

Solving Challenges of Archiving Globally Distributed Microsoft Exchange™ Environments

Joseph Winchester, MCSE

This whitepaper focuses on solving the challenges of archiving a fully distributed global enterprise. Geographic distance, time zones, languages, and political restrictions are but a few of the challenges that can be resolved with the right archiving solution. EAS 3.0 provides the most efficient method available on the market for archiving Microsoft Exchange environments with multiple, geographically dispersed sites. When used with EAS-StoRM, the total solution maximizes the ROI for storage devices, and minimizes the amount of data flowing over Wide Area Network (WAN) links. This paper provides an in-depth look at the architecture of this highly scalable enterprise solution.

Executive Summary

Many civilizations throughout history have grown beyond their original borders to encompass a much broader geographic area. In some cases the growth was due to natural processes like population growth and expansion, and in others the growth was forced through military expansion or a policy like “manifest destiny”. Whatever the stimulus that caused the borders to expand and the area under the civilization’s control to increase, one fact is evident in each expansion, namely the challenges that come with governing a distributed populace and their associated assets.

A centralized government would typically establish smaller governing bodies in remote locations to assist in controlling and providing for the local population. In the Roman Empire military outposts and governorships were set up in remote locations with communications infrastructure (runners, chariots and message carriers) linking these to the central command in Rome. While the physical distance and geography provide one type of problem to overcome, other challenges like indigenous languages, monetary exchanges, religious beliefs and cultural traditions posed additional trials for the leadership. Failure to resolve the challenges presented by a geographically dispersed infrastructure often led to the downfall of the governing empire and its replacement by more efficient rulers or smaller, more manageable territories.

Management of a global corporation with a distributed communications infrastructure can present similar challenges as those faced by the leaders of large empires. Wide-area network communications between global corporate offices pose similar problems as the Internet architecture itself. In an early article on web caching, researchers at UC San Diego commented on the topic of global infrastructure, saying,

“The Internet’s sustained explosive growth calls for an architected solution to the problem of scalable wide area information dissemination. While increasing network bandwidths help, the rapidly growing populace will continue to outstrip network and server capacity as they attempt to access widely popular pools of data throughout the network.”¹

The importance of the researchers’ statement is their realization that simply throwing more bandwidth at the problem was not the proper solution, rather, the solution is to architect applications and solutions in such a way as to efficiently utilize the existing network infrastructure and deliver the desired value. The Exchange Archive Solution (EAS) 3.0 from EDUCOM TS does exactly this for globally distributed email management and archiving of Microsoft Exchange™ environments.

Email is probably the most important communications application in a global enterprise. Because email is tied directly to the problem discussed in the previous paragraph it is essential that any solution related to email management or archiving be designed to alleviate wide-area network communications concerns. Corporations using Microsoft Exchange and implementing an email management solution for archiving and long-term retention need to ensure that their chosen archive solution is truly designed for globally distributed environments. EAS 3.0 delivers a Parent-Child architecture with Advanced Clustering that is highly scalable and provides for the most efficient archiving, retention, search and retrieval of messages in a globally distributed corporate environment.

The following document discusses the challenges that face corporations with globally distributed email environments and reviews the technical qualities that should be considered in selecting an archive solution. EAS 3.0 architecture benefits will be presented with some example illustrations of environments ranging in size from a single Exchange site to a global multi-site Exchange environment with EAS 3.0 archiving implemented.

¹ “*Evolution of the NLANR Cache Hierarchy: Global Configuration Challenges*”, Duane Wessels and K. Claffy, UC San Diego, (1996)

Going Global: Distributed, Multi-Site Enterprises

When a company steps foot outside of its first location it begins to experience the challenges of “going global” even if they have not yet left their native country. This is particularly evident in the area of information technology (IT) infrastructure and communications applications like Microsoft Exchange™. While operating in one location, assets like bandwidth are not really a concern, with most networks having plenty of available bandwidth on the Local Area Network (LAN). Once a company goes to even two locations in the same general area, they have to create a new LAN at the second location, then establish Wide Area Network (WAN) links between the two LANs, establish secure network routing, and get the computers and applications at the two sites to communicate with each other. Even if that were easy to accomplish, the next item of focus will likely be optimizing the efficiency or effectiveness of the communications between computers and applications on the WAN as many applications are not really designed for distributed environments.

Microsoft has developed and improved Microsoft Exchange to operate more efficiently in multi-site environments, one reason it has become the leading corporate email application in the world. Corporations still have to consider whether or not their selected email management or email archiving solution will also operate efficiently in a multi-site environment, and if the application will properly serve their business needs in a multi-site environment. As we will see later in the discussion, the Exchange Archive Solution (EAS) 3.0 from EDUCOM TS is designed to accommodate either the single-site company or the multi-site corporation, as it is architected to provide the most efficient utilization of network (LAN or WAN), storage and computing assets in the Exchange environment it serves. This continues to be true as we enter the challenging arena of the global corporation,

Going beyond a multi-site corporation to a truly global corporation one will encounter a new set of challenges for the IT environment. It is necessary that any application that is deployed in this unique environment will adequately handle the associated challenges. Some of the factors that will put applications to the test in global environments are:

- Multiple languages and localization
- Multiple currencies, conversion and fluctuation
- Political restrictions and regulatory requirements
- Available computing and storage hardware
- Available bandwidth and telecommunications infrastructure
- Adequate provisions for mobile (laptop) users

Of these listed items, all can be a factor in the implementation of an email management solution except the multiple currencies, as an email solution will probably not be concerned with currency other than the proper display and handling of currency characters. One of the leading email archiving applications, EAS 3.0, has developed a number of features to directly answer the challenges of a globally distributed corporation. Some of these features include:

- Parent-Child architecture for multi-site archiving
- Support for over 30 languages in the EAS client
- Support for Mobile users
 - PST Migration and archiving
 - OST integration
 - Outlook Web Access (OWA) support
 - Optional Browser-based EAS client
- Advanced Clustering
- Performance Monitoring and Thresholds
- Up to 80% compression for bandwidth and storage savings
- Storage and Retention Manager (EAS-STORM) supporting a global storage infrastructure

A Flexible, Highly Configurable, Open Architecture

EAS 3.0 is leading the evolution of email management. One of the reasons for this leadership is the open architecture upon which the solution is based. Fast, efficient, and limitlessly scalable for any enterprise environment, EAS leverages the customer's existing information technology (IT) environment and accepted computing standards to deliver enterprise-class email management.

The open-architecture design makes EAS highly flexible in its implementation, with the following components typically involved in a complete EAS 3.0 environment:

- Microsoft Exchange Server(s)
- EAS 3.0 Server(s)
- Storage and Retention Manager (STORM) Server
- Storage and Retention Manager (STORM) Client
- EAS Search
- IIS Server(s)
- Database Server
- Outlook Client
- Outlook Web Access (OWA) Client
- EAS Search Web Client

It is important to note that each "server" above does not necessarily require a separate physical server, rather is representative of a server application. Many of the applications can co-exist peacefully on the same physical server. For example, EAS 3.0 Server, the STORM Server and STORM client could reside on the same server while the EAS Search and IIS might be installed together. If the IIS Server were hosting additional high traffic websites or intranet sites, perhaps EAS Search would be installed on a separate server to minimize the impact. Typically Exchange Servers and Database Servers are installed as single-purpose servers due to the usually high overhead the application places on the server hardware. EAS is designed as an open architecture, giving the enterprise customer choices for their deployment and the ability to utilize existing assets to deploy the total solution. For example, the EAS solution can leverage IIS Servers and Database Servers (Oracle or SQL) within the existing enterprise environment. Additionally, the EAS applications have been efficiently coded with highly efficient programming languages like C++ to allow them to run on minimal hardware. The CTO (and Chief Architect) at EDUCOM is fond of saying that an organization could effectively re-deploy existing server assets to minimize the cost of the total solution hardware, as EAS 3.0 will "run on just about any hardware".

EAS 3.0 can be deployed in a number of configurations, depending upon the particular needs of the enterprise customer. Some of the factors that will determine the most effective configuration include:

- Number of Exchange Servers
- Number of Exchange Sites
- Number of email clients in the enterprise
- Amount of daily email traffic
- Size of daily email traffic
- Lifecycle and retention requirements of the enterprise
- Available storage
- Available LAN and WAN bandwidth

As the discussion of the EAS 3.0 architecture continues, example configurations will be presented to illustrate the flexibility available to customers deploying the solution. One additional benefit of the open and modular architecture of EAS 3.0 is that customers can implement the entire solution, or the just the parts that make sense for their business needs. Customers can also grow their EAS deployment along with the needs of the organization, or roll out the application for part of the users or geographic sites, and

later expand the implementation to additional sites, users or groups. The first implementation example that will be covered in the next section is EAS 3.0 and EAS Search in a single site without the Storage and Retention Manager (EAS-StoRM).

Simple EAS Model (Single Site)

In a rather simple Microsoft Exchange environment involving a single physical location with one or more Exchange Servers, EAS 3.0 will provide significant value as an email management solution. With EAS 3.0 and EAS Search the customer will be able to archive, search, and retrieve Exchange message class data while greatly improving the performance and scalability of their Exchange environment. Some of the technical benefits of a simple EAS environment include:

- Greatly reduce the size of the Exchange Information Store
- Reduce the Exchange backup window due to reduced Information Store size
- Improve performance of Exchange Servers by offloading older message data to Document Store
- Support additional users with each Exchange Server or reduce the number of Exchange Servers required to support the existing users.
- Reduce overall Exchange storage requirement through reduction of Information Store and up to 80% compression of data archived to Document Store

Building upon the open architecture design of the EAS 3.0 environment, only standard network protocols are used to facilitate communications among the various components of the solution. In this simple EAS environment, only TCP/IP, HTTP, MAPI and ODBC protocols are used. The communications protocols used in the EAS environment are illustrated below in Figure 1.

The following diagram provides an example configuration for a single-site EAS environment:

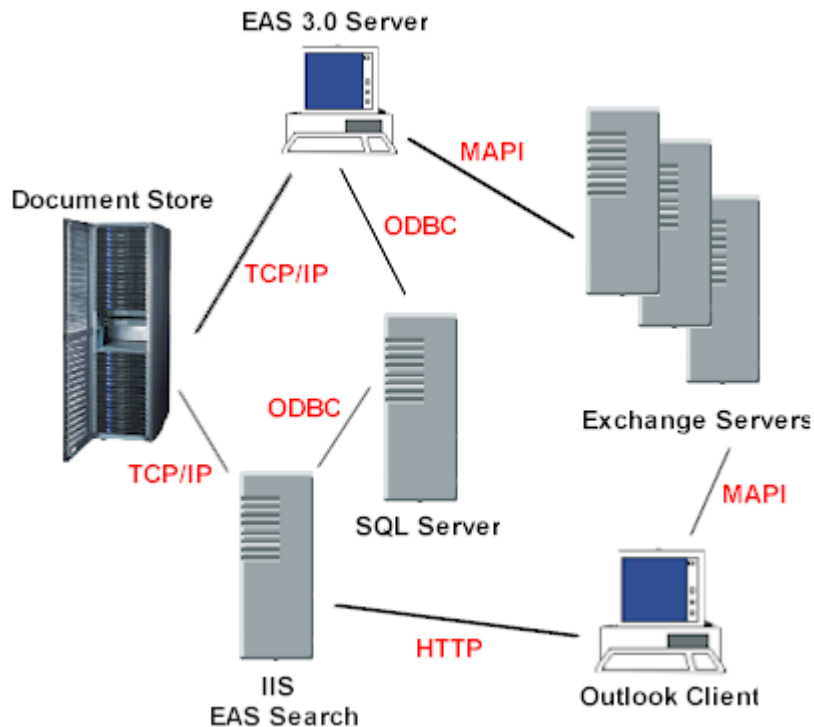


Figure 1: Single-Site EAS 3.0 with EAS Search

It should be noted that the configuration shown in Figure 1 is only representative of a “simple” EAS 3.0 environment, and that there is considerable flexibility available in a single-site deployment to serve the enterprise messaging and retention requirements. For example, if the single site were sufficiently large or the business continuance requirement was such that additional EAS 3.0 Servers might be deployed to provide load balancing and failover. EAS 3.0 can monitor its performance and hardware resource utilization and administrators can set thresholds to maintain certain performance levels. The additional EAS Servers could be configured to take over EAS processing when the initial server reaches a particular performance threshold. EAS 3.0 can also have multiple Document Stores to which message data can be archived. Document Stores can also be configured for a failover situation, should one of the Document Stores become unavailable for any reason. IIS and EAS Search work together to locate and serve archived message data to Outlook clients and administrators. In the above diagram, these two components are installed on the same server. In environments with high search usage or traffic, multiple IIS servers can be used to improve the performance of archive searches. In smaller environments (or with the right multiprocessor server), the administrator might decide to combine the installation of the SQL and IIS with EAS Search. Again, the key with the EAS 3.0 environment is open architecture, flexibility and enterprise scalability.

For enterprises with a requirement or desire to retain message data for longer periods of time, EDUCOM TS introduced the new Storage and Retention Manager (EAS-STORM). EAS-STORM is an archive engine that manages the EAS Document Stores or original archives created by the EAS Server. It enables easy and flexible task-based policy administration of the Document Stores to comply with record retention requirements, and to effectively manage storage resources. Using granular criteria (right down to full content management of messages and attachments) items can be purged, copied, or migrated from the online archive. EAS-STORM's comprehensive, policy-based automation can be applied at various levels: individual user, group, distribution list, and EAS Document Store.

The following diagram builds upon Figure 1, adding EAS-STORM and a direct-attached tape device:

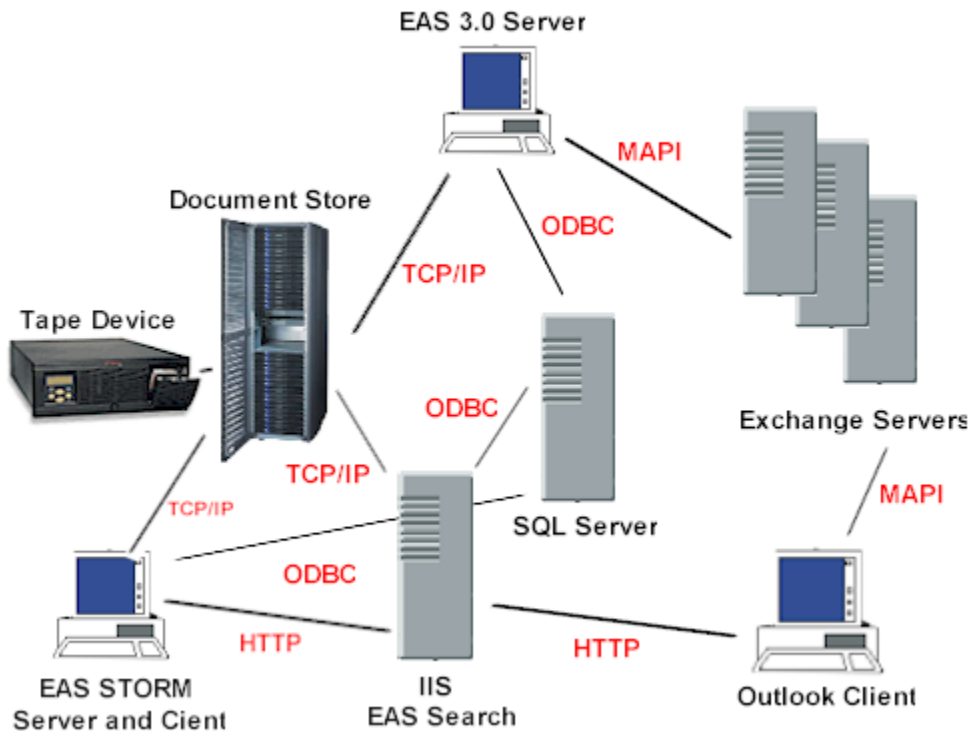


Figure 2: Single-Site EAS 3.0 with EAS-Search and EAS-STORM

For enterprises using EAS-STORM, the open architecture of EAS further lends itself to scaling the environment without the deployment of additional server hardware. As Figure 2 displays, it is easy to add EAS-STORM to an existing EAS environment to leverage the storage and retention management capabilities. The EAS-STORM Server can be installed on the same hardware as the EAS Server for example, depending upon the archiving and storage management tasks that will be utilized. EDUCOM TS and its reseller and service partners can assist the customer in determining the optimum setup to match the enterprise business and technical requirements. As with the other components of the total solution, EAS-STORM also leverages an open architecture, and uses standard communications protocols to make deployment fast and readily compatible with most enterprise IT environments. STORM is a client-server application, where the EAS-STORM client communicates with the EAS-STORM Server using RPC, enabling efficient communication either on the same computing device or across the network.

The EAS-STORM module of EAS 3.0 manages the movement or further archiving of email data from the Document Stores located on the secondary storage devices, to the tertiary or near-line storage and further archives to off-line storage devices. In the above example, the EAS Server archives message data to the Document Store located on a NAS device, and the EAS-STORM Server can further “archive the archives” to the magnetic tape device attached to the NAS server for longer term retention and management. Even when archive data is taken offline, EAS-STORM will continue to retain the location of each message object until its retention period is expired. The flexibility of the module allows administrators to design and implement the levels of archive that are appropriate for their email retention management needs. The EAS-STORM Server can automatically process all archived data according to the prescribed administrative rules, executing a copy, migration, or purge of identified message data throughout the EAS architecture. (See EAS-STORM whitepaper titled “Beyond the Email Archive...”)

Enterprise EAS Model (Multi-Site, Parent-Child Architecture)

The open architecture of EAS becomes even more valuable to the enterprise with multiple sites, particularly when geographic distances restrict the amounts of data flowing over the Wide Area Network (WAN) links. In this section we will continue to build beyond the single-site EAS 3.0 environment and begin to truly leverage the scalable nature of EAS 3.0.

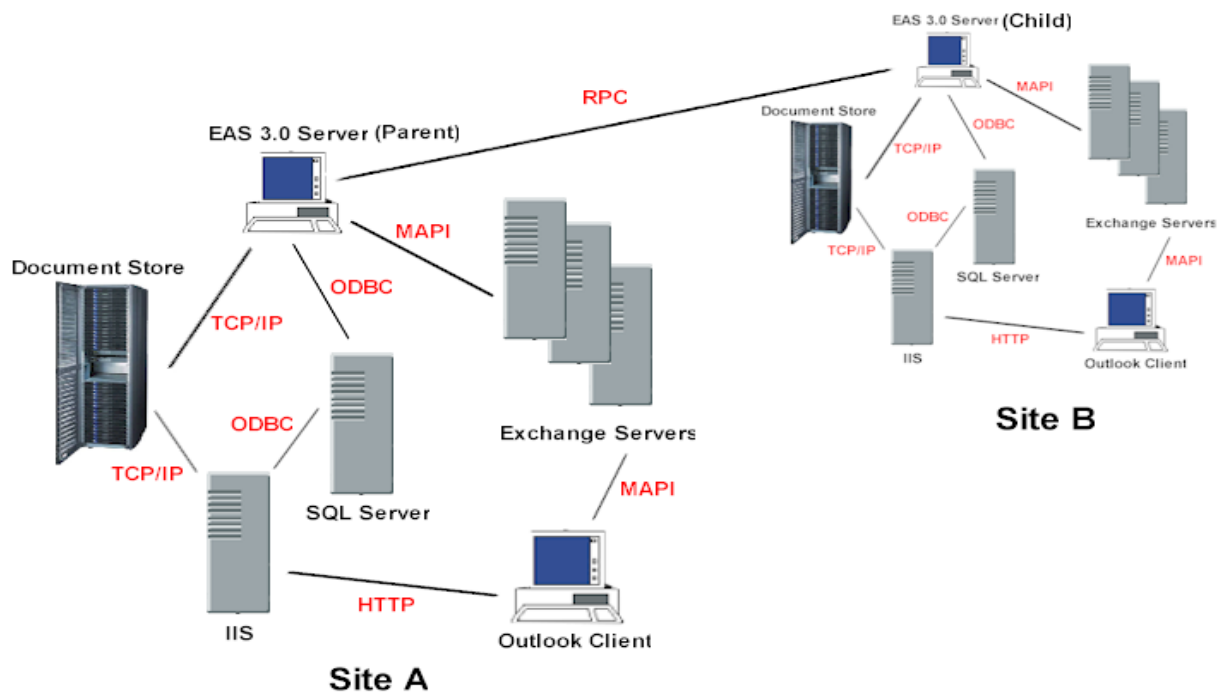


Figure 3: Dual-Site EAS 3.0 and EAS-Search

EAS 3.0 is the first email management application to truly provide the flexibility in deployment and scaling that is needed for a globally distributed enterprise. EDUCOM has developed a “Parent-Child” architecture within EAS 3.0, which allows for centrally managed and administered control of email policies while still allowing for regional dependencies and administration of messages. While there will be only one Parent Server in an EAS 3.0 environment, there can be any number of geographically dispersed EAS Child Servers, providing for limitless scalability. Users can access, retrieve and store email regardless of their geographic location. This open and distributed architecture reduces server overload and eliminates bandwidth concerns even as it allows employees around the world to work from their desktops or mobile laptops.

Flexibility of deployment is one of the crucial characteristics of any application used in an enterprise environment. IT administrators need to be able to configure the application to use their existing IT assets whenever possible in order to reduce the Total Cost of Ownership (TCO) of the solution. EAS 3.0 provides this flexibility. For example, in the following diagram, a third site is added to the configuration that is shown in Figure 3, however the third site does not have a dedicated storage server for the EAS implementation. EAS 3.0 can utilize any UNC addressable storage for the Document Store, so for a smaller site, EAS can use a share of the available storage on one or more of the application servers, perhaps on the IIS Server, EAS Server, or both as shown below.

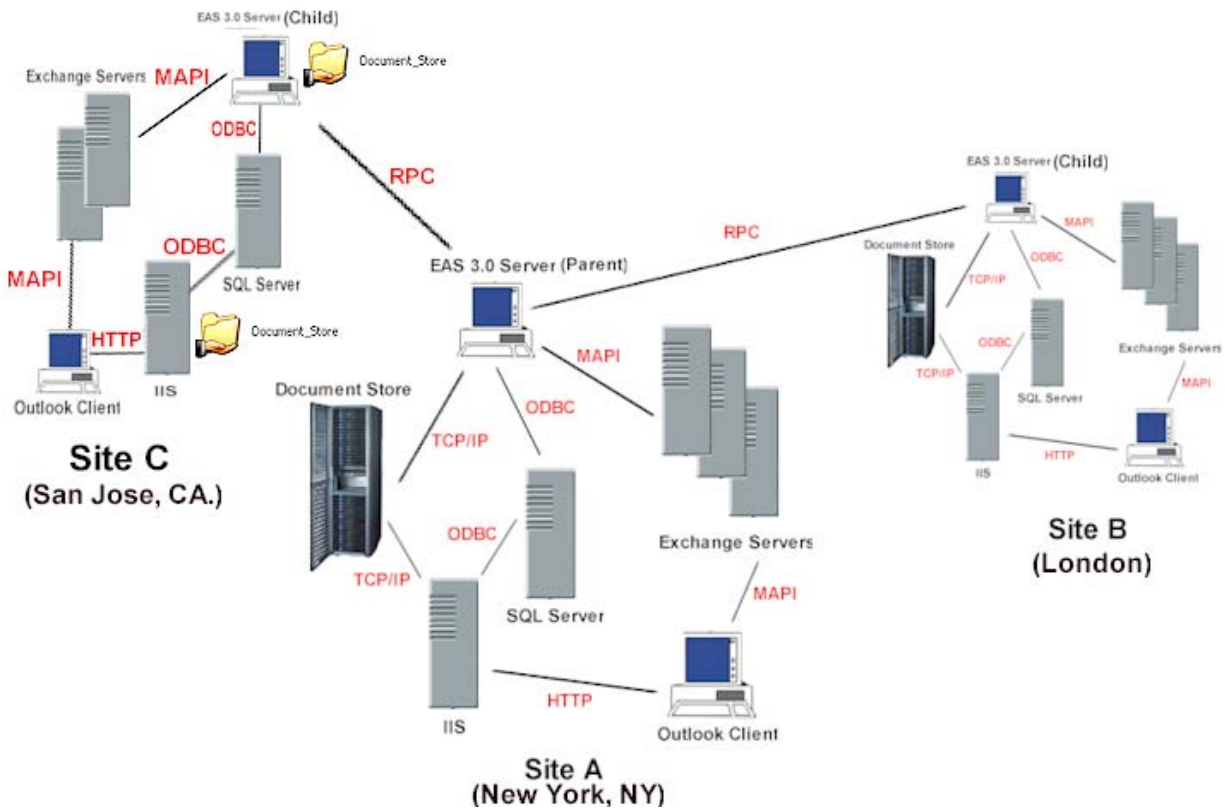


Figure 4: Multiple Geographically Dispersed Sites with EAS 3.0 and EAS-Search

Adding to the flexibility of deployment and the optimization of WAN link usage, EAS 3.0 can utilize read-only copies of the SQL Server database in the remote sites as shown in these diagrams. In this scenario, local user search and retrieval is carried out with a minimum of traffic crossing the WAN links between the Child Sites B or C and the Parent Site A. Although quite flexible in “simple mode”, further optimization and flexibility of deployment becomes available when administrators use the Advanced Clustering features within EAS 3.0.

Advanced Clustering Saves Space and Time

The CTO at EDUCOM TS is also fond of saying, "**the classic computer science quandary is the balance between SPACE and TIME**". Every enterprise organization will have certain expectations that drive their IT implementations. EAS 3.0 gives the administrator the flexibility to design an archive environment for Exchange Server that will fit their particular business needs, balancing the long-term usage of "space" or storage, and maximizing the effective use of "time" through a multitude of performance optimizations for geographically dispersed deployments.

EAS 3.0 further supports geographically dispersed enterprises with a feature called **Advanced Clustering**. Advanced Clustering allows the EAS Administrator to define "clusters" within the EAS environment. A cluster in a distributed, multi-site EAS environment is a logical grouping of EAS Servers, IIS Servers, Document Stores and Exchange Servers along with their **preferences** and **priorities**. Typically EAS Child Servers will belong to a particular cluster, and will direct the archiving activities of that cluster, and "roll up" the information to the EAS Parent Server under which it processes. It is within this Advanced Clustering mode that the preferences, priorities and performance monitoring capabilities of EAS deliver strong flexibility to the administrator. The following examples illustrate the flexibility of this architecture using preferences and priorities.

- Within a cluster, an EAS Child Server will be primarily responsible for particular Exchange Servers. The Child can be assigned a priority for each of the Exchange Servers within the cluster, and be assigned performance thresholds for each to control the impact on the network and processing environment.
- While an EAS Child Server is primarily responsible for the Exchange Servers within its cluster, it can be configured to assist other EAS clusters with their archiving tasks as well.
- An administrator can assign a preferred Document Store for a cluster, meaning that the Exchange Servers in that cluster will only be archived to that "preferred" Document Store".
- If an administrator assigns multiple Document Stores to a cluster, then he can also assign a priority to a particular Document Store, or even restrict the archiving of a particular Exchange Server to a certain Document Store.
- IIS Servers can also be assigned priorities and preferences to control the routing and retrieval of message data from the archives to maximize the efficient use of network bandwidth.

Advanced Clustering allows Parent and Child EAS servers to carry out archiving of their preferred Exchange Servers (typically those local to the EAS server). Simple archiving on the other hand says that there is typically a single storage location...where all archived records will be stored.

Advanced Clusters will allow for multiple document stores, each of which will serve as a Document Store for specific Parent or Child EAS servers. Additionally, Advanced Clustering allows EAS servers to archive particular Exchange servers, by order of set preference, and to assist with the archiving of other Exchange Servers following the completion of their assigned archive tasks. All of these tasks are assigned via archiving "rules". As the EAS Server archives its assigned Exchange servers, it will deposit the messages into the Document Stores based upon rules that determine the desired storage location(s).

Advanced Clustering is also instrumental within the EAS architecture for ensuring the highest possible level of true **single instance storage**, while simultaneously applying archiving rules that will serve the physical or geographic needs of the organization. Once again, the design of the archiving rules, the physical locations of the document stores, and the location of the Exchange users, will dictate the best deployment of advanced clustering.

Don't Forget Your "Road Warriors"

Any organization that has multiple geographic sites will most probably have a few "road warriors" or mobile network clients. It is imperative that the organization provide easy access to all of a particular user's applications, documents, email, and other digital accessories while that user is on the road or

visiting one of the other enterprise sites. Any application that serves this distributed enterprise should include provisions for this remote access. EAS 3.0 builds in a number of features to support the mobile computing user within an Enterprise EAS environment. Some of these features include:

- Integration with Microsoft's Outlook Web Access (OWA) Client
- Automated PST Gathering and Migration
- PST Archiving
- Off-line (OST) integration
- EAS Web Client

Most mobile clients will likely use a PST file or an OST file to retain their email, calendar, contacts and other message class data in order to maintain access to this information while they are on the road. As these users carry valuable corporate information, the enterprise will want to ensure that these users' message data are archived sometime when they are connected to the corporate network. Additionally, the users will require that their personal folder structure is maintained, and that they can search and retrieve the archived items when needed.

PST Gathering and Migration- EAS can search profiles on individual workstations to locate PST files on the users' local drive. EAS can then move or "gather" the PST files to a central network location where EAS will archive the messages and place stubs in the PST file for archived messages (typically all messages). In addition EAS can then migrate the stubs from within the archived PST file into the appropriate user's Exchange mailbox if desired, maintaining the existing folder structure from the original.

PST Archiving – Once the users' PST files have been identified and located, they can be archived in the same manner as the users Exchange inbox. With the new PST Stub Migrator, EAS will populate any archived files that exist in a users PST file in the same folder within that user's mailbox. If that folder is not found in the user's mailbox, EAS will create one to match the one found in the PST file. This is particularly useful if the PST files are to be eliminated in the enterprise following the migration.

OST Synchronization – EAS supports offline synchronization. If there is a archived item in a folder marked to be available "offline", Outlook will keep a synchronized copy of it in the associated OST file. EAS will also make a local copy of the original email and attachments so it can be retrieved when the mobile user is working offline.

All of the above features ensure that the mobile user has continual access to their email data, even the Exchange data that has been archived by EAS. The web access clients allow the user to connect via the Intranet or Internet (depending upon the implementation) to search, view, and retrieve archived message data should their Outlook client be unavailable for any reason.

Summary

Throughout the discussion we have approached global enterprise challenges with a particular interest in mind, namely solving these while archiving geographically distributed Microsoft Exchange Server environments. EDUCOM TS' EAS 3.0 is the only product available on the market at the time of the writing of this document that has the necessary architecture and features to properly serve this kind of distributed environment. The features discussed that solve the global archiving challenges include:

- The Parent-Child architecture that scales limitlessly from a single site to multiple geographically dispersed sites, while minimizing the impact on WAN traffic, server and storage resources.
- Storage and Retention Manager (EAS-STORM) that supports a global storage infrastructures for the efficient storage, long-term retention, search, retrieval and expiration of archived message data. EAS-STORM additionally provides tools that support regulatory compliance requirements.
- Support for over 30 languages in the EAS client.
- Strong support for Mobile users including PST Migration, PST archiving, OST synchronization and web accessible clients.

- Advanced Clustering with configurable preferences, prioritization, load balancing and failover.
- Performance Monitoring and Thresholds allow the administrators to monitor and control the usage of valuable network and computing resources throughout the implementation.

As mentioned in the discussion, EAS 3.0 is a highly configurable email management solution for Microsoft Exchange Server environments. Its flexible, open architecture allows for deployment in a manner that will suit the business needs of any organization, from a small site to a global enterprise. It is exactly this flexibility and scalable architecture that makes the solution ideal for the enterprise client. Additional documentation is available on the various aspects mentioned in this paper, and readers are encouraged to contact their EAS reseller representative or EDUCOM TS if more information is desired.

Joseph Winchester, MCSE

Joseph currently assists EDUCOM's marketing team following his successful entry into marketing with almost three years in the data storage industry focused on enterprise solutions and strategic alliances. Before that Joseph was an independent network consultant and served with systems integrators as a network services manager and systems engineer. Joseph is a Microsoft Certified Systems Engineer (MCSE) with specializations in TCP/IP and Microsoft Exchange. He has also earned his Master Accredited Systems Engineer (MASE) from Compaq (now HP) with a concentration in Enterprise Systems Management. Joseph has a BS degree from the University of Houston and is presently working on an E-Commerce MBA.

WHO IS EDUCOM?

EDUCOM TS Inc. is an industry leader in the development of software solutions focused on the mission-critical management of corporate email archiving and retrieval. Through our proven technology, our strategic relationships with partners and resellers, and our keen grasp of global regulatory issues, EDUCOM delivers a second-to-none solution to the international business market. The company's flexible products enable clients to efficiently meet various email retention requirements, protect their intellectual property, enhance user access to information, and reduce email server overload. Its flagship product, Exchange Archive Solution (EAS), offers intelligent management for Microsoft Exchange email stores. EAS-Wireless is designed to manage the email of mobile users. EDUCOM is located in Ottawa, Canada. For more information, visit www.educomts.com.