Deploying Databases on Cluster Servers

Additional Material

The Need for Cluster Servers

- As data size and processing requirements exploded, businesses sought ever more powerful computer systems to handle the processing loads
- Businesses started demanding 99% to 99.999% uptime
  - 99% “uptime” allows for 12 minutes of downtime in 30 days
  - 99.999% trims that to 26 seconds downtime in 30 days
  - Requires multiple computers with fully redundant components
- At some point, electrons cannot move faster, so parallel processing is required for additional performance

Cluster Server Architecture

- ***Computer cluster***: multiple computers (servers), connected (coupled) together such that they can be viewed and managed as a single system
- Each server in a cluster is commonly called a ***node***
  - Usually, each node uses the same hardware and operating system
  - Node connected by network, loosely or tightly coupled
- Three basic architectures:
  - MPP (massively parallel processing), or “shared nothing”
  - SMP (symmetric multiprocessing)), or “shared everything”
  - Hybrid, which I call “shared disk”

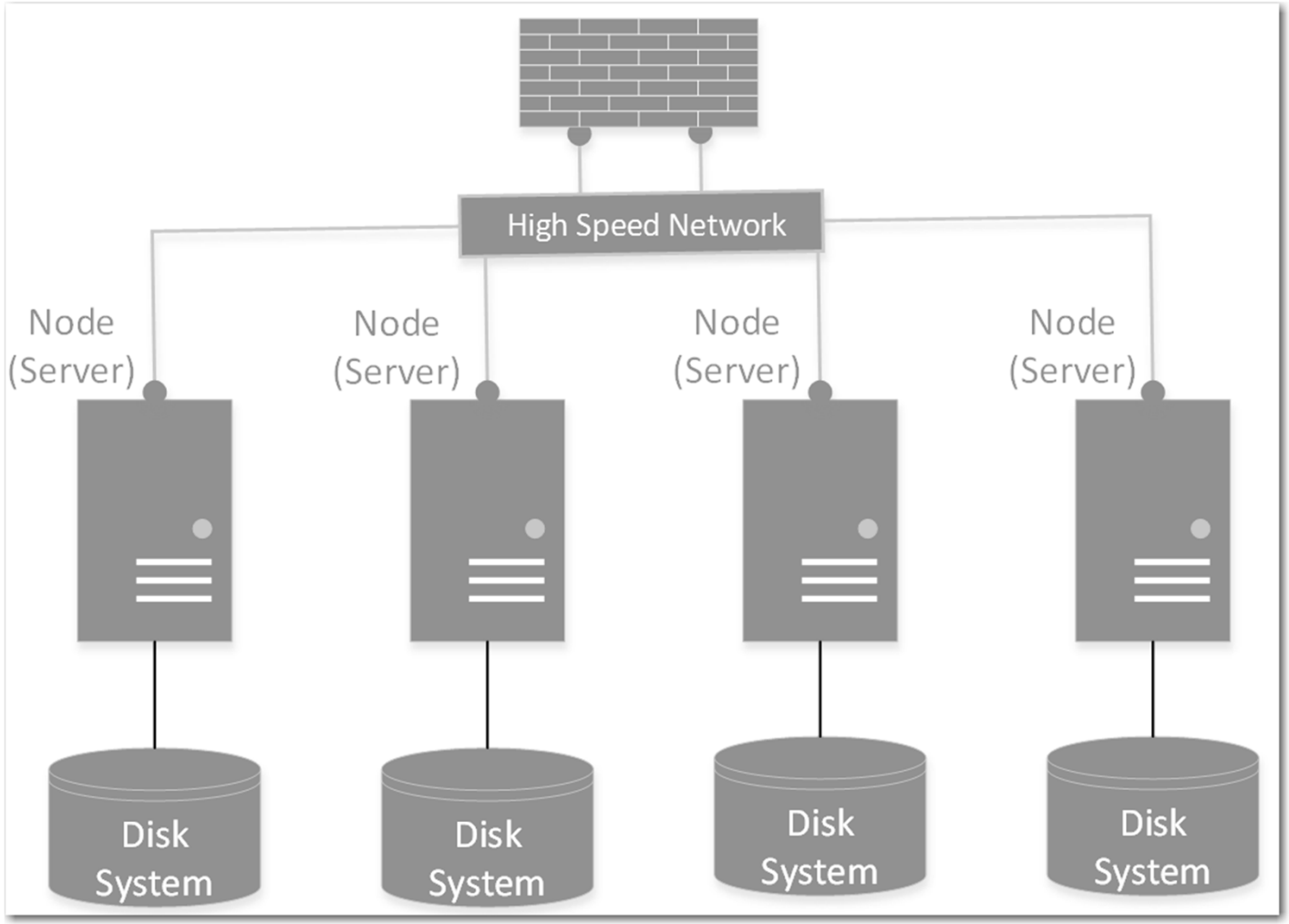
MPP Architecture

- Each node has its own CPU (central processing unit), memory, and dedicated disk storage system
- Data must be divided across nodes
  - Subsets called partitions, segments, or shards
  - Segmentation usually determined by one or more columns
  - Hash function often used to determine node
    - A *hash function* is a software module that maps data of arbitrary size to data of a fixed size
- Supported by MySQL, Microsoft SQL Server, HP Vertica, Bizgres MPP (based on PostgresSQL), and most NoSQL products

MPP Architecture

- Disk systems can be:
  - Simple disk drives
  - Disk storage systems (typically RAID devices)
- To support fault tolerance, each data segment must be redundantly stored (replicated) on another device

MPP (Shared Nothing) Architecture



MPP Advantages

- **Cost**: Hardware/software costs are minimized through use of commodity (commercial off the shelf) systems
- **Scalability**: Cluster can be expanded easily by simply adding more nodes
- **Widely Supported**: Many DBMS vendors that support this architecture

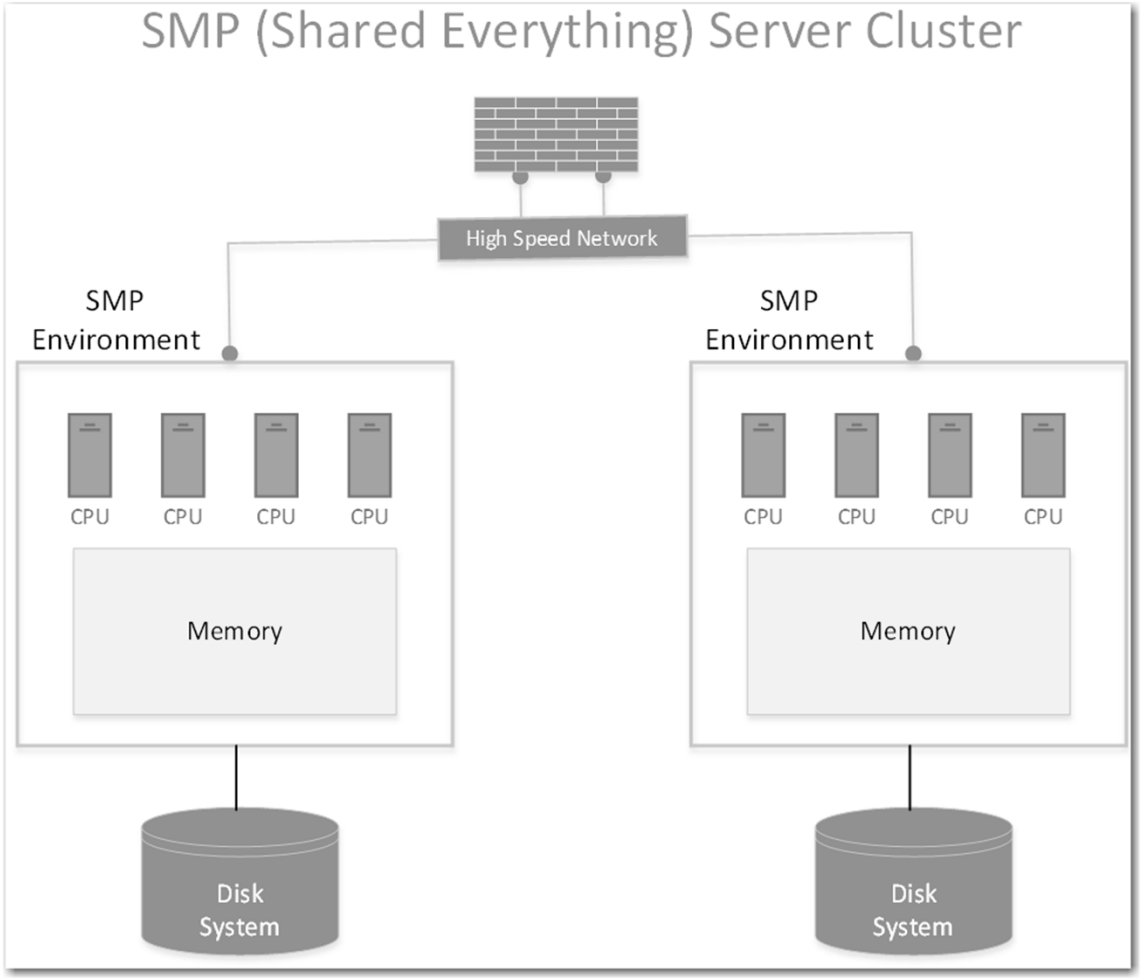
MPP Disadvantages

- **Cross-node Performance**: Joins that require data from multiple nodes cause data to be transmitted from other nodes to the node performing the join
  - Known as a “broadcast” join
  - Performance suffers severely
- **Processing Must Be Asymmetric**: Global queries must be split into queries that can be processed independently on each node
  - Each node queries its part of the data)
  - Results sent to one of the nodes for merge into the final query result set.
  - Data must be distributed evenly for optimized performance

SMP (Shared Everything) Architecture

- SMP (symmetric multiprocessing) systems:
  - Commonly known as “shared everything” servers , as the name suggests, server systems
  - Contain:
    - Many CPUs
    - Common bank of memory shared by all the CPUs
    - A common disk system

Two SMP Environments



Advantages of SMP

- **Support for Complex Symmetric Operations**: When performing complex calculations that require a lot of memory and CPU resources, this solution is hard to beat.
- **Higher Performance**: SMP environments are generally high-powered units that perform well above the levels of comparable MPP systems.

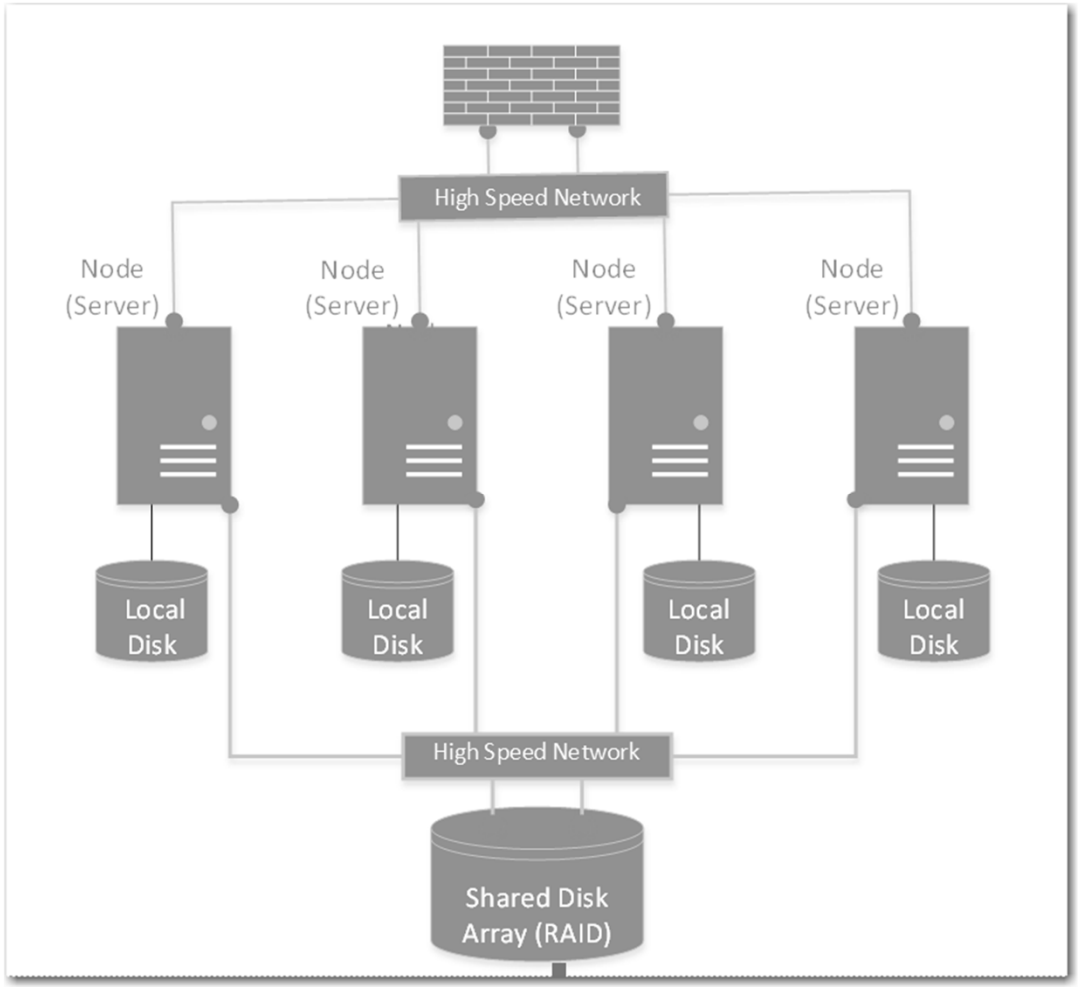
Disadvantages of SMP

- **Cost**: Higher hardware/software costs are because of specialized hardware
- **Scalability Has Limits:** Servers can be expanded by adding or upgrading internal CPU and memory units, but there is always a maximum amount that will fit
- **Limited DBMS Support:** The only vendor that supports an SMP database architecture is IBM.
  - However, Oracle uses the shared disk hybrid of the SMP architecture (presented next)

Shared Disk (SMP Hybrid) Architecture

- SMP modified so nodes can access a shared disk system as well their local disk
  - Shared disk system connected to all cluster nodes via a high-speed network
  - Currently, Oracle is the only DBMS vendor that uses this architecture, which they call Oracle Real Application Clusters (RAC).

Shared Disk Architecture



Shared Disk Architecture

- Attempts to optimize the advantages of both the MPP and SMP architectures
- To minimize contention for data contained on the shared disk system
  - Oracle RAC maintains a shared cache on each node through a feature they call “Cache Fusion”, which uses the high-speed network to keep the caches in alignment.
- As with all SMP architectures, the main disadvantage is higher cost.

Physical Design Considerations

- To optimize databases deployed on cluster servers:
  - Follow vendor guidelines
  - Minimize global indexes
    - A *global index* contains values for all the rows in a relational table (across all the nodes)
    - Primary key and unique indexes must be global (to permit duplicate checking)
    - *A local index* is segmented across nodes in the cluster
      - Help query performance by allowing index searches to be performed in parallel
      - Reduces network traffic between nodes when indexed column data is maintained

Physical Design Considerations

- Avoid Joins Across Nodes by:
  - Large tables to be joined should be segmented by the join key columns. This allows joins to be done locally on each node.
  - Joining a large table with a “small” one can be done by replicating the small table across all nodes. This again allows joins to be done locally on each node.