

RE: Slow Catalog Build

Source:

<http://www.tech-archive.net/Archive/Internet-Server/microsoft.public.inetserver.indexserver/2004-06/0135.html>

From: George Cheng [MSFT] (GCheng_at_online.microsoft.com)

Date: 06/21/04

Date: Mon, 21 Jun 2004 15:13:07 GMT

See the following

317586 HOW TO: Optimize Indexing Service Performance in Windows 2000

<http://support.microsoft.com/?id=317586>

Also

Using Registry Keys

In addition to the steps above, several registry keys can be used to control resources dedicated to indexing. These registry keys control how Indexing Service behaves and how it responds to system resource requests by users or other applications. On a machine dedicated to indexing content, these registry keys can be set to aggressively keep the index up to date, regardless of other activities occurring on the machine.

The following table details the relevant registry keys and the recommended values for optimal indexing performance. Changes to the settings for these keys will be honored without a restart of the service, though in some cases it may take several minutes for the change to take effect.

All of these registry keys are fully documented in the Platform SDK. You may want to consult that documentation if you have questions about these values.

Key Data Type Recommended Setting Comments

ScanBackoff REG_DWORD 0 This key controls the length of time IS periodically pauses while scanning a disk looking for new files. It's useful to configure how aggressively IS should use system resources to complete a scan. A value of 0 tells IS to not pause at all. This is especially useful on multi-processor machines.

MaxFreshCount REG_DWORD 100,000 This controls how many documents have to change before a master merge is started. When a catalog grows beyond a few hundred thousand documents indexing time is dominated by master merges rather than actual indexing. Setting this key to a large value like 100,000 decreases the frequency of master merges. Depending on the catalog, values as high as 400,000 may substantially improve index build time.

FilterTrackers REG_MULTISZ "" Filter trackers are a way for an application to hook into the indexing pipeline. As a document is indexed all filter

trackers listed in this key will be invoked so they can do whatever additional processing required. By default in Windows 2000 one filter tracker is installed that generates thumbnails of documents that are displayed in the file explorer. It's non-trivial to generate the thumbnails, so if you don't require them, delete the value of this key. The value is empty by default on Windows XP.

MaxWordlistIO REG_DWORD 0xffffffff When system IO exceeds this value indexing is temporarily paused. With the recommended value IS won't pause regardless of system IO.

MaxWordlistIODiskPerf REG_DWORD 100 This is another metric used for system IO. Setting the value to 100 means indexing will never stop due to high system IO.

MinimizeWorkingSet REG_DWORD 0 At regular intervals IS will attempt to flush its memory usage to make room for other applications. When this value is set to 0 the behavior is disabled.

MaxWordlistUserIdle REG_DWORD 0 When someone is actively using the keyboard or mouse on a machine IS will by default pause indexing until the machine is again idle. Setting this value to 0 disables this behavior.

FilterIdleTimeout

REG_DWORD 2 Setting this is useful if you are using an IFilter DLL that is buggy or prone to hanging. When the filter daemon (cidaemon.exe) has been idle for this long it is terminated and restarted. The default is 15 minutes, and the recommended value is 2 minutes for the impatient.

MaxIndexes

REG_DWORD 100 This controls how many indexes are created as documents are being filtered. When an index is first being created, a large value like 100 is optimal. Once an index is initially built, queries will run faster with a smaller value, like 25. You should change this value after the initial index build is complete.

MaxWordlists REG_DWORD 50 This controls how many in-memory indexes are built before they are merged and written to a shadow index on disk. The higher the value, the faster the indexing. A setting of 50 means lots of documents will be indexed before they are merged.

MinSizeMergeWordlists REG_DWORD 1024 This limits how much RAM in 4k pages is used by wordlists before they are merged and written to a shadow index on disk. More is better at the cost of RAM.

Each of these keys can be set in two places in the registry. The first location, hklm\System\CurrentControlSet\Control\ContentIndex, sets the default for all catalogs. You can override these values for each catalog at the subkey Catalogs\CATALOGNAME. After a default installation, some values may not exist and may need to be added.

The default values of these keys vary depending on the product installed (Windows 2000 Professional vs. Windows 2000 Server) and the RAM and CPUs available at the time of installation. Some of the values change when you tune performance in the Indexing Service MMC administration tool. The recommended values enable better performance than can be achieved via the MMC tool.

Once an index is up to date, a different set of registry changes can be used to help keep the index current as changes are made to the files being

indexed. These registry keys include:

Key Data Type Recommended Setting Comments

FilterDelayInterval REG_DWORD 3 When fewer than FilterRemainingThreshold documents remain to be filtered, Indexing Service waits for the filter delay interval before indexing them. This is done to minimize collisions with other applications that repeatedly write to files. If your applications don't have this behavior (for example you use xcopy to get files to your system instead of Microsoft Word) you can safely make this value three seconds.

FilterRemainingThreshold REG_DWORD 5 This is the number of documents that remain to be filtered below which FilterDelayInterval is honored.

FilterRetries REG_DWORD 2 This is the number of times Indexing Service will retry a failed document. Don't set this to 1. Due to a bug the true value for this is the value you set minus one. 2 tells IS to retry documents only once.

FilterRetryInterval. REG_DWORD 2 This is the minutes Indexing Service waits between attempts a delayed retry (as opposed to a normal retry) to a document that failed.

DelayedFilterRetries REG_DWORD 1 This is the number of delayed retries for failed documents. The value indicates that delayed documents will only be retried once.

USNReadMinSize REG_DWORD 0 This tells IS to pick up all changes from the USN Journal as soon as they occur, so changes to the disk are picked up as soon as possible.

Thank You

George Cheng

Microsoft Application Center & Index Server Support

Note: This article has no warranties implicit or explicit.

All the content is given on the "as is" basis and the user takes full responsibility for its use and assumption.

Microsoft Corporation Copyright 2004

All Rights Reserved

| From: "CJF2000" <chris1307@fasttoyota.com>
| Newsgroups: microsoft.public.inetsrv.indexserver
| Subject: Slow Catalog Build
| Lines: 9
| X-Priority: 3
| X-MSMail-Priority: Normal
| X-Newsreader: Microsoft Outlook Express 6.00.2800.1409
| X-MimeOLE: Produced By Microsoft MimeOLE V6.00.2800.1409
| Message-ID: <U6fBc.96\$8S5.90@newsfe2-gui.server.ntli.net>
| Date: Sun, 20 Jun 2004 12:09:34 +0100
| NNTP-Posting-Host: 81.99.69.211
| X-Complaints-To: <http://www.ntlworld.com/netreport>

RE: Slow Catalog Build

microsoft.public.inetserver.indexserver: RE: Slow Catalog Build

| X-Trace: newsfe2-gui.server.ntli.net 1087733300 81.99.69.211 (Sun, 20 Jun 2004 12:08:20 GMT)

| NNTP-Posting-Date: Sun, 20 Jun 2004 12:08:20 GMT

| Organization: ntl Cablemodem News Service

| Path:

cpmsftngxa10.phx.gbl!TK2MSFTNGXA01.phx.gbl!TK2MSFTNGP08.phx.gbl!newsfeed00.s
ul.t-online.de!t-online.de!fr.ip.ndsoftware.net!216.196.110.149.MISMATCH!bor
der2.nntp.ams.giganews.com!nntp.giganews.com!news-out.ntli.net!newsrout1.ntl
i.net!news-in.ntli.net!ntli.net!newspeer1-win.server.ntli.net!newsfe2-gui.se
rver.ntli.net.POSTED!53ab2750!not-for-mail

| Xref: cpmsftngxa10.phx.gbl microsoft.public.inetserver.indexserver:29188

| X-Tomcat-NG: microsoft.public.inetserver.indexserver

|

| We have Win 2000 server that acts as the company intranet. We are currently

| cataloging a new index with some 500,000 docs.

| If we let the Indexer rip it has a detrimental affect on the IIS server due

| to the disc activity.

| We have now scheduled the indexer to run "out of hours" however the catalog

| build is still has not complete after 2 weeks.

| Is this a feature of the indexer or is there something else going on!

|

|

|