

# nuix Email Archive Migration Manager

## User Guide

Version 1.2

December 2017

# DISCLAIMER

© 2017 Nuix. All rights reserved.

This publication is intended for informational purposes only. The information contained herein is provided "as-is" and is subject to change without notice. Although reasonable care has been taken to ensure that the facts stated in this publication are accurate and that the opinions expressed are fair and reasonable, no representation or warranty, express or implied, is made as to the fairness, accuracy or completeness of the information or opinions contained herein, and no reliance should be placed on such information or opinions. Neither Nuix nor any of its respective members, directors, officers or employees nor any other person accepts any liability whatsoever for any loss arising from any use of such information or opinions or otherwise arising in connection with this publication. Furthermore, this publication contains the confidential and/or proprietary information of Nuix which may not be reproduced, redistributed, or published in any form or by any means, in whole or in part, without the express prior written consent of Nuix. The use, reproduction, and/or distribution of any Nuix software described in this publication requires an applicable software license.

## Revision History:

The following changes have been made to this document

Version Number	Revision Date	Author	Description
1.2.3	December 2017	Alex Chatzistamatis	Initial release

# Content

<b>INTRODUCTION .....</b>	<b>6</b>
About Email Archive Migration Manager.....	6
About this Guide .....	6
Document Conventions.....	6
<b>SYSTEM REQUIREMENTS.....</b>	<b>7</b>
Architecture.....	7
Hardware .....	8
CPU.....	8
Memory (RAM) .....	8
Storage.....	8
Virtualization.....	10
Example System Configurations .....	11
Software.....	12
<b>INSTALLATION .....</b>	<b>13</b>
Files .....	13
Setup .....	14
<b>GETTING STARTED .....</b>	<b>15</b>
Main Menu.....	15
Interface Overview .....	15
Global Settings .....	16
Interface Overview .....	16
Nux License.....	17
Nux Directories .....	18
Lightspeed Settings.....	19
Exchange Web Services .....	20
Database Settings .....	22
Extracting Email Data from Legacy Email Archives .....	24
Interface Overview .....	24
Working with Veritas Enterprise Vault .....	25
Working with EMC EmailXtender/SourceOne .....	42
Working with HP/Autonomy Zantaz EAS.....	55
Working with Daegis AXS-One.....	68

Converting Legacy Email Data .....	75
Interface Overview .....	75
Performing a User NSF to User PST Conversion.....	76
Ingesting Email Data into Exchange .....	77
Interface Overview .....	77
Working with Exchange Web Services .....	78
Extracting Email Data from Exchange .....	83
Interface Overview .....	83
Working with Exchange Web Services .....	84
<b>APPENDIX I: BACKUP/RESTORE SQL DATABASES.....</b>	<b>90</b>
SQL Backup Workflow .....	90
Accessing SQL Server Management .....	90
Targeting Desired Archive Database(s) .....	90
Configuring Desired Archive Database(s) .....	91
Assigning Database Location and Confirming Drive Space ....	92
Correctly naming the Database and choosing the proper File Extension.....	93
SQL Restore Workflow .....	93
Accessing SQL Server Management .....	93
Restoring Desired Archive Database(s) .....	94
Configuring Desired Archive Database(s) .....	94
<b>APPENDIX II: ARCHIVE SQL QUERIES .....</b>	<b>96</b>
Veritas Enterprise Vault .....	96
EMC EmailXtender .....	97
EMC SourceOne.....	98
HP/Autonomy Zantaz EAS.....	98
<b>APPENDIX III: ARCHIVE METADATA .....</b>	<b>100</b>
Veritas Enterprise Vault .....	100
EMC EmailXtender/Source .....	100
HP/Autonomy Zantaz EAS.....	100
Daegis AXS-One .....	101
<b>APPENDIX IV: EWS BEST PRACTICES.....</b>	<b>102</b>
Leveraging Azure Virtual Machines .....	102
Reduced Number of Nuix Workers .....	102
EWS Throttling Workarounds .....	102

**ABOUT NUIX ..... 104**

# Introduction

Welcome to the Nuix Email Archive Migration Manager User Guide.

## About Email Archive Migration Manager

Nuix Email Archive Migration Manager (NEAMM) provides functionality to efficiently manage legacy email archive migration projects including:

- Extracting email-based content from legacy email archive platforms such as Veritas Enterprise Vault, EMC EmailXtender/SourceOne, HP/Autonomy Zantaz EAS, and Daegis AXS-One.
- Converting legacy NSF data to modern email formats such as PST, MSG or EML.
- Extracting data from Exchange on-premise or Exchange Office 365 mailboxes/personal archives.
- Ingesting data to Exchange on-premise or Exchange Office 365 mailboxes/personal archives.
- Real-time statistics and progress of the migration project.

Once setup, NEAMM can be configured to perform these actions on a single system. NEAMM can be installed and managed on multiple systems to expedite the progress of the migration project.

## About this Guide

This guide provides step by step instructions to help you configure and use NEAMM.

## Document Conventions

The following conventions are used in this guide:

```
//This is a line of code
```

This is a Menu > Option.

### Note

This is a note.

### Tip

Tips, written in highlighted text boxes, are useful pieces of information on how to apply what is in the guide into practice or provide an example.

### Warning

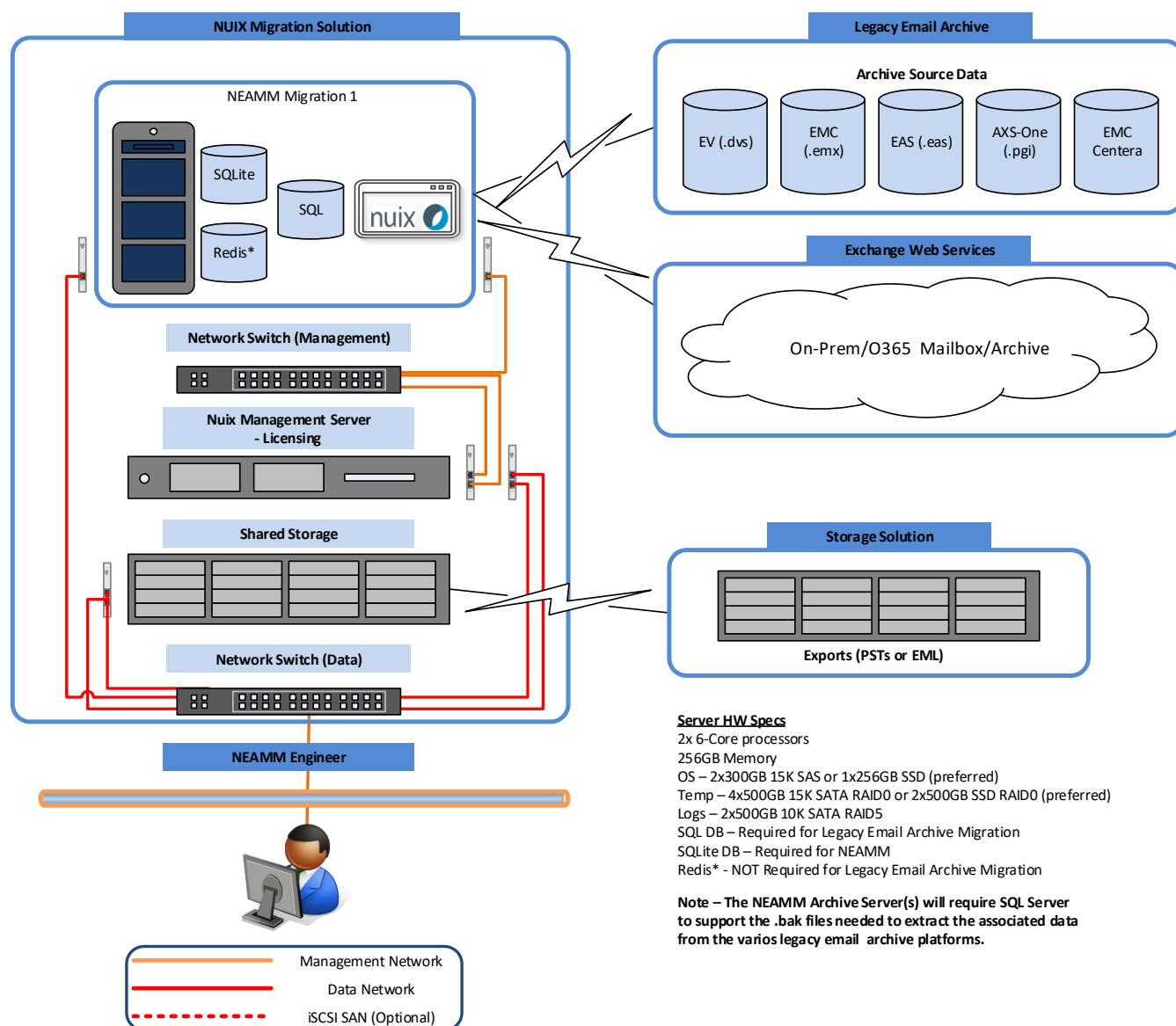
This symbol is used to indicate information that is critical, which must be reviewed.

# System Requirements

NEAMM has been designed to be as lightweight and portable as possible while working on the most common Windows Operating Systems currently available. For more details, please see the detailed sections below.

## Architecture

The architecture of NEAMM installed in an on-premise environment may look like this:



# Hardware

Please refer to the section below for sizing the hardware required for NEAMM. Several factors must be taken into consideration for appropriate sizing including, but not limited to: source data type, source data volume, expected processing throughput and the project timeline. Please speak to your Nuix Account Manager and Nuix Solutions Consultant for more details.

This sections outlines the system configuration required to extract optimal performance from Nuix systems. This includes specifications for:

- CPU
- Memory (RAM)
- Storage
- Latency
- Virtualization

## CPU

The number of physical CPU cores available should always be equal to or greater than the number of Nuix workers. When comparing CPUs faster clock speeds should be preferred over additional cores.

The current processor architecture of Advanced Micro Devices (AMD) is based on multiple, low-efficiency processor cores working in tandem. Nuix Workstation is optimized and licensed around fewer, faster processor cores like those found in Intel processors. In our most recent benchmarks, Intel Xeon series processors provide more than double the core per worker performance of comparable AMD Opteron series processors. Because of this, Intel CPU's provide better value and performance with Nuix Workstation. We will continue to test and evaluate both AMD and Intel's offerings as they become available.

## Memory (RAM)

NEAMM runs a number of independent processes called Workers. Each worker runs as a separate system process to perform the required task with the main application acting as a broker to distribute the items being actioned.

The number of workers which available to you is determined by your project.

When determining the amount of RAM which is needed for your system:

- A minimum of 8 GB of RAM for each Nuix worker is acceptable, however, Nuix recommends allocating 16GB of RAM for each Nuix worker for optimal performance.
- $1\text{GB of RAM} + (1\text{GB} * \text{Number of Workers})$  should be available to the main application.
- A minimum of 4GB should be left unallocated to be used by the Operating System.

## Storage

NEAMM processing is very I/O intensive with a number of processes occurring simultaneously which utilize your storage. At a high level these are:



- **Worker Temporary Directory** – The copies of the files to be processed are saved to the worker temp directory to process. This ensures the original data is preserved and not altered during processing.
- **Source Data** – The original source location from which the data is unpacked into the temporary directory ready for processing.
- **Export Location** – The location used to export data from the case.

The type of drives and their configuration will play a major role in the overall processing and export performance.

## Storage Types

The cost of storage increases as the performance of the drive increases. The following table provides examples of the I/O Operations per second (IOPS) each type is capable of.

Drive Type	IOPS
7.2k SATA	100
10k SAS	150
15k SAS	200
Desktop/Laptop MLC SSD (cheaper)	2,000–10,000
Server MLC SSD	10,000–50,000
Write Intensive SLC SSD (expensive)	100,000
PCIe IO Card	1,000,000+

Solid State Drives (SSD) provide excellent performance especially when placed into a Redundant Array of Inexpensive Discs (RAID) configuration, however, using SSD drives for all storage needs may not be necessary. It is recommended to always balance CPU, RAM and Storage requirements.

## Storage Optimization

Maximizing the potential of a suitably powerful system requires a well configured storage solution. Fast disk speeds and dedicated logical units for different Nuix locations are vital to achieving the best possible performance. Storage size will depend heavily on project scope and the volume of data.

To achieve optimal performance from your setup, be sure to keep separate locations (LUNs) for the following storage areas:

### Worker Temp Directory

- Very fast disk (SSD if possible or RAID 0 array) with Low Latency.
- Minimum capacity: largest single job size x 5 (i.e. 500 gigs to process 100 gigs).
- No redundancy as the required data is only accessed at the time of processing or export and then deleted.

### Source Data

- Heavy Read activity during processing or export, Light Read activity during a review.
- RAID 5/6 provides high Read performance and maximum storage capacity with moderate redundancy.

## Export Directory

- If you plan to move data to another location (off site, Review system and so on) after Export, you can maximize export performance by using a RAID 0 storage configuration (similar to Worker Temporary Directory).
- If the Export location holds data for an extended period a redundant storage configuration with good Write performance is recommended.

Finally, if you are looking to configure your storage for optimal Nuix performance, you can also look at array stripe size. A smaller stripe size improves performance for reading or writing smaller files. Conversely, if you deal with primarily larger files, a larger stripe size can also significantly affect performance for both reading and writing.

## Example Configuration

The following table shows an example configuration which provides a compromise between cost and performance and would be suitable for high levels of processing.

Usage	Disc Type	Configuration
Worker Temporary Directory	Local SSD	RAID 0
Source Data	SAN with Fiber Connection or 10 Gigabit Ethernet	-
Export	Local SSD	RAID 0

## Latency

Latency represents a delay in sending or receiving data between two devices. In the case of a Nuix system and a storage volume it is seeking to Read or Write to, latency can significantly impact performance. Network connections to storage, especially Windows File Share connections (NTFS/SMB) have inherently higher latency than directly attached devices (DAS).

To minimize latency, Nuix recommends:

- Configuring less workers with more RAM per worker when there is high latency with large/spanned/compressed file types. This helps minimize the amount of disk traffic involved.
- Reviewing the array stripe size. A smaller stripe size improves performance in reading or writing smaller files. Conversely, for primarily larger files, a larger stripe size will significantly affect performance for both reading and writing.

Nuix system utilizes all the available RAM, CPU, and I/O resources when processing data, up to the limits of your license. Beyond that, large amounts of RAM provide you an excellent environment for a multi-user concurrent review system.

As above, optimizing disk configuration is often an overlooked element to maximize processing performance.

## Virtualization

To install Nuix in a Virtual Machine (VM) environment, ensure the VM host has better hardware than the proposed Nuix VM specification to account for the 5-10% virtualization performance degradation.

## Processor Resource Pool

Directly mapping the VM's Virtual Cores to the Host's Physical Cores is preferred, however, it is recommended to maintain 2 - 4 Physical Cores without a Virtual Core Mapping in order for the VM Host OS and VM Host Software to function properly. It is strongly recommended to isolate the Nuix VM to its own resource pool or set its Share value to "High" if the VM Host is running multiple VM Guests.

## Virtual Memory Resource Pool

For VMWare/VMsphere, Nuix recommends that:

- The VM Host should have at least 25% more RAM (not page/disk swap) than the RAM allocated to the Nuix VM.
- The VM Host should be capable of hardware assisted memory virtualization.
- The Nuix VM should be configured with a memory Share value of "High" to maximize its access to VM Host memory resources.
- The Nuix VM should be configured with a Reservation level of at least 50% its total available memory allocation.

## Virtual Hard Disk Drive Configuration

The storage input/output (I/O) must be paired with a Virtual Hard Disk Drive (VHD) capable of supporting a minimum average sustained Input/output Operations Per Second (IOPS).

### Note

If the Logical Unit Number (LUN) hosting the VHD for the NUIX VM is shared with other VM VHD's, the Nuix VM Share Level should be set at high or equivalent. This helps the Nuix VM to receive higher priority on reading and writing to the LUN than other VM's accessing VHD's on that LUN.

## Example System Configurations

The following tables outline sample configurations which provide a generic starting point for designing Nuix processing environment.

### Note

Performance will vary based on hardware configuration, source data type, source data volume and Case settings selected at the time processing.

## Physical Servers

Nuix License	CPU Cores	Minimum / Recommended RAM	Minimum non-OS Storage
2 Workers	2 i7/Xeon	16 GB / 32 GB	300 READ & 300 WRITE IOPS DAS
4 Workers	4 i7/Xeon	32 GB / 64 GB	600 READ & 600 WRITE IOPS DAS
8 Workers	8+ Xeon	64 GB / 128 GB	1200 READ & 1200 WRITE IOPS DAS
12 Workers	12+ Xeon	96 GB / 192 GB	1800 READ & 1800 WRITE IOPS DAS
16 Workers	16+ Xeon	128 GB / 256 GB	2400 READ & 2400 WRITE IOPS DAS

## Virtual Servers

Nuix License	VM Host CPU Cores	VM Host Minimum / Recommended RAM	VM Guest Minimum / Recommended RAM	Minimum non-OS Storage
2 Workers	4 i7/Xeon	20 GB / 40 GB	16 GB / 32 GB	300 READ & 300 WRITE IOPS DAS
4 Workers	8 i7/Xeon	40 GB / 80 GB	32 GB / 64 GB	600 READ & 600 WRITE IOPS DAS
8 Workers	12+ Xeon	80 GB / 160 GB	64 GB / 128 GB	1200 READ & 1200 WRITE IOPS DAS
12 Workers	16+ Xeon	120 GB / 240 GB	96 GB / 192 GB	1800 READ & 1800 WRITE IOPS DAS
16 Workers	20+ Xeon	160 GB / 320 GB	128 GB / 256 GB	2400 READ & 2400 WRITE IOPS DAS

## Software

NEAMM requires that following pre-requisites are met:

- **Operating System (OS) :**
  - o Server: Windows Server 2008 (minimum), Windows Server 2012/2016 (recommended)
  - o Desktop: Windows 7 (minimum), Windows 10 (recommended)
- **Nuix:**
  - o Nuix Management Server: 7.0 and above
  - o Nuix Workstation: 7.0 and above
- **Other:**
  - o Microsoft .NET Framework 3.5+
  - o Microsoft SQL Server 2008/2012/2016 (for Legacy Archives)
  - o Redis: 3.2
    - Redis is not required for all deployments. Please consult with your Nuix representative to understand if this is necessary for your project.

# Installation

## Files

NEAMM requires installation and consists of the following files/folders:

1. Application Files
2. NEAMM.application
3. Setup.exe

These files should be copied to a folder within a directory.

Name	Date modified	Type	Size
1 Application Files	8/21/2017 5:14 PM	File folder	
2 NEAMM	8/21/2017 5:14 PM	Application Manif...	6 KB
3 setup	8/21/2017 5:14 PM	Application	416 KB

### Tip

NEAMM can be installed on any system which can:

- obtain a local Desktop license
- obtain a Server license using **Nuix Management Server**
- has **Nuix Workstation** installed
- has the necessary access to the email archive source data/databases
- has access to the Exchange on-premise/Office 365 mailbox/personal archives.

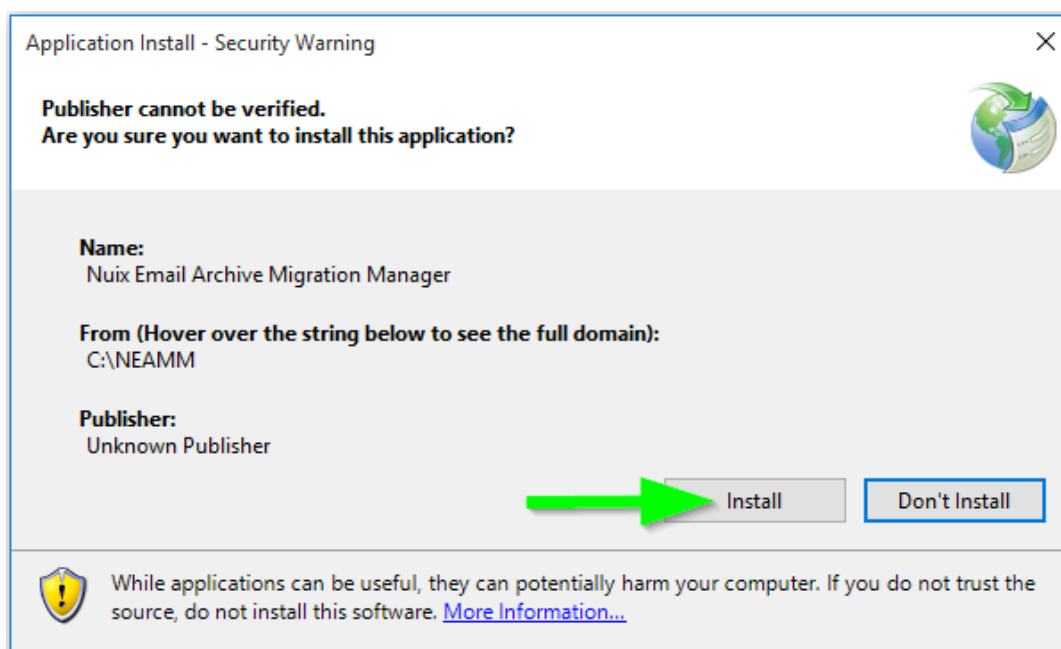
# Setup

In order to install NEAMM, please perform the following steps:

1. Launch "setup.exe"

Name	Date modified	Type	Size
Application Files	8/21/2017 5:14 PM	File folder	
NEAMM	8/21/2017 5:14 PM	Application Manif...	6 KB
setup	8/21/2017 5:14 PM	Application	416 KB

2. Click Install



3. When installation completes, you will see the NEAMM Main Menu.

## Note

Before performing any migration work, you must configure the embedded SQLite database. Please follow review the **NEAMM Global Settings** section to perform this mandatory step.

# Getting Started

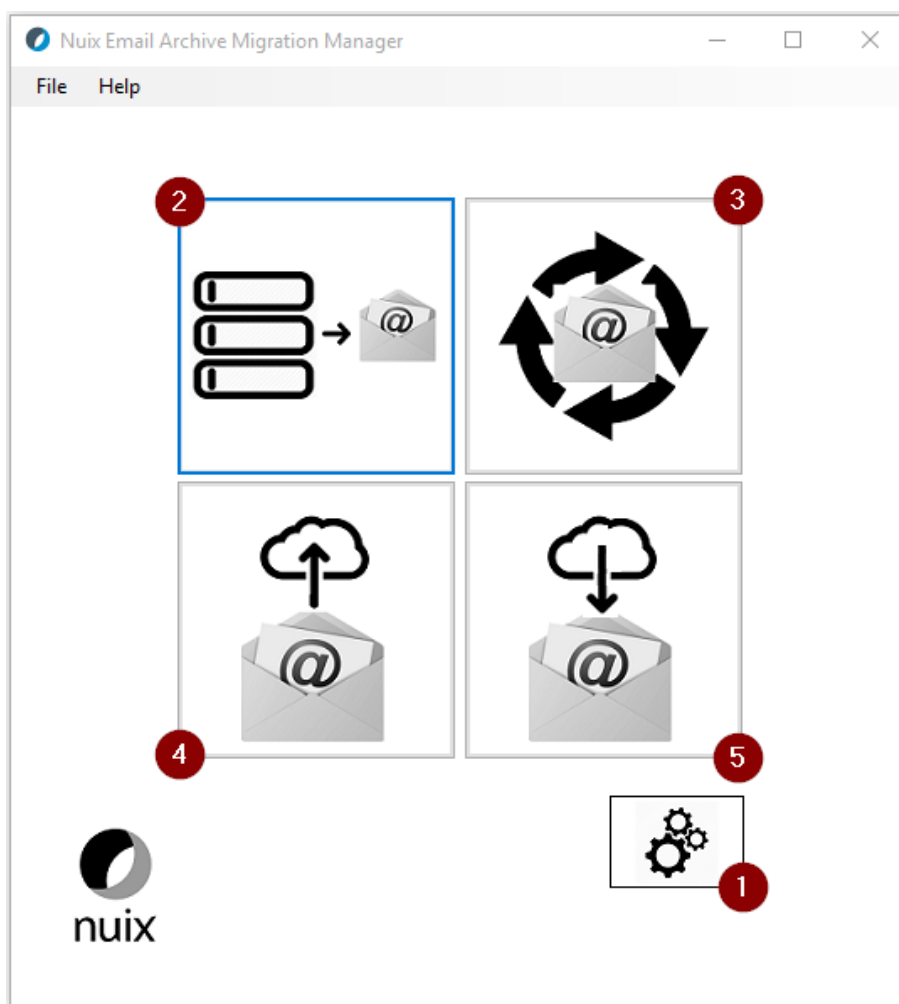
Before you start using Nuix, we will go through the primary components of the user interface, the menu systems and all other major components.

Once you are familiar with the interface and layout, it will be easy for you to complete migration tasks.

## Main Menu

### Interface Overview

Upon launching NEAMM, you will see the main menu which is broken down into several components. Aside from the Global Settings, each of large buttons has a primary function, with a specific workflow associated with it.



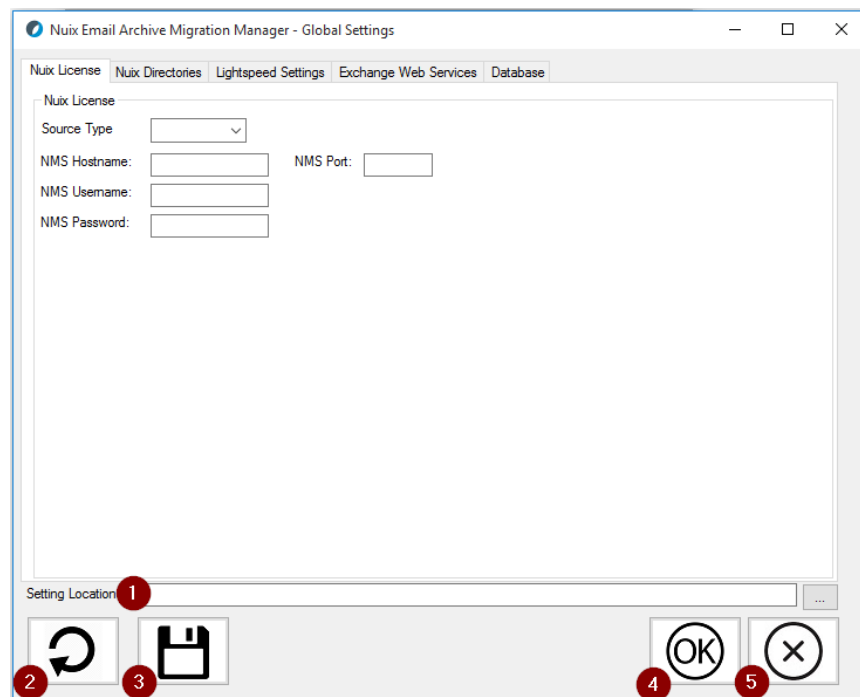
	Name	Description
1	Global Settings	The component which controls settings globally for each of the different components that make up NEAMM.
2	Email Archive Extraction	Allows you to extract email data to various formats from the most common legacy email archive systems.
3	Email Conversion	Allows you to convert email data from one format to another.
4	EWS Ingestion	Allows you to ingest email data into Exchange an Exchange mailbox or archive.
5	EWS Extraction	Allows you to extract email data from an Exchange mailbox or archive.

# Global Settings

## Interface Overview

Global Settings allows you to configure and control many different aspects required for a migration ranging from Nuix licensing to directories, processing, extraction and more. These settings must be configured every time NEAMM is launched and should be checked prior to starting any migration work.

NEAMM Global Settings are saved a standard XML file which can be reloaded into NEAMM whenever it is launched.





	Name	Description
1	Settings Location	Enter a location to save your re-usable Settings XML. Use the ellipsis to browse the file system.
2	Reload Settings	Reload an existing Settings XML file.
3	Save Settings	Save a Settings XML file that can be reused.
4	OK	Accept your Global Settings.
5	Cancel	Cancel and close Global Settings.

## Nuix License

The Nuix License tab is used to allow NEAMM to obtain a license from an existing Nuix Management Server deployment.

The screenshot shows the 'Nuix Email Archive Migration Manager - Global Settings' window with the 'Nuix License' tab selected. The tab contains the following fields:

- Source Type:** A dropdown menu with 'Server' selected.
- NMS Hostname:** A text box containing '127.0.0.1'.
- NMS Port:** A text box containing '27443'.
- NMS Username:** A text box containing 'nuixadmin'.
- NMS Password:** A text box with masked characters (asterisks).

Name	Description
Source Type	Desktop or Server license
NMS Hostname	The IP address or FQDN of your NMS instance
NMS Port	The port for your NMS instance (default: 27443)
NMS Username	Username for your NMS instance
NMS Password	Password for your NMS instance

### Warning

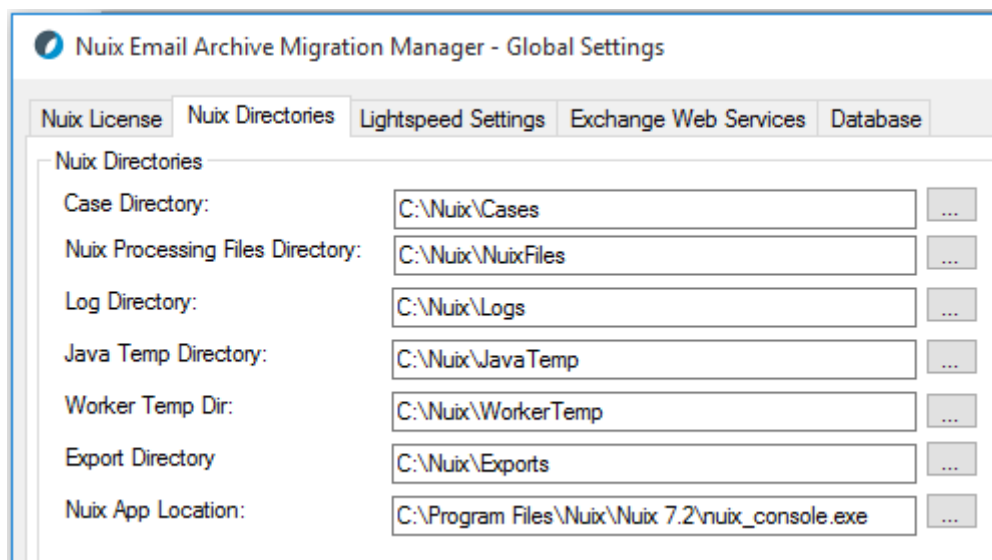
You must enter the IP Address and Port Number used by the previously configured Nuix Management Server (NMS).

### Tip

Create an account to use with the toolkit only as this will ensure that you can easily determine which Audit Events were taken by the toolkit and which by other users.

## Nuix Directories

The Nuix Directories settings tab is used to configure all of the different components needed for any extractions or ingestions, ranging from licensing, directories, processing, extraction and more. These settings must be configured before starting any migration work.



**Nuix Email Archive Migration Manager - Global Settings**

**Nuix Directories**

Case Directory: C:\Nuix\Cases

Nuix Processing Files Directory: C:\Nuix\NuixFiles

Log Directory: C:\Nuix\Logs

Java Temp Directory: C:\Nuix\JavaTemp

Worker Temp Dir: C:\Nuix\WorkerTemp

Export Directory: C:\Nuix\Exports

Nuix App Location: C:\Program Files\Nuix\Nuix 7.2\nuix\_console.exe

Name	Description
Case Directory	the location where Nuix Cases will be stored. The Nuix Case will contain critical information for each batch, including Summary Reports and Success/Exception Reports.
Nuix Processing Files Directory	the location where NEAMM-related files will be stored. The Nuix Files are related to each batch that is run, including: BAT files, Ruby scripts, JSON files, etc.
Log Directory	the location where logs will be stored. The Nuix Logs can be used to troubleshoot errors or other issues. This directory can be purged on a routine basis if necessary.
Java Temp Directory	the location where Java Temp will be stored. This is a temporary location that will be used/cleared during each processing job.
Worker Temp Directory	the location where Worker Temp will be stored. This is a temporary location that will be used/cleared during each processing job.
Export Directory	the location where exports will be stored. This should be treated as a critical location and where extracted data will exist, as well as critical reports for each export job.
Nuix App Location	the location where Nuix App is installed. The location of your Nuix installation is something that should have been completed prior to reaching this step.

**Tip**

It is very important to ensure that each Nuix Directory location has the proper storage configuration in place. Please review the System Requirements section for more details.

## Lightspeed Settings

The Nuix Lightspeed Settings tab is used to configure the necessary Nuix instance settings when performing an extraction from a legacy email archive platform.

**Nuix Email Archive Migration Manager - Global Settings**

**Lightspeed Settings**

System Memory (RAM): 11999.55

Number of Nuix Instances: 1

Nuix App Memory: 1000

Available Memory (RAM): 9996.41

Available After Nuix Instances: 00000

Number Of Nuix Workers: 2

Memory Per Worker (MB): 1000

Worker Timeout: 3600

**MAPI Export Options**

PST Export Size (GB): 2

☒ Add Distribution List Recipients to MAPI-EXPANDED-DL field

**EML Export Options**

☐ Add Distribution List Recipients to Individual EXPANDED-DL field

**Lightspeed Settings:**

Name	Description
System Memory (RAM)	displays the total amount of RAM available on the system
Number of Nuix Instances	controls the total number of concurrent Nuix instances that is able to run on the system
Nuix App Memory	controls the amount of maximum memory each Nuix instance will utilize
Available Memory (RAM)	displays the total amount of memory left before factoring in memory for each Nuix instance
Available After Nuix Instances	displays the total amount of memory left after factoring in memory for each Nuix instance
Number of Nuix Workers	controls the total number of Workers per Nuix instance
Memory Per Worker (MB)	controls the amount of maximum memory that each Nuix Worker will utilize
Worker Timeout	controls the amount of time in seconds that a worker will attempt to process an item before it times out, flags the item as poisoned, and moves on.

#### MAPI Export Options:

Name	Description
PST Export Size (GB)	controls the size of each PST that Nuix exports
Add Distribution List Metadata	includes any Distribution List recipients from the Exchange Journal envelope in the "mapi-expanded-dl" metadata of each MSG.

#### EML Export Options:

Name	Description
Add Distribution List Metadata	includes any Distribution List recipients from the Exchange Journal envelope in the "Expanded-DL" Delivered-To header of each RFC822

#### Note

You must properly balance RAM across the OS, any other running applications (such as SQL), the Nuix application and the Nuix Workers. The majority of your memory should generally be allocated to your workers, as these will be performing the most intensive work. Insufficient memory for the Workers will cause inconsistencies, item poisoning or other memory related errors.

## Exchange Web Services

The Nuix EWS settings tab is used to configure the necessary Nuix instance settings when performing ingestions or extractions with Exchange, either on-premise or Office 365.

The screenshot shows the 'Nuix Email Archive Migration Manager - Global Settings' window with the 'Exchange Web Services' tab selected. The window is divided into several sections:

- EWS Connection:** Includes fields for Exchange Server (https://outlook.office365.com/ews/exchange.asmx), Domain (nuixpa.onmicrosoft.com), User Name (nuixappimp@NuixPA.onmicrosoft.com), and Password (masked). There is a checkbox for 'Enable Impersonation' which is checked.
- Lightspeed Settings:** Includes fields for System Memory (RAM) (11999.55), Number of Nuix Instances (2), Nuix App Memory (1000), Available Memory (RAM) (9996.41), Available After Nuix Instances (00000), Number Of Nuix Workers (1), Memory Per Worker (MB) (1000), and Worker Timeout (3600).
- EWS Upload Control:** Includes a field for Maximum Message Size (MB) (25), a checkbox for 'Enable Bulk Upload' (unchecked), a field for Bulk Upload Size (MB) (10), and a field for 'Remove' (0) levels from folder path.
- EWS Download Control:** Includes a checkbox for 'Enable Bulk Download' (checked), a field for Maximum Message Size (MB) (25), a field for Maximum Download Count (100), a checkbox for 'Enable Collaborative Fetching' (checked), and a checkbox for 'Enable Mailbox Slack Space' (unchecked).
- EWS Throttle Control:** Includes fields for Retry Count (5), Retry Delay (5), and Retry Increment (5).

**EWS Connection:**

Name	Description
Exchange Server	the complete URL to the Exchange server
Domain	the domain for the Exchange environment
Username	the account used to connect to Exchange (this can either be a user account, an account with delegate access or an account with the Application Impersonation role assigned to it)
Password	the password for the account connecting to Exchange
Enable Impersonation	designates whether the authenticating account has the Application Impersonation role assigned it

**Tip**

It is recommended that an account with the Application Impersonation role assigned to it is used when working with multiple Exchange mailboxes/personal archives in parallel.

**Lightspeed Settings:**

Name	Description
System Memory (RAM)	displays the total amount of RAM available on the system
Number of Nuix Instances	controls the total number of concurrent Nuix instances that is able to run on the system
Nuix App Memory	controls the amount of maximum memory each Nuix instance will utilize
Available Memory (RAM)	displays the total amount of memory left before factoring in memory for each Nuix instance
Available After Nuix Instances	displays the total amount of memory left after factoring in memory for each Nuix instance
Number of Nuix Workers	controls the total number of Workers per Nuix instance
Memory Per Worker (MB)	controls the amount of maximum memory that each Nuix Worker will utilize
Worker Timeout	controls the amount of time in seconds that a worker will attempt to process an item before it times out, flags the item as poisoned, and moves on.

**EWS Upload Control:**

Name	Description
Maximum Message Size (MB)	controls the maximum individual message size of each top-level item being ingested to Exchange
Enable Bulk Upload	enables bulk uploads instead of single item uploads
Bulk Upload Size (MB)	controls the maximum size of messages being ingested to Exchange in bulk
Remove [ ] levels from the folder path	trims the folder path of the PST data that may be ingested into an Exchange mailbox/personal archive

#### EWS Download Control:

Name	Description
Enable Bulk Download	enables bulk downloads instead of single item downloads
Maximum Message Size (MB)	controls the maximum message sizes of bulk downloads (works together with Maximum Download Count)
Maximum Download Count	controls the maximum number of messages to download in bulk (works together with Maximum Message Size)
Enable Collaborative Fetching	used to download messages faster from mailboxes with folders less than 1 GB in size
Enable Mailbox Slack Space	used to extract deleted items from EWS mailbox or archive slack space - may be used when looking to capture all aspects of a mailbox or archive.

#### EWS Throttle Control:

Name	Description
Retry Count	controls the number of items an item is retried after it is first flagged for being throttled
Retry Delay	controls the amount of time (in seconds) to wait before attempting to retry the item the first time
Retry Increment	controls the amount of time (in seconds) to wait before attempting to retry the item for each subsequent retry

#### Tip

It is recommended that the throttling controls are set to a minimum of: 5 retries, 5 second delay /5 second subsequent delay. For Office 365 migrations, it may be necessary to set the number of retries much higher.

## Database Settings

The Database tab is used to configure different aspects of the external databases that are integrated with NEAMM.

The screenshot shows the 'Nuix Email Archive Migration Manager - Global Settings' window with the 'Database' tab selected. The 'SQLite Database Location' is set to 'C:\Nuix\Database'. The 'Redis Configuration' section includes fields for 'Redis Host Name' (localhost), 'Port' (6666), and 'Auth' (password).

Nuix Email Archive Migration Manager - Global Settings	
Nuix License   Nuix Directories   Lightspeed Settings   Exchange Web Services   Database	
SQLite Database Location: C:\Nuix\Database	
Redis Configuration	
Redis Host Name:	localhost
Port:	6666
Auth:	password

Name	Description
SQLite Database Location	a path to a folder where the NEAMM SQLite database will be stored. This database will maintain all historic job activity within NEAMM.
Redis Host Name	used to specify the hostname of the system running Redis
Port	used to specify the port Redis is listening on
Auth	used to specify the password for the Redis instance

#### Note

This SQLite database can be copied for Backup/DR purposes. It is recommended to back this database up when there is no Nuix activity.

#### Note

Redis is not installed or configured by NEAMM. This must be setup prior to using NEAMM. Please consult with your Nuix representative to understand if this is necessary for your project.

## Creating the NEAMM Database

NEAMM uses an embedded SQLite database to store any job-related activity. In order to configure this database upon installation, perform the following:

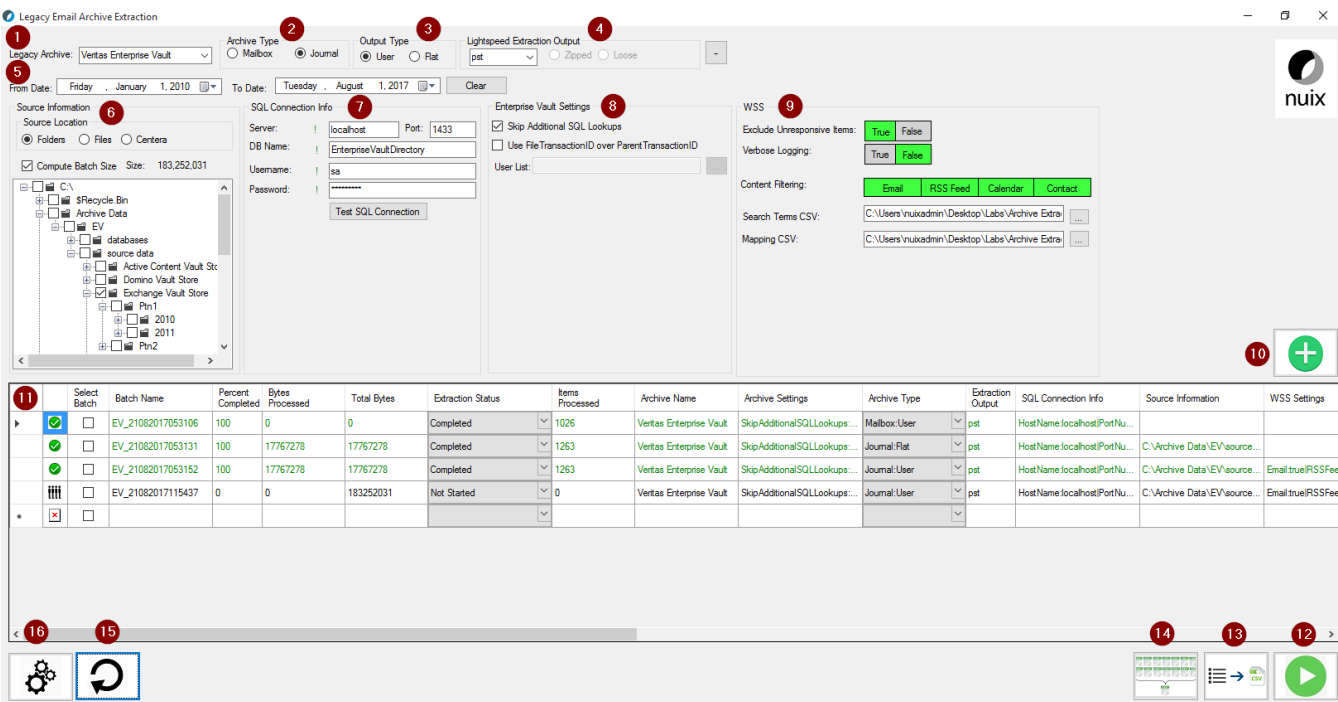
From **Global Settings**, click on the **Database** tab:

1. Enter a location for the SQLite Database location.
2. Enter a location for the Settings XML settings file.
3. Click **Save**.
4. Upon completion of steps 1-3, a pre-configured SQLite database will be created and ready for any migration work.

# Extracting Email Data from Legacy Email Archives

## Interface Overview

Upon launching the Email Archive Extraction module, you will see the interface as shown below. Many of these options in this interface will be enabled/disabled based on selections made. This is by design several may options may not be relevant or necessary for specific source archives or workflows.



	Name	Description
1	Legacy Archive	Select the source email archive that data will be extracted from.
2	Archive Type	Select whether the email archive is Mailbox or Journal based ( <i>source archive dependent</i> ).
3	Output Type	Select whether you want to perform a flat extraction or user-based extraction.
4	Lightspeed Extraction Output	Select whether the format of the extracted data (PST, EML or MSG).
5	From/To Date	Used to filter the email data by date.
6	Source Information	Select if the source data is located on a folder, a physical file itself, or the data is stored on an EMC Centera storage device.
7	SQL Connection Info	Connect to a backup copy of the Email Archive SQL



		databases ( <i>source archive dependent</i> ).
8	Archive Specific Settings	Additional settings for each supported legacy email archive.
9	WSS (Worker Side Script)	Custom filter settings used for Journal Archives that require export User output, or used with AXS-One for User output ( <i>source archive dependent</i> ).
10	Add Job to Grid	Add the pre-configured job to the grid.
11	Grid	Where added jobs will be displayed.
12	Start Job	Start the selected job in the grid for processing.
13	Export Grid to CSV	Exports out the current grid view to CSV format.
14	Lightspeed Exporter Report Consolidator	Will consolidate all Lightspeed Exporter Metrics and Exporter Error into a single report.
15	Reload Grid	Reloads jobs in the grid from previous migrations.
16	Global Settings	View/Change previously configured Global Settings.

### Tip

Before attempting to perform any migration work, be sure to check **Global Settings** and make sure that the **Nuix Directories** and **Archive Extraction** tab are configured correctly.

## Working with Veritas Enterprise Vault

When working with Veritas Enterprise Vault source data, you can use NEAMM in various methods to target the data directly on the file system in its proprietary, .DVS file format (or on Centera when applicable) and export it out to disk in PST, MSG or EML format. Below are several workflows that NEAMM can handle.

### Overview

Enterprise Vault archiving solutions have two components:

- Source Data on Disk (.DVS)
- SQL Database (EnterpriseVaultDirectory + Mailbox and/or Journal vault store databases)

Nuix must have access to the SQL database while processing EV files on disk in order to gather additional metadata and, most importantly, in order to reconstitute emails and their single-instanced parts.

### Note

SharePoint, File System and Instant Message vault stores are not currently supported for legacy archive migrations.

### Prerequisites

The following prerequisites must be met before Nuix can process Enterprise Vault

data:

1. EV database restoration<sup>1</sup>
  - A copy of the production EV database(s) must be restored to a local MSSQL instance on the Nuix server.
  - These include EnterpriseVaultDirectory and all of the mailbox and/or journal vault store databases.
2. EV database configuration
  - The dbo.PartitionEntry table must be modified to reflect the correct Vault Store Partition file system path.
3. Source data identified with a specific EV Vault Store<sup>2</sup>

## Source Data

### Physical Files

Enterprise Vault can be considered one of the most complex email archive solutions in terms of data handling—it can be heavily compressed, especially when collections are enabled and reconstituting the single-instanced parts adds additional overhead. Enabling sharing at the Vault Store Group level adds further complexity.

Vault Stores are created in EV Admin Console and can be compared to a physical volume.

Vault Store partitions are created in EV Admin Console and are containers within the Vault Store.

Enterprise Vault Stores each archived item and its associated SIS parts to the currently open vault store partition for the archive into which the item is being archived. There are three main types of files used in the archive:

- **\*.DVS** - This file contains all un-sharable (per user) properties relating to the archived item.
- **\*.DVSSP** - This file (or files) contains an individual SIS part
- **\*. DVSCC** - This file contains the HTML content conversion of the SIS part

Large files, in Enterprise Vault, are items being archived that are larger than 50 MB in size and these are processed differently from smaller files. To begin with, the archived item will not be compressed (regardless of the compression setting) and it will not be eligible for collection. If Enterprise Vault is configured to migrate data, then the SIS parts and content collections are migrated directly rather than being placed into a CAB file first. Large files will also have a content conversion copy (DVSCC) created but it will only be indexed if the copy is less than 30 MB in size. File extensions for large files are also slightly different:

- If sharing is not enabled, large files are stored as **\*.DVF** files.
- If sharing is enabled, large files are stored as **\*.DVFSP** files.

---

<sup>1</sup> EV data can be processed without a database connection; however, the database is required for expansion of additional metadata and reconstituting single-instanced parts.

<sup>2</sup> Many large EV archive deployments include multiple Vault Stores and archive databases, each of which will have a separate database.

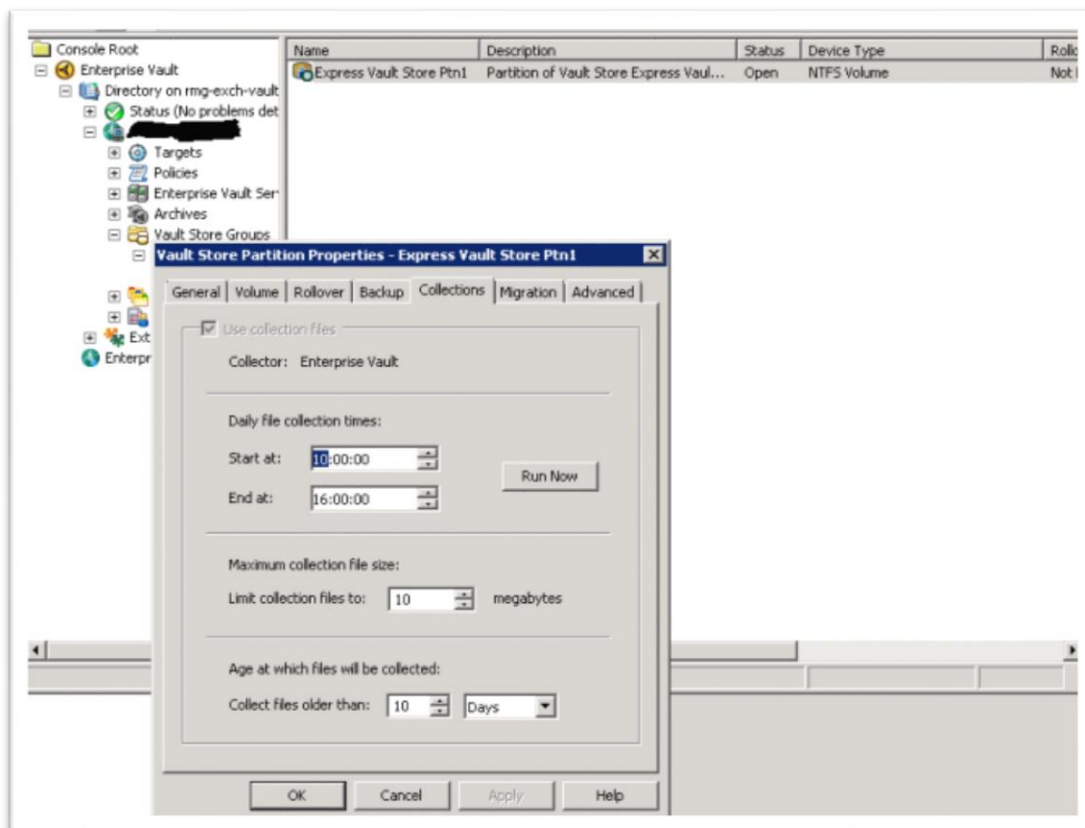
- Large file converted content files are stored as \*.DVFCC files.

Name	Date modified	Type	Size
200A01AFB25C883DD091653FD71CE6E1.DVS	1/6/2014 10:02 PM	DVS File	10 KB
200A01AFB25C883DD091653FD71CE6E1~AE~977E40FE~00~1.DVSCC	1/6/2014 10:02 PM	DVSCC File	12 KB
200A01AFB25C883DD091653FD71CE6E1~AE~977E40FE~00~1.DVSSP	1/6/2014 10:02 PM	DVSSP File	69 KB
200A1A39CF22001792FF611F6FD37D01.DVS	1/6/2014 10:02 PM	DVS File	18 KB

## Collections (.CAB Files)

In many environments, collections can be enabled in a vault store partition. This will cause DVS and Vault Stream files (.DVSSP/.DVSCC) to be containerized in .CAB (cabinet) files. Collections are typically enabled in order to further reduce the overall size of the archive. While the intention is good, this complicates migrations downstream. While Nuix can process these .CAB files, it can reduce processing throughput, since these are compressed files. Additionally, these .CAB files present challenges if needing to perform mailbox-level extractions.

- .CAB files can reduce processing throughput due to their heavily compressed nature.
- If performing a journal archive extraction with Lightspeed, the .CAB is only processed once.
- If performing a mailbox archive extraction with Lightspeed, data will be processed from Enterprise Vault by user, thus it will only extract the specific data that is located within each .CAB.

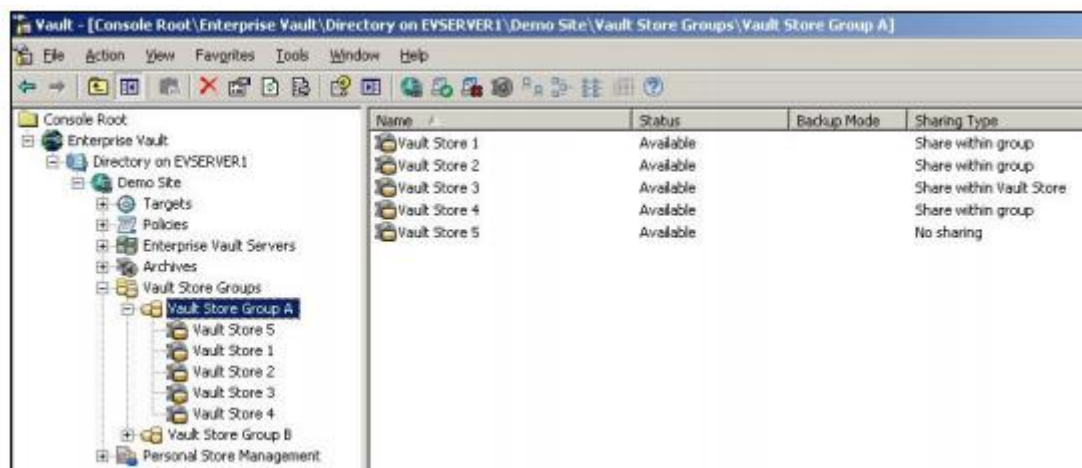


## Sharing (single-instancing / SIS)

Enterprise Vault uses Vault Store Groups to configure item sharing. A Vault Store Group consists of one or more Vault Stores. A Vault Store Group is also a single instance sharing boundary and, as such, any items stored within a Vault Store in the same Vault Store Group can potentially be eligible for sharing with other items stored within that Group. This sharing option does not apply between Vault Stores in different Vault Store Groups.

The Vault Stores contained within Vault Store Groups can each be set to one of three levels of sharing, depending on storage requirements:

- **Share within group** - Items are eligible for sharing with other items from any other Vault Store in the Group that is set to the same level of sharing.
- **Share within Vault Store** - Items are only eligible for sharing with other items archived into the same Vault Store.
- **No sharing** - Do not initiate any sharing at all.



When processing Enterprise Vault, Nuix must have access to ALL data in order to ensure correct reconstruction of single instanced emails and attachments. When Nuix processes an item in the .DVS file, Nuix will then query the item in the database and determine whether it has any single-instanced parts. A .DVSSP file may exist in the same directory as the parent .DVS file or it could be in a completely different Vault Store Partition or Vault Store.

Failure to ensure the following steps occur will result in missed or incomplete data:

1. The entire EV source data corpus must be presented to Nuix during processing.
2. All of the appropriate EV databases must be correctly restored.
3. Nuix must be able to connect to SQL successfully.

### Note

.DVSSP/.DVSCC files were not implemented in EV until EV version 8 and above. Prior to EV v8, top-level items and all attachments were stored in a single .DVS file. If the EV environment does not contain any data that is archived with EV v8 or above, an SQL database may not be needed.

File Distribution & Naming

Prior to EV 8.0, files are named in a hash and date-based format:

(XXXXXXXXXXXXXXXX~**YY**Y**MM**DD**HH**MMSS**XXXX**~**X**.DVS)

For example: 1820000000000000~**20060811**1431200000~0.DVS

In order to identify these items in the SQL database, these file names are queried against the *IdChecksumHigh*, *IdChecksumLow*, *IdDateTime*, and *IdUniqueNo* in various tables and databases to determine the EV Transaction ID. This Transaction ID is then used for any remaining queries that need to occur.

In EV 8.0 and above, the convoluted naming scheme was replaced with a much simpler one using the actual Transaction ID. Having the files named using the Transaction ID removes a lot of the complexity with previous versions. Files are named in a hash format:

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.DVS

For example: 6085D3C8F2E240B1BF80A518B2D68670.DVS

Name	Size	Type
50BACC0E626DE74E9422921811B69E31.DVS	8 KB	Enterprise V
50BACC0E626DE74E9422921811B69E31~21~4C642596~00~1.DVSCC	3 KB	DVSCC File
50BACC0E626DE74E9422921811B69E31~21~4C642596~00~1.DVSSP	12 KB	DVSSP File
50BACC0E626DE74E9422921811B69E31~25~236A1E23~00~1.DVSCC	5 KB	DVSCC File
50BACC0E626DE74E9422921811B69E31~25~236A1E23~00~1.DVSSP	9 KB	DVSSP File
50BACC0E626DE74E9422921811B69E31~C4~E2471232~00~1.DVSSP	32 KB	DVSSP File

Typically, each EV archive server will write EV files to a specific Vault Store Partition location as specific in the EV Admin Console. It is also common for the Vault Store Partition to include date-based subfolders (2010, 2011, or 2012). All EV files are stored in Vault Store Partitions.

The names for each of these files and the path to the items within the Vault Store partition are generated from the Enterprise Vault Transaction ID of the item, the current date, and a number of other attributes. The first level folder under the Vault Store Partition root is named using the current year, the second level uses the current month and day (hyphenated), the third level uses the first character of the item's Transaction ID, and the final level is the next three characters of the Transaction ID.

The actual names of the DVS, DVSSP, and DVSCC files start with the full Transaction ID; in this case, 50BACC0E626DE74E9422921811B69E31.

<

Typical best practice is to process all Enterprise Vault data from the original storage locations. However, it is possible to move the data, provided the following conditions are met:

**Note**

- Data from individual Vault Store partitions is always kept separate.
- Any moved files are identified with a specific Vault Store Partition.
- If single-instancing is in place, ALL data must be moved and presented to the Nuix server to ensure all email can be fully rehydrated.

**Tip**

A typical best practice is to break up processing of EV data into logical subsets based on server, date, and/or overall volume. If moving EV data for processing, the database will require edits to reflect the updated Vault Store Partition location. Depending on the sharing level, all data may have to be moved together in order for SIS content to be reconstituted correctly.

## Centera

It is common for EV deployments to be archived using Centera storage. In order to target data on a Centera device, you must produce a list of the C-Clips that correspond to your data. For Enterprise Vault archives, the list of Clips is available in the "dbo.SavesetStore" table in each individual EV Vault Store DB.

## Prerequisites

The following prerequisites must be met before EMC Centera data can be processed using Nuix:

1. **Pool Entry Authorization (PEA) File** - *if applicable*

- If the EMC Centera device has been configured with additional storage pools, a PEA file will need to be obtained and placed in a dedicated location on the Nuix server. A PEA file is an encrypted file used to communicate and distribute authentication credentials to Centera and contains the default key, key, and credential.

2. **Environment User Variable** - *if applicable*

- If a PEA file is necessary, it must be obtained and placed in a dedicated location on the Nuix server. Next, an Environment Variable must be created on the Nuix server with a value of the PEA file path.

3. **List of Centera Access Nodes (IPs)**

- The Centera Access Nodes are simply the IP addresses of the nodes available on Centera. This is typically a list of two or four IP addresses. If replication is enabled, more addresses may be available.

4. **List of C-Clips**

- C-Clips, also known as Clips, are unique alphanumeric strings that reference specific source data within Centera. The source data varies depending on the archive; however, the concept is always the same. A list

of C-Clips can be generated using several techniques and passed in to Nuix in order to process data.

### Pool Entry Authorization

A Pool Entry Authorization (PEA) file, generated while creating or updating an access profile, is a clear-text, XML-formatted, non-encrypted file that can be used by system administrators to communicate and distribute authentication credentials to application administrators. A PEA file is optional for profiles with non-encoded secrets (created using the File and Prompt options) but is mandatory for profiles with base-64 encoded secrets (created with the Generate option).

#### Note

A PEA file may not be necessary in all environments unless Centera has been specifically setup and configured to require one. For example, a Centera setup with only the default pool may not require a PEA file for authentication. If a PEA file is not needed for authentication, the Environment Variable is not necessary.

### Configuration

If the EMC Centera device has been configured with additional storage pools, a PEA file will need to be obtained and placed in a dedicated location on the Nuix server. An Environment Variable will also need to be created for Nuix to reference the PEA file.

1. Obtain the PEA file from the customer.

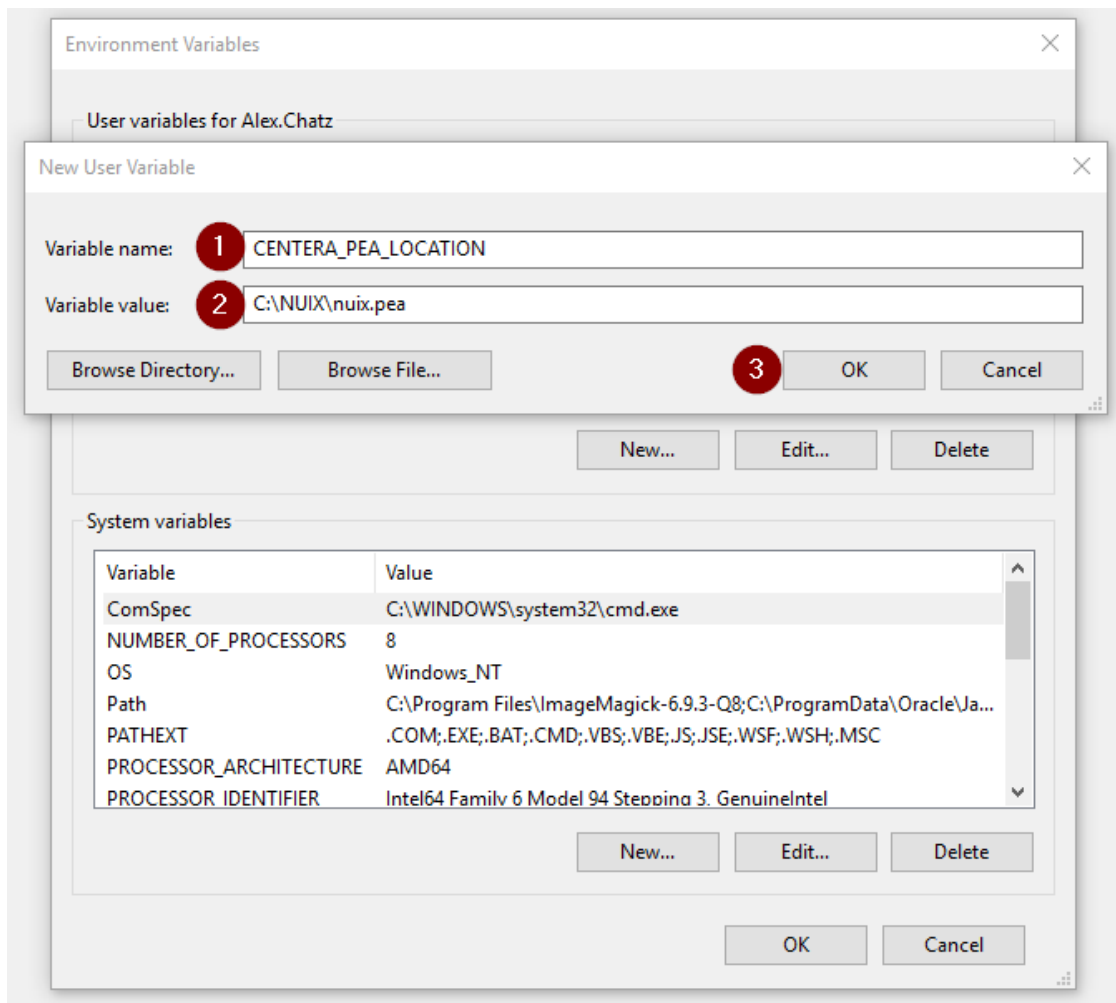
```
<pea version="1.0.0">
<defaultkey name="nuix">
<credential id="csp1.secret" enc="base64">bnVpeDEyMw==</credential>
</defaultkey>
<key type="cluster" id="1f778c7e-1dd2-11b2-ac5c-d3a0eb16f02e" name="nuix">
<credential id="csp1.secret" enc="base64">bnVpeDEyMw==</credential>
</key>
</pea>
```

2. Place the PEA file in a dedicated directory on the Nuix server.
3. Create an Environment Variable that references the location of the PEA file.

#### Warning

It is critical that the Environment Variable is configured properly. Refer to the image below for more information.

- The Variable name must be: **CENTERA\_PEA\_LOCATION**
- The Variable value **must be the absolute path** to the physical location of the PEA file on the Nuix server
- Click OK to save the variable



## Access Node IP Addresses

An EMC Centera access node is a node that has the access role applied to it. Access nodes are gateways to the data stored in Centera. These nodes have IP addresses on the network and are responsible for authentication. If you successfully connect to one such node, you have access to the entire cluster. However, to connect faster, you can specify several available nodes with the access role.

## Configuration

The Centera access nodes are simply the IP addresses of the nodes available on Centera. This is typically a list of two or four IP addresses. If replication is enabled on Centera, more addresses may be available.

After a list of Centera access nodes has been obtained from the customer, next steps include:

1. Create an .IPF (standard text file - "IP Address File") file with each line representing the IP of each Centera access node.

```
1 10.162.111.17
2 10.162.111.18
3 10.162.111.19
4 10.162.111.20
```



2. Place this .IPF file in a dedicated directory on the Nuix server.

- This .IPF file will be required when processing data on Centera.

### C-Clip List

Centera-based data is referenced by C-Clips, which must be passed in to Nuix at time of processing. Unless the client provides these Clips, you will need to generate them using the methods outlined in this section.

#### Generating the Clip List(s) for Enterprise Vault

The list of C-Clips are available in the 'dbo.SavesetStore' table in each of the Vault Store SQL databases.

- Copy the list of clips out of each data Vault Store database table and paste them into a .CLP file (standard text file - "Clip List file").
- The clip list can now be carved up into manageable sets of .CLP files.

For Enterprise Vault, a common workflow may include create Clip Lists containing around 6,000,000 clips per file. This should average out to around 1 TB of compressed source data.

#### Note

Generating a list of C-Clips is dependent on the number of pools that exist in Centera. If only the default pool exists, the list of C-Clips can only be generated for the default pool. Every additional pool that is created will have its own set of C-Clips associated with it.

#### Warning

Understanding the differences between legacy archive file types is critical when processing the same archive data on Centera. For example, processing EMC EmailXtender or EMC SourceOne C-Clips is fundamentally different than processing Symantec EV C-Clips.

A single EmailXtender C-Clip is equivalent to one .EMX file, which could contain hundreds or thousands to-level emails, whereas a single Enterprise Vault C-Clip is equivalent to one .DVS file which can only contain a single top-level email.

Processing 10,000 EmailXtender C-Clips would take substantially longer than processing 10,000 Enterprise Vault C-Clips.

## Database

EV email archive deployments require at least a single database (EnterpriseVaultDirectory) and often have multiple associated databases (one per Vault Store). Additionally, if the archive is stored on a Centera device, database access will be required for retrieval of Centera Clips.

A best practice is to restore a copy of all EV archive databases locally on the Nuix server(s).

**Tip**

Always restore local database copies for best performance.

## Authentication

The credentials used to access SQL will be passed via the Nuix startup file. It is recommended that the account provided uses built-in SQL authentication. Since the SQL instance is local and used only for Nuix processing, we recommend simply using the 'sa' (SQL Administrator) account for Nuix access.

## Configuration

After the EV SQL databases are restored, the dbo.PartitionEntry table will need to be updated in order to point to the location of the source data as it presented on the Nuix server(s). More information on .dbo.PartitionEntry can be found below.

## Critical Tables

It is important to understand the data contained in several tables, and the way Nuix uses these tables to interpret EV data.

### dbo.PartitionEntry

The full list of Vault Store Partitions are located in the dbo.PartitionEntry table under the PartitionRootPath column. The value in this column must be updated to reflect the path to the Vault Store Partitions on the Nuix Server(s).

	PartitionEntryId	IdPartition	VaultStoreEntryId	PartitionRootPath
1	17D1CD5B3441C2245936D8AB4...	0	1D367EA9AD88EA7469E17CD45CBEE34BF1210000EV01	G:\EVStorage\Express Vault Store\Express Vault Store Ptn1
2	12E2D8D462ADE1B4BAAFB8E8...	0	18C4069B6CAD11A44A6C502C69A508B6C1210000EV01	G:\EV Storage 2

### dbo.Archive

The full list of users within the SQL Database can be found within the dbo.Archive table under the ArchiveName column. Use this list to cull down to the set of users to be exported. This information is used in conjunction with the EV Manifest Workflow.

	RootIdentity	VaultStoreEntryId	ArchiveName
1	1	1D367EA9AD88EA7469E17CD45CBEE34BF1210000EV01	EVJournal
2	2	1D367EA9AD88EA7469E17CD45CBEE34BF1210000EV01	Bernard.Schilling
3	4	1D367EA9AD88EA7469E17CD45CBEE34BF1210000EV01	Bernad.TEspey
4	6	1D367EA9AD88EA7469E17CD45CBEE34BF1210000EV01	Bob.Hawks
5	8	1D367EA9AD88EA7469E17CD45CBEE34BF1210000EV01	Brenda.OFreitas
6	10	1D367EA9AD88EA7469E17CD45CBEE34BF1210000EV01	Cheryl.OStafford
7	12	1D367EA9AD88EA7469E17CD45CBEE34BF1210000EV01	Deidre.Holtzmann
8	14	1D367EA9AD88EA7469E17CD45CBEE34BF1210000EV01	Ginger.Trombley
9	16	1D367EA9AD88EA7469E17CD45CBEE34BF1210000EV01	Heike.Weissmuller
10	18	1D367EA9AD88EA7469E17CD45CBEE34BF1210000EV01	Henry.Herzog
11	20	1D367EA9AD88EA7469E17CD45CBEE34BF1210000EV01	Jeffery.Land

## Supported Workflows

### Mailbox Archive to User PST

1. In the **Legacy Archive** dropdown box, select **Veritas Enterprise Vault**
2. Select **Mailbox** from the **Archive Type** radio button selections
3. Select **User** from the **Output Type** radio button selections
4. **PST** will automatically be selected as the **Lightspeed Extraction Output**

#### Note

Due to the complex nature of mailbox folder structures that may exist in EV Mailbox Archives, PST is the only option. PSTs are much easier to manage since emails will be grouped by User on the file system, and long file path names issues will be avoided.

5. If a date filter is requested, enter a date in the **From Date:** and **To Date:** date chooser field.
6. Enter the SQL connection string details in the **SQL Connection Info** settings. Be sure to click **Test SQL Connection** to ensure you can properly connect to SQL.

#### Warning

Testing the SQL is critical, as you will want to ensure that you can correctly query the SQL databases for additional information.

#### Tip

It is recommended to use a built-in SQL service account (instead of Windows authentication) with **READ** access to **ALL** of the Enterprise Vault databases.

7. The Enterprise Vault Settings will default to:

Name	Default Value	Description
Skip Additional SQL Lookups	True	Using the connected EV SQL database, Nuix will not perform any additional EV SQL queries, other than the basic required for reconstituting single-instanced attachments (SIS).  When set to <b>False</b> , additional lookups will be performed and will slow down processing items substantially.  It is recommended to <b>ALWAYS</b> set this to True unless otherwise noted.
Use FileTransactionID over ParentTransactionID	False	When enabled, EV SQL lookups will look use the FileTransactionID column name instead of the usual ParentTransactionID.
User List:	empty	A User List is relevant only when

		<p>extracting data from an EV Mailbox Archive.</p> <p>The <b>User List</b> CSV file must contain 1 EV Mailbox Archive Name per line, for example:</p> <p>John.Doe</p> <p>John.Smith</p>
--	--	---

#### Note

The EV Mailbox Archive Names in the **User List** CSV file must match EXACTLY the way they do in the **ArchiveName** column located in the **EnterpriseVaultDirectory.dbo.Archive** table.

#### Warning

Worker Side Script (WSS) is not available for EV Mailbox Archive extractions. EV Mailbox Archive extractions typically require all data to be exported with any content filtering.

8. Review the settings you've selected for this job and click the **Add Batch to Grid** button.
  - a. If you have more mailbox archives you need to process, select a new list of users and click **Add Batch to Grid**, continuing this until all of users have been added to a Batch on the Grid.

#### Warning

Performance may vary, especially based on the settings configured on the **Archive Extraction** tab in **Global Settings**. Be sure to **ALWAYS** review settings prior to starting Jobs.

9. When you are ready to begin processing, click on the **Start Job** button.

### Journal Archive to User PST

1. In the **Legacy Archive** dropdown box, select **Veritas Enterprise Vault**
2. Select **Journal** from the **Archive Type** radio button selections
3. Select **User** from the **Output Type** radio button selections
4. Select **PST** as the **Lightspeed Extraction Output**

#### Note

It is highly recommend that you select **PST** when your **Output Type** is set to **User**. PSTs are much easier to manage since emails will be grouped by User on the file system, and long file path names issues will be avoided.

5. If a date filter is requested, enter a date in the **From Date:** and **To Date:** date chooser field.
6. Choose whether the Source Data exists in **Folders**, **Files** or on **Centera**.
  - a. If **Folders** or **Files**, use the navigation pane to select the drive letter

where the source data exists. You may also choose to **Compute Batch Size**. This will allow you to obtain additional metrics while the job is processing, like *Percentage Completed* and *Total Bytes* to process.

**Tip**

It is recommend that for each Job, target **no more** than 500 GB - 1 TB of compressed, source data per Job. This is for optimal performance as well as an advantage in system failure scenarios. If a system crashes mid-processing, it would be more efficient to restart a Job that will take 24 hours to complete, versus a Job that may take 7 days to complete.

- b. If **Centera** is selected, select the **Clip List** (\*.CLP) you want to process, browse to the **PEA File** (\*.PEA) and browse to the **IP File** (\*.IPF).

**Tip**

It is recommend that for each Job, target **no more** than 10,000 Clips of compressed data per Job. This is for optimal performance as well as an advantage in system failure scenarios. If a system crashes mid-processing, it would be more efficient to restart a Job that will take 24 hours to complete, versus a Job that may take 7 days to complete.

7. Enter the SQL connection string details in the **SQL Connection Info** settings. Be sure to click **Test SQL Connection** to ensure you can properly connect to SQL.

**Warning**

Testing the SQL is critical, as you will want to ensure that you can correctly query the SQL databases for additional information.

**Tip**

It is recommended to use a built-in SQL service account (instead of Windows authentication) with **READ** access to **ALL** of the Enterprise Vault databases.

8. The Enterprise Vault Settings will default to:

Name	Default Value	Description
Skip Additional SQL Lookups	True	Using the connected EV SQL database, Nuix will not perform any additional EV SQL queries, other than the basic required for reconstituting single-instanced attachments (SIS).  When set to <b>False</b> , additional lookups will be performed and will slow down processing items substantially.  It is recommended to <b>ALWAYS</b> set this to

		True unless otherwise noted.
Use FileTransactionID over ParentTransactionID	False	When enabled, EV SQL lookups will look use the FileTransctionID column name instead of the usual ParentTransactionID.
User List:	empty	Not applicable to a Journal Archive workflow.

### Warning

For emails that were archived from a journal, Nuix will add the distribution list recipients to a new metadata field called "**Expanded-DL**", however, based on your Output Type (PST, MSG or EML), the metadata may not be preserved unless **Add Distribution List Recipients** is enabled under **MAPI Export Options** or **EML Export Options** on the **Lightspeed Settings** tab in **Global Settings**.

9. The Worker Side Script (WSS) settings will allow you to:

Name	Default Value	Description
Exclude Unresponsive Items	True	When enabled, Nuix will not export any items that do not respond to your Search Terms and Mapping CSV.  If you set to <b>False</b> , you will need to adjust your Mapping CSV to include an entry for: <i>unresponsive,unresponsive.pst</i>
Verbose Logging	False	When enabled, Nuix will not include any verbose logging at the WSS-level for troubleshooting purposes.  If set to <b>True</b> , this will allow for easier troubleshooting, however, the size of the logs will be substantially larger.
Content Filtering	Email, RSS Feed, Calendar, Contact	When enabled, all top-level emails (RSS Feeds included), Calendar and Contact items will be extracted.  If you want to filter any of these kinds out, simply de-select the item kind you wish to filter.
Search Terms CSV	empty	You must select a list of Search Terms in CSV format. <b>Search Terms</b> is a 2 column CSV that includes a Flag in Column A and a Search Term in Column B. You do not need a header column.  Each line should include an SMTP address, X400 or X500 address similar to this. You can also add multiple Search Terms to the same Flag, for example:  <a href="#">Alex.Chatzistamatis,alex.chatzistamatis@nuix.com</a>

		<a href="mailto:Alex.Chatzistamatis,alex@nuix.com">Alex.Chatzistamatis,alex@nuix.com</a> Nuix will scan for these search terms across <b>Communication Metadata</b> (From, To, Cc, Bcc) and the <b>Expanded-DL</b> metadata field.
Mapping CSV	empty	You must select a Mapping in CSV format. Like <b>Search Terms, Mapping</b> is a 2 column CSV that includes a Flag Name in Column A and an Output Name in Column B. You do not need a header column.  Each line should include the Flag from your Search Terms CSV and the Output Name. If you add multiple Search Terms to a single Flag, you do not need to add multiple Flags in the Mapping CSV, for example:  Alex.Chatzistamatis,Alex.Chatzistamatis.pst

10. Review the settings you've selected for this job and click the **Add Batch to Grid** button.
  - a. If you have more mailbox archives you need to process, select a new list of users and click **Add Batch to Grid**, continuing this until all of users have been added to a Batch on the Grid.

#### Warning

Performance may vary, especially based on the settings configured on the **Lightspeed Settings** tab in **Global Settings**. Be sure to **ALWAYS** review settings prior to starting Jobs.

11. When you are ready to begin processing, click on the **Start Job** button.

### Journal Archive to Flat PST, MSG or EML

1. In the **Legacy Archive** dropdown box, select **Veritas Enterprise Vault**
2. Select **Journal** from the **Archive Type** radio button selections
3. Select **Flat** from the **Output Type** radio button selections
4. Select **PST, MSG or EML** as the **Lightspeed Extraction Output**

#### Note

It is highly recommend that you select **PST** when your **Output Type** is set to **User**. PSTs are much easier to manage since emails will be grouped by User on the file system, and long file path names issues will be avoided.

5. If a date filter is requested, enter a date in the **From Date:** and **To Date:** date chooser field.
6. Choose whether the Source Data exists in **Folders, Files** or on **Centera**.
  - a. If **Folders** or **Files**, use the navigation pane to select the drive letter where the source data exists. You may also choose to **Compute Batch**

**Size.** This will allow you to obtain additional metrics while the job is processing, like *Percentage Completed* and *Total Bytes* to process.

**Tip**

It is recommend that for each Job, target **no more** than 500 GB - 1 TB of compressed, source data per Job. This is for optimal performance as well as an advantage in system failure scenarios. If a system crashes mid-processing, it would be more efficient to restart a Job that will take 24 hours to complete, versus a Job that may take 7 days to complete.

- b. If **Centera** is selected, select the **Clip List** (\*.CLP) you want to process, browse to the **PEA File** (\*.PEA) and browse to the **IP File** (\*.IPF).

**Tip**

It is recommend that for each Job, target **no more** than 10,000 Clips of compressed data per Job. This is for optimal performance as well as an advantage in system failure scenarios. If a system crashes mid-processing, it would be more efficient to restart a Job that will take 24 hours to complete, versus a Job that may take 7 days to complete.

7. Enter the SQL connection string details in the **SQL Connection Info** settings. Be sure to click **Test SQL Connection** to ensure you can properly connect to SQL.

**Warning**

Testing the SQL is critical, as you will want to ensure that you can correctly query the SQL databases for additional information.

**Tip**

It is recommended to use a built-in SQL service account (instead of Windows authentication) with **READ** access to **ALL** of the Enterprise Vault databases.

8. The Enterprise Vault Settings will default to:

Name	Default Value	Description
Skip Additional SQL Lookups	True	Using the connected EV SQL database, Nuix will not perform any additional EV SQL queries, other than the basic required for reconstituting single-instanced attachments (SIS).  When set to <b>False</b> , additional lookups will be performed and will slow down processing items substantially.  It is recommended to <b>ALWAYS</b> set this to True unless otherwise noted.
Use	False	When enabled, EV SQL lookups will look use



FileTransactionID over ParentTransactionID		the FileTransctionID column name instead of the usual ParentTransactionID.
User List:	empty	Not applicable to a Journal Archive workflow.

**Warning**

For emails that were archived from a journal, Nuix will add the distribution list recipients to a new metadata field called "**Expanded-DL**", however, based on your Output Type (PST, MSG or EML), the metadata may not be preserved unless **Add Distribution List Recipients** is enabled under **MAPI Export Options** or **EML Export Options** on the **Lightspeed Settings** tab in **Global Settings**.

9. Review the settings you've selected for this job and click the **Add Batch to Grid** button.
  - a. If you have more mailbox archives you need to process, select a new list of users and click **Add Batch to Grid**, continuing this until all of users have been added to a Batch on the Grid.

**Warning**

Performance may vary, especially based on the settings configured on the **Lightspeed Settings** tab in **Global Settings**. Be sure to **ALWAYS** review settings prior to starting Jobs.

10. When you are ready to begin processing, click on the **Start Job** button.

# Working with EMC EmailXtender/SourceOne

When working with EMC EmailXtender/SourceOne source data, you can use NEAMM in various methods to target the data directly on the file system in its proprietary, .EMX file format (or on Centera when applicable) and export it out to disk in PST, MSG or EML format. Below are several workflows that NEAMM can handle.

## Overview

EMC archiving solutions have two components:

- Files on disk
- SQL database

Nuix must have access to the database while processing EMC files on disk in order to expand distribution lists.

## Prerequisites

The following prerequisites must be met before EAS data can be processed using Nuix:

1. EMC database restoration<sup>3</sup>
  - A copy of the production EMC database(s) must be restored to a local MSSQL instance on the Nuix server.
2. Source data identified with a specific EMC database<sup>4</sup>

## Source Data

### Physical Files

EMC data is written to the file system in the form of .EMX files. These files are compressed containers that store individual messages with attachments. Each EMX file can contain a few messages or up to several thousand messages.

### File Structure

EMX files are highly compressed files that contain potentially thousands of emails, calendar entries, or contacts. The first item under each EMX file is a single .VOL file that contains metadata about the container itself and holds very little value. Below that, each top-level item will be nested within the EMX wrapper. This wrapper is used by EMC itself for identification purposes and contains no value. Below the EMC wrapper, the top-level item is nested. If ingesting into a case these files

---

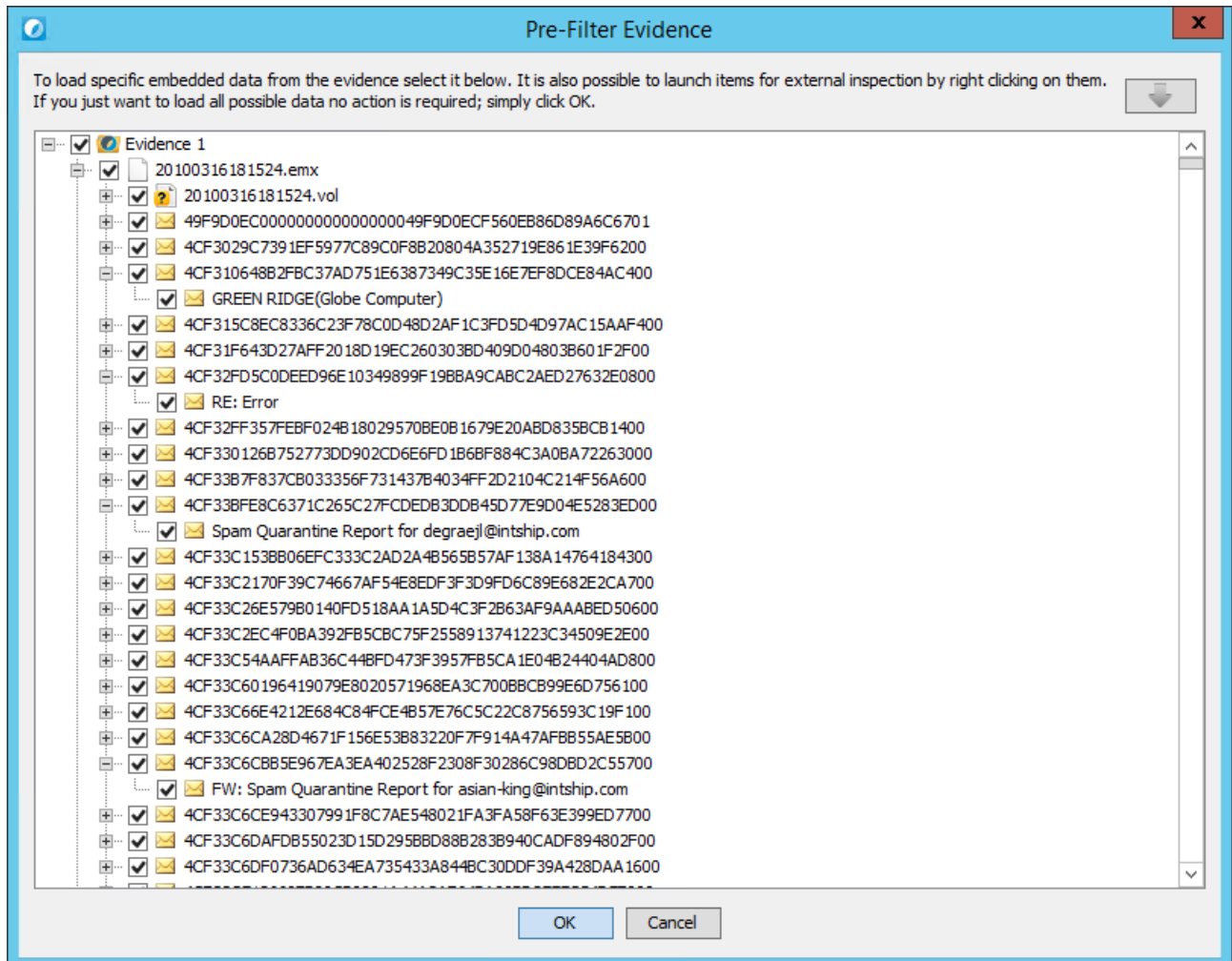
<sup>3</sup> EMC data can be processed without a database connection; however, the database is required for expansion of distribution lists.

<sup>4</sup> Many large EMC archive deployments include multiple archiving servers, each of which will have a separate database.

should be excluded.

**Note**

The .VOL file is not critical to email migrations and will be automatically skipped by Lightspeed.



## Distribution List Expansion

Nuix is able to expand distribution lists (DL) at time of processing if connected to the EMC database. This is often an important component of an archive extraction and should be discussed with the client prior to beginning any work. There are several options for where to inject the expanded addresses in the email metadata.

Designate the Field for Expanded Addresses

Nuix default behavior is to expand distribution lists and write the addresses to an 'Expanded-DL' metadata field. It is also default behavior to push these addresses into the 'To' field. Therefore, it is necessary to adjust system properties in your Nuix startup file either to redirect DL expansion from the 'To' field to the 'BCC' field, or avoid this all together.

## File Distribution and Naming

Files are named in the following date-based format:

YYYYMMDDXXXXXX.emx

For example, 20100316181524.emx

It is common for multiple containers to be created on the same date, so the additional characters XXXXXX are necessary. It is believed that these are also timestamp based (HHMMSS), but it is uncertain precisely how these are calculated.

Typically, each EMC archive server will write all EMX files to a single root directory. It is also common for the root directory to include date-based subfolders (2010, 2011, 2012), or to write the journal mail to a distinct folder. However, the data is laid out on disk and the database provides a map. You can view the path to each individual EMX file in the database at **dbo.Volume.CurrentUNCPath**.

VolumeName	VolumeSize	MsgCount	CurrentUNCPath
20100316181524	102212	2721	\\SOURCE-ONE\\MainArchiveFolder\\MainArchiveFolder\\201003\\20100316181524.emx
20100316182115	90210	4339	\\SOURCE-ONE\\MainArchiveFolder\\MainArchiveFolder\\201001\\20100316182115.emx
20100316182116	102373	5113	\\SOURCE-ONE\\MainArchiveFolder\\MainArchiveFolder\\200912\\20100316182116.emx
20100316182117	100553	4909	\\SOURCE-ONE\\MainArchiveFolder\\MainArchiveFolder\\200910\\20100316182117.emx
20100316182118	102359	3301	\\SOURCE-ONE\\MainArchiveFolder\\MainArchiveFolder\\200908\\20100316182118.emx

### Note

Each .EMX file can range in size, but they are generally 80-100 MB size each. More detail on this is provided in the EMC Database section.

### Tip

Typical best practice is to break up processing of EMC data into logical subsets based on server, date, and/or overall volume. It is possible to move EMC data for processing without requiring any edits to the database.

## Centera

It is common that EMC data is archived to an EMC Centera storage deployment. In order to target data on a Centera device, you must produce a list of the C-Clips that correspond to your data. For EmailXtender this must be done by running a utility from the original archive server; for SourceOne, you can retrieve the Clips from the database.

### Note

A C-Clip in Centera is equivalent to one .EMX file on the file system. Each clip will range from 80-100 MB each.

## Prerequisites

The following prerequisites must be met before EMC Centera data can be processed using Nuix:

1. **Pool Entry Authorization (PEA) File** - *if applicable*
  - If the EMC Centera device has been configured with additional storage pools, a PEA file will need to be obtained and placed in a dedicated location on the Nuix server. A PEA file is an encrypted file used to communicate and distribute authentication credentials to Centera and

contains the default key, key, and credential.

**2. Environment User Variable** - *if applicable*

- If a PEA file is necessary, it must be obtained and placed in a dedicated location on the Nuix server. Next, an Environment Variable must be created on the Nuix server with a value of the PEA file path.

**3. List of Centera Access Nodes (IPs)**

- The Centera Access Nodes are simply the IP addresses of the nodes available on Centera. This is typically a list of two or four IP addresses. If replication is enabled, more addresses may be available.

**4. List of C-Clips**

- C-Clips, also known as Clips, are unique alphanumeric strings that reference specific source data within Centera. The source data varies depending on the archive; however, the concept is always the same. A list of C-Clips can be generated using several techniques and passed in to Nuix in order to process data.

### Pool Entry Authorization

A Pool Entry Authorization (PEA) file, generated while creating or updating an access profile, is a clear-text, XML-formatted, non-encrypted file that can be used by system administrators to communicate and distribute authentication credentials to application administrators. A PEA file is optional for profiles with non-encoded secrets (created using the File and Prompt options) but is mandatory for profiles with base-64 encoded secrets (created with the Generate option).

**Note**

A PEA file may not be necessary in all environments unless Centera has been specifically setup and configured to require one. For example, a Centera setup with only the default pool may not require a PEA file for authentication. If a PEA file is not needed for authentication, the Environment Variable is not necessary.

### Configuration

If the EMC Centera device has been configured with additional storage pools, a PEA file will need to be obtained and placed in a dedicated location on the Nuix server. An Environment Variable will also need to be created for Nuix to reference the PEA file.

1. Obtain the PEA file from the customer.

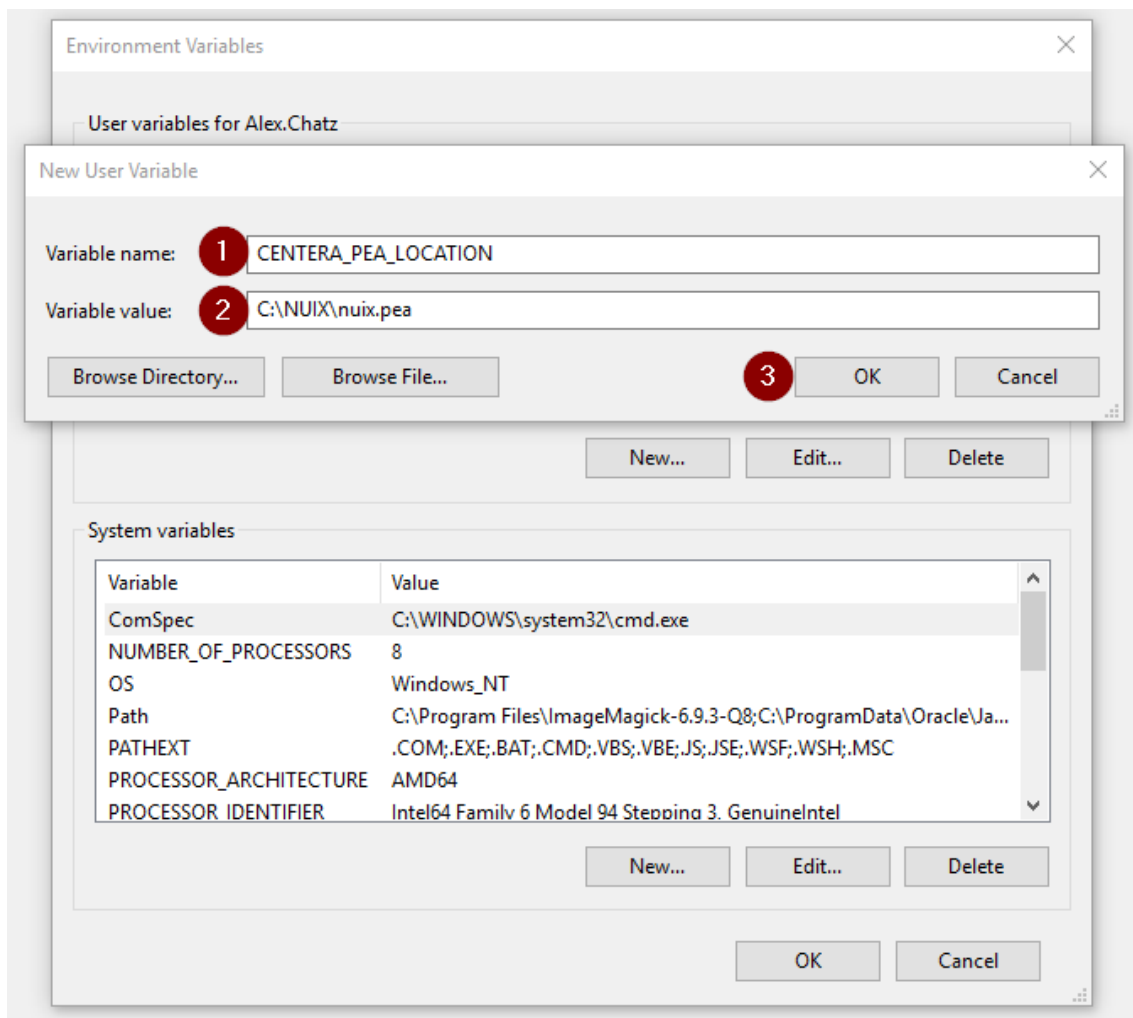
```
<pea version="1.0.0">
<defaultkey name="nuix">
<credential id="csp1.secret" enc="base64">bnVpeDEyMw==</credential>
</defaultkey>
<key type="cluster" id="1f778c7e-1dd2-11b2-ac5c-d3a0eb16f02e" name="nuix">
<credential id="csp1.secret" enc="base64">bnVpeDEyMw==</credential>
</key>
</pea>
```

2. Place the PEA file in a dedicated directory on the Nuix server.
3. Create an Environment Variable that references the location of the PEA file.

It is critical that the Environment Variable is configured properly. Refer to the image below for more information.

### Warning

- The Variable name must be: **CENTERA\_PEA\_LOCATION**
- The Variable value **must be the absolute path** to the physical location of the PEA file on the Nuix server
- Click OK to save the variable



### Access Node IP Addresses

An EMC Centera access node is a node that has the access role applied to it. Access nodes are gateways to the data stored in Centera. These nodes have IP addresses on the network and are responsible for authentication. If you successfully connect to one such node, you have access to the entire cluster. However, to connect faster, you can specify several available nodes with the access role.

## Configuration

The Centera access nodes are simply the IP addresses of the nodes available on Centera. This is typically a list of two or four IP addresses. If replication is enabled on Centera, more addresses may be available.

After a list of Centera access nodes has been obtained from the customer, next steps include:

1. Create an .IPF (standard text file - "IP Address File") file with each line representing the IP of each Centera access node.

```
1 10.162.111.17
2 10.162.111.18
3 10.162.111.19
4 10.162.111.20
```

2. Place this .IPF file in a dedicated directory on the Nuix server.

- This .TXT file will be required when processing data on Centera.

## C-Clip List

Centera-based data is referenced by C-Clips, which must be passed in to Nuix at time of processing. Unless the client provides these Clips, you will need to generate them using the methods outlined in this section.

### Generating the Clip List(s) for EmailXtender

The EmailXtender deployment contains everything you need to retrieve the Centera Clips, but you will require access to the EmailXtender archive server to perform the necessary steps. This is typically performed in conjunction with a client tech that can facilitate the necessary access. The following is a walkthrough of the process:

1. Identify DxMChk.exe utility
2. Identify source data for Clip retrieval<sup>5</sup>
3. Execute configured DxMChk utility
4. Repeat Steps 2 and 3 as necessary to retrieve all Clips.

### Generating the Clip List(s) for SourceOne

The SourceOne database contains a list of all Centera Clips in the archive. This list can be found in the Volume table in the VolStubXML column. Each row of the Volume table corresponds to a specific EMX file, each of which will be archived to a single Centera Clip. If you wish to scope lists of Clips based on date range or archive server, you can reference the other columns in this table to do so.

#### Note

Generating a list of C-Clips is dependent on the number of pools that exist in Centera. If only the default pool exists, the list of C-Clips can only be generated for the default pool. Every additional pool that is created will have its own set of C-Clips associated with it.

---

<sup>5</sup> You will often want to retrieve the Clips in batches according to date range or volume, as opposed to retrieving all Clips at once.

### Warning

Understanding the differences between legacy archive file types is critical when processing the same archive data on Centera. For example, processing EMC EmailXtender or EMC SourceOne C-Clips is fundamentally different than processing Symantec EV C-Clips.

A single EmailXtender C-Clip is equivalent to one .EMX file, which could contain hundreds or thousands of top-level emails, whereas a single Enterprise Vault C-Clip is equivalent to one .DVS file which can only contain a single top-level email.

Processing 10,000 EmailXtender C-Clips would take substantially longer than processing 10,000 Enterprise Vault C-Clips.

- The clip list can now be carved up into manageable sets of .CLP files.

For EMC archives, a common workflow may include create Clip Lists containing around 10,000 clips per file. This should average out to around 1 TB of compressed source data.

## Database

EMC email archive deployments require at least a single database and often have multiple associated databases (one per email archive server). Nuix can process EMC data without these databases; however, a database is required for expanding distribution lists. Additionally, if the archive is SourceOne and the data is stored on a Centera device, database access will be required for retrieval of Centera Clips.

Best practice is to restore a copy of all EMC archive databases locally on the Nuix server(s). However, it is possible to use production SQL instances if this is desired, since, unlike with other archives, EMC databases require no editing for use with Nuix.

### Tip

Always restore local database copies for best performance.

## Authentication

The credentials used to access SQL will be passed in via the Nuix startup file. It is recommended that the account provided uses built-in SQL authentication. Since the SQL instance is local and used only for Nuix processing, we recommend simply using the 'sa' (SQL Administrator) account for Nuix access.

## Configuration

EMC databases do not require any special configuration for use with Nuix. EmailXtender and SourceOne databases use nearly identical structure; however, there are differences. The only requirement for the Nuix engineer is to ensure that, in the case of SourceOne data, the Nuix startup includes a switch<sup>6</sup> to tell Nuix to use the SourceOne schema.

---

<sup>6</sup> Dnuix.data.xtender.addressDbSchema=sourceOne



## Critical Tables

It is important to understand the data contained in several tables and the way Nuix uses these tables to interpret EAS data.

### dbo.Volume

The Volume table contains several columns of interest. Number of messages, date range, and volume can all be tracked in this table on a per-EMX file basis.

VolumeId	VolumeName	VolumeSize	VolumeStartDate	VolumeEndDate	MsgCount	ServerID
1	20100316181524	102212	2010-03-01 01:33:16.000	2010-03-16 21:08:11.000	2721	1229193002
2	20100316182115	90210	2010-01-01 00:00:36.000	2010-01-31 22:59:01.000	4339	1229193002
3	20100316182116	102373	2009-12-01 00:03:16.000	2009-12-31 23:59:32.000	5113	1229193002
4	20100316182117	100553	2009-10-01 00:07:41.000	2009-10-31 23:14:06.000	4909	1229193002
5	20100316182118	102359	2009-08-01 00:02:27.000	2009-08-31 23:53:42.000	3301	1229193002
6	20100316182146	102512	2009-07-01 01:00:14.000	2009-07-31 23:51:40.000	3412	1229193002

Additionally, for SourceOne data on Centera, the Centera Clip listing is generated from the VolStubXML column in this table.

ServerID	CurrentUNCPath	FolderNodeID	VolStubXml
1229193002	\\SOURCE-ONE\\MainArchiveFolder\\MainArchiveFolder...	1706124426	
1229193002	\\SOURCE-ONE\\MainArchiveFolder\\MainArchiveFolder...	2101562106	
1229193002	\\SOURCE-ONE\\MainArchiveFolder\\MainArchiveFolder...	-227173718	
1229193002	\\SOURCE-ONE\\MainArchiveFolder\\MainArchiveFolder...	-1462631378	
1229193002	\\SOURCE-ONE\\MainArchiveFolder\\MainArchiveFolder...	-820577755	
1229193002	\\SOURCE-ONE\\MainArchiveFolder\\MainArchiveFolder...	-1580046614	

### dbo.EmailAddresses

The EmailAddresses table contains a listing of all individual email addresses in the archive. The majority of these entries are SMTP, but SYS, EX, and PST addresses are also stored here.

- **SYS:** These entries are system-generated addresses by EMC.
  - **Example:** SYS:"ES1ExchJournal"<es1exchjournal>
- **EX:** These entries are objects from the Exchange Server. They can include the display name of a user, email address, or a full object value (LegDN).
  - **Example:** EX:"Alex Chatzistamatis"</o=nuix/ou=na/cn=conshohocken/cn=achatz01>
- **SMTP:** These entries are the SMTP addresses that were on the messages at the time the item was sent/received.
  - **Example:** SMTP:"Alex Chatzistamatis"<alex.chatzistamatis@nuix.com>
- **PST:** These entries are items imported from a PST.
  - **Example:** PST:"archive.pst"\\\\"nuixfs01\\pst\\achatz01\\archive.pst

## Supported Workflows

### Journal Archive to User PST

1. In the **Legacy Archive** dropdown box, select **EMC EmailXtender** or **EMC SourceOne**
2. Select **Journal** from the **Archive Type** radio button selections
3. Select **User** from the **Output Type** radio button selections
4. Select **PST** as the Lightspeed Extraction Output

#### Note

It is highly recommend that you select **PST** when your **Output Type** is set to **User**. PSTs are much easier to manage since emails will be grouped by User on the file system, and long file path names issues will be avoided.

5. If a date filter is requested, enter a date in the **From Date:** and **To Date:** date chooser field.
6. Choose whether the Source Data exists in **Folders**, **Files** or on **Centera**.
  - a. If **Folders** or **Files**, use the navigation pane to select the drive letter where the source data exists. You may also choose to **Compute Batch Size**. This will allow you to obtain additional metrics while the job is processing, like *Percentage Completed* and *Total Bytes* to process.

#### Tip

It is recommend that for each Job, target **no more** than 500 GB – 1 TB of compressed, source data per Job. This is for optimal performance as well as an advantage in system failure scenarios. If a system crashes mid-processing, it would be more efficient to restart a Job that will take 24 hours to complete, versus a Job that may take 7 days to complete.

- b. If **Centera** is selected, select the **Clip List** (\*.CLP) you want to process, browse to the **PEA File** (\*.PEA) and browse to the **IP File** (\*.IPF).

#### Tip

It is recommend that for each Job, target **no more** than 10,000 Clips of compressed data per Job. This is for optimal performance as well as an advantage in system failure scenarios. If a system crashes mid-processing, it would be more efficient to restart a Job that will take 24 hours to complete, versus a Job that may take 7 days to complete.

7. Enter the SQL connection string details in the **SQL Connection Info** settings. Be sure to click **Test SQL Connection** to ensure you can properly connect to SQL.

#### Warning

Testing the SQL is critical, as you will want to ensure that you can correctly query the SQL databases for additional information.

**Tip**

It is recommended to use a built-in SQL service account (instead of Windows authentication) with **READ** access to **ALL** of the EmailXtender/SourceOne databases.

8. The EMC Settings will default to:

Name	Default Value	Description
Address Filtering:	SYS, EX, PST	With all three (3) enabled, non-SMTP addresses will be filtered from metadata. Non-SMTP addresses include addresses EMC may have added including: system addresses and PST ingestion information, as well as X400/X500 addresses. Examples of addresses are listed below:  SYS:"ES1ExchJournal"<es1exchjournal>  EX:"Alex Chatzistamatis" "</o=nuix/ou=na/cn=conshohocken/cn=achatz01>  PST:"archive.pst"<\\\\\\nuixfs01\\pst\\achatz01\\archive.pst>
Expand DL to:	"Expanded-DL"	Using the connected EMC SQL database, Nuix will query for any distribution list recipients on every email. Any responsive recipients will be added depending on the value selected in this dropdown.  <b>To: + "Expanded-DL"</b> = distribution list recipients will be added to the To: field AND the "Expanded-DL" field.  <b>Bcc: + "Expanded-DL"</b> = distribution list recipients will be added to the Bcc: field AND the "Expanded-DL" field.  <b>"Expanded-DL"</b> = distribution list recipients will be added to the "Expanded-DL" metadata field.

**Warning**

The **Expand DL to:** selection will only ADD the metadata to a new metadata field called **"Expanded-DL"**, however, based on your Output Type (PST, MSG or EML), the metadata may not be preserved unless **Add Distribution List Recipients** is enabled under **MAPI Export Options** or **EML Export Options** on the **Lightspeed Settings** tab in **Global Settings**.

9. The Worker Side Script (WSS) settings will allow you to:

Name	Default Value	Description
Exclude Unresponsive Items	True	When enabled, Nuix will not export any items that do not respond to your Search Terms and Mapping CSV.  If you set to <b>False</b> , you will need to adjust your

		Mapping CSV to include an entry for: <i>unresponsive,unresponsive.pst</i>
Verbose Logging	False	When enabled, Nuix will not include any verbose logging at the WSS-level for troubleshooting purposes.  If set to <b>True</b> , this will allow for easier troubleshooting, however, the size of the logs will be substantially larger.
Content Filtering	Email, RSS Feed, Calendar, Contact	When enabled, all top-level emails (RSS Feeds included), Calendar and Contact items will be extracted.  If you want to filter any of these kinds out, simply de-select the item kind you wish to filter.
Search Terms CSV	empty	You must select a list of Search Terms in CSV format. <b>Search Terms</b> is a 2 column CSV that includes a Flag in Column A and a Search Term in Column B. You do not need a header column.  Each line should include an SMTP address, X400 or X500 address similar to this. You can also add multiple Search Terms to the same Flag, for example:  <a href="#">Alex.Chatzistamatis,alex.chatzistamatis@nuix.com</a> <a href="#">Alex.Chatzistamatis,alex@nuix.com</a>  Nuix will scan for these search terms across <b>Communication Metadata</b> (From, To, Cc, Bcc) and the <b>Expanded-DL</b> metadata field.
Mapping CSV	empty	You must select a Mapping in CSV format. Like <b>Search Terms</b> , <b>Mapping</b> is a 2 column CSV that includes a Flag Name in Column A and an Output Name in Column B. You do not need a header column.  Each line should include the Flag from your Search Terms CSV and the Output Name. If you add multiple Search Terms to a single Flag, you do not need to add multiple Flags in the Mapping CSV, for example:  Alex.Chatzistamatis,Alex.Chatzistamatis.pst

10. Review the settings you've selected for this job and click the **Add Batch to Grid** button.
  - a. If you have more mailbox archives you need to process, select a new list of users and click **Add Batch to Grid**, continuing this until all of users have been added to a Batch on the Grid.

**Warning**

Performance may vary, especially based on the settings configured on the **Lightspeed Settings** tab in **Global Settings**. Be sure to **ALWAYS** review settings prior to starting Jobs.

11. When you are ready to begin processing, click on the **Start Job** button.

**Journal Archive to Flat PST, MSG or EML**

1. In the **Legacy Archive** dropdown box, select **EMC EmailXtender** or **EMC SourceOne**
2. Select **Journal** from the **Archive Type** radio button selections
3. Select **Flat** from the **Output Type** radio button selections
4. Select **PST, MSG or EML** as the Lightspeed Extraction Output

**Note**

It is highly recommend that you select **PST** when your **Output Type** is set to **User**. PSTs are much easier to manage since emails will be grouped by User on the file system, and long file path names issues will be avoided.

5. If a date filter is requested, enter a date in the **From Date:** and **To Date:** date chooser field.
6. Choose whether the Source Data exists in **Folders, Files** or on **Centera**.
  - a. If **Folders** or **Files**, use the navigation pane to select the drive letter where the source data exists. You may also choose to **Compute Batch Size**. This will allow you to obtain additional metrics while the job is processing, like *Percentage Completed* and *Total Bytes* to process.

**Tip**

It is recommend that for each Job, target **no more** than 500 GB - 1 TB of compressed, source data per Job. This is for optimal performance as well as an advantage in system failure scenarios. If a system crashes mid-processing, it would be more efficient to restart a Job that will take 24 hours to complete, versus a Job that may take 7 days to complete.

- b. If **Centera** is selected, select the **Clip List** (\*.CLP) you want to process, browse to the **PEA File** (\*.PEA) and browse to the **IP File** (\*.IPF).

**Tip**

It is recommend that for each Job, target **no more** than 10,000 Clips of compressed data per Job. This is for optimal performance as well as an advantage in system failure scenarios. If a system crashes mid-processing, it would be more efficient to restart a Job that will take 24 hours to complete, versus a Job that may take 7 days to complete.

7. Enter the SQL connection string details in the **SQL Connection Info** settings.

Be sure to click **Test SQL Connection** to ensure you can properly connect to SQL.

**Warning**

Testing the SQL is critical, as you will want to ensure that you can correctly query the SQL databases for additional information.

**Tip**

It is recommended to use a built-in SQL service account (instead of Windows authentication) with **READ** access to **ALL** of the EmailXtender/SourceOne databases.

8. The EMC Settings will default to:

Name	Default Value	Description
Address Filtering:	SYS, EX, PST	With all three (3) enabled, non-SMTP addresses will be filtered from metadata. Non-SMTP addresses include addresses EMC may have added including: system addresses and PST ingestion information, as well as X400/X500 addresses. Examples of addresses are listed below:  SYS:"ES1ExchJournal"<es1exchjournal>  EX:"Alex Chatzistamatis" </o=nuix/ou=na/cn=conshohocken/cn=achatz01>  PST:"archive.pst"<\\\\\\nuixfs01\\pst\\achatz01\\archive.pst>
Expand DL to:	"Expanded-DL"	Using the connected EMC SQL database, Nuix will query for any distribution list recipients on every email. Any responsive recipients will be added depending on the value selected in this dropdown.  <b>To: + "Expanded-DL"</b> = distribution list recipients will be added to the To: field AND the "Expanded-DL" field.  <b>Bcc: + "Expanded-DL"</b> = distribution list recipients will be added to the Bcc: field AND the "Expanded-DL" field.  <b>"Expanded-DL"</b> = distribution list recipients will be added to the "Expanded-DL" metadata field.

**Warning**

The **Expand DL to:** selection will only ADD the metadata to a new metadata field called **"Expanded-DL"**, however, based on your Output Type (PST, MSG or EML), the metadata may not be preserved unless **Add Distribution List Recipients** is enabled under **MAPI Export Options** or **EML Export Options** on the **Lightspeed Settings** tab in **Global Settings**.

9. Review the settings you've selected for this job and click the **Add Batch to Grid** button.

- a. If you have more mailbox archives you need to process, select a new list of users and click **Add Batch to Grid**, continuing this until all of

users have been added to a Batch on the Grid.

### Warning

Performance may vary, especially based on the settings configured on the **Lightspeed Settings** tab in **Global Settings**. Be sure to **ALWAYS** review settings prior to starting Jobs.

10. When you are ready to begin processing, click on the **Start Job** button.

## Working with HP/Autonomy Zantaz EAS

When working with HP/Autonomy Zantaz EAS source data, you can use NEAMM in various methods to target the data directly on the file system in its proprietary, .EAS file format (or on Centera when applicable) and export it out to disk in PST, MSG or EML format. Below are several workflows that NEAMM can handle.

### Overview

The Zantaz EAS archiving solution has two components:

- Files on disk
- SQL Database

Nuix must have access to the database while processing EAS files on disk in order to correctly parse the uniquely formatted EAS data.

### Prerequisites

The following prerequisites must be met before EAS data can be processed using Nuix:

1. EAS database restoration
  - A copy of the production EAS database must be restored to a local MSSQL instance on the Nuix server.
2. EAS database configuration
  - The restored EAS database must be configured per the instructions in the "EAS Database" section of this document.
3. Source data identified with a specific Docstore
  - If source data has been moved from the original archive storage location, it must be clearly marked as belonging to a specific Docstore.

## Source Data

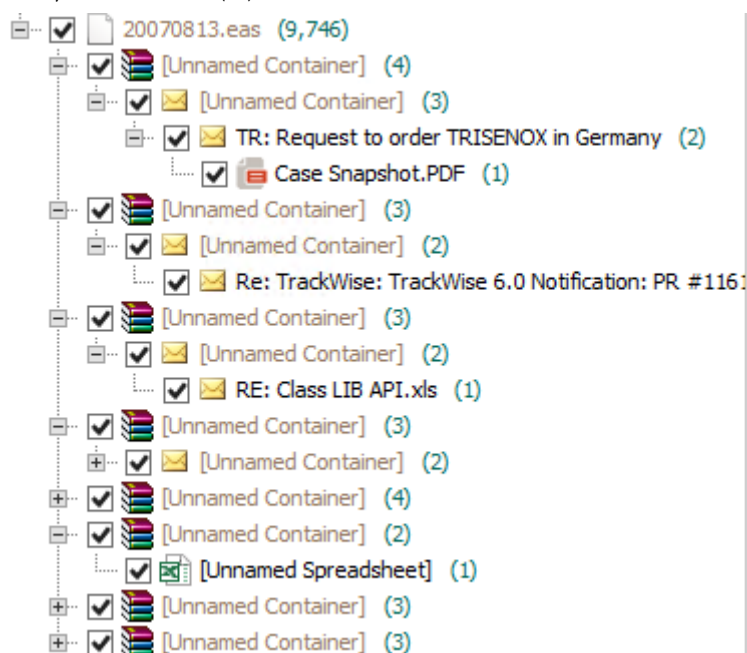
### Physical Files

At the storage locations indicated for each Docstore, Zantaz EAS data is written to the file system in the form of .EAS files. These files are compressed containers which house individual messages and attachments.

### File Structure

Typical EAS files contain single copies of each message with attachments, as seen

in (2) below. If single-instancing has been enabled, loose attachments may also be found in the container, such as (3) below.



## Single-Instancing

It is possible for EAS admins to enable true single-instancing in their archive. A quick look at the ContainsAttachment table in the database will confirm whether this option has been enabled at any point in the history of the archive. If this table contains values, single-instancing was previously enabled. If the table is blank, it was not.

As of Nuix 6.2.7, support for single-instancing is in place. Processing this data correctly requires that ALL EAS Docstores are correctly listed in dbo.DocumentServer and that they remain available to the Nuix server at all times. This is due to the fact that single-instancing can occur across Docstores. In other words, the attachment for an email in file 20120103.eas on Docserver 1 might reside in file 20110904.eas on Docserver 5, or it might live in 20121205.eas on Docserver 3. There is no way to know in advance.

As of Nuix 6, EAS data is automatically presented as individual user-based email. Default Nuix behavior is to process user-based copies for each email in an EAS container. This is necessary to ensure all metadata can be maintained at export, such as user folder structure or read/unread flags.





## File Distribution & Naming

Files are named in a date-based format YYYYMMDD.eas—for example, 20150108.eas. If multiple containers are created on the same date, an additional character will be added to subsequent files. So you may see something like the following on the filesystem: 20150108.eas, 20150108A.eas, 20150108B.eas.

Each active Docstore will archive one of these containers each day; therefore, it is expected that multiple individual EAS files will have the same filename. This is why Nuix can only process data from one Docstore in a single run and why we must provide the DocServerID of the target Docstore at the time of processing.

### Note

Typical best practice is to process all EAS data from the original storage locations. However, it is possible to move the data, provided the following conditions are met:

- Data from individual Docstores is always kept separate.
- Any moved files are identified with a specific Docstore.
- If single-instanting is in place, ALL data must be moved and presented to the Nuix server to ensure all email can be fully rehydrated.

## Centera

It is common for EAS deployments to be archived on Centera storage. In order to target data on a Centera device, you must produce a list of the C-Clips which correspond to your data.

## Prerequisites

The following prerequisites must be met before EMC Centera data can be processed using Nuix:

#### 1. **Pool Entry Authorization (PEA) File** - *if applicable*

- If the EMC Centera device has been configured with additional storage pools, a PEA file will need to be obtained and placed in a dedicated location on the Nuix server. A PEA file is an encrypted file used to communicate and distribute authentication credentials to Centera and contains the default key, key, and credential.

#### 2. **Environment User Variable** - *if applicable*

- If a PEA file is necessary, it must be obtained and placed in a dedicated location on the Nuix server. Next, an Environment Variable must be created on the Nuix server with a value of the PEA file path.

#### 3. **List of Centera Access Nodes (IPs)**

- The Centera Access Nodes are simply the IP addresses of the nodes available on Centera. This is typically a list of two or four IP addresses. If replication is enabled, more addresses may be available.

#### 4. **List of C-Clips**

- C-Clips, also known as Clips, are unique alphanumeric strings that reference specific source data within Centera. The source data varies depending on the archive; however, the concept is always the same. A list of C-Clips can be generated using several techniques and passed in to Nuix in order to process data.

### **Pool Entry Authorization**

A Pool Entry Authorization (PEA) file, generated while creating or updating an access profile, is a clear-text, XML-formatted, non-encrypted file that can be used by system administrators to communicate and distribute authentication credentials to application administrators. A PEA file is optional for profiles with non-encoded secrets (created using the File and Prompt options) but is mandatory for profiles with base-64 encoded secrets (created with the Generate option).

#### **Note**

A PEA file may not be necessary in all environments unless Centera has been specifically setup and configured to require one. For example, a Centera setup with only the default pool may not require a PEA file for authentication. If a PEA file is not needed for authentication, the Environment Variable is not necessary.

### **Configuration**

If the EMC Centera device has been configured with additional storage pools, a PEA file will need to be obtained and placed in a dedicated location on the Nuix server. An Environment Variable will also need to be created for Nuix to reference the PEA file.

1. Obtain the PEA file from the customer.

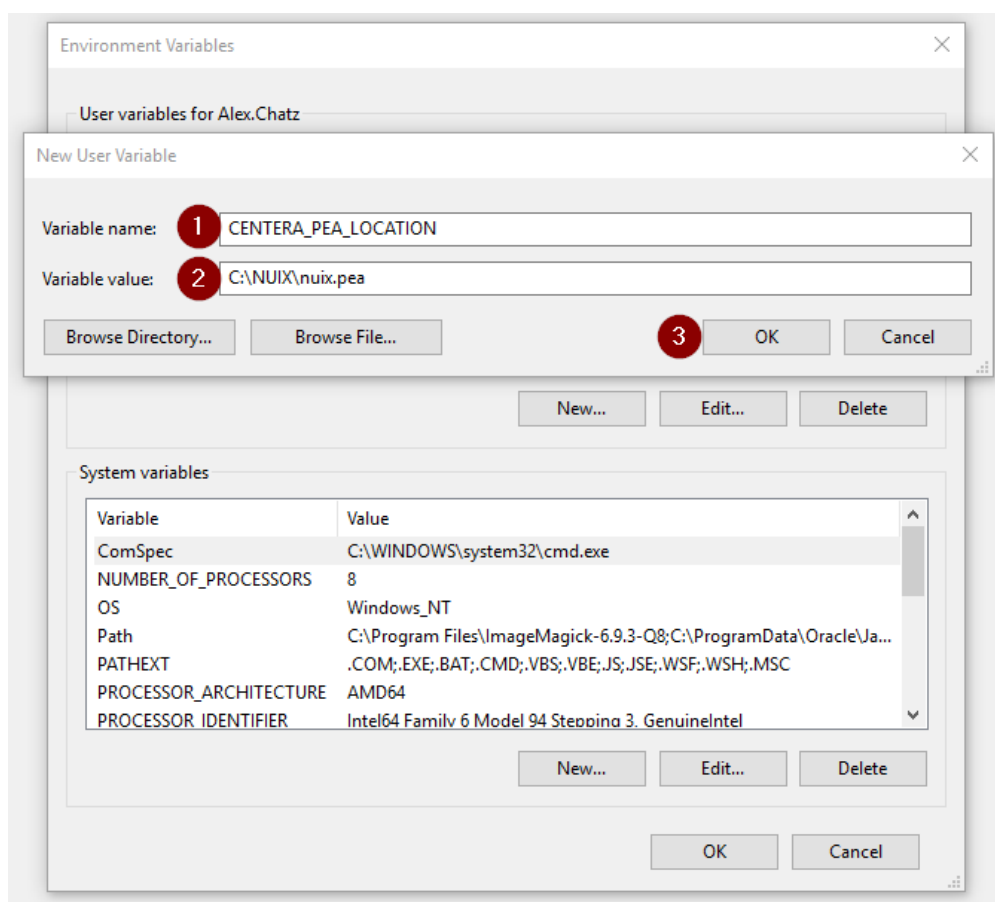
```
<pea version="1.0.0">
  <defaultkey name="nuix">
    <credential id="csp1.secret" enc="base64">bnVpeDEyMw==</credential>
  </defaultkey>
  <key type="cluster" id="1f778c7e-1dd2-11b2-ac5c-d3a0eb16f02e" name="nuix">
    <credential id="csp1.secret" enc="base64">bnVpeDEyMw==</credential>
  </key>
</pea>
```

2. Place the PEA file in a dedicated directory on the Nuix server.
3. Create an Environment Variable that references the location of the PEA file.

### Warning

It is critical that the Environment Variable is configured properly. Refer to the image below for more information.

- The Variable name must be: **CENTERA\_PEA\_LOCATION**
- The Variable value **must be the absolute path** to the physical location of the PEA file on the Nuix server
- Click OK to save the variable



### Access Node IP Addresses

An EMC Centera access node is a node that has the access role applied to it. Access nodes are gateways to the data stored in Centera. These nodes have IP addresses on the network and are responsible for authentication. If you successfully connect to one such node, you have access to the entire cluster. However, to connect faster, you can specify several available nodes with the access role.

### Configuration

The Centera access nodes are simply the IP addresses of the nodes available on Centera. This is typically a list of two or four IP addresses. If replication is

enabled on Centera, more addresses may be available.

After a list of Centera access nodes has been obtained from the customer, next steps include:

1. Create an .IPF (standard text file - "IP Address File") file with each line representing the IP of each Centera access node.

```
1 10.162.111.17
2 10.162.111.18
3 10.162.111.19
4 10.162.111.20
```

2. Place this .IPF file in a dedicated directory on the Nuix server.

- This .TXT file will be required when processing data on Centera.

### C-Clip List

Centera-based data is referenced by C-Clips, which must be passed in to Nuix at time of processing. Unless the client provides these Clips, you will need to generate them using the methods outlined in this section.

#### Generating the Clip List(s) for Zantaz EAS

1. Install the jre-6-27-windows-i586-s.exe instance of Java on a machine

- It can be a Nuix machine, but it doesn't have to be
- Install the package to a distinct path: C:\Nuix\Java

2. Unzip the JSCSScript-win32-3.2.35.zip into the C:\Nuix\Java\bin folder

3. Launch the command prompt

- Start | Run | Cmd <return>

4. Move to the root:

- cd \

5. Change to the Nuix\Java\bin directory:

- cd nuix\java\bin

6. Launch the EMC Centera API

- java -jar JCASScript.jar

7. If it launches, it will return something like:

- CASScript>

8. Connect to the desired Centera Pool. The following example uses a PEA file:

- poolOpen 10.10.13.203?C:\Nuix\Dxprofile.pea

9. If you are connected you will get a response that says:

- Connected to: 10.10.13.203?C:\Nuix\dxprofile.pea

10. Run the following command:

- **queryToFile** <filePathToClipOutput> <maxClipsToReturn>
- If a date filter needs to be applied, the **querySetLowerBound** and **querySetUpperBound** commands can be set first, followed by the **queryToFile** command.

#### Note

Generating a list of C-Clips is dependent on the number of pools that exist in Centera. If only the default pool exists, the list of C-Clips can only be generated for the default pool. Every additional pool that is created will have its own set of C-Clips associated with it.

#### Warning

Understanding the differences between legacy archive file types is critical when processing the same archive data on Centera. For example, processing EMC EmailXtender or EMC SourceOne C-Clips is fundamentally different than processing Symantec EV C-Clips.

A single EmailXtender C-Clip is equivalent to one .EMX file, which could contain hundreds or thousands of top-level emails, whereas a single Enterprise Vault C-Clip is equivalent to one .DVS file which can only contain a single top-level email.

Processing 10,000 EmailXtender C-Clips would take substantially longer than processing 10,000 Enterprise Vault C-Clips.

- The clip list can now be carved up into manageable sets of .CLP files.

## Database

A Zantaz EAS installation only requires a single database. In order for Nuix to correctly interact with the EAS database, several conditions have to be met. Namely, certain tables require a specific schema, and it is likely that at least one table will need to be edited. For this reason<sup>7</sup> it is required that a COPY of the EAS database be used (to be restored locally on the Nuix server), rather than the production SQL instance.

### Authentication

The credentials used to access SQL will be passed in via the Nuix startup file. It is recommended that the account provided uses built-in SQL authentication. Since the SQL instance is local and used only for Nuix processing, we recommend simply using the 'sa' (SQL Administrator) account for Nuix access.

---

<sup>7</sup> It is also best practice to avoid interacting with the production database to minimize impact on the client environment and maximize performance of Nuix processes.

## Configuration

The DocumentServer table must be edited to correctly reflect the path to all source data.

### Schema

Ensure the following tables use the **dbo** schema:

- Distlist
- DistlistRef
- DocumentServer
- EmailAddresses
- EmailMessages
- Folder
- Refer
- Users

Ensure the following tables use the **easadmin** schema:

- DataArchive
- ProfileLocation

## Critical Tables

It is important to understand the contents of several tables and the way Nuix uses these tables to interpret EAS data.

### easadmin.DataArchive

The DataArchive table contains three columns: DataArchiveID, Filename, and DocServerID. In order for Nuix to correctly parse an EAS file, it must retrieve the correct DataArchiveID for each EAS file being processed from this table. Since filenames can be repeated between Docstores, the DocServerID passed in at startup provides the cross-reference to ensure Nuix retrieves the correct DataArchiveID.

	DATAARCHIVEID	FILENAME	DOCSEVERID
1	7905	20031006.eas	1
2	7906	20050128.eas	1
3	7907	20061217.eas	1
4	7908	20040818.eas	1

### easadmin.ProfileLocation

The ProfileLocation is one of two tables in the EAS database that contain an item-level listing of every message in the archive (dbo.Refer is the other).

	MSGID	DATAARCHIVEID	STARTINGPOS	COMPRESSEDSize	LASTACCESSED	UNCOMPRESSI
1	67589	7905	0	1768	2007-02-23 15:06:13.000	11264
2	67881	7905	1768	2120	2007-02-23 15:07:06.000	8704
3	67882	7905	3888	1960	2007-02-23 15:07:06.000	9728
4	67883	7905	5848	3029	2011-07-27 14:11:57.000	9728
5	67884	7905	8877	3744	2011-10-26 18:19:30.000	10752

The ProfileLocation table is a critical path element for Nuix to process EAS data. The StartingPos and CompressedSize columns tell Nuix where each individual email (or attachment) begins and ends in the compressed EAS container. Without this, only the first message in any given EAS file can be processed.

The UncompressedSize column is useful for queries that seek to retrieve volume totals for user archives. This can be helpful in estimating export volume prior to beginning a Nuix run.

## Supported Workflows

### Journal Archive to User PST

1. In the **Legacy Archive** dropdown box, select **HP/Autonomy EAS**
2. Select **Journal** from the **Archive Type** radio button selections
3. Select **User** from the **Output Type** radio button selections
4. Select **PST** as the **Lightspeed Extraction Output**

#### Note

It is highly recommend that you select **PST** when your **Output Type** is set to **User**. PSTs are much easier to manage since emails will be grouped by User on the file system, and long file path names issues will be avoided.

5. If a date filter is requested, enter a date in the **From Date:** and **To Date:** date chooser field.
6. Choose whether the Source Data exists in **Folders**, **Files** or on **Centera**.
  - a. If **Folders** or **Files**, use the navigation pane to select the drive letter where the source data exists. You may also choose to **Compute Batch Size**. This will allow you to obtain additional metrics while the job is processing, like *Percentage Completed* and *Total Bytes* to process.

#### Tip

It is recommend that for each Job, target **no more** than 500 GB - 1 TB of compressed, source data per Job. This is for optimal performance as well as an advantage in system failure scenarios. If a system crashes mid-processing, it would be more efficient to restart a Job that will take 24 hours to complete, versus a Job that may take 7 days to complete.

- b. If **Centera** is selected, select the **Clip List** (\*.CLP) you want to process, browse to the **PEA File** (\*.PEA) and browse to the **IP File** (\*.IPF).

**Tip**

It is recommend that for each Job, target **no more** than 10,000 Clips of compressed data per Job. This is for optimal performance as well as an advantage in system failure scenarios. If a system crashes mid-processing, it would be more efficient to restart a Job that will take 24 hours to complete, versus a Job that may take 7 days to complete.

7. Enter the SQL connection string details in the **SQL Connection Info** settings. Be sure to click **Test SQL Connection** to ensure you can properly connect to SQL.

**Warning**

Testing the SQL is critical, as you will want to ensure that you can correctly query the SQL databases for additional information.

**Tip**

It is recommended to use a built-in SQL service account (instead of Windows authentication) with **READ** access to **ALL** of the EAS databases.

8. The EAS Settings will default to:

Name	Default Value	Description
Doc Server ID	none	An integer value must be entered which correlates to the EAS DocStore that is being selected for each Job.

**Warning**

For emails that were archived from a journal, Nuix will add the distribution list recipients to a new metadata field called "**Expanded-DL**", however, based on your Output Type (PST, MSG or EML), the metadata may not be preserved unless **Add Distribution List Recipients** is enabled under **MAPI Export Options** or **EML Export Options** on the **Lightspeed Settings** tab in **Global Settings**.

9. The Worker Side Script (WSS) section will allow you to:

Name	Default Value	Description
Exclude Unresponsive Items	True	<p>When enabled, Nuix will not export any items that do not respond to your Search Terms and Mapping CSV.</p> <p>If you set to <b>False</b>, you will need to adjust your Mapping CSV to include an entry for: <i>unresponsive,unresponsive.pst</i></p>



Verbose Logging	False	<p>When enabled, Nuix will not include any verbose logging at the WSS-level for troubleshooting purposes.</p> <p>If set to <b>True</b>, this will allow for easier troubleshooting, however, the size of the logs will be substantially larger.</p>
Content Filtering	Email, RSS Feed, Calendar, Contact	<p>When enabled, all top-level emails (RSS Feeds included), Calendar and Contact items will be extracted.</p> <p>If you want to filter any of these kinds out, simply de-select the item kind you wish to filter.</p>
Search Terms CSV	empty	<p>You must select a list of Search Terms in CSV format. <b>Search Terms</b> is a 2 column CSV that includes a Flag in Column A and a Search Term in Column B. You do not need a header column.</p> <p>Each line should include an SMTP address, X400 or X500 address similar to this. You can also add multiple Search Terms to the same Flag, for example:</p> <p><a href="#">Alex.Chatzistamatis,alex.chatzistamatis@nuix.com</a>  <a href="#">Alex.Chatzistamatis,alex@nuix.com</a></p> <p>Nuix will scan for these search terms across <b>Communication Metadata</b> (From, To, Cc, Bcc) and the <b>Expanded-DL</b> metadata field.</p>
Mapping CSV	empty	<p>You must select a Mapping in CSV format. Like <b>Search Terms</b>, <b>Mapping</b> is a 2 column CSV that includes a Flag Name in Column A and an Output Name in Column B. You do not need a header column.</p> <p>Each line should include the Flag from your Search Terms CSV and the Output Name. If you add multiple Search Terms to a single Flag, you do not need to add multiple Flags in the Mapping CSV, for example:</p> <p>Alex.Chatzistamatis,Alex.Chatzistamatis.pst</p>

10. Review the settings you've selected for this job and click the **Add Batch to Grid** button.
  - a. If you have more mailbox archives you need to process, select a new list of users and click **Add Batch to Grid**, continuing this until all of users have been added to a Batch on the Grid.

#### Warning

Performance may vary, especially based on the settings configured on the **Lilghtspeed Settings** tab in **Global Settings**. Be sure to **ALWAYS** review settings prior to starting Jobs.

11. When you are ready to begin processing, click on the **Start Job** button.

### Journal Archive to Flat PST, MSG or EML

1. In the **Legacy Archive** dropdown box, select **HP/Autonomy EAS**
2. Select **Journal** from the **Archive Type** radio button selections
3. Select **Flat** from the **Output Type** radio button selections
4. Select **PST, MSG or EML** as the **Lightspeed Extraction Output**

#### Note

It is highly recommend that you select **PST** when your **Output Type** is set to **User**. PSTs are much easier to manage since emails will be grouped by User on the file system, and long file path names issues will be avoided.

5. If a date filter is requested, enter a date in the **From Date:** and **To Date:** date chooser field.
6. Choose whether the Source Data exists in **Folders, Files** or on **Centera**.
  - a. If **Folders** or **Files**, use the navigation pane to select the drive letter where the source data exists. You may also choose to **Compute Batch Size**. This will allow you to obtain additional metrics while the job is processing, like *Percentage Completed* and *Total Bytes to process*.

#### Tip

It is recommend that for each Job, target **no more** than 500 GB - 1 TB of compressed, source data per Job. This is for optimal performance as well as an advantage in system failure scenarios. If a system crashes mid-processing, it would be more efficient to restart a Job that will take 24 hours to complete, versus a Job that may take 7 days to complete.

- b. If **Centera** is selected, select the **Clip List** (\*.CLP) you want to process, browse to the **PEA File** (\*.PEA) and browse to the **IP File** (\*.IPF).

#### Tip

It is recommend that for each Job, target **no more** than 10,000 Clips of compressed data per Job. This is for optimal performance as well as an advantage in system failure scenarios. If a system crashes mid-processing, it would be more efficient to restart a Job that will take 24 hours to complete, versus a Job that may take 7 days to complete.

7. Enter the SQL connection string details in the **SQL Connection Info** settings. Be sure to click **Test SQL Connection** to ensure you can properly connect to SQL.

#### Warning

Testing the SQL is critical, as you will want to ensure that you can correctly query the SQL databases for additional

information.

**Tip**

It is recommended to use a built-in SQL service account (instead of Windows authentication) with **READ** access to **ALL** of the EAS databases.

8. The EAS Settings will default to:

Name	Default Value	Description
Doc Server ID	none	An integer value must be entered which correlates to the EAS DocStore that is being selected for each Job.

**Warning**

For emails that were archived from a journal, Nuix will add the distribution list recipients to a new metadata field called "**Expanded-DL**", however, based on your Output Type (PST, MSG or EML), the metadata may not be preserved unless **Add Distribution List Recipients** is enabled under **MAPI Export Options** or **EML Export Options** on the **Lightspeed Settings** tab in **Global Settings**.

9. Review the settings you've selected for this job and click the **Add Batch to Grid** button.

- a. If you have more mailbox archives you need to process, select a new list of users and click **Add Batch to Grid**, continuing this until all of users have been added to a Batch on the Grid.

**Warning**

Performance may vary, especially based on the settings configured on the **Lightspeed Settings** tab in **Global Settings**. Be sure to **ALWAYS** review settings prior to starting Jobs.

10. When you are ready to begin processing, click on the **Start Job** button.

# Working with Daegis AXS-One

When working with Daegis AXS-One source data, you can use NEAMM in various methods to target the data directly on the file system in its proprietary, .PGI file format and export it out to disk in PST, MSG or EML format. Below are several workflows that NEAMM can handle.

## Overview

The Daegis AXS-One archiving solution has two components:

- DATA files on disk
- SIS files on disk

Nuix must have access to the DATA and SIS store pairs while processing AXS-One files on disk in order to correctly parse the uniquely formatted AXS-One data. AXS-One does not store any critical information related to the archive in a relational database, like SQL.

## Prerequisites

The following prerequisites must be met before AXS-One data can be processed using Nuix:

1. Source data identified with a specific DATA and SIS store pairs.
  - If source data has been moved from the original archive storage location, it must be clearly marked as belonging to a specific DATA and SIS store pair.

## Physical Files

It is common for an AXS-One installation to archive emails to designated partitions called "DATA" and "SIS" stores. Together, these stores can be referred to as a "pair" or "DATA/SIS pair." Each DATA and SIS store created by the administrators must have a unique storage location and can have unique settings applied to it. Generally, the DATA partition contains the emails themselves, while the SIS partition contains any single-instanced data. DATA and SIS stores can be "paired" with each other, or multiple DATA stores can have one or more SIS stores.

Name	Date modified	Type
 DATA99	1/15/2016 4:37 PM	File folder
 SIS99	1/15/2016 4:37 PM	File folder

In order to correctly process AXS-One data, the paths for each SIS store must be presented to Nuix. Therefore, one or more SIS store must be available locally or mapped to the Nuix server.

The path to each SIS store must be passed to Nuix at startup via the following switch (if more than one SIS volume exists, be sure to separate them with a pipe (|) followed by the next value):

```
nuix.data.axsone.sisFolders="T:\AXSONE\SIS99|U:\AXSONE\SIS22"
```

At the storage locations indicated for each DATA store, Daegis AXS-One data is written to the file system in the form of a "set" of files. In this "set," the .PGI

and the .DCM are the most critical for processing. The .PGI file is essentially a pointer or map to the .DCM file. The .DCM file is a compressed container that houses one or more messages and their attachments. Multiple "sets" may exist for data that has archived content. A "set" is easily identifiable when files all share the same base name.

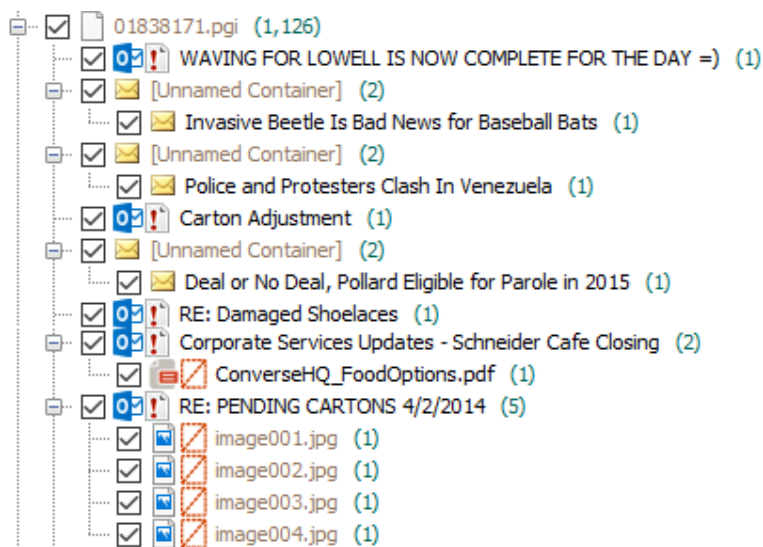
There are other files created by AXS-One that may exist in the "set" such as in the screenshot below; however, they are not relevant.

- .\_HDR files contain header information from all messages within the .DCM file.
- .00n files are companion files to the .DCM.
- .LCZ files, if present, are small Lucene text indexes
- .SIS files contain single-instanced data, all associated to the data stored within the .DCM

Name	Date modified	Type	Size
00120392._hdr	11/18/2010 8:21 AM	._HDR File	67 KB
00120392.002	11/18/2010 8:21 AM	002 File	318 KB
00120392.003	11/18/2010 8:22 AM	003 File	136 KB
00120392.dcm	11/18/2010 8:22 AM	DCM File	38,350 KB
00120392.pgi	11/18/2010 8:22 AM	PGI File	25 KB
00120392.sis	11/18/2010 8:22 AM	SIS File	1 KB

## File Structure

Typical AXS-One files contain single copies of each message with attachments, as seen below. If single-instancing has been enabled, loose attachments may also be found in the container, such as below.



## Single Instancing

It is possible for AXS-One admins to enable single-instancing in their archive. When enabled, a SIS store will exist alongside the DATA store.

As of Nuix 6.0, support for single-instancing is available. Processing this data correctly requires that ALL AXS-One SIS are presented to the Nuix Server and their locations have been passed into Nuix using the Nuix startup switch that was previously discussed in the "AXS-One Source Data" section of this document.

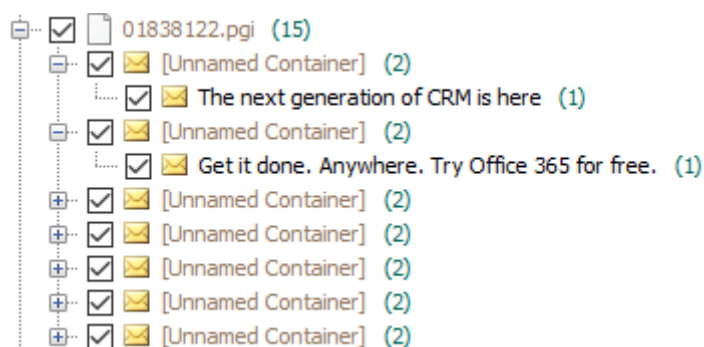
The DATA folder is where the archived messages primarily are stored. The SIS folder is where the single-instanced attachments are stored. Nuix reads the hash of the attachment from the email and attempts to find it in the paired SIS folder.

- The SIS folder has a structure where the first few characters of the hash of the attachment are used to nest the files.
  - » For example: "W665MEFS6RUYEWYN3IHYHP2UNS.sis" would be found in ..\SISxx\2014\W6\65\ME.
- The .sis\_log file, located in the directory mentioned above, will indicate the documents that reference this single-instanced item and their location in the DATA store.

As long as the switch is enabled when Nuix is loaded with the SIS locations, Nuix has the necessary logic hard-coded that will automatically find any single-instanced attachments and rehydrate them with the appropriate message.

## File Distribution & Naming

As of Nuix 6, AXS-One data is automatically presented as it was archived. Default Nuix behavior is to process a single copy of each message in an AXS-One container. This is necessary to ensure all relevant metadata can be maintained at export. Identifying the mailbox that the message was archived in as well as other relevant information can be identified using the AXS-One metadata which Nuix will gather from the AXS-One wrappers [Unnamed Container].



Folders are created in a date-based format (YYYYMMDD). AXS-One file "sets" are named in a hash-based format. For example, a file path could look like C:\DATA99\20150201\00840119.pgi. If multiple "sets" are created on the same date, all subsequent files will have a uniquely named hash. If multiple DATA stores exist, multiple "sets" may get created, each uniquely named.

### Note

Typical best practice is to process all AXS-One data from the original storage locations. However, it is possible to move the data, provided the following conditions are met:

- Data from individual DATA stores is always kept separate.
- Any moved files are identified with their specific DATA store and folder names are maintained.
- If single-instancing is in place, SIS stores can either be left in their original location or ALL data must be moved. The exact folder structure must be maintained to ensure all single-instanced data is rehydrated correctly.

## Supported Workflows

### Mailbox Archive to User PST

1. In the **Legacy Archive** dropdown box, select **Daegis AXS-One**
2. Select **Journal** from the **Archive Type** radio button selections
3. Select **User** from the **Output Type** radio button selections
4. Select **PST** as the **Lightspeed Extraction Output**

#### Note

It is highly recommend that you select **PST** when your **Output Type** is set to **User**. PSTs are much easier to manage since emails will be grouped by User on the file system, and long file path names issues will be avoided.

5. If a date filter is requested, enter a date in the **From Date:** and **To Date:** date chooser field.
6. Choose whether the Source Data exists in **Folders** or **Files**.
  - a. If **Folders** or **Files**, use the navigation pane to select the drive letter where the source data exists. You may also choose to **Compute Batch Size**. This will allow you to obtain additional metrics while the job is processing, like *Percentage Completed* and *Total Bytes* to process.

#### Tip

It is recommend that for each Job, target **no more** than 500 GB - 1 TB of compressed, source data per Job. This is for optimal performance as well as an advantage in system failure scenarios. If a system crashes mid-processing, it would be more efficient to restart a Job that will take 24 hours to complete, versus a Job that may take 7 days to complete.

#### Warning

Daegis AXS-One is currently not supported on EMC Centera storage platforms.

Please see your Nuix representative for more details.

7. Select the location of the AXS-One SIS folders:

- a. One or more SIS folders can be selected.

**Note**

Since Daegis AXS-One does not utilize a relational database like SQL, there is no requirement for connecting to a SQL database, however, it is critical to select the correct **SIS** folders that match the appropriate **DATA** folders.

8. The AXS-One Settings will default to:

Name	Default Value	Description
Skip SIS Lookups	False	If enabled, Nuix will skip SIS lookups and attempt to process the .PGI file directly.  This should be enabled if SIS folders are not enabled or there are issues retrieving attachments from the SIS folders.

9. The Worker Side Script (WSS) section will allow you to:

Name	Default Value	Description
Exclude Unresponsive Items	True	When enabled, Nuix will not export any items that do not respond to your Search Terms and Mapping CSV.  If you set to <b>False</b> , you will need to adjust your Mapping CSV to include an entry for: <i>unresponsive,unresponsive.pst</i>
Verbose Logging	False	When enabled, Nuix will not include any verbose logging at the WSS-level for troubleshooting purposes.  If set to <b>True</b> , this will allow for easier troubleshooting, however, the size of the logs will be substantially larger.
Content Filtering	Email, RSS Feed, Calendar, Contact	When enabled, all top-level emails (RSS Feeds included), Calendar and Contact items will be extracted.  If you want to filter any of these kinds out, simply de-select the item kind you wish to filter.
Search Terms CSV	empty	You must select a list of Search Terms in CSV format. <b>Search Terms</b> is a 2 column CSV that includes a Flag in Column A and a Search Term in Column B. You do not need a header column.  Each line should include an SMTP address, X400 or X500 address similar to this. You can also add multiple Search Terms to the same Flag, for example:  <a href="mailto:Alex.Chatzistamatis@nuix.com">Alex.Chatzistamatis,alex.chatzistamatis@nuix.com</a>



		<a href="mailto:Alex.Chatzistamatis,alex@nuix.com">Alex.Chatzistamatis,alex@nuix.com</a> Nuix will scan for these search terms across <b>Communication Metadata</b> (From, To, Cc, Bcc) and the <b>Expanded-DL</b> metadata field.
Mapping CSV	empty	You must select a Mapping in CSV format. Like <b>Search Terms, Mapping</b> is a 2 column CSV that includes a Flag Name in Column A and an Output Name in Column B. You do not need a header column.  Each line should include the Flag from your Search Terms CSV and the Output Name. If you add multiple Search Terms to a single Flag, you do not need to add multiple Flags in the Mapping CSV, for example:  Alex.Chatzistamatis,Alex.Chatzistamatis.pst

10. Review the settings you've selected for this job and click the **Add Batch to Grid** button.
  - a. If you have more mailbox archives you need to process, select a new list of users and click **Add Batch to Grid**, continuing this until all of users have been added to a Batch on the Grid.

#### Warning

Performance may vary, especially based on the settings configured on the **Lightspeed Settings** tab in **Global Settings**. Be sure to **ALWAYS** review settings prior to starting Jobs.

11. When you are ready to begin processing, click on the **Start Job** button.

### Mailbox Archive to Flat PST, MSG or EML

1. In the **Legacy Archive** dropdown box, select **Daegis AXS-One**
2. Select **Journal** from the **Archive Type** radio button selections
3. Select **Flat** from the **Output Type** radio button selections
4. Select **PST, MSG or EML** as the **Lightspeed Extraction Output**

#### Note

It is highly recommend that you select **PST** when your **Output Type** is set to **User**. PSTs are much easier to manage since emails will be grouped by User on the file system, and long file path names issues will be avoided.

5. If a date filter is requested, enter a date in the **From Date:** and **To Date:** date chooser field.
6. Choose whether the Source Data exists in **Folders** or **Files**.
  - a. If **Folders** or **Files**, use the navigation pane to select the drive letter where the source data exists. You may also choose to **Compute Batch**

**Size.** This will allow you to obtain additional metrics while the job is processing, like *Percentage Completed* and *Total Bytes* to process.

**Tip**

It is recommend that for each Job, target **no more** than 500 GB - 1 TB of compressed, source data per Job. This is for optimal performance as well as an advantage in system failure scenarios. If a system crashes mid-processing, it would be more efficient to restart a Job that will take 24 hours to complete, versus a Job that may take 7 days to complete.

**Warning**

Daegis AXS-One is currently not supported on EMC Centera storage platforms.

Please see your Nuix representative for more details.

7. Select the location of the AXS-One SIS folders:

- a. One or more SIS folders can be selected.

**Note**

Since Daegis AXS-One does not utilize a relational database like SQL, there is no requirement for connecting to a SQL database, however, it is critical to select the correct **SIS** folders that match the appropriate **DATA** folders.

8. The AXS-One Settings will default to:

Name	Default Value	Description
Skip SIS Lookups	False	If enabled, Nuix will skip SIS lookups and attempt to process the .PGI file directly.  This should be enabled if SIS folders are not enabled or there are issues retrieving attachments from the SIS folders.

9. Review the settings you've selected for this job and click the **Add Batch to Grid** button.

- a. If you have more mailbox archives you need to process, select a new list of users and click **Add Batch to Grid**, continuing this until all of users have been added to a Batch on the Grid.

**Warning**

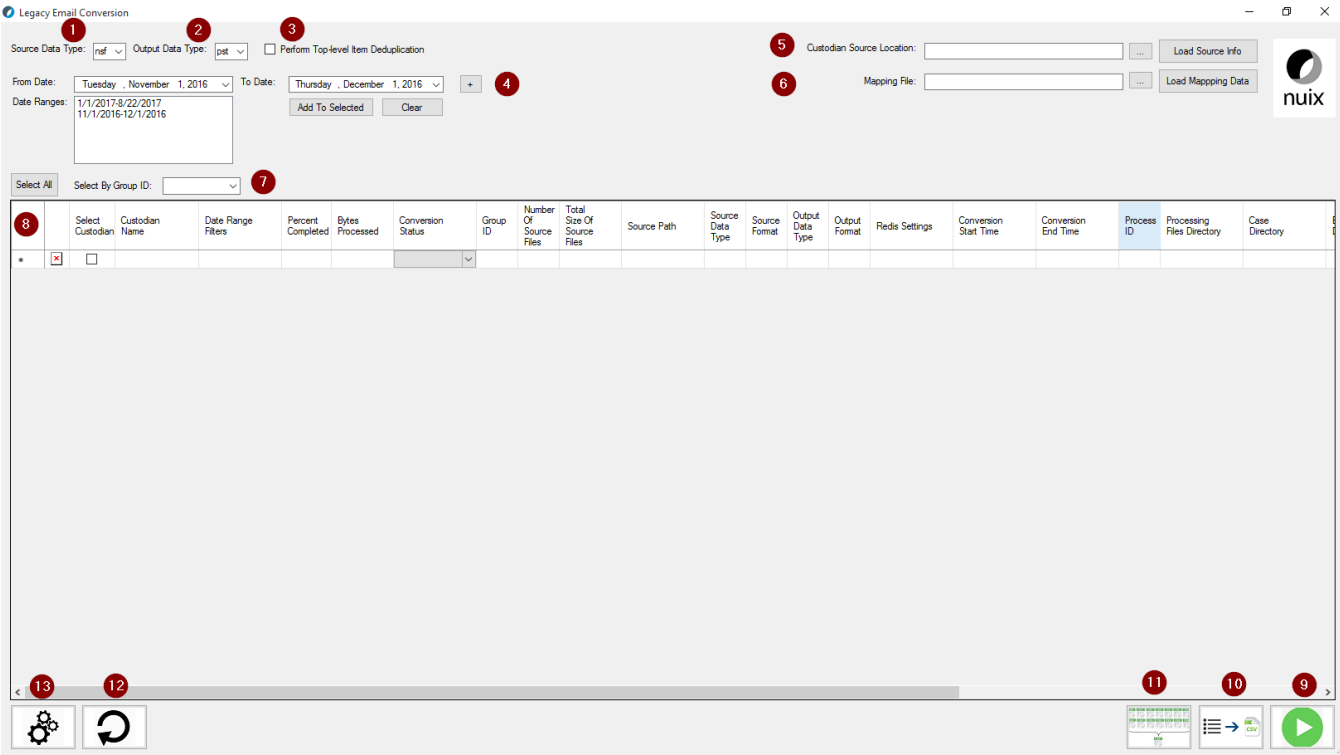
Performance may vary, especially based on the settings configured on the **Lightspeed Extraction** tab in **Global Settings**. Be sure to **ALWAYS** review settings prior to starting Jobs.

10. When you are ready to begin processing, click on the **Start Job** button.

# Converting Legacy Email Data

## Interface Overview

Upon launching the Email Conversion module, you will see the interface as shown below. Many of these options in this interface will be enabled/disabled based on selections made. This is by design several may options may not be relevant or necessary for specific source archives or workflows.



	Name	Description
1	Source Data Type	Select the format of the source data (NSF).
2	Output Data Type	Select the format of the final conversion (PST, MSG or EML)
3	Perform Top-Level Item Deduplication	Use Redis to perform a global deduplication based on the MD5 hash of the top-level item.
4	From/To Date	Used to filter the email data by date.
5	Custodian Source Location	Select the location of the Source Data.
6	Mapping File	Select the Mapping File to be used by Lightspeed.
7	Select / Select By Group ID	Select all jobs in Grid or Select by Group ID
8	Grid	Where added jobs will be displayed.

9	Start Job	Start the selected job in the grid for processing.
10	Export Grid to CSV	Exports out the current grid view to CSV format.
11	Lightspeed Exporter Report Consolidator	Will consolidate all Lightspeed Exporter Metrics and Exporter Error into a single report.
12	Reload Grid	Reloads jobs in the grid from previous migrations.
13	Global Settings	View/Change previously configured Global Settings.

### Tip

Before attempting to perform any migration work, be sure to check **Global Settings** and make sure that the **Nuix Directories**, **Lightspeed Settings** and **Database** tabs are configured correctly.

## Performing a User NSF to User PST Conversion

1. In the **Source Data Type** dropdown box, select **NSF**
2. Select **PST** as the **Output Data Type**
3. If top-level item deduplication is necessary, enable the **Perform Top-Level Item Deduplication** checkbox.

### Note

Performing Top-Level Item Deduplication requires the use of Redis.

Redis is not installed or configured by NEAMM. This must be setup prior to using NEAMM. Please consult with your Nuix representative to understand if this is necessary for your project.

4. Select the directory where your NSFs are stored in the **Custodian Source Location** field and click **Load Source Info**.
5. Select the appropriate **Mapping File** and click **Load Mapping Data** to map your NSF to PST output.
6. If a date filter is requested, enter a date in the **From Date:** and **To Date:** date chooser field.
  - a. You may add multiple date ranges using "+" sign.
7. To begin extracting all of the selected EWS Mailboxes, click **Select All**, otherwise, **Select By Group ID** from the dropdown.

### Warning

Performance may vary, especially based on the settings configured on the **Lightspeed Settings** tab in **Global Settings**. Be sure to **ALWAYS** review settings prior to starting Jobs.

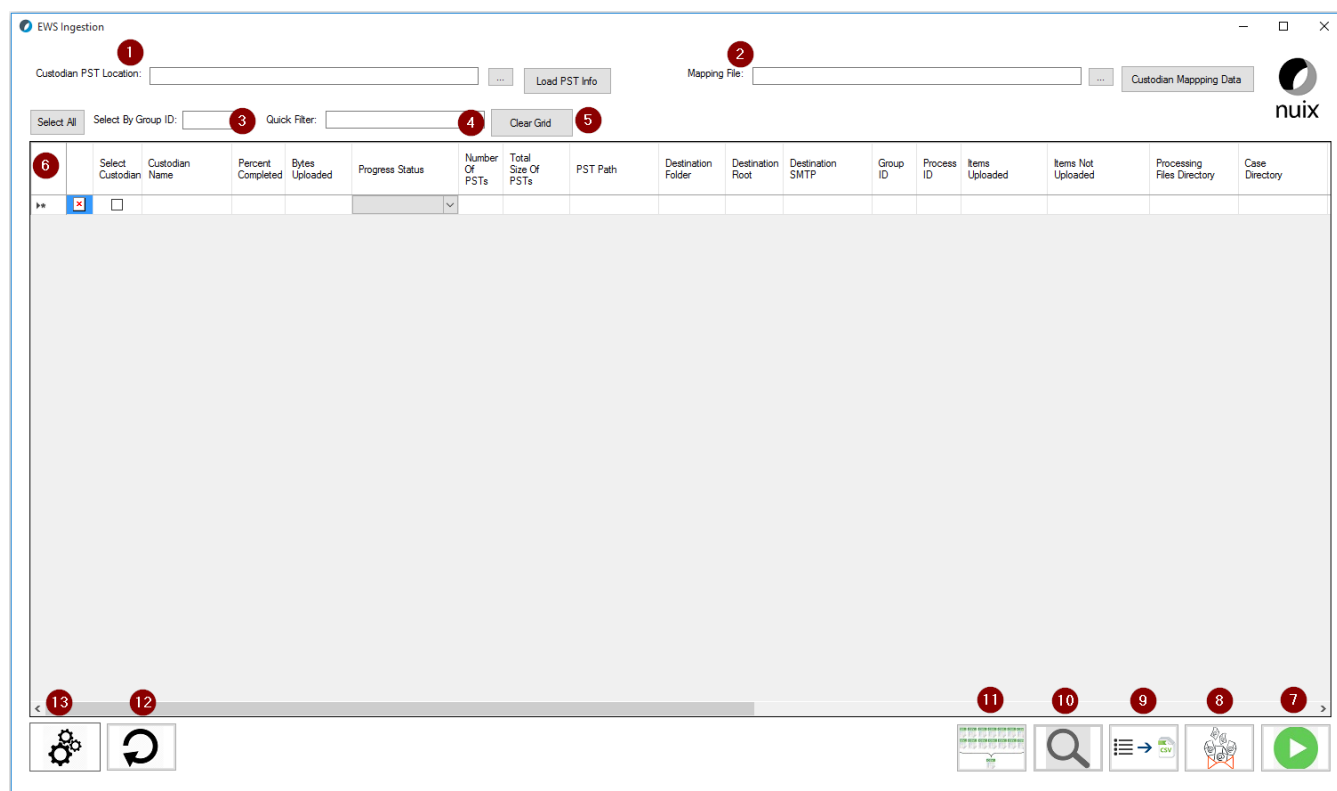
8. If a date filter is requested, enter a date in the **From Date:** and **To Date:** date chooser field.

- a. You may add multiple date ranges using "+" sign.
  - b. Click **Add to Selected** to add your date ranges to the selected batches.
9. When you are ready to begin processing, click on the **Start Job** button.

# Ingesting Email Data into Exchange

## Interface Overview

Upon launching the EWS Ingestion module, you will see the interface as shown below. Many of these options in this interface will be enabled/disabled based on selections made. This is by design several may options may not be relevant or necessary for specific source archives or workflows.



	Name	Description
1	PST Location	Select the PSTs from the File System
2	Mapping File	Select the Lightspeed Mapping File
3	Select / Select By Group ID	Select all jobs in Grid or Select by Group ID
4	Quick Filter	Search filter on any column in the grid
5	Clear	Clears the grid

6	Grid	Where added jobs will be displayed
7	Start Job	Start the selected job in the grid for processing
8	Export Exceptions to PST	Export items not uploaded to EWS to PST format
9	Export Grid to CSV	Exports out the current grid view to CSV format
10	Show Processing Details	Shows success and exception details per custodian
11	Lightspeed Exporter Report Consolidator	Will consolidate all Lightspeed Exporter Metrics and Exporter Error into a single report
12	Reload Grid	Reloads jobs in the grid from previous migrations
13	Global Settings	View/Change previously configured Global Settings

**Tip**

Before attempting to perform any migration work, be sure to check **Global Settings** and make sure that the **Nuix Directories** and **Exchange Web Services** tab are configured correctly.

## Working with Exchange Web Services

### Overview

Microsoft's Exchange email solution allows access to software vendors using their Exchange Web Services (EWS) API. Using this API, Microsoft provides the ability to either extract or ingest data into EWS. An Exchange mailbox consists of two primary components:

- Mailbox
  - Information Store
  - Recoverable Items
    - Purges
    - Deletions
- Archive
  - Information Store
  - Recoverable items
    - Purges
    - Deletions

In order to connect to an EWS mailbox, you must first connect with a user or service account. After successful authentication, Nuix will be able to extract or ingest data.

## Prerequisites

The following prerequisites should be met in order for Nuix to interact with EWS:

1. EWS Service Account created
  - Username (SMTP address) and Password
2. EWS environment prepared
  - A valid EWS SMTP address must be available for each mailbox that will be targeted
  - If targeting an EWS personal archive, be sure that it is online and available to the mailbox

## Source Data

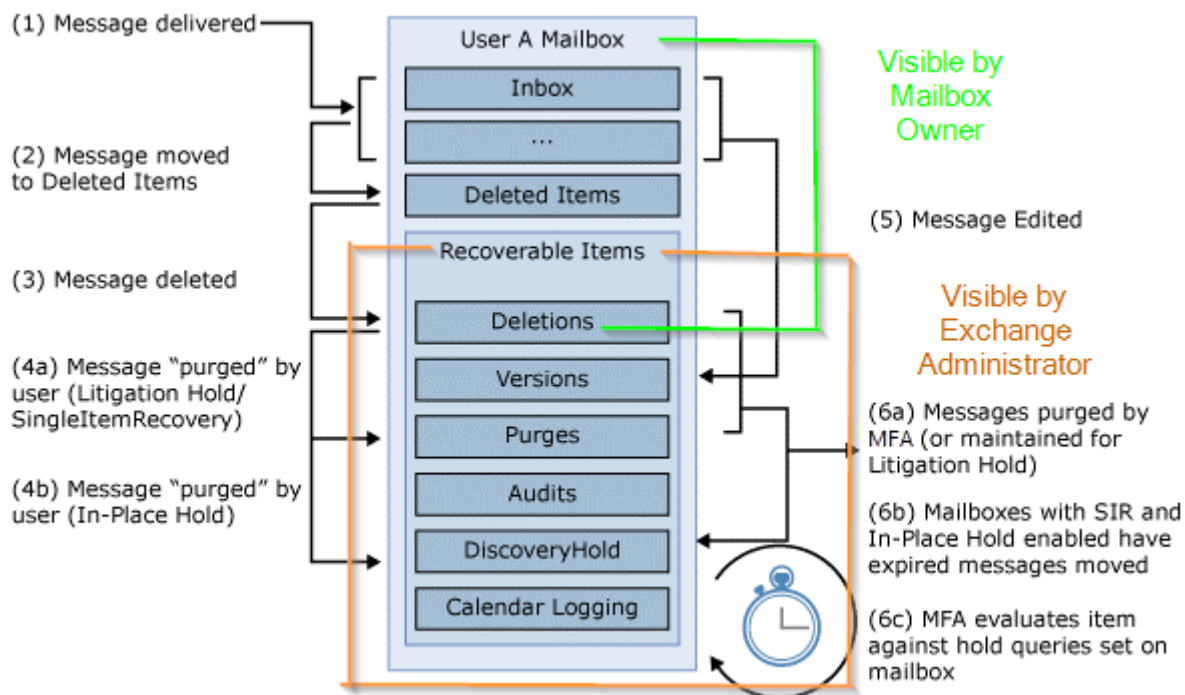
An Exchange mailbox or archive includes two different "partitions":

1. The first "partition" is the standard mailbox folder that is exposed to the users such as: Inbox, Deleted items, Sent Items, Drafts, any customer folder a user can create and more.
2. The second "partition" is the Recoverable Items folder partition which may include: Deletions, Versions, Purges, Audits, DiscoveryHold and Calendar Logging.

Only Exchange Administrators can access and/or view all of the different folders in Recoverable Items. The Recoverable Items folder is hidden from users. The only exception to this rule, is a specific folder in Recoverable Items called "Deletions". This "Deletions" folder does not appear in the standard mailbox folder that the mailbox owner can see, however, the mailbox owner can access the content of this folder by using the "Recover Deletion Items" option in Outlook.

The Recoverable Items folder contains the following subfolders:

- **Deletions** - this subfolder contains all items deleted from the Deleted Items folder. (In Outlook, a user can permanently delete an item by pressing Shift+Delete.) This subfolder is exposed to users through the Recover Deleted Items feature in Outlook and Outlook Web App.
- **Versions** - if In-Place Hold or Litigation Hold is enabled, this subfolder contains the original and modified copies of the deleted items. This folder isn't visible to end users.
- **Purges** - if either Litigation Hold or single item recovery is enabled, this subfolder contains all items that are purged. This folder isn't visible to end users.
- **Audits** - if mailbox audit logging is enabled for a mailbox, this subfolder contains the audit log entries. To learn more about mailbox audit logging, see Mailbox audit logging.
- **DiscoveryHolds** - if In-Place Hold is enabled, this subfolder contains all items that meet the hold query parameters and are purged.
- **Calendar Logging** - this subfolder contains calendar changes that occur within a mailbox. This folder isn't available to users.



## Access Requirements

In order for Nuix to ingest data to EWS mailboxes, the following access requirements will need to be provisioned:

1. EWS fully qualified tenant name (FQTN), such as [example.onmicrosoft.com](https://example.onmicrosoft.com).
2. EWSservice accounts, such as [NuixAppImp@example.com](mailto:NuixAppImp@example.com) (UPN, internal/external addresses, and passwords).
3. EWS test accounts, such as [TestUser@example.com](mailto:TestUser@example.com) (UPN, internal/external addresses, and passwords)
  - These should be standard mailboxes that would mirror production with a main mailbox and/or archive, if available.
4. Configuration of service accounts
  - These accounts should be given delegate access or application impersonation of the production mailboxes.
  - Multiple accounts may need to be made available depending on final architecture specifications.

### Connect to EWS via PowerShell

```
$cred = Get-Credential

$Session = New-PSSession -ConfigurationName Microsoft.Exchange -ConnectionUri
https://ps.outlook.com/powershell/ -Credential $cred -Authentication Basic
AllowRedirection

Import-PSSession $Session
```



## Assigning Mailbox Delegation Rights to Service Accounts

```
Add-MailboxPermission -identity Test1@example.com -user NuixDelegatel@example.com -AccessRights FullAccess
```

### Note

Nuix will require the following PowerShell command to be executed, which will give each upload account delegate access to all of the mailboxes in the environment.

```
Get-Mailbox | Add-mailboxpermission -user  
NuixDelegatel@example.com -AccessRights FullAccess
```

## Assigning Application Impersonation to Service Accounts

```
New-ManagementRoleAssignment -Name: NuixImpersonation -  
Role:ApplicationImpersonation -User: NuixAppImp@example.com
```

### Tip

Nuix **strongly** recommends the usage of service accounts with the Application Impersonation role when running multiple Nuix instances across multiple systems. Service accounts with Mailbox Delegation access will be throttled at a rate much higher than service accounts with impersonation enabled.

Using Impersonation also removes the need for providing Service Accounts with FULL access to all of the Exchange mailboxes.

## Supported Workflows

### Ingesting PST Data into an EWS Mailbox/Archive

1. Browse to your PSTs using the **Custodian PST Location**
2. When prompted to consolidate your PSTs, be sure to do so if you used Nuix to create your PSTs. If your PSTs were provided in user folders already, there is no need to consolidate.
3. Select the **Mapping CSV** which maps your PSTs to an EWS mailbox or archive.

### Tip

You must select a CSV file that maps the Custodian PSTs to the destination EWS mailbox/archive. **Mapping CSV** is a 5 column CSV that includes the Custodian Name in Column A, a folder you want to place the data into in Column B, the destination EWS partition in column C, the custodian's Exchange SMTP Address in Column D, and Group ID in Column E. You do not need a header column.

Each line should look similar to this:

```
Alex Chatzistamatis,ARCHIVE-  
DATA,alex.chatzistamatis@nuix.com,archive,1
```

The list of all possible locations for Column C include:

- mailbox

- archive
- purges
- archive\_purges

4. To begin extracting all of the selected EWS Mailboxes, click **Select All**, otherwise, **Select By Group ID** from the dropdown.

**Warning**

Performance may vary, especially based on the settings configured on the **Exchange Web Services** tab in **Global Settings**. Be sure to **ALWAYS** review settings prior to starting Jobs.

5. When you are ready to begin processing, click on the **Start Job** button.

## Reprocessing EWS Exceptions

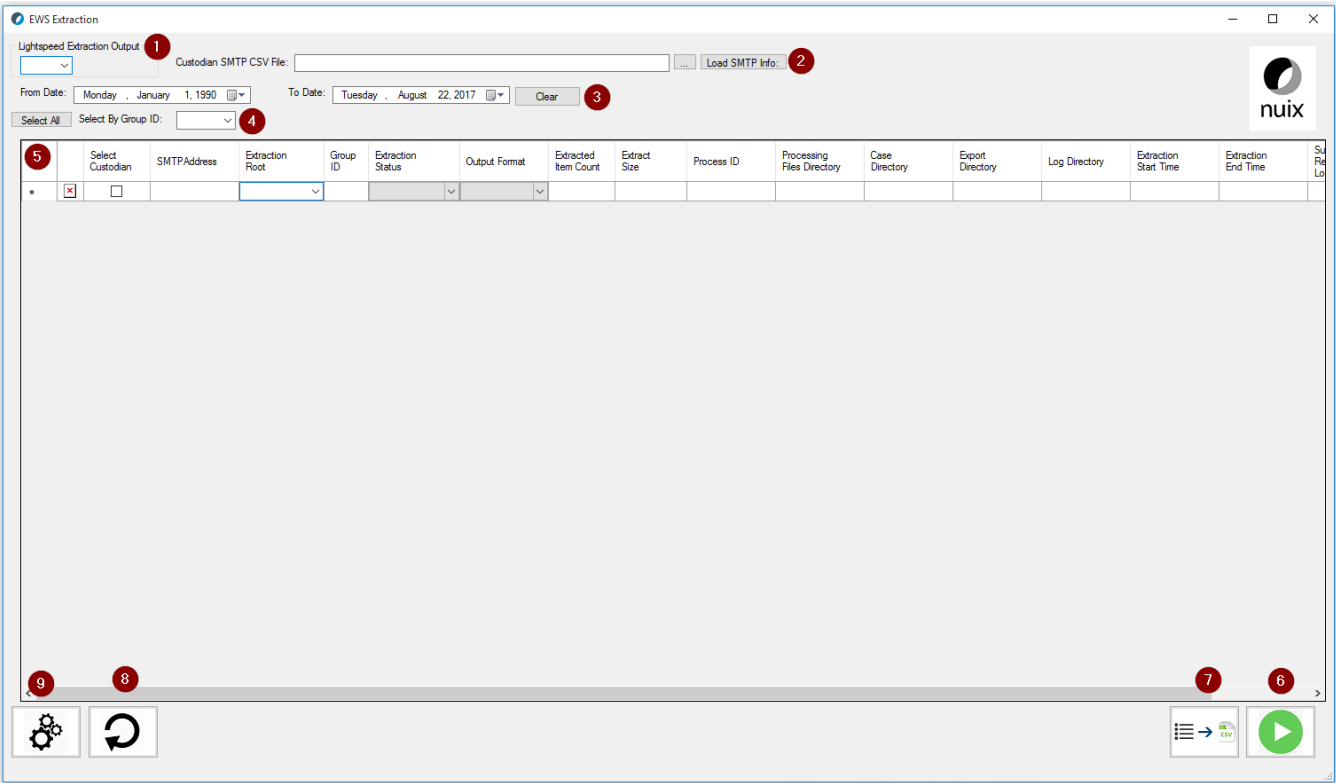
Below is a list of common errors when working with EWS and proposed solutions:

- `microsoft.exchange.webservices.data.ServiceRequestException`: The request failed. Server responded with 500 - Internal Server Error
  - o The Server 500 error is the standard throttling error.
  - o Throttling can be controlled with the switches mentioned above.
- `java.lang.IllegalArgumentException`: FTS data is too large
  - o This error occurs when the data that is being pushed into EWS is larger than the allotted size for the tenancy.
  - o Size limits can be controlled with the switch mentioned above.
  - o The default size will need to be increased in the tenancy; however, Microsoft needs to approve this.
- `java.io.IOException`: Upload failed [Mailbox has exceeded maximum mailbox size]
  - o This error occurs when an individual's mailbox has exceeded its maximum size (usually 50GB)
  - o If this occurs, the processing run should be considered invalidated, and a full run of the same exact data should occur again, AFTER the previously ingested data has been deleted.
- Skipping item for deactivated exporter for destination "user.name@mailbox.com"
  - o This error occurs when the mailbox has not been set up correctly, does not exist, or the admin account does not have delegate access to push information into it.

# Extracting Email Data from Exchange

## Interface Overview

Upon launching the EWS Extraction module, you will see the interface as shown below. Many of these options in this interface will be enabled/disabled based on selections made. This is by design several may options may not be relevant or necessary for specific source archives or workflows.



	Name	Description
1	Lightspeed Extraction Output	Select whether the format of the extracted data (PST, EML or MSG)
2	Custodian SMTP CSV	The list of custodian EWS mailboxes to be processed
3	From/To Date	Used to filter the email data by date
4	Select / Select By Group ID	Select all jobs in Grid or Select by Group ID
5	Grid	Where added jobs will be displayed
6	Start Job	Start the selected job in the grid for processing
7	Export Grid to CSV	Exports out the current grid view to CSV format
8	Reload Grid	Reloads jobs in the grid from previous migrations
9	Global Settings	View/Change previously configured Global Settings

**Tip**

Before attempting to perform any migration work, be sure to check **Global Settings** and make sure that the **Nuix Directories** and **Exchange Web Services** tab are configured correctly.

## Working with Exchange Web Services

### Overview

Microsoft's Exchange email solution allows access to software vendors using their Exchange Web Services (EWS) API. Using this API, Microsoft provides the ability to either extract or ingest data into EWS. An Exchange mailbox consists of two primary components:

- Mailbox
  - Information Store
  - Recoverable Items
    - Purges
    - Deletions
- Archive
  - Information Store
  - Recoverable items
    - Purges
    - Deletions

In order to connect to an EWS mailbox, you must first connect with a user or service account. After successful authentication, Nuix will be able to extract or ingest data.

### Prerequisites

The following prerequisites should be met in order for Nuix to interact with EWS:

1. EWS Service Account created
  - Username (SMTP address) and Password
2. EWS environment prepared
  - A valid EWS SMTP address must be available for each mailbox that will be targeted
  - If targeting an EWS personal archive, be sure that it is online and available to the mailbox

### Source Data

An Exchange mailbox or archive includes two different "partitions":

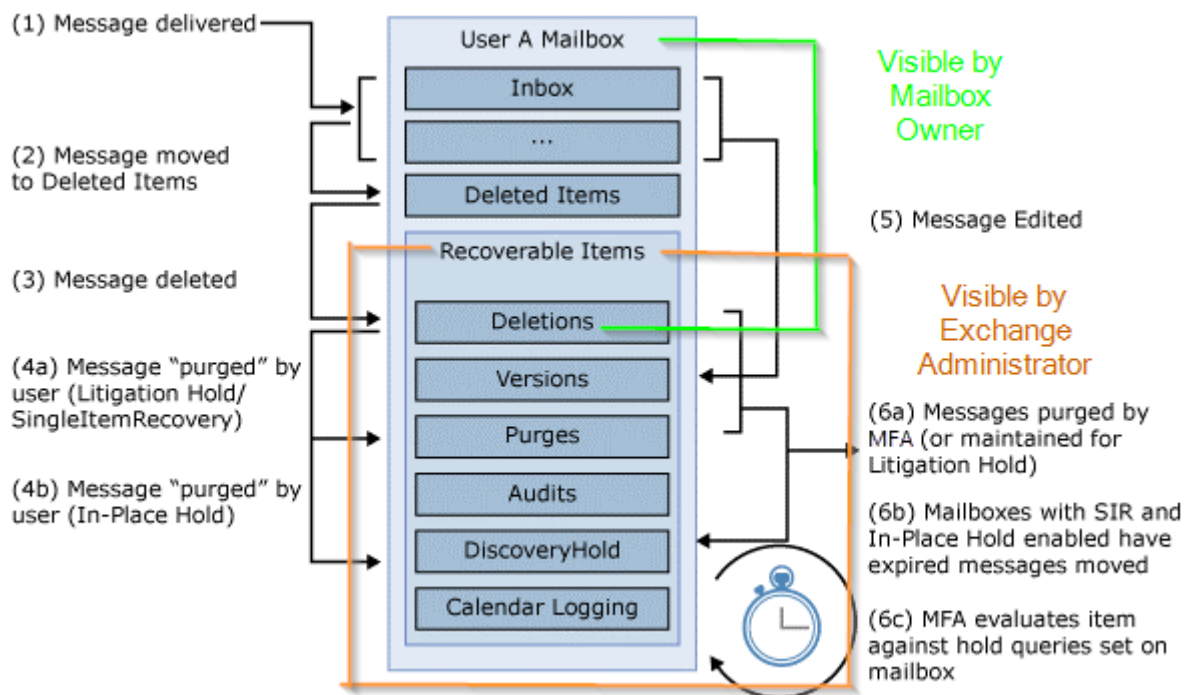
1. The first "partition" is the standard mailbox folder that is exposed to the users such as: Inbox, Deleted items, Sent Items, Drafts, any customer folder a user can create and more.
2. The second "partition" is the Recoverable Items folder partition which may

include: Deletions, Versions, Purges, Audits, DiscoveryHold and Calendar Logging.

Only Exchange Administrators can access and/or view all of the different folders in Recoverable Items. The Recoverable Items folder is hidden from users. The only exception to this rule, is a specific folder in Recoverable Items called "Deletions". This "Deletions" folder does not appear in the standard mailbox folder that the mailbox owner can see, however, the mailbox owner can access the content of this folder by using the "Recover Deletion Items" option in Outlook.

The Recoverable Items folder contains the following subfolders:

- **Deletions** - this subfolder contains all items deleted from the Deleted Items folder. (In Outlook, a user can permanently delete an item by pressing Shift+Delete.) This subfolder is exposed to users through the Recover Deleted Items feature in Outlook and Outlook Web App.
- **Versions** - if In-Place Hold or Litigation Hold is enabled, this subfolder contains the original and modified copies of the deleted items. This folder isn't visible to end users.
- **Purges** - if either Litigation Hold or single item recovery is enabled, this subfolder contains all items that are purged. This folder isn't visible to end users.
- **Audits** - if mailbox audit logging is enabled for a mailbox, this subfolder contains the audit log entries. To learn more about mailbox audit logging, see Mailbox audit logging.
- **DiscoveryHolds** - if In-Place Hold is enabled, this subfolder contains all items that meet the hold query parameters and are purged.
- **Calendar Logging** - this subfolder contains calendar changes that occur within a mailbox. This folder isn't available to users.



## Access Requirements

In order for Nuix to extract data from EWS mailboxes, the following access requirements will need to be provisioned:

1. EWS fully qualified tenant name (FQTN), such as [example.onmicrosoft.com](https://example.onmicrosoft.com).
2. EWSservice accounts, such as [NuixAppImp@example.com](mailto:NuixAppImp@example.com) (UPN, internal/external addresses, and passwords).
3. EWS test accounts, such as [TestUser@example.com](mailto:TestUser@example.com) (UPN, internal/external addresses, and passwords)
  - These should be standard mailboxes that would mirror production with a main mailbox and/or archive, if available.
4. Configuration of service accounts
  - These accounts should be given delegate access or application impersonation of the production mailboxes.
  - Multiple accounts may need to be made available depending on final architecture specifications.

### Connect to EWS via PowerShell

```
$cred = Get-Credential

$Session = New-PSSession -ConfigurationName Microsoft.Exchange -ConnectionUri
https://ps.outlook.com/powershell/ -Credential $cred -Authentication Basic
AllowRedirection

Import-PSSession $Session
```

### Assigning Mailbox Delegation Rights to Service Accounts

```
Add-MailboxPermission -identity Test1@example.com -user NuixDelegatel@example.com -AccessRights FullAccess
```

**Note**

Nuix will require the following PowerShell command to be executed, which will give each upload account delegate access to all of the mailboxes in the environment.

```
Get-Mailbox | Add-mailboxpermission -user  
NuixDelegatel@example.com -AccessRights FullAccess
```

## Assigning Application Impersonation to Service Accounts

```
New-ManagementRoleAssignment -Name: NuixImpersonation -  
Role:ApplicationImpersonation -User: NuixAppImp@example.com
```

**Tip**

Nuix **strongly** recommends the usage of service accounts with the Application Impersonation role when running multiple Nuix instances across multiple systems. Service accounts with Mailbox Delegation access will be throttled at a rate much higher than service accounts with impersonation enabled.

Using Impersonation also removes the need for providing Service Accounts with FULL access to all of the Exchange mailboxes.

## Supported Workflows

### Extracting EWS Mailbox/Archive data to a PST

1. In the **Lightspeed Extraction Output** dropdown box, select **PST**, **MSG** or **EML**

**Note**

Due to the complex nature of mailbox folder structures that may exist, we recommend extracting to **PST**. PSTs are much easier to manage since emails will be grouped by User on the file system, and long file path names issues will be avoided.

2. Browse and select a **Custodian SMTP CSV**. Click **Load SMTP Info** after you've selected the file.

**Tip**

You must select a CSV file that contains a list of Custodian SMTP mailboxes and the locations you are looking to target. **Custodian SMTP CSV File** is a 3 column CSV that includes the Exchange SMTP Address in Column A, a list of locations to target in Column B, and Group ID in Column 3. You do not need a header column.

Each line should look similar to this:

[alex.chatzistamatis@nuix.com,mailbox/mailbox recoverable/archive/archive recoverable,1](#)

[marty.mcfly@nuix.com,mailbox/mailbox recoverable/archive/archive recoverable,1](#)

The list of all possible locations for Column B include:

- root\_mailbox
- root\_archive
- mailbox\_purges
- archive\_purges
- mailbox\_recoverable
- archive\_recoverable
- public folders
- mailbox/archive
- mailbox/mailbox\_recoverable
- archive/archive\_recoverable
- mailbox/mailbox\_recoverable/archive/archive\_recoverable

3. If a date filter is requested, enter a date in the **From Date:** and **To Date:** date chooser field.
4. To begin extracting all of the selected EWS Mailboxes, click **Select All**, otherwise, **Select By Group ID** from the dropdown.

**Warning**

Performance may vary, especially based on the settings configured on the **Exchange Web Services** tab in **Global Settings**. Be sure to **ALWAYS** review settings prior to starting Jobs.

5. When you are ready to begin processing, click on the **Start Job** button.





# Appendix I: Backup/Restore SQL Databases

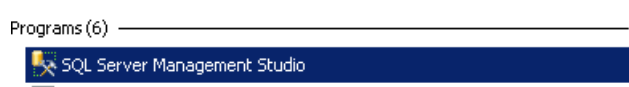
The SQL database contains critical information associated with the email archive software. Nuix uses information from the SQL DB to maintain single instancing as well as extract Distribution List and BCC recipient information, otherwise unattainable from the source data. Once the SQL database is restored on the Nuix server, the source data and SQL DB will be targeted together during the processing phase.

## SQL Backup Workflow

5. Accessing SQL Server Management Studio
6. Targeting Desired Archive Database(s)
7. Configuring Desired Archive Database(s)
8. Assigning Database Location and Confirming Drive Space
9. Appropriately Naming Database and Selection Proper File Extension

## Accessing SQL Server Management

- **Open Microsoft SQL Server Management Studio via the program menu:**



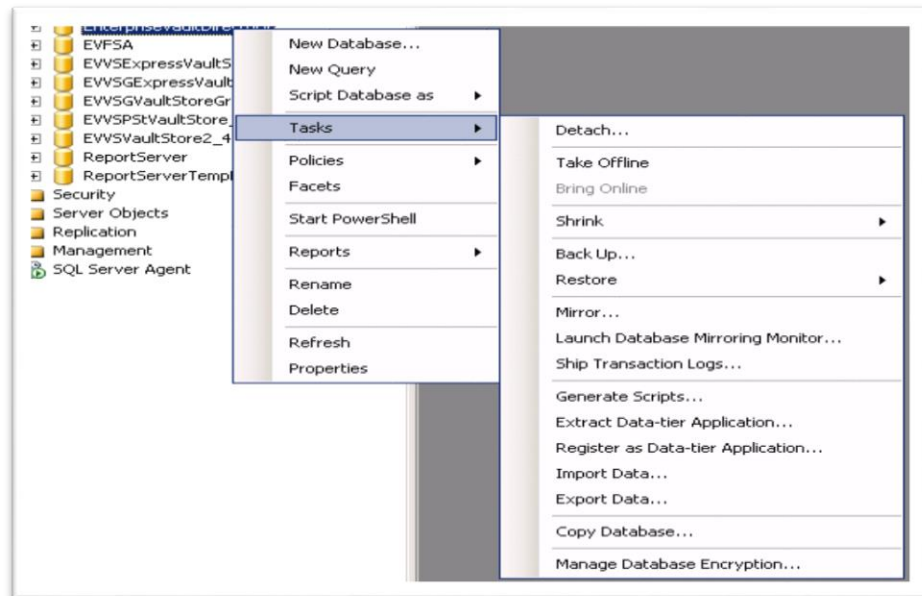
- **At the Connect to Server window:**
  - Confirm Authentication selection is "SQL Server Authentication"
  - Type in Login\Password credentials.



- Select OK

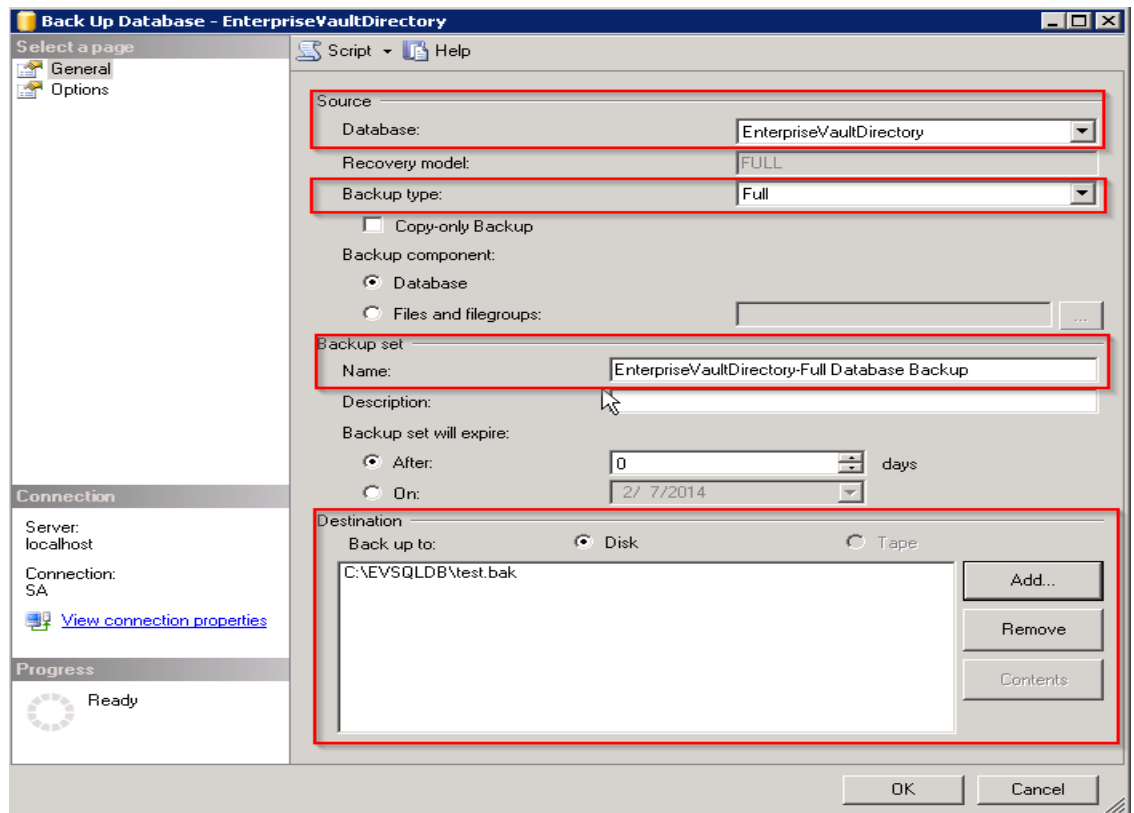
## Targeting Desired Archive Database(s)

- **In Microsoft SQL Server Management Studio:**
  - Expand Database.
  - Right click on desired database to backup and select Tasks\Back Up...\.



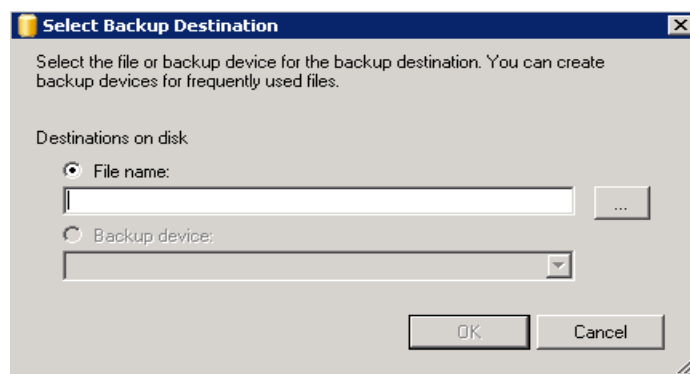
## Configuring Desired Archive Database(s)

- **At the Back Up Database window:**
  - Confirm "Source Database" name
  - "Backup Type" is Full
  - Confirm "Backup Set" name matches Source Database name with -Full Database Backup following
  - Leave "Backup set will expire:" at the default, which is 0
  - Select "Add" in the "Destination" section

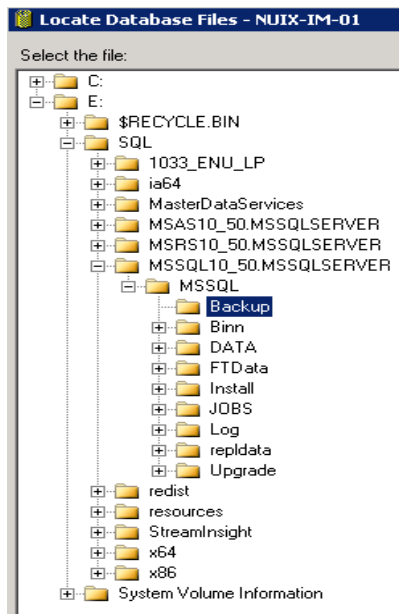


## Assigning Database Location and Confirming Drive Space

- **At the Selection Backup Destination screen:**
  - Select File name and then the ellipsis to set desired location.

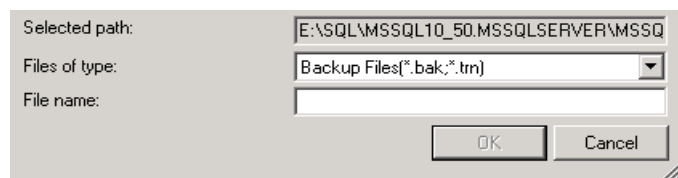


- **At the Locate Database Files screen:**
  - Set a location that will store the data by expanding the appropriate drive.



## Correctly naming the Database and choosing the proper File Extension

- Name the file name according to the database name and include .bak after the file name.



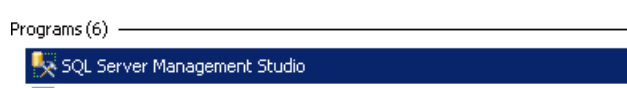
- Select OK.
- Select OK at the Select Backup Destination screen.
- Select OK at the Back Up Database screen.

## SQL Restore Workflow

10. Accessing SQL Server Management Studio
11. Restoring Desired Archive Database(s)

## Accessing SQL Server Management

- *Open Microsoft SQL Server Management Studio via the program menu:*



- *At the Connect to Server window:*
  - Confirm Authentication selection is "SQL Server Authentication"

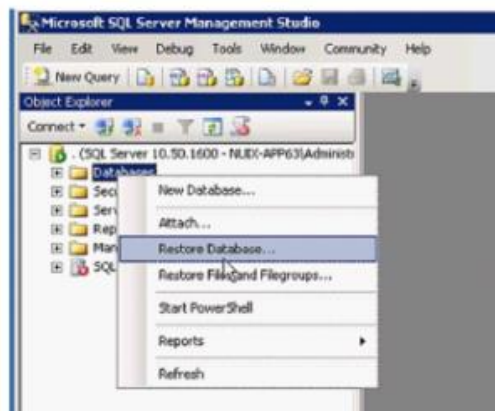
- Type in Login\Password credentials



- Select OK.

## Restoring Desired Archive Database(s)

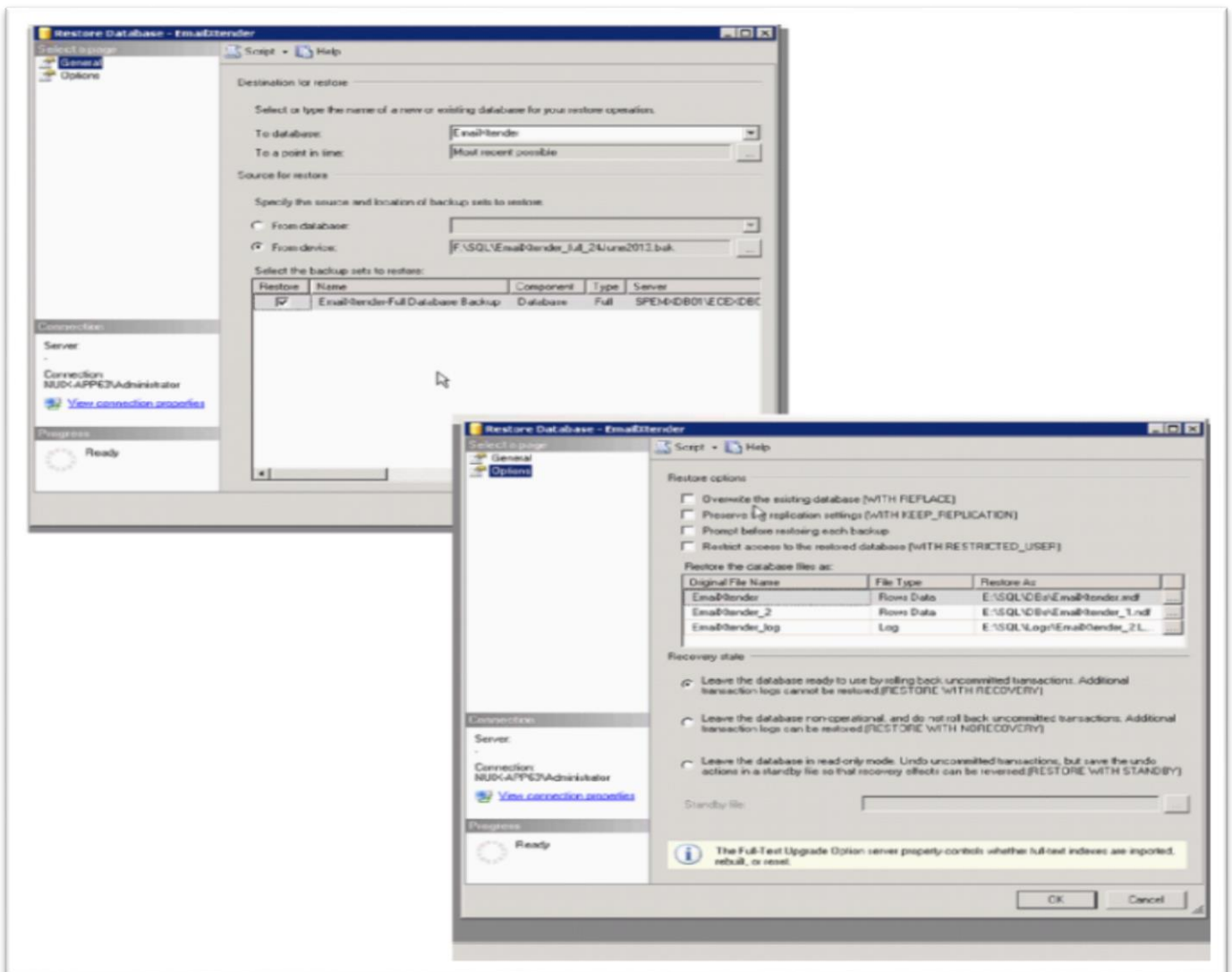
- **In Microsoft SQL Server Management Studio:**
  - Expand Database
  - Right click on desired database to backup and select Tasks\Restore Database...\



## Configuring Desired Archive Database(s)

- **At the Restore Database window:**
  - Select the database you wish to restore in the "From device:" section.
  - Select the database name you wish to apply to this database in the "To database:" section.
  - Confirm the database restore by selecting the checkbox.
- You can configure Restore Options by choosing "Options" in the left pane.

- Click Ok



# Appendix II: Archive SQL Queries

## Veritas Enterprise Vault

The following queries are hard-coded directly into the Nuix Engine when handling Enterprise Vault data.

Main SQL queries used during Nuix processing of Symantec Enterprise Vault data to confirm that the Vault Store databases exist and the Vault Store Partitions have a valid path on the system:

```
"SELECT DatabaseDSN, VaultStoreIdentity, VaultStoreEntryId " +  
"FROM [" + directoryDatabase + "].[dbo].[VaultStoreEntry] " +  
"ORDER BY VaultStoreIdentity";
```

```
"SELECT IdPartition, VaultStoreEntryId, PartitionRootPath " +  
"FROM [" + directoryDatabase + "].[dbo].[PartitionEntry] " +  
"WHERE VaultStoreEntryId = ?";
```

SQL query used during Nuix processing of Symantec Enterprise Vault data to find all associated parent Transaction IDs, SIS Parts (DVSSP files) to reconstitute the message and all single-instanced attachments:

```
"SELECT sp.ParentTransactionId, sp.IdPartition, sp.VaultStoreIdentity,  
s.CollectionIdentity, sp.ArchivedDateUTC " +  
"FROM Saveset s, Saveset_SISPart ssp, SISPart sp " +  
"WHERE s.SavesetIdentity = ssp.SavesetIdentity AND ssp.SISPartIdentity =  
sp.SISPartIdentity " +  
" AND sp.FPDistinctionByte = ? AND sp.FPHashPart1 = ? AND s.IdTransaction = ?" +  
"ORDER BY s.CollectionIdentity DESC";
```

SQL query used during Nuix processing of Symantec Enterprise Vault data to find the folder name and path that the item is archived in:

```
"SELECT af.FolderName, af.FolderPath " +  
"FROM" +  
" [" + directoryDatabase + "].dbo.ArchiveFolder af," +  
" [" + directoryDatabase + "].dbo.[Root] r," +  
" [" + vaultStoreEntry.databaseName + "].dbo.Vault v," +  
" [" + vaultStoreEntry.databaseName + "].dbo.Saveset s " +  
"WHERE" +  
" s.VaultIdentity = v.VaultIdentity AND" +  
" v.VaultID = r.VaultEntryId AND" +  
" r.RootIdentity = af.RootIdentity AND" +  
" s.IdTransaction = ?";
```

SQL query used during Nuix processing of Symantec Enterprise Vault data to find the user Exchange/AD details for each transaction:

```
"SELECT eme.mbxNtUser, eme.ADMbxDN " +  
"FROM " +  
" [" + directoryDatabase + "].[dbo].[ExchangeMailboxEntry] eme, " +  
" [" + vaultStoreEntry.databaseName + "].[dbo].[view_Saveset_Archive_Vault] sav " +  
"WHERE " +  
" eme.DefaultVaultId = sav.VaultId AND " +
```



```
" sav.IdTransaction = ?";
```

SQL query used during Nuix processing of Symantec Enterprise Vault data to find the archive details (Archived Data, Vault Store Entry ID, Archive Name, Archive Description, Archive Point ID, Saveset ID) for each transaction:

```
"SELECT " +  
" [" + vaultStoreEntry.databaseName +  
"].dbo.view_Saveset_Archive_Vault.ArchivedDate, " +  
" [" + directoryDatabase + "].dbo.ArchiveView.VaultStoreEntryId, " +  
" [" + directoryDatabase + "].dbo.ArchiveView.ArchiveName, " +  
" [" + directoryDatabase + "].dbo.ArchiveView.ArchiveDescription, " +  
" [" + directoryDatabase + "].dbo.ArchiveView.[SID] " +  
"FROM" +  
" [" + directoryDatabase + "].dbo.ArchiveView " +  
"INNER JOIN " +  
" [" + vaultStoreEntry.databaseName + "].dbo.view_Saveset_Archive_Vault " +  
" ON" +  
" [" + vaultStoreEntry.databaseName +  
"].dbo.view_Saveset_Archive_Vault.ArchivePointId = [" + directoryDatabase +  
"].dbo.ArchiveView.VaultentryId " +  
"WHERE" +  
" [" + vaultStoreEntry.databaseName +  
"].dbo.view_Saveset_Archive_Vault.IdTransaction = ?";
```

## EMC EmailXtender

Main SQL query used during Nuix processing of EMC EmailXtender data to lookup distribution list recipients and append to '**Expanded-DL**' metadata property:

```
"SELECT [EmailAddress]" +  
" FROM [" + databaseName + "].[dbo].[EmailAddress] " +  
" WHERE [EmailId] IN (" +  
" SELECT [EmailId] " +  
" FROM [" + databaseName + "].[dbo].[Route]" +  
" WHERE [MD5HashKey] = ? AND [RouteTypeId] = ?) ";
```

Additional notes:

- '**MD5HashKey**' is equal to the value of the '**Xtender Hash Key**' located in an email's metadata.
- Here is a list of all the '**RouteTypeId**' options that are available in EmailXtender:
  - 0 = All Recipients
  - 1 = To Recipients
  - 2 = From (sender)
  - 4 = CC Recipients
  - 8 = BCC Recipients
  - 16 = Distribution List
  - 32 = Discovered
  - 64 = Routeable (DL Recipients)

# EMC SourceOne

Main SQL query used during Nuix processing of EMC SourceOne data to lookup distribution list recipients and append to '**Expanded-DL**' metadata property:

```
"SELECT [EmailAddress]" +  
" FROM [" + databaseName + "].[dbo].[EmailAddress] " +  
" WHERE [EmailId] IN (" +  
"   SELECT [EmailId] " +  
"   FROM [" + databaseName + "].[dbo].[Route]" +  
"   WHERE [MessageId] = ? AND [RouteType] = ?) ";
```

Additional notes:

- '**MessageId**' is equal to the value of the '**0x**' + '**Xtender Hash Key**' located in an email's metadata.
- Here is a list of all the '**RouteType**' options that are available in SourceOne:
  - 1 = To Recipients
  - 2 = From (sender)
  - 3 = CC Recipients
  - 4 = BCC Recipients
  - 5 = Distribution List
  - 6 = Routeable (DL Recipients)

# HP/Autonomy Zantaz EAS

Main SQL query used during Nuix processing of Zantaz EAS data to get the Archive ID:

```
filename.contains("%") ? "LIKE " : "= ";
```

```
"SELECT TOP 1 DATAARCHIVEID  
FROM easadmin.dataarchive  
WHERE filename " + like + "?"  
AND docserverid = ?";
```

SQL query used during Nuix processing of Zantaz EAS to get embedded Centera IDs:

```
"SELECT pl.msgid, pl.dataarchiveid, r.userid, r.folderid " +  
"FROM easadmin.dataarchive da " +  
"JOIN easadmin.profilelocation pl " +  
" ON da.dataarchiveid = pl.dataarchiveid " +  
"JOIN dbo.refer r " +  
" ON pl.msgid = r.msgid " +  
"WHERE da.filename = ? AND docserverid = ?";
```

SQL query used during Nuix processing of Zantaz EAS to get embedded IDs:

```
"SELECT pl.msgid, r.userid, r.folderid " +  
"FROM easadmin.profilelocation pl " +  
"JOIN dbo.refer r " +
```

```
" ON pl.msgid = r.msgid " +  
"WHERE pl.dataarchiveid = ?";
```

SQL query used during Nuix processing of Zantaz EAS to get MsgId:

```
"SELECT pl.msgid, pl.dataarchiveid, r.userid, r.folderid " +  
"FROM easadmin.profilelocation pl " +  
"LEFT JOIN dbo.refer r" +  
" ON pl.msgid = r.msgid " +  
"LEFT JOIN easadmin.dataarchive da" +  
" ON pl.dataarchiveid = da.dataarchiveid " +  
"WHERE pl.msgid = ? AND da.docserverid = ?";
```

SQL query used during Nuix processing of Zantaz EAS to get MAPI metadata:

```
"SELECT " + "msgread, categories, pr_importance, msgflagtext, flagstatus,  
flagcomplete, contacts, pr_expiry_time, replytime " +  
// ",flagdueby, flagduebynext, remindersettings " +  
"FROM dbo.refer " +  
"WHERE msgid = ? AND userid = ? AND folderid = ?";
```

SQL query used during Nuix processing of Zantaz EAS to get folder structure:

```
"SELECT foldername FROM dbo.FOLDER WHERE folderid = ?";
```

SQL query used during Nuix processing of Zantaz EAS to get user information:

```
"SELECT OBJDISTNAME, USERNAME " +  
"FROM dbo.USERS " +  
"WHERE userid = ?";
```

SQL query used during Nuix processing of Zantaz EAS to get recipients:

```
"SELECT a.emailaddress, m.typefld " +  
"FROM dbo.EMAILMESSAGES m, dbo.EMAILADDRESSES a " +  
"WHERE a.emailid = m.emailid AND m.msgid = ?";
```

SQL query used during Nuix processing of Zantaz EAS to get message region:

```
"SELECT * FROM easadmin.profilelocation WHERE msgid = ? AND dataarchiveid = ?";
```

# Appendix III: Archive Metadata

## Veritas Enterprise Vault

..... Date	Wednesday, August 14, 2013 at 6:46:32 PM Eastern Daylight Time
..... Enterprise Vault Archive Date	Friday, August 16, 2013 at 12:00:00 AM Eastern Daylight Time
..... Enterprise Vault Archive Description	EV Journal
..... Enterprise Vault Archive Name	EVJournal
..... Enterprise Vault Archive SID	S-1-5-21-2534329463-2701829754-3966392267-1171
..... Enterprise Vault DVS ArchivePointID	173B5773D41FF3C47868B989F232E45881110000EV01
..... Enterprise Vault DVS FolderPath	Inbox
..... Enterprise Vault DVS Retention Category	14B79436C0CFC0F4EB063A1263FA631321b10000EV01
..... Enterprise Vault DVS Transaction ID	B14239423E8E4A42CC219265CC8B0D01
..... Enterprise Vault DVS VaultID	173B5773D41FF3C47868B989F232E45881110000EV01
..... Enterprise Vault NT Username	journal01
..... Enterprise Vault Saveset ID	666754312013081~201308140400000000~Z~b1423942-3e8e-4a42-cc21-9265cc8b0d01
..... Enterprise Vault VaultStoreEntryId	1D367EA9AD88EA7469E17CD45CBEE34BF1210000EV01
..... File Created	Friday, August 16, 2013 at 1:50:31 PM Eastern Daylight Time

## EMC EmailXtender/Source

..... Xtender Hash Key	5c7f9ec98811d2fa5982f9204b1df8a34cb1ed6300
..... Xtender Message ID	4CA526905C7F9EC98811D2FA5982F9204B1DF8A34CB1ED6300
..... Xtender Timestamp	1285891728

## HP/Autonomy Zantaz EAS

..... Zantaz Archive ID	12,155
..... Zantaz Folder ID	30,556
..... Zantaz Message ID	14,454,085
..... Zantaz Message Identifier	12155-14454085-489-30556
..... Zantaz User CN	/O=██████ INC/OU=FIRST ADMINISTRATIVE ...
..... Zantaz User ID	489
..... Zantaz User Name	EAMILJOURNALUSER2

# Daegis AXS-One

AXS-One-attachment	charset=utf-8;file=body.txt;isBody=yes;mimeType=te...
AXS-One-axsmailboxfolder	Long-Term_Correspondence
AXS-One-category	MIS0002
AXS-One-charset	utf-8
AXS-One-conversationid	01CF5430F7C16AEA4A7BA8A44BEA957B208AB420156C
AXS-One-externalsender	Emily.Frye@tanger.com
AXS-One-folder	/exchange/Andrew.Brimlow@tanger.com
AXS-One-form	IPM.Note
AXS-One-from	Emily M. Frye
AXS-One-generator	ExCrawler
AXS-One-hash	d06b8dc006e6195a496bcdabc6d4ad58-01ffMD500
AXS-One-isinbound	no
AXS-One-isinternalonly	no
AXS-One-mailbox	Andrew.Brimlow@tanger.com; ABRIML
AXS-One-messageid	d06b8dc006e6195a496bcdabc6d4ad58-01ffMD500
AXS-One-originalmessageid	<19582F73A76FF44FB98D628D09785E0D011F1C65@E...
AXS-One-origmessagesize	669015
AXS-One-posteddate	04/09/2014 21:55:20
AXS-One-recipient	Andrew.Brimlow@tanger.com; ABRIML
AXS-One-subfolder	
AXS-One-subject	
AXS-One-userinrecipients	no
AXS-One-version	15.1.10.0

# Appendix IV: EWS Best Practices

## Leveraging Azure Virtual Machines

Whether extracting data from O365 or ingesting data into O365, it is highly recommended to provision Azure servers for the Nuix workflow. Azure and O365 are essentially hosted in the same Microsoft data centers, which remove many of the complications that would be present if performing the extraction or ingestion from an on-premise Nuix server. These complications include server-level and network-level throttling (mentioned below), but also bandwidth limitations and the network distance data will have to travel in order for it to be extracted or ingested.

Multiple Azure servers can be used in order to scale Nuix vertically on one machine and horizontally on multiple machines to achieve the desired throughput.

### Note

Consult the Azure pricing guide to determine the appropriate Azure environment size: <https://azure.microsoft.com/en-us/pricing/calculator/>

## Reduced Number of Nuix Workers

Keeping the number of Nuix workers to a minimum during extractions or ingestions is detrimental to your overall performance. Too many workers can create excess connections and create throttling.

### Tip

Nuix **strongly** recommends using 1, 2 or 4 workers when ingesting to or extracting from EWS. More than 4 workers per Nuix instance may cause increased levels of throttling which will cause the ingestion/extraction to take longer and may also lead to a higher than normal number of failed items.

## EWS Throttling Workarounds

It is well documented that Microsoft will throttle connection uploads and downloads into a particular environment based off of:

- IP Address
- Mailbox
- Connection account
- Number of connections / amount of data being pushed at once

To workaroud this throttling, it is recommended to:

- Adjust the Exchange Web Services Settings in your Global Settings accordingly.
- Use single item downloads or uploads over bulk downloads or uploads.
- Use Application Impersonation service accounts over Delegate accounts.

- Scale Nuix vertically on a single machine and horizontally across multiple machines in Azure.
- Consolidate your data per custodian if ingesting data, and push this data into O365 from a single Nuix instance.
- Never overlap data across multiple Nuix instances. This overlap will cause unnecessary throttling, which in turn will increase the likelihood of exceptions.

# About Nuix

Nuix ([www.nuix.com](http://www.nuix.com)) protects, informs, and empowers society in the knowledge age. Leading organizations around the world turn to Nuix when they need fast, accurate answers for investigation, cybersecurity incident response, insider threats, litigation, regulation, privacy, risk management, and other essential challenges.