



Storage Optimization for SharePoint

by David Chernicoff

sponsored by



CONTENTS

Chapter 1

Dealing with Existing and Legacy Data

page 3

Chapter 2

Optimizing SharePoint Database Storage Performance

page 5

Chapter 3

Dealing with the Success of your SharePoint Installation

page 7

Chapter 1

Dealing with Existing and Legacy Data

In many ways, Microsoft SharePoint Server has changed the way people work and how business processes are handled. While collaborative computing technologies have been around since the early integration of the desktop computer in the workplace, SharePoint use has greatly expanded the number of businesses looking for a competitive advantage by gaining better control over company information and workflow processes.

Much of the advantage gained by using SharePoint is due to the powerful data manipulation and management capabilities provided by the SQL Server backend that SharePoint exploits. Data management is what databases are for. By utilizing a full-blown relational database as its backend SharePoint can seamlessly deliver data manipulation capabilities to end users. However, this also highlights the biggest impediment to getting the full benefit of SharePoint; legacy data in the enterprise does not reside in SQL Server databases.

The traditional method of dealing with legacy data is to migrate necessary data directly into SharePoint. Done manually, this is a time-consuming practice even though many tools are available to do this, both within SharePoint and from third-party vendors. A significant industry has sprung up that is dedicated to migrating data into SharePoint. Automating the process and allowing existing business information to be placed into the SharePoint database as quickly as possible is the goal of these third-party tools, and many of them do an excellent job of it.

Unfortunately, this doesn't address a fundamental question that all businesses planning on implementing SharePoint should ask: "What legacy data needs to be moved into the SharePoint database?" For an established business, this becomes a critical question to answer. While storage itself is relatively inexpensive, you have to consider how much data you want to put directly into the SharePoint database as active information. The larger

and more complex the database backend, the more powerful the server hardware you need to deliver acceptable performance. And this translates into how much a business is willing to spend on the hardware infrastructure for SharePoint.

If this was the only issue it would be easily resolved; IT has a long history of the brute force approach to problem solving. Even though budgets have gotten tighter, money can still get thrown at projects to provide solutions when there is a clearly identifiable result from that approach. And since more powerful hardware (servers and storage) to run the SharePoint infrastructure will show benefits in the performance of the application and server software, this approach often gets a lot of adherents.

But what this doesn't address is the fundamental question of what data needs to be migrated into SharePoint and the new workflow model. While the easy answer is "everything," that is rarely practical in an established business of any size. There is simply too much data, often located in too many places, for a SharePoint installation to be all inclusive.

So what is the most practical solution? Let's look at the main reasons that SharePoint is adopted:

- **Document management**
- **Workflow management**
- **And, to a lesser extent**
- **Versioning**
- **Alerting**

SharePoint provides powerful tools to address these issues. From an IT perspective, any solutions delivered once SharePoint is adopted have to meet the same level of usability, manageability, and availability as the core SharePoint service. At first glance, putting everything end users need into the SharePoint database seems like the most reasonable solution. After all, SharePoint has been adopted not just to improve workflow but also to give IT control over the management of the data as it flows through the organization.

But from the end-user perspective it is important to remember this one thing: users don't care how the data is handled, managed, or provided. They just want it available when, and where,

they need it. They don't want to hear that the data they need has yet to be migrated into the system or that it won't be available when it is needed. This is especially true if users know how to access that data outside of the workflow process (i.e., the data resides in the enterprise at an accessible network location).

Because this data is available in the enterprise, the simplest solution would seem to be migrating the necessary data to the SharePoint environment. That seems simple enough until you think of users making dozens if not hundreds of data migration requests during a normal work week. Handling the migration requests then becomes a project of its own and it certainly isn't moving the business forward if users are waiting for IT to respond to these requests to do their own jobs.

Because it is likely impossible to respond to user requests for data migration in a timely enough fashion not to impede their workflow, users will end up working around SharePoint and finding the content they need on the company network. While this doesn't sound inherently bad it tends to defeat the purpose of SharePoint. Individual users end up developing their own workflow processes to address what they see as a flaw in the SharePoint experience.

To avoid this problem, you need a tool such as AvePoint DocAve Connector, which allows IT to manage resources that are not stored in the SharePoint database as if they were, and to present those resources to the SharePoint user transparently. You need the ability to connect users to any data that they would normally be able to see. This means that network shares, filers, FTP sites and any other network-connected storage medium (including cloud storage) needs to be made available within the context of the SharePoint service and, from the point of view of IT, needs to be able to be manipulated using existing SharePoint data management tools.

Doing this successfully isn't simply a win for IT; it provides significant business benefits as well. At the backend, it isn't necessary

to have a SQL Server that can handle every piece of legacy data that might be necessary, along with ongoing business data. This means that you can reduce significantly the amount of storage dedicated to SQL Server and you can implement a fast, efficient SQL Server backend for SharePoint at a lower capital expense. If existing servers were already available, you can direct the SQL Server capacity at growing business requirements rather than at addressing legacy concerns.

When SharePoint treats all the legacy data as if it were stored within the database it means that the tools SharePoint provides for workflow and document management are extended to include the legacy data even though it resides elsewhere. End users are able to treat this legacy data the same way that data in the SQL Server backend is handled, and apply all of the features of the SharePoint system to this legacy information. This means that searches, workflows, alerts, tags, and so on can all be applied to the data as it currently exists without the time and expense of moving it all into the SQL Server backend.

And let's not forget the bottom line; the entire issue of migrating data into your SharePoint installation is made moot. The expense—in terms of time, IT resources, and actual money spent—of delivering the full suite of SharePoint services to users is reduced significantly. IT and business decision makers can focus on getting the competitive advantage that they are looking for from their SharePoint installation rather than spending time and effort on moving existing data into the SharePoint environment.

The right tool means a faster and more effective implementation of SharePoint that allows seamless integration of existing business data and information. More easily realized benefits to end users with a major reduction in the IT effort necessary to deliver the services that those end users need is the result of taking advantage of products such as DocAve Connector.

Chapter 2:

Optimizing SharePoint Database Storage Performance

To a very large extent what makes Microsoft SharePoint successful is the end user experience. Obviously, this can be said about a lot of application software, but the difference is that SharePoint has a positive impact on the way that end users go about everyday work tasks. SharePoint allows users to get their work done more easily and, therefore, more effectively. This provides a better experience for the end user and a measureable benefit to the business, be it directly on the bottom line or in more indirect ways related to improved user productivity.

Once an organization fully utilizes the benefits of SharePoint in its business processes, anything that impacts SharePoint performance in a negative way often has unexpected impact on the overall business workflow. Issues that slow SharePoint performance can often have a cascading effect, as the small delays at different project steps build into noticeable delays in the delivery of work.

This brings us to a fundamental issue with the SQL Server backend used by SharePoint; relational databases are designed to handle structured data. Unstructured data, such as that created in bulk by SharePoint (known as BLOBs or Binary Large Objects), can be difficult for the database server to process efficiently. Because as much as 80 percent of the data being stored by SharePoint can be in the form of BLOBs, this leads to poor database performance, which has a direct impact on the end user experience. The problem for SQL Server is twofold; first, it needs to deal with a huge amount of unstructured data and second, the necessary storage for those BLOBs can be huge.

Microsoft is aware of this potential problem and addresses it by allowing developers to make use of the external binary large object store provider (EBS Provider), which makes it possible to store and retrieve BLOBs from an external (to the SharePoint

installation) storage device. The service provided by Microsoft is a very basic one, and it has noticeable operational limits and trade-offs that are well documented on the Microsoft SharePoint Developer site.

The EBS Provider adds significant housekeeping and management chores when used in its most basic form. It does, however, allow the problems of dealing with the storage and management requirements of BLOBs within the SQL Server database to be addressed. Removing a BLOB from the database and leaving only a stub that allows SharePoint to link to the externally stored data within its applications has a positive impact on the performance of the SharePoint installation, at the cost of additional management effort by the IT staff.

AvePoint's DocAve Extender builds on top of the basic Microsoft EBS Provider to deliver a flexible BLOB manager that goes far beyond the constraints of the Microsoft solution. The result is a flexible tool that gives IT detailed control over how BLOBs are stored within their computing environment. The first point to note is that regardless of where the BLOB is stored it is still completely usable and manageable from within the standard set of SharePoint features. Even though it has been offloaded from SQL Server storage the BLOB is fully addressable with all collaboration and management tools within SharePoint. From an end user perspective, there is no difference between a BLOB that has been offloaded and one that is still stored within the SQL Server database. All security and file permissions are applied as if the BLOB had never been moved, and unique and inherited permissions will work the same as would be expected. As far as SharePoint is concerned, all metadata, both standard and custom metadata types, remains unaffected by the actual location of the BLOB.

Data in the BLOBs remains fully searchable by SharePoint's full-text search and search APIs used by other products. Microsoft Office and third-party applications integrated with SharePoint continue to behave as if the BLOB were stored within SQL Server.

All of these features apply even though the BLOB itself never resides within the SQL Server database storage. The BLOB is analyzed and acted upon by the AvePoint software as it is being uploaded to SharePoint by the end user. And this isn't an "all or nothing" process; Extender allows for the creation of fully customi-

zable triggers to give SharePoint administrators complete control over which BLOBs will reside in local SQL Server storage and which are stored external to the database. This rules-based approach to BLOB management, based on file size, allows BLOB storage to be tailored to meet the exact needs of the business process.

This flexibility extends to the BLOB offload site as well. Any network-addressable storage location can be used, be it local network storage or an Internet-delivered cloud-based offsite storage solution. This very flexible approach means that the most appropriate storage medium can be used for the content, giving administrators detailed control over the location of the BLOBs being stored.

Utilizing this external storage capability with the enhanced functionality of a tool such as DocAve Extender brings both technical and business benefits to the SharePoint environment. The potential for disruption of the end user experience due to performance issues with the SQL Server database is limited by moving BLOB storage out of the database. SQL Server performance is more easily optimized and the SQL Server environment is more easily tailored for the SharePoint server needs when the necessity of dealing with

the huge amount of data that BLOBs represent is removed from the database. Without the need to support storage of BLOBs, the SQL Server can be sized appropriately for a new installation.

The actual location of the BLOB data is of no matter to the end user. The end user experience is the same regardless of the storage location. While performance issues with external storage will impact the overall performance of the SharePoint installation, it is far simpler to diagnose performance problems of a dedicated storage device than of the combination of SharePoint and SQL Server. Hardware storage performance issues are generally more easily defined and solved than those relating to complex software installations.

For an existing SharePoint installation it is possible that users will see an improvement in overall performance as the workload on the SQL Server is reduced for normal, line-of-business processes. For a new SharePoint deployment, installations utilizing the EBS Provider and a good third-party tool may prevent this common performance bottleneck from ever becoming an issue, removing the IT headache and limiting the potential for a negative user experience due to SQL Server performance problems.

Chapter 3:

Dealing with the Success of your SharePoint Installation

The Catch-22 of a successful SharePoint deployment is the success of the SharePoint deployment. When an organization really gets buy-in from SharePoint users it will find itself in the position of having to deal with a veritable data flood. Everyone will be putting all sorts of documents and data files into their site to make them available for other SharePoint users and to glean the benefits of the workflow and collaboration features that SharePoint provides.

With success the SharePoint environment may quickly reach the limits of its SQL Server backend, and the typical solution is to deploy additional SQL Servers to handle the additional load and data, although a major percentage of the need is simply to handle the amount of data and not the number of database records. Adding SQL Servers to your environment is not an inexpensive task in any situation, requiring storage, servers, and software along with the capital investment and the necessary IT resources to add and integrate additional servers.

The solution to this problem is a simple one: free up space on your SharePoint SQL Servers. How to do that is the tricky part. Archiving data from your SQL Servers is a common solution to aging data out of a production SQL Server environment. Generally considered as part of a backup solution, archiving has the advantage of being able to move data from the production server to a near-line or off-line data storage solution and still be able to more easily recover the data to the production server than if it had been moved strictly to a data backup. Even the minimal use of archiving to move older data off the production server can be used to free up large amounts of space in a SQL Server database, giving that server a corresponding increase in overall performance.

An ideal archiving solution allows for any archived data needed to be available to the end user with little, if any, action beyond selecting the file needed. With SharePoint, an archiving solution

also has the benefit of limiting the chance that an end user will exceed quota limits. Regardless of the size of the data (and when you consider the prevalence of video or image data in the workplace, that can be quite large) the impact on the user quota remains the same once the data has been moved to an archived state.

Archived data should be searchable and available to any of the normal SharePoint data management and manipulation tools an end user would utilize. If this isn't the case there would tend to be an "out-of-sight, out-of-mind" approach to the data which would often result in end users duplicating efforts to re-create something that had already been done, but not easily located. This means that full content index and search support is a requirement.

The process of using data that has been previously archived has to be seamless. Users should have flexibility in restoring their own data that may have been moved to archived locations; conversely, administrators should have detailed control over who can restore what archived data. Users should be able to view archived data without having to restore it to the production server, which allows administrators the security of knowing that users will have had the opportunity to double check that only needed data is being moved back onto the production environment. Moving an archived item back to the production server should be a separate task from simply viewing a document.

It shouldn't be a multi-step process to move data back to the production server. Authorized users should be able to do so with as little impact on their workflow as possible. And the information should be restored as quickly as possible, with no intermediate steps or a need for further administrative approval.

The archiving technology needs to be business rule-driven, well beyond the "if it's more than X days old, archive it" approach, and it should allow detailed control over what is being archived, whether it's an entire site collection, a single document, or a specific version of a document. Rules should be easily defined and there should be good administrative control over the archiving process. Reporting and web-based overall management tools that allow administrators to get an overview of what is happening with the archiving process are also important, along with

complete support for internal SharePoint tools and features. While the archiving solution should provide reliable analysis and automation of the archiving process it should also allow for manual archiving, with a simple process for a document or groups of documents to be easily archived or marked for future automated archiving without needing to adjust the business rules driving the process.

Tools such as DocAve Archiver meet the requirements for a truly usable and practical archiving tool for SharePoint. Like the other DocAve tools that we have talked about, DocAve Archiver is able to make use of any visible storage as the location of the archived data, be it locally attached, network attached, or even cloud storage. DocAve Archiver can integrate with EMC Centera storage systems and when used in this fashion it is able to make use of the data de-duplication feature of the EMC storage model and realize further efficiencies in data storage. Making use of the tiered storage model that may already be in place for other network data provides additional cost savings and, in some cases, better integration of the SharePoint environment with other network applications.

The business case for an efficient real-time archiving solution is a simple one: it saves money in the long run by making better use of expensive SQL Server and first-tier storage resources. Short-term benefits exist as well, with a reduction in administrative costs when perceived problems cause end users to contact IT about the SharePoint server.

Improved SQL Server performance and additional headroom in the database translates into improved SharePoint performance and additional flexibility for the SharePoint environment,

enabling it to support additional users, sites, and projects without additional backend expenses.

Implementing an archiving scheme that empowers end users by allowing them to archive and restore their own data without the need to contact an administrator improves the user workflow process, lets users be more productive, and reduces the load on expensive IT personnel.

Comprehensive infrastructure management for your SharePoint deployment will allow your business to get the best from your SharePoint environment in the most cost-effective and business enhancing way possible. Staying ahead of potential problems in the management of your SharePoint data prevents bottlenecks and performance issues that can interrupt your business workflow.

Storage optimization means getting the most from what you have, not simply as an economy measure but as part of the basic utilization of your SharePoint infrastructure. Optimizing your SQL Server storage has the happy result of improving the overall performance of your SharePoint environment and improving the end user experience. By making use of applications like the AvePoint DocAve suite, which extend the capabilities of SharePoint while improving the user experience, SharePoint administrators are able to deploy more effective and usable environments.

Technology alone doesn't solve the problem. Understanding how technology should be applied and how it can best be used, in this case achieving the major payback gained by optimizing your SharePoint storage environment, is the difference between a usable SharePoint environment and one that can become a game changer for your business.