Server Concepts

Jordan Antonov
Power Servers L1

IBM

Jordan Antonov

- Education :

  TU Sofia – Bachelor's Degree,
  Machine Engineering 2005-2010

- Experience :

• Hewlett Packard Enterprise , 2011-2015
- IBM i / i Series / AS-400 Software Support Engineer

• IBM , 2015-now
- Power System Hardware Engineer – Subject Matter Expert

Server Concepts - Agenda

- Purposes
- Server vs PC
- History
- Backward Compatibility
- RAS
- Redundancy – Redundant and non redundant components
- What is a Rack – U
- Hardware
  - Backplane Concept
  - Service processor concept – Monitoring / Hardware controlling / Error handling and reporting
  - Hardware Deconfiguration / FW
  - CPU : RISC / CISC / SMP / Cache / SCM & DCM / Recovery
  - Memory – Error correction
  - Platform Hardware – PSU / AMD / Voltage Regulators / Planars
  - I/O Resources – I/O expansion units
  - Storage - Direct Attached / FC / Tape
  - High Availability – RAID / Mirroring / Controller duplication / Looping
  - Minimum Configuration / CoD / MIPS
  - Locations
- Types of servers
- HyperVisors

What is a Server?

- Server : In computing, a server is a computer program or a device that provides functionality for other programs or devices, called "clients".
- Server can be a Physical server/Virtual/Application to which multiple users are connected simultaneously and request/edit data.

Why We need a Server?

- To request and imput information and data.
- To change/edit and add data/information that can be access and used by others.
- To request any kind of service.
- To place work that has to be done without our interaction and attention.

Server : A server can be either a Physical system , Virtual system or even an application (Service) to which users can connect , interact with it , request information , change/edit information , save new data.

.

https://en.wikipedia.org/wiki/Server_(computing)

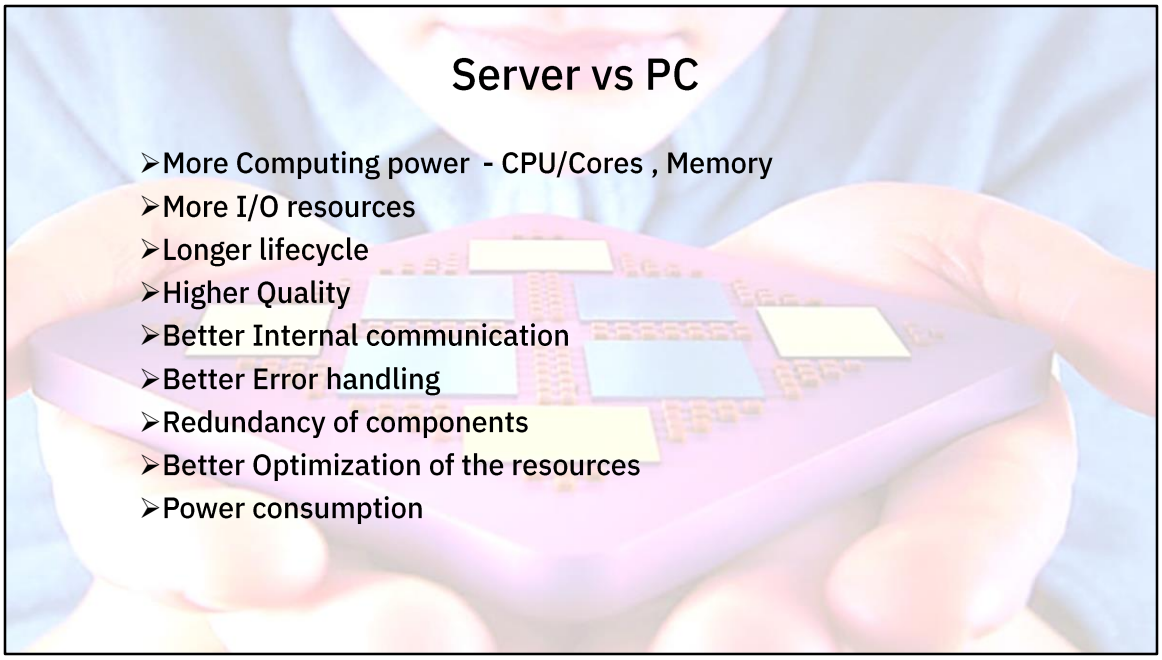
.

Physical servers are actually Computers with more computing power and I/O resources.

Virtual server. Environment that acts as a server , but is using virtual resources

Application server - An application server is a software framework that provides both facilities to create web applications and a server environment to run them. Application Server Frameworks contain a comprehensive service layer model. An application server acts as a set of components accessible to the software developer through an API defined by the platform itself
https://en.wikipedia.org/wiki/Application_server
Clusters - A server farm or server cluster is a collection of computer servers

maintained by an organization to supply server functionality far beyond the capability of a single device


Server vs PC
More Computing power - CPU/Cores , Memory
More I/O resources
Longer lifecycle
Higher Quality
Better Internal communication
Better Error handling
Redundancy of components
Better Optimization of the resources
Power consumption

History of servers

- In the beginning servers were used to store , manage and operate with data.
- Data was inserted with Punch Cards.
- Server was back then created from Input device , Storage Device , Printing Device , Consoles and CPU Node.
- There was no OS. All the instructions that were required were Punched.

https://www.youtube.com/watch?v=YXE6HjN8heg - Computer Punch Cards - Historical Overview

https://www.youtube.com/watch?v=YnnGbcM-H8c - 1964 IBM 029 Keypunch Card Punching Demonstration

https://www.youtube.com/channel/UCOyJD0RHtF_77_oAf5tT1nQ/featured - Computer History AP

https://www.youtube.com/watch?v=KG2M4ttzBnY - Punch Card Programming - Computerphile

https://www.youtube.com/watch?v=VQueCt114Gk - Ken Ross and Paul Laughton demo the IBM 1401

https://www.youtube.com/watch?v=V4kyTg9Cw8g&index=2&list=RDVQueCt114Gk - IBM System 360 Mainframe Computer History Archives 1964 SLT

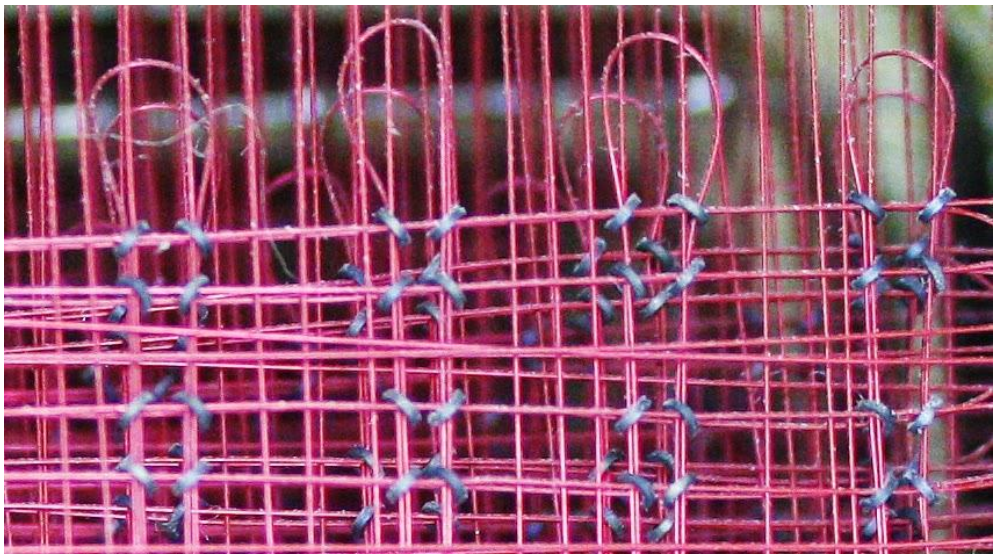
https://www.youtube.com/channel/UCHDr4RtxwA1KqKGwxgdK4Vg - Computer

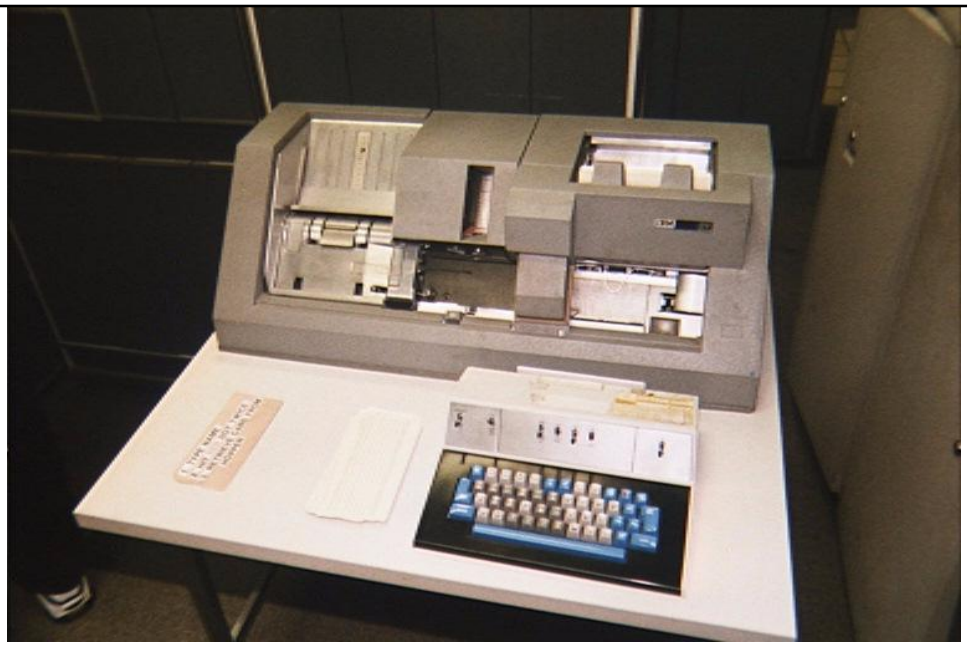
History Museum

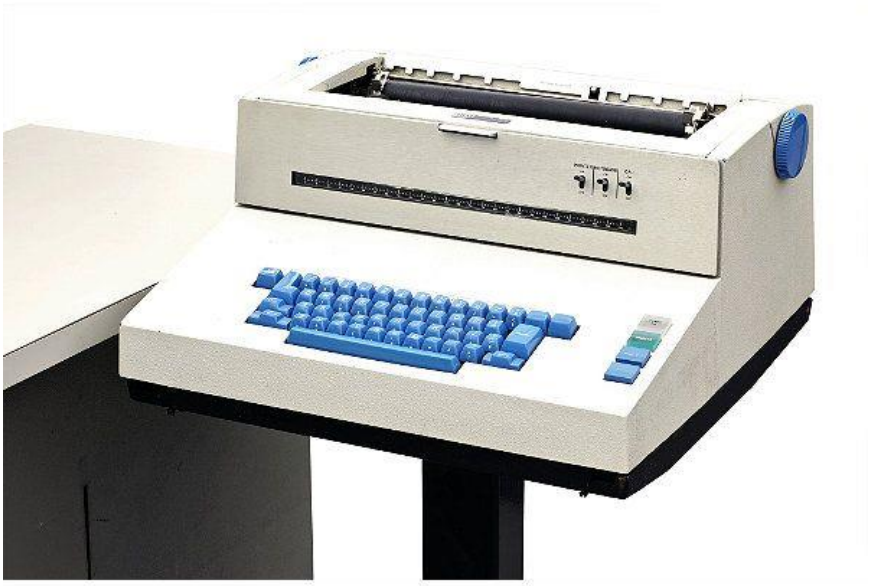
http://www.righto.com/2015/08/examining-core-memory-module-inside.html - Examining the core memory module inside a vintage IBM 1401 mainframe

Computer History Archive Project
Computer History Museum
Computerphile

Picutre – IBM Mainframe system







To give a software product a good chance of success, it has to be compatible with older versions of the same product. This is known as **backward compatibility.**

Backward compatibility is part of the general problem called **Legacy problem** which is a very famous term in computer industry.

It means that there is a history relevant to a certain product that has to be considered.

The legacy problem is not restricted to software. It is also present in hardware especially in mainframes where you could have a mainframe that has parts from 1990 and other parts from 1970.

RAS – Reliability , Availability and Serviceability

- Reliability – can be defined as the probability that a system will produce correct outputs up to some given time.
- Availability - means the probability that a system is operational at a given time, i.e. the amount of time a device is actually operating as the percentage of total time it should be operating.
- Serviceability - is the simplicity and speed with which a system can be repaired or maintained.
- Cold/Hot Swap

Reliability can be defined as the probability that a system will produce correct outputs up to some given time *t*.[5] Reliability is enhanced by features that help to avoid, detect and repair hardware faults. A reliable system does not silently continue and deliver results that include uncorrected corrupted data. Instead, it detects and, if possible, corrects the corruption, for example: by retrying an operation for transient (soft) or intermittent errors, or else, for uncorrectable errors, isolating the fault and reporting it to higher-level recovery mechanisms (which may failover to redundant replacement hardware, etc.), or else by halting the affected program or the entire system and reporting the corruption. Reliability can be characterized in terms of mean time between failures (MTBF), with reliability = exp(-t/MTBF).[5]

Availability means the probability that a system is operational at a given time, i.e. the amount of time a device is actually operating as the percentage of total time it should be operating. High-availability systems may report availability in terms of minutes or hours of downtime per year. Availability features allow the system to stay operational even when faults do occur. A highly available system would disable the malfunctioning portion and continue operating at a reduced capacity. In contrast, a less capable system might crash and become totally nonoperational. Availability is typically given as a percentage of the time a system is expected to be available, e.g.,

99.999 percent ("five nines").

Serviceability or *maintainability* is the simplicity and speed with which a system can be repaired or maintained; if the time to repair a failed system increases, then availability will decrease. Serviceability includes various methods of easily diagnosing the system when problems arise. Early detection of faults can decrease or avoid system downtime. For example, some enterprise systems can automatically call a service center (without human intervention) when the system experiences a system fault. The traditional focus has been on making the correct repairs with as little disruption to normal operations as possible

https://en.wikipedia.org/wiki/Reliability,_availability_and_serviceability

Redundancy of components

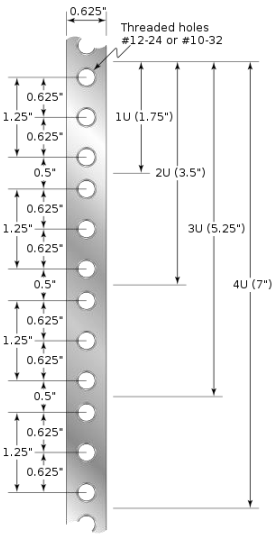
- Every server hardware redundancy is presented.
- Redundancy concepts that IBM uses is N+1.
- Redundancy depends on the type of server
- Platform redundant hardware:

  PSU , Fans , FSP's , Voltage Regulator Modules*, CPU and Memory IF available

- From I/O point of view - I/O controllers (storage controllers , Communication controllers) , Disk drives , cables , SAS Expanders.
- Some parts are not redundant.

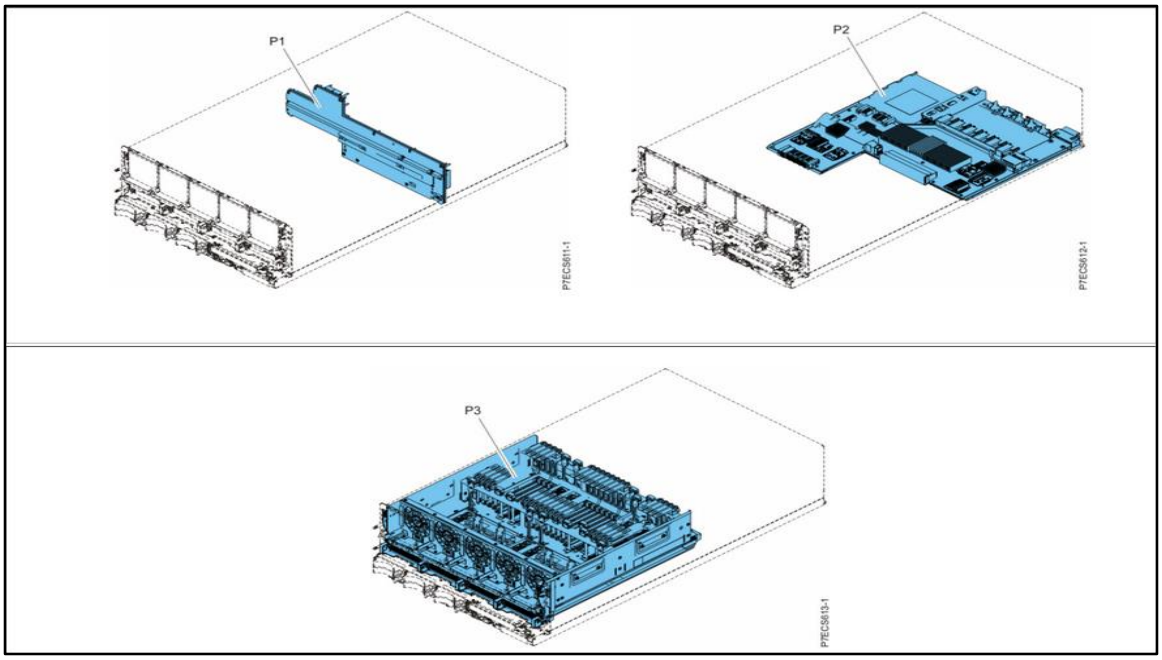
System Racks

- A rack unit (abbreviated U or RU) is a unit of measure defined as 44.50 millimetres (1.752 in)
- 19-inch and 23-inch rack frames are the most common racks.
- A typical full-size rack cage is 42U high, while equipment is typically 1U, 2U, 3U, or 4U high. This means that it holds just over 6 feet 1.8M of Equipment. A typical "half-height" rack is 18–22U, which is around 3 feet (0.91 m) high.
- Every Rack has
  - 2x PDU – power distribution Unit
  - Top on Rack Switch
  - Patch Pannel



Hardware

- Backplane Concept
- Service processor concept - – Monitoring / Hardware controlling / Error handling and reporting
- CPU : RISC / CISC / SMP / Cache / SCM & DCM
- Memory – Error correction
- Platform Hardware – PSU / AMD / Voltage Regulators / System Planar / Power and cooling
- I/O Resources – I/O expansion units etc
- Storage - Direct Attached / FC / Tape
- High Availability – RAID / Mirroring / Controller duplication / Looping
- Minimum Configuration / CoD / MIPS

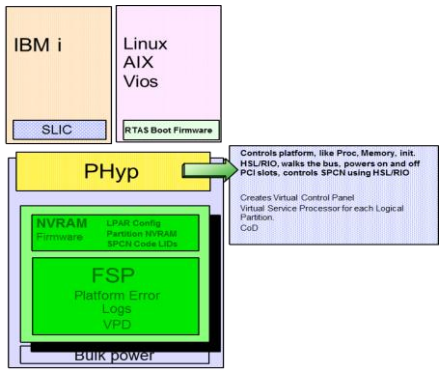


What we mean when we say Backplane or Planar
The Planar Concept
A planar/Backplane/board is a component that encapsulates an entire hardware subsystem. Example : CPU backplane – on this board are placed all the CPU's Memory Dimms and its cooling. I/O backplane is the board on which all I.O devices are attached and controlled.
A Backplane can also be used to connect 2 hardware subsystems.
It is presented in Middle Range and Enterprise servers , because of the complexity of the server.
Also the presence of a backplane , allows more effective and non-interuptable part replacement in some situations.
All connections between the backplanes are with the same speed as if we have one Motherboard with all of the components attached to it.

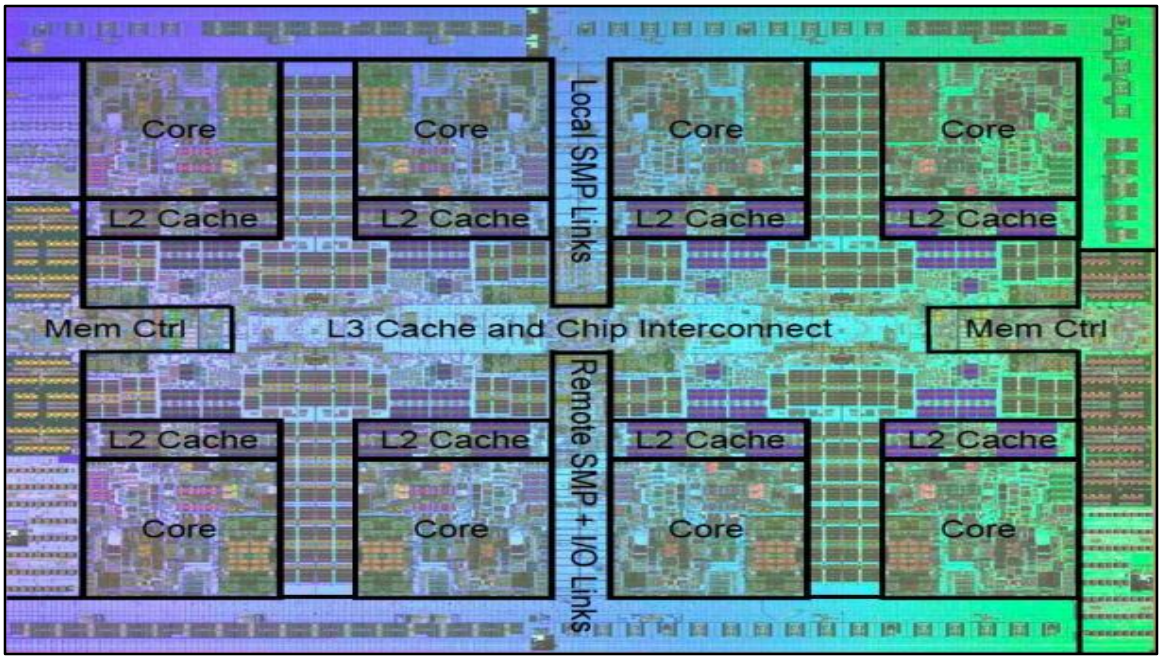
Picture – IBM POWER7 System 9117 Layout.

FSP – Flexible Service Processor

- Hardware that provides diagnostics, initialization, configuration, run-time error detection and correction.
- Error Reporting.
- FSP loads the Hypervisor for IBM machines.
- Own CPU /RAM / NVRAM / Linux Based OS
- System FW
- Hardware Deconfiguration



Consider the Flexible Service processor or only Service processor to be a small Computer inside the server. It has its own memory NVRAM and its own code (Linux Based small OS) Also it has a Flash Storage (NVRAM) that contains the Hypervisor code.
When Power is applied the FSP code – the mini Linux OS is loaded. This does not mean that the entire server is started , actually only the FSP has partially loaded.
The FSP provides Diagnostics , initialization error correction. Also the FSP is the one who inform us about an error. The System Microcode is also part of this guy.
One more feature of the FSP is the embedded redundancy for Microcode. In case the New version is not working as it should be , then the user can load the previous and well known version of a microcode

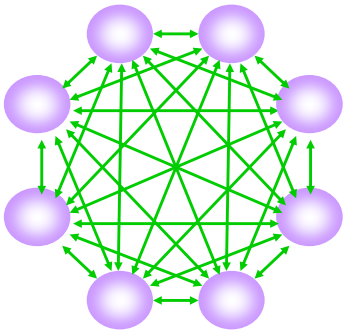


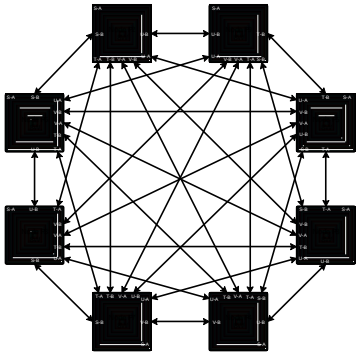
CPU – Servers can contain from 1 up to 48+ SCM.
SCM – Single Chip Module . DCM – Dual Chip Module.
Each Chip can contain numerous Cores.
Memory Controllers are embedded in the Chip.
SMP – Symmetric Multi Processing – All cores and Chip Modules are interconnected , meaning each core has full communication with all other cores in the chip and as well with other chips
Caching - CPU caches are small pools of memory that store information the CPU is most likely to need next.
Connection to the Input/Output devices.

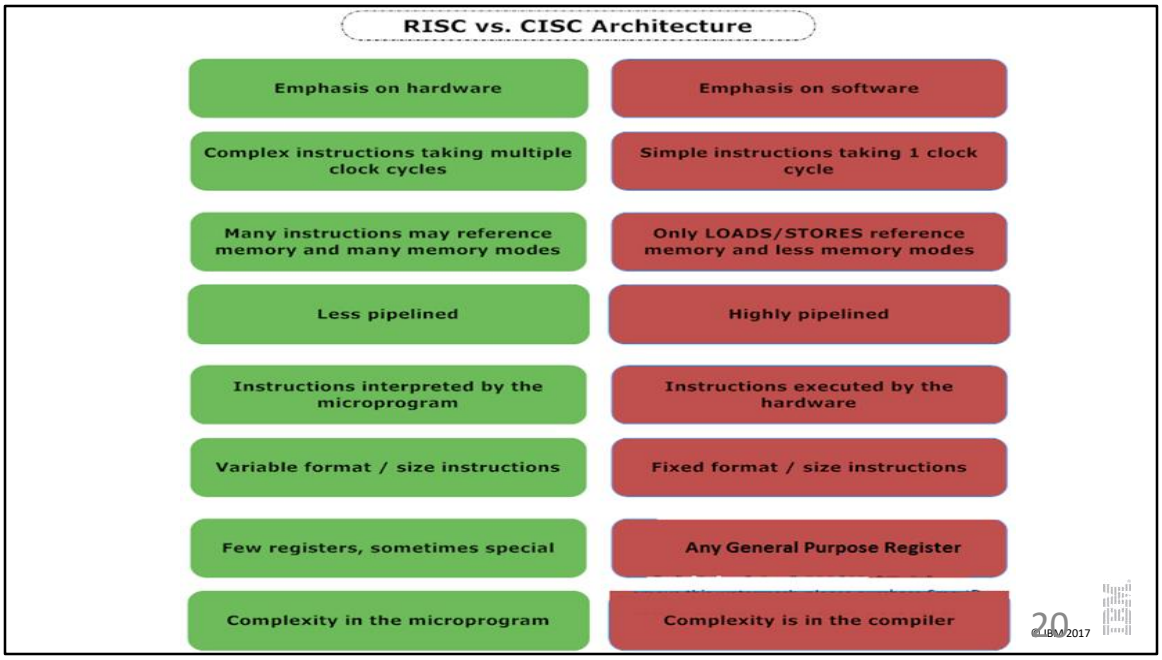
https://www.extremetech.com/extreme/188776-how-l1-and-l2-cpu-caches-work-and-why-theyre-an-essential-part-of-modern-chips
https://www.extremetech.com/computing/55662-top-tip-difference-between-l2-and-l3-cache

SMP :
Symmetric Multiprocessor connection

- All Cores Connection
- All CPU Modules connection





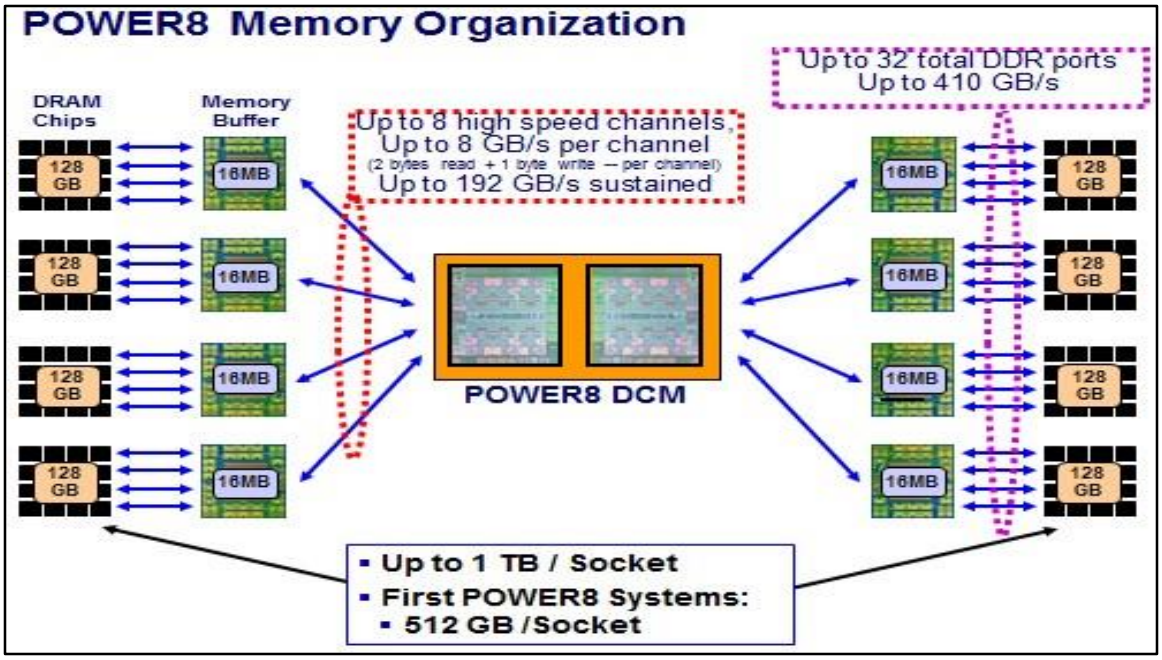

RISC vs. CISC Architecture
Emphasis on hardware
Emphasis on software
Complex instructions taking multiple clock cycles
Simple instructions taking 1 clock cycle
Many instructions may reference memory and many memory modes
Only LOADS/STORES reference memory and less memory modes
Less pipelined
Highly pipelined
Instructions interpreted by the microprogram
Instructions executed by the hardware
Variable format / size instructions
Fixed format / size instructions
Few registers, sometimes special
Any General Purpose Register
Complexity in the microprogram
Complexity is in the compiler
20
©IBM 2017

Server RAM – Random Access memory

- Higher Amount of memory vs PC.
- ECC
- Service Switch for fail DIMM identification

Memory Controllers are embedded into the POWER CPU Chip.
Memory subsystem has different layouts depending on the design and type of the server.
In every server there is a Service Switch that is used to identify faulty DIMM.

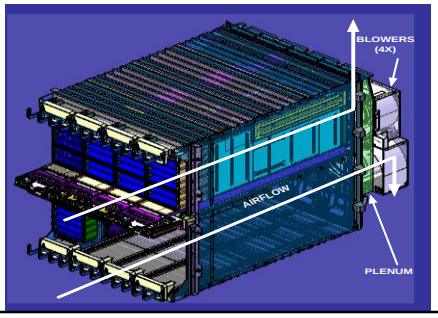
Picture – IBM CDIMM memory Module


POWER8 Memory Organization
Up to 32 total DDR ports
Up to 410 GB/s
DRAM Chips
Memory Buffer
Up to 8 high speed channels
Up to 8 GB/s per channel
(2 bytes read + 1 byte write – per channel)
Up to 192 GB/s sustained
128 GB
16MB
POWER8 DCM
▪ Up to 1 TB / Socket
▪ First POWER8 Systems:
▪ 512 GB /Socket

Picture – IBM

Power Subsystem and Cooling

- 2-4 PSU per CEC (computing node)
- Every PSU has a FAN
- All Fans and PSU are hot swappable
- Airflow passes thru the whole frame.
- Airflow moves from front to back



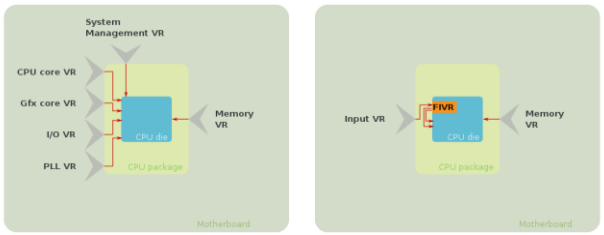
Server cooling is normally divided into subsystems . In IBM servers we have CPU and memory cooling and we have I/O resources cooling. The cooling is done using fans, however they are not just fans , they have some small logic inside them so the FSP to monitor their condition. In some Servers Water cooling is also presented.
Air direction is from front to back. The cold air enters the system from the front and exits the system from the back.
All Power supplies are also equipped with Fans , which are part of the whole system cooling.

Picture 1 - https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcTprEextqzu4nup3R2WoBJplgnpbRxu1IvJqiPuFkQxFPo3LQVacQ
Picture 2 – IBM POWER6 9119-FHA

Voltage Regulator Modules

- Converting +5 V or +12 V to a much lower voltage required by the CP
- Embedded on the CPU – FIVR
- Embedded on the System Planar
- Pluggable VRM's.
- VID – Voltage Identified

Fully Integrated Voltage Regulator (FIVR)



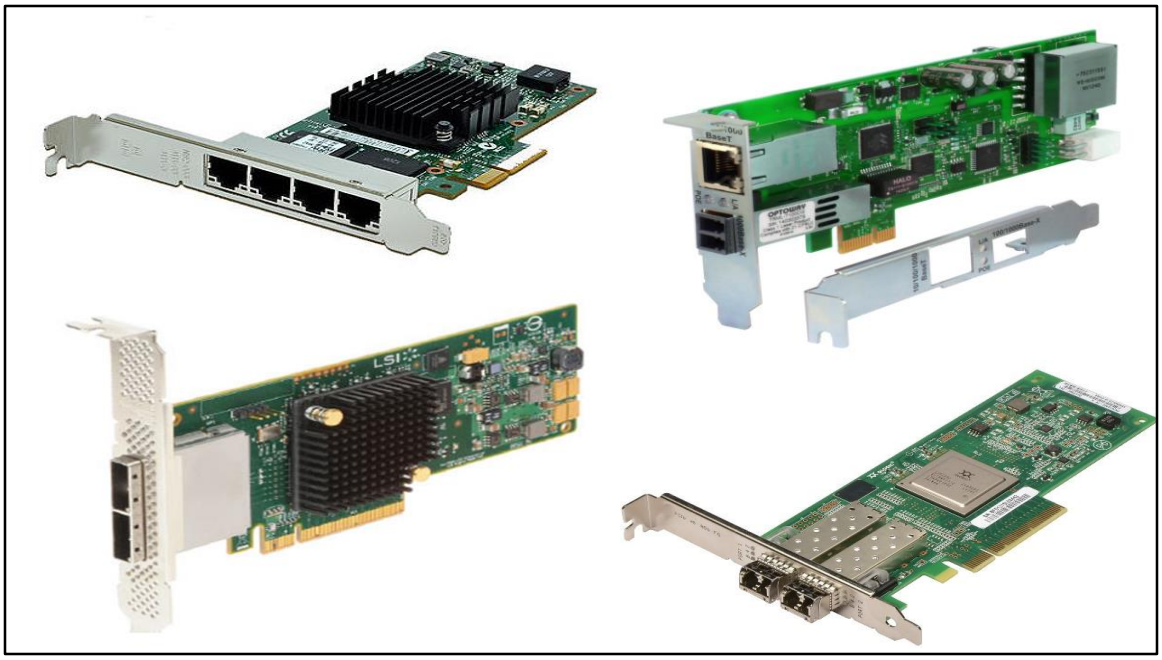
- the ATX PSU delivers the voltages 12V, 5V and 3.3V
- the Voltage Regulators on the motherboard transform 12V to all required voltages
- FIVR integrates legacy power delivery onto processor pkg/die
- better arch flexibility and finer-grain on-die processor delivery control

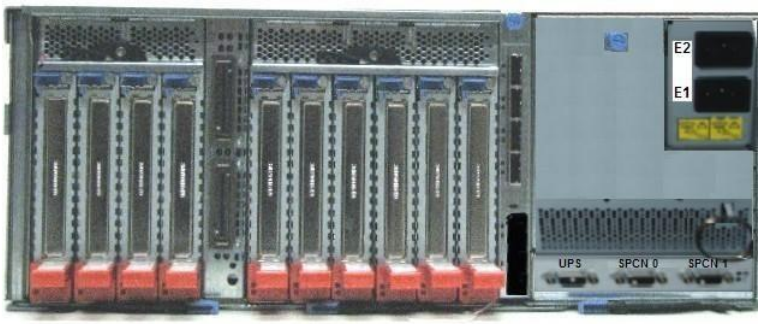
The correct supply voltage is communicated by the microprocessor to the VRM at startup via a number of bits called *VID* (voltage identification). In particular, the VRM initially provides a standard supply voltage to the VID logic, which is the part of the processor whose only aim is to then send the VID to the VRM. When the VRM has received the VID identifying the required supply voltage, it starts acting as a voltage regulator, providing the required constant voltage supply to the processor.

Input / Output Resources

- I/O – what is?
- What is I/O Controller and I/O processor
- Which resources are I/O devices
- I/O Virtualization
- MI/O adapters can work as one.
- I/O Expansion Units

An I/O resources can be all resources that interact with the server , are used as storage / communication line / Printers etc.
For example a Disk drive is an I/O resource because the server use the disk to put some data on it. Network card is also an I/O
Printers as being an output device, the mouse and keyboard , the display and so on
I/O processor. An I/O processor is a I/O device that controls I/O resources. The I/O processor is the person who stays between the CPU and the I/O resources.
In server environment we can merge I/O resources to work as one device. For example we can have 10 Network adapters that are configured to work as 1 .The same can be with Storage Controllers. 2 Storage controllers are can transmit data between the server and the storage devices twice as fast.
Input Output resources in one Server environment are operating at much higher speed that in normal PC's
As bigger one server is , as more I/O resources can be attached to this server.
I/O virtualization. This is a feature where 1 I/O resource can be used among couple of logical/Virtual servers. That way a customer can save money for hardware and to receive the same function, depending of his needs.

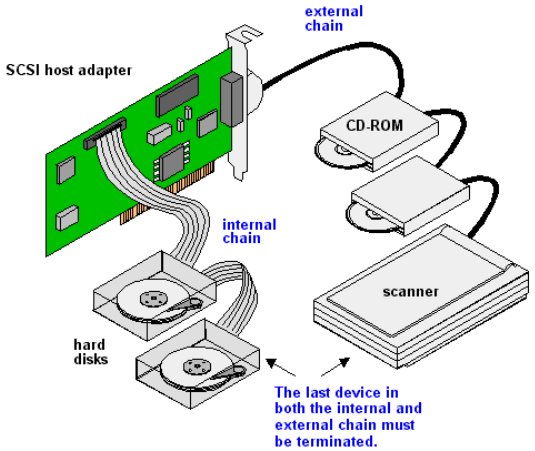

LSI



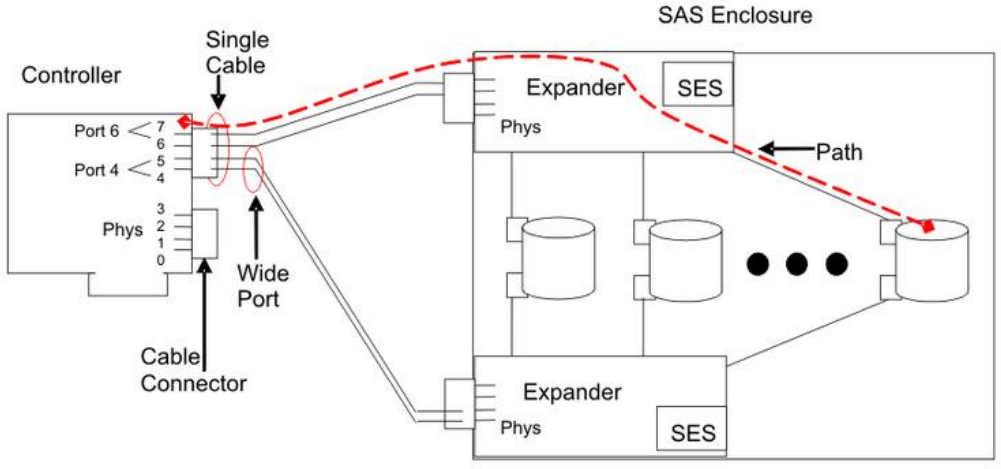
Storage

- DASD – Direct Attached Storage device
  SCSI - Small Computer System Interface
  SAS – Serial Attached SCSI
  SATA – Serial ATA
- External Storage – Fiber Channel attached Storage
  - External Storage solutions
  - Flash Systems
  - Tape Libraries
- Protection

SCSI - Small Computer System Interface

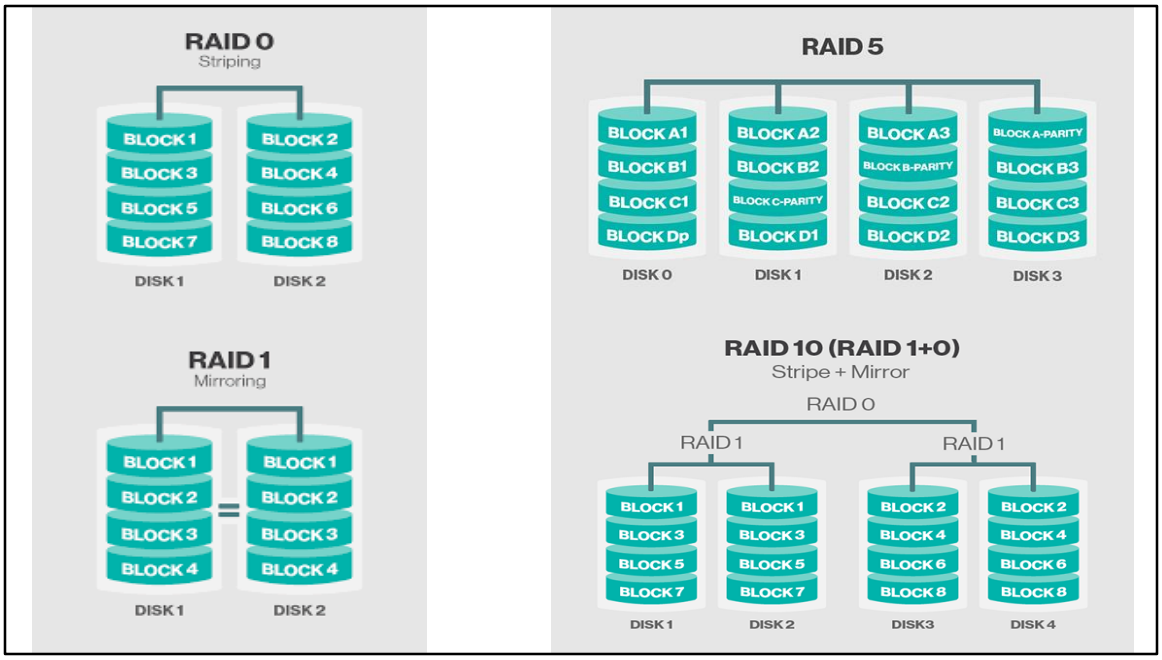


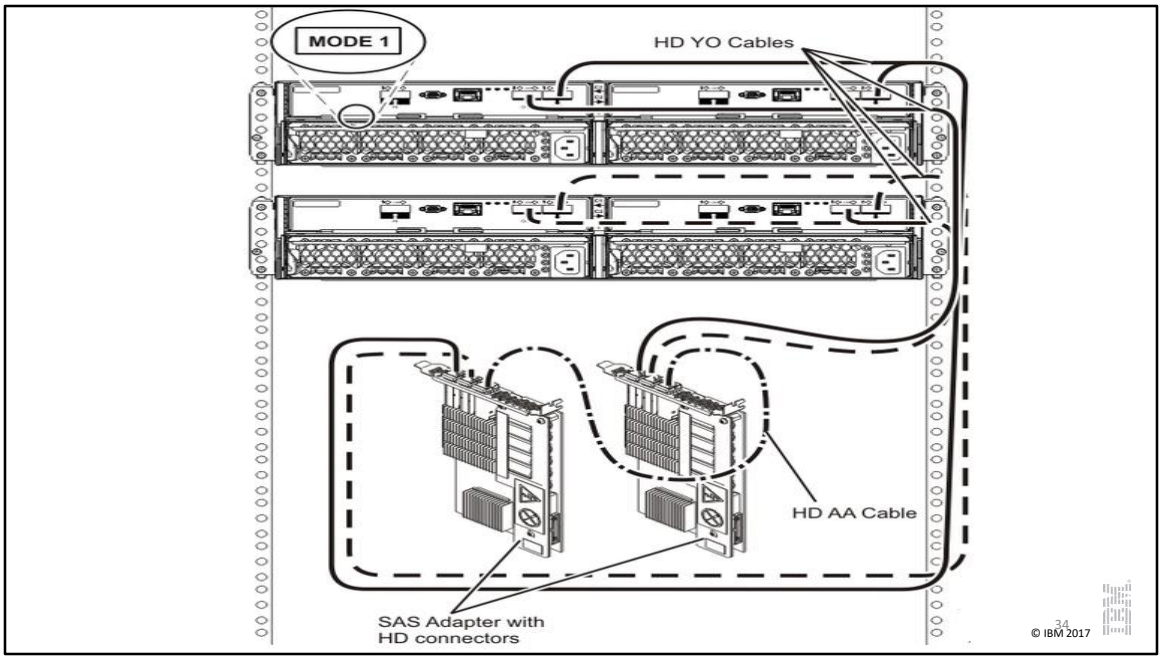
Serial Attached SCSI



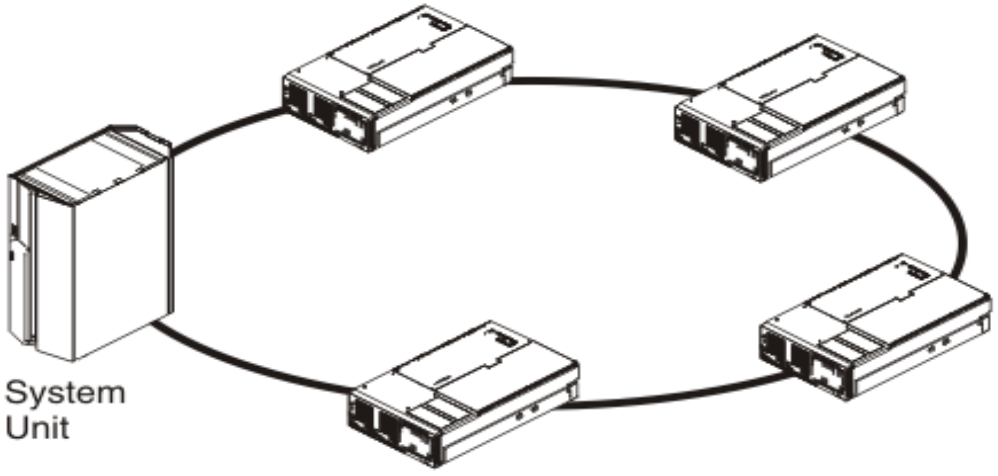
High Availability

- What is ?
- RAID – Redundant Array of Independent Disks , Caching
- Mirroring
- IOA duplication
- Looping
- Ethernet Failover

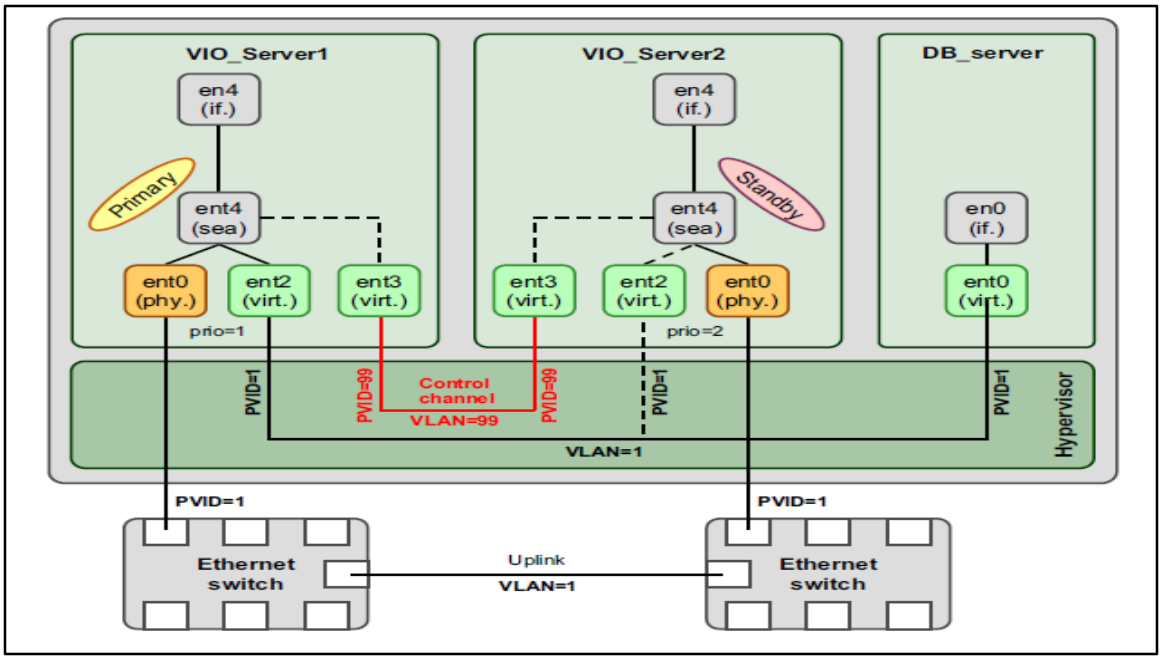

RAID 0
Striping
BLOCK 1
BLOCK 3
BLOCK 5
BLOCK 7
BLOCK 2
BLOCK 4
BLOCK 6
BLOCK 8
DISK 1
DISK 2
RAID 1
Mirroring
BLOCK 1
BLOCK 2
BLOCK 3
BLOCK 4
=
BLOCK 1
BLOCK 2
BLOCK 3
BLOCK 4
DISK 1
DISK 2
RAID 5
BLOCK A1
BLOCK B1
BLOCK C1
BLOCK Dp
BLOCK A2
BLOCK B2
BLOCK C-PARITY
BLOCK D1
BLOCK A3
BLOCK B-PARITY
BLOCK C2
BLOCK D2
BLOCK A-PARITY
BLOCK B3
BLOCK C3
BLOCK D3
DISK 0
DISK 1
DISK 2
DISK 3
RAID 10 (RAID 1+0)
Stripe + Mirror
RAID 0
RAID 1
RAID 1
BLOCK 1
BLOCK 3
BLOCK 5
BLOCK 7
BLOCK 1
BLOCK 3
BLOCK 5
BLOCK 7
BLOCK 2
BLOCK 4
BLOCK 6
BLOCK 8
BLOCK 2
BLOCK 4
BLOCK 6
BLOCK 8
DISK 1
DISK 2
DISK3
DISK 4


MODE 1
HD YO Cables
HD AA Cable
SAS Adapter with
HD connectors

Looping



AREAG502-0


VIO_Server1
en4
(if.)
Primary
ent4
(sea)
ent0
(phy.)
ent2
(virt.)
ent3
(virt.)
prio=1
VIO_Server2
en4
(if.)
Standby
ent4
(sea)
ent3
(virt.)
ent2
(virt.)
ent0
(phy.)
prio=2
DB_server
en0
(if.)
ent0
(virt.)
PVID=1
PVID=99
Control
channel
VLAN=99
PVID=99
PVID=1
PVID=1
VLAN=1
Hypervisor
PVID=1
PVID=1
Ethernet
switch
Uplink
VLAN=1
Ethernet
switch

Minimum Configuration and CoD

- Minimum Configuration – The minimum required Hardware needed for a CEC to start.
- CoD – Capacity on Demand – Service provided by IBM for additional resource activation

Server types

- IBM CEC – Central Electrical Complex
- Low End
- Middle Range
- High End or Enterprise

Low End – Single box or Single CEC

- Single CEC or Stand Alone
- All resources are attached to a single Backplane/systemplaner.
- Consolidate up to 2 CPU modules.
- I/O expansions available.
- Less redundant components .
- 1 to 4 U rack space is required when rack mounted
- Different Internal Storage Solutions

A rack unit (abbreviated U or RU) is a unit of measure defined as 44.50 millimetres (1.752 in). It is most frequently used as a measurement of the overall height of 19-inch and 23-inch rack frames, as well as the height of equipment that mounts in these frames, whereby the height of the frame or equipment is expressed as multiples of rack units. For example, a typical full-size rack cage is 42U high, while equipment is typically 1U, 2U, 3U, or 4U high.

https://en.wikipedia.org/wiki/Rack_unit

Pictures – IBM POWER7 Low end servers.

Middle range servers

- Modular Servers , 1 up to 4 CEC's.
- Up to 4 Chip Modules per CEC.
- More Memory and I/O resources can be accommodated .
- More I/O expansion drawers can be attached.
- More Redundant Components.
- Midrange servers are using the multiple Planar architecture.
- Only rack mounted.

Pictures – IBM POWER6 and POWER7 mid-range servers.

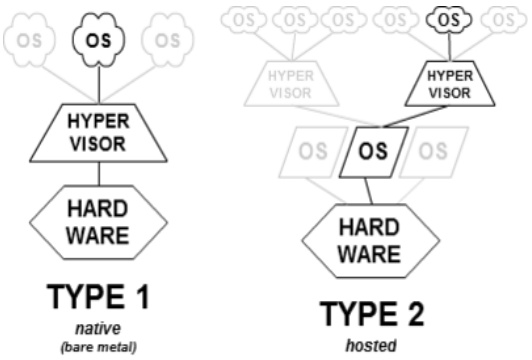
High End or Enterprise Systems

- Designed for High Speed computing.
- Full Redundancy.
- 24x7 Availability
- Again modular Design accommodating the backplane concept.

Picture – IBM Mainframe System Z

Hypervisors and Virtualization

- A hypervisor or virtual machine monitor (VMM) is computer software, firmware or hardware that creates and runs virtual machines
- Type-1, native or bare-metal hypervisors
- Type-2 or hosted hypervisors
- VIOS – Virtual I/O server – Provides Virtualization for I/O resources



Type-1, native or bare-metal hypervisors

These hypervisors run directly on the host's hardware to control the hardware and to manage guest operating systems. For this reason, they are sometimes called bare metal hypervisors. The first hypervisors, which IBM developed in the 1960s, were native hypervisors.[4] These included the test software SIMMON and the CP/CMS operating system (the predecessor of IBM's z/VM). Modern equivalents include Xen, Oracle VM Server for SPARC, Oracle VM Server for x86, Microsoft Hyper-V and VMware ESX/ESXi.

Type-2 or hosted hypervisors

These hypervisors run on a conventional operating system (OS) just as other computer programs do. A guest operating system runs as a process on the host. Type-2 hypervisors abstract guest operating systems from the host operating system. VMware Workstation, VMware Player, VirtualBox, Parallels Desktop for Mac and QEMU are examples of type-2 hypervisors.

https://en.wikipedia.org/wiki/Hypervisor