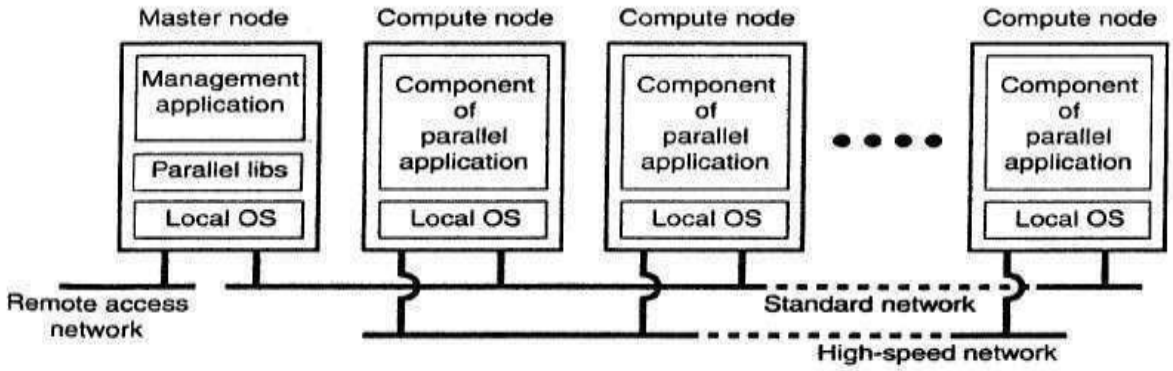
WEEK # 2 - TYPES OF DISTRIBUTED SYSTEMS

Distributed Computing Systems

- Clusters
- Grids
- Clouds

Cluster Computing

- Cluster computing the underlying hardware consists of a collection of similar workstations or PCs, closely connected by means of a high speed local-area network. In addition, each node runs the same operating system.
- Cluster computing systems became popular when the price/performance ratio of personal computers and workstations improved.
- At a certain point, it became financially and technically attractive to build a supercomputer using off-the-shelf technology by simply hooking up a collection of relatively simple computers in a high-speed network.
- Cluster computing is used for parallel programming in which a single (compute intensive) program is run in parallel on multiple machines.
- Well-known example of a cluster computer is formed by Linux-based **Beowulf clusters.**



An example of a cluster computing system.

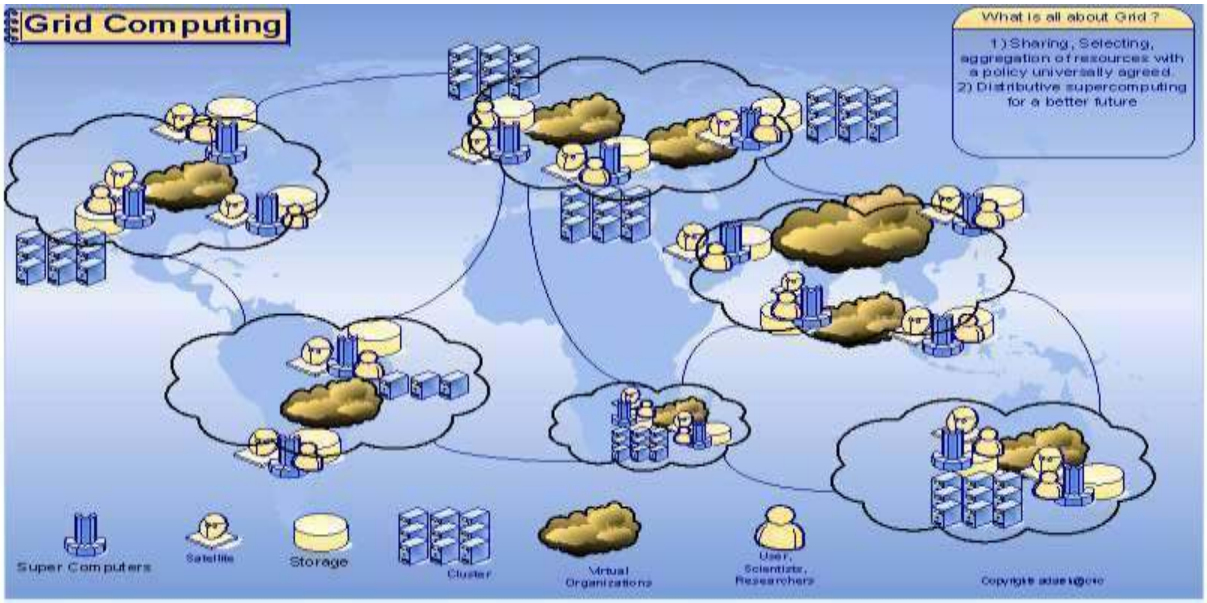
Cluster Types & Uses

- **High Performance Clusters (HPC)**
  - run large parallel programs
  - Scientific, military, engineering apps; e.g., weather modeling
- **Load Balancing Clusters**
  - Front end processor distributes incoming requests
  - server farms (e.g., at banks or popular web site)

- **High Availability Clusters (HA)**
  - Provide redundancy – backup systems
  - May be more fault tolerant than large mainframes

Grid Computing

- Grid computing consists of distributed systems that are often constructed as a federation of computer systems, where each system may fall under a different administrative domain, and may be very different when it comes to hardware, software, and deployed network technology.
- A key issue in a grid computing system is that resources from different organizations are brought together to allow the collaboration of a group of people or institutions.
- Such collaboration is realized in the form of a virtual organization. The people belonging to the same virtual organization have access rights to the resources that are provided to that organization.

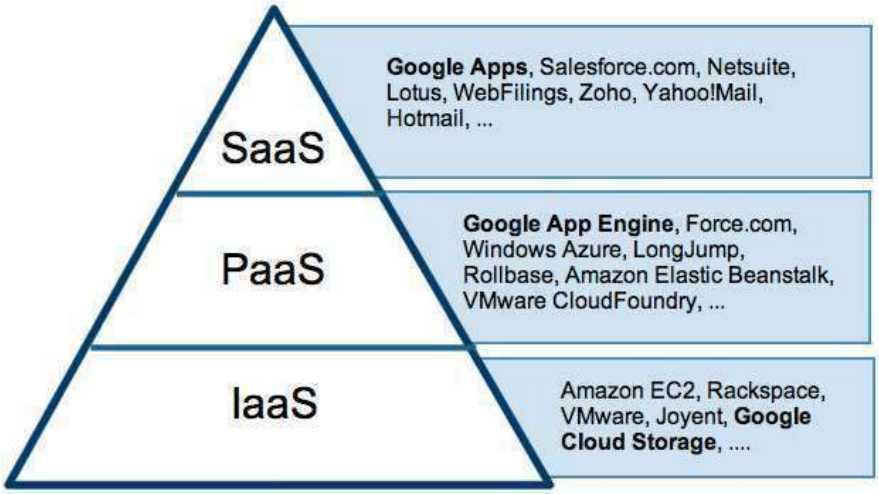


Cloud Computing

- Provides scalable services as a utility over the Internet.
- Cloud computing means you develop/run your software remotely on remote platform. This can be either using remote virtual infrastructure (amazon EC2), remote platform (google app engine), or remote application (force.com or gmail.com).

- Using Cloud Computing, companies can scale upto High capacities immediately without investing in new infrastructure, training the people or new software licensing. It is more useful for small and medium scale businesses who want to outsource their Data Center infrastructure, or some larger companies also prefer if they want to cut down the costs of building data-centers internally in order to get peak load capacity. In short, consumers use what they need and pay accordingly.

- Additional advantages for the consumer are, they no longer has to be at a computer to use the application. They can access it by using smart phones, PDA or which the medium Cloud supports. By adopting cloud, consumer does not need to own the infrastructure, software and also consumer doesn't need to worry about the network maintenance. By choosing this he can reduce the capital expenses, upfront costs and operating expenses.



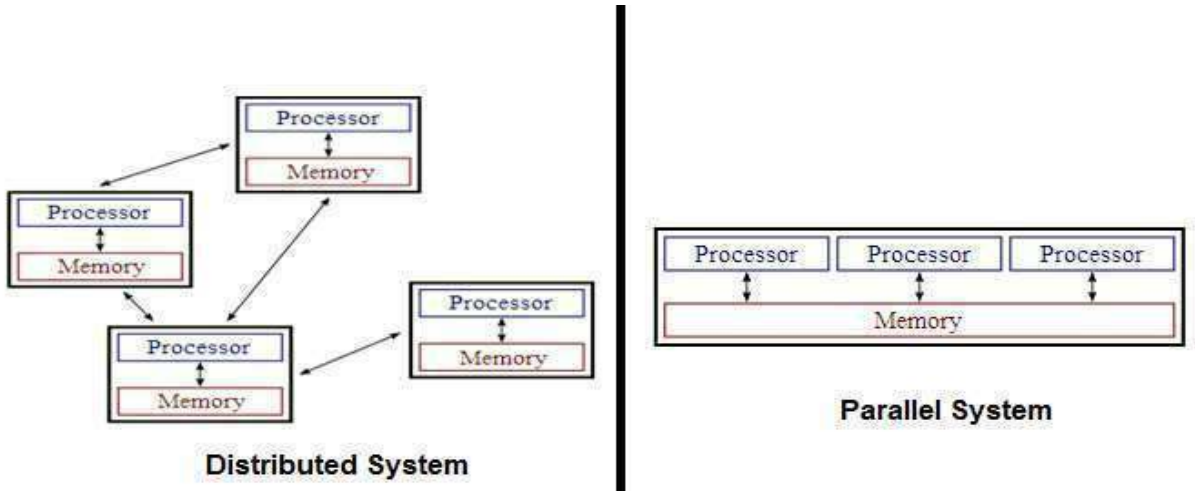
(SaaS=Software as a service, PaaS=Platform as a service, IaaS=Infrastructure as a Service)

Difference between Cluster Computing Systems and Grid Computing Systems

- A characteristic feature of cluster computing is its homogeneity. In most cases, the computers in a cluster are largely the same, they all have the same operating system, and are all connected through the same network.
- In contrast, grid computing systems have a high degree of heterogeneity: no assumptions are made concerning hardware, operating systems, networks, administrative domains, security policies, etc.

Difference between Parallel and Distributed Computing

- In parallel computing, all processors may have access to a shared memory to exchange information between processors.
- In distributed computing, each processor has its own private memory (distributed memory). Information is exchanged by passing messages between the processors.



Difference between Cloud and Grid computing

The difference between a cloud and a grid can be expressed as below:

1. Resource distribution: Cloud computing is a centralized model whereas grid computing is a decentralized model where the computation could occur over many administrative domains.

2. Ownership: A grid is a collection of computers which is owned by multiple parties in multiple locations and connected together so that users can share the combined power of resources. Whereas a cloud is a collection of computers usually owned by a single party.

Examples of Clouds: Amazon Web Services (AWS), Google App Engine, Dropbox, Gmail, Facebook, Youtube, Rapidshare, etc

Examples of Grids: FutureGrid, Berkeley's Open Infrastructure for Network Computing (BOINC)

WEEK # 2 - TYPES OF DISTRIBUTED SYSTEMS

Distributed Information Systems

- Business-oriented
- Systems to make a number of separate network applications interoperable and build "enterprise-wide information systems".
- Two types discussed here:
  - Transaction processing systems
  - Enterprise application integration (EAI)

Transaction Processing Systems

- Provide a highly structured client-server approach for database applications
- Transactions are the communication model
- Obey the ACID properties:
  - Atomic: all or nothing
  - Consistent: invariants are preserved
  - Isolated (serializable)
  - Durable: committed operations can't be undone

Primitive	Description
BEGIN_TRANSACTION	Mark the start of a transaction
END_TRANSACTION	Terminate the transaction and try to commit
ABORT_TRANSACTION	Kill the transaction and restore the old values
READ	Read data from a file, a table, or otherwise
WRITE	Write data to a file, a table, or otherwise

Example primitives for transactions

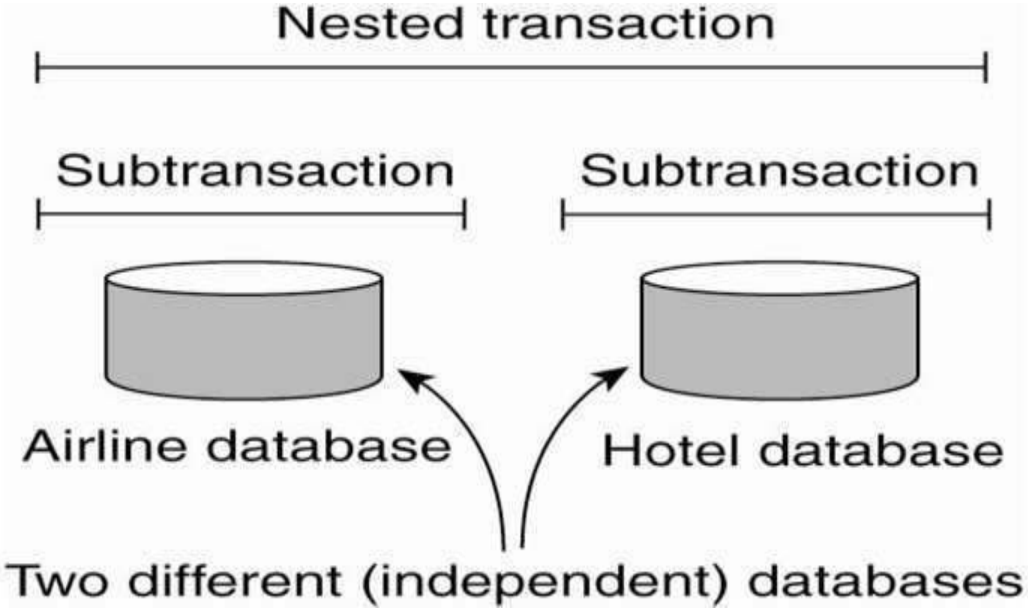
Transactions

- Transaction processing may be centralized (traditional client/server system) or distributed.
- A distributed database is one in which the *data storage* is distributed – connected to separate processors.

Nested Transactions

- A nested transaction is a transaction within another transaction (a sub-transaction)

  - Example: a transaction may ask for two things (e.g., airline reservation info + hotel info) which would spawn two nested transactions
- Primary transaction waits for the results.
  - While children are active parent may only abort, commit, or spawn other children

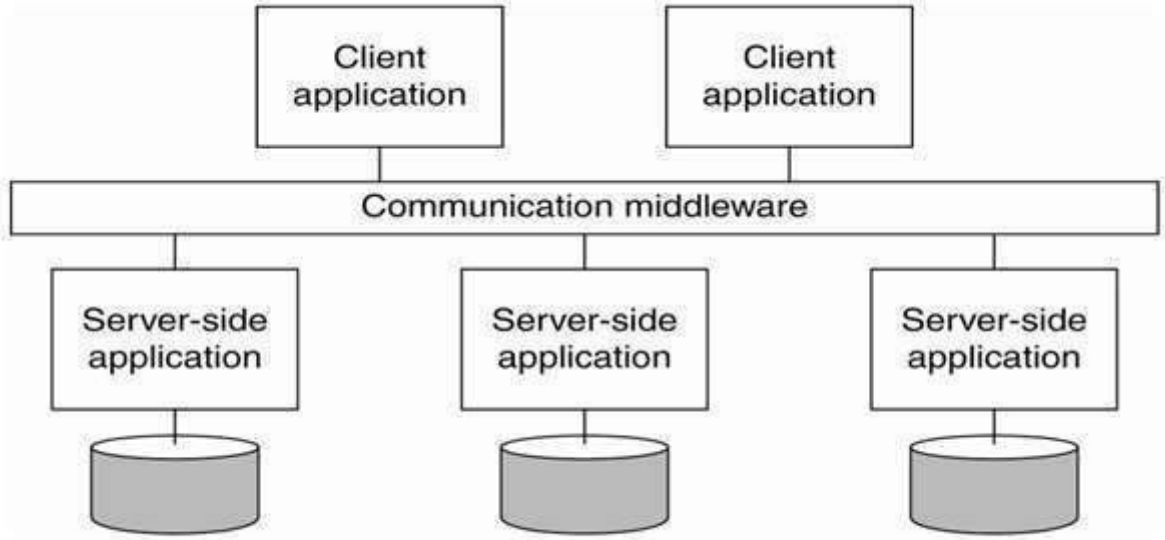


Implementing Transactions

- Conceptually, private copy of all data
- Actually, usually based on logs
- Multiple sub-transactions – commit, abort
  - Durability is a characteristic of top-level transactions only
- Nested transactions are suitable for distributed systems
  - Transaction processing monitor may interface between client and multiple data bases.

Enterprise Application Integration

- Less structured than transaction-based systems
- EA components communicate directly
  - Enterprise applications are things like HR data, inventory programs, …
  - May use different OSs, different DBs but need to interoperate sometimes.
- Communication mechanisms to support this include CORBA, Remote Procedure Call (RPC) and Remote Method Invocation (RMI)



Middleware as a communication facilitator in enterprise application integration.

LOOSELY-COUPLED SYSTEMS

- Each system was a completely autonomous independent system, connected to others on the network
  - FTP (rcp): file transfer program
  - telnet (rlogin/rsh): remote login program
  - mail (SMTP)
- Even today, most distributed systems are looselycoupled (although not that loosely!):
  - each CPU runs an independent autonomous OS
  - computers don't really trust each other
  - some resources are shared, but most are not
  - the system may look differently from different hosts

CLOSELY-COUPLED SYSTEMS

- A distributed system becomes more "closely-coupled" as it
  - appears more uniform in nature
  - runs a "single" operating system
  - has a single security domain
  - shares all logical resources (e.g., files)
  - shares all physical resources (CPUs, memory, disks, printers, etc.)

TIGHTLY-COUPLED SYSTEMS

- A "tightly-coupled" system usually refers to a multiprocessor
  - runs a single copy of the OS with a single workload queue
  - has a single address space

- usually has a single bus or backplane to which all processors and memories are connected
- has very low communication latency
- processors communicate through shared memory

	Cluster	Grids	Clouds
Nodes	Tightly-coupled	Loosely-coupled	Loosely-coupled
Business model	No	No	Yes
Allocation	Centralized	Decentralized	Both
SLA	Limited	Yes	Yes
Virtualization	Little	Little	Yes
Heterogeneity	No	Yes	Yes
Areas	Educational resources, Medical search etc.	Predictive modelling and simulation, energy resources exploration etc.	Banking, Insurance, Weather forecasting, Space exploration etc.
Services	Scientific simulation 3D modelling, DNA sequence analysis, Molecular nanotechnology	Intergrids, Intragrids, Extragrids	IaaS, PaaS, SaaS
Applications	Weather modelling, Life sciences, Protein explorer, Computational fluid dynamics, Nuclear simulations, Image processing, Data mining etc. and as Internet applications are Database servers, Data mining, Email, Proxy, Security etc.	NEESgrid Earthquake Engineering Clooaboratory, The virtual Observatory World-Wide Telescope etc.	Google Docs, Basecamp, Campfire, Xero, Springpad etc.
Projects	Beowulf, Bereley NOW, HPVM, Solaris MC, etc.	Tera grid, Globus project, NKN, Fusion Collaboratory, EGEE etc.	CERN, Unified cloud interface(UCI), OpenNebula, CESWP, TClouds

Comparison of Cluster / Grid / Cloud

To conclude, we can say that cluster computing is pillar architecture of cloud and grid. Many common names in the market already use the cloud like YouTube, Google etc.More and more we see private computing services in the domain of clouds. Cloud computing uses VPNs to provide services to the consumers. Some companies which provide these services and tools and required protocols are Amazon, Microsoft, Yahoo, IBM, Google etc. But in all three of the computing technologies, some issues like privacy, data safety and vendor lock-in are a concern. Most research identifies the only disadvantage of cloud computing as reduction in security