

Incremental Advances in Backup

By Ira Goodman

In my last column, I addressed the new snapshot backup and restore technology, and in that context, glancingly mentioned the concept of "incrementals forever" backups. This is another emerging technology, which I think holds a lot of promise.

The last time I checked, the day still only had 24 hours. Yet the productivity expectations of employers and CIOs for their staffs have ballooned. Staffing cuts the last number of years means that there are now fewer tech people available to meet these higher productivity standards. Similarly, cheap disk has enabled a higher tolerance for sloppy, or no, housecleaning in data storage. There is no time for or expectation of daily cleanup. This means that files can just sit in storage, taking up room for days, months or years, without changing or being accessed, or even being needed any more. But no one knows. This causes data volume to be backed up enormously. One shop I talked to had a backup of 100 GB a day five years ago, and is now faced with 949 GB a day, soon to hit a terabyte. This isn't even possible with most backup windows, so backup becomes a real issue. We could take this in the direction of Hierarchical storage management (or HSM, as it is referred to), and that is one way to deal with this issue. However, that takes time, planning and personnel—what we've just said we were short of. So, finding different ways of managing backups is another solution. Many shops turn to traditional incremental or differential backups in this situation.

Most Sys Admins are familiar with the concept of traditional incremental backups, along with differential backups. Incremental backups backup all the information that has changed since the last full or incremental backup, which saves time and space on the backup side, especially when multiple incrementals are done in between each full backup. However, restores may require restoring multiple incrementals as well as the base, and can be quite slow. Differentials on the other hand, backup all the data that has changed since the last full backup, and if multiple differential backups are performed in between full backups, some data is going to be copied multiple times. Differentials take longer and use more space on backup than incrementals, but are faster to restore, because only the base and the relevant single differential are needed. Both of these can be done to disk or to tape.

However, the new exciting technology using block-level incrementals in the "incrementals forever" strategy I mentioned only works with backup to disk; tape can't spool fast enough, and specific disk hardware is necessary. With this technology, only the blocks of data that have changed since the last base or incremental backup are backed up. The advantage of this is immediately apparent—it's extremely fast and takes up very little room on the disk. However, the true beauty of the

method is that if it is necessary to restore the data, the restore is just as fast as if an original file level base backup were being used.

Thanks to special management software with the hardware, each block-level incremental backup is synthesized into a "virtual full backup" at the time of backup. This doesn't take additional time or space, since it is performed outside of the backup itself. However, when the time comes to restore, it is just as fast as a traditional real full backup. What a bonus! And, if a full backup isn't needed, a logical or file level restore is also possible.

Think of something like an Exchange Server, on which all the data backs up as one, *enormous*, file. That means that every backup with traditional incrementals must backup the whole thing even if only 100 out of 100 million emails have changed—and of course, most of the data won't have changed. So that is a lot of duplicate data being repetitively backed up—taking up space and even more important, taking up time. With a block-level incremental backup, only those blocks that have changed are backed up in each instance saving probably 99% of the volume and time needed, depending on backup frequency.

With backups that are this fast and so low volume, the backup window has just shrunk drastically. So a company has the potential to do backups once an hour instead of nightly. Think of how much less would be lost in the event of a disaster if the users only lost the last 45 minutes of work, rather than six or so hours of effort. So the loss window has just closed up enormously.

Further, with this method, since each backup instance is a virtual full backup, no base backup is needed beyond the first one; just incrementals for as long as is needed—hence the "forever" in the incrementals forever of the concept. Now companies may want to go in and do base backups again, maybe monthly, maybe yearly, whatever, but it'll be their call, and it won't be necessary. In the meantime, they'll have fast, efficient, low-volume backup and equally fast and efficient restore in the always unwished-for event of a disaster forcing recovery.

So the combination of the block-level backup speed and the ability to synthesize the full backup instance for restore seems a sure winner to me—more than incremental advances, so to speak! It hasn't spread widely yet in the industry, but keep an eye out, I think the vendors offering it will be increasing. 🌀

Ira Goodman is software services manager for Syncsort, Inc. He has 26 years of experience in information processing and has managed backup product support for more than a decade.