white paper

Best Practices for SharePoint Backup and Recovery

April 2009

This white paper is intended to aid IT administrators responsible for managing Microsoft® SharePoint® deployments in planning and implementing a comprehensive, reliable, and efficient data protection strategy appropriate to their organizational needs. It outlines the planning, guidelines, and implementation considerations for SharePoint backup and disaster recovery, then briefly reviews the singular attributes of DocAve® *Backup and Recovery*.

AvePoint®

Unleashing the Power of SharePoint™

Table of Contents

Executive Summary

Organizations large and small are rapidly adopting Microsoft SharePoint as a standard platform for their online collaboration, portal, and other mission-critical services. As end-users increasingly utilize SharePoint for regular business activities, organizations are confronted with the consequences of exponential growth in the volume of business-critical data residing in the platform, as well as its overall footprint. As a result, it has become crucial for today's competitive organizations to be vigilant, and prepare for unplanned disasters via the creation of robust data protection and recovery solutions.

With a complex and distributed SharePoint deployment model, and with IT resources becoming increasingly constrained, organizations need an efficient and comprehensive disaster recovery solution to protect the wide variety of data and components that constitute their SharePoint farms. This solution must be able to satisfy the most stringent business requirements and service level agreements, and have the capacity to scale effectively to maintain performance as the SharePoint footprint expands.

DocAve *Backup and Recovery* addresses these challenges by providing a granular, full-fidelity, item- through platform-level backup and disaster recovery solution for SharePoint. This white paper first outlines the planning, guidelines, and implementation considerations for SharePoint data protection and disaster recovery, then briefly reviews the singular attributes of DocAve *Backup and Recovery*. This document is intended to aid IT administrators responsible for managing SharePoint deployments in planning and implementing a comprehensive, reliable, and efficient data protection strategy appropriate to their organizational needs.

Evaluating Your Disaster Recovery Strategy

Successful SharePoint deployments require effective management, vigorous protection, diligent compliance protocols, and continuous availability. As companies increasingly rely on SharePoint to store business-critical digital assets, these requirements become crucial catalysts for optimal production and minimized exposure to costly downtime and data loss. The fundamental question for every organization operating in today's 24/7 economy is: How prepared are you for interruptions to your mission-critical SharePoint environment? Success in today's demanding business environment requires organizations be properly equipped to maintain continuous, uninterrupted access to business-critical digital assets.

Anchoring a Successful SharePoint Backup Plan

How confident are you with the disaster recovery strategy you currently have in place to protect your SharePoint environment? A comprehensive SharePoint disaster recovery strategy must seek to achieve the following goals:

- Reduced backup windows;
- Satisfaction of acceptable Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO);
- Comprehensive protection of all SharePoint farm-level components

Reduced Backup Windows

With the escalating amount of data being stored within the SharePoint repository of today's organizations, getting backups to complete in a timely manner has become an ongoing challenge. Long running backups often have serious ramifications, including performance-loss, inefficient use of system resources, and increased likelihood of data loss as fewer recovery points are taken. Additionally, because 'open' files residing in system memory will likely be excluded from backups, elevated levels of data inconsistency are a consequence of long running backup processes. In order to reduce backup windows, there are several tactics that can be implemented, independently, or collectively:

- Reduce the volume of data being backed up;
- Perform backups incrementally;
- Execute backup jobs based upon a more business-appropriate schedule.

The most business-appropriate strategy for a given organization generally includes some combination of all the above-listed tactics, so let's review ways to approach each:

One of the most effective ways to reduce backup windows is to reduce the amount of SharePoint data included in a given backup. To achieve this – while still robustly protecting digital assets – organizations must granularly determine the business-criticality of each specific site, list, and/or document library within the

deployment. When the criticality of each of these datasets has been determined, then appropriate levels of protection (i.e. frequency of backup executions) can be prescribed. This 'tiering' process empowers organizations to appropriately differentiate their data, so content more critical to the organization is backed up more frequently, while data of relatively less business-importance will be backed up with less frequency.

A second tactic for minimizing backup windows is to perform full backups less frequently. During incremental backups, only content that has been updated *between* backup intervals will be selected for backup, drastically reducing the size of the backup file created. This ensures data remains protected, but reduces execution time, thereby allowing for shorter backup windows.

A final tactic we'll discuss for reducing backup windows is the use of appropriate scheduling. When backups are executed during off-production hours, network resources can be exclusively dedicated to the task. In contrast, backups performed during production must share resources with end-users and other infrastructure processes. It is, however, also possible to schedule backup processes during production hours with the appropriate choice of a lightweight backup methodology that does not place additional strain on the system. A necessary component of this tactic, of course, is the ability to schedule backups. Therefore, a robust data protection strategy should include tools with appropriate scheduling capabilities.

Satisfying Recovery Time and Recovery Point Objectives

One of the primary objectives of data protection and disaster recovery planning is to mitigate disruptions to business-continuity. When analyzing tactics to ensure business-continuity, recovery of lost or corrupted data is the fundamental imperative. The key concern is to get the data back, so work can commence. Interrupted access to content residing on SharePoint can – and usually does – have serious implications for revenue generation, end-user productivity, and other 'bottom-line' business objectives. Two recognized measures with regard to data recovery are Recovery Time Objective (RTO) and Recovery Point Objective (RPO).

- Recovery Time Objective (RTO): The duration of time within which a business process must be restored after a disaster or data loss in order to avoid unacceptable consequences associated with a break in business continuity;

- Recovery Point Objective (RPO): The acceptable amount of data loss measured in time.

Occasionally, data loss is the result of catastrophic failure, such as fire, flood, or hardware malfunction. For these occurrences, it is critical for administrators to maintain a platform-level recovery strategy, which we will address later in this document. But more often, content loss is a result of accidental deletion or discrete corruption, and can be resolved without a full-system recovery. In these instances, SharePoint administrators can minimize recovery time and meet a business-appropriate RTO by implementing recovery strategies that deal at a granular level. If only select data needs recovery, full-system recoveries derived from a complete backup are inefficient, as they take an unnecessarily long time to perform, and

utilize system resources uneconomically. Rather, in these circumstances, an optimal recovery strategy would allow administrators to target and recover only those lost or corrupted objects directly to production. This granular control with regard to data recovery is a key to meeting aggressive RTO's.

Satisfaction of a business-appropriate RPO, alternatively, is a function of backup frequency, and strategies to satisfy business-appropriate RPO's center primarily upon the intervals between backups. Whether the situation involves a catastrophic failure, or the targeted loss of content, satisfaction of RPO is dependent upon backup interval. This dictates that administrators implement data protection routines that allow for backups at intervals in-line with the organization's acceptable data loss. Depending on a given organization's business needs, this might mean that hourly backups are performed and can be retrieved on demand, thereby establishing an RPO of 1 hour.

As we will discuss later, it is important – when discussing RTO and RPO objectives – to keep in mind that true data protection includes not only recovery of primary content, but also all associated metadata, version histories, and access permissions. Because SharePoint is largely a collaborative platform, established RPO's and RTO's must include full fidelity data recovery in order to be truly effective. Administrators must ensure their data protection strategies provide for complete preservation of all content metadata.

Comprehensive Support of Farm-Level Components

A SharePoint farm consists not only of the sites and content that reside in the SQL Server databases, but also additional farm-level elements that must be backed up accordingly, in order to provide comprehensive protection against disasters. These elements include the configuration and central administration databases, Web applications, Shared Services Providers, the Search Index, IIS configurations, and other customizations likely residing in locations on the front-end Web servers commonly referred to as the "12 Hive". Without properly protecting these components, a complete recovery of a SharePoint environment would require extended downtime, complex reconfiguration, and countless manual tasks prone to human-error. It is therefore recommended that any disaster recovery strategy provides adequate protection for the full spectrum of SharePoint farm-level components.

Now that we have reviewed the key considerations an organization must take into account during its SharePoint data protection planning, we can now look more closely at both the anatomy of the SharePoint platform and its corresponding content, and discuss how each can be appropriately protected.

Determine What Needs to be Protected

Everything within a SharePoint environment needs to be protected, but no single strategy will serve as an appropriately efficient solution for the entire deployment. Rather, strategies should be applied to aspects of the deployment based on an understanding of its role, its business importance, and its means of restoration. The first step in taking these measures is to break down your SharePoint assets as follows:

- **Core content**: This refers to all the site collections, lists, document libraries, list items, documents, discussion threads, etc., residing in the deployment. Core content is stored within a SQL Server content database, and is accessible and viewable by the end-user, regardless of whether that end-user is an administrator, a site owner, content owner, or a limited-rights user. From the organizational/business perspective, core content represents the most critical data within SharePoint. It is an organization's sales leads, contact information, financial statements, and other mission-critical data that is stored in the content database, and any loss of such data would likely result in significant pain for the organization.

- **Platform components**: This refers to all the infrastructure elements required for SharePoint to function properly. These components are generally not accessible by end-users, but rather represent the functional architecture of SharePoint. These include the configuration and central administration databases, Web applications, Shared Services Providers, the Search Index, IIS configurations, deployed solutions and other customizations located on each of the front-end Web servers. The SharePoint administrator is typically responsible for managing and configuring these elements, and loss of such would likely result in platform access disruption.

In developing an appropriate backup strategy, it is recommended to approach these two classes of SharePoint assets independently. Put simply: The protection strategies appropriate for one class might not be so for the other. In light of the concerns we addressed in the previous section, we can see how relying on only farm-wide 'snapshot' backups – though simple to execute – would not deliver the protection required (as measured in RTO and RPO) for core content at a typical organization. Alternatively, a strategy that can recover core content efficiently might not enable a quick restoration of the platform's complex components and related configurations. Any robust disaster recovery plan should provide optimal support for protecting both the core content as well as the complex platform-level components.

Analyzing SharePoint's Taxonomy

A properly designed taxonomy is a prerequisite for an efficient SharePoint deployment. By implementing an effective content taxonomy, knowledge workers will be better equipped to harness the value of SharePoint to organize an unstructured Web of information. An appropriately designed taxonomy is straightforward, and relatively easy to implement with some proper planning. Before content sprawl is allowed to occur, simple tasks can frame an effective taxonomy: Sites and sub-sites should be logically grouped; document libraries and lists should be properly categorized, and information should be properly targeted to the appropriate group of users.

Change is constant in today's business climate, and a well-defined taxonomy in SharePoint serves as an effective framework for organizing the ever-growing and shifting information within an organization. Not only can the many dimensions of a well-formed taxonomy facilitate information access for end-users, but it can also greatly simplify the backup strategy with respect to core content. Administrators

can build granular backup plans more efficiently, when they are targeted as subsets of sites, sub-sites, or libraries whose content are logically grouped by taxonomy. By doing so, administrators can appropriately allocate system and storage resources needed to protect different categories of their functional content. SharePoint content that is deemed of high business importance (Sales Productivity Applications, for example) can be backed up more aggressively, while content of relatively less business importance (facilities manuals, for example) can be backed up at less frequent intervals.

We will discuss this concept in further detail later in this paper, when we review concrete procedures and tools available to target and execute backups based on criticality.

Determine the Volume of Core Content

Content growth places ever-evolving burdens on existing backup strategies. Because of this, the 'ideal backup strategy' is always a moving target, and must be reassessed at regular intervals. It is therefore imperative that SharePoint environments are analyzed in detail regularly, to maintain an accurate understanding of how much, and what type of content needs to be protected. In this process, it is important to consider a variety of factors, including:

- The nature of each content database – its overall size, and the type of content that is stored (e.g. list items, documents, discussion threads, mp3's, etc.);
- Rate of change of content – usage/access frequency, and the volume of content that gets generated weekly or monthly.

While considering the above factors, it is important to keep in mind that content from a given database can be backed up to different tiers of storage. To implement an optimized backup strategy, content objects must be appraised granularly, and backup tools that allow for granular precision are required.

Defining Your Service Level Agreements

As we briefly discussed earlier, a viable SharePoint content and platform protection strategy requires the generation of a business-appropriate Service Level Agreement (SLA). This is often the most difficult aspect of protection strategy development, since one must solidify terms and conditions – in the form of an RTO and RPO – in coordination with management and end-users. A formal document should be prepared that communicates agreed-upon expectations with regard to production SharePoint performance. For the administrator, these agreements will:

- Serve as the principles that frame the details of a backup plan;
- Justify the budget and resource allocation requirements for the implementation of such a plan.

Throughout the process of developing a formal SLA document, the administrator must also serve as the 'reality check'. Often management and end-users maintain unrealistic expectations of achievable service levels. In these circumstances, the level of service desired does not match the level of financial and time commitment these parties are willing to invest to achieve such service levels. Managers and end-users often expect a guarantee of near-zero data loss, or continuous uptime, without committing the energy and resources needed to realize such guarantees. It is often the role of the administrator to inform management and end-users of the true costs related to a given service level agreement.

In addition to explicit RTO and RPO measurements, a well-defined SLA for SharePoint should identify:

- *Stakeholders* – These individuals and groups include not only management and end-users, but also the personnel that will play key roles in the event of a disaster. The roles of IT personnel should be clearly defined, and appropriate communication channels established by and between all affected business groups. (E.g. executive staff, end-users, IT personnel.)
- *Recovery Scenarios* – All nature of comprehensible recovery scenarios should be explicitly defined to the greatest extent possible. At the very least, these scenarios should include: What will be done if the entire production environment is down? If only a single site is lost, how will a single site restore be performed? Can lost documents be recovered without performing full farm recoveries? If extended downtime is expected, can critical documents be restored to a file system for temporary access?
- *Data Integrity* – End-users must know what to expect when content is recovered to the SharePoint environment. The SLA document must explicitly define what users should expect with regard to object metadata perseveration, access permissions, and versions. For business-productivity and compliance-related purposes, management should have an explicit understanding of how both security and content (including metadata) will be affected.

- *Measurability* – All services described in an SLA document should be described in measurable terms. With regard to RPO and RTO measurements, service availability should be defined as a percentage of uptime, while recovery time should represent the average time end-users can expect content to be restored. A helpful measurement with regard to platform availability is presented as a function of *Mean Time Between Failure* (MTBF) and *Mean Time To Recover* (MTTR) using the formula below:

$$Availability = \frac{MTBF}{(MTBF + MTTR)}$$

As depicted, minimizing MTTR maximizes availability. This relationship helps illuminate the importance of strategies that include the ability for granular recovery of content. In addition, during downtime, how much content will end-users expect to lose? This can be determined by the backup frequencies covered by each plan. For example, if a document library is protected by an incremental backup plan that runs every hour, end users can expect to lose at most an hour's worth of data should a disaster occur.

With these factors addressed, an SLA document will serve as a valuable blueprint and guidebook for all affected parties in the case of data loss or platform failure. But an SLA is a 'living document', and to be effective it must never remain static for too long. Appropriate and regular revision of an SLA – taking into account changing volumes of content and the business-criticality of each dataset – is vital for maintaining optimal data protection and disaster recovery strategies in an environment of constant deployment and organizational evolution.

Deploying DocAve Backup and Recovery

DocAve *Backup and Recovery* provides full spectrum item through platform-level data protection for Microsoft SharePoint. This 'best of breed' solution protects entire SharePoint environments, from individual items and sites to Web applications, content databases, index servers and IIS settings. Importantly, DocAve *Backup and Recovery* delivers full fidelity recovery of all metadata, securities, version histories, and customized layouts. Unique, granular backup scheduling allows organizations to differentiate data and execute backup processes according to business needs and organizational timetables. Additionally, DocAve *Backup and Recovery* automatically detects new SharePoint content and performs scheduled or on-demand full, incremental, and differential backups upon this content. With DocAve *Backup and Recovery*, organizations can be confident their SharePoint environment is optimally protected, and their SLA's can be adequately satisfied.

Overview of the DocAve System Architecture

DocAve *Backup and Recovery* employs redundant and fully distributed components to ensure continuous uptime and failover for any DocAve processes being executed. While employing an agent-server deployment model that allows it to scale across multiple SharePoint instances and versions, each of the individual components can be broken down into services for further workload distribution, while enabling centralized management and control from a single browser-based interface, accessible from anywhere within the organization's internal and external networks. This architecture is depicted in Figure 1 below.

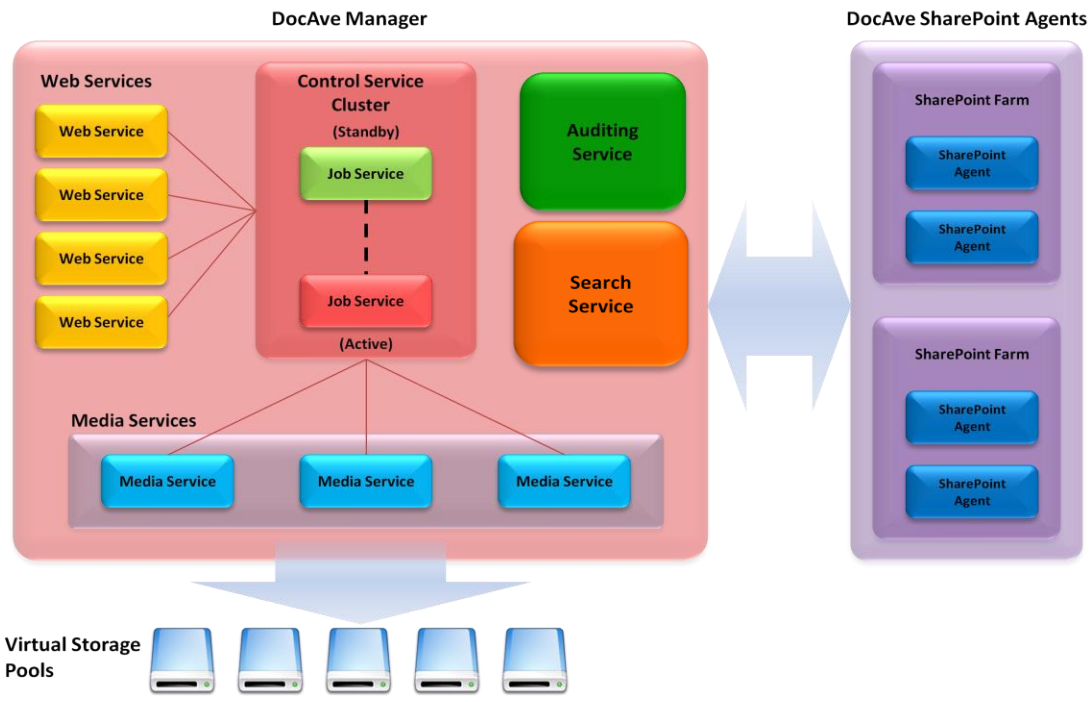

Figure 1: Fully Distributed DocAve v5 System Architecture

The solution is composed of three main components, each designed to maximize workload distribution while providing flexible, intelligent, and granular backup and restore functionality.

As depicted in the figure above, the DocAve Manager Service consists of load-balanced Web services that manage requests/responses, and an active-passive pair of control services. This highly available architecture for the control service allows DocAve to failure to a standby Job service to continually process any DocAve jobs that have been queued up.

DocAve Manager

The DocAve Manager centrally administers all activities performed by the DocAve Software Platform, and allows administrators to control routines, schedule jobs, and view and act upon any error conditions raised by the agent processes. The Manager can be installed on any Microsoft Windows-based hardware with appropriate processing power and access to its member agents throughout the network. It is recommended to install the DocAve Manager on a machine with high availability, although it does provide an active-passive control services cluster to ensure continuous availability. This highly available architecture for the control service allows DocAve to failure to a standby control service to continually process any DocAve jobs that have been queued up. The DocAve Manager is also the server process responsible for delivering the DocAve Web-based interface to the end-user's browser. The DocAve client agents communicate with the DocAve Manager via lightweight XML control messages over TCP/IP socket connections, allowing one Manager to control multiple agents and representing a distributed network of SharePoint servers across multiple SharePoint farms.

DocAve Media Service

The DocAve Media Service is designated to provide dedicated computing resources to read and write to storage media. For largely distributed deployments, it is recommended to have the Media Service(s) deployed within close proximity of the front-end Web servers and the physical storage, but not on the same hardware. It is also recommended to host the Media Service(s) on hardware with high reliability. Multiple dedicated Media Services can be distributed across multiple SharePoint environments to handle additional process workload, thus improving the overall performance of backup and recovery jobs. Additionally, Media Services also support the ability to persist data to multiple logical storage devices. Media Services can be associated to a limited subset of virtual storage devices, so DocAve can automatically attribute the backup processes through the designated Media Services. By doing so, DocAve can further optimize the storage pools according to both business activity and established SLA's.

DocAve Agent

Each DocAve Client Agent is a lightweight, application-specific client component. In the case of DocAve Deployment Manager, the Client Agent resides on a SharePoint front-end Web server. The operations of the agent are controlled by the DocAve

Manager via XML messages over a network. The client agent reads and streams selected data via the SharePoint front-end Web server.

Multiple DocAve client agents can be directed to communicate with one another, allowing the transfer of data between SharePoint environments. This is the means by which Deployment Manager routes custom elements from one location in a SharePoint environment to a different location. Unique capabilities within the DocAve architecture ensure that communication between agents can endure very noisy or even intermittent data channels. Successful results from strenuous data transfer performance and stress tests have showcased the strength of DocAve's data packet level fault tolerance features.

Considerations for Large-Scale Deployments

For larger SharePoint farms, the distributed architecture of DocAve *Backup and Recovery* allows the solution to scale with the growing needs of the organization. Performance and management of large deployments can be maximized by addressing the following considerations:

Plan for SharePoint Boundaries and Performance Guidelines

The first step to ensuring that data protection and disaster recovery plans can scale effectively is to deploy the SharePoint environment according to Microsoft's recommended performance and scalability guidelines documented on TechNet: http://technet.microsoft.com/en-us/library/cc262787.aspx. This article provides various performance test scenarios and corresponding deployment recommendations for SharePoint environments. The guidelines outline many important considerations, including:

- Recommended size limitations of each content database;
- Maximum number of site collections per Web application or content database;
- Maximum number of documents per document library.

These guidelines help administrators provide management and end-users with accurate platform response times for common user operations. Though each particular deployment necessarily results in variable (actual) response measurements, these guidelines help provide 'baselines' with which to develop appropriate approximations.

Limiting Backup Sizes Per Backup Plan

As we discussed earlier, the amount of data being backed up is inversely related to the duration of the subsequent backup window. Longer backup windows can lead to performance-loss, inefficient use of system resources, and increased data loss/inconsistency. To prevent such issues, a generally accepted 'rule of thumb' is to limit each backup plan to accommodate at most 50GB worth of data. The out-of-the-box granular capabilities and unlimited number of backup plans that can be created within DocAve *Backup and Recovery* provide ample flexibility to support this deployment scenario.

Leverage Multiple Media Services

DocAve *Backup and Recovery* supports the deployment of multiple DocAve Media Services on separate physical hardware. As SharePoint grows and more content requires backup, there is increased likelihood that multiple backup processes will be scheduled for execution at the same time. In order to optimize the Media Services' writing of backup data into storage, it is recommended to deploy additional Media Services to limit each media server's load processing to five concurrent jobs. Additionally, because media services should be deployed within close proximity to the front-end Web servers and physical storage locations, deploying a minimum of one Media Service per geographic region is also strongly suggested.

Use Load-balanced Agent Groups

DocAve *Backup and Recovery* provides built-in capabilities to handle agents within agent groups. Within heavily distributed SharePoint farms, one agent should be deployed and enabled within each load-balanced front-end Web server, so DocAve can automatically identify and assign a backup process to the least loaded server available, in order to accelerate any job already in process.

Optimizing Your Backup Plans

Once DocAve *Backup and Recovery* has been deployed, there are several recommended approaches with regard to implementing backup plans. By fully utilizing DocAve's built-in capabilities - such as granularity, storage pooling, and the automated backup planning "Criticality Matrix" – administrators will be able to deliver more robust protection coverage and satisfy even the most stringent SLA's.

Evaluating Your Storage Requirements

After having determined all assets within the SharePoint environment needing protection, and calculating both the amount and the criticality of content, the next step is to establish specific backup plans in DocAve *Backup and Recovery*. The first step in this process is to ensure that the appropriate storage resources are in place to support anticipated data protection routines. To determine storage requirement, the following questions must be answered:

- How much hardware will be needed to host the backup software, and how much disk space will full backups consume on the storage media?
- What additional type of security options will be placed on the storage media?
- How should pruning-rules be defined for the best possible use of valuable storage space when backups are no longer required?

Answering these questions with accurate estimates is critical to optimizing backup plans. This will also influence how frequently full backup jobs can be executed, and how often incremental backup jobs can run to fill in the gaps. Similarly, by pooling available storage resources, multiple volumes can be managed as a single target backup destination to more effectively handle large data backups.

Estimating Storage Space for DocAve Backups

The actual amount of physical storage space required for SharePoint backups is dependent upon two factors:

- The volume of data (core content plus platform components);
- The pruning rules in DocAve that define the number of backup cycles to be retained.

SLA's should specify the length of time and number of backups that should be maintained to adequately support recovering the variety of content being protected. These time periods will directly affect – and be influenced by – the storage resource requirements described in the section above.

A good starting point for estimating the storage space requirements is the following equation:

$$Required\ Storage\ Space \cong 1.5S \times C + S$$

Where S represents the volume of content to be backed up, and C represents the number of backup cycles to be retained.

- For each backup cycle retained, ensure storage space is allocated for one full backup cycle of all data, plus an additional half-volume to compensate for any data updated during the cycle;

- Using DocAve *Backup and Recovery*, pruning jobs can be executed either before or after a backup job, so additional storage should be resourced for consideration of additional full backup cycle(s) that complete before a pruning is executed.

Another factor which affects storage allocation is the use of compression. When utilized efficiently, compressing backup files can effectively reduce the storage requirements by up to 75%. Of course, how much data can be compressed depends on the type of content. For example, the new Office 2007 XML-based document formats can be compressed extremely well, so significant reduction in file sizes can be obtained. Conversely, certain files that are already compressed, such as JPEG, MP3, or other media files, cannot be reduced significantly further.

A final factor in considering storage allocation requirements is the level of optimization among the various backup plans. Often, content selected for backup can also be present in other 'overlapping' backup plans. This means that the content will be duplicated in storage. Eliminating unnecessary backup jobs, (ie. optimizing backup plans to reduce redundancy) can reclaim storage space and reduce further outlays on wasted storage.

Setting up Business-Aware Backup Plans

Because SharePoint is a business-oriented technology platform, data protection and disaster recovery solutions should always take into consideration the needs of the business. Many organizations may think they have a comprehensive disaster recovery plan in place, but simply having a plan is not enough. These plans need to be applicable to the organization's real-world situation and needs, and appropriately validated. Without such validation, satisfying applicable SLA's and maintaining critical business-processes becomes difficult.

Not all Data is Created Equal

The most challenging business decisions to be made by a SharePoint administrator during the design of an appropriate backup architecture relate to the scheduling of backups for various data types. To achieve this, two criteria need to be evaluated:

- The business criticality of the content;

- The usage frequency of the content.

A useful means for visualizing these criteria, and how they inter-relate, is to map them out on a 2-dimensional matrix. This matrix – known as the *SharePoint Backup*

Planning Criticality Matrix – plots business criticality on the *x*-axis and usage activity on the *y*-axis. Figure 2 (below) demonstrates how a sample organization might construct a *SharePoint Backup Criticality Matrix*. As displayed in this example, sales leads and customer records stored in a Sales Productivity Application (SPA) in SharePoint are extremely important to this organization, these are the 'lifeblood' of the company, and any downtime or loss of such information would have drastic consequences for the company. This type of content is also very frequently updated, perhaps several times each hour. This means that a more aggressive backup plan will be required to limit any amount of content loss due to the frequency of content updates. In the matrix, this content will be associated with the top-right-most cell. Conversely, at the other end of the spectrum there resides content updated relatively infrequently – probably once every six months or so – and of low relative importance to business processes. This content would reside in the lower-most-left cell. In this example, employee guides and vacation policies are represented, as they are generally of less consequence, and modified with less frequency, than sales leads or company financial information.

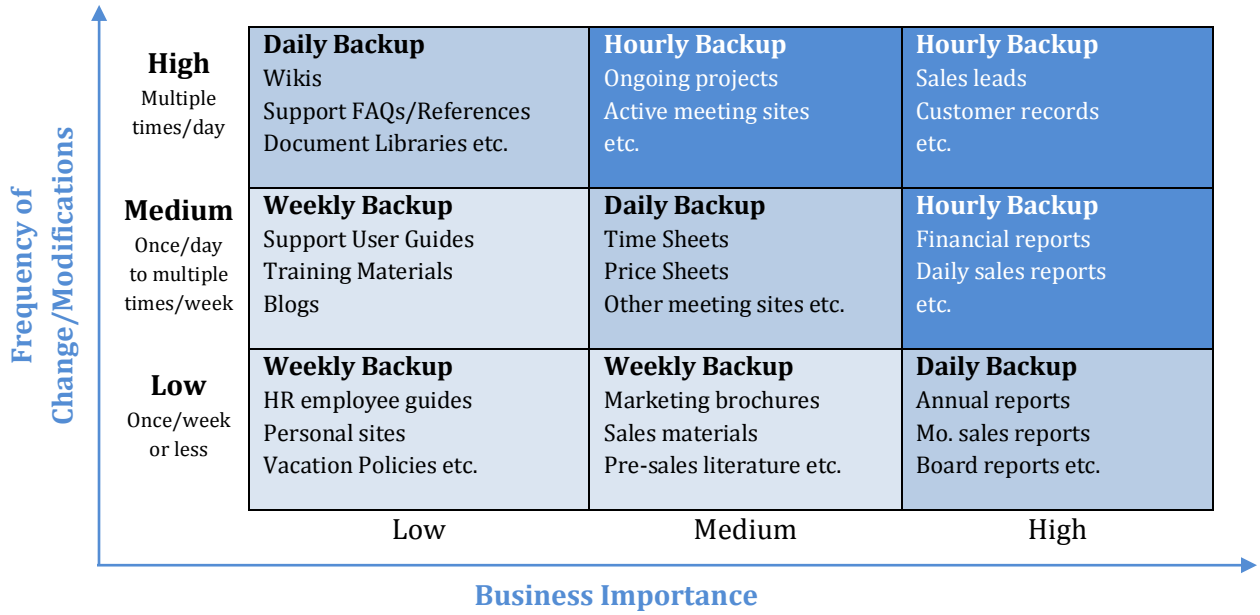

Figure 2: SharePoint Backup Planning Criticality Matrix

Once the criticality of all content has been identified, constructing a backup plan becomes relatively simple. DocAve allows administrators to build an unlimited number of backup plans, consisting of up to six customizable backup schedules per plan. Each backup schedule can perform a full, differential, or incremental backup job, and can be scheduled at specified hourly, daily, weekly, or monthly intervals. This provides administrators with the utmost flexibility in generating the perfect combination of backup schedules in order to cover the entire SharePoint environment.

In the example we used above, the SPA-related content might be protected via a single backup plan. This plan could be generated with DocAve, all SPA-related content – whether it be an entire site collection, individual lists, or document

libraries – could be granularly selected, and multiple backup schedules for this backup plan could be established. The first backup schedule would execute a full backup every Sunday night. A second schedule would cover any content updates more frequently by executing an incremental backup on the hour.

Keeping with the example above, the HR-site residing in the lower-left-most cell of the matrix could be protected by a separate backup plan that executes full backups every week, with incremental backups every Wednesday evening.

By extending this approach to all of the other sites within the SharePoint environment, a modular and flexible backup system can be created that runs according to organizational timetables, mitigates duplication of backups, and ensures the security and integrity of all SharePoint core content.

Expanding on the concepts represented by the Backup Criticality Matrix depicted in Figure 2 of this document, DocAve's *Backup and Recovery* module offers the *Business Criticality Matrix*, (Figure 3, below) which automatically classifies discrete SharePoint content according to business importance and usage frequency. This lets administrators optimize storage and system resources with automated classification, and execute rule-based backups based on real-time item-level data analyses.

Using DocAve *Backup and Recovery*, content owners will have the ability to specify business importance for each site, and the software will automatically prescribe each site to the appropriate cell within the matrix based on both business importance and access frequency. Based upon end-user activity, DocAve will automatically adjust each site's placement within the matrix as criticality or activity changes. For each cell, administrators can apply more aggressive backup plans (hourly or daily, for example) to content covered by more aggressive SLA's.

Figure 3: DocAve 5.0 Backup Criticality Matrix

By automating this classification process and monitoring SharePoint usage activity, DocAve ensures that multi-tiered service level agreements (SLA's) are satisfied and resources are optimized.

Granularity of Backups

Backup plans can be created at the site collection, site, or item levels. Selecting the appropriate level depends on a number of different factors:

The first factor concerns the granularity with which to differentiate discrete data for backups. There are occasions when sites are already logically segregated, and associating a backup plan for each individual site collection or site is simple and straightforward. This would reduce the number of backup plans that need to be managed, and will also more effectively limit the possibility of duplication.

The second factor concerns the restore options appropriate for the content. DocAve *Backup and Recovery* indexes backup files for restoration according to the 'level' of backup chosen. For example, a site-level backup will only allow entire sites to be restored, while individual content items as well as item-versions can be restored from item-level backups. It is important to bear in mind that item-level backup plans also cover entire sites or site collections, as well as the nested sites that get created. So selecting an appropriate level of backup is really dependent upon: How much granularity is required upon restoration, and; the time allotted (as defined in the SLA) for execution of the backup job.[1] We recommend always beginning with item-level backups, as this provides the most flexibility in not just data-differentiation, but also by providing item-level and item version restore capability.

One final consideration in determining backup levels is exactly what data is being restored at each level. Backup files generated from site-level backups are typically larger than similarly targeted item-level backups, as these backups encompass *all* of the assets within that site. This can affect storage resource allocations, as the execution of incremental backup jobs of entire sites will be backed up again, even if only individual content items were updated since the last full backup was executed.

Security Trimmed Access to Backup Data

Security trimmed access to the restoration of SharePoint backup data helps organizations distribute the administrative workload previously held by Farm level administrators. This helps to further optimize IT resources and reduce recovery time and point objectives.

DocAve offers not only tools to perform real-time captures of site-level deletion events (DocAve SiteBin), but also by providing security trimmed access to DocAve for site collection owners, as well as a security trimmed Restore Webpart. DocAve SiteBin performs a real-time capture of sites as they are deleted in SharePoint, thereby reducing the RPO of an accidentally deleted site to zero. Site owners and

[1] Due to the indexing required during item-level backups, time necessary for such execution is longer than a site-level backup of the same site. However, in the case of incremental backups, site-level backups might possibly require longer time periods, as the job will attempt to backup the entire site even if deltas are present in only a single document within that site.

administrators may also perform restores of their own content directly thru DocAve, or via a DocAve installed web-part directly in SharePoint to ensure decreased recovery times and minimize data loss during an accidental deletion event. Site collection owners' AD permissions can be integrated into DocAve to ensure they only see the content that they own. Leveraging this tool to recover a single site, list/library, item, or document from a deleted site means that end-users need not go through the farm or DocAve administrator.

Plan for Probable Recovery Scenarios

With comprehensive backup plans in place to provide adequate protection of the SharePoint environment, administrators should also plan for possible recovery scenarios. Again, granularity is crucial, as any type of recovery activity should not only satisfy agreed upon SLA's, but also negatively impact environment activity in the most minimal possible way.

It is important to keep in mind that recovery procedures can also be delegated to other users with appropriate access permissions to DocAve. For example, IT helpdesk staff - or even the site/content owners themselves - can optionally perform simple recovery of lost or corrupted items, while full platform recovery might remain the sole responsibility of the SharePoint administrator(s).

Some common events that trigger recovery and should be planned for:

- *Hardware failure* – Hardware failure typically means environment downtime. In this case, it is important to be prepared to bring a standby SharePoint farm online as quickly as possible, in accordance with agreed upon SLA's. If the production environment has been configured for high availability, with technologies to update content on the standby environment, content loss can be minimized. (DocAve High Availability provides this functionality.) Otherwise, performing an out-of-place content restore to a separate SharePoint environment can serve as a temporary solution.

- *User error* – Users, developers, and administrators may accidentally delete content, remove entire site collections, or make incorrect configuration changes. User errors are the primary reason for site or content recovery.

- *Viruses, data corruption, etc.* – SharePoint's native *Recycle Bin* feature will not be able to safeguard end-users from corrupted content. Corrupted content must be addressed by overwriting the current file with uncorrupt backups.

- *Data center destruction* – Destruction of a data center requires a full platform recovery, likely onto a separate SharePoint farm that can be brought online in a separate location. Alternatively, in order to better prepare for such a disaster, a high availability strategy, using technologies such as SQL mirroring or log shipping, can be established to create a warm standby SharePoint environment. In this circumstance, content loss would be determined by the frequency of content replication.

With enhancements such as automated, business-centric SharePoint backup and recovery, and a point-in-time restore controller interface, DocAve is poised to again revolutionize the way administrators manage and protect SharePoint.

Protecting Your Farm-Level Components

As we discussed previously, core content is the most significant asset within a SharePoint deployment, as it represents intellectual property critical to operations. Of significant importance, however, is the underlying infrastructure of the SharePoint deployment, consisting of many platform-level components coupled tightly together to create a working environment. Since there are so many 'moving parts' to a SharePoint platform, a failure in any one component could adversely affect the entire environment.

DocAve *Backup and Recovery* provides comprehensive protection for not just the core content residing in the content database(s), but also for the platforms full spectrum of farm-level components and their configurations. All farm elements, including the configuration database, web applications, search and index servers, IIS settings and other file system artifacts on the front-end web server, are robustly protected.

Performing a Full Farm Backup

The user experience during platform-level backup creation in DocAve *Backup and Recovery* is similar to that of site collection, site, and item-level backups. Administrators can schedule backup plans to execute full, incremental, or differential backups on the servers where DocAve agents are installed. Because this is a platform-level backup intended to safeguard the environment against catastrophic events, the resulting backup file(s) should be stored in a location different than the primary SharePoint farm. This ensures that even if the full farm is not functional, access to the SharePoint backups is maintained. Options such as data retention (pruning), encryption, and compression are the same as those described in our discussion of core content backup options. There are, however, a few options specific to platform-level backup:

Selecting VDI or VSS

DocAve *Backup and Recovery* platform-level backup provides an option for selecting either SQL Server Virtual Device Interface (VDI) or Microsoft Volume Shadow Copy Service (VSS) snapshot technology. VDI is a stream-based method for protecting SQL databases, whereas VSS leverages various VSS writers from the Windows Server platform and applications. There are important differences between VDI and VSS that administrators should be aware of before selecting an appropriate backup method. For example, because VSS is a snapshot technology, it has minimal impact on the production SQL Servers, and does not require an interruption of the index crawl during a backup job. On the other hand, VSS does not currently support out-of-place restores, meaning all farm-level components may only be restored to their original locations.

For a comprehensive comparison between VDI and VSS, please refer to the DocAve Backup and Recovery User Guide.

Coexistence with SQL Server, and IBM Tivoli backups

Database administrators typically have their own SQL Server backup plans in place to protect the database(s) in the case of any unplanned disaster. These database servers include those used for custom applications or other operational data repositories, as well as all of the SharePoint content and configuration databases. DocAve *Backup and Recovery* platform-level backups can coexist with institutional database backup utilities, and automatically prevents any conflicts from taking place. Therefore, a recommended approach is to keep SQL Server backups as they are currently staged, and protect only the SharePoint content databases via DocAve.

With the 'Copy Only' option provided in DocAve *Backup and Recovery*, full SharePoint platform-level backups can be performed by taking a 'copy' of the database and associated transaction log rather than committing and truncating the transaction log in the existing database. By doing so, SharePoint administrators have the flexibility to holistically capture a full SharePoint environment via a DocAve backup, as opposed to having to manage SQL Server backups/recovery with a separate utility.

In addition to this functionality, DocAve allows administrators to restore content directly from pre-existing SQL Server backups, as well as backups on IBM's Tivoli Storage and Microsoft's Data Protection Manager (DPM). DocAve is also compatible with EMC Centera Storage, which automatically dedupes data, further minimizing the storage footprint of your SharePoint backups.

Identify and Select Components to Protect

Similar to the site collection, site, and item-level backup functionality in DocAve, the platform-level backup builder provides a data tree populated by the various components available for granular selection for each backup plan. These platform level components include:

- SharePoint farm configuration database;
- Web applications / content databases;
- Central administration content database;
- Global search setting database;
- Shared Services Provider related components, including the SSP database, Project Server databases, search database and index;
- SharePoint solutions installation files;
- InfoPath Form Services and all the form templates installed on the front-end web server and Form Services configuration;
- Single Sign-On configuration and database;

- Front-end web server components, including IIS settings, SharePoint templates under 12 hive, custom features, custom site definitions, and other file system folders;
- SharePoint Learning Kit.

With respect to the front-end Web server components, there are a number of SharePoint configurations that are not stored within the configuration database but reside on the front-end Web server file system itself. Configurations such as SSL configuration, forms-based authentication, and Web Part configuration are located under the IIS settings web.config file. Additionally, custom templates, site definitions, and features are deployed in the "12 hive" [2]. These customizations should be included in backups - so administrators need not take the time to perform the customizations again once the environment is restored. As an added capability, the data tree provided in DocAve expands further into the front-end Web server file system, so any additional dependencies external to the SharePoint configuration can likewise be protected within the same backup plan.

Determining restore granularity level

The platform-level data tree in DocAve lets administrators drill down and granularly select the farm-level components to be protected per backup plan. Administrators have the flexibility to group certain Web applications, content databases, and other individual components into their own backup plans. Furthermore, DocAve lets administrators simply select an entire SharePoint farm, and any child components it includes will be automatically protected within a single backup plan. Even new components that are created will be automatically protected.

Regardless of the granularity with which backup plans are defined, DocAve *Backup and Recovery* provides the capability to index the platform-level backups for item-level or even item version-level recovery, by allowing for selection of the appropriate restore granularity level when the plan is defined. If the restore granularity level is set to 'Site' level, the backup process will index the backup file to allow restoration at the site-level. If the restore granularity level is set to the 'Item' level, then even individual items can be restored from a full farm backup.

Full Farm Recovery Process

In the event of a catastrophic SharePoint farm failure, it may be necessary to recovery the entire farm using a platform-level backup. In order to perform such a restore, there are some notable prerequisites:

- Windows Server 2003 with Service Pack 2;
- IIS with ASP.NET enabled;
- .NET Framework 3.0;

[2] The "12 Hive" is located on the file system at `C:\Program Files\Common Files\Microsoft Shared\web server extensions\12\TEMPLATE`

- Microsoft Office SharePoint Server 2007 installed but not configured (if an existing farm already has front-end Web servers deployed, they should all be disconnected through the SharePoint Products and Technologies Configuration Wizard);
- SharePoint patch level should remain the same;
- Server name(s) and topology should remain the same;
- SQL Server disk layout should remain the same;
- The same Domain account for SharePoint administration should be used.

Once these prerequisites have been satisfied, the farm can be recovered by simply loading the appropriate platform backup through the DocAve Restore Controller. Due to dependencies between various elements of the SharePoint farm, this restoration process is usually a multi-step procedure. Before any front-end Web servers can be attached, the configuration database and the Central Administration databases must first be restored. After this has completed, the front-end Web servers can be brought online and connected to the restored configuration database using the SharePoint configuration wizard. It is important to keep in mind that one of these front-end Web servers should host the central administration Web application.

After the front-end Web servers are online, additional farm-level components can be restored. These include any customized IIS (e.g. SSL or forms-based authentication, web.config) customizations, custom solutions, Shared Services Providers, Single Sign-On, etc.

Note that any restored front-end related components (e.g. IIS settings, custom site definitions and features), should be restored on all of the front-end Web servers of the SharePoint farm.

For more information on DocAve, please visit our website at www.avepoint.com.

AvePoint
3 Second Street
Jersey City, NJ 07311
USA