# *Proven Practices for Optimizing SharePoint Storage Management Strategies with DocAve*

# Table of Contents

*Introduction* ..... 2

*About AvePoint* ..... 2

*Unstructured Data* ..... 3

*Legacy Data* ..... 4

*Inactive Data* ..... 5

*DocAve for a Comprehensive Storage Optimization Strategy* ..... 6

*Conclusion* ..... 7

# Introduction

*Microsoft SharePoint Server 2010, the latest release of Microsoft's best-selling server product, is quickly revolutionizing how organizations worldwide connect their people, processes, and information. It will mark SharePoint's evolution from a server "application" into a full-fledged platform – the world's first enterprise-class "virtual ecosystem" to fulfill the promise of truly unifying business' information management.*

*Organizations considering SharePoint Server 2010 should do appropriate planning with regard to the effective management of SharePoint storage as well as the intelligent management of SharePoint content lifecycles. Because of its best-in-class collaboration and information management capabilities, SharePoint is the ideal platform for the presentation and management of enterprise content.*

*However, SharePoint uses a unified storage infrastructure that utilizes the SQL Server database. While SQL is an efficient database technology, its use as SharePoint's backend can pose unique challenges for organizations that are looking to centralize terabytes' worth of legacy data. SQL is a relatively expensive storage media compared to file- and cloud-based storage, and SQL's performance can also be compromised when burdened with unstructured, non-relational data such as Binary Large Objects (BLOBs).*

*Since throwing money at the problem is simply unacceptable in today's ultra-competitive global economy, IT administrators are working on a cost-effective yet robust strategy to ensure effective, system-wide content lifecycle and storage management. All the while, the solution must ensure user experience is not diminished. End-users must be able to access and manage content in the same manner they normally would within SharePoint. They are doing this by addressing three types of SharePoint data: unstructured data, legacy data, and inactive data.*

*In this paper, we'll examine how organizations can properly manage each type of SharePoint data with AvePoint's DocAve Storage Optimization Suite in order to unleash SharePoint's full potential and unlock the key to quick cost savings for existing storage investments.*

## About AvePoint

*AvePoint is a global technology company and software innovator headquartered in the United States. Since its founding in 2001, AvePoint has become the world's largest provider of infrastructure management software solutions for Microsoft SharePoint Products and Technologies. Propelled by the world's largest SharePoint-exclusive research & development team, AvePoint is the premier provider for EPG, SMB, Mid-Market and Government organizations demanding the most powerful and flexible infrastructure management solutions for their SharePoint environments and assets. AvePoint's award-winning DocAve Software Platform is recognized as the industry standard for comprehensive and scalable SharePoint backup and recovery, administration, replication, migration, archiving, deployment management, reporting, storage optimization, and content lifecycle management.*

*AvePoint is headquartered and maintains its principle engineering center in Jersey City, NJ, with wholly owned sales and engineering centers in the USA - San Jose, Los Angeles, Seattle, Chicago, Washington DC, Houston, Boston; Ontario, Canada; Melbourne, Australia; London, United Kingdom; Munich, Germany; Johannesburg, South Africa; Tokyo, Japan; Singapore; and China - Beijing, Changchun, Dalian. AvePoint's global team, fortified by an expansive network of certified partners, helps more than 6,000 enterprise customers – including many Fortune 500 companies and government agencies – to protect, manage, optimize, and integrate their mission-critical SharePoint environments. AvePoint is a Depth Managed Gold Certified Microsoft Partner and GSA provider.*

# Unstructured Data

*All data uploaded into SharePoint is stored by default within the platform's SQL Server database. SQL is a relational database, which is fantastic for storing structured data. However, it is not as efficient when dealing with larger, non-relational data streams – also known as Binary Large Objects (BLOBs) – and includes Word documents, PDF files, and video files. Considering more than 95 percent of the data a typical organization uploads into SharePoint are BLOBs, the following negatively consequences can result:*

- *Overall time to remove unstructured content from a structured database is inefficient, resulting in longer index times and slower response to SharePoint end-users;*
- *Longer backup windows and reduced ability to stage during restores from database-level backups;*
- *Performance degradation in the time to index and retrieve large files in SharePoint is slowed due to performance of large data;*
- *Decreased performance due to fragmentation in SQL Server, resulting in an overall platform slowdown;*
- *Developers experience substantial stability issues due to BLOBs when sites exceed maximum Microsoft-recommended limits, rending typical developer APIs and commands ineffective.*

*Microsoft helped address this issue by providing an External BLOB Store (EBS) Provider with Microsoft Office SharePoint Server (MOSS) 2007 SP2, which enabled organizations to extend SharePoint storage to other media. The EBS provider can take ownership of BLOBs and move them off to cheaper, more efficient file-based storage while leaving a token or stub in the SQL Server, so SharePoint can retrieve the object if necessary. One problem, though, is that EBS isn't granular – only deployable at the farm-level. Consequently, administrators had to deploy the EBS solution on every SharePoint web front end server. It also required deep knowledge of scripting and coding in order to successfully deploy the provider.*

*SharePoint Server 2010 is looking to improve upon this with the support of the Remote Blob Storage (RBS) API, a SQL-based API. More flexible than EBS, RBS enables storage of all content in a Site Collection on the file system (with metadata retained in the SQL content database). However, it also requires significant coding in order to utilize the RBS effectively.*

*Whether an organization decides to do custom coding or looks to a third-party provider, it is imperative it utilizes Microsoft's EBS and RBS APIs. Keep in mind that the SQL Server is the engine that powers the SharePoint platform. If it is overburdened with unstructured data it is not natively equipped to efficiently handle, it can have potentially disastrous consequences at the precise time your business needs the platform most.*

*AvePoint's DocAve Storage Optimization suite offers a comprehensive solution for the offloading of BLOBs from SQL Server with the free tool DocAve Extender for SharePoint, offering the following features:*

- *Secure offload of BLOBs to non-database storage – based on customizable file-size triggers – for prevention of BLOB data from ever entering SQL;*
- *Transfer of BLOBs to any network file shares or cloud storage system;*
- *Full integration with SharePoint search;*
- *Seamless management, using all of SharePoint's management and collaboration tools, including versioning, workflows, and alerts, all as if it was residing directly in SQL;*
- *Total interactivity with client Office applications and other third-party tools.*

*Benefits organizations realize with DocAve Extender include:*

- *Optimized SQL performance, enhancing end-user experience and drives further platform adoption for intended business initiatives;*
- *Expensive SQL Server licensing is then saved for alternate, cheaper storage devices.*

# Legacy Data

*Many organizations have volumes of legacy content stored on myriad file shares, legacy databases, and other storage devices. For governance and productivity, most organizations would like to unify management and presentation of this legacy data with SharePoint Server 2010. However, for financial, compliance, or logistical reasons, many companies do not necessarily want to migrate this data into their SharePoint environments. Specifically, these challenges can include:*

- *A substantial load on network bandwidth and SharePoint servers during migration projects;*
- *Inefficient means to discover content organization-wide when it resides in multiple data stores;*
- *Disparate categorization, tagging, mapping, and migration of legacy data to appropriate business units;*
- *Loss in usable server and storage resources due to retired legacy projects;*
- *Media and large video files migrated to SharePoint can pose a performance risk and bottleneck within SharePoint for end-users.*

*In order to help companies connect their legacy databases and line of business applications – such as ERP and CRM – to SharePoint, Microsoft included Business Data Catalog (BDC) with its release of MOSS 2007. BDC provided IT administrators a way to present business data from back-end server applications, such as SAP or Siebel, within MOSS without writing any code – provided that it was stored in a database.*

*Even though the BDC made it easier to create read-only solutions that display data in the Business Data List Web Part, it was impossible to make changes and write that data back to the external store via SharePoint. In SharePoint Server 2010, Microsoft unveiled the new BDC, now called Business Connectivity Services (BCS). BCS provides read/write access to external systems – through Web services, databases, and Microsoft .NET Framework assemblies – from within SharePoint Server 2010 and Microsoft Office 2010 applications. However, while developers can now use SharePoint Designer 2010 and Visual Studios 2010 to access external data via the BCS, the BCS can't deliver the same functionality to legacy file-share content.*

*For companies truly looking to streamline its enterprise-wide content through one platform and connect its people, processes, and information, it is imperative they find a means to either migrate all the content and data into SharePoint or expose it via SharePoint.*

*To ensure seamless access to any network or cloud file-share content directly through SharePoint – without the need for migration – organizations can utilize DocAve Connector for SharePoint, enabling the migration-free SharePoint presentation, management, and streaming of legacy content. Furthermore, with DocAve Connector, IT administrators can:*

- *Manage all custom metadata, securities, and alerts directly through SharePoint;*
- *Fully integrate legacy file-share content with SharePoint search, making it fully discoverable;*
- *Connect existing Home Drives to user My Sites in SharePoint, preventing unwanted platform clutter;*
- *Stream media seamlessly to SharePoint users from non-SQL storage locations.*

*Potential return on investment with DocAve Connector includes:*

- *Easily discoverable, efficiently managed content within SharePoint without spending the time and money necessary to migrate the content directly into SQL;*
- *Repurpose and give second life to existing storage purchases;*
- *Save on future SQL expenses due to growth of non-critical storage.*

*Developers wishing to create custom applications for bulk streaming of documents, audio files, and video clips to selected SharePoint locations can do so with* **DocAve Connector Software Development Kit**. *For organizations that already made the decision to migrate all of its content and data into SharePoint and eliminate their legacy data systems, DocAve Migrator for SharePoint offers a lossless, secure transfer of all content and its requisite metadata from various legacy systems – including EMC Documentum and Lotus Notes – to SharePoint Server 2010.*

*The objective of Content Lifecycle Management is to make sure that data is maintained and available in the most efficient and cost-effective way possible throughout its life – from authorship to disposition. When SharePoint Server 2010 is used as the enterprise-class information management system it is destined to be, it becomes home to a variety of content contributed by end-users throughout the organization. While adopting social networking features, utilizing team-building functionality, or using SharePoint's publishing and versioning capabilities, a trail of dormant and unused data is left behind that can begin to decrease performance in SQL, decrease search and indexing speeds, and discourage broader SharePoint use due to clutter as well as the lack of availability to important data.*

*Consider this: 100 versions of a 10-megabyte Microsoft Word document will take up 1 gigabyte of storage in a content database. In reality, only the last five versions or so of this document are actively used and referenced. That leaves 95 other versions of the document sitting in the content database, slowing down the SQL Server.*

*Due to compliance or regulatory objectives, an organization may have to keep all of these versions. However, there is no reason why companies must keep all versions in their SQL content databases. Furthermore, organizations should offload this stale data onto their existing hierarchal storage management (HSM) investments and create specific business rules that will automate the process of offloading this dormant data from the SQL Server.*

*There is no real native archiving or content lifecycle management capability in SharePoint 2010 to help with this. SharePoint Server 2010 offers a greatly enhanced Records Center, but the name really describes the function. It is a SQL Server-based storage repository for SharePoint objects marked as records. But Records Center in SharePoint Server 2010 still does not take dormant content from SQL databases and move it off to cheaper storage. Acting as a separate site collection within SharePoint, it doesn't help when it comes to relieving SQL of content no longer actively utilized by end-users, or help an organization take full advantage of their tiered storage resources. Another new feature in SharePoint Server 2010, in-place records management, allows certain SharePoint documents, blogs, wikis, web pages, and list items to be declared records. The key here is that when declared a record, the content does not move to an archive. Rather, it stays where it is so the end-users can still find and interact with the content. If all data is stored within SQL's content databases, this still does not solve the problem of alleviating SQL Server or utilizing HSM investments.*

*For true content lifecycle management, SharePoint administrators should look to third-party solutions. AvePoint's DocAve Archiver for SharePoint empowers administrators to automate the process of offloading dormant SharePoint content to cheaper file- or cloud-based storage based upon fully customizable business rules, with features including:*

- *Remove dormant Sites, Site Collections, Lists, and Libraries as triggered by the completion of a project;*
- *Prune unused versions from SharePoint, leaving only the most active versions accessible via the SharePoint interface;*
- *Maintain effective SharePoint grooming for list items;*
- *Move large files, lesser-used files, or clutter and contents to cheaper storage;*
- *Present an end-user driven interface for removing and restoring archived contents.*

*In leveraging all of these features of DocAve Archiver, return on investment can include:*

- *The ability to clean up SharePoint contents without removing them using end-user driven archiving;*
- *Improve SharePoint search experience by removing unneeded clutter;*
- *Proactive response to compliance data retention initiatives via customizable pruning rules;*
- *Reduce backup windows by only targeting business-critical content;*
- *Savings on total cost of ownership of SharePoint throughout its lifespan, via pruning unwanted content to existing HSM investments.*

*While each individual solution in DocAve's Storage Optimization Suite can singlehandedly begin to alleviate the storage challenges organizations have with regard to their burgeoning SharePoint deployments, the integrated nature of the DocAve Software Platform, with more than 25 independently deployable modules piloted via a unified, browser-based interface, is what enables companies to take the next step in truly optimizing the entire lifecycle of its enterprise-wide storage.*

*Most third-party solutions simply take care of one aspect of data – unstructured, legacy, or inactive. This may take care of one immediate pain point, but it simply removes a problem without adequately addressing its underlying causes – exposing organizations to the likelihood that they will have to address the issue again and again. DocAve is truly unique in that it comprehensively addresses all three types of data, and also ensures that it is properly protected, managed, integrated, and optimized in order to address all organization-specific service level agreements, compliance regulations, best practices, and IT Governance policies. As an example, let's take a look at how each of the tools in the DocAve Storage Optimization Suite – DocAve Extender, Connecter, and Archiver – as well as additional infrastructure management solutions from AvePoint can help not only to stabilize the current storage issues SharePoint deployments face, but also ensure it is robustly protected and seamlessly managed:*

- ***DocAve Extender*** *automatically routes BLOBs out of the SQL Server content databases, immediately alleviating SharePoint's engine from unnecessary unstructured content, enabling it to run more efficiently.*
- ***DocAve Connector*** *attaches existing data and content throughout the enterprise – be it on legacy file shares, cloud storage locations, or local Home Drives – to SharePoint* **without the need for import**, *exposing the content to all of SharePoint's industry-leading presentation and management capabilities without unnecessarily filling valuable space in SQL Server content databases. The* ***DocAve Connector SDK*** *enables the development of custom applications to stream documents, audio files, and video clips in bulk to selected SharePoint locations.*
- ***DocAve Archiver*** *then adds intelligent content lifecycle management to the storage picture, using SharePoint-aware archiving rules to move any SharePoint data or connected data to lower-cost storage or HSM investments already made by an organization.*

*Regardless of whether the data is still in SharePoint's SQL Server content databases, connected from existing legacy file shares, or archived on HSM devices, DocAve enables organizations to further optimize its SharePoint infrastructure management, including:*

- *Relocate and restructure all documents, sites, and sub sites with ease, either within the same SharePoint farm or multiple farms with* ***DocAve Content Manager****;*
- *Synchronize content – and all its requisite metadata – among multiple, geo-distributed SharePoint locations according to pre-defined business schedules or on-demand with* ***DocAve Replicator****;*
- *Backup and restore all externalized data, content, customizations, and farm-level components of your SharePoint deployment – from the platform level all the way down to individual items – with* ***DocAve Backup and Restore****;*
- *Administer permissions and configurations quickly and easily across sites, site collections, and farms from a single interface with* ***DocAve Administrator for SharePoint****;*
- *Record and monitor all SharePoint user interactions – from the entire portal down to a specific item – with* ***DocAve Auditor****;*
- *Search all active and archived content for legal review, and place legal holds on discrete artifacts with ease with* ***DocAve eDiscovery****;*
- *Store all SharePoint data and content in a structured, auditable way while proactively complying with evolving legal and regulatory requirements with* ***DocAve Vault****;*
- *Gather and act upon real-time intelligence regarding SharePoint infrastructure, platform, and usage with* ***DocAve Report Center****.*

*While storage optimization may be a starting point, DocAve ensures that as your SharePoint use evolves and grows, it can nimbly address your changing needs.*

# Conclusion

*As companies look to fully embrace SharePoint Server 2010, it is critical that proper planning takes place with regard to efficiently managing storage resources and content lifecycles. When first deploying SharePoint, organizations might not consider this a top concern. However, as usage grows, proper planning in this regard can reap demonstrable benefits that can be immediately realized enterprise-wide.*

*Native functionality within SharePoint Server 2010 helps to address many of the challenges, but they are only some of the required tools, not the complete solution to help solve the challenges of dealing with unstructured data, legacy data, and inactive data. AvePoint's Storage Optimization Suite completes the storage and content lifecycle management picture, so organizations can take full advantage of SharePoint for success. Furthermore, by utilizing the rest of the DocAve Software Platform, organizations can ensure that all its content – whether remaining in SQL content databases or offloaded to HSM investments – can be properly stewarded:*

- **Data Protection** – *Deliver comprehensive data protection for SharePoint Server 2010, with support for granular, item-through-platform level backup, and swift, full-fidelity recovery of SharePoint content and architecture regardless of whether it is stored within SQL content databases or HSM investments.*
- **Administration** – *Enable the unified management of all SharePoint Server 2010 permissions, users, objects, and content from a single interface for multiple SharePoint environments and instances.*
- **Reporting & Testing** – *Utilize a fully customizable dashboard to access all mission-critical analytics regarding SharePoint Server 2010 infrastructure, health, and activity, as well as resources to create fully customizable SharePoint testing environments.*
- **Replication & Integration** – *Regardless of where content is stored, DocAve offers real-time one-way, two-way, and one-to-many synchronization of SharePoint Server 2010 content and configurations, within or across farms, so all end-users have access to the most up-to-date information.*
- **Regulatory Compliance** – *Comprehensively audit, flexibly report, and automate lifecycle management in order to meet all regulatory obligations and best practice protocols, as well as supporting a culture of proactive compliance for SharePoint Server 2010 deployments.*
- **Migration** – *Ensure that you can perform a secure and lossless migration to SharePoint Server 2010 from prior versions of SharePoint or a variety of legacy platforms – including Lotus Notes and Documentum.*

*AvePoint Global Headquarters*
*3 Second Street*
*Jersey City, NJ 07311, USA*

*Phone: 1.201.793.1111*
*Fax: 1.201.217.8709*

*For more information, visit www.AvePoint.com.*