A step back in time with Windows 8's File History

Microsoft reinvents Time Machine for the second time.

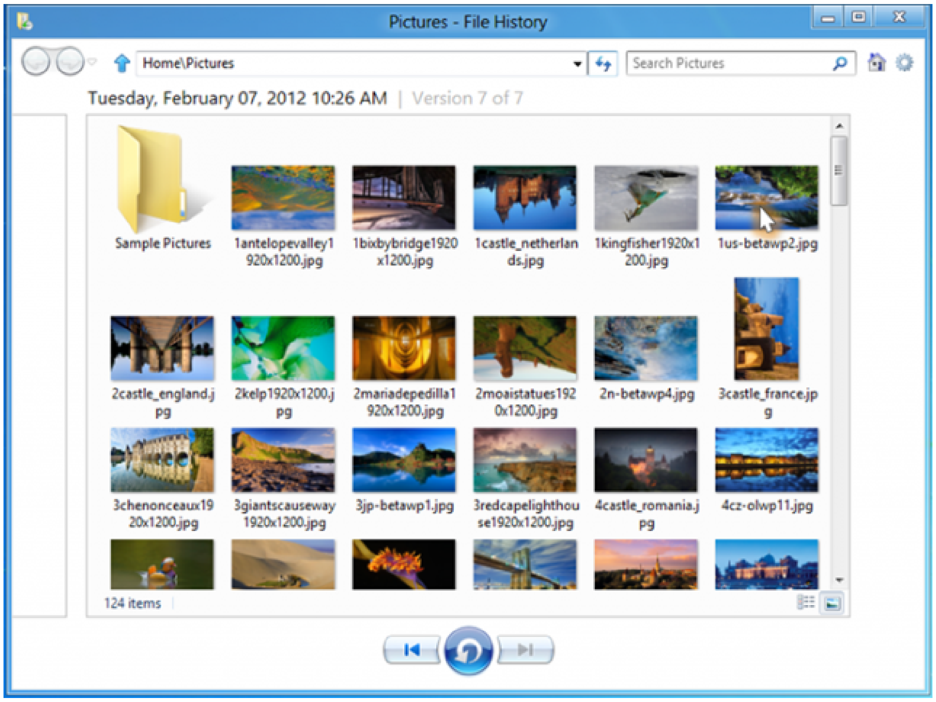
by **Peter Bright** - July 10 2012, 11:00pm EDT

WINDOWS 128

Backing up your data is an important task that most of us neglect to do. Windows has included backup software of some kind for a long time now, but few people actually use it, because they forget, don't understand it, don't know it's there, or simply can't be bothered.

Microsoft's latest attempt to get Windows users to back up their files is Windows 8's File History. File History is an automatic point-in-time backup system that periodically saves snapshots of your data to a separate location (either a network file share or a directly attached hard disk).

Every hour, by default, any modified data files get safely archived away, and there's a reasonably simple user interface to browse all the different versions of a file or folder that the system has stored, and from there they can be opened or restored.



Click back and forward to browse through snapshots.

Microsoft

If this backup concept sounds familiar, that's not entirely surprising. Apple's Time Machine backup system works in an almost identical manner, the only major difference being that Time Machine has a rather more whimsical user interface for restoring files.

What may surprise people, however, is that Windows 8 is not the first Windows version to include a file history feature. Since Windows Server 2003, Windows has had the ability to automatically store historic file versions in a feature known as Shadow Copies. Windows Vista (Business, Enterprise, and Ultimate editions) and Windows 7 (all versions) both include the same capability, calling it Previous Versions.

LATEST FEATURE STORY

FEATURE STORY (3 PAGES)

In his words: How a whitehat hacked a university and became an FBI target

David Helkowski set out to be a whistle-blower; he now faces the feds and unemployment.

WATCH ARS VIDEO

NASA's 747 shuttle carrier on the move

Disassembled and spread out across a thousand foot convoy, one of NASA's Shuttle Carrier Aircraft makes its final journey.

STAY IN THE KNOW WITH

This makes File History a new version of an old feature; it's being done in a new way, using old technology. Is this the solution to all our backup woes?

A trip down memory lane

File History (and Time Machine) work in a very different way to Previous Versions/Shadow Copies, and this has both good and bad aspects. Microsoft has blogged about File History, and how it works, paying particular attention to some of its performance characteristics.

Previous Versions depended on a file system feature called the Volume Snapshot Service (VSS). VSS was designed to solve a handful of common problems faced by backup applications: how do you back up a file that an application has open and is actively modifying, and how do you copy files that have been locked open? To ensure that the backed up file is actually useful, the backup software has to guarantee that it never makes a copy in the "middle" of a write. The same problem also occurs across multiple files: if two files have to be updated to save a piece of information, the backup software must be sure that it records *both* updated files.

VSS solves the problem by letting the file system make static snapshots that freeze the file system at a given point in time. This allows multiple "views" of the file system: the current, live, modifiable view that regular applications see and use, and the frozen snapshots that backup software can safely copy. In tandem with this snapshot capability are hooks to let applications know when a snapshot is being made. Software such as SQL Server and Exchange use these hooks to solve the multiple file problem: they make sure that all the different files that make up their databases are updated and consistent before the snapshot can be made.

The locking issue is solved by dint of the fact that applications only ever open (and lock) "live" files; the frozen versions in the snapshot aren't opened and hence aren't locked.

VSS is integrated deep into the file system, and works at the block level. When tracking space on a physical disk, file systems generally divide that space into small units called blocks or clusters (on NTFS, usually 4096 bytes per block) to make the bookkeeping easier. When snapshots aren't being used, modifications to files are simply made in-place: the blocks are directly changed. When volume snapshots exist, modifications to files are no longer made in-place. Instead, the original blocks are saved safely away, and a *new* block is used as the target of the write.

This system is quite efficient, because the snapshots only need to store the blocks that change. A small write to the end of a large file, for example, will only change the blocks at the end of the file.

In the first version of VSS, which made its debut in Windows XP, the snapshots were all temporary. Reboot the computer and they'd disappear. This was fine for their original purpose, since you normally wouldn't want to make a backup across a system reboot. Shadow Copies, in Windows Server 2003, extended VSS to allow persistent snapshots to be made. They're the same as the temporary snapshots in every way, except one: they don't disappear when you reboot.

Shadow Copies would take a snapshot on a periodic basis; typically about twice a day. Each snapshot would show the entire filesystem as it existed at that point in time, and the system would carefully preserve any blocks that get modified.

LATEST NEWS

DIGITAL RESURRECTED THE VINYL STAR

Largest vinyl record pressing plant in the US is expanding

#SADSTOCK

After lockups expire, Twitter stock price tumbles 18 percent in a day

NO COAL IN THESE STOCKINGS

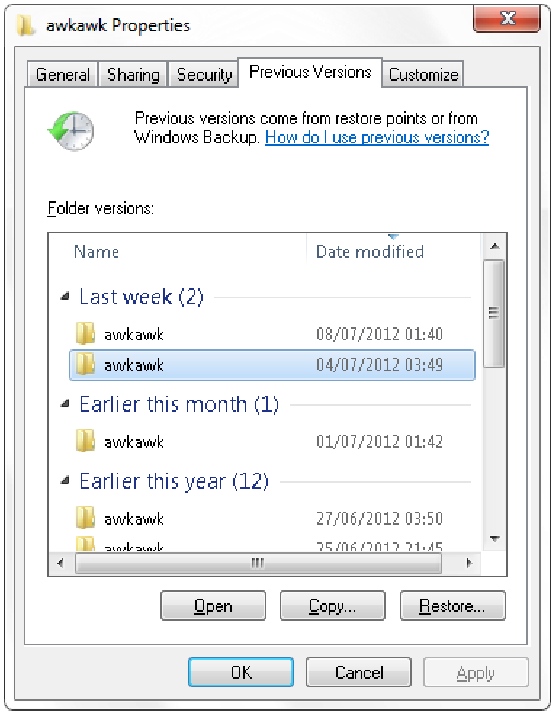
Stanford to divest its endowment from coal stocks

BILLION DOLLAR TICKET

Chinese e-commerce giant Alibaba files for $1 billion IPO in the US

White House's sprawling climate report: Change is already here

House panel voting to end NSA bulk phone metadata program



Previous versions come from restore points or from Windows Backup. How do I use previous versions?

Folder versions:

Name	Date modified
Last week (2)	
awkawk	08/07/2012 01:40
awkawk	04/07/2012 03:49
Earlier this month (1)	
awkawk	01/07/2012 01:42
Earlier this year (12)	
awkawk	27/06/2012 03:50
awkawk	25/06/2012 21:45

A user interface that's annoying to use, and that most people don't even realize exists

The technology behind VSS was all quite clever. But it was let down by its user interface, and this is the main reason that most people have never heard of it. Tucked away in Explorer's properties for files and folders is a tab called Previous Versions. The cramped dialog box shows the various snapshot copies that exist of a file or folder, and they can be opened or restored. While the user interface does work, it's not at all user friendly and it's awkward to use.

Back to the future

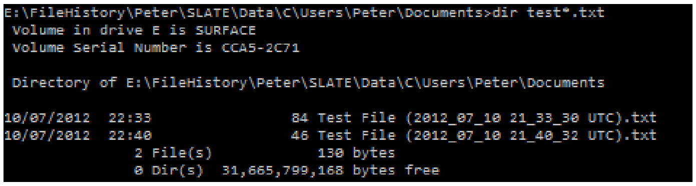
File History leverages an even older file system feature, called both the USN Journal (USN standing for Update Sequence Number) or the Change Journal (they mean the same thing, but Microsoft uses both terms in different documents). The USN Journal was introduced in Windows 2000, and it stores a very simple record of every change made to the file system. Specifically, whenever a file is created, modified, deleted, or renamed (or has certain other special things done to it), an entry is made in the USN Journal to record that the specific operation was made to the file. Each entry is given a number, the Update Sequence Number, hence the name.

The USN Journal doesn't include information about what was actually *changed*; it will note that a file was written to, for example, but doesn't specify what the new data is. It just tracks the list of changes.

The USN Journal was introduced to allow applications such as virus scanners and content indexers to wake up periodically and see in a quick and efficient way all the changes that had been made to a file system since they last ran. A content indexer, for example, might wake up every hour and examine the USN Journal. If the USN Journal has had entries added since the indexer last ran, the indexer can read each entry to see if any files have been modified. If they have, the indexer can then reindex them as necessary.

It's easy to see how this is useful for something like File History. It allows the File History service to lie dormant in the background (Microsoft says it uses less than 10MB RAM when inactive), waking occasionally to check the USN Journal. If the USN Journal says there are changes, File History can quickly see if any of the files that it has backed up are among the changes, and if so, make a new backup copy.

Unlike Shadow Copies and their block-level operation, File History works with whole files. You nominate a disk or network share to be the File History repository, and it creates a replica of all your backed-up locations. Each time a new version of a file is copied to the replica, it gets renamed to include a timestamp in its name. This allows multiple versions of the same file to coexist within the same folder. Files are stored whole, so a minor change to a large file will result in a lot of disk usage.



```
E:\FileHistory\Peter\SLATE\Data\C\Users\Peter\Documents>dir test*.txt
 Volume in drive E is SURFACE
 Volume Serial Number is CCA5-2C71

 Directory of E:\FileHistory\Peter\SLATE\Data\C\Users\Peter\Documents

10/07/2012  22:33                84 Test File (2012_07_10 21_33_30 UTC).txt
10/07/2012  22:40                46 Test File (2012_07_10 21_40_32 UTC).txt
               2 File(s)            130 bytes
               0 Dir(s)  31,665,799,168 bytes free
```

Two versions of a file named Test File.txt showing how the filename gets modified to incorporate timestamps

Alongside these copies, a database is maintained that records which files were visible at each given point in time. It also maintains information for files which can't be stored in the replicated directory tree, such as those with very long filenames. Apart from that, the system seems very naive: if you rename a file, it gets detached from its history (the history still exists, under the old name, but is no longer shown as an older version of the renamed file). Conversely, if you delete a file and then at a later date create a new file with the same name as the deleted file, the histories are joined, as if the newer file were a later version of the older one.

Two implementations, one OS

File History and Previous Versions use very different technology to achieve a very similar goal. So why has Microsoft thrown out the old system and reinvented Apple's Time Machine?

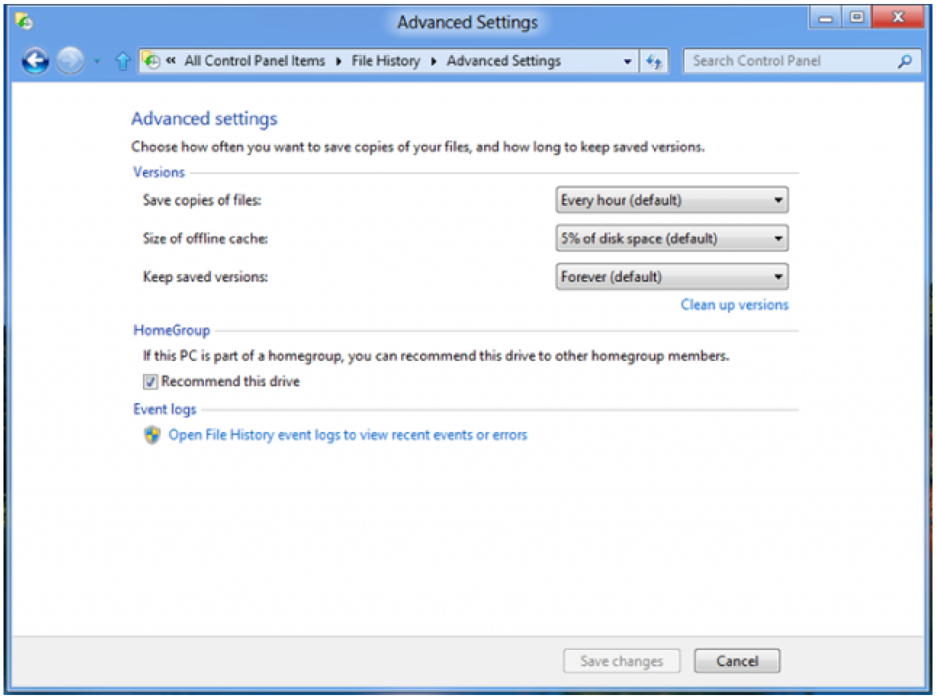
The biggest single difference between the systems is where they store the information. For Shadow Copies, the old snapshot blocks are almost always stored on the same volume as the current, up-to-date file. *Technically*, it is possible to store them on different volumes, but in practice, this is rarely done (and, in any case, was only ever possible on server versions of Windows).

This has two important repercussions, one very good, one very bad. The good repercussion is that the system is self-contained. The only thing you need is a hard disk; it can contain *its own* historic versions. The bad repercussion is that it's self-contained: if that hard disk should stop working, you lose both the current files and all the previous versions too.

File History, on the other hand, *never* uses the same disk to store historic versions—it requires the use of either a network share or a separate physical disk. (Again, technically that's not quite true, as you can use a file share hosted on the same physical disk as the one being backed up, but this defeats the purpose). This means that if the disk should fail, the backups should all be safe. You can recover the backups simply by installing the operating system to a new hard disk, giving the system the same name as the broken one, and using the same location for File History; the history will pick up exactly where it left off.

One consequence of this is that the File History storage location might not be available—if you use network storage on a laptop computer, for example. To cope with that, the system can additionally use a portion of the local disk as a cache into which it will make replicas whenever the history location is unavailable. As soon as the history becomes available (so as soon as you return to your home network or plug in your history USB disk) the cache will be copied to the history location.

File History only tracks files in certain locations; Libraries, the Desktop, and a couple of other places. Shadow Copies, in contrast, track almost the entire disk. This means that if you keep your files in an unusual location, File History won't protect them. While files in the protected locations can be excluded (to, for example, avoid burning lots of space on podcasts or other readily re-downloaded data), there's no provision to *include* extra locations.


Advanced Settings
« All Control Panel Items ▸ File History ▸ Advanced Settings
Search Control Panel
Save changes
Cancel

Advanced settings

Choose how often you want to save copies of your files, and how long to keep saved versions.

Versions

Save copies of files: Every hour (default)

Size of offline cache: 5% of disk space (default)

Keep saved versions: Forever (default)

Clean up versions

HomeGroup

If this PC is part of a homegroup, you can recommend this drive to other homegroup members.

☑ Recommend this drive

Event logs

Open File History event logs to view recent events or errors

Scant configuration options

Microsoft

This could be a big problem for some people. Outlook, for example, defaults to saving PSTs in a location that File History doesn't track, so if you're using a PST to archive mail, you'd better make sure to move it. Similarly, I keep my source code in a separate folder that is not part of any of my Libraries so that I can avoid filenames with spaces (as to this day they prove an inconvenience for many software build scripts). Again, File History can't protect these files.

Update: It appears that the Outlook behavior is a consequence of my Outlook install being an upgrade from Office 2007. A fresh Outlook 2010 installation will put PSTs into the Documents Library. Still, many applications continue to put important information in places that aren't included in Libraries.

There is a workaround, as you can create custom libraries that include all the random locations where important files might be kept. But this means creating pointless libraries whose sole purpose is to make the backup tool do its job, which greatly undermines the simplicity and elegance that File History should have.

Time Machine, by way of comparison, tracks whole disks.

File History can store more snapshots, and can make them more frequently, than Shadow Copies. The frequency at which it copies your files can be varied, from once every ten minutes, to once a day, with numerous intervals in between. It can also keep historic versions "forever," or at least until the volume used to store them is full. In comparison, you can only create 512 concurrent snapshots of a single volume.

Shadow Copies work "automatically" insofar as once a snapshot has been created, the file system itself will always maintain the integrity of that snapshot (until and unless it needs to reclaim the space used by the snapshot). File History relies on a service to work. If the service is stopped or broken, no backups are made. More vexingly, the service is designed to pause if it detects that the system is busy, meaning that it will defer making backups when the system is under load.

Whether File History is better or worse than Shadow Copies is going to depend on how exactly you use your system. The ability to create networked backups is obviously very valuable, at least for systems that can rely on the presence of a suitable network connection. For systems that can't rely on network storage, the situation is less satisfactory. Plugging in a USB disk does work, but it's not especially elegant, and Shadow Copies feel like a better fit for that scenario. File History's inability to track arbitrary locations is also a significant flaw for some usage patterns.

On the other hand, the "infinite" history that File History makes possible is highly desirable. In particular, it plays very nicely with Storage Spaces, as they provide an easy facility to create near-infinite volumes to store backups on.

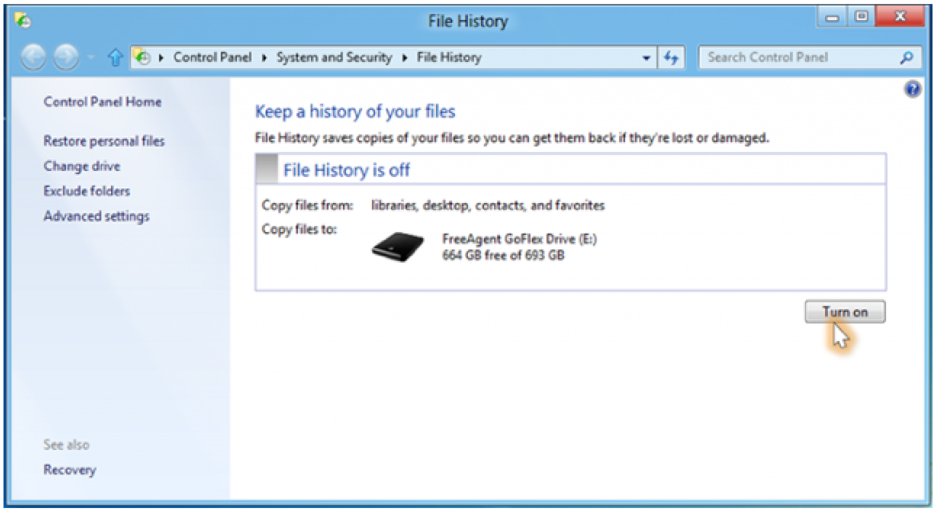
Unfortunately, even if Shadow Copies are a better fit for how you use your system, they're not an option in Windows 8. Citing unspecified "performance issues," the ability to make persistent snapshots has been removed from Windows 8 and Windows Server 2012 (though temporary snapshots, for use by backup software, remain supported). The performance of Shadow Copies has been good enough for the past 9 years, since they were introduced with Windows Server 2003, so it's a little difficult to see where the performance problem is, but the feature is now gone.

Update: The wording on MSDN that said that persistent snapshots have been removed from Windows 8 and Windows Server "8" (as it was once known) appears to have been removed. This is encouraging, as it implies that the underlying capability will be retained, even if the user interface is no longer available. Not a perfect solution, but better than outright removal.

As someone who used Shadow Copies regularly, and whose usage pattern doesn't fit well with File History, this is a great pity. The old system image-based backup tool is still a part of Windows 8, so there is *some* alternative for usage patterns that don't play well with File History, but it's not ideal.

The weakest link

Ultimately, the real challenge for any backup system isn't technological; it's human. File History is off by default (as it rather has to be, since it cannot rely on the presence of a network location or separate hard disk), and so will do nothing for most users, most of the time. For those who are a good fit for its capabilities, and who choose to enable it, it will serve as an effective and easy-to-use backup system.



Keep a history of your files

File History saves copies of your files so you can get them back if they're lost or damaged.

File History is off

Copy files from: libraries, desktop, contacts, and favorites

Copy files to: FreeAgent GoFlex Drive (E:)
664 GB free of 693 GB

While easy to enable, File History is off by default.

Microsoft

Nonetheless, I find it difficult to regard it as a step forward. A whole-disk system, like Shadow Copies and Time Machine, and a system that is at least optionally self-contained, as in Shadow Copies, can do more things for more people, and I suspect that most Windows 8 users, like most Windows 7 users before them, will end up with no backups at all.

Listing image by Microsoft

READER COMMENTS 128

Peter Bright / Peter is Technology Editor at Ars. He covers Microsoft, programming and software development, Web technology and browsers, and security. He is based in Houston, TX.

Follow @drpizza

← OLDER STORY NEWER STORY →

YOU MAY ALSO LIKE

SITE LINKS

About Us
Advertise with us
Contact Us
Reprints

SUBSCRIPTIONS

Subscribe to Ars

MORE READING

RSS Feeds
Newsletters

CONDE NAST SITES

Reddit
Wired
Vanity Fair
Style
Details

Visit our sister sites ▼

Subscribe to a magazine ▼

VIEW MOBILE SITE