



# Intelligent SharePoint Architecture and Optimizing Storage

**Nick Carr**

*Director of Sales – West Region  
AvePoint*



# AvePoint : Who we are?

## The #1 SharePoint Infrastructure Management Solution

- Founded in 2001
- Headquartered in Jersey City, NJ, with global offices in:
  - ✓ USA: Chicago, San Jose, Dallas, Washington D.C., Seattle
  - ✓ International: Canada, UK, Germany, Australia, South Africa, Japan, Singapore
- R&D team of 450+ → Largest SharePoint dev team in the world
- Winner of 2008 *Best of Tech Ed* Award for Best SharePoint Product
- Exclusive OEM relationships with IBM, NetApp, Trend Micro
- The Only SharePoint Focused ISV Partner who is a global MTC Alliance member
- A depth Managed Microsoft Gold Certified ISV Partner; MTC Alliance Member; Notes Transition Partner; Office TAP 14 Member

# Agenda

Content Organization & Storage

---

Storage Optimization

---

Content Access

---

Archiving

---

# Content Organization & Storage

# Information Architecture

Web application:

**Published Intranet Content**



- Determine the business goals
- What will your site structure and taxonomy look like?
- Standardize branding with templates and master pages

*Source: Governance Resource Centre on Microsoft TechNet*

# Planning for SharePoint Storage

- Recycle bin
  - Versioning
  - Search and index information
  - Growth
- 
- Good rule of thumb for initial planning is: 3.5 x file system

## Basic Storage Management Methods

- Set site quotas and alerts!
  - 10 GB quota, 8 GB alert is recommended
- Monitor growth trends
  - Sites: slow over time or large jump in size?
  - Overall content DB size
- Split Content DBs if they get “too big”

# How SharePoint “chooses” a Content DB for a site

- Highest remaining allotment rule, which Content DB gets the site?

Name	Status	Curr. sites	Max. sites
WSS_Content	Started	1	1
WSS_Content_2	Stopped	0	15000
WSS_Content_4435253d41864f8583ef0203fa1fd89b	Started	0	15000
WSS_Content_Asset_Management	Started	0	15000
WSS_Content_dac26d1dc9f044df93b08dd2df8a1e22	Started	1	15000
WSS_Content_ECM	Started	0	5
WSS_Content_Main	Stopped	0	1
WSS_Content_Projects	Started	2	2
WSS_Content_QA	Started	1	1001

- The Content DB with the most availability in alphabetic order
- This can be sidestepped by either “Stopping” all other DB’s (no affect to end-users) or by manually pushing using the STSADM (stsadm -o createsiteinnewdb) command

\*SharePoint Site Content BD selection process: <http://blog.jesskim.com/kb/293>

# Optimal Content DB Sizing

- Backup & Recovery operations(<50-100 GB)
- Performance (<500 GB... nervous at 300 GB)
  - # of objects
  - size of objects
  - Hardware (servers and storage)
- Storage Cost (as small as possible!)

So what is too big?

## BLOBs-- What's the Issue?

- BLOBs = Binary Large Objects
- SharePoint Content = BLOB + Metadata
- Content DB = database of ... BLOBs + Metadata
- SQL DB storage needs high IOPS (input/output operations per second) and low latency
- High IOPS + low latency storage = \$\$\$\$
- BLOBs do not participate in query operations, so no real reason to have BLOBs in a DB
- DB full of BLOBs = wasted \$\$\$

# Database Size Implications

**BLOBs increase DB size, creating issues with:**

- Backup & Recovery operations
- Performance
- Storage Costs

# Storage Optimization

# SharePoint Storage Optimization Methods

- Move the BLOBs out of the database
- Archive content

# Planning for Data Use & Growth

## What does SharePoint 2010 offer OOTB?

- No native archiving tools
- EBS extended to include RBS
  - Available only in SQL Server 2008 R2
  - Only accessible via API
- Filestream
- BCS (BDC in 2007) extended to allow for easier connectivity with legacy data systems

# Storage Optimization

*Extending BLOBs out of the database*

# Available APIs for Extending

SharePoint External BLOB  
Service (EBS)

SQL Remote BLOB  
Service (RBS)

# EBS

- Implemented by SharePoint
- Only 1 EBS Provider per SharePoint farm
- Orphaned BLOBs- no direct method to compare BLOB store and Content DB
- Compliance- what if I don't want to allow SharePoint to delete the object?

# RBS

- Implemented by SQL
- Only 1 RBS Provider per Content DB
- Orphaned BLOBs much less of an issue
- Can lock down operations, from a unified storage perspective
- Can be managed via Powershell

# RBS: SQL Server 2008 Feature Pack API

## Handled natively by database

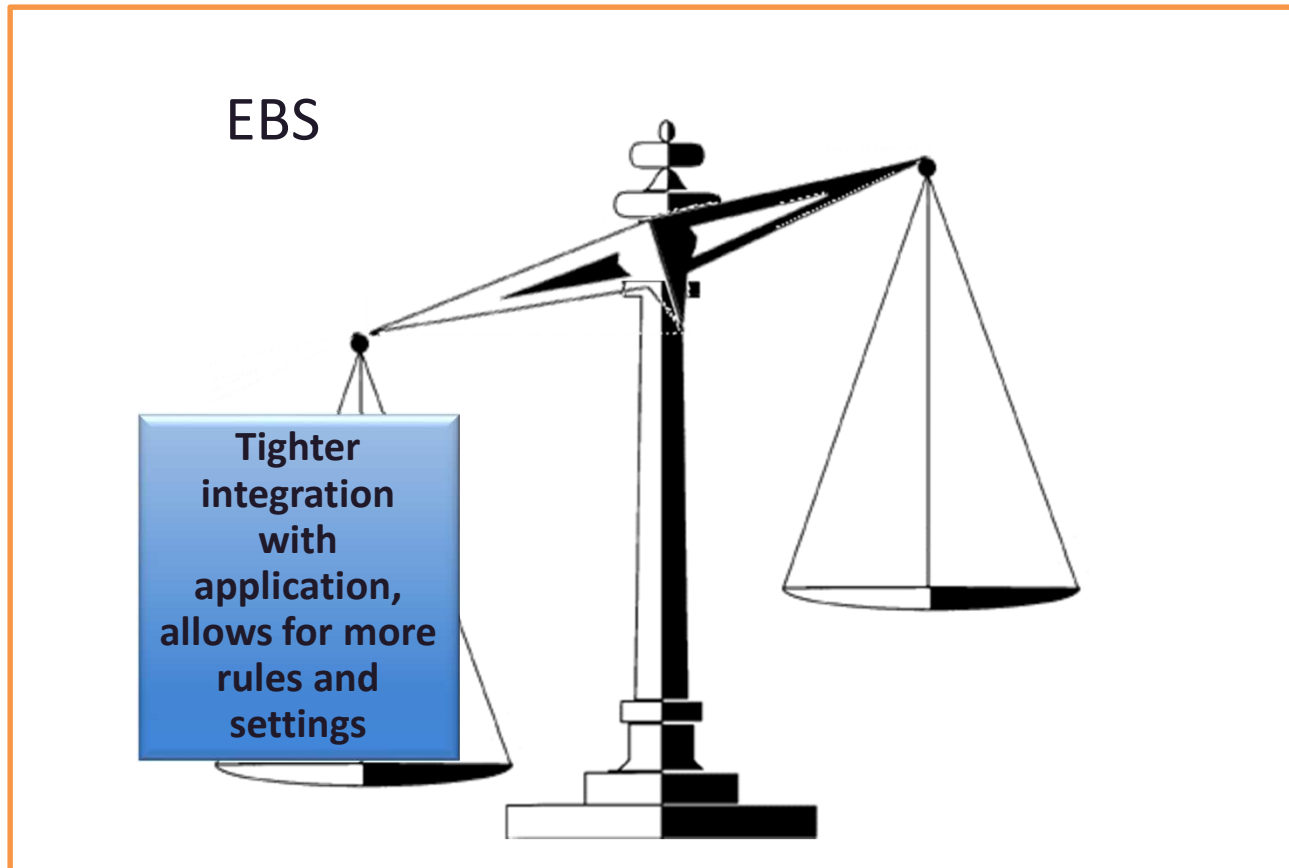
Default Provider: FILESTREAM

1. Enable FILESTREAM provider on SQL
2. Provision data store and set storage location
3. Install RBS on all SP Web and App servers
4. Enable RBS

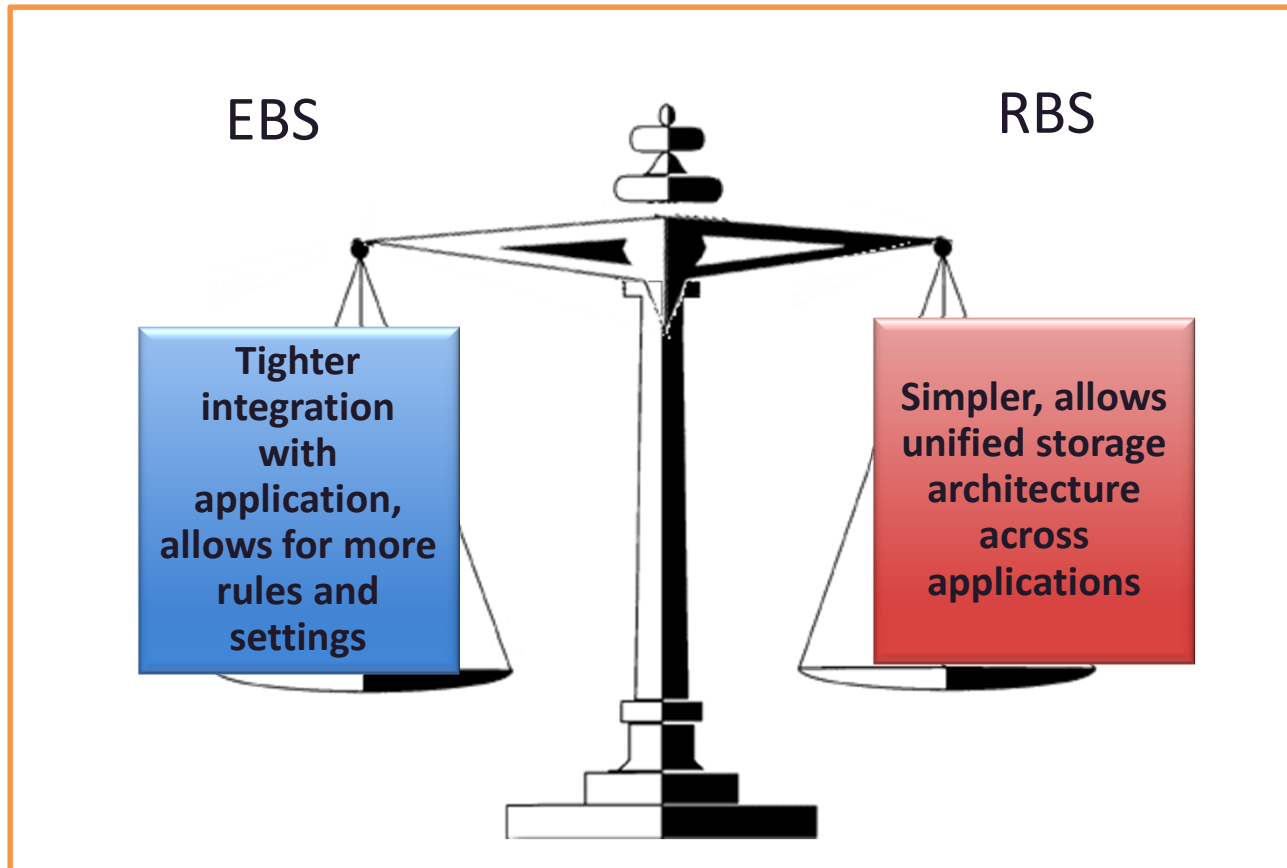
## RBS versus SQL Filestream

- Filestream storage must be file system locally attached to the SQL server
- RBS is an API set that allows storage on external stores - physically separate machines that may be running custom storage code

# EBS versus RBS, which is better?



# EBS versus RBS, which is better?



<http://www.codeplex.com/sqlrbs>

# Benefits of Extending BLOBs

- Backup & Recovery operations
  - Databases are 60-80% smaller
  - Need a method to backup BLOBs synchronously
- Performance
  - Databases are 60-80% smaller
- Storage Cost
  - “Not as expensive” storage
  - Archiving still needed for true savings
- No impact to the End User!

# To BLOB or not to BLOB...

- Proper planning must be put in place
  - How will orphaned BLOBs be managed
  - How does extending BLOBs affect your Disaster Recovery strategy
  - What teams are affected by storing SharePoint content outside of the SQL

# Content Access

# Connecting Legacy Data

- Where is it in it's lifecycle?
- Do you want to expose it in SharePoint?

## SharePoint 2010 Support

- BCS is intended for connecting LOB's (Databases, Windows Communication Foundation (WCF) or Web services, .NET connectivity assemblies, Custom data sources) into SharePoint, without migrating the data
- No OOTB solutions for getting content out of users desktops, file shares, or other ECM systems

# Options for Exposing Legacy Data

(File Shares, Notes, Exchange Public Folders, eRoom Documentum, LiveLink... etc?)

- Migrate

- Manually download/upload, losing author, time, security, history, other metadata
- 3<sup>rd</sup> Party Tool

- Connect

- BCS Mechanisms
- Most major ECM Vendors
- AvePoint's DocAve Connector

*EBS/RBS API's preferred*



# Connect to SharePoint: BCS Mechanisms

- .NET Assembly Connector
  - Provided with Microsoft Business Connectivity Services (BCS)
  - Each .NET connectivity assembly is specific to an external content type
  - Provides no Administration interface integration
- Custom Connector
  - Connect to external systems not directly supported by Business Connectivity Services
  - Agnostic of external content types that connect to a kind of external system (all databases or all Web services)
  - Provides an Administration UI integration

<http://msdn.microsoft.com/en-us/library/ee554911.aspx>

# Which BCS Mechanism Should I Use?

- The *.NET Assembly Connector* approach is recommended if the external system is static. Otherwise, for every change in the back end, you must make changes to the .NET connectivity assembly DLL. This, in turn, requires recompilation and redeployment of the assembly and the models.
- *Custom connector* approach is recommended if the back-end interfaces frequently change. By using this approach, only changes to the model are required.

<http://msdn.microsoft.com/en-us/library/ee554911.aspx>

# Migrating vs. Connecting...Which One?

## Migrating

- Data is available **in** SharePoint
- Data is moved into SharePoint
- SharePoint replaced legacy system
- Burden of storage is on SharePoint
- Changes saved in SharePoint
- Migrate and decommission

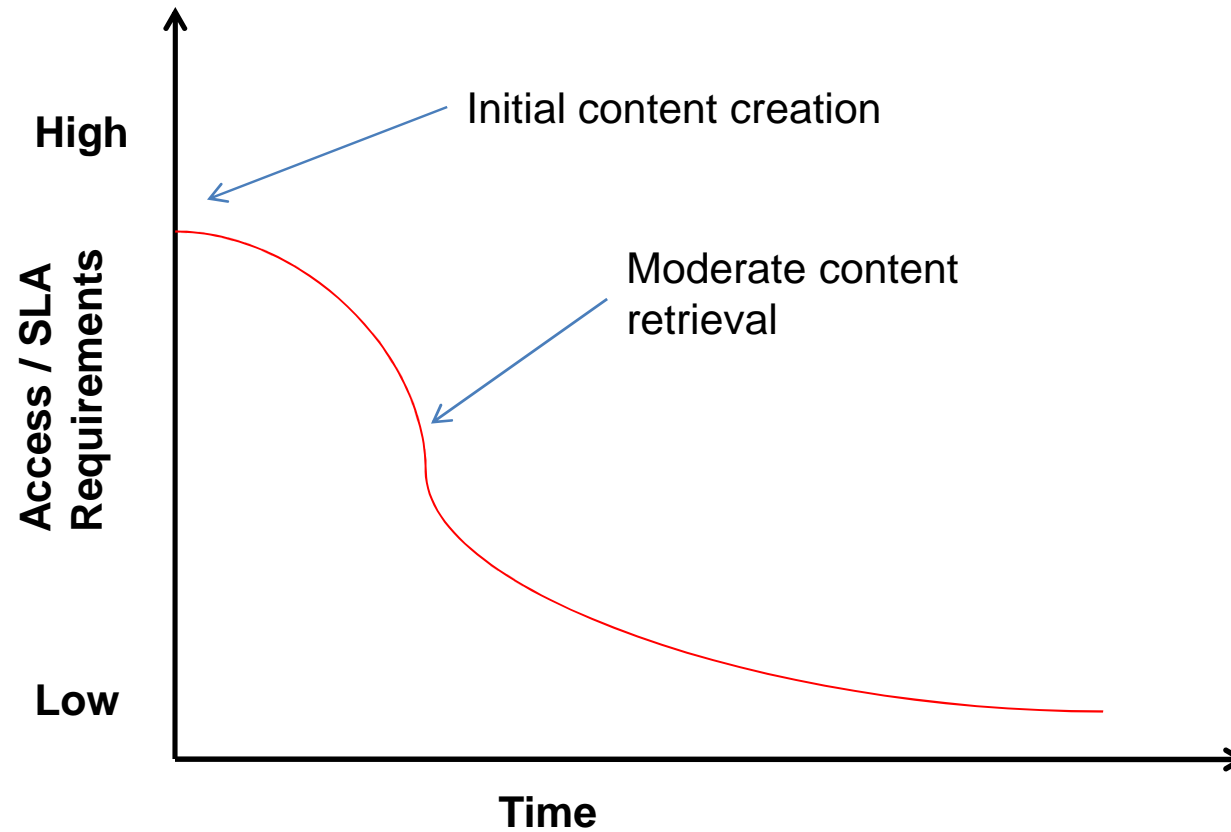
## Connecting

- Data is available **through** SharePoint
- Data is left in source (legacy) system
- Give legacy system second life by increasing its value
- Burden of storage is on legacy system
- Changes propagate to source
- Connect and forget

# Archiving

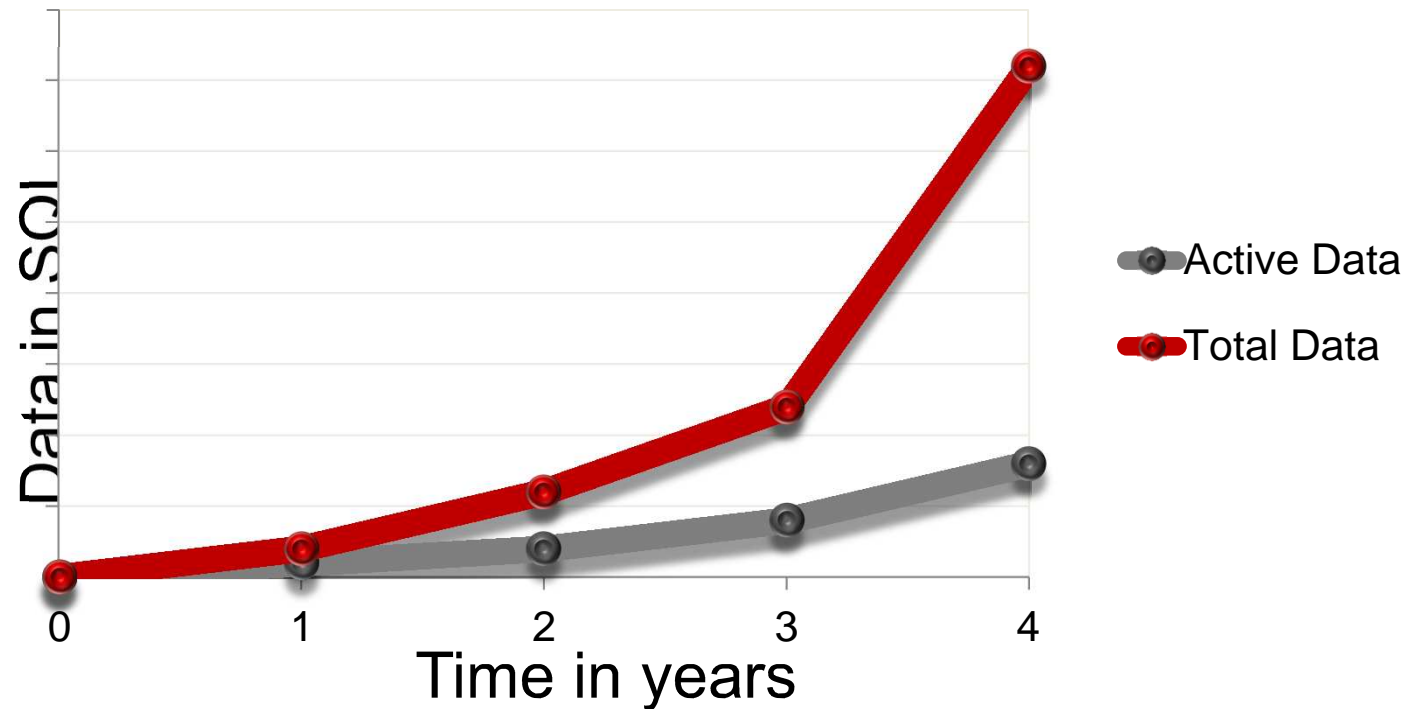
*Adding Lifecycle Management to the picture*

# Lifecycle of a Typical Item



# Storage in a Content Repository

Increase in % of inactive data over time



# Data Lifecycle Management

- Records Center
  - Another SharePoint site
  - Higher % inactive content
  - Consider separate Content DB, with an RBS provider implemented for this DB
- Archiving
  - Backup and delete
  - Workflow
  - 3<sup>rd</sup> Party tools solutions

# 3<sup>rd</sup> Party Archiving Tools

- What rules are available?
  - Last modified time
  - Author
  - Versions
- What scope can I apply rules to? (farm to item)
- Does it use RBS/EBS APIs?
- Does it integrate with other infrastructure management tools? (backup, replication, etc.)

# Summary

**1 Think carefully about organization and storage**  
*Consider where content will be stored and how it will grow over time*

**2 Leverage BLOB Services APIs to Optimize SharePoint Storage**  
*EBS/RBS API's can be leveraged to store BLOBs outside of SQL with little impact on end-users, to save \$\$ and optimize storage*

**3 Content access is key**  
*Develop strategies to handle access to legacy data and content access from remote locations*

**4 Archive content**  
*Plan for long term growth and optimal system performance*

# DocAve® Software Platform

One platform for comprehensive protection and precision control of SharePoint

## Data Protection

### Backup and Recovery

Intelligent item-through-platform level data protection

### High Availability

A one-switch disaster recovery solution for continuous platform availability

### SiteBin

Swift and complete recovery of sites and site collections following deletion events

## Administration

### SharePoint Administrator

One interface to discover and manage all configurations, securities, and settings

### Replicator

Real-time one-way, two-way, and one-to-many synchronization of SharePoint content and farms

### Content Manager

Real-time, point and click movement of SharePoint content for swift reorganizations

### Deployment Manager

Automated lifecycle management of customizations and solutions, from development through production

### SharePoint Antivirus \*

Comprehensive protection of SharePoint platforms from malicious content

## Compliance

### Compliance Vault

Retain all SharePoint content in immutable form, and search with ease for total compliance

### Auditor

Record and report upon all SharePoint events and activity

### Content Shield \*

Scan all content prior to upload to ensure proactive compliance

### eDiscovery

Search and report upon all SharePoint content for precision legal response

## Migration

### Migrator Suite

Seamless, fully-mapped transfer of content from legacy systems to SharePoint including:

- Exchange Public Folder
- File Systems
- eRoom
- Open Text Livelink
- Lotus Notes
- Vignette
- Oracle/Stellant
- Previous Versions of SharePoint

### Website Migrator \*

Migrate any HTTP/HTTPS web content to SharePoint 2007 with precision and ease

## Reporting

### Report Center

The industry's most powerful and flexible SharePoint analytics toolkit, providing single-pane access to a wide variety of mission-critical data and trending statistics about your platform infrastructure, health, and activity

### SharePoint Monitor

A free tool delivering real-time graphical topology of your entire SharePoint environment, including all your servers and their functions, relationships, roles, services, and system health

## Storage Optimization

### File Share Connector

Migration-free attachment of network and cloud file share to SharePoint. Leverage all of SharePoint's presentation and management functionality on File Share content without importing it

### Media Connector

Stream audio and video media files directly in SharePoint from any network or cloud File Share, without the need for import

### Archiver

Business-rule aware archiving for optimal platform performance

### Extender

Instantly offload BLOB content to file-based storage to optimize SharePoint's SQL database resources, improve performance, and increase scalability

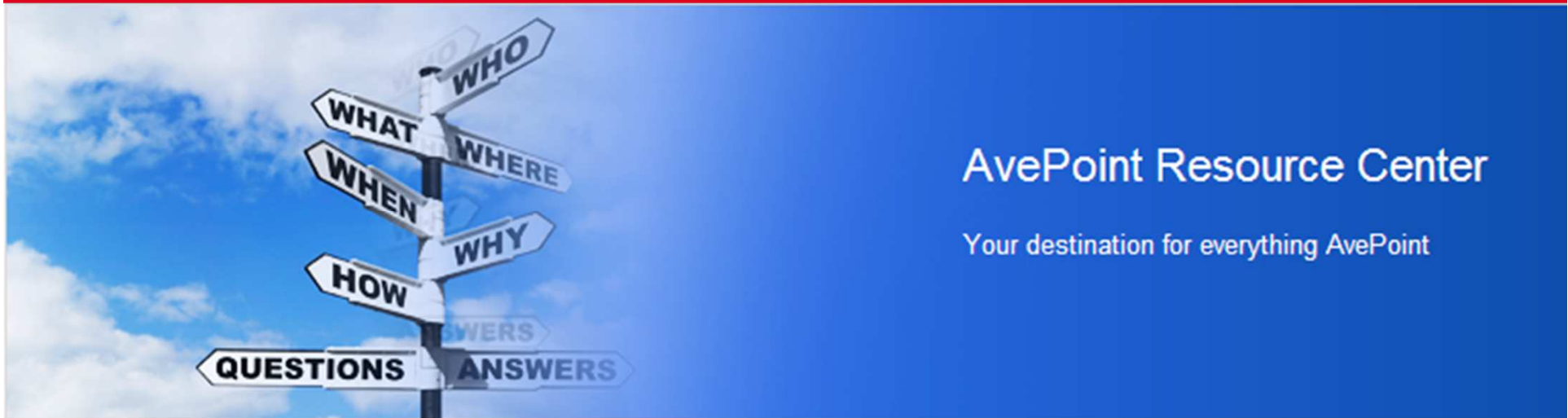
\* Operates independent of DocAve Platform

#### Additional AvePoint Solutions:

**DocAve for Salesforce:** The industry's most powerful and flexible solution, with three independently deployable modules, for the protection and integration of Salesforce CRM data.  
**Lotus Notes Migrator for Exchange:** Seamless, automated transfer of content from Lotus Notes to Microsoft® Exchange® 2003, 2007, 2010, and Exchange Online.

Thank You!

*Q&A*



Visit us: <http://www.AvePoint.com>

Email us: [sales@avepoint.com](mailto:sales@avepoint.com)

Follow us: @AvePoint\_Inc

Download a FREE, fully-enabled 30 Day trial of DocAve at [www.avepoint.com/download](http://www.avepoint.com/download)



*Resource CD includes all white papers, case studies, download links, datasheets, etc!*

# Additional Resources

- Storage Optimization for SharePoint Whitepaper :  
[http://www.avepoint.com/assets/pdf/sharepoint\\_whitepapers/Storage\\_Optimization\\_Technical\\_Advisor.pdf](http://www.avepoint.com/assets/pdf/sharepoint_whitepapers/Storage_Optimization_Technical_Advisor.pdf)
- Configure Content Database for RBS: [http://technet.microsoft.com/en-us/library/ee748641\(office.14\).aspx](http://technet.microsoft.com/en-us/library/ee748641(office.14).aspx)
- FILESTREAM RBS:  
<http://blogs.msdn.com/opal/archive/2009/12/07/sharepoint-2010-beta-with-filestream-rbs-provider.aspx>
- Whitepaper about FILESTREAM:  
<http://msdn.microsoft.com/en-us/library/cc949109.aspx>

- DocAve Extender
  - BLOB controller, uses EBS provider today for SP2007, RBS also supported in SP2010
  - Easy to deploy and configure
  - Can be applied to both new uploads and retroactively to existing data
  - Utilizes DocAve Media Services to support numerous storage devices

- DocAve Connector
  - Takes content already in external stores and makes it accessible in SharePoint
  - BCS is to external databases as Connector is to external file stores
  - Implemented via special Document Library types
    - File System Document Library
    - Media Library

# DocAve Archiver

- Archive SharePoint content (from SQL and/or extended to BLOB stores) to lower tiered storage
- Customizable business rule-based archiving engine
- Utilizes BLOB APIs when appropriate
  - Items versus sites/site collections
  - Maintain seamless access through SharePoint
  - Fully indexed and integrated with native SharePoint search
  - Full support for SharePoint workflows, alerts, Office applications, 3<sup>rd</sup> party apps, etc.

# Connecting to SharePoint: .NET Assembly

- Write code as Microsoft .NET Framework classes and compile the classes into a primary DLL and multiple dependent DLLs.
  - Publish the DLLs into the Business Data Connectivity (BDC) service database.
  - Use Microsoft SharePoint Designer to discover the .NET Connectivity Assembly and create a model.
  - Map each entity to a class in the DLL, and map each BDC operation in that entity to a method inside that "Class".
- At run time, when a user executes a BDC operation, the corresponding method in the primary DLL is executed.

<http://msdn.microsoft.com/en-us/library/ee554911.aspx>

# Connecting to SharePoint: Custom

- Implement `ISystemUtility`, `IConnectionManager`, and `ITypeReflector` interfaces.
  - Implementing `IAdministrableSystem` provides Administration UI property management support and implementing `ISystemPropertyValidator` provides import time validation of `JobSystem` properties (not on the Microsoft Office client).
  - Compile the code into a DLL and place it in the global assembly cache (GAC) on the server and clients.
  - Author the model XML for the custom data source (SharePoint Designer 2010 does not support a model authoring experience for custom connectors).
- At run time when a user executes a BDC operation, this invokes the `Execute` method in the `ISystemUtility` class. The responsibility of executing the back-end method is given to the `Execute` method.

<http://msdn.microsoft.com/en-us/library/ee554911.aspx>