

UNDERSTANDING THE SEARCH ADMIN PAGE

Michele Buccarello

19/08/13

HOW TO USE THE PAGE

Sometimes it happens that the search application don't work properly for specific application such as wikis or blogs. To understand how many applications are impacted by search problems there is a useful page to check index status and the search status for every application in the different type of topologies, cluster and single node.

For accessing the page you must put your administrator in the search-admin role of the search application, after you add the user, you can check the page from this url:

- <http://connections45.example.com/search/serverStatus>

The serverStatus page is divided in to different sections:

- **Checking the installed services:** This section explains how many features are enabled in the installation.

Checking the installed services

The following service is installed: forums
The following service is installed: blogs
The following service is installed: wikis
The following service is installed: profiles
The following service is installed: dogear

- **Checking the resume tokens:** This section explains what is the last update for every application index.

Checking the resume tokens

The resume token for: files
The resume token for: communities Wed Jul 24 03:00:15 EDT 2013
The resume token for: profiles Wed Jul 24 02:31:00 EDT 2013
The resume token for: dogear Wed Jul 24 02:31:00 EDT 2013
The resume token for: status_updates Wed Jul 24 02:47:18 EDT 2013
The resume token for: activities Wed Jul 24 02:31:00 EDT 2013
The resume token for: calendar

- **Checking that the seedlists are available:** This section explains if every seedlist are reachable via http for all applications.

Checking that the seedlists are available

Checking that the seedlists are available files

Checking that the seedlists are available communities

Checking that the seedlists are available profiles

Checking that the seedlists are available dogear

Checking that the seedlists are available status_updates

Checking that the seedlists are available activities

The seedlist could not be obtained for: calendar

Checking that the seedlists are available blogs

Checking that the seedlists are available wikis

Checking that the seedlists are available forums

- **Checking that the schedule calendars have been configured correctly:** This section explains if the scheduler for calendar is enabled and running correctly.

Checking that the schedule calendars have been configured correctly

Schedule calendar check passed.

- **Checking that the WebSphere variables have been set:** This section explains if the WebSphere Application Server variables are set correctly

Checking that the WebSphere variables have been set

Variable: SEARCH_DICTIONARY_DIR /opt/ibm/IC45/shared/search/dictionary

Variable: SEARCH_INDEX_SHARED_COPY_LOCATION /opt/ibm/IC45/shared/search/staging

Variable: SEARCH_INDEX_DIR /opt/ibm/IC45/Connections/data/local/search/index

Variable: SEARCH_IDBC_DRIVER_HOME /opt/ibm/IDBC

Variable: SEARCH_SEEDLIST_TIMEOUT \${SEARCH_SEEDLIST_TIMEOUT}

Variable: FILE_CONTENT_CONVERSION /opt/ibm/IC45/shared/search/stellent/ids/solex/port/exporter

Variable: SEARCH_INDEX_BACKUP_DIR /opt/ibm/IC45/Connections/data/local/search/backup

Variable: CRAWLER_PAGE_PERSISTENCE_DIR /opt/ibm/IC45/Connections/data/local/search/persistence

- **Checking for enabled indexing tasks:** This section explains if the common running tasks are enabled and running

Checking for enabled indexing tasks

The following task is enabled: nightly-optimize-task

The following task is enabled: 15min-search-indexing-task

The following task is enabled: nightly-sand-task

The following task is enabled: 20min-file-retrieval-task

- **Checking the index folder for required files:** This section explains if the search folder (`data/local/search/index`) contains the two files `CRAWLING_VERSION` and `INDEX.READY`

Checking the index folder for required files

The following files are in the search index folder: `CRAWLING_VERSION.1374649215622`

The following files are in the search index folder: `INDEX.READY`

- **Reviewing the logs for errors:** This section explains the errors in the `SystemOut.log`, normally this log is present in the `Infracluster_Server1`.

Reviewing the logs for errors

Review the following error messages: Your path variables are not set up correctly for indexing. Run "env" from the terminal to verify the following paths (Note these are example paths).
`PATH=$PATH:/opt/IBM/LotusConnections/data/local/search/stellent/dcs/oiexport`
`LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/opt/IBM/LotusConnections/data/local/search/stellent/dcs/oiexport`
 Make sure that the paths exist
 Verify that the WebSphere variable `FILE_CONTENT_CONVERSION` is set up correctly, pointing to:
`/opt/IBM/LotusConnections/data/local/search/stellent/dcs/oiexport/exporter`
 This variables should be included in the `setupCmdLin.sh` file on each node

- **Reviewing the configuration files:** This section explains if the files `LotusConnections-config.xml`, `SystemOut.log`, `search-config.xml` and `search-config.xsd` are in their own directories.

Reviewing the configuration files

The following configuration files have been found: `LotusConnections-config.xml`

The following configuration files have been found: `SystemOut.log`

The following configuration files have been found: `search-config.xml`

The following configuration files have been found: `search-config.xsd`

- **Checking the number of dictionaries that have been enabled:** This section explains how many dictionaries are enabled, more dictionaries improve the search quality.

Checking the number of dictionaries that have been enabled

The number of dictionaries enabled is: 1

- **Checking that the nodes are in sync:** This section explains if all nodes are present in the table `homepage.sr_index`, the number of nodes dependent on the connections topologies.

Checking that the nodes are in sync

No nodes are blocked

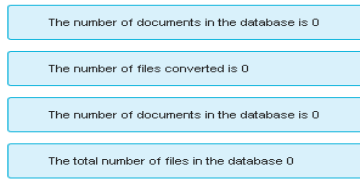
All the nodes are in date

All nodes are in sync

- **Checking the database records:** This section explains some information message "*The number of documents in the database*" is the value of how many records are in the table `Homepage.sr_index_docs`, "*The number of files converted is*" explains the number of files that have been marked as converted in the table `Homepage.sr_filescontent`, "*The number of documents in the database is*" explains the number of files that have been marked as unconverted in the table `Homepage.sr_filescontent`, "*The total number of files in the database*" explains the

total number of files in the table Homepage.sr_filescontent.

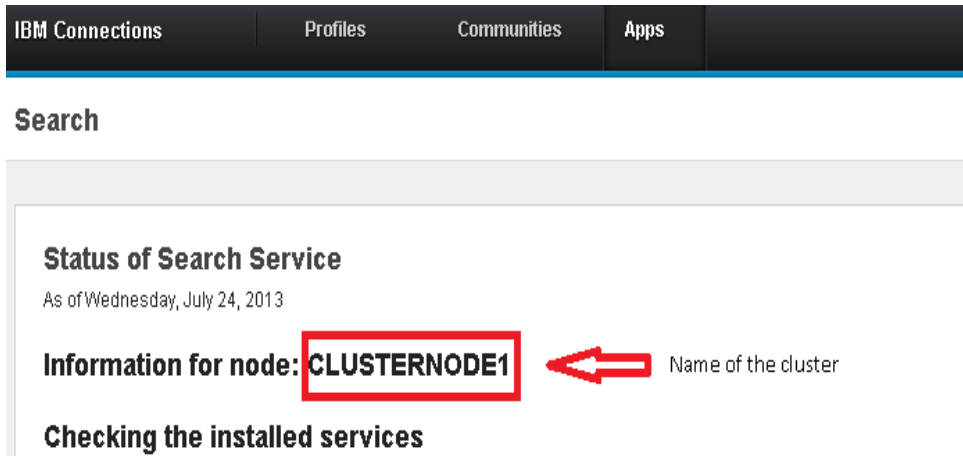
Checking the database records



BEST PRACTICES FOR SEARCH

In this section you can find some best practices to manage the search index for every applications in single node and multiple node environment, the two difference in multiple node environment are:

- Every time you need to run administrative jython command, you need to select what is the node where you want to run the command.
- The search admin page shows you what is the node where you run the checks and their own files and systemout.log



The rule you must follow to avoid problems with the search application in single and multinode are:

- You must copy the search conversion tool from the shared directory in every local node. Normally you can find the search conversion tool in the directory “/SHARED-DIRECTORY/data/shared/search/stellent”. Copy the stellent folder in the directory “/opt/ibm/Connections/data/local/search/”. After the copy you must change the WebSphere environment variable “FILE_CONTENT_CONVERSION” from the shared directory path to local path.
- For linux platform you must add two export lines in the setupCmdLine.sh located in the “/opt/ibm/WebSphere/AppServer/bin”

Example command:

```
export PATH=$PATH:/opt/ibm/Connections/data/local/search/stellent/dcs/oiexport
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/opt/ibm/Connections/data/local/search/stellent/dcs/oiexport
```

After this change you need to stop all websphere application server in every node, stop every node agent and **run the startNode.sh on every node**. If you run the restart command from deployment manager, the java process of the node agent don't read the updates of the system environment variables. After the startNode command, you must check if the java process takes the update by checking this file “/proc/node_agent_pid/environ”, if the check is positive clean the SystemOut.log in the InfraCluster_serverN and then start every nodes, after the start is completed check the serverStatus page if there are errors.

- If the resume token sections for some application don't have the date updated or are

corrupted, follow these steps.

1) Create a background index for every applications that need the rebuild index, for example this commands create a fresh index for all applications.

```
execfile("searchAdmin.py")
```

NOTE: When you run the command in a multinodes enviroment you must choose the node

NOTE: Before running the command, you must create the folders `"/tmp/background/backgroundCrawl"`,

`"/tmp/background/backgroundExtracted"`, `"/tmp/background/backgroundIndex"` on the node you choose.

NOTE: During the “creation process”, check the systemout to see when the process is complete. This process can take more time if you have a lot of information in your connections installation.

```
SearchService.startBackgroundIndex("/tmp/background/backgroundCrawl",  
"/tmp/background/backgroundExtracted", "/tmp/background/backgroundIndex", "activities, blogs, calendar,  
communities, dogear, files, forums, profiles, wikis, status_updates", "communitymembership, graph")
```

2) After the index creation process completed follow these steps

SearchService.disableAllTasks(): This command disables the task for preventing the index during the restore process.

SearchService.notifyRestore("true"): This command cleans the nodes information in the homepage database and clean the cache. After running this command, you can check the SystemOut.log to see the information below.

```
[8/5/13 4:12:55:567 EDT] 00000a96 PurgeIndexDbC I
```

```
com.ibm.connections.search.admin.work.PurgeIndexDbCacheWork run CLFRW0798I: Purge of Index database cache has started.
```

```
[8/5/13 4:12:55:582 EDT] 00000a96 PurgeIndexDbC I
```

```
com.ibm.connections.search.admin.work.PurgeIndexDbCacheWork run CLFRW0796I: Purge of Index database cache has finished.
```

```
[8/5/13 4:12:55:584 EDT] 00000a96 PurgeFilesCon I com.ibm.connections.search.admin.work.PurgeFilesContentWork  
run CLFRW0799I: Purge of files content cache has started.
```

```
[8/5/13 4:12:55:586 EDT] 00000a96 SearchTaskPer I
```

```
com.ibm.connections.search.service.tasks.utils.impl.SearchTaskPersistenceHelperService safelyClearFileContentCache  
CLFRW0324I: The files database cache was cleared successfully.
```

```
[8/5/13 4:12:55:587 EDT] 00000a96 PurgeFilesCon I com.ibm.connections.search.admin.work.PurgeFilesContentWork  
run CLFRW0797I: Purge of files content cache has finished.
```

Clean the content of these folders

NOTE: In a multi node enviroment you must do copy and clean operation on every node

NOTE: During this procedure the search application don't work, the suggestion is to plan a maintenance period

```
/opt/ibm/Connections/data/local/search/index
```

```
/opt/ibm/Connections/data/local/search/persistence
```

```
/SHARED_FOLDER/Connections/data/shared/ExtractedText
```

copy the content of the folder backgroundIndex in the folder

```
“/opt/ibm/Connections/data/local/search/index”
```

copy the content of the folder backgroundCrawl in the folder

```
“/opt/ibm/Connections/data/local/search/persistence”
```

copy the content of the folder backgroundExtracted in the folder

```
“/SHARED_FOLDER/Connections/data/shared/ExtractedText”
```

After the copy is completed on every node you must run the **SearchService.reloadIndex()**

command on every node and to do this on every node, you can simply restart the

wsadmin.sh and run the command `execfile("searchAdmin.py")`. After you run this command, the script asks what is the node you want to run the reloadindex command. To check if the reload index is completed successfully, check the SystemOut.log of the server to confirm that the search admin application is installed.

After all nodes has successfully reloaded the index, you must enable the task with the command `SearchService.enableAllTasks()` and check the `SystemOut.log` to confirm that the task runs correctly.

If you have completed all steps in the server status page successfully, you can see all resume token updated correctly and the node sync sections are all green.