Planning and Designing Data Centres

ABSTRACT

There are many schools of thought on data centre design and implementation. Each has its own benefits and challenges. Organisations will have to consider the specific business needs to come up with their requirements and designs for their data centres. Usually, these data centres are based on unique design concepts, taking into account industry best practices and infrastructure constraints. This article shares some of the design considerations and implementation approaches for data centres.

Yee Keen Seng
Wu Xianghua
Lim Hwee Kwang
Wong Marn Chee
Ang Choon Keat

Planning and Designing Data Centres

INTRODUCTION

Data centres house two of an organisation's most valuable assets – data and IT systems. Well-designed data centres based on a robust strategy will enhance the availability and reliability of the IT services to the end users. This allows organisations to respond efficiently and effectively to business demands and challenges, which is a critical success factor for organisations.

This article discusses some of the basic but key considerations for planning and designing data centres, which are dependent on the organisation's business needs, tolerance for failures and availability of budget. However, certain aspects of the data centre such as total space and structural loading cannot be changed without major redevelopment of the data centre and disruptions to day-to-day operations once it has been built. Thus, it is critical that these aspects are given due consideration during the planning phase.

NEED FOR DATA CENTRE STRATEGY

Some organisations use commercial data centres to host their IT systems while others prefer to build their own data centres for the purpose of security and control, among other reasons. As part of the organisation's overall IT masterplanning, a data centre strategy must be developed, including the decision on the number of data centres and their locations.

For operational efficiency and cost effectiveness, it seems attractive to consolidate IT systems into a single data centre. When space, power and cooling capacities are optimised, developmental and operational costs of the data centre are reduced. Centralising the IT systems also minimises the human resources used for managing the IT systems. When operations are conducted in close proximity to one another, processes can also be streamlined more effectively. However, if the single data centre fails, the entire organisation may be brought to a halt.

Distributing the IT systems over multiple data centres at different sites provides the required resiliency to the organisation. The organisation will not lose all of its operational capabilities should one of its data centres fail. Certain business units or functions may not be available but the rest of the organisation can continue with their daily operations. However, multiple data centres require IT infrastructure to be set up at multiple sites. More manpower may be required to conduct and coordinate operations at the various sites which would increase the developmental and operational costs.

Table 1 illustrates the various considerations in deciding between building a single data centre and four smaller data centres.

Components	Single Data Centre	Four Data Centres
Building Construction	Assumed similar cost	Assumed similar cost
Mechanical and Electrical (M&E) Systems Infrastructure	One Set	Four Sets
IT Common Services Infrastructure (Transmission/Network, Security, Storage, etc)	One Set	Four Sets
Usable Space for Business IT Systems (Assuming IT Infrastructure Common Services take up Y rack space)	X – Y	X – 4Y

Table 1. A comparison of infrastructure requirements

It can be seen that the option of building four data centres requires an addition of three sets of M&E systems infrastructure and IT common services infrastructure. Space will also be taken up by the IT common services infrastructure built at each data centre, effectively reducing the total usable space for the business IT systems of the organisation. With four data centres, the M&E systems and IT common services infrastructure at each data centre may be smaller in capacity, but the total cost of ownership will be higher as compared to having a single data centre.

To balance resiliency and cost, most organisations will have two to four data centres to support their business requirements, and a reliable disaster recovery plan for these data centres.

SPACE CONSIDERATIONS

The organisation has to determine the amount of space required for each data centre. Undersizing the data centre may result in it running out of space quickly and new data centres may need to be built frequently. This may go against the organisation's data centre strategy and incur additional costs. On the other hand, oversizing may lead to wastage of resources and space that could be put to better use.

It is relatively easy for an organisation to know what it needs now and probably for the next three years. However, it is very difficult to forecast the requirements for the next 10-15 years given the rapid changes in technology and business.

One possible way of estimating the space requirement is to look at the historical IT growth of the organisation and assume the same rate of growth for the next 10-15 years. However, this approach has a potential pitfall – the focus and initiatives undertaken by the organisation in previous years may differ significantly from those of the next few years. In that case, there may be undersizing or oversizing of the data centre space requirements.

Besides analysing the historical IT rate of growth, it is useful to review the initiatives that the organisation has carried out over the last few years and study its envisioned state of IT maturity. This will help to determine the future initiatives of the organisation, normalise the spikes and troughs of the IT requirements and give a better projection of the space required for the data centres.

For example, if the organisation is planning to implement projects such as knowledge management or video content management, the data centre space will increase significantly due to large storage requirements of these projects. On the other hand, if the organisation is embarking on an IT consolidation or virtualisation path, there may be a reduction in space requirements for IT infrastructure over time.

ARCHITECTURAL CONSIDERATIONS

The practice of architecture is primarily concerned with the principles of durability, functionality and human comfort. While the first two principles can be quantitatively satisfied, the solution to the third is not as straightforward. Designing a building for servers presents a unique challenge as it may not be intuitive to the architect, who designs spaces for humans instead of machines.

In a data centre, optimal working environment of computer servers takes precedence over human comfort. Thus, M&E requirements drive the design of the data centre and its organisational principles. However, the needs of operators should not be neglected. The architectural design serves to complement the layout of the data centre in ensuring that circulation patterns are logical and conform to building codes and regulations. For example, circulation spaces can be designed to allow natural lighting – this can relieve operators of the stress from working in a hermetic environment.

It may be useful to separate administrative areas from data centre rooms as they have vastly different spatial requirements. While the data centre rooms require a hermetic and closed environment with large slab-to-slab heights, the administrative areas should be open to natural lighting without large vertical spaces. One possible way of resolving the different requirements is to consolidate all the administrative areas on a separate floor or in a wing. This strategy will also aid in compartmentalising the building according to use, security requirements and susceptibility to fire hazards.

STRUCTURAL LOADING CONSIDERATIONS

The required loading on the structure of the data centre is an important consideration for the design of the data centre building. This is normally specified in terms of distributed floor loading and expressed in kilo Newtons per square metre (kN/sqm). The building structure must be adequate to support the weight of the equipment required in the data centre.

Structural loading considerations are critical for the safety of the entire building where the data centre resides. Overloading the building structure may eventually render the building unsafe for occupancy. In the worst case scenario, it may even cause the building to collapse. Hence, it is important to consult a qualified structural engineer before placing an exceptionally heavy piece of equipment in the data centre during the operational phase.

There are several methods to manage the scenario where the weight of equipment is heavier than the structural loading. One way is to load each rack with less equipment and increase the number of racks used. An alternative is to use spreaders – usually a metal plate strong enough to hold the rack and equipment – to distribute the weight over a wider area. These two methods will reduce the usable space for IT in the data centre. A qualified structural engineer will also need to be consulted before any modification is made to the building structure.

Today, a rack fully occupied with equipment such as blade servers and storage area networks may weigh 600-1,000kg. The structure must be adequate for supporting a loading of approximately 10-13kN/sqm. This is in contrast to the designed loading of 5kN/sqm adopted in many buildings. In the future, IT equipment is expected to be heavier due to the incorporation of more electronic components. This means that future data centres should have a structural loading higher than 13kN/sqm. New data centres built today are generally designed to support 15-25kN/sqm in structural loading.

POWER AND COOLING CONSIDERATIONS

M&E systems can easily make up more than 60% of the total developmental cost of a new data centre and is thus a major cost component. M&E systems include electrical substations, chillers, backup generators, uninterrupted power supplies (UPS) and computer room air-conditioning (CRAC) units. Careful consideration of the power and cooling requirements is required to implement M&E systems of an appropriate size to avoid wastage of resources.

The power and cooling capacities to cater for a data centre depend on the type of IT equipment to be deployed. A single rack can pack more blade servers providing more computing resources per rack than rack-mount servers. However, blade servers generally require higher power and cooling capacities than the rack-mount servers. More space may be required for cooling which in turn reduces the advantage of space efficiency when adopting blade servers.

For instance, a standard IT rack can pack about 32 full-height blade servers or 64 half-height blade servers, which require power and cooling capacities of approximately 20-25 Kilowatts (kW) per rack. A similar rack can pack about 20 rack-mount servers, requiring power and cooling capacities of approximately 10kW per rack. Thus, there is a two or three-fold increase in power and cooling requirements for racks of blade servers as compared to racks of rack-mount servers.

To determine the right size of the power and cooling requirements for the data centre, it is important for the organisation to set a direction for the equipping of IT infrastructure. The organisation can chart the power and cooling requirements in phases over the life cycle of the data centre. However, when the data centre is first built, space needs to be provisioned for the amount of M&E systems to support the eventual power and cooling requirements.

Tier Requirement	Tier 1	Tier 2	Tier 3	Tier 4
Power Supply to Data Centre	1	1	2 (One Primary, One Backup)	2 (Both Primary)
M&E Systems Supporting IT Equipment (e.g. Generators, Chillers, UPS, CRAC Units)	N	N+1	N+1	2(N+1)
Downtime Required during Maintenance	Yes	Yes	No	No
Data Centre Availability	99.67%	99.74%	99.98%	99.99%
Data Centre Downtime Per Year	~28.8hrs	~22.0hrs	~1.6hrs	~0.4hrs
Relative Cost*	Low	Medium (~22 Tier 1s)	High (~2 Tier 2s)	Very High (~2 Tier 3s)

* Relative Cost is a general industry assessment and not part of the TIA-942 Standards.

Table 2. Data centre availability tiers by the Uptime Institute

REDUNDANCY CONSIDERATIONS

Although not mandatory, most of the data centres today adopt the TIA-942 Telecommunications Infrastructure Standards for Data Centre in their designs. Four data centre availability tiers, defined by the Uptime Institute, are incorporated into the TIA-942 guidelines which specify the redundancy and tolerable unplanned downtime for a data centre. Table 2 provides a summary of the four data centre availability tiers.

Table 2 shows the various configurations for the design of M&E systems e.g. *N+1*. '*N*' refers to the minimum number of M&E equipment required to support all the operations of a data centre and '+ *X*' means that there are *X* spare systems to back up the main systems. It is very expensive to set up a *2(N+1)* configuration at Tier 4. M&E systems contribute significantly to the cost of a data centre. Unless there are very critical operations to support, most organisations will settle for a lower data centre requirement or just the *N+1* configuration.

The *N+1* configuration means that maintenance of the M&E systems will be carried out one at a time. The single spare system will fill the gap while the maintenance is ongoing. This is feasible if the data centre is small. However, if the data centre is large with numerous units of M&E equipment, the entire maintenance process may spread over days, weeks or even months since only one system can undergo maintenance at a time.

Most importantly, the *N+1* configuration means that the data centre cannot afford to fail during the maintenance of the M&E systems. During maintenance, when one of the M&E systems shuts down, the spare M&E system will be activated to take on the load of the inactive system. If the spare M&E system fails at this point in time, there will not be another M&E system available to activate and support the load from the IT systems. Hence, some IT systems will be affected.

To overcome these potential shortcomings, many organisations design their data centres with *N+2* or *N+3* configurations. In case a failure occurs during maintenance, the second or third spare system can be activated to support the power and cooling loads.

GREEN DATA CENTRE CONSIDERATIONS

There is an increasing focus for data centres to be green. A green approach to design ensures that a building is environmentally friendly and resource-efficient throughout its life cycle from construction to demolition. Going green brings tangible savings to the organisation not only through reduced power consumption, but also in the form of greater durability and economical usage of natural resources. In addition, a green environment also benefits the occupants of the building. For example, maximising natural lighting in administrative office areas or using low volatile organic compound paints can lead to a more conducive working environment and increased productivity.

Energy efficiency in server rooms can often be achieved with simple and inexpensive measures. For example, installing blanking panels in racks and arranging the racks into hot and cold aisles will improve the airflow dynamics and raise the cooling efficiency. In administrative areas, natural lighting can be used to lessen lighting loads.

It is important to determine the right size of M&E systems to ensure requirements are met adequately. In anticipation of future needs, some organisations tend to oversize data centres while equipping M&E systems. Oversizing results in higher developmental costs when purchasing more and bigger M&E equipment, as well as higher operational and maintenance costs. During operations, higher energy consumption will be required to power the oversized M&E systems performing at an inefficient level. The higher energy consumption to support an oversized M&E system is a waste of energy. The organisation should adopt a modular approach and equip the M&E systems when required.

With advancement of technologies and the manufacturers' increasing focus on being environmentally friendly, most M&E systems and IT systems are designed for energy efficiency. For example, linking temperature sensors to smart CRAC units to vary the cooling capacity of the data centre can lead to substantial savings in energy consumption. Rainwater can be harvested to cool plant systems so as to conserve water. These technologies may incur slightly higher data centre developmental costs but great savings will be reaped in the life cycle operational costs. Thus, they should be deployed for the data centre as far as possible.

The standards for green data centres are evolving, spearheaded by organisations like the Green Grid at the international level as well as the Building and Construction Authority (BCA), Infocomm Development Authority of Singapore (IDA) and SPRING Singapore on the local front. SPRING Singapore has published the "Singapore Standard for Green Data Centre – Energy and Environment Management Systems (SS 564 : 2010)" in January 2011, enabling an organisation to plan, build, operate and maintain a green data centre, by establishing the systems and processes to manage and improve the data centre's energy performance. These standards should be considered for adoption if they do not compromise security or cardinal business requirements.

SECURITY AND OPERATIONAL CONSIDERATIONS

A well-designed data centre will consider not only the availability and reliability of the data centre facilities, but also security and operational requirements. Crucial organisational information is housed in the data centre. Thus, it is imperative that the data centre is kept secure to prevent data leakage and sabotage.

Multi-layer security implementation is recommended. All personnel at the data centre including staff and visitors must be screened and cleared by the security department. Access to the data centre should be restricted such that only authorised personnel are allowed to enter the areas that they are required to work in or visit. All equipment in the data centre should also have security locks or equivalent safety measures to control physical access to them. Activities in critical operational areas of the data centre should also be monitored and recorded.

In addition to these basic security measures, data centres should also consider implementing more advanced technologies, such as active radio frequency identification and sensors, to track the actual location of people in the data centre. It should come with an alarm system to detect people who have entered areas without authorisation. Some consultants may recommend triple authentication access measures using cards, passwords and biometrics. However, the actual additional security achieved should be measured in the context of multi-layer security architecture. The cost incurred in implementation, security administration and operations should be balanced with the additional security achieved.

Operational workflows have to be formulated around security policies and implementation. The internal data centre has to be designed to meet these operational workflows. Essentially, form should fit function. For example, areas outside servers should be created as service corridors to house CRAC units. Thus, CRAC service personnel do not need to have physical access to the servers during maintenance of the CRAC units. Printers and backup devices should be housed in separate rooms from the server areas. Such measures enhance operational workflows as the personnel servicing different subsystems do not need to share physical space and obstruct one another.

Security measures and operational efficiency may not always be aligned. Depending on how crucial the data centre is to the organisation, different levels of security implementation may be needed. In all cases, the data centre design should give due consideration to the operations of the data centre. Where there are conflicts, security usually takes precedence over operational efficiency.

VIRTUALISATION

In addition to data centre design considerations, IT strategies and implementation plans also play an important role in enhancing the efficiency of a data centre in terms of space, power and cooling utilisation.

Most servers are operating at a utilisation rate of only approximately 15-20% when running a single business application. Most servers also use approximately 40-50% of their power capacities when idle. An organisation can address this inefficient use of IT resources using virtualisation. With virtualisation, multiple business applications can run on a single physical server. This increases utilisation and power efficiency of servers. At the same time, it reduces the number of servers deployed, leading to fewer idle servers expending data centre power and cooling capacities, as well as better utilisation of space in the data centre.

PUTTING IT TOGETHER

Organisations can consider starting with two data centres. It is not advisable to have just one data centre as its failure will affect the survivability of the organisation. Cardinal IT systems that house critical business applications and data can be split between two data centres. This will maintain at least 50% of the operations if one of the data centres fails. A modular build-up approach can be used to size data centres.

M&E systems should be built and equipped based on anticipated requirements for the next five years. Sufficient space has to be catered for future expansion of the data centre. Assuming that future technologies have a smaller requirement for space as compared to current technologies, it is likely that future equipment can fit into the space reserved for expansion. This reserved space can be used as temporary offices or storage rooms until the need for data centre expansion arises.

The challenge remains in the accurate estimation of power and cooling capacities. The increasing emphasis on being environmentally friendly may lead to the emergence of future technologies with lower power and cooling requirements. However, it is more likely that smaller footprints would lead to more equipment packed into the data centre, leading to increased demand for power and cooling capacities. Fortunately, not every piece of equipment has high power and cooling requirements e.g. network equipment. Based on industry trends, many organisations are designing their data centres for an average power utilisation rate of 10kW per rack for current requirements and possibly provisioning for 20kW per rack for future requirements.

CONCLUSION

Given the large investments in data centres, organisations should plan their data centres properly, consider their available budget and tolerance for failures and disruptions. The organisation must anticipate its immediate and long-term business needs and align them with its IT strategy.

REFERENCES

Telecommunications Industry Association. 2010. TIA-942 Telecommunications Infrastructure Standards for Data Centres.

BIOGRAPHY

Yee Keen Seng is a Senior Engineer (Infocomm Infrastructure). He is currently involved in planning, designing and developing the first purpose-built data centre for the IT systems of the Ministry of Defence (MINDEF) and the Singapore Armed Forces (SAF). He oversaw several best-sourcing projects including the provision of shared services and end-user IT support to MINDEF and the SAF. Keen Seng served in the SAF Chief Information Officer Office where he managed the development and governance of the SAF Enterprise Architecture Framework as well as pioneered the Ops-Admin Systems Integration initiative. Keen Seng is certified by Enterprise Products Integration (EPI) as a Data Centre Professional and a Data Centre Specialist. He holds a Bachelor of Science (Information Systems and Computer Science) degree from the National University of Singapore (NUS).

Wu Xianghua is an Architect (Building and Infrastructure). He is working on the future expansion of the SAFTI Military Institute. Xianghua received his Bachelor of Architecture degree with First Class Honours, with a minor in Architectural History from Carnegie Mellon University. He also obtained numerous awards for his design work and for being the top graduate in the architecture programme.

Lim Hwee Kwang is an Assistant Director (Infocomm Infrastructure). Working with MINDEF and the SAF, he strategises and develops the masterplan for key Infocomm infrastructure ranging from data centres to messaging infrastructure. He leads, directs and manages the development, implementation and engineering support of information infrastructure. Hwee Kwang co-developed the Civil Service-wide Secure Messaging System with the Infocomm Development Authority of Singapore. He also designed and implemented enterprise-wide Public Key infrastructure and authentication solutions. Hwee Kwang holds a Master of Science (Information Security) degree from Royal Holloway, UK and a Master of Science (Management of Technology) degree from NUS. He further attained the Chief Information Officer Certificate as a Top Distinguished Graduate in the National Defense University, US in 2007.

Wong Marn Chee is a Senior Engineer (Infocomm Infrastructure) and the Head of Defence Technology Tower Computer Centre. He oversees the development, engineering support and IT operations of the computer centre. Marn Chee plays a key role in the development of Enterprise Server Farms for MINDEF and the SAF. He has many years of experience in designing, architecting, implementing and managing the engineering support for the server farms and their common network services. Marn Chee received his Bachelor (Electrical Engineering) degree with Honours from Nanyang Technological University in 1995.

Ang Choon Keat is a Senior Engineer (Building and Infrastructure) and works on planning, designing and managing protective building and infrastructure developments. He is a registered Professional Engineer (Civil) in Singapore. Choon Keat was previously Manager (Plans) at the Defence Research and Technology Office where he was involved in the management of the MINDEF Research and Technology portfolio. He graduated with a Master of Engineering (Civil Engineering) degree from Imperial College London, UK in 1999 under the DSTA Undergraduate Scholarship. He further obtained a Master of Science (Operations Research) degree from Columbia University, US in 2005 under the DSTA Postgraduate Scholarship. Choon Keat is currently pursuing a Doctor of Philosophy (Management of Technology) degree from NUS.