Energy efficient IT and infrastructure for data centres and server rooms

PrimeEnergyIT
EFFICIENT DATA CENTERS

Imprint
Responsibility: PrimeEnergyIT Project consortium, July 2011
Project coordination: Dr. Bernd Schäppi, Austrian Energy Agency, Vienna
Reprint allowed in parts and with detailed reference only. Printed on non-chlorine bleached paper.

Efficient technology for energy and cost savings in data centres and server rooms

Energy consumption in data centres and server rooms has been increasing significantly during the last decade. More powerful equipment and more complex IT services have been driving power demand. Since infrastructure and energy costs in data centres have become a central factor in facility and IT management, a range of technologies has been developed to increase energy efficiency. New hardware and power management options support energy saving strategies.

Overall, the energy saving potential in data centres and server rooms is high and may exceed 50% in many cases, depending on the specific IT and infrastructure. In the past the focus of energy saving measures has been on efficient solutions for power supply and cooling. More recently also measures addressing IT hardware efficiency are considered. Current studies show that efficiency measures already lead to a significant reduction of energy demand, compared to a business as usual scenario[1]. Nevertheless, the remaining energy saving potential is still large and new technologies allow even more effective deployment of saving options.

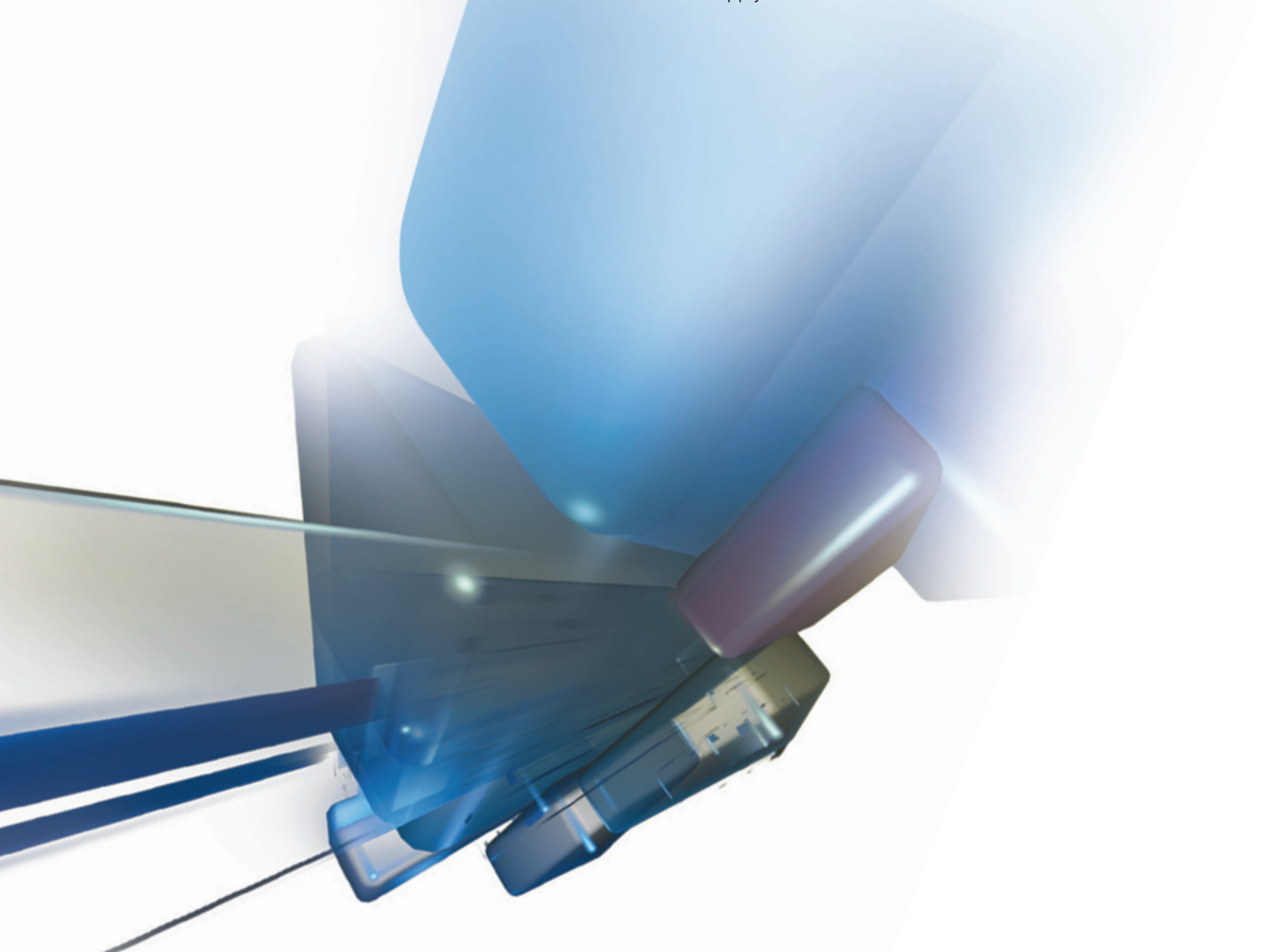
This brochure provides a short overview of current technologies supporting energy efficiency both for IT and infrastructure, with a focus on IT technology. It covers all essential IT technologies in the data centre, including servers, data storage and network equipment. Efficiency approaches include effective system design, power management from hardware to data centre level as well as consolidation and virtualisation approaches.

Recommendations for best practice highlight promising options to be considered in management and procurement. A number of resources for further reading are indicated. The brochure provides a source of basic information for IT and infrastructure managers to support energy- and cost-efficiency.

This brochure has been produced as
part of the international project PrimeEnergyIT
(www.efficient-datacenters.eu)
which is conducted within the framework of the
EU programme Intelligent Energy Europe.

1) Koomey, J. (2011): Growth in Data center electricity use 2005 to 2010, Jonathan Koomey, Analytics Press, Oakland, CA, August 1, 2011

Content



1 Monitoring of energy consumption in server rooms and data centres

Carlos Patrao, University of Coimbra

1.1 Monitoring concepts

Monitoring of energy consumption in server rooms and data centres is essential to detect energy saving potentials and evaluate the effectiveness of efficiency measures. Monitoring concepts should be designed with care to ensure that right data is collected, supporting effective measures.
The following aspects are to be considered [1]:

- Required accuracy and resolution of data
- Breakdown of data collection, ability to collect data from all desired devices
- User friendliness and ease of integration of data across devices and time scales
- Scalability for mass deployment and multi-site capability
- Adaptability to new measurement needs
- Data analysis options and integration with control systems
- Ability to detect problems and notify data centre operators
- Investment costs and pay-back

The following typical approaches for monitoring may be applied:

Minimum Monitoring – Performing periodic spot measurements with portable equipment is mainly an approach for very small facilities. Some data is acquired from manufacturer's information (power input, etc). The approach does not require investment in permanently installed measurement equipment and infrastructure.

Advanced Monitoring – Data is logged in real time by using permanently installed equipment not necessarily supported by online tools. Limited modifications to the infrastructure should be expected.

State-of-the-Art Monitoring – Data is collected by the use of automated/permanent recording systems in real time, with the support of online software with extensive capability for analysis. Modifications in the infrastructure are needed and support from expert technical staff mostly will be required.

The monitoring system must have the necessary number of "info nodes" (or monitoring points) to provide the required information for comprehensive energy consumption analysis. In larger facilities the "info nodes" selection should start with the most representative subsystems (in terms of power usage). Figure 1.1 shows the most important subsystems for which energy consumption should be monitored. These subsystems can also be considered as "info nodes" or "monitoring points".

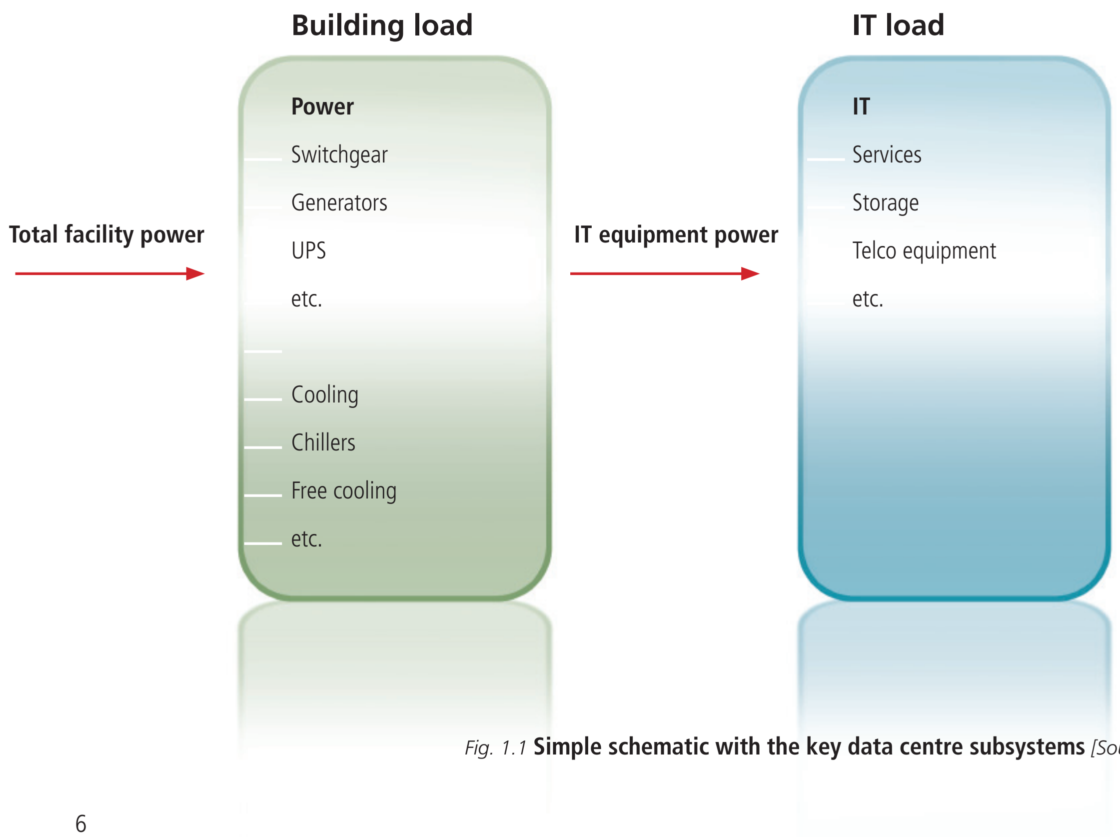


Fig. 1.1 **Simple schematic with the key data centre subsystems** *[Source: ASHRAE [2]].*

Data collection, processing and evaluation is commonly supported by software tools. For example the Save Energy Now Program (US Department of Energy) has developed a tool suite called "DC Pro". The tool suite provides an assessment process, tools for benchmarking and performance tracking as well as recommendations for measures. It is available for free.

http://www1.eere.energy.gov/industry/datacenters/software.html

OTHER EXAMPLES OF USEFUL SOFTWARE TOOLS ARE:

- Power usage tool:
  http://estimator.thegreengrid.org/puee
- PUE reporting tool
  http://www.thegreengrid.org/en/Global/Content/Tools/PUEReporting
- PUE Scalability Metric and Statistics Spreadsheet
  http://www.thegreengrid.org/library-and-tools.aspx?category=MetricsAndMeasurements&range=Entire%20Archive&type=Tool&lang=en&paging=All#TB_inline?&inlineId=sign_in
- PUE and DCiE Data Centre Efficiency Measurement
  http://www.42u.com/measurement/pue-dcie.htm

RECOMMENDATIONS FOR BEST PRACTICE

Proper understanding of the overall goals for the energy monitoring is essential for designing an effective monitoring concept.
Typical goals may be:

- Assessment of total IT and infrastructure energy consumption
- Analysis of energy consumption trends over time
- Understanding the instantaneous power demand of key equipment within the facility
- Billing
- Calculating energy efficiency indexes and energy efficiency metrics

The software/hardware concept for energy monitoring shall provide the following capabilities (Source ASHRAE):

- Reliable data collection and data storage at the required rates and accuracy
- Normalization of data from different devices, interfaces and protocols
- Data storage for long measurement periods
- Analysis and visualization of data in form of tables and graphs
- Scaling of the architecture with the data centre expansion

Key aspects to be taken into account when choosing the devices for the monitoring system among others are instrument range, resolution and accuracy.

Tab. 1.1 **Examples for energy metering devices**

Designation	Example	Description	Monitoring approaches
Portable meter	 Source: Chauvin Arnoux	Portable power meters cover a range of products including hand-held single phase multi-meters to sophisticated three phase power analyzers with recording and triggering capabilities. Most of them have a built-in display where the user can access the measured or recorded data.	Minimum and advanced monitoring.
Panel meter	 Source: Chauvin Arnoux	Panel meters are usually permanently installed in the switchgear measuring UPS systems, generators or other devices. These meters have a display that shows the instantaneous measurements and cumulative variables such as the total energy consumption. They can be installed to measure overall and individual energy consumption of devices.	These meters can be used for best practical and state-of-the-art monitoring.
Revenue meter	 Source: Itron	The revenue meters are mostly used by the electrical utilities, landlords and others who bill their customers. They are rarely used in data centre monitoring systems, but they can provide data about the overall energy consumption of the facility. In some cases utilities can provide the access to the digital communication port, which gives the ability to acquire and store it in a database for future analysis (for instance every 15 minutes).	Can be used in all the approaches.
Intelligent power distribution units	 Source: Raritan	Intelligent or metered rack power distribution units (PDUs) provide active metering to enable energy optimization and circuit protection. Metered rack PDUs provide power utilization data to allow data centre managers to make informed decisions on load balancing and right sizing IT environments to lower total cost of ownership. PDUs may be equipped with real-time remote unit-level and individual outlet-level power monitoring of current, voltage, power, power factor and energy consumption (kWh) with ISO/IEC +/– 1% billing-grade accuracy. Users can access and configure metered rack PDUs through secure Web, SNMP-, or Telnet interfaces.	Can be used in all the approaches.
Server-embedded power metering feature		Server-embedded power metering feature	Minimum and advanced monitoring.
Power transducer	 Source: Chauvin Arnoux	The power transducer is usually referred as an equipment with no display that is permanently connected in a switchgear like the panel meters. Such devices are often used by monitoring systems to acquire power measurements from various points of a data centre.	Can be used in all the approaches.

1.2 Measurement devices

A large number of types of measurement devices is available for measuring key variables such as energy consumption, temperature, flow rate and humidity.
Some examples for energy measuring devices are presented in table 1.1 (left, on page 8). For further reading see the sources indicated in the following section or access the "Technology Assessment Report" available at the PrimeEnergyIT website.

Further Reading

ASHRAE (2010): Real-Time Energy Consumption Measurements in Data Centres, ASHRAE – American Society of Heating, Refrigerating and Air-Conditioning Engineers, 2010.
ISBN: 978-1-933742-73-1

Stanley, J. and Koomey, J. (2009): The Science of Measurement: Improving Data Centre Performance with Continuous Monitoring and Measurement of Site Infrastructure, Stanley John and Koomey Jonathan, October 2009
www.analyticspress.com/scienceofmeasurement.html

Ton, M. et al (2008): DC Power for Improved Data Centre Efficiency,Ton, My, Fortenbery, Brian and Tschudi, William, Ecos Consulting, EPRI, Lawrence Berkeley National Laboratory, March 2008
http://hightech.lbl.gov/documents/data_centres/dcdemofinalreport.pdf

The Green Grid (2008): Green Grid Data Centre Power Efficiency Metrics. White Paper 6, The Green Grid, White Paper 6. December 30, 2008
http://www.thegreengrid.org/Global/Content/white-papers/The-Green-Grid-Data-Centre-Power-Efficiency-Metrics-PUE-and-DCiE

Rasmussen N. (2009): Determining Total Cost of Ownership for Data Centre and Network Room Infrastructure, Neil Rasmussen, APC by Schneider Electric, White paper #6 – Revision 4
http://www.apcmedia.com/salestools/CMRP-5T9PQG_R4_EN.pdf

Rasmussen N. (2010): Avoiding Costs From Oversizing Data Centre and Network Room Infrastructure, Neil Rasmussen, APC by Schneider Electric, 2010. White paper #37 – Revision 6
http://www.apcmedia.com/salestools/SADE-5TNNEP_R6_EN.pdf

Schneider Electric (2011): E-learning website (Energy University) that provides the latest information and training on Energy Efficiency concepts and best practice
www.myenergyuniversity.com

Webinar: „The Data Centre in Real Time: Monitoring Tools Overview & Demon"
http://www.42u.com/webinars/Real-Time-Measurement-Webinar/playback.htm

References

[1] **Stanley, J. and Koomey, J. (2009):** The Science of Measurement: Improving Data Centre Performance with Continuous Monitoring and Measurement of Site Infrastructure. October 2009.
[2] **ASHRAE (2010):** Real-Time Energy Consumption Measurements in Data Centres: ASHRAE- American Society of Heating, Refrigerating and Air-Conditioning Engineers, 2010.
ISBN: 978-1-933742-73-1.

2 Server Equipment

Bernd Schäppi, Thomas Bogner, Hellmut Teschner, Austrian Energy Agency

Server equipment consumes about 30–40% of the total energy used in data centres and server rooms. Therefore, it is one of the primary areas to implement effective energy saving measures. Typical server equipment in common server rooms and data centres includes standard rack servers, blade servers, as well as pedestal servers and multi-node servers.
The energy efficiency potential is high and depending on the type of IT system and the measures applied, energy savings of 20–60% or even beyond can be achieved. The primary approaches for improving energy efficiency involve energy efficient hardware selection and system design, power management at all levels from the hardware component to the total system and last but not least hardware consolidation and virtualization.

The following chapter provides information on power saving technologies and options from the component to the system level. Energy efficiency issues and possible measures for improvement are provided from the server to the rack and data centre level. Two specific sections address blade server technology and server virtualization as potential efficiency strategies. Specific recommendations for best practice options are highlighted in boxes.

Tab. 2.1 **Energy Star Idle Power Criteria**

Category	Number of installed processors	Managed server	Base Idle State Power Allowence (W)
A	1	No	55
B	1	Yes	65
C	2	No	100
D	2	Yes	150

Tab. 2.2 **Concept of SERT assessment tool**

	Server	
Benchmark result system	CPU	Benchmark result
	Memory	Benchmark result
	Storage	Benchmark result
	IO	

2.1 Energy efficiency and power management at the server and component level

Energy efficiency of servers has been strongly improved in the last years, mainly due to development of effective power management for hardware components. To date, server energy efficiency is assessed and declared based on Energy Star requirements and the SPECpower benchmark (SPEC: Standard performance evaluation corporation).
The current ENERGY STAR requirements for enterprise servers [1] stipulate energy efficiency criteria for rack and pedestal servers containing up to 4 processor sockets. The requirements define maximum levels for power consumption in On Idle Mode for 1- and 2-CPU socket servers as well as criteria for power supply efficiency and power management features (see Table 2.1 and Table 2.4). The idle mode criteria are primarily useful as an efficiency indicator for low average load conditions close to idle operation. Such low loads on servers (e.g. < 15%) are still quite common, although hardware consolidation to achieve higher load levels should be a general goal.
Server energy efficiency at higher workloads and for consolidated systems is addressed with the SPECpower–benchmark, which however is focused more on CPU-related efficiency and CPU intense workloads (see information below). A comprehensive Server Efficiency Rating Tool (SERT) addressing all major server hardware components at different load levels is currently in development by SPEC [2] and will be available in winter 2011/2012. The SERT tool will assess server efficiency based on partial benchmarks for CPU, memory, storage and system (Table 2.2). The tool will support IT managers in selecting energy efficient hardware for specific applications.

SPECpower_ssj2008 [2] was the first standard benchmark supporting energy efficiency assessment of volume class servers. It addresses mainly CPU related efficiency and thus provides a good assessment regarding CPU intensive workloads. The benchmark is published by manufacturers only for selected hardware though. Fig. 2.1 shows one example of SPECpower results for a volume server. The typical SPEC graph provides information on the average performance per watt across the range of loads as well as values for ten different load levels. Thus servers can be compared at different load levels from idle to 100%. For procurement purposes the complete SPECpower information (also containing detailed configuration information) should be requested from suppliers. Furthermore it should be considered that products are often tested in low configurations.

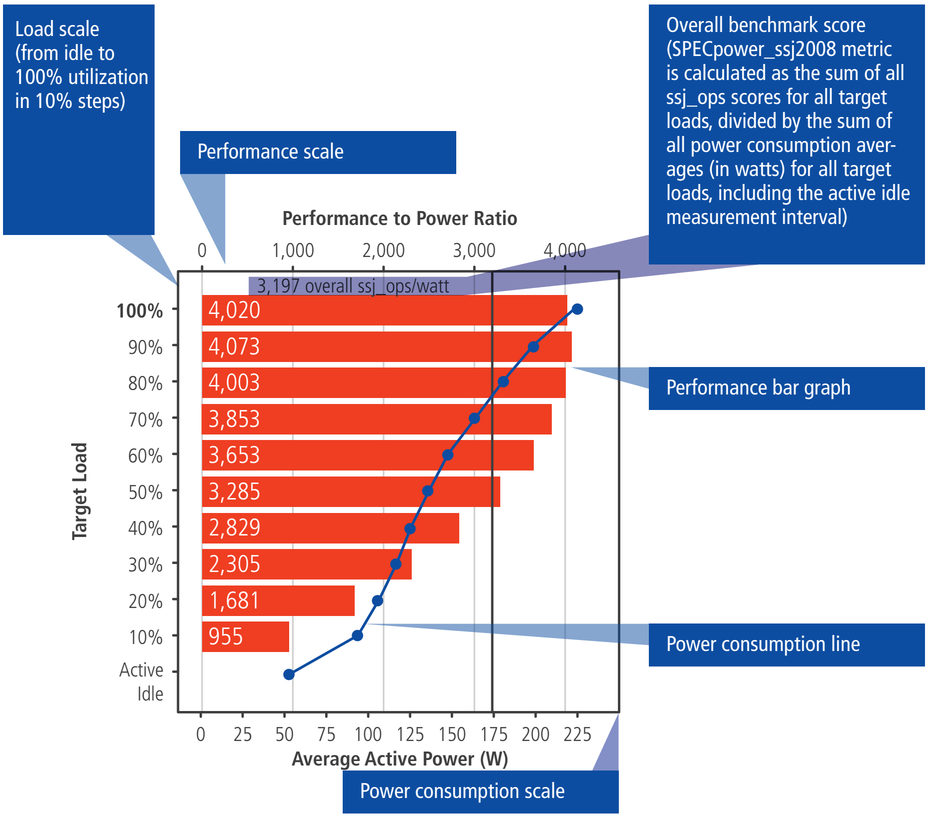


Fig. 2.1 **SPECpower diagram and key information**

Recommendations for best practice

Energy efficiency criteria and benchmarks for hardware selection

- Use the efficiency criteria from Energy Star for procurement if applicable. For servers operated at low loads, Energy Star Vers.1 requirements for idle mode may serve as reasonable efficiency indicators. Requirements for power supplies can be used for any type of equipment.
- Request SPECpower_ssj2008 (and SPEC-SERT as soon as available) benchmarking results from manufacturers. For SPECpower, consider the following issues:
  - It is a CPU centric benchmark, thus most representative for CPU intense workloads.
  - Servers may have been tested at rather low configuration (thus check configuration).
  - To arrive at robust interpretation, consider not only the overall score (overall operations per watt) but also the detailed benchmarking data.

2.1.1 CPU efficiency

CPUs are the most energy consuming components in servers, thus energy efficient CPU models with effective power management can strongly support efficiency.
CPU energy consumption depends on the specific voltage and the clock frequency. Power management at CPU or core level therefore is based on Dynamic Voltage and Frequency Scaling (DVFS) or switch-off of cores. Energy consumption of CPUs is often compared on the basis of the thermal design power (TDP) that indicates the maximum power the cooling system in a server is required to dissipate. However, TDP provides only limited information, as overall efficiency also strongly depends on power management. Manufacturers offer specific low power CPU versions that allow significant energy savings in practice, if the specific performance requirements can be met.
Energy efficiency of CPUs strongly depends on the effective implementation of power management. Common operating systems support power management based on the Advanced Configuration and Power Interface (ACPI) specifications for processor performance states and power consumption (P-States) and thermal management states (C-States). The new system and component controls enabled by ACPI Vs 3 provide higher-level power management engines allowing a finer grained power and performance adjusting based on demand. In many recent server models, predefined power profiles can be applied, e.g.:

- "High performance" (appropriate for servers that run at very high utilisation and need to provide maximum performance, regardless of power costs)
- "Power saver mode" / "Minimum power usage" (applied to servers that are run at low utilization levels and have more performance capability than really needed, using this mode may provide incremental power savings)
- "Balanced power and performance"

Figure 2.2 shows the positive effects of modern CPU power management in benchmark results (SPECpower) for the server product family HP ProLiant DL 380: the ratio of idle power to full load power has been strongly reduced from generation G5 to G7 of the specific server model. For the DL 380 G5 server, idle power (no load) was 33% (170 watts) lower than full load power (253 watts). For G7, it is about 75% lower compared to maximum power. This shows that new server technology is much more energy efficient at low load or idle operation, due to intelligent power management at CPU level. The computing performance for the specific server model on the other hand has been increased by more than a factor of three.

For hardware configuration in procurement, it is generally essential to check for concrete performance requirements to be met by the hardware

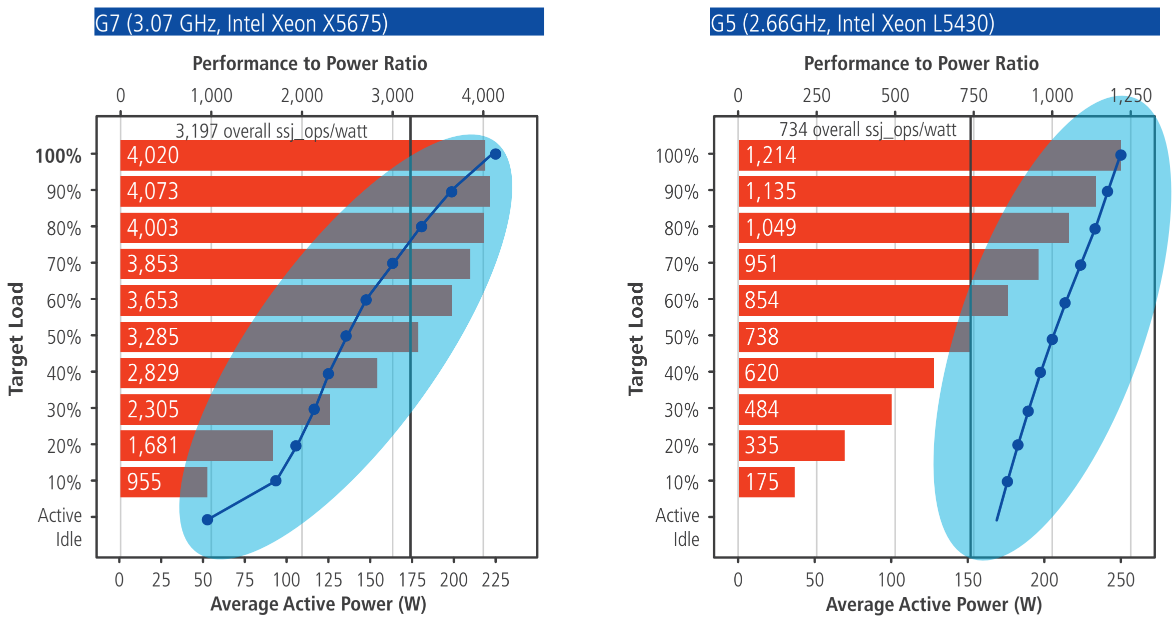


Fig. 2.2 **Example of SPECpower-benchmark for different server generations** *(G5, G7 Server from HP) [SPEC (2010, www.spec.org)]*

components. Different types of server workloads set different requirements regarding hardware performance that should be considered for efficient hardware configuration. A rough indication of hardware performance requirements for different workloads is given in Table 2.3.

Tab. 2.3 **Performance requirements of different server applications** *[5]*

Category	CPU	RAM	Hard disks	IO
File/print server	0	+	++	+
Mail server	+	+	++	0
Virtualization server	++	+++	++	++
Web server	+	+	0	+
Database server	++	++	+++	+
Application server	++	++	0	+
Terminal server	++	++	+	+

2.1.2 Power supply efficiency

The Energy Star program for servers [1] has set requirements for power supply efficiency defining levels for 10%, 20%, 50% and 100% load. The 80 PLUS Certification Scheme [3] also provides energy efficiency requirements for server power supplies but excludes the 10%-load level. For practical purposes and procurement, it is recommended to order power supplies that meet at least the 80 PLUS Gold level, which corresponds to 88% efficiency at 20% load and 92% efficiency at 50% load.
Standard rack servers commonly operated at low loads are often equipped with over-provisioned redundant power supplies. This results in significant energy losses due to a very low operating point of the equipment. Thus right sizing of power supplies is essential. It is supported for example by online power configuration tools offered by manufacturers and by tools for power capping assessment.
Some manufacturers (e.g. HP ProLiant G6 and G7 server series) provide specific hardware features to overcome unnecessary losses for redundant power supplies. Such hardware offers an operation mode that allows use of only one power supply until load exceeds a certain threshold. The second power supply stays in standby maintaining redundancy. This mode provides full power redundancy in case of a power supply or circuit failure.

Tab. 2.4 **Efficiency requirements for power supplies in the Energy Star programme and the 80 PLUS initiative [1, 3]**

	Power Supply Type	Rated Output Power	10% Load	20% Load	50% Load	100% Load
Energy Star Vs1	Multi-output (AC-DC & DC-DC)	All Output Levels	N/A	82%	85%	82%
	Single-output (AC-DC & DC-DC)	≤ 500 W	70%	82%	89%	85%
		>500–1,000 W	75%	85%	89%	85%
		> 1,000 W	80%	88%	92%	88%
Energy Star Vs2 Draft	Multi-output (AC-DC & DC-DC)	All Output Levels	N/A	85%	88%	85%
	Single-output (AV-DC & DC-DC)	All Output Levels	80%	88%	92%	88%
80 PLUS	Bronze	All Output Levels	N/A	81%	85%	81%
	Silver	All Output Levels	N/A	85%	89%	85%
	Gold	All Output Levels	N/A	88%	92%	88%
	Platinum	All Output Levels	N/A	90%	94%	91%

2.2 Power management at rack to data centre level

Going beyond hardware components and single server units, power management at the system level is also important to optimize overall energy efficiency.
As indicated above, the majority of servers are still utilised at modest workloads, thus there is large potential for energy savings to be achieved by hardware consolidation (see next chapter) or by power management at system level. As for the component level, power management at higher levels adjusts performance and power draw to the actual demand and powers off or throttles resources if not needed. Table 2.5 shows the various approaches of power management at different levels [7]. Some of the options are addressed in the following sections and in later chapters.

2.2.1 Capacity planning and energy management

Server management software provides essential tools for secure server operation but also for holistic power management. Server management tools can effectively help to reduce energy consumption as they facilitate the implementation of energy policies throughout the server system and provide features like provisioning, monitoring and configuration management that can strongly support system efficiency. Major features commonly are:

- provisioning
- monitoring
- deployment
- configuration management
- update control
- power management
- workload management

All larger hardware suppliers offer powerful server management tools. IBM (Systems Director) and HP (Systems Insight Manager including Insight Dynamics) offer very comprehensive management solutions capable of integrating third party systems. Fujitsu (Server View Site) offers products with basic functionalities that can be integrated in established management consoles from other suppliers. DELL is using the Altiris Total Management Suite. Sun and Acer provide consoles for their own environments.

Energy Management Suites (e.g. IBM Energy Manager)
Among many other features, this type of tool supports monitoring and collecting power consumption data, managing power including setting power savings options and power caps as well as automating power-related tasks. The latter include configuration of metering devices such as PDUs and sensors, setting thresholds, creating and setting power policies, calculating energy costs. For further information on energy management suites, see below.

Tab. 2.5. **Power management options from component to data centre level [7]**

Component Level	System Level	Rack Level	Data centre Level
• CPU (Package/core C-states, P-states, T-states, Thermal throttle) • Other components (D-states, L-states)	• S-states • Platform-based power management • Workload schedulers • Fan speed control	• System or node management • Application/load balancing • Chassis management	• Application/load balancing • Facilities and equipment monitors • Data de-duplication, etc. • Multi-rack management, dynamic consolidation

Capacity planning tools
(e.g. HP Capacity Planner)
Capacity planners, among other features, support IT managers with increasing server utilisation, reducing energy consumption and enhancing application performance. They allow collection of utilisation data for CPU cores, memory, network, disk I/O and power. Furthermore, they support workload planning or system changes and assessment of impact on resource utilisation. They also evaluate trends to forecast resource needs. For further information on capacity planning tools, see below.

Based on utilisation logs, the HP tool for example provides a good decision basis for consolidation measures by assessing the resource demand for merged applications. Figure 2.3 shows the example of a comparison of the utilisation of two systems indicating that peak performance is occurring at different times and the average load would increase only modestly in case of hardware consolidation.

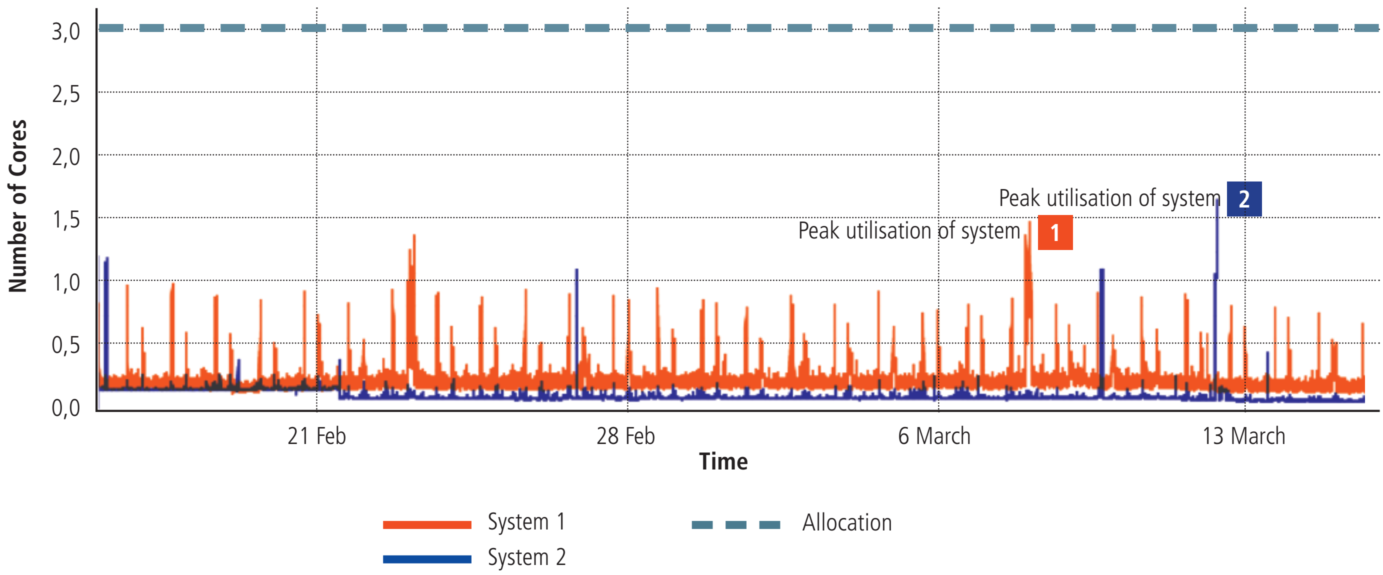


Fig. 2.3 **Comparison of CPU utilisation for "system 1" and "system 2" (see HP Capacity Planner)**

2.2.2 Power capping

Active allocation of power budgets to servers is also known as power capping. IT managers can specify power caps for servers according to real power requirements. Dynamic power capping reduces the maximum power demand of the system and thus optimises power provisioning beyond the level typically supported by power configurators offered by manufacturers.

The concrete savings achieved in practice depend on the level of the cap. The caps should be set in a way that power peaks are capped but computing performance is not visibly affected. Optimised capping requires an assessment of the workload and power consumption pattern. For relatively uniform workloads, caps can be set at average server load without significantly affecting performance. As a rule of thumb, caps should not be set lower than about midway between minimum and maximum power consumption of the servers. Some management tools also provide the option of time dependent specification of capping that defines different caps for different periods in the day depending on load pattern, power costs etc.

RECOMMENDATIONS FOR BEST PRACTICE

Energy efficient DC planning and management

- Use server management tools for capacity planning, workload and power monitoring and specific power management. Detailed descriptions and recommendations on use of power management features are supplied with the technical documentation of the server management suites.
- Use application and load balancing to optimise use of hardware resources.
- Use power capping to keep power demand at desired levels for the whole system.
- Benefit from optimised IT hardware resilience levels. Evaluate the level of hardware resilience actually justified in view of expected business impact of service incidents for each deployed service.
- Decommission unused services and completely remove the hardware. Assess the options for decommissioning low business value services by identification of those services which do not justify the financial and environmental cost.

2.3 Specific power management options for blade servers

Blade server technology is deployed both in data centres and server rooms. The blade server market has been the fastest growing market segment in the last few years and it is therefore important that the technology is as energy efficient as possible.

Blade chassis (see Figure 2.4.) typically include 7, 14 or more blade server modules, one or more management modules as well as KVM interfaces. Chassis support server-, storage- and network modules and may be optimised for specific applications and user types. Compared to standard rack servers, blade technology allows a reduction of some hardware components like power supplies, network I/O and wiring which are shared by several servers in the common enclosure.

Major benefits of blade systems are:

- High computing density and low space demand
- Reduced time for maintenance and upgrade of the system due to hot-plug replacement of modules and integrated management features
- Slightly higher energy efficiency as compared to rack servers if power management and cooling is optimised

Fig. 2.4 **Blade chassis**

Fig. 2.5 **Dual-node blade server**

However, if high blade densities are implemented, this results in high demand for infrastructure and cooling. High computing density increases power densities to 10–25 kW/rack. Consequently, standard cooling in data centres and server rooms is often not sufficient and specific cooling concepts are required. Thus energy efficiency of a blade concept also strongly depends on the overall system design.

Dual-node and multi-node concepts are partly based on a similar philosophy as blade servers. In the multi-node concept, a fixed number of server units (commonly 2 or 4) is combined in one rack-mounted chassis. Similar to blades, the servers share powersupplies and fans, however there are few expansion options. Thus multi-node technology is an approach to implement higher computing density at comparably low cost, often designed for purposes of small and medium enterprises. However, there are also special high performance dual-node servers available for example for blade systems which combine two server nodes in one blade. The main benefits of standard dual and multi-node systems are:

- Lower cost and space demand as compared to standard rack servers
- Slightly lower energy consumption due to shared power supplies and fans

2.3.1 Blade chassis and blade components

Larger power supplies are often more efficient, thus a lower number of larger power supplies in blade systems can increase energy efficiency compared to rack servers. However efficiency in practice also depends on the power demand in relation to the power supply capacity. Figure 2.6 shows the efficiency curve of a platinum labelled power supply [3] of 2,990 W rated power for a blade chassis indicating efficiencies between 92% and 95% across the load range. Efficient power supplies for blades should reach energy efficiency levels above 90% between 20% and 100% load.

For new product generations of blade and multi-node servers, some manufacturers provide several power supply models with different rated power that allow right-sizing according to the power demand. Power supply selection is supported by online power configurators offered by manufacturers.

Fewer and more efficient power supplies, more efficient fans and extended energy management options in the blade chassis offer higher energy efficiency as compared to standard rack servers in principal. However, efficiency in practice strongly

RECOMMENDATIONS FOR BEST PRACTICE

Selection of blade technology based on clear decision criteria

- Define and assess the main reasons for implementing blade technology in the data centre, e.g. space restrictions.
- Assess the benefits that are expected in comparison to rack technology and check if expectations are realistic.
- Check if virtualization may be an alternative solution considering the defined objectives.
- Evaluate the expected Total Cost of Ownership (TCO) and energy efficiency compared to other options (based on information provided by suppliers).

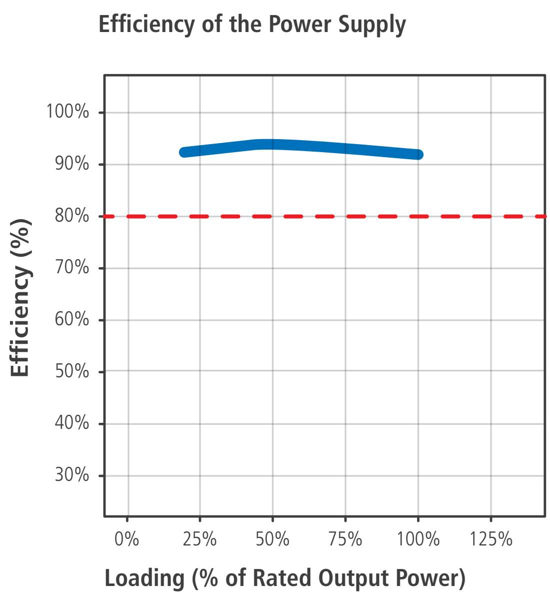


Fig. 2.6 **Blade power supply efficiency [3]**

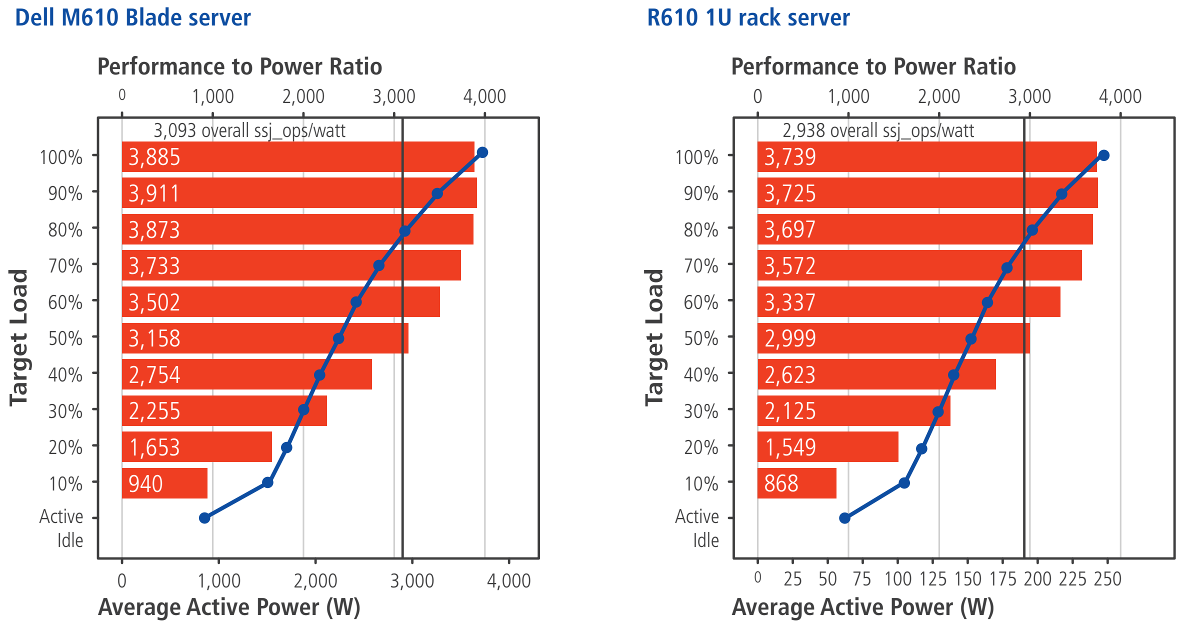


Fig. 2.7 **SPECpower_ssj2008 for a Dell M610 Blade server and R610 1U rack server.** The blade system includes 16 blades with identical processor configuration as the rack server (2 x Intel Xeon 5670, 2.93GHz). SPEC (2010, *www.spec.org*)

depends on the configuration of the chassis as well as on the use of the power management options. Chassis configured with only a few blades will clearly be less efficient, due to over-provisioning of cooling, power and network capacity.

For approximate comparison of the energy efficiency of blade servers versus standard rack servers, a fully configured blade system may be considered. Such a rough comparison based on energy efficiency data published by Dell is shown in Figure 2.7. Dell has published SPECpower-data (SPECpower_ssj2008) for blade systems and comparable rack servers in 2010 (www.spec.org).

The SPEC results show a maximum performance of 3,885 ops/watt at 100% load for the blade system and 3,739 ops/watt for the rack server system, indicating that the performance per watt or the energy efficiency at maximum load is 4% better in the blade system than in the rack solution. The difference increases to about 8% for low loads (10% load) and to 11% for idle operation.

Although this simple comparison must not be over interpreted (as SPECpower does only assess part of the server efficiency), it suggests that blade systems, even if fully configured and optimised for testing, show only slightly better energy efficiency than standard rack servers, especially at high loads. The difference is more significant at low load levels indicating a better overall power management in the blade system at low load.

Thus, blade solutions seem to offer only limited potential for increasing energy efficiency for example compared to virtualization. Similar to rack servers, there is also the option to combine blade hardware with virtualization, which allows for strong improvement of energy efficiency.

Challenges related to high heat densities at rack and row level are addressed in section 2.3.2 below.

Modern blade chassis contain management hardware and software that in combination with remote access controllers in the server blades allow a power inventory and power management of the individual blades. Specific management cards support a hardware and power demand inventory of the different blades. The remote access controller communicates the power budget information to the chassis management card that confirms the availability of power from the system level, based upon a total chassis power inventory. The CMC can set power policies at the system level and actual power consumption at each server module is monitored ensuring that instantaneous power consumption does not exceed the budgeted amount.

The basic functions of the power management in automatic mode are normally not visible for the system administrator. However, priorities for each server module can also be set manually, for example by selecting the lowest priority blades as the first to enter any power saving mode.

In blade chassis, dynamic power capping can be used even more effectively than for standard rack servers, since the dynamic power cap can be specified across multiple servers. Power caps can be dynamically adjusted by the onboard administrator and the service processor. Blades running lighter workloads receive lower caps. Since workload intensity and dynamics are normally different for the different blades, power peaks occur at different times. Consequently, the overall cap for the chassis can be set lower compared to the sum of individual caps for single blades. HP has calculated power savings and reduced TCO for a blade centre where power supply design was based on power capping. Maximum power and power provisioning cost was reduced by about 20% as compared to the approach without power capping [HP2011].

RECOMMENDATIONS FOR BEST PRACTICE

Consider procurement criteria for selecting energy efficient blade hardware

- Define the workloads and expected workload levels to be run on the blade systems.
- Compare costs and energy efficiency of blade systems from different vendors.
- Request product information from suppliers regarding.
  - Total Cost of Ownership (TCO) .
  - Overall energy efficiency (e.g. SPECpower_ssj2008, SPEC-SERT as soon as available).
  - Energy efficient hardware components, e.g. efficiency and right sizing of power supplies.
  - Management tools especially addressing power management and optimization of system design.
- Select equipment offering highest energy efficiency for the workload types and levels you are addressing and adequate power management options.

2.3.2 Blade system – power and cooling issues

In practice, design of efficient blade server systems is often an underestimated challenge, especially if large high-density systems are implemented. The main challenges are:

- Sufficient cooling capacity and appropriate cooling design to cope with high heat densities
- Sufficient power capacity and distribution (local PDU capacity, power wiring etc.)

Traditional cooling concepts often allow only 2-3kW/rack, which is 10 times less than the power of a fully populated blade rack. This means that standard cooling concepts of data centres and server rooms are often not appropriate for larger blade systems and have to be modified.

RECOMMENDATIONS FOR BEST PRACTICE

Use of management tools to optimize energy efficiency of blade systems

- Use management tools and intelligent network and power devices for monitoring of power consumption and load for your blade system.
- Analyse options to balance and manage loads and power consumption within and across blade chassis and racks.
- Use power capping and power balancing features of blade chassis.
- Do a first order estimation on power/cooling capacity demand based on power calculators offered by manufacturers.
- Assess real power demand with available management tools for complete duty cycles and set power caps according to peak load. Adjust power and cooling to fine-tune system based on power caps.

Table 2.6 shows typical options for the design of different blade densities depending on business requirements and constraints such as infrastructure and cooling capacity. Different blade density levels allow the following options for cooling concepts [Rasmussen 2010]:

- Spreading the heat load of blade chassis to different racks: Individual blade chassis are mounted to different racks to spread heat load. For this concept, the percentage of blade chassis in the total system has to be very low.
- Dedicating cooling capacity: Excess cooling capacity is specifically dedicated to the blades. For this approach the percentage of blades in the system has to be relatively low, as only the existing cooling capacity is used.
- Installing supplemental cooling: Supplemental cooling is provided for the blade racks. Power density per rack can be up to 10kW. The approach allows for good floor space utilisation and high efficiency.
- Definition/design of high-density area: Specific area in the data centre is dedicated to blades (high density row or zone). High efficiency and high floor space utilisation. Density up to 25kW. Area has to be planned and re-designed.
- Design of high density centre: High density blade racks throughout the data centre. An extreme and rather uncommon approach, which for most situations leads to significant costs and strong underutilisation of infrastructure.

In existing data centres, there are often certain limits for the deployment of blade technology defined by the specific infrastructure. For example, a standard raised floor system may not allow a higher power density than 5kW per rack. Proper specification of power and heat density is an important prerequisite to allow energy-, space- and cost-efficient system design.

Another essential point regarding energy efficiency at the system level is to avoid over-provisioning of infrastructure and cooling. Density specification should take into account both spatial and temporal variability, e.g. different local power densities in data centres regarding blade racks and standard racks and variation over time where density may increase. Thus, power density has to be specified either at rack or at row level. For larger systems the row level is more appropriate, since cooling and power distribution is mainly row-based. As far as possible, it is recommended that density specifications are defined for a rack or row. They should be left unchanged for the time of operation of the specific rack or row. Thus, implementation of a new technology with different density level should be done in a new rack or row. However, there are also alternatives to this approach allowing some variation of power densities in installed racks or rows:

- Adding hot pluggable UPS modules
- Using hot swappable rack PDUs
- Adding cooling capacity with rack-mounted devices

Tab. 2.6 **Configuration of blade systems at rack level and related requirements for cooling** *[after Rasmussen 2010]*

No Chassis/rack	Spreading load across racks	Dedicating cooling capacity	Additional cooling	High density area	High density centre
1	Possible in most DCs	Possible in most DCs	Possible in most DCs	Not cost efficient	Not cost efficient
2	Rarely practical	Possible in most DCs	Possible in most DCs	Not cost efficient	Not cost efficient
3	Not possible	Possible in most DCs	Possible in most DCs depending on specific solution	Maximum for optimised efficient raised floor systems	Not cost efficient
4	Not possible	Rarely practical	Depending on specific solution	Hot air scavenging systems	Hot air scavenging, room redesign
5	Not possible	Not possible	Not possible	Hot air scavenging systems	Hot air scavenging, room redesign
6	Not possible	Not possible	Not possible	Extreme costs	Extreme costs

For the definition of densities for rows, it has been recommended to define a maximum ratio of peak to average power of 2 for typical row designs. Where double average power is exceeded by specific racks, IT loads should be redistributed within the row or to other rows. Overall, it obviously makes sense to distribute higher density racks in the row. Power and cooling management systems can be used to define rules for deploying installed capacities e.g. allowing a rack to exceed average power only if the power demand of a neighbour rack is significantly below average.

An important issue is how to deal with future developments regarding future needs for IT extension. It is clearly not advisable to implement infrastructure covering maximum future capacity from the beginning, as this would mean over-capacity and high costs over a longer period of time. It is generally recommended to install all piping and wiring for full expansion of capacity but to install the power and cooling equipment at later stages based on specific demand. This approach allows to prepare all the basic infrastructure of the building but to implement the specific equipment according to power and cooling demand of the IT when needed.

2.4 Server virtualization

Server virtualization offers great potential for energy savings. The technology allows for the consolidation of workloads on less physical hardware, thereby strongly reducing power and cooling demand. Overall virtualization offers a number of advantages for the effective design of IT systems in server rooms and data centres , as for example:

- Reduction of hardware and space requirements via deployment of virtual machines (VMs) that can be run safely on shared hardware, increasing server utilisation from 5–15% to 60–80%.
- Test and Development Optimisation – Rapidly provisioning test and development servers by reusing pre-configured systems enhancing developer collaboration and standardizing development environments.
- Reducing the cost and complexity of business continuity (high availability and disaster recovery solutions) by encapsulating entire systems into single files that can be replicated and restored on any target server.

Established virtualization platforms like VMWare, Microsoft Hyper-V and Citrix XEN offer many features like high availability, failover, distributed resource scheduling, load balancing, automated backup functions, distributed power management, server-, storage- and network VMotion etc.

The primary technology options for server virtualization include:

- Physical partitioning
- Virtualization based on an underlying operating system
- Application virtualization e.g. Microsoft Terminalserver, Citrix XenApp
- Hypervisor-based virtualization:
  - VMware ESX
  - Citrix /Open-Source: XENServer 5
  - Microsoft Hyper-V

Considering the market which is dominated by only a few products, the following chapter focuses on the hypervisor-based products: VMware ESX, Microsoft Hyper-V and Citrix XEN Server.
The market leading virtualization platforms VMWare ESX/ESXi/Vsphere4, Microsoft HyperV and Citrix XEN offer support for most common standard guest operating systems. They provide management consoles for administration of smaller server environments as well as data centre level administration.

VMware was the first product on the market in 2001. Its architecture predates virtualization-aware operating systems and processors such as Intel VT and AMD-V. VMware ESX/VSphere4 offers powerful administration tools like VMotion of virtual machines across servers, storage VMotion, storage overprovisioning, desktop and network virtualization, virtual security technology, and it delivers a complete virtualization platform from desktop through data centre up to cloud computing.

Microsoft Hyper-V Server contains the Windows Hypervisor, Windows Server driver model and virtualization components. It provides a small footprint and minimal overhead. It plugs into existing IT environments, leveraging existing patching, provisioning, management, support tools, and processes. Some of the key features in Microsoft Hyper-V Server 2008 R2 are live migration, cluster shared volume support and expanded processor and memory support for host systems. Live migration is integrated with Windows Server 2008® R2 Hyper-V™. Hyper-V™ live migration can move running virtual machines without downtime.

Depending on user requirements, Citrix XENServer may offer a cost effective way of implementing virtualization, since basic elements like the bare hypervisor, resilient distributed management architecture, XENServer management and conversion tools come for free. Advanced management and automation features like virtual provisioning services, distributed virtual switching, XENMotion, live migration, live memory snapshots and revert, performance reporting and dynamic workload balancing make the XENServer comparable to the other two products. However, these features are part of the advanced commercial editions.

BMC Software, Eucalyptus Systems, HP, IBM, Intel, Red Hat, Inc. and SUSE announced the formation of an Open Virtualization Alliance, a consortium committed to fostering the adoption of open virtualization technologies including Kernel-based Virtual Machine (KVM). The consortium complements the existing open source communities managing the development of the KVM hypervisor and associated management capabilities, which are rapidly driving technology innovations for customers virtualizing both Linux and Windows® applications. The consortium intends to accelerate the expansion of third party solutions around KVM and will provide technical advice and examples of best practice.

2.4.1 Energy saving potential of virtualization

Virtualization is one of the most powerful technologies for reducing energy demand in data centres and server rooms. Consolidation of server hardware by concentrating workload on a lower number of physical servers often allows energy savings of 40% to 80% and sometimes more, depending on the specific case. Current technology provides the possibility to implement virtualization with consolidation factors of at least 10–20, depending on the specific systems and requirements.

Figure 2.8 shows the example of a server consolidation by virtualization in the German Federal Ministry for the Environment. The specific measures allowed energy savings of about 68%. The case involved a reduction of hardware to 2 physical servers running VMware ESX [4].

Another example from IBM [5] for a virtualization project involving blade server technology suggests energy savings of more than 90% if all relevant measures at hardware and infrastructure level are considered.
Such examples illustrate that consolidation by virtualization is one of the major options to significantly increase energy efficiency in data centres. However, as for the other IT based approaches the full saving potential can only be accessed if the infrastructure including power supply and cooling is addressed in parallel.

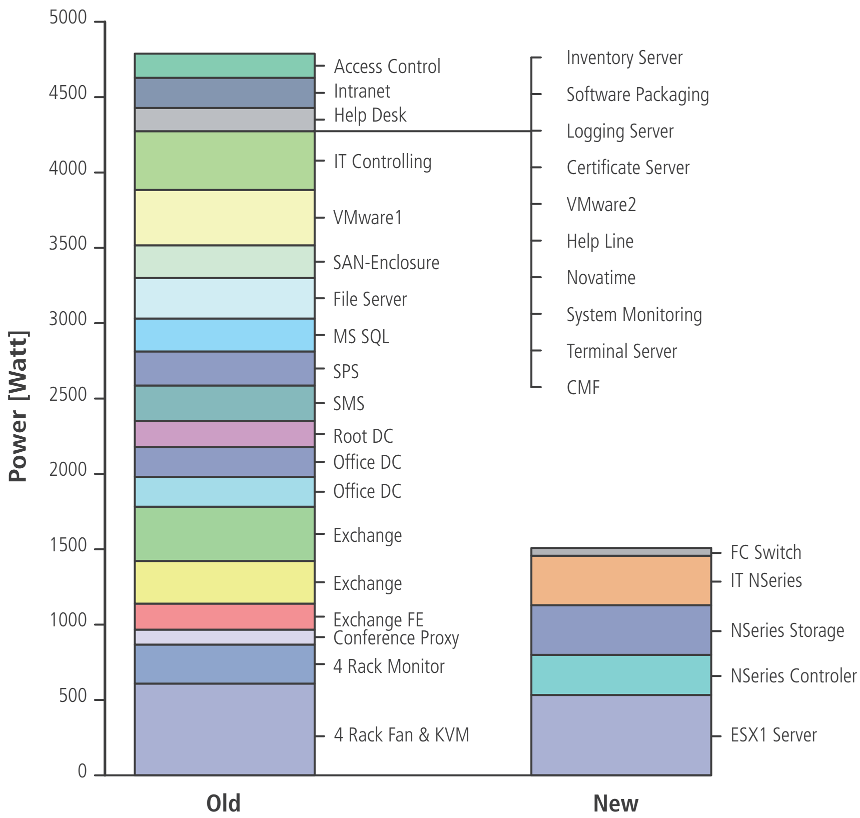


Fig. 2.8 **Reduction of energy demand by virtualization in a case study [4]**

2.4.2 Requirements and tools for virtualization planning

Virtualization in data centres should be based on a virtualization strategy that involves an evaluation and identification of appropriate server candidates.
For such an evaluation, data on performance, system utilisation, end-of-service timelines, business area and application specification is collected. Once the candidates for virtualization have been identified, application specifications and machine load are analysed. Performance evaluation is conducted to assess among others the following requirements as a basis for hardware selection:

- CPU performance
- Required memory
- Disk I/O intensity
- Network requirements
- OS configuration

Several applications can typically be consolidated to a single physical server, which is immune to hardware failure and power interruptions while possessing the ability to load-balance. To achieve this goal, host servers may contain dual power supplies, mirrored hard drives and teamed network interface cards. For a centralized storage solution, a Storage Area Network (SAN) with full fault-tolerant capabilities can be used. Load balancing can further be supported by virtual machine migration between physical servers.

Depending on the type of workloads, a consolidation ratio between 10:1 and 20:1 can be considered. Regarding memory requirements, many virtualization environments offer the feature of memory over provisioning. By means of this feature the sum of memory allocated to all virtual machines can exceed the available physical memory by a factor of 2 to 3.

Virtualization is rarely done for energy saving purposes only. Thus although high energy savings are normally guaranteed, successful virtualization projects typically require thorough planning, which also involves ROI and TCO calculations.

Summing up the relevant cost factors, the TCO for the new virtual server deployment is calculated. A short-term and long-term ROI calculation can be done to assess time-related costs.
The key to successful ROI calculation is to understand virtualization costs. The obvious expenses for virtualization projects are hardware, software (incl. licensing) and labour. Virtualization may involve buying of new, more powerful servers, upgrade of storage, network and security etc. Costs for staff training and management are an additional issue. All these aspects have to be factured into the ROI calculation.

Different software tools available on the market support virtualization planning as well as ROI and TCO calculation. For example, the Assessment and Planning (MAP) Toolkit from Microsoft supports planning for migration including TCO and ROI calculation. The MAP-Toolkit is an inventory, assessment, and reporting tool that can assess IT environments for various platform migrations and virtualization without the use of software agents. MAP's inventory and readiness assessment re-

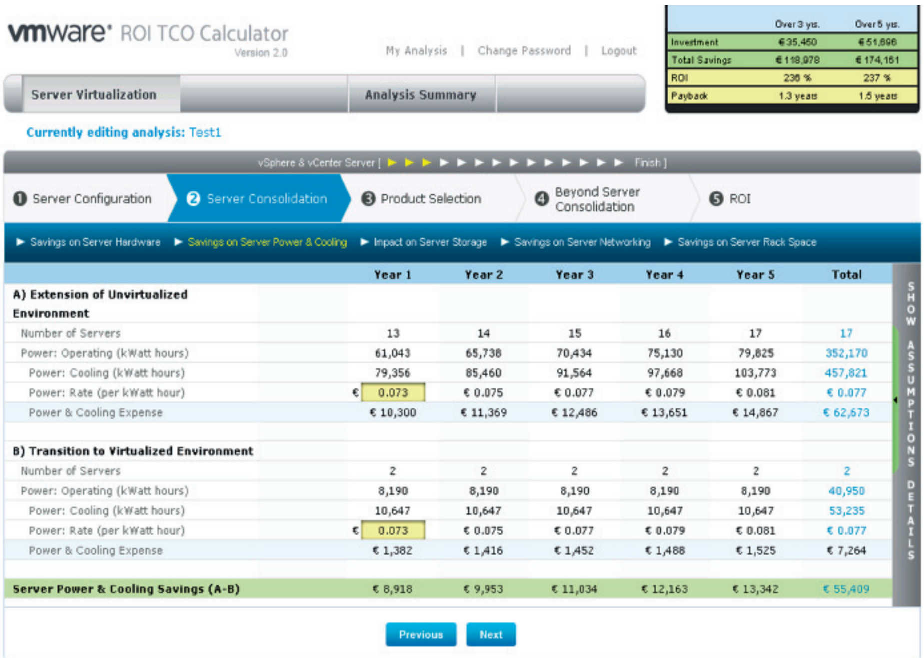


Fig. 2.9 **Example of a ROI/TCO Calculator from VMWare [6]**

ports generate specific upgrade recommendations for migration to Windows Vista and Windows Server 2008 operating systems and also for virtualization. It provides recommendations on how physical servers can be consolidated in a Microsoft Hyper-V virtualized environment. In addition the Microsoft Integrated Virtualization ROI Tool supports the calculation of potential power cost savings with Hyper-V prior to deployment. The tool provides support for examining current production and development servers, desktop and application virtualization opportunities by quantifying potential savings, service level benefits, investments and ROI.

The TCO/ROI methodology offered by VMware (available as an online tool) allows to compare TCO savings, required investments and business benefits of virtualization solutions. It is based on standard financial techniques, VMware field and customer data and user metrics. Based on user-specific data, key figures like savings, investments, ROI, NPV savings, TCO opportunities and payback periods are calculated. Where specific user data is not available, statistical data from industry is provided and may be used for calculations.

2.4.3 Power management in virtualized environments – virtual server migration

Current software solutions for server virtualization support the migration of virtual machines and a temporary shut-down of hosts to reduce power demand. One example providing such features is VMwareVsphere4 (Distributed Power Management DPM). DPM monitors the resource use of the running virtual machines in the cluster. If there is excess capacity, DPM recommends to move some virtual machines between hosts and to put some hosts into standby mode to save power. In case of insufficient capacity, DPM powers on standby hosts again.
Power management can be operated in either manual or automatic mode. In automatic mode virtual machines are migrated and hosts are moved into or out of standby mode automatically. Automatic settings can be overridden on a per-host basis and power management can also be enabled by a scheduled task.

The goal of VMware DPM is to keep the utilisation of ESX hosts in the cluster within a target range. DPM must meet the following requirements to be an effective power saving solution:

- Accurate assessment of workload resource demands. Overestimating can lead to less than ideal power savings. Underestimating can result in poor performance and violations of DRS resource level SLAs.
- Avoiding powering servers on and off too frequently even if running workloads are highly variable.
- Rapid reaction to sudden increase in workload demands so that performance is not sacrificed when saving power.
- Selection of the appropriate hosts to power on or off. Powering off a larger host with numerous virtual machines might violate the target utilisation range on one or more smaller hosts.
- Redistribution of virtual machines intelligently after hosts are powered on and off by seamlessly leveraging DRS.

RECOMMENDATIONS FOR BEST PRACTICE

Effectively assessing and selecting virtualization solutions:

- Develop a virtualization strategy and assess servers to select good candidates for virtualization.
- Assess requirements regarding CPU performance, memory, Disk I/O intensity, Network requirements, OS configuration.
- Consider the appropriate virtualization ratio and mix of workloads (1:6 to 1:20 depending on workload characteristics).
- Check products from different suppliers regarding required features for your specific purposes; consider licensing policies, power management features and price. The different main products on the market have different advantages depending on the specific application needs.
- Do TCO and ROI calculations to identify the benefits of reduced cost for power supply and cooling. Models provided by suppliers shall be refined according to the needs of the specific organisation.
- Consider power management options allowing VM migration and temporary shut-down of server hardware.
- Consider changed requirements for cooling and power supply (reduced and dynamically changing power and cooling demand) and check options of some redesign for cooling.

The basic way to use DPM is to power on and shut down ESX hosts based on typical utilisation patterns during a workday or week. For example, services such as email, fax, intranet, and database queries are used more intensively during typical business hours from 9 a.m. to 5 p.m. At other times, utilisation levels can dip considerably, leaving most of the hosts underutilised. Their main work during these off hours might be performing backup, archiving, servicing overseas requests etc. In this case consolidating virtual machines and shutting down unneeded hosts reduces power consumption.

The following approaches may be used to manually adjust DPM activity:

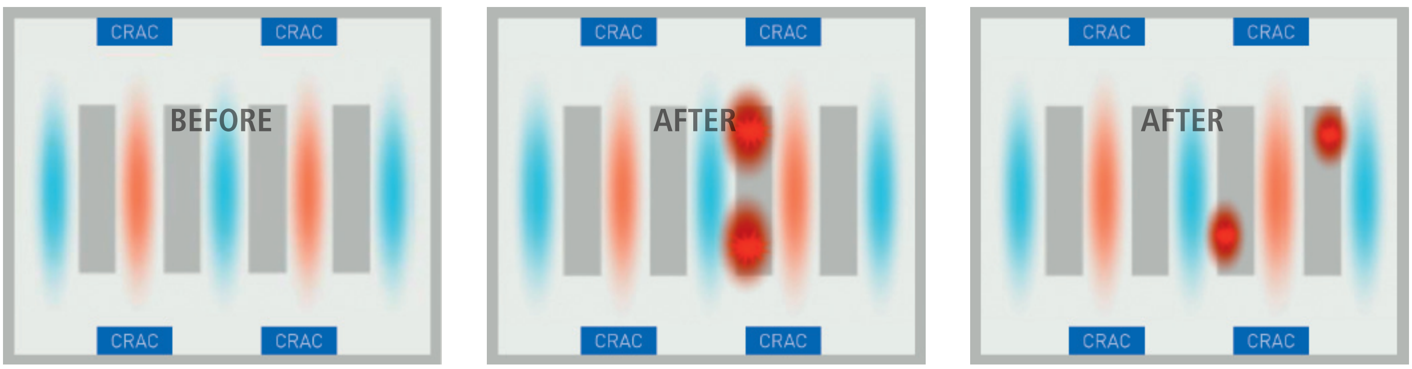
- Increasing the Demand-Capacity-Ratio Target: to save more power by increasing host utilisation (consolidating more virtual machines onto few hosts) the value for the Demand-Capacity-RatioTarget could be increased from default (e.g. from 63% to 70%).
- Using VMware DPM to force the powering on of all hosts before business hours and then selectively shut down hosts after the peak workload period. This is a more proactive approach that would avoid any performance impact of waiting for VMware DPM to power on hosts in response to sudden spikes in workload demand.

Each ESX host's resource utilisation is calculated as demand/capacity for each resource (CPU and memory) where demand is the total amount of the resource needed by the virtual machines currently running and capacity is the total amount of the resource currently available on the host. Thus power management of hosts is executed depending on CPU and host's memory resource utilisation compared to the defined utilisation range. For each host evaluated for a power-off recommendation DPM compares costs, taking into account an estimate of the associated risks with a conservative projection of the power-savings benefit that can be obtained.

2.4.4 Cooling and infrastructure for virtualized systems

While significantly reducing overall power demand, virtualization especially in larger systems may cause increased rack power density. Power management by migration of virtual machines furthermore leads to dynamic spatial change of power and heat density, thus locally increasing the demand for power and cooling. Appropriate power and cooling concepts have to be used to meet the demand of virtualized environments and to avoid hot spots.
If the total power and cooling capacity is not adapted to the lower power demand, PUE will worsen after virtualization. Virtualization can reduce cooling load in a data centre to very low levels which can cause negative effects. Thus right-sized power and cooling is crucial for exploiting energy saving potentials. It is also essential to reduce fixed losses by considering the following measures:

- Scaling down power and cooling capacity to match the load
- VFD fans and inverter pumps that are controlled by cooling demand
- Using equipment with higher efficiency
- Cooling architecture involving shorter air paths (e.g. row based)
- Capacity management system to adopt capacity to demand
- Blanking panels to reduce in-rack air mixing



Constant loads → Stable cooling

Migrating high-density loads → Unpredictable cooling

Fig. 2.10 **Heat density before and after virtualization [5]**

In a conventional environment involving traditional raised floor, room-based cooling can be configured to adequately cool hot spots by rearranging vented floor tiles. Changing requirements due to the dynamic migration of virtual servers, however, also require dynamic cooling solutions. A solution to this challenge is to position cooling units within the rows and get them equipped to sense and respond to temperature changes. Placing cooling units close to the servers allow short air paths between cooling and the load. Dynamic power variation in virtualized environments is a major reason for moving towards row- or rack-based cooling.

Accurate information about demand for power and cooling capacities is crucial in order to respond to changing load profiles over time. Capacity management provides instrumentation for real time monitoring and analysis of power-, cooling- and physical space capacities and enables the effective and efficient use throughout the data centre. Areas of available or dangerously low capacity can be identified. Capacity management systems should be able to handle the following issues:

- Change of load density and location – Virtualization can create hotspots e.g. by VM migration.
- Dynamic system changes – Maintaining system stability may become a challenge if multiple parties are making changes without centralized coordination.
- Interdependencies – Virtualization makes the shared dependencies and secondary effects in the relationship between power, cooling and space capabilities more complex.
- Lean provisioning of power and cooling – During virtualization the power and cooling load goes down and rises again as new virtual machines are created. This can be handled by usage of scalable power and cooling systems.

RECOMMENDATIONS FOR BEST PRACTICE

Energy efficient management of virtualized systems:

- Implement a strict policy for implementing and managing virtualised servers. Avoid uncontrolled server sprawl.
- Use virtual machine migration tools to shut down hardware at times of low loads. Use automatic power management settings for the start and develop own customised settings in a subsequent stage based on the typical operation patterns.
- Reduce cooling according to demand and implement equipment for dynamic local cooling if needed. Address the demand for dynamic spatial changes.
- Adapt IT processes and workflows regarding deployment of virtual machines, data recovery/backup processes, patch administration, availability considerations.

Further reading

HP (2011): HP Power capping and HP Dynamic power capping for ProLiant servers. Hewlett Packard Development company.
SPEC (2011): Server Efficiency Rating Tool (SERT) TM Design Document. 3rd draft. Standard Performance Evaluation Cooperation
Rasmussen, N. (2010): Strategies for deploying blade servers in existing data centres. White paper 125. APC Schneider Electric
80 PLUS (2011): 80 PLUS power supplies. *www.plugloadsolutions.com*
Schäppi B. et al (2009) Energy and cost savings by energy efficient servers. IEE E-Server best practice cases. Brochure 2009
IBM (2011) Server Management suite, Module Active Energy Manager *www-03.ibm.com/systems/software/director/aem/*
HP (2011) Server managment suite «Systems Insight Manager» *www.hp.com*
VMware DPM: Information Guide: VMware Distributed Power Management Concepts and Use. *www.vmware.com*
VMware TCO: VMware ROI TCO Calculator, Overview and Analysis. *http://roitco.vmware.com/vmw/*

References

[1] **EPA (2010):** Energy Star ENERGY STAR® Program Requirements for Computer Servers (vers 1.1)
[2] **SPEC (2010):** SPEC power and performance. Benchmark methodology 2.0. Standard Performance Evaluation Cooperation
[3] **80 PLUS (2011):** 80 PLUS power supplies. *www.plugloadsolutions.com*
[4] **Schäppi B. et al (2009):** Energy and cost savings by energy efficient servers. IEE E-Server best practice cases. Brochure 2009
[5] **BITKOM (2010):** Bitkom/Beschaffungsamt des Bundesministeriums des Innern, Leitfaden Produktneutrale Leistungsbeschreibung x86-Server, 2010
[5] **Comtec Power:** Overcoming the Challenges of Server Virtualization. *www.comtec.com*
[6] **VMware TCO:** VMware ROI TCO Calculator, Overview and Analysis. *http://roitco.vmware.com/vmw/*
[7] **The Green Grid (2010):** White paper Nr. 33 "A roadmap for the adoption of power-related features in servers", Pflueger, J., et al., The Green Grid, 2010

3 Data Storage Equipment

Marcos Dias de Asuncao, Laurent Lefevre, INRIA

Information is at the core of any business, but storing and making available all the information required to run today's businesses has become a real challenge. With the storage needs of organisations expected to grow by a factor of 44 between 2010 and 2020 [1], strategies for high efficiency have never been so popular. The constant fall in the price per MB of storage led to a scenario where it is simpler and less costly to add extra capacity than to look for alternatives to avoid data duplicates and other inefficiencies.
However, as the cost of powering and cooling storage resources becomes more of an issue, inefficiencies are no longer accepted. Studies show that large enterprises are currently faced with the difficult task of providing sufficient power and cooling capacity, while midsize companies are challenged with finding enough floor space for their storage systems [2]. As data storage accounts for a large part of the energy consumed by data centres, it is crucial to make storage systems more energy efficient and to choose the appropriate solutions when deploying storage infrastructure.

This chapter discusses a few technologies that support the energy efficiency of data storage solutions. Moreover, it provides recommendations for best practice that, in addition to the use of the discussed solutions, can improve the energy efficiency of storage infrastructure in enterprises and data centres.
Storage solutions such as disk arrays include drives that provide the raw storage capability and additional components to interface with the raw storage and improve overall reliability. We refer to the individual medium components that form the raw storage as devices (e.g. tape loaders, hard disk drives and solid state drives). Composite storage solutions such as network attached products are referred to as storage elements. When discussing schemes for improving the energy efficiency of storage solutions, these are mainly the two levels at which most techniques apply. Hence, we first present energy efficient concepts for individual devices, and then analyse how these techniques are used and combined to improve the energy efficiency of elements.

3.1 Storage Devices

3.1.1 Tape based systems

Tapes are often mentioned as one of the most cost-efficient types of media for long-term data storage. However, analyses [3][4] indicate that:

- Under given long-term storage scenarios, such as backup and archival in mid-sized data centres, hard disk drives can be on average 23 times more expensive than tape solutions and cost 290 times more than tapes to power and cool.
- Data consolidation using tape-based archival systems can considerably decrease the operational cost of storage centres. Tape libraries with large storage capacity can replace islands of data via consolidation of backup operations, hence reducing costs with infrastructure and possibly increasing its energy efficiency.

With an archival life of 30 years and large storage capacity, tapes are an appealing solution for data centres with large long-term backup and archival requirements. Hence, for an environment with multiple tiers of storage, tape-based systems are still the most power-efficient solutions when considering long-term archival and low retrieval rate of archived files. There are disk library solutions that attempt to minimise the impact of the energy consumption of disk drives by using techniques such as disk spin-down. These technologies are further discussed below.

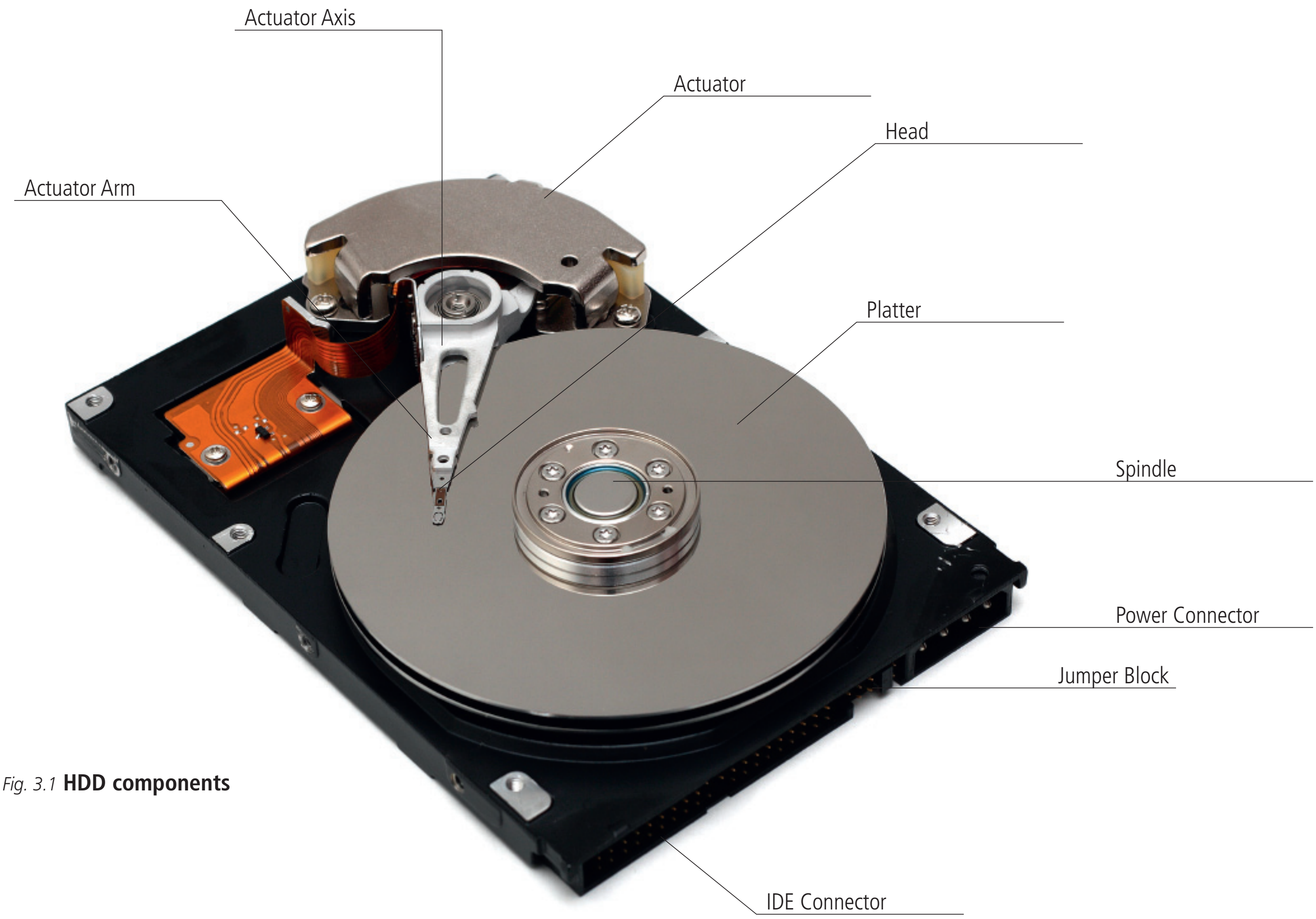


Fig. 3.1 **HDD components**

3.1.2 Hard Disk Drives (HDDs)

HDDs have long been the preferred media for non-volatile data storage that offers fast write and retrieval times. Moving parts such as motors and actuator arms account for most of the power that HDDs consume (see Figure 3.1). To improve data throughput of HDDs, manufacturers increase the speed at which platters rotate, thus further increasing their power consumption. Platters spinning at speeds of 15K RPMs are common for current high-throughput HDDs.

Several techniques are used to improve the energy efficiency of HDDs, including storing data in certain regions of the platters to reduce the mechanical effort when retrieving data, controlling the rotation speed of the platters, and reducing the power consumption during idle periods. A common technique, termed as disk spin-down, consists in spinning platters down and parking the heads at the secure zone after a factory-set period of inactivity. Moreover, instead of stopping platters completely, some drives spin the platters at variable speed according to the read/write load.

Some HDDs implement multiple idle and standby states. Different actions are taken as the period of inactivity increases (e.g. initially the servo system is disabled, then heads are parked, and later platters are spun down). Seagate's PowerChoice technology [5] is an example, where the number of disabled components increases as the drive reaches certain idleness thresholds. The intermediate idle states have recovery times that are generally shorter than recovering from a disk spun-down situation.
Table 3.1 shows that the consumption in standby is about 50% less than the idle consumption. Such approaches may allow substantial savings on RAID systems and Massive Arrays of Idle Disks (MAIDs).
As spinning disks down can compromise performance, manufacturers explore additional

Tab. 3.1 **PowerChoice technology profile for a Constellation 2.5-Inch drive**

State	Power (W)	Power Savings* (%)	Recovery Time (sec.)	Default Timer to Enter
Idle	2.82	0	0	n/a
Idle_A	2.82	0	0	1 sec.
Idle_B	2.18	23	0.5	10 min.
Idle_C	1.82	35	1	30 min.
Standby_Z	1.29	54	8	60 min.

* Power savings estimates and recovery times are preliminary; figures based on Seagate Constellation SAS 2.5-inch hard drive.

techniques such as larger cache sizes and read/write command queuing. Furthermore, to benefit from techniques such as spin-down and variable spinning speed, schemes have been proposed at the operating system and application levels to increase the length of periods of disk inactivity. Some of these approaches consist of rescheduling data-access requests by modifying the application code or data layouts. There are also less intrusive techniques that provide compiler customisations that re-schedule the data access requests at compilation without the need of modifying the application source code. Although these techniques can reduce power consumption, it has also been argued that frequent on-off cycles may reduce the life time of HDDs.

Fig. 3.2 **Picture of a 2.5-inch HDD atop a 3.5-inch HDD (from wikipedia)**

As motors and actuators are responsible for most power consumed by hard disk drives, an approach for making drives more energy efficient is to use Small Form Factors (SFFs). As 2.5-inch HDDs are around a quarter the size of larger hard-disk drives (3.5-inch, see Fig. 3.2), a chassis designed with enough volume for 16 3.5-inch drives might be redesigned to hold up to 48 2.5-inch hard-disk drives without increasing the overall volume. High-performance hard drives in 2.5-inch enclosures show reduced power consumption as motors and actuators are smaller and thus also emit less heat. Manufacturers claim that for Tier-1 2.5-inch hard disk drives, IOPS/W can be up to 2.5 times better than comparable 3.5-inch Tier-1 drives [6]. In addition, less power is required for cooling due to smaller heat output and reduced floor space requirements.

Table 3.2 shows the approximate power consumed by two models of high performance hard disk drives produced by Seagate. It is evident that the smaller form factor takes substantially less power. When active, it consumes approximately 46% less power than its 3.5-inch counterpart, whereas this difference can reach 53% when the disk is idle. Considering the cost to power only 24 drives over a year, based on active power consumption and a price of 0.11€ per KWh, the difference between 3.5-inch drives and 2.5-inch HDDs would be about 140€ per year. In data centres with storage systems with hundreds or thousands of disks, the savings can amount to thousands or ten thousands of Euros.

Tab. 3.2 **Power consumption of two of Seagate's high performance HDDs**

Specifications	Cheetah 15K.7 300GB*	Savvio 15K.2 146GB*	Difference
Form Factor	3.5"	2.5"	–
Capacity	300GB	146GB	–
Interface	SAS 6Gb/s	SAS 6Gb/s	–
Spindle Speed (RPM)	15K	15K	–
Power Idle (W)	8.74	4.1	53% less
Power Active (W)	12.92	6.95	46.2% less

* Data obtained from the specification sheets available at the manufacturer's website.

3.1.3 Solid State Drives (SSDs)

SSDs are equipped with, among other components, flash memory packages and a controller responsible for various tasks. SSDs rely on NAND-based flash memory that employs one of two types of memory cells according to the number of bits a cell can store. Single-Level Cell (SLC) flash stores one bit per cell and Multi-Level Cell (MLC) memories can often store 2 or 4 bits per cell. Most of the affordable SSDs rely on MLC while high-end devices are often based on SLC.

SSDs are more energy efficient and reliable due to the lack of mechanical parts such as motors and actuators. Moreover, they create less heat and can be packed into smaller enclosures, thus decreasing the floor space and cooling requirements. Table 3.3 presents a simple comparison between a Seagate's Pulsar enterprise SSD and a high performance SAS 15k-RPM HDD. The SSD consumes approximately 87% less power than the 15k-RPM HDD in active mode, and around 82% less in idle mode. In practice, however, the energy savings will depend on how the storage solutions use the SSDs and HDDs and the characteristics of the workload applied to the storage equipments.

Tab. 3.3 **Comparison of Seagate's Pulsar enterprise SSDs and Savvio 15K HDDs**

Specifications	Savvio 15K.2 73GB*	Pulsar SSD 50GB*	Difference
Form Factor	3.5"	2.5"	–
Capacity	73GB	50GB	–
Interface	SAS 3Gb/s SAS 6Gb/s	SATA 3Gb/s	–
Spindle Speed (RPM)	15K	–	–
NAND Flash Type	–	SLC	53% less
Power Idle (W)	3.7	0.65	82.4% less
Power Active (W)	6.18	0.8	87% less

* Data obtained from the specification sheets available at the manufacturer's website.

3.1.4 Hybrid Hard Drives (HHDs)

HHDs are HDDs equipped with large buffers made of non-volatile flash memories that aim to minimise data writes or reads on the platters. Several algorithms have been used for using the buffer [7]. By providing a large buffer, the platters can remain at rest for longer periods. This additional flash memory can minimise the power consumed by storage solutions by reducing the power consumed by the motors and mechanical arms. These drives can present potentially lower power requirements when compared to HDDs, but the offerings for enterprise storage are limited.

RECOMMENDATIONS FOR BEST PRACTICE

Consider advantages of different storage technologies in procurement and system design

- Tapes have the best energy efficiency for long-term storage.
- Current hard disk drives have platters that can rotate at various speeds thus saving energy at lower speeds.
- The multiple idle states implemented by HDDs allow considerable energy savings when employed in composite storage solutions such as disk arrays and massive arrays of idle disks.
- Although more expensive, SSDs are much more efficient than HDDs.
- Consider using SSDs as a high performance storage layer.

3.2 Storage Elements

This section presents device-level techniques that may be used and combined to improve the energy efficiency of composite storage solutions such as disk arrays, direct attached storage and network storage (i.e. storage elements). Concepts specific to the storage-element level are also analysed.

3.2.1 Large Capacity Drives and small form factor

For applications that do not demand high-performance storage it is usually more energy efficient to use drives with larger capacity. Typical SATA disk drives consume up to 50% less power per terabyte of storage than Fibre Channel drives [8]. As discussed beforehand, SSF enclosures can save floor space in data centres and decrease the energy footprint by using more power-efficient 2.5-inch HDDs. As an example, using the industry-standard Storage Performance Council (SPC) SPC-1C benchmark, Dell compared two of its disk arrays, one with 3.5-inch HDDs and another with 2.5-inch HDDs [9]. Results showed that in addition to providing 93% higher performance than the array with 3.5-inch drives, the array equipped with 2.5-inch drives consumed 40% less energy.

3.2.2 Massive Arrays of Idles Disks (MAIDs)

MAID is a technology that uses a combination of cache memory and idle disks to service requests, only spinning up disks as required. Stopping spindle rotation on less frequently accessed disk drives can reduce power consumption (see Figure 3.3).

All disks spinning full-speed; high performance but no power saving

25% disks spun down; up to 25% power saving but some performance penalty

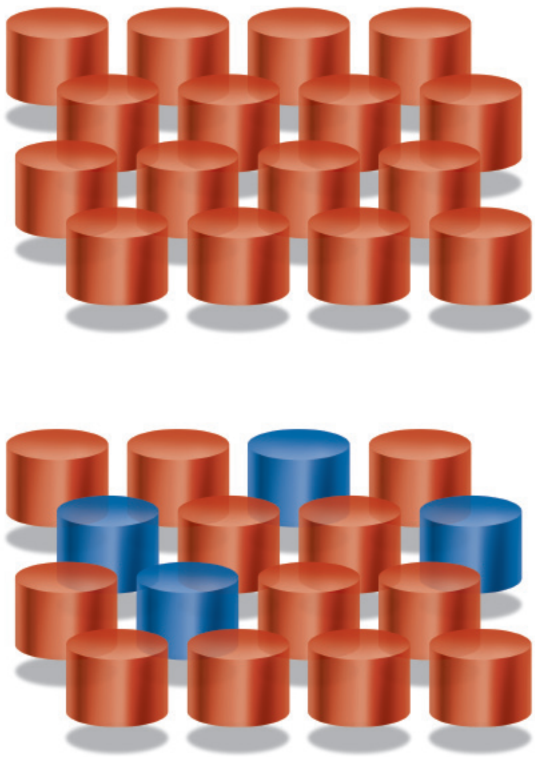


Fig. 3.3 **Pictorial view of MAIDs**

Fujitsu, for example, allows customers to specify schedules with periods during which the drives should be spun down (or powered off) according to the workload or backup policies.

How much power MAID features can save depends on the application that uses the disks and how often the disks are accessed. The criteria used to decide when drives are spun down (or put into standby mode) or spun up, have an impact on energy savings as well as on performance. When initially conceived, MAID techniques enabled HDDs to be either on or off, which could incur considerable application performance penalties if data on a spun-down drive was required. However second generation MAID techniques shall allow Intelligent Power Management (IPM) with different power saving modes and performance. MAID 2.0 as it is often called has multiple power saving modes that align power consumption to different QoS needs. The user can configure the trade-off between response times and power savings. The multiple power saving modes use for example the different HDD idle states described beforehand.

Other power conservation techniques for disk arrays are the Popular Data Concentration (PDC) [10] and other file allocation mechanisms [11]. The rationale behind this approach is to perform consolidation by storing or migrating frequently accessed data to a subset of the disks. By skewing the load towards fewer disks, others can be transitioned to low-power consumption modes.

3.2.3 Efficient RAID Levels

Different RAID levels provide different storage efficiency. When considering data protection, some RAID levels such as RAID 6 present a significant amount of overhead processing. However high performance RAID 6 implementations can provide the same performance as RAID 5 and up to 48% reduction in disk capacity requirements compared to RAID 10.

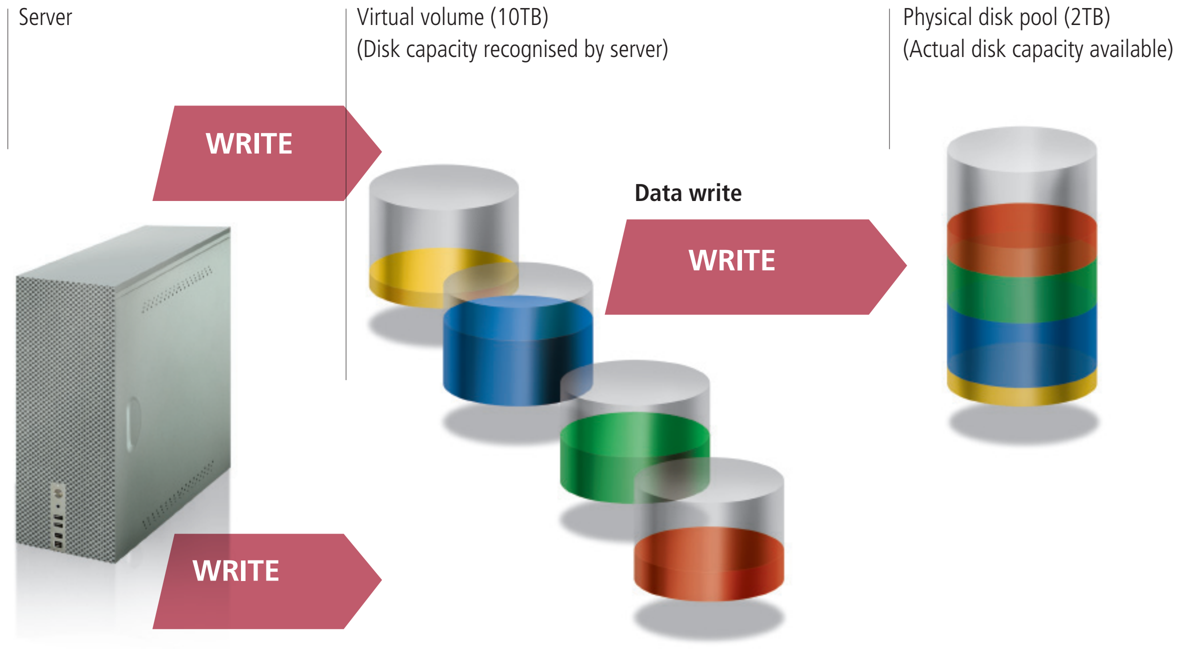


Fig. 3.4 **Thin provisioning (from Fujitsu ETERNUS solutions)**

3.2.4 Horizontal storage tiering, storage virtualization and thin provisioning

For efficient use of storage infrastructure, it is important to design and enforce sound data management policies that use different tiers of storage according to how often the data is accessed, whether it is reused and for how long it has to be maintained (for business or regulatory purposes). Manufacturers of data storage solutions have proposed software systems that allow for seamless and automatic tiering by moving data to the appropriate tier based on ongoing performance monitoring. Examples are EMC2's Fully Automated Storage Tiering (FAST), IBM's System Storage Easy Tier, Compellent's Data Progression and SGI's Data Migration Facility (DMF).

By combining server virtualization with storage virtualization it is possible to create disk pools and virtual volumes whose capacity can be increased on demand according to the applications' needs. Typical storage efficiency of traditional storage arrays is between 30–40%. According to certain reports [12], storage virtualization may increase efficiency to 70% or higher reducing storage requirements and increasing energy savings.

Storage tier virtualization, also known as Hierarchical Storage Management (HSM), allows data to be migrated automatically between different types of storage without users being aware of it. Software systems for automated tiering are used for carrying out such data migration activities. This approach can reduce cost and power consumption as it allows only data that is frequently accessed to be stored on high-performance storage, while data less frequently accessed can be placed on less expensive and more power efficient equipment that use techniques such as MAID and data de-duplication.

Thin provisioning, a technology that generally complements storage virtualization, aims to maximise storage utilisation and eliminate pre-allocated but unused capacity. With thin provisioning, storage space is provisioned when data is written. Reserve capacity is not defined by the maximum storage required by applications but it is generally set to zero. Volumes are expanded online and capacity is added on the fly to accommodate changes without disruption (see Figure 3.4). Thin provisioning can lead to energy savings because it reduces the need for over provisioning storage capacity to applications.

3.2.5 Consolidation at the storage and fabric layers

Storage consolidation is not a recent topic as Storage Area Networks (SANs) have been providing some level of storage consolidation and improved efficiency for several years by sharing arrays of disks across multiple servers over a local private network, avoiding islands of data. Moving direct attached storage to network storage systems offers a range of benefits, which can increase the energy efficiency. Consolidation of data storage equipments can lead to both substantial savings in floor space requirements and energy consumption. Some manufacturers argue that by providing multi-protocol network equipment, the network fabric can be consolidated on fewer resources, hence also reducing floor space, power consumption and cooling requirements.

3.2.6 Data de-duplication

Storage infrastructures often store multiple copies of the same data. Several levels of data duplication are employed in storage centres, some required to improve reliability and data throughput. However, there is also „waste" that can be minimised, hence recycling storage capacity. Current SAN solutions employ data de-duplication (de-dupe) techniques with the aim to reduce data duplicates. These techniques work mainly at the data-block and file levels.

In addition to the level of data de-duplication, de-duplication techniques also differ depending on when the data de-duplication is performed: before or after data is stored on disk. Both techniques have advantages and shortcomings. Although leading to reduced storage-media requirements, de-duplication after the data is stored on disk requires cache storage that is used for removing duplicates. However, for backup applications, de-duplication after storing the data usually leads to shorter backup windows and smaller performance degradation. Moreover, data de-duplication techniques differ on where data de-dupe is carried out: at the source (client) side, target (server) side, or by a de-duplication appliance connected to the server.

As data de-duplication solutions enable organisations to recycle storage capacity and reduce media requirements, they are also considered a common approach to reduce power consumption. The actual storage savings achieved by data de-duplication solutions vary according to their granularity. Solutions that perform hashing and de-duplication at the file-level tend to be less efficient. However, they pose a smaller overhead. With the block-level techniques, the efficiency is generally inversely proportional to the block size.

Although data de-duplication is a promising technology for reducing waste and minimising energy consumption, not all applications can benefit from it. For example, performing data de-duplication before the data is stored on disk, could lead to serious performance degradation, which would be unacceptable for database applications. Applications and services that retain large volumes of data for long periods benefit more from data de-duplication. The more data one organisation has and the longer it needs to keep it, the better results data de-duplication technologies will yield. In general, data de-duplication works best for data backup, data replication and data retention.

Further Reading

McClure T. (2009): Driving Storage Efficiency in SAN Environments, Enterprise Strategy Group - White Paper, November 2009.

Craig B. and McCaffrey T. (2009): Optimizing Nearline Storage in a 2.5-inch Environment Using Seagate Constellation Drives, Dell Power Solutions, Jun. 2009.

SNIA (2010): Storage Power Efficiency Measurement Specification: Working Draft Version 0.2.10, SNIA Green Storage Initiative, August 2010.
Storage Tiering with EMC Celerra FAST, EMC2
www.snia.org/sites/default/files/Storage_Power_Efficiency_Measurement_Spec_v0.2.10_DRAFT.pdf

Clark T. and Yoder A. (2008): Best Practices for Energy Efficient Storage Operations Version 1.0, SNIA Green Storage Initiative, October 2008.

Freeman L. (2009): Reducing Data Centre Power Consumption Through Efficient Storage. White Paper. NetApp, July 2009.

References

[1] **IDC (2010):** The Digital Universe Decade – Are you ready? IDC, May, 2010.

[2] **McClure T. (2009):** Driving Storage Efficiency in SAN Environments, Enterprise Strategy Group - White Paper, November 2009.

[3] **Reine D. and Kahn M. (2008):** Disk and Tape Square Off Again – Tape Remains King of the Hill with LTO-4. Clipper Notes, February 2008.

[4] **ORACLE (2010):** Consolidate Storage Infrastructure and Create a Greener Datacentre. Oracle White Paper, April 2010.

[5] **Seagate (2011):** PowerChoice Technology Provides Unprecedented Hard Drive Power Savings and Flexibility - Technology Paper, Seagate, 2011.

[6] **Seagate (2010):** Seagate Savvio 15K.2 Data Sheet, Seagate, 2010.

[7] **Bisson T., Brandt S., Long D. (2006):** NVCache: Increasing the Effectiveness of Disk Spin-Down Algorithms with Caching, 14th IEEE International Symposium on Modeling, Analysis, and Simulation, pp. 422-432, 2006.

[8] **Freeman L. (2009):** Reducing Data Centre Power Consumption Through Efficient Storage. White Paper. NetApp, July 2009.

[9] **Craig B. and McCaffrey T. (2009):** Optimizing Nearline Storage in a 2.5-inch Environment Using Seagate Constellation Drives, Dell Power Solutions, Jun. 2009.

[10] **Pinheiro E. and Bianchini R. (2004):** Energy Conservation Techniques for Disk Array-Based Servers. 18th Annual International Conference on Supercomputing (ICS 2004), pp. 68-78. Malo, France, 2004.

[11] **Otoo E. D., Rotem D. and Tsao S.C. (2009):** Analysis of Trade-Off between Power Saving and Response Time in Disk Storage Systems, IEEE International Symposium on Parallel Distributed Processing (IPDPS 2009), pp. 1-8, May 2009.

[12] **Blade Network (2009):** Storage Consolidation for Data Centre Efficiency, BLADE Network Technologies White Paper, Jun. 2009.

4 Network Equipment

Alexander Schlösser, TU Berlin, **Lutz Stobbe,** Fraunhofer IZM

According to current information, the energy consumption allocated to switches, routers, and other networking equipment is approximately 8% to 12% of the total data centre's energy footprint. Due to this rather low percentage of the total energy demand, networking equipment has not been the focus of improvement measures. Nonetheless, this perception and situation is changing, particularly in medium and large sized data centres. There are a couple of reasons why the power consumption of networking equipment and the energy effects of the implemented network architecture are now becoming significant considerations in the design and operation of data centres.

With increasing quality-of-service (QoS) requirements in conjunction with delay-critical applications, the functional importance of networking equipment and networks in data centres is growing. The power consumption varies according to the selected technology and architecture including cabling, power supply, and cooling.

4.1 Technical and operational framework

4.1.1 Functional model

Figure 4.1 provides a simplified functional model of energy-related aspects with respect to networks and network equipment in data centres. The functional model helps to visualize the overlapping aspects of the power and cooling infrastructure as well as the interrelation of the network with the main IT equipment including server and storage systems. The model also outlines the main elements for improvement at the network level. This includes the selected network architecture and actual topology, the physical infrastructure, hardware components and cable as well as software configuration and virtualization capability.

The energy efficiency of network infrastructure and networking equipment is also influenced by the applications, service level agreements, bandwidth and latency performance requirements that have been defined by the operator of the data centre. These performance-related aspects have to be considered in the planning process to improve energy efficiency.

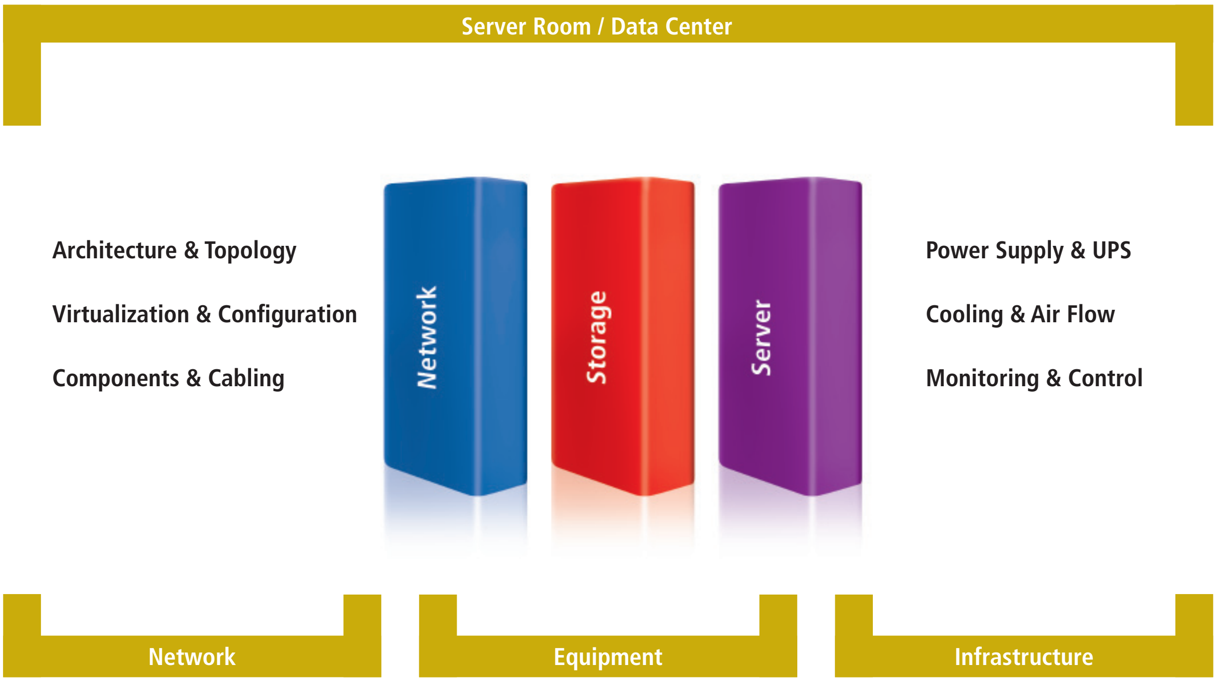


Fig. 4.1: **Data centre networks functional model**

4.1.2 Network attributes

The improvement of energy efficiency with respect to the network infrastructure in data centres requires a structured approach. Planning should incorporate a strategic or long-term perspective due to the fact that networking infrastructure is typically somewhat longer in place. It is assumed that the basic network infrastructure is used for more than 8 years. Changing the basic network architecture and actual topology including equipment etc. is a considerable investment and risk factor. Nevertheless, the improvement of the network not only increases the performance characteristics of the data centre but in many cases the energy efficiency as well. The planning for improvement starts with a strategic analysis.

The data centre operator needs to define network attributes and performance requirements. This task should include a market analysis. The IT world is currently (2011) experiencing a tremendous shift towards a centralized production of applications resulting in new traffic volumes and patterns. In other words, applications are not produced at the end-user side with considerable computing power and software packages. By utilizing Software-as-a-Service (SaaS) and Cloud Computing, applications and traffic is produced in data centres and data centre clouds. A necessary condition for this trend is broadband connectivity and low latency. This general trend leads not only to increased data traffic between client and server, but also to an increased server to server and storage to server data flow. Enterasys [1] indicates in this respect that the network architecture and configuration will change in order to support the growing server to server and storage to server traffic. In order to increase performance (IT productivity), the technical trends are to aggregate networking (bottom-up) and virtualized networking (top-down).

The network architecture will consist of fewer tiers by merging access and aggregation as well as aggregation and core network to some extent (see also Figure 4.3). This trend has the potential to reduce energy consumption due to unified networking. However, this is a balancing act. Little information and data is available, and there is not a single solution visible on the market.
The virtualization will also spread further incorporating network equipment and local area networks (VLAN). Virtualization has the advantage of consolidating physical equipment and has therefore also the potential of increased energy efficiency.
According to Enterasys [1], the common design goals for data centres networks include:

- Bandwidth and low latency (selection of network technology)
- Scalability and agility (network architecture)
- Flexibility to support various services (this aspect is addressing consolidation)
- Security (increasingly important and influencing overhead)
- High availability and redundancy (quality of service requirements)
- Manageability and transparency (this aspect is supported by virtualization solutions)
- Costs optimization (the objective is always lower CAPEX and OPEX)

4.1.3 Balancing network performance and energy consumption

Bandwidth, high-speed, low latency, and lossless traffic are important network performance criteria. Customer satisfaction or what is called Quality of Service (QoS) is an additional performance requirement. QoS is defined in terms of Service Level Agreements (SLA) with characteristics such as minimal throughput, maximal response time or latency time. A responsive, converged, and intelligent network architecture capable of managing traffic dynamically to agreed SLAs, is not only important for future competitiveness, it might also define the basis for a systematic energy efficiency approach.

However, implementing QoS can actually increase the total network traffic and respective energy consumption of the data centre. Individual network technologies and respective equipment fea-

RECOMMENDATION FOR NETWORK DESIGN

Due to the fact that there is a large variety of products and network options available on the market, it is recommended that data centre operators or IT administrators develop a priority list with respect to network attributes, such as:

- network services,
- latency requirements,
- quality of service,
- virtualization support and
- other performance or interoperability aspects.

The best approach is a system optimization. It reflects the interaction of the network infrastructure and performance with the other IT-equipment and support infrastructure.

ture advantages and disadvantages in this respect. As a general trend, it has been observed that 10 Gigabite Ethernet networking (10GbE) is becoming the technology of choice in data centres. Ethernet is not only linking servers (LAN), it is increasingly applied in storage networks (SAN).

Low latency and lossless networks are however basic requirements for storage traffic. According to Lippis (2011) [2] today's 10GbE, switches produce 400 to 700ns of latency. By 2014, it is anticipated that 100GbE switching will reduce latency to nearly 100ns. This shows that with increasing bandwidth the latency improves. From an energy consumption point of view, it is necessary to balance latency improvement (network technology) with potentially higher power consumption of the high bandwidth capacity (component). Component selection and I/O consolidation are aspects that have to be addressed in that respect.

Similarly it is necessary to investigate lossless networking (availability) versus bandwidth performance and subsequent energy efficiency. For instance, lossless networking usually means more complex protocols (overhead) and additional latency in conjunction with more processing power and less bandwidth efficiency. Lossless networking is however a necessary precondition for storage area networks. In the past, (lossy) Ethernet was the hindrance for its application in the storage area network.

Fiber Channel (FC) and Infiniband (IB) were the most common network technologies. However, today multiple storage network options exist based on available Ethernet such as Converged Enhanced Ethernet (CEE), Fibre Channel over Ethernet (FCoE), Internet Small Computer System Interface (iSCSI) over Ethernet, ATA over Ethernet (AoE) and Network-attached Storage (NAS). These options help to unify the networking (and avoids additional adapters), however create additional overhead which results in less bandwidth efficiency. The energy efficiency tradeoffs (if they exist) are currently unknown.

In conclusion, the operator must consider the energy impacts of increased performance, scalability, and adaptability of new consolidated network solutions. It is likely that positive energy tradeoffs result from new solutions. But proper dimensioning is essential. It is recommended that operators who purchase new equipment or complete network solutions ask for the overall energy impact / tradeoff resulting from a new solution.

4.2 Improvement of energy efficiency

4.2.1 Merging traffic classes (I/O consolidation)

Data centre networks have to transmit different types of traffic within different types of application areas. This has led to specialized protocols and network architectures. As a result, considerably complex networks often do not share their resources. The basic improvement objective is the physical reduction of components and the sharing of network capacity by different functional units. The general technical trend toward simpler, fewer tiers, and I/O converged networking based on Ethernet is also driven by energy efficiency considerations. The overall topic is network consolidation. It addresses server and storage networks as well as the network distribution architecture. Convergent Network Adapter (CAN) is merging former separate interfaces:

- Host Bus Adapter (HBA) in support of SAN traffic
- Network Interface Controller (NIC) in support of LAN traffic
- Host Channel Adapter (HCA) in support of IPC traffic

I/0 consolidation is the capability of a switch or a host adapter to use the same physical infrastructure to carry multiple types of traffic, each typically having unique characteristics and specific handling requirements. From the network side, this equates to having to install and operate a single network instead of three as shown in Figure 4.2.
From the hosts and storage arrays side, this equates to having to purchase fewer Converged Network Adapters (CNA) instead of Ethernet NICs, FC HBAs, and IB HCAs. A typical Fibre Channel HBA consumes about 12.5 W [3]. In terms of network redundancy, an adequate consideration of several options has to be regarded to design reliable networks.

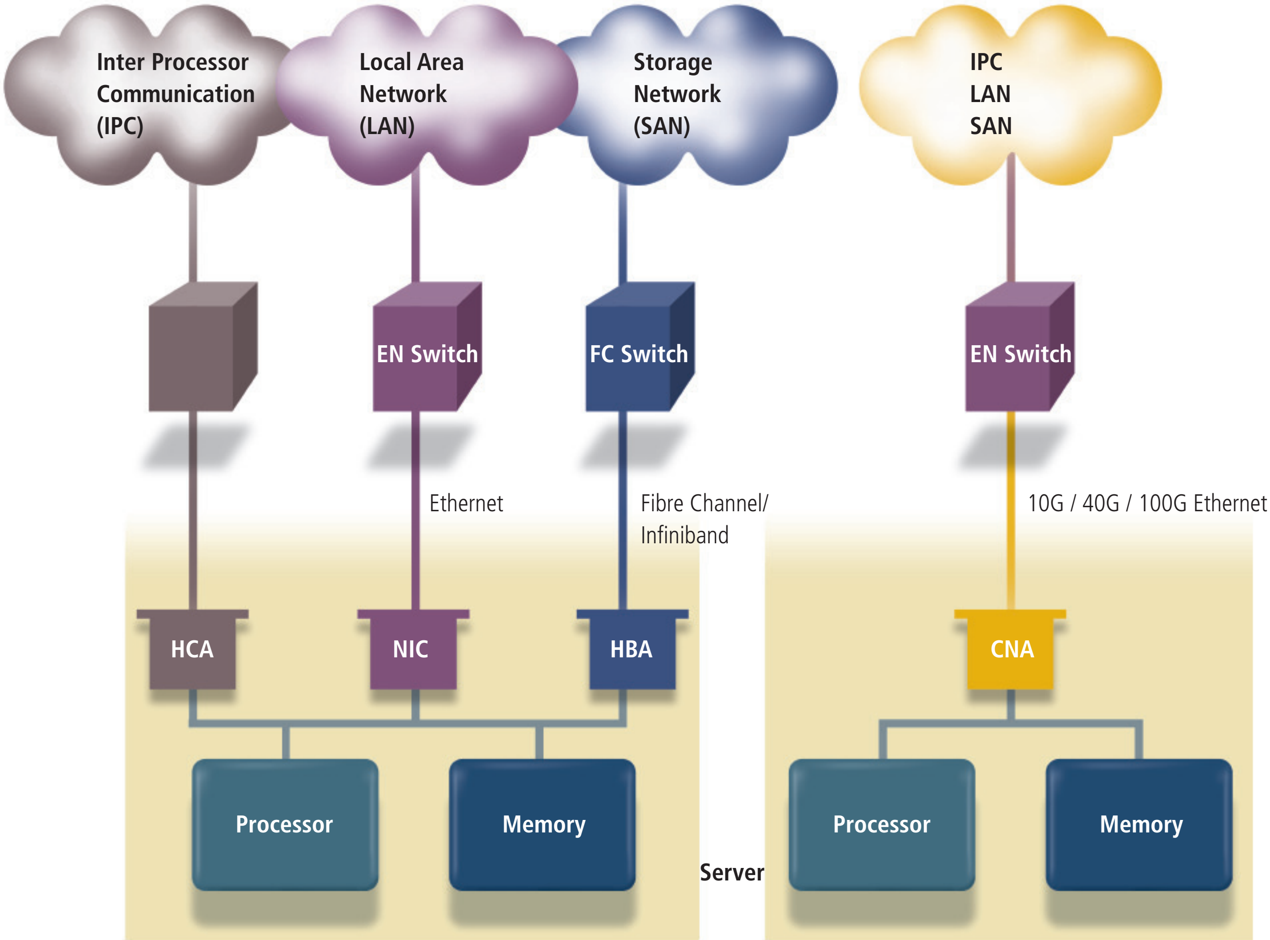


Fig. 4.2 **I/O consolidation and network convergence in data centre networks**

BENEFITS OF CONVERGED NETWORKS

I/O consolidation will enable the consolidation of different network types (LAN, SAN) at a higher level as a preparatory measure for system virtualization. Furthermore, it will significantly reduce the amount of physical infrastructure including switches, ports, connections and cables among different networks. Converged networks will result in:

- Up to 80% decreased adapters and cables
- Up to 25% reduction in switches, adapters and rack space
- Up to 42% reduction of power and cooling costs [4]

4.2.2 Network consolidation

The main approach to optimize the power consumption of the data centre network includes new network architectures and the convergence of formerly separated networks (into a single technology). A typical architecture consists of a tree of routing and switching equipments (multiple tiers/layer) with more specialized and expensive equipment on the top of the network hierarchy. The goal should be the consolidation of the network infrastructure by creating a flat network architecture based on a functional network fabric.

Measures that can be taken include:

- Aggregate switches. Multiple physical switches that operate in one logical device.
- Reduce tiers (layers). Use an aggregated switch to do the work of multiple switch layers. Considered network services and security.
- Create unified network fabric. This combines the two approaches and allows operational simplicity and high performance. Again, considered network services and security.

The convergence of server (LAN) and storage (SAN) networks is a general trend with energy saving potential. Maintaining two separate networks would increase the overall operation costs and power consumption by multiplying the number of adapters, cables, and switch ports required to connect every server directly with supporting LANs and SANs. To simplify or flatten the data centre network structure, converged networking technologies such as iSCSI, Fibre Channel over Ethernet (FCoE), and Data Centre Bridged (DCB) are currently being implemented in data centres.

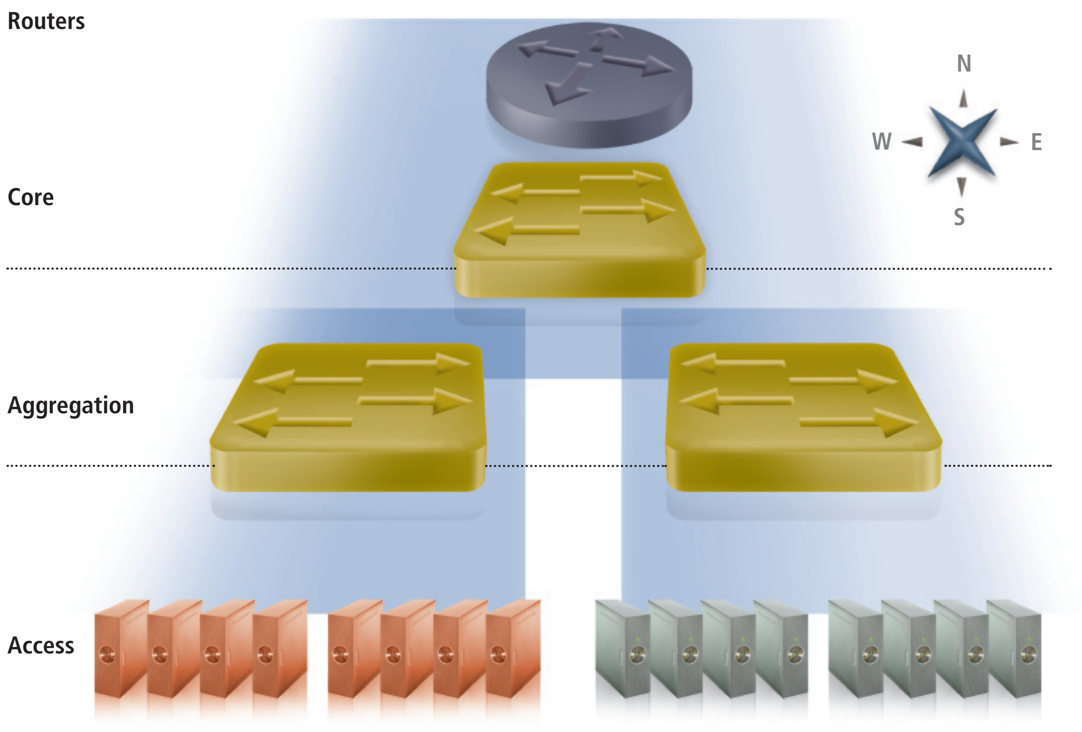


Fig. 4.3 **Network consolidation**

4.2.3 Network virtualization

Virtualization is a well established technology to consolidate physical server with multiple virtual machines. Network virtualization follows the same principle and describes various hardware and software approaches to manage network resources as logical units independent of their physical topology. This results in reduced network traffic, simplified security and improved network control. Key elements for high efficient networks are network level awareness and visibility of the virtual machine (VM) lifecycle. The ability to configure network and port level capabilities at the individual VM level as well as dynamically tracking VMs as they move across the data centre are important for an efficient management of virtualized environments. Energy efficiency is mainly achieved by consolidation of routers, physical adapters for I/O ports, and additional hardware for specific network services.

Extending system virtualization to the network includes:

- Virtual router (software with routing functionality, multiple systems on 1 real machine)
- Virtual links (logical interconnection of virtual router)
- Virtual networks (Virtual routers connected by virtual links)

The increase in server virtualization will result in additional complexity and overhead for the network. Obsolete networking switches are not aware of Virtual Machines and this exposes the risk of service outage and security breaches due to incorrect network configuration. Networking is a key area that also needs to be virtualized to achieve the same level of agility, bandwidth and performance.

Network service virtualization is a strategy to simplify the network operations and consolidate multiple appliances. The virtualization of a firewall module or IPS by providing a software image to different applications via single network hardware would reduce the need of separate devices by utilizing the software in the same hardware.

Reduced power consumption is achieved by consolidating multiple services into a single physical device without requiring deployment of dedicated hardware for each instance. Eliminating the need for additional physical devices effectively removes the need for additional power supplies, cooling, and rack space which would otherwise have been required.

Summarized benefits for network service virtualization:

- Management interfaces are more flexible
- Reduced acquisition cost by use of software
- Increased application performance by simplified service extension and allocation
- Potential decreased power consumption by equipment consolidation

A successful implementation of network virtualization depends on aspects like capital expenditure, the definition of precise objectives or the compatibility with existing hardware. Therefore, virtualization projects require a well balanced cost-benefit analysis, a comprehensive project management and a consequent consideration of possible security risks.

RECOMMENDATIONS FOR SMALL TO MEDIUM DATA CENTRES

For small to medium businesses, the choice between FCoE and iSCSI largely depends upon application requirements and availability of personnel trained in Fibre Channel.

- When high capacity and performance oriented databases are the business critical applications, FCoE and iSCSI are suitable solutions to improve the service level and reduce the power consumption.
- Centralized storage provisioning and disaster recovery require a common SAN → iSCSI is preferred
- Fibre Channel dominant networks adaptation of FCoE is recommended [5].

4.2.4 Components and equipment selection

The power consumption of network equipment is generally influenced by the component selection and actual configuration of the system. The main influence has the supported network technology standard (e.g. 10GbE). The chip-design and system integration level has the highest leverage. The trend is driven by the performance improvement in semiconductor technology which still follows Moore's Law. This also includes the thermal performance of the chip and the interconnection technology. Reliability is a growing issue in that respect. Other factors are the system configuration meaning the types and number of ports deployed in the equipment. Finally, power consumption of network equipment is influenced by the efficiency of the power supply unit and power management options.

RECOMMENDATIONS FOR BEST PRACTICE

Consider procurement criteria for selecting energy efficient network hardware and particularly power supply units.

- Choose equipment with power management functionalities and compare power consumption of different devices in idle and standby states.
- Compare costs and energy efficiency of network systems from different vendors
- Request product information from suppliers regarding:
  - Ioverall energy efficiency (e.g. ECR, TEER as soon as available)
  - Iefficiency and modularity of power supply units
  - Iefficiency and scalability of blower units (variable fan speeds, etc.)

Power Management

The magnitude of the network equipment's energy consumption is related to active use and periods of idling. The difference in power consumption between active (100% load) and idle (with established link) is typically about factor 1.1 (less than 10% difference). If the link is deactivated, the power consumption drops by factor 2 (50% of active).

However, it is expected that in smaller installations (e.g. server room, small DC) idle phases could especially occur during night time. Advanced power management including a type of "networked standby" is not yet common. The term networked standby has been coined by the European EuP/ErP framework directive preparatory study for ENER Lot 26. This study argued that "resume-time-to-application" is the key criteria for the implementation of networked standby. The power management of network equipment is closely related to the server and storage systems which they connect.

In case a cluster of server or storage devices would be powered down into a standby (sleep) mode, it would be possible to also power down parts of the access switches. Again, the critical factor is latency and reliability of the system wake-up. With the introduction of the standards IEEE 802.3az "Energy Efficient Ethernet" and Standard ECMA-393 "proxZzzyTM for sleeping hosts" specific approaches for low power management are underway.

Power Supply Unit

The reliability and conversion efficiency of Power Supply Units (PSU) influences the overall energy consumption. The conversion efficiency of larger PSU (>500W output) has been improved in past years to typical levels of over 85% and in excess of 90%. Due to the fact that larger core switches and routers consume up to a few KW, even the smallest improvements in conversion efficiency (even if only 1%) will result in noticeable energy savings. Further, product specifications do not necessarily disclose information on the PSUs conversion efficiency.

4.2.5 Floor-level switching

There are two basic types of switch distribution on the floor or application level: End-of-Row and Top-of-Rack. End-of-Row (EoR) switching is a conventional networking approach, featuring a single large chassis-based switch support of one or more racks. From an energy efficiency point of view, there are two considerations in respect to EoR:

- Advantage: Centralized switching with good scalability and energy savings compared to suboptimal ToR solution
- Disadvantage: Considerable cabling effort with inefficiency in dense systems

Top-of-Rack (ToR) switching defines a system with a switch integrated in each rack. This concept ensures short latency and high data transmission. The advantage and disadvantage of ToR regarding energy efficiency are:

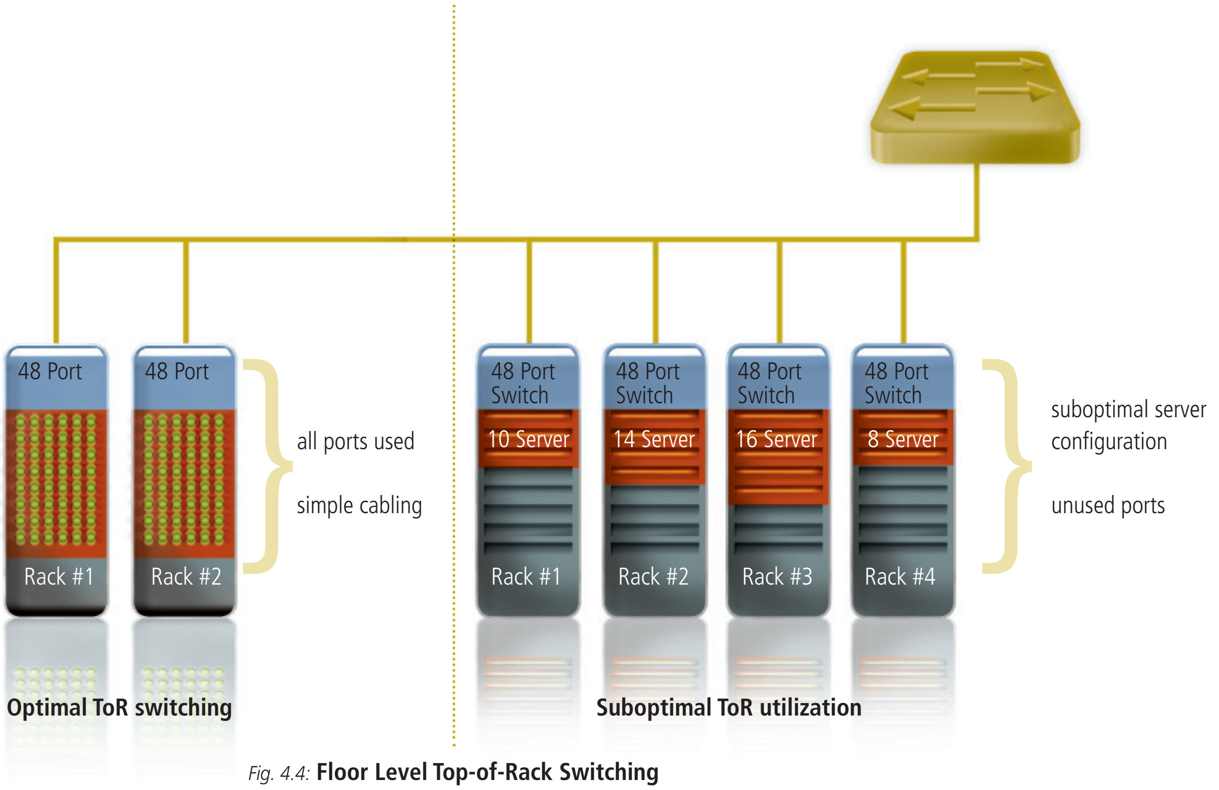


Fig. 4.4: **Floor Level Top-of-Rack Switching**

- Advantage: Decentralized switching for dense server environments (I/O consolidation) which reduces cabling effort. The shorter cabling distance between server and switch improves transmission speed and reduces energy consumption for this transmission.
- Disadvantage: If ToR is utilised in less dense computing (few servers in a rack), the system is over-dimensioned. Energy efficiency is low due to suboptimal utilisation of available ports.

In conclusion, ToR has eco-advantages when applied in properly dimensioned systems.
Figure 4.4 illustrates the ToR switching concept and its proper utilisation

Further Reading

Hintemann R. (2008): Energy Efficiency in the Data centre, A Guide to the Planning, Modernization and Operation of Data centres, BITKOM, Berlin, online available:
http://www.bitkom.org/de/publikationen/38337_53432.aspx

EC JRC ISPRA (2011): Best Practices for the EU Code of Conduct on Data centres
European Commission (2011), EC Joint Research Centre, Ispra, online available:
http://re.jrc.ec.europa.eu/energyefficiency/html/standby_initiative_data_centres.htm

Juniper (2010): Government Data centre Network Reference Architecture, Using a High-Performance Network Backbone to Meet the Requirements of the Modern Government Data centre
Juniper (2010), Juniper Networks, Inc., Sunnyvale, available online:
http://www.buynetscreen.com/us/en/local/pdf/reference-architectures/8030004-en.pdf

References

[1] **Enterasys (2011):** Data centre Networking – Connectivity and Topology Design Guide; Inc Enterasys Networks, Andover.

[2] **Lippis (2011):** Open Industry Network Performance & Power Test Industry Network Performance & Power Test for Private and Public Data centre Clouds Ethernet Fabrics Evaluating 10 GbE Switches; Lippis Enterprises, Inc, Santa Clara.

[3] **Cisco (2008):** Converging SAN and LAN Infrastructure with Fibre Channel over Ethernet for Efficient, Cost-Effective Data centres; Intel, Santa Clara.

[4] **Emulex (2008):** Sheraton Case Study. Virtual Fabric for IBM BladeCentre Increases Server Bandwidth, Reduces Footprint and Enables Virtualization for High-performance Casino Applications; Emulex, Costa Mesa 2010.

[5] **Blade.org (2008):** Blade Platforms and Network Convergence; Blade.org,White Paper 2008.

5 Cooling and power supply in data centres and server rooms

Andrea Roscetti, Politecnico di Milano, **Thibault Faninger,** Bio Intelligence Service

Cooling can be responsible for up to 50% of the total energy consumption in server rooms and data centres. Concepts for energy efficient cooling are therefore essential in both small and larger IT facilities.

The following section shows a number of general options to reduce energy consumption.

Fig. 5.1: **Split cooling system room unit**

external unit (Source: Daikin)

5.1 Cooling in server rooms

Server closets or small server rooms are usually equipped with comfort cooling systems (typically office HVAC[1] systems). Small data centres are typically equipped with 1-5 racks of servers, with a total IT power of maximum 20 kW.

5.1.1 Split systems and portable systems

Split cooling systems are commonly used in small server rooms. The cooling power range of this family of systems is 1–100 kW. Generally split/DX[2] cooling systems have several advantages:

- Investment costs are typically low.
- Design and installation are quite simple.
- Floor space required for the installation is small (units are typically wall mounted).
- Installation is possible in almost all situations.
- Maintenance and replacement of the systems is quite simple and fast.

On the other hand the following drawbacks have to be considered:

- Overall efficiency is quite low for small, older or oversized systems.
- Comfort cooling has poor humidity control.
- Piping between external and internal units has limitations in length and height.

Portable systems may be installed for example to prevent hot spots. The technology provides the following advantages:

- Investment costs are very low.
- Installation is simple.
- Floor space required for the installation is small.
- Maintenance and substitution of the system is simple and fast.

1) HVAC: Heating, ventilation and air conditioning
2) DX: direct expansion

The following disadvantages are to be considered:

- Overall efficiency is quite low: A-class mobile systems are less efficient than D-class split systems.
- Cooling has poor humidity and temperature control.
- Installation is only possible if air can be vented to the outside.

5.1.2 Measures to optimize energy efficiency in server rooms

Oversizing of cooling is common practice for small server rooms. To avoid over-sizing of cooling for well-insulated server rooms, a rule of thumb suggests that the cooling power should not exceed 120% of the IT installed power.

When buying new appliances up to 12 kW cooling power, the EU Energy Label can be considered to support the selection of energy efficient equipment. A high EER[3] /SEER[4] and A-class level efficiency or above are the right choices. SEER and the kW/h annum estimated from the label are the most important criteria for comparison. Table 5.1 shows the efficiency for the current best available technology.

The label is implemented with a transition period until 1 January 2013. Before that time the label can be used by manufacturers already but is not mandatory. During the transition also the old label for air conditioners may still be used (2002/31/EC).

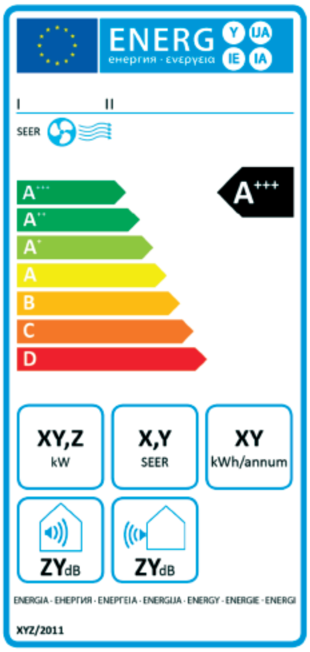


Fig. 5.2: **Energy label for cooling only air conditioners** (Source: regulation supplementing Directive 2010/30/EU of the European Parliament and of the Council with regard to energy labelling of air conditioners)

Tab. 5.1: **Best available technology efficiency values for small cooling systems <12 kW** (source: Ecodesign regulation requirements for air conditioners and comfort fans)

Benchmarks for air conditioners		
Air conditioners, excluding double and single duct	Double duct air conditioner	Single duct air conditioner
SEER	SEER	SEER
8.50	3.00	3.15

3) Energy Efficiency Ratio: ratio of output cooling to input electrical power at a given operating point (indoor and outdoor temperature and humidity conditions)

4) Seasonal EER: represents the expected overall performance in a given location (test method)

RECOMMENDATIONS FOR BEST PRACTICE

Existing server rooms

- Eliminate solar gain, heat transmission and ventilation losses to other rooms/outside space.
- Control and manage the environmental conditions (set points): inlet air to IT (not setpoint) must be 18–27°C however suggested range is 24 to 27°C.
- Verify the ducts/pipes insulation (cold and hot air/water/liquid).
- Evaluate the substitution of obsolete or less efficient components of the cooling system (compare the efficiency class of existing systems with the most efficient ones available on the market).
- Control and verify the layout of the installed cooling system (e.g. distance between cooling systems and loads).
- Turn off lights and remove other mechanical/electrical loads and sources of heat if possible.

New server rooms

- Evaluate the use of precision cooling systems (in order to remove sensible heat from IT and avoid over-dehumidification).
- Define and assess the room and the IT characteristics, taking into account space restrictions and distance between load and external units.
- Avoid the use of mobile units or ducted units with low EER (note: A-class mobile systems are less efficient than D-class split systems!).
- Compare different systems:
  - Opt for the higher energy label class (mandatory for small systems).
  - Maximise the cooling efficiency (SEER), see BAT table.
- Consider the use of free cooling.

5.2 Cooling for medium to large data centres

5.2.1 General aspects

The traditional approach for cooling in medium and large data centres has been based on air cooling. A standard data centre is designed to cool on average 7.5–10 kW/m^2, which translates to 1–3 kW/rack. Newer data centres are designed to cool an average 20 kW/m^{2,} which still limits the power density per rack to 4–5 kW (recall that full rack capacity in consolidated systems or blade server systems may be higher than 25 kW/rack).

IT equipment is arranged in rows with air intakes facing the cold aisle. Cool air is supplied to the cold aisle, passes through the equipment and then is discharged to the hot aisle.

Important elements to consider are the airflow characteristics. The recommended air flow directions are front to rear, front to top, or front+top to rear (see reference). If different equipment with different operating conditions or airflow directions is installed in the same room, a separate area should be created. In case the equipment has different environmental requirements, it is preferable to provide separate environmental controls in order to avoid inefficiencies due to the lower set point or poor air flow control. For further details see reference [1].

5.2.2 Temperature and humidity settings

Data centres should be designed and operated at their highest efficiency possible under given climate conditions (dry bulb[5]). Recommended temperature is between 18 and 27 °C and relative humidity lower than 60% (inlet air to IT equipment). Respectively the dew point should be between 5.5 and 15 °C. Studies on inlet air temperatures suggest 24–27 °C as the optimal range. At higher temperatures the energy consumption of internal fans in servers and other IT equipment will prevail over the improved efficiency of the data centre cooling system (see references). Lower temperature settings waste energy through overcooling.

In addition to temperature settings, airflow optimization (e.g. hot aisle/cold aisle, blanking plates and sealing leaks) is essential to ensure high efficiency. See reference [2] for hot-aisle cold-aisle optimization. Especially higher temperature settings require optimised air-flow to avoid hot spots. At very high power densities (e.g. 25 kW per rack), traditional room cooling based on CRAC/CRAH systems is no longer sufficient to prevent hot-spots. See references [3], [4] and [5] for more detailed information. In this case special rack- and row-based cooling may be appropriate.

RECOMMENDATIONS FOR BEST PRACTICE

Management of cooling systems:

- Control and manage the environmental conditions (set point, schedule, position and number of sensors).
- Replace obsolete or less efficient components of the cooling system (compare the efficiency class of existing systems) with more efficient ones available on the market.
- Verify the ducts/pipes insulation (cold and hot air/water/liquid).
- Locate CRAC at the end of the hot aisle (units are to be placed perpendicularly to the hot aisles).
- Segregate equipment with different airflow/temperature requirements.
- Air flows:
  - Place air supplies (perforated floor tiles or diffusers) in cold aisles only, near the active IT equipment.
  - Install airflow barriers as hot aisle and/or cold aisle containment to reduce mixing of hot exhaust air with cooler room air.
  - Install blanking panels at all open rack locations and within racks to prevent recirculation of hot air
- Cable order:
  - Use overhead cable tray.
  - Control the positioning and the sealing of cable openings and floor tiles.

Criteria for selecting new energy efficient cooling systems:

- Compare efficiency of chiller units, (see references for cooling requirements).
- Compare the different air flow design options (cold/warm aisles, raised floor/return plenum concepts).
- Evaluate the use of:
  - rack based cooling (for high density systems)
  - free cooling (direct/indirect)
  - free water cooling
  - installation of liquid cooling (direct/indirect)
  - waste heat recovery
- Set up a modular cooling system (linked to the IT design concept and management).
- Use Computational Fluid Dynamics (CFD) simulation software for optimisation of the cooling process.

5) the value measured by a thermometer freely exposed to the air but shielded from radiation and moisture level, typically the air temperature

5.2.3 Component efficiency – chillers, fans, air handling units

Air cooled and liquid cooled chillers differ regarding their EER (Energy Efficiency Ratio[6]) which is typically around 3.5 for water systems and around 2.5 for air systems. The "rated energy efficiency ratio" (EERrated) expresses the declared capacity for cooling [kW] divided by the rated power input for cooling [kW] of a unit when providing cooling at standard rating conditions. Eurovent provides data which allows a comparison of the characteristic efficiency of several cooling and ventilation systems and components (*www.eurovent-certification.com*). Water-cooled chillers are a first choice over air-cooled and DX, thanks to the higher thermodynamic efficiency. The opportunity to decrease condensing temperature or increase evaporating temperature should be evaluated. Reducing delta-T between these temperatures means that less work is required in the cooling cycle, hence improving efficiency. The temperatures are dependent upon the required internal air temperatures (see Temperature and humidity settings).

Efficiency of fans primarily depends on the motor efficiency. The use of fixed speed fans consumes substantial power and makes management of data floor temperature difficult. Variable speed fans are particularly effective in case of high redundancy in the cooling system or highly variable IT load. Fans may be controlled by the return air temperature or the chilled air plenum pressure.

6) Energy Efficiency Ratio: ratio of output cooling to input electrical power at a given operating point (indoor and outdoor temperature and humidity conditions)

5.2.4 Free cooling

"Free cooling" is a technique providing cooling by use of the lower level of external air/water temperatures compared to the indoor required conditions. The lower the average external temperature is over a year, the higher the opportunity for free cooling and the efficiency level. Waterside and airside economisers may provide an alternative for supplemental cooling. Climatic conditions define the economic efficiency and payback of investments. Full free cooling operating mode can be used if the difference between the cooling water's return temperature and the ambient temperature is bigger than about 11 K. Consequently, the higher the designed inlet temperature, the higher the energy savings. If a higher server room temperature is chosen for a cooling system's design, free cooling can be used for a longer period of time per year. Free cooling implementation requires a feasibility check and an economic evaluation. For estimated savings also see the evaluation tool for free cooling developed by The Green Grid.
Recommendations regarding sources with specific information on free cooling are provided in the section on further reading.

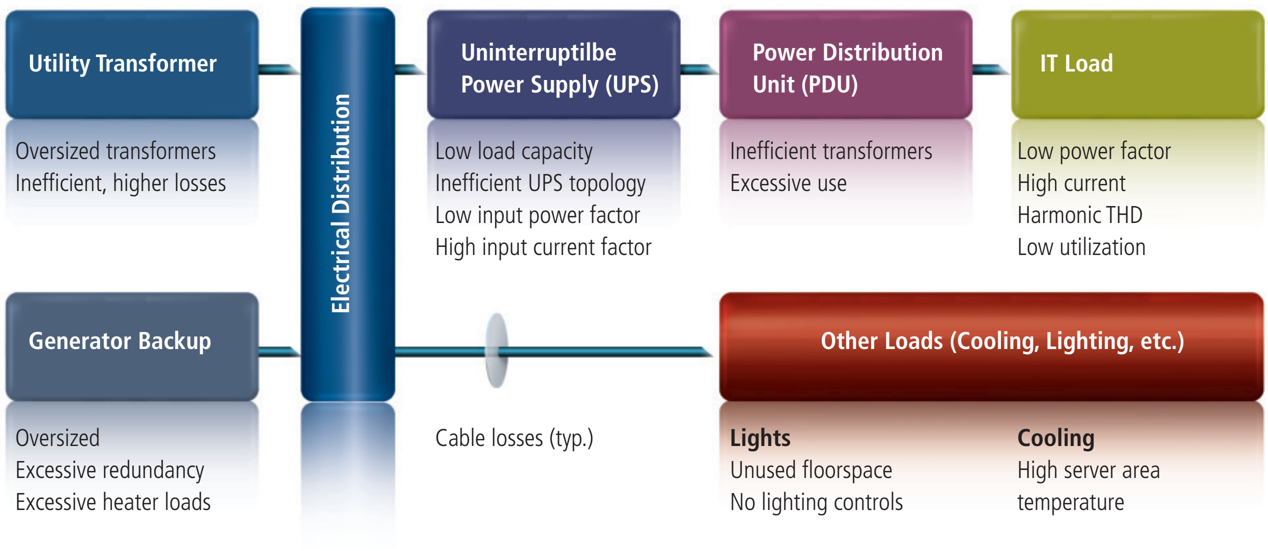


Fig. 5.3: **Electrical infrastructure components and inefficiency in a data centre** *(ASHRAE: Save Energy Now Presentation Series, 2009).*

5.2.5 Rack based cooling / in row cooling

If the power density of modern IT equipment is over 25 kW per rack, traditional room cooling based on CRAC/CRAH systems is no longer sufficient to prevent hot-spots. See references for more detailed information.

5.3 Power supply and UPS in data centres

The power supply system in a data centre primarily transforms the current from alternate (AC) to direct (DC). Losses due to transforming vary depending on the load level. The highest efficiency is typically reached between 80 and 90% of the total load while for levels below 50% energy efficiency decreases significantly.
Figure 5.3 shows the typical power chain in data centres. Typical sources of inefficiency are indicated for all components.

Uninterruptible Power Supply (UPS) systems often provide a large potential for energy savings. UPS is continuously operated to provide standby power and power conditioning for IT equipment and parts of the infrastructure.

Besides their primary function, which is to provide short-term power when the input power source fails, UPS also provide different features to correct utility power issues. Three main system topologies are available, depending on the application desired:

- Passive standby, also called Voltage and Frequency Dependent (VFD), is solely capable of protecting the load from power disruptions (power failures, voltage dips, surge voltages). In a normal electric supply situation the UPS has no interaction with the utility power. When the input supply is outside UPS design load tolerances, an inverter engages the energy storage mechanism to provide power to the load, bypassing utility electrical supply. This topology is more common in low-power applications.
- Line interactive, also called Voltage Independent (VI), is capable of protecting the load like a VFD UPS and in addition provides protection to the load by regulating frequency within optimal limits. In particular, it protects from undervoltage applied continuously to the input, or overvoltage applied continuously to the input. This topology is not commonly used above 5,000 VA [7].
- Double conversion, also called Voltage and Frequency Independent (VFI), is capable of protecting the load against adverse effects from voltage (like a VI) or frequency variations without depleting the stored energy source, as it continuously supplies total load power by regulating utility electricity before it reaches the load. This topology is rare for loads below 750 VA.

Each topology has its advantages and drawbacks. In the range 750 VA–5,000 VA line-interactive UPS tend to have longer operating lives and increased reliability with a lower total cost of ownership, while double conversion on-line UPS occupy less space and can regulate output frequency. UPS can also offer different energy storage mecha-nisms to supply power to the attached load in the event of power disruption:

- Electrochemical batteries, storing and discharging electrical energy through the conversion of chemical energy;
- rotary (flywheel), providing short term energy storage in the form of a spinning massive disk.

Tab. 5.2: **Characteristic efficiencies of UPS topologies**

UPS Topology	Efficiency at 25% load	Efficiency at 50% load	Efficiency at 75% load	Efficiency at 100% load
Double-conversion	81–93%	85–94%	86–95%	86–95%
Line-interactive	n.a.	97–98%	98%	98%

Tab. 5.3: **Minimum average efficiency requirements for AC-Output UPS proposed in EnergyStar UPS** (P is the Real Power in Watts (W), ln is the natural logarithm)

Minimum Average Efficiency Requirement (EffAVG_MIN),				
UPS Class	Output Power	Input Dependency, as specified in the ENERGY STAR Test Method Product Class		
		VFD	VI	VFI
Data centre	P > 10 kW	0.97	0.96	0.0058 x ln (P) + 0.86

Two options are available for delivering the energy to the load:

- Static UPS: no moving parts in the power path (except the fans for cooling). It converts AC power into DC (rectifier for storage in batteries to provide continuity in case of mains loss) and then into AC again for power supply units installed in servers.
- Rotary UPS: transfers power via a motor/generator and is used for applications requiring ride-through of short-duration power system outages, voltage dips, etc.

The UPS energy losses are due to electrical power conversion inefficiencies (in the charger and inverter) and battery charging losses or energy losses in inertial systems (flywheels). The electric losses (and the heat generation) are more important in double conversion UPS (rectifier, inverter, filter, and interconnection losses), than in line-interactive and standby UPS (filter, transformer, and interconnection losses). DC-output UPS (also known as rectifiers) and combined AC-DC-output UPS can be used for some applications and may avoid losses in the inverter and rectifier.

RECOMMENDATIONS FOR BEST PRACTICE

Criteria for new installations

- Assess your needs and size the UPS systems correctly (evaluate multiple or modular UPS, scalable and expandable solutions): battery back-up time, cost, size, number of outlets, etc.
- Analyse the UPS technology and efficiency. Take into account the partial load efficiency of UPS.
- Select correct topology of the power supply systems.
- Select UPS systems compliant with the EU Code of Conduct for UPS or Energy Star.

Criteria for optimisation

- Analyse the UPS technology and efficiency.
- Evaluate options and benefits of replacement of old equipment.
- Evaluate costs and benefits of redundancy.

Most UPS manufacturers quote UPS-efficiency at 100% load. However efficiency drops off significantly at partial load conditions (see Most UPS run at 80%, and in case of redundancy, the load may drop to 50% and below. At loads of 50% or lower both modern and legacy UPS systems run less efficiently with significant dips occurring at loads below 20%. For best practice UPS loads shall be matched as closely as possible to the data centre IT loads. Scalable UPS solutions are available for efficient sizing of UPS capacity.

Minimum efficiency requirements for UPS are specified in the EU Code of Conduct for UPS (new edition 2011) and in the Energy Star programme requirements (draft version 2011). New Energy Star energy efficiency requirements for AC-Output and DC-Output UPS are currently under development (see Table 5.3).

The programme also considers to include requirements for multi-modes UPS. This type of UPS operates with more than one set of input dependency characteristics (e.g. can function as either VFI of VFD). Multi-mode UPS can run in more efficient less protective modes and switch to less efficient higher protective modes, when necessary. Thus significant energy savings are possible.

Further Reading

ASHRAE (2011): Thermal Guidelines for Data Processing Environments – Expanded Data centre Classes and Usage Guidance – ASHRAE, 2011, online available at: *http://tc99.ashraetcs.org/documents/ASHRAE%20Whitepaper%20-%202011%20Thermal%20Guidelines%20for%20Data%20Processing%20Environments.pdf*

EU Code of conduct for data centres (2009): Full list of identified best practice options for data centre operators as referenced in the EU Code of Conduct:
http://re.jrc.ec.europa.eu/energyefficiency/pdf/CoC/Best%20Practices%20v3.0.1.pdf

The Green Grid (2011): Evaluation tool for free cooling.
http://cooling.thegreengrid.org/europe/WEB_APP/calc_index_EU.html

ENERGY STAR (2011): UPS efficiency
http://www.energystar.gov/index.cfm?c=new_specs.uninterruptible_power_supplies

The Green Grid (2011): Evaluation tool for power supply systems
http://estimator.thegreengrid.org/pcee

High Performance Buildings: Data centres Uninterruptible Power Supplies (UPS)
http://hightech.lbl.gov/documents/UPS/Final_UPS_Report.pdf

EU CODE of CONDUCT (2011): EU code of conduct on Energy Efficiency and Quality of AC Uninterruptible Power Systems (UPS):
http://re.jrc.ec.europa.eu/energyefficiency/html/standby_initiative.htm

References

[1] **ASHRAE:** Save Energy Now Presentation Series, 2009.
[2] **Niemann, J. et al. (2010).** Hot-Aisle vs. Cold-Aisle Containment for Data centres; APC by Schneider Electric White Paper 135, Revision 1.
[3] **Rasmussen, N. (2010).** An improved architecture for High-efficiency High-density data centres; APC by Schneider Electric White Paper 126, Revision 1.
[4] **Blough, B. (2011).** Qualitative analysis of cooling architectures for data centres; The Green Grid White Paper #30.
[5] **Bouley, D. and Brey, T. (2009).** Fundamentals of data centre power and cooling efficiency zones; The Green Grid White Paper #21.
[6] **Rasmussen, N. (2011).** Calculating Total Cooling Requirements for Data centres; APC by Schneider Electric White Paper 25, Revision 3.
[7] **ENERGY STAR Uninterruptible Power Supply Specification Framework (2010).** Available at:
www.energystar.gov/ia/partners/prod_development/new_specs/downloads/uninterruptible_power_supplies/UPS_Framework_Document.pdf
[8] **Ton, M. and Fortenbury B. (2008).** High Performance Buildings: Data centres - Uninterruptible Power Supplies. Available at
http://hightech.lbl.gov/documents/UPS/Final_UPS_Report.pdf
[9] **Samstad, J. and Hoff M.;** Technical Comparison of On-line vs. Line-interactive UPS designs; APC White Paper 79. Available at
http://www.apcdistributors.com/white-papers/Power/WP-79%20Technical%20Comparison%20of%20On-line%20vs.%20Line-interactive%20UPS%20designs.pdf

PrimeEnergyIT
EFFICIENT DATA CENTERS

Partners

Supported by

Contact: Austrian Energy Agency | Dr. Bernd Schäppi | Mariahilferstrasse 136 | A-1150 Vienna |
Phone +43 1 586 15 24 | bernd.schaeppi@energyagency.at | www.efficient-datacenters.eu

foorfour Agentur für Kommunikation