Topic 1: Cluster Computing

1. Computer Cluster - Definition and Architecture

What is a Computer Cluster?

- **Definition**: A collection of computers (nodes) connected through high-speed network that work together to simulate a single much more powerful computer system
- Each node is controlled by its **own operating system**
- Each node performs a **different version of the same task**

Key Distinction: Cluster vs Grid

- **Computer Cluster**: Nodes perform different versions of the same task
- **Computer Grid**: Nodes perform different tasks

Architecture Range

- Simple: Two-node system connecting two personal computers
- Complex: Supercomputer with cluster architecture

2. How Computer Clusters Work

Core Mechanisms

1. **Shared Nothing (Sharding) Partitioning**
    - Data is partitioned across nodes
    - Speeds up computing through data distribution
2. **Parallelization**
    - Data processing occurs simultaneously on multiple nodes
    - Dramatically increases processing speed
3. **High Availability**
    - Automatic replacement of failed nodes with replica nodes
    - Ensures continuous operation

Advantages of Computer Clusters

- **Faster processing speed**: Parallel computing capabilities
- **Larger storage capacity**: Distributed storage across nodes
- **Better data integrity**: Replication and redundancy
- **Greater reliability**: Fault tolerance mechanisms
- **Wider availability of resources**: Distributed resource access

3. Sample Cluster Configuration

Example Cluster Setup

Regular Compute Nodes (54 nodes):

- Two 32-Core Intel 8358 processors
- 1.6TB of local NVME storage
- 512GB of memory each

GPU Nodes (5 nodes):

- Two 24-Core AMD EPYC 7413 processors
- Eight A100 GPU cards
- 960GB of local storage
- 512GB of memory each

4. Cluster Computing - Process

What is Cluster Computing?

- **Definition**: The process of sharing computation tasks among multiple computers in a cluster
- Attractive paradigm for processing large-scale science, engineering, and commercial applications

Advantages

- **Cost efficiency**: Cheaper than supercomputers
- **Processing speed**: Parallel processing capabilities
- **Expandability**: Easy to add more nodes
- **High availability**: Fault-tolerant design

Requirements for Optimization

Specialized algorithms needed:

- **Load balancing**: Distribute work evenly across nodes
- **Resource sharing**: Efficient use of cluster resources
- **Resource scheduling**: Optimize task allocation

Simplest Configuration

- **Master Node**: Controls and coordinates cluster operations
- **Slave Nodes**: Execute tasks assigned by master node

5. Linux Cluster

- A collection of connected computers viewed and managed as **a single system**
- Provides unified management interface
- Standard platform for big data processing

Key Points for Exam

Critical Concepts

1. **Cluster Definition**: Collection of networked computers working as one
2. **Master-Slave Architecture**: Fundamental cluster organization
3. **Shared Nothing Architecture**: Data partitioning strategy
4. **High Availability**: Automatic failover capabilities
5. **Parallelization**: Core performance advantage

Important Distinctions

- Cluster vs Grid (same task variations vs different tasks)
- Vertical vs Horizontal scaling
- Cost-efficiency compared to supercomputers

Remember

- Clusters use commodity hardware (not high-performance computers)
- Each node has its own OS
- Automatic replacement handles failures
- Used extensively in big data processing (Hadoop, Spark, etc.)