Hardware and Operating Systems for Data warehousing

→ H/w and OS make up the computing environment for data warehouse.

→ Here are some general guidelines for h/w selection:

Scalability: When DW grows in terms of the no. of users, the no. of queries, and the complexity of the queries, ensure that your selected h/w could be scaled up.

Support: Vendor support is crucial for h/w maintenance. Make sure the support from the h/w vendor is at the highest possible level.

Vendor Reference: It is important to check vendor references with other sites using h/w from this vendor.

Vendor stability: Check on the stability & staying power of the vendor.

→ All the data extraction, transformation, integration & staging jobs run on the selected h/w under the choosen OS.

→ There are two key points in the h/w requirement:

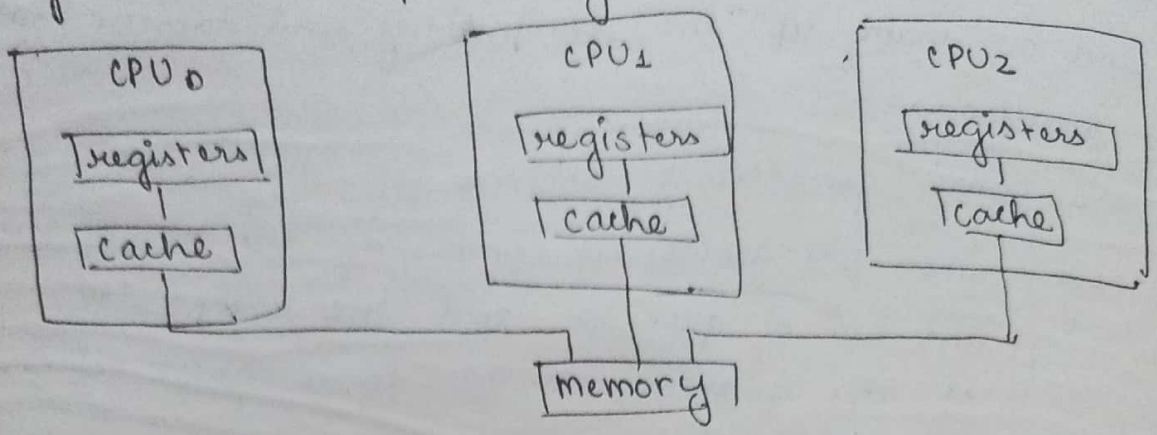
1. Hardware Architecture &
2. Hardware Management.

→ Required H/W. are as:

1. Server H/W
   - Symmetric Multiprocessing
   - Cluster Technology
   - Massive Parallel Processing
   - New and Emerging Technology
2. Network H/W
3. Client H/W

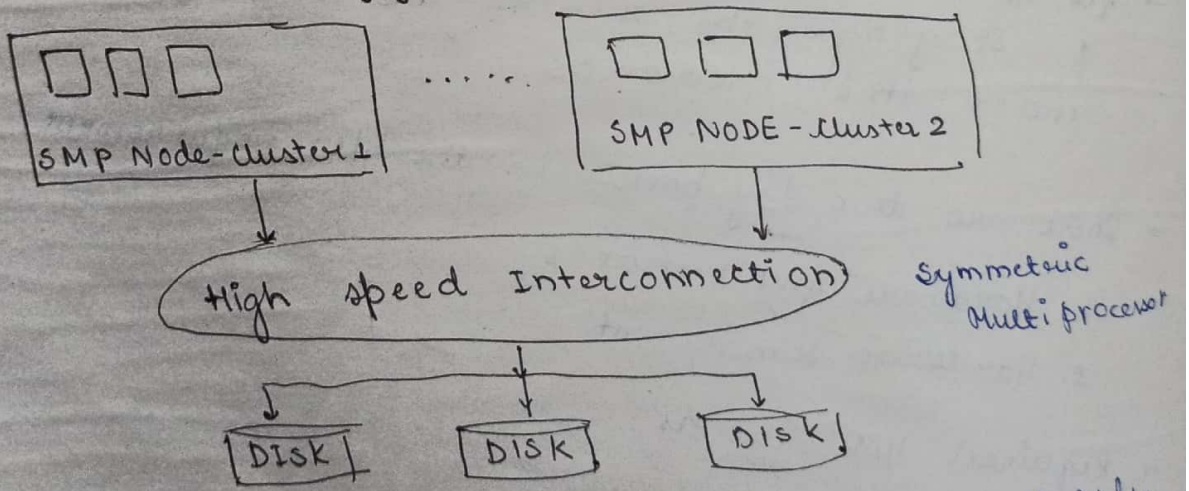
Server H/w and OS

Symmetric Multiprocessing Architecture



- Set of CPU shared memory & disks.
- Shared everything environment.
- Vendor have facility to allow processes to have CPU affinity
- Length of comm^n bus connecting to CPUs is in natural limit.
- Multiple CPUs allow operations to processed in parallel.

Cluster Technology

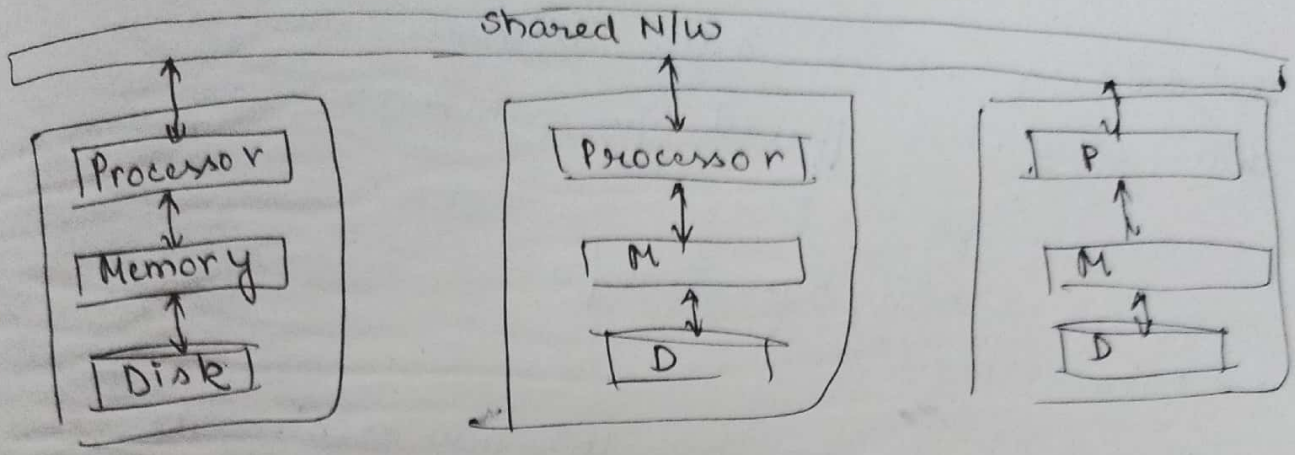


Symmetric Multiprocessor

→ A computer cluster is a set of loosely or tightly connected SMP that work together so that, in many respect s, they can be viewed as a single system.

→ Each m/c has its own CPUs and memory but they can share disk.

Massive Parallel Processing



MPP is the co-ordinated processing of a program by multiple processors that work on different parts of the program, with each processor using its own OS and memory.

Typically, MPP processors communicate using messaging interface. In some implementations, up to 200 or more processors can work on the same appⁿ.

An "interconnect" arrangement of data paths allow msgs to be sent b/w processors.

New $ Emerging Technology

→ It includes NUMA (Non-uniform memory Architecture) $ High speed memory interconnect.

→ It is basically tightly coupled cluster SMP nodes with high speed interconnect.

→ Interconnect enables os to run across the whole m/c as a single distributed instance.

N/w Hardware and OS

→ The main aspects of data warehouse design that may be affected n/w architecture as:

- User Access
- Source System Data Transfer
- Data Extraction.

→ We need to ensure that the n/w architecture and bandwidth are capable of supporting data transfer & extraction in acceptable time.

Client H/w and OS

→ Client mgmt & client tools aspects need to be consider during warehouse architecture design phase.

→ At design stage it is necesary to consider what user side tools will be used.

→ If these tools have special ~~arrangement~~ requirements such as data being summarized in certain ways.

→ These requirements must be given to user side.