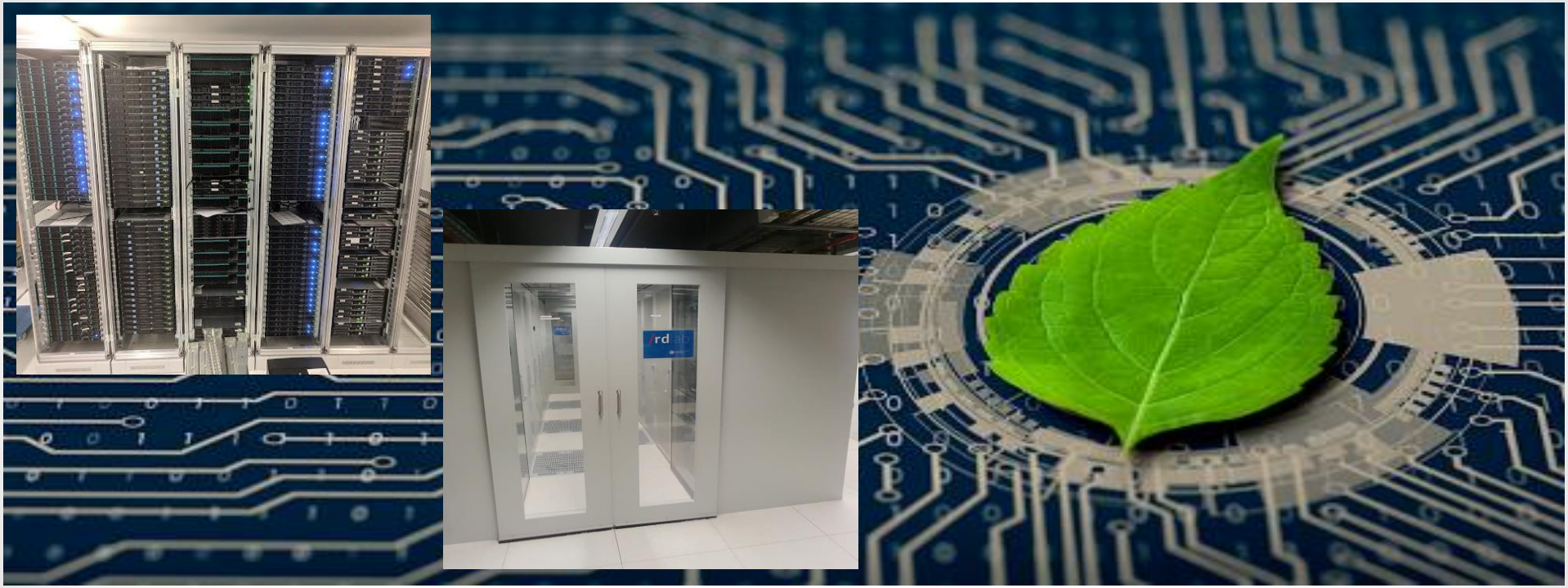
Efficient energy management for data center servers



17th SIG-NOC meeting - Paris, France
November 2022

Special Interest Group
Network Operations Centre

Gabriel Verdejo Álvarez
Computer Science Dept.
gabriel@cs.upc.edu
https://rdlab.cs.upc.edu

Index

First things first

- **Some context, FYI:**

  ➢ OVO Energy UK study "***Think before you thank***"

  16.433 Tonnes of carbon a year!!

  ➢ Outsourcing/external cloud?

  OK, but their carbon footprint is yours too!

- **Expectations and reality**

  ➢ **Frustration** = Expectation – Reality

A holistic approach

- **Efficiency management levels**

Physical

Servers, key components and network

Service

What is provided for users and the internal system staff

Architecture

How we model our facilities, the IT infrastructure and services interdependencies

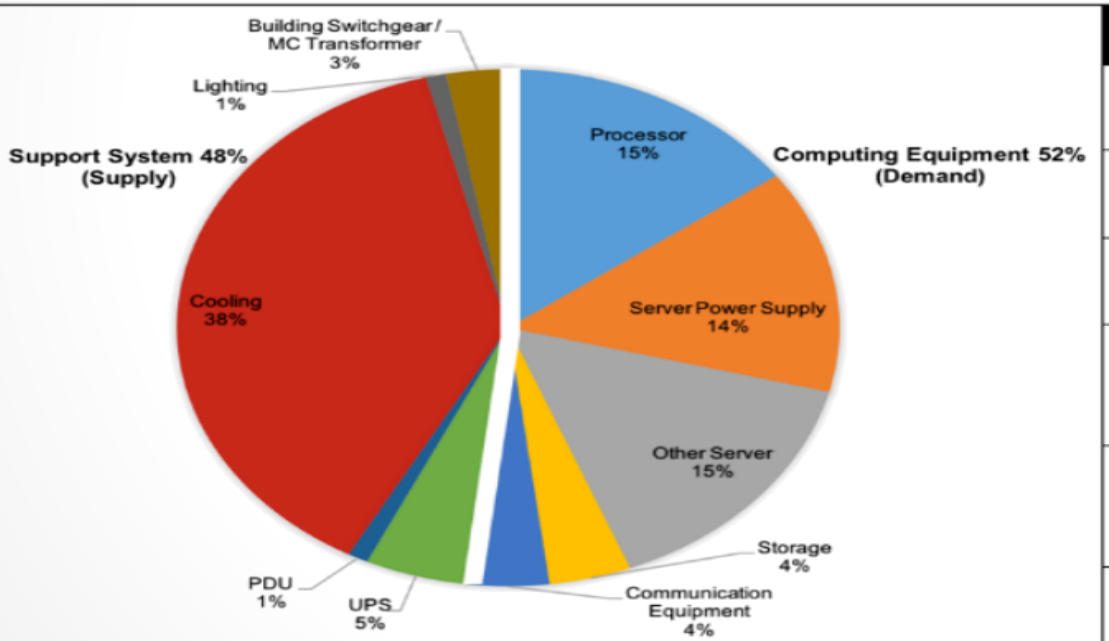
A holistic approach II

- **STRATEGY = Architecture + Service**

➢ Medium/long term → slow changes + change resistances

IT energy strategy ⊂ Institutional energy strategy

- **Third best option: The physical level**

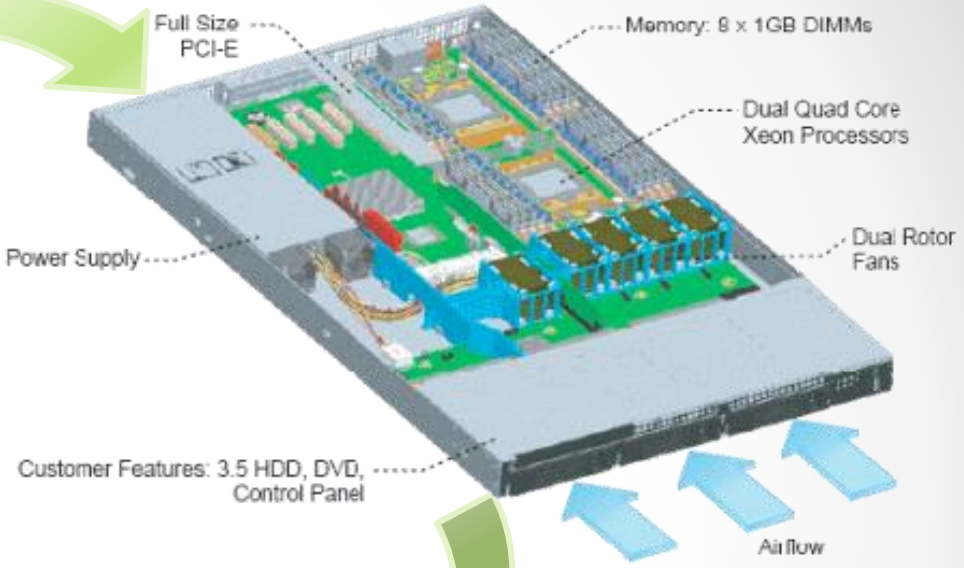


Category	Power Draw
Computing	588 kW
Lighting	10 kW
UPS and distribution losses	72 kW
Cooling power draw for computing and UPS losses	429 kW
Building switchgear/MV Transformer/other losses	28 kW
Total	1,127 kW

Figure 1. Analysis of a typical 465 m² data center (Emerson 2015).

We deal "only" with 50% of energy consumption!

The physical level



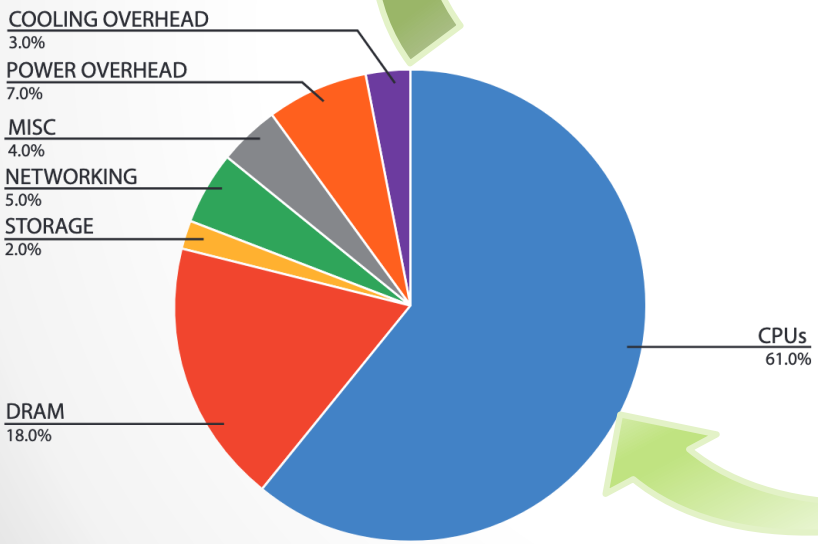
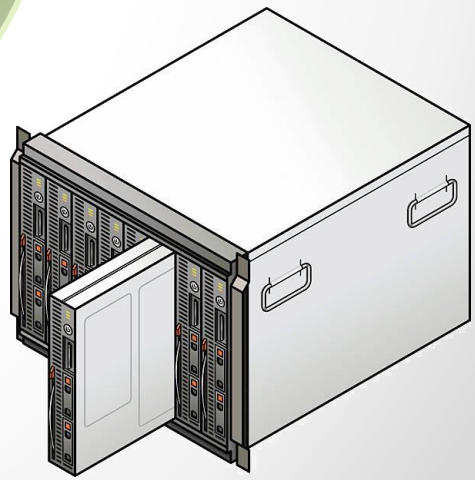


Figure 1.8: Approximate distribution of peak power usage by hardware subsystem in a modern data center using late 2017 generation servers. The figure assumes two-socket x86 servers and 12 DIMMs per server, and an average utilization of 80%.

Rack servers vs blades



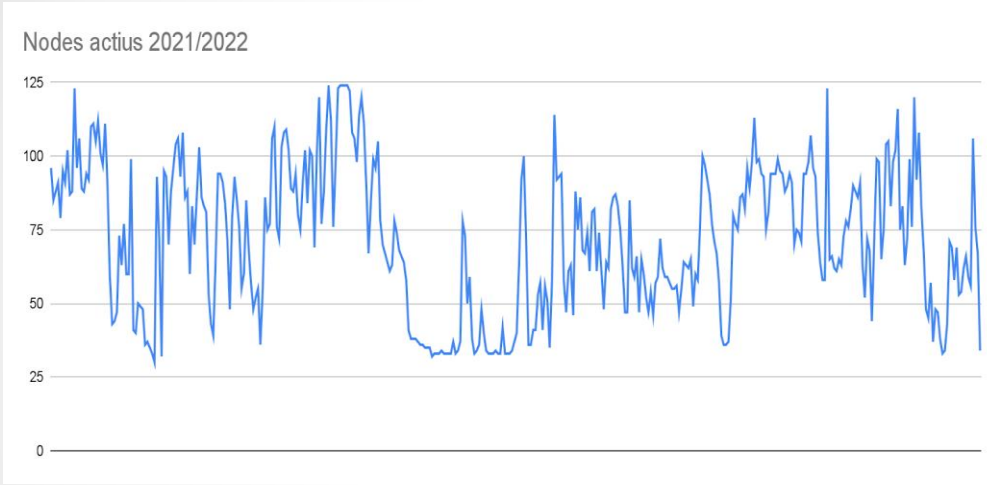
"…because without (proper) data, you are just another person with an opinion."

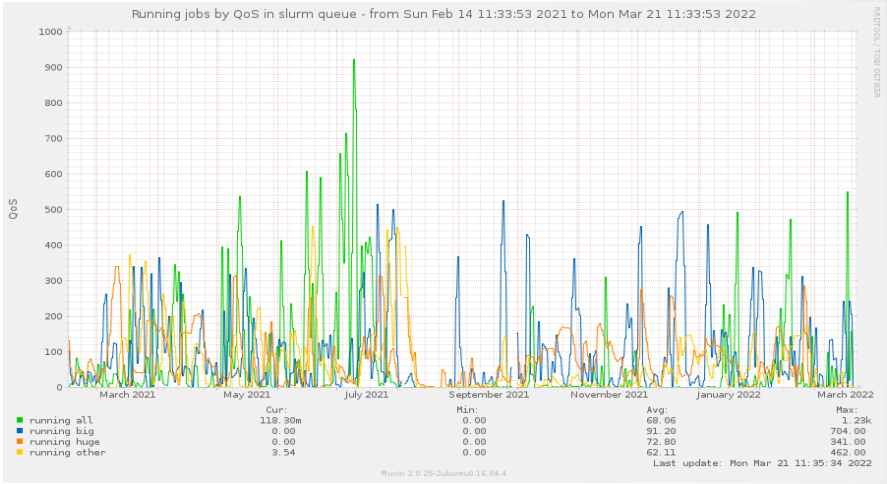
Featuring William Edwards Deming

MythBusters: idle vs power off

- **HPC system: 122 nodes and ~2.500.000 jobs per year at DAC**

➢ Slurm queue system power management → self node power on/off when needed





- **Power consumption**

➢ 1 x 460 W PSU
➢ 2 x CPU 6C/12T
➢ 16 x 8 GB ECC = 128 GB
➢ 1 x 1 TByte HDD SATA Disk

(*) 0,142 € KWh

State	Power	KWh	€ / hour*	€ / day 100 nodes	€ / year 100 nodes
Unplugged	**0 W**	**0**	**0**	**0**	**0**
Connected → power **off**	**10 W**	**0,01**	**0,00142705**	**3,4**	**1241**
Connected → **idle**	**81 W**	**0,08**	**0,0114164**	**27,4**	**10001**
Stress 1 CPU	103 W	0,1	0,0142705	34,2	12483
Stress 1 IO	91 W	0,09	0,01284345	30,8	11242
Stress 10 CPU	151 W	0,15	0,02140575	51,4	18761
Stress 10 IO	118 W	0,12	0,0171246	41,1	15002
Stress 24 CPU	160 W	0,16	0,0228328	54,8	20002
Stress 24xCPU+IO+HDD	182 W	0,18	0,0256869	61,6	22484

MythBusters II: hdd vs ssd

2015 server:

- 2 x Intel E5-2430 (12C/24T)
- 1 x SATA 1TBytes
- 32 GB RAM

# Disks	Type	State	Model	Power (MAX)
0	-	Connected → **idle**	-	55.3w
1	**HDD**	Connected → **idle**	Seagate Constellation 1Tbyte 7.2k rpm	62.2w (**+12%**)
1	**HDD**	I/O Stress	Seagate Constellation 1Tbyte 7.2k rpm	71.5w (**+29%**)
1	**SSD**	Connected → **idle**	Samsung 860 EVO 1TB	58.0w (**+5%**)
1	**SSD**	I/O Stress	Samsung 860 EVO 1TB	66.3w (**+19%**)

MythBusters III: old servers lifespan

2010 server:

- 2 x Intel Xeon 5120 (4C)
- 2 x SATA 500GBytes
- 2 x PSU 460w
- 4 GB RAM
  (RAID* hardware controller+ cache)

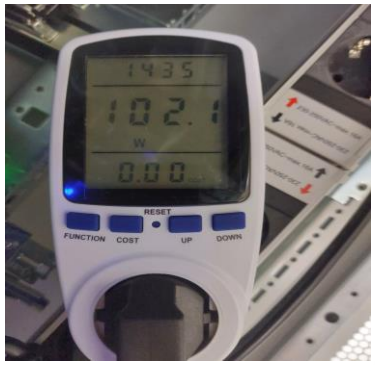
# Disks	Type	State	Model	Power (MAX)
2	**HDD**	Connected → **idle**	Western Digital 500GB RE3 7.2k rpm	**169w**
2	**HDD**	I/O Stress	Western Digital 500GB RE3 7.2k rpm	171.8w (**+2%**)*

"old" server power consumption is +2.5 times a 2015 server power consumption!!!

MythBusters IV: dual power supply

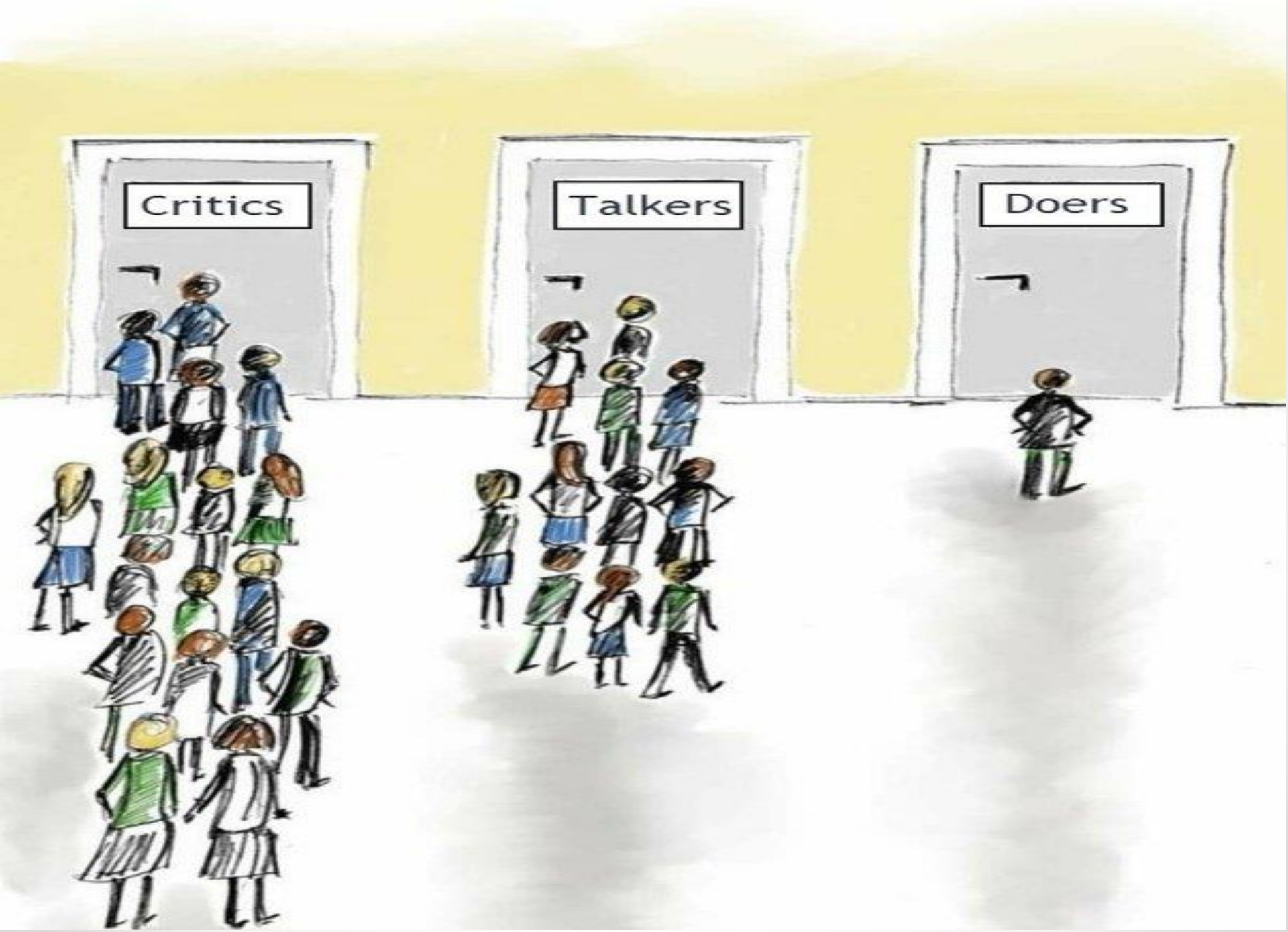
2020 Server:

2 x Intel Xeon 4214R
2 x PSU 500w
1 x HDD SATA 1TByte
64 GB RAM



PSU-1	PSU-2	State	Power (MAX)
Connected	**Unplugged**	Connected → power **off**	**17.5w**
Connected	Connected	Connected → power **off**	23w (**+31%**)
Connected	**Unplugged**	Connected → **idle**	**102.1w**
Connected	Connected	Connected → **idle**	110.5w (**+8%**)

Physical vs Architecture/Service (logical) level redundancy!!!


Critics
Talkers
Doers

How to be a doer in 3 lessons: I

1. Fuente de alimentación

Seleccionad una fuente de alimentación que proporcione una potencia inferior al 125% del consumo máximo del sumatorio de componentes del servidor (CPU+RAM+Discos+Placa base) y que tenga eficiencia dentro de la norma 80Plus. Evitad, si es posible, fuentes de alimentación redundantes. Se recomienda buscar la flexibilidad a nivel de servicio y arquitectura TIC, a excepción de casos de sistemas de alta disponibilidad (HA) o equipos críticos.

Puntuación: 0 para las 80 Plus y añadid 2 puntos adicionales para cada mejora de la categoría energética.

Referencia: https://en.wikipedia.org/wiki/80_Plus

2. Procesador

- Servidores genéricos : Seleccionad la CPU dentro de la familia compatible con el socket de la placa base que tenga menor consumo y que proporcione los cores/threads requeridos. Son preferibles los sistemas con refrigeración pasiva versus los tradicionales con ventiladores. Recordad que generalmente es necesario configurar/habilitar los estados de bajo consumo energético a la BIOS.
- Servidores HPC o específicos (GPUs…): Según necesidades.

CM = Consumo Máximo = Máximo(consumo en watts de todos los procesadores ofertados)
Cm = Consumo mínimo = mínimo(consumo en watts de todos los procesadores ofertados)
VE = Varianza Energética = CM - Cm

Puntuación: 10 * ((CM - Consumo procesador) / VE)

Referencia para CPUs Intel: https://ark.intel.com/content/www/us/en/ark.html Referencia para CPUs AMD: https://www.amd.com/en/processors

NOTA: Son preferibles sistemas con 2 o más procesadores físicos en un mismo equipo, que servidores múltiples desde el punto de vista de ahorro energético para propósito general.

3. RAM

- Servidores con procesadores de equipos portátiles o bajo consumo: 2 DIMMs por procesador físico. Servidores con procesadores para servicios genéricos: 4 DIMMs por procesador físico.
- Servidores con procesadores para equipos HPC/específicos: Según el fabricante, arquitectura del procesador (dual channel, quad channel…) y número de CPU sockets.

Puntuación: 10 – Valor Absoluto (#DIMMs recomendados para la gama del procesador - #DIMMs ofertados)

NOTA: En el caso de ofertar configuraciones con menos DIMMs de los recomendados por el fabricante del procesador, se puntuará con 0 puntos por ineficiente.

4. Disco

- Servidores genéricos: Seleccionad preferentemente discos SSD hasta los 2TBytes. Tened en cuenta la activación en el sistema operativo de la gestión SMART de los discos y TRIM si es necesario.
- Servidores de disco/HPC/específicos: Según necesidades.

Puntuación: 10 * (#discos SSD ofertados / #discos totales ofertados)

Referencia: https://en.wikipedia.org/wiki/S.M.A.R.T.
Referencia: https://en.wikipedia.org/wiki/Trim_(computing)

Fórmula Puntuación Energética = Procesador * 0.4 + FA * 0.3 + RAM * 0.1 + DISC * 0.2

How to be a doer in 3 lessons: II

- **The 10 – 10 – 10 Challenge:**

  ➢ **10% Reduction of electrical power in data centers servers**

  - Measurable (benchmarking)

  - Real impact

  ➢ **10% Of electrical budget investment**

  - Self-financing

  - Continuous **R**eturn **O**n **I**nvestment

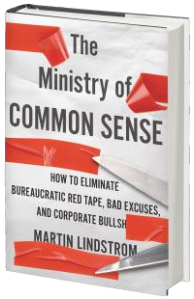
  ➢ **10 Months for execution time**

  - Finite time (vs multi-year)

  - Procurement period

CAMPUS UPC SOSTENIBLE 2030

How to be a doer in 3 lessons: III



Spending vs **Investing**
Buy vs **Use**
Local vs **Institutional Consortium**

References

[www1] The Problem of Power Consumption in Servers
https://www.infoq.com/articles/power-consumption-servers/
https://www.ovoenergy.com/ovo-newsroom/press-releases/2019/november/think-before-you-thank-if-every-brit-sent-one-less-thank-you-email-a-day-we-would-save-16433-tonnes-of-carbon-a-year-the-same-as-81152-flights-to-madrid

[www2] The Datacenter as a computer, 3rd edition
https://www.morganclaypool.com/doi/pdfplus/10.2200/S00874ED3V01Y201809CAC046

[www3] Energy efficiency insight into small and medium data centres
https://www.eceee.org/library/conference_proceedings/eceee_Summer_Studies/2017/7-appliances-products-lighting-and-ict/energy-efficiency-insight-into-small-and-medium-data-centers-a-comparative-analysis-based-on-a-survey/2017/7-086-17_Vasques.pdf/
https://www.serverstack.in/2019/01/19/difference-between-rack-servers-and-blade-servers/

[www4]. The Datacenter of the future (Dell)
https://education.dellemc.com/content/dam/dell-emc/documents/en-us/2021KS_Ganesh-The_Data_center_of_the_Future.pdf

[www5] Breakdown of power consumption in servers (Emerson 2015)
https://www.researchgate.net/figure/Breakdown-of-power-consumption-in-servers-Emerson-2015_fig4_329278608

[www6] Thermal design power (TDP)
https://en.wikipedia.org/wiki/Thermal_design_power

[www7]. Performance per watt (PPW) + RAM Consumption
https://en.wikipedia.org/wiki/Performance_per_watt
https://www.servethehome.com/ddr4-dimms-system-power-consumption-tested/

References II

[www9] PSU efficiency
https://en.wikipedia.org/wiki/80_Plus

[www10] Storage Area Network (SAN)
https://en.wikipedia.org/wiki/Storage_area_network

[www11] Network-Attached Storage (NAS)
https://en.wikipedia.org/wiki/Network-attached_storage

[www12] HDD vs SSD
http://www.amplicon.es/docs/SSD-vs-HDD-white-paper.pdf

[www13] HDD vs SSD power consumption
https://ssdsphere.com/ssd-power-consumption-comparison/

[www14] HDD vs SSD power consumption 2
https://www.anandtech.com/show/7682/the-wd-black2-review/9

[www15] HDD vs SDD lifetime and reliability
https://www.backblaze.com/blog/are-ssds-really-more-reliable-than-hard-drives/

[www16] Nvidia RTX 3090 power consumption analysis – GPU power consumption
https://www.profesionalreview.com/2020/10/23/nvidia-rtx-3090-cuatro-nvlink/
https://www.tomshardware.com/features/graphics-card-power-consumption-tested

References III

[www17] GPUs power consumption
https://www.igorslab.de/en/nvidia-geforce-rtx-3090-founders-edition-review-between-even-between-evaluate-and-common-decadence-if-price-is-not-all/16/

[www18] Slurm HPC Green guide
https://slurm.schedmd.com/power_save.html

[www19] CPU cores vs CPU cores + threads (AMD)
https://www.anandtech.com/show/16261/investigating-performance-of-multithreading-on-zen-3-and-amd-ryzen-5000/5

[www20] CPU cores vs CPU cores + threads (Intel)
https://www.intel.com/content/www/us/en/gaming/resources/hyper-threading.html

[www21] Hyperthread vs Multithread
https://www.pcpartguide.com/is-hyperthreading-worth-it/

[www22] Hyperthread vs Multithread 2
https://www.cgdirector.com/cpu-cores-vs-logical-processors-threads/

[www23] Martin Lindstrom – The ministry of common sense
https://www.martinlindstrom.com/books-by-martin-lindstrom/

[www24] Cream skimming – Prototype vs Minimum Viable Product
https://en.wikipedia.org/wiki/Cream_skimming
https://www.boldare.com/blog/first-version-of-your-app-mvp-vs-prototype/