



+ 이종복 · (주)일근인프라 차장

서버에 적용된 기술들

우리가 현업에서 사용하는 서버들은 전자계산기원론과 같은 이론적인 서적에서 보아왔던 중앙처리 장치, 주기억장치, 보조기억장치, 입출력장치로 크게 구분할 수 있습니다. PC와도 크게 다르지 않지만, 각각의 장치들은 기본적으로 PC들과는 다른 성능과 내구성을 가지고 있습니다. 첫 회에는 서버에서 사용되는 각종 장치(device)들에 대하여 알아보고 2회에서는 가장 기본적인 각 벤더사의 서버들을 간략히 이야기하도록 하겠습니다.

이번 회에서는 크게 CPU, MEMORY, Disk, I/O Bus에 대해 알아보겠습니다.

1. CPU

갤러틴, 우드클레스트, 올프데일, 하퍼타운 등 여러 세대를 거치면서 지금 주로 사용되는 CPU(Intel계열) 계열은 아래의 같습니다.

Intel Xeon CPU Transition Roadmap

| | ~2011년 | 2011년 8월 현재 | 2011년 말 예정 |
|------|---------------------------------------|-------------------------------------|-------------------------------------|
| 4P이상 | Xeon 7500/6500 Series (Nehalem-EX) | Xeon E7 Series (Westmere-EX) | Xeon E7 Series (Westmere-EX) |
| 2P | Xeon 500 Series (Westmere-EP) | Xeon 5600 Series (Westmere-EP) | Xeon E5 Series (Sandy Bridge EP) |
| 1P | Xeon 3400 Series (Lynnfield) | Xeon E3 Series (Sandy Bridge UP) | Xeon E3 Series (Sandy Bridge) |

그림 1 Intel CPU Roadmap

올해 말 새로운 CPU가 등장할 예정이지만 논외로 하고 현재 주로 사용되는 Westmere의 주된 아키텍처와 특징은 다음과 같습니다.

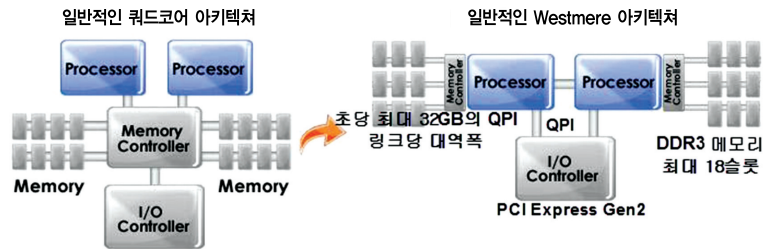


그림 2 프로세서 아키텍처

프로세서 아키텍처

네할렘의 후속프로세서로 공정기술을 45nm에서 32nm 공정으로 향상했으며 Quad-Core 또는 Hexa-Core 모델을 제공합니다. 소켓당 core 개수가 최대 6개로 증가하면서 성능이 크게 향상되면서도 소모 전력은 그대로 사용하기 때문에 단위 전력당 성능(Performance/Watt)이 더욱 향상되었습니다.

공유 L3 Cache의 경우 종전 세대 Xeon이 8MB였으나, Westmere는 12MB로 증가하였습니다. 기존 FSB(Front Side Bus) 방식보다 더 높은 대역폭과 줄어든 대기시간 특성을 가지는 메모리 컨트롤러가 내장된 구조로 프로세서당 32GB/s의 메모리 대역폭을 제공하고 각 프로세서에 독립된 2개의 Quickpath Interconnect Link를 연결하여 Link당 32GB/s의 대역폭을 제공합니다.

I/O Controller Hub 10을 통하여 PCI Express 2.0 인터페이스를 제공하며 병렬화를 강화하여 종전보다 연산 효율성이 33% 좋아졌고 XML 가속을 지원하는 SSE4.2 인스트럭션을 제공합니다.

Dynamic Power Management 기능으로 부하에 따라 프로세서 Clock 속도를 변경하는 방식으로 전력 소모를 줄이고 대기상태의 소모전력을 줄였습니다.

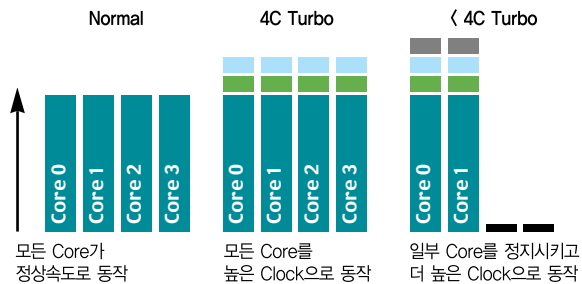


그림 3 Turbo Boost

특징 (Turbo Boost)

Turbo Boost 모드를 지원하여 Threading이 잘되지 않고 프로세서 Clock 속도에 민감한 Apps에 대한 성능 향상이 가능합니다. Turbo Boost 모드에서는 추가적인 성능이 필요한 경우에 프로세서의 전력, 냉각에 대한 허용 범위에서는 Clock 속도를 표기된 정상 Clock 이상으로 높을 수 있도록 해줍니다.

모든 Core는 동일한 속도로 동작하며 허용범위 내에서 133MHz 단위로 Clock이 조정됩니다. BIOS에서 특정 Core를 Disable 시킬 경우 모든 Core를 쓰는 경우보다 더 높은 Clock 속도 조정을 할 수 있습니다. 예를 들면, DB Application은 실수연산이 적어서 발열이 적게 됩니다. 이러면 모든 Core의 속도를 높일 수 있습니다.

2. Momory

메모리아키텍처

메모리 아키텍처는 칩셋을 통해 프로세서와 메모리를 연결하는 Front Side Bus 방식에서 프로세서에 메모리 컨트롤러를 내장시켜 메모리를 직접 연결하는 Integrated Memory Controller 방식으로 변경되었습니다. 총 18개의 메모리 슬롯이 있으며 1개 프로세서만 장착하면 9개의 메모리 슬롯만 사용할 수 있습니다.

메모리 액세스 방식도 모든 프로세서에서 동일한 액세스 속도를 보장하는 UMA(Uniform Memory Access) 방식에서 프로세서 관점에서 볼 때 지역 메모리와 원격 메모리에 대한 액세스 속도가 달라지는 NUMA(Non Uniform Memory Access) 방식에서 변경되었습니다.

DDR3 메모리는 DDR2 메모리에 비해 버스 속도와 전송 속도가 개선되어 채널당 대역폭이 최대 2배 증가하고 서버당 메모리 채널 수도 4개에서 6개로 증가하여 총 메모리 대역폭이 최대 3배 정도 증가 했습니다. 이로써, 메모리 성능에 민감한 Application을 운영할 경우 추가적인 성능 향상을 기대할 수 있습니다
 메모리 작동 전압은 1.8V에서 1.5V 정도로 줄어들어 동일 데이터 전송 속도 기준시 약 25% 정도의 전력 소모량이 줄었습니다.

메모리의 종류

메모리 종류는 크게 Registered 메모리(R-DIMM)와 Unbuffered 메모리(U-DIMM)로 구분되며 동일 서버에서 혼용할 수는 없습니다. Registered 메모리는 DRAM 칩에 메모리 컨트롤러와 연결되는 Register를 가지고 있으며 Register는 일종의 Pass thru Buffer 역할을 합니다. 메모리 컨트롤러는 Register에 명령어를 전달하고 Register는 명령어를 해당하는 DRAM 칩에 전달하는 방식으로 72bits 데이터가 직접적으로 DRAM 칩에 전달되는 것을 막기 때문에 신호 손실 위험이 적어서 메모리 버스에 더 많은 메모리 모듈을 연결할 수 있게 해 줍니다.

Unbuffered 메모리는 데이터 버스에 직접 연결되는 구조이기 때문에 신호 손실을 막기 위해 제한된 슬롯만 지원합니다.

R-DIMM은 서버에 장착된 18개의 메모리 슬롯을 모두 쓸 수 있지만 U-DIMM은 12개의 메모리 슬롯만 사용할 수 있습니다. R-DIMM은 Advanced ECC와 Memory Mirroring, Memory Lockstep과 같은 메모리 보호 기능을 제공하며 U-DIMM은 작은 용량의 메모리를 장착할 경우 R-DIMM보다 메모리 Latency가 작은 장점이 있어 성능 향상을 기대할 수 있습니다.

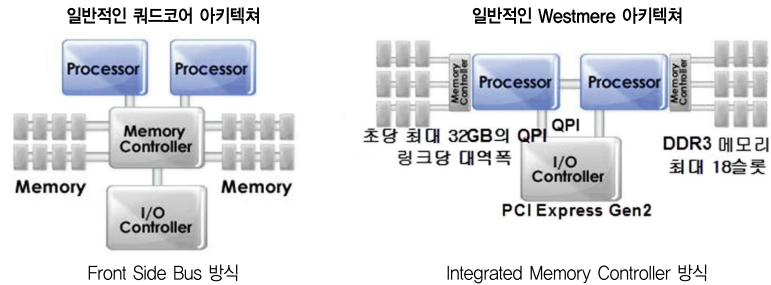


그림 4 메모리 아키텍처

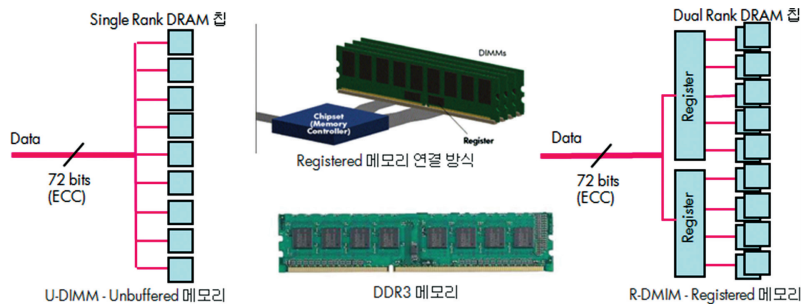


그림 5 메모리의 종류

R-DIMM은 Enterprise급 Application 운영에 적합하고 U-DIMM은 하드웨어 가용성에 덜 민감한 Scale Out 형태의 서버 운영에 적합하며 가격 측면에서도 U-DIMM이 R-DIMM보다 저렴합니다.

메모리구성에 따른 성능차이

DDR3 메모리는 채널에 몇 개의 메모리를 장착하느냐에 따라 실제 메모리 동작 속도가 달라지는 특성이 있습니다. 현재 최대의 메모리 속도인 1333MHz(10.6GB/s)를 얻기 위해서는 채널당 1개의 메모리만 장착해야 하는 제약점이 있습니다.

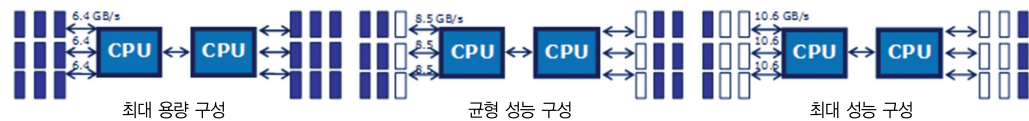


그림 6 메모리 구성방법

3. Hard Disk

주기억장치인 메모리는 전원을 끄면 저장된 내용이 모두 없어지므로 전원을 끄더라도 Data가 저장될 장치가 필요하며 이를 보조기억장치라고 합니다. 보조기억장치 중 가장 많이 사용되는 것이 하드디스크입니다. OS 및 각종 Data의 저장과 IO를 담당하는 보관영역이며 시스템을 운영하는 데 있어 중요한 요소입니다.

물리적인 구조로 본다면 헤드, 헤드암, 스피들모터, 스테핑모터, 플래터(원형디스크), 카트리지(케이스), 인터페이스 부분으로 나누어집니다.

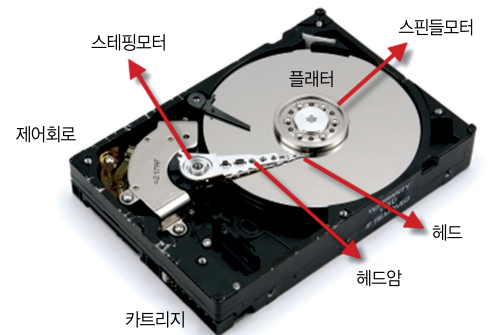


그림 7 Disk 구조

플래터를 트랙과 섹터로 구분하고 그 위로 헤드가 이동하며 자성을 이용하여 0과1의 이진수를 기록하며 바깥쪽에서 안쪽으로 데이터를 저장합니다.

주로 서버용으로 사용되는 HDD는 Interface, RPM, Capacity에 따라 여러 종류의 HDD를 사용하고 있습니다. 과거에는 속도 및 안정성을 바탕으로 3.5인치 SCSI Disk를 사용하였지만 가격적인 문제, DATA의 증가에 따른 Disk 수요증가, 소모전력 등의 문제로 2005년경부터 다양한 방식의 Interface를 가진 HDD로 대체되어 사용 중입니다. 하지만 Disk의 성능은 Interface보다는 RPM에 따라 가장 큰 영향을 받습니다. SSD > 15K Disk > 10K Disk > 7.2K Disk 의 순서로 느려집니다.



그림 8 여러종류의 Interface

인터페이스별로 HDD의 스펙을 표로 정리하면 다음과 같습니다.

| 구 분 | FC HDD | SAS HDD | SATA (FATA) |
|----------------------|-------------------------------------|-----------------------|--------------------------------|
| RPM | 15000 or 10000 | 15000 or 10000 | 7200이하 |
| Interface | Dual Port 4Gbps Fibre Channel | Dual Port 6Gbps SAS | Single Port 3Gbps SATA |
| 연결거리, 최대연결수 | 수 km이상, 8백만개 | 8m, 128 | 1m, 16 |
| Avg. Access Time | 5.4 / 5.9 | 5.4 / 5.9 | 12.7 / 13.7 |
| Max Transfer Rate | 170MB/s 이내 | 170MB/s 내외 | 100MB/s 이내 |
| Error Bit 비율(Sector) | 1/10 ⁶ | 1/10 ⁶ | 1/10 ⁵ |
| MTBF | 160만 시간 | 160만 시간 | 120만 시간 |
| 용 량 | ~ 600GB | ~ 1TB | ~ 2TB |
| 소비전력 | 높음 | 높음 | 낮음 |
| 사용처 | OLTP, DW, ERP 등 Enterprise Storage | Entry, Midrange Array | Archive, Backup 등 Low-end Disk |
| 비 고 | 구체적인 수치는 제조사별, 신규제품에 따라 달라질 수 있습니다. | | |

그림 9 HDD 비교

RAID(Redundant Array of Independent Disks)

Disk가 서버용으로 사용 시에는 성능적인 면도 중요한 요소이지만 안정성도 중요한 요소입니다. 따라서 RAID Card를 이용하여 여러 Disk의 Pool을 하나의 Disk처럼 사용하고 장애발생 시 복구할 수 있도록 하고 있습니다.

- RAID0 (Striping)

Data의 빠른 입출력이 가능하도록 여러 disk에 분산하여 저장하고 어느 한 disk에 장애가 발생하면 data가 손실됩니다. 빠른 입출력 속도가 요구되나 장애복구 능력은 필요 없는 시스템에 적합합니다. 최소 3개의 disk가 필요하며 실가용량은 Disk용량 × N(Disk개수) 입니다.

- RAID1 (Mirroring)

동시에 2개의 Disk에 같은 data를 저장하고 Data의 안정성이 중요한 곳에 사용됩니다. 최소 2개의 disk가 필요하며 짝수개로 구성되며 실가용량은 Disk용량 × N/2 입니다.

- RAID5

Data의 빠른 입출력이 가능하도록 여러 disk에 분산하여 저장합니다. 패리티 비트를 이용하여 Disk의 장애가 있을 때 복구를 합니다. 최소 3개의 디스크가 필요하며 실가용량은 Disk용량 × (N-1)/N 입니다.

4. I/O 인터페이스 (PCI Express)

PCI는 Peripheral Component Interconnect의 약자로, CPU와 I/O devices의 data 전송을 위한 인터페이스 규격입니다. PCI 표준은 1990년 인텔에 의해 도입되어 1955년에 업계에 널리 사용되었습니다. PCI 및 그 확장 표준은 900개 회사 이상의 협의체인 PCI-SIG (PCI Special Interest Group)에 의해 관리되고 있습니다. 크기는 작아졌고 IRQ공유, 플러그 앤 플레이 등의 새로운 기능도 추가되었습니다. 새로운 규격이 나올 때마다 거의 2배의 빠른 성능 향상을 가져왔습니다.

대략 발달 순서 별로 본다면 ISA, VESA → PCI → PCI-X → PCI-Express 정도 됩니다.

주요 동작 속도 및 데이터 버스 폭으로는,

33MHz/32-bit, 66MHz/32-bit, 66MHz/64-bit가 있으며, 그 대역폭은

$33\text{MHz} \times 32\text{-bit}(4\text{byte}) = \text{약 } 125\text{MByte/s}$ (손실 고려하여 계산치 차이)

$66\text{MHz} \times 32\text{-bit}(4\text{byte}) = \text{약 } 250\text{MByte/s}$, $66\text{MHz} \times 64\text{-bit}(8\text{byte}) = \text{약 } 500\text{MByte/s}$ 입니다.

PCI-X는 그래픽 성능이 발전하면서 AGP가 탄생했듯이, 대용량을 필요로 하는 주변기기들(스토리지, 네트워크)을 위해 탄생한 버스방식이며 기본적으로 PCI의 슬롯 형태와 동일하고, PCI와 호환성도 유지하였습니다.

단지 동작 속도를 올린 것인데, 현재 PCI-X 1.0 규격은 133MHz, 64-bit로 전송 대역폭이 1GB/s에 이릅니다. 그리고 규격만 릴리즈한 것으로 PCI-X 2.0이 나왔는데, 여기서는 DDR과 QDR을 사용하여 대역폭을 2배, 혹은 4배로 늘려서 2GB/s, 4GB/s까지 대역폭을 확장하였습니다. 하지만 서버용으로는 PCI-X 2.0이 아니라 다른 규격인 PCI-Express가 사용되기 시작하였습니다.

PCI-Express 방식은 기존까지의 병렬 버스 방식에서 탈피하여 스위치 패브릭(switched fabric) 구조와 점 대 점(point-to-point) 패킷 연결망 개념을 도입한 I/O 버스로 속도의 문제를 개선한 방식입니다. PCIe 장치와 시스템 간의 연결은 'link', 이는 'lane'이라고 하는 양방향성, 직렬(1bit) point-to-point 연결구조로 되어 있습니다. 한 lane은 동시에 초당 250MB의 데이터를 양방향으로 전송하는 성능을 가지고 있습니다. 한 link는 동시에 여러 lane을 사용할 수 있고, 모든 PCIe 규격의 link는 'x1'으로 표시하는 하나의 lane을 최소한 지원 해야 합니다. 더 높은 대역폭을 지원하기 위해 PCIe 장치 및 시스템은 동시에 여러 lane을 사용하는 link를 지원할 수 있습니다. 예를 들어 'x16'은 16 lane을 사용하는 link입니다. x1, x4, x8, 및 x16 link의 카드와 슬롯이 통용됩니다.

현재 x16 link를 사용하는 장치는 그래픽 카드가 유일합니다. PCI-Express 2.0은 lane 당 500MB/s의 전송속도를 가집니다.

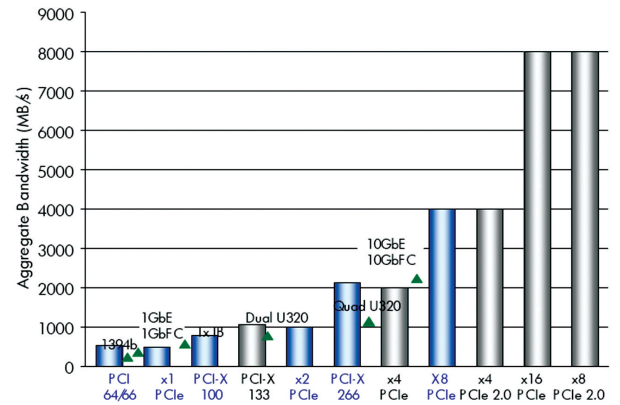


그림 10 PCI 슬롯 전송속도 비교

이상으로 서버에 사용되는 기술들을 간략히 살펴보았습니다. 다음 호에는 각 벤더사의 대표적인 제품의 스펙, 특징 등을 살펴해보도록 하겠습니다.

참조 : [Technology implementation in HP Proliant G7 Server](#)

[DDR3 Memory technology](#)

[표준제안서](#)