Microsoft Exchange Migrations and the HP StorageWorks Reference Information Storage System — white paper

Introduction

The HP StorageWorks Reference Information Storage System (RISS) is an active archiving solution for the long-term storage of reference data including email messages and their attachments.

This white paper describes how the RISS solution can be deployed as part of a Microsoft® Exchange Server 2003 migration to simplify the final infrastructure design, lighten the workload on Exchange servers, and reduce total cost of ownership for enterprise email. HP recommends that a RISS active archiving solution for email is included in the final infrastructure design for an upgraded Microsoft Exchange 2003 system. The establishment of a permanent, centrally managed repository for archived email delivers short- and long-term benefits such as increased end-user productivity, faster access to archived information, and enhanced data security leading to reduced business risk.

This white paper explains why Microsoft users are migrating from Exchange 5.5 (or earlier) to later versions of Microsoft Exchange, discusses some of the challenges involved with the migration of enterprise email, and describes how RISS can help the solution design and provide a cost-effective and risk-reduced platform for long-term storage of data with improved access to valuable information contained in the email environment.

The importance of email

Email provides fast and standardized communications for the global community of IT users. Companies rely on email for communications with their customers, their suppliers, and their employees. And today, many important business transactions are executed **entirely** in the email domain. As a result of these trends, email is now recognized by many organizations as a business-critical tool.

This dependency on email-based communications is reflected in the growth of email traffic. By the end of 2004, the worldwide number of corporate mailboxes will reach 482 million, according to The Radicati Group, a firm of independent analysts that tracks the email market. Each mailbox user sends and receives an average of 84 email messages per day with a total capacity per user of around 10 MB per day. By 2007, it is estimated that the number of corporate email users will grow to 773 million with storage capacity per user reaching 14.3 MB per day, resulting in a staggering 10.3 PB of global email traffic each day.

And it is not just about quantity. Because a lot of email traffic represents important company information regarding sales transactions, contracts with suppliers, legally binding communications, and important competitor intelligence, long-term access to this valuable company "information asset" is yet another factor to consider when planning a major upgrade of the corporate email infrastructure.

About Microsoft Exchange

Microsoft Exchange Server is one of the most widely deployed solutions for messaging and collaboration in the corporate world today. It is estimated that worldwide, over 100 million users depend upon Microsoft Exchange Server for email communications.

In October 2003, Microsoft began shipping the latest release of Exchange, named Exchange Server 2003. This is the successor product to Exchange 5.5, which first appeared in February 1998, and Exchange 2000, which began shipments in November 2000. But until now many user organizations have been reluctant to move from Exchange 5.5 because of questions regarding the cost and risks of migration versus the benefits of adopting the later Microsoft Exchange releases. As recently as August 2003 it was estimated that up to 60% of Exchange systems were still running Exchange 5.5.

Microsoft's announcement that Exchange 5.5 will reach the end of its support life on December 31, 2005 has forced a decision on the migration issue for many customers. The question is no longer

whether to migrate, but when. But because the migration of an enterprise messaging infrastructure, together with all of the Exchange clients, is a non-trivial exercise, many customers are looking to trusted service providers, such as HP, to help them with the difficult task of migrating their email platform.

HP Services for Microsoft Exchange

HP has unparalleled experience in providing Microsoft Exchange—related services. Microsoft and HP have been business partners for over 20 years with a relationship that has its roots in a deep technological foundation: Microsoft runs their business on HP servers and HP runs its business on Microsoft software.

HP demonstrates leadership in applying Microsoft technology, with:

- 23,000+ Microsoft-trained professionals
- 5,000 Microsoft-certified engineers (including MCSEs, MCSDs, MCSAs, and MCDBAs)
- 3,200 Windows® 2000–trained consultants and support engineers
- 2,600 Microsoft-dedicated practice personnel
- 1,400 Windows and messaging specialists
- 1,200 dedicated .Net solutions practice professionals
- 850 dedicated Exchange consultants
- 1,100 SQL-certified engineers
- 2,100 PMI-trained and 400 PMI-certified project managers averaging 10 to 15 years experience
- Over 1,500 managed projects annually
- HP has over 12 million Microsoft Exchange Server 2000/2003 seats under contract through HP Services; plus over 6 million seats through HP authorized business partners
- HP ProLiant servers are the platform of choice for Exchange with more market share than our three largest competitors
- Microsoft Information Technology Group has deployed Exchange 2000/2003, servicing over 60,000 mailboxes on ProLiant servers and HP StorageWorks storage within Microsoft's infrastructure
- HP is the largest integrator of Microsoft technologies worldwide, even larger than Microsoft Consulting Services

Microsoft's Exchange business unit manager acknowledges, "HP dedicates more people to working on Exchange than Microsoft has developing the product. When it comes to deployment--what works and what doesn't for real customers in the real world--they are the clear experts."

HP and Microsoft—Frontline Partnership

The Frontline Partnership is a long-standing relationship, dedicated to developing and delivering industry-standard information technology in both core businesses and new strategic markets while offering unmatched ownership satisfaction and value.

Through this partnership, HP and Microsoft combine their respective strengths as industry leaders to bring complete, integrated technology solutions that are easy to set up, stay up all the time, and include world-class service on demand.

For over two decades, HP and Microsoft executives, engineers, and sales and service teams have worked side by side to research, develop, test, implement, and support solutions that meet the most demanding business objectives and help drive new business development efforts.

From the beginning, HP has shared technology and engineering resources to develop and advance the Windows 95/98, Windows NT®, and Windows 2000 operating systems; Exchange 2000; Datacenter Server; and most recently, the complete suite of .NET Enterprise Servers.

HP and Microsoft implement the solutions they offer to you within their own companies. HP ProLiant servers comprise 95% of Microsoft's 6,000 servers and were used as the development platform for Windows 2000 Server, Commerce Server 2000, Exchange 2000 Server, SharePoint Portal Server 2000, SQL Server 2000, and .NET. Microsoft and HP, respectively, represent the first and second largest Windows 2000 implementation in the world.

Because we can deliver all of the service and support necessary to implement Frontline Partnership solutions, Microsoft endorses HP Services as the only Worldwide Prime Integrator for Windows 2000, Windows NT, Exchange 2000 Server, and BizTalk Server 2000.

Migrating Microsoft Exchange

The benefits of migration

As a recognized authority on the Microsoft Exchange platform, HP recommends migration to HP Exchange Server 2003 for the following reasons:

1. Customers will see an immediate increase in end-user productivity. Improvements in Microsoft Office Outlook and Outlook Web Access offer:
    - Faster email access and intuitive user interface
    - Improved Outlook and Outlook Web Access performance that enables higher productivity for mobile workers connecting over slow, latent, inconsistent network connections
    - Secure Internet access from Outlook that reduces the need for a virtual private network (VPN) connection when used with Microsoft Windows Server 2003 and that enables single-URL access for Outlook, Outlook Web Access, and mobile devices
2. IT organizations will see increased productivity for IT staff. Administrators can save time and increase productivity with:
    - Single-seat administration for Exchange and infrastructure through the Active Directory service
    - Single mailbox restore
3. Faster, easier migrations and deployment. Exchange 2003 includes handy wizards that make upgrading easier with:
    - Checks to ensure that your earlier environments have been correctly prepared before native Exchange Server 5.5 upgrades
    - Active Directory Connector wizard that automatically creates connection agreements between Exchange 5.5 and Active Directory
    - Public folder migration and new administration tools
4. Reduced cost. Exchange 2003 offers better server consolidation with more mailboxes per server for Exchange 5.5 users as well as:
    - Increased performance between Outlook clients and Exchange 2003 that increases productivity and enables Exchange mailbox server consolidation scenarios
    - Support for four-node and eight-node clusters when used with Windows Server 2003, Enterprise Edition

5. Enhanced security. Exchange 2003 and Windows Server 2003 combined provide better security:
    - Improved junk mail controls that provide support for real-time blacklists and help to minimize spam
    - S/MIME support and automatic logoff after inactivity in Outlook Web Access
    - Connection filtering
6. Increased server availability and reliability. Improved virtual memory management and support for four-node and eight-node clustering with Windows Server 2003, Enterprise Edition.

    Speedier mailbox recovery and improved support with:
    - End-to-end Outlook monitoring
    - Improvements in the Exchange System Manager tools
    - Mailbox Recovery Center
    - Recovery Storage group
7. Improved Mobile Access with:
    - Built-in remote synchronization support for Windows Mobile–based devices such as Pocket PCs, Pocket PC Phone Editions, and Microsoft Windows Mobile based Smartphones
    - xHTML (WAP 2.x), cHTML, and HTML browser-based devices and mobile devices that can securely browse Exchange 2003 mailboxes
    - Ubiquitous client access using a single URL for all external client types such as Outlook, Outlook Web Access, and mobile devices
8. Investment protection. Capitalizing on existing technology and infrastructure, Exchange 2003 can:
    - Co-exist with Microsoft Windows NT, Windows 2000, Exchange 5.5, and Exchange 2000
    - Run on either Windows 2000 Service Pack 3 (SP3) or Windows Server 2003
9. Collaborative application development. Enable line-of-business applications and other programs to securely access Exchange data and provide integrated workflow scenarios and Web services.
10. The final factor is the impending removal of support for Exchange Server 5.5, currently set at December 2005.

Microsoft Exchange migration process overview

This section briefly details the simple steps required to migrate from an Exchange 5.5 server and NT4 environment to an Exchange Server 2003 environment running on Windows 2003.

The HP recommended standard, simple process for Exchange migration requires the following steps:

1. Establish trust relationship between Windows NT 4.0 and Windows server 2003
2. Migrate Windows NT accounts to Windows Server 2003 using migration tools such as Active Directory Migration Tool (ADMT), supplied by Microsoft.
3. Prepare the Active Directory forest and domain for Exchange Server 2003.
4. Deploy the Active Directory Connector (ADC) to synchronize Exchange Server 5.5 directory with Active Directory.
5. Install the first Exchange server 2003 in an Exchange 5.5 site.
6. Use the Move Mailbox Wizard to move mailboxes.
7. Eventually, use the Active Directory cleanup tool to merge contacts with mailbox.
8. Remove the last Exchange 5.5 server.
9. Switch to native mode.

Microsoft Exchange challenges and opportunities

The steps in the previous section can be easily carried out in a test lab environment. However, the real issues arise when you increase the number of Exchange 5.5 servers and the number of end-user mailboxes being migrated and add in the complexity of systems found in large companies. Suddenly the logistics and the volume of data to be migrated can become overwhelming.

The task of migrating thousands of users and many terabytes of data is a very time-consuming process with significant business risk should anything go wrong. As mentioned in the introduction, the Radicati Group predicts that by 2004 the typical user will send 84 messages a day with a combined capacity of 10 MB per user per day. Given these statistics and the complexity involved in large email migrations, companies need to consider and address a number of key challenges.

Microsoft Exchange Challenge 1—Data growth and data migration

The volume of data under management in a typical Microsoft Exchange environment is a big challenge. Newer systems are better able to cope with larger volumes of data, but first the data has to be migrated to the new environment. After the data has been migrated, Exchange 2003 allows mailbox data to be distributed across a number of storage groups. This essentially breaks the single large database store that was seen in Exchange 5.5 into a number of smaller and more manageable stores. However, as detailed in the previous section, the relentless growth of email data means that the storage capacity problem has only been temporarily relieved by migration to a newer system, and at some point will need to be addressed again.

Microsoft Exchange Challenge 2—Personal archive folder anarchy

The stop gap measures taken by IT organizations to slow down Exchange server capacity growth have simply masked the underlying problem. If email capacity quotas are applied, end users simply create personal archive files (.pst) that are then stored either on local or networked drives. Unfortunately, these workarounds create new problems of archive management, data protection, and capacity consumption. Key questions to consider include: Where are the .pst files stored and what local or networked resources are consumed? Are the archive files properly managed, backed up, and archived? What is contained in personal archives? Is the information being stored securely and is the content appropriate to the needs of the business?

Microsoft Exchange Challenge 3—Shrinking backup windows and data recovery SLAs

Data in user mail accounts must be protected from accidental loss or corruption. Well-managed email environments are regularly backed up to allow data recovery and to ensure business continuity. However, as a result of unprecedented data growth, routine data protection procedures are exceeding the available backup window. In addition, it is not always possible to recover large Exchange databases in line with recovery time objectives defined by the relevant Service Level Agreements (SLAs).

Microsoft Exchange Challenge 4—Compliance

Legislation, due diligence, and IT governance is playing an increasingly important role in the corporate IT landscape. In certain industries it is now a legally enforced requirement to record all emails sent and received by all or selected groups of users. These enforced retention requirements accelerate the growth of Exchange data stores and raise further question regarding the management, security, and protection of data regardless of where the data is stored: in the Microsoft Exchange server, on a shared network drive, or on a user's local C drive.

Deploying RISS as part of your Microsoft Exchange migration

In this section, you will see how an active archiving solution, such as the HP StorageWorks Reference Information Storage System (RISS), can be deployed as part of a Microsoft Exchange migration project. We will show how RISS solves the major Exchange challenges described in the previous section and how RISS can amplify the benefits of upgrading to Microsoft Exchange 2003.

Introducing RISS

The HP StorageWorks Reference Information Storage System (RISS) is a complete active archiving solution for the long-term storage of reference information. Active archiving refers to long-term storage of information *with accessibility*. Reference information is defined as "information that is not expected to change" or alternately "static data or "fixed content."

RISS is a cornerstone of the HP Information Lifecycle Management (ILM) vision, a strategy that provides for the automated management of information throughout its lifecycle according to the value of the data and the needs of the business and as a result turns data into information.
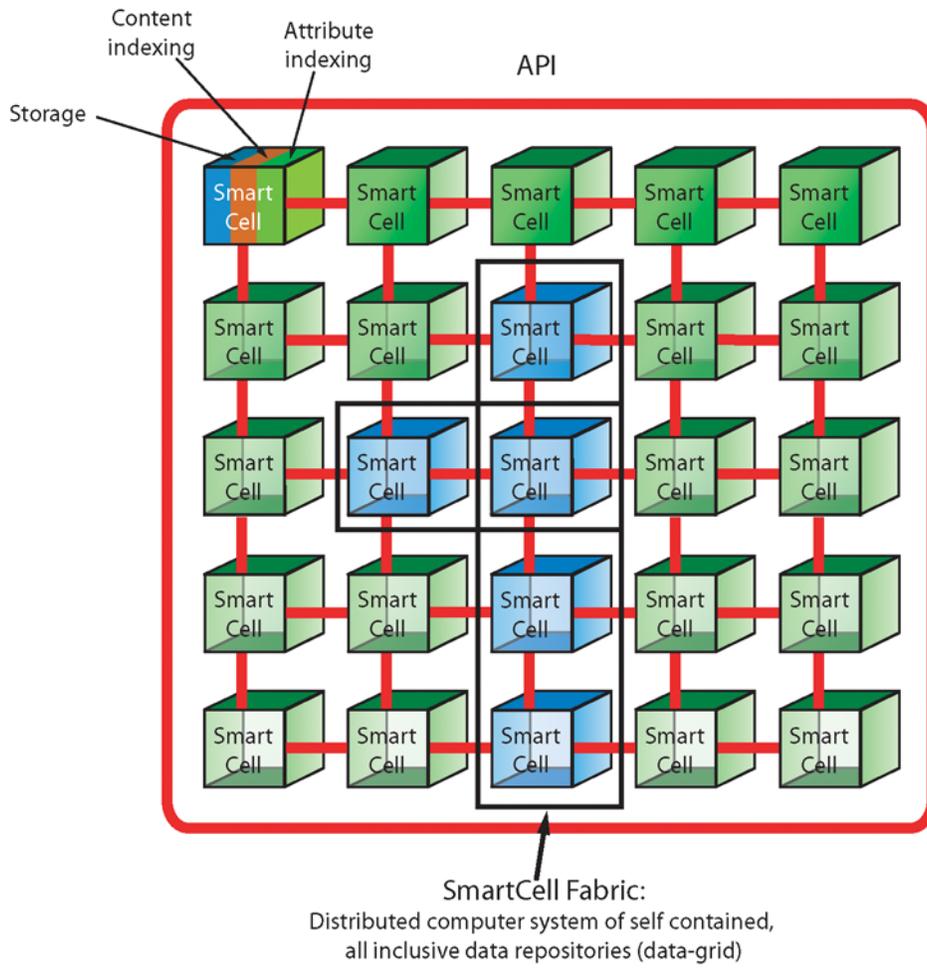
Deployment of a RISS system reduces the total cost of ownership (TCO) for enterprise messaging environments by automatically migrating older email to lower cost storage. This relieves pressure on overworked Exchange server and reduces the time spent on routine storage tasks such as tape backups and capacity upgrades.

Data that has been archived to a RISS active archiving system is protected from modification or deletion and includes digital signature and time-stamping features that can be used to prove the authenticity of stored data. RISS has all the features required to assist customers to meet compliance requirements in regulated industries as well as providing centralized management for old emails that allows organizations to take back control over how email is used.

In addition, RISS uniquely addresses the "scalability gap" that appears in traditionally architected email archiving systems as the number of data objects under management begins to grow. Many customers have discovered that storage architectures that are designed for the active management of a few million information objects do not scale well to address the challenge of storing and tracking billions of objects. Realization is growing that what at first appears to be a data retention challenge, quickly becomes an information access nightmare. This is because traditional architectures cannot support the indexing and retrieval of billions of objects without either performance degradation (and reduced Quality of Service) or alternately, escalating cost, as more money is thrown at providing "bolted on" data access tools.

The RISS system was designed from day one to be a highly scalable solution for active archiving of millions or billions of information objects. Starting with the basic concept of the storage smart cell, HP breaks down a formerly large, indigestible problem into small bite size chunks. Each storage smart cell includes both storage capacity and processing power, allowing dynamically generated content indexes and authentication data to be held locally with the content itself. The individual storage smart cells are then combined in a "storage grid" to allow the searching of multiple cells in parallel (see Figure 1). By leveraging the power of the storage grid, RISS ensures that searches across millions or billions of objects achieve a response time measured in seconds.
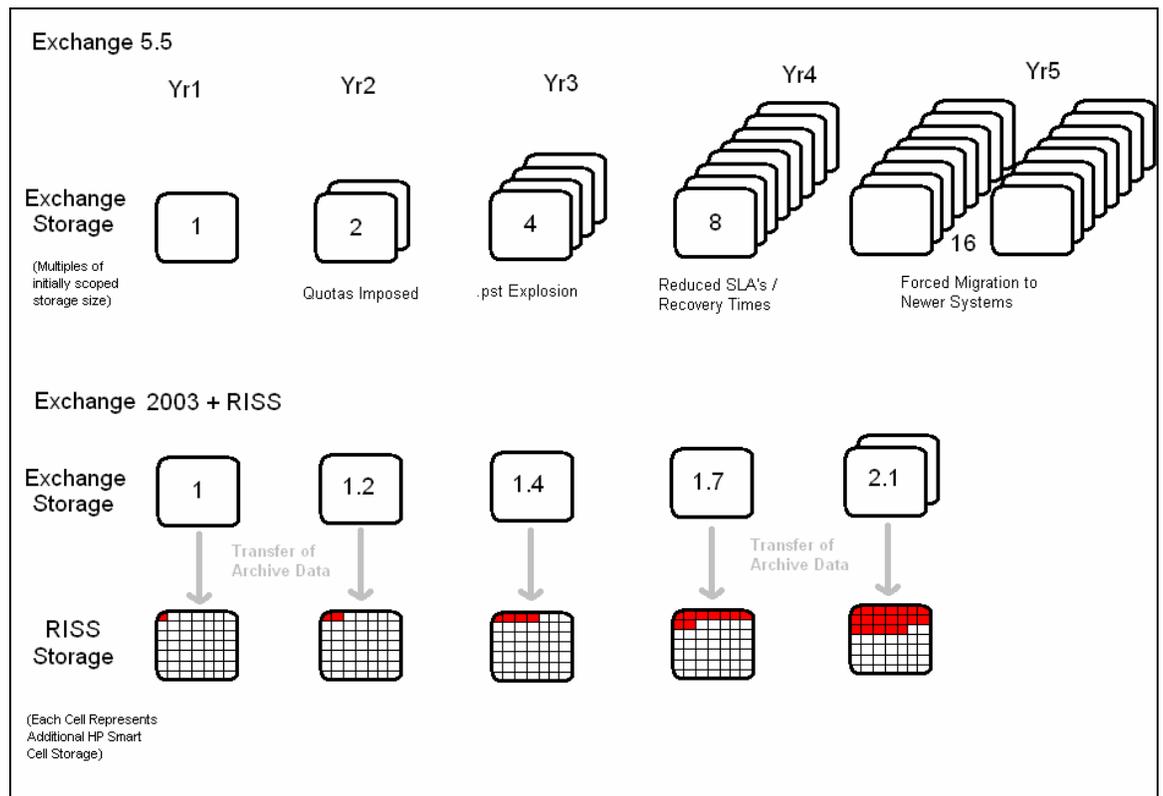
Figure 1.

Content indexing
Attribute indexing
Storage
API

Smart Cell

SmartCell Fabric:
Distributed computer system of self contained,
all inclusive data repositories (data-grid)

How RISS optimizes the Exchange infrastructure design

Incorporating RISS into the final design of an Exchange 2003 solution can address all the challenges identified earlier. The following diagram shows the typical expansion in volume of data in existing Exchange 5.5 installations. With data doubling every year, IT organizations soon resort to draconian measures such as email quota setting. This immediately drives users to "hide" the data in local archive folders or on stored network drives. But as organizations grow and the average email message increases in size, increasing costs and declining quality of service will eventually force a migration. Alternately, the inclusion of a RISS system in the final design radically reduces the volume of data stored on Exchange 2003 by moving selected messages and attachments to purpose-designed and highly scalable, "active archiving" storage.

Figure 2.

Source: Hewlett-Packard, Consulting and Integration

As shown in Figure 2, when Exchange 2003 and RISS are combined, the data stored on the Exchange system is limited to the data that is most active, or in other words "operational data." The bulk of data is less frequently accessed "reference data," and this is automatically migrated to the RISS active archiving system. As seen Figure 3, for most companies, the bulk of data falls into the "reference data" category.

Figure 3.
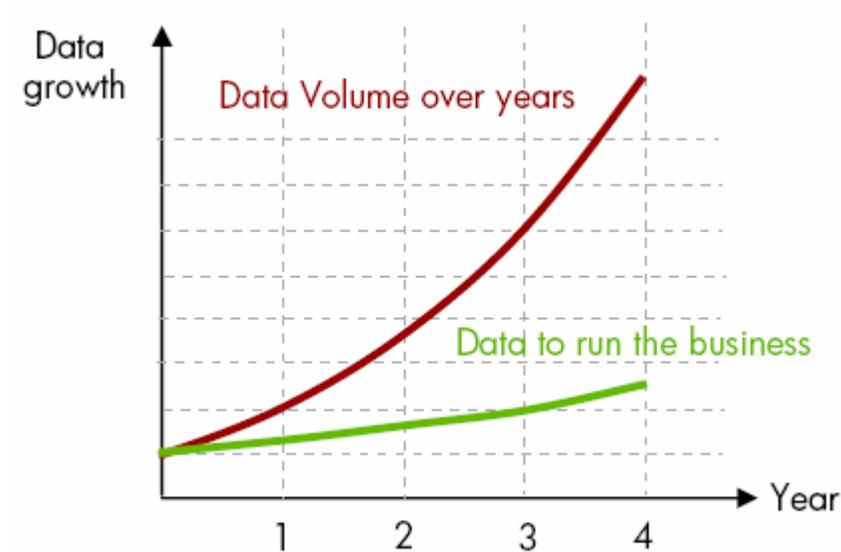
Source: Hewlett-Packard, Consulting and Integration

In addition, because the RISS system records only a "single instance" of email messages, additional space savings are possible in the archive as each unique email is only recorded once.

HP recognizes that the value of data changes over time. In many cases the data value declines, but in some cases the value of data can increase, for example, if required as evidence to defend against a lawsuit. One of the challenges is that users cannot know in advance what data may be needed, in the future, and so companies need to implement data retention polices to reduce risk and, in certain industries, to ensure compliance with data retention regulations. Historical information can also be regarded as an information asset that can be used to support informed decision making, if easy access is provided to that "corporate memory." Figure 4 illustrates the changing use of data over time.

Figure 4.

Data usage over time
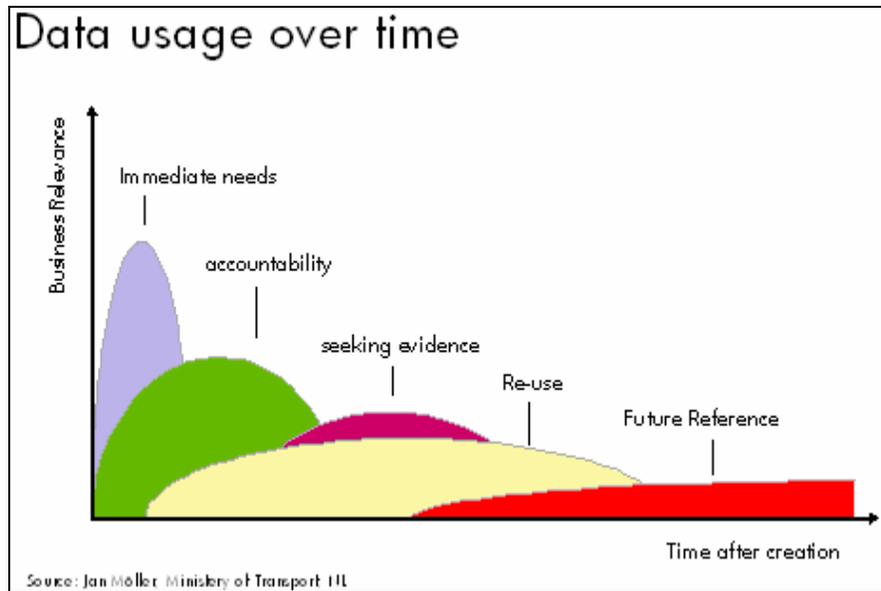
How the email challenges are addressed by deploying RISS

In this section you will see how each of the Microsoft Exchange information management challenges previously identified can be addressed by implementing active archiving using RISS as part of the Exchange Server 2003 deployment.

Microsoft Exchange Challenge 1—Data growth and data migration

The RISS appliance can be configured to automatically archive email messages and attachments, depending upon policies and rules applied by the RISS system. In "selective archiving" mode (see following sections for a full description), as soon the emails and their attachments have been safely migrated to the RISS system, they are automatically deleted from the users mailbox on the Exchange server (a reference [tombstone or stub] is left in the mailbox that typically contains the original e-mail header and a handle for the archived content and is used to retrieve the archived e-mail content), thus freeing up valuable space. Suddenly the new Exchange 2003 environment can cope with increasing email traffic and growing message stores with little or no impact to the Exchange infrastructure. In addition, because end-user mailboxes are now much smaller, more users can be managed from the same Exchange server, reducing investment in servers as well as primary storage.

Microsoft Exchange Challenge 2—Personal archive folder anarchy

RISS obviates the need for users to continually create and store .pst files. Centrally managed archives in a RISS environment require no end-user administration and provide faster access to information by using intuitive search tools. The legacy user .pst files can be easily migrated from local C drives or shared network volumes to RISS and thereafter any message can be searched and retrieved directly from the Microsoft Outlook graphical user interface (GUI). The customer experience is greatly enhanced by removing the concept of email quotas and the need for manual archiving of messages.

Microsoft Exchange Challenge 3—Shrinking backup windows and data recovery SLAs

Because the amount of data stored on the Exchange server has been reduced to the bare minimum of operational data, the time required to perform routine backup tasks is radically reduced. In addition, should it be necessary to restore data from tape, the recovery time SLAs can be met as there is now much less information to recover.

Microsoft Exchange Challenge 4—Compliance

RISS assists companies to meet their regulatory compliance requirements for email systems. The system can be configured to automatically store all emails sent or received by specified user groups and can discover and recover vital information quickly when it is needed for audit purposes, or when requested by a regulatory authority or by a court of law.

In addition, the RISS archive can preserve the integrity of data by using digital signatures and time stamps, and it provides several layers of data protection including logical WORM functionality (data cannot be changed or deleted), redundant copies of data, and a remote replication option.

How RISS works with Exchange Server

This section gives an overview of how RISS integrates with the Microsoft Exchange Server 2003 environment.

Unregulated user environments

Figure 5 shows RISS installed as part of a Microsoft Exchange infrastructure in an unregulated user environment. "Unregulated" describes e-mail that is not subject to explicit regulations by government or industry bodies in charge of policing corporate governance.

Figure 5.
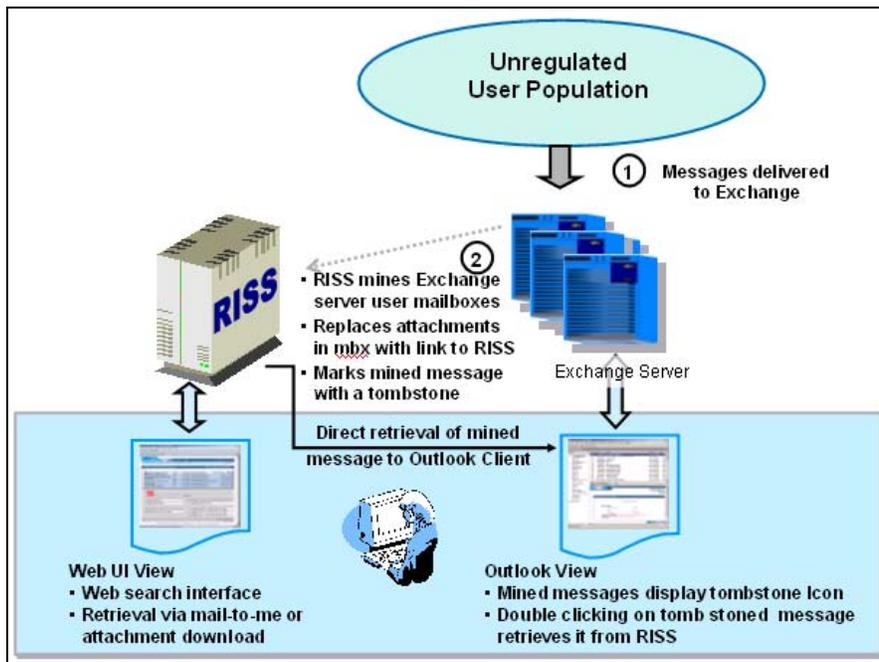
In this scenario, RISS is used to mine messages from the end-user mailboxes and archive the content to the RISS system. Each message is mined in its entirety including any attachments that may be present. Each mined message is marked with a "tombstone" within Outlook. Users simply double-click tombstoned messages and they appear within Outlook as a normal message.

Users can also search and retrieve archived messages using a RISS Web Browser. Searching includes full word searching of all message contents including full text searching of attachments.

Full text searching of attachments is currently supported for the following document types:

- MS Word (.doc)
- MS PowerPoint (.ppt)
- MS Excel (.xls)
- MS Access (various file extensions)
- MS Outlook personal folders (.pst)
- Adobe Public Distribution Format (.pdf)
- Rich text format (.rtf)
- Outlook encapsulated .rtf (TNEF)
- Hypertext Markup Language (.html)
- Text files (.txt)
- ASCII files (.asc)

NOTE:
Files of other types can be moved to the archive but will not be automatically keyword indexed, but will be retrievable by file name, file extension, or other standard meta data.

More about mining

The rules that determine when messages are migrated to the RISS archive can be finely tuned to an organization's business requirements and can be designed to maximize positive impact on the existing Exchange infrastructure. Most of the attributes of an email message can be used to define mail migration policies. For example, polices can be applied such as:

- Only mine items that are older than 90 days **and** greater than 10 KB in size
- Mine only the attachment **or** the attachment and the message body.

These and other policies can be combined to reduce mailbox sizes by up to 80% on a typical Exchange server.

Regulated user environments

RISS can also be utilized where users or corporations work in a capacity that must adhere to pre-specified government or corporate regulations such as SEC, FDA, HIPPA, and so on.

Figure 6 shows RISS in a regulated environment.

Figure 6.
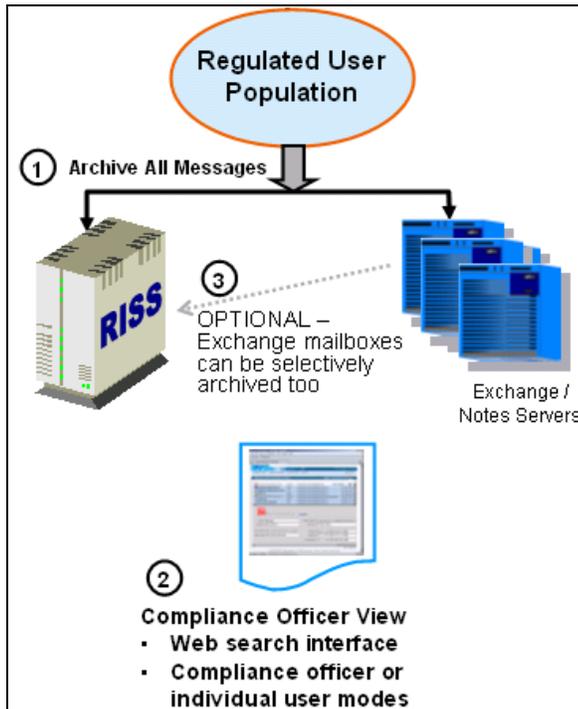
In this scenario, RISS captures all messages that go to the Exchange Information Store. The messages may be retained for a pre-defined period or retained indefinitely.

Each message within the RISS is digitally fingerprinted, time-stamped, and digitally signed so that the data cannot be tampered with and so that the authenticity of the content can be proven at a later date.

Powerful tools are provided to allow authorized personnel such as compliance officers or regulators to review the mails stored in end-user archives.

Additionally, regulated end user mailboxes can also be selectively "mined," thus offering the twin benefits of enforced retention and cost reduction.

RISS can be deployed for a mixed community of regulated and unregulated users. Therefore, large organizations that need to adhere to compliance regulations for certain departments and sections can selectively regulate and archive mailboxes as required.

On-going benefits of the RISS active archiving system (amplifying Exchange 2003 benefits)

When the RISS migration is complete, the establishment of a structured, business-aligned email archiving process in effect **amplifies** the benefits offered by Microsoft Exchange Server 2003 (as previously discussed).

Enhanced end-user productivity:

- End users no longer need to manage mailbox quotas.
- End users do not need to manage a .pst archive.
- Users spend less time searching for information because of the powerful and intuitive search tools provided by RISS, including full text search of email header information, email, text, and attachments.
- End users can initiate restores of individual or multiple messages together with all attachments.

Increased IT staff productivity:

- Use of RISS email archiving reduces storage capacity of Exchange server by around 80%.
- Less frequent capacity upgrades are required.
- Backup windows are reduced; disaster recovery is faster (only 20% of original data needs to be recovered).
- End user archived data is centralized and properly managed.
- End users no longer call IT for individual message restores (old process: find the right tape, recover whole mailbox to separate server, recover the mailbox, folder, or individual message).

Lower cost (reduced equipment):

- RISS archive system uses lower cost storage than enterprise email servers.
- All archiving operations are automated and transparent.
- Exchange server capacity upgrades are less frequent.
- Costs of managing the email environment are reduced.
- Exchange server backup process improvements:
  - Smaller backup window
  - Less tape media (reduced capacity)
  - Restore only for Disaster Recovery (individual emails restored by users)

Security:

- The RISS product has several features designed to help customers meet their compliance requirements including:
  - Managed retention periods
  - Digital signatures/time stamps
  - Logical WORM media
  - Duplicate copies of all data
  - Data cannot be altered or deleted
- Data in the archive can be audited by compliance officers, supervisors, and so on.
- In the case of a lawsuit, evidence can be quickly produced and content authenticity substantiated.
- Access controls—users can only access emails that they are authorized to review.

Conclusion

This white paper has demonstrated how the deployment of the HP StorageWorks Reference Information Storage System can amplify the benefits of the migration from Exchange 5.5 while simultaneously provide many lasting benefits, including permanent cost reduction, risk mitigation, and improved information access by establishment of a structured, active archiving platform for storing reference information.

For more information on RISS, contact your local HP sales office or an HP Authorized Reseller or visit http://h18006.www1.hp.com/products/storageworks/riss/index.html.

Appendix A

RISS compatibility

Currently RISS is compatible with the following email systems:

- Exchange 5.5
- Exchange 2000
- Exchange 2003
- Lotus Domino V5 and V6 in compliance mode only
- Sendmail in compliance mode only

Email Client compatibility:

- Microsoft Outlook 2000
- Microsoft Outlook XP
- Microsoft Outlook 2003

Instant Messaging Archive Server

- IMlogic (covers most popular IM formats)
- Akonix (covers most popular IM formats)