



Optimize Your SharePoint 2010 Content Using Microsoft's New Storage Guidance

By Corey Milliman
October 2011

Copyright © 2011 Metalogix International GmbH. All rights reserved. Metalogix is a trademark of Metalogix International GmbH. StoragePoint is a registered trademark of BlueThread Technologies, Inc. Microsoft, Exchange Server, Microsoft Office, SharePoint, and SQL Server are registered trademarks of Microsoft Corporation.

Contents

Abstract.....	4
Introduction	5
Microsoft’s Solution for SharePoint Performance.....	6
New Guidelines: 200 GB to 4 TB	7
The Details of the SP1 Sizing Guidelines	8
Metalogix Brings Simplicity to Active Sites	9
Metalogix Brings Simplicity to Archive Sites.....	10
Metalogix Brings Simplicity to Disaster Recovery.....	11
Conclusion.....	12
About the Author	13
About Metalogix	13

Abstract

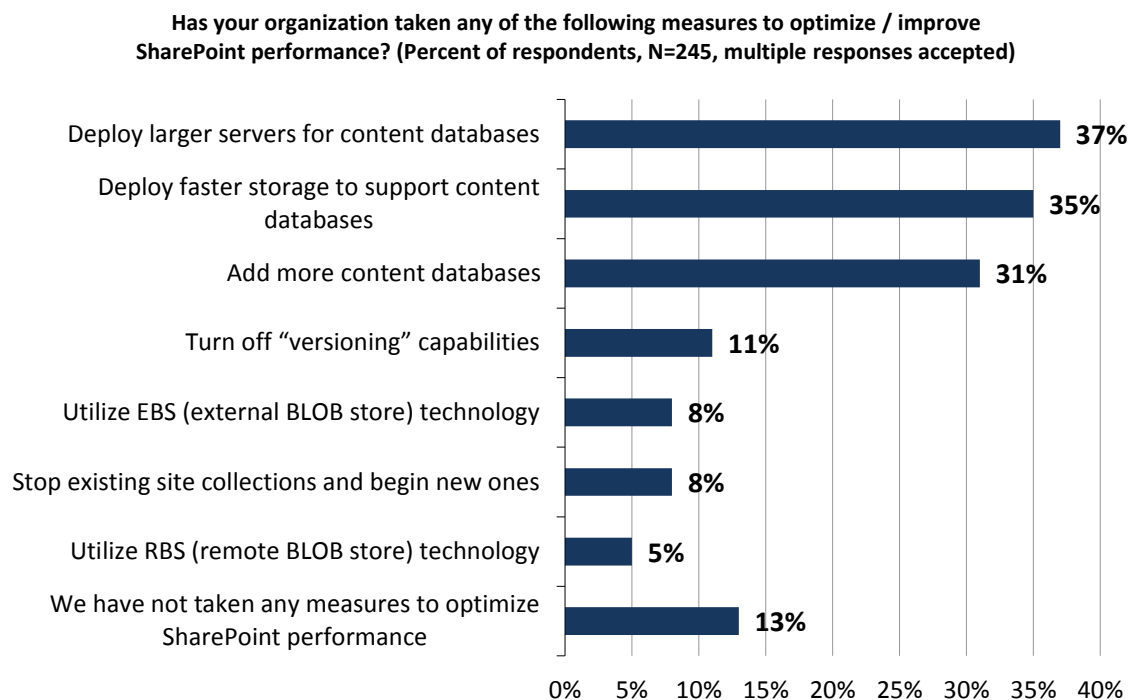
Microsoft recently released new storage and database sizing guidelines along with the release of Service Pack 1 for SharePoint 2010. This guidance marks a shift in how Microsoft looks at SharePoint content and is important to organizations looking for an efficient, cost-effective means of providing a more robust end user experience by fully leveraging RBS technologies. In this paper, we will explore how to work within the new guidelines to provide a truly autonomous way to manage enterprise content and position SharePoint to be your go-to ECM system.

Introduction

SharePoint 2010 provides a rich set of collaboration features to help information workers more efficiently create, manage and find content to make better business decisions. Since SharePoint was designed to empower users to work via a collaboration platform, it has allowed users to easily create and store tremendous amounts of content.

We are currently experiencing an explosion of digital content, with volumes set to grow from 1.2 trillion GB in 2010 to 35 trillion GB by 2020.¹ We create and consume more content than ever before, and IT teams are expected to make that content highly available to meet business needs and keep workers as productive as possible. It's expected that SharePoint content will grow at a similar rate for two reasons. First, increasingly tech-savvy end users are embracing SharePoint at an unprecedented rate, and therefore store their content on the platform. Secondly, as SharePoint continues to mature in an organization's application infrastructure, its content continues to grow as decision makers consolidate content from various legacy systems into SharePoint.

Historically, governance plans and strict design considerations have been the answer for IT to meet strict SLAs for their SharePoint farms. The difficulty in managing SharePoint becomes apparent in the way organizations respond to the technical challenges brought on by increased adoption of SharePoint as a platform. According to recent research from Enterprise Strategy Group (ESG), many organizations are forced to react in various ways that actually remove some of the features users are looking for in a collaboration platform.



Source: ESG

¹ IDC, May 2010. *The Digital Universe Decade*. Available at: http://gigaom.files.wordpress.com/2010/05/2010-digital-universe-iview_5-4-10.pdf

Microsoft's Solution for SharePoint Performance

Microsoft SharePoint products and technologies rely on SQL Server as the foundation for data storage. SQL Server is a high performance tool for management of structured data, and gives users the capabilities that drive adoption of SharePoint, such as search, metadata, workflow and versioning.

The problem is that as users create unstructured data, such as Office documents and media files, that data is also stored in SQL. When these objects are stored in SharePoint, they are stored as unstructured binary data streams associated with necessary metadata. These streams, often referred to as binary large objects or BLOBs, represent the majority of the data residing in SharePoint content databases.

SQL Server is not the best option for storing unstructured binary data streams due to the inherent nature in which SQL actually works with data. Store and retrieve operations of BLOB data in SQL Server are little more than simple "get and put" statements. BLOB operations suffer because the unstructured streams cannot be exposed to the full set of one of SQL Server's most powerful features, its query engine.

Microsoft released two solutions for handling the BLOB operation issue: External BLOB Storage or EBS and Remote BLOB Storage or RBS. Both of these technologies allow for the performance-robbing BLOBs to be stored on data stores designed to handle unstructured data, while seamlessly allowing SQL to handle the structured data.

The impact of deploying an RBS solution is immediate and quantifiable: smaller content databases, better performance for end-users, and the ability to make flexible decisions about where and how you store your unstructured data – all without changing the way your end users work.

Microsoft recently released new storage and database sizing guidelines along with the release of Service Pack 1 for SharePoint 2010. This guidance marks a shift in how Microsoft looks at SharePoint content and is important to organizations looking for an efficient, cost-effective means of providing a more robust end user experience by fully leveraging RBS technologies.

With Service Pack 1 and complementary solutions from Metalogix, organizations can significantly reduce the total cost of ownership (TCO) associated with SharePoint. In this paper, we will explore how to work within the new guidelines to provide a truly autonomous way to manage enterprise content and position SharePoint as your go-to ECM system.

New Guidelines: 200 GB to 4 TB

Microsoft's previous content database sizing guidelines, recommending no more than 200 GB in a single content database, made managing SharePoint difficult because that 200 GB included both metadata and BLOBs regardless of where they were located (in SQL or externalized to other storage). In July of 2011, Microsoft posted the following guidance on storage:

"With the release of SharePoint 2010 SP1 and some new guidance we are changing the supported data size limits for SharePoint content databases. Prior to SP1 the content database limit was 200 GB for collaboration and 1 TB for document archive. The content database size includes both metadata and BLOBs regardless of where the BLOBs are located and use of RBS does not bypass or increase these limits.

The new guidance for supported content database size details outlines specific guidance for SharePoint administrators as the data size grows. If this new guidance is followed SharePoint can support up to 4 TB of data in all usage scenarios and has no imposed size limit for document archive scenarios."

As you would expect there are certain details and caveats in Microsoft's guidance that have to be considered in order to provide 4 TB of storage in a single content database. But before digging deeper into the Microsoft guidance, what does this new data size limit really mean?

The guidance dictates that the 4 TB of storage includes any content offloaded by RBS. That means that you can use RBS to bring the actual SQL content down by 90-95% depending on how you externalize data. Since third-party RBS solutions, such as Metalogix StoragePoint, allow for the use of multiple storage devices while increasing performance, the smaller content databases are cheaper to store, maintain and backup.

When looking at *SharePoint Server 2010 capacity management: Software boundaries and limits*² published by Microsoft, a single web application can have 300 content databases associated with it and each content database can hold 2,000 site collections comfortably. Based on the new guidance, that means one web application can be associated with 1.2 Petabytes (PB) of information. 1.2 PB of SQL storage would be difficult to manage, backup, maintain, and would likely be cost prohibitive for many organizations. SharePoint backups would also be very difficult, and ensuring 99.9% uptime for a SQL cluster of that magnitude would require a large capital investment and a complicated high availability strategy.

Without utilizing an RBS solution, the guidance makes for a good read but doesn't make for a cost-effective or practical SharePoint implementation. Increasing database size limits isn't going to solve the fundamental problems we face with SharePoint: backup, ease of recovery, high availability, the pure cost of SQL hardware to support all of that content, and most importantly end-user experience. If you build it, they will come. If it doesn't work, they will leave!

Just because you can now officially store 4 TB of BLOB content in a single database doesn't mean it is going to be feasible. Every time a user browses a document library, or views a list, SQL places a lock on the resources associated with every document or list attachment presented to that user. If you follow the Microsoft guidance, and support a single content database with 4 TB of storage with thousands of site collections, users will very likely see a significant decrease in performance. If we externalize our BLOBs to another device, we can improve the SharePoint experience without straining our budgets.

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

The Details of the SP1 Sizing Guidelines

Let's review the recommendations from Microsoft on how to achieve the 4 TB content database:

"Requires disk sub-system performance of 0.25 IOPS per GB, 2 IOPS per GB is recommended for optimal performance."

- It is assumed that the IOPS requirement is for the transaction database (SQL Server) and not the externalized BLOBs. Keep in mind that with Microsoft's updated FileStream RBS Solution, Microsoft can natively support NAS appliances via iSCSI.
- Calculation
 - Total Content DB-4096 GB (Externalized Content + Content Database)
 - Content Database Size-409.6 GB (Assuming 90% reduction)
 - Externalized Data Size- 3686.4 GB
 - IOPS per GB-.25
 - Total IOPS Required-102.4
- 102.4 IOPS is comfortably within the performance benchmarks of many storage devices on the market today, including Fiber Channel, etc.
- For more details on IOPs calculations and how various disk configurations can meet these guidelines, please read the blog, "[Revisiting SharePoint Remote BLOB Storage](#)."

"Requires the customer to have plans for high availability, disaster recovery, future capacity, and performance testing."

- When planning a SharePoint deployment, high availability, disaster recovery, performance testing and benchmarking and future capacity considerations are critical.

Furthermore, Microsoft has *removed* the size limits for archived content. We are free to have unlimited archive, as long as the following conditions are met:

- SharePoint sites must be based on Document Center or Records Center site templates and must be an archive scenario where less than 5% of content is actively read from each month and less than 1% of content is actively written to.
- Do not use alerts, workflows, link fix-ups, or item-level security on any SharePoint objects in the content database. Document archive content databases can be the recipient of documents as a result of Content Routing workflow.

Other changes with the new SP1 guidance:

- You are now limited to 60 million items in any one content database
- The 5 TB limit per SQL Server instance has been removed

There are a few other key items to consider:

- RBS providers are required to return the first byte of data in a request in 20ms. This applies for all requests between SharePoint and the RBS provider storage layer.
- RBS is most likely to be used for document archive scenarios where documents are written and not updated. BLOBs in RBS are never updated once they are written; instead a new BLOB is created for any update. BLOBs are immutable, old BLOBs are garbage collected later.
- Backup strategy must be carefully considered. Both document metadata and document BLOBs must be backed up at exactly the same point in time. This means any third-party backup solution needs to be capable of restoring both the SQL database used by SharePoint and the BLOBs used by SharePoint as a set where no variance occurs which would have the database reference BLOBs that are not available from the same backup.

Metalogix Brings Simplicity to Active Sites

There is a lot of information to consider in the new SP1 guidance, and no previous announcement from Microsoft better highlights the need and importance for every organization to consider RBS as an integral part of their ECM strategy. SP1 is just as much about intelligent content management as it is about the evolution of the backend storage architecture. With the right RBS toolset at your disposal, you can turn SharePoint into the best ECM tool for your organization, and easily stay within the Microsoft recommended boundaries of today and the future.

As discussed earlier, there are two different database sizing limits to consider: 4 TB for active content and “unlimited” for inactive content based on the Document or Record Center Templates.

To optimize your storage for active content, Metalogix StoragePoint provides a seamless way to optimize your storage resources. StoragePoint shines because of the direct insertion into SharePoint’s Central Administration Site. There are no additional servers to deploy and make highly available, nor are there additional open-source technologies, web parts or custom code inserted into your environment. StoragePoint itself is architected to be highly available, just like SharePoint. Since StoragePoint deploys directly to SharePoint, the loss of a Farm Server will not cause an outage for your users. There’s no new interface to learn or maintain; only an extension of what you already know how to use.

StoragePoint also supports multiple storage scopes based on the location in the SharePoint hierarchy or the metadata associated with a document or site. This allows for smart externalization of content to different SAN, NAS, Cloud or enterprise storage devices via adapters that are included with the product. This is important because different business rules for the management of content might dictate that different types of content be externalized differently. Highly sensitive content can be encrypted utilizing 256-bit AES encryption, with the user providing a hash key. Other content may need to be compressed, and other types of content may need to have both the BLOBs and the metadata externalized. Finally, different types of content might need to be stored in different locations based on business value. All of these scenarios are fully supported in a straightforward interface from Central Administration.

As just mentioned, different tiers of storage can be utilized based on the business value of content. If a project has been closed, it most likely can be moved to a lower-cost tier of storage. Photos of the company Holiday Party probably don’t need to sit on your most expensive storage, while your business continuity documents are probably of the utmost importance. StoragePoint gives you the ability to turn your business rules for document management into intelligent storage profiles that will save your organization precious dollars on storage without impacting the end user experience. You can immediately start saving money on storage and dramatically increase performance without changing the way your users work.

After deploying StoragePoint, the end user experience improves significantly. When a user uploads a document, the BLOB never hits the content database, as only the metadata is inserted. Large files can be uploaded without timing out. It is important to note that with this guidance, Microsoft still supports a maximum upload size of 2 GB per file. Note that vendors that bypass this limit are putting your organization, and your data, in an unsupported state! List views are presented faster, metadata interaction is lightning fast, and any SharePoint operation that doesn’t actually have to retrieve a file is able to run at peak performance with minimal SQL resources.

Metalogix Brings Simplicity to Archive Sites

Microsoft has also indicated that archive sites have no limit from a database perspective, as long as the content is based on certain templates. Of that content, only 5% can be actively read from, and 1% actively written to. While there is some functionality within SharePoint 2010 to automate the delivery of documents to the document center or records center sites, Metalogix Migration Manager for SharePoint can automate this process for you.

Metalogix Migration Manager for SharePoint, while primarily thought of as a migration tool, also provides extensive functions for ongoing management of your SharePoint farms. The solution is StoragePoint aware, so managing content throughout the lifecycle is simple and painless. Migration Manager for SharePoint and StoragePoint can work together seamlessly to identify, move, and classify content ready for “archive” status.

To perform this process, jobs can be setup to easily move content around your farms based on the metadata associated with content. If a SharePoint site hasn't been accessed in 6 months, some organizations would consider it inactive. In other organizations, a project could be marked for archive as soon as someone indicates the project is complete through filling in a site column or completing a workflow.

A job can automatically run to move the site while re-templating it to a document or record center site, or move the previous site collection to a sub-site of a record center site. As part of the move, the job can automatically connect to the “archive” database. Since the tool is StoragePoint aware, the BLOBs will either be moved to fit the profile of the destination, or stay at rest with only the site collection moving. Either way, this is a very easy way to stay within the Microsoft Guidelines for active vs. inactive content with consideration to database sizing limits.

Another significant feature is that other jobs can be defined to pull sites out of archive. What if an administrator has moved content into the archive database and a user starts interacting with content consistently? Automated jobs can be created to move content back out from archive to live content databases, while maintaining the integrity of the metadata, permissions, look, feel, etc.

The procedures above can be done via a GUI, or automated via PowerShell. Metalogix Migration Manager for SharePoint is a 64-bit application with multi-threading, item-level support, and the ability to run multiple jobs at once. With this robust tool, jobs can run unattended, stacked, and even during production without impact on end users. Since we've externalized the BLOBs and are only moving BLOB references, a site collection move representing 400 GB of data can take place in minutes, giving the SharePoint administrator the freedom to perform other tasks.

Metalogix Brings Simplicity to Disaster Recovery

Introducing RBS into your environment does introduce another layer in data protection and disaster recovery scenarios. By utilizing products available from Metalogix and a back-to-basics approach to disaster recovery, you can remove the complexity and deploy a solution that meets even the most aggressive service level agreements.

Since StoragePoint's RBS/EBS tools are utilizing SharePoint APIs, we are able to fully function with traditional high availability practices such as SQL Log Shipping and SQL Mirroring. Once the BLOBs are offloaded with RBS, you will probably have the bandwidth required to actually perform SQL Mirroring of their content databases.

The content databases themselves contain BLOB references. You have to make sure that whatever device you've externalized to is also highly available. At first, many customers view this as a problem. However, IT organizations have been dealing with file systems for years. As IT administrators, we already know how to make our file systems highly available. As an industry, we've already solved the file system backup problem. You can use Windows Distributed File Systems, Storage Based Replication, or Microsoft Azure for cloud storage of BLOBs. You can also use hardware vendors, such as EMC, that take care of block-level replication and de-duplication of data to multiple data centers. In the event of a failure, you can easily perform a DNS switch and be up and running in minutes.

Microsoft also indicates that both the BLOB and the corresponding BLOB reference must be backed up at the same time. Synchronized point-in-time backups of SQL and file systems are attainable, and StoragePoint-aware restore capabilities are present within Metalogix Selective Restore Manager.

With Selective Restore Manager, SharePoint administrators can utilize native SQL backups to restore data at varying levels within the SharePoint hierarchy. The lightweight tool supports both in and out-of-place data restores from the site collection down to the document version.

Conclusion

With Microsoft's new storage guidance and the release of Service Pack 1 for SharePoint 2010, organizations have more flexibility in managing their SharePoint environments. However, SharePoint administrators must make sound decisions to properly manage these environments and overcome the challenges associated with rapidly expanding content databases, manageability of content, and increasing storage costs.

While it becomes more and more evident that SQL Server is not the best option for managing unstructured content, administrators face issues in offloading and storing this content in a cost-effective and performance-enhancing way. A solution, such as Metalogix StoragePoint, gives administrators the freedom to meet these challenges by taking advantage of Microsoft's RBS BLOB offloading method, which results in optimized SharePoint performance and reduced total cost of ownership by utilizing lower-cost tiered storage.

About the Author

Corey Milliman – SharePoint Systems Engineer

Corey joined Metalogix as a Systems Engineer in the Federal Sector. Corey has been a prominent speaker for Metalogix, presenting at conferences and trade shows on BLOB offloading and migration scenarios to SharePoint 2010.

Prior to joining Metalogix, Corey was a Sales Engineer with another ISV focused on the Public Sector. Corey has spent 12 years in consulting roles covering the architecture, design, implementation and support of various ECM Systems and Microsoft Products.

About Metalogix

Metalogix is the trusted provider of innovative 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 provides global support to thousands of customers and strategic partners and is a Microsoft Gold Partner and GSA provider. 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

1.877.450.8667