The Guide to Edge Computing

Why it matters to your customers' experience and your operations

A Tech Research Asia Report
Commissioned by Schneider Electric

Introduction and Executive Summary

The opportunities presented by emerging technologies such as those that come under the Internet of Things (IoT) or digital transformation are exciting. We can do things better or create entirely new experiences or outcomes. It is imperative for all business and IT leaders to ensure they are exploring this new world. In our view, part of this will also necessarily involve challenging the status quo assumption that all IT applications should only be moved into centralised data centres or cloud computing centres. While the ICT industry has been heavily promoting a centralisation to cloud or colocation mantra for several years and many enterprises have enacted cloud-first programs, there are legitimate business and technical reasons to now consider placing equipment and applications at the edge of networks to complement cloud approaches. This TRA whitepaper will outline why this is the case and why that is important for your customers' experience in future. This report provides you with:

- An explanation of what edge computing is and isn't
- An analysis of TRA and Schneider Electric research on Edge in Australia
- The business drivers for adopting edge computing
- The technical reasons for adopting edge computing
- Examples of the possible applications of edge computing
- A checklist of strategy questions to ask of your own organisation's approach

Key points:

- Edge computing is relevant to many types of IoT- or digital business-related projects and outcomes. Not everything can or should be delivered from a centralised data centre or cloud computing service. But edge computing is typically complementary to centralised approaches.
- Regardless of type of project, the customer's experience should be an influential factor in determining whether edge computing is adopted or not. In most cases it will be the most important factor to consider. Emerging technology projects should not fail to deliver a great customer experience. This may not always be possible from existing cloud computing or data centre colocation services.
- Edge computing is not a product and will be viewed in a different light depending on the unique circumstances of the organisation, its desired outcomes and the project being pursued.
- Many Australian IT and business leaders say they already have an edge computing strategy in place. However, we would argue many of these organisations are conflating older distributed computing approaches with contemporary edge computing. There are many differences to consider.
- Regardless of whether you agree with the ideas economy and innovation hype, Australian IT and business leaders are trying to deliver more value than just keeping the IT lights on. Edge computing will have a role to play considering many projects pivot on the customer experience and IoT or digital transformation.

Data Centre and Cloud Computing Strategy in Australia Today

To better understand the data centre and cloud computing strategies being pursued by Australian organisations and the possible role of edge computing, Tech Research Asia undertook a program of quantitative and qualitative research. This included over 20 in-depth interviews with IT and business leaders, a survey of 320 organisations that had more than 50 employees, and peer-group discussions at events in Perth and Melbourne.

The dominant theme emerging from this research echoes TRA's other work and experience – the market remains in a period of ongoing consolidation of applications and infrastructure either to colocation or cloud services. Approximately 75% of the market are expecting to do some level of consolidation of their data centres and infrastructure in the next 24 months, with 42.5% saying they will move out of all their facilities in favour of colocation or cloud services. In contrast, a minor percentage of the Australian market expects to expand the number of data centres they have (2.19%) while 10.3% will upgrade and modernize their existing facilities.

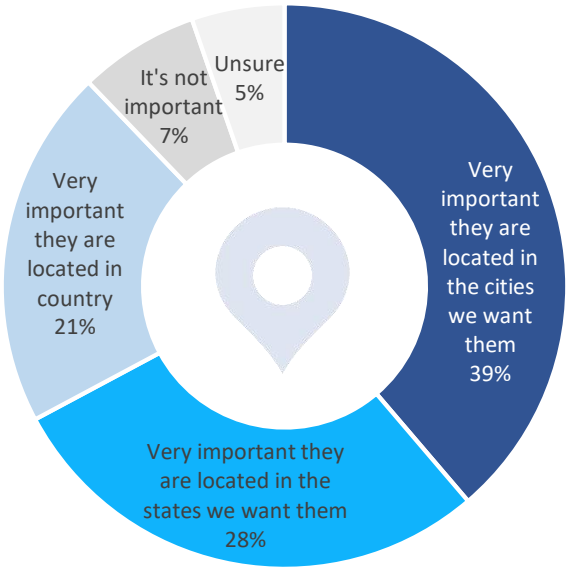
This result should come as no surprise to anybody. There has been a consistent push from both the supply side of the market, and also from business and IT leaders to adopt cloud, and to a lesser extent colocation services. Indeed, 33% say they have a "cloud first" strategy, which means they will consider cloud computing options (IaaS, PaaS, or SaaS) ahead of all others (but not always to the exclusion of other options). That's all expected to happen in the next 24 months and comes on top of the existing period of consolidation in the market that has been happening in the last five years. Clearly, these two approaches are becoming mainstream and we have some way to go until it is completed.

However, it is important to note that many organisations are retaining some of their own facilities or using a mixture of options. Despite the hype, not everything is "in the cloud", and not everyone is getting out of data centres completely. The research further showed that no single type of application workload has more than 50% of Australian organisations running it from a cloud service or is likely to in the next 24 months.

Indeed, the reality is there are many approaches being taken including on-premises and colocation. Yes, cloud computing is a destination for many and there is strong migration activity likely to occur over the next 24 months in all of the most common workloads investigated as part of this research. However, the majority are not moving everything to cloud – at least not for the foreseeable future. Hybrid approaches are and will remain prominent.

Among all this, in TRA's view there are an increasing number of organisations making decisions on a workload-by-workload basis and mapping them not just to the best hardware, but also the best data centre or cloud service, whether this is internal or external. This often follows the step of evaluating and identifying what the organisation's DNA is. Or in other words, what is the entity's purpose and which application workloads are most critical to achieving this core goal? Answering this often means undertaking the mapping activity mentioned above. There are other drivers helping to inform this process.

Q. How important is physical location to you when deciding where to locate a workload?



For instance, one persistent discussion raised when considering data centre and cloud computing decisions is whether the physical location of workloads matters. There are many facets to this discussion including data sovereignty, application latency, end user or customer experience, and network costs to name a few. The survey results (see

chart above) – where 67% think it is important to have workloads in the cities or states they desire – along with the ongoing expansion of major cloud computing service and colocation providers into Australian cities should now have put the view that it 'doesn't matter' and you can locate anywhere, to bed. It won't as many providers will continue to push the 'cloud from anywhere' view. But it is prudent in TRA's view to proceed on the premise that physical location does matter.

We have many reasons for saying this, one of the most important being that where data resides will continue to be a critical decision and influenced by government regulation as well as the preferred data architecture. Some of the other survey results also support the idea that physical location *is* important. For instance, over 50% of Australian organisations said application latency is very important (giving it an 8,9 or 10 rating out of 10), 47% said customer experience was very important, and 44% say Internet of Things (IoT) projects are very important.

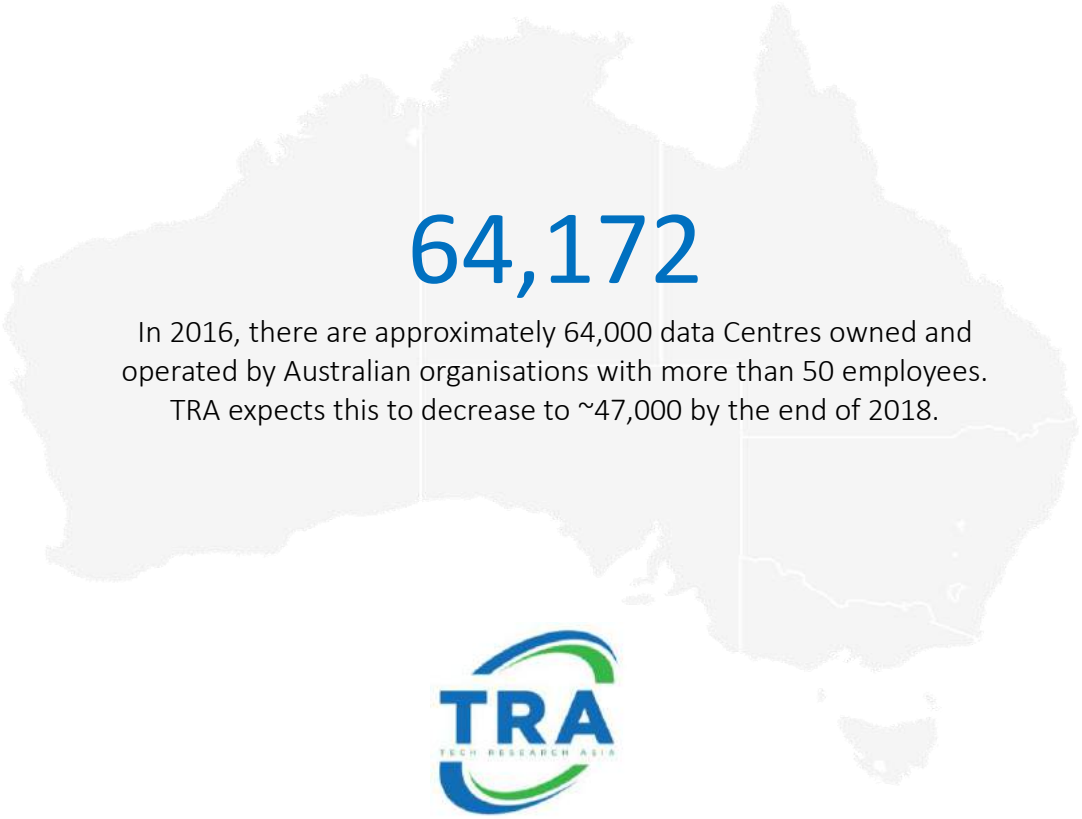
Why do we call out these specific areas? The application latency question is the easiest to grasp and encapsulates the main thrust of the overall argument – if latency is too high, performance of the application may be insufficient and a potential risk to the organisation. Indeed, this idea permeates the other areas as well. Will customers wait for content or for an application to respond? Can you make real time decisions using the data captured from IoT-related projects that are at the centre of innovation efforts? It all comes down to whether the outcome being sought is achieved or hindered by the choice of hosting location of the workload. This is where the contemporary concept of edge computing comes into the picture.

Edge Computing: What is it?

It's not new but is still hard to pin down with one ultimate definition. One thing edge computing certainly isn't, is a product or one type of IT architecture. There will be suppliers that attempt to frame edge computing in product terms – with all the associated marketing terminology – but this should be ignored and actively discouraged.

In reality, edge computing is always going to be contextual to each organisation. For example, someone with a global outlook or business may say a colocation facility somewhere in Australia is the edge. Someone with a national view might say capital cities are the edge. In any one city, it may be the actual site of a project or operation or office/building. In TRA's view, these are all reasonable positions to take.

However, we would encourage IT and business leaders to look at today's edge computing discussion as being about the locating of applications, data, and related IT infrastructure at specific, localised physical sites. This differentiates it from the popular use of primary and



64,172

In 2016, there are approximately 64,000 data Centres owned and operated by Australian organisations with more than 50 employees. TRA expects this to decrease to ~47,000 by the end of 2018.

secondary centralised colocation data centres or cloud services. Indeed, in TRA's view it is a different enough approach from what we've been doing to date to seriously consider when looking at your data centre and cloud computing strategy.

We acknowledge that some will ask whether today's discourse around edge computing is any different from distributed computing of the past. In some ways it is very similar. You have IT infrastructure located in distributed environments that are not centralised data centres or cloud computing services. Yet, there are some important differences we would highlight that should mean it is at the least evaluated afresh by IT and business leaders. Some of this is outlined below in the different business and technology trends sections. We would draw immediate attention, however, to the different use cases involved**.** Edge computing deployments today are often supporting single, specific – and mostly new – workloads or outcomes that may be related to IoT-style projects. (See below for some examples.)

Today's edge deployments are not necessarily for the generic IT services (such as file servers, email, collaboration, etc.) that have been maintained from distributed data centres (but don't exclude them) previously. These new deployments may also be for data- or video-heavy applications that have real time output requirements.

It is important to note that new edge computing projects may be via captive or on-premises data centres (bricks and mortar or pre-fabricated containers/closets), it may be through colocation facilities, or it may be through cloud computing providers. For the latter, this could either be localised providers or even the big global players that are increasing placing their own nodes (IT infrastructure) in more distributed locations to address the same or similar issues facing IT and business leaders.

TRA's research uncovered that 16% of Australian organisations with more than 50 employees say they already have an edge computing strategy in place. 63% of this say they use their own data centres, and 15% say they use cloud computing services near customers. A further 22% say they are about to undergo a program of works regarding edge computing with 55% of this group saying they will build a facility or use a pre-fabricated data centre option and 14% say they would use a cloud service.

We think this combined 38% of organisations doing or about to do edge computing may be too ambitious in their self-evaluation. In that it is too optimistic for this much activity to take place at the moment and perhaps instead means the market is unsure of what edge computing really is today and continues to conflate it with the old distributed computing concept.

Some possible types of edge computing we'd like to propose to help with this required education, include:

- **The Latency-Sensitive Edge** – This type of edge computing project focuses predominantly on customer or user experience and/or application performance. Latency becomes one of the driving forces in decision making. If it is too high, the outcomes of the project(s) are jeopardised. As a result, IT infrastructure and the applications that run on it need to be placed closer to where the customer is or where the activity is taking place to ensure outcomes are met. This has already been a reality for many media organisations or large websites that cache content at the edge of their networks, often via content distribution networks (CDNs). However, it will also now be the case for many IoT projects along with digital business programs where customer experience or real time action is the priority.

"We have invested heavily in edge computing and are now taking it as far as we can into the devices our customers' use."

Utility Provider CIO at TRA roundtable in November 2016.

- **The Proximity Edge:** This type of edge computing approach is for when an organisation has a requirement for ensuring the services it offers don't go down as a result of incidents happening in other locations. They need to be completely self-reliant. For example, when a natural disaster or other major incident occurs there are several types of organisations that need to continue to provide services such as hospitals and first responders. What is their acceptable "blast zone" for the services they offer? There are what might be called special circumstance organisations that provide critical services like these that have a different risk profile from other enterprises. It may be an unacceptable risk for them to rely on a network provider to access a centralised data centre or cloud at all times, or even to rely on the actual centralised data centres or cloud services themselves.

- **The Data Edge** – A Data Edge is when you need to carefully consider the way you handle data and decide if you need to transmit all or even a little bit of it over a network back to a central repository, hub, lake, or warehouse. Is it smarter to do the collecting, storing, parsing, analysis, and decision making or actioning on site? Questions that will arise during this evaluation relate to: data sovereignty and compliance – are you legally allowed to transmit the data; network performance – will sending the data result in lower performance that impacts other areas; and economics – does it actually make financial sense to send the data over a network link?

- **The Beyond Supply Edge** – This type of edge computing is the most similar to traditional forms of distributed computing. It is to be considered when there are no reliable or available (even in the financial sense) cloud computing or colocation services from which to deliver application services from. It will be more common in regional and remote locations. Moreover, it will be the type of edge computing that is adopted in remote places where technology hasn't been deployed before or isn't maintained regularly – such as for exploration in some industries and remote site operations in others. Often it may be beyond the reach of existing terrestrial telecommunications network coverage.

The above types of edge computing are not fixed nor exhaustively definitive in any way and we don't intend them to be. There may be other ways to consider edge computing and we welcome any addition to the discussion. These are merely guides to help frame things: we believe they may help those investigating edge computing to understand the different ways the concept may be applied beyond conflating it to past iterations of distributed computing.

The business side of edge

There will be many factors to consider when evaluating edge computing, or indeed a typical data centre or cloud computing strategy. We offer the following business factors into the mix as a starting point and encourage proper due diligence and "outside of the box" thinking:

- **The user or customer experience:** We've touched on this already and for good reason – it should be a significant driving factor in all decision making across the organisation regardless of whether they are in a business or IT role. User and customer experience can be significantly impacted because of where an application is located. Competitive advantage and market share can be won or lost depending on whether competitors or your organisation are closer to customers and providing a better experience. Further, application performance may be "okay" today but as we already know, customer expectations change quickly and important new service features with higher bandwidth or lower latency requirements can be just around the corner. It is prudent to consider whether a centralised and/or edge computing approach allows you to best adapt.

- **Risk and business continuity:** Undertaking an edge computing project or indeed any change to your data centre environment, regardless of what the final result is, will have an impact on your risk profile and your business continuity planning. Greater resilience is possible if things are done right, but so too is greater risk, including with your regulatory requirements. The impact on this area must be front of mind during any investment evaluation.

- **Business model opportunity:** Just as some organisations have realised they can leverage their own data centre assets to provide services to the market, the same can also be said for edge computing. Some business and IT leaders will immediately baulk at this suggestion considering their adoption of "as-a-service" IT in recent years. But TRA encourages everyone to investigate what new businesses can be developed via an edge computing investment, especially if it involves an IoT project or a data-driven service. The cost and time required to realise innovation or new services based on technology has never been lower. Plus, there are many gaps in the market and buyers consistently express a willingness to consider start-ups or non-traditional technology suppliers in their own strategies.

- **Attracting or retaining great people**: Edge computing projects are different and may involve some of the most exciting emerging technologies or efforts to solve important issues. The same was true of cloud computing for the past few years, although this is now becoming mainstream. It is prudent to see whether your organisation can establish a "virtuous circle", whereby pursuing rewarding and different projects from what is happening in the mainstream allows you to attract or retain great people, that then has a cascading effect on organisational outcomes.

- **It is not necessarily Capex:** Although we are effectively talking about data centre facilities, edge computing does not necessarily mean you need to consider a 20-year+ capital investment like we did with bricks and mortar facilities in the past. Today, there are operational expenditure or leasing options for facilities and infrastructure.

- **It's not the same time frames:** While captive or on-premises data centres may have taken many months or even a couple of years to build in the past, many edge computing facilities – especially pre-fabricated containerised data centres – and all the associated IT equipment (potentially along with the application and data layers) can be installed within weeks or a few months. This is, of course, typically slower than using a public cloud computing provider, but it is much faster than traditional data centre construction.

59%

Is the average perceived amount of time spent on keeping the lights on versus working on new innovative projects or programs according to Australian IT and business leaders. This should be viewed as aspirational as it is very unlikely to reflect the reality in the average Australian organisation. In TRA's view this is the percentage of time most leaders "want" to spend between the two options, but not the actual. However, it is a good indicator that Australian organisations want to be innovative and try new approaches. Many of these will likely require an evaluation of edge computing.

The technology side of edge

Just as there are many additional business factors to consider, there are also many technology factors. The below is, again, a starting point and not a comprehensive list of technology factors to consider:

- **Latency:** Like customer or user experience we've already mentioned this a lot, and that's because it is a very important component that will continue to influence decisions for many years to come. Consider that in addition to trying to understand what the appropriate latency is for a given application, we should also accept that: not all data centres have the same performance, they don't have the same performance over time, they won't always perform the same at the same time on the same day but in a different week or month, one test or simplified evaluations may not give you the best view of what good performance is, and most applications will continue to require better performance – or perhaps that should be the other way – users will need and expect better performance in future. There is a lot more to the latency question than just speed. To be sure, there are many things that affect application performance and what is acceptable can be subjective. That said, the lower the latency the better position you are in for achieving the desired outcome.

- **Data management:** Many edge computing projects will be pursued because of the value of

the data involved. The issue will be how well the data is managed and whether it can or should be sent back to a centralised hub directly or whether it makes more sense to deal with it (to some degree) at the edge. For example, in an edge computing approach instead of sending collected data in raw format to a central data hub, it could instead be collected, stored, parsed, possibly analysed and actioned (or shared) at the edge, and then forwarded, stored, analysed, actioned, and archived at the main central hub. There are many possible approaches, but the point is that edge computing can offer more alternatives to pursue.

- **Data Centre and Cloud Management:** Adding edge computing into the mix will mean ensuring the organisations has full visibility and ability to manage the environment effectively with a new architecture. This will include software tools like DCIM but also people and vendor management.

- **BCDR:** Processes and practices for ensuring reliable BCDR (business continuity and disaster recovery) may in some instances be more complex when involving an edge computing strategy. But at the same time the distributed nature of these approaches can mean each individual site becomes self-reliant. Thus, improving the overall level of resilience in the organisation. Regardless, a new BCDR strategy will need to form part of the overall strategy.

- **Networking:** Many edge computing projects will involve different networking approaches including low power WAN networking or CORS networking for agricultural IoT projects, or even Software Defined WAN for moving data around over the Internet. In some instances, it may mean a change from the traditional MPLS-based hub and spoke model.

- **Scalability:** It will be critical to ensure that whatever edge computing approach is adopted that it is able to scale and adapt. This will be important particularly for those pursuing "smart" projects (i.e. a smart city) where the scale of projects can grow quickly as well as the volume of sensors and data. Indeed, by their very nature, many of the workloads that could be appropriate for edge computing are evolving quickly. By scalability we refer to scaling up as much as scaling down.

- **Integration:** Unless completely outsourcing the environment, integration will always be necessary for any edge computing project – many are, after all, attempting something relatively new with emerging technology. However, the IT hardware component sitting in the data centre facility may possibly be less onerous than the past as a result of converged and hyper-converged infrastructure. There are also many service providers that can do this work on behalf of an organisation, reducing the requirement for internal integration or implementation skills.

- **Security**: There is no understating the importance of security in contemporary organisations. It is effectively a board-level priority as much as improving customer experience is a requirement. Edge computing and the associated technology and facilities will expand the number of attack vectors and require advanced security capabilities and culture be established. This cannot be lip service.

"Edge computing is simply about getting closer to your customers and ensuring they have a great experience."

Investment fund CEO at TRA roundtable event. November 2016

Example Edge Scenarios

There are many possible examples of the application of edge computing. The following list is provided as an introduction and should not be considered exhaustive. Indeed, whether edge computing is used to complement your own strategy will be dependent on your unique circumstances. In many cases a centralised option may be just as viable as an edge computing approach. Regardless, we encourage business and IT leaders to examine edge computing and ensure it is part of any project planning or evaluation and to not fall into the trap of only considering centralised cloud computing or data centre services – there are many options available to you to ensure the best customer experience and operational outcomes.

Possible edge-related projects include:

- Smart- buildings or offices for digitally-enabled workforces and automated / integrated building management. This may include video monitoring and sensor-based networks for triggering actions that only occur within the premises based on occupancy movements.
- Smart farms – not only precision agriculture – that include significant levels of data collection and analysis, along with tech-enabled actions that help with production, safety, and lifestyle.
- Mine sites and other primary industry campuses and their associated living quarters in remote locations that aren't able to access common services.
- Smart fishing vessels that use video and/or sensors for catch analysis and machinery for production. Similarly, cruise and other ships will likely require edge computing in order to provide digital services.
- Hospitals and healthcare providers that require lots of bandwidth for sharing medical imagery and data. Considering the role that IT plays in contemporary care, they may also have a high aversion to reliance on 3rd party providers that have data centre or cloud facilities located off site.
- Stadiums, concert halls, and other event venues which depend on digital platforms to provide services to large numbers of visitors at the one time. This is also applicable to other locations like airports or train stations that have high volumes of visitors.
- Autonomous vehicles that require on-board computing capabilities in order to function effectively.
- Utility plants and infrastructure that have high numbers of sensors and data requirements, sometimes in highly distributed locations.
- Contemporary factories or production facilities that use autonomous machinery, robotics, and automated processes.
- Retail outlets and malls that have multiple environment management systems, smart car parks, CCTV, location-based services, and content delivery platforms.
- Marinas or ports that enable IoT-driven logistics such as those embedded with sensors or RFID chips and require tracking.
- Bandwidth heavy content or online games delivery.
- Biometric or facial recognition systems for transportation or other identification gateways.
- 4K CCTV either on a city-wide or one premises basis.
- Cyber security or other critical infrastructure command centres such as that relied on by public safety organisations.
- Defence operation theatres where self reliance is of critical importance.

A Decision-Making Framework

The relevance of edge computing will become clearer the more that Australian organisations move into the worlds of digital business and IoT. It is, in TRA's view, one ingredient for ensuring optimum customer experience and operational outcomes. It should not be viewed as opposing cloud computing or other centralised data centre strategies. Instead, it should be complementary and entirely unique to the project circumstances and outcomes sought. We strongly advise including edge computing in any evaluation of emerging technology projects.

Some things to do when it comes to edge computing:

- Make it about your customers' experience first and foremost. This really should go without saying.
- Decide your organisation's DNA and understand which application workloads are most critical to achieving this purpose. Edge computing may have a role to play.
- Make sure you have Intelligent Data Centre and Cloud Computing Foundations that focus on the outcomes and can adapt as you need. Don't prescribe solutions before you know what the problem or opportunity is.
- Consider the business model opportunity – can you turn a cost centre into a profit?
- Explore, Experiment – there are many opportunities to do things better or create entirely new experiences.
- Factor in the ongoing management and people skills you will need with any emerging technology project, especially if it includes edge computing.
- Push your providers to listen and work towards your desired outcomes – not just their quarterly sales targets.
- Encourage co-innovation and collaboration with partners and suppliers. Start with proof of concepts and encourage shared responsibility for outcomes.
- Push your providers to offer solutions based on the types of budgets that you have. You should have choice and flexibility in deciding how to consume edge computing.

Things we recommend not to do:

- Don't use new acronyms or fancy terms for your strategy when creating a business case or sharing the narrative with a broader computing. Terms like 'fog' computing – which is sometimes used to describe edge computing – only create frustration and alienate stakeholders. Keep it simple and put things in business terms.
- Don't think centralised clouds or colocation will be the answer for everything. At the same time, don't think edge computing is the only way forward. It is not a dichotomy.
- Don't think that your old distributed data centre facilities like server closets will be able to support the modern (high density) infrastructure used in many IoT and digital projects. Often they won't be able to sustainably and efficiently host the infrastructure nor be secure.
- Don't be complacent. The strategies we have taken with our underlying infrastructure over the past few years may not necessarily support the opportunities on the horizon.

Methodology

Schneider Electric commissioned TRA to undertake a quantitative survey of 320 Australian organisations in April 2016. Respondents had to be responsible for, or have intimate knowledge of the technology used in their organisation and have more than 50 employees. A total of 320 responses were achieved distributed across the country; 50% have more than five locations in Australia and 20% more than 5 locations across Asia Pacific. TRA also undertook multiple in-depth interviews with CIOs and CxOs at Australian organisations regarding their edge computing and data centre / cloud computing strategies. In addition, TRA hosted two roundtable events discussing the same topic in November 2016 in Melbourne and Perth. For further information on the methodology used, please contact Tech Research Asia.

This report was commissioned by Schneider Electric. For more information on Schneider Electric and its Edge Computing offerings, please visit http://www.schneider-electric.com.au/en

About Tech Research Asia

TRA is a fast-growing IT analyst, research, and consulting firm with an experienced and diverse team in: Sydney | Melbourne | Singapore | Kuala Lumpur | Hong Kong | Tokyo. We advise executive technology buyers and suppliers across Asia Pacific. We are rigorous, fact-based, open, and transparent. And we offer research, consulting, engagement and advisory services. We also conduct our own independent research on the issues, trends, and strategies that are important to executives and other leaders that want to leverage the power of modern technology. TRA also publishes the open and online journal, TQ.