Enterprise Vault™ Indexing Best Practices

14.2

VERITAS™

Enterprise Vault™ Indexing Best Practices

Legal Notice

Veritas Technologies LLC
2625 Augustine Drive.
Santa Clara, CA 95054

http://www.veritas.com

Technical Support

Technical Support maintains support centers globally. Technical Support's primary role is to respond to specific queries about product features and functionality. The Technical Support group also creates content for our online Knowledge Base. The Technical Support group works collaboratively with the other functional areas within the company to answer your questions in a timely fashion.

Our support offerings include the following:

- A range of support options that give you the flexibility to select the right amount of service for any size organization
- Telephone and/or Web-based support that provides rapid response and up-to-the-minute information
- Upgrade assurance that delivers software upgrades
- Global support purchased on a regional business hours or 24 hours a day, 7 days a week basis
- Premium service offerings that include Account Management Services

For information about our support offerings, you can visit our website at the following URL:

www.veritas.com/support

All support services will be delivered in accordance with your support agreement and the then-current enterprise technical support policy.

Contacting Technical Support

Customers with a current support agreement may access Technical Support information at the following URL:

www.veritas.com/support

Before contacting Technical Support, make sure you have satisfied the system requirements that are listed in your product documentation. Also, you should be at the computer on which the problem occurred, in case it is necessary to replicate the problem.

When you contact Technical Support, please have the following information available:

- Product release level
- Hardware information
- Available memory, disk space, and NIC information

- Operating system
- Version and patch level
- Network topology
- Router, gateway, and IP address information
- Problem description:
  - Error messages and log files
  - Troubleshooting that was performed before contacting Technical Support
  - Recent software configuration changes and network changes

Licensing and registration

If your product requires registration or a license key, access our technical support Web page at the following URL:

www.veritas.com/support

Customer service

Customer service information is available at the following URL:

www.veritas.com/support

Customer Service is available to assist with non-technical questions, such as the following types of issues:

- Questions regarding product licensing or serialization
- Product registration updates, such as address or name changes
- General product information (features, language availability, local dealers)
- Latest information about product updates and upgrades
- Information about upgrade assurance and support contracts
- Advice about technical support options
- Nontechnical presales questions
- Issues that are related to CD-ROMs, DVDs, or manuals

Support agreement resources

If you want to contact us regarding an existing support agreement, please contact the support agreement administration team for your region as follows:

Worldwide (except Japan)	CustomerCare@veritas.com
Japan	CustomerCare_Japan@veritas.com

Contents

Technical Support 4

Chapter 1 About this guide 9

Scope of the document 9
Intended audience 9
Related documents 9

Chapter 2 About the Elasticsearch indexing engine 11

Overview of the 64-bit Elasticsearch indexing engine 11

Chapter 3 Enterprise Vault 14.2 design considerations 13

Server sizing considerations 13
Storage considerations 14
Choosing a suitable storage device for index location and index snapshot location 15

Chapter 4 Enterprise Vault 14.2 architectural considerations 17

Overview of Index Server Groups 17
When to use Index Server Groups 18
Splitting out indexing services from storage services 18
Archive > Index Volume > Elasticsearch index mapping 19
Elasticsearch index rollover 20

Chapter 5 Upgrading from an earlier version 22

Planning for the upgrade 22
What happens immediately after the upgrade 23
Upgrading existing indexes (32-bit Alta-vista or 64-bit Velocity) to Elasticsearch indexes 23
Index upgrade process 25

Chapter 6 Other deployment considerations 26

Elasticsearch JVM Heap size 26
Index location and index snapshot location 28
Pagefile size 28
Index backup 28

Chapter 7 Tuning 29

Search criteria 29
Federated searching 29
Adding more CPU vs. Adding more memory 30
Performance counters 30

Chapter 1

About this guide

This chapter includes the following topics:

- Scope of the document
- Intended audience
- Related documents

Scope of the document

This document aims to provide guidance on designing and sizing of the new 64-bit Elasticsearch indexing engine found in Enterprise Vault 14.2 and later. The recommendations in this document do not apply to earlier versions of Enterprise Vault.

This document should be used in conjunction with other performance and best practice guides as outlined in the Related documents section of this document.

Intended audience

This document is aimed at customers, consultants, and support staff. It is assumed that the reader has a thorough understanding of the architecture and operational aspects of Enterprise Vault 14.2.

Related documents

Use this guide in conjunction with the following documents:

- *Enterprise Vault Performance Guide*: https://www.veritas.com/support/en_US/article.100000918

- *Discovery Accelerator Best Practices Guide*: https://www.veritas.com/support/en_US/doc/DA_142_Best_Practices_Guide
- *Enterprise Vault Installation and Configuration Guide*: https://www.veritas.com/content/support/en_US/doc/115743999-151277577-0/

Chapter 2

About the Elasticsearch indexing engine

This chapter includes the following topics:

- Overview of the 64-bit Elasticsearch indexing engine

Overview of the 64-bit Elasticsearch indexing engine

Enterprise Vault 14.2 introduces a new 64-bit Elasticsearch indexing engine. The new features and functionality of this engine need to be considered by both existing customers while upgrading and new customers installing a fresh deployment.

Enterprise Vault offers backwards compatibility with existing 32-bit Alta-Vista indexes, 64-bit Velocity indexes seamlessly federating searches between old and new data indexed by the new 64-bit Elasticsearch indexing engine.

In addition to offering backwards compatibility, the 64-bit Elasticsearch indexing engine offers the latest features of indexing technology, including:

- Modern, highly scalable and delivers better performance in terms of indexing, searches and optimizes memory utilization.
- Storage and memory utilization.
- Seamless federated searching across legacy 32-bit, 64-bit non-Elasticsearch indexes and new 64-bit Elasticsearch indexes transparent to customers following the upgrade to Enterprise Vault 14.2 and later.
- No downtime during backing up index data. Indexing and searching can be performed during backing up index data. The backing strategy for indexes is changed in Enterprise Vault 14.2 and later. For more information about backing up index data, see the following technical white paper:

https://www.veritas.com/support/en_US/doc/EV_WHITEPAPER_BACKINGUP_EV

- A new set of PowerShell commands have been developed to aid in managing both day-to-day and upgrade scenarios, including workflows with progress and status indicators.

Chapter 3

Enterprise Vault 14.2 design considerations

This chapter includes the following topics:

- Server sizing considerations
- Storage considerations
- Choosing a suitable storage device for index location and index snapshot location

Server sizing considerations

The Elasticsearch indexing engine is more scalable and provides better performance, but it does require additional hardware resources to achieve this. For Enterprise Vault 14.2 and later, the recommended hardware requirements for an EV server are 16 CPU cores and 32GB of RAM. Customers using this recommended hardware should expect to see equal or better-archiving performance when compared to a server running an earlier version of Enterprise Vault on the recommended hardware specified for that version. The following table shows the new hardware requirements for Enterprise Vault:

Table 3-1

Minimum and recommended hardware	Details
Recommended, if running Indexing service	16 CPU cores, 32GB RAM
Minimum (servers not running Indexing service)	8 CPU cores, 16GB RAM

The memory requirements have also increased, but it will vary depending on the search and ingestion requirement. Some percentage of memory is allocated and reserved for Elasticsearch indexing engine. Default 8GB memory is allocated and reserved for the Elasticsearch indexing engine. Enterprise Vault Administrator can configure reservation; but be careful about performance impact if reduced from default value.

Storage considerations

Careful consideration should be given to storage changes outlined in this section of the document; Enterprise Vault 14.2 has different index sizing and performance characteristics as compared to earlier versions.

Single index location

Elasticsearch indexing engine does not support multiple index locations. Enterprise Vault Administrators can only specify single index location to store Elasticsearch indexes. We recommend providing enough disk space considering this requirement. Also, it is better to have a separate index location for storing index data of Elasticsearch indexes rather than using existing non-Elasticsearch index locations.

As multiple index locations not support by Elasticsearch indexing engine. Please refer here for recommendation from Elasticsearch to design Elasticsearch data storage architecture for scale.

Indexing footprint

Similar to earlier versions, Enterprise Vault 14.2 and later provide two levels of indexing ; Brief and Full. Index data footprint for different content sources are mentioned in the table below.

Table 3-2

Index Type	Exchange Mailbox	File System	SMTP
Full	9%	6%*	7%

* Indexing footprint for File System might vary based on type of attachments/documents/files being indexed. If you have a large proportion of binary files (images, audio, video, executables) then the indexing footprint will be considerably smaller than the figures given above.

Provision disk space for index data based on the future requirement.

Index snapshot location

Elasticsearch indexing engine does not support the traditional file-system way of backing up indexes. It supports a snapshot mechanism for backing up indexes. Enterprise Vault 14.2 and later, the newly created indexes should be backup using snapshots mechanism.

Enterprise Vault Administrators must specify an additional location to store snapshots of Elasticsearch indexes. Therefore, there is an additional disk requirement to store snapshots. Elasticsearch index data is already compressed and therefore no further compression during snapshots. Therefore, snapshots of Elasticsearch index data require the same amount of disk space allocated or provisioned for storing index data of Elasticsearch indexes. We recommend to provision disk to store snapshots as 2 times more than that you have provision for storing index data. Also, provision disk space for storing snapshot data considering amount of data being archived and indexed in future.

Choosing a suitable storage device for index location and index snapshot location

Similar to earlier versions, a high-speed storage device is recommended for index location used to store Elasticsearch indexes. Lower tier NAS devices are not recommended for index locations and should not be used to host indexes where any type of Accelerator search is used. NAS devices connected over CIFS shares are not suitable to host index volumes for certain environments due to the slower connectivity speed.

Elasticsearch indexing engine performance varies based on storage used. Refer to the following link to learn about the recommendations from Elasticsearch indexing:

https://www.elastic.co/pdf/elasticsearch-sizing-and-capacity-planning.pdf

Index snapshot location

We recommend storing snapshot of all index servers in same location. Better to have same index snapshot location configured on all Enterprise Vault index servers in the Enterprise Vault site. Index snapshot location can be configured at the site level. Considering number of index servers and amount of index data store per index server, please provision sufficient amount of disk space for index snapshot location if all snapshots are stored in single location. See “Storage considerations” on page 14.

We recommend using Fiber and iSCSI over CIFS. Before using partitions on a SAN, consider the I/O load together with any other applications that are already using the SAN to ensure that the performance can be maintained. Ideally, the

implementation should be discussed with the SAN hardware vendor to ensure that optimum performance is achieved.

Note: Mapped network drive is not supported.

Chapter 4

Enterprise Vault 14.2 architectural considerations

This chapter includes the following topics:

- Overview of Index Server Groups
- When to use Index Server Groups
- Splitting out indexing services from storage services
- Archive > Index Volume > Elasticsearch index mapping
- Elasticsearch index rollover

Overview of Index Server Groups

A new feature in Enterprise Vault 10 is the ability to split out Indexing and Storage roles. Index Servers can be grouped together into logical containers called Index Server Groups. Index Server Groups can be used to share resources for more scalable indexing and searching.

A dedicated Indexing server (that is, a server that does not have a local Storage service or agents) must be a member of an Index Server Group in order to host indexes. A combination EV server (that is, one that has both Index and Storage services) can also be part of an Index Server Group; this will however defeat the point of Index Server Groups as the server's resources will not be dedicated to indexing and might reduce performance.

Note: The Index Server Group membership of associated index servers and vault stores cannot be changed as soon as the first item is archived. Index Server Groups and vault stores association can be changed or removed, but once data is archived you cannot change components, whether index server, vault store or indexes, in the group.

When to use Index Server Groups

Index Server Groups are a scalability feature and as such should be considered carefully. In most cases a cost/benefit analysis will help determine whether Index Server Groups should be included in the design; weighing up the customer requirements for search performance and archiving throughput against the added complexity and cost of additional hardware and Windows server licenses.

In environments where frequent eDiscovery searches are performed on large index volumes such as journal or file system archives, and the speed of which results are returned is important, the use of Index Server Groups will likely be of benefit. Index Server Groups provide the option to scale out index volumes, and use specialized fast storage and provide fast results when it matters.

The impact of increasing the number of servers in the Index Server Group does not improve archiving or expiry performance significantly, but will have a positive impact on Discovery Accelerator search performance.

Mailbox archiving and SharePoint archiving are unlikely to benefit from using Index Server Groups. This means that in the majority of cases we anticipate that Index Server Groups will not be required.

Splitting out indexing services from storage services

By adding additional hardware you can split storage services and indexing services between servers, freeing up resources that would have been consumed by other processes. This can be particularly beneficial where an Enterprise Vault server is already running at capacity.

Distributed architecture of having separate storage and index server remains unchanged.

Archive > Index Volume > Elasticsearch index mapping

There was one-one mapping between Enterprise Vault index volume and index at indexing engine side in earlier versions. This mapping has been changed in Enterprise Vault 14.2 and later due to architecture of Elasticsearch indexing engine. Every Elasticsearch index must have at least one shard and the total number of shards per node (Enterprise Vault index server) is capped to 600 by default. A general thumb rule is to have a max 20 shards for every GB assigned to Elasticsearch JVM. Elasticsearch recommends capping memory of JVM at either 50% of physical RAM or 30GB, whichever is less. Consider maximum limit 30 x 20 = 600 (max cap for JVM * number of shards per GB) shards can be created at max per node (Enterprise Vault index server). It means maximum 600 Elasticsearch indexes (minimum 1 shard per index) can be created per node. Like earlier versions, if existing one-one mapping had been used in Enterprise Vault 14.2, there would be an implicit limitation to support only 600 index volumes per node. In that case, the existing recommended load of 5000 user mailboxes per server would have been reduced to 600 user mailboxes per server (considering 1 user mailbox => 1 index volume)

To overcome above limitation of Elasticsearch shards limits per node (Enterprise Vault index server), Enterprise Vault 14.2 introduce an alternate way of mapping index volumes to Elasticsearch indexes known as “one to many mapping” (one Elasticsearch index containing data of multiple Enterprise Vault archives). Enterprise Vault will be storing index data of multiple archives in a single Elasticsearch index based on the following rules.

- **Naming conventions**:
  - Name of newly created Elasticsearch index will be of format “**<EV server alias>_<Archive type>_<Index number>**”.
    - Archive type is a shorthand notation for type of archive. For example, "mb" for mailbox archives, "sd" for shared archives, "smtp" for smtp archives, and so on.
    - Index number is an incremental counter tracking indices of similar archive type. For example, first index created for any archive type will have the index number as 1 followed by 2 for the second index, and so on.
  - Elasticsearch has the flexibility to create multiple aliases for any index. Enterprise Vault creates an alias for every Enterprise Vault index volume and associates it to an Elasticsearch index while indexing. Name of every index alias will be of format "**<EV Index Volume name>_<Index number>**"
- **Constraints**:

- "MaximumShardSizeMB" : Maximum allowable size of a shard in MB. Default value: 50000
- "MaximumArchivesPerIndexForLargeData" : Maximum number of archives of type "Large" that can be stored in an Elastic Index. Default value: 7
- "MaximumArchivesPerIndexForSmallerData" : Maximum number of archives of type "Small" that can be stored in an Elastic Index. Default value: 600
- "ShardsPerIndexForLargeData" : Maximum number of shards to be linked with Elastic Index storing archives of type "Large". Default value: 10
- "ShardsPerIndexForSmallerData" : Maximum number of shards to be linked with Elastic Index storing archives of type "Small". Default value: 5
- "CommaSeparatedLargeDataArchiveTypes" : Comma separated list of archives types based on the enum “DV_DS_E_VAULT_TYPE” that are considered as "Large". Default value: "DV_DS_VT_JN_ARCHIVE,DV_DS_VT_SMTP_ARCHIVE,DV_DS_VT_LOTUS_JN_ARCHIVE"

The above architectural changes introduced in Enterprise Vault 14.2 will not have any impact to Enterprise Vault Administrator, Discovery Accelerator Administrator and End Users. Searches and Indexing will be seamless at Enterprise Vault archive and index volume level.

Elasticsearch index rollover

Enterprise Vault performs Elasticsearch index rollover if any of limits mentioned above is reached. During rollover it might create or re-use an existing Elasticsearch index.

Elasticsearch index selection

Any Enterprise Vault Index volume can either be:

- A newly created Index volume and the archive type of the associated archive of index volume is either:
  - Already mapped to an Elasticsearch Index. In this case Enterprise Vault will either:
    - Use the same Elasticsearch index to add new data if it is within limits mentioned above
    - Chose another Elasticsearch index if any of limits mentioned above is reached either
      - From other existing Elasticsearch indices (within limits) related to the same archive type

        - By creating a new Elasticsearch index for that archive type in case there is no existing Elasticsearch index within limits. The name of Elasticsearch index will be as mentioned above

    - Not mapped to any Elasticsearch Index
        - Enterprise Vault creates a new Elasticsearch index for this archive type with index number 1. The name of Elasticsearch index will be as mentioned above.

- An existing Index volume and thus the archive must have already been mapped to some Elasticsearch Index. In that case:
- Enterprise Vault will either:
    - Use the same Elasticsearch Index to add new data
    - Chose another Elasticsearch Index either
        - From other existing Elasticsearch Indices related to the same archive type
        - By creating a new Elasticsearch Index for that archive type

---

Note: Elasticsearch index rollover and Enterprise Vault index volume rollover are completely different. Enterprise Vault index volume rollover happens when its configured limit is reached and Elasticsearch index rollover happens when one of limits mentioned above is reached. Enterprise Vault index volume rollover can create a new Elasticsearch index or can be rolled over to an existing Elasticsearch index.

Chapter 5

Upgrading from an earlier version

This chapter includes the following topics:

- Planning for the upgrade
- What happens immediately after the upgrade
- Upgrading existing indexes (32-bit Alta-vista or 64-bit Velocity) to Elasticsearch indexes
- Index upgrade process

Planning for the upgrade

Consider the following points and act before you upgrade to Enterprise Vault 14.2 and later:

- Ensure your server is appropriately specified to meet the minimum or recommended CPU and Memory requirements. If not, you must plan for a hardware upgrade.
- New indexing engine does not allow multiple index location. So, all index data will be stored in one location. Plan proper storage location and disk for storing index data. Please refer to the *Indexing footprint* table in Storage considerations and plan disk space accordingly. Also, consider future growth of archive and index data.
- The backing up strategy of indexing data is changed in Enterprise Vault 14.2 and later. It does not support traditional file-system backup of indexing data. It supports a snapshot mechanism; additional storage is required for index snapshot location. It is better to have a separate disk altogether for storing snapshots of indexing data, so that data can be recovered if indexing data disk failures. The

index data is already compressed, so there is no additional compression in case of snapshot. We recommend having 2 times the storage space that of indexing location for storing snapshots. Also, consider future growth of archive and index data.

- Ensure index location and index snapshot location are part of the Anti-Virus exclusions (both real-time and scheduled) to avoid scanning those areas and causing open file locks. If adding new server cache locations, ensure those locations are excluded from scanning.
- Ensure you specify index location for storing Elasticsearch index data. If not specified “Enterprise Vault Indexing Service” will not start. Please follow the steps mentioned in the warning message displayed during upgrade through installer or by `Start-EVDatabaseUpgrade` PowerShell command.
- Ensure you specify index snapshot location in case you want to backup index data. Please refer to the Storage considerations section in this document.

What happens immediately after the upgrade

The following processes will take place after the upgrade:

- All existing index locations will be closed.
- “Enterprise Vault Indexing Service” will not start unless you specify new Elasticsearch index location.
- When first item is indexed during normal indexing workflow or rebuild or upgrade workflow, new Index Volume (Type = 64-bit and IndexingEngineType = 1) will be created for Elasticsearch. Please refer to “Archive => Index Volume => Elasticsearch index mapping” section in this document.
- Newly archived items go into Elasticsearch index.
- Legacy index volumes co-exist with Elasticsearch index volumes.
- Searches are federated across 32-bit Alta-vista index, 64-bit Velocity index and 64-bit Elasticsearch index.

Upgrading existing indexes (32-bit Alta-vista or 64-bit Velocity) to Elasticsearch indexes

For customers upgrading from earlier version to Enterprise Vault 14.2 and later, the best search performance will be obtained if they upgrade existing indexes to Elasticsearch indexes. There is however no reason to immediately perform the upgrades, searches are transparently federated across both 32-bit Alta-vista, 64-bit

Velocity and 64-bit Elasticsearch indexes, so it recommended that initially only a small number (between 5 and 10) of indexes are upgraded and monitored. Once the administrator is comfortable with the process and an indication of upgrade performance is obtained a larger number of archives can be scheduled for upgrading. During the index upgrade process end-users and Accelerators can continue to search the old 32-bit Alta-vista, 64-bit Velocity index volumes until the upgrade is complete at which point it will be switched over to the 64-bit Elasticsearch volume, and old index volume deleted. It is up to the administrator to decide which indexes they want to upgrade, and when the upgrades should take place. The default schedule is to run index upgrades during the day. Index upgrades can be performed throughout the night if required, including during archive runs. Index upgrades cannot be performed while the Index service is in backup mode, therefore check the schedule of backups and when the backup mode PowerShell scripts are set to run.

After the upgrade, the old 32-bit Alta-vista or 64-bit Velocity index will automatically be deleted, unless errors are reported. The status console will clearly tell you if there was a problem upgrading the index, and you can investigate the logs and try again.

The upgrade does not require the old index to be functional as it is performing a rebuild index from storage. Slower storage (particularly partitions migrated to lower tier storage such as tape devices) will take longer to rebuild. If the upgrade fails, then it is due to issues with the original data in storage. A failed upgrade will generate a report and based on the content of the report you can either try and repair the storage or ignore the error and allow it to complete.

The following table shows a **Rebuild rate** on a single Enterprise Vault server based on our performance run.

Table 5-1

Exchange Mailbox (Items per/hour)	SMTP (Items per/hour)	File System (Items per/hour)
310K	260K	390K

The above numbers are calculated on the following factors of test data used in our performance run and rebuild rate might vary in your environment.

- Average item size: 165KB for email | 100KB for FSA
- Number of Archives: 5K mailbox archives | 1 SMTP archive | 1K FSA archive.
- Number of items in index: 1.5M for Exchange & SMTP | 0.5M for FSA
- Storage used for partition: DAS (Non-SSD)
- Storage used for indexing data: DAS (Non-SSD)

- Other operations running on Enterprise Vault server: Only Rebuild Indexing operation was in progress.

It is recommended to run the rebuild operation during off-peak hours and no other operation is running on the Enterprise Vault index server.

Index upgrade process

The Index Upgrade task enables index volumes to be upgraded from 32-bit Alta-vista, 64-bit Velocity to 64-bit Elasticsearch volume. The old index volumes remain searchable whilst the new volumes are being rebuilt. During the upgrade process existing old index volumes will be upgraded to Elasticsearch index volumes. There will be separate index volumes in Enterprise Vault, but at the backend, all might be indexed to the same Elasticsearch index. Please refer to the Archive > Index Volume > Elasticsearch index mapping section in this document to know what the Elasticsearch index is created.

Chapter 6

Other deployment considerations

This chapter includes the following topics:

- Elasticsearch JVM Heap size
- Index location and index snapshot location
- Pagefile size
- Index backup

Elasticsearch JVM Heap size

Some percentage of memory is allocated and reserved for Elasticsearch indexing engine, by default 8 GB memory is allocated and reserved. Enterprise Vault Administrator can change this default value; but be careful about performance impact if reduced from default value.

The following table shows impact on Accelerator search rate when Elasticsearch JVM Heap size is increased.

Table 6-1

Search Client	8 GB	16 GB
VAS (hotword search)	31 Hits/Sec	51 Hits/Sec
Discovery Accelerator (Blanket search)	3208 Hits/Sec	3465 Hits/Sec

The above numbers are calculated on the following factors of test data used in our performance run and search rate might vary in your environment.

- Average item size: 165KB
- Number of Archives: 1 SMTP for DA | 5 SMTP for VAS
- Number of items in index: 1.5M for DA | 7.3M for VAS
- Storage used for partition: DAS (Non-SSD)
- Storage used for indexing data: DAS (Non-SSD)
- Number of hotwords for VAS search: 1000
- Other operations running on Enterprise Vault server: DA search | VAS search, one at a time.

The following table shows impact on end user searches when Elasticsearch JVM Heap size is increased.

Table 6-2

Search Client	8 GB	16 GB
Enterprise Vault Search response time*	39 sec	8 sec

*The response time for end user search might decrease if multiple users execute search simultaneously.

Above numbers are calculated on following factors of test data used in our performance run and end user search response time might vary in your environment.

- Average item size: 165KB
- Number of Archives: 1 SMTP
- Number of items in index: 1.5M
- Storage used for partition: DAS (Non-SSD)
- Storage used for indexing data: DAS (Non-SSD)
- Other operations running on Enterprise Vault server: EVS search only.

We observed 65% improvement in VAS hotword searches when we increase the JVM heap size to 16GB.

We recommend increasing Elasticsearch JVM heap size to 16GB in the environments where there are intensive Accelerator searches. For example searches with criteria having large number of hotwords, departments or employees.

Note: Increasing JVM heap size may have impact on other Enterprise Vault services if those are running on same server having Enterprise Vault Indexing service.

Index location and index snapshot location

- If indexes are stored on NetApp devices and other NAS systems, opportunistic locking must be turned off for volumes that contain indexes.
- Better to have index location local to index server.
- Better to store snapshots of all the index servers in the same location. Specify index snapshot location at Enterprise Vault site level.
- Given indexes can be distributed across an Index Server Group, this means that every index server in the group can then demand data from a single storage server, potentially overloading the server. It is recommended that Index Server Groups have no more than 15 members for Journal archives.
- Disable Windows file indexing on the drives that contain Enterprise Vault indexes.
- Exclude index location and index snapshot location from anti-virus scanning.
- Failing to exclude the indexes from scanning can cause index corruption and will result in reduced server performance.

Pagefile size

The Pagefile should conform to Microsoft's recommendation of 1.5 times the amount of memory. It is recommended that the page file is placed on a different partition and different physical hard disk drive from the system partition so that Windows can handle multiple I/O requests more efficiently.

Additionally, if you set the "Initial size" of the Pagefile to the same as the "Maximum size" it will prevent unnecessary fragmentation file. An un-fragmented paging file leads to faster virtual memory access.

Index backup

Detailed recommendations on how to back up the index volumes can be found in the article below:

https://www.veritas.com/support/en_US/doc/EV_WHITEPAPER_BACKINGUP_EV

Chapter 7

Tuning

This chapter includes the following topics:

- Search criteria
- Federated searching
- Adding more CPU vs. Adding more memory
- Performance counters

Search criteria

Due to limitation from Elasticsearch indexing engine on wild card searches, it is recommended to provide a minimum of 3 leading characters along with the wildcard when dealing with searches involving phrases.

For example, search criteria “Enterprise Vault indexing engine migrated to **ela***” is better than specifying “Enterprise Vault indexing migrated to **e***”

Federated searching

When searches are performed (using Discovery Accelerator, Browser Search, Outlook Integrated Search, and so on), all Index Volumes for an archive (both 32-bit and 64-bit) will be searched automatically. Discovery Accelerator will always search all Index Volumes separately and combine the results. End-user search applications (such as Browser Search) will only search a maximum of five index locations concurrently by default. This setting can be changed by modifying the FederatedSearchMaxVolSets value in the webapp.ini file. As with all performance tuning care should be taken when changing default values – changes should be made in a controlled fashion and any impact measured against baseline performance testing done beforehand.

Adding more CPU vs. Adding more memory

Scaling up an index server can help improve indexing performance; generally more performance will be obtained by adding more CPU as opposed to more memory. If disk I/O is proving to be a bottleneck adding more memory will help improve performance.

Performance counters

A number of Windows performance counters have been added under an “Enterprise Vault Indexing Volumes Processing” category. These counters will capture the ingestion state of affairs since the last Indexing service restart.

- **Additions Items**: The number of items successfully indexed to the local indexing engine.
- **Processed Items**: The number of items processed for indexing. This includes items that have been successfully indexed as well as the failed items and items that have been indexed but are not yet confirmed.
- **Deleted Items**: The number of items deleted from the local indexing engine.
- **Updated Items**: The number of items updated in the local indexing engine.

The following table described counters tracking activity within the agents.

Table 7-1

Performance Counter	Description
Item fetched	Items retrieved from storage
Item submitted	Items submitted to the index engine but not yet confirmed as indexed
Item acknowledged	Items confirmed by indexing engine as indexed
Item audited (failed)	Item recorded as failed
Item audited (successful)	Items recorded as successful
Item persisted (pending)	Items recorded as pending acknowledgment (because of shut down or re-queue). These will be reloaded and acknowledged later
Item reloaded (pending)	Items previously persisted as pending acknowledgment that have been reloaded for acknowledgment

Table 7-1 *(continued)*

Performance Counter	Description
Item in the pipeline	Total items currently in the index pipeline
Item in the pipeline	Items that have re-fetched
Item missing from the pipeline	Items that have been re-acknowledged
Item queue count	Total items currently queued up waiting to be traversed ready to send
Item queue size (KB)	Total memory used by items queued up waiting to be traversed