White paper

Five steps to creating a more flexible data center

Gordon Moore, co-founder of Intel, famously observed back in 1965 that "The number of transistors incorporated in a chip approximately doubles every 24 months." The timeframe has since been shortened to 18 months, and transistor count has faded as the metric for computational power. Today, the law is taken to mean that a device of size X will have twice the computing power in 18 months as a comparable device today.

Moore's Law has proven to be so accurate that technologists, including data center managers, rely on it as they plan expansions and new construction.

Another fact you can count on is that demand on data centers will continue to grow. McKinsey & Company[1] project 40% growth in global data generated per year. The question of how technology center owners can ramp up to meet these demands efficiently can be difficult to answer.

Factor into the equation unknowable variables like the evolving electric utility landscape, new regulations and carbon taxes, and general tax implications related to "green" initiatives. Disruptive changes created by market forces or now-unknown technology are other variables that can have an exponential effect on the equation and quickly obsolete a "state of the art" facility.

This calculus drives data center planners to devise expansion and construction plans that incorporate considerable flexibility – flexibility that provides them with the ability to adjust for the many unknowns that lie over the horizon, while continuing to provide reliable transaction processing for their organizations or clients.

Scalability the top concern

In a recent survey,[2] 64% of the data center experts polled said that concern about scalability is their number one issue.[3] They fret about how to best lay the foundation for continued growth while maintaining up time and operational efficiency.

"Taking a scalable approach to system planning, including considerations of both IT and facilities, helps avoid drastic – and typically costly – modifications in the future," according to Mark Reed, the North America Manager for data centers at ABB. "Planning and constructing with scalability in mind may require a slightly higher initial investment, but you gain tremendous flexibility which will reduce you long term installation costs and disruptions to operations. Those higher costs can be spread over the life of the facility."

Mega-providers like Facebook and Apple have big enough checkbooks to start with a huge shell building and build it out

1. http://www.itracs.com/category/data-centers/
2. 2011/2012 Data Center Market Insights survey conducted by DataCenterKnowledge.com
3. http://www.itracs.com/category/data-centers/

over time as demand increases. For most organizations, making that big up-front capital investment is probably unwise and often impossible.

Instead, they construct current facilities with scalability in mind, starting relatively small, but sizing the upfront infrastructure for the ultimate expected load. That allows them to make investments in smaller amounts but with an eye toward how the facility can be expanded later to bring more technology equipment online.

Scalable data center construction requires far more than just having enough real estate available for expansion.

"Planners need to consider the expandability of all systems and processes, beginning with the substations and including HVAC, power distribution and the automation systems used to manage the facility," Reed says. "The inability of any one of these components to expand and meet increased demand will create a costly impediment to center growth."

The modular method

The concept of scalability might best be exemplified by the building-block approach represented in modular and container-based data centers. Both make it possible to bring blocks of new capacity online in a relatively short time.

Because they are engineered products, the technology is very tightly integrated, resulting in high efficiency in both power and cooling, with PUEs[4] of 1.1 available with modular technology. This approach also overcomes the common contention between IT and Facilities because the two systems arrive fully integrated. These solutions can readily achieve the Uptime Institute Tier 3 or Tier 4 standards.

There are significant differences between the modular and container approaches.

[5]The modular approach is typically based on a prefabricated module housing the technology required to add a certain amount of capacity. A module may be delivered as an intact unit, ready to be moved into a facility. It may also arrive Lego® block fashion – a kit of all of the parts ready for assembly on site. The timeframe to bring additional servers on line is significantly shorter than with a build-from-scratch approach. Time from order to online is a matter of months, with some suppliers promising that their platforms can be online in just weeks.

The modules may rely on the site's infrastructure for power and cooling or may come as a self-contained environment that incorporates the supporting HVAC systems, ready to integrate with existing systems. Customers must be wary of high PUE claims if the container or modules do not include the necessary backup power and cooling system components in the stated PUE values. Modular solutions for power, cooling and back up generation are readily available.

The archetype for the container approach may have been Sun Microsystem's Blackbox data center container solution that was introduced in 2007. Like the current generation, it was a plug-and-play unit of new computing capacity housed in a standard intermodal container like those seen stacked on container ships or railroad freight cars.

The containers may also be totally self-contained mini-data centers or configured for connection to external infrastructure. The facilities-related systems are sometimes also housed in containers. This creates a true building-block approach, making it possible to add container-housed servers and related containers for cooling and power as needed.

The containers, repurposed from their original duty in intercontinental shipping, are very durable and able to resist the elements. In most cases, they are housed within a facility, but with additional weather hardening are sometimes placed outside. A cluster of containers surrounded by the appropriate fencing to provide security may be visually inelegant but is highly cost effective. This approach changes the face of "bricks and mortar" facilities.

The distinction between "modular" and "container" is blurring as the size and configuration of modular offerings has expanded. While there are relatively small, server-only modular solutions, there are also turnkey modular offerings. The only clear distinction is that containers are literally housed in a repurposed shipping container; Modules may arrive "naked," ready for installation inside your facility or enclosed inside a weatherproof housing.

These two methods of scaling-up capacity aren't the right approach for everyone, but in light of increasing demand for modular and container-based solutions, it's an approach that's obviously meeting many organizations' needs.

The flexibility of reliability

Reliability is the sine qua non of data centers, and there's a connection between increased reliability and improved flexibility.

"The two go together to a certain extent," says Frank Catapano, Vice President of Operations at Validus DC Systems. "Higher levels of reliability also buy you more flexibility in most cases."

Center planners build in as much reliability as they can afford, including redundant systems and over-capacity to accommodate usage spikes and technical problems. But few, if any, organizations have an open checkbook. Hard decisions regarding the quality and capabilities of the selected technology have to be made. Inevitably, concessions are made as costs are shaved.

4. Power Usage Effectiveness, a measure of how efficiently the data center uses electrical power, calculated as the ratio of total power used to the power actually delivered to computing equipment. A PUE of 1.0 is optimal.

5. Most of the data regarding modular and container solutions was taken from Data Center Knowledge White Paper, Guide to Modular Data Centers.

"Commercial-grade equipment is a step up from consumer-grade, but it still doesn't deliver the capabilities of true industrial-grade equipment," says Reed. "The commercial technology often lacks not only some of the durability and features of higher-quality equipment, it also lacks the adaptability."

That places limitations on how the equipment can be installed, configured, operated and maintained, reducing the flexibility of the technology.

Considerations regarding technology also apply to power distribution equipment, as well as building management and DCIM software.

"You need to think about whether your critical building infrastructure and DCIM systems have enough flexibility to enable you to do an upgrade or support modifications without having to shut them down or put the critical load at risk," Catapano says. "A lower-cost system may save you money up front, but those savings quickly vanish if you have repeated maintenance shutdowns. A new rack of servers could mean costly changes to two, three or even more disparate systems. It's a pay-me-now-or-pay-me-later proposition."

In many facilities, N+1 or 2N power topologies are constructed in order to maintain redundant power paths to critical equipment. This scenario requires extra power conversion equipment in the delivery path; and in the case of a true 2N system, twice the number of components are required to support the load from a power consumption perspective.

"When redundant equipment is added, it drives down the overall system efficiency, since the power electronics operate in a lightly loaded state and there are more standby losses," Catapano says. "Recognizing that power electronic components tend to have higher efficiencies at higher loads and that servers don't always run at 100% load, we pay an efficiency penalty in redundant system scenarios based on today's system configurations."

Solutions being brought to market can address the aforementioned issue by actively managing power conversion modules and directing higher levels of power through fewer components to optimize efficiency. Actively managing power conversion in both AC and DC systems makes it possible to increase system efficiency while maintaining required levels of redundancy. This methodology can also be used to selectively manage onboard server components to quiesce or even turn off major energy-consuming devices in the server when these assets aren't required.

Planning for power distribution

While it's possible to estimate future power needs, precisely predicting these loads is impossible; so sizing the power requirements is a challenge. One approach is to simply include spare breakers and distribution equipment, but no one wants to invest capital until it's necessary. To reduce the cost of capital but still provide for future growth in power distribution, data centers rely on advanced switchgear that enables safe, power-on maintenance and upgrades.

"In most facilities, work on equipment is done during 'quiet times' so power shutdowns occur when traffic is minimal," Catapano says. "Maintenance people are allocated a service window of a set number of hours when there's reduced disruption to service, and less risk of an outage due to power loss during maintenance.

"Selecting the right power switchgear provides center managers with much greater flexibility. Cabinets are available that feature isolated power control units that can be removed, serviced and replaced while the equipment remains energized."

This type of switchgear also provides a safety benefit.

"When you have people, hands and tools inside the equipment, accidents can happen that put the center and personnel at risk," Catapano says. "In higher-quality switchgear, the manufacturer can isolate and insulate major power distribution elements. This makes it much more difficult for someone to make accidental contact with live power. Being able to safely work on power distribution without disrupting center operation provides the flexibility to work on the equipment as needed or when convenient. That's better than being tied to working within short, off-hour windows of opportunity."

A unified window on operations - DCIM

Knowledge is power. In a data center, accurate knowledge regarding energy utilization gives you the power to make better decisions.

"Having a Data Center Infrastructure Management (DCIM) system that allows you to monitor how and where power is being used throughout your facility gives you a great deal more flexibility," Reed says. "With more information at hand, you can better evaluate your options and take action with less risk of something going wrong. Some DCIMs can even take action for you, capitalizing on the intelligence built in to these more-advanced systems."

To provide the needed decision-making support, DCIM process sensoring must be extended as close to the server as possible. Some facilities are limited to measuring only top-level power metrics; the total IT power compared to the amount of power from the UPS. While useful when looking at overall energy consumption trends, this level of detail is useless when making day-to-day operations decisions.

What's needed is more granular output data coming directly from the servers, such as power, temperature, CPU utilization, etc. Newer technology may include energy-use feedback capabilities that can also provide data to the DCIM.

Sensors can be located to monitor power use at the rack, or better yet, at the server level. Incorporating these sensors adds both expense and complexity to the system, so center operators often take a phased approach to deployment. Sensors can be added to the locations that are of the greatest interest. In highly standardized facilities, it may be possible to extensively sensor a few racks or aisles and extrapolate data to the facility at large.

An alternative to physical sensor devices are agentless modules that extract data from every available source in the IT family, stores it in a database and turns it into useful business data that can be incorporated in the overall DCIM automation strategy. This is a growing trend that can reduce the capital costs associated with monitoring.

This information can also serve multiple purposes. It can determine efficiency and end-of-life for servers based on current technology available, performing "what-if" analyses to guide placement of servers, health monitoring, cost accounting and charge backs. It can also drive dynamic adjustment of cooling capacity as well as power capping and throttling. This option provides the most flexible of options for measurements because it is not hardware dependent as far as the IT equipment is concerned.[6]

Each data center manager must assess their "need to know" regarding granularity of data, but there's no such thing as too much data when it comes to DCIMs.

"The more you can monitor your power – the more you know where power is being used – the more flexibility you have in operating your facility," Reed says.

Switching to direct current

Direct current is receiving considerable attention these days because of its increased simplicity, enhanced reliability and other benefits. Switching from tried-and-true AC power is sometimes perceived as a radical change from well-known standards, yet it provides a simplified architecture for integrating many different power sources. Among the many advantages touted for DC is the significant increase in flexibility it adds to data center operation.

In the present AC world, power is converted many times between the utility and server. Those conversions require a series of devices that increase the cost of building and operating a data center while simultaneously reducing overall reliability. Aside from the fact that all of those devices waste energy, they also require physical space and maintenance clearances, further reducing flexibility for data center managers.

Consider the overstuffed closet or the cluttered garage. The fact that they offer little free space limits your ability to work in them and arrange their contents most efficiently. The same is true in data centers. Less space means less flexibility to operate the facility. It also limits expansion options. The relatively smaller number of power transmission/conversion devices required in a DC direct current power systems can provide valuable free space or mitigate significant portions of the initial construction cost.

The corollary benefit to reduced device count is reduced maintenance. There are simply fewer devices to breakdown or require preventative maintenance, and this creates long-term operational savings for the end user. Fewer equipment components also drives an increase in reliability.

The reduction in power transmission/conversion devices has a multiplying effect regarding reduced maintenance. Each of those devices generates considerable heat that increases the workload on HVAC systems. Eliminating large heat generators like uninterruptible power supplies and PDU transformers can bring about a significant reduction in the HVAC load. It logically follows that the reduced output required from the HVAC system will drive comparable savings opportunities within the mechanical cooling infrastructure as well.

An additional DC flexibility benefit comes from the ability to more easily integrate alternative energy sources.

"Right now, all stored and alternative energy sources are based on DC," Catapano says. "Wind turbines, photovoltaic and solar cells, ultra-capacitors, fuel cells and flywheel systems all generate energy in DC or have a DC link before their power is converted to AC."

Those sources can be added to an existing AC system with the repeated AC/DC conversions – inversions and rectifications – or can simply be integrated into a DC system through highly efficient, simple and reliable DC-DC voltage converters.

"If you have two different AC sources or systems in a data center, you have to design in very costly and complex controls to make sure that, if you switch between those two sources, the frequencies and the phase angles of the waveforms are in sync," Catapano explains." You don't have to worry about those issues in the DC world. DC is much more forgiving when you switch between sources. You can combine multiple DC sources with considerably fewer controls. This is one of the main drivers of DC system reliability."

Even with the flexibility gain and other benefits documented for direct current power, there is considerable debate about the wisdom of converting from AC.

"If you put 10 engineers in a room, they will come up with 10 different approaches to facility design," observes Catapano. "But most people planning data centers today are considering DC as a primary power source or as a cost-effective and highly reliable source for backup power to an AC system. With modern power control devices, transforming DC power has become much easier and more efficient than it used to be. In light of the significant opportunity to reduce space, increase efficiency and enhance flexibility, DC systems deserve consideration."

There are solutions available that don't require a wholesale conversion, instead providing a flexible, phased approach.

6. From Mark Reed

One system uses a standardized AC input to create a DC link that can support multiple stored or prime-energy inputs. These inputs can provide a +/-190 VDC, an AC or even a hybrid output to technology loads. This solution features non-volatile software management of the conversion modules to maximize efficiency while allowing a customer to scale the system over time to minimize up front capital costs.

Conclusion

"The incredible growth rate of data center transactions shows no sign of leveling off," according to Reed. "To keep up with that growth and reliably meet user demands, successful data center managers need to build in as much flexibility as possible. For current operations and data centers still in the planning phase, operators have a long list of options available to ensure the flexibility of their technology, systems and processes. By building flexibility into the technology center infrastructure and DCIM, operators will be better prepared to adapt both to technology and industry changes."

9AKK105408A8193

ABB Data Centers
125 East County Line Road
Warminster, Pennsylvania, U.S.A.
Phone: +1 800 HELP 365

www.abb.com/datacenters