CENTRES DE PROCESSAMENT DE DADES (CPD)

David Carrera
David López

2017-2018/Q1

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

DATA CENTER INFRASTRUCTURES

David López

V.1.15
Spring 2016

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Outline

Acknowledgements:

If I have seen further it is by standing on ye sholders of giants.

Sir Isaac Newton. Letter to Robert Hooke, circa 1676

Part of these slides have been borrowed from "The unstoppable transformation (revolution) of IT sector", a talk by Jordi Torres

Another part, from the course "Administración Avanzada de Sistemas Operativos" by Guillermo Pérez Trabado, from the Universidad de Málaga.

Key technical concepts

- Colocation centers
- Warehouse-scale Computers
- Servers and racks
- Building elements:
  - Raised floor
  - HVCA
  - Power / UPS
  - Physical access
  - Fire control
- Efficiency and cost
  - PUE
  - DCE
  - SPUE
  - TCO

References

- The Datacenter as a Computer: An introduction to Design of Warehouse-Scale Machines
  - Luiz André Barroso and Urs Hölzle. 2009 (free on the net)

Additional references

- Computer Architecture, A Quantitative Approach
  - J.L. Hennessy & D.A. Patterson
  - Chapter 6 of the 5th edition. 2012
- Enterprise Data Center Design and Methodology
  - Rob Snevely, Sun microsystems 2001

Type of Data Centers

Heterogeneous data centers (Your own. Colocation centers)

- Individual computing nodes / servers
- Individual / shared networks and storage systems
- Shared cooling, physical security and power distribution
- Can host servers from different clients

Homogeneous or *Warehouse-scale Computers* (Google terminology)

- Belong to a single organization
- (Relatively) homogeneous hardware & system software platform
- Run a small number of very large (data centered) applications
- Workloads are designed to tolerate a large number of component faults maintaining the service level of performance and availability
- Examples: Facebook, Google, E-bay, Amazon

High-Performance Computing Centers (or *Supercomputing Centers*)

- Belong to a single organization
- Homogeneous machines (usually clusters and grid)
- Intensive computational tasks
- Examples: Marenostrum, all Top500 computers facilities

Colocation centers

Equipment space and bandwidth are for rental

- Provide space, power, cooling, and physical security for server, storage and networking equipment of other firms
- Connect them to telecommunications and network service providers

Special features:

- You can rent cabinets or cages
  - Customers escorting, PIN, proximity card, biometric recognition (fingerprint, voice, …)
- Usually remote access, plus 7/24 technical staff available

Advantages:

- You pay power installed (kWh) + a rent depending on your space
- You safe the cost of building the infrastructure
- Fast to start, fast to scale

Inconvenient:

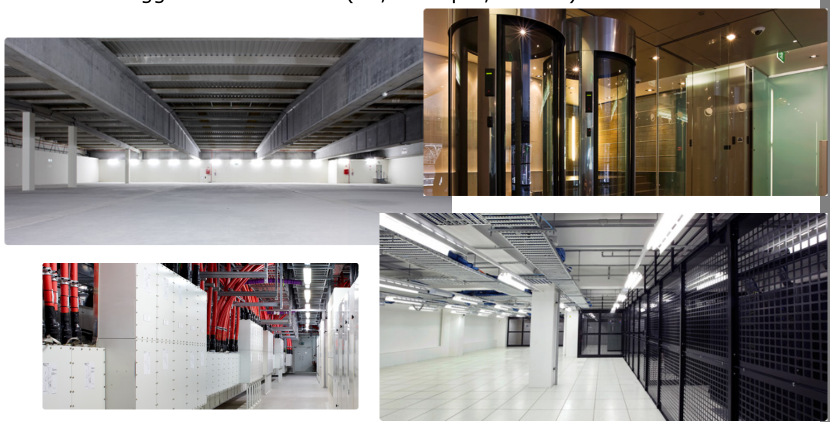
- Can you data be outside your building?

An example of Colocation Center: Global Switch

http://www.globalswitch.com/en

Nine centers:

- The biggest: East London (65,543 sq m, 45 MW)



Examples: Arsys (Logroño)

EL PAÍS

ECONOMÍA

ECONOMÍA EMPRESAS MERCADOS BOLSA MIS AHORROS VIVIENDA TECNOLOGÍA OPINI

En las tripas de la nube

- Arsys tiene en La Rioja un centro de datos de los más punteros del mundo

ANDRÉS GARCÍA DE LA RIVA | Logroño | 1 NOV 2014 - 22:48 CET

Archivado en: España Empresas Tecnología Economía Política Informática Ciencia Industria

Centro de datos de Arsys en Logroño

16 Arsys tiene uno de los centros de datos más innovadores y

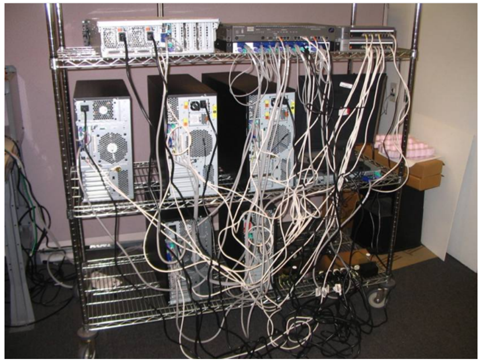
Common elements to all DCs

- Servers
- Building
  - Raised floor, suspended ceiling and cable trays
  - HVCA (Heating, Ventilation and Air Conditioning)
  - Power infrastructure
    - Including UPS – Uninterruptible Power Supply
    - Redundant systems, generators
  - Physical access control
  - Fire protection systems (active and passive)
  - Technical staff
    - And some room for these people
    - “Lights-out” or “dark” DCs are increasing
- Network + Storage

Servers

Tower servers

- It’s OK when you have an small DC
- A lot of space in case >5 servers
- Problems with wires



Servers

Racks

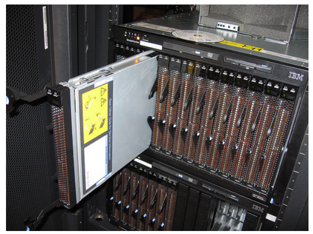
- There is a standard measure: rack unit or U
- To mount on 19-inch (data equipment and servers) or 23-inch racks (telecommunications equipment)
- 1U: 1.75 inches (44.45mm) high
- Typical rack: 42 U (up to 48U)
- Pieces of 1U, 2U, … and so

- Interconnection using local Ethernet
  - Inside the rack
- Storage out of the rack

Servers

Blades

- The servers share the power supply, fan and I/O connectors
- An enclosure is required to host blades
  - Blade center, or chassis
- High servers density (8 to 16 servers in 6U or 7U)
- For increasing processing capacity without storage
- Can be connected to a central storage system using Fiber Chanel

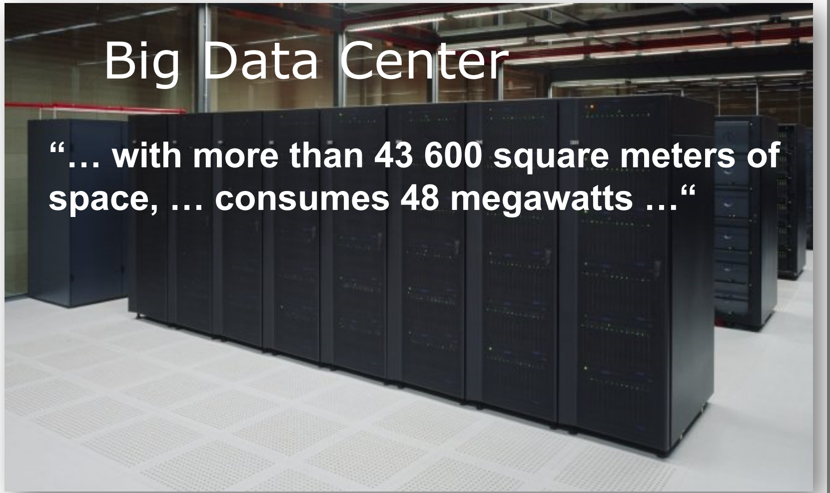


Servers

Mainframes

- A whole computer
- Business oriented
  - High reliability and security
  - I/O extensive facilities
  - Backward compatibility
- IBM dominates the market
  - Actually IBM System z

Building

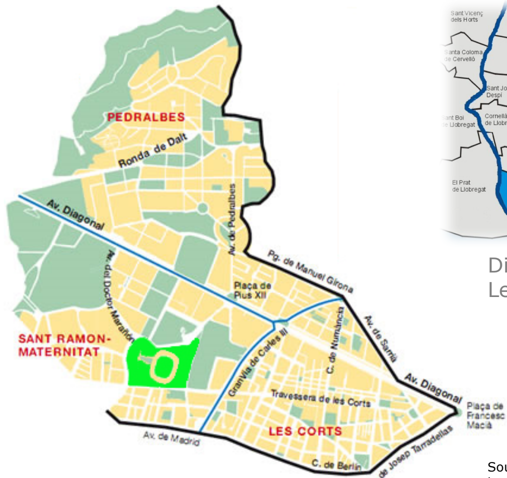


Source: *Tech Titans Building Boom By Randy H. Katz. IEEE Spectrum, February 2009*

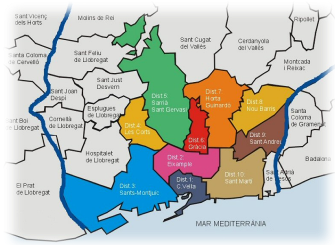
43.600 sq. meters

48 MW

Les Corts neighborhood (30,000 families)



Source:http://www.bcn.es/bcnbarris/ca/barrixbarri_districte4.html



Districts of Barcelona,
Les Corts coloured yellow.

Source:
http://en.wikipedia.org/wiki/Les_Corts,_Barcelona

Building a DC

Thermal and acoustically insulated

Geostrategic considerations:

- Good access to the net (Internet backbone optical fibers)
- Close to power generators
- Capex considerations (land cost, taxes power cost, average outside temperature,…)
- Physical communications
- Natural-disaster proof
  - Depends on the place (Tokyo vs. Barcelona)

Raised floor

- Provide an elevated structural floor above a solid substrate to create a hidden void for the passage of wiring, electrical supply and under floor air distribution
- Conditioned air is provided under the floor and dispersed upward into the room through regularly spaced diffuser tiles and blowers
- Wiring holes must be as adjusted as possible to the wire diameter

Raised floor

Diffuser tiles (cold air 20ºC)

Wiring entering the racks through holes beneath them

Raised floor details

- Minimum 50 mm high, can be raised enough for a person to crawl or even walk beneath
- Sometimes provide lighting and smoke detectors
- Rarely cleaned!
- Must be designed to support the weight of racks and people working in the DC (taking into account future equipment)

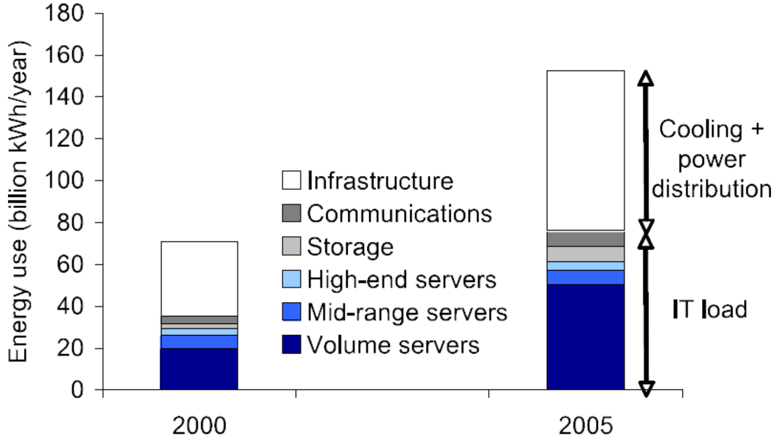
HVCA (Heating, Ventilation & Air Conditioning)

Basic ideas:

- The electrical power used by the electronic equipment is converted to heat
- Unless the heat is removed, it can result in hardware malfunction
- When designing the DC, you must know the peak HVAC consumption
  - Measured in kWh or BTU/h (British Thermal Units/hour)
  - 1 kWh = 3413 BTU/h
- Important idea: air flow
- Humidity is important!
- Not all points in the room have the same requirements!
- An important part of the final DC consumption

HVCA (Heating, Ventilation & Air Conditioning)

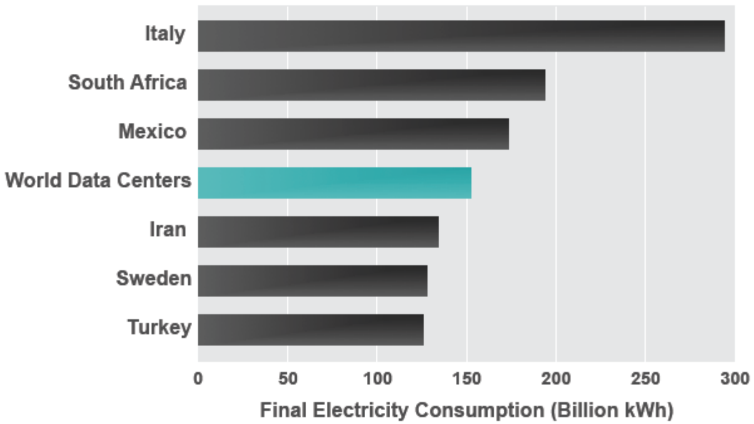
- A very important issue!



Source: Koomey, Jonathan. 2008. "Worldwide electricity used in data centers." Environmental Research Letters. vol. 3, no. 034008.

HVCA (Heating, Ventilation & Air Conditioning)

- How much is 152.000.000.000 kWh?



Source for country data in 2005: International Energy Agency, World Energy. Balances (2007 edition) (Koomey, Jonathan)

Some recommendations about HVAC

The maximum temperature must not exceed the limits given by the manufacturer of servers

- Between 20 º and 24 º C in the room
- More than 30 degrees reduces the life of electronic components
- Many units include temperature sensors with automatic shutoff

The optimum relative humidity should be around 45% to 55%

- Dry air influences the occurrence of electrostatic discharge (ESD)
- Excessive moisture condenses on internal components increasing energy consumption
- The relative humidity depends on air temperature

The flow must meet manufacturer's specifications

- The computer must be able to absorb a minimum amount of air per minute in m^3 / s
- The temperature inside the equipment depends on the external temperature and the air flow

The room must be clean of dust and particles

- Use anti-dust paint
- Vacuum cleaning before the DC start working
- Prevent dust

… but, there are not precise models and studies on the relationship between temperature, humidity and malfunction

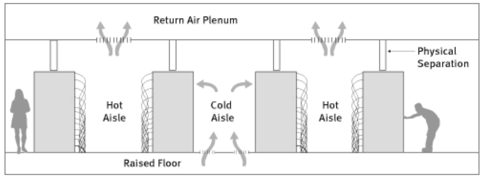
Cooling Systems

Small / domestic refrigeration equipment

- For small DC (with low power dissipation and low air flow)
- Relative humidity is not controlled
- It can not provide cold air flow where needed
  - It cools the room, not the servers

Raised floor room

- The coolers take hot air from the top of the room and push cold air into the raised floor
- The ground pressure is higher than in the room and cold air comes out where there is a hole or grid
- Racks draw air from the front and expel it from behind
  - Ground grids are placed in front of the racks
  - Two rows of racks must be faced to aspire cold air from the cold aisle
  - There are hot aisles
  - The flow is not perfectly predictable (but there are models)



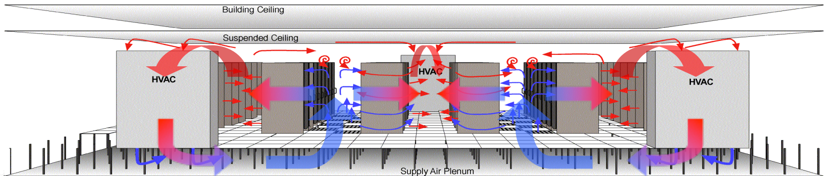
Source:hightech.lbl.gov

Hot aisle / cold aisle

Arrange the IT equipment so that all heat is exhaust into hot aisles, and all air intakes draw from cool aisles. Cool air is supplied only into the cold aisle, with return air being drawn directly from the hot aisle.

Good practices:

- Build rigid enclosures to fully separate the heat rejected from the rear of IT equipment from the cool air intakes on the front
- Use flexible strip curtains to improve the separation by blocking open space above the racks
- Blank unused rack positions to prevent hot air recirculation



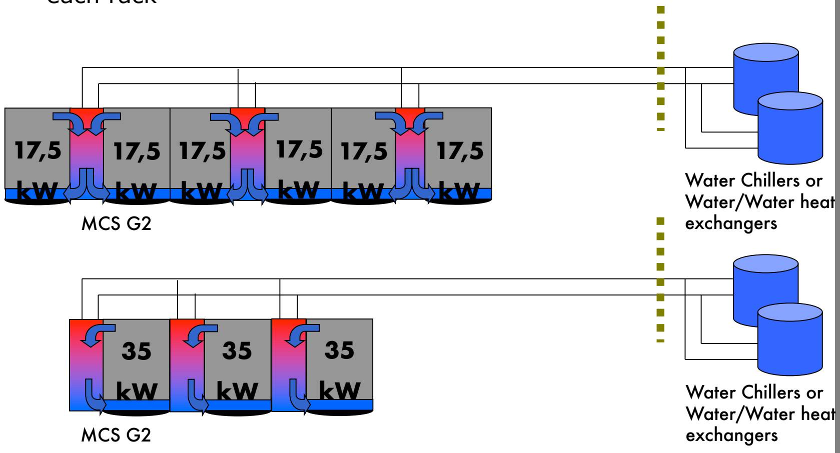
Source: Guillermo Pérez

Cooling within the rack

- To cool the milk, do you refrigerate the kitchen or just the inside of the refrigerator?
- Hot air does not escape the rack (door sealed)
  - Higher thermal efficiency (130% of capacity with the same amount of cooling)
  - If the cooling fails it can burn in minutes!
  - The rack doors open in case of failure of the refrigerator
  - You need to have emergency cooling on the room

Self-refrigerated racks

- Requires a cooling unit to provide cold water (using pipes) to each rack



Routing wiring and cables

Installed during construction phase
- Power wiring form *breakers* to *outlets* (under raised floor)
- Network cable from **PODs** (points of distribution) under raised floor to the network room

Available to connect by data center personnel
- Power wiring from outlets to racks
- Network cables form PODs to devices

Important idea:
- LABEL ALL YOUR CABLES!

UPS (Uninterruptable Power System)

Sistema d'alimentació ininterrompuda (SAI)

Three functions:
- Contains a transfer switch that chooses the active power unit (utility power or generator power)
  - A generator can take about 10-15 seconds to start and assume full rated load
  - Usually diesel generators (for 48 MW?)
  - Utility power usually have redundancy accepting two independent power sources (typically called "A side" and "B side")
- Contains batteries or flywheels to bridge the time between the utility failure and the availability of generator power
  - AC-DC-AC double conversion
- UPS conditions the incoming power feeds
  - Removing voltage spikes or sags
  - Harmonic distortions in the AC feed

Going green: flywheels

Flywheel energy storage (FES) works by accelerating a rotor (flywheel) to a very high speed and maintaining the energy in the system as rotational energy. When energy is extracted from the system, the flywheel's rotational speed is reduced

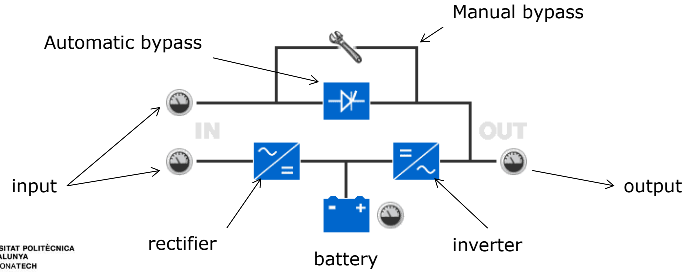
PROS:
- Flywheels can operate at a much wider temperature range, and are not subject to many of the common failures of chemical rechargeable batteries
- Made of inert or benign material (no environmental damage)
- Costs of a fully installed flywheel UPS are about $330 per 15 seconds at one kilowatt (one half of a chemical battery)
- Consumes 75% less space
- 20 year design life (against 3 years)

CONS:
- Use of flywheel accumulators is currently hampered by the danger of explosive shattering of the massive wheel due to overload

From wikipedia

UPS scheme

1. Input: Monofasic or trifasic
2. Output: Monofasic, with an electronic controlled interrupter
3. Rectifier
   - Transform AC in DC
   - Battery charge
   - Feeds the inverter that generates stabilized AC
4. Battery
   - Can be a chemical battery, modern ones based on flywheels
   - Autonomy depends on the battery capacity (in kWh)
5. Inverter
   - Generates AC from DC
   - Generated power must have stable voltage and frequency
   - It feeds from rectifier or from batteries
6. Automatic Bypass
   - Feeds directly input from output in case of UPS malfunction
7. Manual Bypass
   - Direct connection for UPS maintenance



Additional features

UPS MUST communicate with systems to shut down them in case of battery discharging

- Monitoring units in UPS
- Daemon in the O.S
- Communication system
- Control message protocol

Less critical in case of having a generator power system

Real-world data centers usually include "paralleling" of generators of UPS units: multiple devices feed a shared bus so that the load of a failed device can be picked up by other devices

Typical configurations

- N+1: allows one failure or maintenance
- N+2: allows one failure even when one unit is offline for maintenance
- 2N: redundant pairs

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

UPS pros and cons

PROS

- Protection against power failures
- Stable voltage and frequency

CONS

- Costs: acquisition and maintenance
- Usually three years lifetime
- Difficult to scale (overestimation for the future?)
- Automatic operation complexity (for automatic shutdown)

FUTURE?

- Some companies are thinking about non-UPS DC
  - http://www.datacenterknowledge.com/archives/2010/07/14/data-centers-with-no-ups-or-generator/
  - http://www.datacenterknowledge.com/archives/2010/06/24/blue-waters-awesome-power-awesome-efficiency/
- Based on data and operations redundancy (only available for huge companies, like Yahoo!)

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

UPS: from small ones (up to 8 kVA)

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

To big ones (>8kVA)

More datacenter power and HVAC system elements

- Power Distribution Units (**PDUs**): the UPS output is routed to PDUs, that sit on the datacenter floor. They take higher-voltage feed (typically 240V, 30A) and break it up into many 110- 220-V, 15A circuits that feed the servers
- **CRAC** units (computer room air conditioning): pressurized the raised floor plenum by blowing cold air. Hot air produced by the servers recirculated back to the intakes of the CRAC that cool them
- **Heat rejection devices**: cooling towers outside the building, radiators, water-water heat exchangers
  - More problems: cooling equipment is also backed up by generators (and sometimes UPS units) because the data center cannot operate without cooling for more than a few minutes
  - According to Barroso & Hözler, in a typical datacenter chillers and pumpers can add 40% or more to the critical load that needs to be supported by generators
- **Emergency diesel generators**
  - Have you think about the fuel storage tanks?
  - And the fuel provision?

Power and HVCA summary

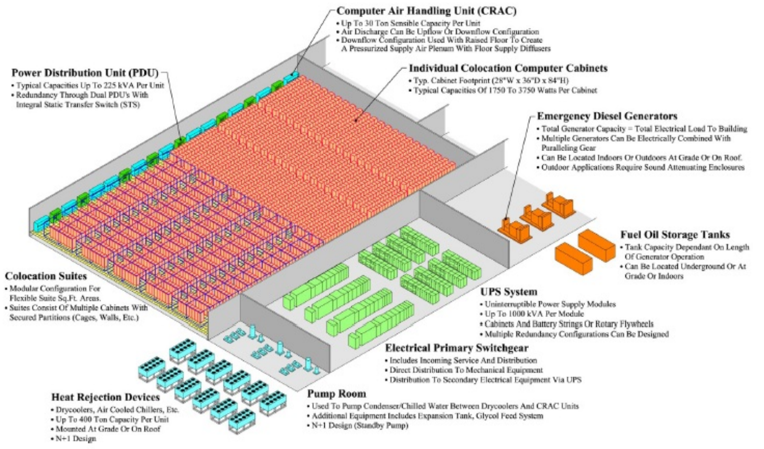


FIGURE 4.1: The main components of a typical datacenter (image courtesy of DLB Associates [23]).

Image from Barroso and Hölzle, The Datacenter as a Computer

Datacenter staff

Different roles and responsibilities

Head of Operations / Operations director

- Make decisions about the DC design
  - Choose the technologies used
- Design policies of maintenance, operation, backup, etc..
- Organize work schedule and assigns tasks to staff

System / network administration

- Install, configure and maintain networks and OS

Technical applications administration

- Install, configure and maintain applications on servers
- They can be hired to the software company

Operation

- Routine maintenance: hardware changes, changing magnetic tape robots …
- Monitoring environmental conditions of the room

Others

- Help desk, web design, …

<u>THEY NEED SOME FACILITIES</u>

Physical access control

Access to the computer room must be restricted

- Locked doors, with electronic access control system (ACS)
  - Access card readers or biometric sensors (e.g. fingerprint reader)
- Video camera surveillance
  - In particular, pointing to the computer room doors
- Permanent security guards?
- Mantraps

Inside the computer room

- Chassis opening sensor (alarm)
  - Avoid data theft about subtracting discs

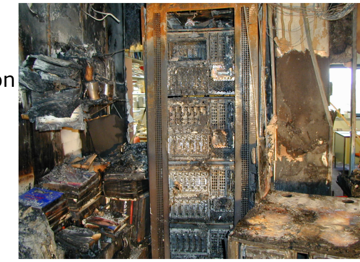
UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Fire protection system

A fire can mean the total destruction of the equipment

Early detection = minimal damage

- Heat sensors
  - Detect high temperatures (fire)
- Early Warning Smoke Detector (EWSD)
  - Detects the first signs of smoke
  - Automatic sensors on the ceiling and cooling ducts
  - Disconnect the cooling equipment at the first sign of smoke
- Very Early Warning Smoke Detector (VEWSD)
  - Detects overheating hours or days before the fire
  - Vacuum system for sampling the air in the room searching for smoke particles
- Several sensors placed separately
  - 1 active sensor triggers the alarm
  - 2 active sensors triggers the extinction



UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Fire protection system

Extinction with inadequate means (water, foam or powder extinguishers) may involve the destruction of equipment

Based on water (sprinklers on roof)

- Can leak! Normal system in office buildings
- Solutions
  - The pipes are not filled until the alarm
  - Sprinklers are opened by fusion of a heat-seal
  - Dense fog: absorb heat with microscopic droplets (100-120 microns) that do not damage the equipment

Based on gas

- Leave no residue on equipment or in the room
- Reduction of oxygen:
  - Lower the oxygen concentration available with inert gas (argon or nitrogen).
- Heat removal
  - Different HCFCs (*hydrochlorofluorocarbon*) absorb heat from the environment by eliminating combustion

Hybrids

- First you use a gas system and if it fails to activate the water to save the lives of people and the building

Appropriate extinguishing = no need to stop operation

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

(Not so) minor details

- Don't forget ramps and lifts (scaled for new equipment)
- Weight is important!
- When scaling your UPS, don't forget basic security elements (like emergency lighting or electronic equipment required for DC access –like access card readers), and reserve power enough for them before battery is exhausted
- Avoid unnecessary loads from the datacenter power source (like elevators)

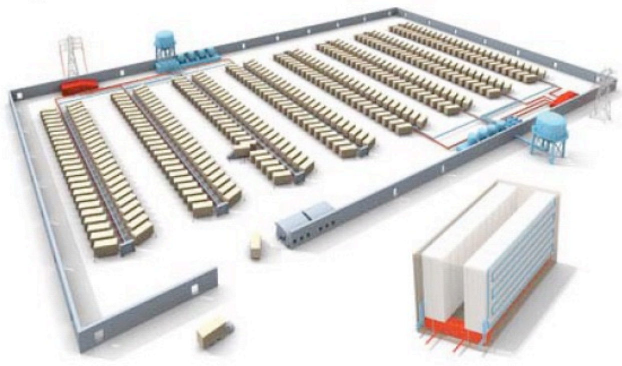
UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Future buildings: time for containers?

- **Provides excellent energy efficiency by offering more precise control of airflow within the container**
- **Examples: Microsoft and Google**

Estimated 14% consumption reduction

Next generation Datacenters?

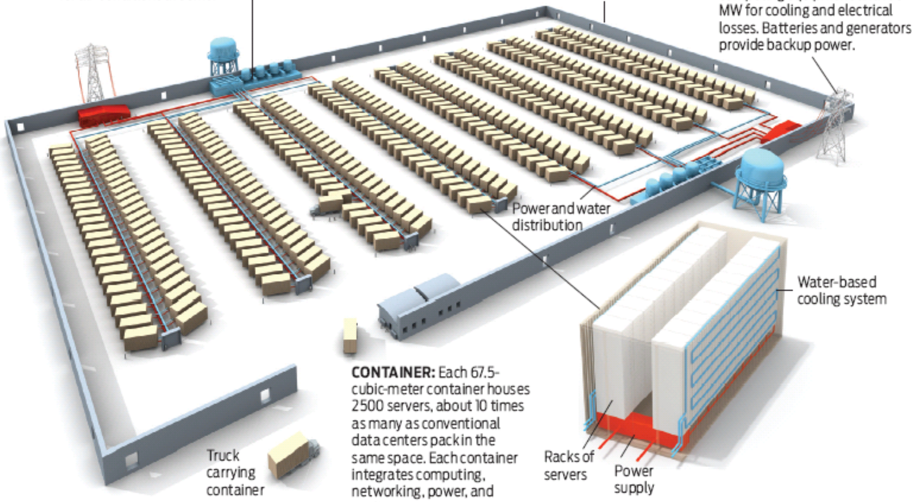


Source: Tech Titans Building Boom , Randy H. Katz.
IEEE Spectrum, February 2009
http://spectrum.ieee.org/green-tech/buildings/tech-titans-building-boom

COOLING: High-efficiency water-based cooling systems—less energy-intensive than traditional chillers—circulate cold water through the containers to remove heat, eliminating the need for air-conditioned rooms.

STRUCTURE: A 24 000-square-meter facility houses 400 containers. Delivered by trucks, the containers attach to a spine infrastructure that feeds network connectivity, power, and water. The data center has no conventional raised floors.

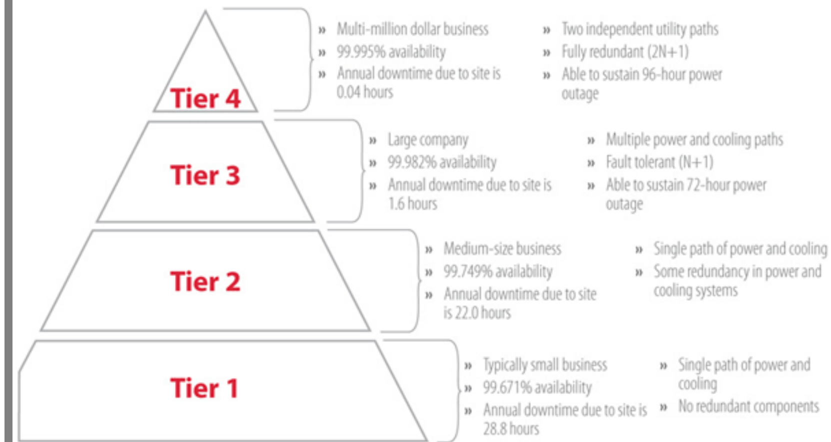
POWER: Two power substations feed a total of 300 megawatts to the data center, with 200 MW used for computing equipment and 100 MW for cooling and electrical losses. Batteries and generators provide backup power.



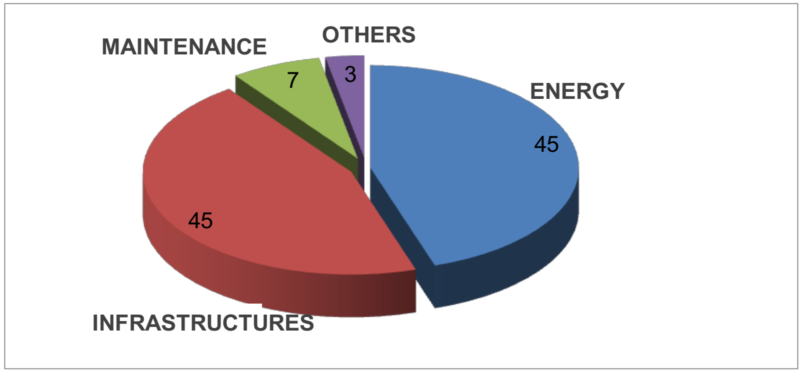
CONTAINER: Each 67.5-cubic-meter container houses 2500 servers, about 10 times as many as conventional data centers pack in the same space. Each container integrates computing, networking, power, and cooling systems.

Tiers

Data Center Tiers



Costs of a Datacenter



Data courtesy of
AST IT Infrastructures
(www.AST-global.com)

PUE: Data Center efficiency metrics

Data center supporting infrastructure has a major impact on the energy use

A common measure of how efficiently a DC uses its power is called power usage effectiveness ratio (**PUE**).

PUE DC UPC= 1.7 (*)

$$PUE = \frac{\text{Total Facility Power}}{\text{IT Equipment Power}}$$

(*) Source: Javier Hidalgo

2.6. MEASURING COSTS AND EFFICIENCY

Are we efficient?

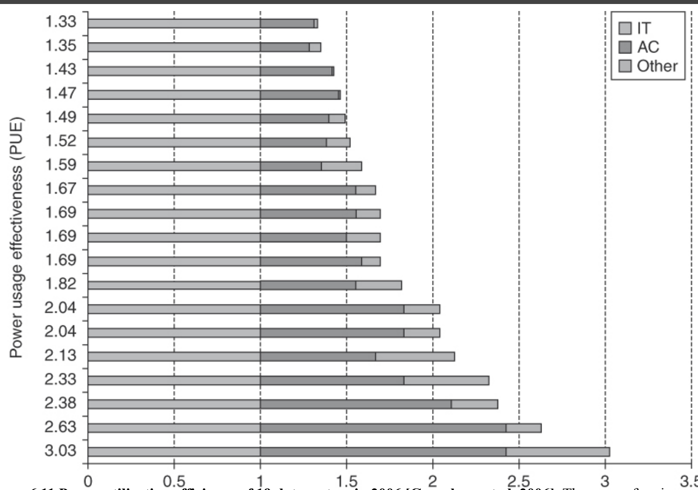


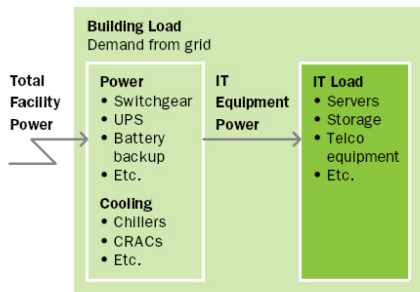
Figure 6.11 Power utilization efficiency of 19 datacenters in 2006 [Greenberg et al. 2006]. The power for air conditioning (AC) and other uses (such as power distribution) is normalized to the power for the IT equipment in calculating the PUE. Thus, power for IT equipment must be 1.0 and AC varies from about 0.30 to 1.40 times the power of the IT equipment. Power for "other" varies from about 0.05 to 0.60 of the IT equipment.

Borrowed from Hennessy & Patterson: Computer Architecture, A Quantitative Approach

2.6. MEASURING COSTS AND EFFICIENCY

PUE & DCE

PUE: Power Usage Effectiveness
DCE: Data Center Efficiency



$$PUE = \frac{\text{Total Facility Power}}{\text{IT Equipment Power}}$$

$$DCE = \frac{1}{PUE} = \frac{\text{IT Equipment Power}}{\text{Total Facility Power}}$$

Source: The Green Grid, 2008, "The Green Grid Metrics: Data Center Infrastructure Efficiency (DCiE) Detailed Analysis"

2.6. MEASURING COSTS AND EFFICIENCY

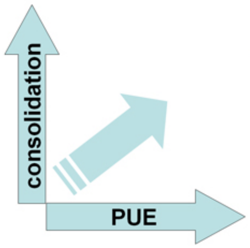
Is PUE a good measure?

Usually the goal is to reduce the data center PUE

- Nevertheless any project that improves an IT load alone will yield a worse PUE!

Example:

- 100 Mw coming into a facility and 50 Mw are taken up by the IT load → PUE = 2
- A consolidation strategy reduce the IT part to 40 Mw
- The PUE is now 2.25 (90/40), which is worse than the PUE of 2 we had before our virtualized/consolidated strategy



PUE is not enough!

So?

No actual metric has been defined yet

Ideas:

Efficiency = Computation / Total energy

$$\text{Efficiency} = \underset{(a)}{\frac{1}{\text{PUE}}} \times \underset{(b)}{\frac{1}{\text{SPUE}}} \times \underset{(c)}{\frac{\text{Computation}}{\text{Total energy to Electronic Components}}}$$

We know the (a) term of equation (DCE)

SPUE

Term (b): PUE does not account for inefficiences within the server (server's power supply, voltage regulator modules -VRMs-, cooling fans)

- SPUE= Server PUE = total server input power / power consumer by components directly involved in computation (motherboards, disks, CPUs, DRAM, I/O cards, and so)
- Current SPUE values (Barroso & Hölzle)= 1.6 – 1.8

Some companies use **tPUE** (total PUE)

$$\text{tPUE} = \underset{(a)}{\frac{1}{\text{PUE}}} \times \underset{(b)}{\frac{1}{\text{SPUE}}}$$

And term (c) of equation?

How much of electricity delivered to electronic components is translated in useful work?

- Difficult to measure
- Not all companies run the same workloads
  - How can you compare two WSC?
- There are <u>experimental</u> results

$$\text{Efficiency} = \underset{(a)}{\frac{1}{\text{PUE}}} \times \underset{(b)}{\frac{1}{\text{SPUE}}} \times \underset{(c)}{\frac{\text{Computation}}{\text{Total energy to Electronic Components}}}$$

CPU utilization in Google servers

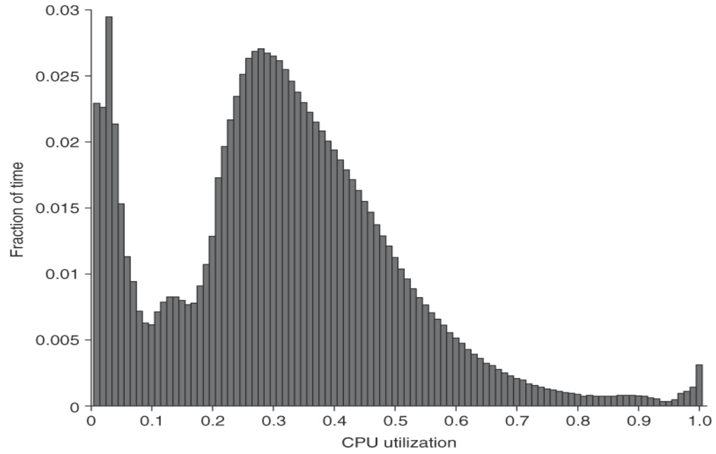


Figure 6.3 Average CPU utilization of more than 5000 servers during a 6-month period at Google. Servers are rarely completely idle or fully utilized, in-stead operating most of the time at between 10% and 50% of their maximum utilization. (From Figure 1 in Barroso and Hölzle [2007].)

Borrowed from Hennessy & Patterson: Computer Architecture, A Quantitative Approach

TCO (Total Cost of Ownership)

TCO analysis was popularized by the Gartner Group in 1987

- TCO tries to quantify the financial impact of deploying an information technology product over its life cycle
- An associate idea is the Return of Investment (**ROI**)

Technology deployment can include the following as part of TCO:

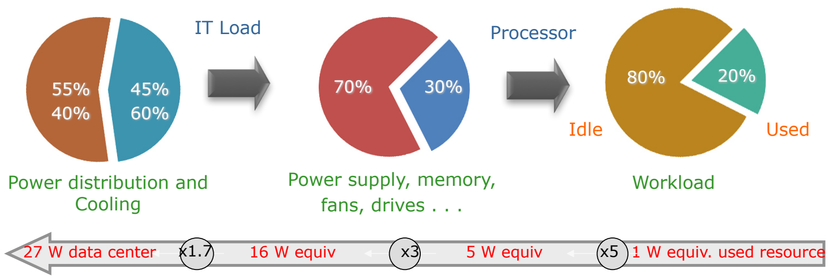
- Computer hardware and programs
  - Network hardware and software
  - Server hardware and software
  - Workstation hardware and software
  - Installation and integration of hardware and software
  - Purchasing research
  - Warranties and licenses
  - License tracking – compliance
  - Migration expenses
  - Risks: susceptibility to vulnerabilities, availability of upgrades, patches and future licensing policies, etc.

(continue)

TCO (Total Cost of Ownership)

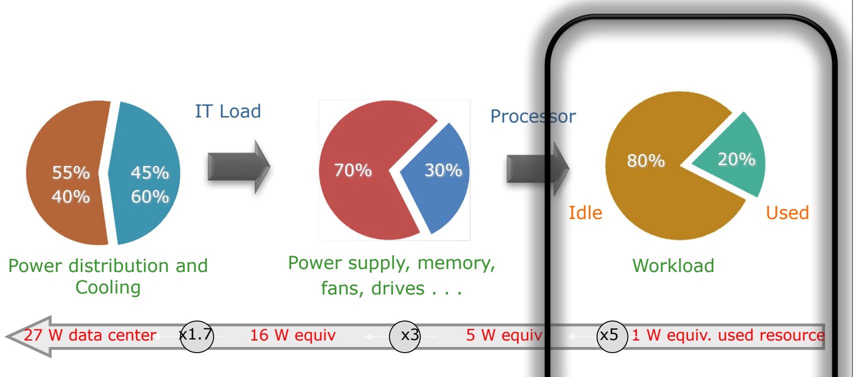
- Operation expenses
  - Infrastructure (floor space)
  - Electricity (for related equipment, cooling, backup power)
  - Testing costs
  - Downtime, outage and failure expenses
  - Diminished performance (i.e. users having to wait, diminished money-making ability)
  - Security (including breaches, loss of reputation, recovery and prevention)
  - Backup and recovery process
  - Technology training
  - Audit (internal and external)
  - Insurance
  - Information technology personnel
  - Corporate management time
- Long term expenses
  - Replacement
  - Future upgrade or scalability expenses
  - Decommissioning

How do datacenters spent energy?



Idea: IBM and Dynamic Infrastructure, Doug Neilson, IBM Systems Group, 2009

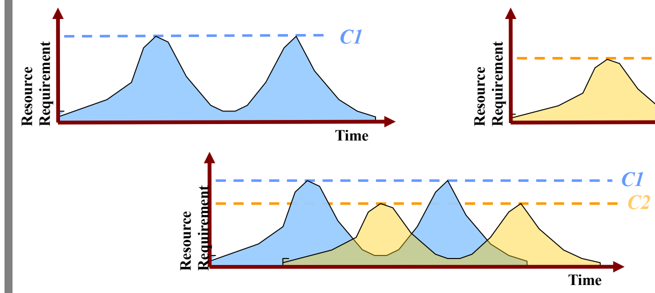
How do datacenters spent energy?



Idea: IBM and Dynamic Infrastructure, Doug Neilson, IBM Systems Group, 2009

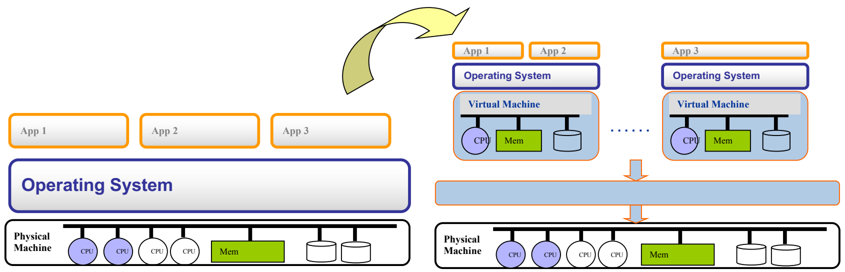
Consolidation

- Servers are underused: around 10 to 20 percent
- Consolidate into a single machine
  - Resource multiplexing if different peak times

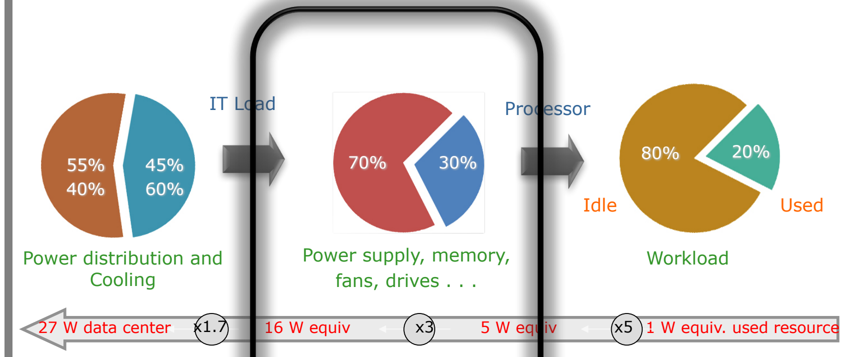


Virtualization

Virtualization divides a physical server into isolated virtual environments, enabling organizations to run multiple applications or OS on a single server



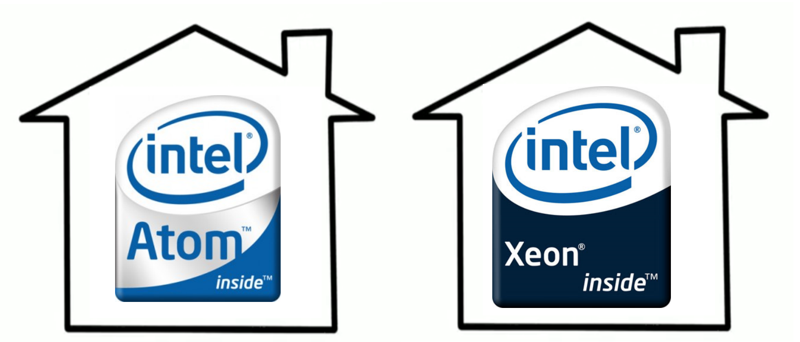
How do datacenters spent energy?



Idea: IBM and Dynamic Infrastructure, Doug Neilson, IBM Systems Group, 2009

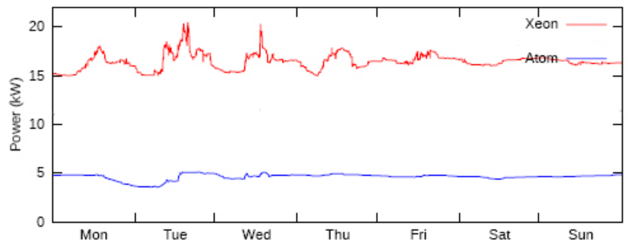
Use of low power processors?

Atom vs Xeon?



Good approach for transactional workloads (webs)

Experiments at BSC

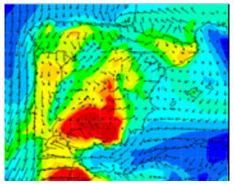


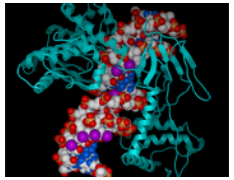
However the numerical applications are not like that

Experiments at BSC with 4 HPC tasks:

- 1 Xeon * 1 hour → 317 Watts
- 2 Atom * 5 hours → 398 Watts

The most energy-efficient approach in this environment is running jobs very fast and then power the system off.





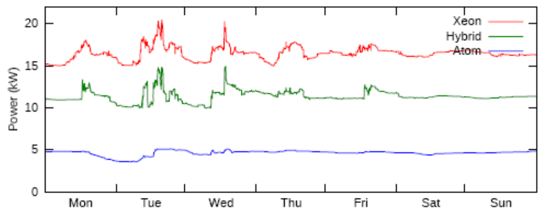
SO?

If for HPC Xeons works more efficiently than Atoms
And for Transactional Atoms works more efficiently than Xeons

Why not build the two together?

Better tradeoffs

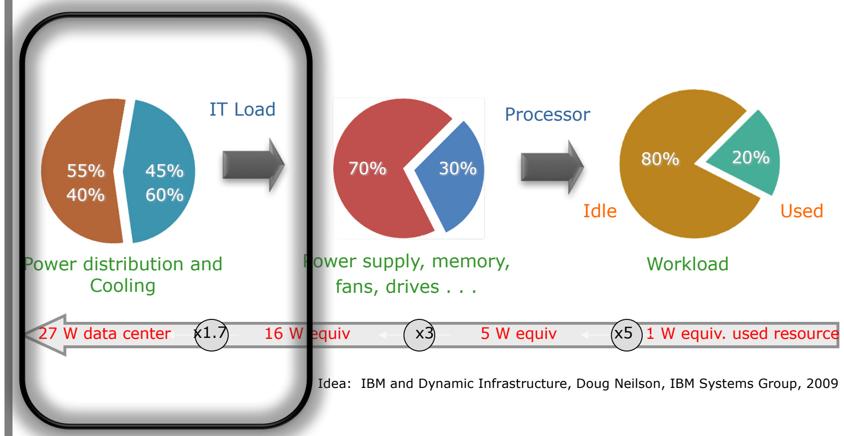
Experiments at BSC



Nodes	Work/ON	Power (kW)	*DL* (%)	*R* (%)	B (€)
Xeon	25.1 / 27.1	2960.2	59.8	84.2	816.5
Heterogeneous	38.1 / 40.1	2053.7	52.2	83.7	944.7
Atom	52.7 / 54.7	849.9	0.0	83.6	709.6

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

How do datacenters spent energy?

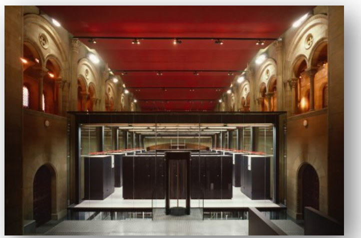


Idea: IBM and Dynamic Infrastructure, Doug Neilson, IBM Systems Group, 2009

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Best Practice at Marenostrum

Can we Minimize losses and thermal/cooling overheads in Marenostrum?



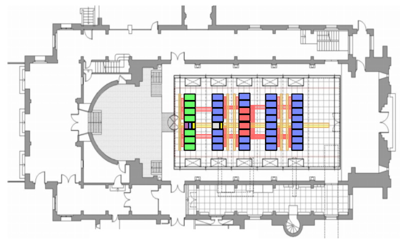
Source: BSC

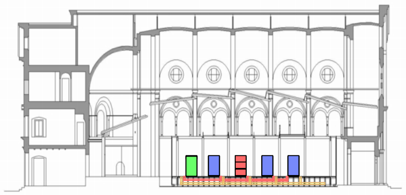
Source:
"El reto operacional de dirigir el tercer supercomputador más grande de Europa"
Sergi Girona,
Director Operaciones
BSC-CNS, 2009.

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Some Marenostrum facts

- Peak Perf: 94.21 Teraflops
- 10,240 Power PC 970MP at 2.3 GHz (2560 JS21 blades)
- 20 TB of main memory
- 280 + 90 TB of disk storage
- Interconnection networks:
  - Myrinet and Gigabit Ethernet
- Linux: SuSe Distribution
- Space 160 square meters
- About 26 tons of steel (19 glass).
- Power aprox. 1.071 kW

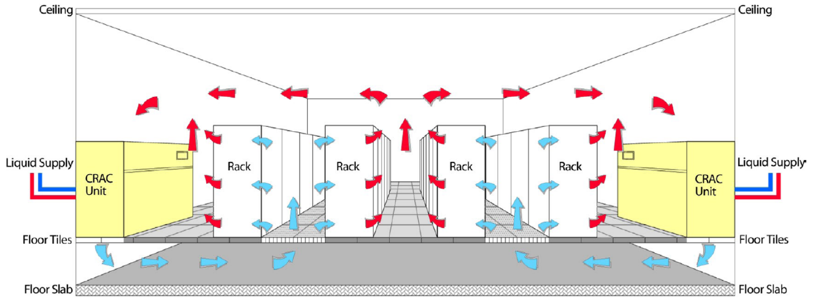




UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Improving air flow management

Raised floor with hot – cold aisles
Estimated PUE around 1,45



Font: Luiz Andre Barroso, Urs Hoelzle, "The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines", 2009.
(Image courtesy of DLB associates , ref [23] of the book)

Under-floor pressure

Problem measured: to much under-floor pressure
- Solution: Move some floor tiles

Benefits observed:
- Improvement of AC equipment performance
- Improved the rack-bottom temperature

Source: BSC

Substituting floor tiles

- Composite
  - 20% opening

- Metallic tile
  - 40% opening

Benefits observed:
- Less working pressure for the Cooling components
- All bladecenters temperature reduced by 2º C
- Cold barrier that prevents the reflux of hot air

Source: BSC

Temperature map

Problem measured:
- not all the Racks have the same temperature

Idea:
- Force the air flow

28.50	28.00	25.50	28.00	25.50	25.50	25.50	31.00
27.00	24.00	24.00	25.00	23.00	24.00	23.00	28.50
26.50	25.50	23.50	25.00	24.50	23.00	24.00	29.50
28.00	24.50	23.50	26.00	24.00	24.50	23.50	28.50
22.50	26.00	25.00	27.00	24.50	25.00	25.00	27.50
27.50	27.00	27.00	27.00	25.50	28.50	25.00	27.50
27.50	26.50	27.00	27.50	26.50	26.00	28.00	27.50
27.50	25.00	23.00	25.00	24.00	26.00	24.50	30.00
29.00	24.00	24.50	23.50	23.00	23.50	24.50	29.00
28.50	24.00	23.50	25.50	24.50	25.50	24.50	28.00
27.00	23.50	25.00	25.50	21.50	25.00	25.50	26.00
27.00	24.50	24.50	22.50	22.50	25.50	24.00	27.50
26.00	25.50	MYRI	MYRI	MYRI	MYRI	24.00	30.00
28.50	24.00	MYRI	MYRI	MYRI	MYRI	25.00	27.50
26.00	24.00	MYRI	MYRI	MYRI	MYRI	23.50	30.50
27.00	24.00	19	19	19	19	25.50	31.00
28.50	21.50	19	19	19	19	22.00	29.00
25.50	25.50	19	19	19	19	21.50	27.00
28.50	25.50	26.00	26.00	NET	26.50	26.00	27.50
29.00	25.50	24.50	23.50	NET	24.50	23.00	25.50
26.50	24.50	24.50	24.00	NET	24.50	23.00	26.50
27.50	25.50	25.50	25.00	NET	22.50	23.00	23.00
25.50	26.00	26.50	27.50	NET	24.50	25.50	21.00
27.00	26.00	27.00	28.00	NET	26.50	27.00	24.00

Forcing air flow

Methacrylate screens installed in front of each rack

- to guide the cold air flow directly to the computer, instead of each rack having to take cooling air from the general environment

Benefits observed:

- All BladeCenters Rack has an equal temperature +/- 1ºC
- BladeCenter fan speed reduced

Source: BSC

Results

Including other improvements:

- reduction of **10%** of power consumption (and CO_2)

Marenostrum **power consumption**: approx. 1.2 Mw

→ aprox 1.100.000 €/year

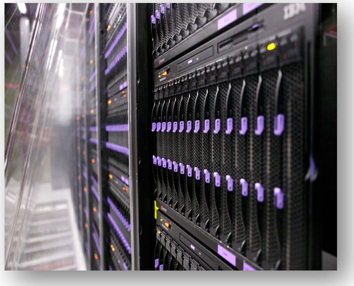
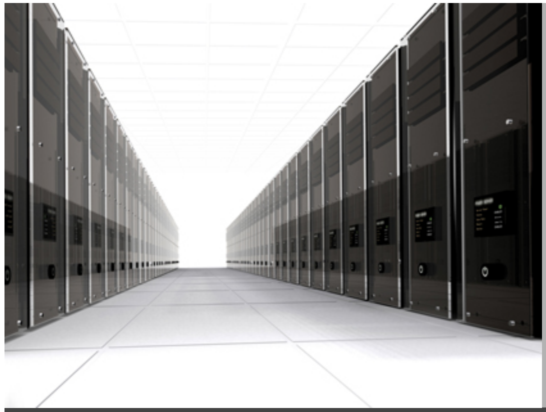


Image courtesy of UPC

PUE ~ 1,3

Google good practices

Part I: Introduction (3:30)
http://www.youtube.com/watch?v=O96PwWkJdUo
Part II: Manage Airflow (2:37)
http://www.youtube.com/watch?feature=endscreen&NR=1&v=3OVqjJjNtAE
Part III: Adjust the thermostat (1:05)
http://www.youtube.com/watch?NR=1&feature=endscreen&v=yzJfNP5sAao
Part IV: Utilize free cooling (1:55)
http://www.youtube.com/watch?v=_uVeDwleAEo&feature=endscreen&NR=1
Part V: Optimize power distribution (2:05)
http://www.youtube.com/watch?v=dk95NV01dAY&feature=endscreen&NR=1



DATA CENTER INFRASTRUCTURES

David López

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH