CS147 - Lecture 18

Kaushik Patra
(kaushik.patra@sjsu.edu)

1

- Cluster Organization
- NUMA Organization
- Vector Processing

Reference Books / Source:

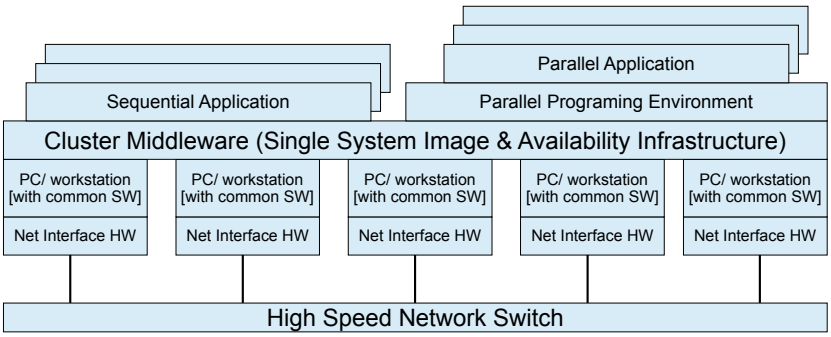
1) *Chapter 17 of 'Computer Organization & Architecture' by Stallings*

Cluster Computer ...

2

Cluster Computer

- Group of interconnected individual computers working together as unified computing resource that can create illusion of being one machine.



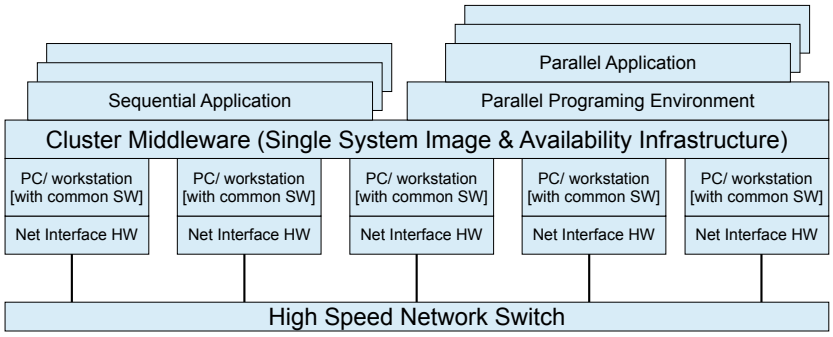
3

- In a cluster organization multiple machines are connected to each other through interconnect network in LAN or WAN. One single system image is presented to end user using cluster middleware. For an end user cluster system acts as a single machine. The user does need to specify the individual machine at which the program needs to be executed. Sequential applications directly can use the cluster through middleware. On the other hand, the parallel application usually needs to go through additional parallel programing environment (libraries in terms of programing implementation).

- This architecture has been developed as an alternative to SMP architecture to provide high performance and availability. This approach is attractive for server applications.

- Each individual machine is called a node is the cluster system.

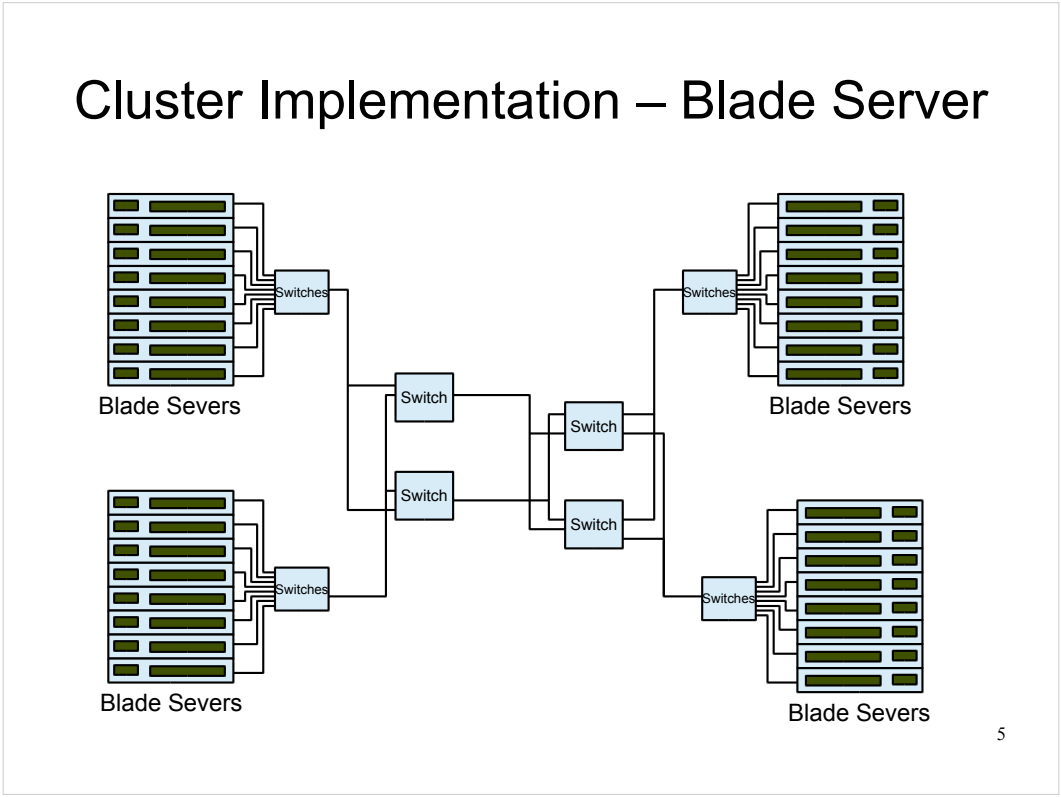
Cluster Architecture



- Single Entry Point
- Single File Hierarchy
- Single Control Point
- Single Virtual Networking
- Single Job Management
- Single User Interface
- Single IO Space
- Single Memory Space
- Single Process Space
- Check Pointing
- Process Migration

4

- Since cluster provides single point entry to the system, user logs onto the cluster rather than individual machine. Access to specific machine is controlled by access control mechanism deployed as a part of cluster.
- Even if files may reside in different hard disk across the network, cluster provides single file hierarchy to the end user, as if the directory structure is residing under the same root.
- Only single workstation is used as single control point to manage a cluster.
- Any node can access any other point of the network through single virtual networking, even though machines in the cluster are connected through hierarchies of network.
- Distributed shared memory enables programs to share variables.
- Under cluster job scheduler, user can submit a job without specifying the host to execute the program. This gives user a feeling of single job manager in the system.
- A GUI is used as a single user interface to user job management and monitoring independent of the workstation at which user is performing the task.
- Cluster gives feeling of single IO space where any node can access any IO device without knowledge to device's physical location.
- A single process space is created so that a process on a cluster node can create or communicate with any other process on a remote node.
- Check point service saves process state and intermediate computing result to allow rollback recovery after a failure.
- Process migration enables cluster system manager to taken one process out of a node in middle of execution and restart it in another node without altering result of the process. This is used for load balancing the cluster.



- A common implementation of the cluster approach is the blade server. A blade server is a server architecture that houses multiple server module or blades in a single chassis. It is widely used in data center to save space and improve system management.The chassis provides the power supply and each blade has its own processor, memory and hard disk. Multiple blade servers in a single chassis are connected via network switch. These network switch can communicate to other network switch in LAN/WAN to extend the cluster.

Cluster Implementation – Blade Server

Blade Servers in single chassis

Cluster Computer in Data Centre

Cluster Benefit

- Absolute Scalability
- Incremental Scalability
- High Availability
- Superior price/performance

7

- It is possible to create large clusters that far suppress the power of even the largest standalone machine. A cluster can have thousands of individual machines.
- A cluster is configured in such a way that it is possible to add new systems to cluster in small increments. Thus an user can start with a modest system and expand it as needed.
- Because of each node in the cluster is a standalone computer, cluster is fail safe computing environment. If one node fails, the work load is immediately distributed among other nodes.
- A cluster does not need very high end powerful computers. By using commodity building blocks, it is possible to put together a cluster with equal or greater computing power than a single large machine, at a much lower cost.

NUMA Organization ...

8

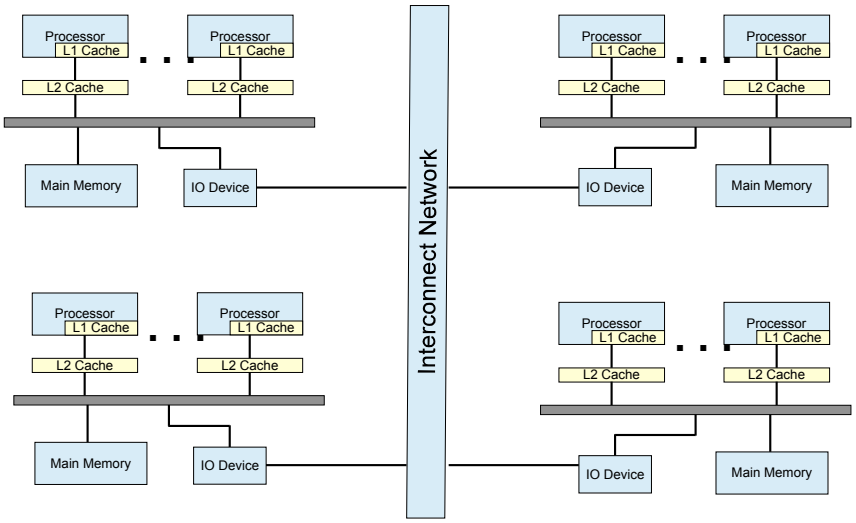
Shared Memory Machines Types

- Uniform Memory Access (UMA)
- Nonuniform Memory Access (NUMA)
- Cache-Coherent NUMA (CC-NUMA)

9

- For a UMA system, all processors have uniform access to all parts of the main memory using load/store operation. The memory access time to all the processor to all region of the memory is the same. The SMP organization is an example of UMA system.
- For NUMA system all the processors have access to all parts of the main memory using load/store operation. However, the access time of a processor may differ from other depending on which region is being accessed.
- For a CC-NUMA system, processor local cache memories are maintained with cache-coherency.

NUMA Organization



10

- NUMA is a hybrid approach of cluster and SMP. Multiple SMP machines are connected to a interconnect network to form cluster. In this system a process can access all the memory attached to each of the individual system through a single memory space. As a result the process access to local main memory portion is faster compare to a remote memory (thus non-uniform memory access). If a process in a node wants to access a memory space one a remote node, it must be done through network IO communication.

NUMA Motivation

- SMP systems can not expanded without limit.
  - A cluster system can.
- For a cluster system application does not see a larger global memory.
  - A SMP system does.
- NUMA is a hybrid of SMP and cluster.

11

- Typically a SMP system has limitation on expansion for additional processor. Usually such system it typically limited to 16 to 64 individual processors. This limitation occurs due to performance degradation for increasing interconnect traffic and usually data transfer through interconnect (bus or crossbar) becomes slower as we increase number of processors. A cluster system does not suffers from this limitation. It can be extended as much as needed.
- In a typical cluster system, process typically can access local main memory only. Hence in this system, memory space for a process is limited to local main memory only. On the other hand, a SMP system, all the processor has access to the entire available memory.
- NUMA tries to make a hybrid of these two concepts to take advantage of both of them. It connects multiple SMP system in cluster and make all the memory in each individual SMP system available to all the processors of all the nodes in the cluster.

Vector Processing ...

12

Different Paradigm of Array Operation

Scalar addition of array

```
float a[8], b[8], c[8];
int i;
for(i=0; i<8; i++)
   c[i] = a[i] + b[i];
```

Vector addition of array

```
float a[8], b[8], c[8];
c = a + b;
```

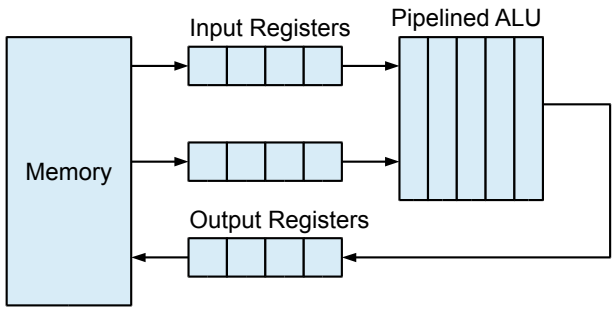
Parallel addition of array

```
float a[8], b[8], c[8];
int i;
#pragma omp parallel for
for(i=0; i<8; i++)
   c[i] = a[i] + b[i];
```

13

- We can write three different flavor of writing addition of two arrays. In the scalar version, it takes a loop to compute addition of individual elements in the array.
- In the parallel version (using Open MP, for example) we uses compiler directive 'pragma' to parallelize the loop where each individual addition is done in parallel on different processor in the execution environment. You can get more on OMP based parallelization in http://https://computing.llnl.gov/tutorials/openMP/#API
- In the vector version of the program the array operation is written just as its single variable version. Compiler schedules the operation as scalar operation on a scalar only system or uses vector computation capability if it is present in the hardware. Vector computation is more related to expensive floating point capability.

Vector Computation Organization I

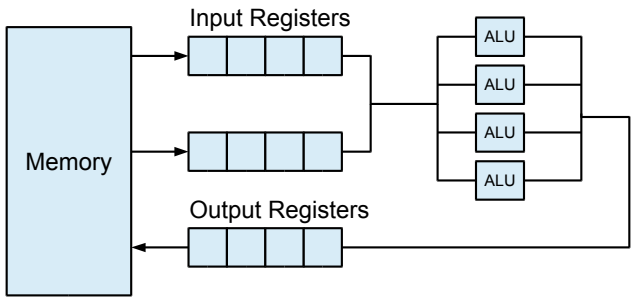


Pipelined ALU Vector Processor

14

- Vector computation can be implemented in two ways. First is using a pipelined ALU. Typically for a floating point operation involves multiple independent stages. Therefore a floating point ALU (usually called a floating point processor or FPU) can be pipelined (similar to pipelined processor execution). In this organization, operand registers (or input registers) are loaded with multiple operands and then pushed into the ALU pipeline. The output from the pipeline is stored into the output registers before it is stored back into the memory.

Vector Computation Organization II

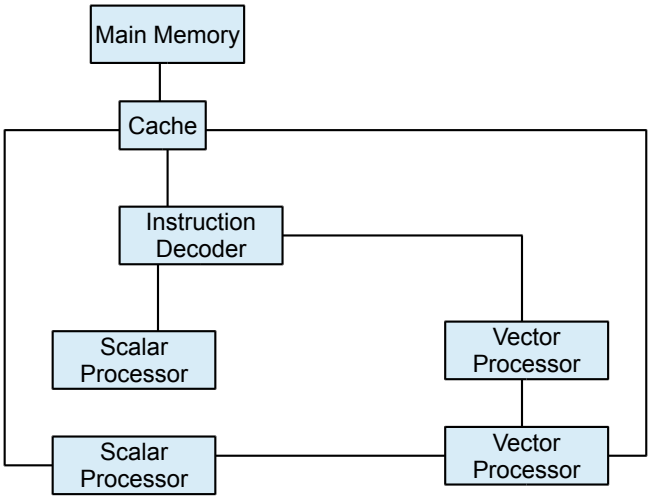


Parallel ALU Vector Processor

15

• The second approach is to use multiple ALUs and perform operation on multiple operands at same time. The advantage of this approach is that integer type array operations can also be vectorized.

Hybrid (scalar + vector) Architecture



IBM 3090 with Vector Facility

16

- Typically any vector computation system provides both scalar and vector processor. Based on the instruction decoding result, the instruction is scheduled into vector processor or the scalar processor.

CS147 - Lecture 18

Kaushik Patra
(kaushik.patra@sjsu.edu)

17

- Cluster Organization
- NUMA Organization
- Vector Processing

Reference Books / Source:

1) *Chapter 17 of 'Computer Organization & Architecture' by Stallings*