VIRTUALIZATION AND CLOUD COMPUTING

CSE 423

Computing is being transformed into a model consisting of services that are commoditized and delivered in a manner similar to utilities

LET'S BREAK IT DOWN

Overview of Distributed computing

Virtualization techniques

Introduction to Cloud Computing

Migrating into a Cloud

Understanding cloud architecture

Cloud mechanisms

EVALUATIONS

Three evaluations

Test 1 and 2 will be analytical MCQ questions relevant to scenarios and Technology covered in syllabus

Term Paper
The student has to write an original paper on the latest technological advancement

The Internet plays a significant role in cloud computing for representing a transportation medium of cloud services that can deliver and accessible to cloud consumer.

HISTORY

- The last decades have reinforced the idea that information processing can be done more efficiently centrally, on large farms of computing and storage systems accessible via the Internet
- When computing resources in distant data centers are used rather than local computing systems, network-centric computing and network-centric content are accessed.
- Advancements in networking and other areas are responsible for the acceptance of the two new computing models and led to the grid computing movement in the early 1990s and, since 2005, to utility computing and cloud computing.
- In utility computing the hardware and software resources are concentrated in large data centers and users can pay as they consume computing, storage, and communication resources.
- Cloud computing is a path to utility computing embraced by major IT companies such as Amazon, Apple, Google, HP, IBM, Microsoft, Oracle, and others.

NETWORK-CENTRIC COMPUTING

The concepts and technologies for network-centric computing and content evolved through the years and led to several large-scale distributed system developments:

- The Web and the semantic Web are expected to support the composition of services (not necessarily computational services) available on the Web.
- The Grid, initiated in the early 1990s by National Laboratories and Universities, is used primarily for applications in the area of science and engineering.
- Computer clouds, promoted since 2005 as a form of service-oriented computing by large IT companies, are used for enterprise computing, high-performance computing, Web hosting, and storage for network-centric content.

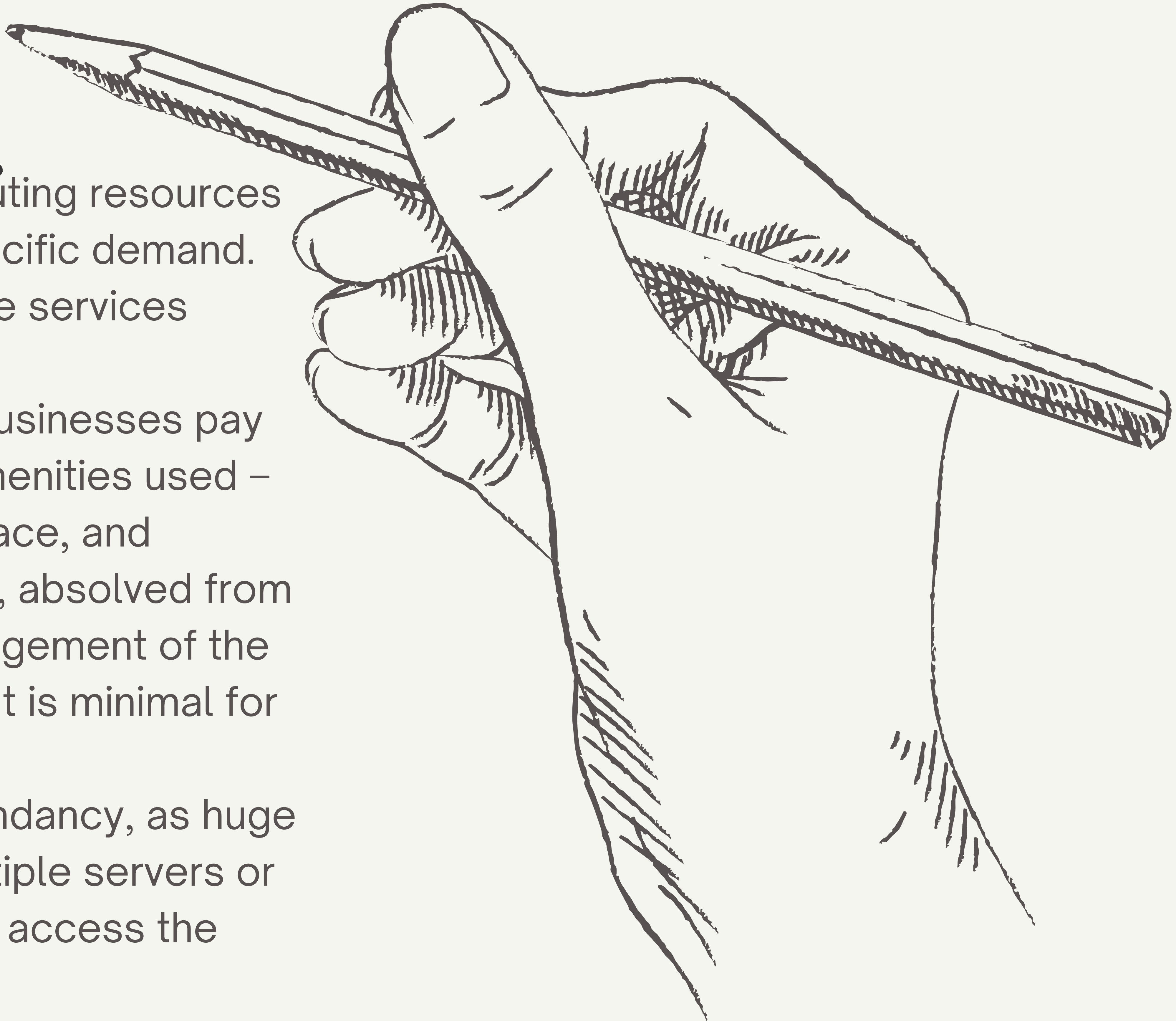
PEER-TO-PEER SYSYEMS

The user-centric model, in place since the early 1960s, was challenged in the 1990s by the peer-to-peer (P2P) model. P2P systems can be regarded as one of the precursors of today's clouds. This new model for distributed computing promoted the idea of low-cost access to storage and central processing unit (CPU) cycles provided by participant systems; in this case, the resources are located in different administrative domains.

P2P systems exploit the network infrastructure to provide access to distributed computing resources. Decentralized applications developed in the 1980s, such as Simple Mail Transfer Protocol (SMTP), a protocol for email distribution.

UTILITY COMPUTING

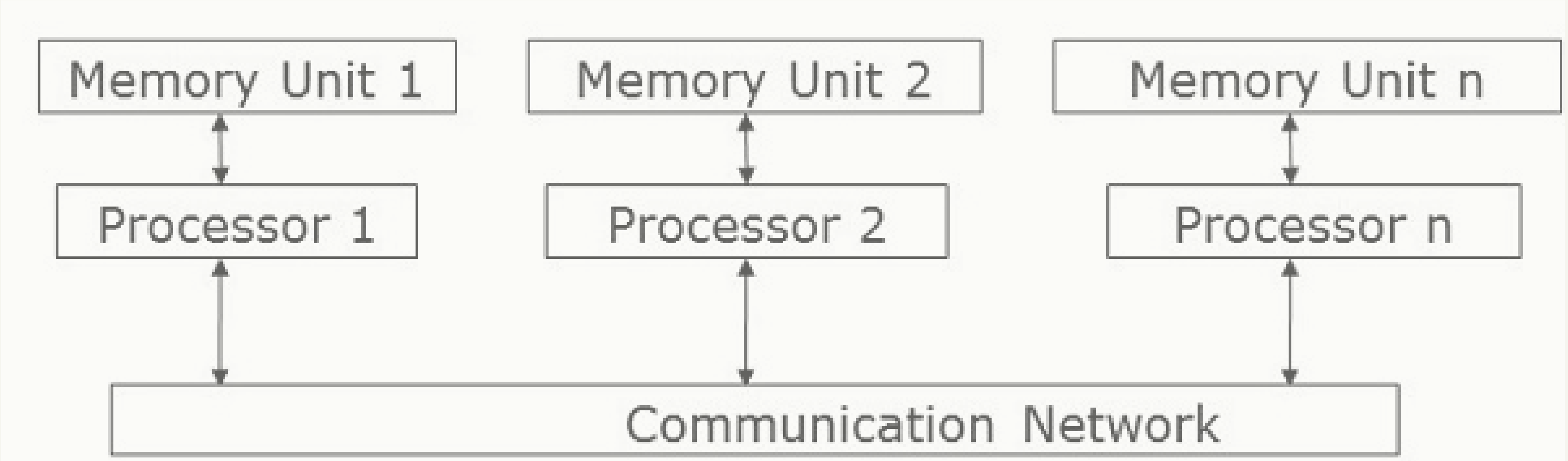
- Utility computing is a model in which computing resources are provided to the customer based on specific demand. The service provider charges exactly for the services provided, instead of a flat rate.
- The foundational concept is that users or businesses pay the providers of utility computing for the amenities used – such as computing capabilities, storage space, and applications services. The customer is thus, absolved from the responsibility of maintenance and management of the hardware. Consequently, the financial layout is minimal for the organization.
- Utility computing helps eliminate data redundancy, as huge volumes of data are distributed across multiple servers or backend systems. The client, however, can access the data anytime and from anywhere.



PARALLEL SYSTEMS

- The concepts introduced in this section are very important in practice. Communication protocols support coordination of distributed processes and transport information through communication channels
- Parallel computing allows us to solve large problems by splitting them into smaller ones and solving them concuParallel hardware and software systems allow us to solve problems demanding more resources than those provided by a single system and, at the same time, to reduce the time required to obtain a solution.
- The speed-up measures the effectiveness of parallelization; in the general case, the speed-up of the parallel computation is defined by Amdahl's Law
- Coordination of concurrent computations could be quite challenging and involves overhead, which ultimately reduces the speed-up of parallel computations. Often the parallel computation involves multiple stages, and all concurrent activities must finish one stage before starting the execution of the next one

ELEMENTS IN PARALLEL COMPUTING MODEL



- **Modularity:** A complex system is made of components, or modules, with well-defined functions. Modularity supports the separation of concerns, encourages specialization, improves maintainability, reduces costs, and decreases the development time of a system.
- **Client-Server Paradigm:** The ubiquitous client-server paradigm is based on enforced modularity; this means that the modules are forced to interact only by sending and receiving messages.
- **Remote Procedure Call (RPC):** RPC is often used for the implementation of client-server systems interactions.
- **Atomic Actions:** a multistep operation should be allowed to proceed
- to completion without any interruptions, and the operation should be atomic.

DISTRIBUTED SYSTEMS

- A distributed system is a collection of autonomous computers that are connected through a network and distribution software called middleware, which enables computers to coordinate their activities and to share the resources of the system. A distributed system's users perceive the system as a single integrated computing facility.
- A distributed system has several characteristics: Its components are autonomous, scheduling and other resource management and security policies are implemented by each system, there are multiple points of control and multiple points of failure, and the resources may not be accessible at all times.
- Distributed systems can be scaled by adding additional resources and can be designed to maintain availability even at low levels of hardware/software/network reliability.
- Distributed systems have been around for several decades. For example, distributed file systems and network file systems have been used for user convenience and to improve reliability and functionality of file systems for many years.

DISTRIBUTED SYSTEMS

- The remote procedure call (RPC) supports inter-process communication and allows a procedure on a system to invoke a procedure running in a different address space, possibly on a remote system
- A communication channel provides the means for processes or threads to communicate with one another and coordinate their actions by exchanging messages
- These two abstractions allow us to concentrate on critical properties of distributed systems without the need to discuss the detailed physical properties of the entities involved

KEY ADVANTAGES

Distributed computing makes all computers in the cluster work together as if they were one computer. While there is some complexity in this multi-computer model, there are greater benefits around:

- Scalability. Distributed computing clusters are easy to scale through a “scale-out architecture” in which higher loads can be handled by simply adding new hardware (versus replacing existing hardware).
- Performance. Through parallelism in which each computer in the cluster simultaneously handles a subset of an overall task, the cluster can achieve high levels of performance through a divide-and-conquer approach.
- Resilience. Distributed computing clusters typically copy or “replicate” data across all computer servers to ensure there is no single point of failure. Should a computer fail, copies of the data on that computer are stored elsewhere so that no data is lost.
- Cost-effectiveness. Distributed computing typically leverages low-cost, commodity hardware, making initial deployments as well as cluster expansions very economical.

DIFFERENCE BETWEEN

PARALLEL COMPUTING

- Multiple processors perform multiple tasks assigned to them simultaneously. These tasks are broken down from a single main problem.
- Done using a single computer
- Memory in parallel systems is either shared or distributed between processors
- Multiple processors perform processing
- Parallel computing helps to significantly increase the performance of the system, provides concurrency, and saves time

DISTRIBUTED COMPUTING

- Multiple computers perform tasks at the same time to achieve a single decided goal through networked computers.
- Requires multiple computers
- Each computer has its own memory
- Multiple computers perform multiple operations
- Allows scalability, to share resources and helps to perform computation tasks efficiently

DIFFERENCE BETWEEN

PARALLEL COMPUTING

- To provide synchronization all processors share a single master clock
- Environments are tightly coupled
- Used in places requiring excessively higher and faster processing power
- Example: Supercomputers

DISTRIBUTED COMPUTING

- There is no global clock, uses synchronization algorithms
- Environments might be loosely coupled or tightly coupled
- Used when computers are located at different geographical locations and speed doesn't matter
- Example: Facebook