

## Improvements to Backup Applications Enabled By Expedite

January 2014

The Information Asset Management functionality provided by Expedite benefits many systems within computer architectures, including backup systems. The following list are some of the ways backup could be improved, corrected, or expanded if it was integrated with an Expedite system.

1. The first thing Expedite does is improve the integrity of the files that are being backed up through its extensive information assurance capabilities. Important files can't simply be accidentally deleted unknowingly and anonymously. The integrity of the backup data is only as good as the integrity of the files being backed up. With Expedite, the integrity of the file system is now translated to the backup copies.
2. Because of the improved integrity of the file system, backup can now, for the first time, actually validate the integrity of its copies. It can check that 1) all the files on the file system exist on the backup, 2) that the recorded file signatures match, 3) can recomputed the file signatures and make sure they match, or 4) can actually validate the contents of the files for consistency.
3. The validation process outlined above can be a time consuming task. These can be run based upon the asset types, business requirements of the data, etc. For example, copies of financial records can be checked against content, not the entire backup set.
4. Most backup systems are forced to do a directory structure traversal in order to identify which files have changed since the last backup. The storage subsystem can be excessively strained running this scan impacting overall system performance and the user experience. With Expedite, the list of files that have changed could be a simple, fast, and efficient database query greatly speeding up the backup process.
5. More efficiency in scanning for files that have changed means less of the backup window is used for scanning allowing more time for data transfers.
6. Since file systems don't know anything about the importance of the data stored, backup systems are forced to treat every file with the same level of importance even though some files are clearly more important for the company than others. Expedite can now provide information to backup systems they can use to know which files need to be backed up before others. This is very important for things like disaster recovery where important files should be restored first.
7. One of the more frustrating aspects of backup is the inability of backup to know when it should run a restore. Every backup solution requires a human to "notice" a problem and manually initiate the restore process. This is pretty easy if an entire file system disappears due to a hardware failure. However, if a file was accidently deleted or, harder yet, corrupted, it is quite likely a human won't notice it. Backup has no option but to keep making copies of it (or not if it was deleted!). Expedite can actually detect this condition and put the file back from its own mirrored copy. It would make sense to actually get it back from the backup copy.
8. When a file is deleted or corrupted, if it is not restored quickly enough, every backup system will eventually roll the media or available backup capacity and throw away the last good copy of a file. If that file is needed after it is rolled off, backup gets a bad reputation for its inability to recover the file. With Expedite, the file won't disappear in the first place so backup copies of the correct data will continue to be saved.
9. Because the backup process is so inefficient, backup execution is often limited to times where the systems are not used. This creates a widow where, should data be lost, it won't be saved on

the backup. With Expedite, the option to run backup on just certain key data is now possible. The decision to do this and when it should be done can be controlled based upon the business requirements of the information.

10. Because Expedite is controlling the processes, any backup process could be synchronized so the file set being backed up could be in a consistent state. This is similar to the database backup problem but for files. If there is a case where more than one file needs to be changed as a set, Expedite can ensure the set is backed up in a stable and consistent state.
11. Expedite is controlling the modification process of the information assets it manages. The most important time to back up a file is when its business context changes. Doing backups only on a context change can greatly reduce the number of backups and these match what the user needs should the information need to be restored.
12. Expedite implements the concept of Master Copy Management. This means that a master copy of the information asset is maintained and protected. Any copies of the information are made to support the needs of the technology, business process requirements, or specific for users. Expedite controls the creation, location, repair, and rescinding of these copies automatically. Backup systems could use this information to prevent these extra copies from being included in the backup process reducing the quantity of data copied during the backup window.
13. Expedite can support the idea of an Asset Mirror. If configured, Expedite will keep a separate copy of the master copy of all the assets in a container. There are many ways this could be expanded with backup systems. One is that the actual backup could be performed by accessing the mirror, not the main file system preventing adding to the load on the primary storage device.
14. Every company has data that probably doesn't need to be backed up at all. Expedite could be configured to allow backup systems to quickly query Expedite to see if the information should even be backed up. There are many reasons for this including data that has been downloaded from the internet, data that already has a second source, etc.
15. Backup systems have directory scanners or directory change notification systems that could be used for additional capabilities beyond just detecting what files have changed.
  - a. Look for rogue copies of controlled files. These are files that are already controlled in Expedite but are just copies sitting out on the user's file system or on file servers. How the system reacts to these can be specified based upon the business requirements of the asset. For example, if the file is a blank petty cash form, then it is not a big deal. However, if it is copy of confidential customer data, then notices could be made, the file removed, redacted, replaced with a link to the master copy, etc.
  - b. Can be used to identify files that should not be in certain locations such as dropbox directories, on user's systems that should not have access to the files, etc.
  - c. Could extract the data files and detect key phrases or sequences that part of a data loss prevention strategy.
16. The backup frequency of assets could be specified based upon the requirements of the information. Some information could be backed up once a week where others should be backed up the instant it is changed. Expedite can now provide ways for backup to know this.
17. Testing the backup process (actually the restore process that is so dependent on the backup process) is very difficult to perform. Many people simply don't do this until it is actually needed only to get a nasty surprise. Expedite can be used to quickly test that a file could be restored quickly. Also see the backup integrity checking outlined above.
18. Backup isn't the only independent system that does scanning of directories and files. Anti-virus, data loss prevention, full content search scanners, etc., can also perform directory scans. Having these all start at once can bring any system to its knees. The ability of Expedite to

provide backup the list of files/directories that have changed can avoid interactions with these other systems. (Note: Expedite already controls the operations of full content indexing, only triggering the indexing process when an asset's business context changes, thus avoiding the lengthy directory scans those systems usually perform)

19. Because of the ways Expedite can improve backup, some of the advantages of disk-based backup over tape-based backup could be reduced making tape more competitive.
20. Expedite maintains collections of assets with the same business content is collections called containers. Within a container, additional files such as previous revisions, obsoleted files, etc. are also saved. Backup could know this and backup the main asset differently than these other files that have a lower level of importance than the main assets.
21. What should be done is to transition backup from a file-based system to an information asset based solution. Instead of simply backing up a pile-of-files, it should backup complete assets, including the metadata, business context, etc. This would greatly enhance the ability to search for and retrieve assets should the master copies be lost. (Business users want to search on asset type, context, etc., not exact file path/filename.extension)
22. If the backup system understood information assets, a backup copy of assets could actually be a viable archive. The information in that archive would be much more useful to business users in the future than a pile of files.
23. Backup could control the number of copies of the file saved on various media based upon the business requirements of the data. Some data may be important enough to have multiple copies of it saved off should there be faults within the backup system or the media used.
24. An important part of any disaster recovery plan is the movement of the data to a different physical location. Moving every file offsite is rarely required but some information is important enough to have it copied offsite the instant its business context changes. Expedite knows this information and can trigger backup to perform the offsite copies as defined by the business requirements of the information.
25. The Expedite asset mirroring scheme requires additional storage to maintain a copy of the assets. The bandwidth requirements of this second copy is much less than the primary copy which could allow backup to be used instead. When the business context changes, instead of making a copy on another disk, backup could be triggered that would then make a copy of the file. Should Expedite automatically detect a file has been deleted or corrupted by mistake, restore could be called to return the proper copy. Such a scheme would reduce the disk requirements for the additional copy.
26. If backup is configured to access media at an offsite location, the automatic asset mirroring outlined above would provide an immediate offsite backup copy of the important assets, along with automated immediate recovery.
27. When doing offsite copies, the cost/capacity limits may mean that not every single file at a company can be saved this way. Expedite can be used to guide backup system to know which of the files should be protected this way. The different asset configurations could be setup based upon the different business requirements of the information.
28. Many backup companies also support the traditional continuous data protection schemes where every write operation is saved for later access. These systems tend to require rather significant capacity and bandwidth resources so they can't realistically be applied to all storage systems as a whole. However, there are still situations where CDP would be valuable and yet not impact overall system performance. Expedite drives any modification processes by putting copies of files into a different directory allowing modification by the user without putting the master copy at risk. This secondary directory could be configured to use CDP thus protecting all edits from getting lost. Yet, at the same time when the master copy is to be updated with the changes,

the file is removed from the directory. The CDP process can now remove all the saved updates for that file greatly reducing the capacity requirements.

29. Disaster recovery means getting the business back up and running as quickly as possible. Expedite can provide backup systems with the files that are critical to the recovery process. Backup could maintain copies of these files in different forms or locations that could quickly facilitate the recovery process (e.g. keep a standby VM ready with the files configured that could be lit upon a disaster situation). Since the amount of files that would make up such a list could be much less than the total file set, the cost of maintaining the copies could be manageable.
30. One challenge in managing backups is to ensure that when configurations are changed, added, etc. , that backup is actually configured to protect the data. It can be a nasty surprise to find out that new disk, server, storage system that was added didn't get added to the backup configuration. While Expedite can't do this for every configuration, the data Expedite does track could check to see if some type of data protection is configured and may even require something configured when created.
31. Reporting and auditing how each information asset type is protected would be much easier with a combination of backup and Expedite.
32. Should backup fail to run, Expedite could be setup to detect this and warn admins, users, etc., that the data is not currently protected by backup.
33. Many backup systems, especially the online backup solutions, support some type of file versioning. Unfortunately, the versions are created by the backup system itself and can have little to do with the changes in business context so important for business users. Expedite keeps copies of the versions based upon business context changes (e.g. which version did we send to the client?). The backup system could use this information to maintain the correct business context versions of the files.
34. The preservation market has always been challenged by the problem that it is cost prohibitive to save everything and yet currently, there is no real way to know what is important. Expedite now can drive the preservation process identifying the important information. Backup could use this to provide a cost effective preservation solution beyond simply backup.

## Conclusion

As can be seen by the list above, the integration of backup with Expedite could greatly improve backup. The backup industry has been severely limited to what it could do by the fact that file systems have not changed or improved in many decades. Backup has been at the mercy of what file systems could provide. It is time now to use the new Expedite technology to make much-needed enhancements and fixes to backup. In doing so, the value of the backup solution to a business user would be greatly improved. Backup can expand its solution scope to different solutions previously limited by the problems and limitations of the computing systems of today. A partnership with Expedite is one of the few methods backup companies can take the next valuable technological and market advances in its product offerings taking their businesses into the future.