IBM Systems magazine

Martin Kennedy, Citigroup Inc.'s managing director of enterprise-systems infrastructure, says the mainframe provides the highest levels of security.

A $ound Investment

Citigroup tames its backup environment with dedicated mainframes

By Jim Utsler

Whether for internal or external reasons, most organizations require data be backed up and available for disaster-recovery or business-continuity purposes. In particular, financial institutions are often held to a higher standard because of the sensitive nature of the data they handle.

Although most customers "don't see the detailed processing that happens in the background, the fact that we can tell them that their data is being managed with the highest levels of integrity, availability and security is a big plus," says Martin Kennedy, Citigroup Inc.'s managing director of enterprise-systems infrastructure.

That's why Citi created a backup environment many would consider best-in-class, using four IBM System z10* Business Class (BC) machines as dedicated mechanisms for backing up the company's myriad servers and systems. Combined with INNOVATION Data Processing's FDR/UPSTREAM and UPSTREAM/SOS, their design not only ensures proper backups, but also optimizes CPU usage and network bandwidth.

"The traditional backup model limits backups to an off-hour window because of the high CPU and bandwidth requirements," says Kennedy. "With our model—with the data moving through a dedicated SAN [storage area network] fabric and not over TCP/IP—you can pretty much schedule backups whenever you want without impacting production applications and workflow."

PHotography by Matt Greenslade

A Cost-Effective Mechanism

Citi is a global provider of consumer, corporate and investment-banking products and services. Its many brands include Citibank, CitiMortgage, Citi Cards and Women & Co. Citi operates in more than 160 countries and handles around 200 million customer accounts.

With such a large presence, you'd expect it to have a large IT infrastructure, and it does. It runs servers of all types, including IBM mainframes and Power Systems*, Linux* (stand-alone on System z*), Windows* and UNIX* OS-based machines. These boxes handle core functions: serving customer accounts, running business-intelligence applications or supporting internal back-office accounting functions. They all produce a huge amount of data, most of which is backed up, stored and ready for recovery for regulatory or internal purposes.

This data is vital to Citi's everyday operations. That's why the company put so much effort into creating a seamless data-retention model: from point-of-origin to vaulted storage. But things weren't always so efficient, as Kennedy explains.

"Back in 1996, we were dealing with our first large data warehouse with around 20 terabytes of data, which was large for that time," he says. "We would have to break backups into 20 separate components, 20 separate servers and 20 separate tape libraries. We only had one window in which to do that, typically on Sunday afternoons: the only time we could take a machine down for system backup. We struggled with that quite a bit."

That's probably an understatement given the scale of the organization and the growth in the amount of data it was collecting. Network bandwidth, for example, became an issue, forcing Citi's IT staff to tune the network continuously to accommodate increased traffic. This affected everyday operations as data movement competed with typical application transactions.

"We revisited our backup architecture," Kennedy recalls. "As a result, we considered using INNOVATION tools, including FDR/UPSTREAM and UPSTREAM/SOS, both of which would allow us to move data in a more robust environment. No sooner had we converted to EMC and INNOVATION than our issues related to backups almost literally disappeared."

These improvements resulted in part because a SAN fabric and transfer devices had been introduced into the storage environment. Using this SAN, Citi IT staff could bypass the production network and run backups whenever they were scheduled, without worrying about a network bottleneck and reduced application response times. This also held true with its Linux on System z instances, with UPSTREAM/SOS passing these mainframe-based backups over System z FICON* and Fibre Channel Protocol (FCP) channels rather than over the network.

Ordinarily FDR/UPSTREAM pushes distributed backups to mainframe-based tape or DASD, but

UP CLOSE

CUSTOMER: Citigroup Inc.

Headquarters: New York City

Business: Global provider of financial products and services for consumers, corporations, governments and institutions

Challenge: Improving data-backup processing

Solution: Using four IBM System z10 Business Class machines as dedicated backup servers and INNOVATION Data Processing's FDR/UPSTREAM and FDRSOS to automate backups

Hardware: Four IBM System z10 Business Class servers

Software: INNOVATION Data Processing's FDR/UPSTREAM and FDRSOS

they allowed Citi to use LPARs on its mainframe systems to channel backups. This meant the organization's previously distributed backup model for its largest and most complex servers could be consolidated to fewer systems and therefore fewer backup media.

"This is a very cost-effective backup mechanism," Kennedy says. "We were able to leverage the mainframe tape infrastructure, which allowed us to make the most of our cartridges by filling them with data that had like retention requirements." That can be difficult to do in a distributed backup environment, where overall cartridge utilization of unexpired data tended to be lower. Managing cartridge utilization is key to managing the overall cost of a backup solution.

> "IBM has some **strong algorithms** built into the machines, as well as crypto engines and certificate management."
>
> —Martin Kennedy
> Managing director of enterprise-systems infrastructure, Citigroup Inc.

An Important Characteristic

Other organizations may have been satisfied with these improvements, but not Citi. When IBM announced the System z10 BC machines, Citi decided to further bolster its backup model by dedicating these boxes—one located at each of the corporation's primary North American data centers—purely for backup purposes. This would allow the organization to move all of its backup loads for its largest servers off its production systems.

Citi installed one z10 BC at each of its main North American data centers, with the boxes supporting the local backup requirements for large servers at these data centers. When a group wants to add a server to the backup schedule, it fills out a change-management template that identifies what it wants backed up, how often they want it backed up and what the retention period is for the backup. At that point, someone in the distributed backup team creates an FDR/UPSTREAM job and schedules it according to user requirements.

Initial backups are of the full system. Following that, daily incremental backups take place, with FDR/UPSTREAM looking to copy only the files that have changed since the previous backup. Once a week, merged backups combine the incrementals to create a new, full-system weekly backup.

"Those weekly backups can take place outside the typical window for server backups, and we don't have to be connected to the server to do that," Kennedy says. "This is in contrast to many backup models, where you have to be connected to the server. Now we can drive those business class machines 24 hours a day with productive work without impacting the servers that are being managed by those machines. It's all very efficient."

Citi can now also aggregate backups based on retention. In the past, different backups with different retention rates were stacked on top of each other. Sometimes tapes were kept in storage because, even though one server's retention-based backup may have expired, others hadn't.

"You might end up with a situation where 60 or 70 percent of the data on a tape might be expired, but you couldn't reuse that tape because the other 30 or 40 percent had not yet expired," Kennedy says. "You ended up with tape capacity you couldn't reclaim. Since FDR/UPSTREAM uses more traditional mainframe tape-management mechanisms, we can now stack data so when data on the tape has expired, the tape is immediately available for reuse."

This data is encrypted as it's written to tape, some 650 TB a month, according to Citi's Jim Jurasin, vice president of enterprise systems infrastructure. The encryption process occurs on the organization's IBM tape drives instead of on the System z10 BC machines. This reduces CPU overhead, which allows the mainframes to run their jobs without taking performance hits. Citi has also implemented disk encryption on the backup servers "to ensure the environment is completely secured from the time the data goes to disk and then to tape," Kennedy adds.

"That's an important characteristic for us," he continues. "We have a lot of very sensitive data around here, so it's critical that we make it as secure as possible, and IBM has some strong algorithms built into the machines, as well as crypto engines and certificate management."

Exceeding Expectations

Citi's customers probably don't think about this type of back-end processing, but if they did, they'd be impressed at the measures the corporation has taken to make sure their data is backed up, safe and secure. Additionally, this streamlined backup model has also benefitted the corporation itself. Backups that run on the BC machines no longer bog down the corporation's traditional networks, servers don't have to be taken offline when they're being backed up, backup media is being used more efficiently and automated processes have reduced backup administration. In the end, it's all about Citi's customers and their service expectations—which are now more than being met. Z

Jim Utsler, *IBM Systems Magazine* senior writer, has been covering technology for nearly 20 years.