Cluster Computing

Cluster computing refers to a group of loosely or tightly connected computers that work together in such a way that they can be viewed as a single system. The components of a cluster are usually connected to each other through fast local area networks. Each node (individual computer) in a cluster is set to perform the same task, controlled and scheduled by software.

Hardware Components

- Nodes (Computers/Servers): These are the individual computers that make up a cluster. Each node can process its part of the larger task.
- Networking Hardware (Switches, Routers, Cables): These components connect the nodes within the cluster to enable fast communication. This can be especially important for 'tightly' coupled clusters where tasks are split into many small parts and require frequent communication.
- Storage Systems: The nodes in a cluster typically need to access the same data. This could be stored on each node ('distributed' storage), or on a dedicated storage system accessed by all nodes.
- Load Balancer: This is crucial in a cluster to distribute the tasks among the nodes evenly. A well-balanced load can significantly enhance the performance of a cluster.
- Power Supplies/UPS: Since the cluster is often made up of many nodes, maintaining consistent power to all components is essential. Power supplies, often with backup UPS systems, are key to ensuring the cluster remains operational.

Software Components

- Cluster Management Software: Software like Kubernetes or Apache Mesos manages the deployment of tasks, schedules them, and monitors the health of the nodes and tasks.

- Operating System: An OS that supports clustering capabilities, like Linux, is required on each node. They can manage the resources of each node and allow the cluster management software to run.
- Networking Software: This helps in establishing secure and efficient communication between nodes. It should be able to handle the high-speed exchange of data.
- Middleware: This software provides a unified view of the distributed resources and hides the complexity of the network from the user.
- Parallel Programming Environments: Tools and libraries that support the development of parallel applications (like MPI and OpenMP) are essential to efficiently use the cluster.

SAN and NAS in Cluster Computing

- SAN: In a clustered environment, SAN can provide high-speed access to shared block storage. It can be useful when large amounts of data need to be shared among the nodes, and the applications running on these nodes require block-level access to the stored data. This could be crucial for data-intensive applications, like databases or big data analytics, running in a cluster.
- NAS: NAS could be a good fit for clusters where data sharing at the file level is required among nodes. It provides a single, consistent view of the data to all nodes. A use case could be a web server cluster, where all nodes need to serve the same set of files.
- Mass Storage: This can be useful in cluster computing when dealing with massive amounts of data, where data durability and availability are crucial. Cloud-based mass storage solutions can provide scalable storage that can be accessed by all nodes in the cluster, making it suitable for large-scale, data-intensive applications. For example, a Hadoop cluster processing big data could use mass storage for its data.