e-shelter

The CIO's Dilemma

How to meet the challenges of the modern data center

> Whitepaper

In corporation with FROST & SULLIVAN

The CIO's Dilemma

Executive summary

Data centers are essential to modern organisations and to their ability to perform mission-critical tasks. Given the vast amounts of very different types of data that organisations work with, the emphasis in a modern data center has shifted from data processing and data storage to data management.

We have found that data loads are increasing rapidly and that this trend is likely to continue. At the same time, data centers are under pressure because organisations are seeking to comply with a long list of rules and regulations pertaining to data retention which have come into force in recent years. These phenomena translate into a number of "hard" challenges:

- Business Continuity
- Scalability
- Financials

Well-managed organisations have generally been able to meet the "hard" challenges, but a number of additional challenges are emerging – we can these "soft challenges" – which ever the best managed organisations will be struggling with:

- Energy Efficiency
- Security
- Public Relations

We have described and analysed both types of challenges. An organisation failing to meet all these challenges excludes itself from adopting many innovative business models that might have generated substantial benefits.

In order to meet the challenges, data center capacities must increase. The first decision a CIO needs to take is whether to "Make" or "Buy". However, the "make" scenario – developing the data center capacity in house – has become very difficult, because of the unavailability of suitable construction sites, skilled staff and, not least, because of the difficulty of realising economies of scale and high levels of energy efficiency. Consequently, organisations will be looking at alternatives to the traditional in-house data center. We have analysed three such alternatives – outsourcing, application outsourcing and colocation.

Outsourcing will allow organisations to meet most of their IT and/or data center challenges, but there will be serious issues in terms of employee relations and loss of control. Application outsourcing is a quick fix that may address a peak in IT needs at a given time but will not provide a long-term solution to the challenges. Colocation can be a good alternative to an in-house data center, but the space available in colocation centers may not be sufficiently customisable to suit an organisation's specifications. Moreover, there may not be enough spare capacity in the colocation center to ensure that sufficient scalability is built into the solution.

We have found a fourth alternative – the Campus Model. The campus is owned and operated by a thirdparty data center, but the floor space is fitted to the customer's specifications. Moreover, organisations are able to rent cages, suites or even entire buildings on the campus. Power, cooling and security are provided by the third-party data center. Because of the size of the campus, the third-party provider is able to maintain an infrastructure that guarantees an extremely high availability of sufficient power. It also assumes the responsibility for ensuring that sufficient power will always be available to meet future demands. The customer organisation retains responsibility for all other elements of the data center.

On the basis of an analysis of the in-house data center and of the four alternative "buy" models, we conclude that the campus model will enable most organisations to effectively meet all their data center challenges.

The CIO's Dilemma

Table of contents

The CIO's Dilemma

1. The evolving role of the data center

The modern data center came into being with the introduction of the mainframe computer from the late fifties and onwards. Because of the sheer size of the early mainframes, a significant amount of space was necessary to house them, and it became practical – and therefore increasingly common – to place all an organisation's mainframes in a purpose-built data center. At the time, the computer industry was dominated by the big mainframe vendors – IBM, Burroughs, Honeywell, Univac, ICL, Bull etc. – their advantage being the ability to handle large amounts of transactions. Consequently, the emphasis in early data centers was on data loads (in a process that could be illustrated as INPUT > COMPUTING > OUTPUT). The early adopters of data centers were organisations that had an almost constant and very sizeable amount of transactions to handle, banks being the prime example.

With the paradigm shift to distributed computing and, especially, with the introduction of the personal computer into the workplace – a development which began in the early eighties – the role of the data center started evolving. The emphasis moved to archiving (i.e. data storage). The information that needed to be archived was directly liked to the data generated locally in the organisation, in other words the organisation was in complete control of its data storage needs. These needs would increase in line with the growth of the organisation (or in line with rate at which the organisation decided to embrace distributed computing). When a company added more clients or new product lines, more information would need to be processed and archived, but this was a largely linear development, resulting in only a modest increase in the need for data storage. Moreover, as it was completely in control, an organisation was able to plan ahead and to expand its data center as required.

The advent of the internet and the redefinition of work processes in most organisations have led to a veritable explosion in the amount of data available to an organisation. Access to the internet is now so readily available in the industrialised world that most interaction between organisations and their users has become electronic in some form. Social networking sites have taken off, web 2.0 has been introduced, and companies are trying to defend and increase their competitiveness by reinventing themselves as customer-centric organisations. Organisational strategies and day-to-day work processes are now fuelled by customer data which must obviously be available and shared constantly in a reliable manner. Much of the information organisations need to handle and share today is multimedia-based. The challenge of dealing with all that data has, again, changed the role of the data center, shifting the emphasis from data storage to data management.

When most processes that take place in an organisation are dependent upon data, the facility where that data is managed – the modern data center – becomes mission-critical. Modern companies like Amazon and eBay have made the active use of customer data an integral part of their business concept, in order to gain and sustain competitive advantage. This sophisticated use of data involves activities such as data mining and complex reporting. The undisputed success of such business concepts has inspired traditionally conservative organisations (e.g. banks, insurance companies and even public administration) to follow in Amazon's and eBay's footsteps. This means that the data center has never been more important: If it suffers an outage, the organisation grinds to a halt.

The CIO's Dilemma

2. Current data center trends

The mission-critical nature of the data center in a modern organisation is underpinned by the fact that most processes no longer have manual back-up procedures. Manual (often paper-based) procedures have been abolished because they are costly to maintain, and because they potentially invalidate the complex security measures that organisations are putting in place. Credit card transactions are an excellent example of this development. Ten years ago, offline credit card payments were completely normal. When the X.25 network (or whatever network the point-of-sale terminal was connected to) failed, the shop assistant would simply print the credit card details on a paper slip. Today, there are even payments that will not be authorised, if the point-of-sale terminal is not equipped to read the chip on the credit card.

Data loads never cease to increase

The development of data loads (i.e. the computational loads of data needing to be processed) in recent years has been truly exceptional. Not only has the increase in data loads been enormous, it also appears that the rate at which data loads increase every year is not slowing down.

In his article "The Future of the IT Organisation: Do You Run a Green Machine," from December 2006, Richard Edwards, a senior research analyst at Butler Group, points out that the density of IT equipment in companies is up to fifty times higher today than what it was fifty years ago. In addition to that, the residential IT density is also going up. IT equipment has been – and is continuing to be – deployed in the residential environment at a phenomenal rate. The prices of IT equipment have fallen so much that it is accessible to most people in the industrialised world. Even so, there are often government incentives available to those who might not be able to afford IT equipment, because Western governments are beginning to regard computer ownership and internet access as an indicator of wealth, much the same way the ownership of motor vehicles or the number of telephone lines in operation might have been fifty years ago.

On this background, it is easy to understand why data loads are increasing and why they will continue to increase. There has never been a decrease in data loads from one year to the next, and there is no reason to believe that there ever will be. One of the areas where data loads are increasing the most is electronic mail. In its whitepaper The Toxic Terabyte, published in 2006, IBM estimates business e-mail to grow at a rate of 25-30% every year. In the same whitepaper, IBM estimates that just three years from now, the World's combined information base will double in size every 11 hours. Even our vocabulary will struggle to keep up as gigabytes become terabytes; terabytes become petabytes etc.

Although electronic mail may seem like an innocuous, almost banal application, it is an excellent example of what is happening to data centers at the moment. The handling of a single e-mail gives rise to a series of computational activities that must be performed with minimal delay, creating data load peaks at specific times of the day, seeing that so much essential information circulates via e-mail, all arriving e-mails must be stored and multiple back-ups kept of the storage medium. Very typically, parallel storage takes place, because the addressee will be likely to store attachments both locally and on shared network drives. The consequence is an almost explosive increase in the data center needs (i.e. the combined needs for data center resources relating to data loads and storage), making e-mail a prominent driver behind today's evolution of data center needs.

Small to medium-sized enterprises managing erp and mail systems

SMEs are, of course, diverse, but they will certainly all use email and experience the drain on their data center resources which was illustrated on the previous page. In recent years, the tendency has been for downrated versions of Enterprise Resource Planning (ERP) systems (e.g. SAP) to be introduced into the SME space. Although they are downrated, they are still somewhat over engineered for many SMEs. They will not only integrate previously separate systems, they will create entirely new processes (or move manual processes onto an IT platform) which will increase the data center needs. As ERP becomes mission critical, redundancy and disaster recovery procedures must be adopted.

Internet-based businesses managing customer search profiles

Data centers are essential to companies like Google and Yahoo. Google anticipates "a significant rise in cost of revenues in 2007 primarily as a result of anticipated increases in traffic acquisition and data center costs". Yahoo, "in an effort to reduce the likelihood of a geographical or other disaster impacting its business", is continuing to invest in additional data centers located around the world. Companies are increasingly considering the cost of their services going offline and are willing to invest in solutions that will guarantee business continuity. Because so much essential business is now conducted online, this tendency is spreading outside the dot com companies.

Financial trading require low latency data center solutions

The difference between the success or failure of a financial trade is now counted in milliseconds. Low latency connectivity is becoming a must, as market players develop more sophisticated trading strategies and demand direct access to liquidity pools. By placing data centers near execution venues, companies can reduce distance-related delays, which will, in turn, save significant amounts of money on missed markets. Similarly, with a data center close to a banks headquarters (but not so close as to subject the data center to the same risks that might threaten the headquarters),

The CIO's Dilemma

staff is able to reach the data center in a minimum of time, when this becomes necessary.

Healthcare managing patient records

Healthcare organisations are under pressure to modernise their procedures, because the tax-payer is not willing to foot the bill from new treatments and an ageing population. A good example is the Picture Archiving and Communications System (PACS) of the National Health Service in England. PACS will enable centralised storage of images, full inter-operability and compatibility across the NHS with ensured access to images throughout. It will be tightly integrated with patient records, removing the traditional barrier between images and other patient records, and providing a single source for clinical information. PACS eliminates the need to physically exchange pictures, saving money and improving the service to patients.

2.1. The different data center needs of different organisations

The underlying data center needs are increasing in virtually every organisation in the World, but the data center needs do not always increase for all the same reasons. The business model of a company – or the operational model of a non-commercial organisation – drives the development of data center needs. It immediately follows that a change in business model at a company can result in radically different data center needs, practically from one day to the next … and the data center needs usually go up. The opposite argument is also true. If a company underlies capacity and development constraints at its data center(s), it is effectively prevented from adopting certain business models that could have been beneficial.

It will be a struggle for organisations to fulfil all their future data center needs, but the consequences of not doing it, would be disastrous and, as the example from the English National Health Service shows, even potentially a matter of life or death.

2.2. External factors influencing data center needs

Many factors that are completely outside the control of an organisation can, nonetheless, be the main drivers behind the development of the data center needs – notably the data management needs – of an organisation.

Regulatory compliance as a main driver for data management

The last couple of years have seen the introduction of numerous new regulatory measures. Some of these are already effective, the remaining measures coming into effect over the next three years. Many of these regulatory measures are promulgated by supra national organisations such as the European Union, in addition to which practically all countries in the industrialised world will also have national legislation covering data protection, terrorism, money laundering and the like. What these rules and regulations all have in common is that they require sophisticated data management to be performed by the targeted organisations. Often – and this is especially true of anti terrorist measures and rules and regulations targeting the financial sector – there are requirements for capturing large amounts of data, storing this data for a long period of time, and making it available (to the authorities, typically) with minimal delay. This adds a significant data storage requirement to the already considerable data management requirement. Because organisations have no way of knowing exactly how much data they will end up having to store, and because rules and regulations are dynamic and will be modified frequently, organisations are faced with the challenge of introducing wide-ranging flexibility and scalability into their data storage solutions. Rules and regulations are often a result of political compromises – especially the supra national ones. The well-publicised dispute between the European Union and the United States about the amount of data regarding transatlantic passengers that airlines must capture, store and make available (and the lawfulness of doing so) to US authorities is a first class example of the predicament in which organisations can find themselves.

> "The data manager's job has never been more difficult. End users demand continuous access to application while data and business are more conscious than ever of the need for rapid restore and effective disaster recovery capabilities. At the same time, the growth in data far exceeds the growth of the business. Regulatory compliance is forcing companies to store more of the day data than ever before. IT budgets are so tight that it can seem impossible to keep everyone happy"
>
> **Christo Conidaris, Vertriebsleiter, Sepaton Inc.**
>
> Sepaton is a data security company, specialising in data storage and retrieval requirement compliance.

Frost & Sullivan believes that regulatory compliance will become the single most important driver behind the development of data center needs over the next couple of years. Private enterprises and public organisations are all struggling to meet the various deadlines for compliance. The predicted growth in data loads will only make

The CIO's Dilemma

this struggle worse. Some rules and regulations target specific sectors, whereas others are generic:

- Generic (Sarbanes-Oxley): Sarbanes-Oxley (or SOX, as it is often called) is focused on providing transparency to business processes, decision-making and communication. As a result, information on business processes and all communication must be stored and made available upon request. Systems that can capture all types of communication require massive storage capacities as well as sophisticated data management capabilities to ensure compliance.
- Financial Services (Basel II, MiFID, Solvency II): Basel II sets up rigorous risk and capital management requirements designed to ensure that a bank holds capital reserves appropriate to the risk the bank exposes itself to through its lending and investment practices. MiFID is about creating a single European market for investment services. Solvency II extends Basel II to the insurance industry.
- Telecommunications (EUDRD): The European Union Data Retention Directive requires all calls, SMS, MMS, e-mails and instant messaging conversations to be captured and mapped to the personal details of the originators, their locations and IP addresses as well as the first and last data package. The storage capacity necessary to ensure compliance for telecom operators and internet service providers is extremely high.
- Retail (PCI DSS): All payment card transaction details must be captured and stored. This will create a massive storage requirement for all retailers, large high street chains or small internet-based shops alike.

The increasing complexity of technology

The increasing complexity of the technologies employed by data centers are becoming an issue with CIOs and data center managers around the world. To cater for the more sophisticated requirements that data centers must live up to, organisations are under pressure to acquire, enhance and retain specialist know-how and experience – a task which is getting more difficult. AFCOM, the American organisation for data center professionals, estimates that, by 2015, the talent pool of qualified senior level technical and management data center professionals will have shrunk by 45%. These difficulties will even be felt in the shorter term. In a survey conducted by AFCOM in 2006, 47% of respondents believe that finding senior-level data center professionals with the right skill sets will become more challenging over the next five years. The situation in Germany is no better. The German IT association BITKOM has conducted a survey (December 2006) in which 56% of the participating companies report that they already suffer from a shortfall of skilled IT personnel. BITKOM estimates that there are currently 20,000 open IT positions in Germany that companies have trouble filling. It sees this situation as so threatening to the development of the German IT industry that it has written a whitepaper – "Standpunkte zur Zuwanderung hochqualifizierter Arbeitskräfte – Den Wettbewerb um die besten Köpfe gewinnen" – in which it argues that immigration from outside the European Union should be made easier for workers with IT skills.

Power and Cooling in a data center are inextricably linked. The high power requirement from the highdensity IT equipment in modern data centers increases the requirement for cooling which, again, boosts the demand for power. Additionally, managing cooling is becoming increasingly complicated. Research into areas such as dynamic cooling will create energyefficient cooling solutions, but these will require highly specialist skills to manage – skills that will be in short supply. Moreover, many data centers will be forced to relocate closer to a power source, in order to guarantee a sufficient supply of power. This limits the choice of location, but it also limits an already strained supply of top-notch specialists because the power sources are rarely found in or near the major cities where people tend to want to live. A more detailed discussion of the power and cooling challenges facing data centers is found in sect. 2.5 of this whitepaper.

PCI DSS defines a requirement for capturing and storing payment card transaction details.

EUDRD creates a requirement of European telecom operators to retain data relating to calls and other traffic.

The data management tasks performed by data centers are becoming increasingly complicated due to the sophistication of many new applications. A data center needing to provide its organisation with PCI DSS or EUDRD compliance, for example, must employ highly specialised professionals. On top of that, the more generic business process related applications are also acquiring higher levels of sophistication that will require specific skill sets to be maintained. Although business process outsourcing might alleviate this problem, Frost & Sullivan does not believe company will be willing to outsource the running of their mission-critical applications, least of all outsource these to offshore locations. To some extent, legislation will prevent them from even considering his option.

Building up and maintaining know-how in all the specialist areas required to run a modern data center will be an almost insurmountable challenge to most organisations.

2.3. Implications for the modern organisation

We have shown that, unless an organisation downsizes and scales back its activities, its data center needs will increase. Most organisations are acutely aware of this, but they do not always have a solution readily available.

The immediate implication of rising data center needs is a necessity to expand. In Symantec's 2007 survey, two thirds of respondents believe that they will run out of data center capacity within the next

The CIO's Dilemma

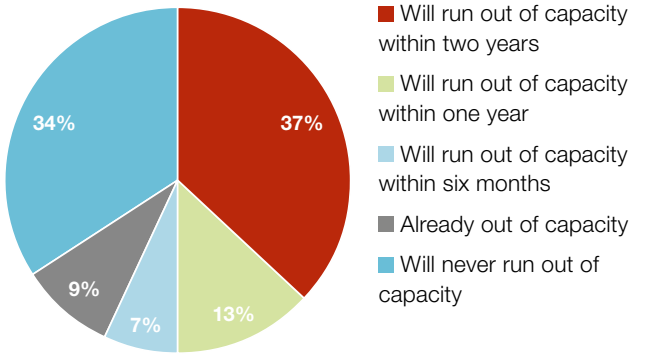
two years. This is a very serious problem. To address it in-house, companies can either refit and redevelop an existing environment or buy land to build a new data center. Refitting existing environments is challenging. Due to the cooling and power requirements, generic office buildings are generally unfit for data center space, an even an older building, purpose-built as a data center in its time, could be unfit by today's standards. If an operational data center is refitted, the risk of outages is considerable. A purpose-built new data center would be a better solution, except … it isn't always possible. Suitable sites are scarce, authorities are restricting the building of new sites through environmental legislation, the capital costs are very high, and the lead times are long between the decision to build and actual relocation of the data center.

"We're just about out of capacity and using any piece of sello tape and string to patch up the data center. We're even converting old meeting rooms into data centers."
Senior Architect for a Global Financial Firm, in Symantec's 2007 survey.

Whichever solution an organisation opts for, it is essential that its data center should be future-proof, so that it may be adapted to meet all technological and regulatory challenges within a foreseeable future. Frost & Sullivan believes that organisations should look at least ten years ahead when making decisions regarding their data centers and that minimising costs in the short term, could become very costly in the long term.

Every CIO's recurring nightmare is moving his or her data center, and this is the one reason why the importance of future-proving a data center cannot be overstated. There are tremendous complications associated with moving a data center, and the risk of outages

How much of an issue is data center capacity for your organisation?



Source: Symantec's Data Center Capacity Survey, 2007

and other business disruptions in the transition phase is large. Even if an organisation runs several data centers, taking one off line temporarily will remove an element of redundancy that is crucial to business continuity.

So, if organisations know that they are running out of space, why are so many seeking stop-gap solutions, rather than doing something about the problem once and for all? Frost & Sullivan believes that there are a variety of reasons for this. First of all, there is business continuity: Most companies cannot afford any down-time, and they may not have the expertise and/or resources in-house to manage a transition. Similarly, companies may not have the expertise to articulate demands for a new data center and to write the specifications. Also, they may lack resources for the technical planning and design of a data center, something which is hugely important in relation to power supply, cooling and the overall cost-efficiency. Last but not least, there is the element of cost. A stop-gap solution will be cheaper in the short term, whereas a future-proof solution will necessitate high capital expenditure and probably also generate higher operational costs in the short term. CIOs know this, of course, but that does not mean that they can always obtain board approval to make the necessary investments at the appropriate time.

2.4 "Hard" challenges

When an organisation makes decisions concerning the future of its data center(s), meeting three, well-defined challenges is paramount. We call these challenges "hard" because an organisation has little discerning power in relation to meeting them.

Business continuity

We have already shown several examples of why business continuity is important to modern organisations, and we have also shown that ensuring Business Continuity is, in many cases, a legal requirement. We shall discuss four separate challenges that must be met in order to ensure Business Continuity: Power, Cooling, Communications and Transport Infrastructure.

Power density (typically defined as the amount of power delivered per square metre of data center floor space) requirements are increasingly rapidly as a result of the increasing IT density discussed earlier. Most utility companies already cannot deliver the power densities that modern data centers require, so finding an adequate supply of power is a major challenge which often involves re-locating the data center outside the major cities, closer to a source of power. What is worse, utility companies cannot guarantee that an adequate power supply will be available all the time, as they are even struggling to provide ordinary households with power at certain times. Power outages and blackouts are becoming more and more common. Just from the year 2000 and onwards, there have been 38 serious power outages in the industrialised world, including the 28 September 2003 blackout in all of Italy and parts

The CIO's Dilemma

of Switzerland, affecting 56 million people and lasting up to 10 hours. On 4 November 2006, approximately ten million people mainly in Germany and France experienced a blackout caused by a chain reaction that had begun when a 380 kV power line across the river Ems had been turned off so that the cruise ship Norwegian Pearl could pass safely. Power outages happen every summer because networks are unable to cope with the sudden surge in demand from air conditioners. When exactly that happened in Los Angeles on 24 July 2006, MySpace, one of the world's most successful social networking sites, was off line for half a day. Major data centers in San Francisco and London have suffered outages because of power failures, affecting companies such as BP. In the London outage, the emergency power generators failed to kick in, suggesting that only connection to multiple, independent sources of adequate power will ensure business continuity.

Cooling consumes tremendous amounts of power. Traditional cooling systems were not designed for the IT densities we see at data centers today, and a data center outage can just as well be caused by a failure of the cooling system as by a power cut. Broad Group estimates that half of today's data centers do not have sufficient cooling capacity to cater to high-density equipment. It is necessary to study alternative methods of cooling (e.g. in-rack cooling and free cooling). The challenge is to ensure sufficient cooling without increasing the power consumption.

Communication networks must be available to ensure that back-up communications options are available for data centers to ensure continuity. This could involve alternate routing and ducting and preferably also the use of multiple carriers. High-speed data access is another essential requirement which is why a number of data centers are built on or near internet exchanges. Proximity to fibre routes is a definite must.

"Scalability is an absolute priority to EDS. Ensuring sufficient scalability in our data centers means being far away from any issues. We know that there will be no restrictions to the ability of our data centers to serve EDS and its customers in the future."

Ralf Schlebusch, Manager
Facilities & Asset Services EMEA, North & Central Hub, EDS

Transport infrastructure cannot be overlooked. There are occasions where an organisation needs to despatch staff to the data center without delay. The IT equipment itself is normally stable, whereas cooling equipment is inherently maintenance-heavy. If the cooling of a rack fails, or worse still, if the cooling of an entire section of a building fails, then all affected IT equipment must be shut down. When a data center plays a part in an organisation's disaster recovery procedures, staff must be able to rapidly travel to the data center to continue working after the main site has become unserviceable.

Scalability

A data center must be scalable, because an organisation cannot know precisely how much its data center needs will grow and what the timeline will be for that growth. A data center which is not scalable could run out of capacity almost overnight.

Financials

Striking the right balance between cost and performance is always a challenge, and in the case of data centers, meeting that challenge is closely linked to obtaining economies of scale. Obtaining economies scale means optimising operational expenditure. A large, new, purpose-built data center is significantly cheaper to run than several small, old data centers, to a large extent because the higher energy efficiency of a new data centers translates into large savings on the cost of power. Few organisations are under any illusion that they will ever be able to save costs on their data centers, and it is certainly true that the cost of maintaining a data center is almost irrelevant compared to the disastrous consequences of a data center failure.

"A couple of years ago, I would measure a data center in square footage. Now I look at megawatts of power. It is a new way of measuring technology."

Debra Chrapaty, Senior Manager, Microsoft
MS Infrastructure in Fortune, August 2006

Generally speaking, Frost & Sullivan finds that most well-managed organisations have been able to cope with their hard challenges up until now. The problem is, however, that organisations may not be able to cope in the future, as the challenges mount.

2.5 "Soft" challenges

A number of new challenges are complicating the picture. We define these challenges as "soft", because they are not easy to define, and because they partly relate to perceptions. The "soft" challenges are evolving, and are destined to rival the importance of the "hard" challenges when an organisation makes decisions regarding its data center(s).

Energy efficiency (going green)

Establishing and communicating a green profile to the public is becoming increasingly important to all industries. IT equipment vendors and data center operators have all recognised the need to go green, and an organisation called The Green Grid has been established. The Green Grid is a consortium of information technology companies and professionals seeking to improve energy efficiency in data centers around the globe. Existing data

The CIO's Dilemma

centers consume an incredible amount of conventionally generated power and release hot air to the surroundings. Because it would be both politically and practically impossible to let the power consumption rise in line with data loads, it is universally recognised that future data center strategies will be primarily influenced by green issues Sustainability, CO_2 emission reduction and renewable energy will be important challenges for all data centers to meet. To this end, the Green Grid has developed a standard measurement of energy efficiency – the Power Usage Effectiveness (PUE). By creating a common benchmarking reference, the Green Grid hopes to improve the energy efficiency of data centers by measuring over time whether designs and processes are improving.

Power Usage Effectiveness (PUE)

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

A PUE factor of 1.5 is regarded as very efficient, whereas a factor of 3.0 is regarded as very inefficient.

Source: The Green Grid

Going green can also generate cost savings. IBM will be investing $1bn a year in "greening" it massive data centers, just to make sure that their carbon footprint does not increase. However, IBM does estimate that, with significantly greater energy efficiency, its average power consumption will be reduced by 42%. IBM operates 750,000 sqm of data center space on six continents, and its power savings would be approximately five billion kilowatt hours per year. If we estimate the average price of a kWh at $0.10, then it is clear that the $0.5bn saved every year will not cover the $1bn investment. However, it is still good business. A green profile will be an effective sales argument or, the other way around, without a green profile one will not sell anything at all. Moreover, the IT industry is trying to pre-empt the promulgation of new legal requirements. The fear is that, if the IT industry is not able to "go green" on its own, then legislation will be brought in to force it.

Security

Traditionally, security risks have been related to accidents and natural disasters. A data center should not be placed under the approach path to an airport; in areas where seismic activity could occur; in areas at risk of flooding etc. In recent years, however, acts of terrorism have become the prime concern. A data center should clearly not be placed near military or other facilities that could become the direct target of a terrorist attack. Industrial espionage is another risk that cannot be overstated. Kevin D. Murray, an industrial counter espionage expert for 35 years, believes that, at some point, if a company holds information that represents a value to someone, then that company will be exposed to industrial espionage. What this means in practical terms is that data centers need to put heavy physical security measures in place. This is a significant challenge, not just because of the cost it represents, but because security is so incredibly far from the core competence areas of most organisations.

"It makes an awful lot of sense that data center facilities should be provided by experts. The modern data center is not only the physical location of core data, it adresses a modern energy challenge. I expect a data center provider´s capability to offer energy efficient services to not only be a cost issue but an environmental necessity."

Prof. Dr. Tim Weitzel, Department of Information Systems and Services, University of Bamberg, Team Member eFinanceLab-Frankfurt

Public relations

The public relations challenge goes beyond raising a green profile from increased energy efficiency. Some data centers belong to very large organisations that must carefully manage their public image.

A good example of this trend is Nike. The company is working towards the stated goal that all its facilities and business travel must be climate neutral by 2011, and it has already inaugurated a distribution facility in Belgium which runs on wind power. In terms of human resources, Nike has acknowledged corporate responsibility for all the workers in its supply chain. Again it is working towards reaching a number of specific goals (e.g. abolishing excessive overtime) by 2011.

"We see corporate responsibility as a catalyst for growth an innovation, an integral part of how we van use the power of our brand, the energy and passion of our people, and the scale of our business to create meaningful change."

Nike Inc.

Decisions taken with regard to the data center could have a negative impact on an organisations image. This is especially true of decisions relating to outsourcing, because large numbers of staff are concerned. When there are redundancies, it is almost certain that there will be a negative impact on the public image, and this must be avoided or minimised at all cost.

Frost & Sullivan finds that even very well-managed organisations are struggling to cope with their "soft" challenges.

The CIO's Dilemma

3. "Make" or "buy"?

In the previous sections, we have concentrated on the "make" scenario (i.e. companies running their data centers in-house). We have shown that all data centers are faced with challenges of such a magnitude that continuing to run data centers in-house is just not an option any longer for many organisations. Consequently, they will be looking at a variety of alternatives – "buy" scenarios – to the traditional "make" scenario.

3.1. In-house benefits

With all the challenges attached to building and running in-house data centers, why are organisations still doing it? Inertia could be one explanation, but a more thorough analysis will reveal that the "make" scenario does generate a number of benefits. Control is the number one benefit of an in-house data center. By running a data center with staff on one's own payroll, in buildings one owns on the basis of development scenarios made by one self, there is control of almost every detail. Controlling every detail makes it easier for an organisation to vouch for the safety of its data center operations – both internally and externally. It gives an organisation a perception of security which translates into a benefit that could be termed the "sleep-well-at-night benefit". First of all, continuing to run a data center in house is not a controversial choice. Everything stays the same so there is no perception of risk. A CIO who suggests an alternative data center solution must go to great lengths to defend the choice internally, as it will be questioned by everybody. If something should go wrong, it is quite likely that the CIO's job would be in jeopardy, whether or not he or she had been in any way to blame. Then there is the problem of legacy. Existing staff, buildings, installations and IT equipment could be made redundant. There would need to be an action plan for the disposal of all such assets which would almost certainly run into opposition. Finally, depending on the circumstances of an organisation, keeping the data center on the balance sheet can be an advantage, especially in companies that are well capitalised but challenged on the profit side.

3.2. "Buy" alternatives

Traditionally, all data centers were proprietary – owned and operated by the organisation itself. This is no longer the case. There are several "buy" scenarios that an organisation can consider, and organisations can opt for a combination of different models. Here we shall examine three models: Outsourcing, Application Outsourcing and Colocation.

Outsourcing

Outsourcing involves the transfer of the management and day-to-day execution of an entire business function to a third-party service provider, specialising in the management of that particular function. In the simplified pyramid of the functional elements of a data center, we have illustrated how extensive outsourcing is. Outsourcing agreements are based on the transfer from the client organisation to the third-party service provider of the assets and other resources required to execute the business function. This would include human resources. In the case of a data center, the third-party service provider would normally buy (or lease) the data center buildings and the existing IT equipment, would take over the existing data center staff, and would take over all related service contracts in existence (telecommunications, power, security etc.).

Outsourcing

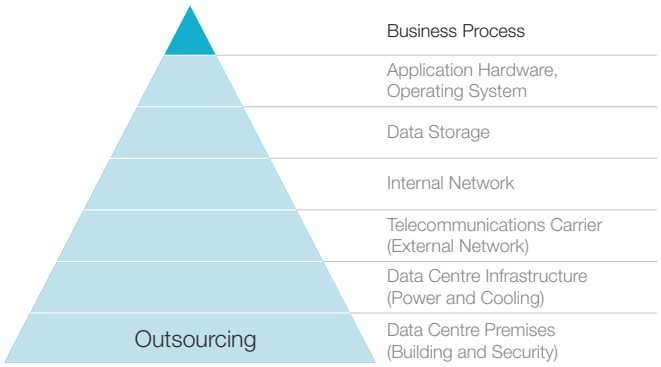


Figure 1

In principle, any business function can be outsourced, as long as that business function is not perceived as a core function – meaning essentially that the function is central to an organisation's position in the marketplace. In recent years, the trend has been for much narrower definitions of core functions to become accepted, meaning that organisations, today, are willing to outsource functions they would have kept in-house ten years ago. IT is – and has traditionally been – one of the functional areas most prone to outsourcing. EDS, Capgemini and LogicaCMG are three prominent service providers that have been successful in insourcing functions from large IT-heavy organisations such as banks.

An outsourcing agreement is governed by contracts, service level agreements (SLA) and a number of specific documents that regulate the transfer of assets. Outsourcing agreements are long-term commitments. The tendency is for outsourcing agreements to be renewed when they reach the end of the contract term.

Although the service provider takes over the assets, an outsourcing agreement rarely means that business simply carries on as normal. The whole point of outsourcing is that the specialist service provider is able to perform the business functions better than the organisation itself by introducing best practice processes and procedures, by reorganising the work and, also, by achieving

The CIO's Dilemma

economies of scale. Outsourcing agreements are fulfilled with staff dedicated to that particular client, but with more efficient practices, not all the original staff members may be needed. In such a situation, the service provider will use its expertise in enhancing the skills of certain staff in order to employ them in other parts of the service provider organisation or it will put those staff members through redundancy programmes. Still, it cannot automatically be assumed that outsourcing will lead to cost savings.

Application outsourcing (managed services)

Application outsourcing is the contracting of specific business tasks – not entire business functions, as opposed to the previous outsourcing model – to a thirdparty service provider. Usually, Application outsourcing is a cost-saving measure that can involve both back and front office activities. We shall concentrate on back office activities, as our focus needs to be activities that an organisation could outtask from its current data center. The third-party service provider will perform its allocated tasks with its own staff, there will generally not be staff dedicated to specific clients, and the details will be contractually regulated, typically backed up by a SLA. Offshore outtasking refers to the situation where the third-party service provider is located in a country far from the country of origin of its client organisation. This is mostly chosen when there is a significant cost advantage over onshore outtasking (i.e. using a thirdparty service provider in one's own country). Also, having a service provider based in a different time zone makes it easier to maintain around-the-clock operations.

Application Outsourcing

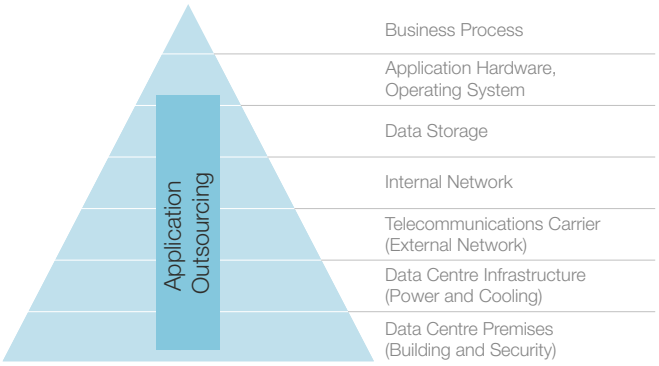


Figure 2

Many service providers already offer service portfolios that target specific challenges that an organisation's data center may have, the company Computacenter being a good example. Challenges that relate to rules and regulations are being addressed to an increasing extent by service providers, and EUDRD is an excellent example of this. Because telecom operations are struggling to achieve EUDRD compliance, many services providers are developing appropriate solutions that they are pitching to the telecom operators with success. If an organisation's data center is incapable of coping with a particular new task, then application outsourcing will be a quick fix that will enable an organisation to retain its existing data center which will then only perform the core corporate tasks. Application outsourcing does not involve a transfer of staff, nor is there any capital outlay, and the perception of risk will be low, meaning that the solution will be popular, especially with CFOs. Frost & Sullivan believes that this trend will continue with force, and that the supply of specific application outsourcing solutions will increase.

Colocation

Colocation is a "buy" model where an organisation rents pure space in a third-party data center which is external to the organisation itself. In a third-party data center, multiple organisations locate their IT and network equipment, continue to run the applications, and they can generally connect to multiple communication providers. Most third-party colocation centers offer different solutions varying from single racks (or even partial racks in some cases) to dedicated rooms, generally referred to as cages. In a colocation center, power, cooling and security will be provided by the data center provider, and many data center providers will also offer valueadded services relating to the maintenance of the IT equipment or consulting services relating to the definition and running of the equipment. Client organisations will mostly place their own equipment in the space they rent at a colocation center. The space available at a colocation center will have been fitted to the standard specifications of the colocation provider. Uptimes on power and cooling will be defined in a SLA.

There are many benefits associated with colocation. Power, service and security can all be delivered more effectively at a colocation center than at an organisation's own, much smaller data center, and the client organisation does not need to keep the relevant expertise in-house. There are typically cost savings on telecommunications services as well, because the local loop is eliminated as

Colocation

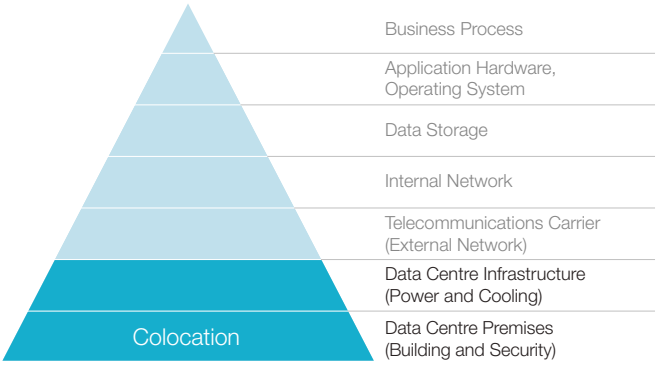


Figure 3

The CIO's Dilemma

most important operators offer connection points directly at the colocation center. It is also common that an internet exchange will be located inside a colocation center. Finally, there are financial benefits: A straight forward benefit is the lack of high upfront capital expenditure. There could be tax benefits in many countries, and an organisation will benefit from increased transparency of its IT operational expenditure.

Not surprisingly, colocation is on the rise, both in terms of the space taken up by customers and the revenue that it generates. The consulting company Broad Group suggests that the overall increase in the supply of space in colocation centers across Europe will be 10-15% per year.

3.3. The CIO's Dilemma

We have shown that the CIO has a dilemma. He or she must meet numerous difficult challenges ... and must do so fast. The first choice the CIO needs to make is whether to stick with the traditional "make" scenario or to embrace one of the many "buy" alternatives available. Unfortunately, neither alternative looks particularly attractive. The CIO will need to decide on the basis of his or her organisation's particular characteristics and strategies, and on the basis of the political climate (viz. the willingness of the CEO and CFO to support innovative solutions).

Outsourcing v. in-house

If everything goes well and the third-party service provider lives up to the SLA, then outsourcing will certainly address the challenges of business continuity, scalability, financials and security. One would also assume that the energy efficiency would be addressed, but it would be difficult for an organisation to raise its public green profile, on the basis of a green data center that it does not own or operate. The two major drawbacks, however, are control and labour relations. Outsourcing is a very drastic solution by way of which an organisation no longer has any direct control of a function as vital as the data center. Control can only be exerted in relation to contracts and SLAs (viz. via the courts, if things go wrong). What is more, labour relations are likely to suffer with public relations problems as an immediate consequence. Outsourcing is inherently unpopular with staff, and the labour laws of many countries can make the outsourcing model difficult to implement.

Application outsourcing v. in-house

Application outsourcing will address the challenges of business continuity, scalability and financials, but only with regard to the application and related tasks covered by the agreement with the third-party service provider. This is probably the main issue that a CIO would have with application outsourcing – it will not enable him or her to meet the wider data center challenges but would only deal with the immediate peak in IT needs. Application outsourcing should therefore be regarded as a quick fix or stop-gap solution.

Energy efficiency and security are two challenges that application outsourcing is not likely to address, and especially security is a concern. The organisation has very little influence over how the tasks are performed (and by whom). The organisation only sees the output, but has little visibility over the processes. Consequently, a problem is not likely to be detected by the CIO before it is too late, as service providers will not flag difficulties if they think they might still fix them.

Colocation v. in-house

Colocation will address the challenges of business continuity, financials, energy efficiency and probably also security, although there can be many differences between colocation centers that a potential client should examine. Whereas all colocation centers will have strict first level access control (to the building), access to individual racks or cages is not always as strictly controlled. It is also relevant to question the security of the building itself. Sarah D. Scarlet of CSO, in her article "19 Ways to Build Physical Security into a Data Center", suggests that a data center should be at least 30 metres from the main road, and that there should be a 30 metre buffer zone around the data center. She also advocates avoiding windows at all cost. Most colocation centers cannot offer that, especially not colocation centers that are located in the major cities. It should also be noted that not all colocation centers are equipped to offer high power densities. Colocation centers offer scalability, but only to the extent that the colocation center itself does not run out of space. Consequently, it is worth checking how much free floor space is available, and it is probably wise to avoid small facilities.

The CIO's Dilemma

4. The campus model

We have illustrated the pros and cons of the "make" scenario and three common "buy" models. Some models will work well in specific situations, but on a whole we have not found a "buy" alternative that will comfortably enable organisations to meet all their existing and future challenges effectively.

What is needed is a model that will incorporate the benefits of the "make" model with those of the colocation model.

4.1 Introduction to the campus model

In the innovative Campus Model, Frost & Sullivan has identified a strong alternative to the traditional in-house data center.

One of the pioneers behind the campus model is the German company e-shelter. Although e-shelter has traditional colocation assets in Hamburg, Frankfurt and Munich, the campus model in its purest form is offered at e-shelter's Frankfurt and Berlin data centers, indeed purpose-built as campuses. The Frankfurt campus is the largest single data center facility in Europe, at which it potentially has 60,000 sqm of floor space available. Despite its uniqueness, the e-shelter campus model is a proven concept, as it has been operating successfully since the year 2000.

The campus model differs from the colocation model in the important areas of customisation, scalability and security. Colocation involves installing racks within a shared space or buying a cage fitted to the service provider's general specifications, if one is available. The campus model is far more flexible, offering very high levels of scalability. Customers are able to rent cages or suites – or even an entire building on the campus – which are then fitted to the customer's specifications, meaning that the solution is extremely customisable. If customers are not able to define their own specifications, e-shelter will provide that service under a consulting arrangement.

Another unique feature is that the company offers business continuity office space on its campus. In a disaster recovery situation where an organisation's main site has become unserviceable, it can despatch staff to the campus, from where it can continue to work, almost as though nothing had happened. This effectively means that organisations can set up their own data center permanently that continue to be run and maintained by their own staff, just on the campus where power, cooling and security, are all provided.

4.2 Campus model benefits

Frost & Sullivan's qualitative analysis of the campus model shows that organisations will be able to realise numerous benefits which are unique and which are both "hard" and "soft".

The business continuity challenge is met owing to the high quality of the campus facilities. Power is supplied via two redundant intake points by two major power networks. Moreover, the campus has its own transformer station. e-shelter is classified as Tier 4 in relation to the tier classification system introduced by the US-based Uptime Institute, meaning that the company will guarantee a power availability of 99.991%. The e-shelter campuses are telecommunications carrierneutral, allowing free access to telecom operators so that the customers can choose from a multitude of carriers available at the campuses. The network connectivity is fully redundant. Because organisations continue to run their own data centers, they retain all the control-related benefits that enable the CIO to sleep well at night. Also, the transition from a traditional in-house solution is easy, because all that really changes is the physical location.

e-shelter's business model is really simple, as it devotes itself to the real estate elements of the value chain. This enables the company to focus on the long-term development, building and running of the data center facilities, without diverting attention to the many value-added services that are often upsold by colocation centers.

The vast space available, the customisable nature of that space and, perhaps most important, the extremely high availability of sufficient power and cooling, enable organisations to meet all the scalability and flexibility challenges. They get a future-proof data center solution that the CIO will not have to worry about.

Campus model

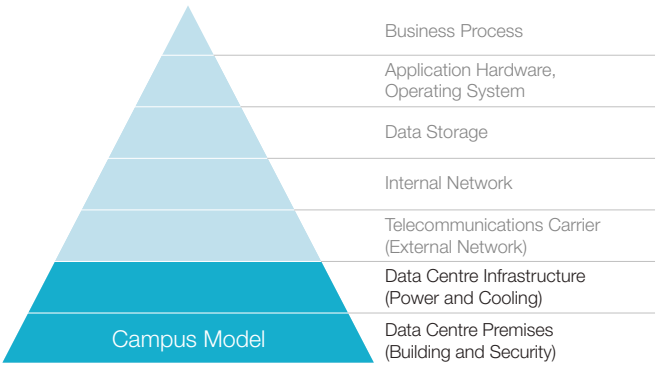


Figure 4

The CIO's Dilemma

The financial challenges are met by the campus model, because customers are able to realise the benefits of economies of scale without the fixed costs and initial capital outlay. The operational expenditure will, of course, increase but the IT spending will be transparent.

"The campus model offers the flexibility to adapt the required space to customer specific requirements in terms of LAN/WAN design, physical an IT security while leveraging existing campus infrastructure, power/cooling, facility management an real estate. Clients that do not favour managed services for their infrastructure can retain their flexibility in this model while limiting their investments for best-in-class facilities."

Dr. Michael Fritsch, Principal Telecomunication and IT, Booz & Company

With a PUE factor of 1.7, the e-shelter campus is a much more environmentally friendly facility than most data centers in the world. What is more, as illustrated by the graph to the left, a favourable PUE factor translates into significant cost savings. Typically, the cost of an in-house data center will end up being spread over a number of business units, making it difficult to understand important elements such as the real cost of power consumption and all the required manpower costs. With the campus model, the data center spending can be easily defined and captured, as there are no hidden costs or overheads. Furthermore, it can be accurately forecast and expansion hypotheses accurately ascertained. All in all, the operational and financial risks are minimised by the campus model.

Power-Usage-Effectiveness (PUE)

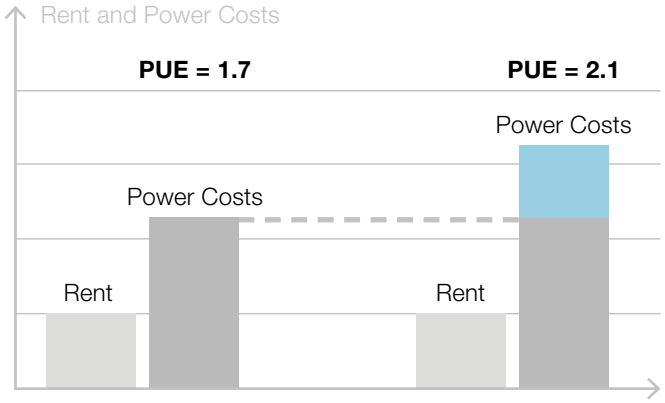


Figure 5

On the basis of the German tax regulations, KPMG has ascertained that a number of tax benefits would be associated with the campus model:

- There is no need to show an investment in the customer's balance sheet.
- The customers' profit and loss accounts will not show depreciation in respect of the data center investment provided the customer will not be the economic owner of the space.
- From 2008 onwards 25% of the interest on long term debt arising with an investment of the customer will be added back for Trade Tax purposes. Since there is no investment, no add back is to be considered for Trade Tax purposes at the level of the customer.
- From 2008 onwards the interest ceiling rules have to be considered. If certain requirements are not met, interest expenses are not fully deductible. Since the customer's payments to e-shelter are not considered interest expenses, the new interest ceiling rules will not apply insofar.

e-shelter operates its own security force, and armed guards patrol the campus constantly. There is physical separation between office and data center space, and the space where the power and cooling systems are housed is also physically separate and off limits.

Finally from the point of view of a CIO, the campus model is absolutely uncontroversial. No changes are made to an organisation's data center staff – in fact, staff will appreciate the model because it guarantees their job security and gives them a more modern, functional workplace – and the obvious green benefits discussed above will make a positive contribution to the public relations effort of the organisation. When all is said and done, maybe the greatest benefit is that the CIO will sleep well at night.

The CIO's Dilemma

5. Conclusion

Data loads have never decreased from one year to the next, and there is no reason to believe that this will ever happen. We have even found that the rate of growth of data loads is likely to accelerate. This, in itself, provides a formidable challenge to CIOs, as most data centers have no alternative but to expand.

Additional challenges are posed by factors completely outside an organisation's control. Compliance with rules and regulations such as Sarbanes-Oxley, Basel II and the European Union Data Retention Directive is and has been a struggle for many organisations. There is every reason to believe that the future will bring even more rules and regulations and that legislation could be introduced to curb the power consumption of data centers.

Different business models demand different data center solutions, but generally the most modern, innovative business models are the ones that require the most sophisticated data management capabilities. In order to not be stuck with old fashioned business models without the ability to innovate, organisations really have no choice: They must look for ways to address their data center challenges, and they must do so now.

We have identified a number of alternatives open to organisations, including expanding an existing in-house data center. Because the traditional alternatives will not sufficiently address all an organisation's data center challenges, a new model – the Campus Model – has emerged.

On balance, it is Frost & Sullivan's opinion that the Campus Model provides all the advantages of a traditional in-house data center, all the while avoiding the disadvantages. The Campus Model is not a short-term solution, as it will enable most organisations to address all their data center challenges within a foreseeable future. Consequently, we believe the Campus Model to be a viable response to the CIO's dilemma.

About Frost & Sullivan

Frost & Sullivan, a global growth consulting company, has been partnering with clients to support the development of innovative strategies for more than forty years. The company's industry expertise integrates growth consulting, growth partnership services and corporate management training to identify and develop opportunities. Frost & Sullivan serves an extensive clientele that includes Global 1000 companies, emerging companies and the investment community by providing comprehensive industry coverage that reflects a unique global perspective and combines ongoing analysis of markets, technologies, econometrics and demographics.

For more information visit http://www.frost.com

About e-shelter

High-availability data centers are the backbone of the digital economy and the Internet. Since 2000 e-shelter designs, builds and operates high-availability data centers, whose infrastructure guarantees the highest standards of physical security and operational reliability. e-shelter operates approximately 90,000 sqm of data center space, on eight sites. 60,000 sqm space alone are located at e-shelter's main site in Frankfurt, Germany, making it Europe's largest single data center site. Additional sites are located in Berlin, Frankfurt, Hamburg, Munich as well as in Zurich, and Vienna. Among e-shelter's clients are fi nancial services companies, telecoms operators and IT service providers as well as cloud service providers. As a part of NTT Communications Corporation e-shelter provides access to a network of 140 data centers around the world.

Our own security personnel of e-shelter security guarantees the security of our data centers and develops safety concepts according individual requirements.

With our extensive experience in operations of data centers we are a sought-after contact especially when it's about individual solutions for complex projects and high power density. Because of the size of our data centers we offer the necessary flexibility for hybrid IT solutions and direct access to cloud providers in particular.

www.e-shelter.com