Cloud Computing

Tentative List of Topics

- **Introduction to Cloud Computing and its Enabling Technologies :** Evolution of Computing, Grid and Utility Computing, The vision of Cloud Computing, Characteristics and Benefits, Distributed Computing, Virtualization, Web 2.0, Service Oriented Architecture
- **Cloud Computing Architecture** : Introduction, Cloud Service Models, Cloud Deployment Models, Open Challenges
- **Virtualization:** Introduction, Characteristics of virtualized environments, Taxonomy of virtualization techniques, Virtualization and cloud computing, Pros and cons of virtualization, Technology examples, Containers and Applications
- **Cloud Platforms:** Amazon Web Services, Google Cloud Platform, Microsoft Azure, Aneka, OpenStack, Cloud Automation using CHEF/Ansible
- **Introduction to Bigdata:** Bigdata concepts, terminology, NoSQL; Distributed File Systems-Hadoop File System, GFS, Introduction to MapReduce and applications
- **Cloud Security:** Security Issues in Cloud Computing, Hypervisor and VM Security, Data Security in Cloud Environment, Identity and Access Management in Cloud

Course Outcomes

- To describe the different cloud computing models and underlying technologies.
- To develop real world applications using cloud computing platforms and containerization technologies.
- To implement solutions to complex problems using distributed computing technologies.
- To identify and analyze security issues in cloud computing.

Tentative Evaluation Policy

- **Examinations: 70%**
  - Mid Exam: 20%
  - End Exam: 50%
- **Project: 20%**
- **Scheduled Quiz: 10%**

Module 1

Introduction to Cloud Computing and its Enabling Technologies

What is Computing?

- **Computing** consists of three things:
  - *Managing,*
  - *Processing, and*
  - *Communicating information*

- Over the years, computing has evolved – passing through a number of computational paradigms
  - More computational capacity
  - More efficiency
  - Less cost
  - Less size

Vacuum Tube

Transistors

Integrated Circuit

Microprocessor

Quantum Computer

1st Generation Computer

2nd Generation Computer

3rd Generation Computer

4th Generation Computer

5th Generation Computer

The Evolution of Commercial Computing

Centralized Era

first, large and expensive Computers

(since 1940s)

Microprocessors & Mass Production

Personal Computers

(since 1970s)

Miniaturization & Mobility needs

Notebook Computers

(since 1980s)

Timesharing Operating Systems & Local Area Networks

dedicated, on-premise Data Centers

(since 1980s)

Internet

Decentralized Era

Public Data Centers

(since late 1990s)

Service-oriented Architecture

Public Cloud Computing

(since early 2000s)

Blockchain

Distributed Computing (with limited use cases)

(in development since 2016)

Artificial Intelligence

Distributed Era

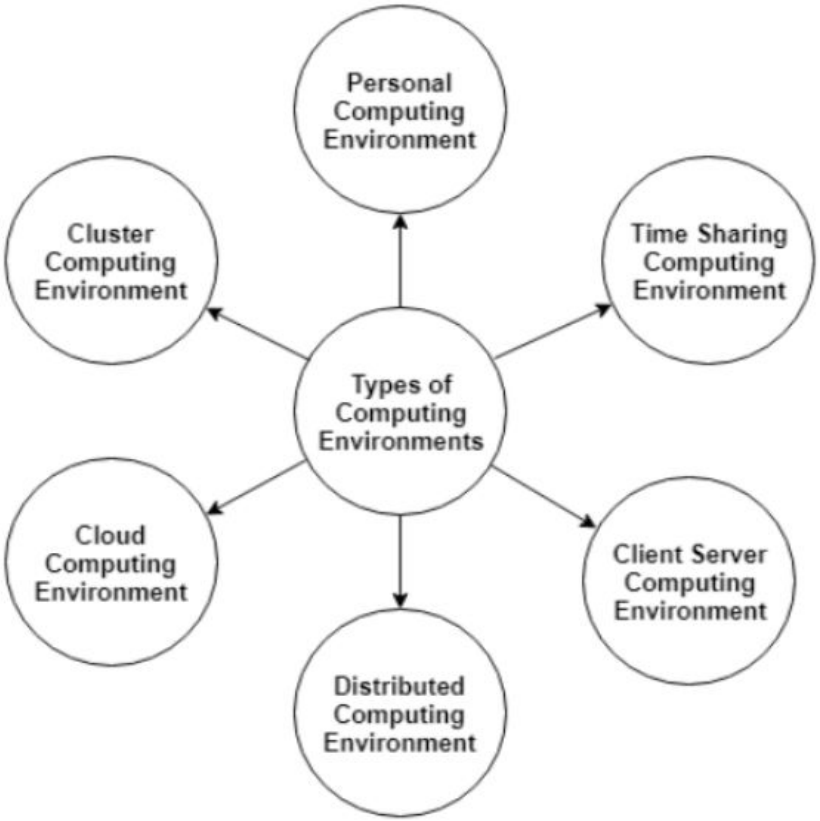
Distributed Computing (with flexible use cases)

(in development since 2018)

HiveNet

Types of Computing Environments

- A computer system uses many devices, ***arranged in different ways*** to solve many problems. This constitutes a computing environment where many computers are used to process and exchange information to handle multiple issues.



Personal Computing Environment

- In the personal computing environment, there is a single computer system. All the system processes are available on the computer and executed there.
- Smaller size machines – semiconductor technology.
- Designed for personal use.
- Every system now had dedicated computing power, memory and storage – available locally.
- The size of PCs has reduced considerably since their inception, along with an increase in computational power and storage capacity.

Time Sharing Computing Environment

- The time sharing computing environment allows multiple users to share the system simultaneously.
- Each user is provided a time slice and the processor switches rapidly among the users according to it.
- Because of this, each user believes that they are the only ones using the system.

Networked Computing

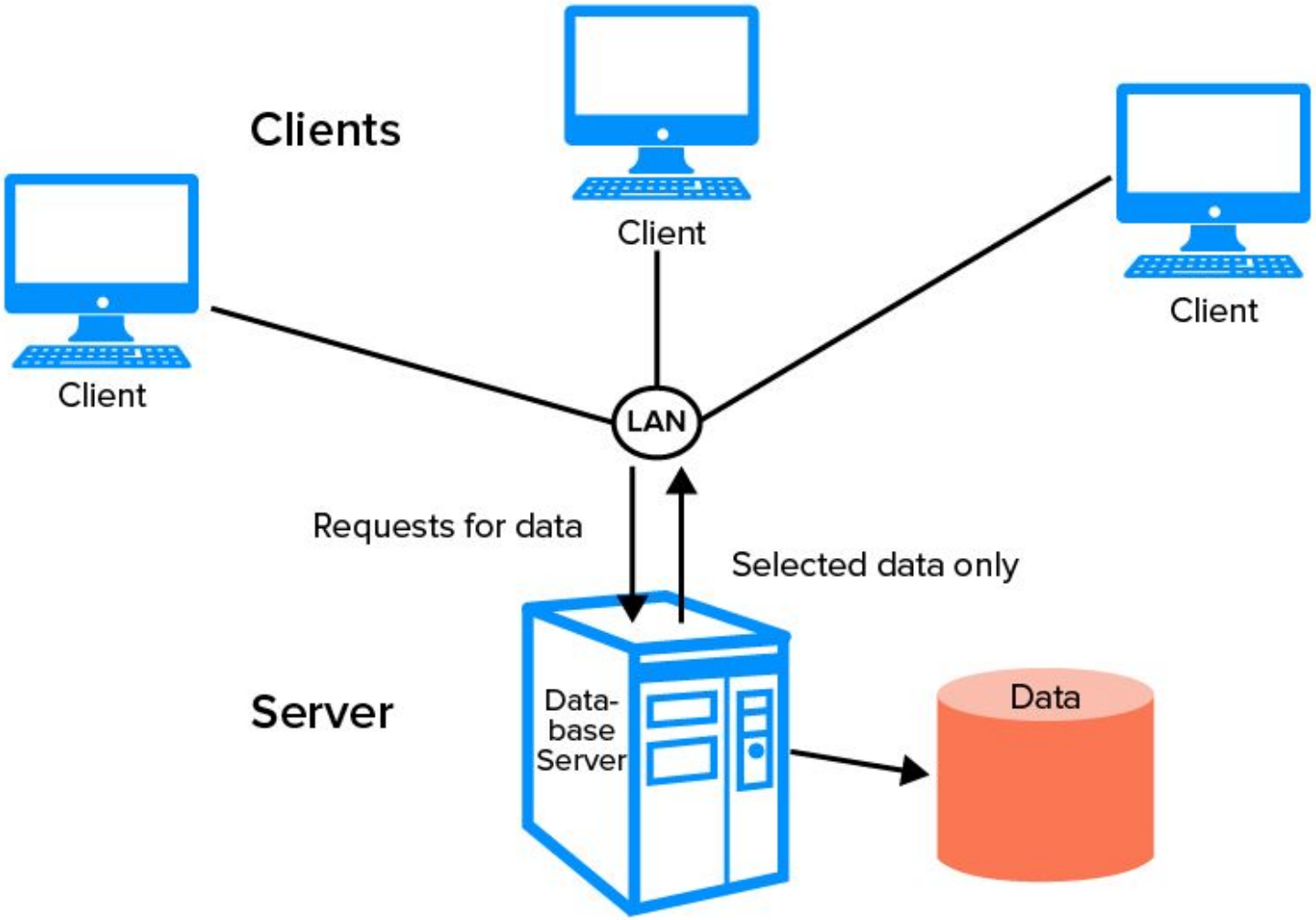
- Need for networked computing:
  - Better communication between systems
  - Resource sharing
- Systems and associated resources, such as printers, were interconnected to form local networks (LANs)
- These small networks evolved to form ARPANET, which finally evolved to the Internet.

ARPANET: It stands for Advanced Research Projects Agency Network. ARPANET was first network which consisted of distributed control. It was first to implement TCP/IP protocols.

Client Server Computing Environment

- In client server computing, the client requests a resource and the server provides that resource.
- A server may serve multiple clients at the same time while a client is in contact with only one server.
- Both the client and server usually communicate via a computer network but sometimes they may reside in the same system.

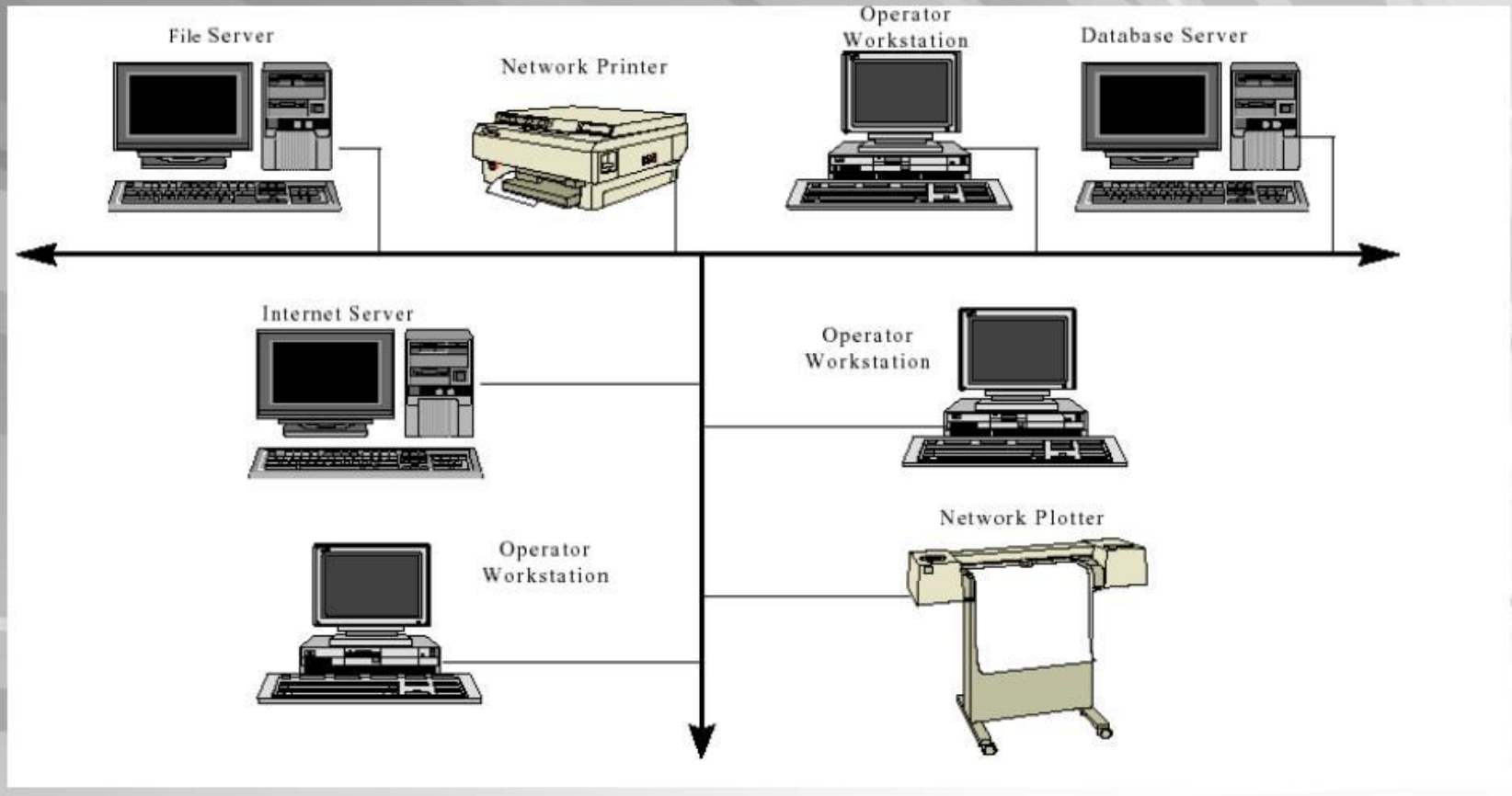
Client/ Server Architecture



Distributed Computing Environment

- A distributed computing environment contains multiple nodes that are physically separate but linked together using the network.
- All the nodes in this system communicate with each other and handle processes in tandem.
- It uses a centralized resource manager and all nodes cooperatively work together as a single unified resource
- Distributed system architectures are bundled up with components and connectors.
- Components can be individual nodes or important components in the architecture whereas connectors are the ones that connect each of these components.
- Component: A modular unit with well-defined interfaces; replaceable; reusable.

Distributed System Architecture



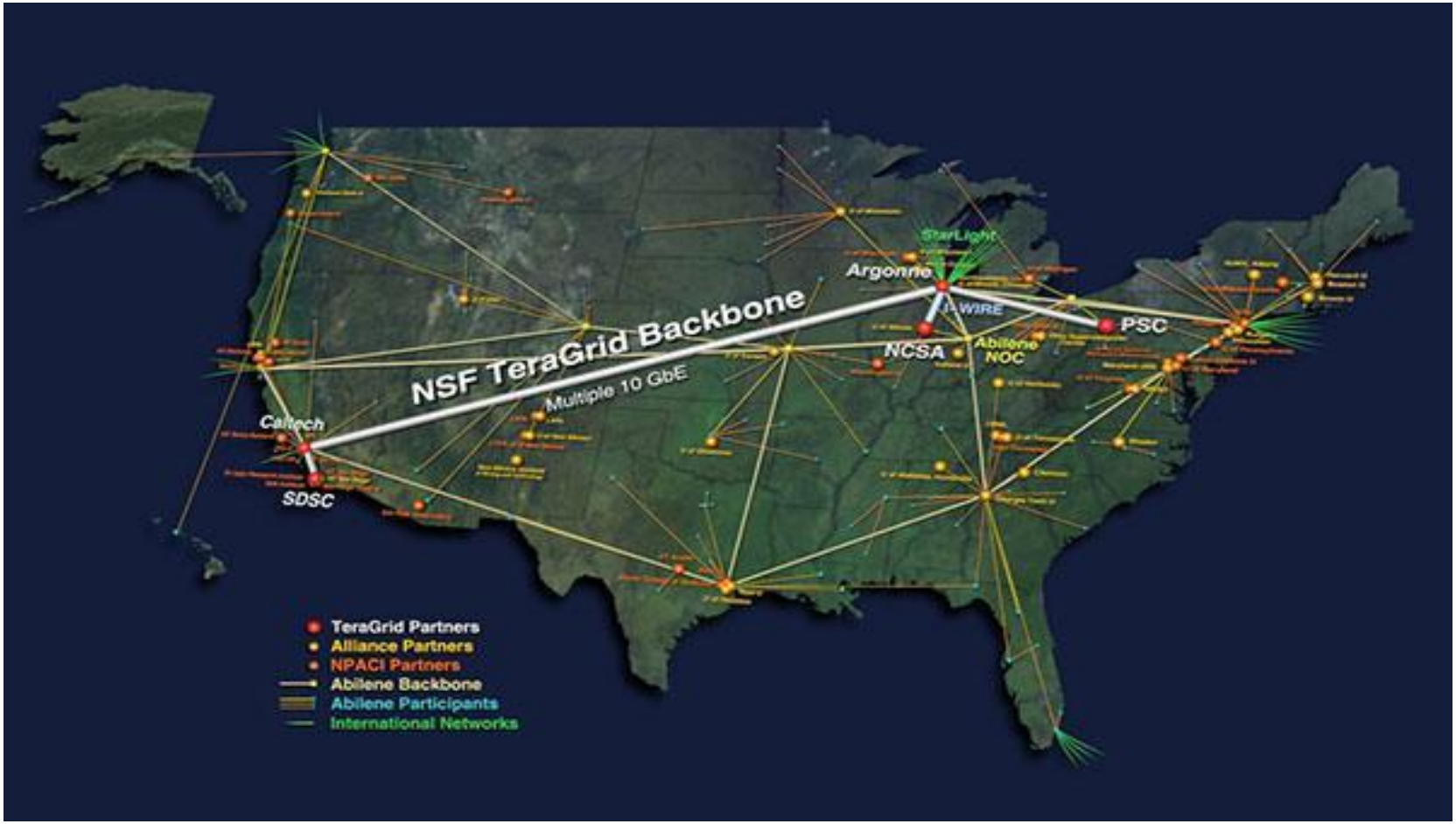
Grid Computing

- The volunteer computing projects are often compared to electric grids
  - *Large number of connected nodes which act like a single entity.*
  - *Users do not know which node serves their requests.*
- Multiple physically separated systems share data and resources for **performing a common task.**
- Often, the original **task is split and distributed** among different systems or nodes.
- Requires **special software** to be installed on the systems ("middleware").

TERAGRID

- e-Science grid computing **spread over 11 sites** across USA.
- Funded by the National Science Foundation (NSF) in the United States.
- **Petaflops of computing capability** and more than **30 petabytes of online and archival data storage** connected by **high speed optic fibre (10 Gbps).**
- Coordinated TeraGrid Software and Services (CTSS).
  - single-sign on
  - remote job submission
  - workflow support
  - distributed accounting and account management software
  - verification and validation software
  - set of compilers, programming tools etc

TERAGRID



TeraGrid users primarily came from U.S. universities. There are roughly 4,000 users at over 200 universities. Academic researchers in the United States can obtain exploratory, or development allocations (roughly, in "CPU hours") based on an abstract describing the work to be done.

Advantages of Grid Computing

- Improved resource utilization
- General performance increase – parallel processing
- Easier collaboration
- Increased robustness