COMPS267F Chapter 11

Distributed Systems

Dr. Andrew Kwok-Fai LUI


COMPS267F STUDENTS ARE DOING FINE!
Well Done!

Aim of the Chapter

- Discusses distributed systems as a configuration for running applications that involves multiple computers connected to a network.
- Distributed systems are almost everywhere
  - Computer networks and mobile devices are ubiquitous

introduction

Distributed Systems

- Multi-computer systems
  - Computers are networked and work together
  - Providing a service to its users
- Computers distributed in different locations or in the same site
- Characteristics different from single computer system

Virtual System

- A distributed system is a collection of independent computers
  - Networked
  - Providing a coherent service
  - Presenting as a single system
    - Software such as OS can create this illusion

Computer Network or Distributed Systems

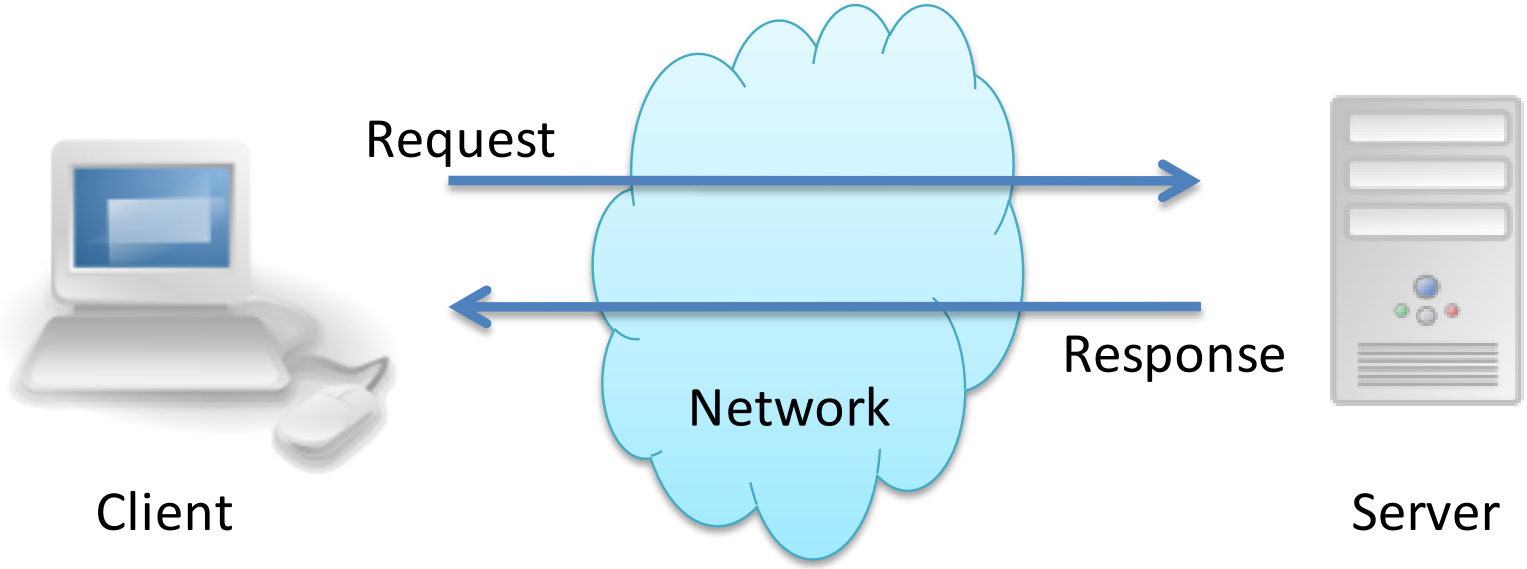
- Differences between a computer network and a distributed system:
  - A computer network: a collection of independent computers that are interconnected
  - Distributed systems are built on top of computer networks
    - The network is one of the resources of a distributed system
    - The system parts work together for a common objective

Transparency to Users

- The existence of multiple autonomous computers in a distributed system is transparent to its users
  - The How and Where of the systems are not seen by the users

simple model of distributed systems

Client-Server Systems



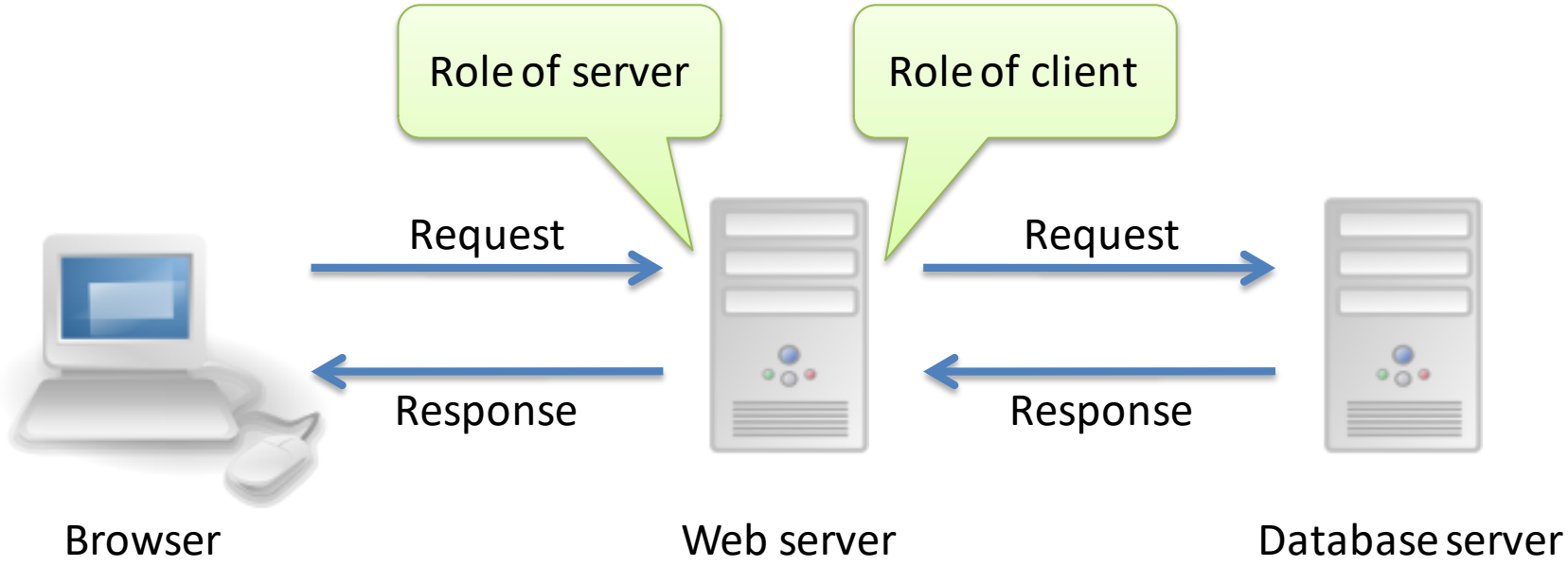
Client-Server Systems

- A common way to share resources.
- Client and server are two systems connected
  - Server manages certain resources such as printers, files or database records.
  - Client accesses these resources by sending a request.
  - The server, upon receiving the request, provides the required service by sending a response to the client.

Client-Server Systems

- Client and server are the roles in the systems.
  - Server: service provider
  - Client: service requestor
- Benefit is clear when there are multiple clients.
  - The resources of server are shared.

Client-Server Systems



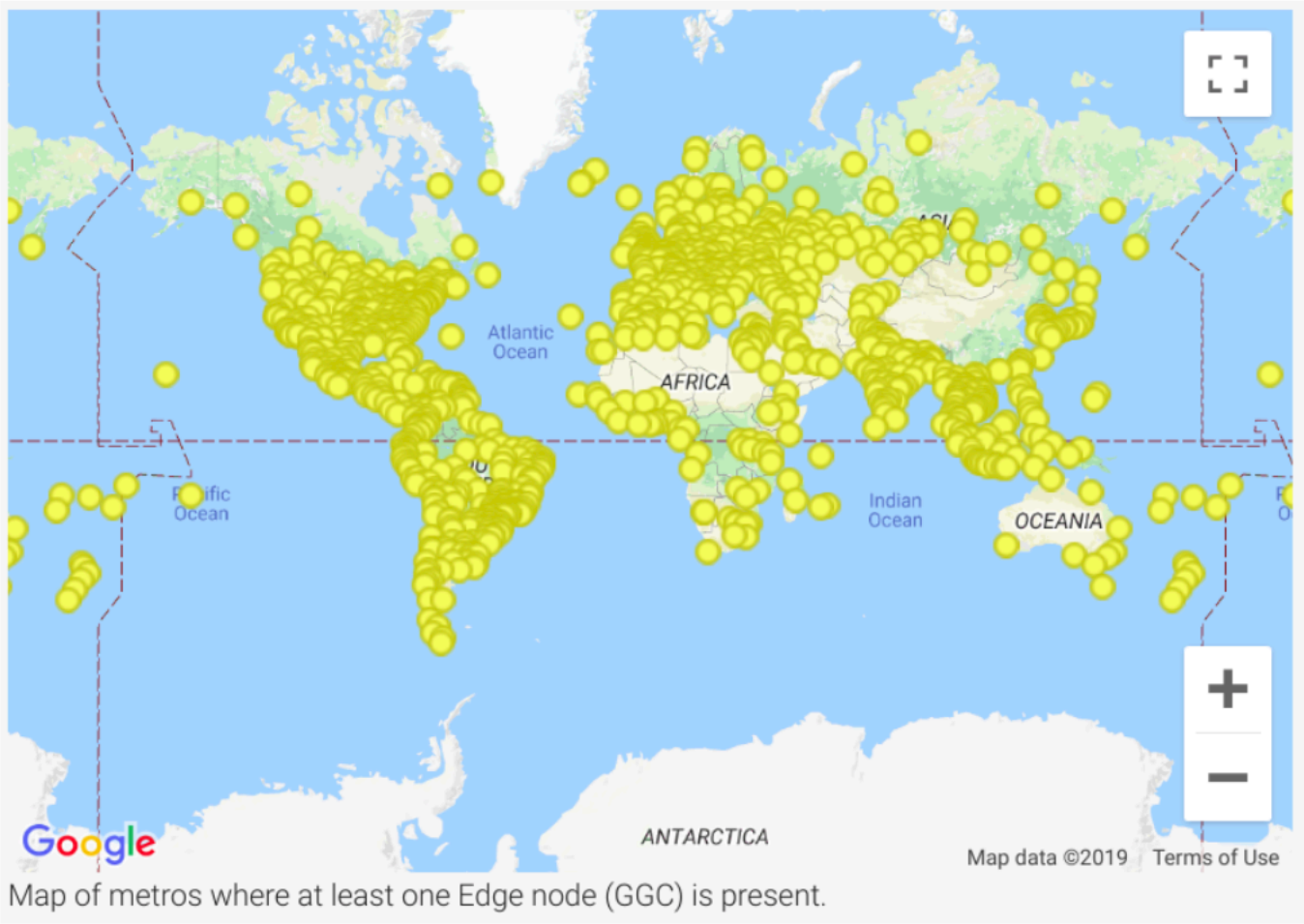
Characteristics of Distributed Systems

- Creating an illusion that the collection of independent computers is simply a single coherent system.
  - Heterogeneity
  - Location transparency
  - Access transparency
  - Migration or mobility transparency
  - Replication transparency
    - Increased reliability
  - Concurrency transparency
    - Handling synchronization
  - Failure transparency

Characteristics of Distributed Systems

Features	***Remarks***
Heterogeneity	Computers of the distributed system may vary with respect to computer hardware (e.g. CPU type, memory capacity, and I/O facility), operating systems, and interconnection networks.
Location transparency	Resources can be accessed without knowing their physical locations. Users might need to know the names of resources but not their locations.
Access transparency	Local resources and remote resources are accessed in the same way. Whether the resources are from local or remote machines is irrelevant, as the way to access them does not differ.
Migration (mobility) transparency	Movement of resources and users' submitted jobs within the distributed system without affecting how the users operate the system

Characteristics of Distributed Systems



Map of metros where at least one Edge node (GGC) is present.

Trending videos are moved and cached at closer Edge node

Characteristics of Distributed Systems

Replication transparency	Multiple instances of resources to be present throughout the system without the user or application programmers knowing anything about the replicas. Increases the reliability and performance of the services provided by the system.
Concurrency transparency	Multiple users can use the same resource at the same time without external mechanisms, to avoid interference. The users may not even be aware that there are other users who access the same resource. An example of this transparency is multiple processes, or users accessing a shared database simultaneously.
Failure transparency	Service is as usual when some failures occurred within the system Failure types include hardware, software, or communication link failures Users are not aware of the failures, to a certain extent, within the system.

computer clusters

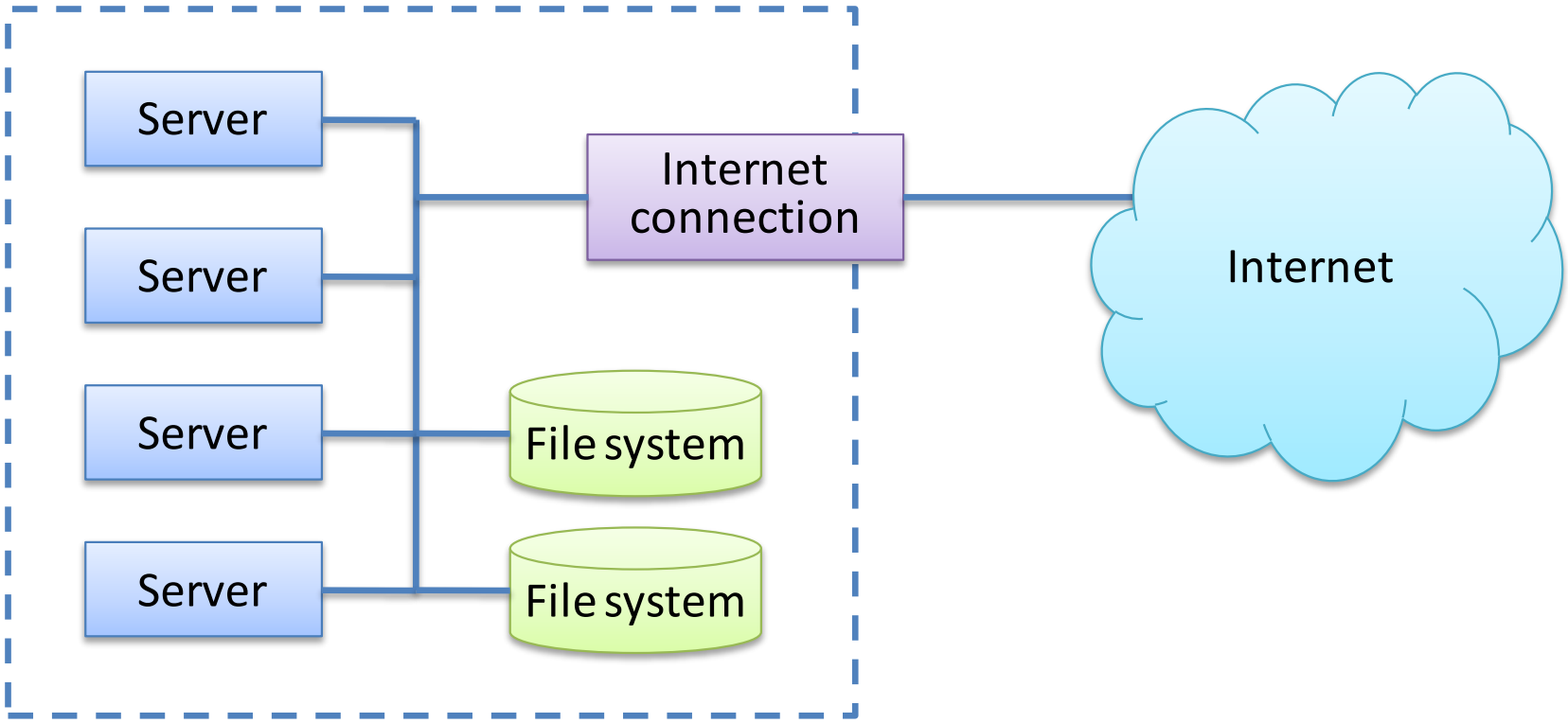
Computer Clusters

- A group of dedicated computers connected and configured to work together as a unit.
- Users of the cluster are not aware of multiple computers present in the system

Computer Clusters: High Performance Web Servers

- Web servers for enterprises satisfy non-functional requirements:
  - Scalability
  - Availability
- Single computer usually cannot satisfy the requirements
- A cluster for a web server contains a group of server machines and other resources, such as file systems
  - Presented as a single server system

Computer Clusters: High Performance Web Servers



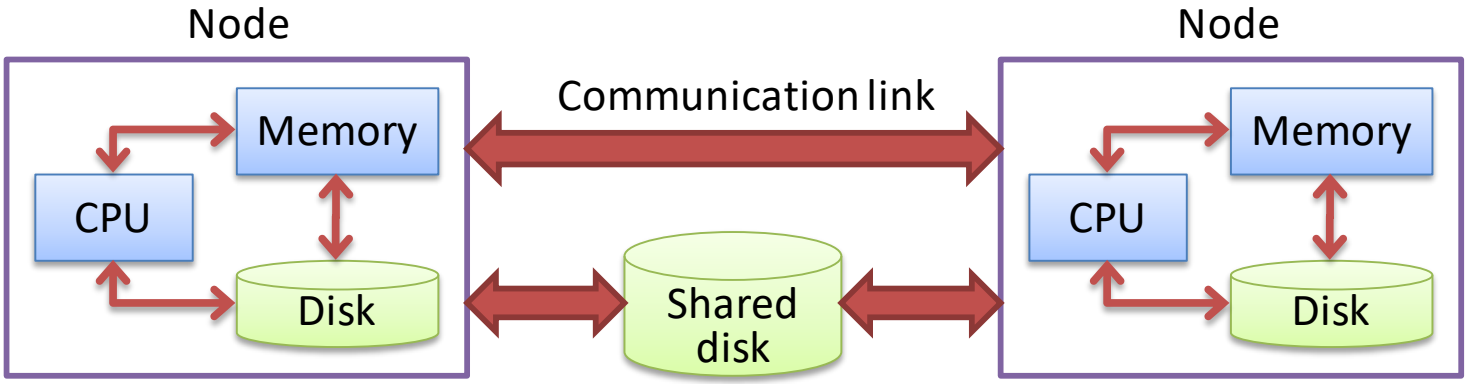
A cluster for web server

Computer Clusters: High Performance Web Servers

- Each server within the cluster handles connections independently.
- Failure: will not affect overall web service
- A failure in any server will not affect the overall Web
  - Only causes a gradual degradation in response time
    - One less server is used to provide the Web services

Major Configurations of Computer Clusters

(a) Shared-nothing configuration



(b) Shared-disk configuration

Major Configurations of Computer Clusters

- Shared-nothing configuration
  - Own local disk and no access of shared resources.
  - Links for communication of control and data messages
    - A lot of data transfer
  - Inter-node communication should be minimized

Major Configurations of Computer Clusters

- Shared-disk configuration
  - Share a disk on top of their individual local disks
  - No need to transfer data using the communication links that connect the nodes.
  - Data integrity during data update

Characteristics and Benefits of Clusters

- Computers in a cluster work for each other.
  - Owned by a single company or organization.
  - Situated in a single location or scattered in multiple locations.
  - Usually homogeneous
    - Identical or very similar in configuration.
    - Easy for both purchase and control.

Characteristics and Benefits of Clusters

- Scalability
  - Absolute scalability
  - Incremental scalability: adding new computers
- Fault tolerance
  - Fail-over and switch
- High availability
  - Adding/removing nodes happen without shutting down system
- Load-balancing
- Superior performance-to-price ratio
  - Inexpensive computers

PC Cluster

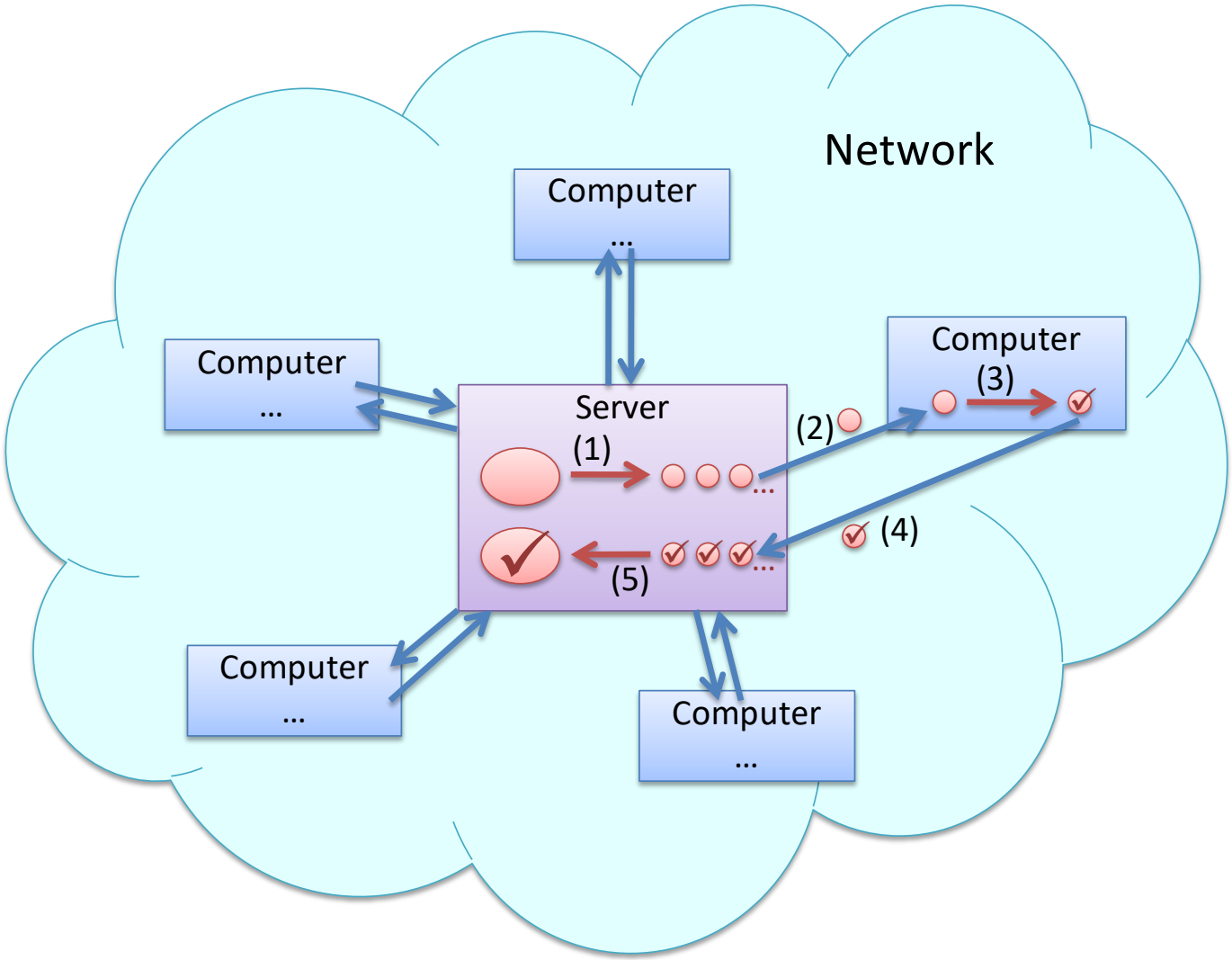
- A NASA Project
  - Commodity hardware
  - Private system network
  - Linux

grid computing

Grid Computing

- Individual computers are serving the users, and also playing a role in a distributed system.
  - Task its associated computers for executing in spare CPU time.
  - Improve computing resource utilization
  - Perform useful tasks
- Computer power is rarely utilized in many computers

Mechanism of Grid Computing



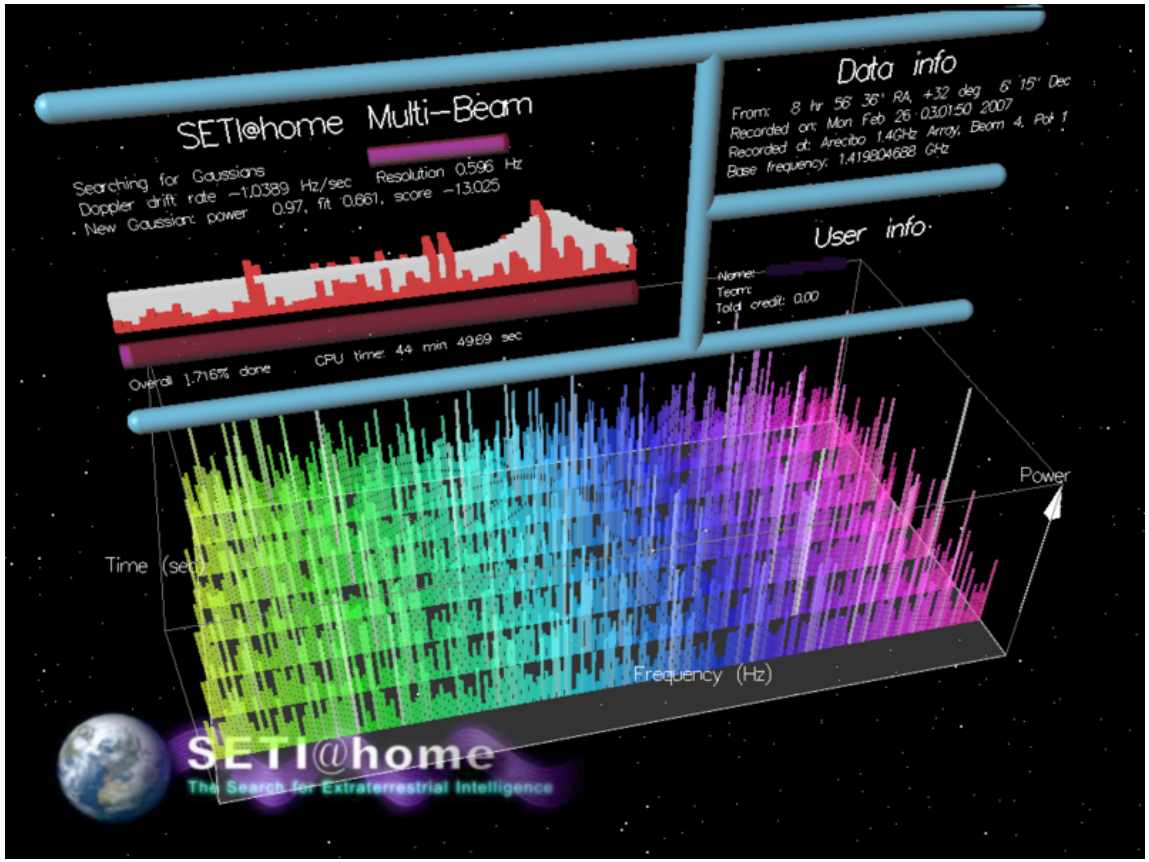
Key Features of Grid Computing

- Computers in grid systems are often heterogeneous, scattered, and owned by different parties.
  - Similar set of properties compared to clusters
  - Security concern

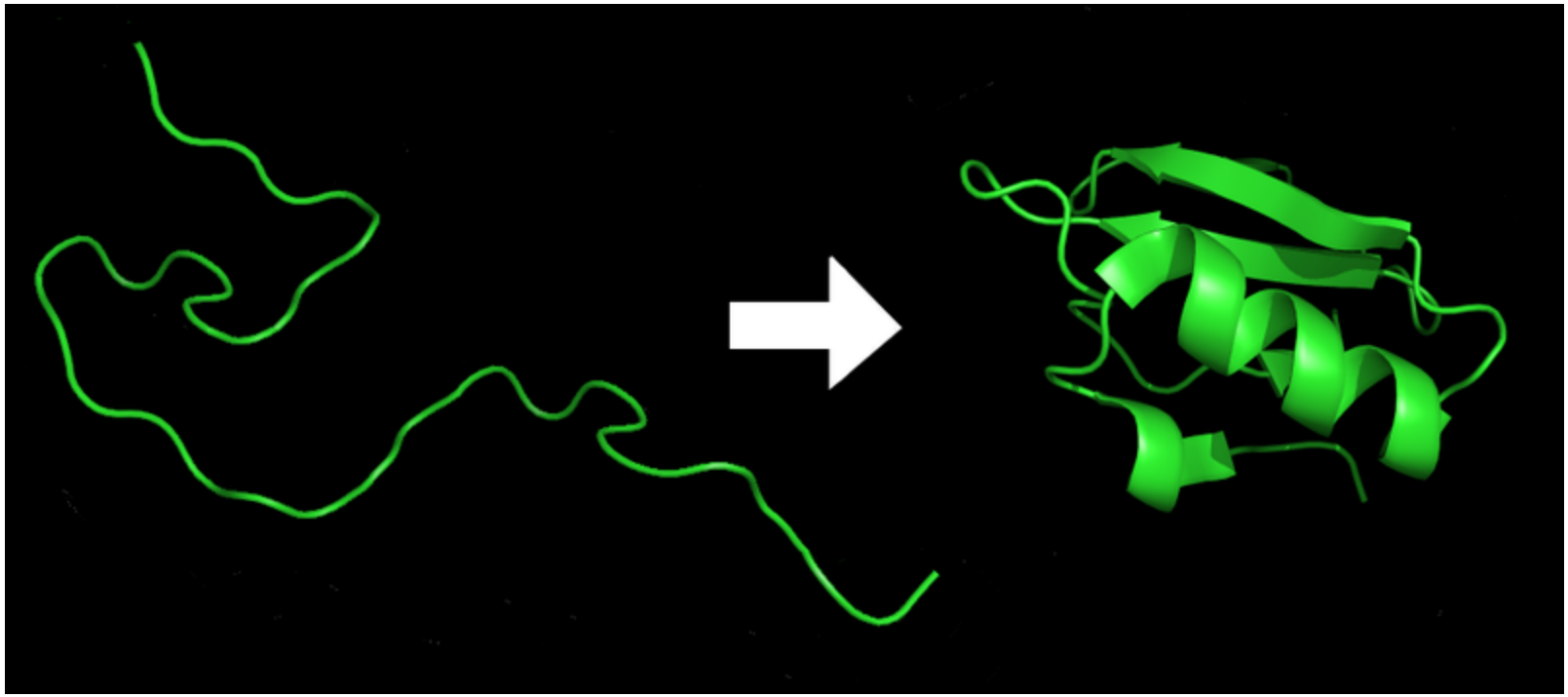
Examples of Grid Computing

- SETI: Search for Extraterrestrial Intelligence at Home
- Folding@home project

Examples of Grid Computing



Examples of Grid Computing



cloud computing

Cloud Computing

- Traditional view of hardware and software provisioning
  - Directly owned
- Cloud computing view
  - Rented or acquired in an ad hoc manner
  - Concept of utility computing: like electricity and water

Benefits of Cloud Computing

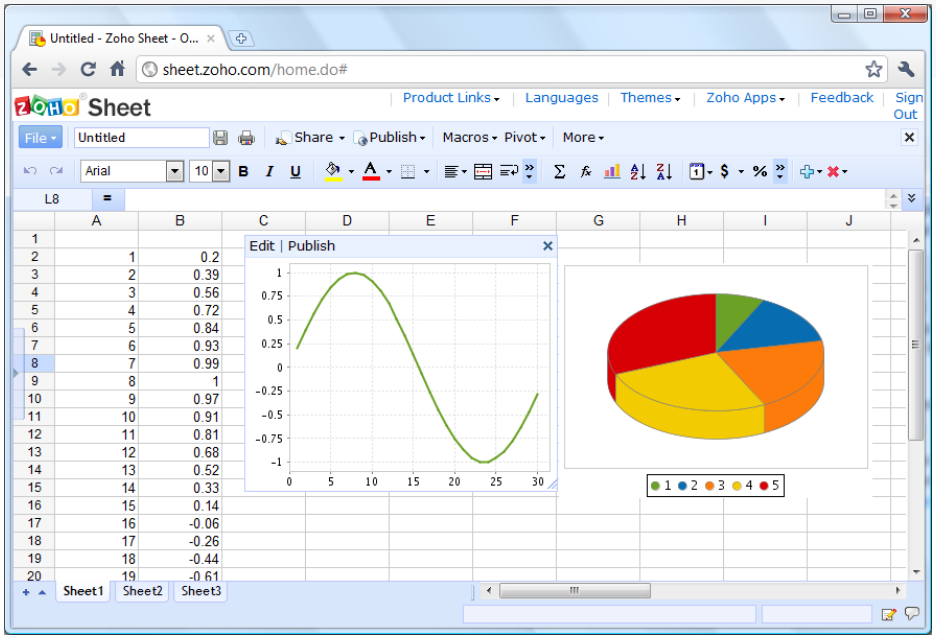
- Economy of scale and leading to overall financial benefits
- Servers for cloud computing must be highly powerful and reliable.
  - Often mainframe computers

Types of Cloud Computing

- Application level
  - Software as a Service (SaaS)
- Platform level
  - Platform as a Service (PaaS)
- Infrastructure Level
  - Infrastructure as a Service (IaaS)

Application Level

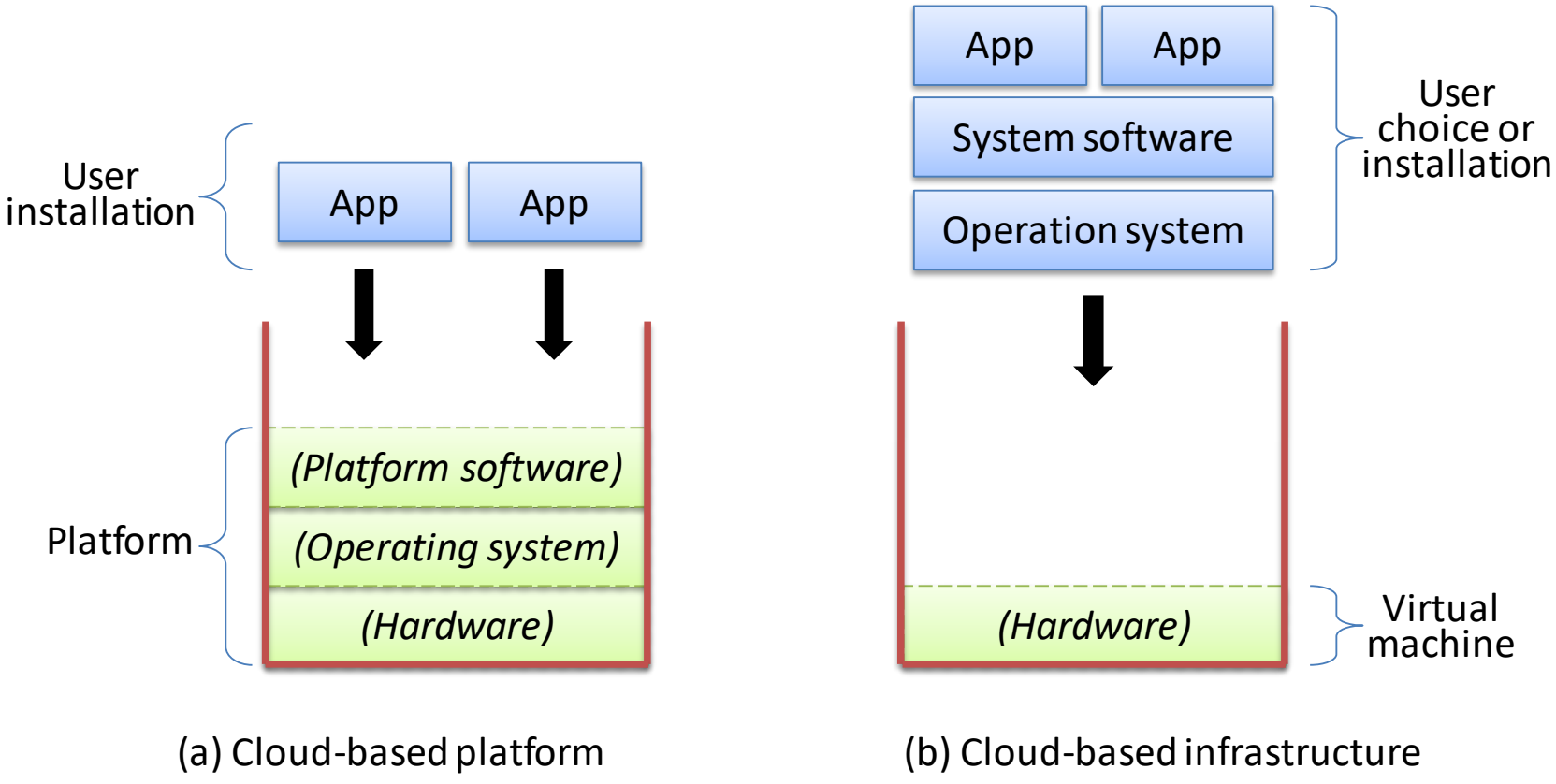
- Provision of application software via the Web
- Examples:
  - Google Docs
  - Microsoft Office Web Apps



Platform Level

- Software environments for running
  - Online application is running in an environment.
- Example:
  - Google App Engine

Platform Level and Infrastructure Level



(a) Cloud-based platform

(b) Cloud-based infrastructure

Infrastructure Level

- Virtual computers or virtual machines to the users
- Examples:
  - Amazon EC2

Benefits of Cloud Computing

- Cost saving
  - Charged on usage or even free
  - Low start-up cost
- Maintenance
  - Carried out by cloud service providers
- Scalability
  - Managed by the cloud service providers
- Universal access
  - Assessed through the Internet and data is stored with providers

Benefits of Cloud Computing

- Automatic software update
- Sharing and collaboration
  - Documents and other resources
- Reliability for hosted applications
  - Managed by cloud service provider

Concerns of Cloud Computing

- Online requirement
- Reliability of cloud-based applications
  - Comparatively speaking
  - Should be more reliable than in-house
- Security
  - Confidential information
  - Data transfer

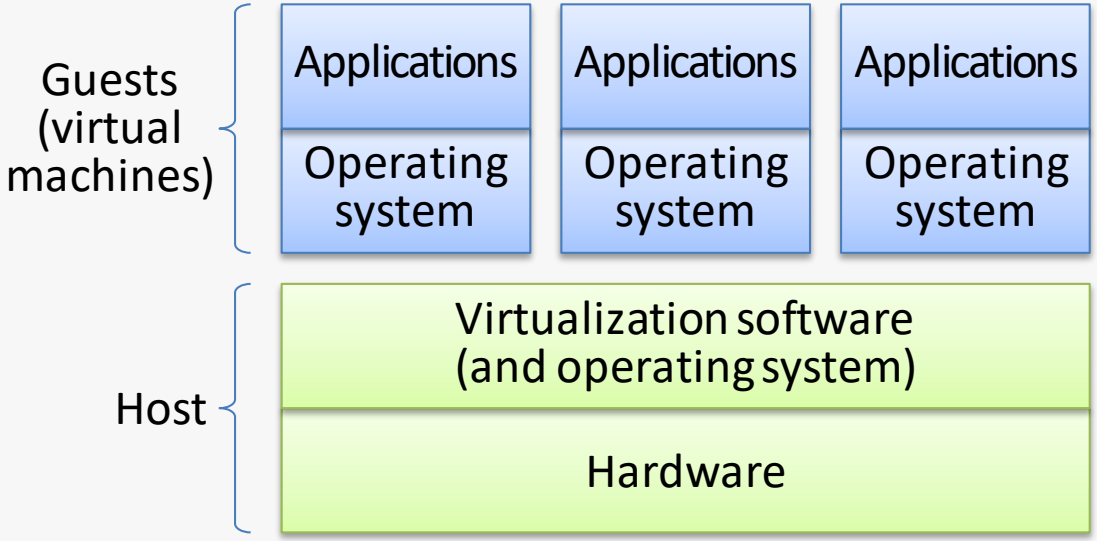
virtualization

Virtualization

- A common concept in computing
  - A logical view is provided to the users irrespective of the underlying real and physical system
  - An illusion

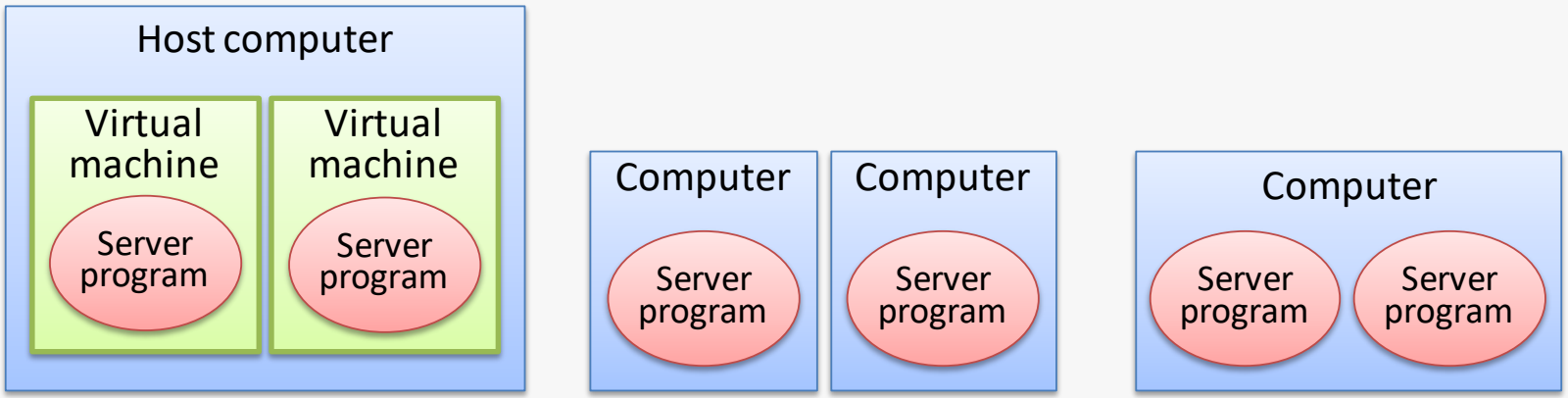
Platform Virtualization

- Machine virtualization
  - Virtual machines from a physical system (the host)
  - Operate independently of each other
  - Users see the virtual machines as individual computers



Platform Virtualization

- Server virtualization
  - One application can be set up in each virtual machine
  - Save hardware and operating costs
  - Improve reliability by avoiding interference.
  - Widely used in data centers and web hosting



(a) Server virtualization (b) Using two computers (c) Using one computer

Platform Virtualization

- Desktop virtualization
  - Multiple OS simultaneously

Software Virtualization

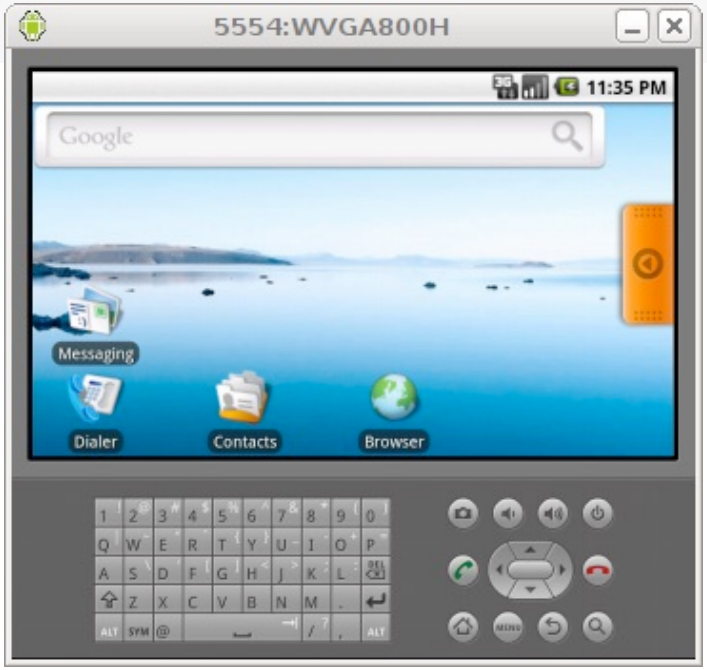
- Virtual software environments for executing applications
  - Operating system virtualization: guest root account
  - Kernel-level virtualization: guest kernel
  - Application virtualization: sandbox
  - Library emulation: system calls
    - Running Windows in Linux systems

Benefits of Platform Virtualization

- Cost savings of hardware and software
- Reliability
  - No interference between virtual machines
- Improved utilization
- Disaster recovery
- Ease of development
- Ease of deployment

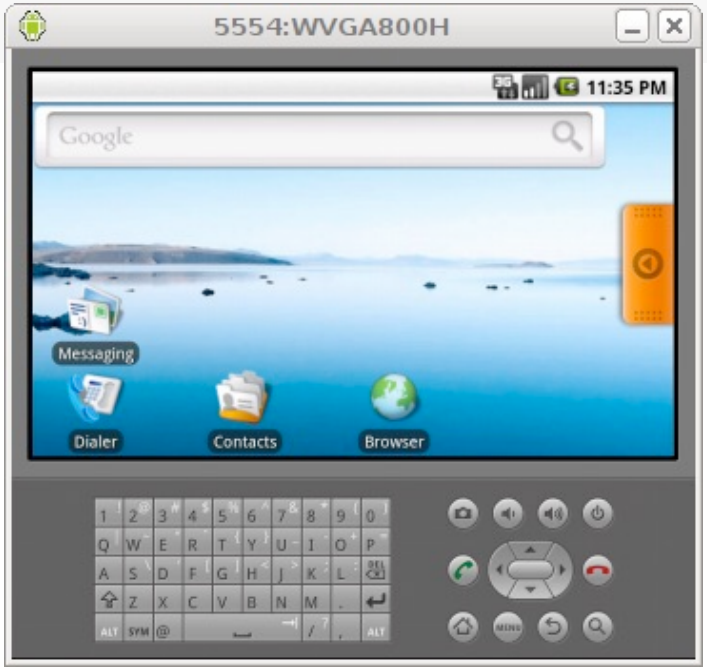
Virtual Machine Technologies

- Emulation: a software application that pretends to be a processor
- Example: Android emulator



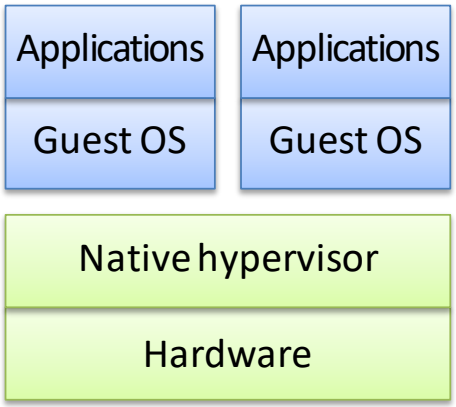
Virtual Machine Technologies: Emulation

- Emulation: a software application that pretends to be a processor
- Example: Android emulator

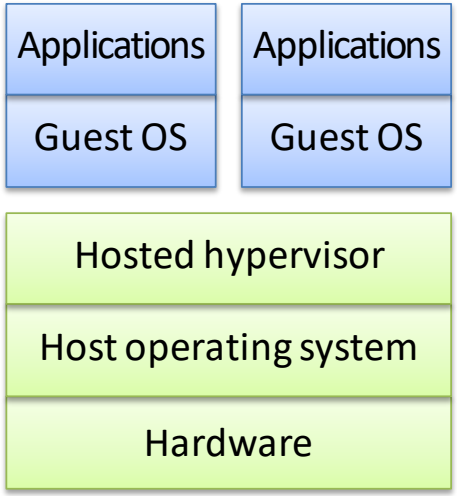


Virtual Machine Technologies: Hypervisors

- A hypervisor is a virtualization software layer
  - Interfaces with guest systems on one side
  - Computer hardware or host operating system on the other side.
  - Most instructions of guest systems be performed natively therefore more efficiently



(a) Native hypervisor



(b) Hosted hypervisor

Virtual Machine Technologies: Paravirtualization

- Paravirtualization is based on hypervisors
  - Guest systems running modified operating systems
  - The modifications change privileged (kernel-mode) instructions to special calls to the hypervisor.
  - May be faster