

Five Lessons from a Data Centre's Crisis of Capacity

By Robert Lemos , CIO , 15/10/2009

In 2005, problems in the data center at Pacific Northwest National Laboratory came to a head.

Unscheduled outages were occurring almost monthly, bringing down the data center for hours at a time. Groups were buying an increasing number of rack-mounted servers - which had recently become cheaper at the time - to boost the computing resources, says Ralph Wescott, data center services manager for the government laboratory, which is managed by the U.S. Department of Energy. In July, 2005, the server room had reached its capacity limit.

"Groups would go buy a server and throw it over the wall to me, saying, 'Hey, install this,'" Wescott says. "But I didn't have any space, power or cooling (capacity) left. If I installed (one more), the whole room would go dark."

Wescott and PNNL embarked on a broad project to revamp their data center without breaking the budget. Every quarter for three years, the data center group spent a weekend shutting down the server room and replacing a row of old servers and tangled network cables under the floor with more efficient, yet more powerful servers connected by fewer cables running in the ceiling. The new configuration allowed for more efficient cooling under the floor.

The result? PNNL moved from 500 applications on 500 servers to 800 applications running on 150 servers.

During a tight economy, tackling such information-technology projects require a tight grip on the purse strings, says Joseph Pucciarelli, the program director of technology, financial and executive strategies for analyst firm IDC, a sister company to CIO.com.

"The situation is a very common one," he says. "Companies are making just-in-time investments. They have a problem, and they are looking at the problem in a constrained way."

Here are some lessons PNNL learned in bringing their data center back from the brink.

1. Plan, don't react The first problem Wescott needed to solve was the data center group's habit of reacting to each small problem as it arose, rather than seeing the systematic issues and creating a plan to create a sustainable service. In addition to the 500 servers, the data center had some 33,000 cables connecting those servers to power, networking and security systems.

"We decided what the data center should look like and what its capacity should be," he says.

The group concluded that the current trajectory would result in 3,000 applications, each running on its own server, in 10 years. Now, the data center has 81 percent of applications virtualized - and average of 17 per server - and Wescott plans to reach the 90 percent mark.

Companies should focus on three areas to increase capacity, says IDC's Pucciarelli. Reducing the number of physical servers and running applications on virtual systems helps reduce power requirements, as does more efficient cooling systems and improvements in electrical distribution.

"That's typically the one-two-three that you go to when updating the data center," he says.

Pucciarelli has encountered many companies that have replaced up to 50 servers with just two or three larger capacity systems and used virtualization to run their applications.

2. Measure to manage Data center managers need ways to monitor the state of the data center, but all too frequently they don't have the right tools, PNNL's Wescott says. Prior to the changes, Pacific Northwest National Labs had no way to measure the efficiency of its data center. Power problems were discovered when the room went dark, or through a more seat-of-your-pants method.

"If there was too much amperage through our power supplies, the way I found out was to put my hand on the circuit breaker and if it was warm, then I knew we had a problem," he says. "That's proof that you need tools."

Now, PNNL has sensors in place on every fourth cabinet at the low, medium and high points to create a 3-D heat map of the server room. The data allowed Wescott to change the way he cools the data center, increasing overall temperatures and applying cooling where he needed it.

"I think that is going to save me a lot of money, and wear and tear, on my air conditioners," he says, adding that current estimates are that the data center will be 40 percent more efficient with cooling.

3. Take small steps Radically reconfiguring the data center without disrupting operations is a major problem, says Wescott. The manager advocates taking small steps to minimize outages, but left the decision to his managers, he says.

"I presented two choices to the management," Wescott says. "We take the entire campus for seven days and we go from scratch; the other is that we take an outage over a weekend every quarter."

* Source: http://www.cio.com/article/493746/5_Tools_to_Prevent_Energy_Waste_in_the_Data_Center.

By taking small steps, the group prepared to replace the data center a row at a time. On the first three-day weekend, the 30-person team spent 14 hours a day in the data center, replacing a row of server racks and testing the new configuration. Immediately, the data center became more reliable and stable, Wescott says.

If management cannot agree to allowing a data center outage, remind them that it's better to have a planned outage than a sudden, unplanned failure, he says.

"You can't paint the bottom of a boat as it is sailing across the ocean, but if you don't paint it, it's going to sink," says Wescott.

4. Accept short-term pain for long-term gain Management also cannot be shy of spending a little extra to save money down the road.

In order to reduce the energy requirements of his cooling system, Wescott's group evaluated waterside economizers, which use water and the outside temperature to cool racks of servers. While they estimated that using ambient cooling systems would save them money in the long run, the waterside economizers put the price of the cooling units 10 percent over budget. Wescott worked with the vendor, however, to reduce the price to within budgetary limits.

"They have paid for themselves over and over again," he says.

5. Find out what you don't know In revamping data centers, managers also need to look for places where energy is being consumed with little or no gain. A common flaw in data centers are ghost and rogue servers.

Ghost servers are machines that have been deployed but remain unused. They still eat up energy, but do not help the data center with its core job. A rogue server is a machine that someone has put in his office, outside of the data center, to skirt any restrictions that may be enforced by data center personnel. Such servers can waste a lot of energy budget, Wescott says.

"Buildings that should have shut down their air conditioning every night were running it to keep their rogue servers going," he says.

While the data center has only had a single unplanned outage since he started revamping the facility - due to an extremely hot day and a cooling system failure - Wescott knows that he has not finished the job, just pushed off the inevitable.

"We've calculated the wall," he says. "In five years from now, I'm going to run out of room because of storage, and I will probably run out of space in that room."