CAT3053/N Distributed Computing
Week 12

Grid Computing

What is Grid Computing?

- Computational Grids
  - Homogeneous (e.g., Clusters)
  - Heterogeneous (e.g., with one-of-a-kind instruments)
- Cousins of Grid Computing
- Methods of Grid Computing

Computational Grids

- A network of geographically distributed resources including computers, peripherals, switches, instruments, and data.
- Each user should have a single login account to access all resources.
- Resources may be owned by diverse organizations.

Computational Grids

- Grids are typically managed by gridware.
- Gridware can be viewed as a special type of middleware that enable sharing and manage grid components based on user requirements and resource attributes (e.g., capacity, performance, availability…)

Cousins of Grid Computing

- Parallel Computing
- Distributed Computing
- Peer-to-Peer Computing
- Many others: Cluster Computing, Network Computing, Client/Server Computing, Internet Computing, etc...

Distributed Computing

- People often ask: Is Grid Computing a fancy new name for the concept of distributed computing?
- In general, the answer is "no." Distributed Computing is most often concerned with distributing the load of a program across two or more processes.

PEER2PEER Computing

- Sharing of computer resources and services by direct exchange between systems.
- Computers can act as clients or servers depending on what role is most efficient for the network.

Methods of Grid Computing

- Distributed Supercomputing
- High-Throughput Computing
- On-Demand Computing
- Data-Intensive Computing
- Collaborative Computing
- Logistical Networking

Distributed Supercomputing

- Combining multiple high-capacity resources on a computational grid into a single, virtual distributed supercomputer.
- Tackle problems that cannot be solved on a single system.

High-Throughput Computing

- Uses the grid to schedule large numbers of loosely coupled or independent tasks, with the goal of putting unused processor cycles to work.

On-Demand Computing

- Uses grid capabilities to meet short-term requirements for resources that are not locally accessible.
- Models real-time computing demands.

Data-Intensive Computing

- The focus is on synthesizing new information from data that is maintained in geographically distributed repositories, digital libraries, and databases.
- Particularly useful for distributed data mining.

Collaborative Computing

- Concerned primarily with enabling and enhancing human-to-human interactions.
- Applications are often structured in terms of a virtual shared space.

Logistical Networking

- Global scheduling and optimization of data movement.
- Contrasts with traditional networking, which does not explicitly model storage resources in the network.
- Called "logistical" because of the analogy it bears with the systems of warehouses, depots, and distribution channels.

Who Needs Grid Computing?

- A chemist may utilize hundreds of processors to screen thousands of compounds per hour.
- Teams of engineers worldwide pool resources to analyze terabytes of structural data.
- Meteorologists seek to visualize and analyze petabytes of climate data with enormous computational demands.

An Illustrative Example

- Tiffany Moisan, a NASA research scientist, collected microbiological samples in the tidewaters around Wallops Island, Virginia.
- She needed the high-performance microscope located at the National Center for Microscopy and Imaging Research (NCMIR), University of California, San Diego.

Example (continued)

- She sent the samples to San Diego and used NPACI's Telescience Grid and NASA's Information Power Grid (IPG) to view and control the output of the microscope from her desk on Wallops Island. Thus, in addition to viewing the samples, she could move the platform holding them and make adjustments to the microscope.

Example (continued)

- The microscope produced a huge dataset of images.
- This dataset was stored using a storage resource broker on NASA's Information Power Grid.
- Moisan was able to run algorithms on this very dataset while watching the results in real time.

Grid Users

- Grid developers
- Tool developers
- Application developers
- End Users
- System Administrators

Grid Developers

- Very small group.
- Implementers of a grid “protocol” who provides the basic services required to construct a grid.

Tool Developers

- Implement the programming models used by application developers.
- Implement basic services similar to conventional computing services:
  - User authentication/authorization
  - Process management
  - Data access and communication

Tool Developers

- Also implement new (grid) services such as:
  - Resource locations
  - Fault detection
  - Security
  - Electronic payment

Application Developers

- Construct grid-enabled applications for end-users who should be able to use these applications without concern for the underlying grid.
- Provide programming models that are appropriate for grid environments and services that programmers can rely on when developing (higher-level) applications.

System Administrators

- Balance local and global concerns.
- Manage grid components and infrastructure.
- Some tasks still not well delineated due to the high degree of sharing required.

Some Highly-Visible Grids

- The NSF PACI/NCSA Alliance Grid.
- The NSF PACI/SDSC NPACI Grid.
- The NASA Information Power Grid (IPG).
- The Distributed Terascale Facility (DTF) Project.

Globus

- A collaboration of Argonne National Laboratory’s Mathematics and Computer Science Division, the University of Southern California’s Information Sciences Institute, and the University of Chicago's Distributed Systems Laboratory.
- Started in 1996 and is gaining popularity year after year.

Globus

- A project to develop the underlying technologies needed for the construction of computational grids.
- Focuses on execution environments for integrating widely-distributed computational platforms, data resources, displays, special instruments and so forth.

The Globus Toolkit

- The Globus Resource Allocation Manager (GRAM)
  - Creates, monitors, and manages services.
  - Maps requests to local schedulers and computers.
- The Grid Security Infrastructure (GSI)
  - Provides authentication services.

The Globus Toolkit

- The Monitoring and Discovery Service (MDS)
  - Provides information about system status, including server configurations, network status, and locations of replicated datasets, etc.
- Nexus and globus_io
  - provides communication services for heterogeneous environments.

The Globus Toolkit

- Global Access to Secondary Storage (GASS)
  - Provides data movement and access mechanisms that enable remote programs to manipulate local data.
- Heartbeat Monitor (HBM)
  - Used by both system administrators and ordinary users to detect failure of system components or processes.

Using the Grid

- Globus
- Condor
- Harness
- Legion
- IBP
- NetSolve
- Others

Grid Security

- Algorithm complexity theory
  - Verifiability
  - Concealment
- Cryptography and checkpointing
  - Corroboration
  - Scalability
- Voting and spot-checking
  - Fault tolerance
  - Reliability

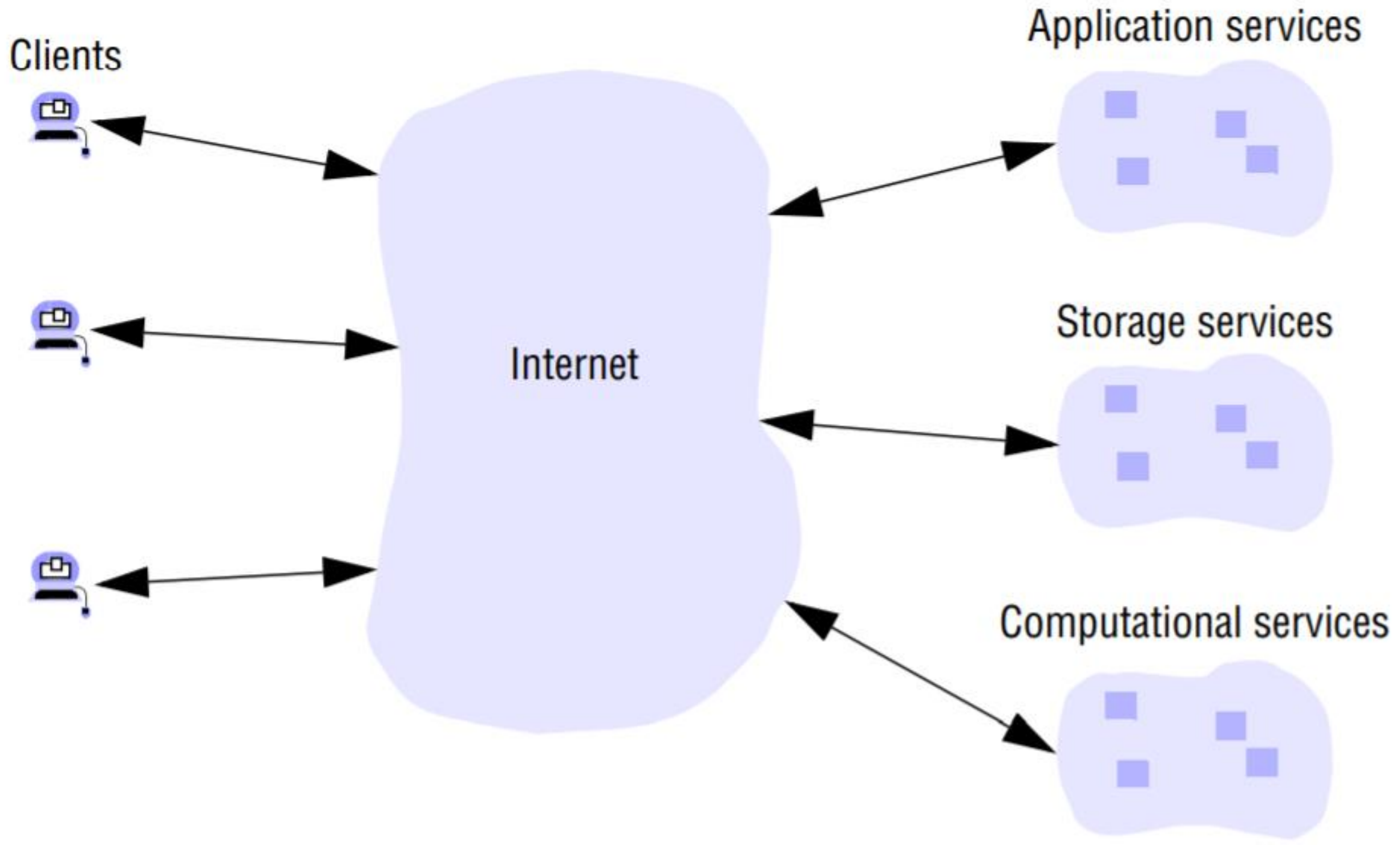
Some General Issues

- Grid architecture.
- Resource management.
- QoS mechanisms.
- Performance monitoring.
- Fault tolerance.

Cloud Computing

- Distributed computing is the use of distributed systems to solve single large problems by distributing tasks to single computers in the distributing systems.
- On the other hand, cloud computing is the use of network hosted servers to do several tasks like storage, process and management of data.

Cloud Computing



Cloud Computing

- Cloud computing has taken over the IT industry in the recent past.
- This is due to the fact that it is cheaper and easier to get services from the cloud.
- The cloud enables its users to choose how they will get and deliver IT services.
- Cloud computing means you can store and access data from the internet rather than the traditional computer hard disk storage.

Cloud Computing (cont.)

- This means you can access the data you have stored in the cloud anywhere anytime. The cloud will help you access the storage, servers, databases and multiple application services all in one place, the internet.

Benefits of Cloud Computing

- **Cost effective**
- The cloud helps you pay for the services you only require. Unlike building servers and databases, which are extremely expensive to build and maintain, the cloud helps you cut down that cost since you will pay for only what you are using.

Benefits of Cloud Computing

- **Economies of Scale**
- By using the cloud, you will gain immensely from benefits of economies of scale. Simply put, you will get more value for money when using cloud rather than going solo.

Benefits of Cloud Computing

- **Access to the global market**
- When using the cloud, you will have the chance of going global with a few clicks. You can reach the global audience without spending lots of cash and that's not all, your customers will get superior services thanks to the cloud

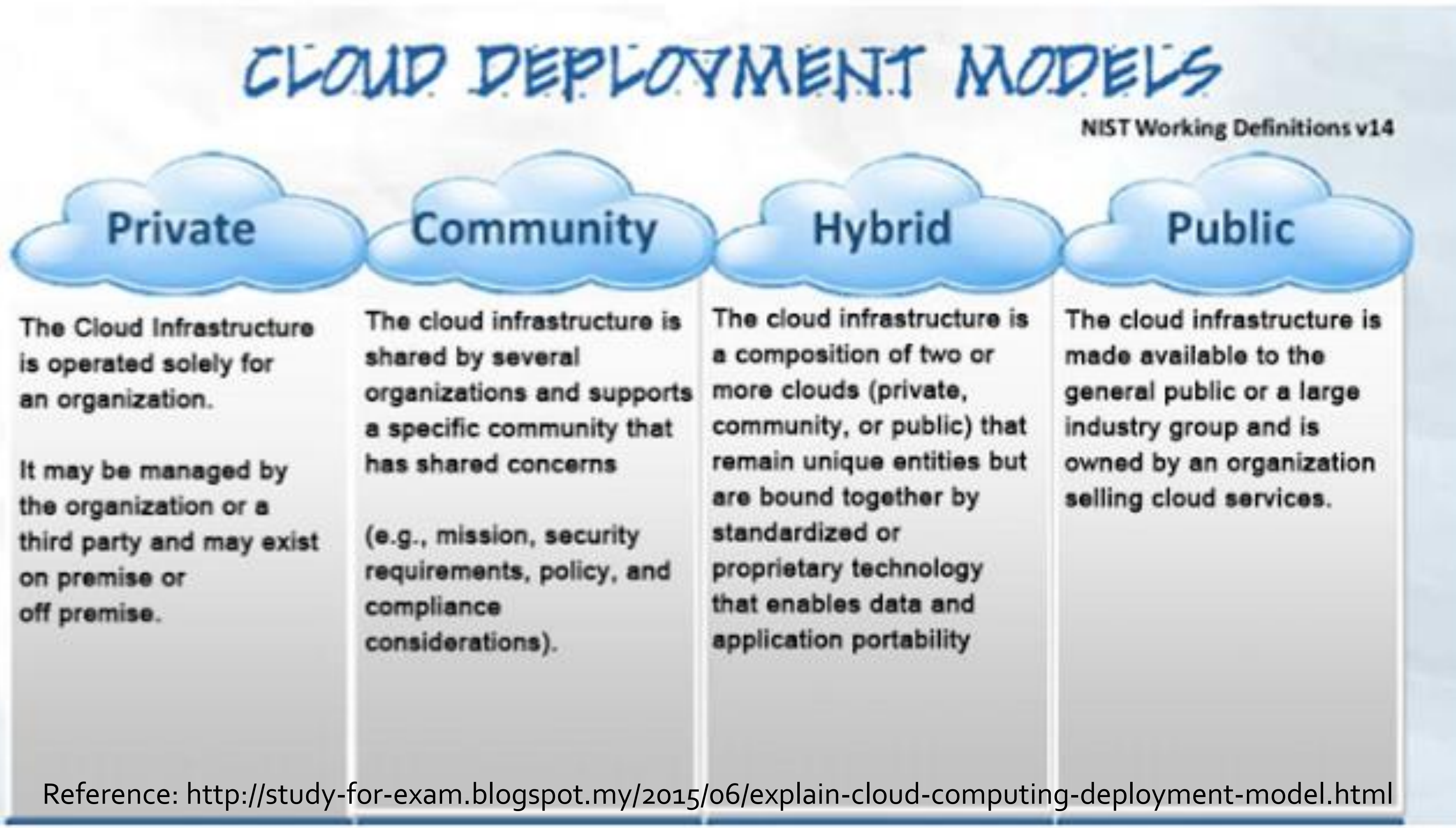
Service models

- Software as a service (SaaS)
  - Designed for end-users, delivered over the web.
- Platform as a service (PaaS)
  - Designed to make coding and deploying those applications quick and efficient.
- Infrastructure as a service (IaaS)
  - Hardware and software solutions –servers, storage, networks, operating systems.

Amazon Web Services

Web service	*Description*
Amazon Elastic Compute Cloud (EC2)	Web-based service offering access to virtual machines of a given performance and storage capacity
Amazon Simple Storage Service (S3)	Web-based storage service for unstructured data
Amazon Simple DB	Web-based storage service for querying structured data
Amazon Simple Queue Service (SQS)	Hosted service supporting message queuing (as discussed in Chapter 6)
Amazon Elastic MapReduce	Web-based service for distributed computation using the MapReduce model (introduced in Chapter 21)
Amazon Flexible Payments Service (FPS)	Web-based service supporting electronic payments

Deployment models



Reference: http://study-for-exam.blogspot.my/2015/06/explain-cloud-computing-deployment-model.html

Integration of Clouds and the IoT

- Combining clouds and the IoT
  - To support required resources to increase heterogamous objects
  - To meet the dynamic computational needs of environmental applications with existing sensor
- Benefits
  - The cloud can work on behalf of the object for increasing availability, maintaining performance and scalability
  - The cloud can support resource continuity so that objects move freely changing access technologies while using resources from the same cloud

Key features of clouds to support the IoT

- On-demand self service
  - Accessed without any special assistance
- Intelligent applications
- Broad network access
  - Several connectivity options (tablets, mobile devices, laptops)
- Resource pooling
  - Information can be shared with those who know where and how to access the resource, anytime and anywhere
- Rapid elasticity
  - Scalability and agility
- Virtualization
- Security

Conclusion

- Both distributed and cloud computing use the same concept but individually they are two distinct things. As a business you can use both to improve your business and in return yield higher profits.
- Some of the examples of distributed computing are Facebook, World Wide Web and ATM.
- Examples of cloud computing are YouTube, Google Docs and Picasa.

GRID COMPUTING

END