



SNS COLLEGE OF TECHNOLOGY

Coimbatore-35

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A++' Grade

Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai



DEPARTMENT OF MECHATRONICS ENGINEERING

UNIT 3 – HADOOP

HORIZONTAL SCALABILITY



HORIZONTAL SCALABILITY



Horizontal scalability, also known as scale-out or adding more nodes, is a key characteristic of distributed systems like Hadoop. It involves increasing the capacity of a system by adding more machines (nodes) to a network, rather than upgrading the resources of individual nodes. Horizontal scalability is a fundamental concept in Hadoop, allowing clusters to handle larger volumes of data and perform more parallel processing. Here are key aspects of horizontal scalability in the context of Hadoop:

Key Characteristics:

1.Adding Nodes:

1. Horizontal scalability involves adding more nodes to a distributed system. In the context of Hadoop, this means expanding the cluster by introducing additional commodity hardware.

2.Linear Scalability:

1. Ideally, horizontal scalability allows the system to scale linearly, meaning that adding more nodes results in a proportional increase in overall capacity and performance. This is crucial for handling growing datasets and processing demands.

3.Fault Tolerance:

1. Horizontal scalability enhances fault tolerance. With multiple nodes, data can be replicated across the cluster, ensuring that if a node fails, the system can continue processing tasks using other available nodes.

4.Load Balancing:

1. Horizontal scaling facilitates load balancing by distributing tasks across multiple nodes. This helps prevent overloading specific nodes and ensures efficient resource utilization.

5.Ease of Expansion:

1. Adding new nodes to a Hadoop cluster is relatively straightforward. The system can dynamically adapt to changing workloads, and administrators can expand the cluster as needed.



HORIZONTAL SCALABILITY



Benefits:

1. Cost-Effective:

1. Horizontal scalability is often more cost-effective than vertical scalability. Commodity hardware is typically less expensive, and organizations can incrementally add nodes based on their budget and requirements.

2. Flexibility:

1. Horizontal scaling provides flexibility in adapting to changing demands. As the volume of data or processing requirements increases, organizations can easily add more nodes to meet the growing needs.

3. Easy Maintenance:

1. Maintenance is simplified in a horizontally scaled system. Individual nodes can be taken offline for maintenance or upgrades without disrupting the entire system.

4. Improved Performance:

1. The parallel processing capabilities of horizontal scalability can lead to improved performance, especially when dealing with large-scale data processing tasks such as those handled by Hadoop's MapReduce framework.

5. Scalability Beyond Hardware Limits:

1. Horizontal scalability allows a system to scale beyond the physical limits of individual hardware components. By distributing the workload across multiple machines, the system can achieve levels of scalability that would be challenging with a single, powerful machine.



HORIZONTAL SCALABILITY



Considerations:

1. Network Bandwidth:

- The efficiency of horizontal scaling depends on the network bandwidth between nodes. A high-bandwidth network is essential for optimal communication and data transfer within the cluster.

2. Cluster Configuration:

- Proper cluster configuration is crucial for horizontal scalability. The Hadoop ecosystem and related tools are designed to distribute and parallelize tasks, and a well-configured cluster ensures balanced resource utilization.

3. Data Partitioning:

- Efficient data partitioning strategies are essential for maximizing the benefits of horizontal scalability. Well-distributed data across nodes helps prevent bottlenecks and ensures parallel processing.

4. Dynamic Resource Allocation:

- Modern Hadoop distributions, along with resource management frameworks like YARN, provide dynamic resource allocation. This allows the cluster to adapt to changing workloads, contributing to horizontal scalability.

Horizontal scalability is a fundamental concept in designing and managing large-scale distributed systems like Hadoop. It provides a flexible and cost-effective solution for accommodating the increasing demands of big data processing and analytics.