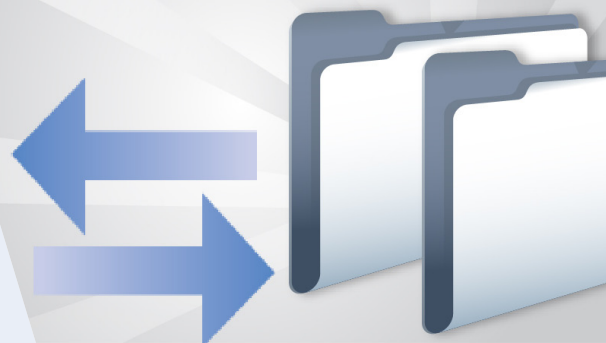


MOSS Content Backup and Recovery Options With StoragePoint

Tom Miller

September, 2009



CONTENTS

Overview And Scope	3
Standard Backup And Recovery Options	4
Backup Options	4
Recovery Options	4
StoragePoint Backup And Recovery Options	5
File System, SAN, Or NAS Externalization.....	6
Distributed File System (DFS)	7
Cloud Storage	7
Content Compression	7
3rd Party Backup And Recovery Options	8
Backup And Recovery Options Summary	8
Configuration Databases	9
Other Considerations	9
Conclusion.....	9

OVERVIEW AND SCOPE

SharePoint natively stores both content BLOBs (Binary Large Objects)ⁱ and their associated metadata within SQL. Since BLOBs are exponentially larger than their metadata counterpart, they add significant overhead which impacts SQL's I/O performance and transactional operations.

StoragePoint dramatically reduces SharePoint's content database size by externalizing content BLOBs to virtually any storage device. This includes Network Addressable Storage (NAS)ⁱⁱ, Storage Area Networks (SAN)ⁱⁱⁱ, and cloud storage iv options such as Amazon S3, Azure, or EMC Atmos.

Because all content in MOSS is stored a SQL server database, backup and recovery options are quite limited. Externalizing the content to a BLOB Storev can present an administrator with choices to ensure their content is available and easily recovered from actions as simple as a user deleting a file to complete restoration from a catastrophic event. The scope of this discussion is limited to backup and recovery. For more information and related white papers please visit our website.

STANDARD BACKUP AND RECOVERY OPTIONS

BACKUP OPTIONS

MOSS provides some out of the box tools via STSADM and Central Admin tools to backup a site collection. These tools will allow administrators to backup site collections either via command line and scripting or an administrator issued command. These tools are not recommended for use on¹ :

- ▶ 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.

With these recommendations in place, administrators often look to 3rd party tools to help with their database backup and recovery strategy.

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 either recover individual items or recover all content due to a catastrophic system failure.

MOSS 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. This is defaulted to 30 days. After the period expires the items are moved into the second tier which is only recoverable by an administrator. There are some important items to keep in mind when configuring your recycle bin settings:

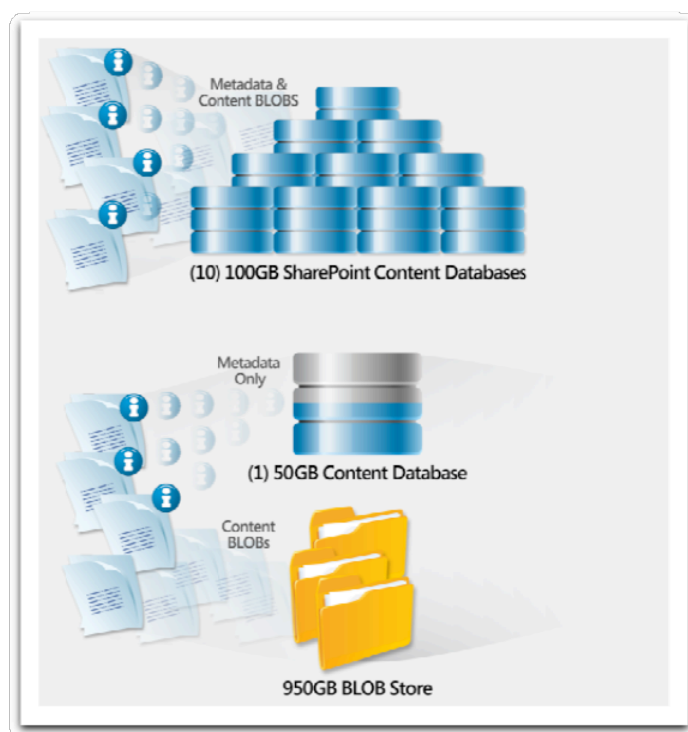
- ▶ Items in the recycle bin still count against a users quota (tier 1)
- ▶ These items are still stored in the content database if StoragePoint is not implemented
- ▶ The second tier of the recycle bin needs to be accounted for when planning database sizes and growth

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

¹ <http://technet.microsoft.com/en-us/library/cc263427.aspx>

One option is to take frequent backups of a site collection using the STSADM to keep the size of the recycle bin smaller. Unfortunately item level recovery beyond the recycle bin constraints is not possible with this method. An administrator would have to restore the content to a separate site collection and recover the item manually and upload it back to its original document library on SharePoint.

STORAGEPOINT BACKUP AND RECOVERY OPTIONS



When StoragePoint is implemented on a MOSS site collection or web application the BLOBs within the content database are moved to an external storage medium. This reduces the content database at least ninety-five 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 anywhere between twenty and eighty+ percent amounting to a significant cost savings. The types of content being externalized will determine how much you benefit from compression.

Using StoragePoint does introduce an additional step when backing up content as there is now a BLOB store that needs to be backed up. While this may seem cumbersome at first it actually creates opportunities that were simply not available to SharePoint Administrators without StoragePoint. One terabyte of content in a MOSS farm will translate to ten separate content databases to backup each sized around 100 gigabytes². Using StoragePoint can radically reduce this complexity by reducing the number of content databases from ten to one.

Externalizing the content to a BLOB store using StoragePoint can address some of the item level recovery challenges MOSS administrators face. It will keep the content database small even if the recycle bin is allowed to hold items for 365 days or more. Since the most frequent type of recovery actions performed on a MOSS system will be item level recovery due to accidental deletion this can also save administrators some of the tedious tasks they would need to perform to recover a proposal from three months prior that has been accidentally removed.

² <http://technet.microsoft.com/en-us/library/cc262787.aspx>

StoragePoint can also help to recover from a catastrophic failure. By externalizing the content to a SAN or NAS administrators can take advantage of many of the built in features of the hardware to create snapshots of the BLOB Store. Content externalization with StoragePoint will typically reduce a MOSS content database by at least 95 percent. For example, if a content database is around 250 GB and the content is externalized to a BLOB Store with StoragePoint then the remaining size of the database would be around 12.5 GB. Restoration of a smaller database will result in significantly improved recovery time.

An additional option to consider is using a cloud storage provider for the BLOB Store. Cloud storage never needs to be backed up since that is handled by the cloud services provider. It is also accessible from anywhere there is an internet connection. Bringing up a DR site will be much faster since the content does not need to be recovered or shipped to the backup site.

When backing up your MOSS farm you should take care to backup the content database prior to the BLOB store. Backing up or taking a snapshot of the BLOB store prior to backing up the content database could lead to orphaned list items. This is a condition where a list item in SharePoint references an externalized BLOB that does not exist in the referenced BLOB store. As long as you follow this one simple requirement, backup content database before you backup/snapshot BLOB store, you will never experience the orphaned list item condition.

You may experience an orphaned BLOB condition, but orphaned BLOBs do not impact SharePoint performance or make content irretrievable. Orphaned BLOBs occur when a content database is restored and the corresponding BLOB store contains BLOBs that were added after the content database was backed up. A BLOB Garbage Collector timer job is available in our 2.1 release that will crawl the BLOB store and remove these orphaned BLOBs. There are other conditions within the normal use of SharePoint that create orphaned BLOBs, but these are cleaned up by StoragePoint in real-time as they occur.

FILE SYSTEM, SAN, OR NAS EXTERNALIZATION

Content externalized to a file system, SAN, or NAS device will need to be included in your SharePoint backup strategy. It is important to ensure that the content database is backed up prior to backing up the BLOB Store. This ensures that any items listed in a list or a library will have their associated blobs available in the event of a recovery.

Backup of a file system BLOB Store can be done using typical backup methods either through native tools provided by Windows Server or 3rd party applications. A SAN or NAS provides additional options using snapshot features (i.e. EMC® SnapView, Hitachi Copy-on-Write, or NetApp® Snapshot) that are typically included with the hardware. These snapshots can then be replicated offsite either through a WAN link or physically moved on magnetic tape or other media. Consult with your hardware vendor for more information.

Restoring a content database that has had its content externalized requires the BLOB store to be recovered as well. If the BLOB Store is recovered to a different physical location the StoragePoint storage profile^{vi} will need to be updated. This can be done through the StoragePoint administration tools in SharePoint Central Administration. Refer to the StoragePoint Installation and Administration Guide for details.

DISTRIBUTED FILE SYSTEM (DFS)

If a BLOB store exists in a DFS namespace then replication to redundant locations occurs automatically based on the DFS configuration settings. This essentially creates your own managed cloud solution within the organization. Since data on a DFS share is replicated to multiple servers there is built in redundancy and offsite protection assuming the servers are located in different data centers.

It is not necessary to recover the blobs when recovering a content database that has content externalized to a DFS share assuming that the DFS share is still available to the MOSS farm. Saving this step in the recovery process can amount to several hours saved when faced with a recovery situation.

CLOUD STORAGE

When configuring content externalization to a cloud, a backup of the content is not necessary. Cloud storage providers are responsible for ensuring the availability of your data. It is important to understand the service level agreement (SLA) for your provider.

Restoring a content database that has had its content externalized to a cloud storage provider has some unique benefits. The content database size is incredibly small so restoration will take significantly less time than if the blobs were included in the database. More importantly, the cloud is accessible anywhere there is internet connectivity. Not only do you not have to restore your content, but you can have your data center located anywhere that has the bandwidth available to serve the content to the user community.

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 backup, 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 MOSS system without StoragePoint. Using compression in conjunction with other technologies and services can also reduce your overall storage costs particularly with cloud storage options as bandwidth and space are both reduced.

If your storage platform has the ability to compress content and it 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 as you will benefit from reduced transmission times and in the case of Cloud platforms, reduced bandwidth utilization charges.

3RD PARTY BACKUP AND RECOVERY OPTIONS

There are many backup and recovery solutions for MOSS available today. StoragePoint will coexist with those tools as long as they are using fully supported Microsoft API's. It may still be necessary to backup the externalized content from the BLOB Store as tools from 3rd parties may just be performing content database backups. Consult with your backup and recovery solution provider for more information.

BACKUP AND RECOVERY OPTIONS SUMMARY

BLOB Store	What to Backup	Benefits	Notes
File System	Content Database File System BLOB Store	Simple Implementation	Schedule the database backup to run prior to the file system backup to ensure all of the content that SharePoint is referencing is available in the BLOB store. This order will eliminate the creation of orphaned list items.
SAN/NAS	Content Database SAN/NAS Volume	Can take advantage of hardware features such as data de-duplication.	Most organizations will already have a backup plan or procedure in place for their SAN/NAS storage. Verify that it coincides with the needs of your MOSS backup plans.
DFS	Content Database	Content does not need to be backed up if it is replicated to offsite datacenters.	If content is not replicated offsite make certain the BLOB store is included in the backup plan.
Cloud Storage	Content Database	Since cloud storage providers ensure high availability for their storage a backup does not need to be performed.	Verify your cloud storage provider's SLA to ensure it coincides with the needs of your data.
Compression	N/A	Smaller physical file sizes result in improved backup and recovery time	Although there is some processing overhead, you will still realize a net improvement in performance when uploading content.
3 rd Party Tools	Consult with your vendor	3 rd party backup and recovery tools will work with StoragePoint as long as they are using supported MOSS API's.	BLOB Store location determines if the BLOBs should be backed up. Consult this chart to help make that determination.

CONFIGURATION DATABASES

SharePoint and StoragePoint configuration databases are not backed up automatically by the out-of-the-box SharePoint tools, so they need to be accounted for separately in your backup/recovery process. Failing to do this could lead to content becoming irretrievable, especially if changes to the StoragePoint storage profiles 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 also 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 MOSS backup strategy, be sure to include IIS settings, configuration databases and any custom changes that may have been made.

CONCLUSION

MOSS has quickly become a core offering from an IT infrastructure perspective. It is critical that the information contained within a MOSS system is available and recoverable in the event of a system failure. Adding StoragePoint to a MOSS deployment gives administrators unprecedented benefits and options for backup and recovery such as:

- ▶ Item Level Recovery – With StoragePoint implemented, MOSS 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.
- ▶ 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 ninety-five percent or greater.
- ▶ 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 backup the content BLOB entirely.

GLOSSARY

- ⁱ BLOB – A Binary Large Object is a collection of binary data stored as a single entity in a database management system
- ⁱⁱ Network Addressable Storage (NAS) – A device connected to a network to provide storage for other network clients
- ⁱⁱⁱ Storage Area Networks (SAN) – A SAN is a network of devices optimized to provide storage capacity for servers and other IT infrastructure hardware
- ^{iv} Cloud Storage – Network storage that is accessible through API's. Cloud storage is typically provided by a 3rd party.
- ^v BLOB Store – A location where BLOB's can be externalized. This can be a file system, SAN, NAS, or cloud storage.
- ^{vi} StoragePoint Storage Profile – StoragePoint storage profiles define how and where the content blobs are externalized. These settings are defined in SharePoint Central Admin.

ABOUT METALOGIX

Metalogix is the trusted provider of content lifecycle management solutions for Microsoft SharePoint, Exchange and Cloud platforms. We deliver high-performance solutions to scale and cost-effectively manage, migrate, store, archive and protect enterprise content.

METALOGIX

5335 Wisconsin Ave NW, Suite 510, Washington DC 20015
sales@metalogix.com | www.metalogix.com | 1.877.450.8667