Preparatory study for implementing measures of the Ecodesign Directive 2009/125/EC

DG ENTR Lot 9 - Enterprise servers and data equipment

Task 4: Technologies

July 2015 – Final report

Document Information

Client	European Commission, DG Internal Market, Industry, Entrepreneurship and SMEs
Specific Contract N°	185/PP/ENT/IMA/12/1110333
Framework Contract N°	ENTR/29/PP/2010/FC LOT 2
Report Title	Preparatory study for implementing measures of the Ecodesign Directive 2009/125/EC DG ENTR Lot 9 - Enterprise servers and data equipment - Task 4: Technologies
Project Name	Preparatory Studies for Product Group in the Ecodesign Working Plan 2012-2014: Lot 9 – Enterprise servers.
Project Code	ENTR Lot 9
Project Team	Bio by Deloitte, Fraunhofer IZM
Date	31 July 2015
Authors	Mr. Anton Berwald, Bio by Deloitte Mr. Thibault Faninger, Bio by Deloitte Ms. Sara Bayramoglu, Bio by Deloitte Mr. Benoît Tinetti, Bio by Deloitte Mr. Shailendra Mudgal, Bio by Deloitte Dr. Lutz Stobbe, Fraunhofer IZM Dr. Nils Nissen, Fraunhofer IZM
Key Contacts	Mr. Anton Berwald, Bio by Deloitte aberwald@deloitte.fr or Mr. Benoît Tinetti, Bio by Deloitte btinetti@deloitte.fr
Disclaimer	This report was prepared for the European Commission. However, it contains the results of research by the authors and is not to be perceived as the opinion of the European Commission. The project team does not accept any liability for any direct or indirect damage resulting from the use of this report or its content.

Please cite this publication as: Bio by Deloitte (2015), Preparatory study for implementing measures of the Ecodesign Directive 2009/125/EC DG ENTR Lot 9 - Enterprise servers and data equipment - Task 4: Technologies

Table of Contents

List of Figures

List of Tables

Glossary

AC	Alternating Current
ACL	Access Control Lists
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
B2B	Business-to-Business
BEOL	Back-End-of-Line
CAPEX	Capital Expenditure
CE	Controller Enclosure
CMOS	Complementary Metal Oxide Semiconductor
CPU	Central Processing Unit
DC	Direct Current
DDR	Double Data Rate
DE	Disk Enclosure
DIMM	Dual Inline Memory Module
DRAM	Dynamic RAM
EC	European Commission
ECC	Error-correcting Code
EEE	Energy Efficient Ethernet
ENIG	Electroless Nickel Immersion Gold
ErP	Energy-related Products
EU	European Union
FC	Fibre Channel
FCoE	Fibre Channel over Ethernet
FEOL	Front-End-of-Line
GPU	Graphic Processor Unit
HASL	Hot Air Solder Levelling
HBA	Host Bus Adapter
HDD	Hard Disk Drive
HPC	High Performance Computing
IB	Infiniband
IC	Integrated circuit
ICT	Information and Communication Technology
IDS	Intrusion Detection System
IHS	Integrated Heat Spreader
ILM	Integrated Load Mechanism
I/O	Input Output
IOPS	Input Output Per Second
iSCSI	internet Small Computer System Interface
LAN	Local Area Network
MEErP	Methodology for the Ecodesign of Energy-related Products
MLC	Multi-Level Cell
MTBF	Mean Time Between Failures
NFS	Network File Service
OEM	Original Equipment Manufacturer

OPEX	Operational Expenditure
OSP	Organic Solder Preservative
PCB	Printed Circuit Boards
PCF	Product Carbon Footprint
PCH	Platform Controller Hub
PGA	Pin Grid Array
PSU	Power Supply Unit
QoS	Quality of Service
QPI	Quick Path Interconnect
RAID	Redundant Array of Independent Disks
RAM	Random Access Memory
RU	Rack Unit
SAS	Serial Attached SCSI
SATA	Serial Advanced Technology Attachment
SCSI	Small Computer System Interface
SDRAM	Synchronous Dynamic Random Access Memory
SLA	Service Level Agreement
SLC	Single-Level Cell
SMPS	Switched-Mode Power Supplies
SMT	Surface-mount technology
SNIA	Storage Networking Industry Association
SPEC	Standard Performance Evaluation Corporation
SRAM	Static RAM
SSD	Solid State Devices
TDP	Thermal Design Power
THT	Through-hole technology
TIM	Thermal Interface Material
UPS	Uninterruptible power supply
VPN	Virtual Private Network

1.Technical product description

1.1. Existing products (working towards the definition of Base-Cases)

1.1.1. Objective and structure of the report

According to MEErP, this task aims to explain in easy-to-understand wording for non-experts what physical or chemical processes are involved in the functional performance of the product, in particular where such processes are responsible for resources use and emissions. At the same time, the explanation is also directed at technical experts, presumably the designers and developers of the industry placing the products on the market. This means that it should be identified and reported what the latest research findings are and what they would imply for the future functional and environmental performance.

The following section 1.1.2 provides a general overview of the main technical and environmental aspects of existing products. The remaining sections are for capacity building and they provide detailed technical descriptions of server and storage system designs, product configurations, and subassemblies. The report also addresses the connectivity and network technologies in conjunction with system architecture of existing server and storage systems.

Finding an effective methodical approach to this technical product description has been a challenge. According to MEErP, it is required to describe all technical and environmental aspects directly on a product level. This would mean to describe enterprise servers, storage and network equipment independently. But this approach has some limitations.

The investigation of existing products indicated a tremendous variety of existing technologies, product features and component configurations. The reason for this product spectrum is related to a number of factors including:

- Very fast technology cycles in the semiconductor / microelectronics industry
- Competing connectivity / network standards
- Functional / technical convergence on product level
- Operational requirements (e.g. these requirements include latency, capacity, reliability, and serviceability and are defined in individual Service Level Agreements - SLA)
- Operating condition and allowances on data centre level
- CAPEX and OPEX considerations

Against that background, it seems important to develop a basic understanding for the complex, functional system interaction between servers and storage equipment (connectivity) and the influence of surrounding operational conditions. Therefore, in the description of connectivity technologies, integrated network interfaces and independent network elements were included directly in conjunction with the description of servers and storage products.

The analysis leads to the conclusion that the environmental impact of the ENTR Lot 9 products depends more upon the system design, than on the product design. The following sections provide background information, data and summaries concerning not only the existing products but also concerning the system design, system interaction and utilisation of the product within a system effort.

1.1.2. Product overview

Enterprise servers and storage equipment are commercial information and communication technology (ICT) sold in the business-to-business (B2B) market. Server and storage products provide individual functionalities. They are two separate product categories and further sub-categories have been defined in order to distinguish certain applications, performance and form factors. It is furthermore important to recognise the fact that server and storage equipment are mostly operated interconnected. The workload is typically distributed to a larger pool of products. The provided IT services (e.g. user benefit, useful work) are not created by a single machine but through a combined (virtual) server and storage environment. Connectivity and interoperability are essential aspects of server and storage equipment. The technologies, software, and specific products applied to facilitate connectivity are manifold and considered a distinct and specific spectrum of products.

Due to the fact that the majority of server and storage equipment are pooled together, specific mounting options and deployment environments have been developed over the past decades. Most of the server and storage equipment is designed to be mounted in standard 19-Inch racks (cabinets) and placed in air-conditioned server rooms or data centres. The rack and room level support infrastructure is considered in this study through the "technical system" approach. This rack and room level infrastructure is not considered part of the server or storage product. However, this support infrastructure is absolutely necessary for safe, secure and reliable operation. It basically includes cooling and air conditioning infrastructure as well as uninterruptible power supply (UPS) and power distribution equipment. Fire safety, condition monitoring, physical security and other aspects are also part of the infrastructure.

Even if the study scope is not the data centre facility and particular infrastructure equipment, it is necessary to recognise that the local conditions (e.g. climate, cooling options) in which the ICT equipment is operated have an influence on the technical performance and the environmental impacts of the products (see Task 3 for more details).

Current enterprise servers and storage equipment are designed in a more and more modular way. Vendors offer many different configuration options in order to adjust performance and price to the needs of the customers. The products can be distinguished by various aspects including:

- Purchasing price: This is an indicator for a certain performance and quality of components and it reflects the hardware and software configuration.
- Hardware configuration: The type and number of processors, memory and storage capacity, and interfaces are indicating the performance spectrum and intended application.
- Software configuration: The type of operation system, applied software (proprietary or open-source) and virtualisation capability as well as the applied network protocols in storage systems for instance are also indicating the performance spectrum and intended application.
- Form factor: This is an indicator for the level of system integration, modularity and scalability of functional components. There is a trend towards smaller form factors and multi-functionality in current product design.
- Allowable operating conditions: Products are offered for different operating conditions including thermal and humidity specifications.

Previous investigations indicate that the primary environmental impact of enterprise servers and storage equipment is related to the electricity consumption in use phase. Nevertheless, the materials and processes used for manufacturing the specific hardware elements are also contributing to the overall environmental impact particularly in impact categories linked to resource use. The manufacturing of electrical, electromechanical, electronic and photonic components is in general resource intensive. The following components can be considered in that respect:

- Large-size, multilayer printed circuit boards (PCB) populated with semiconductor-based active components (ICs) such as processors (CPU, GPU), memory (RAM), storage devices (HDD, SSD) as well as the full range of passive electronic devices.
- Active and passive cooling elements such as fans, heat-sinks or heat-pipes, which typically consists of aluminium, copper or a combination of both.
- One or multiple (redundant) power supply units (PSU) including big caps and coils
- Cables for power supply and communication (copper or glass fibre)
- Global distribution of the supply chain and manufacturing process (transport)

The hardware is to a very large extent produced outside of the European Union. The material composition of the products is changing slightly over time, mostly due to new technologies in the semiconductor and electronic packaging field. The continuously growing demand for bandwidth and higher frequencies will lead to an increasingly larger introduction of photonic technologies and related materials such as Gallium, Germanium, and Indium in III-V semiconductor components. Furthermore, the steady miniaturisation (Moore's Law) is leading to higher and higher energy density. This trend demands also the utilisation of new material compositions and structures on the chip level in combination with novel cooling systems. This in turn could impact the complexity and environmental footprint of the manufacturing processes.

Finally, the material recovery in the end-of-life treatment is not yet at an optimum. There is a variety of materials that are not fully recovered from the products at the present. Nevertheless, the industry already recognises the material value of the products and individual stakeholders are implementing refurbishment, dismantling, and recycling schemes in their business.

Existing product carbon footprints (PCF) of enterprise servers, conducted by leading manufacturers and environmental assessment experts, indicate that over the whole product life cycle, the use phase is the most contributing life cycle phase (more that 80% of the total PCF, see Task 3). The following aspects have to be considered in that respect:

- The products are typically operated 24/7 for a period of 3 to 7 years, and are always online. Standby/off mode is not a common practice due to prolonged reactivation time on a system level.
- Average server utilisation is currently still quite low (typically 20% load). Server utilisation is slowly increasing and primarily driven by virtualisation.
- Moore's Law is still contributing strongly to periodical performance increase and related improvement of energy efficiency. Performance and power consumption are always addressed together in technology development.
- The operating conditions including cooling infrastructure and energy supply are directly related to the power consumption of the products. It is necessary to consider energy trade-offs between the ICT equipment and the supporting infrastructure as being relevant for energy savings on system level (data centre).

The energy consumption in the use phase is influenced by the technology level of the hardware as well as by the conscious deployment of a product for the right purpose. In other words, the specific hardware and software configuration determines the power consumption in relation to the functional performance. This includes not only the power draw of the electronics but also the dimensioning and conversion efficiency of the power supply unit, the cooling requirements, as well as energy and material overhead for the required cooling.

Components selection, thermal design features and air flow characteristics have a substantial influence on the overall energy consumption. The trend goes to higher operating temperatures with allowable inlet temperature of up to 35°C. When addressing the energy consumption and energy efficiency, it is therefore necessary to consider the operating conditions (room level) and procedures (service level). This means for instance that the external temperature and air flow conditions, availability of certain cooling media, as well as the type of energy source will largely influence the environmental impact of the data centre equipment in the use phase.

Energy efficiency has been strongly addressed by the industry over the past seven years in conjunction with initiatives of the U.S. EPA to develop test standards and benchmarks that quantify the performance-to-power ratio. There are standardised benchmark tests available now which measure the energy consumption of server and storage equipment according to a specific applications or workload. These performance-to-power benchmarks are providing data for average and specific power consumption of the equipment (see also Task 1) and a comparison of existing measurement results indicates that the actual hardware and software configuration is influencing the power consumption considerably.

A high utilisation of the server and storage equipment – meaning a high load level – is one aspect that has been stressed over the past years in data centres in conjunction with better resource utilisation and hardware consolidation. With the introduction of virtualisation technologies and respective software tools (middleware), it has now become possible to run multiple "virtual servers" on one physical machine. Virtualisation is also growing in the area of storage and network equipment.

1.1.3. Enterprise servers

1.1.3.1.Server components and subassemblies

As already outlined in the previous section, an enterprise server consists of the following main subassemblies:

- **Chassis**: A housing made of metal and plastic material that encloses the electronic components (server-board) and provides mounting features e.g. for the power supply unit, cooling elements, and other functional components.
- **Server-board and electronic components**: This is the main printed circuit board (PCB) that provides the computing and communication functionality and which is populated with active components including the CPU with attached memory and chipset, storage media and network devices as well as passive electronic components including resistors, capacitors, etc.
- **Cooling system**: The cooling of the active electronic components such as the CPU is typically achieved by a combination of passive and active technologies. In general, passive cooling includes a heat spreader directly attached to the top of the processor chip, a large heat sink or heat pipes that distribute the thermal energy away from the chip and towards the active cooling system which is most often a fan unit.

- **Storage devices and drives**: The storage system derives historically from hard disc drives (HDD) and includes nowadays 3.5-inch HDD, 2.5-inch HDD as well as semiconductor-based solid state devices (SSD) that are integrated into a housing of a 2.5-inch disk drive. Some servers feature other drives (CD-ROM/Blu-Ray) as well.
- **I/O control and network connectors**: Servers are remotely accessible via Ethernet connection and therefore feature multiple network interfaces and links (connectors) on the backside. The current products feature 1 GE (1 Gigabit per second Ethernet) and 10 GE links. The trend goes towards 40 GE and 100 GE connections. Copper and optical fibre cables are utilised in data centres.
- **Power supply unit**: The PSU is typically configured as a single unit or a multiple unit and comes in its own housing (metal cage).

Figure 1 shows the main subassemblies and illustrates the basic design layout of an enterprise server.

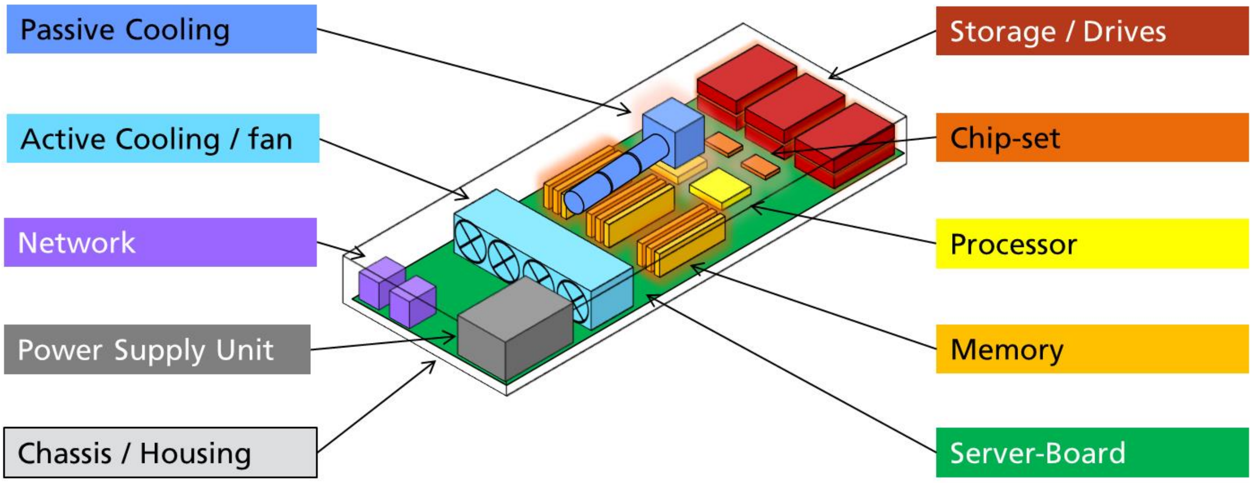


Figure 1 : Enterprise server - main subassemblies

The following subsections describe the function and technology, as well as the material and energy related aspects of the main subassemblies and important components.

1.1.3.2.Chassis

- **Technology:**

The form factors, dimensions, and intended way of mounting dictates the design of the server’s chassis (also typically called enclosure, case or housing). In most cases, the chassis is a simple metal box with a type of frame and mounting parts. The purpose of the chassis is to enclose and mount the main subassemblies including the server’s printed circuit boards, storage devices, integrated fans and passive cooling devices, the power supply unit, and interfaces. The chassis may feature rails and cages (bays) for mounting different exchangeable devices such as disk drives or power supply units.

Enterprise servers can be stand-alone devices, where the chassis includes a pedestal. However, the increasing majority of enterprise servers are nowadays mounted in 19-inch server racks. With respect to the chassis, it is necessary to distinguish between:

- Integrated single servers (e.g. rack server)
- Modular server systems (e.g. blade system)

Integrated single servers are typical rack-optimised servers, where all functional subassemblies are in the main chassis of the server. The dimension of this chassis can vary in height and depth (see chapter 1.1.3.7). The integrated server is inserted and fixed directly in the 19-inch rack. The server rack usually fits multiple rack-servers. The individual servers need to be manually connected to the power distribution and network (cabling). With respect to rack-mounted servers, the front- and back-side sometimes feature perforated metal or plastic covers that contain LEDs or a display as user interface, openings for disk drives, power supply units and network connectors. Many subassemblies including disk drives and power supply units are hot-swappable nowadays, meaning they can be exchanged while in operation. Also, the top covers of the servers are usually removable for easy access to the main subassemblies. The inside of the chassis provides frames or rails for fastening the subassemblies such as the mainboard and cages for power supply units and fan units. The chassis slides into the rack on standard rails and is fixed with quick to open fasteners. Larger units (>4U) have metal or plastic handle bars on the outside for better handling during installation.

Modular server systems have the advantage of shared resources and prefabricated connectivity. Due to the requirement of providing scalable server capacity with exchangeable server and storage modules, new form factors with modular enclosures have entered the market. Modular server systems such as blade systems or micro server systems consist of two housing elements. One element is the individual chassis for the server modules. A server module is basically the server boards with connectors for power and I/O and is called blade, cartridge or book, depending on the manufacturer’s preference. A certain number of server modules are inserted into a larger system enclosure, which is housing the shared resources including e.g. network, storage, cooling and power supply devices. The fastening mechanism is mostly based on rails and clips in support of fast exchangeability. The system enclosure is designed to fit into the 19-inch server rack. Depending on the overall weight of the system, special handles are provided. The market shows a high diversity in terms of dimensions, form factors, and system configurations.

- **Material considerations:**

Depending on the form factor and actual dimensions of the chassis, the material value varies accordingly. Most chassis are made of low-alloyed steel or chromium steel, brass and some plastic parts. The number of fasteners such as screws and clips are varying from product to product. There is no average number or dominant fastener mechanism to name. The designs are manifold. However, the chassis design and fastening mechanisms need to be addressed in an eco-design perspective, due to the feasibility of separating the chassis and its main bulk materials (sheet metal and plastics) in a recycling-oriented product end-of-life treatment.

- **Energy considerations:**

Air flow and thermal environments are key considerations in optimizing PUE and lowering energy consumption. Most air cooled servers receive cool air in from the front and exhaust hot air through the rear. Conversely, some network systems may receive air from the sides. Coordinated air flow is a necessary consideration when optimizing the environmental conditions and reducing the energy consumption used to cool these systems. Some chassis feature baffle plates for better air flow. The chassis functions sometimes as a heat spreader. The placement and utilisation of multiple servers in a rack will also influence the surrounding thermal conditions.

1.1.3.3.Printed circuit board

- **Technology:**

The server’s mainboard (server-board) is the carrier for the active semiconductors devices (processor, memory, etc.), passive components (resistors, capacitors, inductions, etc.), various sockets (CPU socket, DIMM sockets, etc.) and connectors (Ethernet ports, USB ports, etc.) that provide the intended computing and communication functionalities. In essence, the mainboard realises the mounting and conductive paths between the components.

The mainboard is a multi-layer printed circuit board (PCB) on which the electronic devices are directly connected (soldered) mostly by surface-mount technology (SMT) or through-hole technology (THT). A server-board has about 12 layers (plus/minus 4). Due to the fact that many functional circuitries are nowadays realised as integrated circuits (ICs), the number of layers is not increasing.

- **Material considerations:**

The mainboard is made of FR-4 laminates and copper foils (cores). FR-4 is a composite material consisting of fibreglass sheets impregnated with an epoxy resin binder that is flame resistant (FR) according to UL94 V-0 flammability classification. The thickness of the FR-4 prepegs (premanufactured substrate layers) is typically a few hundreds micrometers (µm). The copper cores are foils with a thickness of about 18 µm or 35 µm. The single copper layers are connected through small, copper-plated holes (vias). **The dimensions of the printed circuit board and the number of layers are important parameters for the environmental assessment.**

Furthermore, the mainboard’s outer layer is coated with a protective surface finish. This finish varies in price according to the utilised materials. Common technologies are Electroless Nickel Immersion Gold (ENIG) and Immersion Silver (Imm Ag) in higher end products and Immersion Tin (Imm Sn), Organic Solder Preservative (OSP) or lead-free Hot Air Solder Levelling (HASL) in more economic products. Currently Cu with solderability preservatives (CuOSP) is regarded as the established technology.

Common solders are nowadays Sn-Ag-Cu (Tin-Silver-Copper) alloys also being used in servers. However there is an exemption according to RoHS 2[1] allowing lead for solder in servers (see section 4.1.3 in Task 1). So far the study could not collect quantitative data on the actual use of lead-based solders in servers and storage equipment, but stakeholders report that it is hardly used anymore. An application criterion for certain solders and surface finishes is the intended operation temperature. Lead-based solders are typically used for high temperature applications e.g. chip packages. According to industry, higher density board designs may require lead based solders for dual sided reflow depending on the board design and mass of the (SMT) components.

With respect to the environmental impact of the passive electronic components, the use of tantalum capacitors needs to be reconsidered. The rare tantalum oxide is superior in its technical properties as dielectric in comparison to most electrolytic or ceramic capacitors. However, due to the critical conditions under which tantalum is mined and the relative high price, tantalum capacitors have been substituted over the past years but might still be found in some servers.

- **Energy considerations:**

Galvanic processes in the manufacturing of printed circuit boards as well as the higher temperatures of the lead-free soldering process are energy intensive in general.

1.1.3.4. Processor and memory

- **Technology:**

Processor: The central processor unit (CPU) in conjunction with the employed memory (RAM) and chipset is generally providing the computing functionality of the enterprise server. There are only a few processor architectures utilised in the server market:

- The x86 instruction set architecture is the most widely distributed and supports Microsoft Windows, Unix, Linux, AIX and Solaris operating systems. It is also compatible with virtualization software and most other software stacks employed in the industry. This processor architecture goes back to the original Intel 8086 CPU of the year 1978. The 8th generation of x86 systems with 64-bit linear address space and on-die memory controller has been well implemented over the past 10 years. The 9th generation features a 64-bit linear address with 40 to 48-bit physical address space and integrated on-die graphic processor unit (GPU). The main x86 server-CPU manufacturers are Intel with its Xeon processor family and AMD with the Opteron processor family.
- The SPARC processor architecture is another 64-bit proprietary technology which was originally developed by Sun Microsystems and supports the Oracle Solaris operating system. SPARC processors are offered by Oracle and Fujitsu (the SPARC T and M series by Oracle, and the SPARC64 X series by Fujitsu). The SPARC T-5 features high workload performance through 16 cores and possible 8 threads per core. The SPARC64 X processor features a maximum operating frequency of 3 GHz, 16 cores per chip with 2 threads per core, high capacity cache and integrated functionality as System-on-Chip. The SPARC64 processor is produced by Fujitsu with an advanced 28nm technology node (manufacturing process).
- The POWER processor family is the high performance 64-bit architecture developed by IBM specifically for big-data handling. The specifications for the current generation POWER8 include an operating frequency up to 4 GHz, 12 cores per chip with 8 threads per core for a total of 96 threads of parallel execution. The processor features very high capacity on-chip and off-chip caches as well as a new extension bus called CAPI that can be used to attach dedicated off-chip accelerator chips such as graphics processing units (GPUs), application-specific integrated circuits (ASICs) and field programmable gate arrays (FPGAs). The POWER8 is supposed to be manufactured in a 22nm technology node using a silicon-on-insulator fabrication process.
- There are other architectures including Intel Atom, Intel Xeon Phi and FPGA-type Intel Xeon processors, ARM 64-bit SoC, GPGPUs from AMD and NVIDIA. Upcoming trends of cloud computing being used for gaming purposes and similar applications result in adapted server

---

[1] RoHS 2 : DIRECTIVE 2011/65/EU Annex III 7(b): Exemption "Lead in solders for server, storage and storage array systems, network infrastructure equipment for switching, signalling, transmission, and network management for telecommunications" http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32011L0065&from=EN

configurations and technology implementations that relate to different requirements compared to common server configurations.

The technical performance and features of a CPU is characterised by a number of factors including:

- Technology node, die dimensions and resulting number of transistors
- Number of cores per chip, threads per core, functional specialisation of the core
- Cache configuration and capacity, supported memory
- Number, type and control of I/Os
- System-on-chip or further integrated functionality such as power and memory control
- Maximum operating frequency, frequency scaling per core,
- Operating voltage, power scaling options, intelligent throttling
- Packaging and mounting on the mainboard including type of socket

The semiconductor industry is currently manufacturing processor chips with a technology node of 32 nm, 28 nm, and 22 nm (size of the smallest structure created by the lithography process). Intel's technology roadmaps indicate that CPU production in 14 nm technology is the next phase starting approximately in 2015. The challenges of the ongoing miniaturisation are manifold. Electro-migration and further voltage scaling are to be named. New designs and materials with a trend towards optical systems are part of the solution. The performance improvement will furthermore address the functional specialisation and improved control of the individual processor core (see section 1.4).

The following Table 1 provides a simplified overview of existing variety of CPUs on the example of the current versions of Intel Xeon processors.[2] The list differentiates Xeon processor families E3 and E5 for managed volume servers, and E7 for resilient servers (with significantly more complex RAS functionality and power profiles). The list also provides some technical data regarding the collaborating chipset, the type(s) of socket, number of cores and threads, the operation frequency and the memory configuration. From an environmental point of view, the thermal design power (TDP) value, if available, has been added, which indicates the power draw and heat signature, the die and packaging size as well as the sales price in US Dollar.

---

[2] About 90% of all CPUs for servers are currently shipped by Intel. AMD has a market share of 9% and IBM the remaining.

Table 1: Intel Xeon Processor Overview

Intel XEON Processors	E3 (1-P)	E5 (1-P)	E5 (2-P)	E5 (4-P)	E7 (2-P)	E7 (4-P)	E7 (8-P)
Sandy Bridge-based Xeons (32nm)	E3-12XX v1	E5-14XX v1 (16XX)	E5-2XXX v1	E5-46XX v1			
Chipsets	C600 series	C600 series	C600 series	C600 series			
Socket	LGA 2011/ 1155	LGA 2011/ 1356	LGA 2011/ 1356	LGA 2011			
Cores (Threads)	2-4 (4-8)	4-6 (8-12)	4-8 (8-16)	4-8 (8-16)			
Clock rate	2.2 - 3.6 GHz	1.8 - 3.6 GHz	1.8 - 3.1 GHz	2.0 - 2.9 GHz			
Memory	1x - 2x DDR3	1x - 3xDDR3	3x - 4xDDR3	4xDDR4			
Thermal Design Power (TDP in Watt)	20 W - 95 W	40 W - 130 W	50 W - 150 W	95 W - 130 W			
Die size (mm²)	D2: 216 mm², Q0: 131 mm²	D2: 216 mm², Q0: 131 mm²	D2: 216 mm², Q0: 131 mm²	D2: 216 mm², Q0: 131 mm²			
Placed on the market (year / price)							
Ivy Bridge-based Xeons (22nm)	E3-12XX v2	E5-14XX v2	E5-2XXX v2	E5-4XXX v2	E7-28XX v2	E7-48XX v2	E7-88XX v2
Chipsets	C200 series	C600 series	C600 series	C600 series	C600 series	C600 series	C600 series
Socket	LGA 1155 / BGA 1284	LGA 1356/ 2011	LGA 1356/ 2011	LGA 2011	LGA 2011	LGA 2011	LGA 2011
Cores (Threads)	2-4 (4-8)	4-8 (8-16)	4-12 (4-24)	4-12 (8-24)	12-15 (24-30)	6-15 (12-30)	6-15 (12-30)
Clock rate	1.8 - 3.7 GHz	2.2 - 3.7 GHz	1.7 - 3.5 GHz	1.9 - 3.3 GHz	2.3 - 2.8 GHz	1.9 - 2.8 GHz	2.2 - 3.4 GHz
Memory	1x - 2x DDR3	1x - 4x DDR3	1x - 4x DDR3	1x- 4x DDR3	1x - 4x DDR3	1x - 4x DDR	1x - 4x DDR3
Thermal Design Power (TDP in Watt)	17 - 87 W	60 - 130 W	50 - 150 W	70 - 130W	105 - 155 W	105 - 155W	105 - 155W
Die size (mm²)	160 mm²	160 mm					
Package size	37.5mm x 37.5mm	52.5mm x 45.0 mm / 45mm x 42.5mm	45mm x 42.5mm /52.5 x 45(/51) mm	52.5 mm x 45 mm	52mm x 45mm	52mm x 45mm	52mm x 45mm
Placed on the market (year / price)	(2012/ 264$) - (2013/384$)	(2014/494$) - (2013/785$)	(2013/1302$)	(2014/ 2174$)	(2014/ 4412$)	(2014/ 3490$)	(2014/ 5437$)
Haswell-based Xeons (22nm)	E3-12XX v3	E5-16XX v3	E5-2XXX v3				
Chipsets	C220 series						
Socket	LGA 1150/ 1364	LGA 2011-3	LGA 2011-3				
Cores (Threads)	2-4 (4-8)	4-6 (	6-8 (				
Clock rate	1.1 - 3.7 GHz	3 - 3.7 GHz	1.6 - 3,2 GHz				
Memory	1x - 2x DDR3	4x DDR4	4x DDR4				
Thermal Design Power (TDP in Watt)	13 - 84 W		120 - 160 W				
Die size (mm²)							
Package size	37.5mm x 37.5mm						
Placed on the market (year / price)		Q3, 2013	Q3 2014				

With respect to the performance, the clock frequency, number of individually addressable cores (and threads), cache capacity, input/output (I/O) speed and capacity are very important design features of a CPU. In the past years, CPU development addressed integrated GPUs. In order to further improve the performance of servers, two or more processors units are connected via a high speed bus or routing interface. In the current Intel system architectures, this link is called quick path interconnect (QPI). Server systems with four and more processor sockets feature 2 to 4 QPI-links per processor in order to increase cross-connect capability.

Each processor consists of two or more cores allowing typically two threads per core. Multithreading aims to increase utilisation of a single core by using thread-level as well as instruction-level parallelism. The processor also features caches – memory buffers – on various levels (L1, L2, and L3) and other functional segments. Caches store information to avoid multiple computations and improve access times. The cache can be implemented as a hardware or software element. In a CPU, the cache answers most of the requests drastically lowering the effective load on the processor. Due to big and fast caches being rather uneconomic, there are usually several cache elements in a cache hierarchy. The smallest and fastest cache element is the first that attempts to answer a request; if the requested information is not stored, the request is handed to the next bigger and slower cache element. Caches utilise the faster SRAM to fulfil access time demands.

Static RAM (SRAM) retains data bits in its memory while power supply is connected. Dynamic RAM (DRAM) stores bits by using capacitors and a transistor, that needs to be refreshed periodically. As SRAM does not have to be refreshed, it is much faster than DRAM but uses more parts resulting in a drastically lower memory per chip value. SRAM is also more expensive than DRAM.

Dynamic random-access memory (DRAM):

In current computer architectures, the memory is separated from the CPU and connected via a high-speed bus (I/O) called peripheral component interconnect express (PCIe). The memory supports error-correcting code (ECC). Today most systems feature commodity ICs in form of double data rate (DDR) synchronous dynamic random access memory (SDRAM) which is packaged in a dual inline memory module (DIMM). The DDR3 generation has been introduced in 2007 and is commonly used today. The next generation of DDR4 memory with 8GB capacity in an individual die entered the market in late 2014.

The DDR3 DIMM is a printed circuit board populated with a certain number of DRAM ICs (chips). Each DRAM IC contains arrays of individual bit storage locations. The DRAM IC had 4 data I/O signals or 8 data I/O signals. DDR3 memory is currently made up of 1 Gb (Gigabit), 2 Gb, and 4 Gb chips. One GigaByte (1GB) of memory is made up of eight 1 Gb chips. There are different types of DIMMs available with unique characteristics including Unbuffered with Error Correction Code Memory (UDIMMs), Registered Memory (RDIMMs), Load Reduced Memory (LRDIMMs), and Hyper Cloud Memory (HDIMMs).

The DDR3 DIMM board features a standard interface with 240 connector pins which is plugged into a corresponding slot (socket) on the motherboard. The connector pins socket contacts are gold-plated (various thickness). There is an on-going shift from gold bonding to copper bonding both due to technical feasibility and cost efficiency, which seems to be relevant also for the total environmental footprint of memory modules.

Chipset:

The CPU is connected with a corresponding chipset via the Direct Media Interface (DMI 2.0), four-lane PCIe 2.0-like link with 20 Gbps of bidirectional throughput. The chipset is a complex semiconductor device that functions as a Platform Controller Hub (PCH) for the interaction with the e.g. local area network (LAN), other network elements such as USB, the attached storage devices, and other media (see Figure 2).

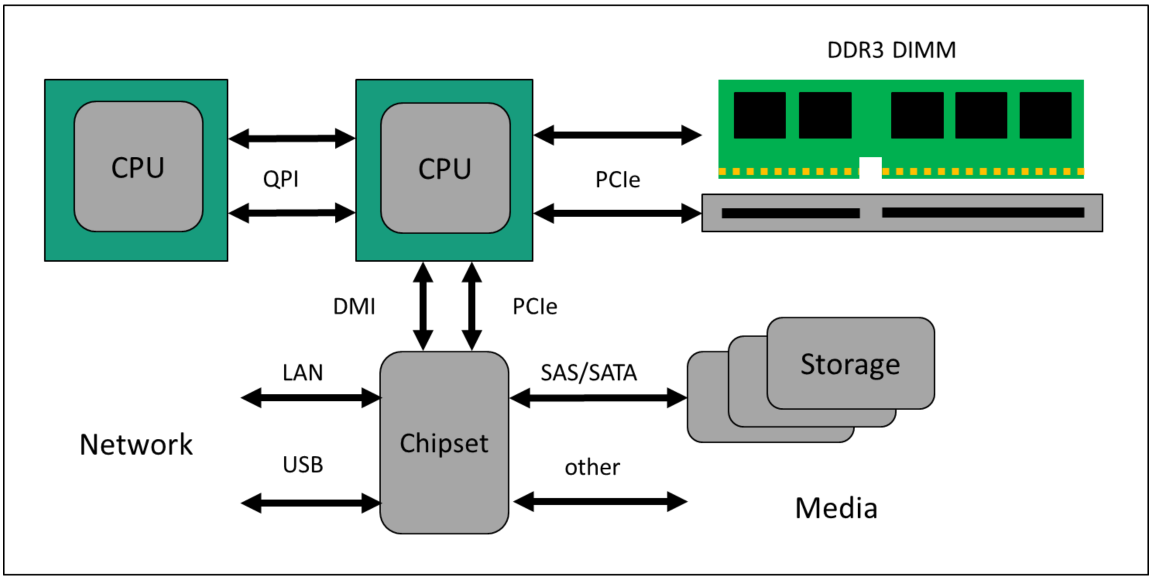


Figure 2 : Principle schemata of CPU, memory, chipset and interfaces

- **Material and energy consideration with respect to the manufacturing process:**

The integrated circuits (IC) of the CPU and memory are mainly made of polycrystalline silicon with high-κ gate dielectrics and fabricated in a technology called complementary metal oxide semiconductor (CMOS). The ICs are fabricated on large 300 mm silicon wafers. There is a distinction between Front-End-of-Line (FEOL) processes for creating the transistor structures and the Back-End-of-Line (BEOL) processes for creating the metal interconnections. The manufacturing process for creating the structured layers consists of four repeating process steps:

- Material deposition e.g. by physical or chemical vapour deposition, electrochemical deposition and growing crystalline layers by epitaxy;
- Removal of material e.g. by wet or dry etching, and chemical-mechanical planarization;
- Patterning the material by lithography; and
- Modification of electrical properties by doping processes.

These steps require clean room conditions, certain atmospheres, complex machinery, water and water treatment facilities as well as various chemicals for the individual processes. About one third of the energy consumption in the semiconductor manufacturing process is related to the machinery and the other two third to the clean room conditioning, utility infrastructure and water treatment. Modern semiconductor factories (wafer fabs) consider the improvement of energy and resource consumption as an important task. The clean room environments are therefore designed more modularly and the energy requirements on the installations and equipment are addressed in the procurement.

The exact material composition of the IC and the resource used in the manufacturing process depend on the individual properties of the chip. The die size and number of process steps (structured layers) indicate roughly the resource intensiveness of IC's manufacturing process. Latest chip generations feature up to 15 structured layers. There are only a few environmental assessments available for CPUs. However, recent environmental assessments of semiconductor manufacturing processes in general indicate that **the overall resource productivity is increasing in conjunction with technological progress**.[3] [4] [5]

After all structures of the die have been processed on the wafer, the die (chip) is diced for further packaging. The die is mounted on top of a small PCB (interposer) which has a Pin Grid Array (PGA) on the back-side (see Figure 3). The interposer nowadays features some electronic components for power control. The number of pins varies according to the CPU. Average pin count is about 1000 to 2000 pins. The pins are typically gold plated (0.4 µm or 0.8 µm). The pin size and thickness of the plated gold is getting smaller due to miniaturisation.[6]

After mounting the die on the interposer, the backside of the die is covered with a Thermal Interface Material (TIM) and connected to an Integrated Heat Spreader (IHS). This ensures that the thermal energy dissipated by the IC during operation is effectively conducted to the heat sink or heat pipe which is later mounted on top of the CPU. The quality of the TIM and HIS are very important for the reliable performance of the CPU due to the considerable thermal energy that has to be conducted away from the chip. The product datasheets provide the respective value. It is called the Thermal Design Power (TDP).

According to the datasheet for the LGA 2011 socket (for Intel Xeon E5) the housing, cap and dust cover is made of a high temperature thermoplastic (UL94V-0). The integrated load mechanism (ILM) is made of stainless steel and nickel-plated carbon steel. The backplate is also made of nickel-plated steel. The LM and backplate insulator is an adhesive coated polycarbonate film. The contact array of the socket for surface mount on the printed circuit board is a solder ball array. The contacts are made of copper alloy with gold plating (either 0.4 µm or 0.8 µm) over nickel plating. The solder balls are tin-silver-copper.[7]

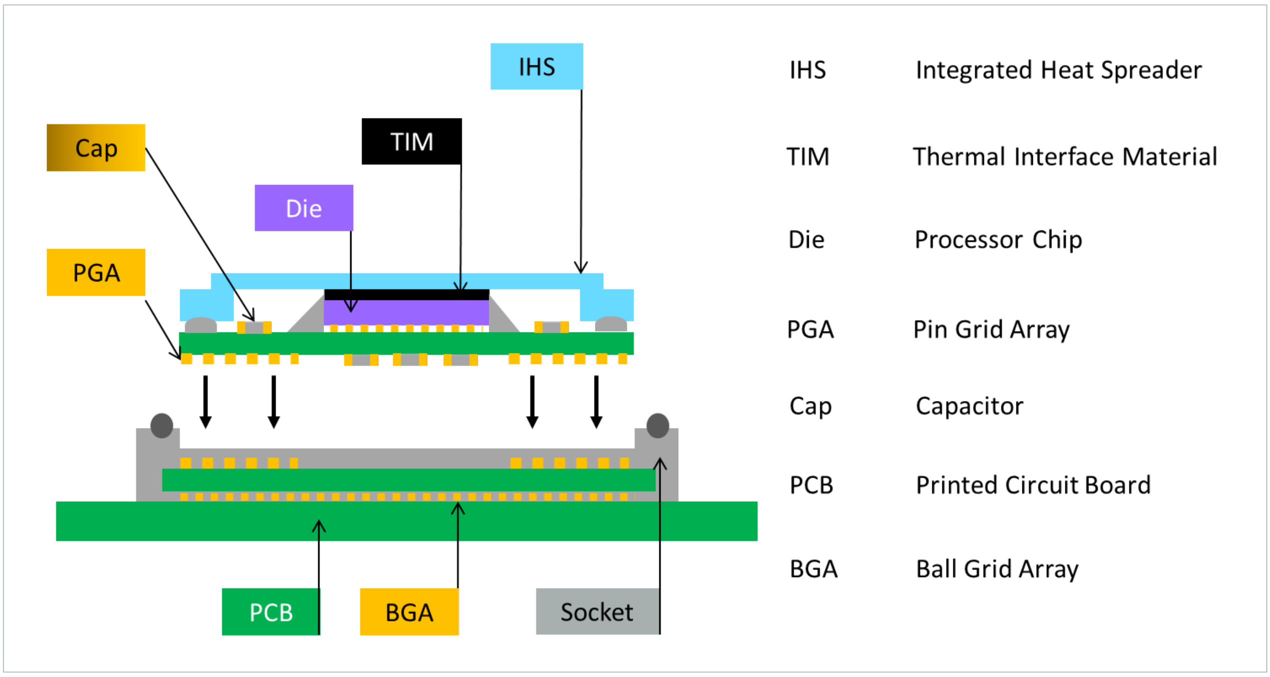


Figure 3 : Principle schemata - CPU package and socket

- **Energy considerations with respect to the use phase:**

The most power-drawing subassembly in a server is usually the processor in conjunction with the chipset and memory. Low energy consumption and better performance-to-power ratio is a high-priority objective

---

[3] Sarah B. Boyd, Life-Cycle Assessment of Semiconductors, Springer, New York, 2012.

[4] Russ, R.; Bipp, H.-P.; Jantschak, A.; Stiewe, M.; Cedzich, A.; Dietrich, M.: Abschlussbericht zur Selbstverpflichtung der Halbleiterhersteller mit Produktionsstätten in der Bundesrepublik Deutschland zur Reduzierung der Emissionen bestimmter fluorierter Gase, ZVEI - Zentralverband Elektrotechnik und Elektronikindustrie e.V., Fachverband Electronic Components and Systems, Frankfurt, November 2011.

[5] Higgs, T.; Yao, M.; Cullen, M.; Stewart, S.: Developing an Overall CO2 Footprint for Semiconductor Products, IEEE International Symposium on Sustainable Systems and Technology ISSST, May 2009.

[6] Intel Xeon E5 Datasheet: http://www.intel.com/content/dam/www/public/us/en/documents/datasheets/xeon-e5-1600-2600-vol-1-datasheet.pdf

[7] LGA 2011 Datasheet: http://www.farnell.com/datasheets/1651052.pdf

particularly in the design of new CPUs. Fortunately, the chip designer and semiconductor industry is achieving further improvements still continuously with each technology generation.

The specification of the thermal design power (TDP) is a practical indicator for the power draw of the chip. The TDP defines the maximum amount of heat (in W) which the CPU is generating in average operation. It is a specification for the design and proper dimensioning of the cooling system. For example, the maximum TDP value for the different Intel Xeon E5-2600 CPUs is ranging from 60 W to 150 W (see also Table 1). The TDP value is influenced by a number of factors including the operating frequency (clock speed), number of cores, cache and I/O configuration as well as integrated peripherals. According to chip manufacturers, the TDP of CPUs for volume servers will not increase over 150 W in the next few years. This limitation will help original equipment manufacturers (OEMs) in their midterm planning of new product designs. The TDP values for individual CPUs are easily found in product documentations online.

The power consumption of the memory in the use phase is influenced by many factors including the memory capacity and number of DIMMs, DIMM type (e.g. registered, load reduced, unbuffered), number of DRAMs per DIMM, DRAM technology, data transfer rates, operating frequency and voltage. The power consumption per GB memory varies according to DRAM technology and configuration (e.g. 4, 8, or 12 DRAM per DIMM), but also the operating voltage and frequency. Standard DDR3 DIMMS operate a 1.5 V, compared to 1.8 V for DDR2 DIMMs. There is a low voltage version available for DDR3 DIMMs that runs at 1.35 V. The resulting power consumption varies accordingly. There is only limited information about the precise power consumption of different memory devices available. However, some manufacturers provide comparative power consumption values for different configurations in their product data sheets. A comment by DIGITALEUROPE stated that there can be significant differences, up to 50%, in energy use and thermal dissipation for memory chips with the same GB capacity but made by different manufacturers.

For example, HP provides a guideline for configuring and using DDR3 memory with HP ProLiant Gen8. This extensive datasheet, published in 08/2012, provides complete power consumption and performance data for all possible product configurations.[8] For the purpose of illustration, some configurations are listed below:

- Three 8 GB LV RDIMMs per channel at 1.35 V (low voltage) vs. 1.5 V (standard voltage) operation in a 24 slot HP ProLiant Gen8
- Two 32 GB RDIMMs per channel at 1.35 V vs. 1.5 V operation

The respective power consumption values are shown in Table 2. For the purpose of better comparison, data from HP (power per module and power consumption per 1 GB) were added. The data indicate that the low voltage configuration is reducing the power consumption in operation by up to 20%. The idle power consumption of 1 GB is about 0.6 W and varies by only 150 mW. In full operation (loaded power), the power consumption increases by about factor 3 to 6 and reaches between 1.7 and 4.1 W depending on the configuration. This comparison indicates the considerable influence of product configurations with respect to power consumption and performance.

Table 2: Exemplary power consumption of memory according to HP

	3 x 8 GB LRDIMM per Channel at 1.5V	3 x 8 GB LRDIMM per Channel at 1.35V
Throughput (GB/s)	60.5	60.5
Idle power per module (W)	16.6	15.5
Loaded power per module (W)	98.6	78.8
Idle power per 1GB (W)	0.69	0.65
Loaded power per 1GB (W)	4.11	3.28

---

[8] http://h20565.www2.hp.com/portal/site/hpsc/template.BINARYPORTLET/public/kb/docDisplay/resource.process/?spf_p.tpst=kbDocDisplay_ws_BI&spf_p.rid_kbDocDisplay=docDisplayResURL&javax.portlet.begCacheTok=com.vignette.cachetoken&spf_p.rst_kbDocDisplay=wsrp-resourceState%3DdocId%253Demr_na-c03293145-2%257CdocLocale%253D&javax.portlet.endCacheTok=com.vignette.cachetoken

Table 3: Exemplary power consumption of memory

	2 x 32 GB LRDIMM per Channel at 1.5V	2 x 32 GB LRDIMM per Channel at 1.35V
Throughput (GB/s)	68.1	68.1
Idle power per module (W)	38.4	35.3
Loaded power per module (W)	139.4	110.8
Idle power per 1GB (W)	0.60	0.55
Loaded power per 1GB (W)	2.18	1.73

Resilient servers employ memory buffers between the processor and the memory DIMM to improve memory performance and access. They incur a significant power debt of 5 to 10 W per buffer with a single buffer typically supporting two DIMMs.

1.1.3.5. Passive and active cooling system

- **Technology:**

The cooling system of enterprise servers is most often a combination of passive and active cooling elements. A proper thermal management supported by a well dimensioned and designed cooling system is essential for a reliable operation of the server. The basic task is to transport the heat generated by the CPU and other active components away from the devices in order to ensure reliable function. The cooling system provides the needed cooling capacity which is specified by W/cm². The cooling capacity is the rate at which heat is removed from a certain space e.g. the surface of the CPU. The technology, design, and material characteristic of the cooling system is determined by various factors including:

- The number, type, and thermal design power of the CPU, memory and other electronic components of the server. The TDP of a CPU can reach up to 150 W and energy density due to higher system integration is increasing constantly. Semiconductor components transform most of the electrical energy into thermal energy but passive electronic components have thermal losses and require cooling as well. The technical properties of electronic components also vary according to production tolerances.
- The type, form factor, actual dimensions, and modularity of the server. There is a considerable difference in the thermal conditions of an integrated rack-server and of a modular blade-server system with shared resources and various population options. The thermal energy density, air volume, internal airflow and air pressure conditions are changing according to the actual population (configuration) of a modular server system.
- The external (ambient) temperature conditions, air intake design, dust filter mats, and other aspects of the rack design will influence the server cooling efficiency. The allowance for ambient temperature and humidity conditions in the room are increasingly more variable (ASHRAE allowances) and higher inlet temperatures up to 35°C are increasingly considered by operators in order to save energy with respect to the data centre cooling infrastructure (see Task 3 for more information).

There are two basic types of cooling system technology, and the combination of both also exist:

- Air cooling system (see Figure 4 : Air cooling system);
- Liquid cooling system; and
- Hybrid cooling system.

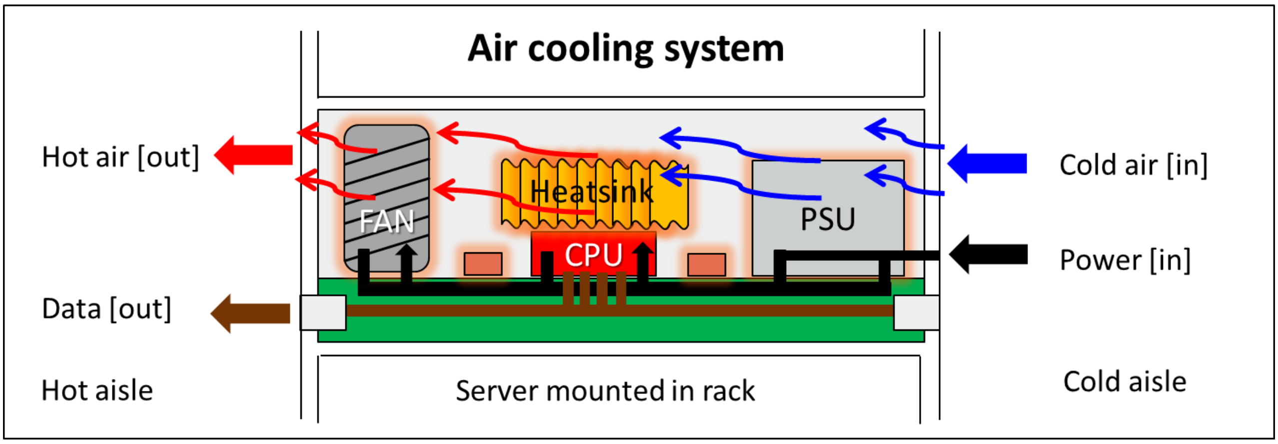


Figure 4 : Air cooling system

Passive cooling elements:

Passive cooling elements are radiators such as heat spreaders, heat sinks and heat pipes which are directly connected via a thermal interface material to the chip. They have the task of radiating away the thermal energy (heat) from the active components. This is basically achieved by the good thermal conductivity of the material that is used such as silver (only for thermal interface layer), copper, aluminium or iron alloys for the larger surface heat sinks and radiators. In operation, the cold air is sucked or pushed into the server enclosure and circulates around the plate fins of the heat sinks. Through the large surface area, the hot air is transported away by the air stream out of the system.

Heat pipes – which have a smaller form factor – are also used to efficiently transfer the generated heat away from the active components towards a separated heat sink or radiator. Heat pipes made of copper, aluminium or steal alloy, contain typically water and small amounts (<10%) of chemicals for effective heat transport. The selection of the fluid depends on the working point (e.g. temperatures, system architecture) and cost considerations. There are two work phases in the operation of a heat pipe. First is the evaporation phase where a fluid is transformed by the heat into a gas and moved to a colder radiator where a condensation occurs and the aggregation state changes back to a fluid. After the condensation, the capillary action is shifting the fluid back to the heat source thus creating an effective circuit.

Besides heat spreaders, heat sinks, and heat pipes, passive cooling elements also include design features that improve the air flow:

- air intake which features mostly a round shape or a more effective honeycombed shape,
- air baffles for channelling the airflow within the server enclosure and leading the cold air stream to all relevant electronic components that need cooling,
- air flaps or air valves at the rear of the chassis that avoid the air leakage when modules (e.g. PSU, Drives, Blades) are pulled out of the system

A general consideration for an effective air cooling system is low air pressure differences and reduction of cooling air volume. Changing air pressure requires additional energy which should be avoided in the design of the cooling system. Therefore, the air intake and air flow design including the placement of components (e.g. placing the PSU and Disk Drives away from the hot air stream, symmetric layout proposed by the open compute consortia) and their thermal properties of the components (e.g. the ambient operating temperature allowance of electrolytic capacitors) are important considerations.

Active cooling elements:

In an air cooling system, the air flow is created by fans inside of the server enclosure.[9] The fans suck in the cool air from outside (the cold aisle in the data centre) through air intake holes typically on the front side of the server. There are two basic types of air moving devices:

---

[9] Please note that there is a large variety of rack-cooling systems available in the market which interacts with the integrated server cooling system

- the centrifugal fan or blower, and
- the axial flow fan

In server architectures, axial flow fans are most common as they are designed to move big volumes of air at low resistance. Fans for server use brushless DC motors, which generate much less electromagnetic interference than other types. Common fan sizes include 40, 60, 80, 92, 120, and 140 mm, but there are larger ones available too.

Fans generally have two published specifications: the free air flow and maximum differential pressure. Free air flow is the amount of air a fan will move with zero back-pressure. Maximum differential pressure is the amount of pressure a fan can generate when completely blocked. To determine the efficiency of the fan is a complex task.

Adjusting the fan rotation speed according to the actual temperature conditions (e.g. the CPU) is already a standard technology. There are two options. The first introduced solution was a multi-step control of the fan rotation speed. A more advanced technology is a fan speed controlled linearly according to the temperature of the CPU. The noise level created by the fans rotation is a factor that needs consideration as well. Data centre operators demand low noise levels in order to allow technicians into the server rooms. The fan design including the form of the blades and housing is also supporting noise reduction and energy efficiency.

Liquid cooling technology:

The necessity for active liquid cooling technology derives from the increasingly higher energy density. The following factors influence this development:

- Highly integrated multi-processor architectures targeted for enterprise applications, virtualisation, and high performance computing (HPC);
- High core, high frequency CPUs with relatively high thermal design power up to 150 W;
- Small form factor of individual server, resulting in high number of servers within the rack, and the energy density in the rack is increasing over 25kW per m^3;
- Utilisation of thermal energy from the servers for secondary use (utility). Such secondary utilisation demands a temperature level of 70°C or higher;
- Higher ambient temperatures of 35 to 40°C.

Liquid cooling technologies are basically systems where the dissipated heat from the CPU is coupled via a heat spreader with enclosed water channels to a liquid cooling loop. The liquid cooling technology utilises basically water for heat fluxes up to 100 W/cm^2. In contrast to simple heat pipes, the liquid cooling system includes active components such as miniature pumps, compressors, and extensive pluming. The heat is transported through the pipes to a radiator (heat exchanger) in the server chassis or outside of the chassis. There are currently only a few examples of implemented liquid cooling systems, basically for HPC as still being very expensive solutions. The long-term reliability and possible aging effects of liquid cooling systems have to be considered.

Product level liquid cooling technologies are usually employed to cool CPUs and similar hot-spots like large memory. However, the cooling of the mainboards and other components needs consideration in the thermal design as well, so air cooling systems are additionally used for the remaining components.

- **Energy considerations:**

The thermal design and the utilised active and passive technologies are influencing the overall energy efficiency of the server. Because the active cooling systems such as fans in air cooling systems and pumps in liquid cooling systems are adding to the energy consumption, the appropriate selection of technology and proper design are of great importance. Regarding air cooling fans, the following aspects might influence the overall energy consumption:

- Number, dimensioning (air throughput characteristic), and positioning of the fans for optimal air flow. It is important to avoid an over- and under-provisioning.
- Load variable power consumption of the fan depending on the fans technology and design features. The optimal working point for conventional fans is quite often in partial loads of 50% to 60% of maximum load.

- Fan speed control and the capability to adjust according to the cooling demand of the heat source.
- Failure detection through condition monitoring has been implemented by some manufacturers

As mentioned already, the ambient temperature conditions and the set-up of the server for operation in higher temperature conditions (outside ASHRAE 2 envelop and therefore >27°C) will considerably influence the selection of components, passive and active thermal design. This means that the integrated active cooling system might be somewhat over-dimensioned in order to add some safety.

- **Material consideration:**

The passive cooling elements such as heat spreader, heat sinks and heat pipes are typically made out of aluminium and/or copper. The material selection and the design of the cooling elements are influenced by the specific thermal conductivity (W/m*K) and the related form factor and costs. Copper is a very good thermal conductor but comparatively expensive and heavy due to its density. Aluminium on the other hand has only 30% of copper's density and is therefore lightweight, but still features 59% of the thermal conductivity of copper. The designers therefore try to reduce the amount of copper in cooling systems. But because small form factor is an increasing requirement in the design of servers the utilisation of pure copper or copper-aluminium hybrids is still common. Industry stakeholders also indicated that iron alloys are used as base materials for heat sinks and heat pipes.

With respect to eco-design, a conscious concept for the selection of the passive and active cooling elements including its material composition is recommended. It seems preferable to have mono-material design. This definitely supports the end-of-life treatment and necessary separation of copper or aluminium in the recycling process.

1.1.3.6. Power supply unit and converters

- **Technology:**

The power supply unit (PSU) receives electrical power via the power distribution unit (PDU) from an AC (alternating current) source and converts this current into DC (direct current) and specific voltages (e.g. 12 V, 5 V, 3.3 V) with which the server operates. PSUs differ by power capacity, conversion efficiency rating, input and output power as well as redundancy, hot-swap capability and failure monitoring options. On the board and component level, this input voltage is further adjusted through DC/DC converters. The number of power conversions and the individual conversion efficiency contributes to the overall power consumption of the server.

In modern enterprise server, storage and network equipment, the typical type of power supplies used are switched-mode power supplies (SMPS) with power factor correction (PFC). They offer many advantages in comparison to conventional linear power supplies including:

- Higher power conversion efficiency and lower heat dissipation due to the demand-oriented, constant switching between different dissipation states. The series regulator element is either on or off and high dissipation transitions are tried to be avoided.
- Smaller form factor due to the use of integrated circuits for power switching and smaller transformers, inductors and capacitors for voltage storage. The reduced thermal losses help to reduce passive cooling elements and increase the compactness of the unit.
- There are some problems with transient spikes that occur from the switching action when filters are not sufficiently designed. The spikes can migrate into other areas and cause electromagnetic interference with other devices.

Further improvements in performance and miniaturisation are achieved by advanced semiconductor-based power electronics and highly integrated control circuitries. This reduces the overall heat dissipation, the number of electrolytic capacitors, board dimensions and housing. Over the past years the development of silicon-carbide (SiC) transistor technology for high efficient power conversion devices made considerable progress. The implementation of this state-of-the-art technology in power supply units has further improvement potential.

Previous improvements of power supply performance and conversion efficiency resulted from developments in power converter topologies, magnetic materials and semiconductors. Significant future improvements are expected to come from system architecture design, power management optimisation, refined packaging and thermal management techniques as well as advanced control algorithms. So the focus shifted from converter-level topology improvements to system-level architecture considerations and power management improvements in terms of load-activity.

- **Energy considerations:**

Enterprise servers including rack-optimised severs and blade servers feature multiple configuration options for their power supply. The manufacturer provides these multiple options in order to give the customer a wider selection. Typical power supply units have a capacity of 250W, 350W, 400W, 450W, 650W, 700W, 750W, 800W, 850W, 1000W, 1200W and 1500W. It is possible to use either one larger PSU or two smaller PSUs in one system.

Another aspect concerning the configuration is the redundancy configuration. Traditionally, servers with higher availability feature a 1+1 redundancy, meaning a rudimentary power supply unit is connected to compensate hardware failures of the other one. There are in general two possible operation configurations for redundant PSUs:

- On / Balanced: both PSU are active providing roughly the same power output (PSU-1 [55%] / PSU-2 [45%])
- On / Standby: PSU-1 is fully active and provides roughly most of the power output, PSU-2 is on standby with minimum power output (PSU-1 [97,5%] / PSU-2 [2,5%])

Concerning the pros and cons of both types of redundant PSU configuration, there is an interesting online forum that discusses them on website “serverfault.com”.[10]

The Climate Savers Computing Initiative already proposed the N+1 approach instead of the traditional 1+1 redundancy, to achieve a power supply optimised load range for network equipment[11]. This type of configuration is reducing the overall number of installed PSUs and demands on the other hand a more complex wiring. Servers with redundant PSUs feature hot-swap capability. Sometimes they also have flaps or blinds that cover the empty bay in order to avoid heat leakage from the server during operation.

The PSU’s rated output power and conversion efficiency in partial loads is influencing the baseline energy consumption of the server. Own calculation indicate that depending on the product type, configuration and load level, the PSU contributes at least up to 15% of the overall power consumption of a server. Due to the fact that larger servers, storage and network equipment consume up to a few kW, the highest conversion efficiency (even if only with 1 % difference) will result in significant energy savings. And because server utilisation is rather low on average, conversion efficiency in partial loads is important.

The conversion efficiency of PSU and particularly higher capacity PSUs has been improved in the past years to typical levels of over 85% and up to 90% at the 20% load point. Efficiencies at the 50% and 100% load points have been increased to the mid90th percentile for gold and higher power supplies. Figure 5 shows the **80plus certified** conversion efficiency factors in correlation to the utilisation level. The 80plus certification differentiates overall five levels of efficiency, including bronze, silver, gold, platinum, and titanium. A switch from silver or gold to platinum or titanium will result in a considerable improvement in energy efficiency.

---

[10] http://serverfault.com/questions/659452/how-do-servers-with-redundant-power-supplies-balance-consumption. 25.02.2015

[11] 2011, Climate Savers White Paper - Energy Efficiency Guide for Networking, http://www.thegreengrid.org/Global/Content/csci-white-papers/CSCIWhitePaperEnergyEfficiencyGuideForNetworkingDevices Retrieved september 12, 2014

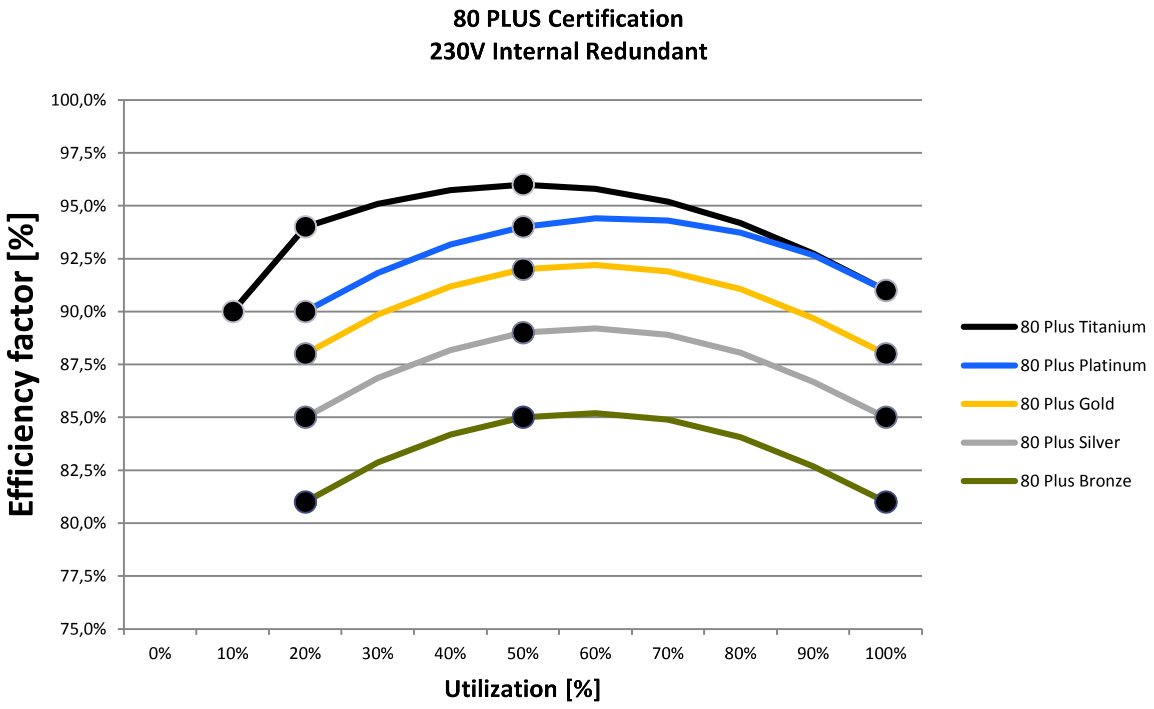


Figure 5 : Interpolated 80 PLUS certification criteria for power supplies (based on 80 PLUS Certification[12])

Equipment vendors in order to vary the product price typically offer different configuration options for PSUs, differentiating maximum output and conversion efficiency. A screening of the product offers of major vendors indicates that different levels of 80plus PSUs are commonly available. As a general trend it was observed that larger PSUs with a Wattage of 1000W and higher feature 80plus platinum and even titanium certificates. There are no data available concerning the market share of the individual level 80plus PSUs.

- **Material considerations**:

The main material contributions in terms of weight derive from the robust metal housing (steel alloy), the larger electronic components including the transformer and inductors (e.g. copper and ferrites), passive cooling elements (aluminium), the fan (plastic) and the printed circuit boards (glass, epoxy, plastics, solder).

- **Sales price:**

The sales price of a PSU is determined by a number of aspects including the number of output currents, connector configuration (pins), conversion efficiency (80plus certification), redundancy, hot plugging capability, and form factor. The quality and performance is influenced by the component selection and thermal design of the PSU as well. Table 4 shows current price ranges for single PSUs according to various internet sources.

Table 4: Current price ranges for single PSUs (Sources: Various websites)

Single PSU	250 - 400 W	450 – 800W	850 – 1500 W
80plus Gold	50 – 80 €	90 – 130 €	140 – 200 €
80plus Platinum	80 - 120 €	90 – 220 €	230 – 400 €
80plus Titanium		200 – 400 €	420 €

---

[12] 80 PLUS Certification: http://www.plugloadsolutions.com/80PlusPowerSupplies.aspx

1.1.3.7. Hardware configuration and form factor

- **Product spectrum and configuration options:**

On a physical level, enterprise servers can be distinguished by their form factor, modularity, performance specifications (e.g. resilience, nodes) and respective hardware and software configuration. The particular hardware and software configuration reflects the intended use or application, supported service levels and redundancy, the use environment and climatic conditions, interoperability and connectivity as well as expected lifetime and upgrade capability. Product vendors provide comprehensive data sheets with detailed information concerning following technical aspects:

- Type, dimensions, weight, and mounting options of the server chassis
- Mainboard type and dimensions
- CPU options including type, cores, threads, I/O, frequency, type and number of sockets, etc.
- Memory options including type, capacity, interfaces, number and type of memory sockets, etc.
- Internal storage (drives) including type, capacity, interfaces, number and storage bay space, etc.
- Connectivity (input/output) including network controller, connectors, expansion slots, etc.
- KVM (keyboard, video, mouse) interfaces, monitoring and service options, etc.
- Power requirements including rated power, power supply unit configuration, redundancy, etc.
- Power consumption values including maximum, sometimes idle or thermal design power of CPU (power-to-performance benchmarks e.g. SPEC are typically only offered on request)
- Active cooling elements such as fans and fan control
- Operational conditions and allowances including min and max temperatures, humidity in room
- Operation system and other software specifications including licenses

The market offers a tremendous variety of server products and even more configuration options. It is this large spectrum of configuration options that translate into the technical and environmental performance of the server as well as in respective purchasing costs (CAPEX) and cost of utilisation (OPEX). For the purpose of this study, it seems important to understand, that the variety of products and configuration options offered on the market are a positive development, because it allows selecting an (almost) optimum product/configuration for a very specific use or application. In the end, the single server is only a small element in a much larger system that provides the intended user benefit. And the user defines the application and conditions of the product utilisation. Against that background, the user needs transparency about the advantages and possible disadvantages of a server in order to make a conscious selection of the right product for a particular task. This selection obviously has an impact on the total cost of ownership and the environmental aspects. In the sections below, some main products and respective configurations are described on a highly aggregated level. For more information concerning the utilisation of servers and respective system aspects, please see the Task 3 Report.

Server design and form factor:

The type and size of chassis (housing, enclosure) is an important aspect (see section 1.1.3.2). Historically the market developed from two angles, which are now merging more and more. There are on the one hand small, stand-alone, PC-type servers (e.g. tower server) and on the other hand large mainframes.

Rack unit: Most servers – except for stand-alone products such as tower servers – are mounted in standardised 19-inch rack cabinets. The height of the server is specified as a rack unit (RU). One rack unit (1U) measures 1.75 Inch, two rack units (2U) 3.5 Inch, and so on. Half units are also distinguished. The dimension of a blade server system is often ten rack units (10U).

With respect to current enterprise server market we can basically distinguish between:

- Integrated single servers (e.g. rack-server)
- Modular server systems (e.g. blade system)

Integrated single servers: Over the past three decades, integrated rack-servers have been developed and they are currently representing the majority of enterprise server products sold in the market. Rack-servers integrate all functional elements of a server including the server-board populated with processor, memory and other active and passive electronics, passive and active cooling devices, power supply unit, interfaces and connectors in a single enclosure. The rack-server is inserted into the 19-Inch standard server cabinets. The individual servers need to be manually connected to the power distribution and network (cabling). The hardware and software configuration of rack-servers is optimised for a certain application spectrum and performance. Scalability of performance is limited to a few options including partial or full CPU configuration respective memory and storage capacity. The number of I/O ports is the real limitation. Scaling such server systems requires extensive cabling.

Figure 6 shows the schematic diagram of a rack-server.

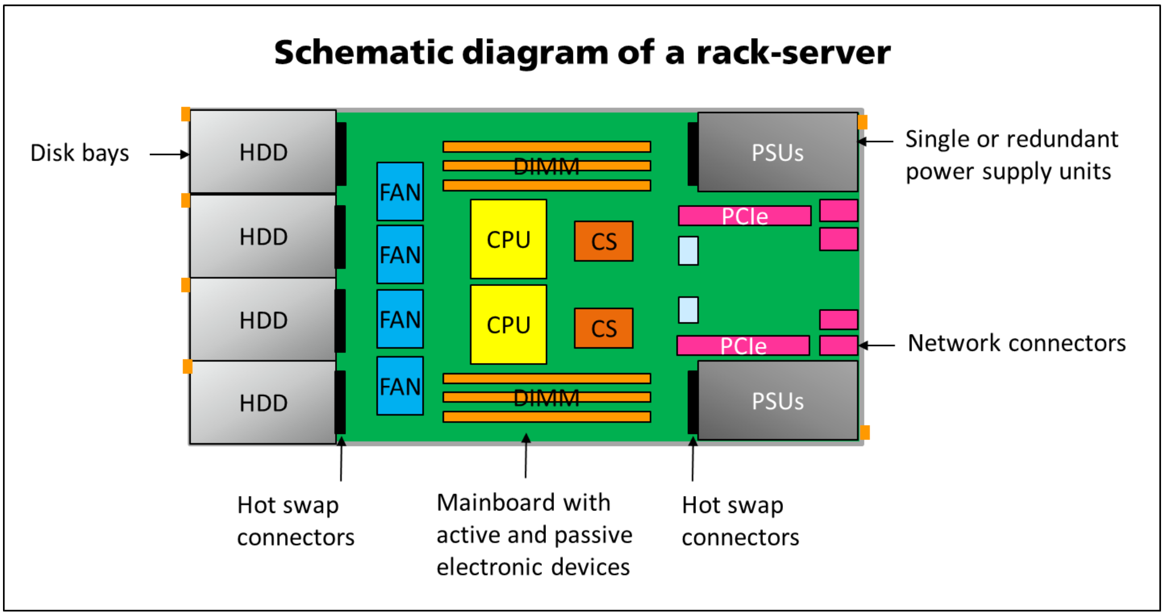


Figure 6 : Schematic diagram of a rack-server

Modular server systems: In order to achieve faster and better scalability for certain applications (e.g. web-services) and therefore to increase the density of the computing performance and reduce the complexity of cabling, the server designers developed modular server systems with partially shared functionality. Examples for this design approach are blade server systems and multi-node server systems. These server systems provide the capability to easily pack a number of full functional server nodes into a chassis without the necessity to manually connect them. In the system chassis they have access to so called “shared resources” including power supply, active cooling, storage capacity and most of all a full operational network mesh. Hardware components are directly mounted on “midplane board” and connected to the individual system boards (sever mainboards), network boards and storage media units. The main advantage of such modular server systems are:

- Reduction in LAN, NAS, and KVM cables by factor 5 to 10 (depending on the configuration)
- Reduction in wiring errors and breaking of cables (critical problem during installation)
- Increased reliability through power supply and cooling redundancy (shared resources)
- Fast and easy exchange and scaling of computing capacity (blades units)
- Smaller overall dimensions and weight reduction

Typical blade systems are populated with 12 to 18 server blades. Current multi-node servers are a mix between a larger blade system and a rack-optimised system. The disadvantage of modular server systems is the considerable hardware overhead and respective energy consumption in the case that the system is not fully populated. Server management software is addressing this problem by actively monitoring and controlling the shared resources. This allows shutting-off certain hardware elements when they are not required for use.

Figure 7 shows the schematic diagram of a blade-server system.

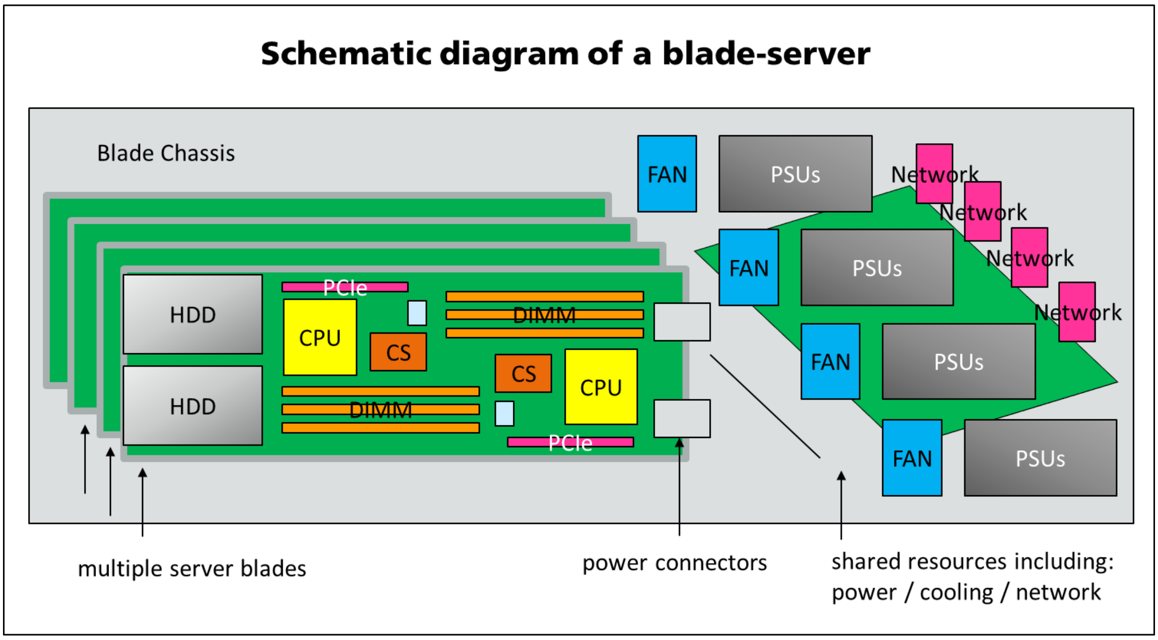


Figure 7 : Schematic diagram of a blade-server

The industry is currently developing also so-called **micro-servers** (see also Task 1 for the definition). Micro-servers are especially designed for highly scalable cloud applications for example front-end webservers. These products feature smaller system-on-chip processors such as 64-bit ARM Cortex-A50 series or 64-bit and 86-bit compatible Intel Atom S1200. The new products are expected to draw less than 10 Watts in operation. A micro-server system will pack several hundreds of micro-servers into one server rack.

As a trend, the server design is addressing very densely packed systems with a much reduced form factor (weight and volume). This trend is supported by the:

- Continuous miniaturisation of electronic components and functional large scale integration (LSI),
- Development of small and highly efficient semiconductor-based power electronics,
- Increase of network bandwidth and speed through advanced photonics,
- Improved thermal properties of hardware components and electronic packaging technologies allowing for higher operating temperatures.
- Sensors and integrated software for managing the available resources according to the actual demand (load)

Server platforms: The computing performance of enterprise servers is mainly influenced by the number and type of CPUs that are employed. Enterprise servers feature 1 or 2 CPU-sockets (entry/service platform), 2 or 4 CPU-socket (efficient/mainstream platform), and 4 or more CPU-sockets (mission critical/performance platform). The market is currently dominated by 2-socket servers. These products have a 73 % market share in terms of unit sales (see Task 2). The second largest market share belongs to 1-socket servers (23%). Servers with four or more CPU-sockets on the mainboard are typically mission critical servers (resilient). They have a small market share of less than 4 %.

Product examples: Table 5 shows some examples of typical enterprise server products differentiated by form factor including a stand-alone tower server, a rack-optimised single server, a blade server system with multiple server blades, and a multi-node server system. The product examples are for illustration purpose only, all taken from the product portfolio of Fujitsu. The provided technical data are rudimentary. The actual product data sheet is of course more extensive.

Table 5: Enterprise servers – typical form factor and product examples

Form Factor	Photo	Comments
Tower		**FUJITSU PRIMERGY TX120** Entry level server platform for offices with high availability Single processor CPU: Intel® Xeon® processor E3-1200v2 family Storage: 2 x 3.5” HDD or 4 x 2.5” HDD PSU: 1 x 250 Watt Weight: up to 10 kg
Rack-optimized (2U)		**FUJITSU PRIMERGY RX2520 M1** Enterprise mainstream platform with scalable storage Dual processor CPU: Intel® Xeon® processor E5-2400 v2 product family Storage: 8 x 3.5” HDD or 16 x 2.5” HDD PSU: 1 x 800 Watt or 2 x 450 Watt Weight: up to 25kg
Blade System (10U)		**FUJITSU PRIMERGY BX900 S2** Enterprise mainstream platform with maximal 18 blades in a 10U chassis. Dual processor CPU: Intel® Xeon® processor E5-2400v2 family (per blade) Storage: 9 x 3.5” HDD or 18 x 2.5” HDD PSU: 3 x 3200 Watt to 6 x 1600 Watt Weight: up to 191kg
Multi-node (2U)		**Fujitsu PRIMERGY CX400 S1** Multi-node server system with 4 nodes (server modules) in a 2U chassis. Dual processor CPU: Intel® Xeon® processor E5-2600 family (per node) Storage: 12 x 3.5”HDD or 24 x 2.5” HDD PSU: 2 x 1400 Watt Weight: up to 38kg

1.1.4. Enterprise storage

1.1.4.1. Storage devices and system elements

Enterprise storage systems including larger data centre deployments are providing non-volatile data storage services to connected servers (host) and/or to remote computing devices (clients) via network connections. The data storage system supplements the server’s internal memory, by providing more capacity, redundancy and flexible data management. The storage system hardware consists basically of a larger number of storage media, a controller that handles the input and output (IO) requests, and the necessary connectivity for data transmission. All these devices are mounted in a chassis together with power supply units, active cooling devices (fans) and management interfaces. The storage system also requires software for the controller (management of the workload) and administration (see Task 3). The storage system is connected to the servers or clients by means of direct connections or network connections. This includes direct attached storage (DAS), storage area networks (SAN) and network attached storage (NAS). The trend goes to shared and easily manageable storage resource (virtualisation) provided through fast and lossless links (network). The main performance feature of storage systems are data protection, capacity, latency, connectivity and manageability.

Data storage systems include following subsystems:

- **Storage media and devices**: These include hard disk drives (HDD), solid state devices (SSD)[13], tape cartridges, and optical disks providing non-volatile data.
- **Storage controller**: External or internal subassembly including a processor (sequencer) and other electronics which autonomously process a substantial portion of IO requests directed to storage devices.
- **Storage elements:** Product configuration (e.g. controller enclosure or disk enclosure) such as redundant array of independent disks (RAID) or a robotic tape library with a number of storage devices and integrated storage controller for handling I/O requests.
- **Connectivity and network elements**: Storage devices can be directly connected to a host or connected through a network. Connectivity and networks are based on various technologies (protocols) including Serial Advanced Technology Attachment (SATA), Serial Attached SCSI (SAS), Fibre Channel (FC), Infiniband (IB) or Ethernet (TCP/IP).
- **Connectors and cables:** The data transmission between servers and storage devices requires interface controllers (integrated or on separated cards), connectors and cables. The functionality, performance and form factor are important aspects.

1.1.4.2. Hard disk drives (HDD)

Hard disk drives (HDD) are the most commonly used volume storage media for high capacity block storage. This is a mature technology. Data are stored on fast rotating disks (platters) by magnetic remanence. HDDs are more cost efficient in comparison to semiconductor-based solid state devices (SSD). Two basic types of HDDs are used in enterprise (business) applications: traditional 3.5-Inch HDD for very high capacity; and more recent 2.5-Inch HDDs for faster access but smaller capacity.[14] Both types of HDDs feature considerable differences in terms of technology and performance.

Performance specifications of HDDs include the storage capacity, average latency, type of connectivity or interface, the maximum sustained transfer rate, spin speed, aerial density, cache, encryption, format, maximum operating shock, form factors, average failure rate, operating temperature, and average operating power[15]. Storage capacity, latency, and data transfer rates are the important characteristics that make HDD the preferred technology for low to high end “online” storage systems with latency below 80 ms.

The mean time between failure (MTBF) ratings differentiates HDDs for various applications such as in MAID or RAID storage systems. Failures are typically related to vibration and thermal cycling causing galling of the spindle’s swivel. Operating temperature for HDD is typically 5°C to 40°C, but current product data sheets indicate even higher operating temperatures of up to 60°C.

The workload is determined by the amount of time the HDD is in active mode, meaning the HDD is seeking, reading or writing data as well as the time in idle (spinning). The main performance indicator of a HDD is the Input Output Per Second (IOPS) value that aggregates the average rotational latency and read/write seek latency: IOPS = 1000/(Seek Latency + Rotational Latency) (for more information on storage utilisation, workload aspects and performance, see Task 3).

General **hardware elements and material aspects** of a HDD are:

- Platters (disks) as storage media: the base material of 3.5-Inch and 2.5-Inch HDDs are considerably different (see below)
- Electromechanical read-and-write-system including a servo motor, actuator arm, and magnetic head (recording transducers)
- ICs and electronics on a printed circuit board (ICs include e.g. storage controller, interface controller, and buffer memory)
- Network and power connectors
- Aluminium cast housing (typically more than 90% of total weight)

---

[13] SSD come in the same housing or form factor such as regular HDDs

[14] 1.8-Inch HDDs are not considered in enterprise applications.

[15] http://www.seagate.com/gb/en/internal-hard-drives/enterprise-hard-drives/hdd/enterprise-capacity-3-5-hdd/#specs

Figure 8 shows the principle design of a hard disk drive in a schematic diagram.

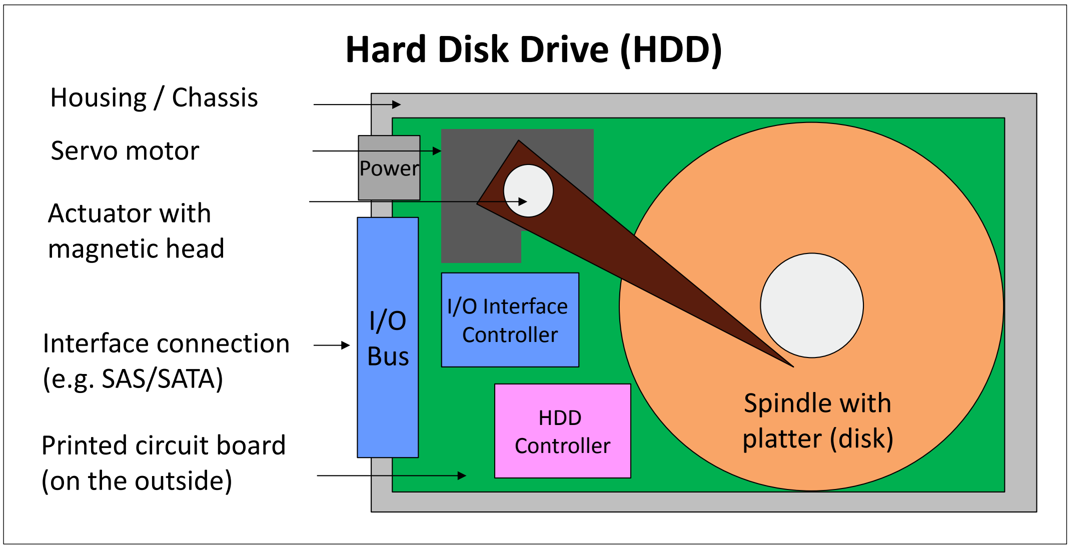


Figure 8 : Schematic diagram of a HDD

3.5-Inch HDDs feature an aluminium platter with multiple magnetic layers that realise data storage. According to industry sources, the platter's base material (aluminium) accounts for more than 90% of overall material weight. Nickel and magnesium are other materials representing between 3 to 4% of total weight.

The magnetic layers of the platters (storage media) are made of a larger spectrum of materials in low concentrations. The soft magnetic (keeper) layer is made from an alloy that contains iron, cobalt, and nickel. On top of this under-layer come two ferromagnetic layers which are separated by very thin layer of ruthenium. The data-storage (recording) layers that follow on top are made from an alloy of cobalt, chromium and platinum (CoCrPt) and these layers are separated again by a thin layer of ruthenium. Whereas the ruthenium functions as buffer layer and is known as synthetic anti-ferro-magnet, platinum provides thermal stability, preventing data loss if the disk is subjected to external magnetic fields or heat. This material is essential for required technical properties of high capacity HDD. The total weight of a single aluminium platter is about 25 grams. Regarding the environmental impacts of the production process it is very interesting to notice that the used materials and manufacturing processes are not the most energy consuming aspect overall, but according to industry it is the final testing of the individual cells of the HDDs. That means that the environmental impact of the manufacturing process is determined by the total storage capacity. This insight derives from the European project "LCA-to-Go" which developed carbon footprint reference values for various electronic products including HDDs.[16]

The storage capacity is high and current products have a capacity from 0.5 Terabytes (TB) up to 6.0 TB. This capacity is provided by a total number of one to seven platters[17]. Typical spin speed, which indicates performance, is 7 200 rounds per minute (rpm). The typical cache size in this segment is either 64 or 128 MB. Table 6 presents some technical data on 3.5-inch HDDs.

---

[16] The LCA-to-go project is lead by Fraunhofer IZM, http://www.lca2go.eu/

[17] For example see http://www.storagereview.com/hgst_6tb_ultrastar_he6_hdd_now_shipping

Table 6: 3.5" HDD - Technical Data

Technical parameter	Specification[18]	Comments
Capacity	0.5 to 6 TB	Average: larger capacity
Spin speed	7 200 rpm / 10 000 rpm	Average: 7 200 rpm
Interface	SAS/SATA 1.5-12 Gbps	Average: 6 - 12 Gbps
Weight	605 to 780 grams	Depends on # platter
Depth / Width / Height [mm]	147.00 / 101.85 / 26.11	Depends on # platter
Platter	1 to 5 platter, disk material: aluminium	Best available technology: 6 - 7
Aerial Density	324 to 643 Gbit/in²	

2.5-Inch HDDs feature not only a smaller form factor but also a different platter base material. The platter substrate is glass on which the magnetic layers are build-up. This base material of silica and crystalline silica accounts for more than 90% of total weight. The magnetic layers are build-up in a similar way like the larger one. The total weight of a single glass platter is about 5 grams.

The storage capacity is lower in comparison to 3.5-inch devices. Current products have a capacity from 0.2 Terabytes (TB) up to 1.0 TB. This capacity is provided by total number of one to three platters. Typical rotation speed is 7 200, 10 000 or even 15 000 rpm. Technical data on 2.5 inch HDDs can be found in Table 7: 2.5" HDD - Technical Data.

Table 7: 2.5" HDD - Technical Data

Technical parameter	Specification[19]	Comments
Capacity	0.25 to 1 TB	Average: larger capacity
Spin speed	7 200 rpm / 10 000 rpm / 15 000 rpm	Average: 7 200 – 10 000 rpm
Interface	SAS/SATA 1.5 to 12 Gbps	Average: 6 - 12 Gbps
Weight	~ 179 to 200 grams	Depends on # platter
Depth / Width / Height [mm]	100.45 / 70.10 / 15.00	Depends on # platter
Platter	1 to 3 platter, disk material: glass	
Aerial Density	417 Gbit/in²	

Power consumption of HDDs: Table 8 below shows specifications and respective average power consumption for the current product portfolio of one of the major storage manufacturers Seagate. According to these data, the average power consumption of current business storage 3.5 inch HDDs is about 9.8 W while the average power consumption per device roughly scales with its capacity. In the case of the 2.5 inch HDD, the average power consumption under active load is 5.3 W. The idle power consumption has not been specified by Seagate.

---

[18] Example data: http://www.seagate.com/gb/en/internal-hard-drives/enterprise-hard-drives/hdd/enterprise-capacity-3-5-hdd/#specs

[19] Example data: http://www.seagate.com/gb/en/internal-hard-drives/enterprise-hard-drives/hdd/constellation/#specs

Table 8: Exemplary enterprise HDD power consumption (Seagate)

Form factor	Interface	Storage Capacity (TB)	Average Power Consumption (W)	Power to Capacity Ratio (W/TB)	Comments
3.5" HDD enterprise storage	SATA	0.50	-		6 Gb/s, 7 200 RPM
		1.00	7.7	7.7	
		2.00	9.6	4.8	
		3.00	10.9	3.6	
		4.00	10.3	2.6	
		5.00	10.5	2.1	
		6.00	11.1	1.9	
	SAS	0.50	6.8	13.5	6-12 Gb/s, 7 200 RPM
		1.00	8.5	8.5	
		2.00	10.5	5.2	
		3.00	11.0	3.7	
		4.00	11.0	2.7	
		5.00	-		
		6.00	11.3	1.9	
	SAS + SATA aggregated	0.50	6.8	13.5	Total average: 9.8 W
		1.00	8.1	8.1	
		2.00	10.1	5.0	
		3.00	11.0	3.7	
		4.00	10.7	2.7	
		5.00	10.5	2.1	
		6.00	11.2	1.9	
2.5" HDD enterprise storage	SATA	0.25	4.6	18.4	7 200 RPM 10 000 RPM
		0.50	5.1	10.2	
		1.00	5.9	5.9	

1.1.4.3. Use of neodymium and other rare earth elements in HDD

Hard Disk Drives (HDDs) contain several permanent magnets inside the chassis. The most dominant magnets are the permanent magnets of the voice coil motor (VCM). These are usually 2 identical kidney-shaped magnets attached to a metal plate each. Some 3.5 inch and most 2.5 inch HDDs only contained one magnet. Further magnets are built into the read/write head and the spindle motor (Zepf 2013).

In a current study, 195 HDDs from 1990 onwards were collected, 147 HDDs thereof have been disassembled and 130 samples were usable. Here a clear trend was obvious, that the weight to storage capacity ratio shrinks with time. The magnets of the 3.5 inch HDDs showed weights ranging from 2.5 g for to 76.6 g. A general trend for 3.5 inch drive magnet weight could be demonstrated with a slight decrease in magnet weight over time. The average weight of the magnets of the 3.5 inch HDDs is 16.3 g; covering all the years, disregarding any specialties, like SCSI or storage capacities. An estimation of 2.5 g per 2.5 inch HDD could serve as a reasonable value for the intent of this analysis (Zepf 2013).

The samples that were analysed by RFA showed an average content of 25 % Nd, 5.3 % Pr and 66 % Fe. Measurements of separated magnets from hard disk drives delivered a neodymium concentration in the magnets of 23–25 % and concentrations of praseodymium, dysprosium and terbium between 0.01% and 4 % (Rotter et al. 2013).

1.1.4.4. Solid state devices (SSD)

Solid State Devices (SSD) are semiconductor-based, non-volatile flash memory devices that are packaged as single-chip or multi-chip module on a printed circuit board. SSDs are not defined as a drive so they do not have to meet the drive standard specifications of a HDD. The SSD includes a memory controller that runs on a firmware (own operation system), an I/O interface controller and respective connector (bus). SSDs are offered in different form factors including PCIe cards and HDD standard housing.

SSDs are using NAND flash technology, which offers relative high storage capacity, high density, fast sequential reads, writes, and erases. Because there are no moving parts (like in the HDD), the access time is much faster, the power consumption lower and an application under slightly higher operating temperatures is possible. A disadvantage is the higher price. SSDs are considerably more costly in comparison to HDDs, due to the semiconductor ICs (flash memory chips), which contribute to about 80% of total costs. Production capacity is still limited and dictates the price.

SSD NAND memory capacity is provided similarly to block devices (HDD). However, each block consists of a number of pages with a defined storage capacity. A few bytes are associated with each page that can be used for storage of an error correction code (ECC). A qualitative distinction is the number of bits saved in a cell. There are single-level cell (SLC) devices which feature 1 bit per cell and multi-level cells (MLC) which feature 2 or 4 bits per cell.

The SSD logic component is the SSD controller. This microprocessor executes a firmware code that controls the I/O requests for reading, writing, erasing, error check, encryption, and additional features to improve the performance of the storage devices. This performance enhancement (based on the individual firmware) includes write-amplification algorithms, error-avoidance and system-level error management beyond the standard error correction code.

Every SSD provides an I/O interface including the I/O controller and I/O connector. The I/O controller is a microprocessor that creates the link between the storage devices (SSD) and the server (host). There are various interface options available. Most common enterprise SSDs are supporting Serial ATA (SATA) and Serial Attached SCSI (SAS).[20]

According to product data sheets, the operating temperature of SSD is between 0°C to 70°C and therefore considerably wider in comparison to HDDs.

Figure 9 shows the main components of an SSD in a schematic diagram.

---

[20] SSDs come in other applications as USB devices or PCIe cards as well.

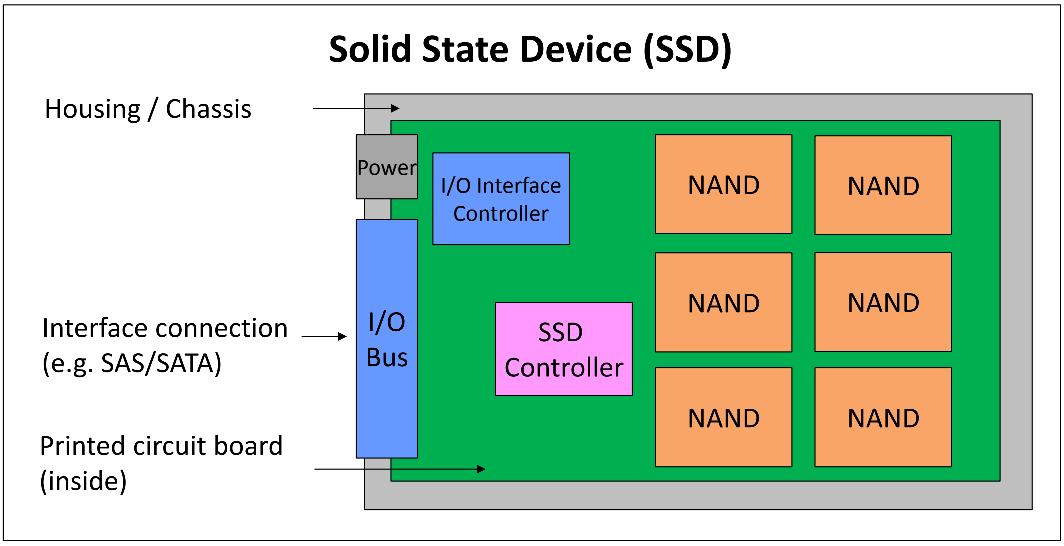


Figure 9 : Schematic diagram of a SSD

- **Material considerations:**

SSD consist of a certain number of semiconductor ICs. They are manufactured under clean room conditions and pass through a high number of process steps (see CPU manufacturing). In comparison to HDD, the SSD production is considerably more energy intensive. For orientation and as approximation, Samsung published an environmental product declaration for a 16 GB memory module of 64 kg CO_2 eq. (cradle to gate). In comparison, the carbon footprint of HDDs is about 10 kg CO_2 eq. (and for much higher capacities of 250 or 500 GB typically).

- **Energy considerations:**

The device level power consumption of enterprise SSDs is varying considerably between manufacturers and type of SSD. The power consumption of enterprise is also not directly comparable to the power consumption of consumer SSDs. According to tests conducted by the known online forum “Tom’s Hardware” consumer SSDs feature considerably low power consumption of about 0.3 to 1.0 watts in idle, 1.0 to 2.0 watts in average active, and 2.0 to 6.0 watts in maximum load.[21] In comparison the power consumption of enterprise SSDs is considerably higher, particularly due to the interface (I/O capability) and control features.

The active load power consumption of enterprise SSDs for larger capacity SSDs (>400 GB storage capacity) is between 3.0 and 12.0 watts and the idle power between 2.0 and 6.0 watts. Exemplary power consumption data for enterprise SSD are compiled in Table 9.[22] As trend the power consumption is further dropping and it seems justified to assume the today an average 800 GB enterprise SSD consumes about 3.0 to 6.0 watts in average active and 1.0 to 2.0 watts in idle.

Most enterprise SSDs provide a device initiated power management (DIPM) that tunes down the interface ports when no I/O activities occur. Energy savings of DIPM are in the area of mW and therefore not considered to be a major factor regarding energy saving potential.

Table 9-Table 11 below show the exemplary enterprise SSD power consumption for different capacities.

---

[21] http://www.tomshardware.com/reviews/crucial-m550-ssd-review,3772-11.html

[22] http://www.enterprisestorageforum.com/storage-hardware/ssd-vs.-hdd-pricing-seven-myths-that-need-correcting-2.html

Table 9: Exemplary enterprise SSD power consumption, 200 GB capacity

Enterprise SSD 200GB				
Manufacturer	Model	Capacity (GB)	Active Load (W)	Idle Load (W)
Toshiba	MK2001GRZB	200	-	6.5
Toshiba	PX02SSF020	200	-	2.7
Hitachi	HUSMM8020ASS200	200	9.00	2.1
Hitachi	HUSMH8020ASS200	200	9.00	2.1
Seagate	ST200FM0073	200	3.93	2.89
Seagate	ST200FX0002	200	4.97	4.04

Table 10: Exemplary enterprise SSD power consumption, 400-800 GB capacity

Enterprise SSD 400-800GB				
Manufacturer	Model	Capacity (GB)	Active Load (W)	Idle Load (W)
Toshiba	MK4001GRZB	400	-	6.5
Toshiba	PX02SSF040	400	-	2.7
Hitachi	HUSMM8040ASS200	400	9.00	2.1
Hitachi	HUSMR1050ASS200	500	9.00	2.1
Seagate	ST400FM0073	400	3.71	2.72
Seagate	ST400FX0002	400	6.67	5.92

Table 11: Exemplary enterprise SSD power consumption, more than 800 GB capacity

Enterprise SSD >800GB				
Manufacturer	Model	Capacity (GB)	Active Load (W)	Idle Load (W)
Toshiba	PX02SSF080	800	-	3.60
Hitachi	HUSMM8080ASS200	800	11.00	2.20
Seagate	ST800FM0053	800	4.05	3.00
Samsung	SM1625	800	4.00	-
Intel	910 series	800	8.00	-
Hitachi	HUSMR1010ASS200	1000	11.00	2.20
Super Talent	RAID Drive II plus	1024	3.00	-
Toshiba	PX02SMB160	1600	-	2.70
Micron Real	P420m HHHL	1400	30.00	-
Hitachi	S846 Encryption SAS SSD	1600	12.00	-

1.1.4.5.Redundant array of independent disks (RAID)

Redundant Array of Independent Disks (RAID) defines generally a storage system in which a number of storage devices (e.g. HDD, SSD) are virtually combined into one logic storage unit. There are two main objectives for combining multiple storage devices. One objective is to improve performance and mainly to increase the storage capacity and operational speed of the system. A second objective – and in an enterprise and data centre context more dominant objective – is redundancy and therefore maintaining high availability in case of a device failure. A third objective would be a combination of both. Please note that a RAID-system does not substitute a data back-up system.

A RAID system basically consists of a central RAID Manager and RAID controllers. Due to the different objectives there are different RAID-system implementations and different hardware and software technologies available to achieve them. The most common RAID-systems are RAID Level 0, 1, 5, and 6. The RAID level basically defines the trade-off between performance and reliability – they provide performance improvements and redundancy on different economical levels. The most commonly utilised RAID systems and their performance characteristics are summarised in Table 12.

Table 12: RAID system comparison

Performance	RAID 0	RAID 1	RAID 5	RAID 6
Redundancy	None (stripping)	1+1 (mirroring)	N+1 (striping with distributed parity)	N+2 (striping with double parity)
Configuration min. (e.g. HDD, SSD)	2	2	3	4
Performance (e.g. transfer speed)	High (read and write)	High (only read)	Medium (depends on controller)	Low (due to compute overhead)
Data processing overhead	Low	Low	Medium	High
Costs	Low	High (number of devices)	Medium	High (controller)

The RAID 0 and 1 are providing very good data transfer rates. With respect to RAID 5 and 6, the writing speed goes down, because the RAID controller calculates special checksums for the data and has to distribute the blocks with the checksum and those with the data over all disks. More redundancy is given as well by mirroring and parity.

- **Why is it necessary to deploy redundant storage systems?**

HDD failures are the most common failures in data centres (see Task 3 for details). One reason is the electromechanical technology of the HDD. Failures typically occur due to the following reasons:

- Vibration: When multiple disk drives are working in close contact, the rotating platters create vibrations which are transferred through the RAID-system and impact the actuators. Read and write errors can occur.
- Thermal stress: In conjunction with the servers' heat dissipation and own thermal losses of the RAID, a certain thermal stress is induced into the system. Due to the different coefficients of thermal expansion of the individual materials, this can influence the precision of the actuator, the spindle and microelectronic.
- Others: This includes e.g. seek errors, scan errors, reallocation and probational counts.

It is noteworthy that the operating temperature of SSDs is with up to 70°C considerably higher than the one of HDDs with only 55°C. The cooling requirements of HDDs have to be considered when higher operation temperatures e.g. >27°C on the inlet (ASHRAE 3) are considered.

RAID-system HDDs are running 24/7 and are therefore often of better quality. This means that they have longer mean time between failure (MTBF) ratings. The cost of such device is higher.

On a technological level one can differentiate three different RAID solutions:

- Hardware-RAID: A dedicated microprocessor with firmware functions as a RAID-controller and handles I/O requests, calculates checksums, and distributes the data. Through this, the CPU utilisation of the server (host) is considerably reduced. The RAID-controller comes on an adapter-card in smaller configuration or as an own device in large configurations.
- Software-RAID: In this case, the CPU of the host is calculating all checksums and handling of data. The various operation systems support different RAID levels. The CPU utilisation is considerably increasing.

- Host-RAID: This is a hybrid of hardware- and software-based RAID. The host requires a special RAID-chipset on the server-board. Some RAID functions are supported by the RAID-chipset and the more complex calculations are provided by the CPU. The RAID-chipsets are therefore not providing the full spectrum of functionality like a RAID-controller.

- **Environmental considerations:**

RAID systems have certain energy and resource impacts. The environmental impacts are basically related to the issue of redundancy and data processing efficiency. For instance, in terms of resource efficiency, RAID 1 solutions lead to a high number of necessary storage devices. In terms of energy consumption, a hardware-RAID with a specific RAID-controller is faster and possibly more energy efficient.

1.1.4.6. Massive array of idle disks (MAID)

Massive Array of Idle Disks (MAID) is a non-RAID storage architecture in which only those disk drives in active use are spinning at any given time. MAID is a technology in support of less time-critical, near-online storage systems and which requires only occasionally read- and write-processes. MAID systems can consist of a very high number of cost-efficient (lower MTBF ratings), SATA connected storage devices (see section 1.1.5.1). MAID solutions are offered as a replacement option to high-volume tape libraries. Their main advantages are low costs, reduced power consumption and cooling, good lifetime due to prolonged idle phases.

1.1.4.7. Storage system form factor and configuration

Storage products are highly complex combinations of a controller and storage drives, both HDD and SSD. Like servers, there are large numbers of possible configurations whose selection for a given customer will be highly dependent on their specific use case(s) and workload profile(s). Storage product performance and efficiency is highly dependent on the application of software based functionality including compression, de-duplication, software defined storage management systems, and workload tiering. The use of these software functions enables a single storage product or group of storage products (software defined storage) to substantially reduce the number of storage systems required to manage a specific set of workloads. On an individual system it may increase the power consumption resulting in a lower performance power score as measured by the SNIA Emerald test, but it reduces the overall power consumption required to perform a given workload because a smaller number of storage systems are required.

RAID-enabled storage systems are roughly standardised towards 19-inch rack form factors. The controller can either be inside the same enclosure as the disks or come in an extra controller enclosure (CE) with attached disk enclosures (DE) generally providing great scalability. Those systems may organise hundreds or even thousands of attached HDDs or SSDs. The configuration concerning type and number of attached drives greatly differs depending on the different requirements of companies and customers. Each system is typically sold in various basic configurations differentiating the number of controllers and drive size (2.5 and 3.5 inch) while modern systems usually support all of the common interfaces like SAS, iSCSI, FCoE or FC. Common supported RAID levels are 0, 1, 5 and 6.

The FUJITSU Eternus DX80 is presented as a mid-level example system[23] which has a separated controller enclosure. Table 13 and Table 14 below show selected specifications for the 3.5 and the 2.5-inch FUJITSU Eternus DX 80.

Table 13: Selected specifications of a 2.5 inch FUJITSU Eternus DX80

Model	Maximum number of attached drives	Maximum capacity [TB]	Maximum power consumption per device [W] (AC 200-240 V)	Maximum total power consumption [W]	Maximum weight [kg] (35 kg per enclosure)	Dimensions (BxTxH) [inch]
2.5 inch	120	480	CE: 630 DE: 570	2,910 (CE + 4xDE)	175	CE: 19x25.4x3.5 DE: 19x21.3x3.5

---

[23] http://www.fujitsu.com/downloads/STRSYS/system/dx80s2_datasheet.pdf Retrieved, 22. July 2014

Table 14: Selected specifications of a 3.5 inch FUJITSU Eternus DX80

Model	Maximum number of attached drives	Maximum capacity [TB]	Maximum power consumption per device [W] (AC 200-240 V)	Maximum total power consumption [W]	Maximum weight [kg] (35 kg per enclosure)	Dimensions (BxTxH) [inch]
3.5 inch	120	480	CE: 610 DE: 550	5,570 (CE + 9xDE)	350	CE: 19x26.4x3,5 DE: 19x21.9x3.5

1.1.5. Server and storage connectivity and architecture

1.1.5.1. Direct attached storage (DAS)

Direct attached storage (DAS) is a cost effective, short distance storage solution. There is a fixed wired, direct connection between the server and the storage system. Direct attached storage is an extension option for one server. However, there are DAS-products available for the connection of two or four servers possible. In the past, the connection of a host with a DAS-system was based on SCSI standard. Nowadays, SAS is the common interface standard for direct attached storage. These standards allow for only short distances of up to 10 meters. There is the (more expensive) option to realise a DAS-system based on Fibre Channel which would allow for longer distances, but this is rather unusual. Figure 10 shows the basic topology and elements of a direct attached storage system.

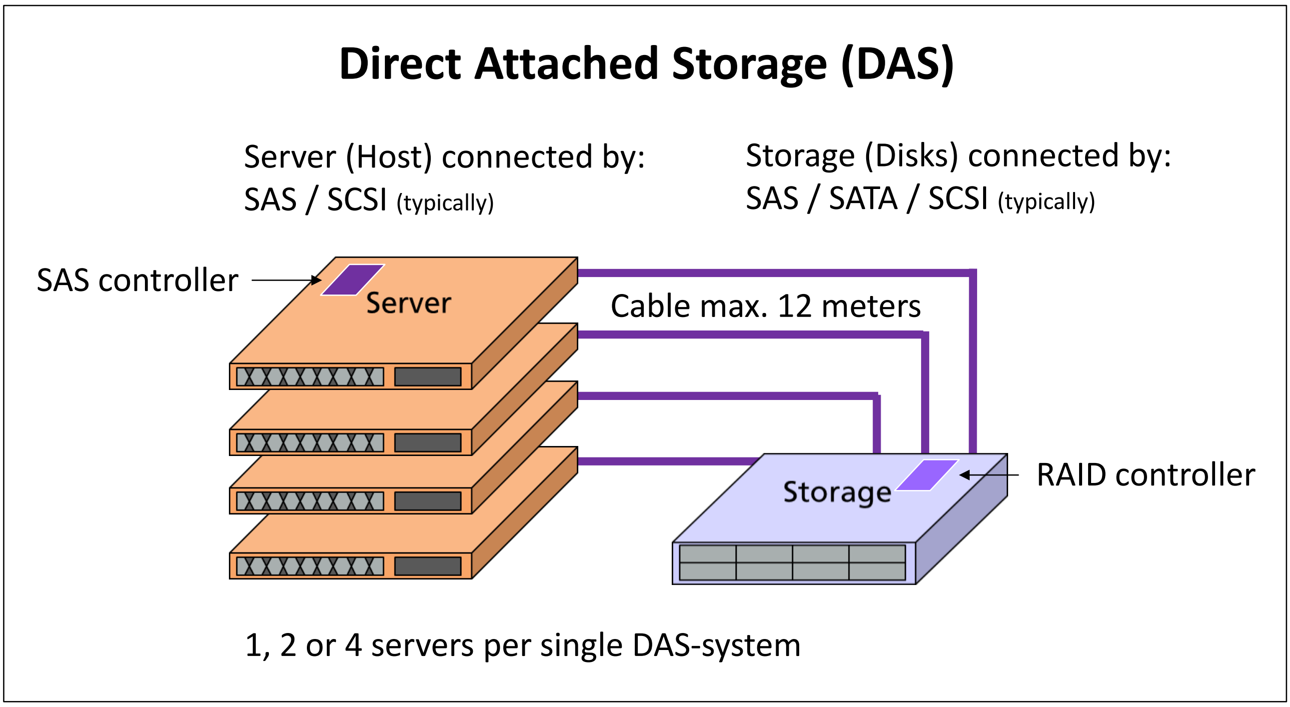


Figure 10 : Schematic diagram of direct attached storage

1.1.5.2.Storage area network (SAN)

Storage area networks (SAN) are fabric attached architectures for fast, secure and lossless communication between servers and storage devices. In comparison to DAS, SAN has the advantage that large amount of computer (servers) can access all storage systems through this network (fabric). This architecture supports allows for mutual administration of all storage elements and improves the utilisation of the existing storage capacity (consolidation). The consolidated storage can be accessed by the servers like a local storage device through the servers own operating system. SAN architecture consists of a server interface, the Host Bus Adapter (HBA) in the case of Fibre Channel (FC), a network element such as switch or router, and arrays of storage devices including RAID systems. SANs are typically used for block I/O services rather than file access services. SANs are fast growing applications and technically very sensitive to packet drops. Network devices need temporarily store frames as they arrive, until these assembled in sequence and delivered to upper layer protocol. Due to the high bandwidth of current available transmission standards it is possible that a device will be overloaded with frames. Therefore the protocol needs a mechanism to avoid such frame overload. SAN therefore utilise high latency and lossless protocols and network interfaces such as Fibre Channel (FC) or InfiniBand (IB). Over the past years, alternative solutions have appeared, including Fibre Channel over

Ethernet (FCoE) and internet Small Computer System Interface (iSCSI) which both allow communication through TCP/IP. Figure 11 shows the basic topology and elements of a storage area network.

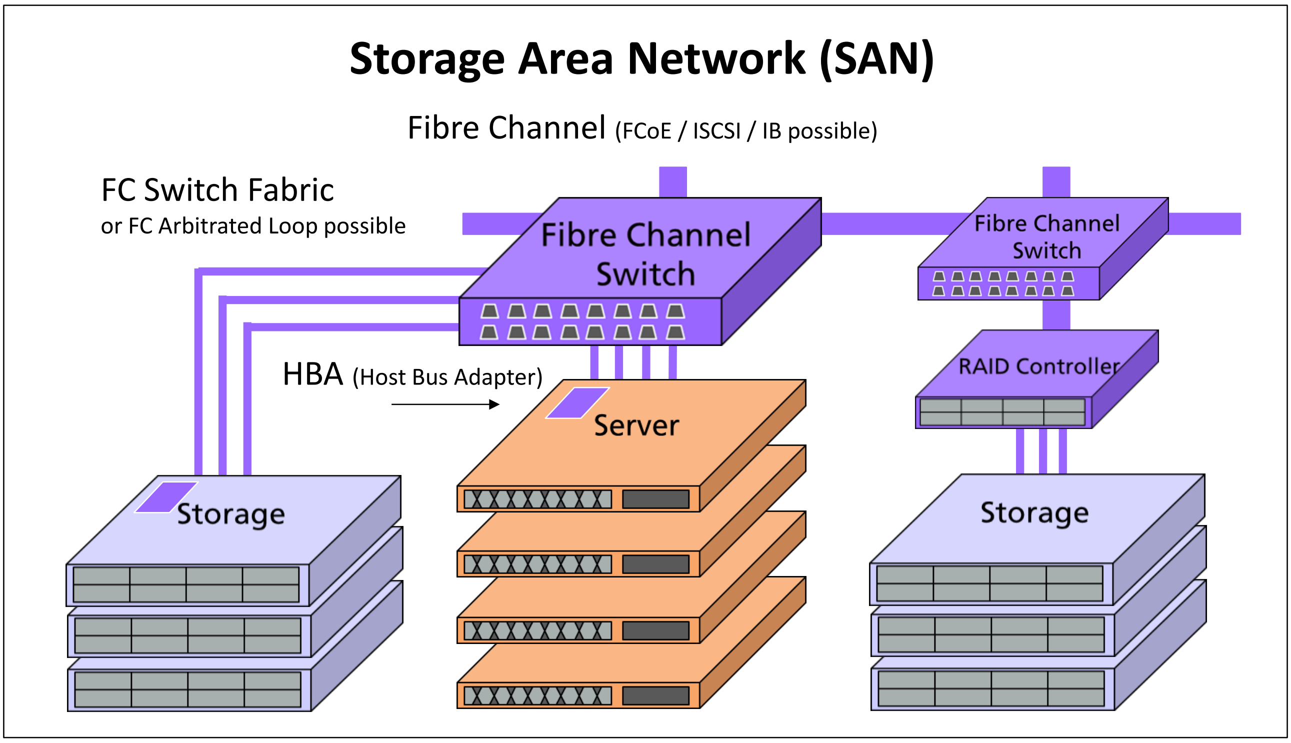


Figure 11 : Schematic diagram of storage area network

1.1.5.3.Network attached storage (NAS)

Network Attached Storage (NAS) is more cost effective solution to extent storage capacity and performance in comparison to SAN. NAS is therefore often used in small business networks, whereas SANs are preferred by large companies. This is due to the fact that adding more storage capacity to a NAS system can only be done by installing additional devices, even though each NAS operates independently. SANs, on the other hand, are deployed to handle very high-speed file transfers or many terabytes of centralised file storage given their high-performance disk arrays but requires its own hardware and interfaces (FC).

NAS are specialised file server or storage systems which are attached through a cost-efficient local area network (LAN) to a server. NAS consist of an engine that implements the file services, and one or more devices, on which data is stored. NAS products feature frequently an own operating system. NAS use different remote file access protocols such as network file service (NFS) in Linux/Unix environments or Common Internet File Services (CIFS) in Windows environments. NAS systems fit the SNIA category Near Online. Figure 12 : Schematic diagram of network attached storageshows the basic topology and elements of a storage area network.

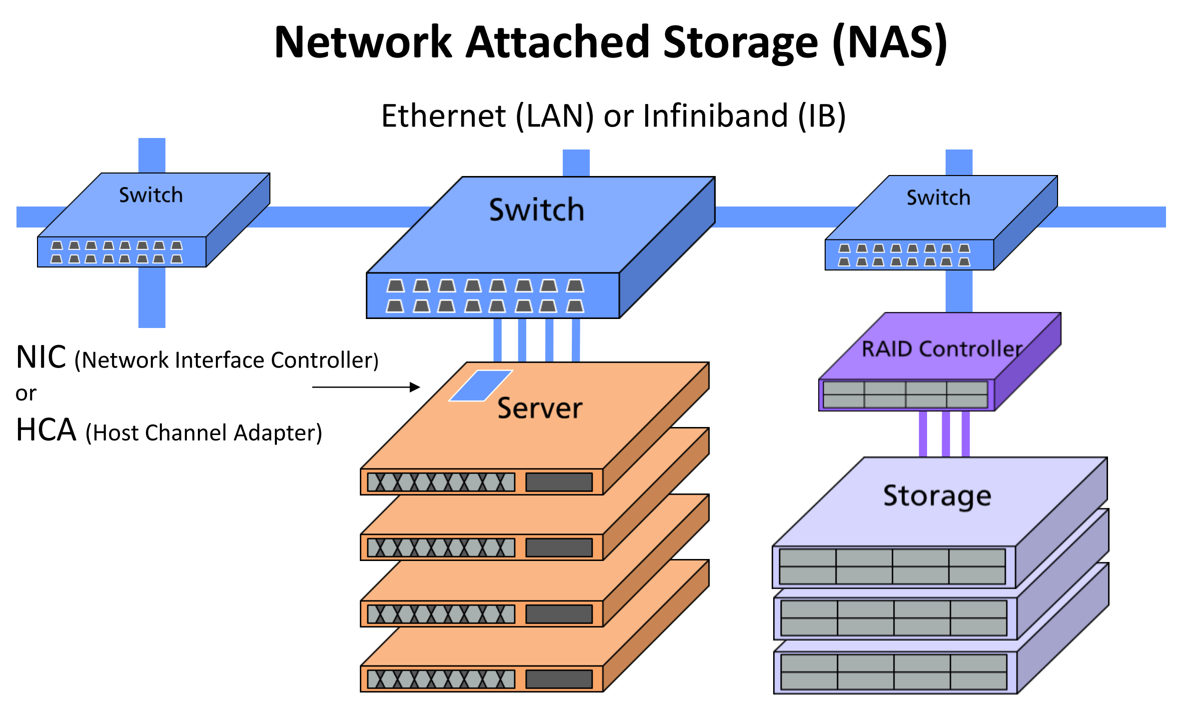


Figure 12 : Schematic diagram of network attached storage

1.1.6. Network protocols

The most commonly used network protocols in a classic data centre are SCSI, SAS, SATA and FC. Typically a classic high performance data centre features an Ethernet LAN for server-client communication and a separate Fibre Channel SAN for server-storage communication resulting in parallel network infrastructure and redundant network devices. Modern solutions employ unified fabrics merging the protocols to a single convergent network. The most common solutions are the lossless Fibre Channel over Ethernet (FCoE) or Infiniband (IB) and Small Computer System Interface over IP (iSCSI) as a low-price solution for small and midsized enterprises. The technology choice is generally driven by performance and cost efficiency demands. The protocols are described in detail in the following subsections.

1.1.6.1. Small Computer System Interface (SCSI)

SCSI is an industry standard that specifies the physical connection (electrical and optical interfaces) as well as the data transport (protocols and commands) between hosts (servers) and storage devices (SAS or SATA disk drives and tape libraries). SCSI has a long history and became a common standard for server and storage connection in the 1990s. The SCSI standard is buffered peer-to-peer interface, a parallel bus-system that comes with a 68-pin (single) and rather large connector which outlines some limitations of this technology. An advantage is that SCISI works independently from the host's operations system. Some technical specifications of the latest generations (in use since 2006 approximately) are listed in the Table 15. SCSI is extremely sensitive to packet drops and there have been further developments such as using the SCSI protocol in conjunction with Fibre Channel serial bus. The SCSI specifications are shown in Table 15: SCSI Specifications.

Table 15: SCSI Specifications

Interface name (Version)	Implementation (Year)	Throughput max. (in Mbit/s)	Distance max. (in meter)
Ultra 160 SCSI	1999	160	12
Ultra 320 SCSI	2002	320	12
Ultra 640 SCSI	2003	640	10

Even though Ultra 640 SCSI (Ultra-5) with a possible throughput of 640 Mbit/s was developed, there were no devices marketed with this technology as the industry was already pursuing the more promising path of Serial Attached SCSI.

1.1.6.2. Serial Attached SCSI (SAS)

SAS is a serial protocol and creates a point-to-point connection instead of the bus-topology of SCSI. The standard specifies all aspects from the physical layer up to the link, transport and application layer. The data throughput of SAS is improving with each generation. The common transfer speeds for SAS 2.1 is 6 Gbit/s and for the new SAS 3.0, 12 Gbit/s. Each SAS device consists of one transceiver with multiple transmitters and receivers for full duplexing. A HDD features two transceivers and a host bus adapter 4 or 8. The SAS standard also specifies various transport protocols for command-level communication with SATA devices (STP), SCSI devices (SSP), and for managing the SAS network (SMP). The SAS expanders are other elements that realise links between many SAS devices similar to a switch. Expanders contain two or more external expander-ports. An edge-expander can command up to 128 devices and a fan-out-expander up to 128 edge-expanders. This allows for a theoretical number of 16 384 connected devices.

There are currently various small form factor (SFF) connector types available. There is a distinction between internal and external connectors. The external connectors facilitate interoperation not only with multiple SAS devices (SFF 8484) but also with other standards including Infiniband (SFF 8470) and SATA (SFF 8482). The smaller form factor is an important feature because it allows not only the use of smaller storage devices such as the 2.5-Inch HDD but also increases the possible number of connectors on the server side. SAS are hot-swap capable.

Mini-SAS HD Active Optical Cable requires 0.8 Watt in use.

Some selected SAS specifications are shown in Table 16.

Table 16: SAS selected Specifications[24]

Interface name (Version)	Implementation (Year)	Throughput max. (in Gbit/s)	Distance max. (in meter)[25]
SAS 1.1	2005	3	6-10 (passive copper)
SAS 2.1	2008	6	20 (active copper)
SAS 3.0	2013	12	100 (optical)

1.1.6.3. Serial Advanced Technology Attachment (SATA)

SATA is the current version of the parallel ATA interface (bus) that features a bit-serial, point-to-point connection for the linking storage devices (disk drives and tape) with the host (server board). SATA is a fast, simple, cost effective, small form factor, and hot-pluggable technology. External SATA interface technology competes with USB and Firewire. SATA is a short distance connection.

The physical SATA connecting is made of a 7-line flat-band cable, which can be up to 1 meter long.[26] The connector is only 8 mm wide.

SATA is also utilizing low voltage differential signalling (LVDS) similar to SCSI in order to avoid signal interference. Table 17 shows selected SATA specifications.

---

[24] HGST (2013) « 12Gb/s SAS: Key Considerations For Your Next Storage Generation » http://www.hgst.com/tech/techlib.nsf/techdocs/9bf6d9a994f2e94586257b8600679a52/$file/12gbpssas_keyconsid_wp.pdf

[25] SNIA (2011) « SAS Standards and Technology Update » http://www.snia.org/sites/default/files2/SDC2011/presentations/monday/HarryMason_SAS%20_Standards_Technology_Updater1.pdf

[26] eSATA cable can be up to 2 meter long and xSATA up to 8 meter long.

Table 17: SATA selected Specifications

Interface name (Version)	Implementation (Year)	Throughput max. (in Gbit/s)	Distance max. (in meter)
SATA 2.0	2005	3	1
SATA 3.0	2009	6	1
SATA 3.2	2013	8 / 16	1

Since SATA 2.0, the interface throughput was typically not stressed by traditional HDDs and therefore was not a relevant performance factor. This changed as SSDs entered the market leading to the development of SATA 3.2 to cover highest transfer rates of modern SSDs. Mainframe SANs use ESCON or FICON.

1.1.6.4. Fibre Channel (FC)

Fibre channel protocol is a standard interface for fast and lossless storage area networks. FC standard defines connection architecture, physical components including connector and cables (not only glass fibre but also copper, data throughput, distances, coding, management and services. An important characteristic of FC is its high availability. FC avoids frame drops through link flow control mechanism based on credits called buffer-to-buffer credit. FC is reducing with its own hardware element, the Host Bus Adapter (HBA), the workload for the central processor unit of the server. FC allows for three basic topologies with respective port types including point-to-point (P2P), arbitrary loop (AL), and switched fabric (SW). Switch fabric is commonly used today. FC is in comparison to SCSI hot-pluggable. FC specifications are shown in Table 18.

Table 18: FC Specifications

Interface name (Version)	Implementation (Year)	Throughput max. (in Gbit/s)	Distance max. (in meter)
8 GFC (SPF+)	2008	12	20 - 150
16 GFC (SPF+)	2011	24	20 - 150
32 GFC (SPF+)	2016 (projected)	48	N/A

1.1.6.5.Fibre Channel over Ethernet (FCoE)

Fibre Channel over Ethernet is an extension of Fibre Channel (FC) providing an encapsulation of Fibre Channel frames over Ethernet networks using 10 Gigabit (or faster) Ethernet networks. FCoE enables merging parallel network infrastructure to one convergent network (unified fabrics). FCoE is usually used in high performance applications with a low tolerance for latency and packet loss and high throughput demands. FCoE runs on Fibre Channel hardware that is typically already implemented in high performance data centres making FCoE the first choice for larger enterprises. Compared to iSCSI, FCoE has a lower package overhead typically resulting in higher data transfer rates. Technically FC uses a stack with five layers (FC-0 to FC-4) and Ethernet uses an Open Systems Interconnection Model (OSI-model) with seven layers. FCoE basically transfers the FC layer FC-2 to Ethernet which enables higher FC-layers (FC-3 and FC-4) to be implemented in Ethernet infrastructures as well. This limits FCoE to layer-2-domains bringing hardware requirements to the data centre infrastructure. As a migration from different systems requires a lot of new hardware it is usually too cost inefficient. Therefore FCoE is typically implemented in existing Fibre Channel infrastructures or is considered as an option in the design of a new data centre.

1.1.6.6. Internet Small Computer System interface (iSCSI)

iSCSI is an IP-based SCSI networking standard operating over TCP. iSCSI allows merging of parallel network infrastructures to a convergent network (unified fabrics) as the server utilises the same interface for both LAN and SAN reducing cabling and required adapters. iSCSI is known to be very cost-effective as no new hardware is required while the performance is often considered lower compared to the more costly FCoE in terms of latency and packet loss, as the utilisation of TCP comes with an additional protocol overhead. The lower price makes iSCSI a common choice for small and medium-sized enterprises. In specific environments iSCSI can even be on par with FCoE performance as for example jumbo packaging is handled more efficiently. There is no clear winner between FCoE and iSCSI as both systems have different advantages and disadvantages, the choice is always bound to individual demands and conditions.

1.1.6.7. Infiniband (IB)

Along with the SCSI Remote Direct Memory Access (RDMA) Protocol (SRP), Infiniband is another high speed, low latency technology used to interconnect servers, storage and networks within the datacentre. Low latency networking with delays around 20 microseconds (end-to-end) is one thousand times less than 20 milliseconds for a data centre Ethernet network.

IB provides 1X, 4X and 12X variations of the basic 2.5 Gb/s rate[27]. Each IB link operates at one of five data rates in a bidirectional way. SDR, DDR and QDR use 8b/10b encoding, i.e. for 8 bits received 10 bits are sent, therefore 20% of sent data do not reach the recipient. FDR and EDR use 64b/66b encoding meaning for 64 bits received 66 bits of data are sent. Resulting effective theoretical unidirectional throughput is shown in Table 19. High Data Rate (HDR) is announced for 2017 providing a signalling rate of 50 Gb/s per lane. Table 19 shows some IB specifications.

Table 19: IB Specifications

Infiniband bond	Single Data Rate (SDR) [Gb/s]	Double Data Rate (DDR) [Gb/s]	Quad Data Rate (QDR) [Gb/s]	Fourteen Data Rate (FDR) [Gb/s]	Enhanced Data Rate (EDR) [Gb/s]
1X	2	4	8	13.64	25
4X	8	16	32	54.54	100
12X	24	48	96	163.64	300

Infiniband tends to be more expensive and complex than comparable solutions like FCoE or iSCSI and often requires new hardware. Therefore it is primarily used in applications where other solutions cannot match performance requirements.

1.1.7. Enterprise network equipment

1.1.7.1. LAN Switch

A switch is a multi-port bridge which forwards data frames based on information (destination address in the data packet) in the data link layer (layer 2) of the Open Systems Interconnection (OSI) model.

- Layer 2 access switches

Access switches (L2) provide connectivity on the application layer. The distinguishing factors are:

  - Connectivity: 10/100 Megabit to 1/10 Gigabit Ethernet (mostly copper cabling)
  - Service: with Up-link port or not
  - Form factor: most common are 1-rack unit or other small form factor
  - Number of ports: configurations with 8, 12, 16, 24 or 48 ports
  - Configuration: standalone, stackable or modular (PoE) configurations
  - Software: Security, Quality of Service, routing

- Layer 3 / layer 2 core switches

Core Switches (L3 / L2PM) provide fast, reliable, and policy defined connectivity on the distribution layer within various application areas in the data centre. Layer 3 switches also process data on the network level (layer 3 and higher) providing additional functions like IP-filtering, prioritisation for QoS, routing, monitoring, and additional controls that may rely on information carried on higher layers than layer 2. Data centres usually implement multilayer switches. The distinguishing factors are:

  - Connectivity: 10/100 Megabit to 1/10 Gigabit Ethernet (copper and fibre cabling)
  - Service: with Up-link port or not

---

[27] http://www.mellanox.com/pdf/whitepapers/InfiniBandFAQ_FQ_100.pdf Retrieved, 23. July 2014

  - Form factor: most common are 1-rack unit or other small form factor
  - Number of ports: configurations with 16, 24, 48 or more ports (incl. SFP)
  - Configuration: standalone, stackable or modular (PoE) configurations
  - Software: Security, Quality of Service, routing
- Router or gateways

Routers or Gateways securely connect the data centre's core distribution area to the outside world. They feature inbuilt security (see further below). The distinguishing factors are:

  - Connectivity: Wide Area Network (WAN) interface and Gigabit Ethernet Local Area Network (LAN) switches. Service: Firewall, Security (IPsec)
  - Service: outward and inward facing applications (e.g. DNS, SMPT).
  - Form factor: 1-2 rack units and stand alone
  - Number of ports: 8 – 196 ports
  - Configuration: standalone, stackable or modular configurations
  - Software: Security, Quality of Service

In order to protect the data centre from the outside and inside routers deploy firewalls, access control lists (ACL), intrusion detection systems (IDS), virtual private network (VPN) and other specific security solutions

1.1.7.2. Cables and cabling

The network equipment is linked via cables. There are two basic media used in data centre – copper and glass fibre. The selection of the transmission media depends on the required performance in terms of speed, bandwidth and range, intended network topology, the durability and ease of installation (pull, break, bend properties) and most of all, the cost factor.

Copper wire considerations include:

  - Lower costs: in comparison to optical fibre
  - Limited range: e.g. practical distance for 10GE is about 100m
  - Power consumption: for 1GbE and 100ME, typical Ethernet PHY[28] consume less than 1W (at 10m)

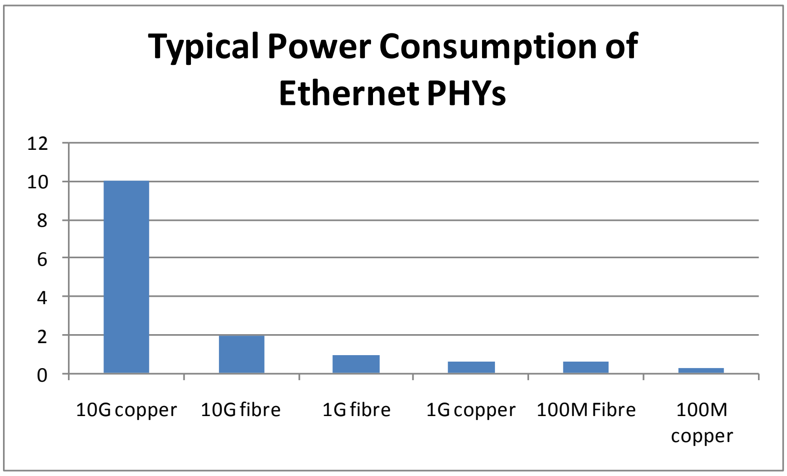


Figure 13 : Average power consumption of different cable ports in 2009 [Schneider 2010]

---

[28] PHY is a short form of the physical layer of the Open Systems Interconnection (OSI) model

Considerations about optical fibre include:

- High CAPEX (initial costs) because of active components (photonics)
- Higher performance: speed, bandwidth over long distance.
- Power consumption: with 10GbE port optical fibres consume with 1 – 2 Watts about 20% of the power than comparable copper PHYs
- Useful application for core switching layer while copper cables remain first choice to connect server with access switches

The small form factor and the recent trend to higher integrated chip solutions with lower power consumption resulted in high efficient devices for both technologies. Significant efficiency improvements have been obtained for the copper PHYs. Even with future generations of 10 GbE copper (less than 4W are expected), the optical fibre solution still require less power to operate compared to copper PHYs.

The IEEE 802.3az Energy Efficient Ethernet (EEE) standard tries to reduce energy consumption of Ethernet devices by defining a low power mode for inactivity periods. A PHY that has no frames to transmit can turn into a low power mode and, as soon as new frames arrive, it is set back into the active mode within a few microseconds. This enables energy savings that are rather transparent to upper protocol layers. Under these conditions, for lightly loaded links, a copper PHY that implements EEE may be more energy efficient than an optical PHY. However a study by Larrabeiti concludes that 10 Gbps Ethernet fibre PHYs will remain more energy efficient than copper PHYs even when EEE is implemented on the copper PHYs, except for links whose load is extremely low. Besides core networks which usually have medium loads at their Links, the copper EEE PHY is not a competitor in terms of power consumption. Only in scenarios where the traffic goes down drastically for long periods, the energy savings can be greater compared to fibre PHYs [Larrabeiti et al. 2011].

A direct comparison of Fibre and Copper cables can be found in Table 20.

Table 20: Comparison of Fibre and Copper cables [Neveux 2008]

10 GbE Media	Fibre (SR/LRM)	Copper (Base-T)
Power Consumption (PHY + Adaptor)	1– 3 Watts	14-17 Watts
Distance	300 meters	100 meters
Future Data Rates	40 – 100 Gb/sec	10 Gb/sec
Density per Rack Unit[29]	32	24
Cable Density	10%	100%

In high density applications, this power usage is significant, not only in consumption, but also in the generation of heat, which requires cooling to protect transmission equipment from rising temperatures. A 10GbE copper system will require more switches and line cards to match the bandwidth capability of a 10GbE optical system.

Large cable carriers or cable bundles can block the airflow on the backside or in the raised floor area. This situation can lead to less effective air flow and leakages. The objective is to keep hot and cold isles strictly separated and have good air flow pressure (cold air into the system).

Other issues related to cables that need to be considered include:

- Clip cables together in order to improve backside airflow
- Consider overhead pathways
- Locate and close leakages (particularly in the raised floor use brushes, special tiles)

---

[29] Based on Cisco Nexus 7000SFP+10GBase-SR module.

1.2. Products with standard improvement (design) options

The baseline for products with standard improvement options is difficult to determine. The industry is in constant progress providing new components generations, system architectures, product configurations, operating and application software in short time intervals of a few months. Through that, the performance increases periodically and up to now without increase in energy consumption. This constant progress is expected to continue in the foreseeable future. A complex improvement option on the system level lays in energy trade-offs considering reduced data centre cooling and air conditioning efforts and an increase of inlet temperatures. This is addressed by the thermal management regarding material use and design options and holds for both enterprise servers and storage equipment.

1.2.1. Latest component generation

A simple indicator for state-of-the-art design on the hardware level is the performance gain of every new processor and memory generation. According to Moore's law the CPU performance is doubled every 18 - 24 months while maintaining the level of power consumption. One example is the new DDR4 memory generation. With lower supply voltage and better termination scheme the DDR4 has shown theoretical power reduction of up to 37%. However, real world testing indicates somewhat lower power reduction in the range of 10 to 15%. Power consumption on the chip-level is reduced by voltage scaling, multi-core designs, new materials, and single core power management as well as power gating. On the board-level power consumption is reduced by integrating passive component functionality into semiconductor ICs thus reducing the number of actual components that need to be powered and generating thermal (energy) loss. System integration is not only reducing the number of electronic components but also the complexity of the conventional circuitry on printed circuit board level. The size of the PCB and number of copper layers is currently not increasing.

The periodic performance improvement on the hardware level is a standard important aspect. A second aspect is the hardware and software configuration of a server or storage product. The energy and resource efficiency can be positively influenced when the configuration of the product matches the intended application. A specific configuration might be optimal for one purpose (application) but less optimal for another. This is why for instance SPEC SERT is testing different performance requirements. On an individual base, such tests provide very useful information in conjunction with procurement processes where performance data for specific product configurations are needed.

1.2.2. Storage media and utilization

With regards to storage power consumption there are a few considerations with respect to storage media selection on the one hand and storage utilization optimisation on the other hand. SSDs have in comparison to HDDs a certain advantage regarding lower idle power for same number of drives. Active power differences depend on total memory size and rate of read/write transactions at storage array. In some cases, SSDs have a higher maximum power demand as compared to HDDs. Performance improvements may be seen in storage worklets in SPEC SERT and in the comparison of maximum and idle power measurements on identical systems with all SSDs or all HDDs.

The current server generation features HDD lower power states which have existed for years. When IDLE, the individual power condition IDLE_C reduces the RPM of the drive and SLEEP spins down the drive[30]. While SLEEP reduces power further, it has a longer latency to resume. The actual utilisation of these low power options strongly depends on the customer's willingness or capability to tolerate access time delays. Also, link level power management is becoming more robust and continues to add more functionality. Front-end and back-end connectivity including PCIe, SAS, SATA Ethernet, etc. support various link power management features which include reducing number of active lanes, reducing speed, and IDLE states. The proper incorporation and testing of these power management options is necessary in order to ensure availability and meet quality-of-service requirements.

Deduplication, data compression and other software based options to optimise the storage capacity by managing redundant data are available as products in the market (NetApp, VMWare, etc.). These options are typically implemented in medium and large size storage systems (Online 3 upwards) but also available for smaller systems (Online 2). Deduplication, data compression and other software based options are often automatically activated and typically operated during low traffic night time hours. According to industry stakeholders, storage capacity improvement through these measures is in a range of 25 to 50% for highly virtualized systems.

---

[30] http://www.seagate.com/files/docs/pdf/whitepaper/tp608-powerchoice-tech-provides-us.pdf

1.2.3. Power management

Power management options are generally provided through ACPI protocol for Microsoft/Windows operating system and similar functionality for other operating systems in conjunction with processor platforms such as SPARC, POWER and others. The utilization of the power management options is according to industry stakeholders very low. Customers are said to disable power management in order to ensure high availability and continuous operation. On the other hand, the utilization of servers in traditional data centres is still rather low and often influenced by meta-stable day-and-night cycles. This situation would allow for a more consequent power management including low power idle. There are two low power states under development:

- Inactive power state 1: 90% of idle power consumption with up to 10 seconds recovery time
- Inactive power state 2: 30% of idle power consumption with up to 20 minutes recovers time

Power management capability is driven by the CPU development. Multi-core-CPUs in the field of personal computer demonstrated low power options through partially suspending individual cores.

1.2.4. Effective data processing

Multiple cores offer the ability to run multiple single threaded activities on the same machine. Single threaded software still dominates the commercial space as multi-threaded updates are introduced. Multi-threading is an on-going software activity that correlates to (but, is not the same as) the activity on parallel computing.

Parallel computing is a software optimization concept that uses parallel compute structures to accelerate processing for large complex problem sets. The use of parallel computing requires programming languages and compilers which are capable of assembling the code to run on parallel compute structures, resolve interdependencies (e.g. conditionals dependent on results of a parallel compute activity, and converge on the processing. Parallel computing utilizes dedicated data structures along the computing path, whereas multi-threading can be employed on a common (e.g. shared dataset). As a result, the intermediate memory structures may be referred to as a type of memory cache for the individual computing path. As implied, the (data and programming) memory structures are quite large and may require dedicated management.

1.2.5. Virtualisation

The improvement of the server and storage utilisation goes hand in hand with an active power management. Virtualisation capability and respective software is a common technology and gets more and more implemented in the field. Virtualisation allows the consolidation of hardware by creating multiple virtual servers on single physical machine as well as by shifting loads between such servers. Multi-threading and parallel computing has some conflicts with virtualization as each thread may consume a virtual machine (VM), thereby reducing the number of single VM jobs the systems can support. Multi-threads generally stay on the same physical machine.

The assessment of the server’s virtualization capability is to some extent possible based on the SPECVirt_sc2013 benchmark. This benchmark determines the optimum configuration to achieve virtualization for a particular architecture and system. However, industry stakeholders indicate that this benchmark requires specialized infrastructure far beyond the server and a test run is therefore resource and time intensive.

1.2.6. Modular design

Modular server systems have the advantage of shared resources and prefabricated connectivity. Due to the requirement of providing scalable server capacity with exchangeable server and storage modules, new form factors with modular enclosures have entered the market. Modular server systems such as blade systems or micro server systems consist of two housing elements. One element is the individual chassis (enclosure) for the server modules. A server module is basically the populated server boards with the processors and memory as well as connectors for power and network. These server modules are called blade, cartridge, or book, depending on the manufacturer’s preference. A certain number of server modules are inserted into a larger system enclosure, which is housing the shared resources including network interfacing, additional storage devices, active cooling and power supply devices. Modular design improves performance scalability and maintenance with regard to resource efficiency. Modular designs allow for more application oriented hardware configurations and therefore a better overall utilization of the existing IT resources.

1.2.7. Efficient power supply

An efficient and properly dimensioned power supply unit is the basic element for an energy efficient server or storage product. The user (customer) has often the option to select the PSU on its own. The electrical, physical and thermal specifications are provided by the equipment manufacturer. The mainboard design considers usually the thermal properties of the PSU (e.g. how cooling management of the PSU is influencing the airflow on the mainboard). In recent years, server vendors have offered a PSU portfolio with different PSU capacities.

Besides optimizing cost, right sizing the PSU improves PSU efficiency through higher average load power thus reducing total system power. Right sizing the PSU, meaning not over-dimensioning the PSU, has the capability for considerable power reduction. However, this may also limit configurability and ability for customers to add additional functionality to system after original purchase.

80plus certified PSUs are good practice already. However, it is unknown to what extent different levels (e.g. bronze, silver, gold, platinum, and titanium) are actually sold in the market. This is important because the performance criteria for platinum and titanium category are considerably more ambitious. An 80plus titanium PSU has to support an energy conversion efficiency of 94% at 20% load. The overall energy improvement potential related to a more efficient PSU e.g. shifting from 80plus silver to 80plus platinum is in an approximate range of 5% of the resulting power consumption. Common server and storage products are currently supplied with 80plus silver and 80plus gold power supply units. Platinum and titanium PSU are currently only available for larger power output typically over 750 watts.

According to stakeholders, Industry currently (2015) is early in the transition to Gold level PSUs in multi-output power supplies and is much more mature in the availability of Gold with some amount of platinum in the single output power supplies. For both single output and particularly for multi-output power supplies the larger, higher output power capable supplies are slower to transition to higher efficiency levels. Storage systems use multi-output supplies to optimize voltage regulation conversion within the system. Managing the various voltage shifts within the power supply optimizes the conversions and can reduce the overall power system losses. As an example, consolidating the voltage shifts into the multi-output supply may decrease the power supply efficiency by 1%, but increases the overall power system efficiency by 4%.

1.2.8. Thermal management

Thermal Management addresses the transport of heat from the component (e.g. CPU) out of the product to the warm aisle of the data centre. The material of the heat sinks and other passive cooling elements are determined by the form factor and thermal performance as well as by price considerations. Most server and storage products feature load adaptive air-cooling fans, baffles for airflow optimisation, blinds for unused modules, and other features. A random sample of server product data sheets indicate that a majority of products can be operated in the ASHRAE A2 envelop up to 32°C ambient temperature.

Fluidic cooling on product level exists, however, it is very expensive and only used in higher performing product configurations. Various sensors on component and product level provide real-time data e.g. for adjustment of the fan speed. The adjustment (control) is today mostly done based on pulse width modulation (PWM).

The thermal management has the task of transporting the heat away from the active components. Improving the thermal design can promote higher resource efficiency by utilising optimised form factors and materials for passive cooling. Energy efficiency can also be improved by properly dimensioning of active cooling elements including all air and fluidic cooling systems. Currently, test procedures for the confirmation of proper thermal design and management are very complex and only address singular aspects. Simulations are used but the comparability is limited as the model parameters are not standardized. Sub-aspects to consider are electromagnetic compatibility, signal integrity, form factors, fan throughput, and size limitations. Furthermore, various configurations of optional and modular designs interact with the thermal management.

1.2.9. Higher inlet temperature

The interaction between the IT equipment internal cooling system and the external (data centre level) temperature conditions and cooling infrastructure has been addressed as an important improvement aspect throughout this study (see for more details chapter 1.2.3 in Task 3). It is assumed that an increase of the inlet temperature could reduce the overall energy consumption of the cooling infrastructure in the data centre, while the power consumption of the IT only slightly increases. These options have been implemented in Europe in conjunction with so called free-cooling. There are however concerns regarding the reliability and aging of the electronic components when servers and storage systems are operated constantly under higher temperature. Further investigations are necessary.

1.3. Best Available Technology (BAT)

Disclaimer

If quantifications of BATs were available it was explicitly mentioned in the text. In most cases quantitative data for the environmental improvement potential of BATs could neither be obtained from stakeholders (see questionnaire) nor from other public sources.

The best available technology is a moving target and difficult to determine. All technologies have a timestamp of market introduction. Therefore the latest technology is usually the best in certain key aspects at the time of introduction. But the latest technology is usually costly and may not yet be readily available (production volume). A BAT in one market segment may be commonplace in a neighbouring segment, or may not yet be used commercially in a third segment. The speed of development in the IT sector is very high such that a clear distinction of technology maturity in combination with a transparent timestamp would be needed. The available sources (including company information) do not always provide this level of structured details.

1.3.1. Power supply unit (PSU)

Currently 80 PLUS Titanium power supply units provide major energy-efficiency advantages as compared to lower rated PSUs. Titanium is the only category that applies requirements to 10% load states while the lowest load requirements for other categories are set for 20%. The titanium conversion efficiency requirements are shown in Table 21.

Table 21: 80 PLUS Titanium conversion requirements

Load level	10%	20%	50%	100%
Efficiency requirement	90%	94%	96%	91%

So far, power supply losses are considered to account for 10-20% of server power consumption. According to DIGITALEUROPE those losses are expected to decrease to 5-10% due to the migration to gold and higher ranked power supplies. Table 22 shows a comparison of server power consumption in regard of differently ranked power supply units.

Table 22: Comparison of power supply efficiencies at different PSU ratings provided by DIGITALEUROPE

		Power Supply Power Consumption as % of Daily Server Power Consumption						
		Bronze	**Silver**	**Gold**		**Platinum**		**Titanium**
Average Server Utilization	Server Power Consumption	1 692 W	1 640 W	1 200 W	1 725 W	900 W	1 927 W	1 925 W
	kwh/day							
10%	9.6	18.7	18.1	9.8	12.7	7.7	9.4	5.2
20%	10.5	18.0	17.0	9.2	12.1	7.4	9.1	5.0
40%	11.4	17.1	15.8	9.0	11.5	7.2	8.8	4.9

The energy saving and customer cost benefit of advanced power supply units is considerable and has been exemplary calculated by Fujitsu. According to the published whitepaper “PRIMERGY Modular Power Supply

Unit for PRIMERGY Servers", the little higher costs for an 80plus Titanium PSU will be compensated by the overall energy saving of 275€ over a 5 year lifetime.[31]

Example: PRIMERGY RX 2540 M1, redundant 800W PSU, PSU load 25% equals 200W net power consumption per PSU, 400W overall power consumption

- 80 PLUS Titanium PSU: 12,7 W loss (94% efficiency @ 25% load)
- 80 PLUS Platinum PSU: 22,2 W loss (90% efficiency @ 25% load)
- Average power saving: 19 W for 2x PSU
- Server runtime: 24 hours / 365 days / 5 years
- Energy costs: 0,15 € / kWh
- Power usage effectiveness (PUE): 2.2

Energy / cost saving calculation:

- 19W * 24h * 365d * 5a * 0,15€ * 2,2(PUE) = 275 €

Highly efficient power supplies require less cooling resulting in less air flow from the PSU that may compromise the air flow design of the server. So server design needs to support the high efficient power supply units in terms of air flow management.

It has to be noted that storage systems may use single output or multi-output power supplies. The two supplies have different efficiency characteristics due to the higher losses associated with multiple voltage adjustments in the multi-output power supply.

Right sizing the PSU, meaning not over-dimensioning the PSU has the capability for considerable power reduction. A majority of server feature today redundant PSUs. Load sharing between redundant PSUs could lead to inefficiency because both PSUs might operate at a lower point on PSU efficiency curve (optimum is typically at 50% load upwards). To minimize the power consumption losses of redundant PSUs, redundant PSUs may be turned off until there is a failure of the primary PSU. This approach results in generally lower system power consumption in redundant PSU configurations for lightly configured systems at low utilization levels. Magnitude of improvement depends on PSU efficiency curves. However, this approach also results in higher PSU cost to support longer hold over requirements and added switch over functionality. In the case where the redundant PSUs are fed from separate, redundant power feeds, the use of this approach unbalances the power delivery and creates instabilities in the power system. Some customers turn this capability off because of the instability and impact on data centre stability.

1.3.2. Microprocessors and memory

Microprocessors are the central aspect of every server design and typically comprise a large percentage of the motherboard's capabilities, cost and energy consumption. Doing more with less is the obvious answer. The design challenge continues to be how to implement more efficient multiprocessing and multithreading. Recently, for chips, the term more efficient has come to mean not just more MIPs per chip but more MIPs per watt as well. The most common technique for implementing multiprocessing is to design chips with multiple embedded processing cores. This enables the software multithreading that allows the server to execute more than one stream code simultaneously. Multithreading also reduces energy consumption as the same computing power on a single thread leads to higher heat dissipation that exponentially increases the cooling requirements. In addition to the multicore/multithread enhancements, processors can now execute 124-bit instructions, which results in a significant performance boost.

One industry stakeholder provided this list of BAT for microprocessors:

- Advanced processor power management such as adaptive voltage operation (also called adaptive clocking). This addresses the processor voltage fluctuations by temporarily reducing clock frequency instead of providing higher voltages to address momentary voltage deviations. It also includes the monitoring of processor voltages which occurs at a fraction of a billionth of a second.
- Newer generation processors support individual power rails for each core. This enables each core to operate at a unique performance state (P-state). This technology reduces typical power consumption in real world applications as each core P-state is optimized based on the core's real time workload. In

---

[31] http://globalsp.ts.fujitsu.com/dmsp/Publications/public/wp-py-modular-psu-concept.pdf

order to recognize this behaviour a benchmark would require an asymmetrical core loading workload. It is unknown if there is any existing benchmarks to support this.

- Improved C-state (CPU core and socket sleep states) implementation will reduce processor power at low utilization and at idle. C-states impact latency as core/socket must “wake” before executing next batch of instructions. Some performance degradation at low utilization levels dependent on workload.
- Smart processor power management featuring an “intelligent boost” microcontroller which tracks application behaviour of the processor in real-time and boosts the processor frequencies only when there is an actual benefit of higher performance. This reduces processor power at times of low utilization.
- Processor and chipset improved power gating which reduces power consumption by shutting off the current to hardware blocks (functions) of the circuit that are not in use. Power gating increases time delays, as power gated modes have to be safely entered and exited.
- Heterogeneous processor cores with functionally tailored hardware elements for a specific task. This improves energy and material efficiency by properly dimensioning that hardware to a specific application. This may include the implementation of GPUs for large data sets, parallel data streams applications.

The design and implementation of the commodity memory architecture has resulted in significant performance and capacity limitations. To circumvent these limitations, designers and vendors have begun to place intermediate logic between the CPU and DRAM. This additional logic has two functions: to control the DRAM and to communicate with the CPU over a fast and narrow bus. The benefit provided by this logic is a reduction in pin-out to the memory system and increased signal integrity to the DRAM, allowing faster clock rates while maintaining capacity. The DDR4 RAM is recently available. With lower supply voltage and better termination scheme the DDR4 has shown theoretical power reduction of up to 37%. However, real world testing indicates somewhat lower power reduction in the range of 10 to 15%.

1.3.3. Hard disk drives and storage

According to Toshiba[32] in the past decade the annual growth rate of HDD areal recording density was between 30 and 50% trending towards ~40% in the following few years[33]. The growth is driven by more and more refined technologies. The latest milestone was the commercialization of Discrete Track Recording (DTR) on HDDs that entered the market in 2009. DTR raised aerial density to about 2 TB/inch² by removing medium between tracks and minimizing magnetic interferences. The next step in capacity growth was marked by Shingled Magnetic Recording (SMR) that entered the market in 2013. By letting tracks overlap and avoiding critical magnetization thresholds the effective area of the medium is increased enabling aerial densities of ~5 TB/inch² at the cost of overwriting performance as overwriting existing data requires additional processing of neighbouring tracks as a result of the overlapping.

The current top-end product using SMR is the 8 TB Archive HDD by Seagate[34] that achieves 25% more capacity on six platters due to SMR. The Archive HDD is a cost optimized product designed for the workload demand of storage systems in cloud data centres. They only allow for a limited annual access of 180 TB. The rotation speed is low at 5 900 RPM. This is a good example for an application specific product development. According to Seagate the 5 TB 4-platter HDD has an idle power consumption of 3.5 Watt and active power of 5.5 Watt. The larger capacity (6/8 TB) 6-plater HDD requests 5 Watt in idle and 7.5 Watt in active.

There are other technologies that increase the effective recording media space. For example helium-filled hard drives: As helium only has 14% the density of air there is less drag on the spinning platters meaning less motor power is required for spinning. The number of platters in a 3.5 inch enclosure is raised from 5 to 7 increasing total capacity accordingly. A 6 TB HDD from HGST is available (Ultrastar He6)[35].

Another approach is to reduce storage disk space demand by different compression technologies on the software or hardware level. Generally compression reduces the required disk space but may decrease

---

[32] Masakatsu et al., 2011, Trends in Technologies for HDDs, ODDs, and SSDs, and Toshiba’s Approach: http://www.semicon.toshiba.co.jp/product/storage/pdf/ToshibaReview_vol66n8_02.pdf

[33] Francis, 2011, Data Storage – Trends and Directions: https://lib.stanford.edu/files/pasig-jan2012/11B7%20Francis%20PASIG_2011_Francis_final.pdf

[34] Seagate, 2014, Archive HDD 8 TB product specification: http://www.seagate.com/www-content/product-content/hdd-fam/seagate-archive-hdd/en-us/docs/archive-hdd-dS1834-3-1411us.pdf

[35] HGST, 2013, Ultrastar® He6 3.5-Inch Helium Platform Enterprise Hard Drive: http://www.hgst.com/tech/techlib.nsf/techdocs/F8B3820BADAD9E6588257C160032F257/$file/HeliumProductSummary_final.pdf

performance (read speed). This requires typically data processing capacity and therefore CPU power. Most solutions aim at compressing files that are used infrequently or not mission-critical. One of the most popular algorithm-based enhancements is data-deduplication which aims at identifying redundant data (duplicates) and replacing those with a small reference before data is written on a non-volatile disk. The processing takes places before data is encrypted. Conventional technologies to deduplication (e.g. target deduplication) require excessive hardware resources (compute and network), create network traffic, and often need additional backup server resources to boost deduplication performance. There are proprietary technologies (products) available that claim to drastically reduce necessary hardware (and respective power consumption) by more intelligent deduplication solutions.

Active power management in enterprise storage is currently not common. Over the past 20 years there have been considerable research concerning HDD power management (mainly in less QoS-depended consumer product) including the investigation of the pros and cons with respect to HDD sleep mode (spin-up and spin-down).[36] Simple methods are the deployment of a timer to monitor the disk access traffic. Analysing the utilization history (profiling) and predicting possible low activity periods is another more complex (and legally problematic) approach. Power mode transition shows a good potential for energy saving. However, latency is an issue as well as reliability. Cache deployment in a storage I/O procedure is another method to reduce energy consumption that has been investigated. Online power management in VM storage has been investigated as well and show considerable results of energy savings up to 34%.[37] Applying virtualization technology in order migrating storage demand to the energetically best option is another methodology. Unfortunately there is no quantification of the results of virtualization to improve energy efficiency in an defined enterprise storage system.

1.3.4. Modular Servers and Microservers

Modular server designs and server fabrics have a few advantages in terms of energy and material efficiency due to prefabricated network connectivity (including switching) and other shared resources including cooling fans, power supplies and storage disks. The modularity allows for server disaggregation by creating larger pools of computing resources from which virtual machines can be provisioned. Modular servers such as blade and other multi-node systems with conventional CPU, RAM and chipsets are standard technology. Hot swapping modules for easy scalability and maintenance are also common. Best available technology is coloured buttons and handles indicating which component can be exchanged etc.

Microservers are a relatively new specialised server design developed to carry out a multitude of lightweight workloads very efficiently. Small form factors are achieved by using system-on-a-chip (SoC) boards that have the CPU, memory and system I/O in a single integrated circuit without a socket reducing space and hardware material (see Figure 14).

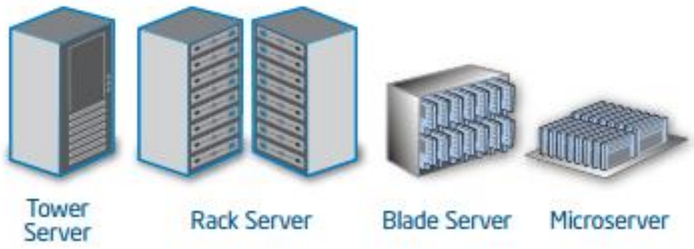


Figure 14: Examples of common server designs

Microserver processors typically have lower single thread performance and more available threads than traditional server systems being more cost- and energy efficient for lightweight workloads that for example commonly occur in networked storage, app servers or web services. More infrastructure devices like fans and power supplies are shared among server devices increasing density and lowering power per node demands.[38] The larger number of cores allows lower power single threading that greatly supports virtualisation efficiency.

The current prototype of microserver design is the HP Moonshot project. HP has released figures claiming that 1,600 of its Project Moonshot Calxeda EnergyCore microservers, built around ARM-based SoCs, packed into just half a server rack were able to carry out a light scale-out application workload that took 10 racks of 1U

---

[36] A good overview of the topic including a list of respective studies is provided by: He Li et all (2014): Deduplication-Based Energy Efficient Storage System in Cloud Environment, In « The Computer Journal Advance Access » published December 7, 2014

[37] Nathuji, R. and Schwan, K. (2007) Virtualpower: Coordinated Power Management in Virtualized Enterprise Systems. Proc. 21st ACM SIGOPS Symposium on Operating Systems Principles, Stevenson, WA, October 14–17, pp. 265–278. ACM, New York, NY, USA

[38] http://www.intel.com/newsroom/kits/atom/c2000/pdfs/Intel_Microserver_Whitepaper.pdf

servers -- reducing cabling, switching, and peripheral device complexity. In comparison to a traditional x86 server system HP announced Moonshot systems (see Figure 15: HP Moonshot outline) to use up to 89% less energy, 94% less space, 63% less cost and being 97% less complex on lightweight workloads.

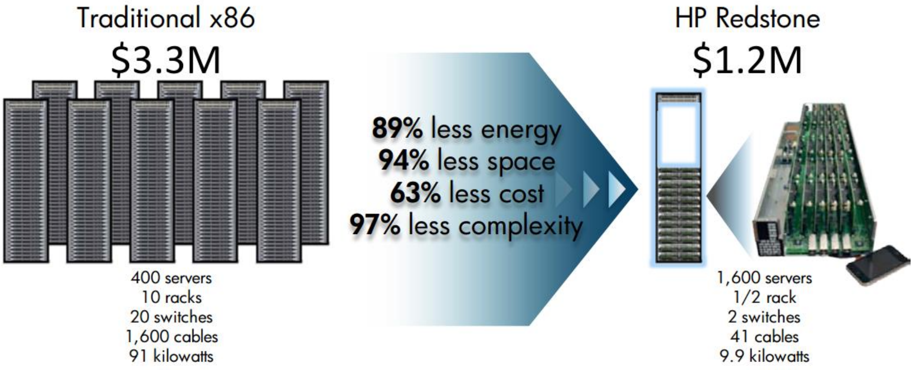


Figure 15: HP Moonshot outline[39]

By miniaturisation and reduction of complexity the number of servers per rack can increase by a factor of one hundred while federated infrastructure scales seamlessly with additional server modules making- up and downgrading more practicable.

Other competing products in this market segment are Dell PowerEdge, Cisco UCS, Lenovo/IBM FlexSystem or AMD SealMicro.

1.3.5. Advanced cooling

Thermal Interface Materials (TIM) are often polymer systems with special filler technologies mostly aluminium-oxide particles, to close the gap between CPU and cooler. Important technology parameters are high thermal conductivity (pads up to 6 W/(mK). Phase changing materials are also use for optimal thermal conductivity during the whole temperature profile. Pure metal interfaces are expensive and not exchangeable. BAT are pastes with up to 25 W/(mK) which is achieved only with directional graphene nanotubes. A low thermal resistance depends on the contact pressure.[40]

Heat spreaders are necessary to spread the heat fast and continuously over large areas, to protect the processor or other ICs from overheating. Basic heat spreaders use a simple rectangle copper plate with a thermal conductivity of 380 W/(mK). Round heat spreaders are somewhat more material efficient, but no data regarding costs or complexity of manufacturing could be obtained.

Heat pipes or diffuser are also used in conjunction with heat spreaders. Their vapour chambers are typically made of copper with a fluid and a defined border structure. Inner structure is sintered (position independence, best performance), groove or meshed (fine fines on border). The distance, which the heat has to be transported, determines the design of the heat pipe. To avoid corrosion, the copper heat pipes are now plated with nickel or a special antioxidation treatment.

Thin finned copper coolers for server systems are state-of-the-art. Typical round pin coolers which aluminium extracted long pins are used to cool down less power consuming parts of the server mainboard, e.g. chip set. Thermally conductive polycarbonate has been introduced for heat sink applications in e.g. in field of LED but not in computer product.

Adaptive passive cooling solutions transport the heat to the chassis or over longer distances to the heat sink. Simplified spreaders of copper or aluminium are used to cool down larger, higher performing RAM modules, which create considerable heat in operation. They are by design directly connected to the air flow of the system. Jetting air (10-100m/s) against a hot surface by forcing it through a nozzle, combined with an external compressor up to 60 W/cm^2 can be cooled – approximately the limit for air cooling in general.

---

[39] http://opengpu.net/EN/attachments/154_HiPEAC2012_OpenGPU_HP_moonshot.pdf

[40] http://www.electronics-cooling.com/2014/12/carbon-based-thermal-interface-material-high-performance-cooling/
http://www.electronics-cooling.com/2014/04/polymer-based-thermal-interface-material-cools-devices-200c/
https://www.electronicmaterials.com/semiconductor/tims/

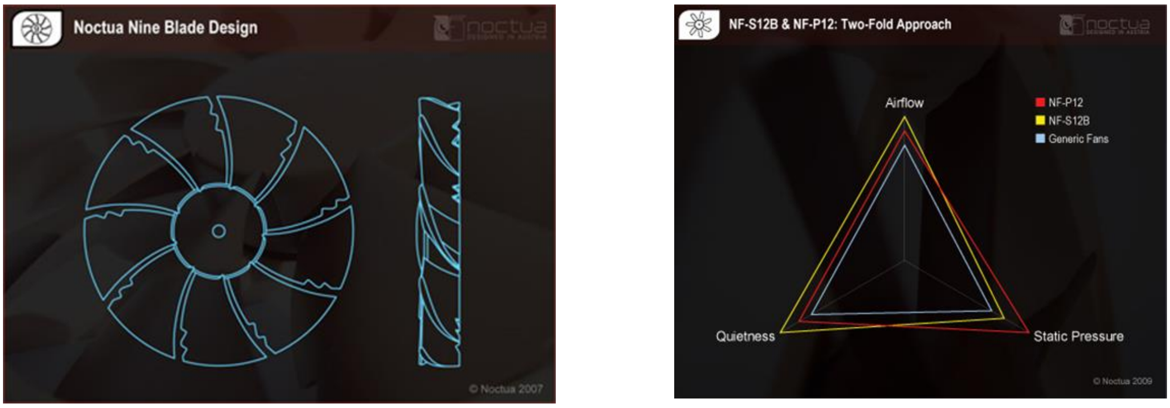
Direct touch cooling replaces air heat sinks with Heat Risers that transfer heat to the skin of server chassis where cold plates between servers transfer heat to refrigerant. This means that no fans are needed, but refrigerate needs < 61°F and cold plates between servers reduce the capacity of a 42U rack to about 35RUs.[41]

Load adaptive fans or blowers are state-of-the-art. There are some differences between the rack fans and the CPU fans which are directly mounted to CPU cooler unit. New paddle serrated designs give the opportunity to make fans more quiet, that they need less power and reaching a high volume flow:

- High volume flow (> 8CFM, pressure drop)
- Less power consumption (~ 3 Watts)
- Low noise level (<17 db(A))
- High reliability (min. 40 000 h)

Table 23 here below shows an example for a new blade design by Noctua.

Table 23: Example for a new blade design by Noctua[42]



Direct liquid cooling systems are highly specialized, require significantly more expensive infrastructure and limit the ability to access, modify, and/or repair the server systems. The direct liquid cooling systems provide both better heat transfer, concentration of the heat load to enable reuse in some situations, and depending on the actual size and configuration of the system lower energy consumption in the overall cooling system.

1.3.6. Data Centre Infrastructure Management (DCiM)

Data Centre Infrastructure Management (DCiM) is a management interface and protocol communicating between data centre resources including IT and environmental controls. DCiM requires sensors, application programming interfaces (APIs) and other software tools to monitor, connect and control IT and infrastructure equipment from multiple manufacturers and appropriate control and management algorithms to manage the data centre infrastructure. Implementation is closer in cloud data centres with homogeneous hardware systems and more difficult in enterprise data centres that utilize a range of IT equipment.

---

[41] http://www.extremetech.com/computing/197029-new-kinetic-cooler-is-a-heatsink-thats-also-a-cpu-fan
http://www.sanyodenki.com/news/newslist/20141110_SanAce60_9ga.html
http://www.storagereview.com/hp_proliant_dl380p_gen8_server_review
http://www.electronics-cooling.com/2012/09/system-level-benefits-of-augmenting-of-fan-cooled-server-with-localized-synthetic-jets/

[42] http://www.noctua.at/main.php?show=nine_blade_design&lng=en

1.4.Best Not yet Available Technology (BNAT)

1.4.1. Processor and memory development

HP currently reinvents the whole computing architecture in their “The Machine” project. The new generation computer system is announced to fuse memory and storage by implementing their *memristor* technology (memory resistor) that shall serve as a “universal memory”. The process of reading data from storage and process it in the memory becomes obsolete greatly improving performance and efficiency. Additionally the data transfers will be based on photonics optimising speed.[43] Memristors have a dynamic electrical resistance and will be capable of both processing and storage tasks while stacking in a three-dimensional array. In 2010 three nanometre memristors were planned which switch on and off within a nanosecond with a memory density of 20 GB per square centimetre. Prototype benchmarks showed very promising results regarding error rates in read and write processes.[44] In 2013 HP announced 100 TB memristor storage drives for 2018. The combination of photonics and memristors is expected to make 160 petabyte racks possible with an access time of <250 nanoseconds for any byte stored.[45] HP announced The Machine for 2017 earliest and 2020 latest.

Other trends are:

- For high performance CMOS logic systems germanium shows promising results for future technologies and might replace the role of silicon.
- Scheduled 2016: 3D-memory – RRAM technology (university of Michigan) 20 times better write performance compared to NAND-flash, 10 time lower energy consumption, 10 times the lifetime and half the die size.
- Quantum computing / spintronic are long-term research activities.

1.4.2. HDD development

There are many promising technologies being currently developed. Seagate announced the release of Heat Assisted Magnetic Recording (HAMR) HDDs end of this year. With HAMR, the recording head includes a tiny laser that heats up the medium before information is written. The technology offers higher thermal stability and could take information density to ~5 TB/inch², but there are still plenty of kinks to iron out. A further development of this technology is Microwave Assisted Magnetic Recording (MAMR). The use of microwaves overcomes some technical disadvantages connected to HAMR, yet MAMR is not expected to enter the market before 2017. A comparable enhancement on the same time schedule is Two-Dimensional Recording (TDMR) that changes the write and read heads of SMR HDDs and will most likely be implemented in Solid-State Hard Disk Drives (SSHDD). Another hot topic is Bit Patterned Media (BPM) that focuses on a special recording media by combining nano-imprinting and self-assembling block co-polymers. The fine surface of the media separates effective bits possibly increasing areal recording density. BPM is expected to enter the market in 2020 earliest.

In the nearer future, Seagate plans to offer larger hard drives – not just in terms of capacity, but in physical size. The enterprise environment has no use for storage devices that are 5mm thin – it needs Terabytes. And at the moment, adding more disks and recording heads seems the best way to grow these Terabytes.

According to a survey done at the TMRC conference 2013 HAMR, MAMR and TDMR are most likely to be implemented expecting products with at least one of those technologies by 2017. Aerial densities are expected to increase from the current 735 Gbpsi (2013) to about 10 Tbpsi. Due to the developments and potential capacity increases HDDs are expected to maintain their place in the market for at least several more years.

1.4.3. Advanced Cooling

Direct immersion cooling is a type of rack-level cooling where the complete server is immersed in a di-electric fluid for cooling. This technology drastically improves heat transfer and capture. It requires considerable chances in the rack / mounting infrastructure and currently in the alpha stage, with many uncertainties regarding the impact of direct immersion of the server on system reliability.

---

[43] HP Labs, 2015, The Machine:
http://www.hpl.hp.com/research/systems-research/themachine/

[44] Engadget, 2010, HP touts memristor development, bleak future for transistors:
http://www.engadget.com/2010/04/08/hp-touts-memristor-development-bleak-future-for-transistors/

[45] The Register, 2014, HP starts a memristor-based space program to launch ... THE MACHINE:
http://www.theregister.co.uk/2014/06/11/hp_memristor_the_machine/

2. Production, distribution and end-of-life

This section presents information about the production, distribution and end-of-life phases of the equipment in scope of the Lot 9 preparatory study. Basic information were collected through literature review and refined during the consultation process of stakeholders. The project team was provided with bills of materials for some products from which average, representative products were constructed.

These average products also serve as inputs for the environmental analysis in Task 5. Detailed information on the end-of-life treatment of server and data equipment is seldom available, for which reason several assumptions needed to be made, based on information summarized in Chapter 3 of Task 3.

2.1. Production, distribution and end-of life phases of enterprise servers

2.1.1. The production phase of a rack server

Below table shows the bill of materials of an average rack server.

Table 24: Bill of materials of an average rack server

Component	Material	Weight (g)	Component	Material	Weight (g)
Chassis	Metal Body	12 265	**PSUs (2*400W)**	Low-alloyed steel	1 027
	Plastics	348		Chromium steel	66
	Plastics	282		Brass	42
	Aluminium	249		Copper	9
	Copper	179		Zinc	7
	Electronic components	131		Aluminium	491
Fans (4)	Steel	386		High Density Polyethylene	184
	Copper	78		Polyvinylchloride (PVC)	92
	Iron based	55		Paper	50
	Plastic (PBT-GF30)	206		Electronic components	1 101
	Plastic (PCABSFR40)	21		Solder	31
	Plastic (undefined)	200		PCB	326
HDDs (4)	Steel	12	**CPU Heat Sink**	Copper	442
	Low alloyed steel	222		Steel	140
	Aluminium	1 335	**Mainboard**	Controller board	1 667
	PCB	179	**Memory**	PCB	97
				IC	38
ODD	Low alloyed steel	115	**Expansion Card**	PCB	349
	Copper	7	**Cables**	Brass	7
	Aluminium	1		Copper	81
	High Density Polyethylene (HDPE)	28		Zinc 0.166 kg	96
	Acrylonitrile-Butadiene-Styrene	12		High Density Polyethylene	104
	Polycarbonate (PC)	7		Polyvinylchloride (PVC)	145
	Electronic components (capacitors,	8		Polyurethane (PUR)	2
	Solder	2		Synthetic rubber	35
	PCB	9	**Packaging**	Cartons	3629
CPUs (2)	Copper	31		HDPE/ unspecified plastics	78
	Gold	0,4		GPPS/ Styrofoam	1 026
	PCB	21			
	IC	2			
Total weight of BC-1: 27 748 g					

The total weight of an average rack server amounts to 27.7 kg, of which the chassis represents the major share (44%). The mainboard does not include memory and CPUs, which are listed and evaluated as separate items. The packaging consists mainly of cardboard and different sorts of plastics.

2.1.2. The production phase of a blade system

The following table shows the bill of materials of a blade system which consist of an enclosure for 16 slots, containing 8 blade servers (50% populated, see Figure 16).

Table 25: Bill of materials of a blade system with 8 servers

Enclosure		
Component	**Material**	**Weight (g)**
Chassis	Steel	87 000
Fans (6)	Steel	964
	Copper	194
	Iron based	137
	Plastic (PBT-GF30)	515
	Plastic (PCABSFR40)	52
	Plastic (undefined)	499
PSUs (4)	Low-alloyed steel	4 981
	Chromium steel	319
	Brass	202
	Copper	43
	Zinc	32
	Aluminium	2 384
	High Density Polyethylene (HDPE)	894
	Polyvinylchloride (PVC)	447
	Paper	245
	Electronic components	5 343
	Solder	149
	PCB	1 581
Total weight of enclosure		**105 981 g**

Figure 16: Illustration of a blade system

8 Blade Servers					
Component	**Material**	**Weight (g)**	**Component**	**Material**	**Weight (g)**
Chassis	Steel	33 600	**HDDs (16)**	Steel	47
CPUs (16)	Copper	244		Low alloyed Steel	888
	Gold	3		Aluminium	5 341
	PCB	170		PCB	717
	IC	15	**Packaging**	Cartons	14 969
CPU Heat Sinks	Copper	1 688		HDPE/ unspecified plastics	321
	Steel	560		GPPS/ Styrofoam	4 233
Memory	PCB	773	**Mainboards**	Controller Board	6 451
	IC	307			
Total weight of 8 Blade Servers: 70 327g					
Total weight of BC-2: 176 308 g					

Data were provided by industry and refined through internet research. The enclosure amounts to around 60% of overall weight, largely dominated by the chassis. An average blade server weights around 6.3 kg (without packaging).

2.1.3. The distribution phase for a rack server and blade system

Table 26: The distribution phase of a rack server and a blade system

	Rack Server	**Blade System**
ICT or Consumer Product <15kg?	NO	NO
Installed Appliance?	YES	YES
Volume of packaged final product in m^3	0.04	0.51

Information on the volume of packaged final products has been obtained from documents received directly from the stakeholders or from product information available on stakeholders' websites.

2.1.4. End-of-life phase of an average rack server and blade system

Only few detailed data on the end-of-life treatment of enterprise servers is available, but from literature review and stakeholder consultations it became clear that the re-use rate of enterprise servers is relatively high. The major companies have all asset recovery services and/or leasing programmes in place that record resale rates of around 80%. However, these programmes do not cover the entire market and other sources discussed in Task 3 show lower figures. For this reason an average reuse rate of 50% has been retained for servers.

As far as the end-of-life of plastics is concerned, only a minor share goes to landfill, while the biggest part is incinerated with heat recovery. The EoL values for metals are fixed in the EcoReport and cannot be adjusted manually. Stakeholders have reported that almost all electronic parts are recycled. However, since the electronics category in the EcoReport also contains multi-material parts like printed circuit boards, a share of 12.5% is considered to be incinerated with heat recovery and 0.5% to be landfilled. The rates for the miscellaneous category are taken from the standard configuration in the MEErP tool.

Publicly available studies on end-of-life treatment of servers are not available at the detailed level required by the EcoReport tool. Therefore, assumptions needed to be made which are in line with the latest available data (e.g. treatment of professional IT and telecommunications equipment in France, see Task 3) and further inputs received from stakeholders. The following table shows the overall assumptions on the end-of-life phases for server and blade systems.

Table 27: End-of-life phase of enterprise servers

	Plastics	Metals	Electronics	Misc.
Re-Use	50%			
Material Recycling	5%	45%	36.5%	43%
Heat Recovery	44%	0%	12.5%	1%
Non-recovery incineration	0.5%	0%	0.5%	5%
Landfill	0.5%	5%[46]	0.5%	1%
Total	100%	100%	100%	100%

2.2.Production, distribution and end-of-life phase of enterprise storage

2.2.1. The production phase of a storage system

The table below shows the bill of materials of a storage system, a virtual storage unit with ½ controller, two disk array enclosures (DAEs) and a storage media mix consisting of 16.66 (+2.69 spare) 3.5 inch HDDs, 12.07 (+1.95 spare) 2.5 inch HDDs and 2.87 SSDs, representing a product belonging to the SNIA 2-3 taxonomy. The cabin was not considered. More detailed inputs of the EcoReport tool are presented in the Annex. The figure on the right serves as an illustration:

The storage media mix amounts to a capacity of around 40 TB.

Figure 17: Illustration of two DAEs

---

[46] This value cannot be changed in the EcoReport tool.

Table 28: Bill of materials of 0.5 controller

Controller (1/2)					
Component	Material	Weight (g)	**Component**	Material	Weight (g)
Controller	Steel	7 450	**PSU controller**	Mainboard	825
	Stainless steel	1 680		Cables	20
	Aluminum sheet	287		Chassis and bulk material	889
	Copper	520	**PSU Fans**	Steel	110
	ABS	510		Copper	65
	PET	39		Iron based	13
	HDPE	87		Nylon 6	9
	PP	18		PC	35
	PC	31		ABS	19
	Nylon 6	5	**Packaging**	Cartons	3 629
	PVC	85		HDPE/ unspecified plastics	78
	Other plastics	12		GPPS/ Styrofoam	1 026
	Printed circuit board	577			
Total weight of controller: 18 020 g					

Table 29 : Bill of materials for two Disc Array Enclosures (DAEs)

Disc Array Enclosures (2)					
Chassis	PC	406	**Fans in PSUs (8)**	Steel	563
	ABS	92		Copper	332
	Steel sheet part	15 374		Iron based	66
	Zinc Part	298		Nylon 6	47
	Steel Machined Part	3		PC	177
PSUs in DAEs (4)	Mainboard	4 217		ABS	95
	Cables	104	**Controller cards (4)**	Electronics	2 308
	Chassis and bulk materials	4 546	**Mid plane boards (2)**	Electronics	920
Total weight of DAEs: 29 554 g					

Table 30 : Bill of materials for an average storage media mix

Storage Media Mix (40.1 TB)					
3.5 HDD (19.35)	Steel	58	**2.5 HDD (14.01)**	Steel	278
	Low Alloyed Steel	1 103		Low Alloyed Steel	211
	Aluminium	6 637		Aluminium	2 562
	PCB	890		PCB	123
				ABS	7
SSDs (2.86)	Electronic components	172			
	IC	5			
Total weight of Storage Media Mix: 12 050 g					
Total weight of average storage system (SNIA Online 2/3): 59 623 g					

The storage system consists of 2 DAEs which together make up for almost 50% of the total weight. 30% of the weight is coming from the controller and the rest can be attributed to storage media (20%). As can be seen from above table, aluminium constitutes the largest share of the total weight of the storage media mix (76%).

A study published by Carnegie Mellon University in 2007 shows that annual disk replacement rates typically exceed 1%, with 2-4% being common and which can go up to 13% for some systems[47]. Further information provided directly by stakeholders suggests that lifetime replacement units are relatively high for HDDs. The EcoReport tool includes a parameter for spare parts, which is fixed to 1% and cannot be adjusted. For this reason, spare HDDs have been added to the bill of materials in order to assure the proper functioning of the storage system over its lifetime and to take into account a higher replacement rate.

---

[47] https://www.usenix.org/legacy/events/fast07/tech/schroeder.html

2.2.2. Distribution phase of the storage system

Table 31: The distribution phase for storage systems

	Storage System
ICT or Consumer Product <15kg?	NO
Installed Appliance?	YES
Volume of packaged final product in m^3	0.072

Information on the volume of packaged final products has been obtained from documents received directly from the stakeholders or from product information available on stakeholders’ websites.

2.2.3. End-of-life phase of the storage system

For the end-of-life phase, information from stakeholders suggests that the re-use rate is lower for storage equipment than for enterprise servers, ranging around 20-30%. For this reason an average value of 25% has been retained. For material recycling, heat recovery, incineration and landfill the same assumptions hold as for enterprise servers described above. The following table shows a summary:

Table 32: The end-of-life phase of storage systems

	Plastics	Metals	Electronics	Misc.
Re-Use	25%			
Material Recycling	5%	70%	50%	68%
Heat Recovery	69%	0%	24%	1%
Non-recovery incineration	0.5%	0%	0.5%	5%
Landfill	0.5%	5%[48]	0.5%	1%
Total	100%	100%	100%	100%

---

[48] This value cannot be changed in the EcoReport tool.

3.Recommendations

3.1. Refined product scope from the technical perspective

No recommendation for amending the general product scope including server and storage systems can be made at this point. However, the ENTR Lot 9 study recognizes the fact that certain eco-design aspects such as amongst others power management may interfere with the following types of products:

- Products with higher reliability, availability, and serviceability (RAS) as well as failover capacities (resilient);
- Products for high performance computing (HPC) which are designed for parallel computing and not for virtualization;
- Large mainframes and high energy density servers systems featuring e.g. novel fluidic cooling systems as well as specialized racks and power supplies;
- Larger / higher performing embedded server and storage systems;
- Multifunctional products platforms which fit the description of a server or storage products depending on the actual configuration.

3.2.Barriers and opportunities for Ecodesign from a technical perspective

From a technical perspective, there are no barriers for eco-design for servers and storage systems. However, the ENTR Lot 9 study recognizes the fact that certain eco-design options may interfere with the real life operation of the products under specific service level agreements (SLA).

There are other hardware and software aspects, which are determined by the customer that may limit the applicability of eco-design options. This includes amongst others a required form factor by the customer (e.g. a certain rack unit) or licences for necessary software (e.g. a certain operation system, middleware, or application software).

The ENTR Lot 9 study also recognizes the fact that certain best performing technologies such as IC-based storage (SSD) are considerably more expensive than conventional technologies and that the production capacity is globally limited.

Form a technical perspective there are on the other hand considerable opportunities for eco-design.

Energy efficiency improvements have occurred from a variety of sources:

1. Improved efficiency of power system components. This includes voltage regulators and power supplies.
2. Improved management firmware and software including improved power management capabilities, hypervisors (virtualization for servers), COMs for storage systems, and other improvements.
3. Increased workload capacity on CPU, memory, storage, and I/O which enables more work to be done per unit of energy consumed. This is the result of miniaturization and architectural innovations.

Eco-design requirements such as low power consumption will show a positive impact on the continuous improvement efforts of the semiconductor industry. This could also result in shorter product cycles and the adoption of new products within shorter time periods.

3.3.Typical design cycle

The typical design cycle for this product and thus an approximately appropriate timing of measures is 2 to 3 years.

Annex

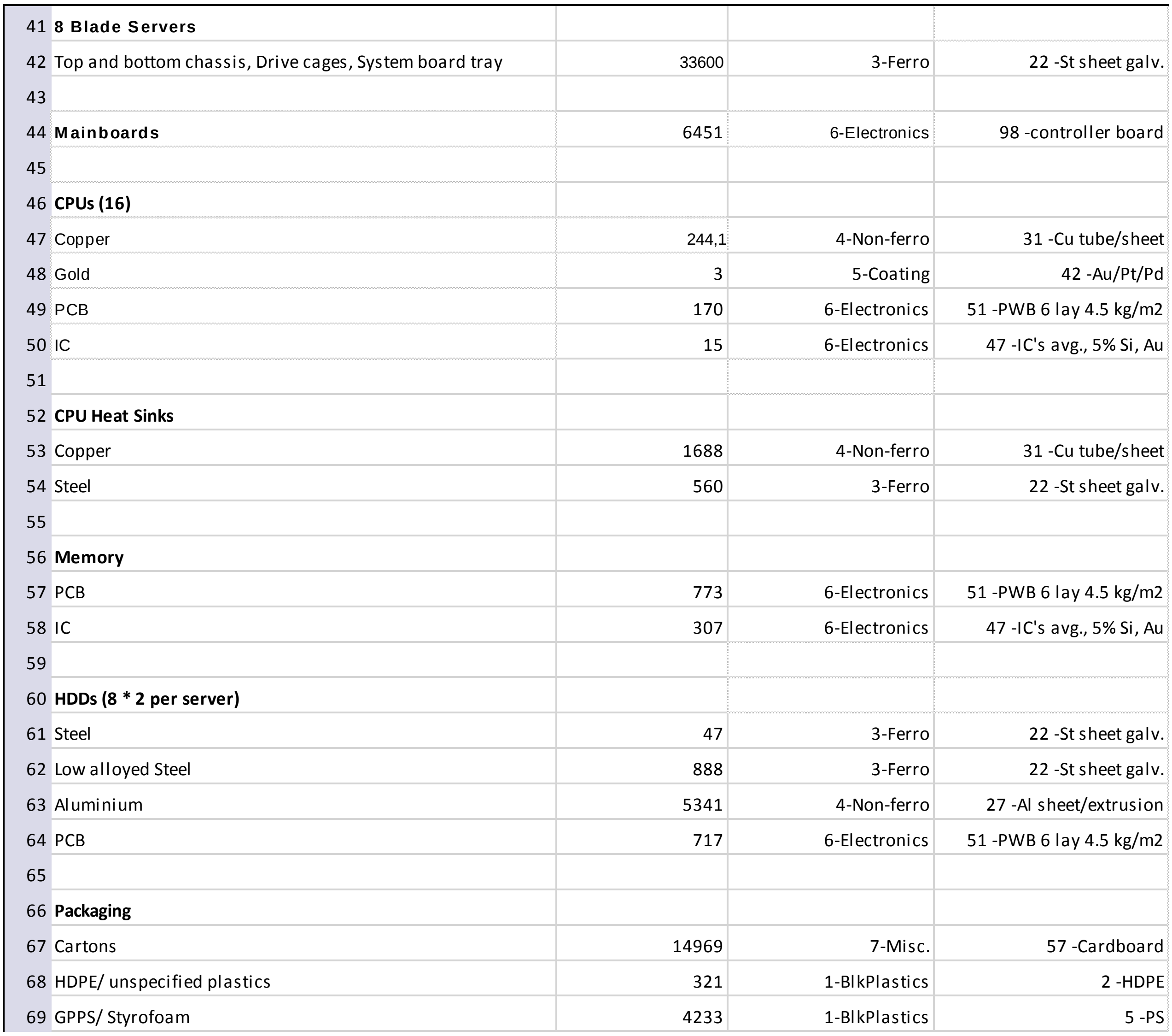
Table 33: Bill of Materials - Rack Server

1	**Chassis**			
2	Metal Body	12265	3-Ferro	22 -St sheet galv.
3	Plastics	348	1-BlkPlastics	11 -ABS
4	Plastics	282	2-TecPlastics	13 -PC
5	Aluminium	249	4-Non-ferro	28 -Al diecast
6	Copper	179	4-Non-ferro	32 -CuZn38 cast
7	Electronic components (capacitors, inductors, printed circuit, resistors,transformers, transistors)	131	6-Electronics	98 -controller board
8				
9	**Fans (4)**			
10	Steel	386,0	3-Ferro	22 -St sheet galv.
11	Copper	78,0	4-Non-ferro	31 -Cu tube/sheet
12	Iron based	55,0	3-Ferro	24 -Cast iron
13	Plastic (PBT-GF30)	206,0	1-BlkPlastics	5 -PS
14	Plastic (PCABSFR40)	21,0	2-TecPlastics	13 -PC
15	Plastic (undefined)	200	1-BlkPlastics	11 -ABS
16				
17	**HDDs (4)**			
18	Steel	12	3-Ferro	22 -St sheet galv.
19	Low alloyed steel	222	3-Ferro	22 -St sheet galv.
20	Aluminium	1335	4-Non-ferro	27 -Al sheet/extrusion
21	PCB	179	6-Electronics	51 -PWB 6 lay 4.5 kg/m2
22				
23				
24	**ODD**			
25	Low alloyed steel	115	3-Ferro	22 -St sheet galv.
26	Copper	7	4-Non-ferro	31 -Cu tube/sheet
27	Aluminium	1	4-Non-ferro	27 -Al sheet/extrusion
28	High Density Polyethylene (HDPE)	28	1-BlkPlastics	2 -HDPE
29	Acrylonitrile-Butadiene-Styrene (ABS)	12	1-BlkPlastics	11 -ABS
30	Polycarbonate (PC)	7	2-TecPlastics	13 -PC
31	Electronic components (capacitors, inductors, printed circuit, resistors,transformers, transistors)	8	6-Electronics	98 -controller board
32	Solder	2	6-Electronics	53 -Solder SnAg4Cu0.5
33	PCB	9	6-Electronics	51 -PWB 6 lay 4.5 kg/m2

41	**Mainboard**	1667,0	6-Electronics	98 -controller board
42				
43	**PSUs (2*400W)**			
44	Low-alloyed steel	1027	3-Ferro	22 -St sheet galv.
45	Chromium steel	66	3-Ferro	26 -Stainless 18/8 coil
46	Brass	42	4-Non-ferro	33 -ZnAl4 cast
47	Copper	9	4-Non-ferro	30 -Cu wire
48	Zinc	7	4-Non-ferro	32 -CuZn38 cast
49	Aluminium	491	4-Non-ferro	27 -Al sheet/extrusion
50	High Density Polyethylene (HDPE)	184	1-BlkPlastics	2 -HDPE
51	Polyvinylchloride (PVC)	92	1-BlkPlastics	8 -PVC
52	Paper	50	7-Misc.	58 -Office paper
53	Electronic components (capacitors, inductors, printed circuit, resistors,transformers, transistors)	1101	6-Electronics	98 -controller board
54	Solder	31	6-Electronics	53 -Solder SnAg4Cu0.5
55	PCB	326	6-Electronics	51 -PWB 6 lay 4.5 kg/m2
56				
57	**PCB (expansion card/other)**			
58	PCB	349	6-Electronics	51 -PWB 6 lay 4.5 kg/m2
59				
60	**Cables**			
61	Brass	7	4-Non-ferro	33 -ZnAl4 cast
62	Copper	81	4-Non-ferro	30 -Cu wire
63	Zinc 0.166 kg	96	4-Non-ferro	32 -CuZn38 cast
64	High Density Polyethylene (HDPE)	104	1-BlkPlastics	2 -HDPE
65	Polyvinylchloride (PVC)	145	1-BlkPlastics	8 -PVC
66	Polyurethane (PUR)	2	2-TecPlastics	17 -Flex PUR
67	Synthetic rubber	35	2-TecPlastics	16 -Rigid PUR
68				
69	**CPUs (2)**			
70	Copper	30,5	4-Non-ferro	31 -Cu tube/sheet
71	Gold	0,4	5-Coating	42 -Au/Pt/Pd
72	PCB	21,2	6-Electronics	52 -PWB 6 lay 2 kg/m2
73	IC	1,9	6-Electronics	47 -IC's avg., 5% Si, Au
74				
75	**CPU Heat Sinks**			
76	Copper	442	4-Non-ferro	31 -Cu tube/sheet
77	Steel	140	3-Ferro	22 -St sheet galv.
78				
79	**Memory**			
80	PCB	97	6-Electronics	51 -PWB 6 lay 4.5 kg/m2
81	IC	38	6-Electronics	47 -IC's avg., 5% Si, Au
82				
83	**Packaging**			
84	Cartons	3629	7-Misc.	57 -Cardboard
85	HDPE/ unspecified plastics	78	1-BlkPlastics	2 -HDPE
86	GPPS/ Styrofoam	1026	1-BlkPlastics	5 -PS

Table 34: Bill of Materials - Blade System

1	**Enclosure**			
2	**Chassis**	87000	3-Ferro	22 -St sheet galv.
3	**4 x PSU**			
4	Low-alloyed steel	4981	3-Ferro	22 -St sheet galv.
5	Chromium steel	319	3-Ferro	26 -Stainless 18/8 coil
6	Brass	202	4-Non-ferro	33 -ZnAl4 cast
7	Copper	43	4-Non-ferro	30 -Cu wire
8	Zinc	32	4-Non-ferro	32 -CuZn38 cast
9	Aluminium	2384	4-Non-ferro	27 -Al sheet/extrusion
10	High Density Polyethylene (HDPE)	894	1-BlkPlastics	2 -HDPE
11	Polyvinylchloride (PVC)	447	1-BlkPlastics	8 -PVC
12	Paper	245	7-Misc.	57 -Cardboard
13	Electronic components (capacitors, diodes, inductors, printed circuit, resistors, transformers, transistors, connectors)	5343	6-Electronics	98 -controller board
14	Solder	149	6-Electronics	53 -Solder SnAg4Cu0.5
15	PCB	1581	6-Electronics	51-PWB 6 lay 4.5 kg/m2
16				
17	**6 Fans**			
18	Steel	964	3-Ferro	22 -St sheet galv.
19	Copper	194	4-Non-ferro	31-Cu tube/sheet
20	Iron based	137	3-Ferro	24 -Cast iron
21	Plastic (PBT-GF30)	515	1-BlkPlastics	5 -PS
22	Plastic (PCABSFR40)	52	2-TecPlastics	13 -PC
23	Plastic (undefined)	499	1-BlkPlastics	11 -ABS
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				



41	**8 Blade Servers**			
42	Top and bottom chassis, Drive cages, System board tray	33600	3-Ferro	22 -St sheet galv.
43				
44	**Mainboards**	6451	6-Electronics	98 -controller board
45				
46	**CPUs (16)**			
47	Copper	244,1	4-Non-ferro	31 -Cu tube/sheet
48	Gold	3	5-Coating	42 -Au/Pt/Pd
49	PCB	170	6-Electronics	51 -PWB 6 lay 4.5 kg/m2
50	IC	15	6-Electronics	47 -IC's avg., 5% Si, Au
51				
52	**CPU Heat Sinks**			
53	Copper	1688	4-Non-ferro	31 -Cu tube/sheet
54	Steel	560	3-Ferro	22 -St sheet galv.
55				
56	**Memory**			
57	PCB	773	6-Electronics	51 -PWB 6 lay 4.5 kg/m2
58	IC	307	6-Electronics	47 -IC's avg., 5% Si, Au
59				
60	**HDDs (8 * 2 per server)**			
61	Steel	47	3-Ferro	22 -St sheet galv.
62	Low alloyed Steel	888	3-Ferro	22 -St sheet galv.
63	Aluminium	5341	4-Non-ferro	27 -Al sheet/extrusion
64	PCB	717	6-Electronics	51 -PWB 6 lay 4.5 kg/m2
65				
66	**Packaging**			
67	Cartons	14969	7-Misc.	57 -Cardboard
68	HDPE/ unspecified plastics	321	1-BlkPlastics	2 -HDPE
69	GPPS/ Styrofoam	4233	1-BlkPlastics	5 -PS

Table 35: Bill of Materials - Storage Unit

1	**0,5 x Controller**			
2	Steel	7 450	3-Ferro	22 -St sheet galv.
3	Stainless steel	1 680	3-Ferro	26 -Stainless 18/8 coil
4	Aluminum sheet	287	4-Non-ferro	27 -Al sheet/extrusion
5	Copper	520	4-Non-ferro	31 -Cu tube/sheet
6	ABS	510	1-BlkPlastics	11 -ABS
7	PET	39	1-BlkPlastics	10 -PET
8	HDPE	87	1-BlkPlastics	2 -HDPE
9	PP	18	1-BlkPlastics	4 -PP
10	PC	31	2-TecPlastics	13 -PC
11	Nylon 6	5	2-TecPlastics	12 -PA 6
12	PVC	85	1-BlkPlastics	8 -PVC
13	Other plastics	12	1-BlkPlastics	11 -ABS
14	Printed circuit board	577	6-Electronics	98 -controller board
15	**PSU of controller (1x in total)**			
16	Mainboard	825	6-Electronics	98 -controller board
17	Cables	20	4-Non-ferro	30 -Cu wire
18	Chassis and bulk material	889	3-Ferro	22 -St sheet galv.
19	**PSU Fans (2x in total)**			
20	Steel	110	3-Ferro	22 -St sheet galv.
21	Copper	65	4-Non-ferro	31 -Cu tube/sheet
22	Iron based	13	3-Ferro	24 -Cast iron
23	Nylon 6	9	2-TecPlastics	12 -PA 6
24	PC	35	2-TecPlastics	13 -PC
25	ABS	19	1-BlkPlastics	11 -ABS
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				
39				
40				

41	**2 x DAE**			
42	**Chassis**			
43	PC	406,4	2-TecPlastics	13 -PC
44	ABS	92,2	1-BlkPlastics	11 -ABS
45	Steel sheet part	15374,0	3-Ferro	22 -St sheet galv.
46	Zinc Part	298,6	4-Non-ferro	34 -MgZn5 cast
47	Steel Machined Part	3,4	3-Ferro	23 -St tube/profile
48				
49				
50	**PSUs in DAEs (4 in total)**			
51	Mainboard	4217,2	6-Electronics	98 -controller board
52	Cables	104,4	4-Non-ferro	30 -Cu wire
53	Chassis and bulk materials	4546,2	3-Ferro	22 -St sheet galv.
54				
55	**Fans in PSUs of DAEs (8 in total)**			
56	Steel	563,04	3-Ferro	22 -St sheet galv.
57	Copper	332,928	4-Non-ferro	31 -Cu tube/sheet
58	Iron based	66,912	3-Ferro	24 -Cast iron
59	Nylon 6	47,328	2-TecPlastics	12 -PA 6
60	PC	177,888	2-TecPlastics	13 -PC
61	ABS	95,064	1-BlkPlastics	11 -ABS
62				
63	**4 x Controller card**	2308,2	6-Electronics	98 -controller board
64	**2 x Mid plane board**	920,36	6-Electronics	98 -controller board
65				
66	**3.5 HDD (19.35 in total)**			
67	Steel	58	3-Ferro	24 -Cast iron
68	Low Alloyed Steel	1103	3-Ferro	22 -St sheet galv.
69	Aluminium	6637	4-Non-ferro	27 -Al sheet/extrusion
70	PCB	890	6-Electronics	51 -PWB 6 lay 4.5 kg/m2
71				
72				
73	**2.5 HDD (14.01 in total)**			
74	Steel	278,41	3-Ferro	23 -St tube/profile
75	Low Alloyed Steel	211,2	3-Ferro	22 -St sheet galv.
76	Aluminium	2562,57	4-Non-ferro	27 -Al sheet/extrusion
77	PCB	123,84	6-Electronics	51 -PWB 6 lay 4.5 kg/m2
78	ABS	7,15	1-BlkPlastics	11 -ABS
79				
80	**SSDs (2.86 in total)**			
81	Electronic components (capacitors, inductors, printed circuit, resistors,transformers, transistors)	172,4	6-Electronics	98 -controller board
82	IC	5,75	6-Electronics	47 -IC's avg., 5% Si, Au
83				
84	**Packaging**			
85	Cartons	3629	7-Misc.	57 -Cardboard
86	HDPE/ unspecified plastics	78	1-BlkPlastics	2 -HDPE
87	GPPS/ Styrofoam	1026	1-BlkPlastics	5 -PS

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee, and its network of member firms, each of which is a legally separate and independent entity. Please see www.deloitte.com/about for a detailed description of the legal structure of Deloitte Touche Tohmatsu Limited and its member firms.

Deloitte provides audit, tax, consulting, and financial advisory services to public and private clients spanning multiple industries. With a globally connected network of member firms in more than 150 countries, Deloitte brings world-class capabilities and high-quality service to clients, delivering the insights they need to address their most complex business challenges. Deloitte has in the region of 200,000 professionals, all committed to becoming the standard of excellence.

This communication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms, or their related entities (collectively, the “Deloitte Network”) is, by means of this publication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser. No entity in the Deloitte Network shall be responsible for any loss whatsoever sustained by any person who relies on this communication.

© 2014 Deloitte SA. Member of Deloitte Touche Tohmatsu Limited