Synthetic Backup - An Alternative Technique for Effective Data Protection

Enterprises big and small have seen their data grow exponentially over the years. Data protection is becoming very complex and requirements for availability are increasing every day. Given the fact that data protection is a primary concern for companies of all sizes, it requires significant investment (depending upon the company size) and surely is a contributing factor to the company’s profitability. With well thought planning in deploying IT infrastructure and implementing best practices for maintaining corporate data, such as Disaster Recovery (DR) [1] planning for data protection and security; companies can save significantly and improve their profit margins. Synthetic backup is one such DR best practice that provides effective data protection at a lower cost of operation. It also offers an edge over conventional file system backup techniques during system backup and restore.
What is Synthetic backup? Synthetic backup is an alternative to creating full backups from periodic incremental backups. It involves one regular full backup (only the first time) and from that point onwards periodic incremental [2] backups are applied to the first full backup. Since this type of backup is synthesized from a full backup, hence the name. Technical explanation The expression "synthetic" in this context refers to the fact that the assembled file is not a direct copy of any single current or previously created file. Instead, a synthetic file is merged or "synthesized" by a specialized application program from the original file and one or more modifications to it.
Figure 1: Conceptual representation of Synthetic versus Full backup This backup procedure is called "synthetic" because it is not a backup created from original data. Figure 1 shows the steps involved in taking a traditional full backup and a synthetic backup. Less data, which is translated in to less network traffic, will be flowing through the pipe in the case of synthetic backup. It is due to the fact that the traditional approach of taking a full volume backup requires transferring all used blocks of data from the volume to the backup location. Whereas with a synthetic backup technique in place, the system will only take incremental changes, transfer them over the network pipe and then run a special program [3] at the storage location to apply those changes to the previous full (or synthetic) backup. The bottom line is after the first full backup you need to take only incremental backups without taking a full backup every time. Each new incremental backup will be synthesized and will maintain an up to date copy of a full backup. Why Synthetic backup? Synthetic backup is not just another way of taking backups; it is a better alternative. Having one Full backup and incremental backups thereafter will not offer a similar advantage as it would by having a synthetic backup. In the former case, the system administrator will be dealing with a multitude of incremental files when it comes to restoring the system using those backups. However, in the case of synthetic backup only one file will hold all the changes since the last incremental backup. This type of backup is suitable for corporate environments where network bandwidth is a premium and it cannot be scaled easily. Business models such as Application Service Providers is a typical example, where hundreds of gigabytes of data flow through a shared pipe, mostly over T1 or T3 lines. Even with a T3 line backing up 20-30 servers with 150 GB data on average on each one of them will require serious investment to cover the bandwidth and leased line cost. Moreover, it will also require extremely large storage capacity to house all those periodic full backups.
A synthetic backup can best be used when time or system requirements do not allow for a full backup. Some of the key benefits are:
- Relatively smaller time and space requirements for taking backups – Since volume changes are periodically merged on top of existing full/synthetic backups, less storage is needed.
- Reduced restore time – System restore times are reduced, as it will only look at one unified backup file versus a combination of a full and multiple incremental backup files.
- Less expensive full backups – Guaranteed availability of full snapshot of the system in one backup file.
- ROI – Less resource hungry; less CPU power, less network traffic, less storage, resulting in more dollars saved.

- Less messy – Avoids the hassle of keeping track of multiple full backups. Featured Users Some of the typical synthetic backup users are: Application Service Providers Application Service Providers (ASPs) host software from centralized data centers, renting access over dedicated, high-speed networks or the Internet. Application services range from Enterprise Resource Planning (ERP) systems, Customer Relationship Management (CRM) solutions and vertical applications, to groupware, personal productivity suites, and full desktop and printing services. In order to ensure data protection, these solution providers are faced with the tough challenge of backing up critical data over time, using either leased lines such as T1. It is a costly operation and requires significant investment to cover leased line cost and network traffic. Improved backup practices such as synthetic backup, allow ASPs to significantly bring the cost of operation down (considering the multiplier factor) and also benefit from the time and space advantage which synthetic backup has to offer. Managed Service Providers A managed service provider (MSP), is a company that manages information technology (IT) services for other companies via the Internet. An MSP is a company that offers continuous outsourcing of an IT function. The MSP provides all technical support for this server including software updates, backups, hardware, bandwidth, firewall, and other technical issues. Given synthetic backup’s cost effectiveness and time and space optimization, the MSP business model can achieve significant gains in revenue overtime. Small and Medium size Businesses (SMB) Data is a crucial intellectual property for any company whether it is an ISV, insurance company, bank, retail, data center, software development house, ASP/MSP, etc. Synthetic backup guarantees optimal performance in terms of backup time and storage requirements. Less harsh on the production systems' resources, it can produce intermittent synthesized full backup with minimal impact on shared and un-shared applications running on those production systems, including file servers, database servers and/or print servers. Another advantage that it has to offer is that it takes less time to restore from the backup. As opposed to dealing with multitudes of incremental backup files there will be just one synthetic file which will contain all those changes recorded over time in each incremental backup.
Summary This article provides a technical overview of the file system synthetic backup technique. Think of it as a best

practice for DR, aimed at providing cost effective and robust solution for data protection. Due to its optimized time and space behavior, it can help companies with large quantities of business-critical data to effectively handle their volume backups and storage. In short, this technique will allow maintaining cost effective data storage and will provide optimal backup sets. About Sonasoft Sonasoft Corp. automates the disk-to-disk backup and recovery process for Microsoft Exchange, SQL and Windows Servers with its SonaSafe Point-Click Recovery solutions. SonaSafe is the only product that provides an integrated backup/recovery and replication solution for Exchange and SQL servers. Designed to simplify and eliminate human error in the backup and recovery process, SonaSafe solutions also centralize the management of multiple servers and provide a cost-effective turnkey disaster recovery strategy for companies of all sizes. For more information, please visit www.sonasoft.com. [1] Disaster Recovery – is an important practice in Information Lifecycle Management arena, which prepares the company in case of disaster and gives peace of mind to the stake holders. [2] It refers to the backup containing changes on the hard disk (volume) since the previous backup. [3] Typically this application runs on the same system with Network Attached Storage (NAS). Alternatively it can also run on any other machine residing in same subnet with NAS.