Паралелно и дистрибуирано процесирање

Домашна 3

Стефан Милев | 206055

2.1

a)

- In a compact cluster, the nodes are closely packaged in one or more racks sitting in a room, and the nodes are not attached to peripherals (monitors, keyboards, mice, etc.).
- In a slack cluster, the nodes are attached to their usual peripherals (i.e., they are complete SMPs, workstations, and PCs), and they may be located in different rooms, different buildings, or even remote regions.
- While a compact cluster can utilize a high-bandwidth, low-latency communication network that is often proprietary, nodes of a slack cluster are normally connected through standard LANs or WANs.

b)

- A cluster can be either controlled or managed in a centralized or decentralized fashion.
- A compact cluster normally has centralized control, while a slack cluster can be controlled either way.
- In a centralized cluster, all the nodes are owned, controlled, managed, and administered by a central operator
- In a decentralized cluster, the nodes have individual owners. For instance, consider a cluster comprising an interconnected set of desktop workstations in a department, where each workstation is individually owned by an employee.
- The owner can reconfigure, upgrade, or even shut down the workstation at any time. This lack of a single point of control makes system administration of such a cluster very difficult.
- It also calls for special techniques for process scheduling, workload migration, check pointing, accounting, and other similar tasks.

c)

- A homogeneous cluster uses nodes from the same platform, that is, the same processor architecture and the same operating system; often, the nodes are from the same vendors.
- A heterogeneous cluster uses nodes of different platforms. Interoperability is an important issue in heterogeneous clusters.
- For instance, process migration is often needed for load balancing or availability. In a homogeneous cluster, a binary process image can migrate to another node and continue execution.
- This is not feasible in a heterogeneous cluster, as the binary code will not be executable when the process migrates to a node of a different platform.

d)

- In an exposed cluster, the communication paths among the nodes are exposed to the outside world. An outside machine can access the communication paths, and thus individual nodes, using standard protocols (e.g., TCP/IP).
- Such exposed clusters are easy to implement, but have several disadvantages: Being exposed, intracluster communication is not secure, unless the communication subsystem performs additional work to ensure privacy and security. Outside communications may disrupt intracluster communications in an unpredictable fashion.
- In an enclosed cluster, intracluster communication is shielded from the outside world, which alleviates the aforementioned problems. A disadvantage is that there is currently no standard for efficient, enclosed intracluster communication

e)

- A dedicated cluster is typically installed in a deskside rack in a central computer room.
- It is homogeneously configured with the same type of computer nodes and managed by a single administrator group like a frontend host.
- Dedicated clusters are used as substitutes for traditional mainframes or supercomputers.
- A dedicated cluster is installed, used, and administered as a single machine. Many users can log in to the cluster to execute both interactive and batch jobs.
- The cluster offers much enhanced throughput, as well as reduced response time.
- An enterprise cluster is mainly used to utilize idle resources in the nodes. Each node is usually a full-fledged SMP, workstation, or PC, with all the necessary peripherals attached.
- The nodes are typically geographically distributed, and are not necessarily in the same room or even in the same building.
- The nodes are individually owned by multiple owners. The cluster administrator has only limited control over the nodes, as a node can be turned off at any time by its owner.
- The owner’s “local” jobs have higher priority than enterprise jobs - The cluster is often configured with heterogeneous computer nodes.

2.2

a)

- The availability of a system is defined by: Availability = MTTF/(MTTF+MTTR)
- In this case the MTTR is 40 seconds

b)

- If only one node at a time is taken, then theoretically it would have 100% availability.

2.4

1.

a)

This draws attention to how well the scalability of a scalable parallel computer with additional processors will improve. Increased resources are CPU, memory capacity or I / O capabilities. For machines there is a maximum number of processors that the system can accommodate. This indicates the upper limit of adaptability over large machines.

b)

This shows how well the system can handle major problems such as data size and workload. It depends on the size of the machine but also on the memory capacity and communication ability of the machine.

c)

Used to get higher performance or functionality by increasing the number of processors, more memory or software upgrades. Three categories should be taken into account and taken into account. Adjustable machine size that shows how well the performance will improve. Then increase the resource which means getting higher performance by investing more memory. Finally, the scalability of the software that shows us how the performance of the system is improved with a new version of the operating system.

d)

This refers to increasing the performance of the system by using the next generation components. Examples are a faster processor, faster memory, a newer version of the operating system, or more powerful compiler.

2.

- Hot standby server clusters In a hot standby cluster, only the primary node is actively doing all the useful work normally. The standby node is powered on (hot) and running some monitoring programs to communicate heartbeat signals to check the status of the primary node, but is not actively running other useful workloads. The primary node must mirror any data to shared disk storage, which is accessible by the standby node. The standby node requires a second copy of data.
- Active-takeover clusters In this case, the architecture is symmetric among multiple server nodes. Both servers are primary, doing useful work normally. Both failover and failback are often supported on both server nodes. When a node fails, the user applications fail over to the available node in the cluster. Depending on the time required to implement the failover, users may experience some delays or may lose some data that was not saved in the last checkpoint.
- Failover cluster This is probably the most important feature demanded in current clusters for commercial applications. When a component fails, this technique allows the remaining system to take over the services originally provided by the failed component. A failover mechanism must provide several functions, such as failure diagnosis, failure notification, and failure recovery. Failure diagnosis refers to the detection of a failure and the location of the failed component that caused the failure. A commonly used technique is heartbeat, whereby the cluster nodes send out a stream of heartbeat messages to one another. If the system does not receive the stream of heartbeat messages from a node, it can conclude that either the node or the network connection has failed.
- Examples: Microsoft Failover Clusters, StarWind Virtual SAN

2.7

a)

From the top 10 supercomputing systems, the first 4 aren’t x86. While the next 6 are Intel or AMD x86 server CPUss. After that, most of the supercomputers on the list are x86 processors either from Intel or AMD.

b)

From the top 10 supercomputing systems, 3 don’t use any kind of accelerator or co-processor, while 6 use some kind of NVIDIA GPU, and 1 uses a Matrix GPU.

2.9

1.

NVIDIA DGX SuperPOD

- Computer: NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband
- Cores: 19,840
- Power Efficiency (GFlops/Watts): 26.20

2.

MN-3

- Computer: MN-Core Server, Xeon Platinum 8260M 24C 2.4GHz, Preferred Networks MN-Core, MN-Core DirectConnect
- Cores: 1,664
- Power Efficiency (GFlops/Watts): 26.04

The common pattern between the top most efficient supercomputer systems is their utilization of GPUs as accelerators as well as AMD Zen-2 (Rome) server CPUs.

2.10

a)

Those are Summit, Sierra and Sunway TaihuLight. They all use infiniband. Their peak floating point performances are:
- Summit: 200,794.88 TFlops/s
- Sierra: 125,712.00 TFlops/s
- Sunway TaihuLight: 79,215.00 TFlops/s

b)

Their Interconnects are:
- Summit: Dual-rail Mellanox EDR Infiniband
- Sierra: Dual-rail Mellanox EDR Infiniband
- Sunway TaihuLight: Mellanox HDR Infiniband

2.16

a)

Single-system image (SSI) is a very rich concept, consisting of single entry point, single file hierarchy, single I/O space, single networking scheme, single control point, single job management system, single memory space, and single process space.

b)

Single memory space gives users the illusion of a big, centralized main memory, which in reality may be a set of distributed local memory spaces. PVPs, SMPs, and DSMs have an edge over MPPs and clusters in this respect, because they allow a program to utilize all global or local memory space. A good way to test if a cluster has a single memory space is to run a sequential program that needs a memory space larger than any single node can provide.

c)

We use the term “single file hierarchy” to mean the illusion of a single, huge file system image that transparently integrates local and global disks and other file devices (e.g., tapes). All files a user needs are stored in some subdirectories of the root directory /, and they can be accessed through ordinary UNIX calls such as open, read, and so on.

d)

It is like a big workstation with four network connections and four I/O devices attached. Any process on any node can use any network and I/O device as though it were attached to the local node.

e)

The cluster exists on only one network (LAN). If there are problems with the network, the whole cluster is damaged. Synchronization has the biggest impact.

f)

Single networking means any node can access any network connection. A properly designed cluster should behave as one system (the shaded area). In other words, it is like a big workstation with four network connections and four I/O devices attached.

g)

The system administrator should be able to configure, monitor, test, and control the entire cluster and each individual node from a single point. Many clusters help with this through a system console that is connected to all nodes of the cluster. The system console is normally connected to an external LAN so that the administrator can log in remotely to the system console from anywhere in the LAN to perform administration work.

h)

In the dedicated mode, only one job runs in the cluster at a time, and at most, one process of the job is assigned to a node at a time. The single job runs until completion before it releases the cluster to run other jobs. Note that even in the dedicated mode, some nodes may be reserved for system use and not be open to the user job. Other than that, all cluster resources are devoted to running a single job. This may lead to poor system utilization.

i)

In the dedicated mode, only one job runs in the cluster at a time, and at most, one process of the job is assigned to a node at a time. The single job runs until completion before it releases the cluster to run other jobs. Note that even in the dedicated mode, some nodes may be reserved for system use and not be open to the user job. Other than that, all cluster resources are devoted to running a single job. This may lead to poor system utilization.

j)

It gives us a lot of computing power, and that's why everyone wants to use this power everywhere. All user processes on different nodes form a single process space and have a process identification scheme. A process on any node can create a new process or communicate with remote node processes.