# Energy consumption of data centers

# Spotřeba datových center

Bc. Jiří Zlámal

Master Thesis
2011

Univerzita Tomáše Bati ve Zlíně
Fakulta aplikované informatiky

*** nascannované zadání str. 1 ***

***   nascannované zadání str. 2   ***

## ABSTRACT

This thesis describes the most used methods for better data center efficiency for the largest contributors of data center inefficiency. To understand the issues of effectivness and ways of masuring it, the used efficiency parameters are explained. Tier levels are described together with designs of data center constructions. Two parts of this thesis are dedicated to energy management in the network and server infrastructure and their possibilities. Subsequently comparison of traditional and modular (container and hybrid designs) centers is developed. Comparison of older and recent technologies in data center construction is revealed in ralation with performance and efficiency.

Keywords: data centers, efficiency, energy consumption and management, traditional and modular data centers

## ABSTRAKT

V této práci jsou shrnuty nejpoužívanější metody zvýšení efektivity provozu datových center u největších přispěvatelů spotřeby energie. Pro pochopení problematiky efektivity a způsobů jejího měření jsou vysvětleny používané parametry. Úrovně Tier klasifikace jsou vysvětleny společně s konstrukcemi datových center. Dvě části této práce jsou věnovány řízení spotřeby a jejich možnostech v síťové a serverové infrasktruktuře. Následně je zpracováno srovnání tradiční a modulárních (kontejnerových a hybridních) datových center. Srovnání starších a novějších technologií je provedeno ve spojení s výkonem a efektivitou těchto systémů.

Klíčová slova: datová centra, efektivita, spotřeba energií, tradiční a modulární datová centra

## Acknowledgment:

I want to thank my supervisor professor Karel Vlček for guidance and technical advices. I would also like to thank to Milan Šimčík and Ing. Antonín Křížek for technical advices, help with procuring materials and last but not least the opportunity to see the data center technology solutions in real application. Great thanks belongs also to my family and friends, without whose support this work could not have incurred.

**Prohlašuji, že**

- beru na vědomí, že odevzdáním diplomové/bakalářské práce souhlasím se zveřejněním své práce podle zákona č. 111/1998 Sb. o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších právních předpisů, bez ohledu na výsledek obhajoby;
- beru na vědomí, že diplomová/bakalářská práce bude uložena v elektronické podobě v univerzitním informačním systému dostupná k prezenčnímu nahlédnutí, že jeden výtisk diplomové/bakalářské práce bude uložen v příruční knihovně Fakulty aplikované informatiky Univerzity Tomáše Bati ve Zlíně a jeden výtisk bude uložen u vedoucího práce;
- byl/a jsem seznámen/a s tím, že na moji diplomovou/bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, zejm. § 35 odst. 3;
- beru na vědomí, že podle § 60 odst. 1 autorského zákona má UTB ve Zlíně právo na uzavření licenční smlouvy o užití školního díla v rozsahu § 12 odst. 4 autorského zákona;
- beru na vědomí, že podle § 60 odst. 2 a 3 autorského zákona mohu užít své dílo – diplomovou/bakalářskou práci nebo poskytnout licenci k jejímu využití jen s předchozím písemným souhlasem Univerzity Tomáše Bati ve Zlíně, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly Univerzitou Tomáše Bati ve Zlíně na vytvoření díla vynaloženy (až do jejich skutečné výše);
- beru na vědomí, že pokud bylo k vypracování diplomové/bakalářské práce využito softwaru poskytnutého Univerzitou Tomáše Bati ve Zlíně nebo jinými subjekty pouze ke studijním a výzkumným účelům (tedy pouze k nekomerčnímu využití), nelze výsledky diplomové/bakalářské práce využít ke komerčním účelům;
- beru na vědomí, že pokud je výstupem diplomové/bakalářské práce jakýkoliv softwarový produkt, považují se za součást práce rovněž i zdrojové kódy, popř. soubory, ze kterých se projekt skládá. Neodevzdání této součásti může být důvodem k neobhájení práce.

**Prohlašuji,**

- že jsem na diplomové práci pracoval samostatně a použitou literaturu jsem citoval. V případě publikace výsledků budu uveden jako spoluautor.
- že odevzdaná verze diplomové práce a verze elektronická nahraná do IS/STAG jsou totožné.

Ve Zlíně ……………………

podpis diplomanta

## CONTENTS

# 1 INTRODUCTION

With continous expansion of IT technologies in different fields of human activities the demands for data center number and performance are growing rapidly. Larger data centers are required by more companies and considering the increase of energy costs, effectivness became one of the most important attributes of the data center. Also environmental concerns became an important issue in business good reputation. Those are the main reasons why companies are getting more interested in the „green“ solutions. It might seem odd, but mostly the greenest solution is also the simplest and the most effective one.

Other reason why should be data center owners concerned about the power consumption is the fact that many data centers are running out of power and space capacity. More servers cannot be added because of their power, cooling and space limits. These data centers can be optimized and free capacity can be released.

# I. THEORETICAL PART

# 2 DATA CENTERS

## 2.1 History

Data centers as we know them have started their existence in the beginning of 1990s with the growth of the Internet. However the actual history of data centers is connected with the appearance of the very first computers.

ENIAC is the first to be considered as a data center. It was composed of 18 000 vacuum tubes, its weight was about 30 tons and it needed 3ms to solve one dividing. This data center was finished in 1946 and it was used strictly for military purposes till 1955 by the US army. This was a very unreliable data center. Almost half of its operational time was needed for repairs.

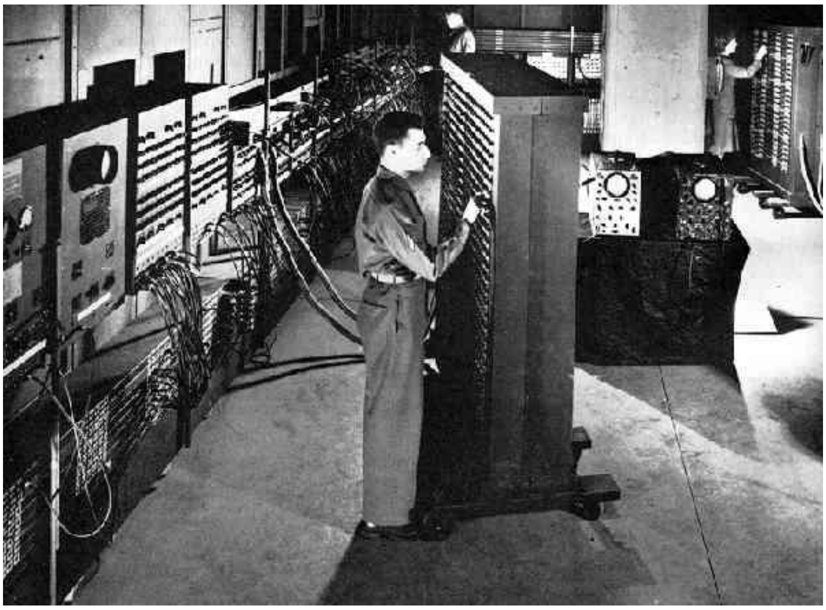


Figure 1 – ENIAC [source: http://www.dvojklik.cz/wp-content/uploads/2010/08/ENIAC02.jpg]

Considerations of security, cooling and cable organization of this data center led to creation of some standards in the data center construction.

In 1959 the very first civil use of data center was performed by data center ERMA for automated production control. In 1965 data center DEC PDP-8 was built, which is considered as the beginning of reducing of data center size.

In the 1980s the computer industry became widely used thanks to the microcomputers. Many computers connected together without specific installation guidelines were built.

Soon it became clear, that organization of all the information and of whole data centers was very difficult and new solutions were required. In the 1990s client-server connection became a standard and this is considered as the beginning of the data centers as we know them. Hierarchical design, dedicated rooms and network cabling standards were born. With the expansion of Internet, companies realized the importance of establishing fast and reliable Internet connectivity. Large data facilities were deployed and terms as “Internet data centers” were used for the first time. These large facilities became the leaders in using new technologies and operating practices within the data center industry. Even nowadays the largest data centers are usually using unique state of the art solutions, which may become standard solutions for smaller data centers in the future.

The need of standards led to establishment of consortiums such as The Uptime Institute. They are providing new solutions and standards for maximizing the availability and efficiency based on the best practices. The Uptime Institute is providing the information about data center cooling and electrical systems since 1993 and the scope is continuously enlarging based on new trends.

### Data center construction

New technologies and practices were required for the expansion of the IT systems. That is how standardized cabinets (Racks), raised floors, and standardized components came to the market. With the use of racks came new trend of separation of the IT components such as storage or networking equipment. Indisputable advantage of using racks is clarity of the whole system, easier installation and much higher efficiency of cooling systems. Use of the raised floors and lowered ceilings became very popular in the data center construction industry. Power distribution, network cables and air distribution are integrated in the raised floors or lowered ceilings, being protected from possible damage. Proper and well organized structure is shown in Figure 2.

Figure 2- Raised floor in server room [source:http://568systems.com/pic/2DataCenterRaisedFloor.jpg]

The raised floors are also used for seismic safety of the data center as it is shown in Figure 3.

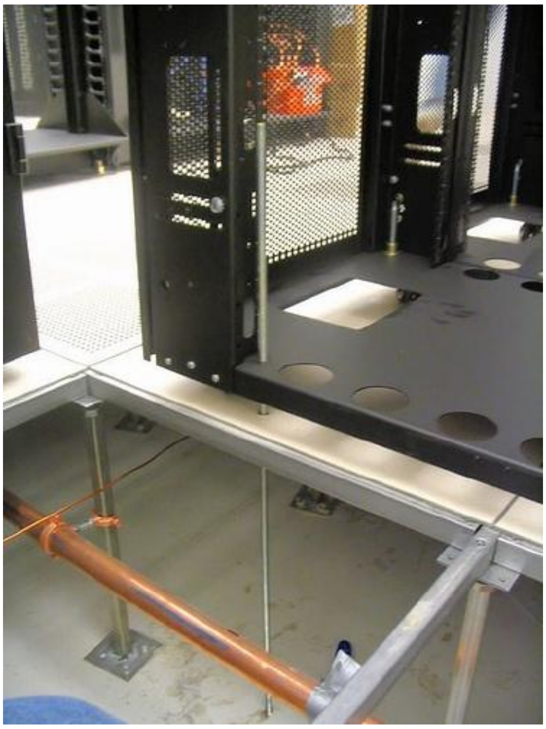


Figure 3 – Seismic safety for racks [source: Data centers lectures, Instituto Superior Technico Lisboa]

# 3 ENERGY MANAGEMENT

One of the most important problems in the global scale is the energy efficiency. With the IT industry interfering with more fields of human work, the demands for the data centers are growing. Data centers consume from 10 to 200 times more energy per square feet then a typical office and if the reduction of consumption would be overlooked, then the costs for running server could become equal or even exceed the costs of acquisition.

Table 1 shows the energy consumption of all USA data centers according to the EPA (Environmental Protection Agency) Data Center Report to Congress from year 2007.

| Sector | Energy consumption [kWh] | Costs [USD] |
|---|---|---|
| **Private** | 61 billion | 4,5 billion |
| **Federal** | 6 billion | 450 millions |

Table 1 – Energy consumption of all USA data centers in 2007

Anticipated growth till year 2011 is up to 100 billion kWh, which will be equal to 2.5% total U.S. energy consumption and costs will be around 7.4 billion USD. Given these facts the DOE (Department of Energy) came with a plan which should result in 10% energy savings in the U.S. data center field between the years 2006-2010.

Several years ago the main characteristics of the data centers were reliability and availability. Given the rise of the energy costs and significant increase of data centers performance, efficiency became one of the most important characteristic as well. “Going green” became a standard in data center construction field.

Organizations such as The Green Grid were established and new terms were born. The most important and widely used terms connected to the energy efficiency are described below.

## 3.1 PUE

This shortcut stands for Power Usage Effectiveness and has been developed by consortium The Green Grid. It is a ratio of total power used by data canter to the power consumed by the IT equipment. An ideal PUE is 1.0, which means that for each 1W of IT power, there is no other power needed for the support systems such as cooling, lightning, UPSs etc. PUE can be also expressed as inversion of DCiE.

$$PUE = \frac{\text{Total facility power}}{\text{IT equipment power}}$$

An average data center has PUE between 2.5 and 2.0, but there are the state of the art data centers with much lower numbers. Google claims that at least one facility is returning PUE of 1.11 and PUE of all their facilities is 1.16.

## 3.2 DCiE

Means Data Center infrastructure Efficiency and it is a percentage value given by the ration of IT equipment power and total facility power. DCiE was developed by The Green Grid.

$$DCiE = \frac{\text{IT equipment power}}{\text{Total facility power}}$$

## 3.3 ERE

ERE is a shortcut, meaning Energy Reuse Effectiveness. This means for example using the heat produced in the data center for heating or other purposes. This re-using of energy cannot be included in PUE, because the definition of PUE does not allow it. ERE can be calculated as follows.

$$\text{ERE} = \frac{\text{Total Energy-Reused Energy}}{\text{IT energy}}$$

## 3.4 CUE

This shortcut representing Carbon Usage Effectiveness was developed by The Green Grid for measuring the negative impacts of data centers on resources and the environment. CUE cores the operation of the data center, but does not cover the full environmental burden like impact of manufacturing all the equipment. The CUE can be calculated as follows.

$$\text{CUE} = \frac{\text{Total}\,\text{CO}_2\ \text{ emissions caused by the Total Data Center Energy}}{\text{IT Equipment Energy}}$$

CUE has its own units which are kilograms of carbon dioxide (kg$CO_2$eq) per kilowatt-hour (kWh).

## 3.5 WUE

Stands for Water Usage Effectiveness and it is a part it the “xUE” Green Grid family. It is a ratio of annual water usage to IT equipment energy and) unlike the PUE has units - L/kWh (liters/kilowatt-hour). The range of WUE is between zero and infinity, where zero represents the best case and infinity represents the worst possible one.

Reduction of water use can result in more energy costs. As it takes significant amount of water to generate electricity, the higher water usage will be just transferred from the site to another. Deciding between higher energy consumption or water usage depends mostly on many factors like location, business strategy and financial considerations.

Therefore there are two metrics of WUE.

- WUE, which considers water used on-site for operating the data center (evaporated water for cooling or energy production or humidification of the data center)
- $WUE_{source}$, which considers water used on-site and water used off-site for the production of power used on-site.

## 3.6 Practical importance of energy efficiency

Using the energy efficiency calculator on www.42u.com immediately shows the results of higher efficiency of a data center.

Input values are current PUE value, the planned PUE value for better efficiency, cost of electricity and total IT load. Let’s consider these values which are similar to the examined data center in this thesis. The results convince everyone that going green is the right way.

- Current PUE: 1.9
- Planned PUE: 1.5
- Total IT load: 300kW
- Cost per kWh: 0.0964 USD

Annual data center efficiency savings are:

| kW of Electricity | Electricity Costs | $CO_2$ Emissions |
|---|---|---|
| 1 051 200 kW | 101 336 USD | 634 Tons |

Table 2 – Annual Data Center Efficiency Savings

Regarding the EPA Data Center Report to Congress from year 2007, the efficiency of average enterprise data center from year 2006 was 2.0 or higher. Expected improvements based on more efficient equipment in 2011 should result in PUE of 1.9. The state of the art data centers should operate on PUE of 1.2.  Summarization of the expected scenarios is shown in Table 3.

| Scenario | PUE |
|---|---|
| **Typical Enterprise** | 1.9 |
| **Improved technologies** | 1.7 |
| **Best practices** | 1.3 |
| **State of the art** | 1.2 |

Table 3 – EPA Estimated Values of PUE for 2011

Nowadays we can easily compare these estimated values. The state of the art data centers of Google, Yahoo or Microsoft usually exceed the EPA goal. Google claims that all their data centers with IT load of at least 5MW have together PUE of 1.16. The lowest PUE acomplished is 1.09. PUE is very important for the data center efficiency, however also usage of renewable power sources to lower the CUE is very actual. For example the brand new 1 Bilion USD 100MW Apple data center in North Carolina was reported with the lowest clean energy index of all rated companies with just 6.7%. For better comparison, Yahoo won with 55.9% clean energy used, followed by Google (36.4%) and Amazon (26.8%).

As there are many smaller solutions of data centers (hundrets of kW), it is harder to find out their actual average PUE. However it becomes usual that even these smaller data centers are very considered with the power consumption and more money are spent in this field as well. In case of examined data center, planned PUE while running on full capacity is 1.5.

# II. PRACTICAL PART

# 4 DATA CENTER CONSTRUCTION AND CLASSIFICATION

## 4.1 Data center construction

Currently there are two basic types of data center construction traditional and modular. Traditional data centers are located in single purpose buildings (only data center) or in a sector of some building. Single purpose buildings data centers are mostly used by hosting providers and large corporations such as Yahoo, Google etc.

Modular data centers are very popular nowadays and they could be categorized into container and modules data centers. Container (can be referred also as a portable) data centers are built in standard shipping containers and provide many advantages such as low PUE (Power Usage Effectiveness), sophisticated cooling, high modularity, extensibility and mobility. Module data centers are different from each manufactures. Considering the growth of the company a short time and even cheaper deployment and extension of a modular data center is probably the biggest advantage. More detailed specification of these types can be found in section no. 6 - Comparison of traditional and modular data centers.

## 4.2 Classification of data center

Enormous variations of data center designs are a result of different point of view of each date center designer even for the same stated requirements. Regarding these facts their classification is hard. There have been various methods introduced, but I am going to focus on three most common known methods: The Uptime Institute’s Tier Performance Standards, TIA 942, and Syska Hennessy Group's Critically Levels$^{TM}$. Basically all Tier standards are formed on the bases of The Uptime Institute’s Tier Performance Standard and so this base standard is the most used one.

### 4.2.1 The Uptime Institute’s Tier Performance Standard

This standard was developed in 1995 and has become widely referenced in the data center construction industry. There are 4 tier levels (Tier 1- Tier 4), which have evolved through different data center projects. This method comes with a high level guideline but it does not provide specific design details for each Tier level. The Uptime Institute’s Tier performance standard is now the most used standard for evaluating the data center tier level.

#### *4.2.1.1 Tier I: Basic site infrastructure*

The fundamental requirement

- There are no redundant capacity components, nor distribution paths for the equipment.

The performance confirmation test

- Any failure of capacity component or distribution path will have direct consequences on the computer systems.
- Complete shutdown will be required for planned work and will impact the computer systems.

The operational impact

- Susceptible to disruption from planned and unplanned activities
- For safe preventive maintenance and repair work the site infrastructure has to be completely shut down. Human failure during performing these maintenance work increases the risk of unplanned disruption and data center down time.

#### *4.2.1.2 Tier II: Redundant capacity components site infrastructure*

The fundamental requirement

- Redundant capacity components and single non-redundant distribution paths for the computer equipment.

The performance confirmation test

- A failure of a capacity component may have consequences on the computer equipment
- A failure of a distribution path will shut down the computer equipment

The operational impact

- Susceptible to disruption from planned and unplanned activities
- Redundant UPS modules and engine generators are required

- For safe preventive maintenance and repair work the site infrastructure has to be completely shut down. Human failure during performing these maintenance work increases the risk of unplanned disruption and data center down time.
- Data center disruption can be caused by operation errors or spontaneous site infrastructure components failures.

#### *4.2.1.3 Tier III: Concurrently maintainable site infrastructure*

The fundamental requirement

- Redundant capacity components and multiple distribution paths for the site's computer equipment. One distribution path serves the computer equipment at any time.
- Engine generators for Tier III and IV level are considered main power supply and cannot have any limitation on consecutive hours of operation.

The performance confirmation test

- Each and every capacity component and element of the distribution paths can be removed without shutdown of any computer equipment.

The operational impact

- Susceptible to disruption from unplanned activities
- When planned site infrastructure maintenance is performed, the remaining equipment can work using redundant capacity components and distribution paths.
- Tier III computer hardware must have dual power inputs to allow concurrent maintainability of power distribution between UPS and the computer equipment.
- Risk of disruption can be elevated during the maintenance activities, due to the human failure.
- Data center disruption can be caused by operation errors or spontaneous failures of site infrastructure components.

#### *4.2.1.4 Tier IV: Fault tolerant site infrastructure*

The fundamental requirement

- Tier IV data center has redundant capacity systems and simultaneously working multiple distribution paths for the site's computer equipment.
- All IT has to be installed properly to be compatible with the topology of the site's architecture and has to be dual powered as well.

The performance confirmation test

- Computer equipment cannot be impacted by a single worst-case failure of any capacity system, component or distribution element.
- Each and every capacity component and element of the distribution paths has to be able to be removed from service on a planned basis not causing shutdown of any of the computers.
- Dual power inputs are required to establish fault tolerance and concurrent maintainability of critical power distribution systems.
- To prevent any event from impacting both complementary systems and distribution paths, these systems have to be physically separated.

The operational impact

- The site cannot be disrupted from a single unplanned worst-case event.
- The site cannot be disrupted from any planned work activities.
- Using the redundant capacity components and distribution paths on remaining equipment, the maintenance of site infrastructure can be performed.
- Risk of disruption can be elevated during the maintenance activities, due to the human fault.
- Fire alarm, fire suppression or the emergency power off (EPO) use may cause disruption of the data center.

Performance standards by tier level comparison are shown in Table 4.

| Tier Requirement | Tier I | Tier II | Tier III | Tier IV |
|---|---|---|---|---|
| Source | System | System | System | System+System |
| System Component Redundancy | N | N+1 | N+1 | Minimum of N+1 |
| Distribution paths | 1 | 1 | 1 normal and 1 alternate | 2 simultaneously active |
| Compartmentalization | No | No | Yes | Yes |
| Concurrently Maintainable | No | No | Yes | Yes |
| Fault Tolerance (single event) | No | No | No | Yes |

Table 4 – Tier Levels

All information about Tier levels are based on ([1]).

### 4.2.2 TIA 942

In the TIA 942 revision 5 there are also 4 tier levels which are based on the Uptime Institute’s Tier Performance Standards. The four tiers described in TIA 942 appendix G are informative and are not considered as requirements of this standard. Nevertheless appendix G provides specific design criteria which help designers to build a specific tier level and the ability to evaluate own design to the owners of data centers.

#### *4.2.2.1 TIA942 levels*

There are currently 4 tier levels, which describe requirements and availability as well. Tier 1 is the most simple data center. It consists of a server room with equipment installed by the basic computer system installation guidelines and has availability 99,671% (28.8 hours of annual down time). The state of the art data centers are tier 4 class, which has multiple cooling and power distribution paths and all the components are redundant 2 (N+1). Availability of this type of data center is 99,995% which means 0.4 hours of annual downtime.

### 4.2.3 Syska Hennessy Group's Critically Levels

There are ten critically levels build on the Uptime's four tiers in this classification. These levels consider also the recent data center trends such as high density computing and flexible architectures. Syska also includes not only the “upfront” components and construction but also more complex elements which evaluates the maintenance and operation of data center. Syska’s levels are described at a high level and lack of the specificity of TIA-942. The 11 items are: power, HVAC, fire and life safety, security, IT infrastructure, controls and monitoring, commissioning and testing, operations, maintenance, operations and maintenance procedures and disaster preparedness. Syska refers to its tier system as a Criticality Levels$^{TM}$.

### 4.2.4 Tier location aspects

There are not only server construction demands related to the tier levels. To fit each tier level, location has to be chosen very carefully taking in consideration physical safety of the data center. High tier level data centers cannot be built in or near locations such as airports, flood area or in non single purpose buildings. Access to the data center area has to be secured and even parking places or roads close to the data center building have to be considered as a threat for high availability.

## 4.3 Real data center implementation:

In the case of examined data center, it is not possible to examine the exact Tier level. The architecture is mostly Tier IV, but several things are not corresponding (such as one independent utility power supply) with Tier IV level, so it cannot be really called Tier IV class data center. Big effort was made to the data center security in many ways.

### 4.3.1 Entrance and doors

At the entrance level there is card reader for the outdoor doors. Then ALUSER chamber door with bulletproof glass is installed. To open these doors you need the entrance card and provided pass code from the security. Each card type has also different access possibilities inside the data center. The aim of the chamber doors is to detect the amount of people entering the data center and to detect that only one person is presented in the chamber. Only if all conditions mentioned above are accomplished the person is allowed into the data center.

All the doors are fireproof (EI/EW30) with security class BT2 equipped with security locks. Double antistatic floor is manufactured from PVC MERO ELAST antistatic and it is covering every room with electronic equipment except the hall, where is the monitoring area.

### 4.3.2 IT power supply

In case of IT equipment, each cabinet is dually powered from independent three phase switchboards (IT-A and IT-B). Each of these switchboards is able to provide 11kW. This system corresponds with the Tier IV level. These IT-A and IT-B switchboards are connected to the superior power supply system by automatic transfer switch in the main switchboard room. All switchboards are installed together with the UPS systems and batteries within one room. This solution will prevent any possible fire of these devices to damage the expensive and sensitive IT equipment.

### 4.3.3 Fire detection and extinguishing

There are two levels of disposal of fire. As electronic equipment is presented everywhere, it is not possible to use common fire extinguishing methods such as water or foam. Therefore gas fire extinguishing systems are installed in every room. Various types of sensors are installed within the room considering possibilities of causes of fire. There are detection tubes in the switchboards, which are highly sensitive to heat. These tubes are installed through all of the switchboards and they can detect fire much faster than the smoke detectors in the ceiling. Once the fire is detected, alarm is set and the personnel have 10 seconds to exit the room before it will be locked. In this time the affected switchboard is turned off. After 10 seconds the fire extinguishing gas is released to the affected switchboard.

In case of the server room with sensible IT equipment, other system is installed as well. VESDA stands for Very Early warning aspirating Smoke Detection. This system is installed in every single cabinet in the server room and it can detect the fire before visible smoke, which can be detected by conventional smoke detectors appears. Usually this system detects fire in such advance, that it can be examined by the operator and appropriate actions can be performed before actual fire or smoke escalates. Wide range of sensitivity is provided by this system so the system can be integrated in many solutions.

### 4.3.4 Redundant support systems

The redundancy of UPSs and other support components such as cooling will be described later in separated chapters about them.

# 5 POWER MANAGEMENT IN NETWORK INFRASTRUCTURE

Most attention in power savings are focused on server, cooling equipment and power distribution as they are contributing with about 70% in the energy budget. The network infrastructure is involved in 10-20% of the energy consumption of the IT equipment. General practice is to leave all the equipment on. As the servers are not usually running full capacity, cost savings in network infrastructure should be considered as well.

## 5.1 Typical energy savings

Mainly network equipment is running on full capacity and the only means of power savings are simply thermo-controlled adjustable speed of fans and improved lower power consumption chips.

Some savings can be performed by turning off the unused equipment. It is very important to say that most of the power is consumed by the switch (difference between zero and full load traffic increases power consumption less than 8%). The fact that unused port saves only around 1-2W leaves us with no other choice than turning off the whole switch for a reasonable power saving (in case of 200W switch the lowest consumption on zero load is around 150W). The situation in the market is slightly improving, but generally there are still no hi-efficient network products available. Therefore maximum efficiency can be done only by combination of improved hardware components and its management. That is why also other concepts such as Elastic tree are developed.

## 5.2 Elastic tree concept

This concept is now more in the experimental phase however it should bring around 30% power savings in the network infrastructure. Principle is pretty simple. Using more complex and richer topology than typical 2N will result in higher redundancy of connections. This fact combined with advanced monitoring system and other features such as prediction of server load allows turning off some network equipment, while the lower load is presented without affecting the performance or reliability. The elastic tree is in fact a fat tree, where the unnecessary switches are switched off (as show Figure 5 and Figure 6). The reliability is higher than in the traditional 2N concept as well. As you can see in Figure 4 while in 2N topology two failures can even disconnect server using the fat tree the durability is higher. On the other hand as Elastic tree concept is more complex, it is also more complex to implement it especially in large server installations and has to be well managed to provide proper functionality.

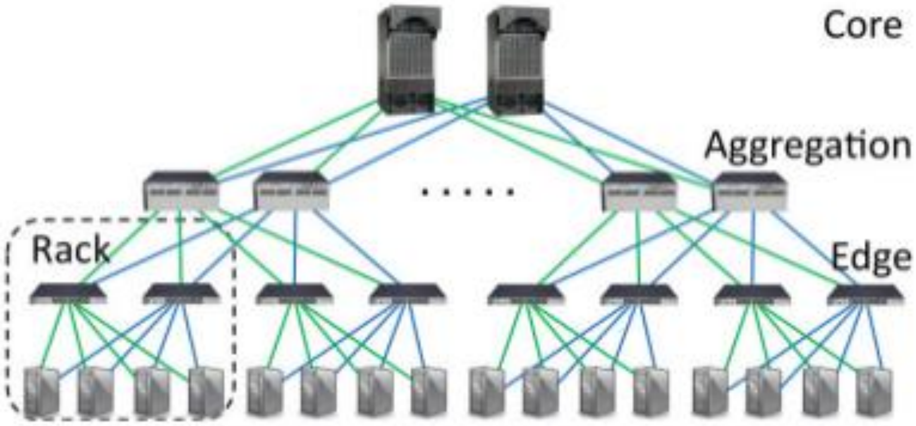


Figure 4 – Typical 2N topology [source: HELLER, Brandon, et al. ElasticTree : Saving Energy in Data Center Networks [online]. [s.l.] : [s.n.], 2010 [cit. 2011-05-09]. Available on WWW: <www.usenix.org/event/nsdi10/tech/full_papers/heller.pdf>.]

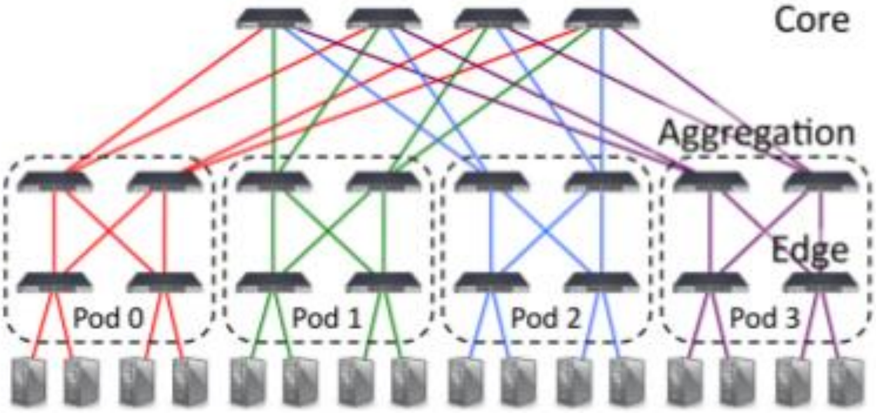


Figure 5 – Fat tree [source: HELLER, Brandon, et al. ElasticTree : Saving Energy in Data Center Networks [online]. [s.l.] : [s.n.], 2010 [cit. 2011-05-09]. Available on WWW: <www.usenix.org/event/nsdi10/tech/full_papers/heller.pdf>.]

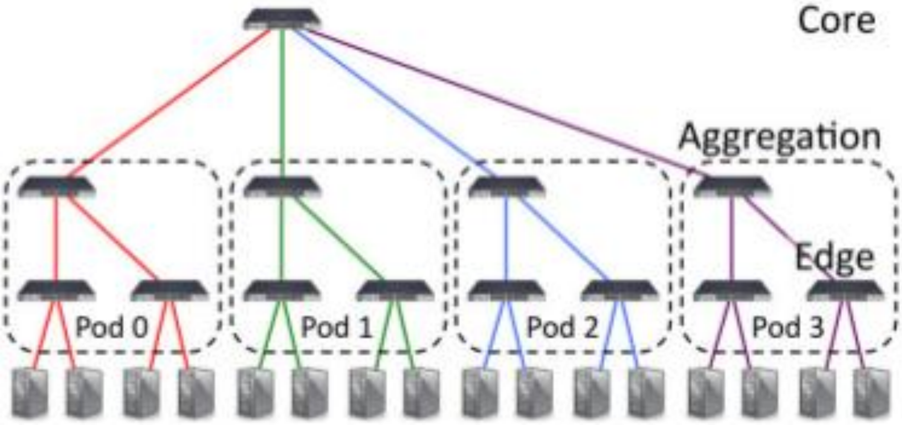


Figure 6 – Elastic tree [source: HELLER, Brandon, et al. ElasticTree : Saving Energy in Data Center Networks [online]. [s.l.] : [s.n.], 2010 [cit. 2011-05-09]. Available on WWW: <www.usenix.org/event/nsdi10/tech/full_papers/heller.pdf>.]

Elastic tree is a system for dynamic adaptation of energy consumption of a data center network and as shown in Figure 7 consists of three modules: optimizer, routing and power control.

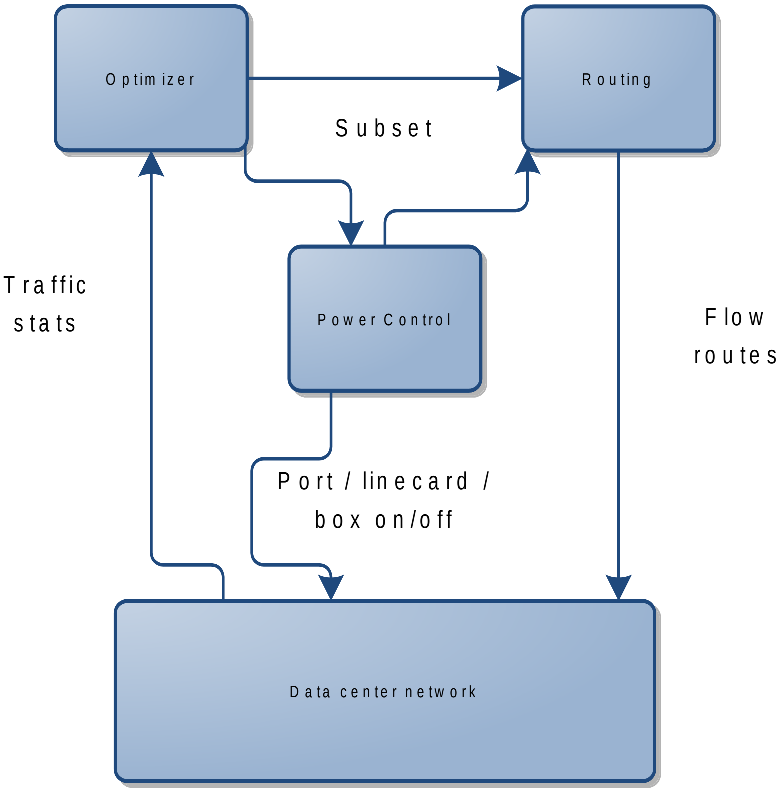


Figure 7 – Elastic tree system diagram

The optimizer finds the minimum power network subset for current traffic conditions. Optimizer decides at the base of topology, traffic matrix, power model for each switch and the required fault tolerance (spare equipment). Power control manages the power states of switches, linecards and ports. Routing chooses the proper paths for all flows based on the information from optimizer and power control and then pushes routes into the network. There are different methods to compute the minimum network subset based. They are summarized in Table 5.

First method is a formal model and due to need of high computational requirements it is used to evaluate quality of other optimizer's solutions. Second method is used for larger topologies to better understand the possible power savings. Topo-aware method is a simple heuristic method used to find quickly subsets in networks with regular structure.

| Type | Quality | Scalability | Input | Topology |
|---|---|---|---|---|
| **Formal** | Optimal | Low | Traffic Matrix | Any |
| **Greedy** | Good | Medium | Traffic Matrix | Any |
| **Topo-aware** | OK | High | Port Counters | Fat Tree |

Table 5 – Optimizer methods summary [source: HELLER, Brandon, et al. ElasticTree : Saving Energy in Data Center Networks [online]. [s.l.] : [s.n.], 2010 [cit. 2011-05-09]. Available on WWW: <www.usenix.org/event/nsdi10/tech/full_papers/heller.pdf>.]

Each method has different ratio of scalability and optimality, but all can be improved while using data center's history of traffic.

Specialized software is needed to maintain the ElasticTree concept. The system requires current network utilization monitoring and control over flow paths. In the ElasticTree prototype Open Flow software is used to complete these two tasks.

### 5.2.1 Power savings Analysis

ElasticTree system is compared to an always-on based system. The main metric inspected is the % of original network power. This percentage which will give us an accurate idea of overall power saved by using means of ElasticTree can be computed by this equation:

$$\% = \frac{\text{Power consumed by ElasticTree x 100}}{\text{Power consumed by original fat-tree}}$$

Only energy saved directly from the network is taken in consideration. Extra energy is spent for the servers required to provide ElasticTree computations. Cooling these servers will increase the amount of energy required for cooling servers, on the other hand powering off the unused switches will result in larger cooling savings. These extra costs/savings are not included in this computation.

There are two extremes which have to be mentioned. Near localized traffic matrices, where servers communicate with each other, using their edge switch. Second one is far (non-localized) traffic matrices, where servers communicate with servers in other pods using the network core. The near traffic represents the best case for energy savings and on the other hand far traffic represents the worst case for power savings. In case of far traffic, the energy savings are highly related to the network utilization. It is clear from the ElasticTree concept, that the highest energy savings can be achieved, when network is at lower utilization level. If the load is close to 100% then all links and switches have to remain active, however lower load can be transferred to smaller number of core links, saving the power by turning off the unused switches. Figure 8 summarizes the possible power savings for both far and near extremes. It is clear that there are no energy savings in case of 100% load for far traffic.

More information about this topic is available in [2].

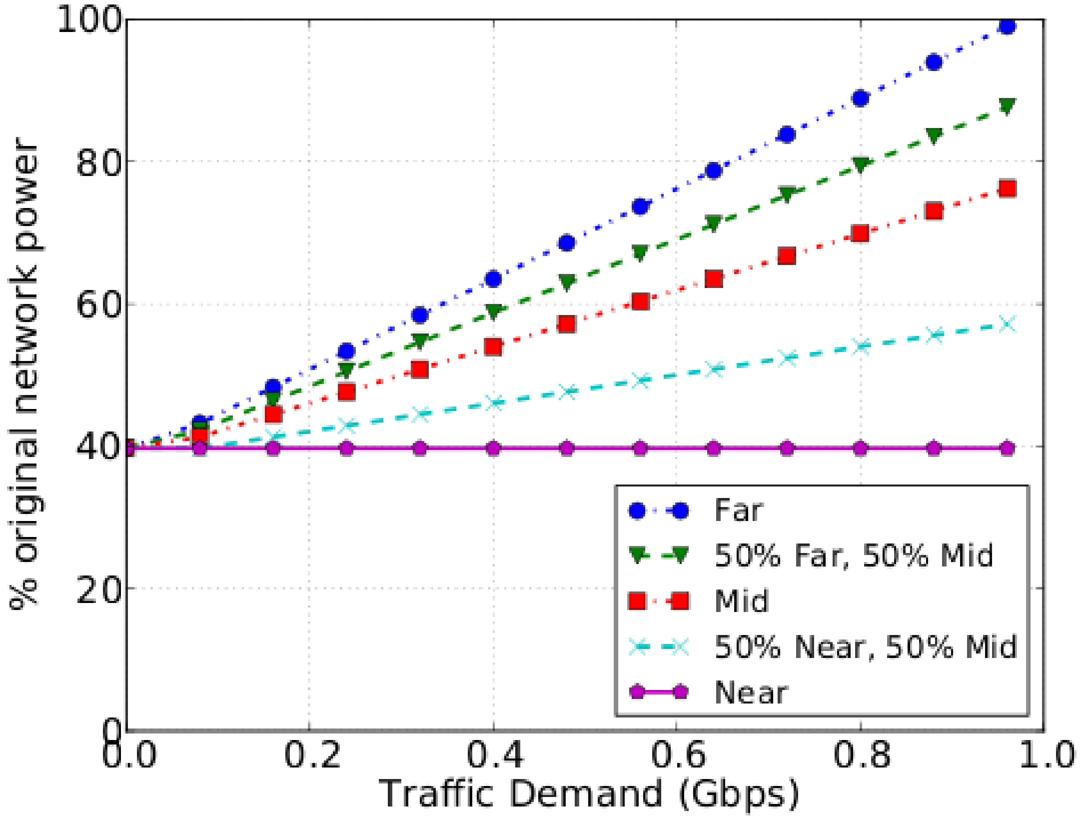


Figure 8 – Elastic Tree energy savings as a function of traffic demand [source: HELLER, Brandon, et al. ElasticTree : Saving Energy in Data Center Networks [online]. [s.l.] : [s.n.], 2010 [cit. 2011-05-09]. Available on WWW: <www.usenix.org/event/nsdi10/tech/full_papers/heller.pdf>.]

## 5.3 Cisco EnergyWise

Cisco EnergyWise is a completely another way to approach lower the energy costs via network. Basic principle is control of various devices which can communicate via network. Not only IT equipment such as IP telephones, access points, PCs, cameras can be controlled, but also other devices such as elevators, HVAC, security equipment etc may be integrated in the Cisco EnergyWise solution. Figure 9 shows the layers of the Cisco EnergyWise and its components.

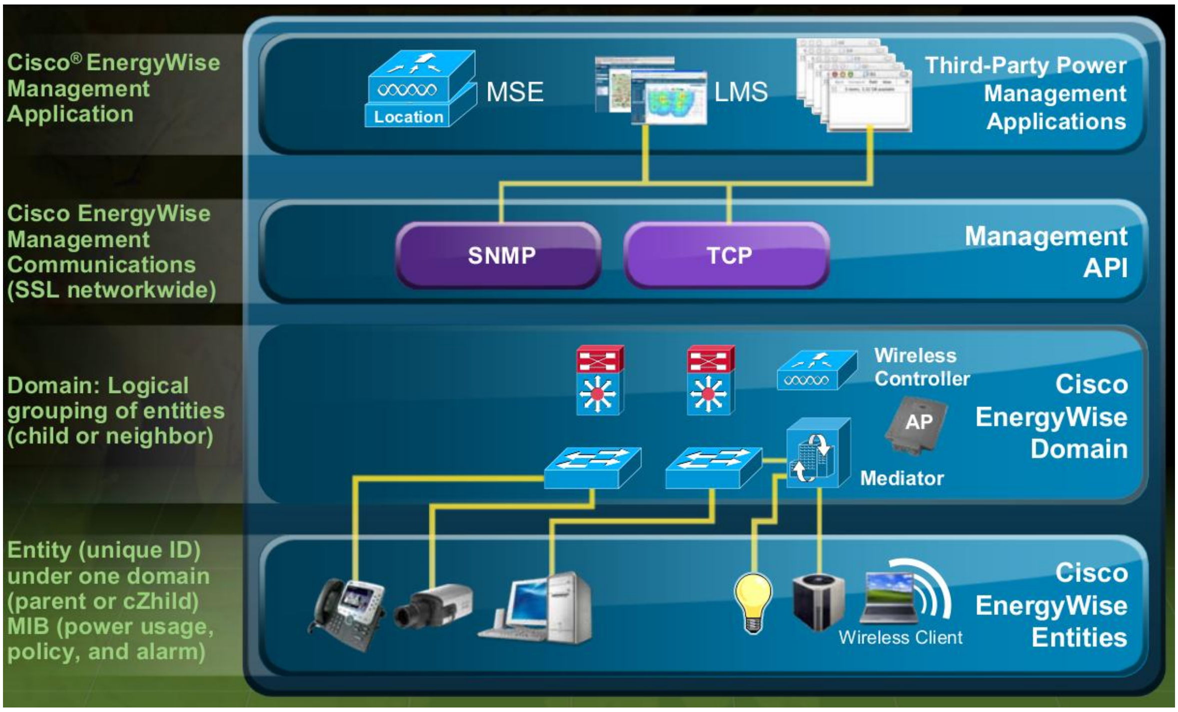


Figure 9 – Cisco EnergyWise components [source: Cisco Systems. Integrated Energy Management Solution [online]. [s.l.] : [s.n.], 2009 [cit. 2011-05-09]. Available on WWW: <http://www.cisco.com/>.]

For easier implementation Cisco EnergyWise cooperates with third party software such as IBM Tivoli or Solarwinds and with many more to come. Using this software makes it very easy and user friendly not only to implement, but also to set up automatic actions based on the monitoring system.

Cisco EnergyWise is able to bring significant energy savings, nevertheless these energy costs are not really related to the data center network core, but more likely to the external network equipment. More information about this technology can be found in [3].

## 5.4 Real data center implementation:

Due to the lack of other external components controlled by the Energy wise, the examined data center uses the always turned on system.

# 6 POWER MANAGEMENT IN SERVER INFRASTRUCTURE

## 6.1 Hardware

There are many hardware manufacturers, but as this thesis has been written with cooperation and consultation of IBM it has focused mainly on the IBM technologies.

### 6.1.1 Processor

IBM has come with a new POWER7 Processor and the whole POWER7 family around it in February 2010. Compared to the POWER6 or POWER6+, there have been many changes, but not all are important for this thesis. Given the energy consumption, two main aspects should be mentioned. The chip is manufactured with 45 nanometers copper/SOI technology (compared with 65 nanometers POWER6) which allows not only 1.2 billion transistors (790 million POWER6), but also reducing the power consumption for each transistor. Second aspect is lowering the frequency. Nevertheless in this case, it does not mean decreasing the performance. The slowest POWER7 core (3GHz) is faster than a 5GHz POWER6 core.

To identify problems in power management continuous collecting of real-time data is necessary. EnergyScale provides this functionality and the data can be displayed by IBM System Director Active Energy Manager and may be used for prediction of data center power consumption. There is no need for additional hardware to measure power consumption and thermal reports, because all POWER 7 processor systems collect data internally via built-in power meters. There are several methods for lowering the consumption of POWER7 processors. Description of power saving methods is based on [4].

#### *6.1.1.1 Static Power Saver Mode*

This feature allows lowering the processor frequency and voltage on a system fixed amount. This reduces the power consumption of the system while it delivers predictable performance. The level (percentage) is predetermined within safe operation limits and it is not user configurable. Recommended software for enabling/disabling Power Saver mode is Active Energy Manager. This method is mostly used in situation with predictable lower server utilization such as over night or weekends. This feature cannot be used while booting or rebooting the system.

### *6.1.1.2 Dynamic Power Saver Mode*

As the title indicates, this feature changes processor frequency and voltage based on utilization of POWER7 system processors. Settings can be changed by user only using Active Energy Manager. Given a constant load as the processor frequency increases its utilization decreases. While measuring real-time system utilization, Dynamic Power Saver uses this relationship to detect opportunities to save power. Performance and utilization of each POWER7 processor cores are monitored and Dynamic Power saver can adjust the frequency and voltage within milliseconds to gain power savings.

### *6.1.1.3 Differences in DPS from POWER 6 TO POWER 7*

The difference between these two models are described in

| Product Family | DPS Mode | Frequency Range | System Behavior |
|---|---|---|---|
| **POWER6** | Favor Power | Limited | System increases frequency only under heavy utilization |
| | Favor Performance | Full | System increases frequency only under heavy utilization |
| **POWER7** | Favor Power | Full | System increases frequency only under heavy utilization |
| | Favor Performance | Full | System defaults to maximum allowed processor frequency and only decreases frequency when processors are lightly utilized or idle |

Table 6 – Differences summary of DPSM between POWER6 and POWER7 [source: BROYLES, Martha, et al. IBM Energy Scale for POWER7 Processor-Based Systems [online]. [s.l.] : [s.n.], 2010 [cit. 2011-05-09]. Available on WWW: <http://www-03.ibm.com/>.]

#### *6.1.1.4 Power Capping*

Usually data centers are planned for power consumption based on the Undewriters laboratories rating placed on the back of servers. This UL rating represents the absolute worst case of power consumption.

Power Capping is a limit on a power consumption set by user by Active Energy manager. Generally in most data centers margin power which is hardly used is installed. The main purpose of Power Capping is to allow the operator to transfer power from the current load to the systems with margin power. This allows better use of current systems. It also means that the system will not use more power then assigned by the operator.

#### *6.1.1.5 Soft Power capping*

Soft power capping is very different from the guaranteed range (described above). It allows the system to be set in power save mode running at lower power/performance mode. This mechanism is used, when particular power consumption limit is required.

**Power capping example:**

There are twelve JS22 servers and each has the UL rating approximately 402W (DC). Based on this, the power needed to feed this data center is around 4.8kW. All servers will be connected to the Active Energy Manager. Using the statistics the actual power consumption reaches maximum of 308W, while running normal workload. The power cap will be set to 368W for each server and that will guarantee that the average power over time for each server will never exceed 368W. Using simple math (402W-368W) x 12 = 408W we will get 408W free. This is equal to consumption of another server which can be installed.

#### *6.1.1.6 Energy-optimized Fan control*

To avoid over-cooling all POWER7 systems are adjusting fan speed in real-time based on the temperatures of system components. In some cases the exhaust temperatures may increase, but this is logical consequence of the fan speed optimization and avoiding the over-cooling.

#### *6.1.1.7 Processor Core Nap*

Core Nap means a low power mode, which is activated, when there is no work for particular processor core. In this case processor will stop execution. When the processor thread is idle, the Hypervisor firmware immediately puts the thread into the Nap mode. When all threads are idle, the whole core enters Nap mode as well.

#### *6.1.1.8 Processor folding*

Additional energy savings can be realized when processor core is idle intently. Processor folding adjusts the number of processors available to match the demanded number by workload. It increases energy savings in periods of low and moderate workload. Unavailable processors remain idle for longer than they normally would. Power savings are similar to virtualization savings, but it can be performed with much greater efficiency and fidelity and even without the impact on virtualization configuration.

### 6.1.2 Cooling

#### *6.1.2.1 Precision and comfort cooling comparison*

Typical data center computer room can have up to five times higher heat density then a typical office as it is shown in Figure 10. Due to high density of rack configurations data center temperatures are increasing faster than ever before.

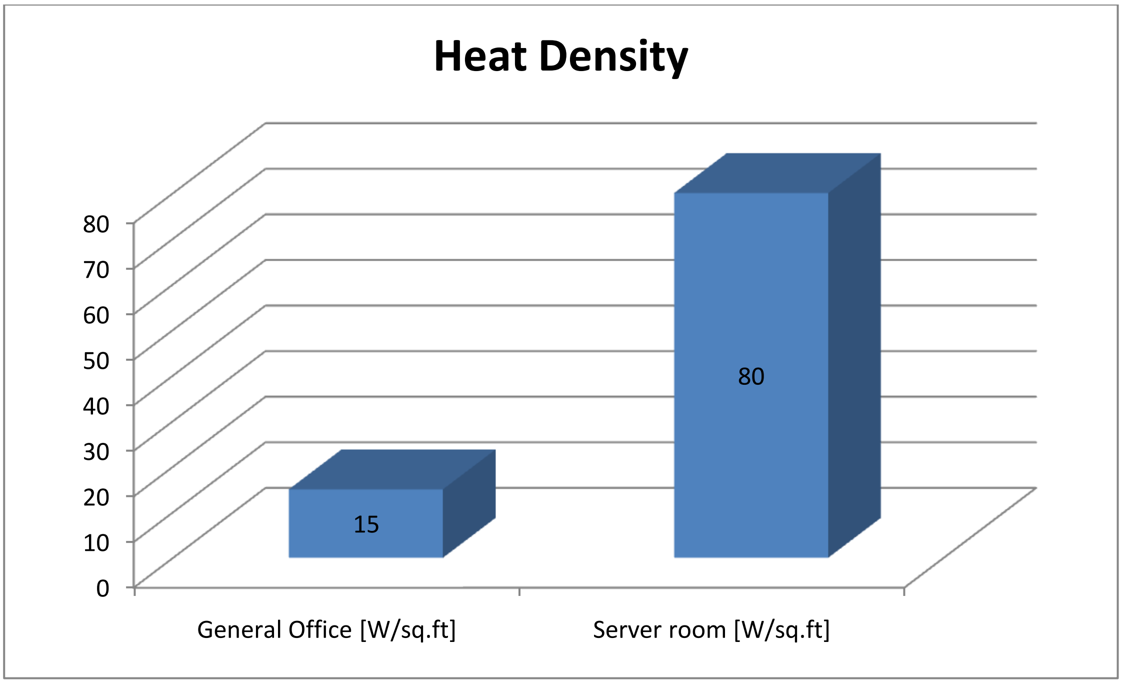


Figure 10 – Heat density Office/Computer room comparison

Therefore precise temperature and humidity control is necessary and as the comfort cooling systems typically do not have built-in humidity controls, precision cooling systems use is necessary. Lower or higher moisture levels can cause damage to the equipment. In case of higher moisture switching circuits can corrode causing malfunctions and failures. On the other hand low humidity causes static discharges, which interfere with normal equipment operation. Dust and small particles can damage charged electronic components. Most comfort cooling systems are using only 10% efficient residential filters which makes them unsuitable for data centers. Precise cooling systems are using chilled water and sealed coils and other mechanisms reducing the possibility of airborne particle contamination.

Purchase costs of precision cooling unit are higher, but this investment is compensated by the lower operation costs and reliability. An example of precise and comfort units will give the best explanation. To better understand this example, SHR (Sensible Heat Ratio) has to be explained. In short way it is the percentage of measureable by thermometer that has to be removed by cooling. In terms of computer room it is between 90-100% compared with 60-70% in an office environment. This example is based on calculations in document [5].

Let’s assume then these facts:

- Each ton of cooling requires 0.800kW
- The compressor motors and fans are 90% efficient
- Electricity costs $0.08kWh
- Humidification is required from November through March (3650 hours)
- The precision cooling system has a SHR of 0.90 and the comfort system has a SHR of 0.60

Calculating the cost per ton of cooling for a year:

$$\frac{0.8\text{kW/ton } x \ 8760 \ hrs./yr. \, x \ \$0.08/\text{kWh}}{0.90 \text{ efficiency}} = \$623 \text{ ton/yr.}$$

Then determining costs per sensible ton of cooling for each system.

Precision cooling system:

$$\frac{\$623}{0.90\text{SHR}} = \$692 \text{ ton/yr.}$$

Comfort cooling system:

$$\frac{\$623}{0.60\text{SHR}} = \$1\ 038 \text{ ton/yr.}$$

As you can see there is a difference of $346 per ton and it confirms the general principle that it takes three tons of comfort cooling capacity to equal two tons of precision capacity.

Second point of comparison is the cost of humidification which is determined by calculating the latent cooling that occurs per ton of sensible cooling (re-humidification).

Precision system:

$$\frac{12\ 000\ \mathrm{BTU/ton}}{0.90\ \mathrm{SHR}} - \ 12\ 000\ \mathrm{BTU/ton} = \ 1\ 333\ \text{latent}\ \mathrm{BTU/ton}$$

Comfort system:

$$\frac{12\ 000\ \mathrm{BTU/ton}}{0.6\ \mathrm{SHR}} - 12\ 0000\ \mathrm{BTU/ton} = 8000\ latent\ BTU/ton$$

This means that comfort system spends extra 667BTU/ton of sensible cooling to remove the humidity that has needs to be replaced to preserve the required data center humidity of 45-50%.

The added cost is computed as:

$$\frac{6667\ \mathrm{BTU/ton\ x\ 3650\ hrs/yr\ x\ \$0.08/kWh}}{3413\mathrm{BTU/hr/kW}} = \$570\ \mathrm{ton/yr}$$

In this case the operating cost of a comfort-based system spends $916/ton/yr more of sensible and latent cooling in comparison with a precision air system.

#### *6.1.2.2 Calculating total cooling requirements for data centers*

In order to make a right decision for cooling systems, it is necessary to know all the heat producing elements and their exact position in the data center. Most of the equipment in the data center is air-cooled and following the best practices in the air-flow management with adequate planning results in much better effectiveness and therefore in energy savings.

Measuring the heat output

Heat output is commonly expressed in BTU per hour, Tons per day and Joules per second. In the North America the BTU is mostly provided unit, however there is an effort to use one standard unit, the Watt. There is a conversion table available for all the units.

| Given value in | Multiply by | To get |
|---|---|---|
| **BTU/hr** | 0.293 | Watts |
| **Watts** | 3.41 | BTU/hr |
| **Tons** | 3.530 | Watts |
| **Watts** | 0.000283 | Tons |

Table 7 – Thermal unit conversion [source: RASMUSSEN, Neil. Calculating Total Cooling Requirements for Data Center [online]. [s.l.] : American Power Conversion, 2007 [cit. 2011-05-09]. Available at WWW: <http://www.apcmedia.com>.]

The total heat output of all the equipment can be fortunately easily determined using simple standardized rules. UPS and power distribution system output heat consist of general fixed losses and proportional loses to the operating power. These losses are usually very similar across models and even brands, so they can be approximately calculated without a significant error. Heat of people and lightning can be also easily predicted using standard values. Then only rated electrical system power and floor area are left to consider while determining the cooling load.

Air conditioning units are producing significant amount of heat, but this heat is exhausted to the outside and therefore it is producing just a small amount of thermal load inside the data center. This heat does not decrease the efficiency of the air conditioning system and so it is only taken in consideration when large air conditioning units are used.

In some cases simplified calculation of the thermal heat can be used instead of detailed thermal analysis. This method is typically within margin of error of the more complicated detailed analysis. Both methods can be used for comparison and revision of each other. Using Table 8 a simple and fast heat load calculation is possible to be determined by almost anyone without specialized knowledge. However this method is applicable for small data centers.

| Item | Data required | Heat output calculation | Heat output subtotal |
|---|---|---|---|
| IT equipment | Total IT load power in Watts | Same as total IT load power in Watts | __________ Watts |
| UPS with Battery | Power system rated power in Watts | (0.04 x Power system rating)+(0.05 x Total IT load power) | __________ Watts |
| Power distribution | Power system rated power in Watts | (0.01 x Power system rating)+(0.02 x Total IT load power) | __________ Watts |
| Lightning | Floor area | 2.0 x floor area (sq ft) or 21.53 x floor area (sq m) | __________ Watts |
| People | Max no. of personnel in data center | 100 x Max no. of personnel | |
| **Total** | Subtotals from above | Sum of heat output subtotals | __________ Watts |

Table 8 – Server room heat output calculation [source: RASMUSSEN, Neil. Calculating Total Cooling Requirements for Data Center [online]. [s.l.] : American Power Conversion, 2007 [cit. 2011-05-09]. Available at WWW: <http://www.apcmedia.com>.]

Figure 11 shows the general approximate percentage of contributors of heat in typical data center environment. The exact data can vary especially depending on the total data center utilization. The higher utilization, the lower percentage of can be produced for example by the UPS systems.

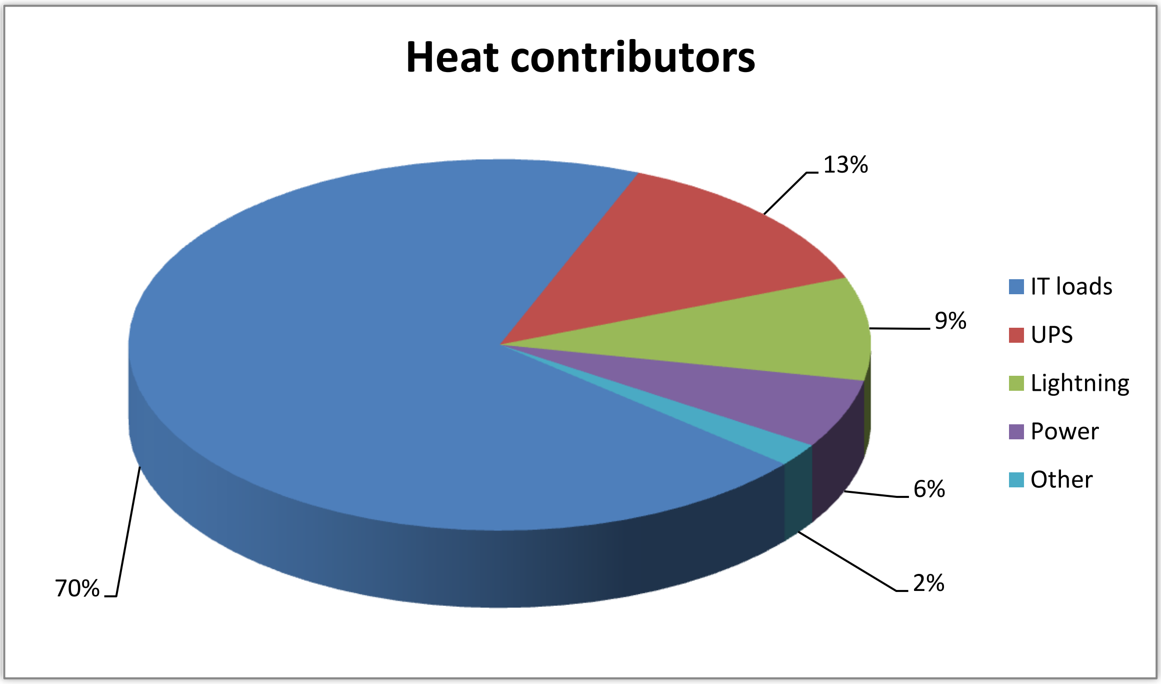


Figure 11 - Approximate percentages of heat contributors

There are various types of other heat sources. Heat can be conducted in from outside walls, as well as heat from sunlight pervading through the windows. In case of small data centers there is usually no need for these calculations, because they are mostly placed in a room without walls and windows to the outside. On the other hand it is very important for larger data centers to consider this heat and calculate the removal of it.

Humidity level is very important for determination of air conditioning units of the data center. To maintain the humidity level, the cooled air is mixed with non-returned air, so the humidity level is as required, saving the energy for additional humidification. In large data center the CRAC units have to deliver air with a very low temperature for mixing and therefore the capacity of the CRAC units is significantly decreasing. Generally CRAC units must be oversized around 30%. More specific information about calculation of total cooling capacity can be found in [6].

#### *6.1.2.3 Types of air conditioned data center systems*

Various types of cooling suiting different purposes were established during the evolution of data centers. Decision which type to use is usually based on data center structure, utilization and budget. The cooling system is using raised floors and sometimes lowered ceiling for the air distribution. Precise planning and construction are required to avoid energy waste and overheating of some parts of the server room. Planning must consider even each rack‘s density and their placement decided on a base of these facts.

Figure 12 shows different approaches for higher loaded racks placement. Figure 10A spreads the highly loaded racks into the whole area of the server room, so the higher loaded racks can use the cold air capacity of the whole row. In case of Figure 10B where all the highly loaded racks are placed in one row (dedicated high-density area), a designated higher-cooling area should be provided for this row. If all racks are highly loaded the whole room should be provided with capability cool any and every rack.

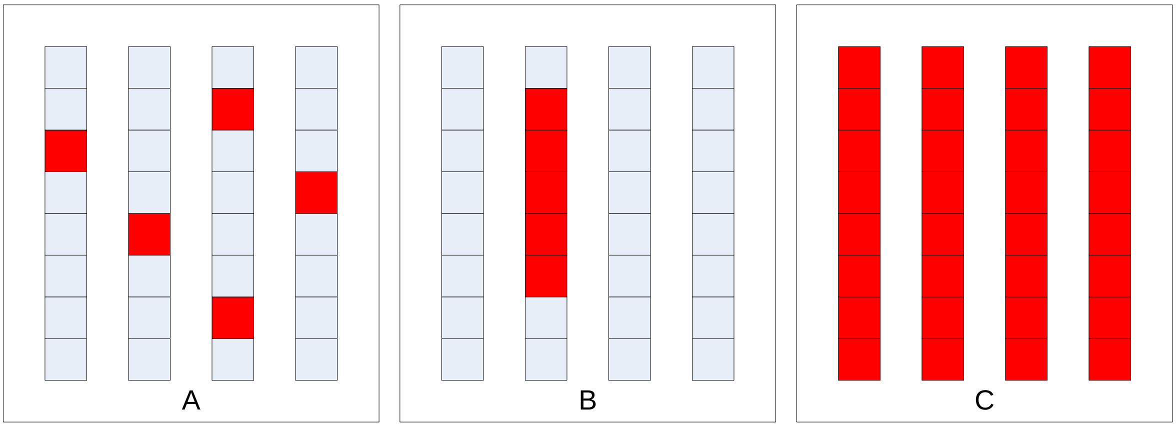
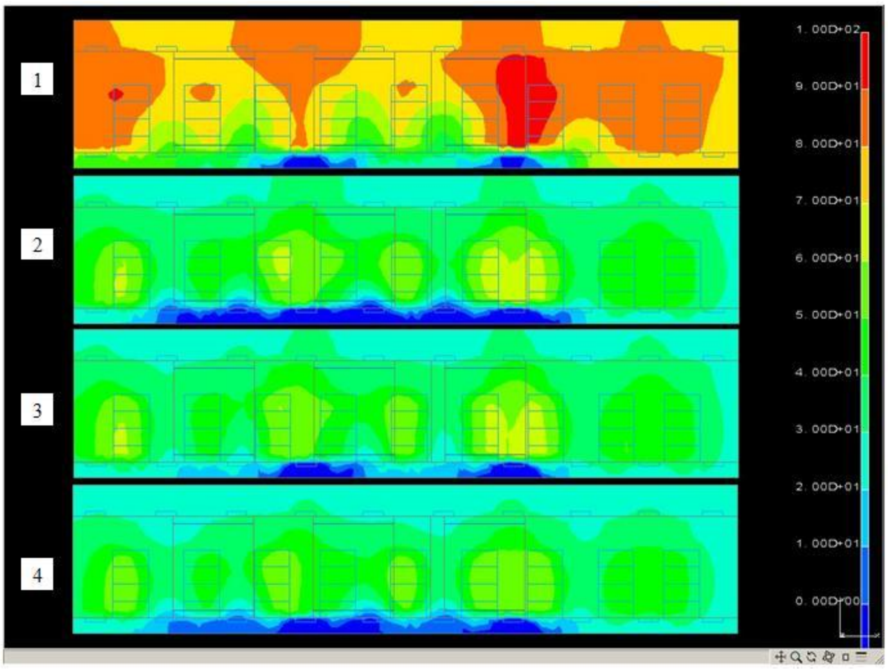


Figure 12 – Power density three configurations

Higher efficiency in cooling can be done also by the different approaches in air flow distribution. Figure 13 clearly shows the temperature differences in case of different methods of air distribution used. The highest effectiveness is clearly the fourth case, when both floor and ceiling fans are used.



- ☐ Case #1: Air temperature distribution in the data center obtained without fan-assisted tiles.
- ☐ Case #2: Air temperature distribution obtained with floor fan tiles.
- ☐ Case #3: Air temperature distribution obtained with ceiling fan tiles.
- ☐ Case #4: Air temperature distribution obtained with both floor and ceiling fan tiles.

Figure 13 – Different aproches in air flow distribution [source: Data centers lectures, Instituto Superior Technico Lisboa]

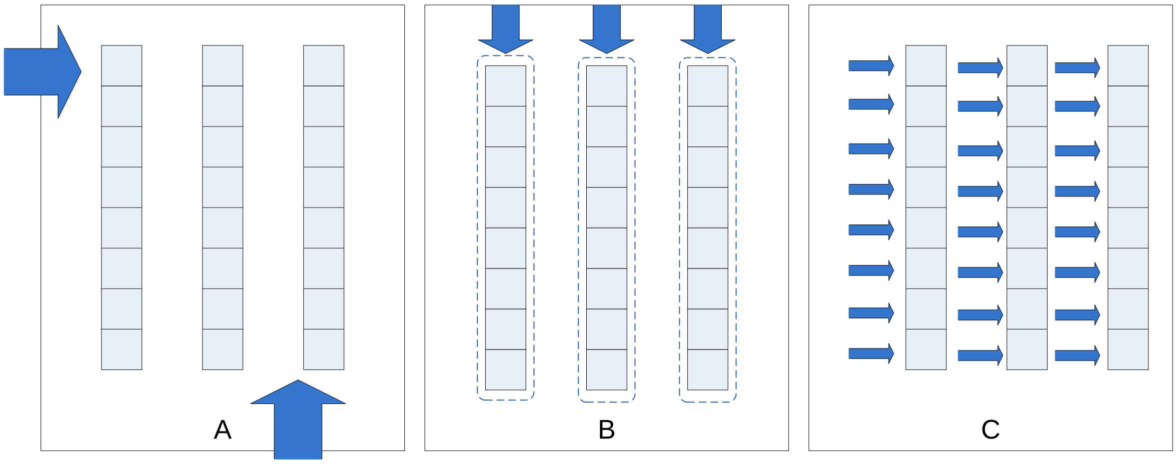


Figure 14 – Cooling oriented architectures A: Room, B: Row, C: Rack

### Room-oriented architecture

Room oriented architecture is the simplest one and the CRAC units are designed to cool the entire room capacity. In small data centers the air can be distributed unrestrictedly by fans ducts. In more sophisticated data centers raised floors can be used for distribution of the cold air into the hot/cold aisle layouts for better control of the air movement. High precision planning for power density is required due to low performance uniformity.

## Row-oriented architecture

In case of Row-oriented architecture the CRAC units are associated only with a row. Mostly the CRAC units are mounted close to the row to avoid performance losses. They can be mounted next to the racks or under the raised floor.

This solution is used for high density solutions, because the airflow is much more predictable. The reduction of path lengths and so the power required for distribution can be a large benefit for lightly loaded data centers. The power used for airflow distribution can sometimes exceed the total IT load power consumption. The row architecture can be targeted to actual needs for each row depending on their load and density.

Row oriented cooling can be also realized without raised floors and this can significantly reduce the installation costs and also data centers can be built inside buildings that would not be suitable for data center installation. It can be done by overhead cooling as showed in Figure 15. On the other hand it is usually not possible to provide this solution for high density installation. These installations have their special requirements.

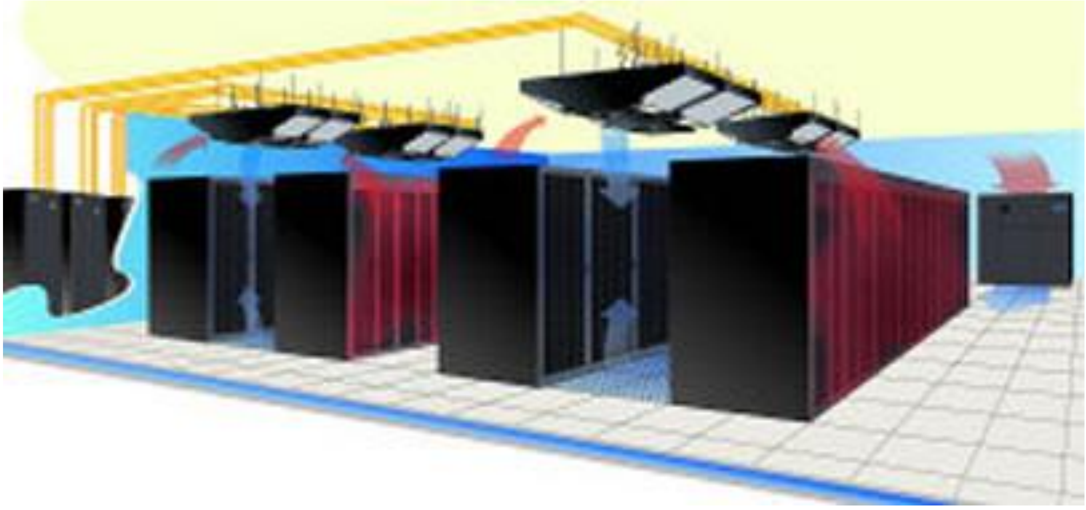


Figure 15 – Overhead cooling [source: http://www.42u.com/images/liebert-xdo-application-example.jpg]

The row-oriented architecture can be simplified in the implementation because it is relatively independent of the server room geometry. In some cases there is need of more CRAC units for the server room while comparing the row-oriented and room-oriented architecture, but this does not have to mean higher costs because of higher efficiency especially for the high density solutions.

**Rack oriented architecture**

In the rack-oriented architecture, CRAC units are associated and directly mounted to or inside the rack. The advantages are obviously very short and strictly defined air paths. This makes the cooling the most effective cooling from already mentioned architectures. These air paths are also absolutely independent of any installation variation and room constraints. This solution is sometimes inevitable for high density (up to 50kW per rack) installations. It is very suitable for data centers, where each rack power density differs a lot (blade server vs. communication equipment) and also where the redundancy has to be targeted to specific racks.

An example of this cooling is shown in Figure 16. There are various types of implementation. Figure 14 shows solution installed inside the rack, other solutions put the CRAC unit for example above or under the rack as well. The main disadvantage of this system is large amount of air condition devices required.

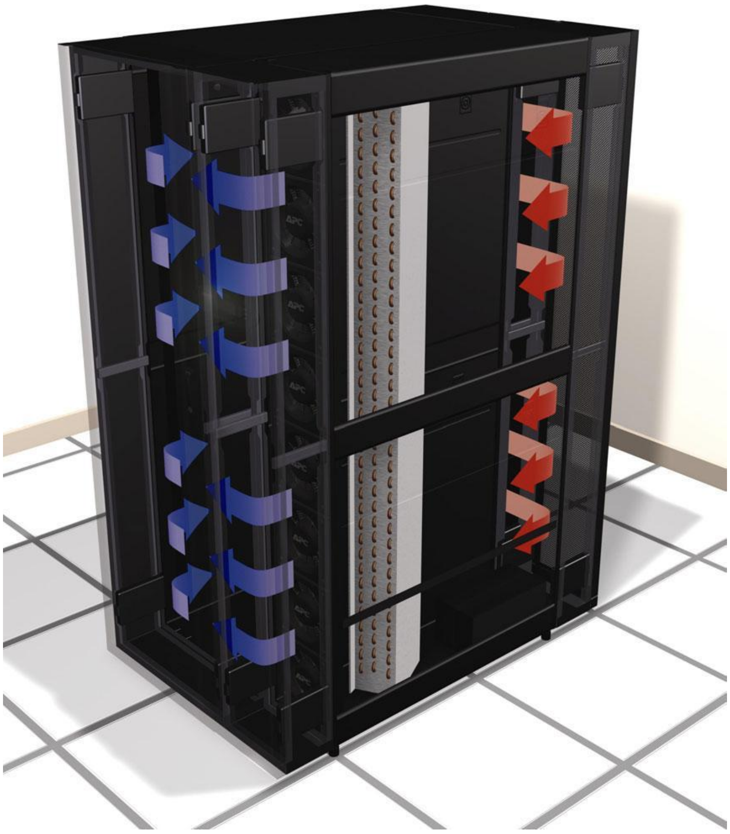


Figure 16 – Rack-oriented cooling solution (inside the rack) [source: DUNLAP, Kevin; RASMUSSEN, Neil. The Advantages of Row and Rack-Oriented Cooling Architectures for Data Centers [online]. [s.l.] : [s.n.], 2006 [cit. 2011-06-09]. Available on WWW: <www.datacentres.com/>.]

Disadvantage of rack-oriented system is lower effectiveness in practical use. When the IT equipment is lightly utilized the dedicated air conditioning is cannot be used for other racks.

**Mixed architecture**

It is not unusual to combine all these three architectures together. Nothing prevents all of them from being used at the same time, contrary it can bring many benefits. This mixed architecture is very suitable for data centers with large spectrum of power densities. An example of mixed architecture is shown in Figure 17, where the highest density racks are served with rack-oriented architecture and on the other hand the lowest density racks are supplied by room-oriented cooling.

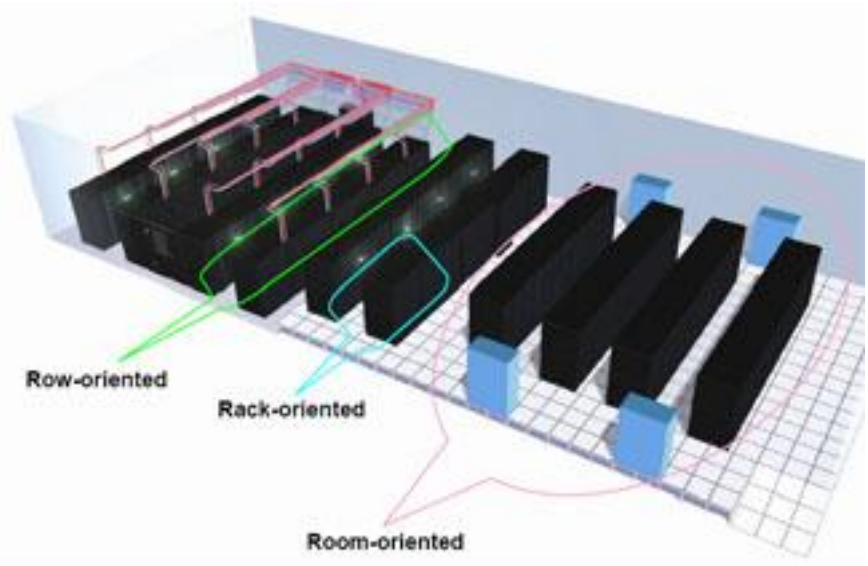


Figure 17 – Mixed cooling architecture [source: http://www.psmtech.com.sg/prodimages/DC%20Cooling9.jpg]

Room-oriented cooling can supply for example communication equipment and storage. Row-oriented cooling supplies high density areas such as blade servers and rack-oriented cooling supplies isolated ultra-high density racks. Row and rack-oriented architectures can be also used while additional capacity is installed in the data center, when room-oriented cooling is no longer sufficient. These approaches make the added capacities thermally neutral for the existing cooling system, so the whole system does not have to be changed.

More information about cooling architectures is available in [7].

Actual capacity of cooling system differs from the performance of each cooling system. If we have 300kW cooling units installed they actually cannot cool 300kW of IT load. The performance differs from each type of architecture and different power density. It is summarized in Figure 18. The capacity of the CRAC unit should never reach 100% capacity. It would mean that the cooling is insufficient and system would have to be resized.

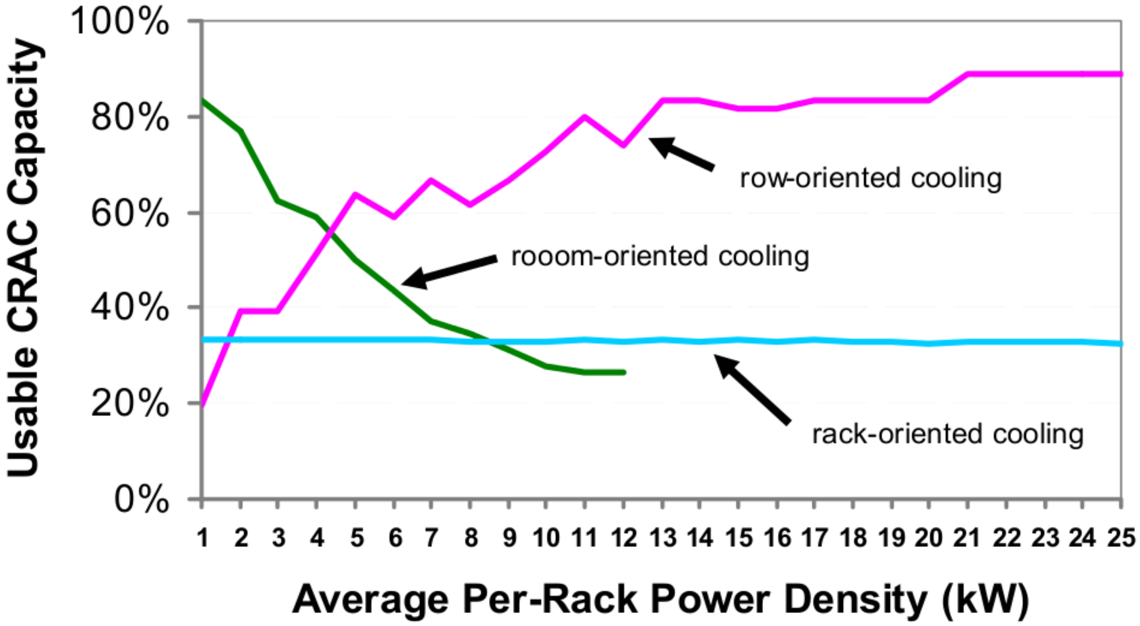


Figure 18 – Usable cooling capacity for different architectures [source: DUNLAP, Kevin; RASMUSSEN, Neil. The Advantages of Row and Rack-Oriented Cooling Architectures for Data Centers [online]. [s.l.] : [s.n.], 2006 [cit. 2011-06-09]. Available on WWW: <www.datacentres.com/>.]

This effect is connected with CRAC units only. In case of outdoor heat rejection systems it is possible operating them at 100% usable capacity for all three architectures.

Another important aspect when considering air conditioning is the cost and electrical efficiency. Regarding the advantages and disadvantages of each architecture summarization of costs is shown in Figure 19.

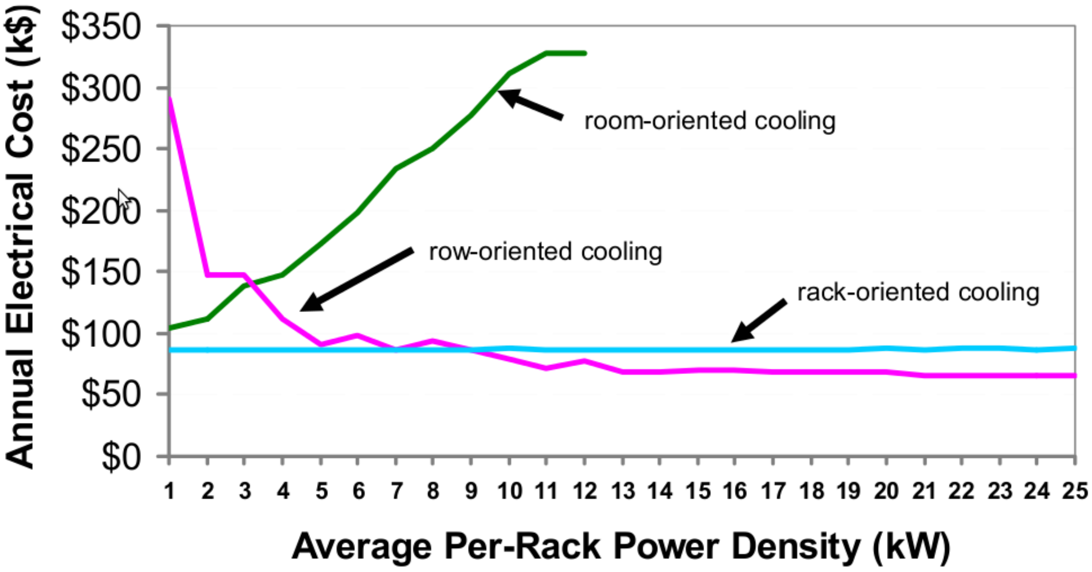


Figure 19 – Annual CRAC electrical costs for different cooling architectures [source: DUNLAP, Kevin; RASMUSSEN, Neil. The Advantages of Row and Rack-Oriented Cooling Architectures for Data Centers [online]. [s.l.] : [s.n.], 2006 [cit. 2011-06-09]. Available on WWW: <www.datacentres.com/>.]

### Hot/Cold Aisle Layout

Hot/Cold Aisle layout can be combined with room and row architecture, improving their characteristics and efficiency. The basic principle is to separate areas with cold and hot air. If it is just the aisle separation there are usually problems with mixing the air show in Figure 20.

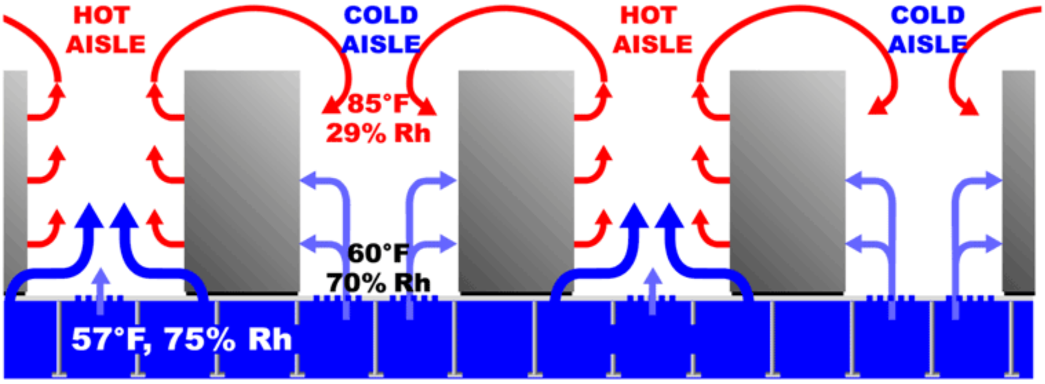


Figure 20 – Hot and cold air mixing problem [source: http://media.techtarget.com/digitalguide/images/Misc/sullivan-hotspots-2-sm.gif]

In all cases the effectiveness can be improved if air-blocking curtains or barriers around the cold/hot areas are installed, to prevent mixing of the air as shown in Figure 21.

This problem can be solved by different approaches. The air blocking curtains or solid barriers installed are the best practices to resolve this problem.

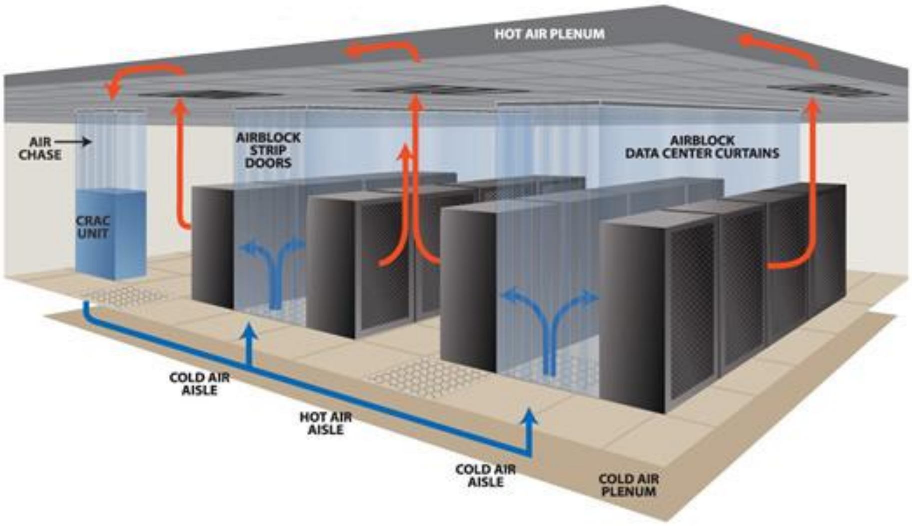


Figure 21 – Air blocking curtains as mixing air solution [source: http://www.dpguardian.com/images/plenum-to-plenum.jpg]

### 6.1.3 Real data center implementation

Server room is mainly equipped with the combination of POWER6+ and POWER7 technology. The POWER7 represents around 60%, while the POWER6+ represents the remaining 40%. The POWER6+ technology was installed as the very first one, when POWER7 was not yet released. The new modules will contain the POWER7 systems for better performance and energy efficiency.

Cooling system is separated into two cooling systems. Outdoor units are System+System configuration and inner units are designed with n+2 redundant. Cooling units Emerson type SQS040 have integrated free cooling and hydro modules. Free cooling is a method, which is using low external air temperatures to cool the chilling liquid with minimum extra energy needed. This method significantly decreases the energy costs during cold weather times. These Emerson cooling units are designed in a way, that one unit running covers the whole cooling capacity and the other is a backup unit, however both of them can run at the same time given the priority of the free cooling. Each cooling unit has its own pipe distribution to the “cooling machinery room”.

There is a primary and secondary circuit. Primary circuit represents the chilling units, assistant gatherers and gatherers. It is served with pumps integrated in the chiller units. The secondary circuit is provided with one primary circulator and one backup. In case of leakage of the cooling liquid or loose of the pressure the system is equipped with automated expander machine with storage tank. This system is also used for degassing. To maintain uninterruptible cooling capacity, there are 5 accumulation tanks with total volume of 4000 liters with a backup running capacity about 5 minutes. Cooling of the liquid tanks is ongoing and it is provided with 10% of total cooling performance.

The server room is designed with two areas (high density and standalone IT system and it is using the row-oriented cooling in both of them.

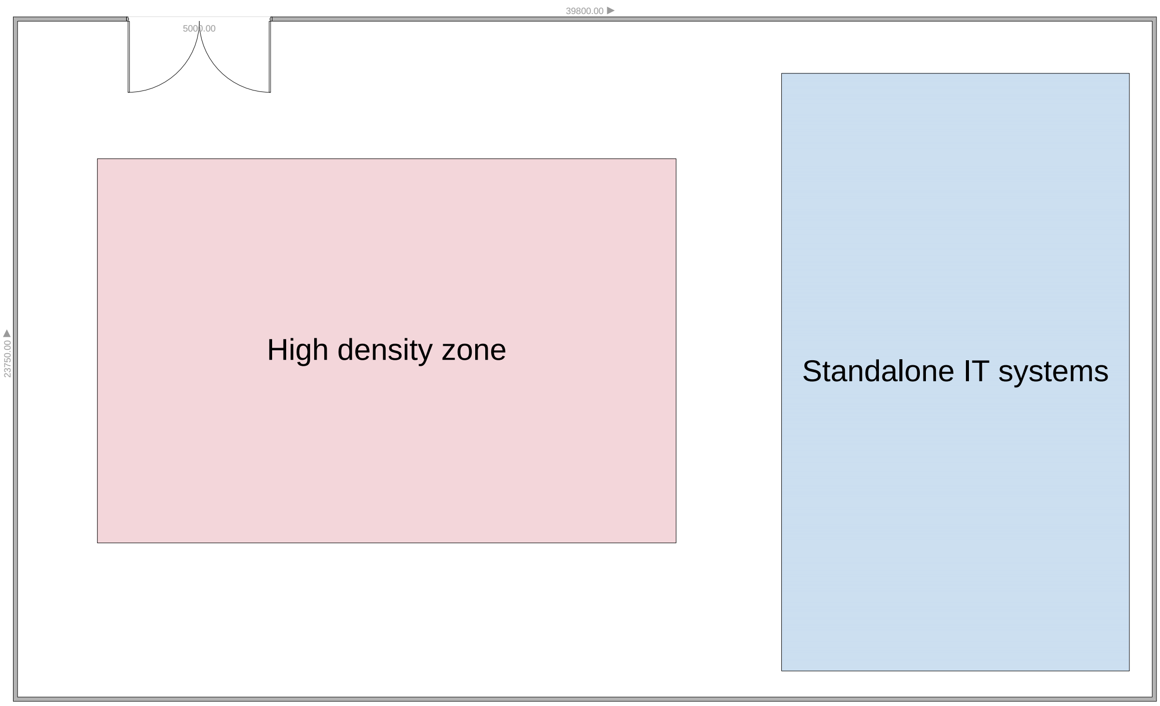


Figure 22 – Server room

Row-oriented architecture used in the Hi-density zone is combined with the cold aisle concept shown in Figure 23 and Figure 24. High density zone is equipped with 12 server racks, 2 cabinets for network equipment and 8 heat exchange units Rittal LCP Inline (each 28kW).

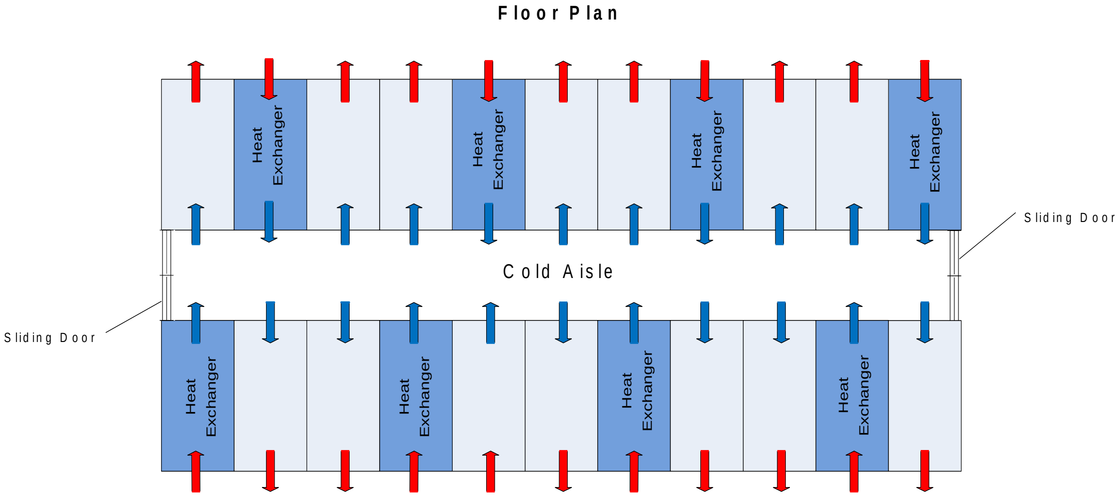


Figure 23 – High density zone: Floor plan

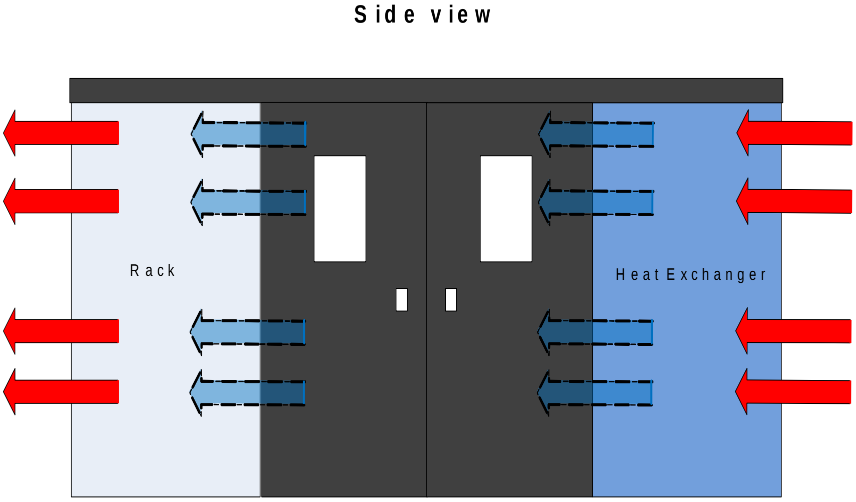


Figure 24 – High density zone: Side view

The air from the hot aisle is extracted by the heat exchangers units and the chilled air (approximately 18 degrees of Celsius) is blown into the cold aisle.

Standalone zone hosts IT systems with non-standard dimensions and also storage racks. There are 6 Rittal LCP Inline (28kW) units placed as well for direct cooling of this zone using the row-oriented cooling. The hot/cold aisle layer is used as well, but regarding some non-standard dimensions of the cabinets there is no closed space. This lowers the efficiency of the cooling, but this IT equipment such as storage does not produce so much heat and so the lower effectiveness can be ignored.

### 6.1.4 UPS

UPS systems are one of the most concerned systems while building green data center. In ways of inefficiency, there are two main contributors. First one is the inherent losses of the UPS modules. The second one is the implementation such as choosing the right product, redundancy etc.

#### *6.1.4.1 UPS efficiency curve*

Most of the manufactures are providing information about the efficiency of the UPS systems at 100% load because this is the level, where the UPS has the very best efficiency. Unfortunately the approximate load of UPS of medium to large scale data center is around 30%. The state of the art UPS systems have 93-98% efficiency. Figure 25 shows the efficiency curve of UPS created by measuring the input and output power at different load levels. At 30% load the efficiency reaches 89%, but it is in a case of single UPS. In cases of data center, which are using redundant systems (2N), the efficiency drops even more due to the load splitting between both UPSs. Regarding these facts the efficiency can decrease to 82%.

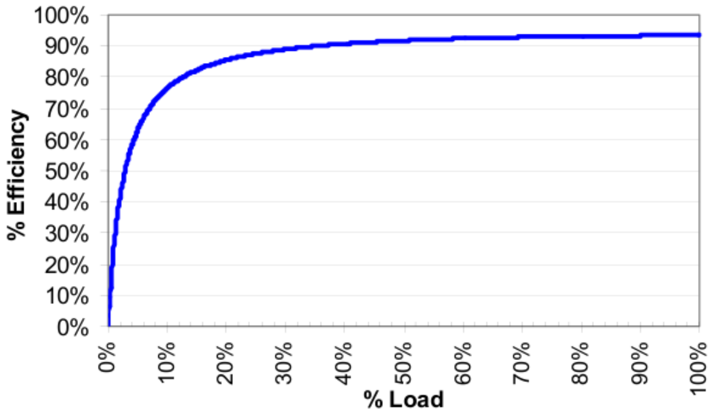


Figure 25 – UPS efficiency curve [source: http://www.zonit.com/storage/ric-product-pictures/ups_energy_efficiency_curve.jpg?__SQUARESPACE_CACHEVERSION=1286046501444]

#### *6.1.4.2 UPS Technologies*

**Offline / standby**

Offline UPS is the simplest solution and offers just surge protection and battery backup. The load is connected directly to incoming utility power and when the voltage falls below specific level it is switched to the internal DC to AC inverter which is powered from a battery.

### Line-interactive

This is a similar construction to the offline UPS. The difference is in using multi-tap variable-voltage autotransformer. This autotransformer can increase or decrease the output voltage by adding or subtracting powered coils of wire. This UPS type can handle continuous undervoltage and overvoltage surges without the need of using battery power.

### Double-conversion / online

The main advantage of the online UPS is the ability to provide a safe and stable source for the power fluctuation sensitive equipment. Basically there is no bypass and switches. The AC input voltage is always transformed to DC, charging the batteries and then inverted back to AC, supplying the output with stable power. This construction costs usually much more due to the need of AC-DC battery charger/rectifier designed for high currents.

### Hybrid topology / double conversion on demand

Hybrid UPS operate as offline UPS, when the power conditions are within set conditions. This allows high efficiency. When the conditions fluctuate UPS is switched to online conversion mode and the UPS is able to adjust the voltage without using battery power, filtering the noise and controlling frequency.

### Rotary DRUPS

The basic principle of the Diesel Rotary UPS is a high-mass spinning flywheel, which provides short term power supply in the event of possible power loss. Typically 10-20s protection is provided, giving time to the diesel generators to start running. The start of the engine is performed either separately, or by transmission gearbox and then the engine is directly spinning the flywheel back, making sure the power supply will not fail. Multiple flywheels can be paralleled without the need for separate engines or to extend the time. Usage of DRUPS is usually when more than 10kW is needed. They also have greater lifecycle (up to 30 years) because of the absence of batteries. Due to the mechanical principle they require periodic downtime for mechanical maintenance, mostly for changing the ball bearing. In some DRUPS the electromagnetic cushion is used to “levitate” the body of the flywheel to prevent any damage to the ball bearings.

DRUPS can be also short time overloaded up to 1000% of nominal performance, which is 17 times more durable to short circuit than typical electronic UPS systems. Considering the environment, DRUPS are much friendlier because of the lack of plumb batteries, which are usually replaced in about five years.

The efficiency of this system is about 98%. Continuous voltage correction is also big advantage of the DRUPS. Basically the utility power is used to keep the generator spinning and thanks to the mass of flywheel the generated voltage is constant. That is why this system can provide power factor correction, distortion filtering and also protection against pikes and other anomalies.

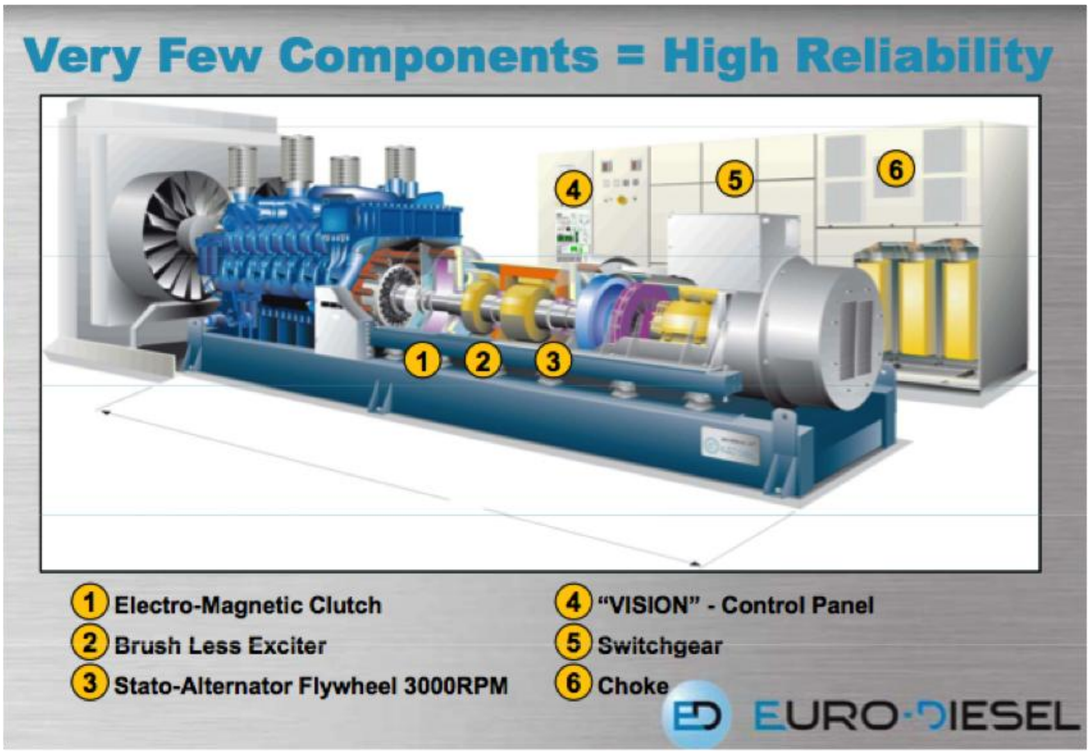


Figure 26 – DRUPS structure [source: http://blogs.cisco.com/wp-content/uploads/DRUPS2-550x378.png]

### Air-DRUPS

This type of UPS was first developed by Pnu Power. The principle is pretty easy as shown in Figure 27. The main part of the Air-DRUPS is scroll expander, which rotates when compressed air is applied. Using a DC generator, its rotation is converted to DC power. Speed of the expander is regulated by adjusting the pressure applied to the scroll expander.

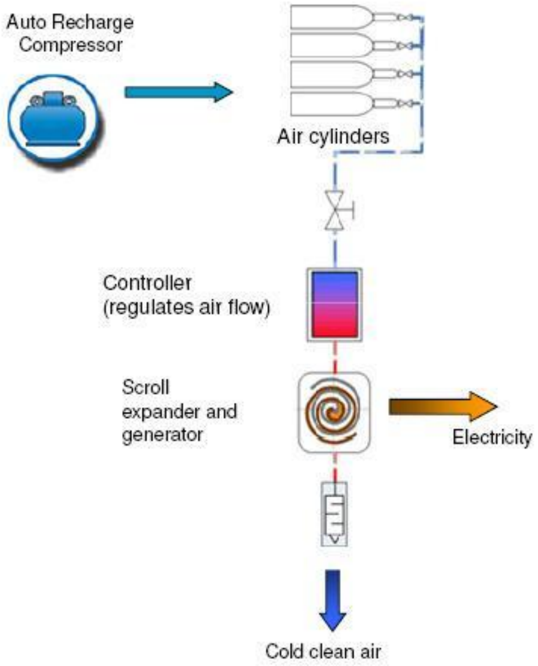


Figure 27 – Air-DRUPS principle [source: GOODWIN, Andrew; DERBY, James. The Road Map Towards 2020 : Eliminating Battery Back-up Power with the Use of Compressed Air [online]. [s.l.] : [s.n.], 2011 [cit. 2011-06-09]. Available on WWW: <pnu-power.com>.]

Air-DRUPS can be producing power in 50 milliseconds but due to the possibility of damage of sensitive equipment, capacitance is added to the DC bus to provide the voltage during the startup.

Probably the biggest advantage is low no-load losses, which make it a great backup source for datacenter not running at full capacity. The possibility of being installed outside removes the need for expensive placing and securing in buildings. Information about the principle of Air-DRUPS was taken from [8] and more information can be found on pnu-power.com.

#### *6.1.4.3 UPS designs*

Information about UPS designs are based on [9].

##### 6.1.4.3.1 "N" (Capacity, non-redundant) system

An N system is a system of modules, which matches to the critical load. It means that for 400kW computer room there would have to be one 400kW or two 200kW UPSs. Larger UPSs have different design. Each UPS module has internal static bypass switch which allows transferring the load to utility source if the UPS module has internal problems. Points of bypass are selected by manufacturer to provide the best possible protection for critical load and guarding the module against self damage. Static bypass can be used in various situations and are different for each model. Here are two examples of static bypass: While performing maintenance or for example when the overload is detected, the UPS can handle 10 minutes overload at 125%. When the internal timer runs out and the load is not lower, it will automatically transfer load to static bypass.

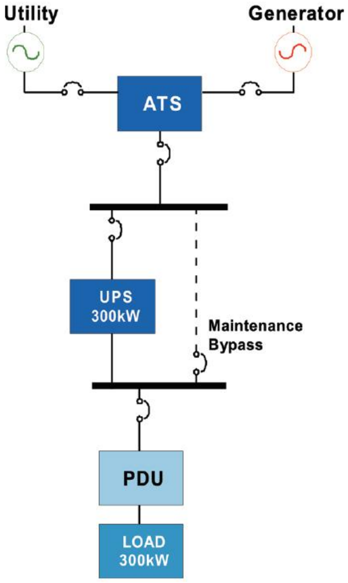


Figure 28 – Single module capacity UPS configuration [source: http://www.todaysfacilitymanager.com/assets/images/2010-Issues-Jan-Dec/07_Jul/1007-fm-issue-fig1.jpg]

**Advantages:**

- Simple concept, low costs
- High efficiency of the UPS because of high capacity utilization
- Expandable (multiple units can be used, depending on manufacturer, 8 UPS modules can be usually used in parallel)

**Disadvantages:**

- Limited availability in case of UPS failure (load transferred to bypass using unprotected power)

- During maintenance the load is exposed to unprotected power as well (2-4 hours a year)
- Limitation in load protection of the UPS due to no redundancy
- Single point of failure (can be the weakest point)

#### 6.1.4.3.2 Isolated redundant (N+1)

This system is referred as N+1 however it is different from parallel redundant, which is also referred as the N+1. Isolated redundant does not require the modules to have the same capacity, nor requires a paralleling bus. It is based on primary UPS module that feeds the load in normal conditions. The secondary UPS is providing the power for primary UPS bypass circuit. The principle is illustrated in Figure 29. This solution is useful for upgrading from non redundancy configuration without need of complete replacement the existing UPS, however separate input for static bypass circuit is needed at the primary module.

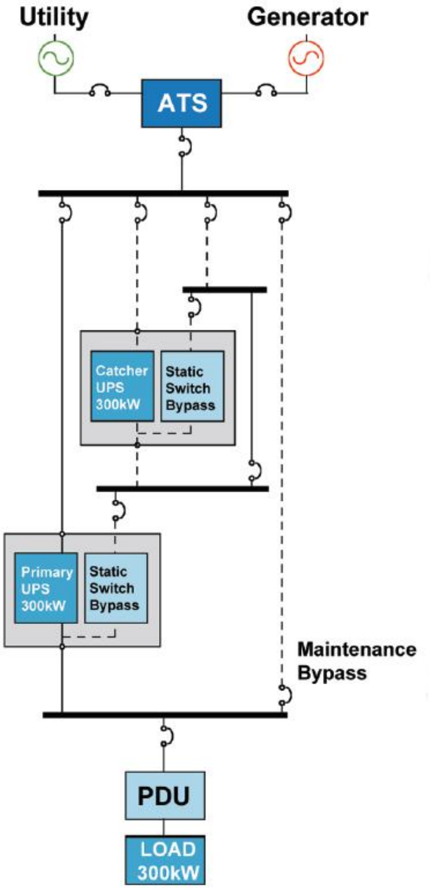


Figure 29 – Isolated redundant UPS configuration [source: http://www.todaysfacilitymanager.com/assets/images/2010-Issues-Jan-Dec/07_Jul/1007-fm-issue-fig2.jpg]

In case of normal operation, primary UPS is feeding the load and the secondary UPS module is completely unloaded. If primary module fails, the load is transferred to static bypass and the secondary UPS will accept the full load. Therefore the secondary module has to be chosen correctly to ensure being capable taking the full load this rapidly. In other case the load would be transferred to the static bypass and supplied by unprotected utility power (if available).

Single point of failure still exists and the system needs to be shut down for 2-4 hours per year for maintenance. The switches and control systems are the most probable point of failure. Regarding the MTechnology consultants (specializing in high reliability power systems) both systems non-redundant (N) and isolated redundant (N+1) gave the same unreliability of 1.8% per year of operation when human error, component aging and environmental effects were ignored. The results are 30 failure modes in case of isolated redundant model vs. 7 failure modes in case of non-redundant system. Probability of those 23 failures is small, nevertheless it proves that complexity rises number of failure modes.

Advantages:

- Flexible use of UPS modules (no need of same capacity or manufacturer)
- UPS fault tolerance
- No synchronization

Disadvantages:

- Need of proper primary UPS bypass operation
- Secondary UPS module capable handling sudden load when primary module transfers to bypass.
- Complex switchgear and associated controls
- Low energy efficiency due to 0% load of the secondary UPS
- Single point of failure (single load bus)

#### 6.1.4.3.3 Parallel redundant

Parallel redundant configuration consists of multiple same capacity and manufacturer UPS modules connected to one output bus. The spare amount of power has to be at least equal to the capacity of one system module to be N+1 redundant. Paralleling board from the manufacturer is required for output voltage synchronization between the UPS modules as well. Usually the parallel bus has a monitoring capacity for displaying the load and other current characteristics. The parallel bus must be able to display how many modules are connected and number of modules needed to maintain redundancy in the system. Maximum number of UPS modules that can be paralleled onto a common bus differs from UPS manufacturers. In parallel redundant configuration UPS modules share the critical load in normal operation situations. In case of failure or maintenance of one UPS module, the other modules have to immediately accept the load of the malfunctioned module. This allows maintenance of any module without connecting the load to the unprotected utility power.

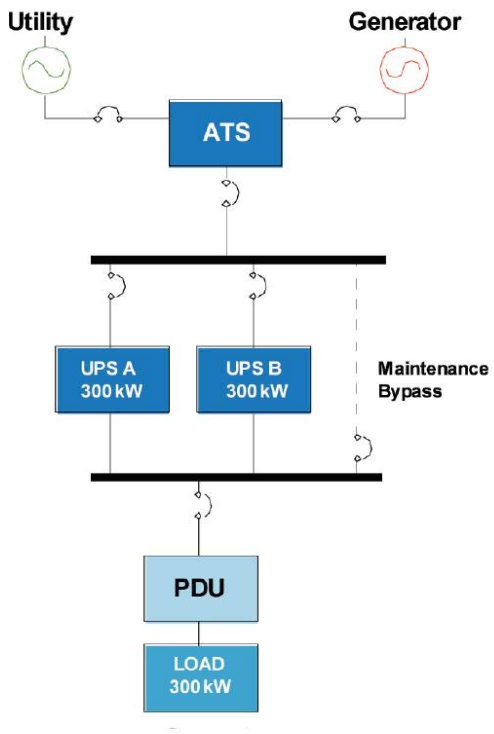


Figure 30 – Parallel redundant (N+1) UPS configuration [source: http://www.todaysfacilitymanager.com/assets/images/2010-Issues-Jan-Dec/07_Jul/1007-fm-issue-fig3.jpg]

System efficiency

As was shown in Figure 25 – UPS efficiency curve, poorly loaded UPS modules are less effective than high loaded modules. Table 9 shows different UPS modules sizes used, all feeding 240kW load. System efficiency can be seriously affected by the chosen modules. These are general values used as an example and real values can differ from each manufacturer.

| UPS Modules in Parallel | Mission Critical Load | Total UPS System Capacity | % of Each UPS Module Load |
|---|---|---|---|
| **2 x 240 kW** | 240kW | 480kW | 50% |
| **3 x 120 kW** | 240kW | 360kW | 66% |
| **4 x 80 kW** | 240kW | 320kW | 75% |
| **5 x 60 kW** | 240kW | 300kW | 80% |

Table 9 – N+1 Configurations [source: MCCARTHY, Kevin. Comparing UPS System Design Configurations [online]. [s.l.] : [s.n.], 2005 [cit. 2011-06-09]. Available on WWW: <http://www.apcmedia.com>.]

**Advantages:**

- Higher level of availability
- Lower probability of failure compared to isolated redundant because modules are online all the time (no step loads)
- Expandable (multiple units configuration in same installation)
- Hardware configuration simple

**Disadvantages:**

- Same design, manufacturer, same parameters and technology and configuration of modules
- Single points of failure before and after the UPS system
- Need of maintenance bypass circuit for maintenance of parallel bus
- Lower operating effectiveness due to lower utilization of each UPS module

#### 6.1.4.3.4 Distributed redundant

Distributed redundant configurations are commonly used UPS system solutions. This design uses three or more UPS modules with multiple PDUs and STSs. This system minimizes single points of failure. Power can be distributed to single-corded loads or to prevent single point of failure dual-corded should be used instead. In case of using single-corded loads, power is distributed with just one STS and STS becomes SPOF. In case that all equipment (including PDUs) is dual-corded, STS can be removed from this design. Distributed redundant systems are usually used for larger solutions, where many loads are single corded and concurrent maintenance is required. STS is a primary weakness of distributed redundant system. STS are very complex devices and can suffer by unexpected failures, from which the worst case is a short circuiting each other. This failure causes two UPSs to be unable to provide power for the load and STS are becoming a SPOF.

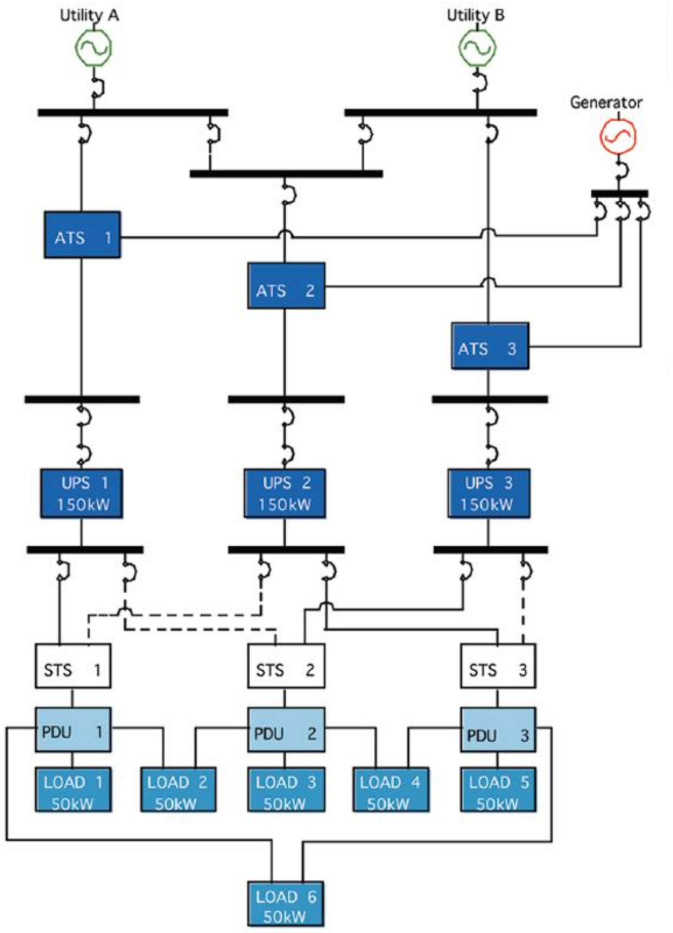


Figure 31 – Distributed redundant UPS configuration [source: http://www.todaysfacilitymanager.com/assets/images/2010-Issues-Jan-Dec/07_Jul/1007-fm-issue-fig4.jpg]

**Advantages:**

- Concurrent Maintenance available if all loads are dual-corded
- Cost savings against a 2(N+1) thanks to lower number of UPS modules
- Two separate power paths for each dual-corded load

**Disadvantages:**

- High cost solution because of a use of many switches

- STS can become SPOF
- Need of multiple source synchronization
- Complex configuration for larger systems
- Harder malfunction detection
- Ineffectiveness of UPSs because of not using on full load

#### 6.1.4.3.5 Systems plus system redundant

This is the most reliable and most expensive design in the field. There are many variations and names of this classification (System plus System, Double-Ended, Multiple Parallel Bus, 2(N+1, 2N+2, [(N+1)+(N+1)]). Using this design the load may be never required to use utility power. System can be designed without any SPOF, but the costs of implementing will rise. While the size of this infrastructure is large, it is common to build these systems in specially designed standalone buildings together with other support facilities such as cooling, electrical distribution or generators).

Designs can vary greatly and it depends on owners requirements, design engineer knowledge and vision how it is going to be implemented. In Figure 32 you can see the 2(N+1) configuration which is based on duplication of parallel redundant UPS systems. To reach high Tier levels it is necessary to distribute power from separated utility powers and separate generators as well. This is sometimes big problem and this solution costs have to be justified by the importance of availability. Costs vary by how deeply duplication is implemented. Each part of electrical equipment can be changed or can fail without transferring critical load to the utility power.

.

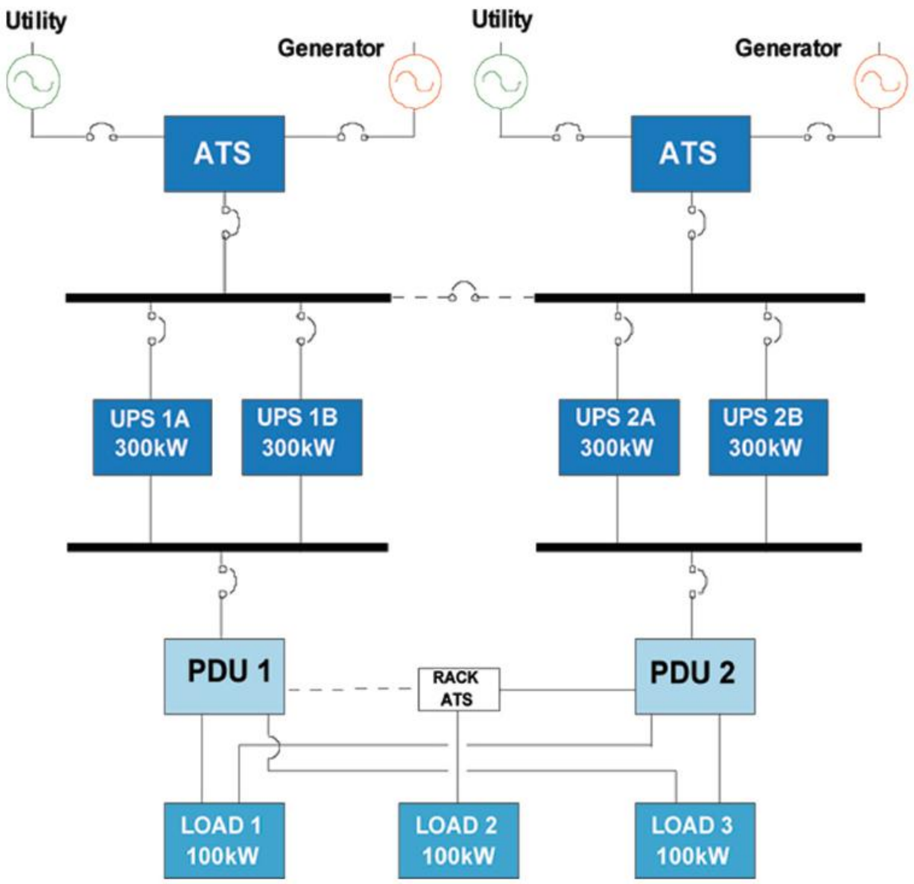


Figure 32 – 2(N+1) UPS configuration [source: http://www.todaysfacilitymanager.com/assets/images/2010-Issues-Jan-Dec/07_Jul/1007-fm-issue-fig4.jpg]

Advantages:

- Very fault proof (two independent power paths for no SPOF)
- Complete redundancy all the way to critical loads even during concurrent maintenance
- Easier configuration and to keep systems equally loaded

Disadvantages:

- Highest cost solution
- UPS ineffectiveness
- Harder to implement in typical buildings

#### *6.1.4.4 Real data center implementation:*

There are 3 independent systems of modular UPS from NEWAVE. All three UPSs are installed in a separate room together with switchboards.

Two UPSs (UPS-A and UPS-B) are supporting all IT equipment. In first phase the installed performance is 150kW with the possibility of enlargement up to 300kW. A redundant module available is always implemented. The backup time is estimated to ten minutes. Each UPS (NEWAVE DPA 250) is consisted of modules with 40kW available. In first phase of installation there will be 4+1 modules available.

The third UPS (UPS-C – type NEWAVE CONCEPTPOWER) with minimum 64kW and redundant module is used for supplying the non IT infrastructure. The backup time is 10 minutes as well.

Batteries for these UPSs are installed in separated battery stands with implemented monitoring system. Each battery is provided with measurement of the voltage, internal resistance, temperature and the discharging current. Temperature of the whole battery section is measured as well. This is very important to maintain the reliability of the whole UPS system and all the measured values can be seen in the monitoring system. Warnings are automatically appearing, when a part of the system reports unusual characteristics. All battery sections for UPS A, B, C will be separately connected for each UPS with separated control unit for each.

This system is designed in the 2(N+1) configuration with one Utility power. There are two generators installed. One is primary, the other one is a backup generator. Each of them can support all installed IT and support systems. In case of utility power interruption, both of them will start, and they will share the load. This will provide the system with security. If the primary generator would fail during supplying and the UPS modules would not have enough power for the IT load till the secondary would start, it would result in major power supply failure, causing the server shut down. This system could also work more efficient if the second generator would be switched off after the UPSs will be charged, providing the generators sufficient time to start running. However the availability of the service is more important than the costs of running spare generator, so both of them remain running. The second reason why to have them both running is the fact that the utility power is reliable and the power does not go down for longer periods of time.

## 6.2 Software

### 6.2.1 Virtualization

Server virtualization brings many advantages and the power effectiveness is just one of many. Data centers are usually designed for peak loads, however normally they operate on really low utilization (from 2-15%). The VMware states that 85-90% of time the processors are idle. When idle, the servers consume almost the same amount of power as they would be utilized.

The principle of virtualization is shown in Figure 33. The virtual infrastructure layer is aware of the physical infrastructure and controls the servers. The unused servers can be shut down for power savings. No application or operating system is aware of the actual physical layer and is provided with the information from the virtual machine.

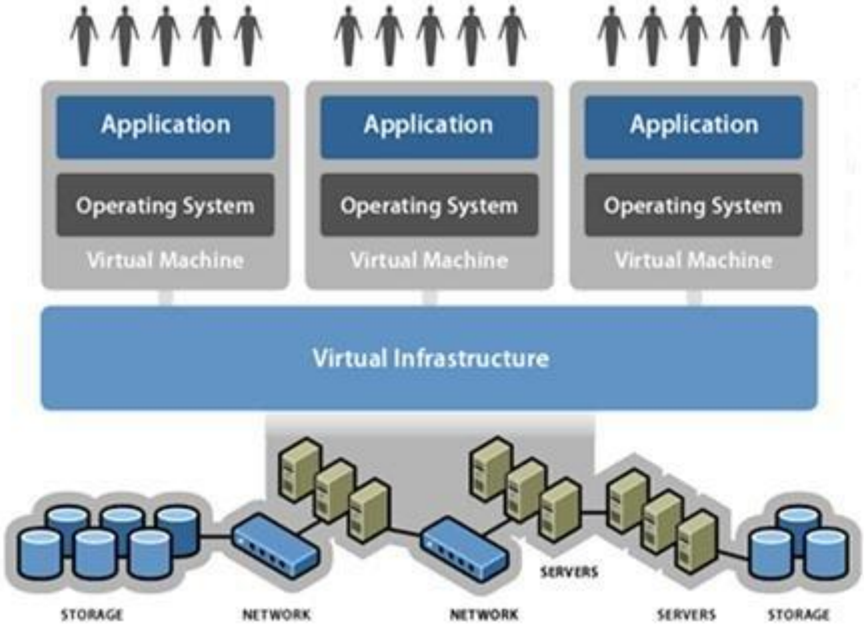


Figure 33 – Virtualization layers [source: http://questivity.com/virtualization_clip_image002_0000.jpg]

This provides the ability to run one application on more servers or more applications on one server, utilizing it to 70-80% average. This will provide us with the option to reduce the servers and dramatically reduce the power consumption also thanks to the fact, that less support equipment can be used to run the data center. The reduction of the energy consumption is typically around 80%.

### 6.2.2 Energy management tools

Various programs are used for better server and energy management. In case of IBM technology, IBM Tivoli is used for monitoring the energy usage, IBM WebSphere is used for application management and virtualization and IBM Systems Director Active Energy Manager is used for monitoring, managing and it can also be integrated with other software mentioned above.

#### *6.2.2.1 IBM Systems Director Active Energy Manager*

Together with mentioned programs and EnergyScale technology provided from the POWER6 systems it can become a great energy saving tool. With the integration in the firmware of the POWER systems these energy savings become even more powerful.

Energy Scale provides continuous information about the real server power consumption. AEM uses this information to predict data center power consumption at different times. Proper actions such as power capping or other features for saving the energy described in the chapter 6.1.1 can be performed by the AEM.

If workloads during longer period of times are appearing server power down should be considered. AEM provides information about the amount of power, which can be saved during these long periods and can be used to determine the right choice. AEM allows this feature (as well as many others) to be scripted and automated.

### 6.2.3 Real data center implementation

Due to the security of the server and confidentiality of these facts, examination of virtualization, either use of System Director Active Energy Manager was not allowed.

# 7 COMPARISON OF TRADITIONAL AND MODULAR DATA CENTERS

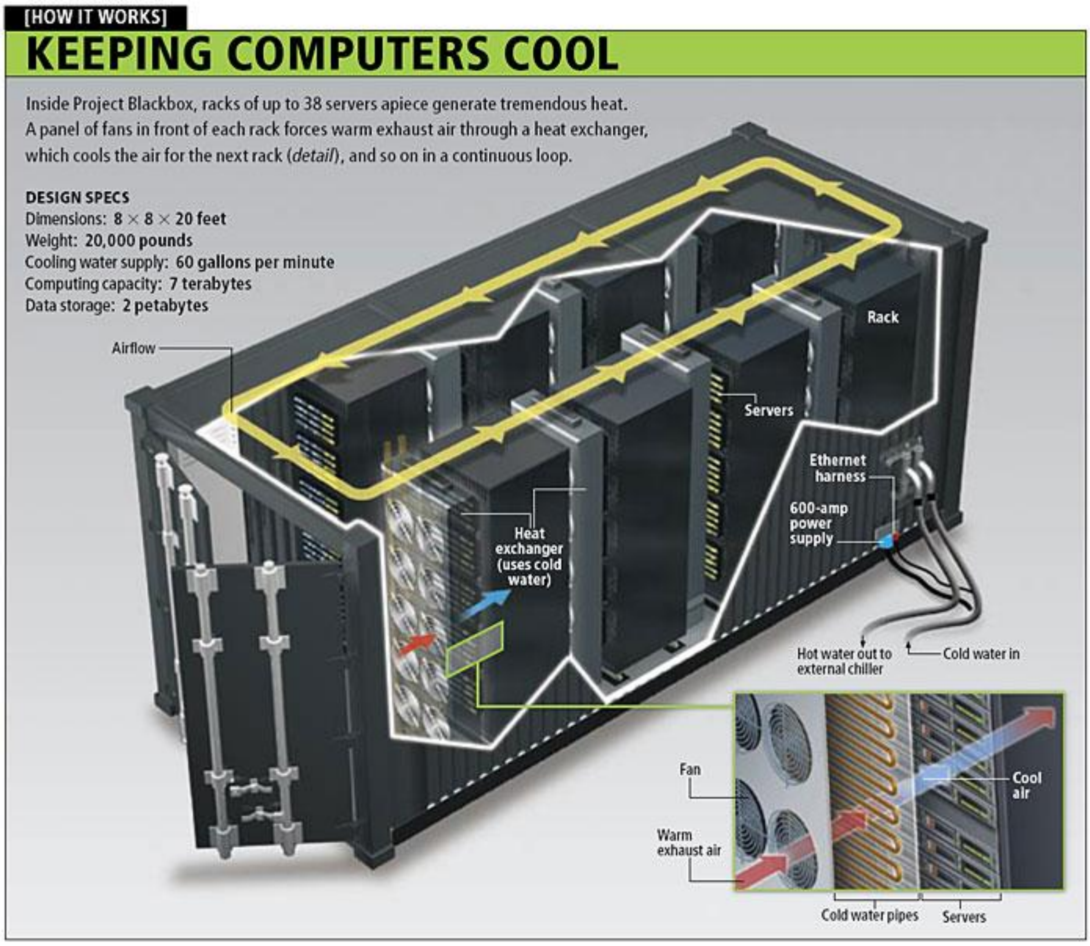
As it was mentioned in chapter 4.1 modular centers (which include modular and container data centers) are getting more popular. Modular data centers can be divided into Modular data centers and container data centers. Some of their characteristics are different so they are described individually.

Common advantages are:

- Fast deployment
- High modularity
- No specialized buildings needed
- Optimized construction (no special planning required)
- Higher efficiency

## 7.1 Container data centers

Container data centers are complex data center solutions installed within standardized 20 or 40 feet shipping container. History of these unique solutions starts in 2005 when Google started with development of their own container data center. This project was abandoned in 2007 and patent was granted to Google for this concept in the same year. In 2006 Sun Microsystems started their own project called the BlackBox, which was successfully finished and publicly presented in 2008. Subsequently other companies started manufacturing the container data centers as well.



[HOW IT WORKS]

KEEPING COMPUTERS COOL

Inside Project Blackbox, racks of up to 38 servers apiece generate tremendous heat. A panel of fans in front of each rack forces warm exhaust air through a heat exchanger, which cools the air for the next rack (*detail*), and so on in a continuous loop.

**DESIGN SPECS**
Dimensions: **8 × 8 × 20 feet**
Weight: **20,000 pounds**
Cooling water supply: **60 gallons per minute**
Computing capacity: **7 terabytes**
Data storage: **2 petabytes**

Figure 34 – BlackBox: internal layout [source: http://www.scientificamerican.com/media/inline/B1027B68-E7F2-99DF-352186A04761EB7F_4.jpg]

The concept usually differs from each company. Sun Microsystems have all the required equipment except chillers built in one container. Other companies like IBM have also each container assigned with specialized function such as storage, UPS modules and all are connected together (shown in Figure 35). Each solution has its own advantage and the choice depends on needs of each individual customer.

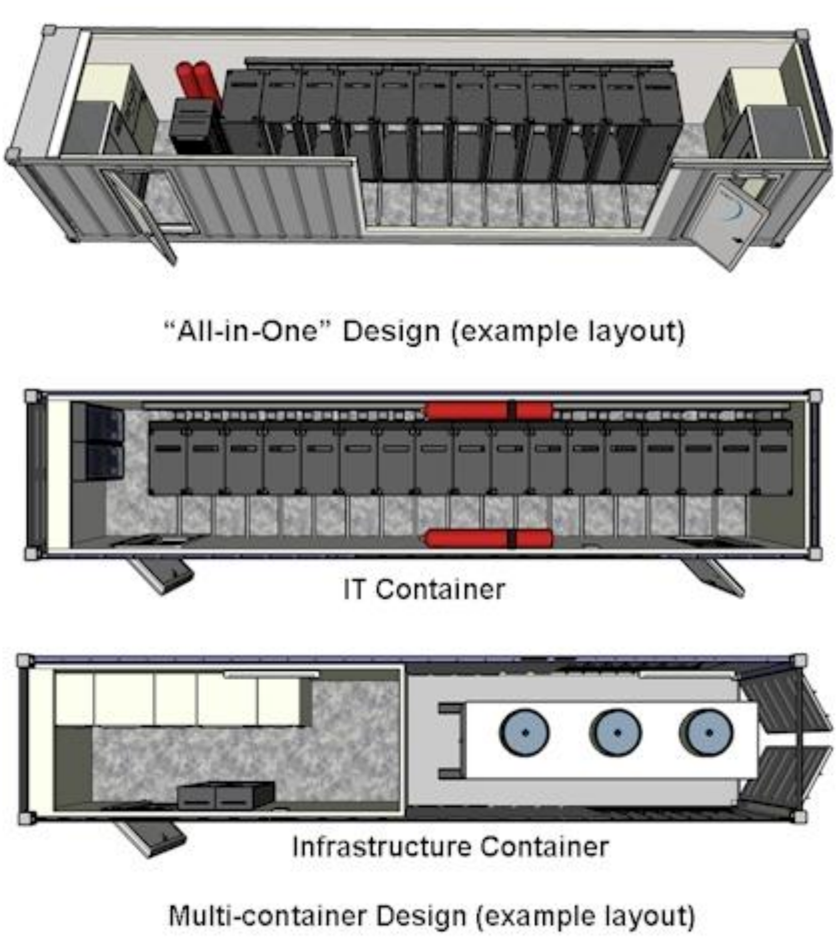


Figure 35 – IBM All-in-One and Multicontainer layout [source: http://regmedia.co.uk/2009/12/07/ibm_pmdc_layout.jpg]

Besides the common advantages mentioned above, container data centers are highly mobile. They also do not need any kind of specialized building to be deployed in, but due to the security reasons, they are usually placed in warehouses or similar purpose buildings. The density in container data centers is usually twice higher than in traditional data center with even higher efficiency.

The deployment time is probably the fastest one thanks to standardization of construction and high mobility (no need of special means of transportation). IBM claims that they can deploy data center between 500 and 2500 square feet in between 8 and 12 weeks, compared with 24-36 six month of traditional data center.

Modularity of container data center is probably the highest of all solutions. The possibilities of enlargement are limited only by available space and the availability of the support systems.

Container data centers can be also used for large facilities. In 2009 Microsoft has started using a large container data center in Chicago. This enormous facility will host 56 forty feet long containers in the first phase. In second phase 112 container data centers will be placed in this facility providing 224 000 servers running. As the Microsoft team says, if additional capacity is required the container can be completely installed within 8 hours.

## 7.2 Modular data centers

Modular data centers can be described as a hybrid model of traditional "brick" data center and the container data center. It is built from prefabricated blocks, which are connected together. The modularity can be demonstrated in Figure 36 on HP "Butterfly".

The name is derived from the shape, which consists of central networking core, four server rooms and power and cooling infrastructure connected on the edges of each server room. If not all four server rooms are required and the capacity growth is expected, customer can buy even one server room with support systems and networking core only. Later when additional capacity is required, other modules can be added. Thanks to the standardized prefabricated blocks it can be quickly deployed and HP claims that it will be also cheaper than building a brand new traditional data center.

The advantage of this hybrid model is the possibility of much greater customization of power and cooling choices of each server room. In case of container data center the solution is complete and customer cannot decide for example if raised floors will be used, as it has to be compressed into a very small area. HP also claims that their state of the art energy-optimized design can result in PUE equal of 1.18, which is comparable with the very best and largest facilities such as Google data centers (based on [10]).

Figure 36 – HP butterfly modular data center [source: http://www.datacenterknowledge.com/wp-content/uploads/2010/07/hp-butterlfy-top.jpg]

## 7.3 Conclusion

Even when modular data centers show their big advantages, traditional data centers are still one of the most used solutions. Therefore summarization of typical facts for decision of the data center construction is following.

Traditional data centers are usually built when a appropriate building is available within the company. It has to be more sophisticated in the planning phase. The growth of business and also the data center capacity has to be carefully considered. If large capacity is installed and not used, the efficiency of the data center rapidly decreases and the energy costs are very high. This is also problem of the examined data center. The data center is running on 20% of its maximum designed capacity. Even when not all UPS modules are presented, the inefficiency of the poorly loaded chiller systems and other systems are increasing the planned PUE=1.5 to PUE 1.9. This results in much higher energy costs.

Modular data center in general are great solution for building brand new stand-alone data center (considering even new building). It will provide the lower costs, effectiveness, much faster deployment and very high modularity needed for growing business.

Container data center are well suited for the same purposes as modular data centers in general, but their biggest advantage is their mobility. When company wants to save energy costs, the whole data center can be transported basically anywhere in the world following the needs of the company. Renting a container data center became also partially advantage of this system. If company needs to temporary increase the capacity of their data center, container data center is thanks to its complex solution the simplest way to provide this capacity.

Hybrid solutions are great solutions for new stand-alone data centers, when customizing of the data center construction is required. More choices can be done also with the support systems as well. Customer can choose between four choices of cooling systems and five different UPS configurations using both DRUPS and battery-based UPSs.

Each solution can be also used for large projects. Google is mostly building large traditional data centers facilities. In 2009 Microsoft has started using enormous container based data center in Chicago. Both of these facilities are very efficient and efficiently comparable. This proves, that no design is going to be replaced soon by the other.

# 8 COMPARISON OF OLDER AND RECENT TECHNOLOGIES REGARDING PERFORMANCE, ENERGY CONSUMPTION AND RETURN ON INVESTMENTS

## 8.1 Processors

This comparison will compare POWER7 processors against other manufacturers and their models. It is based on information available in the beginning of the year 2011.

The biggest competitor is most probably the Intel Nehalem EX. Which was released half a year earlier then IBM POWER7 processor family. Comparison of their main attributes is summarized in Table 10.

| Parameters | IBM POWER7 | Intel Nehalem EX |
|---|---|---|
| **Manufacturing technology** | 45nm | 45nm |
| **Number of cores** | 4,6,8 | 6,8 |
| **Frequency** | ~3.5 to ~4 GHz | ~2.7 GHz |
| **Cache** | 32MB eDRAM L3 | 24MB L3 |
| **Threads /core** | 4 simultaneous | 2 simultaneous |
| **Number of sockets possible** | Distributed switch scale to 32 sockets | Scale directs to 4 sockets |
| **Virtual machines** | >1000 threads | Max. 8 threads |

Table 10 – IBM POWER7 vs. Intel Nehalem EX main characteristics

Distributed switch technology makes the communication between the nodes much faster and efficient while heavily loaded than the cell board architecture. The difference is that cell board architecture uses bottleneck to connect the chips/cores and power distributed switch connects the cores directly.

### 8.1.1 Money savings

The effectiveness is from 4 times to 7 times better then Sun SPARC and HP Integrity. IBM also claims that 83% energy savings can be achieved with use of IBM Power 750 instead of a 64-core HP Integrity Superdome (HP Integrity superdome -11 586W, IBM Power 750

Express 1950W). While these savings are achieved, there is also 28% increase of performance (based on information from [11]).

The energy use difference between POWER5 and POWER7 family is up to 90%. Large saving can be achieved also thanks to the software licensing as the licensing is usually connected with the number of cores running. As 24 POWER7 cores can substitute 64 POWER5 cores, it can save more than one million USD. Maintenance is also contributor to the complete evaluation. The POWER7 family can save up to 600 000 USD in this field compared with the POWER5 family. These information are based on [12].

The prices of the processors vary for each customer. Usually the Intel processors are twice cheaper than the IBM. No matter how attractive this simple comparison is, all the advantages and higher performance of the IBM family must be considered too. In the end IBM processor wins.

It is also very important to mention, that the IBM family was released half a year later than the Intel and therefore we can expect the Intel to deliver new processor family earlier and another comparison will have to be done.

## 8.2 UPS

More efficient UPS systems are developed continually. For example the use of DRUPS is slowly increasing, but as they are quite a new solution in the market (compared with the battery-based) and they usually require more space available for their installation, the number of them is not increasing so rapidly. The long life of DUPS is a great benefit, but the life cycle of the DRUPS is almost twice more than the average data center. They can be also used while building new data center, however different power capacity will be most probably required and therefore the DRUPS may not fit in the new data center.

## 8.3 Cooling

Actual means of cooling are not changing much. Using free cooling is very popular nowadays, thanks to large power savings. Free cooling takes the advantage of cold outdoor air (around 3 degrees of Celsius) for cooling the chilling liquid without using the compressor unit. Dynamic free cooling extends these abilities. The principle of cooling itself remains the same, but it can operate in higher temperatures, while the data center utilization is not high. Water used for air-conditioning units has approximately around 10 degrees of Celsius. If the IT equipment is less utilized (around 60%) the required cooling capacity is also 60%. That means that that the air conditioning system can provide in free cooling mode the 60% cooling capacity from cooling liquid around 15 degrees of Celsius. This cooling liquid temperature can be achieved even with 17 degrees of Celsius of outdoor air. This highly increases the amount of days, when free cooling can be used. Google data center in Belgium is completely based on free cooling. If the weather conditions does not allow the free cooling, some or whole load can be transferred to other facilities in the world.

Another solution used in large facilities is evaporative cooling, which needs even less than one third of electric power less that is consumed by conventional chillers. The principle is shown in Figure 37. The incoming air is pushed through evaporative pads, cooled down and then distributed back to the air distribution system. These systems are widely used in state of the art large data centers of Google, Yahoo etc. However this system decreases the electric energy consumption it brings the disadvantage of large water consumption. This will result in better PUE, but the WUE is getting worse.

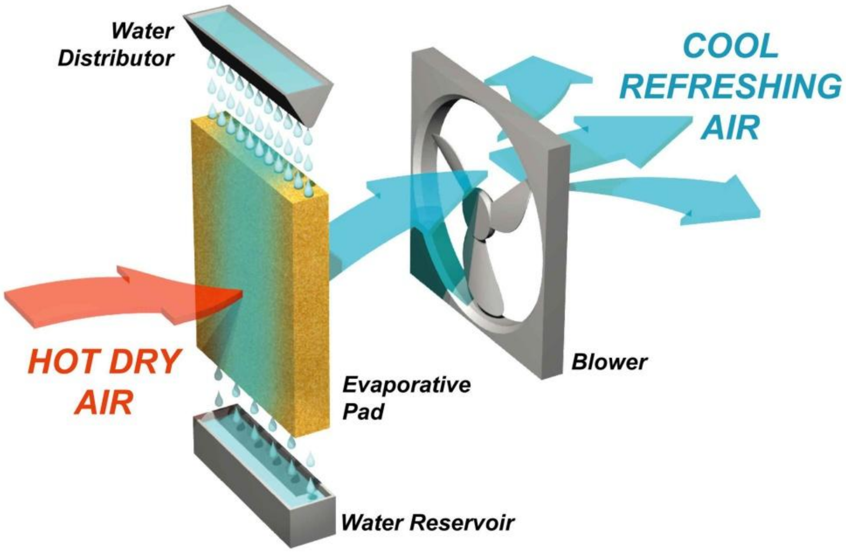


Figure 37 – Evaporative cooling principle [source: http://www.estexpert.com/image/mypic_customize/How-Evaporative-Cooling-Works.jpg]

## 8.4 Virtualization

The virtualization itself is a great tool for increasing the power efficiency. As well as in other fields there are new possibilities evolving. IBM PowerVM using POWER7 family seems to be the most promising one even when compared with the well known VMware virtual machines. IBM claims, that PowerVM outperforms VMware by up to 65% using the Power 750. Many advantages of PowerVM are related to the integration of PowerVM directly into the firmware of power systems servers.

Benchmark results prove that combination of PowerVM and POWER7 platforms gives much better result than VMware vSphere on Intel x86 platforms. The main advantages of PowerVM are higher performance, scalability and flexibility of the system.

| **Parameters** | **VMware ESX 4.0 (in VMware vSphere 4)** | **PowerVM** |
|---|---|---|
| **Virtual CPUs per VM** | 8 | 64, or 256 on POWER7 |
| **Memory per VM** | 255G | 4096G and even more on POWER7 |

Table 11 – VMware and PowerVM comparison

Dynamic changing of the number of virtual CPUs and memory are also important. The PowerVM dispose of complete functionality of these features. On the other hand VM ware is able to increase the number of CPUs and the amount of memory, but it is not able to remove them. The number of live migrations of PowerVM is twice more than in case of VMware (8 vs. 4).

There are also more features where the PowerVM shows their better qualities, but let's do not forget that mainly all the advantages are profiting from the integration and better cooperation among the IBM products. VMware is third party software available for more platforms and therefore it is harder to decide which system to use. Comparison was based on information available in [13].

# CONCLUSION

The goal of this thesis was to summarize the actual data center power consumption trends. Therefore main power consuming systems were analyzed and compared. Most of the analyzed systems were hardware however software capabilities and interconnections with hardware were also mentioned. Implementations of some systems were compared with the previously analyzed systems as well.

At first the energy efficiency importance together with explanation of the main attributes used for evaluation of data centers energy efficiency was summarized. There are many attributes related to the data center efficiency and environmental burden, but for purposes of this thesis PUE was the most important one. During the description of construction of data centers and Tier levels, it was discovered that not only server room solutions have to be considered, but also a lot of attention has to be paid to other aspects such as the construction and placement of used buildings and its parameters.

During elaboration of the chapter “Power Management in the Network Infrastructure“ it was discovered, that nowadays systems are running in always-on mode and no advanced power saving techniques are used. Due to this fact the concept of ElasticTree, which is currently under development was brought up in chapter 5. Due to its concept stage, issues of network equipment power management are going to be elaborated on my possible future dissertation work.

To the issue of power management in the server infrastructure many more information was available. Therefore focus was given on the hardware and software solutions. In hardware section main contributors were mentioned together with possible power savings in their fields. Processors IBM POWER7 were described together with their functions related to low power consumption. According to the available information, there are no better processors both in performance and power savings for server use in these days.

As cooling is one of the main contributors to the data center power consumption, large part of chapter 6 was dedicated to this topic. First importance of precision cooling system for data centers was shown and proved that comfort systems are not suitable for data center purposes. Methods of calculating of total cooling requirements for data center were examined as well. Cooling architectures together with their advantages and possible uses were described in the end. There is no universal architecture for best performance and usually mixed architecture is the best solution.

UPS systems are also big contributors to the data center inefficiency. UPS technologies were described in relation with energy effectiveness. New trends in UPS use such as Diesel Rotary UPSs are the most effective solution however they do not fit every installation requirements. Efficiency and reliability is highly affected by used UPS designs as well and therefore they have been mentioned as well. For each data center some compromise between reliability, budget and efficiency has to be considered.

In software section virtualization effects and energy management tools and their possibilities were elaborated on. The software tools are still evolving and new features are getting available.

In chapter 6 traditional and modular data centers were compared. The assignment of this thesis was to compare only container and traditional data centers, but as the hybrid modular data centers are beginning to gain share in the market, they were examined as well. As it was not possible to get specific studies about exact comparable data centers, this comparison was made on the basis of materials from different manufactures. Each design fits different requirements. That is why in the end of the chapter, advantages and disadvantages of each design are highlighted together with possible usage.

The final part of this thesis was dedicated to comparison of older and recent technologies in their performance, energy consumption and return on investments. Unfortunately, specific business case could not be examined due to the confidentiality of individual customer prices. However, partial comparison was possible thanks to publically available IBM's promotional materials, where IBM processors are compared with other manufacturers. Based on these materials, the IBM processors clearly win above others and not only in their own category, but they can also beat some processors with more sockets. New trends in UPS and cooling fields were mentioned as well. In case of cooling, descriptions of nowadays most promising methods (free cooling and evaporative cooling) are explained. Last part is dedicated to comparison of Virtualization solutions from IBM and VMware. IBM PowerVM is clearly better, but its success is dependent on much better cooperation with IBM POWER technologies, so it is not suitable for all platforms.

The existing data centers are very strictly secured because business of many companies is dependant on their 24/7/365 availability. Combined with the fact, that all of the used technologies are also extremely expensive, it is hard to actually test any technology in real implementation. In this thesis real data center implementations of some technologies are mentioned. More information would bring better insight to the whole system, but this information was confidential and part of the “know-how“ and cannot be published. Unfortunately even tests in laboratories were not possible due to their high utilization.

The data center construction is fast-evolving field and new solutions are provided very often. To understand the state of the art solutions it is necessary to continuously gain new information from specialized communities and white papers provided from companies.

# ZÁVĚR

Cílem této práce bylo shrnutí nynějších trendů spotřeby datových center. Proto byly analyzovány a srovnány nejvíce energeticky náročné systémy. Většina analyzovaných systému byla hardware, avšak možnosti softwaru a vazeb na hardware zde byly taktéž zmíněny.

Prvně byla vysvětlena důležitost efektivity s hlavními parametry používanými pro zhodnocení efektivity datových center. Existuje hodně možností hodnocení datových center vzhledem k jejich efektivitě a zátěži životního prostředí, avšak pro účely této práce byl nejdůležitějším parametrem PUE. Během popisu konstrukce datových center a Tier úrovní bylo zjištěno, že velká pozornost musí být věnována nejen konstrukci serverové části, ale i dalším aspektům jako například konstrukce, parametry a lokace použitých budov.

Během vypracovávání kapitoly řízení spotřeby elektrické energie v infrastruktuře sítí bylo zjištěno, že dnešní síťové systémy jsou používány v tzv. „vždy zapnuto" režimu a reálně nejsou používány žádné pokročilejší metody řízení spotřeby. Vzhledem k tomuto faktu byl popsán ElasticTree koncept, který je nyní ve fázi vývoje. Vzhledem ke konceptuální formě tohoto systému budou otázky řízení spotřeby síťového vybavení pomocí této metody rozvedeny v rámci mé budoucí disertační práce.

Pro vypracování kapitoly řízení spotřeby bylo k dispozici více materiálů a soustředilo se na řešení na úrovni hardware a software. V hardware části byli zmíněni hlavní přispěvatelé společně s možnými úsporami energií v jejich sektoru. Funkce procesorů IBM POWER7 pro řízení spotřeby byly popsány. Dle dostupných informací se v nynější době jedná o nejlepší procesory pro serverové využití a to jak po stránce výkonu, tak i spotřeby energií.

Vzhledem k faktu, že chladící systémy jsou jedním z největších energeticky náročných zařízení v rámci serveru, velká část 6. kapitoly je věnována právě jim. Jako první obhájena nutnost použití precizních chladících systémů v rámci ekonomičnosti sytému a bylo prokázáno, že standartní chladící jednotky nejsou vhodné pro použití datových centrech. Metody výpočtu celkových chladících nároků pro datová centra byly též prozkoumány. Následně byl proveden rozbor chladících architektur spolu s jejich výhodami a možnými využitími. Bylo zjištěno, že neexistuje universální architektura chlazení, která by byla vhodná pro všechny systémy a proto je většinou nejlepším řešením kombinovaná chladící architektura.

UPS systémy patří také do kategorie velkých přispěvatelů spotřeby energie. Technologie UPS byly popsány v souvislosti s efektivitou. Nové trendy jako použití diesel rotačních UPS jsou nejefektivnějšími řešeními, avšak tyto systémy nemusí splňovat všechny požadavky pro konstrukci některých datových center. Efektivita a spolehlivost jsou vysoce ovlivněny také designem UPS systémů a proto byly možné zapojení taktéž popsány. Při návrhu datového centra musí být vždy zvolen určitý kompromis mezi spolehlivostí, efektivitou a náklady.

V sekci softwaru byly popsány možnosti a důsledky virtualizace a programů pro správu energií. Softwarové nástroje pro řízení spotřeby se neustále vyvíjí a stále přibývají nové funkce.

V 7. kapitole bylo provedeno srovnání tradičních a modulárních datových center. V zadání této práce bylo srovnání kontejnerových a tradičních datových center. Vzhledem k faktu, že hybridní modulární datová centra si získávají své místo na trhu, bylo provedeno i jejich srovnání. Vzhledem k faktu, že nebylo možné získat data pro srovnatelná datová centra těchto různých typů, srovnání bylo provedeno na základě materiálů od různých výrobců. Na konci kapitoly je provedeno srovnání jejich výhod a nevýhod každého designu společně s možnými použitími.

Poslední část této práce byla věnována srovnání starších a novějších technologií ohledně výkonu a spotřeby, včetně návratnosti investic. Bohužel vzhledem k důvěrnosti informací o individuálních cenách pro každého zákazníka nemohlo být provedeno srovnání na konkrétním případu. Proto bylo provedeno srovnání IBM procesorů na základě dostupných propagačních materiálů IBM, kde byly IBM procesory srovnány s ostatními výrobci. Na základě těchto materiálů IBM procesory jasně převyšují ostatní výrobce a to nejen v jejich kategorii, ale dokáží porazit i některé procesory s více sockety. Nové trendy v rámci UPS a chladících systému byly zmíněny rovněž. V rámci chlazení byly popsány principy nyní nejslibnějších a nejefektivnějších metod (free cooling a odpařovací chlazení) Poslední část je věnována srovnání virtualizace od společnosti IBM a VMware. Ze srovnání je patrná větší sofistikovanost IBM PowerVM. Tato skutečnost je ale podmíněna spolupráci na hardwarové úrovni technologie IBM POWER a tudíž není vhodná pro všechny platformy.

Jelikož jsou na provozu a neustále dostupnosti datového centra přímo závislé společnosti, nynější datová centra podléhají vysokému zabezpečení. Všechny použité technologie jsou i velice drahé a z těchto důvodů je prakticky nemožné v reálných podmínkách jakkoliv testovat implementaci technologií. V této práci jsou zmíněné uplatnění některých technologií v rámci reálného datového centra. Více informací z praktického prostředí by poskytlo lepší pohled na celý systém, avšak většina z těchto informací byla důvěrná a částí „know-how“ a tudíž nemohla být publikována. Bohužel nebylo možné provést testy v laboratořích, kvůli jejich vysokému vytížení ze strany společnosti.

Konstrukce datových center je velmi rychle se rozvíjející odvětví a velice často se přichází s novými řešeními. Pro porozumění těch nejsofistikovanějších a nejnovějších řešení je nutné stále získávat nové informace ze specializovaných komunit a dokumentací společností zabývajících se touto problematikou.

# REFERENCES

[1] TURNER IV, W. Pitt; SEADER, John H.; BRILL, Kennth g. Tier Classifications Define Site Infrastructure Performance [online]. [s.l.] : The Uptime Institute, [2003] [cit. 2011-05-09]. Available on WWW: <www.iswest.com/colocation/TierClassification.pdf>.

[2] HELLER, Brandon, et al. ElasticTree : Saving Energy in Data Center Networks [online]. [s.l.] : [s.n.], 2010 [cit. 2011-05-09]. Available on WWW: <www.usenix.org/event/nsdi10/tech/full_papers/heller.pdf>.

[3] Cisco [online]. [2010] [cit. 2011-05-13]. Cisco EnergyWise technology. Available on WWW: <http://www.cisco.com/en/US/products/ps10195/products_data_sheets_list.html>.

[4] BROYLES, Martha, et al. IBM Energy Scale for POWER7 Processor-Based Systems [online]. [s.l.] : [s.n.], 2010 [cit. 2011-05-09]. Available on WWW: <http://www-03.ibm.com/>.

[5] Choosing a Cooling System to Support Bussiness-Critical IT Environments [online]. Columbus, Ohio : Liebert Corporation, 2010 [cit. 2011-05-09]. Available at WWW: <emersonnetworkpower.com>.

[6] RASMUSSEN, Neil. Calculating Total Cooling Requirements for Data Center [online]. [s.l.] : American Power Conversion, 2007 [cit. 2011-05-09]. Available at WWW: <http://www.apcmedia.com>.

[7] DUNLAP, Kevin; RASMUSSEN, Neil. The Advantages of Row and Rack-Oriented Cooling Architectures for Data Centers [online]. [s.l.] : [s.n.], 2006 [cit. 2011-06-09]. Available on WWW: <www.datacentres.com/>.

[8] GOODWIN, Andrew; DERBY, James. The Road Map Towards 2020 : Eliminating Battery Back-up Power with the Use of Compressed Air [online]. [s.l.] : [s.n.], 2011 [cit. 2011-06-09]. Available on WWW: <pnu-power.com>.

[9] Today's Facility Manager [online]. 2010 [cit. 2011-05-13]. Backup Power Design. Available on WWW: <http://www.todaysfacilitymanager.com/articles/fm-issue-backup-power-design.php>.

[10] Data Center Knowledge [online]. 2010 [cit. 2011-05-13]. HP Unveils ‘Butterfly’ Data Center Design. Available on WWW:

<http://www.datacenterknowledge.com/archives/2010/07/27/hp-unveils-butterfly-data-center-design/>.

[11] IBM [online]. 2010 [cit. 2011-05-13]. Compare Unix servers. Available on WWW: <http://www-03.ibm.com/systems/power/migratetoibm/compare.html>.

[12] Slideshare [online]. 2010 [cit. 2011-05-13]. Power 7 Introduction - Get the Facts!. Available on WWW: <http://www.slideshare.net/thiededj/power-7-introduction>.

[13] A Comparison of PowerVM and VMware Virtualization Performance [online]. [s.l.] : [s.n.], 2010 [cit. 2011-05-13]. Available on WWW: <http://www-03.ibm.com/systems/power/software/virtualization/resources.html>.

# LIST OF SYMBOLS AND ABBREVIATIONS

| | |
|---|---|
| EPA | Environmental Protection Agency |
| DOE | Department of Energy |
| UPS | Uninterruptible Power Source |
| STS | Static Transfer Switch |
| ATS | Automatic Transfer Switch |
| PDU | Power Distribution Unit |
| SPOF | Single Point Of Failure |
| HVAC | Heating, ventilation and air conditioning |
| BTU | British Thermal Unit |
| CRACK | Computer Room Air Conditioning |
| TPmC | Transaction Processing per minute Complex |
| CPU | Central Processing Unit |
| VM | Virtual Machine |
| AEM | Active Energy Manager |

# LIST OF FIGURES

## LIST OF TABLES