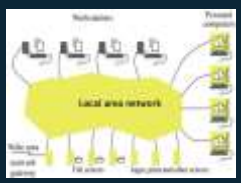
3

Distributed Systems vs. Distributed Computing

- A distributed system is one in which hardware or software components located at networked computers communicate and coordinate their actions only by message passing.

Distributed Systems vs. Distributed Computing Cont…

- In the term distributed computing, the word "distributed" means spread out across space. Thus, distributed computing is an activity performed on a spatially distributed system.



Advantages of a Distributed System

1. Resource sharing
   -- Only one device, share it over the network
2. Computation speedup
   -- Parallel computation on multi-computer
3. Reliability
   -- One node fails but the service goes on
4. Communication
   -- Share info between people & processes that are far apart

Why Distributed Systems Are Needed

- Inherently distributed applications
- Performance/cost
- Resource sharing
- Flexibility and extensibility
- Fault tolerance and availability
- Scalability
- Network connectivity is increasing.
- Combination of cheap processors often more cost-effective than one expensive fast system.
- Potential increase of reliability.

7

Consequences

- Processes within a distributed system
  - Execute concurrently
  - Interact in order to co-operate to achieve a common goal
  - Co-ordinate their activities and exchange information by means of messages transferred over a communication network
- Typical problems of concurrent systems – **deadlocks (a** deadlock is a state in which each process is waiting for another to take an action or release a resource.)
- Some applications may need to be tweaked to take full advantage of the **new model**
- Easy access to shared data raises **security concerns**.

8

Consequences—Contd.

- Absence of a global state
  - In the general case, there is no single process in the distributed system that would have a knowledge of the current global state of the system – Due to concurrency and message passing communication
- New Failures
  - If network fails, how is this detected and recovered

9

Distributed Computing vs Cloud Computing

- In Distributed Computing, a task is distributed amongst different computers in a network for computational functions to be performed at the same time
- In Cloud Computing an on-demand network model is used to provide access to shared pool of configurable computing resources over the Internet.

10

Grid Computing vs Cloud Computing

- Job Scheduling is the core value and aim of grid technology.
- It can divide a huge task into a lot of independent and non related sub tasks.
- Even if any node fails or does not return results, the whole process will not be affected.

- Cloud Computing makes a huge resource pool through grouping all the resources.
- The resources provided by cloud are assigned to complete a special task.
- For example, a user may apply resources from the resource pool to deploy an application.

11

Role of Internet in Cloud Computing

- **Internet** is a global system of interconnected computer networks. It is a medium to deliver resources and services that **Cloud Computing** goes to provide.

SaaS: Software as a Service.
PaaS: Platform as a Service.
IaaS: Infrastructure as a Service.

12

Issues & Challenges

13

1. Heterogeneity of components

- Variety or differences that apply to computer hardware, network, OS, programming language and implementations by different developers.
- All differences in representation must be dealt with to enable message exchange.

14

2. Openness

- System can be extended and re-implemented in various ways.
- Cannot be achieved unless the specification and documentation are made available to software developer.

15

3. Transparency

- **Access transparency** ensures local and remote resources to be accessed using identical operations.
- **Location transparency** enables resources to be accessed without knowledge of their location.
- **Concurrency transparency** enables several processes to operate concurrently using shared resources without interference between them.
- **Replication transparency** enables multiple instances of resources to be used to increase reliability and performance without knowledge of the replicas by users or application programmers.
- **Failure transparency** enables the concealment of faults, allowing users and application programs to complete their tasks despite the failure of hardware or software components.
- **Mobility transparency** allows the movement of resources and clients within a system without affecting the operation of users or programs.
- **Performance transparency** allows the system to be reconfigured to improve performance as loads vary.
- **Scaling transparency** allows the system and applications to expand in scale without change to the system structure or the application algorithms.

16

4. Security

- Security for information resources in distributed systems has 3 components :
  a. **Confidentiality:** protection against disclosure to unauthorized individuals.
  b. **Integrity:** protection against alteration/corruption
  c. **Availability:** protection against interference with the means to access the resources.

17

5. Scalability

- Distributed computing operates at many different scales, ranging from small Intranets to Internet.

- Challenges:
  - ✓ a. controlling the cost of physical resources.
  - ✓ b. controlling the performance loss.
  - ✓ c. preventing software resources running out.
  - ✓ d. avoiding performance bottlenecks.

18

7. Failure Handling

- Failures in a distributed system are partial – some components fail while others can function. This makes the handling of failures a challenge

1. **Detecting failures:** some failures cannot be detected but may be suspected.
2. **Masking failures:** hiding failure not guaranteed in the worst case.

8. Concurrency

- Where applications/services are processed concurrently it may result in a conflict of operations with one another and produce inconsistence results.
- Each resource must be designed to be safe in a concurrent environment.

19

20