



ExtremeZ-IP 4.1 Indexed Search – Information and Configuration

Version: 1.0
Date: 10/18/05
Product Version: 4.0

Introduction:

ExtremeZ-IP 4.1 introduces an all new, high performance file search feature which enables the client user to search extremely large volumes of files up to 600 times faster* than servers running ExtremeZ-IP 4.0, Windows File Sharing/Server Message Block (SMB), or File Services for Macintosh (SFM). This incredibly fast search performance is made possible by a new search indexing component included in ExtremeZ-IP 4.1. This document will explain how this new search feature works and how you can get the most out of it.

Technical Overview:

ExtremeZ-IP 4.1's new search feature functions by creating an index file containing information about every file on your ExtremeZ-IP volumes. Instead of scanning your server's drives each time a client makes a search request and looking through every file in the volume, ExtremeZ-IP 4.1 simply checks the index file to retrieve search results. This process takes a few seconds compared to minutes or even hours, depending on the number of files in your shared volume.

The search indexer monitors changes within ExtremeZ-IP volumes and updates its indexes continuously to keep search results accurate. While ExtremeZ-IP has to perform work keeping volumes' index files up to date, this process consumes far fewer system resources than the act of multiple clients doing traditional searching, which scans an entire volume's files and folders individually on every search.

Searching from a Mac OS client:

With ExtremeZ-IP indexed searching enabled on the server, a Mac client can use the built-in Mac OS search functionality to perform fast searches of ExtremeZ-IP volumes. No client side configuration or applications are necessary.

Note: The Mac OS performs two types of files searches: *catalog searches* and *enumeration searches*. A catalog search issues a search request that is processed on the server side. This is the type of request that ExtremeZ-IP indexed searching responds to. The Mac OS only issues a catalog search request when the user is searching the root of a volume. For this reason, it is only technically possible to provide indexed search results for searches initiated at the root of a volume. **Please instruct your users to search the entire ExtremeZ-IP volume for the fastest results.** If a search is initiated from a sub-folder of a volume, Mac OS performs an enumeration search and scans this folder and all of its subfolders across the network. This combination of non-indexed searching and network latency results in drastically reduced search performance.

The Index File:

A search index file is created when you first configure an ExtremeZ-IP volume or when you first start ExtremeZ-IP after upgrading to version 4.1. Indexed searching will be enabled by default on new and existing volumes. Each volume with indexed searching enabled will contain an index file in the root of its physical location on the server. This file will begin with ".INDEX" and will be hidden from Mac clients. The index file is also set with the "hidden" windows file system property. You will be able to see it by browsing to the volume location on the server itself, if you have "Show Hidden Files and Folders" enabled on the server. If the volume is also shared to PC clients using Windows File Sharing, Windows clients will be able to see the index file if they have "Show Hidden Files and Folders" enabled.

The location of the index file can be modified on a per volume basis, if desired, in the volume's properties dialog in ExtremeZ-IP Administrator. This option will be further detailed in the Indexed Search Configuration Options section of this document.

The Indexing Process:

When you start ExtremeZ-IP 4.1 for the first time after installation, the search indexer will scan all files and folders in each configured volume and add their information to the volume's index file. The time that this initial indexing takes is dependent on the number of files and folders you have shared on your server and the speed of your server. In testing, the indexing of 1 million files has taken around 15 minutes on a reasonably modern server. The availability of files is not impacted in any way by search indexing and ExtremeZ-IP 4.1 File Server continues to make files available immediately after start up.

Once this initial indexing has been performed, the search indexer will update the volume's index with any file changes to the volume. The NTFS change journal is used to monitor changes to the volume's files and folders and can be used to update ExtremeZ-IP indexes with changes that were made by ExtremeZ-IP clients and any non-ExtremeZ-IP service or client. Because the change journal runs independently of ExtremeZ-IP, it can even be used to update search indexes with changes that were made while the ExtremeZ-IP service was stopped for maintenance or a server restart.

The amount of time that ExtremeZ-IP can be stopped and recover gracefully with search index updates from the change journal is dependent on the number of changes that took place while ExtremeZ-IP was shut down. The change journal has a finite length and if the number of file changes that took place on an individual server drive while ExtremeZ-IP was shut down exceeds the capacity of the change journal, ExtremeZ-IP will not be able to pick up where it left off and all volumes on that server drive will have to be fully re-indexed to ensure accurate search results.

During initial indexing of your server, searches will return incomplete results until the indexing has been completed. The indexing status of your volumes can be viewed in the ExtremeZ-IP Administrator's **Volumes** dialog. After the initial indexing occurs, if ExtremeZ-IP is required to fully re-index volumes due to a prolonged service shutdown, the existing index file will be used to satisfy search requests until the new index file has been created. Because the existing index file was up to date when the service was shut down, it should provide very close to 100% search accuracy. Only files added to the volume while the service was shutdown will not be displayed in search results until the re-indexing has completed.

Index File Re-indexing/Compaction:

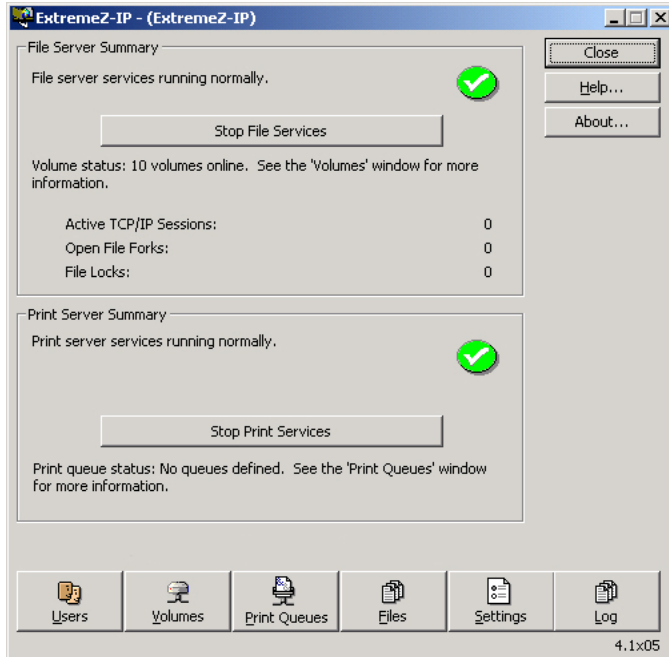
As is the case with many performance-centric database systems, the ExtremeZ-IP index file entries for files that have been deleted or moved from a volume are not physically removed from the index file at the time the actual file is deleted. The indexed search service is designed to account for these deleted entries and to keep search results accurate despite the fact that these entries are left in the index file. However, this will cause the index file to grow over time and, as the file gets larger, slow search performance to a small extent. The rate at which the index file grows is dependent of the number of files being added, moved, and deleted on the file server.

In order to keep ExtremeZ-IP search performing at optimal levels, volumes' indexes will be routinely re-indexed and compacted. The interval at which this occurs is determined by the ratio of deleted (stale) records to valid entries in the index. By default, the ExtremeZ-IP search service will re-index an individual volume when approximately 1/3rd of that volume's index file records are deleted, stale records. This maintenance occurs on a per volume basis and only volumes requiring re-indexing will be processed. While re-indexing, the volume's existing search index is kept up to date and used to provide 100% accurate search results.

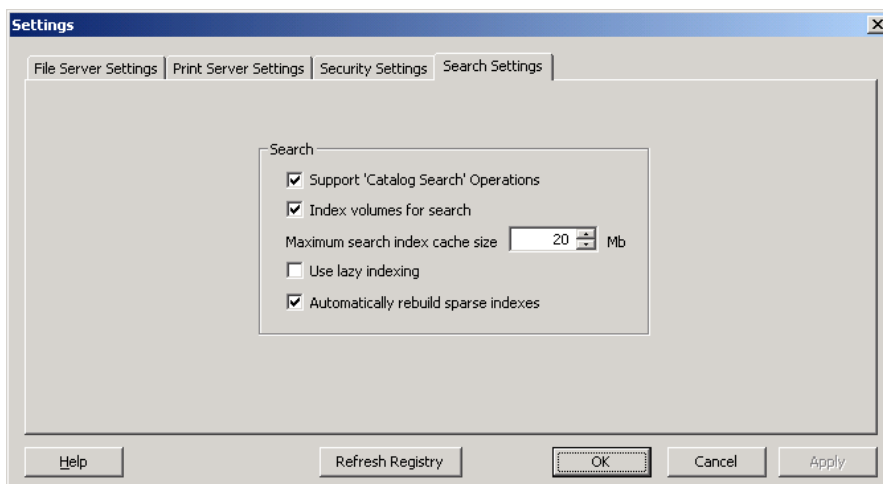
Re-indexing should not have any detrimental effect on other server processes while it is running. While ExtremeZ-IP is re-indexing an individual volume, a status of "Reindexing" will be shown in the **Volumes** dialog of the ExtremeZ-IP Administrator. If you would prefer disable automatic re-indexing and schedule re-indexing on a set schedule during off-hours, an EZIPUTIL command, fully detailed at the end of this document, can be used in a batch file or script and triggered by a scheduling service of your choice.

Indexed Search Configuration Options:

Many aspects of ExtremeZ-IP 4.1's indexed search can be customized through the ExtremeZ-IP Administrator program. Global settings related to search are accessed by selecting the **Settings** button in the main Administrator window. Per-volume settings related to search are accessed by selecting the **Volumes** button.



Global indexed search setting: Indexed searching is enabled by default and can be enabled and disabled globally in the **Search Settings** tab of the **Settings** dialog of ExtremeZ-IP Administrator. To turn indexed searching on or off for all ExtremeZ-IP volumes, simply check or uncheck “**Index volumes for search**”. In order for changes to this setting to take effect, you must stop and restart the *ExtremeZ-IP File Services for Macintosh* service. In addition, **Support ‘Catalog Search’ Operations** defaults to being enabled. This must be enabled for index search to function.



Search RAM cache maximum size setting: Indexed searching uses a RAM cache to help manage indexing and search requests more efficiently. This cache is set to a max size of 20 MB by default. In reality, the amount of RAM used by indexed searching will only grow to approximately the total size of all index files on your server. An index file containing 250,000 files should only be around 8 MB in size. In a majority of situations, leaving the

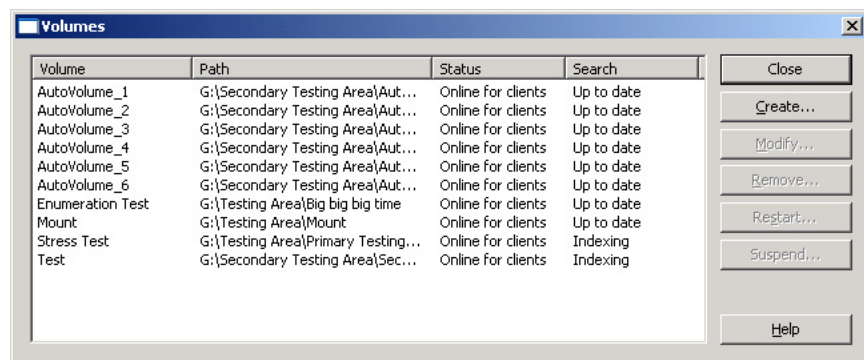
search index RAM cache limit at the default setting and letting ExtremeZ-IP manage it should give the best results. On a server with a large amount of physical memory and over 500,000 files in shared volumes, search performance may improve if the search RAM cache limit is raised to 32 MB x # of millions of files on the server. In cases where a server is running with limited physical memory, the RAM available to indexed searching can be reduced to as little as 8 MB. This setting, “**Maximum search index cache size**”, is available in the **Search Settings** tab of the **Settings** dialog of ExtremeZ-IP Administrator, as shown above. This setting will take effect immediately.

Indexing ‘aggressiveness’ setting: By default, indexed searching uses any available system resources to keep its indexes up to date. Indexing cooperates with other system processes and should not adversely affect overall system performance to a significant extent. In a case where a server is under high load or is running many different services simultaneously, it is possible to limit the system resources that search indexing will consume by configuring ExtremeZ-IP search to use lazy indexing mode. This setting is available in the **Search Settings** tab of the **Settings** dialog of ExtremeZ-IP Administrator, as shown above. To enable lazy indexing, simply check “**Use lazy indexing**”. This setting will take effect immediately.

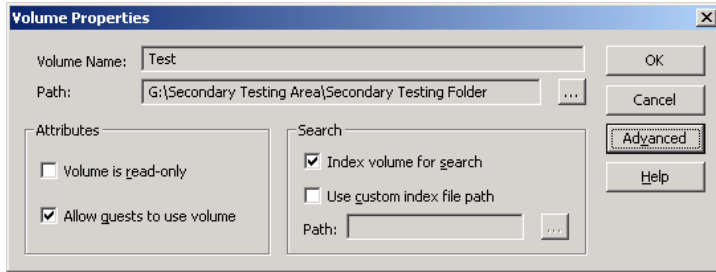
Automatic rebuild of sparse indexes setting: By default, ExtremeZ-IP indexed search automatically monitors and re-indexes individual volumes when their index file’s contents have become sparse. For purposes of re-indexing, this occurs when an index has reached approximately 1/3rd stale entries. These stale entries are created when files are moved and deleted from the volume. Automatic rebuilding of sparse indexes can be enabled and disabled through the **Search Settings** tab of the **Settings** dialog of ExtremeZ-IP Administrator, as shown above. This setting will take effect immediately.

Note: If “**Automatically rebuild sparse indexes**” is disabled, it is recommended that scheduled re-indexing be set up at an appropriate interval by using the EZIPUTIL command detailed below in conjunction with a batch file or script and a scheduling service of your choice. Without routine rebuilding of indexes, index file size will continue to grow over time.

Volume Status: The status of indexed search on each individual volume is shown in the **Volumes** dialog of ExtremeZ-IP Administrator, shown below. These statuses include: Disabled, Up to date, Indexing, Reindexing, and Failed to initialize.



Per-volume indexed search setting: Indexed searching is enabled on all existing and newly created volumes by default. You can disable or enable indexed searching on a per volume basis by checking or unchecking “**Index volume for search**” in the individual volume’s **Volume Properties** dialog in ExtremeZ-IP Administrator, as shown below. This setting is accessed by selecting the volume in the **Volumes** dialog and clicking **Modify...** This property can be set at initial volume creation time or after the volume has been created. In order for changes to this setting to take effect, you must stop and restart the *ExtremeZ-IP File Services for Macintosh* service.



Per-volume index file path setting: By default, a volume’s search index file will be created in the root of the volume’s physical location on the server. An alternate index file location can be specified, on a per volume basis, by setting the “**Use custom index file path**” option in the individual volume’s **Volume Properties** dialog in ExtremeZ-IP Administrator. This property can be set at initial volume creation time or after the volume has been created. In order for changes to this setting to take effect, you must stop and restart the *ExtremeZ-IP File Services for Macintosh* service.

EZIPUTIL Commands:

The following command has been added to the EZIPUTIL utility and can be used to manually trigger the re-indexing of a volume. This might be helpful if you would like to automate re-indexing using scripts or batch files during “off-hours”. EZIPUTIL.exe is located in the chosen ExtremeZ-IP program installation directory on the server.

To rebuild a volume's search index:

```
EZIPUTIL VOLUME /REINDEX
    /NAME:volumename
    /PATH:root directory path
    [/SERVICENAME:servername] (Cluster only).
```

Enabling Search Logging:

If you wish to view a detailed status of search indexing in the ExtremeZ-IP log, you will need to enable search logging. This will also allow you to view the status of search requests, indexing, and incremental index updates in the ExtremeZ-IP log. To enable search logging change the value of the registry key:

```
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\ExtremeZ-IP\Parameters4\DebugLogging\Search
```

To “INFO”. By default, the ExtremeZ-IP log is located in: C:\Program Files\Group Logic\ExtremeZ-IP\Logs\ExtremeZ-IP\ExtremeZ-IP LOG.txt

With search logging at its default setting, error and warning messages only will be written to the log.

Known Issues Related to Mac OS X 10.4 – 10.4.2 ‘Tiger’ Features and Bugs:

Mac OS X 10.4 through 10.4.2 contains several features and bugs which may cause problems while using ExtremeZ-IP indexed search. Group Logic has verified that these issues also occur when searching a Mac OS X Server’s AFP shared volumes from a 10.4.2 client and has submitted bug requests to Apple regarding them. These issues potentially could be addressed in the upcoming 10.4.3 update.

Searches may hang when returning large numbers of results: In testing, this has been found to occur, to an extent, when greater than 200 results are returned. As the number of results reaches the 1000's, extended delays may occur. This seems to occur because 10.4 resolves the file information for each result returned much slower than the results are actually returned to the client. As long as the search is not canceled, it will complete, but possibly after several minutes of delay, which gives the appearance that the search has hung.

Searches may hang when typing slowly into the 10.4 search prompt: This problem is a direct result of the problem returning large numbers of results. Mac OS 10.4 includes a "search as you type" function that will start a search on whatever is in the search prompt if you pause typing for more than ½ second or so. This produces a quick set of results when searching a local hard drive using Spotlight but causes problems searching servers over a network. When entering a search request on the client, if a user intends to type "Tiger" but pauses briefly after typing "T", a search for "T" will be initiated automatically, likely returning 100's or 1000's of results very quickly and hanging the client Mac's Finder for several minutes. The only real work-around for this is to type fast enough to not trigger the unintended search or to cut and paste your search request into the search prompt.

Searches on 'Name' with parameters "ends with," "begins with," and "is" don't work: This is apparently just a bug in Mac OS X 10.4 through 10.4.2.

Sometimes stopping a search crashes the Finder: When experiencing one of the hang/delay conditions above, it may be natural to cancel the search request that appears to be hung. In some instances, this has been shown to crash the Finder. This is apparently just a bug in Mac OS X 10.4 through 10.4.2.

* Based on a comparison test performed with a Mac OS 10.3.9 client searching a 1 million file volume located on a Windows 2003 Enterprise Edition server over a 1000Base-T network.