I. Distributed system:

1. Concept:
    - Distributed system is a collection of independent computers that appears to its users as a single coherent system.
    - Example: parallel computing: if you could design a program that run without serious modifications on distributed systems that appear to have the same performance as non-distributed systems
2. Role of middle ware in a distributed system:
    - To enhance the distribution transparency, improving the single system view
3. Distribution transparency:
    - The concept that distribution aspects of the system are hidden from users and applications
    - Transparency has an **inverse** relationship with performance
    - Including:
        - Access: hide differences in data representation and how a resource is accessed
        - Location: hide where resource is located
        - Migration: hide that resource might be moved to another location
        - Relocation: hide that resource might move to another location while in use
        - Replication: hide that a resource is replicated
        - Concurrency: hide that a resource may be used by several users
        - Failure: hide the failure and recovery of a resource
4. Replicating data advantages and disadvantages:

Availability

If one of sites containing relation A fails, then the relation A can be obtained from another sites. Thus, queries (involving relation A) can be continued to be processed in spite of the failure of one site.

Increased Parallelism

The sites containing relation A can process queries (involving relation A) in parallel. This leads to faster query execution.

Less Data Movement over Network

The more replicas of, a relation are there, the greater are the chances that the required data is found where the transaction is executing. Hence, data replicas reduce movement of data among sites and increase speed of processing.

Disadvantages

Increased overhead on update

When an update is required, a database system must ensure all replicas are updated.

Require more disk space

Storing replicas of same data at different sites consumes more disk space.

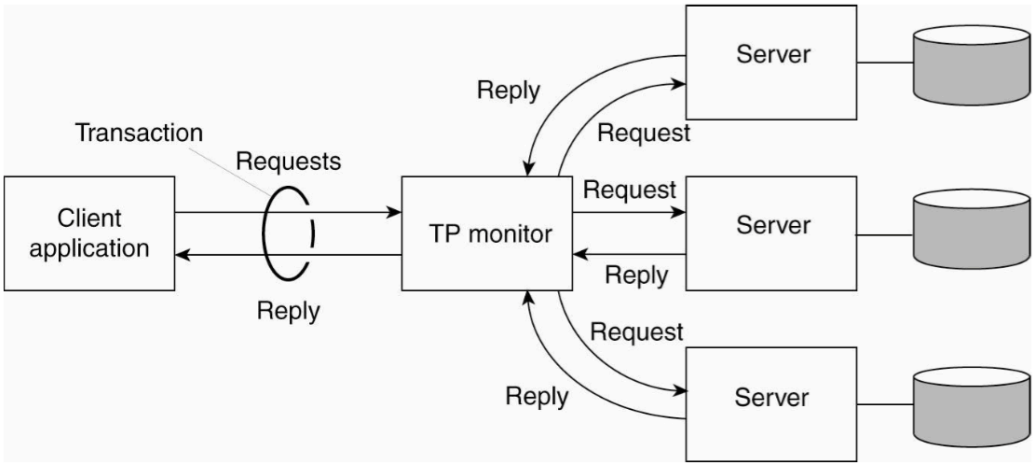
Expensive

Concurrency control and recovery techniques will be more advanced and hence more expensive.

5. Scalable system
    - A system is scalable with respect to either its number of components, geographical size or number and size of administrative domain if it can grow

in one or more of these dimensions without an unacceptable loss of performance
    - How to achieve scalable:
        - Replication: maintains multiple copies of data. It not only increases the availability of resources but also helps to balance the load between components. Besides, it also reduces the access time if the data is replicated on a nearby nodes of a distributed system
        - Hiding communication latencies: the requester can do some useful work while waiting for the response from the client side. When reply comes in, then the application will execute a special handler to complete the request. Another solution is to move part of the computation to the client
        - Distribution: taking a component and splits it into multiple parts and distributed them to nodes in a distributed system. It helps to reduce the load within one server.
        - Caching: a special form of replication. The client makes a copy of the resource so subsequent calls to the cached resource will get the resource from the cache instead of making another call to the server to request for the resource
6. Centralized vs decentralized:
    - Decentralized characteristics:
        - No machine has complete information about the system state
        - Machines make decisions based only on local information
        - No implicit assumption that a global clock exists
        - Failure of one machine does not ruin the algorithm
7. Cluster computing system:
    - A cluster consists of a collection of machines which are of the same type, and are located at the same location
    - Main advantage: much cheaper than supercomputers
    - Main disadvantage: network is a bottleneck in internode communication
8. Grid computing system:
    - Consists of heterogeneous nodes
    - Location of computing nodes can be geographically dispersed
    - Are generally cheaper than compute clusters
    - Administrational roles and heterogeneity pose a problem
9. Grid computing system architecture:
   Layered system architecture:
    - Application: a specific application within a grid
    - Collective layer: deals with access to multiple resources
    - Connectivity layer: protocols for data transfer
    - Resource layer: Act as an interface between connectivity and fabric layer
    - Fabric layer: Local resource at a site
10. Transaction processing system:
    - Operations:
        - BEGIN_TRANSACTION
        - END_TRANSACTION

    - ABORT_TRANSACTION
    - READ
    - WRITE
- Characteristics:
    - Atomic: transaction happens indivisibly
    - Consistent: does not violate system invariants
    - Isolated: concurrent transactions do not interfere with each other
    - Durable: changes are permanent once a transaction commits
- Implemented by a middle ware component called TP monitor, the TP processes each sub transaction so that to the client this appears as an atomic transaction



11. Distributed pervasive system
    - Computer systems applied to environments which change constantly
    - Requirements:
        - Recognize sharing as the default
        - Encourage ad hoc composition. Many devices in pervasive systems will be used in very different ways by different users.
        - Embrace contextual changes
    - Example:
        - Mobile network
        - Sensor network (traffic control systems)
        - Disaster management system
        - Some health care systems

12. Advantages and disadvantages of distributed system:
    Advantages:
    - Make it easier to integrate different applications running on different computers into a single system
    - Scale well with respect to the size of the underlying network

    Disadvantages:
    - Weaker security
    - Degradation of performance
    - Complex software

13. Caching limitation and improvement
    - Limitation:
        - Out of date resource
        - Can be used only for static documents
    - Improvement
        - Extra headers (If-Modified-Since)
        - Expiration (how long ago the document was last modified when it is cached)
        - Server notifies the proxies that a document has been modified by sending an invalidation