Lab Validation Report

Symantec Enterprise Vault 10

Comparing Index Integrity

By Ginny Roth and Brian Babineau

December 2011

Contents

ESG Lab Reports

The goal of ESG Lab reports is to educate IT professionals about emerging technologies and products in the storage, data management and information security industries. ESG Lab reports are not meant to replace the evaluation process that should be conducted before making purchasing decisions, but rather to provide insight into these emerging technologies. Our objective is to go over some of the more valuable feature/functions of products, show how they can be used to solve real customer problems and identify any areas needing improvement. ESG Lab's expert third-party perspective is based on our own hands-on testing as well as on interviews with customers who use these products in production environments. This ESG Lab report was sponsored by Symantec.

Introduction

The "search engine" component of a purpose-built archive solution is frequently brushed over by vendors and customers alike because, although core to the benefits provided by an application, it is not as flashy as some of the other functions highlighted during the buying and selling processes. Most archive buyers view the application as a means of saving information rather than accessing it. But purpose-built archives are now relied upon by knowledge workers (retrieving old messages), attorneys (producing evidence), and compliance officers (fulfilling audit requests) for information access. Disrupting these processes is likely to have a substantial negative impact on the company in terms of productivity and risk.

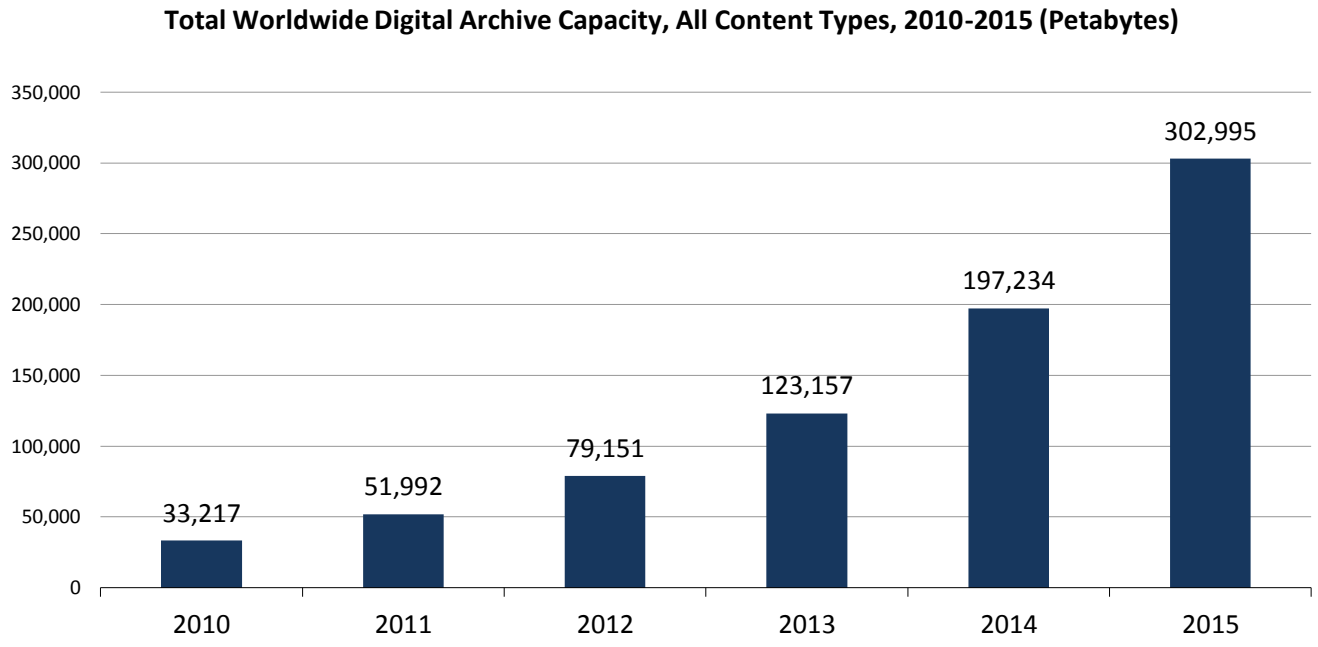
While search engines have to scale to meet more demanding access needs and bigger archives, they must fundamentally function properly in terms of producing accurate results, especially when organizations are using them within compliance and e-discovery processes. The balance between performance of a search engine and accuracy is critical for a maximum payback for any archive investment.

With the release of Enterprise Vault 10 Symantec has delivered a 64-bit archive solution that aims to meet the performance requirements of large archive systems while also maintaining accurate data retrieval results. ESG Lab tested the query results of Enterprise Vault 9 with version 10 indexes to verify the accuracy of data retrievals when upgrading from 32-bit to 64-bit indexes.

Background

The underlying search engine must scale in performance and manageability as archives get bigger, and they will get bigger. ESG estimates that organizations will archive over 44,000 petabytes of e-mail by 2015 (see Figure 1).[1] When including file and database data, the total capacity increases to nearly 303,000 petabytes. Examining just e-mail archives, the average deployment saw a 200% increase in storage capacity between 2007 and 2010.[2]

Figure 1. Total Worldwide Digital Archive Capacity, 2010-2015



Source: Enterprise Strategy Group, 2010.

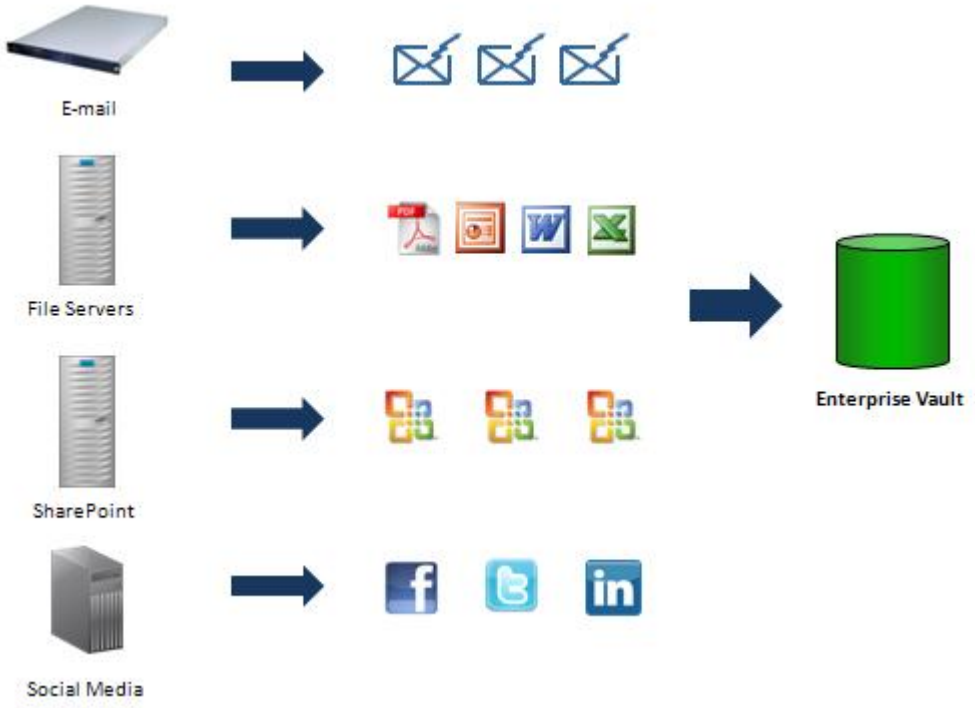
[1] Source: ESG Research Report, *Digital Archive Market Forecast 2010-2015*, July 2010.
[2] Source: ESG Research Report, *E-mail Archiving Market Trends*, May 2010.

Search engines have to keep up in terms of performance for index and retrieval while remaining easy to manage. An archive should not increase back-end operational cost or complexity—the search engine is often a prime driver of those pitfalls.

Symantec Enterprise Vault 10

Symantec's version 10 of Enterprise Vault (EV10) provides a new 64-bit index for archived items across multiple sources. Content can be archived into a single repository where the data is indexed, deduplicated, and compressed, helping to reduce storage costs and backup window time requirements.

Figure 2. Symantec Enterprise Vault 10



Enterprise Vault 10 (EV10) for Exchange offers the following key technology features:

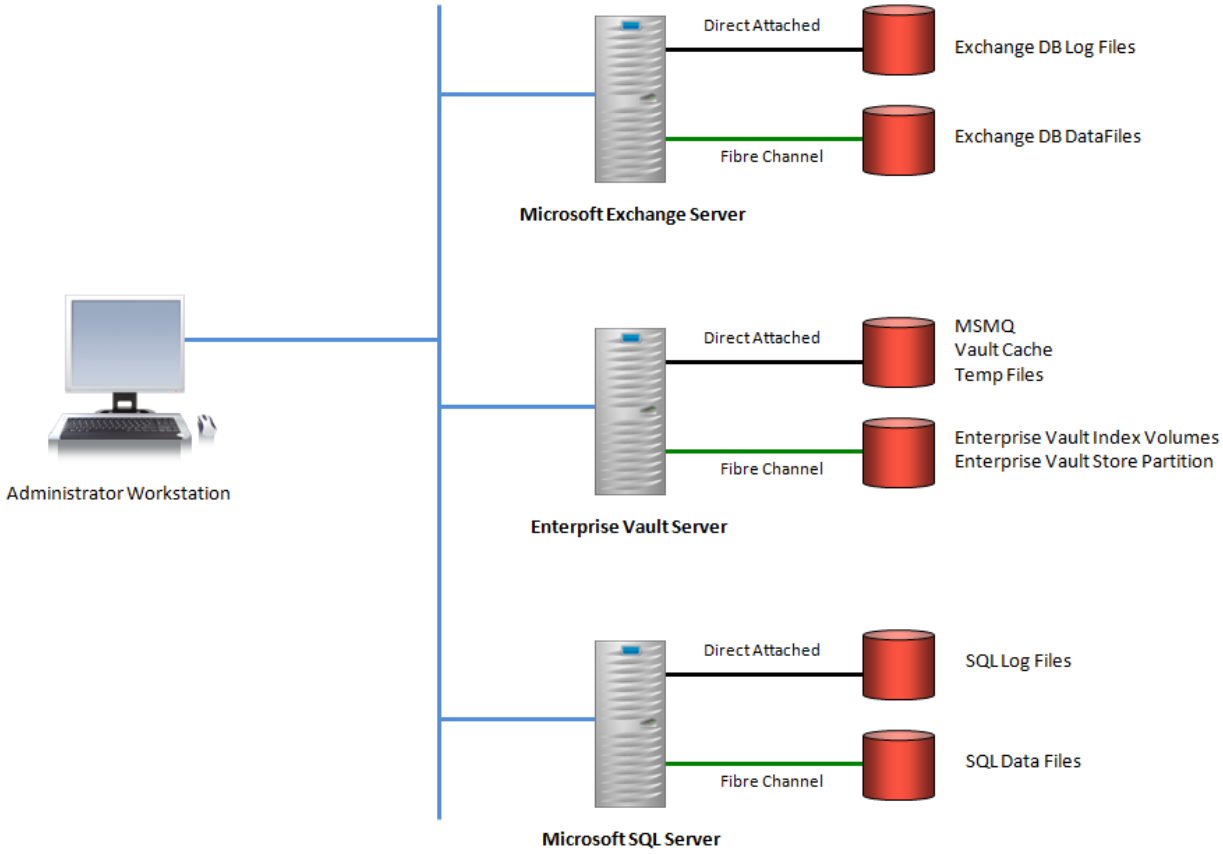
- **Optimized indexing**. Mail items archived in EV10 are indexed with a 64-bit index vs. a 32-bit index with earlier versions.
- **Integrated end-user experience**. Enterprise Vault works seamlessly with Outlook clients for Windows and Mac, OWA, and mobile devices to make searching archived items quick and painless.
- **Simplified installation and administration**. Administrators can leverage wizard-driven installation along with a management program that consolidates all archiving, indexing, and policy functions into a simple to use tool.
- **File system archiving filterpoint.** Customers can deploy filters to enable archiving and retention decisions to be made on the content of the file, and not solely on file metadata.
- **Data classification**. Organizations can meet e-mail management policies by applying retention and expiry rules across multiple classes of e-mail. Unique tags can also be applied to e-mail to help speed up discovery requests, providing a more intelligent, proactive approach to searches for relevant content.
- **Cloud storage support**. Leveraging Symantec's OpenStorage API archived e-mail can be stored in the cloud with Nirvanix's Storage Delivery Network.

ESG Lab Validation

ESG has witnessed several different search engine upgrades and replacements, and few vendor commitments work "as advertised." As such, ESG Lab put Symantec's two Enterprise Vault search deployment options to the test, performing hands-on evaluation and testing of EV10 at Symantec's facilities in Mountain View, California. Testing was designed to examine the consistency and integrity of searches in both federated and fully upgraded indexes.

ESG Lab used ten Microsoft Exchange 2007 user mailboxes containing e-mail imported from PST files from the publicly available Enron data set to test the indexing integrity between versions 9 and 10 of Enterprise Vault. As shown in Figure 3, one server was used to host the Exchange mail system. A second server hosted a Microsoft SQL server used by Enterprise Vault to maintain system configuration information. The third server contained the Enterprise Vault archive indexes of the Exchange mailboxes. ESG Lab used a Microsoft Terminal Services connection to a Windows 7 workstation to control the Enterprise Vault administrative tool and to manage the upgrades and perform searches on the indexes.

Figure 3. ESG Lab Test Bed



Federation Across EV Indexes of Different Versions

With large indexes involving thousands of users, it is often impractical to complete a full upgrade in a short maintenance and outage window. For this reason, Symantec offers a federated model that allows two or more indexes of different versions of Enterprise Vault to work together to provide a unified search. During the upgrade, the current index of data remains intact and an index with the new version is created. After the upgrade, all new data is sent to the new 64-bit index, but data is searchable across both indexes. Since there are separate threads for index updates and searches, administrators can upgrade data incrementally to the new index during normal business hours or maintenance windows, whichever they prefer.

ESG Lab Testing

ESG Lab completed numerous tests with consistent queries to compare the search results of 32-bit and 64-bit indexes. The sequence of tests is as follows:

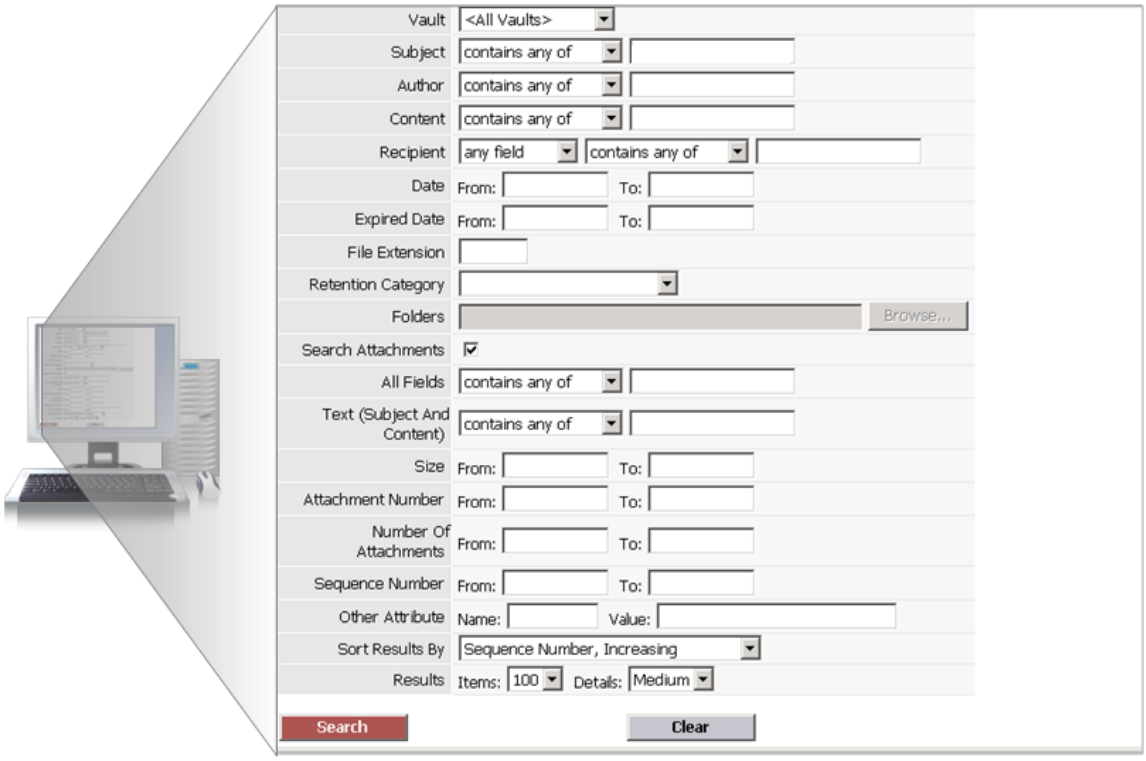
1. A test run in an Enterprise Vault 9 (EV9) system to set a baseline for search results for the queries.
2. A test run in a federated system consisting of both 32-bit and 64-bit indexes to compare the query results with the baseline test from EV9.
3. A test run in a fully upgraded EV10 index to compare the query results with the baseline test from EV9.

The expected outcome from the multiple tests is a consistent result using the same queries across all tests. If the data is consistent in all environments, the tests are considered successful.

EV9 Baseline Test

ESG Lab tested the integrity of data in the federated model to ensure that the search results produced in EV9 are consistent with results in the federated indexes. In order to set a baseline for search results, ESG Lab performed four queries in EV9 using the Enterprise Vault Advanced Web Search program through a standard web browser. Figure 4 shows the search parameters available with Enterprise Vault.

Figure 4. Enterprise Vault Search Parameters



ESG Lab used the following parameters for the four baseline searches.

Query 1: Subject contains any of **natural gas explosion**

Results will return archived e-mail items and attachments that contain either "natural," "gas," or "explosion" in the subject line.

Query 2: Subject contains any of **natural gas explosion**, File Extension .**xls**

Results will return archived e-mail items that contain either "natural," "gas," or "explosion" in the subject line and have a spreadsheet attachment.

Query 3: Content contains phrase **price fix***

Results will return archived e-mail items and attachments that contain any phrase starting with "price fix" in the content of the e-mail.

Query 4: Content contains all of **"natural gas" settlement**

Results will return archived e-mail items and attachments that contain the phrase "natural gas" and the word "settlement" in the content of the e-mail.

Table 1 represents the results with the number of e-mail items found with each search test.

Table 1. Search Results for EV9 Index

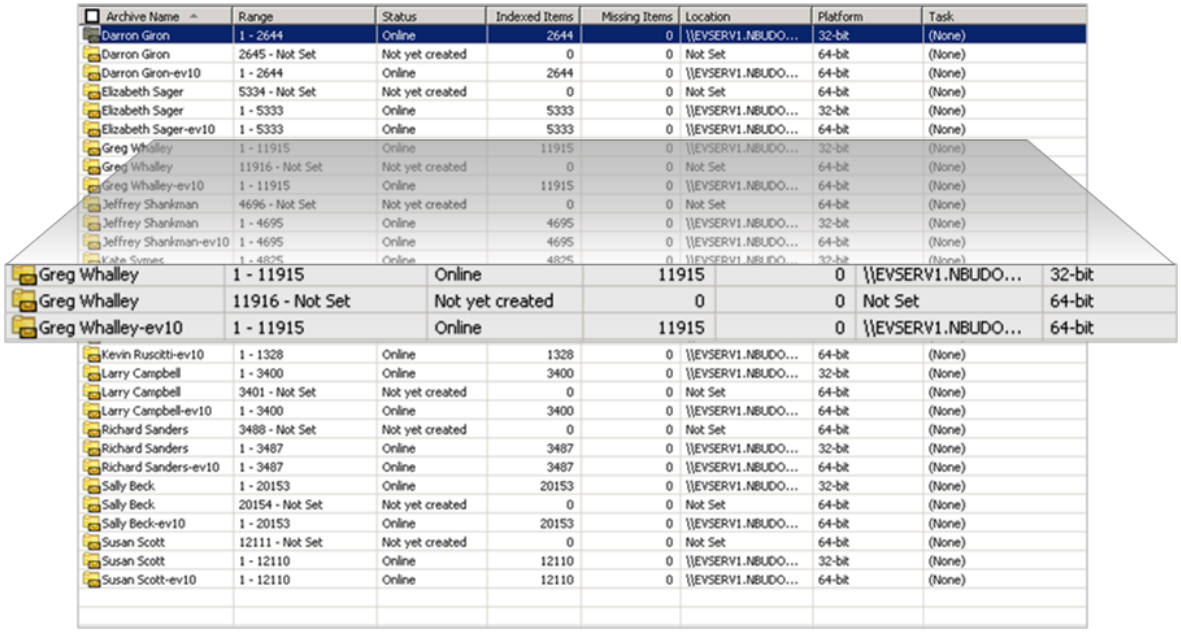
Mailbox	Query 1	Query 2	Query 3	Query 4
Chris Doland	45	4	23	8
Elizabeth Sager	18	6	22	138
Greg Whaley	21	0	6	45
Jeffrey Shankman	32	6	14	35
Kate Symes	17	0	3	43
Kevin Ruscitti	86	3	12	6
Larry Campbell	62	8	3	15
Richard Sanders	41	2	37	59
Sally Beck	190	72	10	179
Susan Scott	421	53	18	1012

Federated Test

After the initial baseline searches, ESG Lab upgraded the EV9 software to EV10 in a federated configuration. With the new EV10 system, all existing archived e-mail remained in the 32-bit EV9 index. Any new archived e-mail was indexed in the new 64-bit EV10 index.

In order to compare results, ESG Lab imported the same ten user mailboxes into Exchange with an extension "-ev10" added to the mailbox name. Figure 5 shows the indexes in a mixed EV9 and EV10 federated environment.

Figure 5. Federated Indexes



When ESG Lab updated the software to EV10, new 64-bit indexes were created for each mailbox. However, the data still resided in the 32-bit indexes. As ESG Lab imported the same ten mailboxes with the "-ev10" extension into the EV10 system, indexes were created with the same number of items as the 32-bit indexes, essentially creating identical mailboxes in both EV9 and EV10 indexes for direct comparison.

With identical mailboxes in both 32-bit and 64-bit indexes, ESG Lab expected to see identical search results when querying across a federated environment.

ESG Lab then repeated the four search queries on the federated system with the results listed in Table 2. As the results indicate, the searches in the 64-bit indexes returned the same number of items as in the 32-bit indexes in the federated environment.

Table 2. Search Results for Federated Indexes

Mailbox	Query 1	Query 2	Query 3	Query 4
Chris Doland	45	4	23	8
Chris Doland-ev10	45	4	23	8
Elizabeth Sager	18	6	22	138
Elizabeth Sager-ev10	18	6	22	138
Greg Whaley	21	0	6	45
Greg Whaley-ev10	21	0	6	45
Jeffrey Shankman	32	6	14	35
Jeffrey Shankman-ev10	32	6	14	35
Kate Symes	17	0	3	43
Kate Symes-ev10	17	0	3	43
Kevin Ruscitti	86	3	12	6
Kevin Ruscitti-ev10	86	3	12	6
Larry Campbell	62	8	3	15
Larry Campbell-ev10	62	8	3	15
Richard Sanders	41	2	37	59
Richard Sanders-ev10	41	2	37	59
Sally Beck	190	72	10	179
Sally Beck-ev10	190	72	10	179
Susan Scott	421	53	18	1012
Susan Scott-ev10	421	53	18	1012

Why This Matters

Many organizations are too large to afford the loss in business continuity required by a forklift upgrade of the entire archive system. They need a solution that upgrades incrementally, allowing upgrades to occur during normal maintenance windows or, better yet, to be performed live as users continue to access the archive system.

Symantec offers a federated system that allows administrators to upgrade individual mailboxes with an incremental approach to produce search results still consistent with the data available in EV9.

ESG Lab tested the federation capabilities of EV10 that makes two searchable indexes available for consolidated searches and found the results consistent with data returned in an EV9-only index.

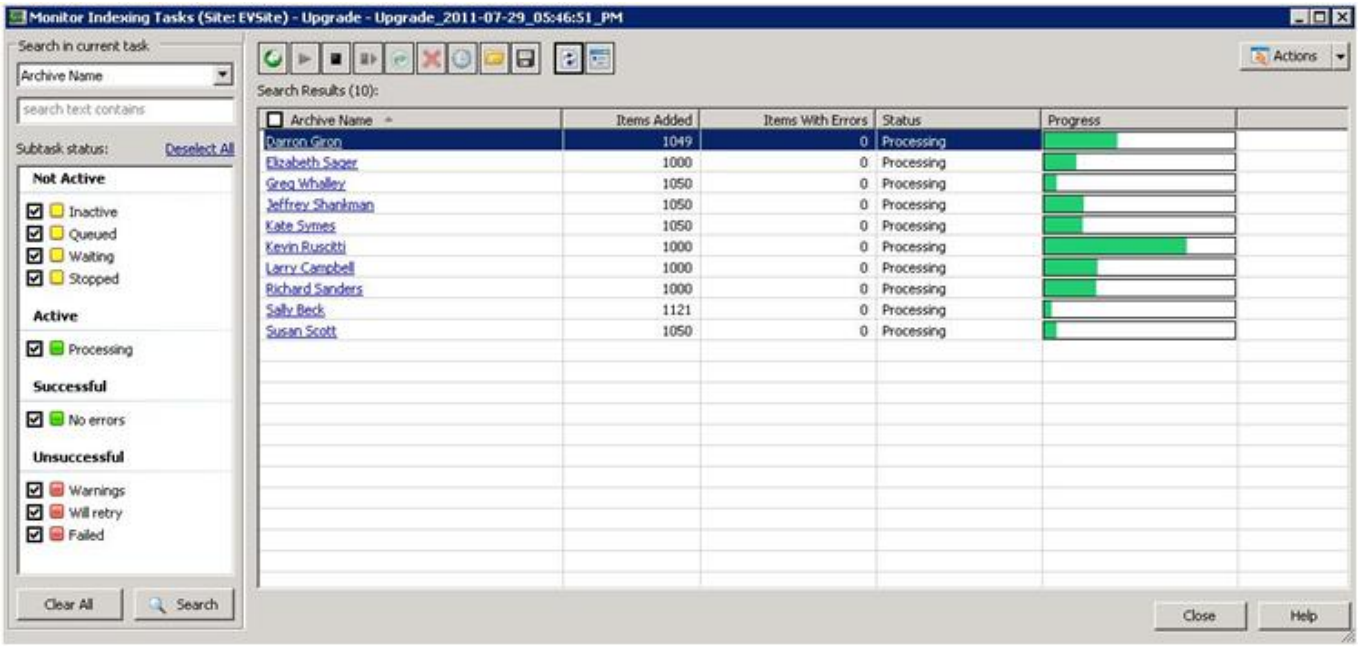
Full Upgrade to EV10 Index

Once Enterprise Vault software is upgraded to version 10, a new 64-bit index is created for each mailbox archive. The old EV9 index still exists for the existing archive until an upgrade on that index is performed. Administrators have the option to upgrade the entire index at once, usually in a single maintenance window, or incrementally. For large organizations, the incremental approach is required due to the sheer volume of mail items that cannot be entirely upgraded during a maintenance window. When upgrading incrementally administrators can choose any number of user mailboxes for each upgrade task scheduled. The index process can be stopped and restarted at any time and the process will resume where it left off.

ESG Lab Testing

ESG Lab tested the consistency of data searches after a full upgrade of EV9 indexes to EV10. To begin, ESG Lab selected the ten mailboxes in the 32-bit index and created an indexing task to upgrade the mailbox archives, and started the task immediately. Figure 6 shows the progress of the upgrade of all ten archives to EV10.

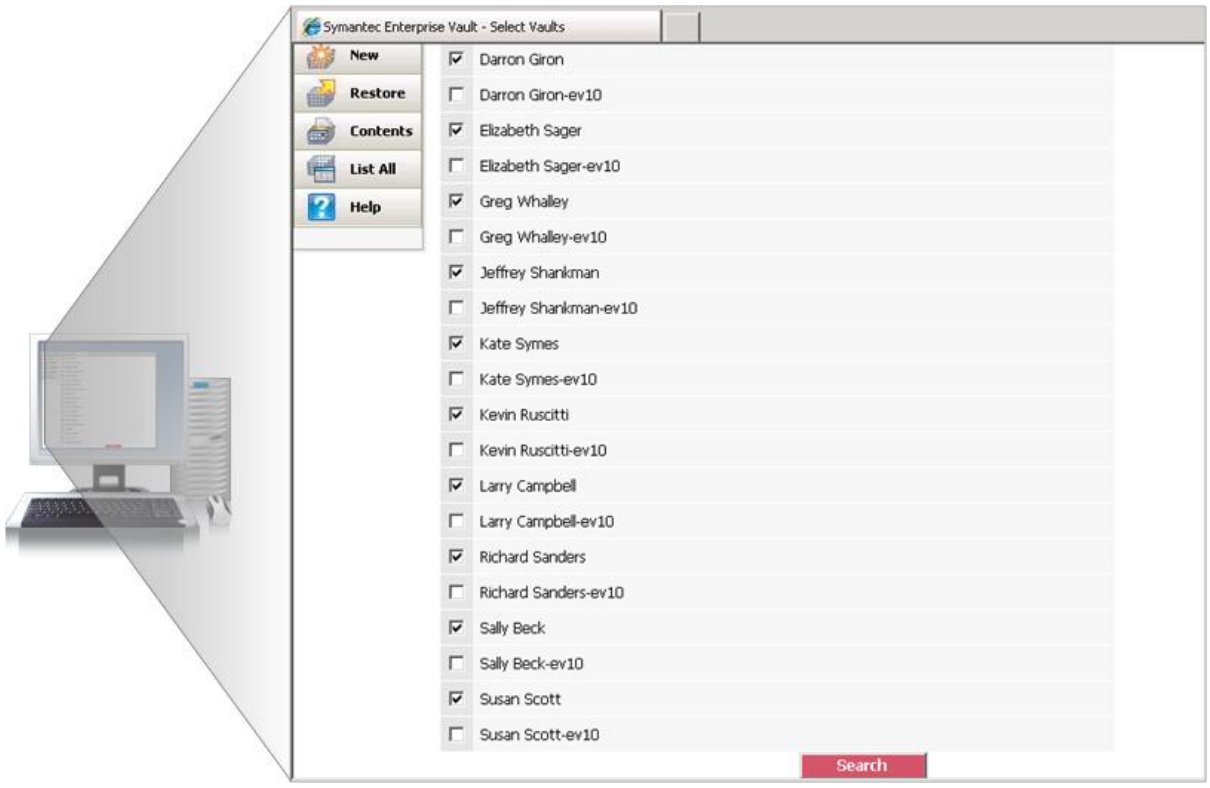
Figure 6. Upgrade Progress



While the upgrade was in progress, ESG Lab tested the system's search capability to ensure that performance would not degrade and to verify that the data returned was still consistent with earlier results in the federated system.

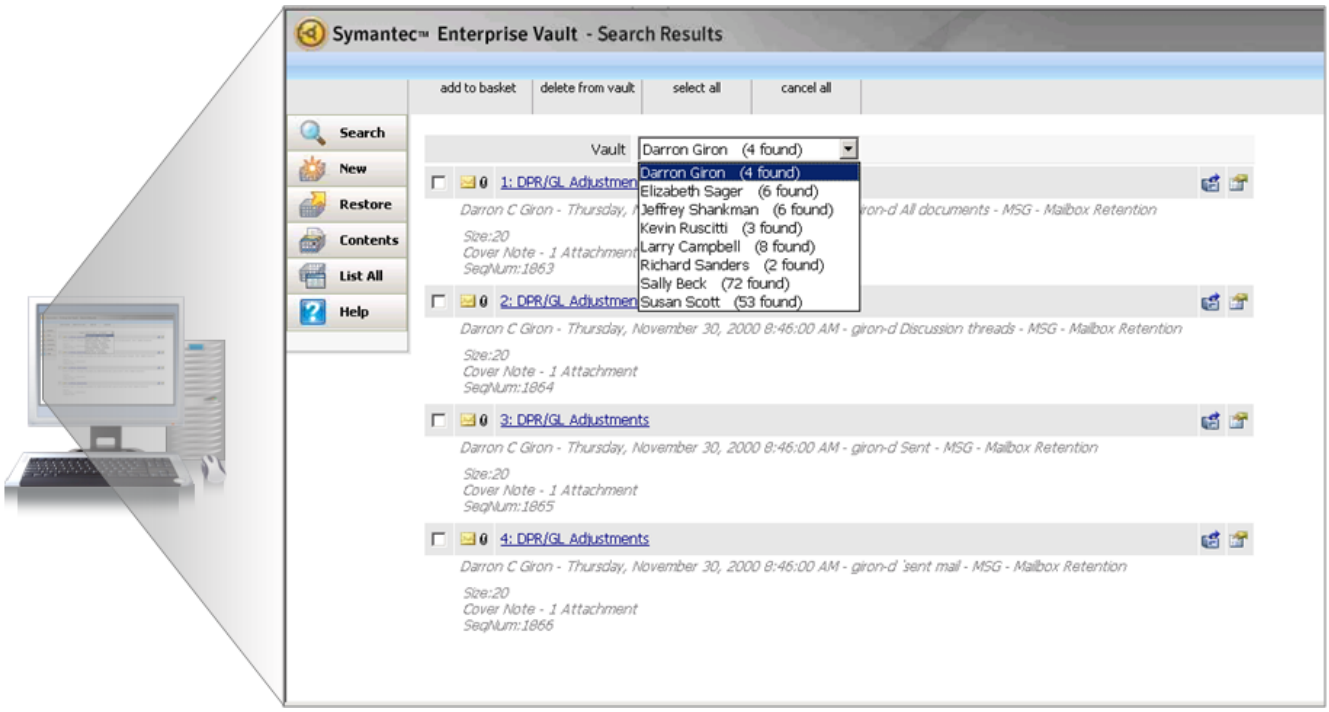
Since the 32-bit index was still searchable during the upgrade, ESG Lab selected the EV9 mailboxes and performed the second search test. Figure 7 shows the mailboxes selected for the search.

Figure 7. Index Search During Upgrade



When compared to the second test in Table 1, the results shown in Figure 8 match the results from earlier tests.

Figure 8. Test 2 Search Results



ESG Lab also performed additional random searches on the index while the upgrade was in progress and found that search performance was not negatively impacted from a user perspective.

After the upgrade completed, ESG Lab repeated the four search tests on the new 64-bit indexes. As the results in Table 3 indicate, the items returned in the four tests were identical to the initial baseline tests run in the EV9 index.

Table 3. Search Results for EV10 Index

Mailbox	Query 1	Query 2	Query 3	Query 4
Chris Doland	45	4	23	8
Elizabeth Sager	18	6	22	138
Greg Whaley	21	0	6	45
Jeffrey Shankman	32	6	14	35
Kate Symes	17	0	3	43
Kevin Ruscitti	86	3	12	6
Larry Campbell	62	8	3	15
Richard Sanders	41	2	37	59
Sally Beck	190	72	10	179
Susan Scott	421	53	18	1012

Why This Matters

As organizations look to upgrade their large archives, they often have two major concerns: how the upgrade will impact the performance of their current environment and whether the archive will maintain its integrity through the upgrade to the new index.

Symantec's EV10 allows administrators to upgrade data incrementally during a maintenance window or while users continue to access data. Providing these options gives administrators the flexibility to manage an upgrade with minimal impact on the user community.

ESG Lab executed multiple searches during a live upgrade of an EV9 index and found no visible performance impact in the returned results. In addition, ESG Lab compared the search results in the EV10 index to the baseline searches created in EV9 and found no variance in the search results, reinforcing that the integrity of the data was not compromised.

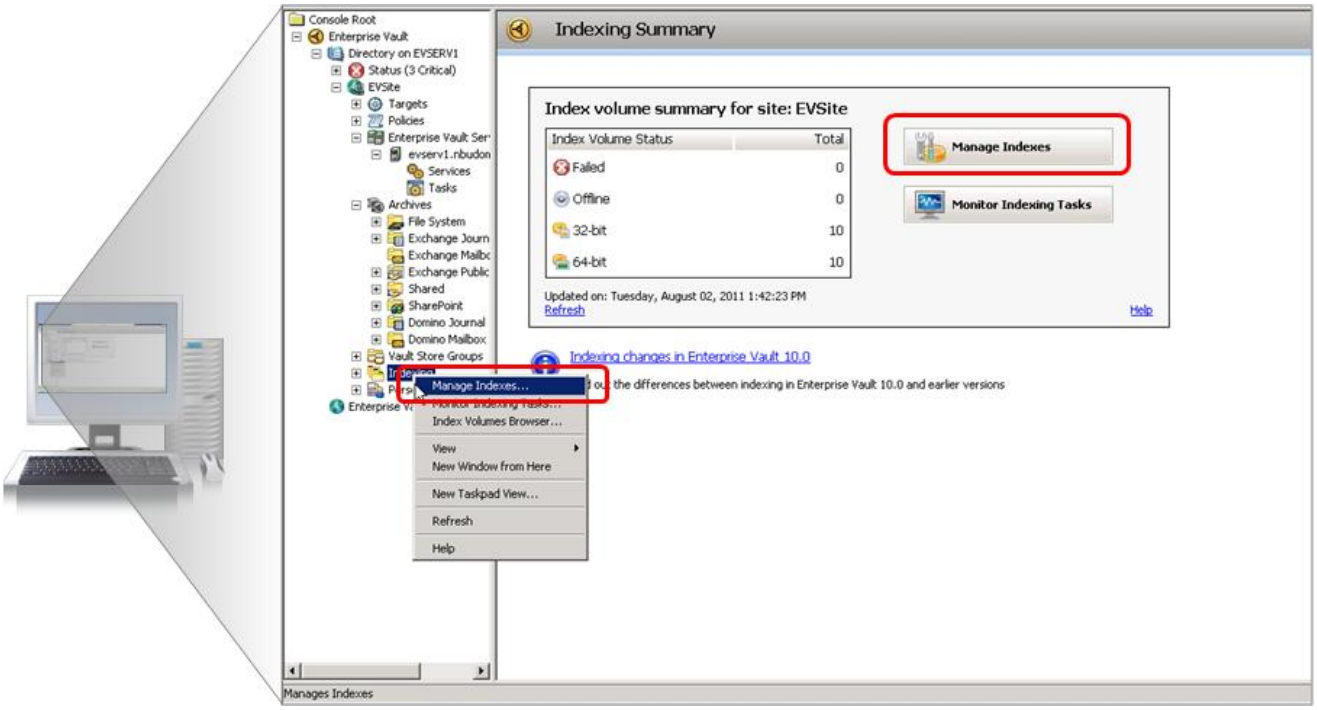
Index Manageability

Enterprise Vault boasts many improvements in its latest version, including data classification services, archive to the cloud, support for Mac Outlook clients, and enhancements in management functions, particularly indexing tasks. ESG Lab looked specifically at the management functionality and its improvements over EV9. Indexing is now fully managed by the Administration Console, which includes a new Manage Indexes wizard and Monitor Indexing Tasks page.

ESG Lab Testing

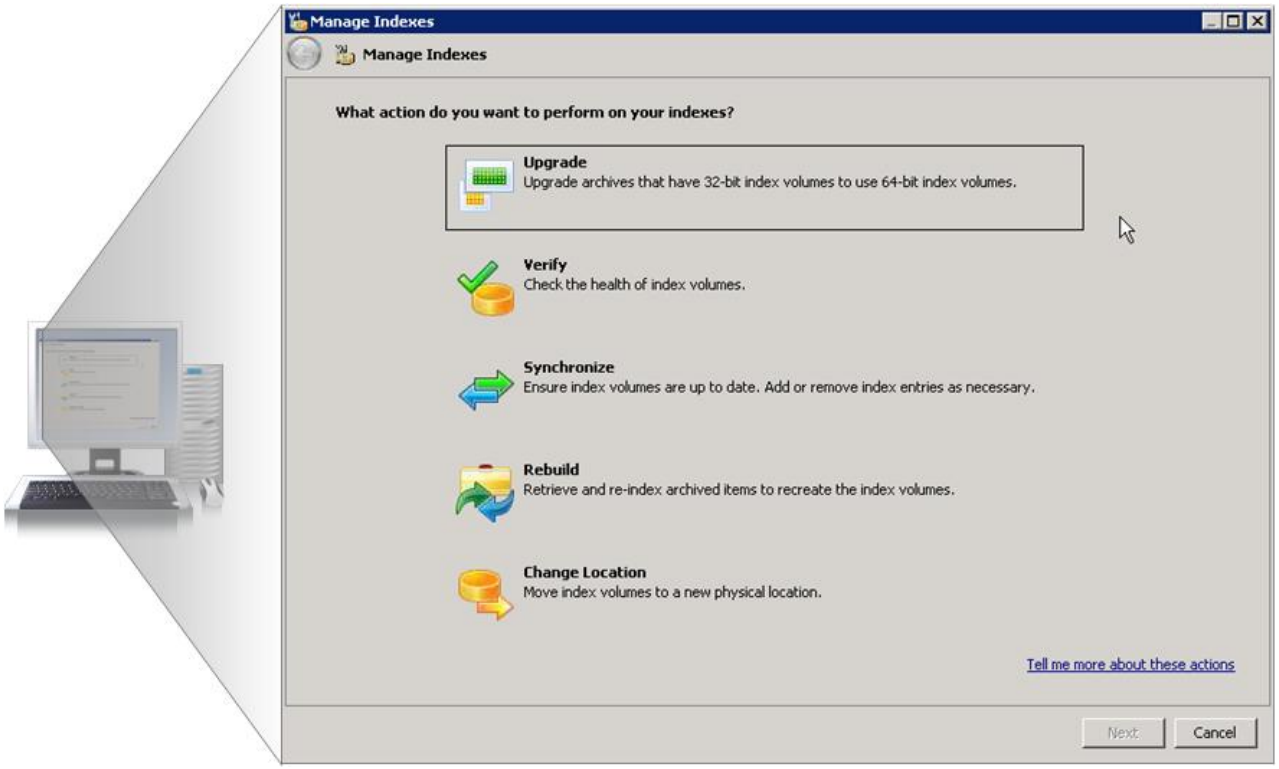
ESG Lab examined the new management functions for the index. The most significant difference ESG Lab observed was the ability to manage indexes through one utility. In the Enterprise Vault administration tool, ESG Lab navigated to the indexing section, which showed a summary for the available index. ESG Lab was able to access the utility by either right-clicking the indexing item or selecting the "Manage Index" tab on the indexing summary page, as shown in Figure 9.

Figure 9. Indexing Summary



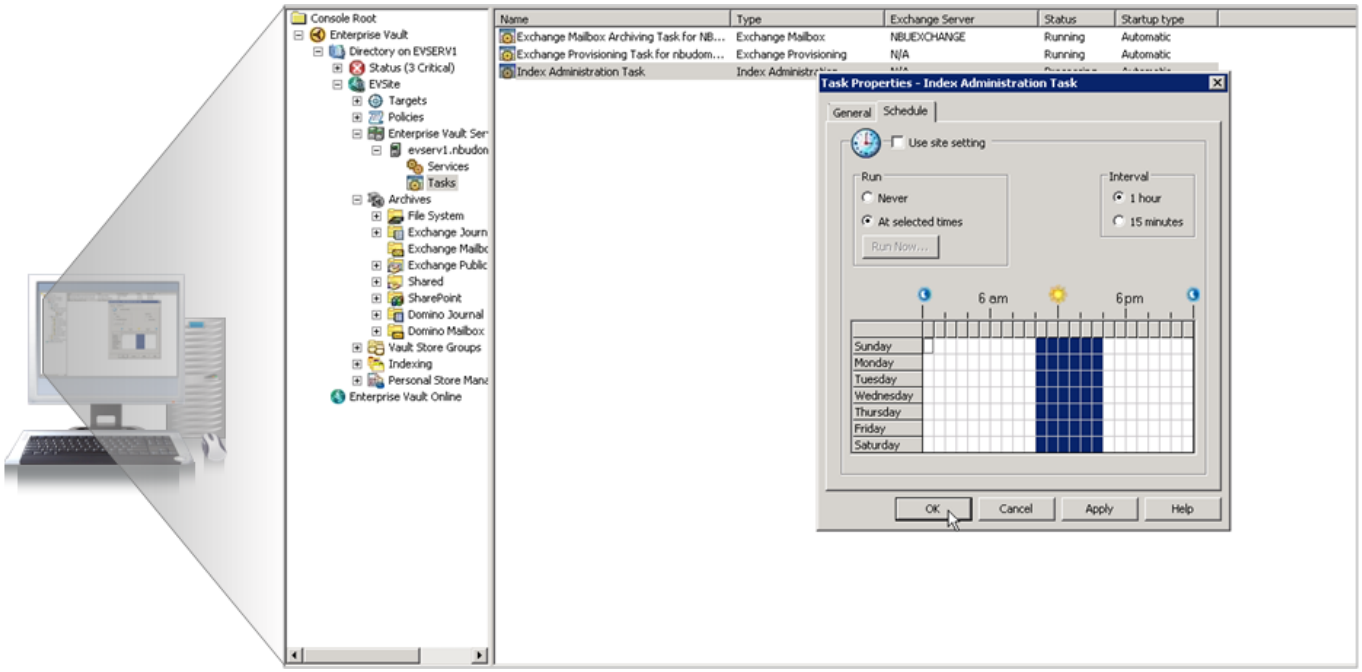
This utility enables users to perform index management tasks, allowing them to rebuild, synchronize, or verify the health of an index. ESG Lab chose the Upgrade option shown in Figure 10.

Figure 10. Manage Indexes



With EV10, indexing tasks can be scheduled, much like mailbox archiving tasks, allowing administrators more flexibility in setting up indexing functions to be performed during maintenance windows. ESG Lab looked at this capability when performing an upgrade on the ten user indexes from EV9 to EV10. As Figure 11 shows, ESG Lab was able to create a schedule in the properties page of the "Index Administration Task," choosing dates and times that the task would be run.

Figure 11. Schedule Indexing Tasks



Why This Matters

The integrity of the data in an archive system is paramount. But it can't overshadow the tools to manage and maintain the environment. IT organizations benefit when management tools are easy to use, allowing administrators to focus on creating effective e-mail retention and e-discovery policies that support the business.

EV10 provides a simple management tool that combines all the functions (indexing, policy, archive tasks) of Symantec's archive system into one consolidation program.

ESG Lab found the upgrade process to EV10 to be simple and straightforward with a deployment scanner that proactively discovered potential issues with an upgrade before any problems arose. Additionally, ESG Lab determined that the consolidated indexing functions were an improvement over previous versions.

ESG Lab Validation Highlights

- ☑ Running in a federated environment, Enterprise Vault 10 was able to return consistent results with searches performed in an EV9 system.
- ☑ After upgrade of EV9 indexes to EV10, search results were consistent with those obtained in EV9 and federated environments.
- ☑ Upgrading to EV10 was simple and intuitive. The deployment scanner provided a proactive approach to analyzing any prerequisites that would need addressing before an upgrade.
- ☑ Indexing functions were consolidated into one utility, making management of the indexes straightforward.

Issues to Consider

- ☑ Enterprise Vault introduced the "ItemGranularity" feature in its index schemas in EV 6.0 based on numerous requests from customer feedback. The ItemGranularity schema combines the information from attachments into the top level parent document in the index, which is the actual mail item. This results in greater search efficiency. However, the net result is that attachments are not shown separately in the search results. Searches with the Enterprise Vault browser or integrated search features will return top level document results only and not individual attachments as separate search results. Therefore, a search which hits on multiple attachments for the same message would return different results when searching against a 32-bit EV9 index and a 64-bit EV10 index, since EV9 does not have ItemGranularity enabled by default for mailbox archiving. For example, if searching the 32-bit index in one query returns 12 hits that include eight attachments from four distinct mail items, the same query in EV10 64-bit index would return four hits because Enterprise Vault only returns the four distinct parent messages. In order to test for consistent results across the two versions of Enterprise Vault, the ItemGranularity option must first be turned on in EV9 and the indexes must be rebuilt so attachments are not separate items in the search results.
- ☑ EV10 has the ability to search on items using wildcards combined with just one matching letter. EV9 required at least three letters with a wildcard to return matching items. As a result, using the search parameter "price f*" in EV10 could return any item that started with that phrase. However, EV9 would return no results since it isn't supported as a valid search option. This is a functional difference between 32-bit and 64-bit indexes. So, when operating in a federated environment where there is a mix of 32-bit and 64-bit indexes, a query using "price f*" would return results starting with that phrase in the 64-bit index but no results in the 32-bit index.
- ☑ Customers may notice that the default sort order of query results has been changed to Date Descending rather than relevance which may affect the look of query results when comparing EV9 and EV10 archives.

The Bigger Truth

Current archive market trends indicate that the next wave of innovation will center on archive access. The current size and projected growth of archives coupled with organizations' desire to search and utilize the retained information, amongst other drivers, are spurring the enhancements and evolution of many products including Symantec's Enterprise Vault 10. With this version, available to the market as of July 2011, Symantec upgraded its search engine to establish a foundation for faster queries, contextual presentation of query results, and smaller index footprint as archives continue to get larger.

Introducing a new search engine into an archive creates plenty of opportunities, especially as it pertains to making information more accessible and usable to the business. However, there are also potential risks as upgrading data to a new search engine requires an upgrade and re-index process or the need for federation where a new search engine can query previously indexed data. Symantec has committed to both options: a non-disruptive upgrade from a previous EV version to 10.0 or a federated deployment where existing indexes created by an older version of EV are searchable via a 10.0 implementation.

ESG Lab tested indexing in both a federated configuration and a fully EV10 index, and found the results returned from multiple queries consistent with the previous EV9 results.

Symantec has shown that it understands the upgrade concerns of IT organizations and has provided a strong solution that not only preserves the integrity of archived data, but eases the burden for organizations with large archives by offering an incremental upgrade path to EV10.

Appendix

Table 4. ESG Lab Test Bed

Symantec Enterprise Vault 10	
Enterprise Vault Server	
Hardware	HP Proliant DL360 G5 Dual Quad Core Intel Xeon E5405 2.00GHz 16 GB RAM
Software	Microsoft Windows 2008 x64 Enterprise R2 SP1 Enterprise Vault 10.0
Microsoft Exchange Server	
Hardware	Dell PowerEdge 19050 Dual Quad Core Intel Xeon E535 2.00GHz 4 GB RAM
Software	Microsoft Windows 2008 x64 SP1 Exchange Server 2003 SP3
Microsoft SQL Server	
Hardware	HP Proliant DL360 G5 Dual Quad Core Intel Xeon E5405 2.00GHz 4 GB RAM
Software	Microsoft Windows 2008 x64 Enterprise R2 SP1 SQL Server 2008 x64 R2

Enterprise Strategy Group | **Getting to the bigger truth.**

20 Asylum Street | Milford, MA 01757 | Tel:508.482.0188 Fax: 508.482.0218 | www.enterprisestrategygroup.com