Powering the data centre

A look at alternative approaches to managing energy for cost and/or sustainability reasons in data centres

January 2012

The cost of energy continues to be variable – but trending ever upwards. With data centres being very energy hungry, there is more focus being placed on how to control energy costs here.

This report, consisting of a series of articles written throughout 2011 for on-line data centre magazine SearchVirtualDataCentre.co.uk looks at various approaches that data centre owners can use to control their energy costs and create long term energy sustainability around their IT platforms.

Clive Longbottom
Quocirca Ltd
Tel : +44 118 948 3360
Email: Clive.Longbottom@Quocirca.com

Powering the data centre

A look at alternative approaches to managing energy for cost and/or sustainability reasons in data centres

Overall energy efficiencies within a data centre are poor	Over-provisioning of IT equipment, low utilisation rates and dependencies on national energy grid systems for energy transmission and distribution result in overall energy efficiencies as low as 1.3%. Fiddling at the edges will not be good enough to change this level appreciably – organisations and data centre managers need to look at greater changes in order to completely alter their data centre energy efficiencies.
The best long-term returns will be gained through big investments	Changes in how energy is generated, transmitted and distributed will give the most sustainable strategy. A move to a mix of e.g. photovoltaic, geothermal and hydro electricity mixed with more local generation where the heat (normally wasted) can be used in the organisation or across the local community will drive efficiencies up massively.
Short term tactical approaches can still provide good returns	The majority of data centres have grown organically and are now sub-standardly constructed both at the facility and at the IT level. Implementing a rationalisation, virtualisation and consolidation strategy combined with hot and cold aisles, structured cabling and more effective cooling provision can all provide major uplifts in energy efficiencies without requiring massive investment.
New compute approaches offer escape routes	There is an on-going move toward the use of cloud computing and to using co-location or hosted services. These approaches will lead toward far greater sharing of available resources, which will drive up utilisation rates and so make shared data centres more effective in their energy usage. The facilities being built by cloud and co-location providers will also tend to be more cooling and energy efficient in their design and use.
Cooling is not all about CRAC units	Although energy savings can be gained through reviewing what type of and how CRAC units are used, other cooling systems – such as free air and water cooling – may offer better efficiencies and result in recoverable heat that can be used elsewhere in the organisation or in the community. Each option has its own pros and cons – but should be considered when looking to build, upgrade or retrofit a data centre
Renewables may not be the long term answer – but may be part of it	Wind, photovoltaics and many other renewable energy sources are intermittent in their capabilities to provide energy and cannot therefore be relied upon as a primary energy source for a data centre. However, judicious use of renewables can offset data centre energy costs and can be a solid part of an organisations overall energy strategy.

Conclusions
The growth in the importance of data centres around the world in supporting the global economy means that there is increasing focus on how well data centres are performing in themselves. The poor levels of overall energy utilisation are being seen by organisations and governments as a cause for concern, and data centre managers must have tactical and strategic plans in place in order to show how energy costs are to be controlled across their IT platform, and how any activity will support an organisation's sustainability activities.

A black day for data centre carbon emissions

If we don't get a grip on carbon emissions, the world will end, right? I doubt that your organisation sees things in this light, yet it is likely that its carbon emissions will come under close scrutiny over the coming years. Business electricity prices have been notoriously variable over the past couple of decades, but the trend underlying the variability has been unavoidably upward.

At just a basic level, as the dependency on IT has grown, the percentage of an organisation's energy bill attributed to the data centre has grown and has created a focal point for the business. When finances are tight, the immediate reaction is to look for soft targets -- and the data centre power bill is one of them.

UK carbon emission reduction

The situation is worsening as governments around the world increase their commitments to reduce carbon emissions. The UK is a good example: The government have taken on heavy carbon emission reduction targets, and have put laws in place to put pressure on organisations to achieve these. The Carbon Reduction Commitment (CRC) Energy Efficiency Scheme was put forwards as a carrot-and-stick approach for heavier energy users. Those making the biggest savings would make money out of the scheme, while those performing badly would pay penalties.

This has now been changed, and all organisations caught in the CRC net will end up paying for their carbon outputs in the form of a "carbon tax". Because the government needs more tax revenues, you can take it as a given that even more organisations will be taxed in the future.

So it doesn't really matter whether your organisation wants to control carbon emissions for global sustainability reasons or not -- the cost of not working on lowering output may be enough to break the camel's back and be beyond capacity.

Power distribution and high levels of carbon emissions

Let's look at a couple of the major reasons for concern about energy usage in the data centre: power distribution and overall efficiency.

In a fossil fuel power station, typically there will be a 10% loss in energy capability just in burning the fuel. In effect, only 90% of the available energy actually creates electricity through becoming steam to drive a turbine. The overall energy efficiency in creating electricity in this manner will be around 45%.The high-voltage transmission losses from the power station to the local substation will also be around 10%, so we're already down to around 36.5% overall efficiency.

From the substation to the data centre, there should be only one transformer in place, with further losses of around 2%. In the data centre itself, each transformer will introduce around 2% losses -- and there may be five or more transformers in play before the electricity gets to a point where the IT equipment can make good use of it. That makes for further losses of nearly 11%, and lowers our overall energy efficiency from 45% to around 32%. All of the losses will be in the form of heat -- which means that even more energy is required for cooling.

How carbon emissions are linked to PUE

But this low-level energy efficiency is attributed to the data centre equipment itself, before a single CPU cycle has been turned or a byte of data processed. How much of the electricity that actually reaches the IT equipment is carrying out useful work? Power Usage Effectiveness (PUE) is increasingly used to measure the overall energy efficiency of a data centre through dividing total data centre power usage by the power used for running IT workloads. Uptime Institute research has shown that the typical data centre has an average PUE of 2.5. This means that for every 2.5 watts *in* at the utility meter, only one watt is delivered *out* to the IT load. Uptime Institute estimates most facilities could achieve 1.6 PUE using the most efficient equipment and best practices.

Only 40% of electricity coming into a data centre is used to run workloads. The rest is “wasted” through cooling, UPS and other secondary systems. A further 60% of the energy is not used for “real” IT work. This means that only 13% of the oil/coal/gas we started with is actually available to drive useful data centre work. Even when the electricity hits the server, there will be more transformers in place, losing even more energy in the process.

But the worst is yet to come. Let’s assume that 80% of the electricity is used in the server environment (the other 20% being in storage and networking equipment). Again, research shows that the “average” utilisation of servers in a standard distributed computing environment is between 5% and 10%. Therefore, 90% of electricity used in running servers is actually wasted because there is no workload being applied to it. Now, our already poor overall energy efficiency has suddenly shrunk to around 1.3% -- if we consider the storage and network environments to be almost 100% efficient.

Process	Efficiency	Overall efficiency
Coal burning	90%	90%
Electricity production	45%	40.5%
High voltage transmission and initial transformation	90%	36.5%
Lower voltage distribution to point of commercial use	98%	35.7%
Data centre distribution	89%	31.8%
PUE impact	40%	12.7%
Workload efficiency	10%	1.3%

Overall efficiency of 1.3% is not sustainable, so organisations are right to look at energy utilisation in the data centre to lower their carbon emissions -- and their energy costs. For every 10 grams of carbon emissions created, directly at the data centre, nearly 1 kg of carbon emissions will be created at the power station.

A lot can be done within the data centre itself to address the generation/transmission losses involved, but it’s not the only place in an organisation where energy consumption can be reduced. IT can help with these reductions, for example, by making energy-intensive business processes more efficient or reducing the need for employees to travel in and out of the office to be able to access the company’s services. Surprisingly, there is more that can be done outside the data centre by looking at newer means of power generation that can impact the 65% energy losses involved before electricity even gets to your organisation.

In the coming months, research and analysis company Quocirca will be studying how organisations can look at various aspects of the data centre to create a more energy-effective environment. This will range from short-term fixes to long-term strategic directions. In addition, it will look at how IT can be positioned to create an energy-efficient organisation, not just an energy-efficient data centre.

How to reduce your data centre CO2 emissions

As data centre managers consider ways to reduce the CO_2 emissions of a data centre, they often begin by targeting data centre IT equipment and how effectively it is used. By introducing new equipment and virtualisation to boost utilisation rates, IT managers can achieve rapid results in reducing total data centre energy use. But, in key ways, this approach misses the point of carbon emission reduction.

For long-term savings rather than short-term gains, any strategy to reduce carbon emissions has to be well-planned and well-implemented. Such foresight and solid execution, however, involve new approaches not only to the IT equipment in use but also to ancillary equipment. Consider, too, the lifecycle management of equipment to reduce the "embedded" carbon footprint—the total energy use associated with a product's manufacture, transport and use.

It takes energy to manufacture and transport equipment, and all items have their own carbon footprint before they even arrive at the data centre. For you, the data centre manager, this means considering factors other than equipment as a means of energy consumption. Take a simple server, for example. Each part of a server is manufactured, shipped, assembled, shipped again, further assembled and then shipped through the channel to get to you. The amount of embedded carbon associated with each of these stages can easily be greater than the energy consumed by the final product in a year.

CO2 emissions and lifecycle management strategies

Data centre managers can't control the entire supply chain, so they can't address embedded carbon directly. They can, however, ensure that their own practices don't result in a higher total carbon footprint because of poor management of the "carbon chain."

So, a full lifecycle management policy is required. Virtualisation allows equipment to be far more easily tiered in its usage, so a server that's used for the most demanding workloads today can still be used, for example, for file and print services in three years' time. You can move 1 gigabit (GB) switches to the periphery as 10 GB and faster switches become available. Use high-performance storage for less critical storage—or even for near-line back up storage over time.

It may be more difficult to manage other equipment in the data centre in this way. For example, chiller units tend to be fit for purpose—or just fit for disposal. Once power distribution systems are in place, it is difficult to change them without impacting the rest of the data centre. So, ensure that the greatest flexibility is allowed for when procuring these items in the first place to ensure that they will have as long a life as possible.

Take the same approach with the data centre facility itself. Yes, cloud is all the rage at the moment, but what about two years from now? Will the future be around the scaling out of an infrastructure—that is, just adding more and more servers as required; will it be further convergence; or will private data centres pretty much die out? Unless these questions are addressed and flexibility is built into the data centre facility from the start, you may find yourself with costly problems going forward as you face the need to build different facilities to keep up with broader changes in technologies. A well-designed data centre will allow for power distribution to keep pace, for cooling to be adequate and targeted and for new architectures to be built alongside the old.

Configuring a private data centre to grow and shrink as needed to free up space for other requirements can improve the overall efficiency—and therefore the carbon footprint – of the data centre. A fully engineered data centre, built for efficiency and combined with higher running temperatures and fully targeted cooling can lead to massive carbon footprint reductions as long as you also balance changes with the probability of a future, shared data centre approach.

Data centre power is the next area to look at for reducing carbon emissions. Facilities owners decide which power supply will be used to generate energy in the data centre. Some may choose to contract directly with a provider of hydro, nuclear, wind or solar power, or through the use of combined heat and power generation local to the facility. Others might sign a contract with a sustainable or low-carbon energy company that ensures that investments are made in long-term sustainable energy generation.

New devices, new strategies

The growth of mobile users and the device types they use, along with the ubiquity of compute devices across the whole organisation, means you many want to consider new approaches, such as server-based computing to provide the main capability of serving the users' desktops. However, such approaches may have an overall negative impact on an organisation's carbon emissions, as more power is required in the data centre to support the centralised desktops—at least until existing “fat” devices (such as existing PCs) have been replaced with more energy-efficient, low-powered “thin” ones (thin client devices, tablets, smartphones, etc.).

With all IT apparatus, disposal of the equipment and its impact on carbon emissions needs to be covered as well. In the UK, the WEEE (waste electrical and electronic equipment) legislation means that equipment should be recycled in a controlled manner. In many other countries, however, this may not be the case. Alternative, planned disposal has to be part of the overall equipment lifecycle planning. Wherever possible, provide channels for reusing complete items of equipment to reduce the impact of the embedded carbon. Where reuse is not possible, recycling components (e.g., hard disk drives and memory modules) may be the best bet. But then the embedded carbon of dumping the old fat clients offsets the benefits of replacing the energy inefficiencies of fat clients, and this is where the complexities come about.

Finally, IT should not be seen in isolation when it comes to a company’s energy use. Any pressure from the business to lower the energy usage of IT should be looked at carefully. Creating a strategy for low CO2 emissions should not be an IT-only issue. In many cases, the IT department can help create a far more energy-sustainable organisation, one where small increases in the amount of energy used by IT leads to massive savings elsewhere in the business. By using electronic delivery in the supply chain, for example, a company can replace employee travel with telepresence.

Issues with power generation and power management

The biggest issue with energy usage in the data centre is not how it is used within the data centre itself, but the way that it’s generated and distributed and the losses incurred before it even reaches the point of use.

The many different voltages used in the transmission and distribution of energy means that it is transformed several times, and each transformation produces efficiency losses. Choosing the right form of power generation and where it is situated can provide greater energy sustainability and lower the carbon impact of a data centre.

If a data centre is powered directly from a national power grid infrastructure, very little can be done about these losses, but at least a choice of generator with less carbon impact (e.g. hydro, nuclear, geothermal) can be chosen in many cases. Even though the electrons used by the equipment may have come from fossil fuel generation, such contracts encourage further investment in non-fossil fuel systems and continue the move towards more sustainable systems. Transmission and distribution losses will be the same, but the elimination of fossil fuels from the equation at least lessens the carbon impact.

High/low voltage and data centre power generation

Transmission losses are based on the resistance of the cables carrying the power, so the shorter the transmission link, the better. High-voltage AC transmission is more efficient than lower-power AC or DC distribution as the power losses are more dependent on current than on voltage. Therefore, ensuring that the data centre has its own substation with optimised low-voltage distribution will minimise losses as the data centre owner has control over the supply.

More data centre owners, however, are looking at changing their approach to the initial power generation needed by moving the data centre closer to the generator – or integrating generation as part of the data centre itself. For example, Yahoo opened a data centre in Quincy, Wash., US, that uses locally produced hydroelectric power that has fewer losses in the short distance between the point of generation and usage. In addition, it uses natural air cooling, which eliminates the need for electricity for cooling – an area which is often more energy-intensive than the IT equipment itself.

Hydro power, geothermal power, solar and wind power management
Quincy has become a focal point for high-tech data centres, with Microsoft also having an air-cooled, hydro-powered data centre there, and Dell and Sabey both building large facilities that can make use of the large amounts of hydropower in the area.

In a similar move, Iceland has been marketing itself as a place to build data centres with access to unlimited cheap energy – but there it's based on geothermal power. A prime example there is the Thor data centre, again being more energy efficient because of lower transmission and distribution losses, and also having zero incremental carbon impact from the generation itself. Additionally, the year-long low external temperatures also make air-cooling of a data centre feasible.

Fujitsu, Verizon and First Bank of Omaha are all running data centres that use hydrogen-powered fuel cells as part of their main energy feed. Touted as a low-impact energy source as the output from a hydrogen fuel cell is just electricity, heat and water, the amount of energy required to create the hydrogen in the first place can make such an approach ineffective overall if the original energy for cracking hydrogen from water or petrochemicals came from fossil fuels originally.

Although not yet used for powering data centres on their own, solar and wind power can at least be used for powering some of the more peripheral aspects of the data centre. For example, Co-operative Insurance (CIS) in the UK has covered one side of its headquarters in Manchester with photovoltaics, with the capacity to generate 180,000 units of energy per year; it's nowhere near enough for a data centre itself, but it's plenty for lighting and ancillary needs.

Other World Computing (OWC) uses wind power to power its data centre in Woodstock, Ill., US, but it has to depend on the grid for power when there is insufficient wind. However, the reduction of both CIS and OWC's carbon footprint is significant enough to make the approaches worthwhile.

Within the data centre itself, the question of only using AC for distribution to equipment or using DC throughout the data centre is still an item for discussion. A couple of years ago, ther seemed to be a far better argument for using DC, but the need for equipment in the data centre to be specifically built for DC use and for data centre wiring and distribution architectures to be specific to DC seems to have pushed data centre owners to stick with AC.
Furthermore, for DC to provide the promised energy efficiencies, it is preferable for all equipment to use the same voltage – and as this is highly unlikely. There will still be the need for multiple transformations to the main 12V, 5V and 3.3V voltages used at the equipment level.

The benefits of structured cabling
Structured cabling can ensure that losses are minimised in many cases and that the right cabling is used for specific jobs. Plus, low-power systems can be powered over Ethernet connections rather than having dedicated power links. In addition, ensuring that cables are not looped or otherwise positioned to create power drains through interaction with each other can save small amounts per feed, which in a modern data centre can soon mount up to considerable overall energy savings.

Making sure that a data centre is both energy efficient and energy effective requires careful thought and implementation. The means of power generation, the most effective means of transmission, preferably over the shortest distances possible with the most modern transformers, and fully structured power management cabling will enable energy to be saved and sourced with a lower carbon impact than standard fossil sources. If done with the right forethought, it will also provide flexibility for the future and enable new equipment and architectures to be embraced as needed.

Optimising the existing data centre for immediate gains

The starting point for the majority of organisations when looking at getting to grips with increasing energy costs is to look at the existing data centre and identify what quick wins can be made. Often, this is just based on making cuts to existing systems, which can result in adverse impacts on the business, as the IT platform become increasingly sparse and incapable of having the flexibility and headroom to deal with changing workloads from the business.

However, through a more careful approach, costs can be controlled through small changes, giving either immediate or rapid returns on the investment required to implement the change.

For the most rapid and continuing savings for the lowest outlay, Quocirca recommends that the following steps are taken:

- Start with a full equipment audit. Identify all existing equipment and wherever possible look at its energy efficiency. Ideally, this should be based on monitoring power usage at the equipment level, but if that is not possible, then take the "plate" values (the Amperage on the label attached to the equipment). Identify which equipment is the least energy efficient and look at whether replacing it with new, more energy efficient kit would provide a rapid payback.
- Do a full software audit and see where there are multiple instances of an application running. Can these be consolidated down into one single instance? If so, running this single instance on fewer servers will drive up utilisation rates and will also reduce energy usage.
- Look to bring in virtualisation – even if it is piecemeal. Even if there is some lingering doubt over whether virtualisation is a tool that should be used across the data centre, areas such as file and print and other base functions can be run with little risk on a virtual pool of servers. Again, this should drive up hardware utilisation rates – and will therefore reduce energy utilisation.
- Consolidate into less space. A large data centre requires more cooling if standard space cooling is used. Using stud walling to create a room more fitted to the amount of equipment in the space will require less cooling, and therefore energy to be in place.
- Check existing cooling architectures. Here. Quocirca is not looking at replacing any hardware, just checking that things are running as they should be. Look at the sub-floor space if raised floors are being used, and tidy up any cabling that is blocking flows. Check all floor tiles to make sure that any broken ones are replaced so that cool air is not wasted, and that pierced tiles are not blocked by dust or other things – including the feet of racks or the bases of equipment itself. Move tiles around so that the cool air is directed where it will have the greatest effect. Even if ducted cooling is already in use, measure spot temperatures and make sure that the cooling air is being used to its best advantage.
- Run the data centre hotter. The majority of data centres are run at temperatures suited more for humans than for technical equipment. As part of the equipment audit carried out above, look at the operating limits for each item. Running a data centre at 28 degrees Celsius is more than feasible these days – and can be higher if desired. Each degree hotter a datacentre can be run at means less energy required for cooling – so more savings.
- Use containment. General space cooling in a data centre is wasteful, as the vast volume of air not actually around the equipment is also being cooled down. Creating contained volumes – for example by fitting the aisles with ceilings and doors at each end - means that each aisle can be cooled as a self-contained entity, and the external volumes can be left to their own devices. Commonly known as hot and cold aisles, this does not necessarily have to be at a highly engineered level – simple doors and roofing can create a relatively contained environment that cuts back on the need for cooling.

All the above is aimed at providing the biggest bang per buck in gaining the fastest savings. Once rationalisation, consolidation and virtualisation has been implemented, the main aspects tend to revolve around saving energy through optimising cooling. Once all of the above quick wins have been done, it may be worth looking at more advanced approaches, such as:

- Computational fluid dynamics – model the data centre and see how the cooling really works. Identify were dangerous hot spots may be building up, and ensure that cooling flows really are optimised. In-row cooling can provide far more targeted cooling and can also provide greater availability as the systems are more distributed.
- Water cooling – fitting rear door cooling systems, or on-chip water cooling means that the effectiveness of cooling can be maximised, and the heat removed can be used for other purposes, such as space heating or for hot water provision.
- Variable speed CRAC systems – conventional chillers are either on or off, whereas variable speed systems can be run to provide a variable amount of cooling, so optimising the energy that is being used.
- Free air cooling – the great end result, where chillers are no longer needed. If the data centre can be run hotter, and highly targeted air flows can be shown to keep equipment within operational thermal limits, the use of filtered air can be considered. The savings against using CRAC systems can therefore be big, as only fans and filters are required.

Energy usage and the variability of its cost are the biggest threats to the data centre at the moment. The business is increasingly aware that IT is a large cost centre, and will increasingly pressurise the IT department for savings. This may be couched as "do more with less", but the evolution of the data centre has tended to leave plenty of room for quick wins that not only save money, but provide a more flexible and sustainable IT platform going forward. The above recommendations should be able to show the business that IT can respond to the financial climate without the need for cuts to the service provided.

Outsourcing IT and cloud models serve as carbon emission escape routes

Data centres are increasingly seen as sub-optimised environments with their low systems utilisation marks, high functional redundancy and over engineering that has led to massive energy needs that are unpredictable on a short-term basis. All that's known for sure is that energy bills will keep going up.

Organisations are figuring out that data centres are now a major issue for them with regards to their overall carbon emissions, and governments are seeing data centres as soft targets when it comes to "saving the planet." In last year's series of articles, Quocirca looked at the impact of carbon emissions legislation, but there are ways to avoid at least the worst excesses of such laws.

Energy savings from outsourcing

Depending on the type of organisation, a data centre can account for up to 40% of its energy use, so anything that minimises that number has to be a good thing. And this is where the cloud and outsourcing IT comes in. To calculate its energy usage, a company must report only on what is directly under its own control. When IT functions are powered and managed by an external facility and are outside the organisation's control, then there's no need to report the energy used in providing the function. True, the facility owner will have to report and pay for the emissions, but that's a different story.

Therefore, if a data centre is rated as a 1 megawatt facility and even 10% of that can be outsourced to an external facility, then an immediate 100 KW of corporate energy usage has been transferred to someone else's books.

This makes for an effective way of managing an organisation's carbon emissions. In essence, very little change is made in how the organisation carries out its business, but its technical functions are run in a third-party facility. The third party must report on the energy usage.

But what if the third-party facility passes the costs on to its customer? Then, the payments to cover the carbon emissions would just become a standard operational expenditure for the outsourcing facility.

What is far more likely, however, is that a well-chosen cloud service provider will have a more energy-efficient environment, will be able to provide economies of scale, have better contracts in place with energy providers and will have a far better environment than the customers could have on their own. Therefore, the customer is able to offload a degree of its emissions to a third party, the third party is able to minimise its emissions in a far more effective manner, and a more sustainable overall model is created.

The capability to remove a proportion of an organisation's measured emissions may well be short-lived, because it's highly likely that governments will rapidly see that plugging such a loophole is financially favourable for them; keeping all emissions within an overall "dirty" organisation means more taxes than seeing a proportion of emissions rolled into a far "cleaner" service provider. Perhaps additional rules will be put in place forcing organisations to obtain details from third parties as to what emissions should be assigned to their own usage. As cloud computing increases its share of the market, this could make for some interesting calculations.

Overall, outsourcing IT can offer a more sustainable platform model and provide business, economic and sustainable benefits for an organisation. While it's possible, being able to make the most of outsourcing IT and minimising any liability for carbon emissions makes good sense. Just don't leave the opportunity too late – governments may make it harder to gain any benefits once they see how such loopholes can be used.

Energy saving tips: Reduce energy consumption with hot and cold aisles

With cooling being a major energy drain in any data centre, it is that an IT manager ensures that the overall approach to cooling is optimised in order to reduce energy consumption. Many existing data centres use computer room air conditioning (CRAC) units to chill and condition air, ensuring that it is at the right moisture level, and then channel this to the bottom of equipment racks in the data centre via the plenum (the gap between the solid floor and the raised floor). The cold air makes its way through the rack and is eventually expelled through the top of the rack (and in many cases along the sides of the rack) into the general volume of the data centre from where it is extracted via fans.

This approach can be wasteful for two reasons. First is the need to count on the temperature of the general data centre remaining within whatever limits are decreed. If a data centre covers a general area of 10,000 square feet and has a raised floor to dropped ceiling height of 10 feet, then it will be required to maintain 100,000 cubic feet of air within a specific temperature and moisture level. Second, after a while, raised floors become somewhat patchy. As equipment is moved, perforated tiles may not be moved as well, non-perforated tiles become dislodged and under-floor cabling can block the effective flow of the cold air.

Energy saving tip: Use containment to minimise the volume of cold air being used. That can reduce the wasting of cold air and therefore can reduce the data centre's energy consumption.

A starting point is to create a "cold aisle", which can be done in many ways -- from relatively simple to highly engineered. Taking the relatively simple as an example, the problem is how to create a contained environment between two equipment rows. You can do that by covering over the rows with some form of an enclosure and placing simple overlapping plastic hanging doors at each end. Blanking plates in the front of the racks can then be replaced with open, semi-open or perforated plates to allow air to enter, with the same at the rear of the racks to allow air to escape. Cold air is pumped into the covered aisle (generally through the cover so that the cold air can fall naturally and use positive pressure to push through the equipment racks), and this makes its way through the equipment racks and is expelled into the open area of the data centre. Figure 1 shows where each two rows create a single cold aisle, and each group of two equipment rows plus the cold aisle is a self-contained environment within the data centre itself.

Figure 1

This is a good start to reducing energy consumption; cold air volumes are being applied across a greater area of equipment and should be more effective than a bottom of rack to top of rack approach.

Energy saving tip: When the cold air route is via the aisle, the chances of any significant blockage, as could be found in a raised floor with cabling and other "out of sight" items, is minimised. This should result in a decreased need for cold air to be produced. However, the exit air is still heating up the general data centre, and this can be dealt with through introducing a "hot aisle" concept – the converse of the cold aisle.

Figure 2 shows the spaces between the cold aisle units are now covered over with doors at the end of the aisles. Cold air is pumped into the cold aisle spaces as before, but now the hot air is contained within smaller volumes in the hot aisles, rather than being vented into the open data centre. What is gained here is that the total volumes involved in the data centre for cold and hot air are both reduced and the rest of the data centre volume can be run at whatever temperature you want. Indeed, no cooling may be required at all for the volumes outside of the cold and hot aisles. From the 100,000 cubic feet of cooling required in an uncontained data centre, it is more likely the cooling will now be down to around 20,000 cubic feet – a saving of up to 80% in cooling required.

At the highly engineered end, vendors provide completely self-contained systems that include the racks, specialised cooling systems, highly targeted cooling air ducting, pressure monitoring to ensure continuous cooling and so on. For complex environments, this will be required to ensure high availability of systems.

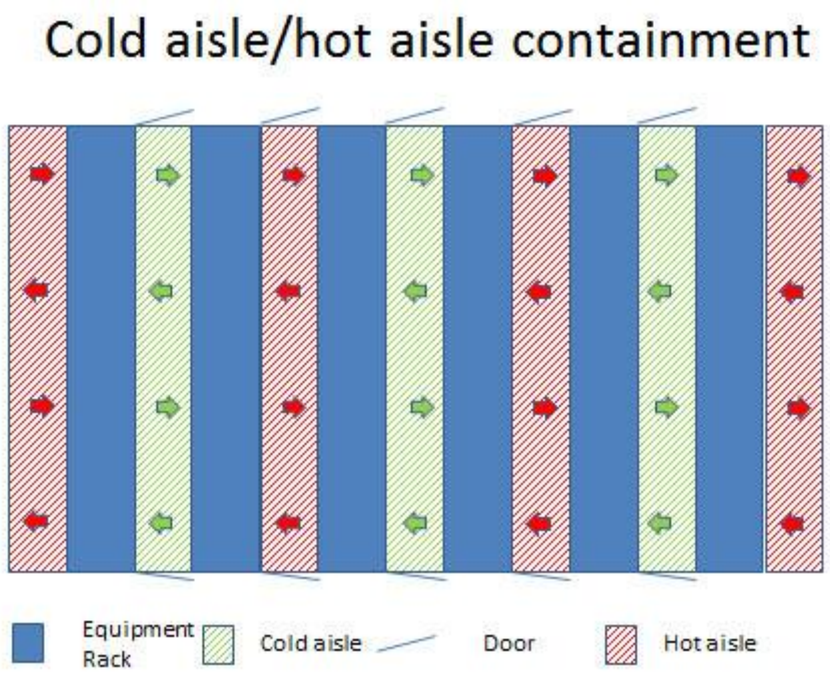

Figure 2

Energy saving tip: Bearing in mind that modern equipment runs far warmer than equipment in the past, it is possible to run inlet cooling air at a higher temperature as well – so reducing the need for energy even further. Indeed, by taking the idea of hot and cold aisles one step further, and making the equipment rack its own self-contained system using in-row cooling, air flow routing and direct hot air ducting outside of the data centre, energy use can be completely minimised to a level where more than 90% savings can be made.

You'll need to address two issues when using containment approaches such as those above. The first is you must fully understand the cooling flows required, and here it is recommended that you use computational fluid dynamics to ensure that flows reach the areas that need cooling and that monitoring of temperatures is used to ensure that hot spots do not build up.

The other issue is that should the cooling fail, there is less cold air sink to keep the equipment cool whilst the problem is being fixed than there would be if the equipment racks were open within a large uncontained data centre. Therefore, be sure that cooling is fully redundant where downtime has to be avoided, or that equipment can be elegantly switched off within a suitable time frame to avoid overheating should cooling become unavailable.
Containment can be a relatively easy means to reduce energy consumption in a data centre, and it's something that can be retro fitted relatively easily.

How to optimise CRAC units for best energy efficiency

Within a data centre, it is often assumed that the IT estate (i.e. servers, storage and network equipment) is the energy hog. Therefore, many organisations think they can optimise their energy costs by consolidating down to a more optimised server, storage and network estate. So, they embark on just such an energy efficiency action plan.

The Green Grid's Power Usage Effectiveness (PUE) scoring compares the total amount of energy used by a data centre facility against the amount of energy that is used in server, storage and networking. It finds that the majority of data centres are running at a PUE of greater than 2. That number means that for every unit of energy used within the server, storage and network equipment, at least another unit is used in "peripheral" equipment, which includes lighting, cooling and backup power supplies.

Organisations look at rationalising the IT estate through extensive use of virtualisation and then find that their PUE has got worse and wonder why.

Assuming that the original data centre had a PUE of 2.0, with half the energy being used by the IT estate and half by peripheral equipment, and with that amount of IT estate being halved through the use of virtualisation, the new case is that an overall energy savings of 25% has been made (half of the IT estate's energy usage).

If nothing has been done about the peripheral equipment, then PUE has suddenly risen to 3.0.

Simply shutting down CRAC units is not a solution
It is tempting for an organisation to just shut down some of the computer room air conditioning (CRAC) units, but this can bring in new problems. For example, most organisations install CRAC units to ensure that there is a degree of resilience in place should any one unit fail and to allow for any unit to be taken out of use to carry out planned maintenance. Therefore, just removing CRAC units from the facility based on a theoretical need for less cooling can mean reducing data centre resilience as well.

Cooling is an energy hog and there are many approaches that can help minimise the amount of cooling required within a data centre. For instance, you can run the data centre at a higher temperature and use highly targeted means of cooling such as hot and cold aisles.

Energy efficiency action plan
Further improvements can be made through changing the way the chiller units themselves work. Most existing CRAC units are fixed-speed. Once on, they tend to run at 100% on a continual basis, only turning off when control systems can guarantee that there is enough cold air available to maintain the temperature in the data centre for a period of time. In most cases, this means that the CRAC units run pretty much continuously – wasting energy. It also means that more frequent maintenance is necessary on the units than should really be required.

Pros and cons of variable speed CRAC systems
Newer CRAC units use variable speed capabilities – which also means variable power. The unit uses a feedback loop to ensure that the amount of chilled air the CRAC unit provides is just sufficient to meet the data centre's needs. If the data centre is already within its thermal envelope, the CRAC units can lower their output and therefore become more energy efficient.

As long as the feedback loop is effective, the units will run far closer to a theoretical optimised limit and energy savings will be highly worthwhile.

But, there are a couple of downsides to moving to variable speed CRAC systems. The first is the obvious one: It means replacing older units, so there is a capital cost associated with it. The second is that even modern variable speed CRAC units have energy overheads associated with them when running at reduced loads. The most energy-efficient CRAC unit is one that is not running at all, whereas the most cooling-efficient one is running at 100%.

These two contradictory aspects can be harnessed if there is a means of storing the output from the CRAC units. For example, the cooling output from the units can be passed through to an intermediary substance, such as water or another liquid, rather than being transferred directly to air.
Indeed, newer systems from the likes of IBM are looking at using a phase change substance, where the cold energy is used to “freeze” a liquid. Then, on thawing, it releases the cold back to the air as required. Using an intermediary store means fixed-speed CRAC units can be run at 100% when they are needed and then turned off completely when a sufficient store of cold has been created. As that store is depleted, the units can be turned back on again – at 100%.

You’ll need to carefully calculate the volume of store to ensure that the overall efficiencies work out. When combined with higher data centre operating temperatures and highly constrained, targeted cooling architectures, this can be more cost effective for many existing data centres than a complete move to variable-speed CRAC units.
An intermediary store also means you can more easily manage the failure of a CRAC unit because there will still be coolant available for a known period of time from the store itself. In addition, in climates where ambient temperatures for even a part of the year are adequate for direct data centre cooling, you can turn off the CRAC systems completely, with external air just needing to be filtered and its humidity levels modified before being used in the data centre.

Don’t become a CRAC addict
Although CRAC systems are an important part of a facility, there is no need to become a CRAC addict.
A sound data centre energy efficiency action plan involves a little forethought and making sure the facility is matched to the needs of the IT estate. This can go a long way to ensure that a fully-optimised environment provides suitable levels of overall availability while minimising energy u se and on-going maintenance costs.

Water cooling vs. air cooling: The rise of water use in data centres

In the deep, dark past of computing, water was the default means of keeping a computer cool. Now, in the 21st century, it is making its presence felt again. If you compare the use of water to air as the cooling technique within data centres, a lot of IT professionals are exploring how to achieve energy efficiency through air cooling. But don’t write off using water for direct cooling in the data centre.

As electronics improved and as computers became smaller, air cooling became the norm. Chilling down vast volumes of air that could be blown across the electronics works where electronic densities are still relatively low and where energy prices are reasonable. Much engineering work has improved heat exchanger design at the CPU level, but design issues still occur as equipment densities are pushed.

Increases in density along with sharp increases in energy prices have forced many IT pros to look at how inefficient existing cooling practices are. The majority of data centre owners are investigating ways to save energy while maintaining adequate cooling through the use of more targeted air circulation using computational fluid dynamics (CFD) and thermal imaging. But, direct cooling through water use in the data centre is raising its head again.

Air cooling: Popular but flawed

Air cooling's faults are becoming more apparent. It's just not very good as a heat transfer medium. As data centre equipment has increased in density, the use of large fans has decreased because of a lack of space; so has the capability to move the large volumes of air required through the equipment. Being a gas, air has poor heat conductivity, so extra means are required to transfer the heat from its source into the air itself. Therefore, fins must be attached to hot spots to maximise the surface area available for heat transfer away from the equipment.

The amount of energy required to chill down air to required temperatures and to move it around and through the data centre is becoming somewhat expensive. Although water may be the most cost-effective way forward, today it's still predominantly a secondary source.

Another means of efficiently using water to chill air is water-side economisers, which are finding increasing use to either replace or supplement standard computer room air conditioning (CRAC)

Other data centres are taking things a little further. For instance, Google is using sea-water cooling for its new data centre in Finland, while PlusServer, a German organisation, is building a new data centre in Strasbourg that will use ground water at a fixed 12 degrees to 14 degrees Celsius as feed water for cooling air in the data centre.

Other similar approaches include falling curtain evaporative cooling (a method where air is forced directly through a falling curtain of water and cooled due to evaporative energy exchange) in hot climates as well as direct river-water cooling in cold climates.

More water-cooling vs. air-cooling facts

Water has between 50 and 1,000 times the capacity to remove heat than air and can therefore be far more effective to cool hotspots if it's engineered and implemented in the correct manner.

With mainframe and certain midsized computers that used water cooling in the 1970s, '80s and'90s, this water was run through copper pipes at positive pressure and used to cool hotspots as needed, specifically at the CPU. Most other electronics within the computer ran at a low temperature and in an open space to be cooled through low-pressure air flows fed by chiller systems. But, if a water leak occurred, the positive pressure would force water out into the heart of the computer, and, unfortunately, water and the insides of a computer do not mix that well.

However, water cooling has matured to a point where leakage should not be the problem that it once was. For example, data centres can use rear door heat exchanges as self-contained systems, such that if there is a leak, it is all contained within the system with no risk of the water getting to any electrical equipment. Here, the rear of a 19-inch rack is replaced with a heat exchanger through which chilled water is pumped. Even as a passive system (i.e. no forced air used), IBM and others claim that such a system can remove 60% of the heat from a 33 kW high-density rack. Used within a self-contained sealed rack, rear door heat exchangers can provide considerable savings against having to put in place new CRAC units along with targeted cooling.

Targeted water cooling

Another approach is to use highly targeted water cooling, in which metal pads (generally copper or even gold-plated copper for additional thermal efficiencies) with micro-channels are used to replace the standard cooling fins used on CPUs. Pure water (no dissolved solids or gases) is passed through the micro-channels, removing heat directly from the CPUs; often it can be used in other parts of the building as output at temperatures that are high enough to heat water.

Although CPU-based water cooling is not particularly new in itself, what is new is that systems are now run at negative pressure, so that the water is sucked round the system, rather than pumped. If the system develops a leak, air is sucked in, rather than water leaking out. Sensors monitor the system continuously so that if this does happen, administrators are informed immediately and can take remedial action.

For those wanting to move to maximum equipment densities, cooling is one of the biggest issues they must deal with. Using forced air only may be short-sighted as energy prices continue to fluctuate but trend upwards. The use of water to more effectively cool hotspots, combined with higher overall data centre temperatures provides the means to optimise energy usage and provide the capability to create and operate a high-availability, high-density data centre for the future.

If you are wondering whether to use air cooling or water cooling in the data centre, you will find that using water is very cost effective in most scenarios.

Free-air cooling – using natural cooling in the data centre

In many circumstances, the most cost-effective means of maintaining an optimum temperature within the data centre is by using natural cooling, or free-air cooling.

In the past, when the accepted temperature for a data centre was in the range 19 degrees Celsius to 21 degrees Celsius, natural cooling only applied to certain geographical environments and only at certain times of the year. Consequently, natural cooling wasn't used much. But, with the American Society of Heating, Refrigeration and Air-Conditioning Engineers (ASHRAE) now accepting temperatures of up to 27 degrees C (and with many IT equipment vendors accepting higher temperatures of up to 34 degrees C) as being within working limits, free-air cooling is worth considering again.

Natural cooling, however, requires more than just opening a few extra windows in the data centre. The air's humidity is also an important factor and sometimes far more important than the temperature. A high moisture level in the air can lead to rapid deterioration of some metals, whilst air that is too dry can lead to issues with static electricity, as well as the growth of dendrites, little whiskers of metal on printed circuit boards (PCBs) that can cause paths to short out in high-density systems.

So, even if air is at a temperature suitable for cooling, it still has to be treated to ensure the right humidity as well as filtered to catch particulates that could also cause problems in the data centre.

But you can still save on energy requirements for chiller units. You can also save on energy by operating large, slowly running fans rather than the high-pressure, fast spinning fans that are needed to push air through the smaller-bore ducting used in standard computer room air conditioning (CRAC) systems.

The most basic free-air system draws in air through filters to remove particulates. This system measures the moisture content and remedies it through drying or by adding small amounts of additional moisture as required. The main use of free-air cooling tends to be around the use of "economisers", using liquid or air-based systems to provide a relatively closed system.

A water-side economiser uses an evaporative tower or a dry cooler to deliver the cooling needed for a heat rejection loop. An air-side economiser uses ducting, filters and dampers alongside sensors and controls to allow filtered, treated cold air into the data centre as required.

Keeping a watch on outside temp

Such systems can lower energy requirements by 30% to 60% over standard CRAC units, but they can be highly dependent on external temperatures for their overall efficiency. When external temperatures are too warm, other cooling will still be required, which makes the financial returns on free-air cooling a little more complex.

Other approaches to natural cooling are heat wheels that exchange the heat from the data centre with cooler outside air via large metal surface areas; and the superior thermal mass of metal to improve overall heat exchange. But even these systems bring a considerable amount of external air into the data centre, and the air still requires filtering and moisture controlling.

However, the Kyoto wheel or "KyotoCell" (an energy recovery heat exchanger) attempts to get around this issue by using an enclosed system that allows just tiny amounts of external air to enter the data centre (less than 4% per wheel revolution). This minimises the need for filtering and moisture control, making the overall system far more energy efficient.
KyotoCooling International BV, the company that sells Kyoto wheels, uses four sections to achieve this energy efficiency. The heat exchange wheel, made from wound corrugated metal, rotates through each section. Internal hot data centre air enters in one section, transfers its heat to the wheel and then exits as cool air to be used back in the data centre through the next section. Cold external air enters in the third section, strips the heat from the hot wheel and exits to the outside air through the fourth section.

Again, external temperatures are an issue, but because of the small surface areas involved in the heat exchange mechanism, less cooling is required compared to using direct cooling via CRAC units or other free-air cooling systems.

Capturing heat through evaporation
Evaporative cooling is another natural cooling approach. As a liquid evaporates, it absorbs heat from its surroundings. Therefore, air circulating past a liquid that can be evaporated at a relatively low temperature will be chilled down as that liquid evaporates. Think of it in terms of a school physics lesson where air blown through a straw into a beaker of chloroform placed on a thin puddle of water causes the water to freeze.

For very warm climates, a contained system with a liquid that evaporates at a low temperature can use solar energy to cause evaporation and to cool the secondary liquid or air. The evaporated liquid can then be condensed to fall back again and maintain the constant cooling during the day. At night, provided the air temperature falls far enough, standard free-air cooling can then take over. Even if the night air is not cool enough, standard CRAC units can be used as necessary, and benefits still exist as the units are only used for 30% to 40% of the time.

I would recommend free-air cooling to all data centre managers building a new facility or retrofitting an existing facility. The potential savings of using natural cooling in the data centre are appreciable, and with energy prices remaining volatile, any approach that gives you some control over cooling systems is worthwhile.

Improve data centre energy efficiency by improving energy generation

Much of the focus on data centre energy efficiency is on how the energy is used within the data centre. Yet this may not be the best way to look at creating a sustainable, fully energy-efficient IT platform. It may be better to look at how the energy is generated in the first place -- where; in what form; and how the by-products are dealt with.

In this tip, we'll consider the realities of energy generation and delivery and examine the impact on data centres.

The energy efficiency chain
Energy generation is subject to efficiency losses along every step of the process. The burning of fuel, the heat losses from that combustion, the efficiency of converting heat into electricity, losses in electrical transmission and distribution, as well as losses in voltage transformation and rectification from AC to DC all introduce losses. In actual practice, around 60% of the original energy produced is being lost before any IT equipment (or, in fact, any business

or consumer equipment) is plugged in. The *Digest of United Kingdom Energy Statistics* (DUKES) comes out with an overall energy efficiency of 38.5%.

Therefore, with only 40% of the available energy being used in the data centre facility and its equipment, improving data centre energy efficiency by 10% will only have a 4% overall efficiency improvement on a total energy basis. On the other hand, improving the energy efficiency based on usage from the point of generation by 10% will result in efficiency gains throughout the whole chain, and 6% at the data centre itself – or actual gains of 50% above what would have been gained purely by messing around at the data centre.

Many new-generation plants use heat recovery as part of their operations, which does lead to higher overall thermal efficiencies, but this heat isn't utilised in any significant manner. Some main generation plants capture the heat for use in complementary endeavours such as agricultural environments. Some power plants are even capturing the carbon dioxide (CO_2) by-product and using this in complementary tasks like tomato production as a means of growing tomatoes more rapidly.

However, not all main generating plant are suitably positioned next to a suitable agricultural outlet, or even near to large conurbations of human dwellings where excess heat can be used for space and water heating. And, although generating thermal efficiency is driven higher, the energy usage efficiencies still remain low because heat and transmission losses continue to be high.

Even though improvements have been made in energy transmission, it is impossible to overcome the basic laws of physics. The longer a wire is, the more energy has to be used to overcome the resistance of the wire. The power loss equation [$P (loss) = P^2R/V^2$] shows that to minimise the energy loss, the voltage (V) needs to be as high as possible, while the resistance of the cable (R) needs to be as low as possible. However, if the voltage is pushed too high, electrical arcing can occur, so wires have to be higher off the ground and spaced further apart. To maintain tensile strength of high tension cables, it is not possible to use extremely low resistance metals (these tend to be more ductile, and, as such, will stretch and finally break). Therefore, a careful balance has to be made between the types of cable used and the voltages the energy is transmitted at.

In general, energy is created at a power station and is transformed into high-voltage alternating current (AC) -- generally above 100 kV -- for transmission. At a substation, it is transformed to a "customer" voltage, which may still be at the kV level or may be at the more normal levels used in homes and data centres of 110 V or 220 V. Once this is delivered to the data centre, it then must be transformed to the various DC voltages used by the computing equipment within the data centre itself -- for example, 3 V, 5 V and 12 V. Each step introduces losses.

Data centre energy efficiency and local energy production

Many data centres now own supplemental power-generation equipment such as diesel generators along with a growing mix of solar, wind and biomass cogeneration technologies. The same limitations that impact a commercial utility provider also affect onsite cogeneration, so although there isn't much you can do to boost the efficiency of a regional utility provider or transmission infrastructure, there are clear strategies that can increase the efficiency of your own power-generating equipment.

The two main areas that data centre managers should consider for data centre energy efficiency are thermal efficiencies and transmission losses. Bringing the point of energy generation closer to the point of use (such as a local generator) will reduce losses. Also, you can ensure that as much thermal energy is captured from the generation process as possible. This is not just looking at electrical energy but also at the heat generated through fuel burning.

One way forward is to look at community combined heat and power (CCHP). The generating plant produces both electrical power and heat. The owner of the plant (for this article, a data centre owner), takes the electrical power it needs as the base load from the generating plant. Any excess electrical energy can be sold off to other local businesses or homes for their use.

As the data centre will then create heat through the thermal losses of electrical energy usage, this needs to be captured and used for space heating or (through the use of heat pumps) for water heating elsewhere in the business. The heat created through burning fuel in the generating plant is captured and sold to local businesses and homes, helping to offset the cost of the generating plant itself.

Tips and tactics for data centre energy efficiency
The key to generating power efficiently is to size things correctly. By building smaller plants, their overall energy efficiencies can be above 80%, which is double the energy efficiency for a basic large generating plant. Equipment costs can be more controlled. High voltage transformation is not required, because the electrical energy does not need to be transmitted over large distances. The generator must cover the base load of the company concerned (not just the data centre's needs, but the organisation's total needs across its associated campus), and consideration should also be given to the future power needs of the business (an often overlooked part of capacity planning).

In many cases, such excess energy can be sold back to the grid at good rates. The UK has a good feed-in tariff covering such a situation. By careful planning, the excess load calculation gives sufficient pay-back from other businesses, homes and feed-in tariffs such that the electrical energy created and used will be at least as cheap, if not cheaper, than electricity taken direct from the grid.

However, a fully available system based on system redundancy may not be financially viable. In that case, the system must have the capability to fail over to the grid should the local system fail.

CCHP can be far more sustainable than a national grid-based energy system. For instance, the University of Warwick introduced a combined heat and power (CHP) system for its campus in 2003, without any "community" aspect (i.e. no sale of excess heat or power to other interested parties. Even so, the payback was calculated at 5.5 years, based on stable 2003 energy prices). Carbon savings were calculated at 3,500 tonnes per annum. Although this is not a data centre example, it demonstrates how CHP is worth considering. And with feed-in tariffs and excess energy resale, ROI would be far faster.

Although headline figures can make it appear that electricity provided can be expensive -- and so counter to the current drive for reduced costs -- careful planning of how excess energy is used can make this a very cost-effective approach to achieve further data centre energy efficiency.

Data centre strategy: Using renewable and sustainable energy sources

Data centre managers consider and explore different approaches to cut down on energy costs as part of their data centre strategy. In a previous tip, we looked at the energy generation process and using local CCHP (community combined heat and power).

In this tip, we explore the issues with using renewable and sustainable energy sources as primary energy sources for data centres and more importantly how to successfully use them.

For many, the question is where do renewable and sustainable energy sources fit in and how can they be used in data centres. Unfortunately, there are many issues around the use of renewable and sustainable energy sources -- such as wind, solar and tidal energy -- which make them unsuitable as primary energy sources for data centres.

The generation capability of wind and solar is intermittent due to lack of wind or sunlight. With tidal power, where the renewable source is completely predictable, the energy produced is cyclical and cannot be counted on as a primary data centre energy source.

Other renewables such as wood or (to a lesser extent) garbage as fuel sources for more standard generation plants may be more useful, but there are problems here too. Those sources remain outside the direct control of a majority of data centre owners, primarily because of their enormous capital expense and regulatory impact. Large power generating companies are more inclined to choose those methods because they provide a mix of generating technologies across their own commercial utility portfolio.

Their use for power generation is only something that makes economic sense for an electricity generator -- there would be no economic case for a small-scale system.

Then there is hydroelectric generation from lakes or river water controlled with a dam. This could be a viable option if the water source is constant and does not dry up or run too sluggishly during dry periods. But few organisations or data centres are built next to a suitable river, so this approach is restricted to a small minority of IT pros looking to improve data centre energy efficiency.

Smoothing the wrinkles in renewable energy availability

Data centre managers that wish to use standard renewables/sustainables in their data centres have to find a viable solution to this quagmire of availability.

One potential solution is to store the energy in batteries to smooth out supply. However, battery technology is expensive and batteries need to be replaced on a regular basis, making the option for many cost prohibitive. Another alternative is to pump water from a low point to a higher one, and then use hydroelectric generation to fill the gaps in output from the primary source.

For a pump-driven hydroelectric solution to work, there has to be sufficient energy to spare when some of the primary source is diverted to pump the water and, as such, the primary source has to be engineered to be larger than the base need -- again, impacting the costs of such a system. Also, unless there is a suitable terrain around for a contained lake and hydroelectric system to be put in place, the cost of creating such a system will be too expensive for all but major generating companies.

Data centre strategy – role of renewable energy sources

So, can renewables be worth considering at all when looking at optimising data centre energy efficiency?

Certainly. One way is to choose an energy supplier that can demonstrate a good mix of renewables in its portfolio. If you're looking for a more sustainable energy policy with greater control over the points of generation themselves, then you should see renewables (wind or solar especially) as a secondary rather than a primary source. In other words, since they are not a predictable energy source, they can't be the main source of energy. But they can be a secondary source, and that provides a means of minimising the amount of primary energy sources used.

So, data centre operators can have their own energy generation capability based on renewables, but unless they are going for a nonrenewable approach (eg, CCHP), then the grid will have to remain their main energy source.

Another strategy is to figure a base load for the data centre and rely upon the commercial utility grid for that essential supply, but then turn to renewable/sustainable alternative energy sources to address additional energy demands.

How to use sustainable/renewable energy sources?

Let's assume that the data centre requires 2 MW of energy overall. And, let's assume that with cyclical workloads and the capability to turn off some parts of the IT estate during certain times, there is a minimum energy need of 1.6 MW, with the 400 kW remaining being used by the more cyclical workloads, therefore only needing power on an intermittent basis.

The organisation concerned negotiates an agreement with a central generator for 1.6 MW of "base" power on a continuous basis on preferential rates (as there will be no variability in the load taken, many generators will offer better rates because it helps them in their energy-generation planning). The additional 400 kW required is then theoretically provided via renewables.

Still, the problem with renewables is that they are not a guaranteed source. Ideally, the energy from the renewables would be available just as the cyclical workloads needed it. In practice, this will not happen. Some of the time, the renewable energy will be available when the workloads aren't running; other times, the workloads will need to run when the renewable energy is not available.

But this can be ironed out. The majority of countries now offer feed-in tariffs, where energy generated "off-grid" (ie, locally to an energy user) can be fed back into the grid which will reimburse the local energy producer, usually at a generous rate. Therefore, any renewable energy produced at times when the data centre (or the rest of the organisation's estate) cannot use it can be sold to the grid.

Where the energy generated by the renewables and the data centre workloads coincide, wonderful! The energy, then, is used in the data centre and everything works as per theory and helps you execute your planned data centre strategy.

However, when the workload requirements and the renewables energy generation are out of sync, other approaches have to be brought to bear Because of the financial risk involved.

Large users of energy will have special deals with energy providers, based on tiered usage. So, if a 10 MW data centre says it will buy 10 MW from a supplier, it will get a better per-KW deal than one that buys just 1 MW.

But if a data centre owner wishes to use renewables for just 50% of the energy and agrees a deal based on 5 MW of energy, then anything that is used above that 5 MW will be charged at a higher price per-KW rate.

At that juncture, if the user agrees on a 6 MW deal to get a lower price still and allows for a bit of overrun, but uses only 5 MW, the energy provider will have the rights to put in place some charge-back mechanism that would undoubtedly be in the supplier's favour.

So, does the data centre owner agree to a tier price that only just meets his need and pay anything over and beyond? Or does he agree to a price for an oversupply and use it come what may, putting more of the renewables energy back into the grid to maximise energy income?

The risks can be pretty big for data centre owners who get their math wrong.

Spot pricing and aggressive negotiation

The best approach is to buy energy from the grid based on spot pricing, rather than on long-term agreements. In many cases, spot pricing may be more expensive than long-term agreements, and, as such, the aim is to minimise energy usage. However, in other cases, good prices can be found for using energy at certain times from certain sources.

For example, a standard fossil-fuel-based generating plant is not good at flexing to follow changing energy usage. Therefore, a degree of energy is generally purposefully lost through shunting the steam required for driving the generators through to cooling towers. The fossil fuel is still burnt, but no useful energy is created from it. Therefore, if a possible customer comes along who is willing to pay for some of the energy that would otherwise be shunted away as lost steam, then a good price may well be negotiable, as anything is better than nothing to the generating company.

The key is to engineer the system so that there is a good balance of how renewables feed into the data centre, how feed-in tariffs provide revenue, and how these tariffs offset the costs of spot energy usage when extra external energy is required.

From my point of view, renewables on their own cannot meet a data centre's energy needs outside of centres that have the wherewithal to be able to build next to massive hydroelectric generating systems or similar, but renewable energy sources can be used as part of an overall approach to cut down on overall energy use from the grid.

This creates a less carbon-heavy, more sustainable energy plan and helps you succeed in your data centre energy efficiency strategy.

Wrapping it all up – 10 steps to better sustainable energy management in the data centre

Over the last 11 articles, Quocirca has provided insights into different approaches for managing the energy usage across data centres, from choosing a more sustainable energy provider through to generating your own energy. To decide what the best overall approach is for your organisation, Quocirca summarises the following points which have been covered in more detail throughout the series:

- **Rationalise, consolidate and virtualise.** If your data centre is running at less than 50% server workload utilisation, then a large proportion of your energy is being wasted. With average utilisation rates still running at 15% or less, 85% of a server's possible work output is being wasted, yet the server is still using a lot more than 15% of the energy of a 100% utilised server, is giving off more than 15% of the heat, and still requires more than 15% of the cooling. Driving up overall utilisation rates will improve energy metrics – not just at the server itself, but across the data centre, including cooling, uninterruptable power supplies (UPSs) and so on. For every Watt that is saved in the data centre, this can be saving up to 4 Watts at the point of generation due to the thermal and transmission losses involved.
- **Review the data centre operating temperature.** If the existing operating temperature is in the low 20 degrees centigrade/70 degrees Fahrenheit, then it is likely that you are running your data centre too cool. Look to move upwards towards 24C/75F – or even higher if the majority of your IT equipment is relatively new. Each degree increase in temperature requires less cooling, so less energy.
- **Use air cooling effectively.** If you are cooling down the whole of the data centre, then there is a massive volume of air that is being cooled for no real reason. Use hot aisle/cold aisle approaches, or in-row cooling to minimise the amount of cool air required and ensure that it is ducted effectively to cool down any hot spots.
- **Don't dismiss water cooling.** Water has a higher coefficient of heat convection than air, and is much more efficient as a coolant. Newer systems, based on negative pressure water systems, have extremely low failure rates and can dramatically lower cooling (and therefore energy) costs.
- **Optimise energy distribution systems.** For every voltage transformation involved in a distribution system, there will be a necessary loss of energy. Therefore, maintain a "flat" distribution system – move from sub-station AC voltages through to data centre AC and DC voltages as fast as possible and route from there. Although only small percentages are involved here, saving even 1% of a large data centre's energy bill is worthwhile.
- **How big is your data centre now?** If you have an existing large data centre, then it may be worth considering generating your own power directly, particularly if the data centre is close to other businesses or housing developments. Excess electricity and heat can be sold into the local community, offsetting the cost of the energy generation. Overall energy efficiency can rise from around 30-40% to 80%+ through the use of community combined heat and power (CCHP) generators as opposed to centralised fossil fuel generators.
- **How big is your data centre likely to be in the future?** If you believe that your data centre is likely to increase its energy requirements in the foreseeable future, then self-generation could make sense, even if

the data centre is only of a medium size now. With energy costs being so variable, attempting to hedge costs now makes sense. Also, if you have a lot of excess power to sell, putting this back into the grid can provide useful feed in tariff revenues. However, if you believe that your data centre will remain the same size or shrink as cloud computing becomes more of a reality, then self-generation may not make as much financial sense.

- **Choosing a site for a data centre.** If you have no choice as to the location of your data centre (maybe as it is an existing facility), you will have to keep thinking as to how to save energy. If a new-build facility is on the cards that needs to be within a defined geography, then building it close to a more sustainable energy source, such as hydro or geothermal will make your data centre more energy sustainable, even if no money is saved. However, if the placement of the data centre is completely flexible, then moving it to a colder climate where free air cooling can be used will not only provide a more sustainable system, but also lower energy costs.
- **Use of renewable energy sources.** Probably not practical as a direct replacement for operating the entire data centre, but as an incremental source of power, then renewable sources should be considered. Wind or solar energy is not predictable enough in most cases to provide core power requirements, but can be used to deal with the cost of energy for peak workloads through using feed in tariffs. Energy created while the wind blows or the sun is out is sold into the grid; energy required to deal with peaks is either drawn directly from the wind or solar systems if available, or bought in at spot pricing if the systems cannot supply it.
- **Don't forget outsourcing.** Not only does outsourcing move energy usage out from the organisation's energy bill onto someone else's, and as such may sidestep certain carbon emission taxes, but an outsourced environment may well be far more efficient than your own. Use of a co-location data centre facility or, even better, a managed hosting provider can provide better energy management capabilities across your total data centre infrastructure requirement, as well as pushing effective utilisation rates up through the use of a highly engineered, multi-tenanted environment.

Data centres will remain a focus for energy costs within all organisations. Making moves to be more energy efficient – and to have more sustainable data centres – does not necessarily mean massive changes to the facility and the equipment housed within it. Small steps can be taken with high levels of payback – both through achieving cost savings and/or in meeting your organisation's corporate social responsibility (CSR) targets.

About SearchVirtualDataCentre.co.uk

SearchVirtualDataCentre.co.uk is a comprehensive online resource for IT professionals seeking cost-saving strategies and help with creating, implementing, managing and updating U.K. data centres. Topics include utilising virtualisation technologies for server consolidation, disaster recovery, lab and testing, desktop centralisation, server selection and other server hardware issues in virtual infrastructures. Our site provides unbiased news analysis, a library of essential tips, informative learning guides, and white papers -- all in an effort to arm you with the tools and tactics you need to do your job successfully and make the right technology purchasing decisions.

Activate your FREE SearchVirtualDataCentre.co.uk membership today!

Topics covered include:

* cloud computing
* data centre design
* disaster recovery
* power and cooling
* server hardware
* server virtualisation
* systems management
* and more

Subscribe to our free newsletters to get our news and advice delivered to your inbox. And make sure to check out all the features available on SearchVirtualDataCentre.co.uk today!

About TechTarget

TechTarget, a leading online technology media company, gives technology providers ROI-focused marketing programs to generate leads, shorten sales cycles, and grow revenues. With its network of more than 60 technology-specific websites and more than 7.5 million registered members, TechTarget is a primary Web destination for technology professionals researching products to purchase. The company is also a leading provider of independent, peer and vendor content, a leading distributor of white papers, and a leading producer of webcasts, podcasts, videos and virtual trade shows for the technology market. Its websites are complemented by numerous invitation-only events. TechTarget provides proven lead generation and branding programs to top advertisers including Cisco, Dell, EMC, HP, IBM, Intel, Microsoft, SAP and Symantec.

For inquiries about SearchVirtualDataCentre.co.uk contact Brent Boswell at +44 (0) 7584311889 or Sean Matthews at 1-617-431-9414.

TechTarget UK
Marble Arch Tower
55 Bryanston Street
London W1H 7AA

SearchVirtualDataCentre.co.UK

REPORT NOTE:
This report has been written independently by Quocirca Ltd to provide an overview of the issues facing organisations seeking to maximise the effectiveness of today's dynamic workforce.

The report draws on Quocirca's extensive knowledge of the technology and business arenas, and provides advice on the approach that organisations should take to create a more effective and efficient environment for future growth.

About Quocirca

Quocirca is a primary research and analysis company specialising in the business impact of information technology and communications (ITC). With world-wide, native language reach, Quocirca provides in-depth insights into the views of buyers and influencers in large, mid-sized and small organisations. Its analyst team is made up of real-world practitioners with first-hand experience of ITC delivery who continuously research and track the industry and its real usage in the markets.

Through researching perceptions, Quocirca uncovers the real hurdles to technology adoption – the personal and political aspects of an organisation's environment and the pressures of the need for demonstrable business value in any implementation. This capability to uncover and report back on the end-user perceptions in the market enables Quocirca to provide advice on the realities of technology adoption, not the promises.

Quocirca research is always pragmatic, business orientated and conducted in the context of the bigger picture. ITC has the ability to transform businesses and the processes that drive them, but often fails to do so. Quocirca's mission is to help organisations improve their success rate in process enablement through better levels of understanding and the adoption of the correct technologies at the correct time.

Quocirca has a pro-active primary research programme, regularly surveying users, purchasers and resellers of ITC products and services on emerging, evolving and maturing technologies. Over time, Quocirca has built a picture of long term investment trends, providing invaluable information for the whole of the ITC community.

Quocirca works with global and local providers of ITC products and services to help them deliver on the promise that ITC holds for business. Quocirca's clients include Oracle, Microsoft, IBM, O2, T-Mobile, HP, Xerox, EMC, Symantec and Cisco, along with other large and medium-sized vendors, service providers and more specialist firms.

Details of Quocirca's work and the services it offers can be found at http://www.quocirca.com

Disclaimer:
This report has been written independently by Quocirca Ltd. During the preparation of this report, Quocirca has used a number of sources for the information and views provided. Although Quocirca has attempted wherever possible to validate the information received from each vendor, Quocirca cannot be held responsible for any errors in information received in this manner.

Although Quocirca has taken what steps it can to ensure that the information provided in this report is true and reflects real market conditions, Quocirca cannot take any responsibility for the ultimate reliability of the details presented. Therefore, Quocirca expressly disclaims all warranties and claims as to the validity of the data presented here, including any and all consequential losses incurred by any organisation or individual taking any action based on such data and advice.

quocirca