

THE OFFICIAL JOURNAL OF AIRAH

OCTOBER 2019 · VOLUME 18.9

RRP \$14.95

PRINT POST APPROVAL  
NUMBER PP352532/00001

# Ecolibrium

**Rack 'em up**  
**Keeping data centres cool.**





# The data rush

Australia's data centre environment is a rapidly evolving one, as worldwide players set up shop to satisfy our appetite for technology. Sean McGowan talks through some of the issues with

**Sean Treweek**, F.AIRAH, chief executive officer of Meinhardt Australia;

**Ross McCarthy**, lead mechanical engineer at Aurecon;

**Mark Toner**, manager of data centre and telecommunications facilities at A.G. Coombs;

**Craig Gianoli**, associate director mission critical at NDY;

and **Marco Hopman**, M.AIRAH, technical director mechanical services ANZ at Jacobs.





Sean Treweek



Ross McCarthy



Mark Toner



Craig Gianoli



Marco Hopman

**Eco:** How does data centre design here in Australia compare with the international experience, and what impact does climate have on these differences?

**Toner:** We compare well in Australia with our global counterparts when it comes to data centre design and construction processes.

Our customers' dependence on their critical ICT systems is universal. They demand an always-on, secure, interconnected data centre facility, that can scale as their business requires. We in turn provide them with designs that incorporate global best practice and are certifiable to global and other local standards such as the Uptime Institute, NABERS, ISO, LEED, etc.

Recent hyperscale facilities designed and built here highlight that Australia has the capability to design, build and operate these massive facilities, to construction programs, and to a standard that is equivalent or better than what we see overseas. This is evidenced by many of the global Tier I cloud providers who have entrusted their local cloud footprint to Australian data centres.

**Hopman:** The design solutions for data centres worldwide are heavily influenced by the exterior environment – temperature, humidity, and air quality – and the internal IT environmental temperature and humidity ranges that must be maintained. In addition to reliability and security requirements, risks tolerance concerning airborne contamination, and operator preferences, data centre mechanical designs in Australia and internationally are influenced by the data centre operators' requirements with respect to the internal conditions and how the local climatic conditions relate to that criteria.

Climatic conditions in Australia vary greatly between the west, east and northern regions, which are key focus of data centre growth. Airside economisation – or free outside air cooling – and indirect free cooling, for example, are tried and tested practices in cooler climates in Australia that help to reduce a facility's PUE (power usage effectiveness), which is critical for data centre investments.

In the Australian context, water supply security and the potential of drought period water use restrictions must also be considered for data centre site selection.

Evaporative cooling and heat-rejection systems, whether cooling towers or indirect evaporative cooling systems, are significant water consumers, and therefore a reliable local water supply system is required to ensure system reliability.

**McCarthy:** Australian data centre design quality is on par internationally, achieving similar uncompromising levels of performance. At the same time, research has shown almost half – 48 per cent – of data centre outages are caused by human error, not equipment failure.

We naturally think about the technical aspects of data centres but less about the place of people within these spaces. Locally we are having more discussions around getting the best performance from people through more human-centred design to reduce human error.

Cold climates may help reduce energy consumption but extreme low temperatures can pose tougher design challenges than warmer climates. The majority of Australian data centre hubs need to accommodate extreme highs up to 46°C, but extreme lows are not typical, making use of free cooling technologies more prevalent in those environments.

**Treweek:** Over the past couple of decades, Australian data centre design has compared well; however the past five years or so has seen the rise of enterprise hyperscale data centres for large end-users such as Google, Microsoft, Apple, Amazon etc., in North America, Europe and Asia. This is a function of their business model and large footprint in those regions, compared to a much smaller-scale footprint in Australia. With cooling being one of the primary concerns in a data centre, climate plays an important role in system selection. Cool climates with the ability to use outdoor air as a mechanism for "free-cooling" will start to define the mechanical system type to achieve a certain annualised PUE.

**Gianoli:** Broadly speaking the data centre internal environmental requirements are largely consistent internationally across the colocation sector. The large corporate multinational providers are starting to broaden the allowable internal conditions in the pursuit of improved energy efficiency.

The effects of climate change have seen increased pressure on peak ambient

design conditions. We've been required to provide supplementary cooling to key electrical plant to preserve system rating.

Within Australia most facilities are two levels, with multistorey data centres being relatively uncommon. This is contrasted with our experience in Europe and Asia where multistorey facilities are more prevalent due to real estate costs or other spatial constraints.

**Eco:** Are our data needs different to those of other regions, so therefore scale and configuration is different also?

**Gianoli:** Data needs are closely related to the population density and network interconnectivity. In Australia our population is relatively spread out, with Sydney being the unofficial hub of data centre activity in Australia.

**McCarthy:** Our data needs do differ from other locations. In some less developed countries, mobile devices account for higher data needs compared to Australia. Wholesale and retail – hyperscale and colocation – facilities are shaping more of local needs. Australia is leading in the roll-out of 5G networks, which can maximise peer-to-peer data transfer, but still require data storage in facilities. We see 5G driving needs for edge computing and small facilities.

**Treweek:** I would argue that our data needs are similar in terms of uptime, reliability, bandwidth etc., but the total volume of data is much smaller in Australia given our much smaller population and distances between major centres of population. As such, the scale of Australian data centres will generally be smaller, but the configuration in terms of system types, redundancy, Tier rating and the like is not dramatically different.

**Toner:** Ultimately, we don't see our data needs to be different to those in other developed countries. We all share the same need for data integrity, availability and privacy, but not all cloud service providers globally have the right processes or maturity to ensure these goals are met.

**Hopman:** Access to data has become critical and we have seen hyperscale developments worldwide along with smaller enterprise and edge compute facilities. There is a drive to reduce CAPEX and OPEX of data centre facilities while increasing flexibility and efficiency. One consistent theme we see from clients

is a request for scalability and modularity in the design. The ability to defer expenditures until the business needs are realised can be critical.

The large hyperscale data centres of over 100MW are driven by the need to service a large population from a hub city, like Singapore or Tokyo. We are starting to see this type of data centre size requirement in the Australian market.

**Eco:** How do large-scale colocation data centres differ from those enterprise data centres owned and operated by their user, such as a bank for example?

**Hopman:** It mostly comes down to the acceptability of risk and security.

Both users take the subjects very seriously, but a bank is more likely to drive toward the highest levels of reliability, including the universal application of 2N utility infrastructure for critical IT applications. The large-scale commercial co-location data centres typically use a N+X infrastructure, most common being N+1 and N+2. Other non-financial data centre users may use a commercial facility for some applications while keeping critical applications in their enterprise data centre. We have seen this approach executed by private business and governmental entities.

**Gianoli:** Corporate data centres have the luxury of working closely with the IT stakeholders to define and agree on conditioned space, redundancy and performance requirements to suit their business needs. Colocation providers, on the other hand, must meet the most onerous subset of from the Service Level Agreements across their portfolio over a wide range of energy densities. In theory the homogenous environmental requirements for corporate data centres should result in improved design optimisation, however, this is not always the case.

**McCarthy:** Enterprise data centres are typically smaller than colocation facilities. In some cases a duplicate (or disaster recovery) facility is constructed in a different location to mirror business needs. If one of the facilities is offline, the recovery facility can take over. The cost for an unplanned outage can be severe, particularly for a financial institution.

Large-scale colocation facilities typically lease space or cabinets – this can range from one to over 50 tenants in any



facility or a data hall. The individual tenants leasing space can be anywhere from a small to medium business, a large institution or a major cloud provider. In either facility, recognised Uptime Institute, TIA 942 or equivalent standards may be required by the tenant.

**Treweek:** Enterprise data centres are generally designed and built for very specific requirements of the end user, whether it is a Big Four bank, a government department or any other type of enterprise. A colocation data centre will generally be designed and certified in accordance with a specific Uptime Institute Tier classification, as these centres need to be benchmarked against industry recognised standards to be able to compete for the hosting dollar on the open market.

**Toner:** Large-scale commercial data centres have many differing attributes to Owner Operator facilities including dedicated, purpose built facilities in industrial parks, not typically in mixed use buildings or CBD locations.

They have bigger capacity data halls (commercial facilities are now being built with multi-MW capacity per data hall) and they typically provide for power and cooling profile, where one size fits all (average kW per rack). Owner operator facilities can typically offer greater flexibility for bespoke ICT solutions.

Other differences include better environmental credentials using latest innovations, greater on-premise cloud





A Tier IV data centre.  
Each Tier incorporates all the  
requirements of the lower ones.

interconnection solutions, greater telecommunication carrier options, and scalable designs. They are also typically Tier III or IV formerly certified. Larger power and cooling plant building blocks also provide scale economies.

**Eco:** The tier rating of a data centre is becoming an industry identifier. Is a different approach taken to data centre design and operation for each of these classifications?

**Gianoli:** Absolutely. Typically dedicated data centres are designed to an equivalent of Tier III whether they seek the formal accreditation or not. Moving up to a Tier IV design requires additional planning of distribution pathways and consideration of the autonomous response to a fault. Earlier designs to meet the Tier IV requirement typically consisted of 2N mechanical and electrical systems but modern designs are this achieving this with as few as N+1 capacity elements electrically and reduced mechanical redundancy delivering a far more cost-effective build out.

Throughout Australia there are five sites with an Uptime Accredited Tier IV design, with only three of these obtaining an As Constructed rating.

**Toner:** The designs for each of the Tiers are definitely different as we are targeting different business availability requirements for the IT systems that are housed in these data centre facilities.

For example, a data centre hosting research high performance computing infrastructure might be designed to Tier I, having minimal to no redundancy in critical power and cooling systems. In the event of a data centre outage, computing would stop. It may cause inconvenience for the research team, but not result in significant loss of revenue, brand damage or loss of life. Whereas a data centre hosting a bank's EFTPOS system is more likely to have Tier III or Tier IV credentials as an outage could have significant ramifications to not only the bank's business but also to those of the bank's customers.

The higher the tier the greater the level of redundancy in the data centre infrastructure. And in the higher tiers, no single points of failure.

**McCarthy:** Data centre designs following the Tier III or Tier IV Uptime Institute or TIA 942 framework need to follow a set of guidelines to allow for certification.

Any electrical or mechanical component that forms part of the facility's infrastructure can fail, so if you build for concurrent maintainability you can prevent failures by providing easy access to ongoing maintenance procedures.

Regardless of the Tier rating, experienced professionals look for single points of failure (SPOFs).

**Treweek:** Each tier will have a vastly different design in order to achieve the required redundancy level. The higher the Tier level the more complexity needs to be introduced into the design and the physical quantum of equipment increases. A Tier IV centre could easily have triple the volume of equipment compared to a Tier I and much more complex interconnectivity in order to achieve a fault-tolerant redundancy level.

Operationally, testing regimes become much more onerous for a higher tier centre.

**Hopman:** We see many data centres that are designed only with redundant production components – pumps, air handlers etc. – which would be similar to an Uptime Tier II; and concurrently maintainable infrastructure, which is similar to an Uptime Tier III. The fault tolerant, which is similar to Uptime Tier IV, is not a common request. As far as the design approach there is a significant increase in effort as one moves up the level of reliability.

At a high level, the approach to a redundant component facility (Tier II) requires the designer to evaluate system operation with the number of components equal to the redundancy being non-operational, that is one unit off in an N+1 system, two units off in a N+2 system, etc. Mechanically this analysis is probably most critical in the air-management system, as the air must still be delivered in an adequate quantity and quality to the IT equipment.

For a concurrently maintainable facility

(Tier III), again at a high level, the same system performance analysis as the redundant component facility must be performed; however, it requires the designer to evaluate for concurrent maintainability. That is every component – e.g., pump, pipe, air handler – must be able to be safely removed from the system without impacting the operation of the IT equipment.

The fault-tolerant facility (Tier IV) goes through the same analysis as the previous levels but adds to it the requirement for compartmentalisation and the ability to automatically detect and isolate a system failure without impacting the operation of IT equipment.

**Eco:** What are the ramifications for best-practice design where there are collocated facilities, with multiple Uptime Institute Tier rating areas?

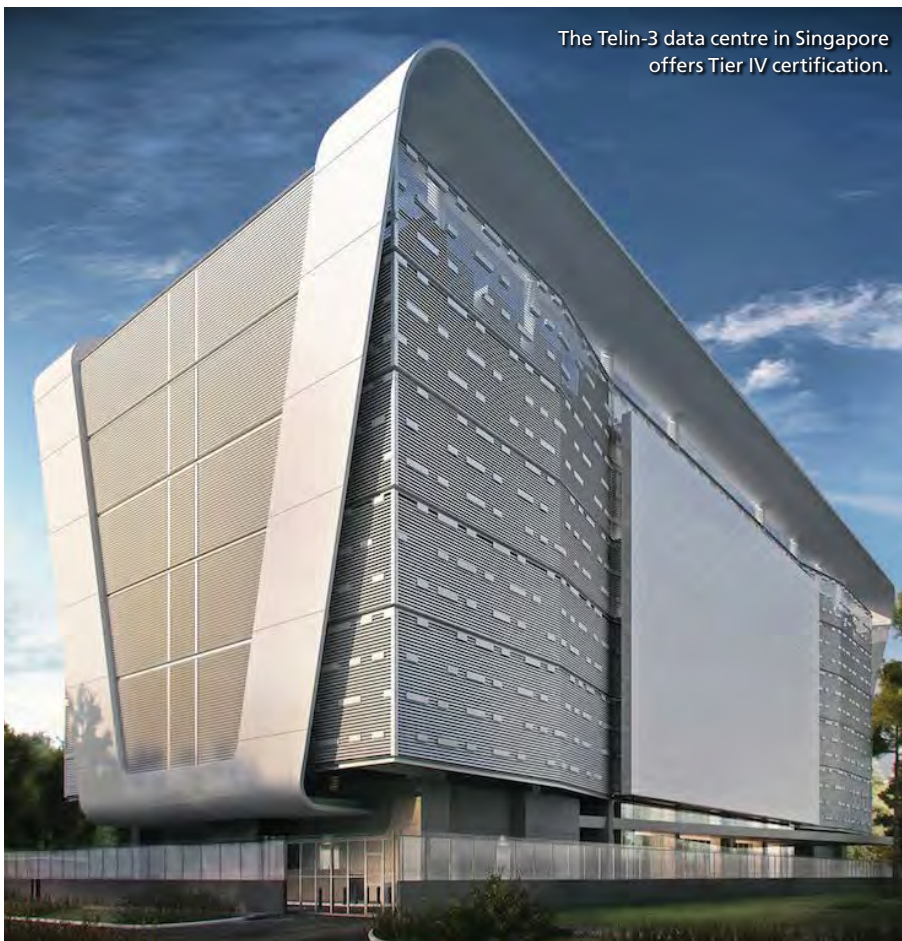
**Treweek:** Sites with multiple tier rating areas can add an additional level of complexity into the design. For example, a site with a Tier III area and a Tier IV area will generally need its main site infrastructure – i.e., incoming power, communications, water etc.) designed to the Tier IV standard, so the Tier III space will also benefit from this increased reliability. The difficulty then becomes the cost competitiveness of the lower tier space compared to a stand-alone centre for that tier.

**McCarthy:** Co-located facilities need to provide for multiple tenants. These are complex environments, even though we strive for simple designs.

Different tier ratings impact the fundamental fault domains for equipment, therefore increasing complexity that may pose challenges for certifying a facility to Uptime Institute.

**Gianoli:** While it is technically possible to have two tier ratings for a given facility it would be considered highly unusual and is unlikely to deliver a commercial outcome for the client. Modern advances in technology and newer design topologies have lowered the CAPEX premium to achieve Tier IV objectives.

**Hopman:** One of the components that must be carefully considered in the design is that a Tier IV (or fault-tolerant) facility requires compartmentalisation of the facility support systems, and the routing of utilities between those facility support systems and the IT space.



The Telin-3 data centre in Singapore offers Tier IV certification.

While there are some unique designs that have met Tier IV requirements with a Tier III level of mechanical equipment, typically a facility that supports a Tier IV data hall would require more building system infrastructure and space than one solely housing Tier III space. However, there are multiple solutions and building configurations that could be deployed that would have different resulting impacts in system and building configuration.

An important factor to keep in mind is that Uptime's Tier system considers the emergency power generation system – including the fuel delivery – as the primary power source. The local utility is an “economic” alternative.

**Toner:** Most commercial data centres don't have Hybrid Tier designs and as such don't offer different Tier service levels within a single data centre facility.

While the business availability requirement of a client's technology can vary from application to

application – e.g., financial systems are more critical than test/development systems – generally clients procure data centre services on a one size fits all basis. Price sensitive clients who truly understand their business availability requirements and cost of downtime could be tempted to split their IT infrastructure across Hybrid Tier facilities, if the option existed.

**Eco:** How are data centres typically certified? What metrics are considered?

**Hopman:** Requirements for certification vary greatly by industry and geographical location. It is not uncommon to see businesses such as colocation providers require a certification, usually Uptime although there are others such as that promoted by International Data Centre Authority (IDCA). We usually see users specifying project goals for PUE and some have also provided goals for WUE (Water Usage Effectiveness).

**Treweek:** Uptime Institute certification is the default standard. It is a performance based approach, not prescriptive, and is technology-neutral and vendor-agnostic. Certification can be done at the design stage but can also be carried through to the construction and operational stages of the centre.

PUE is one key metric that data centres are measured by, given the significant operational energy costs associated with these facilities.

**McCarthy:** Peak PUE, annualised PUE, NABERS, CUE (Carbon Usage Effectiveness), WUE, LEED, European Code of Conduct, TIA 942 and Uptime Institute to name but a few. Some data centres will be externally certified while others can internally certify to achieve the same level of oversight.

Other metrics based on FM Global or SCEC Zoning can also influence certifications.



**Toner:** There are a number of certifications the commercial data centre market is typically aiming to achieve including Uptime Institute Tier certification (typically Tier III or Tier IV), NABERS and LEED. There is also ISO 27001: 2013 Information Security Management System Certification, ISO 9001:2015 Quality Management System Certification and ISO 14001:2015 Environmental Management System Certification.

Other niche certifications exist, that are driven by specific client requirements. The data centre operator may purpose-build a facility to accommodate this, such as PCI DSS Payment Card Industry Security Standards and SSEC security standards.

**Gianoli:** Ultimately it comes down to what is important to the client. For some, the security of the facility is an absolute imperative requiring the highest level of SSEC certification. Others may have had a bad experience with a particular UPS topology so they'll be looking to avoid that in future lease agreements. Universally, energy efficiency is a priority, with both average and maximum PUE key market differentiators for data centre operators. We are regularly engaged to review the operation of facilities and provide advice on strategies to improve their energy efficiency.

NDY is proud to have worked with NEXTDC in the NABERS certification of their M1 and S1 facilities – the only two data centres in Australia to have achieved a 5 star NABERS certification.

**Eco:** In your experience, what fraction of a data centre's energy profile is attributable to cooling? Does it differ according to the scale and profile?

**Hopman:** The amount of energy required is dependent on the climate of location and internal environmental operational requirements; however, the annual average power usage for cooling systems will typically range between 10 to 35 per cent of IT load. There are solutions and internal conditions that can result in average power usage outside – higher and lower – of this range. To understand the applicability of those solutions requires a more in-depth discussion of the allowable operating environmental conditions and the potential impact on personnel and the IT equipment.

The power demand of the mechanical system typically does vary with scale, with larger equipment being more efficient; however, solutions such as direct or indirect airside economisation, and others, can lend themselves just as effectively to smaller-scale facilities and achieve comparable results.

**Gianoli:** Typically 25 to 40 per cent of energy is used in cooling per annum. The differences can be attributed to a range of factors including not only the scale of the facility but the relative ambient conditions, use of free cooling systems, system load as a function of a design load and commissioning of the facility.

**McCarthy:** An important design consideration is part-load performance – even a well-designed low-PUE facility may perform poorly before fully utilised. Staging and phasing of cooling plant in conjunction with maximising free cooling opportunities is critical.

A recent facility in Melbourne was achieving 15 per cent cooling energy of total demand; however, the average in Australia is likely closer to 40 per cent when climate is considered.

Machine learning and server eco functions can help to reduce and optimise a cooling system's performance, reducing carbon footprint and facility overheads.

**Treweek:** The cooling energy profile is highly dependent on the facility and system type. An older facility utilising chillers/CRACs, with other combined uses such as a NOC (network operations centre) might have an HVAC energy profile of, say, 40 per cent.

A new black-site facility utilising liquid cooling or in a cold climate where outdoor air can be used for free cooling through energy storage systems will be significantly less. An annualised PUE of 1.1 to 1.2 is achievable in certain circumstances.

**Toner:** We know 10 to 20 per cent of a data centre's energy profile is attributed to cooling. It differs based on the cooling architecture and location of the facility. For example a data centre located in the cooler southern parts of Australia can be designed with free cooling infrastructure. That won't work for data centres located in the warmer northern parts of Australia.

A purpose-built facility can take advantage of the latest innovations, while often an existing facility cannot, due to spatial, acoustic or other constraints.

## EMERGING TRENDS

### WARM LIQUID COOLING

"In the high-performance computing space, we are seeing some solutions being developed with liquid cooling direct to the server in the rack," says Mark Toner. "This is enabling rack densities of up to 80kW, reducing data centre footprints."

### DIRECT WATER

"With the increased prevalence of machine learning and the associated computer-intensive workloads, we're starting to see direct water or in-rack cooling for individual racks using up to approximately 60kW," says NDY's Craig Gianoli.

### SOLAR FARMS

"Green data centres are also becoming more prevalent, with the use of large-scale solar farms attached to centres, where space permits," says Sean Treweek, F.AIRAH, from Meinhardt Australia.

### NODE COOLING

"Node cooling, or directly cooling high-density servers within traditional air-cooled spaces is becoming more prevalent to accommodate processing needs," says Ross McCarthy from Aurecon.

### GREY WATER

"One of the areas that is starting to get attention, and will be getting more, that will have an impact on cooling designs, is the amount of water used by these facilities," says Marco Hopman, M.AIRAH, from Jacobs. "There are developments in the use of grey water, water reclamation, and just an overall reduction in a reliance on water – especially potable water systems."



**Eco:** Where is change and innovation occurring fastest in the pursuit of reducing the energy profile of cooling? Is it at the ICT equipment end or at the HVAC design end?

**McCarthy:** We see a combination of both – higher temperatures allow for greater free cooling, and syncing server profiles allow for opportunities for predictive control through the use of control neural networks.

We have seen great results from using machine learning on existing facilities to optimise and reduce cooling energy demand.

AI is driving a large portion of our data needs – we like the symbiotic relationship to “train” these tools to optimise cooling systems.

**Hopman:** The simple answer is it’s a bit of both. The HVAC designer reacts to the conditions required by their client – the data centre owner or operator. The owner requirements are based on the IT equipment and a perception of risk.

Historically, data centre operators preferred a low temperature to offset hot spot development. In the mid-2000s hot aisle containment systems started to appear that allowed the HVAC designer to increase supply air temperatures, and other system modifications, as the containment physically blocked server air recirculation.

ASHRAE’s TC9.9 worked with IT equipment manufacturers to define the operating temperature and humidity condition, and in 2004 published the first edition of the Thermal Guidelines for Data Processing Environments. Updates followed in 2008, 2012, and 2016. With the concurrence of the IT equipment manufacturers the environmental criteria expanded, and with the backing of IT manufacturers and ASHRAE research, data centre operators embraced using the ASHRAE guidance. Later in the 2000s as a further push for energy efficiency occurred, there was a marketing effort by AMD to reduce CPU power, and The Green Grid established the PUE metric.

It is in this environment that HVAC designers and equipment manufacturers started to more fully evolve their designs, and data centre operators quickly embraced the PUE metric, both as an element in their marketing and to drive operational efficiencies.

Interestingly, The Green Grid developed the PUE metric to measure improvements in a facility over time but never intended it to be used to compare facilities.

**Gianoli:** This depends on the nature of the facility. The large corporate data centres are pushing extremely hard on the permissible thermal envelope, whereas clients in colocation facilities are generally more conservative and reliant on achieving efficiency through more efficient designs.

**Toner:** It’s definitely a bit of both, but perhaps more in the ICT equipment space. Free cooling is pretty much mainstream now for any new builds in the right climates. As too are elevated data hall temperatures in line with

ASHRAE standards, hot and cold aisle designs and variable speed EC fans.

Meanwhile ICT equipment is being designed to run at higher operating temperatures, being deployed in higher density formats and developing mainstream solutions for liquid cooling.

**Treweek:** The ICT equipment is changing rapidly in terms of its configuration and footprint, as well as its ability to operate across a wider band or under harsher circumstances. This, in addition to energy demand reductions, is driving the uptake of alternative designs to achieve better outcomes – more cooling using less energy with a lower carbon footprint in a higher density and therefore reduced area. ■

## TIER BY TIER

The Uptime Institute created the standard Tier Classification System as a means to effectively evaluate data centre infrastructure in terms of a business’ requirements for system availability.

The system provides the data centre industry with a consistent method to compare typically unique, customised facilities based on expected site infrastructure performance, or “uptime”.

According to the Uptime Institute, the Tiers (I-IV) are progressive, in that each Tier incorporates the requirements of all the lower Tiers.

### TIER I: BASIC CAPACITY

A Tier I data centre provides dedicated site infrastructure to support information technology beyond an office setting. Tier I infrastructure includes a dedicated space for IT systems, an uninterruptible power supply (UPS) to filter power spikes, sags and momentary outages; dedicated cooling equipment that won’t get shut down at the end of normal office hours; and an engine generator to protect IT functions from extended power outages.

### TIER II: REDUNDANT CAPACITY COMPONENTS

Tier II facilities include redundant critical power and cooling components to provide select maintenance opportunities and an increased margin of safety against IT process disruptions that would result from site infrastructure equipment failures. The redundant components include power and cooling equipment such as UPS modules, chillers or pumps, and engine generators.

### TIER III: CONCURRENTLY MAINTAINABLE

A Tier III data centre requires no shutdowns for equipment replacement and maintenance. A redundant delivery path for power and cooling is added to the redundant critical components of Tier II so that each and every component needed to support the IT processing environment can be shut down and maintained without impact on the IT operation.

### TIER IV: FAULT TOLERANCE

Tier IV site infrastructure builds on Tier III, adding the concept of “fault tolerance” to the site infrastructure topology. Fault tolerance means that when individual equipment failures or distribution path interruptions occur, the effects of the events are stopped short of the IT operations.

Source: The Uptime Institute, [www.uptimeinstitute.com](http://www.uptimeinstitute.com)