

The Numbers Don't Lie

To cut costs, Statistics Canada adds a partner and consolidates storage

It was the American author Mark Twain who, in the late 1800s, popularized the statement, "There are three kinds of lies: Lies, damned lies, and statistics." More than a century later, it took just a few Canadian statisticians and a bit of technology to prove him wrong.

In 2002, the Canadian government issued a nationwide directive for its agencies to reduce operating budgets. Among the agencies obliged to follow this directive was Statistics Canada ("StatsCan" for short), the organization that conducts a national census every five years. It also collects and analyzes statistics on everything from population growth to small business performance.

By Minda Zetlin

In response to the mandated budget cuts, Ivan Fellegi, StatsCan's chief statistician (the equivalent of the agency's CEO), challenged the agency to streamline operations. This challenge posed a problem for the agency's informatics technology services department (ITSD), which was experiencing an increased need for storage.

In 2001, Canada offered some citizens the choice of responding to the census over the Internet, rather than on paper.

"We got 7 terabytes of data from that," says Guy Charron, assistant director of StatsCan's infrastructure services. In the 2006 census, all 19 million Canadians old enough to respond had the option of providing their data online, and about 20 percent did so. Australia and New Zealand also offered the public the chance to respond online to a 2006 census, but

neither country had a higher return rate than Canada, according to the U.S. Census Bureau.

The new census method means huge storage needs: Charron's team estimates it will need 50 terabytes of storage from that 20 percent census response alone. The storage will need to be long-term, as the Canadian Parliament mandates census records be kept for 92 years.

StatsCan needed to increase storage dramatically while simultaneously cutting costs. Charron believed it was possible—if the agency could consolidate its data backup and storage into one stack overseen by ITSD.

In 2003, StatsCan had about 100 terabytes of data, most stored and managed locally by eight different operating divisions on more than 750 servers. That year, ITSD conducted a total cost of ownership survey, collecting information on data storage agency-wide. The survey showed that 49 percent of disk space was unutilized. Because backup and storage use was rapidly growing, it also suggested that these would be the principal drivers of future IT costs.

Though ITSD had been offering centralized backup and storage as an option to all of StatsCan since 1999, only 31 percent of stored data had been migrated to ITSD servers.

When Fellegi challenged ITSD to streamline, Charron and his team responded with a proposal to mandate that ITSD handle all backup, storage, and

file and print services for the entire agency. "The initial business case showed savings of close to CAD 3 million [about US\$2.7 million] over five years," Charron says. He knew that figure would be even greater once StatsCan was able to realize IT labor savings due to consolidation. "That's when we started making the case that there were economies of scale."

Zero budget

Controlling costs has special significance at StatsCan because ITSD operates as a pure fee-for-services division. "Unlike most other government departments, we are a complete internal cost recovery organization," Charron says. "Our IT budget essentially starts at zero every year. We sell our services to Statistics Canada's various divisions."

This arrangement grew out of the agency's desire to manage its IT costs by individual divisions—paying only for what was used rather than subsidizing the rest of StatsCan, Charron says.

"The advantage is not having to go back through for funding every time we want to do something. Our rates are set on consumption by users. Want to lower IT costs? Use less. Want more? Here it is."

Another result: efficiency. "I think we're a very lean shop,"

Charron says. "I've looked at what others are charging for the same services in private industry. Strictly from a cost perspective, we are much cheaper. We



Stats On
Statistics Canada
 Founded: 1918 as the Dominion Bureau of Statistics
 Employees: 5,700
 IT staff: 1,200+

could not outsource at anywhere near what we are delivering the service for internally.”

Any major change—such as storage consolidation—requires buy-in from internal customers. Before making its case to senior management, ITSD recruited IT experts from StatsCan’s eight fields to help evaluate the benefits of centralized storage.

“We set up multiple subcommittees,” Charron says. “One looked at storage, one at backups, one at the file and print environment. IT representatives from the client areas participated in developing the numbers sent to senior management. When clients are involved in the process, it’s hard for them to disagree on the decision afterward.”

“The committees had membership from across Statistics Canada,” notes Gary Roberge, who served on two committees as an IT representative from the transportation division. (He has since switched to ITSD, where he is chief of regional offices, IT and Intel server support services.) “It was a pretty hard sell when all this came down. StatsCan has about 65 divisions, and most have their own small IT shops. There was some reluctance and skepticism in moving from a stovepiped environment to a more centralized model. But once the business case was put on the table, it became clear that centralization was something we could not afford to pass up.”

In addition to the savings, consolidated storage offered protections that weren’t accessible to the divisions before. “If you asked if they had a disaster recovery plan, they’d say, ‘No, because it’s too expensive,’” Roberge notes. Though consolidating storage meant giving up some direct control, it also meant a less uncertain future.



Guy Charron, Assistant Director Infrastructure Services, Statistics Canada, with TeraMach Account Executive Jeff Clarke

“We showed we could achieve much lower cost in our enterprise solution,” Charron says. “Then we factored in nontangibles, such as availability and disaster recovery that we bundled with our service. It was clear that consolidation would provide a great value. It was only after we showed them these numbers that we got very senior management buy-in.”

Migration mandate

Buy-in meant ITSD got the mandate it was seeking: In January 2004, the corporate planning committee chaired by Fellegi decreed that the 69 percent of stored data still in servers among the agency’s field operations would migrate to central storage. TeraMach Technologies Inc., an Ottawa-based Symantec partner, helped ITSD select appropriate hardware and software.

Working with TeraMach, ITSD first centralized StatsCan’s storage and backup

environments into a three-tier system that lets data be placed on storage resources that align with the value of the data to the organization. In addition to the cost savings from consolidation, the system is intended to provide additional savings by storing older, less essential data on more cost-effective, less accessible systems. The tiered system consists of fibre disks for Tier 1, serial ATA disks for Tier 2, and tape for Tier 3.

Implementation of the tiered system is still underway, but the benefits of consolidation are already substantial—and greater than the CAD 3 million (about US \$2.7 million) ITSD originally projected. Centralized storage on Veritas NetBackup Enterprise Server allows for backup and recovery cost avoidance of nearly CAD 9 million (about US\$8 million). “Standardization of enterprise software tools such as Symantec NetBackup, CommandCentral, Enterprise Vault, and Storage Foundation Suite enabled the department to

True Partners

For Statistics Canada (StatsCan), a big advantage to its relationship with Symantec partner, TeraMach is cost: TeraMach's consulting services offer substantial savings over other options. This is a priority for the agency's informatics technology services department, which operates on a purely fee-for-services basis.

TeraMach focuses on four areas: storage, optimization, security, and availability. Jeff Clarke, a TeraMach account executive, believes the company has become a trusted advisor to StatsCan in these areas.

TeraMach can also help StatsCan select products that best fit its needs. "They keep themselves aware of our direction, our requirements, and our constraints," says Guy Charron, assistant director of StatsCan's infrastructure services. "And they've often helped us get the best value."

lower overall cost of operations," says Jeff Clarke, a TeraMach account executive.

In addition to hardware and software savings, ITSD estimates that centralization will save nearly CAD 3 million (about US\$2.7 million) in IT operating costs. Achieving these savings will take time due to StatsCan's no-layoffs policy. ITSD is implementing consolidation gradually. The departments will transfer data to central storage from 2004-2008, a range of time chosen because four years is the life cycle of a server.

"Because of that methodical approach, we are able to absorb people through retirement, through redeployment, and through absorbing some of the full-time employees in client areas into the central IT division," Charron says. "We've had no layoffs."

Without consolidation, StatsCan would have had to increase IT staff to deal with growing storage needs. "Beyond the census, they have many other surveys that require additional storage resources," Clarke says. "For example, the labor force survey, the small business survey, tax data, etc. All this increases storage costs on an ongoing basis, but the tools they have implemented enable their IT staff to do more with less." In fact, ITSD was able to add 100 terabytes of storage with only one additional full-time employee.

Creating storage policies

StatsCan's next step is moving data into the three-tier storage system. "Phase one was installation of CommandCentral to understand what our holdings are," Charron says.

Determining storage policies has been hard. At issue is how long each type of data should reside in Tier-1 storage, when it should be moved to Tier 2, when it should be moved to Tier 3, and whether and when it should be deleted. Because StatsCan is a government agency, it must also consider legal restrictions on the confidentiality of data and how long it should be stored.

Once policies are determined, TeraMach will work with StatsCan to install Veritas Enterprise Vault software for storing unstructured data such as email so it can be moved to Tier-2 and Tier-3 storage. Charron and his team plan to recommend moving email from Tier 1 to Tier 2 if it hasn't been accessed in 60 days, and to tier 3 if it hasn't been accessed in 120 days. "We did a data classification exercise in March of 2005," Charron says. "We found that 62 percent of all files over eight weeks old had never been accessed or modified again."

The process may be gradual; a recent Symantec report shows that 75 to 80 percent of StatsCan's Microsoft Exchange files have not been accessed in more than a year.

To keep the migration manageable, ITSD may begin by transferring two-year-old data, then one-year-old, and so on, working its way down to the 60/120-day threshold.

ITSD is establishing policies by working with committees that include representatives from the agency's fields. "We should see some progress in the next 12 to 18 months," Charron says.

Ultimately, all of these projects are part of a master plan. "It's our quest to implement information life cycle management from cradle to grave," Charron says. "One of our starting steps was to consolidate storage, and that's nearly done. We've addressed the infrastructure layer and the operational layer, with regards to centralized storage and backups. Now we're addressing the management layer and the policy layer as we classify data and prepare to implement the data movers to move data from Tier 1 to Tier 2 and Tier 3. Symantec will play a big role in that."

According to Clarke, the same information life cycle strategy and use of the same technologies could work equally well for any large organization. "One of the lessons from this project is that backup and storage are areas where corporations or government agencies can lower their total cost-of-ownership. These changes are not that difficult to implement. There are pretty great savings that can be achieved in a short period of time by focusing on the right areas and using the right resources to implement them. There isn't a single customer that couldn't benefit from using the technologies Statistics Canada used." ■

Minda Zetlin is a business technology writer in Woodstock, New York, and a contributor to Computerworld and EContent.



Read Online:
www.symantec.com/ciodigest/statscan