POWER TO THE CUSTOMER – Q9'S CAPACITY RESERVATION MODEL ENSURES THE RIGHT FIT

Until very recently, most outsourced data centre service providers priced their services based on the amount of physical space a customer required. Not so at Q9. In 2005, Q9 led the industry by introducing a revolutionary pricing model based on how much power and cooling capacity it reserved for the exclusive use of each of its customers.

"Up until 2005 we priced in terms of the physical space our customers occupied," says David Chaloner, Senior VP, Marketing at Q9. "But it became clear to us that there was more customer value in accounting for how power and cooling resources were allocated. That was the genesis of our capacity reservation model."

To understand the shift from space to power and cooling, you have to look at how data centres are designed and constructed. Each data centre has a finite quantity of space, power and cooling resources available. Data centre space is simply the total amount of physical space available for customer equipment. Power resources include back-up power generators, UPS systems and power distribution systems that ensure customer equipment receives reliable, redundant conditioned power. Power systems as a whole are engineered to reliably deliver a specific and finite amount of peak power demand.

Cooling systems remove the heat generated by customer equipment (when it is operating and consuming power). As such, the cooling system capacity is carefully engineered to match the capacity available within the data centre power systems. The result of these three factors (space, power and cooling) is a particular design density: that is, an amount of power and cooling that can be effectively delivered to the space occupied by customer equipment. The simplest way to determine a data centre's design density is to take the amount of power and cooling it has been designed to support and dividing that by the total amount of space available for customer equipment. The result is a measure of density often expressed as a number of watts per square foot.

When customers were allocated space in a data centre, everything worked well as long as the amount of power and cooling resources required by their equipment remained constant. There was a balance between how much power, space and cooling resources a customer was using. As computing equipment became more powerful, however, the space customers required on a data centre floor tended to stay the same or even shrink while the amount of power and cooling they required in that space increased (often dramatically). This meant that customers were requiring more and more power and cooling resources in the same amount of space as a result of regular upgrades to their equipment over time.

FINDING THE RIGHT FIT

Compounding the problem was the fact that the method being used to allocate data centre capacity was also working against it.

"Space-driven pricing models motivated customers to fill their cabinets or cage space with as much equipment as possible," says David Ralston, Q9's Senior VP, Sales. "This wasn't really a problem ten years ago, because most computing devices required a generally constant amount of power relative to their size."

As customers upgraded to newer, smaller, more powerful servers, however, the increased power demand and heat generation in the same or smaller area created new cooling challenges.

"Newer devices are much smaller, so you can fit more of them into the same amount of space. They also consume more power and produce more heat; as a result, they require more cooling capacity for every square foot of data centre space," says Ralston.

As customers in traditionally (space) managed data centres upgraded their equipment, they began to experience operational problems as they exceeded the design density of their data centre. The resulting temperature increases from higher density devices also risked impacting legacy equipment within the same data centre.

"Data centre operators who sold capacity using a space-based approach have found that the newer servers can easily

overload their cooling systems," says Ralston. "In some cases, they've had to tell their customers that they couldn't use all the space that they were paying for."

In an attempt to control how much power is used inside their data centres, some providers have approached the problem by limiting the number of power circuits they provide to their customers. However this approach doesn't deal with the core issue of matching a customer's requirements to the data centre's available power and cooling capacity. In many cases, customers are forced to pay for a circuit as if they were using it at maximum capacity, when their actual requirements could be for only 20-30% of the circuit's capacity.

Enter Q9's Capacity Reservation Model, in which a data centre's total power and cooling capacity is treated as the data centre's inventory. As each customer contracts for a dedicated portion of the data centre's capacity, it is removed from inventory.

"We work with the customer to determine how much critical power and cooling they'll require to support their computing environment," says Ralston. "Whether the requirement is two thousand volt-amps or two million volt-amps, that's what we take out of our available inventory and reserve for their exclusive use."

Customers are allocated an amount of physical space that is matched to their power and cooling requirements based on the design density of that specific data centre.

"With this approach, customers are simply allocated enough physical space to allow them to spread out their equipment so that it can be cooled effectively," says Ralston. "There are nominal one-time charges for setting up the space and providing all the power circuits they require. But the only monthly charge is for the amount of power and cooling capacity that has been reserved for the customer."

This approach allows Q9 to maintain a balance between power, cooling and physical space and ensure that none of these factors is ever over-subscribed. Given that power, and not space, is now the constraining factor, Q9 uses its Capacity Reservation Model to guarantee availability for each customer during peak demand periods and to accommodate future growth.

"Every co-location customer, in every Q9 data centre, contracts for a dedicated portion of the facility's capacity, so power and cooling capacity cannot be oversubscribed" says Ralston. "This, combined with an appropriate level of redundancy in all of our key mechanical and electrical systems, allows us to guarantee 100% power availability."

DRIVING VALUE FROM IT ASSETS

It is one thing to say that a given amount of capacity is reserved for a customer, but how can customers be sure that this reservation matches their requirements? This is another area where Q9 distinguishes itself in the marketplace.

"Balancing a customer's power and cooling requirements with an amount of space only gets them half of the way there," says Chaloner. "They need visibility into their power utilization to ensure that the balance is maintained every day. We've invested in systems to provide power monitoring on a continuous basis. The monitoring tools can pinpoint individual circuits and racks or look at the aggregate demand across a customer's entire environment in a particular data centre."

For Q9, intimate knowledge of a data centre's capabilities and an organizational commitment to its Capacity Reservation Model means that the company's Customer Solutions Architects can work closely with customers to ensure they get exactly what they need.

"When quoting a customer, we first explain our approach to reserving capacity for them," says Chaloner. "Then we do an estimate of capacity – what a customer will need to reserve inside one of our data centres. And we look at how they can distribute their equipment to achieve good efficiency of power and cooling resources within their space."

The Q9 team then shows the customer a final layout designed to meet everyone's needs, and to make certain that enough capacity has been allocated for them. From that point on, the customer can monitor their own environment and see exactly how much capacity has been reserved and how much they are using.

"There is no mystery to this," says Chaloner. "This is the customer's equipment. They can go in themselves and see what the power draw is. The customer has the same ability to find out as we do."

The advantage with Q9's approach is not only the reliability that derives from the Capacity Reservation Model, but from the customer's ability to leverage the information in order to more efficiently use their IT assets. They are now able to plan for projects on a comprehensive capacity basis, and to set targets for individual energy reduction projects such as virtualization.

"We give our customers the tools they need to get the most out of their IT assets," says Chaloner. "They can see the effect of a new storage array, or a new server, and can monitor its impact within a matter of minutes."

This kind of visibility and control encourages customers to take a real interest in power consumption and running their own systems more efficiently. The end result is a more cost-effective, efficient, and environmentally responsible approach to energy use in data centres.