WHITE PAPER

Changing the SharePoint Backup Game: How to Backup Multi-Terabyte SharePoint Farms in Minutes

Trevor Hellebuyck, Chief Technology Officer, Metalogix
2/2013

Metalogix

CONTENTS

EXECUTIVE SUMMARY

SharePoint natively stores Binary Large Objects[i] (BLOBs) — the documents and other types of unstructured data that are the actual content in SharePoint — as well as their associated metadata, within the Microsoft SQL Server relational database. Since BLOBs compose up to 95% of the overall database volume and SQL doesn't efficiently process this non-relational data, they add significant overhead which impacts SQL's I/O performance, including backup operations.

Metalogix StoragePoint dramatically improves SharePoint upload, download and backup performance by relocating BLOBs from SQL to virtually any storage device without affecting SharePoint features or functionality. Supported storage technologies include Storage Area Networks[ii] (SAN), Network Addressable Storage[iii] (NAS), Content Addressable Storage[iv] (CAS), and cloud storage[v] options such as Amazon S3, Windows Azure, or EMC Atmos.

When this paper was written, about half of organizations with SharePoint were managing well over 1 terabyte of content[1] , and content was growing at about 65 percent each year. This volume already exceeds by several times Microsoft's recommended limits for backing up content databases or site collections using native backup tools. Third-party backup solutions may offer additional features, such as item-level restore. However, they will encounter the same challenges meeting an organization's Recover Point Objective (RPO) — which is the maximum tolerable period during which data might be lost — if BLOBs are stored in SQL.

Externalizing content to a BLOB Store[vi] makes it feasible to achieve short Recover Point Objectives — even for larger content databases or site collections. However, this means that BLOBs will have to be backed up using another method. A robust BLOB externalization solution should automatically back up BLOBs and provide granular restore functionality lacking in native tools. The combination of native tools and a best-of-breed BLOB externalization solution, such as Metalogix StoragePoint, enables administrators to achieve short RPOs, improve SharePoint performance and dramatically reduce storage costs, while foregoing the expense of third-party backup solutions. The remainder of this discussion is limited to backup and recovery. For more information on optimizing SharePoint storage with StoragePoint, visit www.metalogix.com.

1 Improving SharePoint Use and Governance, Osterman Research, October, 2012

OUT-OF-THE-BOX BACKUP AND RECOVERY OPTIONS

BACKUP OPTIONS

STSADM backup and SharePoint Central Admin backup are the native, no-cost tools provided with SharePoint. These tools allow administrators to backup content databases and site collections via either command line and scripting or an administrator-issued command in Central Admin. These tools are not recommended for use on[2]:

- Content databases larger than 100 gigabytes (GB).
- Site collections that are larger than 15 GB that you want to back up by using the STSADM command-line tool.

THIRD-PARTY BACKUP SOLUTIONS

With these recommendations in place, administrators often look to third-party tools to help with their database backup and recovery strategy. There are many third-party backup solutions available at considerable expense for SharePoint today. However, these tools encounter the same challenges backing up large databases and site collections if BLOBs are stored in the database.

RECOVERY OPTIONS

Hardware will fail. Users will delete items they need. When that happens, it is up to the IT staff to implement the necessary recovery options. Typically they will need to recover individual items or recover all content due to a catastrophic system failure.

SharePoint has a built-in mechanism for item-level recovery via the recycle bin feature. This allows users to recover their own deleted items from a recycle bin on a site without any administrator intervention. The recycle bin is a two-tiered system. When a user deletes a document, it is available in the site recycle bin for a period of time defined by the administrator. The default is 30 days. After the period expires the items are moved into the second tier, which is only recoverable by a site collection administrator.

SharePoint 2010 SP1 introduced a site and site collection recycle bin, which allows administrators to recover sites and site collections without restoring the entire content database. There are some important items to keep in mind when configuring your recycle bin settings:

- Items in the tier 1 recycle bin still count against a user's quota.
- Content BLOBs in the recycle bin are still stored in the content database if StoragePoint is not implemented.
- The second tier of the item recycle bin and the site recycle bin need to be considered when planning database sizes and growth.

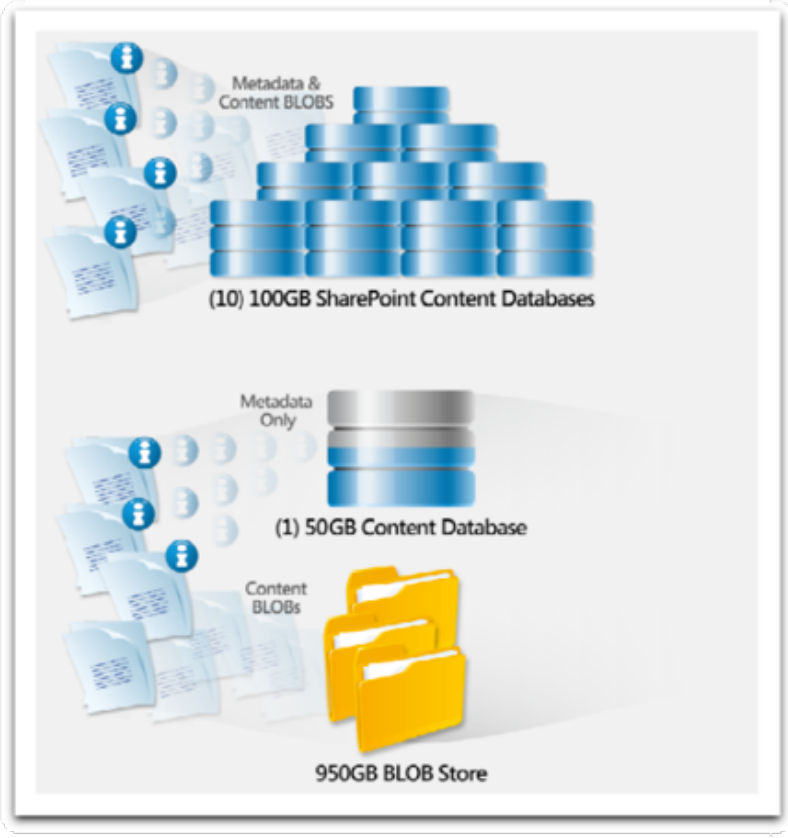
2 http://technet.microsoft.com/en-us/library/cc263441(v=office.12).aspx

These factors all increase the size of the content database. A larger content database will require a longer backup window as well as longer restoration times. Since the item-level recycle bin and the site and site collection recycle bin are set at the site collection level, administrators may be reluctant to increase the length of time items are kept in the recycle bins.

One option is to take frequent backups of a site collection using STSADM to minimize the size of the recycle bins. Unfortunately, item-level and site collection recovery beyond the recycle bin is not possible with this method. An administrator would have to restore the content by creating a recovery site collection, locate the item in the recovery site collection and manually upload it to its original document library in SharePoint.

Most third-party SharePoint backup and restore solutions offer item-level recovery without the need to stand up a SharePoint recovery farm. However, as we'll see, there are more economical alternatives to achieving item-level recovery.

STORAGEPOINT BACKUP AND RECOVERY OPTIONS



StoragePoint was designed to reduce SharePoint storage costs while improving performance. Its methodology for accomplishing this, plus additional features specifically created to enhance out-of-the-box content backup and recovery processes, enable organizations to meet RPOs in the terabyte-scale SharePoint environments that have become the norm — without investing in a separate backup and restore solution.

When StoragePoint is implemented on a SharePoint site collection or web application, the BLOBs within the content database are moved to an external storage medium. This reduces the content database size by up to 95 percent. Database backups can now be completed in minutes rather than hours. Content that is externalized can optionally be compressed as it is sent to the BLOB store. This can reduce disk space requirements between 20 and 80+ percent and lead to a significant cost savings. The types of content being externalized will determine how much you benefit from compression.

The dramatic reduction in database size makes it feasible to back up a large SharePoint environment using out-of-the-box tools without exceeding Microsoft's guidelines for backup using these tools. StoragePoint also works with most third-party backup tools.

AUTOMATIC BLOB BACKUP

Of course, externalizing BLOBs means that this content must be backed using some means other than the native SharePoint tools. The SharePoint backup capabilities in Central Administration as well as database backups will not capture externalized BLOBs as part of the backup process. StoragePoint closes that gap with a BLOB backup capability that automatically and continuously creates copies of BLOBs after they are written to their primary BLOB Store(s). The proactive creation of backup copies of BLOBs allows for a very low Recover Point Objective (RPO) because recurring backup processes need only backup SharePoint content databases. Organizations can use OOTB backup tools in conjunction with the StoragePoint BLOB backup features to provide a low cost, low RPO backup solution for SharePoint content.

(Note: STSADM and PowerShell cmdlets that backup site collections will retrieve externalized BLOBs as part of the backup process. This means that site collection backups created with these tools will contain BLOBs that were previously externalized, eliminating the need to enable StoragePoint BLOB backup for this scenario. Also, SQL database backups have a differential backup option which can shorten backup windows between full backups. However, when recovering, differential backups must be manually recombined to reach the needed recovery point, which can increase the Recovery Time Objective (RTO) beyond acceptable limits. Many administrators will want to take full database backups at least weekly.)

EXTENDING ITEM-LEVEL RECOVERY WITH SELECTIVE RESTORE MANAGER

StoragePoint completes the native tool backup and restore option by providing granular restore capabilities which, unlike native tools, do not require creation of a recovery farm. StoragePoint's Selective Restore Manager picks up where the recycle bin leaves off. It can restore anything from single items to entire site collections from SQL Server backup files or from BLOBs backed up by StoragePoint. Its Windows Explorer-like interface allows administrators to browse and even preview items before restoring to ensure the correct content for restoration has been identified. Thus, StoragePoint eliminates the need to buy a backup solution simply to fulfill the need for granular restore.

Since the most frequent type of recovery performed on a SharePoint system will be item-level recovery due to accidental deletion, this can save administrators many hours of work to set up a recovery farm or site collection to recover a proposal that was accidentally deleted three months ago.

StoragePoint can also help to recover from a catastrophic failure. Since content externalization with StoragePoint will typically reduce a SharePoint content database up to 95 percent, a 250 GB content database externalized to a BLOB Store with StoragePoint would be reduced to about 12.5 GB. A smaller database will result in significantly improved recovery time.

CONTENT COMPRESSION

StoragePoint allows for content to be compressed before it is externalized to a BLOB Store. Compressed content will not only be faster to back up, but faster to recover as well. Compressing content BLOBs as they are externalized does come with some processing overhead, but you will still realize a net improvement in performance over a SharePoint system without StoragePoint. Using compression in conjunction with other

technologies and services can also reduce your overall storage costs — particularly with cloud storage options since bandwidth and space are both reduced.

If your storage platform has the ability to compress content and is on-premise then it may make sense to keep compression off in StoragePoint and allow your storage hardware to perform this operation, especially if the platform is caching the BLOBs and performing the compression as a post-cache operation.

It's a best practice to turn compression on for off-premise or Cloud platforms because you will benefit from reduced transmission times and, in the case of Cloud platforms, reduced bandwidth usage charges.

STORAGEPOINT CONFIGURATION DATABASE

The StoragePoint configuration database is not automatically backed up by SharePoint out-of-the-box tools. However, it can be added to the backup set. Failing to do this could lead to content becoming irretrievable, especially if changes to the StoragePoint Storage Profiles[vii] are not captured in a restored backup. This could be the absence of a storage profile altogether or the modification of a connection string setting. This is especially important if any of your storage profiles use encryption and you chose to have the key generated off a randomly-generated passphrase. It is a best practice to use a passphrase to generate the encryption key, so the key and the supporting storage profile can be recreated from a catastrophic failure where no complete backups are available.

OTHER CONSIDERATIONS

There are many factors to consider that are not in the scope of this document. When planning a SharePoint backup strategy, be sure to include IIS settings, SharePoint configuration databases and any custom changes that may have been made.

THIRD-PARTY BACKUP AND RECOVERY COMPATIBILITY

As mentioned above, there are many backup and recovery solutions for SharePoint available today. StoragePoint will coexist with those tools as long as they are using fully supported Microsoft APIs. It may still be necessary to back up the externalized content from the BLOB Store using StoragePoint because tools from third parties may just be performing content database backups. Consult with your backup and recovery solution provider for more information.

BACKUP AND GRANULAR RESTORE OPTIONS

The following table summarizes the key performance and cost impacts of backup and recovery options discussed in this paper:

	OOTB Backup Tools Alone	3rd Party Backup Tools	StoragePoint with OOTB Backup Tools
Back Up Tool Cost	$0	$$$	$0
Backup Storage Costs	High (Multiple copies of BLOBs with each backup)	High (Multiple copies of BLOBs with each backup)	Much Lower (BLOBs aren't duplicated in each backup)
Small RPO	No (Slow to back up large content DB)	No (Slow to back up large content DB)	Yes (Content DB backup is quick)
Granular Restore	No	Yes	Yes

CONCLUSION

SharePoint has become a core component of enterprise IT infrastructure. Thus, it's critical that the information contained within a SharePoint system is available and recoverable in the event of a system failure. However, the average SharePoint farm has grown to several times the maximum recommended size for backup and recovery with SharePoint's out-of-the-box tools. Adding StoragePoint to a SharePoint deployment makes native tools a viable option for today's terabyte-scale farms through the following features and benefits:

- Automatic BLOB Backup: StoragePoint can automatically and continuously back up externalized SharePoint content. This eliminates the need to set up a separate process to back up externalized content and reduces backup storage requirements by storing only a single copy of each BLOB backed up.
- Item Level Recovery: With StoragePoint implemented, SharePoint administrators can configure the site level recycle bin to hold items longer without concern for an ever-expanding content database. This gives users more ability to service their own undelete requests and provides confidence that their data will be available. Metalogix Selective Restore Manager can restore items beyond the recycle bin window from native SQL Server backups and from BLOBs backed up by StoragePoint.
- Reduced Content Database Size: Reducing the size of the content database through externalization of content BLOBs will decrease your backup and recovery times. Content databases can be reduced as much as 95%.
- Simplified Database Backups: Following the official guidelines provided by Microsoft can result in several large content databases that will need to be backed up. Using StoragePoint reduces the number of content databases that are required.
- Performance: BLOBs are better handled by the underlying file system than in a SQL database. Disk I/O is greatly improved allowing faster performance with your backup software.
- Content Compression: Content that is externalized by StoragePoint can be compressed. This can significantly improve your backup and recovery time as well as save on overall storage costs.
- Cloud Storage: Externalizing content to a cloud or any redundant storage system that is located in separate physical locations may eliminate the need to back up the content BLOBs entirely.

GLOSSARY

i Binary Large Objects (BLOBs) – are the actual content in a SharePoint content database rather than the metadata that describes it. BLOBS are typically unstructured chunks of data such as Microsoft Office documents, PDFs or TIFF document images.

ii Storage Area Networks (SAN) – Tier 1 storage that is physically attached to a server. The fastest and most expensive storage option.

iii Network Addressable Storage (NAS) – Tier 2 storage that is slower but less expensive than SAN storage. A device connected to a network to provide storage for other network clients

iv Content Addressable Storage (CAS): Unlike other storage technologies which retrieve data based on a location in an index without regard to content, CAS assigns a unique identifier to each piece of content. This method was created for archiving and compliance solutions. It's faster but more expensive than cloud storage.

v Cloud Storage: Network storage that is accessible through web-based API's. Cloud storage is typically provided by a third party.

vi BLOB Store: A location where BLOBs can be externalized. This can be a file system, SAN, NAS, or cloud storage.

vii StoragePoint Storage Profile: Define how and where the content BLOBs are externalized. These settings are defined in SharePoint Central Admin.

ABOUT METALOGIX

Metalogix provides content infrastructure software to improve the use and performance of enterprise content. For over a decade, Metalogix has transformed the way commercial and government organizations manage terabytes of content to improve knowledge sharing and collaboration. Today, more than 7,500 customers rely on the company's products to upgrade, migrate, organize, store, archive and replicate content on Microsoft SharePoint, Exchange and Cloud platforms. Metalogix is a privately held company backed by Insight Venture Partners and Bessemer Venture Partners. For more information, please visit: www.metalogix.com

METALOGIX

5335 Wisconsin Ave NW, Suite 510, Washington DC 20015

sales@metalogix.com | www.metalogix.com | 1.202.609.9100