Course Guide

IBM FileNet Content Manager 5.2.1: Optimize Search Performance

Course code F284 ERC 1.0

IBM Training

October 2016 edition

Notices

This information was developed for products and services offered in the US.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
United States of America

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml.

Contents

Trademarks v

Course description vi

Agenda viii

Unit 1. Use searches with bulk actions 1-1
Unit objectives 1-2
Actions on Multiple Objects 1-3
Types of Multiple-object Actions 1-4
Actions and Operations 1-5
Perform Bulk Actions 1-6
Perform Batch Operations (1) 1-7
Perform Batch Operations (2) 1-8
Unit summary 1-9
Exercise: Perform searches and bulk actions 1-10
Exercise introduction 1-11

Unit 2. Configure Content Search Services 2-1
Unit objectives 2-2
What is content-based retrieval (CBR)? 2-3
What is a content index? 2-4
What are IBM Content Search Services? 2-5
Starting and stopping IBM Content Search Services 2-6
What is an index area? 2-7
Use affinity groups to control server assignment 2-8
Enable Content Based Retrieval 2-10
Domain level tasks 2-12
IBM Content Search Server console commands 2-13
Instructor demonstration 2-15
Unit summary 2-16
Exercise: Configure a text search server 2-17
Exercise introduction 2-18

Unit 3. Configure index partitions 3-1
Unit objectives 3-2
What is an index partition? 3-3
Configuring Index partitions 3-4
Settability values 3-5
Selecting a string property for an index partition 3-6
Use a date property for a partition 3-7
Instructor demonstration 3-8
Unit summary 3-9
Exercise: Configure index partitions 3-10
Exercise introduction 3-11

Unit 4. Create content-based indexes 4-1
Unit objectives 4-2
4.1. Create content-based indexes 4-3
Create content-based indexes 4-4

Topics 4-5
Index areas and full-text indexes 4-6
Index area states 4-7
Index area settings 4-9
Full text index settings 4-10
Automatic activation of index areas 4-12
Enable Content Based Retrieval 4-14
Object store level tasks 4-16
Create an index area object 4-17
Index jobs 4-19
How to run a CBR search 4-21
CBR SQL query format 4-22
CBR enabled properties 4-23
Guidelines 4-24
4.2. Optimize CBR queries 4-25
Optimize CBR queries 4-26
Topics 4-27
Combined CBR ad database searches 4-28
How the mixed query works 4-29
CBR Query Optimization 4-30
Configuration 4-31
Handling requests for ranked results 4-32
Configure CBR query optimization options 4-33
Instructor demonstration 4-34
Unit summary 4-36
Exercise: Create content based indexes 4-37
Exercise introduction 4-38

Trademarks

The reader should recognize that the following terms, which appear in the content of this training document, are official trademarks of IBM or other companies:

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide.

The following are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide:

DB2®	FileNet®	Notes®
Redbooks®	Tivoli®	WebSphere®

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Windows is a trademark of Microsoft Corporation in the United States, other countries, or both.

Other product and service names might be trademarks of IBM or other companies.

Course description

IBM FileNet Content Manager 5.2.1: Optimize Search Performance

Duration: 4.5 hours

Purpose

This course is intended to teach the skills that are needed to configure and optimize searching in IBM FileNet Content Manager 5.2.1.

Audience

This course is intended for IBM FileNet P8 Platform administrators.

Prerequisites

- IBM Content Manager concepts and terminology: Object stores, objects, properties, content.
- Be able to navigate the Administration Console for Content Platform Engine.
- F280 - IBM FileNet Content Manager 5.2.1: Introduction or equivalent skills and knowledge.
- F282 - IBM FileNet Content Manager 5.2.1: Work with Object Metadata or equivalent skills and knowledge.

Objectives

- Perform searches with bulk actions.
- Perform a batch action.
- Configure a text search server
- Select a Property for an Index Partition.
- Configure a String Index Partition.
- Configure a Date Index Partition.
- Configure Content Based Retrieval
- Configure an Index Area
- Check Indexing Logs
- Reindex
- Optimize CBR queries

Contents

- Perform Searches and Bulk Actions
- Configure Content Search Services
- Configure index partitions
- Create content based indexes

Agenda

Note

The following unit and exercise durations are estimates, and might not reflect every class experience.

Day 1

(00:15) Course introduction
(00:30) Unit 1. Use searches with bulk actions
(00:30) Exercise 1. Use searches with bulk actions
(00:30) Unit 2. Configure Content Search Services
(00:20) Exercise 2. Configure a text search server
(00:20) Unit 3. Configure index partitions
(00:20) Exercise 3. Configure index partitions
(00:50) Unit 4. Create content-based indexes
(00:40) Exercise 4. Create content-based indexes

Unit 1. Use searches with bulk actions

Estimated time

00:30

Overview

Occasionally, you must administer many documents at one time. You can find the documents by using a search, and then update them by using a bulk action or a batch operation.

How you will check your progress

- Complete lesson exercises.

References

IBM Knowledge Center: P8 Platform

http://www.ibm.com/support/knowledgecenter/SSNW2F

IBM Training

Unit objectives

- Use bulk actions to modify multiple documents.

Use searches with bulk actions

Figure 1-1. Unit objectives

IBM Training

Actions on Multiple Objects

- Administrators sometimes must change multiple objects.
  - Possibly thousands of objects are affected.
  - It is inefficient to change objects individually.
- You can change multiple objects by using bulk actions and batch operations

Figure 1-2. Actions on Multiple Objects

Help path

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Changing objects>Handling bulk processing with queries

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/p8pcc193.htm

IBM Training

Types of Multiple-object Actions

- Bulk actions
  - Performed on objects during a search.
  - Affect all objects that the search returns.
  - Can be combined to perform several actions at the same time.
  - Must be enabled before search.
- Batch operations
  - Accessed from searches and other containers.
  - Affect only selected objects.
  - Can perform only one action at a time.

Figure 1-3. Types of Multiple-object Actions

Help paths

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Changing objects>Handling bulk processing with queries>Applying bulk actions to selected search results

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/searchandbulkops/sb_start_bulk_operations.htm

IBM Training

Actions and Operations

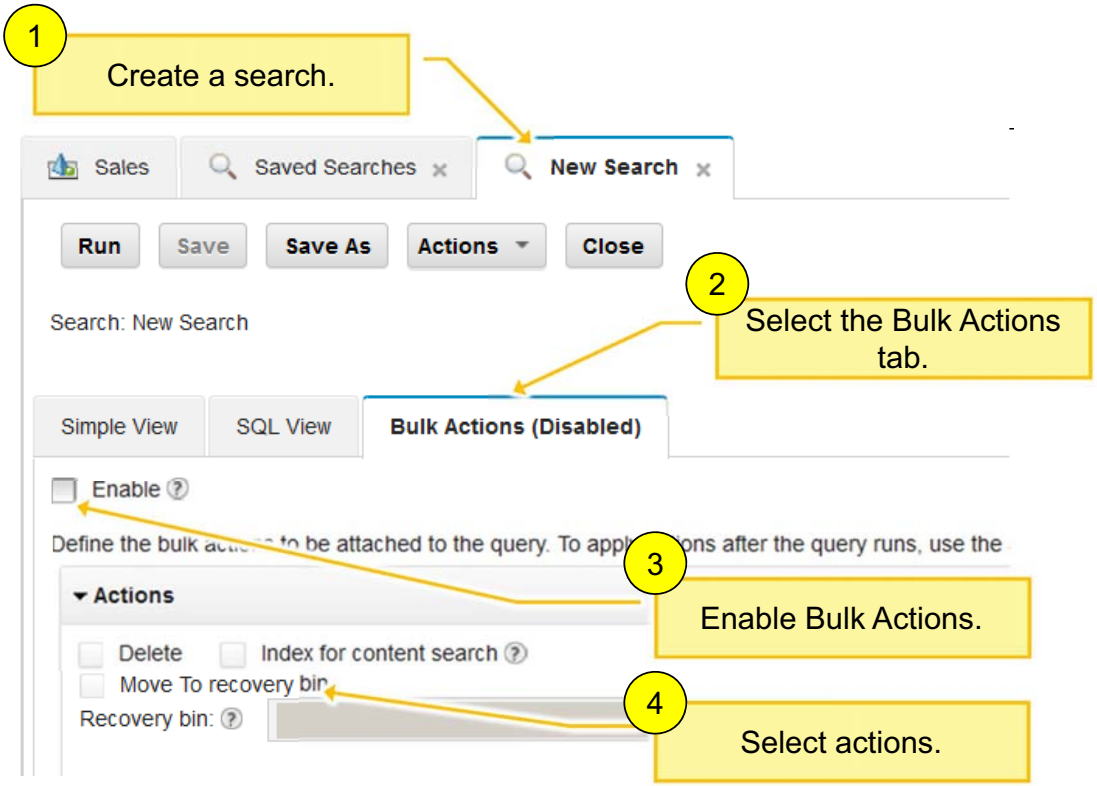
- Actions
  - Delete
  - Index for content search
  - Move to recovery bin
- Versioning
  - Cancel checkout
- Lifecycle
  - Set or clear exception mode
  - Promote or demote
  - Reset
- Referential containment
  - File or unfile
- Replication
  - Add or remove group association
- Script
  - Run a script
- Security
  - Add or remove users or groups
  - Set permissions

Figure 1-4. Actions and Operations

Bulk actions and batch operations can be used to do the same tasks. For batch operations, the script and security tasks are on their own tab.

IBM Training

Perform Bulk Actions



Use searches with bulk actions

Figure 1-5. Perform Bulk Actions

The graphic shows how to start Bulk Actions.

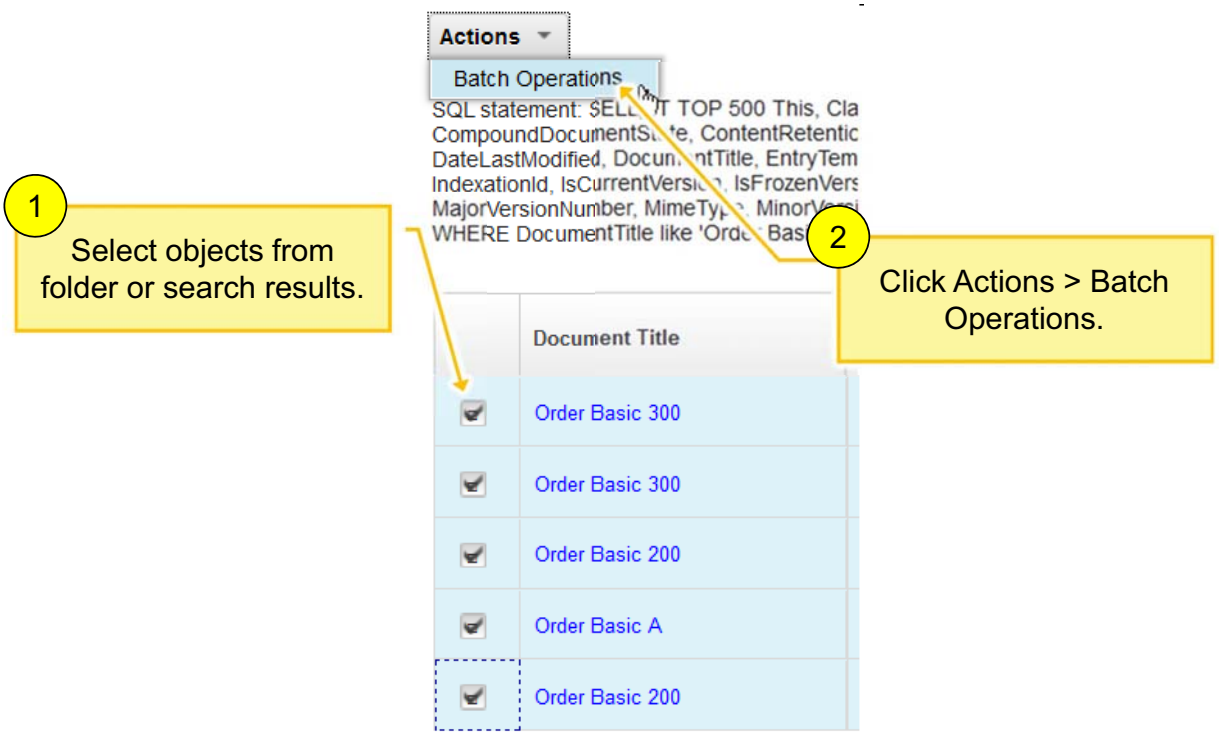
1. Create a search using Administration Console for Content Platform Engine.
2. Select the Bulk Actions tab.
3. Enable Bulk Actions.
4. Select the Bulk Actions to run.

You can select multiple bulk actions, as long as they are not logically inconsistent. For example, you cannot file a document in a folder after deleting it.

Bulk actions are disabled by default to prevent unintended actions, such as deleting objects.

IBM Training

Perform Batch Operations (1)



Use searches with bulk actions

Figure 1-6. Perform Batch Operations (1)

The graphic shows the first two steps for performing a batch operation.

1. Select the objects from the folder or search results.
2. Click Actions > Batch Operations.

IBM Training

Perform Batch Operations (2)

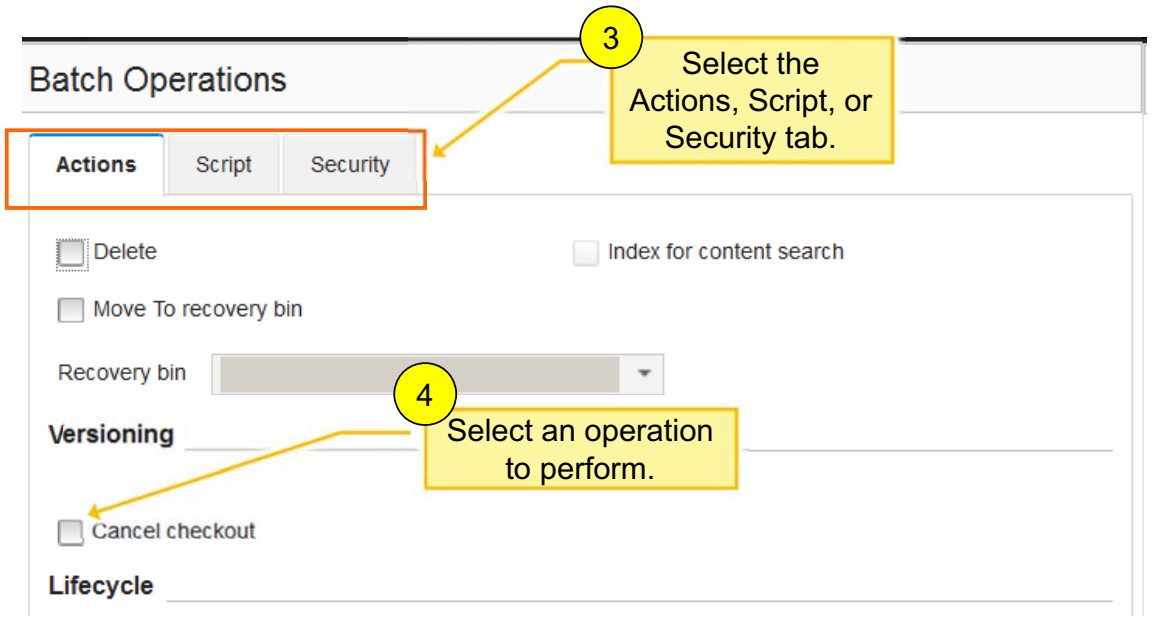


Figure 1-7. Perform Batch Operations (2)

The graphic shows the 3rd and 4th steps for performing a batch operation.

3. Select the Actions, Script, or Security tab, depending on the type of action you want to perform.

4. Select the operation to perform. The operation depends on which tab is selected.

When you select an operation, all of the other operations are disabled.

IBM Training

Unit summary

- Use bulk actions to modify multiple documents.

Figure 1-8. Unit summary

IBM Training

Exercise: Perform searches and bulk actions

Use your student system and the Course Exercises guide to complete the exercise.

Figure 1-9. Exercise: Perform searches and bulk actions

IBM Training

Exercise introduction

- Use bulk actions to modify multiple documents
  - Create a search for Marketing documents
  - Run bulk action to modify security
- Practice using bulk actions
  - Cancel the checkout of documents that are checked out by a user.

Figure 1-10. Exercise introduction

Unit 2. Configure Content Search Services

Estimated time

00:30

Overview

To search for documents based on their content, you must configure an IBM Content Search Services server. This unit shows how to configure IBM Content Search Services server in IBM FileNet Content Manager.

How you will check your progress

- Complete lesson exercises.

References

IBM Knowledge Center: P8 Platform

http://www.ibm.com/support/knowledgecenter/SSNW2F

IBM Training

Unit objectives

- Configure a text search server.

Figure 2-1. Unit objectives

IBM Training

What is content-based retrieval (CBR)?

- Searches content, annotations, and properties for
  - Words, phrases
  - Words in proximity
- Supports most document types:
  - Microsoft Office documents, PDF, HTML, ASCII, and other formats.
  - Can search in XML tags.
- Most Content Platform Engine search utilities support CBR.
  - IBM Content Navigator searches, stored searches, search templates, and Query Builder
- Documents are indexed.
  - IBM Content Search Services provides full-text indexing.

Figure 2-2. What is content-based retrieval (CBR)?

Content-based retrieval (CBR) is also known as full-text searching. Full-text indexing is not related to creating index properties (known as single indexing or database indexing).

CBR finds words or phrases within the text of a document. In addition, CBR searches can find words that are similar to one another. CBR searches can find annotations and string properties that include XML tags. CBR-enabled folders and custom objects can have CBR-enabled string properties that can be used in content-based searches. However, you enable CBR on an entire class, not on specific folders or custom objects.

IBM Training

What is a content index?

- Content index
  - A file that contains pointers to the character-based content in an object store
  - Like an index in the back of a book
  - Instead of page numbers, index stores references to documents.
- What is searched?
  - Content Based Retrieval searches the index file, not the actual documents

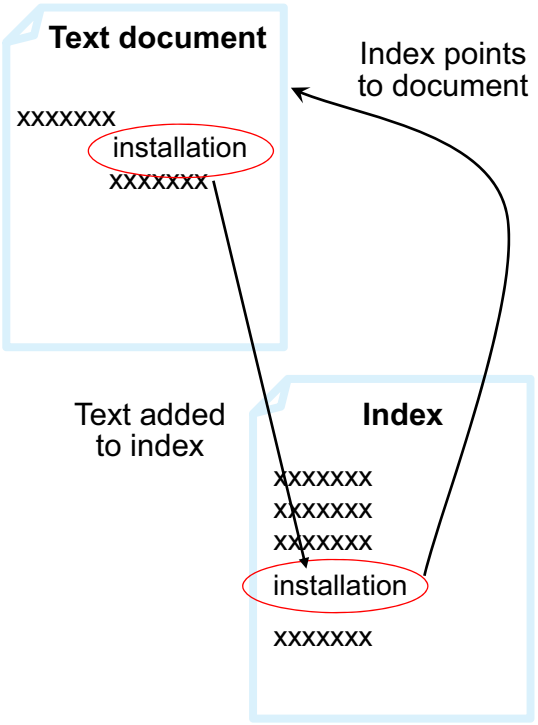


Figure 2-3. What is a content index?

The diagram shows the relationship between an index entry and a document.

IBM Training

What are IBM Content Search Services?

- Indexing and search services for content based retrieval.
- Provided by an IBM Content Search Services server.
- Registered with the FileNet P8 Domain as a Text Search Server.
- Text Search Server Modes:
  - Index only
  - Search only
  - Index and search
- You can change the server status:
  - Enabled
  - Disabled

Figure 2-4. What are IBM Content Search Services?

To distribute workload, a text search server can be registered for indexing, searching, or both indexing and searching tasks. If you have only one server, you must configure it for both indexing and searching.

IBM Training

Starting and stopping IBM Content Search Services

- If IBM Content Search Services is installed on Windows as a service
  - Start > Control Panel > Administrative Tools > Services
  - Can be configured to start automatically.
- Windows (32 and 64 bit) startup and shutdown
  - [css_install_location]\[Server_name]\bin\startup.bat
  - [css_install_location]\[Server_name]\bin\shutdown.bat
- AIX, Linux, Solaris startup and shutdown
  - [css_install_location]\[Server_name]\bin\startup.sh
  - [css_install_location]\[Server_name]\bin\shutdown.sh

Configure Content Search Services

Figure 2-5. Starting and stopping IBM Content Search Services

Help path

FileNet P8 Platform 5.2.1>Administering>Starting and stopping FileNet P8 components>Starting and stopping IBM Content Search Services

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.common.admin.doc/startup_shutdown/content_search_services.htm

IBM Training

What is an index area?

- An index area is a file system folder that contains the indexes that are used for content-based retrieval.
  - Each index area is associated with a particular object store.
- An Index Area **object** Is an object in an object store.
  - It contains the path the index area.
- Multiple index areas can be associated with one object store.
- Load balancing
  - Content Search Servers are assigned to indexes automatically.
  - Provided by a Content Platform Engine algorithm.
- Affinity groups
  - You can override the automatic load balancing algorithm by creating affinity groups.
  - A server in an affinity group serves only indexes that are assigned to that affinity group

Figure 2-6. What is an index area?

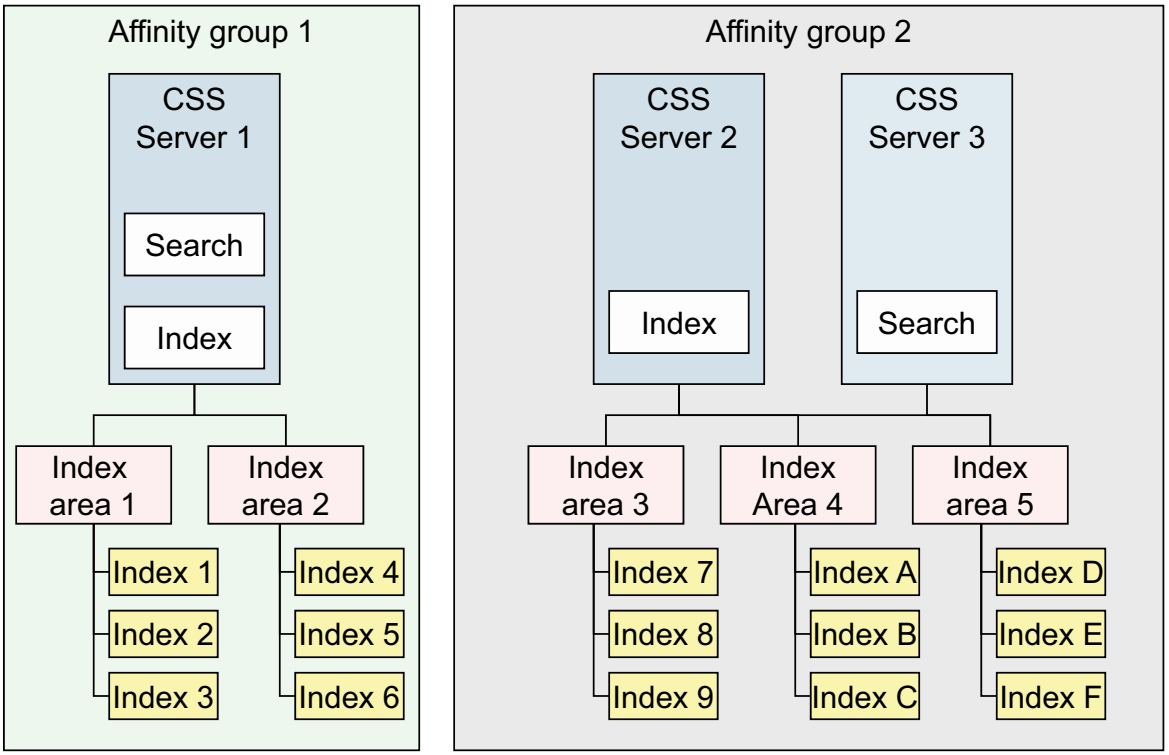
Help path

FileNet P8 Platform 5.2.1>Planning and preparing>Planning and preparing for FileNet P8 installation>Performing the required installation preparation tasks>IT administrator installation tasks>Preparing for IBM Content Search Services>Choosing a load balancing method for IBM Content Search Services servers

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.planprepare.doc/p8ppi255.htm

IBM Training

Use affinity groups to control server assignment



Configure Content Search Services

Figure 2-7. Use affinity groups to control server assignment

Help path

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Indexing object text>Configuring the object indexing process>Preparation phase>Controlling the indexing workload>Affinity groups

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_affinitygroups.htm

The diagram shows how affinity groups are used to limit the assignment of indexes to IBM Content Search Services servers.

By default, all Content Engine servers send data to Content Search Services indexing servers. The amount of indexing work that each server does can be controlled by the FileNet P8 administrator (the maximum threads, batch sizes, and total concurrent text extractions are all exposed in Administration Console for Content Platform Engine).

When multiple Content Search Services servers are configured for search, the Content Engine distributes the workload among them to balance the load.

Indexing is a site-specific job. The location of the indexed data (index area root directory) needs to be in the same site as the object store to prevent performance issues with slow disk access. A search server is always used with the same site designation as the object store.

Each index area allows a new root directory to be specified. The root directory that is used to store indexing data is configured by an administrator by using the Index Area object. The operating system user that runs IBM Content Search Services processes must have read and write access to the directory.

A network share is required to have multiple Content Search Services servers on different computers. Access speed to the disk can impact indexing and search performance. The share must be at the same site as the object store where the index areas are defined. Searching and indexing failures can occur if the disk is not accessible because of a failure.

Administrators can protect against these failures with disk mirroring or other highly available solutions. Lost index data can be regenerated with an index job, but that task can take a significant amount of time to complete. Multiple Content Search Services servers are also useful for failure cases. When one server is not running, another one is used.

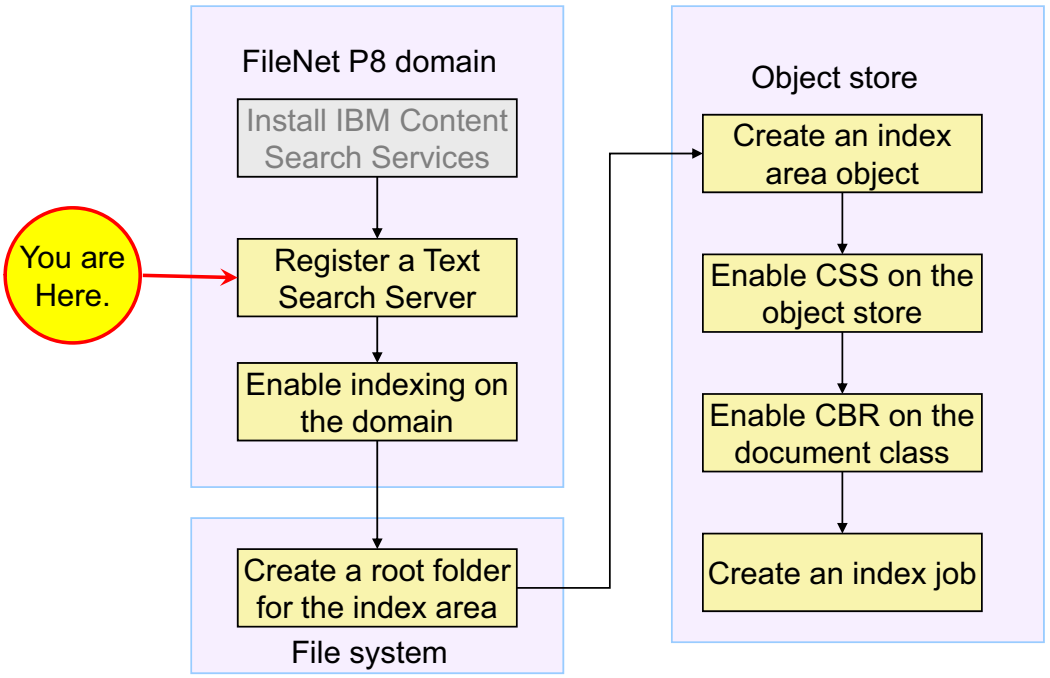
The assignment algorithm considers only active full text indexes. Assignments are determined by server indexing rates. Faster servers are assigned more indexes.

The affinity group improves performance because you can index your data on a disk that is local to IBM Content Search Services.

Without affinity groups, servers are assigned to indexes by using a Content Platform Engine algorithm based on server speed. You can use affinity groups to enhance efficiency by assigning indexes to servers that are collocated. The downside is that Content Platform Engine cannot provide failover. If the local disk that hosts the index area fails, all indexing and search requests to that index area fail.

IBM Training

Enable Content Based Retrieval



Configure Content Search Services

Figure 2-8. Enable Content Based Retrieval

Install IBM Content Search Services

In a production environment, you typically install IBM Content Search Services on its own server. At the completion of the installation wizard, the authentication token is displayed. Record the authentication token so that you can use it when you register the server with the P8 Domain.

Register the IBM Content Search Services server as a Text Search Server on the P8 Domain.

You must have the authentication token when you register the IBM Content Search Services server with the P8 Domain. If you did not get the token during installation, you can request it from the server using a command.

Create an index area on a shared file system.

Avoid the root area where the file stores exist to avoid disk I/O bottlenecks.

Enable indexing at the domain level (or site level).

Sites inherit their settings from the domain, so if you enable indexing at the domain level, it is enabled for all sites unless you choose to override the settings for any particular site.

Create an index area on the object store.

The index area has a pointer to the root directory on the file share. The index area object is also where you configure the index size parameters.

Enable IBM Content Search Services on the object store.

You select the Indexing language and select the partitions if you have them. The index area must exist before you can enable IBM Content Search Services on the object store.

Enable CBR on the document class.

You must enable CBR on each document class that you want to index. You can index subclasses at the same time.

Create an index job.

You initiate the index job on the document class. Schedule index jobs to run during off-peak usage hours. Index jobs can be expensive in time and resources.

IBM Training

Domain level tasks

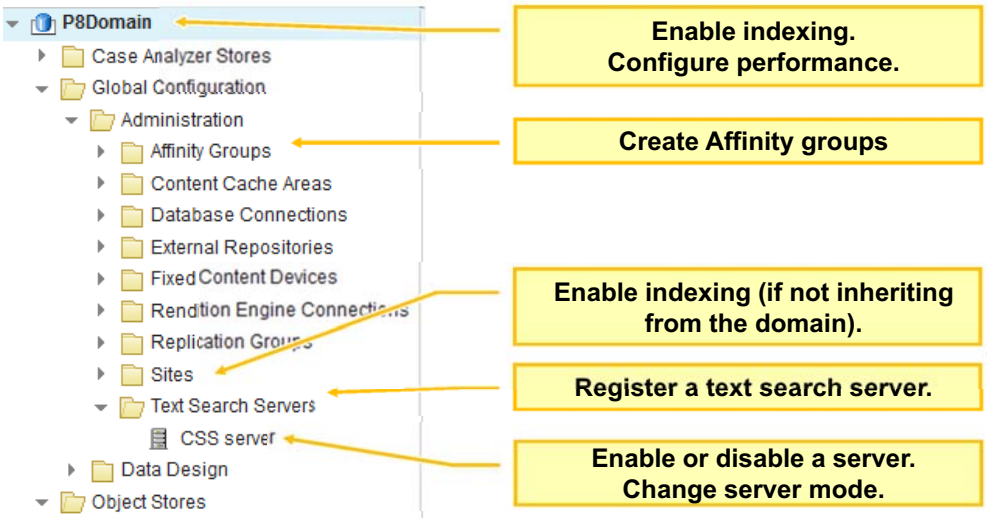


Figure 2-9. Domain level tasks

The diagram shows the P8 Domain tree view in Administration Console for Content Platform Engine. The indicated locations show where the domain level tasks for implementing CBR are performed.

IBM Training

IBM Content Search Server console commands

- **`configTool sysinfo`**
  - Provides information about the server environment
- **`configTool printToken –configPath [local]`**
  - Provides authentication token and encryption key
- **`configTool generateToken`**
  - Generates an authentication token,
- **`adminTool version –configPath`**
  - Provides version information
- Commands run from the CSS Server\bin directory
- configPath is the config folder path for IBM Content Search Services.

Figure 2-10. IBM Content Search Server console commands

Help paths

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Command-line tools: General usage reference>Configuration tool usage

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_configtool_usage.htm

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Command-line tools: General usage reference>Administration tool usage

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_admintool_usage.htm

You use console commands to configure and administer IBM Content Search Services.

Useful Administration tool commands:

- Configure server logging level
- Get software version of the server

Use the Configuration tool commands to do the following:

- Get authentication tokens
- Get system information
- Configure update information for index pipeline statistics

IBM Training

Instructor demonstration

- Create a text search server

Figure 2-11. Instructor demonstration

IBM Training

Unit summary

- Configure a text search server.

Figure 2-12. Unit summary

IBM Training

Exercise: Configure a text search server

Use your student system and the Course Exercises guide to complete the exercise.

Figure 2-13. Exercise: Configure a text search server

IBM Training

Exercise introduction

- Content Search Services is installed, but not yet configured as a text search server on the object store.
- Procedures:
  - Obtain the authentication token
  - Create a text search server
  - Inspect the text search server
  - Verify indexing on P8 Domain
  - Verify indexing at the site level

Figure 2-14. Exercise introduction

Unit 3. Configure index partitions

Estimated time

00:20

Overview

This unit describes how you can make searches faster in some cases creating an index partition. Many documents fall into a few categories that vary on a string property. You can make the searches more efficient by creating an index partition. Users often search for time-sensitive documents. You can make the searches more efficient by creating an index partition that is based on a date property.

How you will check your progress

- Complete the lesson exercises.

References

IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/SSNW2F

IBM Training

Unit objectives

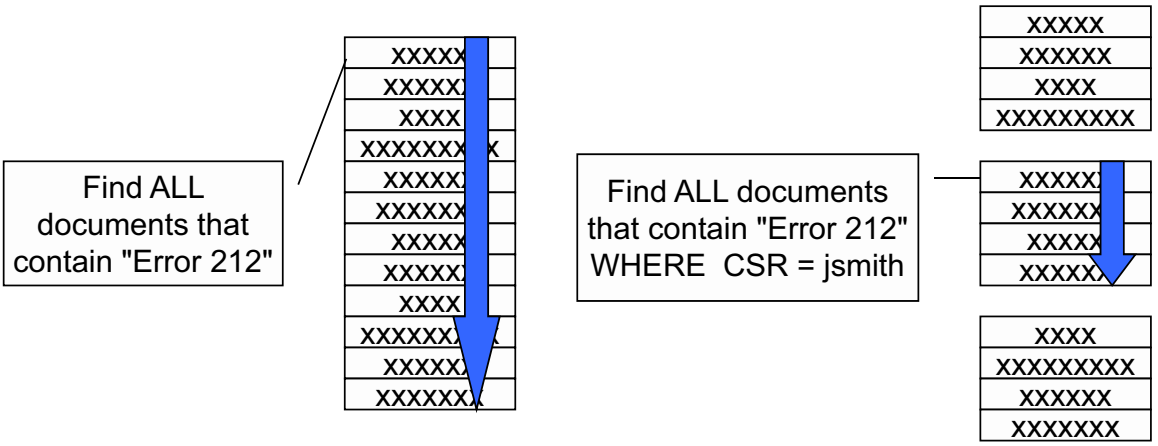
- Select a property for an index partition.
- Configure a string index partition.
- Configure a date index partition

Figure 3-1. Unit objectives

IBM Training

What is an index partition?

- An index partition is a grouping of object index information into separate CBR indexes that are based on object property values.
- It can reduce the amount of index information that must be searched.
- It potentially improves search performance.



Configure index partitions

Figure 3-2. What is an index partition?

Help path

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Indexing object text>Maintaining indexes for the CBR text configuration>Managing index areas and full-text indexes>Partitioning indexes

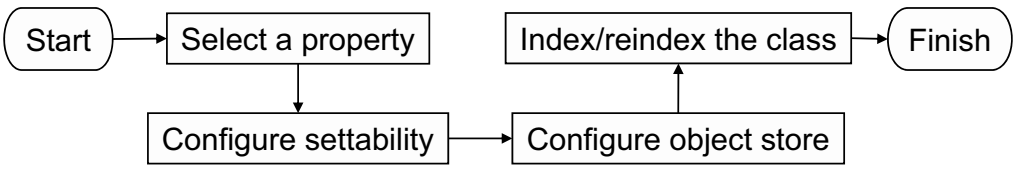
http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_partition_configuring.htm

An index partition can reduce the amount of index information that must be searched by breaking up the index into smaller indexes based on a property value. In the graphic, the first search is for all documents in an object store that contains the phrase "Error 210." Without an index partition, the entire index is searched for the relevant documents. In the second example, the index is partitioned by the "CSR" property. To be effective, the person that does the search must include a value for the CSR property, such as "jsmith." When the search is run, only the index partition for "jsmith" is searched.

IBM Training

Configuring Index partitions

- Create the index partition before you create the index.
  - Otherwise, you must reindex the class.
- Configure index partitions on the object store.
- Only one string-valued property and one date property can be configured for each object store.
- Properties must be set to SETTABLE_ONLY_ON_CREATE to be selected as an index partition.
  - Properties can be set in Administration Console for Content Platform Engine.



Configure index partitions

Figure 3-3. Configuring Index partitions

Settability values

Name	Value	Description
READ_ONLY	3	Indicates that a property is read-only; only the server can set its value.
READ_WRITE	0	Indicates that a property is read/write; you can set its value at any time.
SETTABLE_ONLY_BEFORE_CHECKIN	1	Indicates that you can only set the value of a property before you check in the object to which it belongs.
SETTABLE_ONLY_ON_CREATE	2	Indicates that you can only set the value of a property when you create the object to which it belongs. After you save the object for the first time, the property's value cannot be changed.

Figure 3-4. Settability values

Help path

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Indexing object text>Maintaining indexes for the CBR text configuration>Managing index areas and full-text indexes>Partitioning indexes>Setting index partition properties

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/p8pcc215.htm

In Administration Console for Content Platform Engine, you must use an integer to set the Settability value of the property. The table shows the integer value for each settability setting.

IBM Training

Selecting a string property for an index partition

- To be effective, a string property must be carefully chosen for the index partition.
- Use a string property that has the following characteristics:
  - It must be a custom property.
  - It is often used in searches.
  - It has few, non-unique values.
- Avoid using properties with unique values.
  - A separate partition for every document is inefficient.
- Avoid using properties that are not often used in searches
  - The index partition is only applied when the property is specified in a search.

Figure 3-5. Selecting a string property for an index partition

Help Path

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Indexing object text>Maintaining indexes for the CBR text configuration>Managing index areas and full-text indexes>Partitioning indexes>Index partitioning by string value

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_string_partition.htm

String-partitioned indexes have an associated value for the string partition property of the indexed objects. For example, the associated string value might be email. Such an index contains the index information for any objects with email as the string partition property value. The first object indexed for an index determines the associated string value for the index.

To avoid the creation of superfluous indexes, select a property with few possible values as the string partition property for the object store.

IBM Training

Use a date property for a partition

- An object store has two date partition-related properties:
  - Date Partition Property
  - Date Partition Interval
- Use a Date Partition Property as an index partition if the property is commonly used in searches.
- Set the date property interval to create index partitions for the wanted time interval.
  - Example: Set the interval to one month to partition the index by month.
  - Documents that are added during a particular month are indexed in that month's partition.
  - A search that includes a date searches only the documents that were indexed within that month interval.

Figure 3-6. Use a date property for a partition

Help path

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Indexing object text>Maintaining indexes for the CBR text configuration>Managing index areas and full-text indexes>Partitioning indexes>Index partitioning by date range

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_date_partition.htm

Date partition property

The date partition property is the name of a property that takes a date/time value. For example, suppose that you want to partition documents that belong to an email class based on the received date. You might set the date partition property to DateReceived.

Date partition interval

The interval is the length of time in months of the associated date range of the date partitioned indexes. The possible interval values are 1, 2, 3, 4, 6, 12, 60, or 120. (The default value of null or a value of zero indicates that partitioning is disabled).

IBM Training

Instructor demonstration

- Configure an index partition

Figure 3-7. Instructor demonstration

Configure an index partition.

1. Log on to ACCE as p8admin.
2. On the Sales object store, open the po_number property template.
3. Change the settability value to 2.
4. Save.
5. Open the Text Search tab of the Sales object store.
6. Select po_number as the string property index partition.

IBM Training

Unit summary

- Select a property for an index partition.
- Configure a string index partition.
- Configure a date index partition

Figure 3-8. Unit summary

IBM Training

Exercise: Configure index partitions

Use your student system and the Course Exercises guide to complete the exercise.

Configure index partitions

Figure 3-9. Exercise: Configure index partitions

IBM Training

Exercise introduction

- Configure a string index partition
  - Select a string property
  - Change the property settability option
  - Create the index partition
- Configure a date index partition
  - Practice: optional exercise

Figure 3-10. Exercise introduction

Unit 4. Create content-based indexes

Estimated time

00:50

Overview

In order for users to be able to perform content-based searches, there must exist content-based indexes.In this unit you learn how to create and configure a content-based index area and to start an index job. You also learn how to configure content-based retrieval searches to run more efficiently when users search for documents with a mixed query that includes both metadata and content search elements.

How you will check your progress

- Complete lesson exercises.

References

IBM Knowledge Center: P8 Platform

http://www.ibm.com/support/knowledgecenter/SSNW2F

IBM Training

Unit objectives

- Configure CBR.
- Configure an index area.
- Check indexing logs.
- Reindex.
- Optimize CBR queries.

Figure 4-1. Unit objectives

4.1. Create content-based indexes

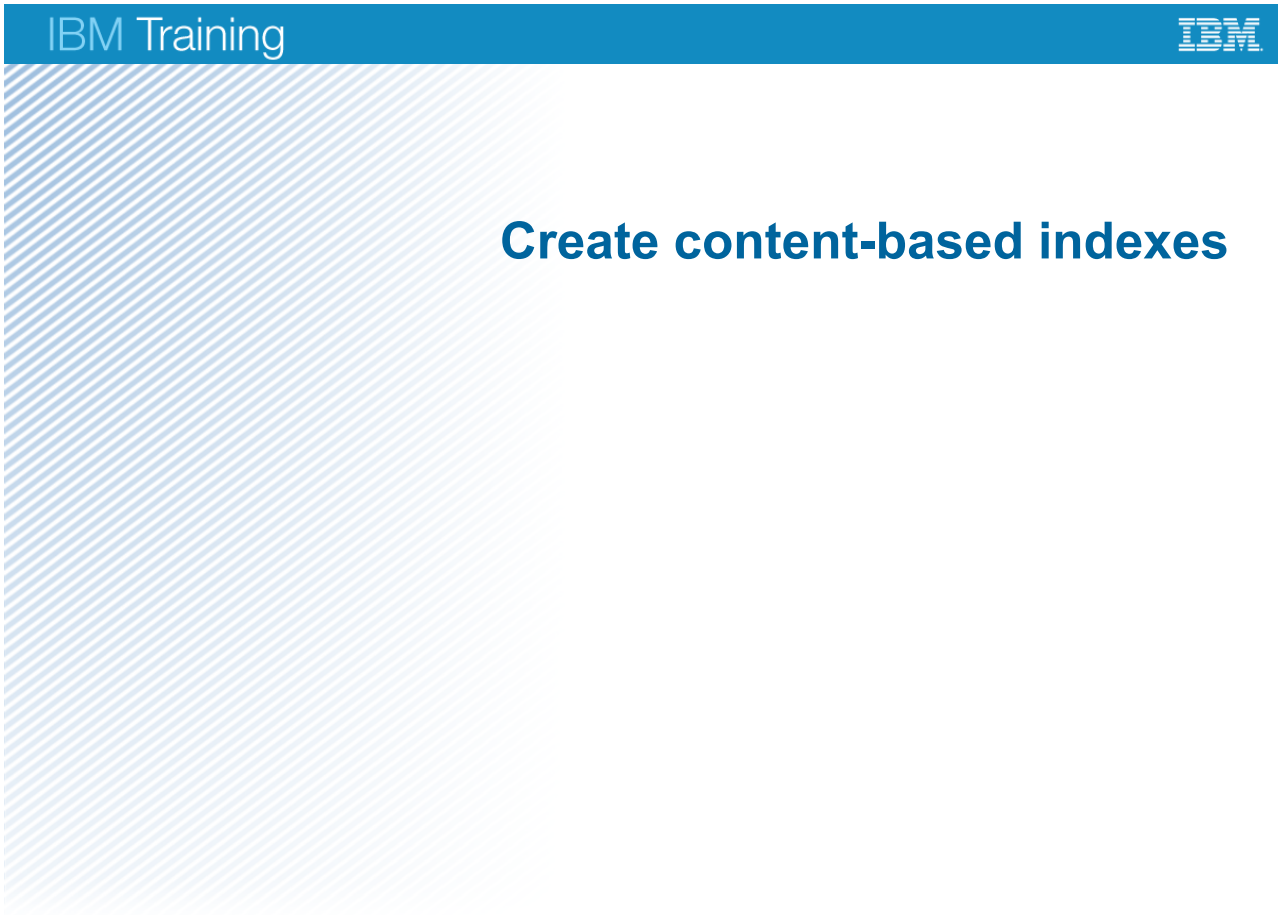


Figure 4-2. Create content-based indexes

IBM Training

Topics

- Create content-based indexes
- Optimize CBR queries

Figure 4-3. Topics

IBM Training

Index areas and full-text indexes

- An object store can have multiple index areas.
  - You create index areas manually.
- An index area can have multiple full-text indexes.
- Objects that belong to a full-text index must share the following characteristics:
  - Belong to the same indexable base class or to some subclass of the indexable base class.
  - Have the same partition property value if a string property partition is configured.
  - Have a partition property value within the same date range if a date partition property is configured.
- Content Platform Engine automatically creates full-text indexes to satisfy these requirements.

Figure 4-4. Index areas and full-text indexes

Help path

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Indexing object text>Maintaining indexes for the CBR text configuration>Managing index areas and full-text indexes>Index areas and full-text indexes

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_indexarea.htm

IBM Training

Index area states

- An index is in one of the following states:
  - Open
  - Closed
  - Full
  - Standby
- Only Open index areas allow full-text index creation and index entry creation.
- You can create an index area in the OPEN or STANDBY state.
- You can manually set an index area to CLOSED.
- Index areas and full-text indexes have separate states.
  - Indexes can be UNAVAILABLE, REPLACING,

Create content-based indexes

Figure 4-5. Index area states

The states of the index determine which actions can be performed on the index area.

Open

- Full-text index creation
- Object index entry creation
- Object index entry update
- Object index entry deletion

Closed

- Object Index entry update
- Object index entry deletion

Full

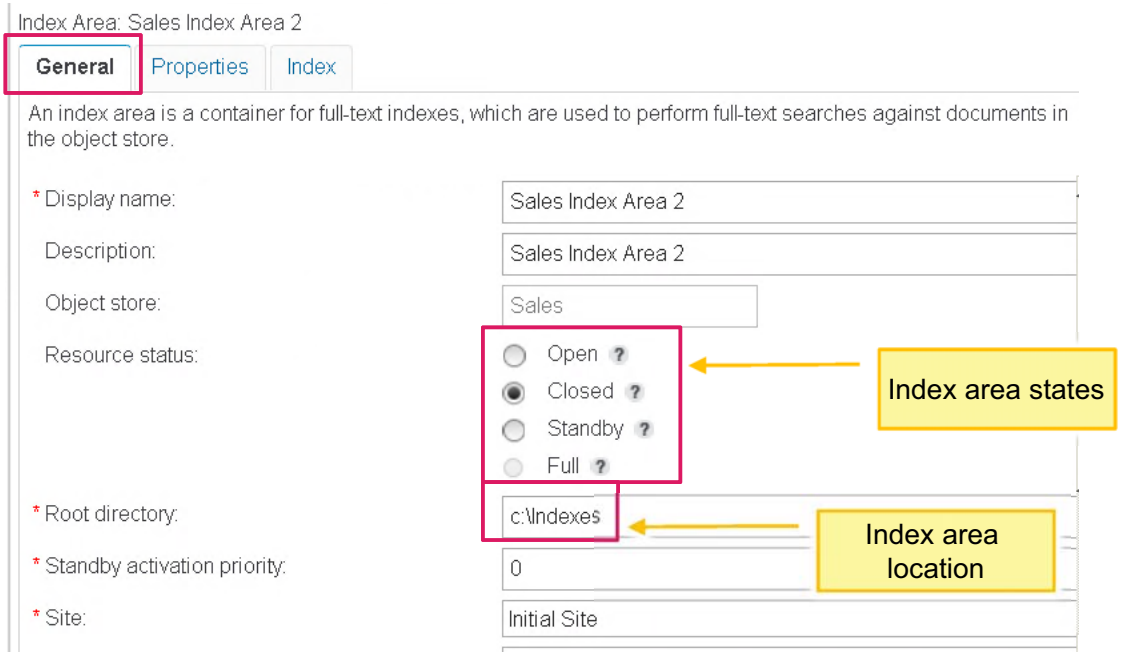
- Object index entry update
- Object index entry deletion

Standby

- Object Index entry update
- Object index entry deletion

IBM Training

Index area settings



Create content-based indexes

Figure 4-6. Index area settings

The graphic shows the index area general settings. From here you can change the resource status, root directory, and other settings.

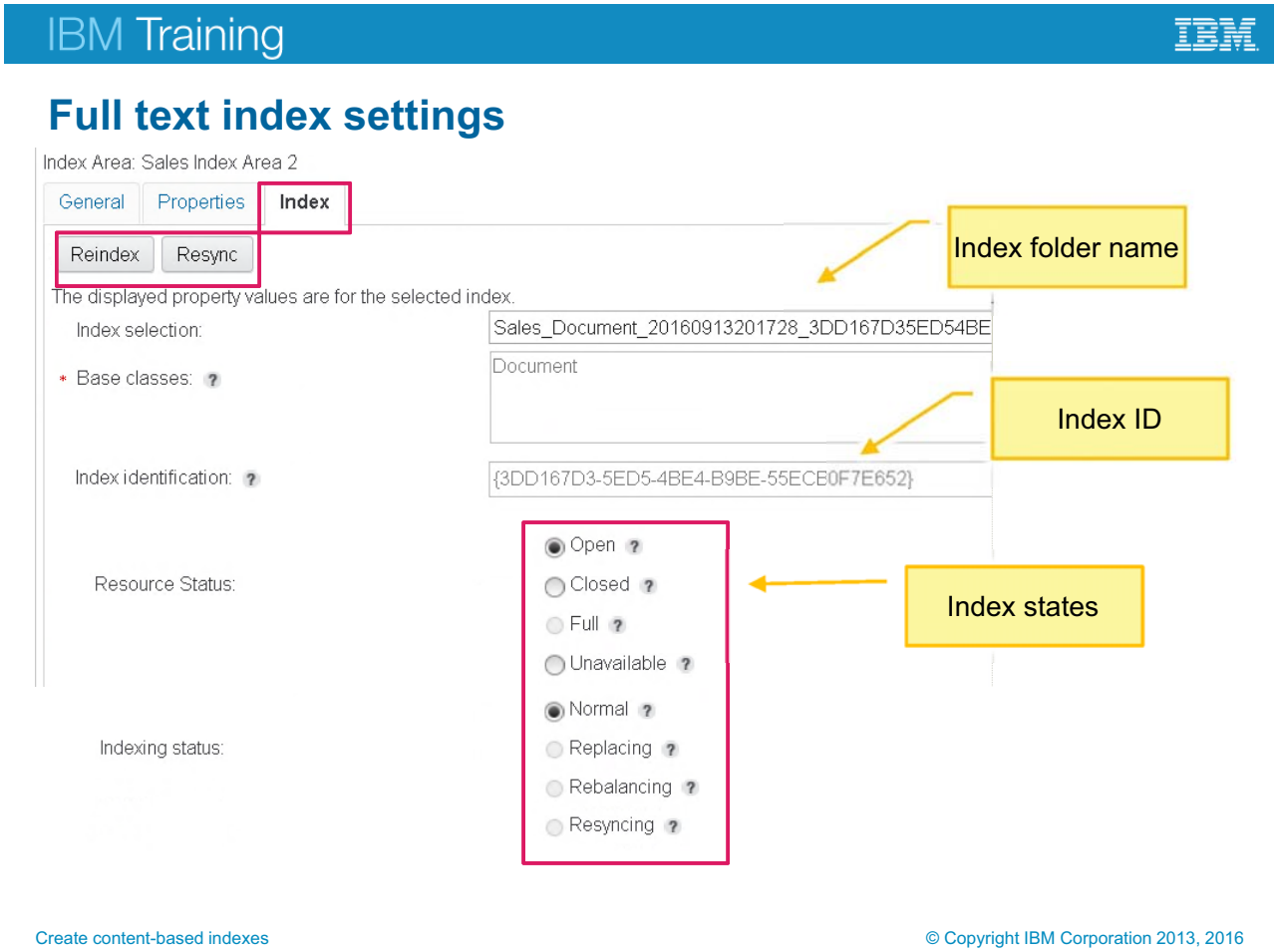


Figure 4-7. Full text index settings

Help paths:

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving Documents>Finding objects with content-based retrieval>Indexing object text>Maintaining indexes for the CBR text configuration>Managing index jobs and index requests>Identifying and resolving indexing failures>Fixing a corrupt full-text index

> http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_corruptindex_fixing.htm

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Indexing object text>Maintaining indexes for the CBR text configuration>Managing index jobs and index requests>Identifying and resolving indexing failures>Fixing an out-of-sync full-text index

> http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/p8pcc212.htm

The graphic shows the settings of the Index tab. The Index tab is where you configure the settings for the full text index. From here, you can also reindex and resync the index.

Reindex: If you suspect that the index has been corrupted, you can set the index to UNAVAILABLE and then reindex. After the index job completes, the index state returns to Open.

Note: If you set the index area to UNAVAILABLE, you cannot change the value again.

Resync: The index can become desynchronized if you restore the object store from a backup. You can resync the index from this interface. During the resynchronizing process, the index state becomes CLOSED. After resynching, it becomes open.

The system updates the Indexing status. For example, during a resynchronizing operation, the resource status is CLOSED and the indexing status is RESYNCING.

IBM Training

Automatic activation of index areas

- You set the capacity of an index area by defining limits.
  - When an index area or full-text index reaches the maximum, its status changes from OPEN to FULL.
- Automatic activation
  - When an index area becomes FULL, Content Platform Engine activates an index area that is in STANDBY.
- Activation priority
  - You can specify the activation priority.
  - Index areas with the highest activation priority are activated first.
- Result:
  - Constant number of index areas with an OPEN status.

Figure 4-8. Automatic activation of index areas

Help paths

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Indexing object text>Maintaining indexes for the CBR text configuration>Managing index areas and full-text indexes>Controlling index area use>Prioritizing index areas for automatic activation

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_indexarea_autoactivating.htm

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Indexing object text>Maintaining indexes for the CBR text configuration>Managing index areas and full-text indexes>Controlling index area use>Resource status automatic transitions

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_resourcestatus_autotransitions.htm

Index area limits:

Maximum number of full-text indexes

Full-text index limits:

Maximum number of index entries

Maximum disk space

Both

Reduced number of OPEN index areas can result in reduced indexing rates. Automatic activation of STANDBY indexes prevents the reduction of OPEN index areas. The system maintains a steady number of OPEN index areas without manual intervention.

For automatic activation to work, Both index areas must be on the same object store.

IBM Training

Enable Content Based Retrieval

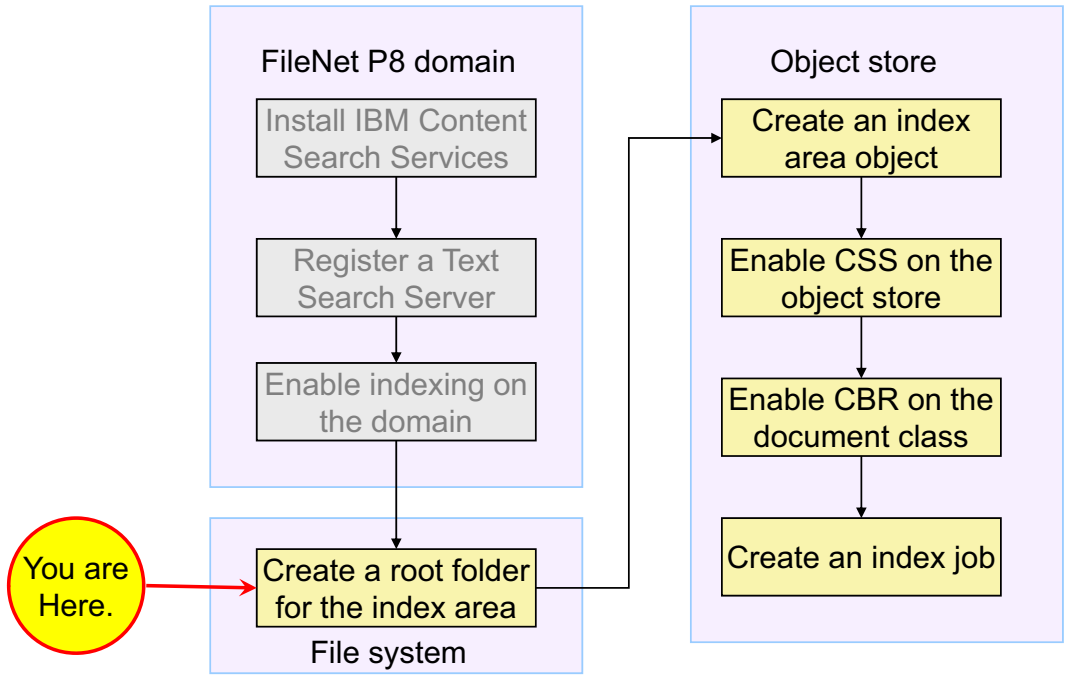


Figure 4-9. Enable Content Based Retrieval

The diagram shows the steps that are required to configure content based retrieval.

Install IBM Content Search Services

In a production environment, you typically install IBM Content Search Services on its own server. At the completion of the installation wizard, the authentication token is displayed. Record the authentication token so that you can use it when you register the server with the P8 Domain.

Register the IBM Content Search Services server as a Text Search Server on the P8 Domain.

You must have the authentication token when you register the IBM Content Search Services server with the P8 Domain. If you did not get the token during installation, you can request it from the server by using a command.

Create an index area on a shared file system.

Avoid the root area where the file stores exist to avoid disk I/O bottlenecks.

Enable indexing at the domain level (or site level).

Sites inherit their settings from the domain, so if you enable indexing at the domain level, it is enabled for all sites unless you choose to override the settings for any particular site.

Create an index area on the object store.

The index area has a pointer to the root directory on the file share. The index area object is also where you configure the index size parameters.

Enable IBM Content Search Services on the object store.

You select the Indexing language and select the partitions if you have them. The index area must exist before you can enable IBM Content Search Services on the object store.

Enable CBR on the document class.

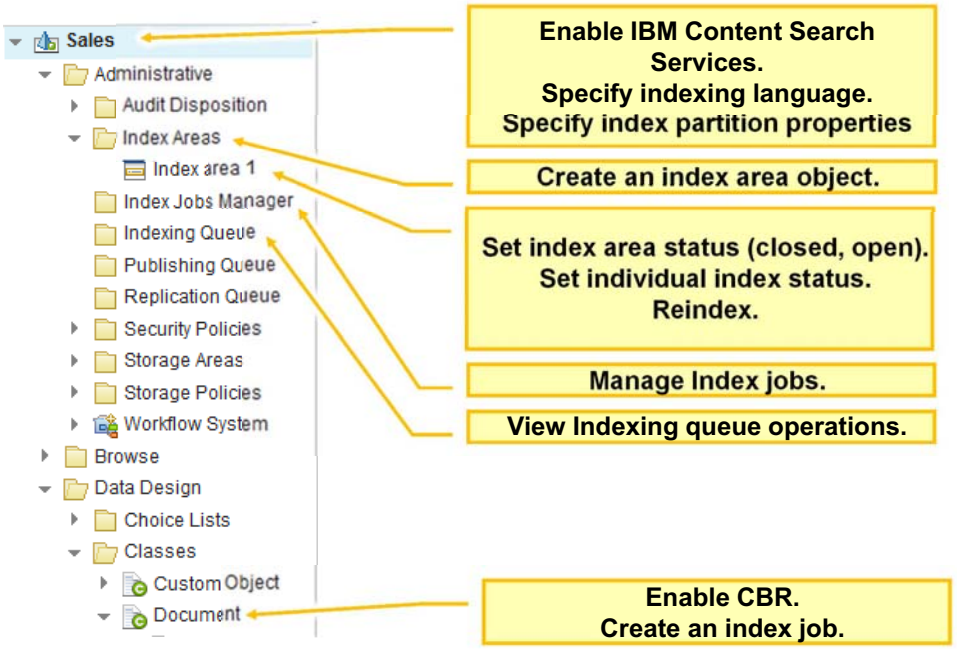
You must enable CBR on each document class that you want to index. You can index subclasses at the same time. Avoid indexing document subclasses that do not need to be indexed. Indexing is expensive in resources and time. Nothing is indexed until you enable CBR on the document class.

Create an index job.

You initiate the index job on the document class. Schedule index jobs to run during off-peak usage hours. Index jobs can be expensive in time and resources.

IBM Training

Object store level tasks



Create content-based indexes

Figure 4-10. Object store level tasks

The graphic shows the object store tree view in Administration Console for the Content Platform Engine. The locations for object store level tasks for enabling content based indexes are indicated.

IBM Training

Create an index area object

- You create a shared folder in the file store.
  - CSS Server service user must have authorization.
- Create an index area object that references the folder.
- Configure index area properties
  - Root directory
  - Maximum index count
  - Index maximum size
  - Resource status
  - Affinity group
  - Standby activation priority

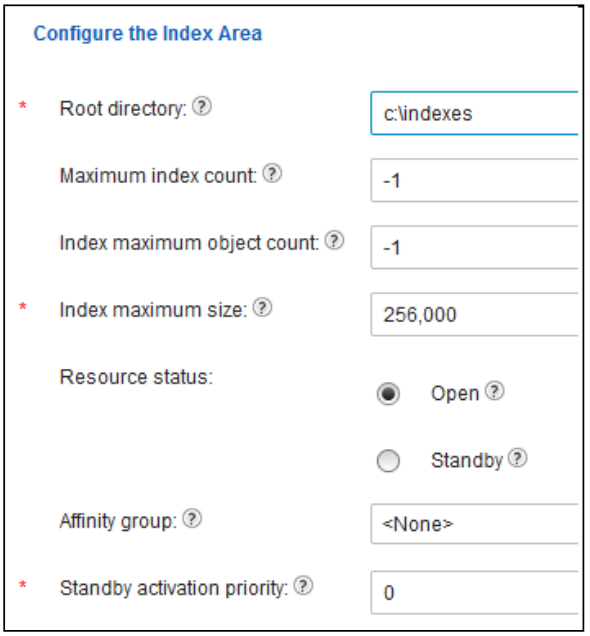


Figure 4-11. Create an index area object

Help path

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Indexing object text>Maintaining indexes for the CBR text configuration>Managing index areas and full-text indexes>Creating an index area

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.install.doc/p8pin329.htm?pos=2

When you create the index area folder on the network share, make sure that the folder grants full control to the CSS Server service account, otherwise, the indexes cannot be created.

Index area properties

- Root directory: the location of the index area in the file share that contains the indexes.
- Maximum index count: A threshold that governs when the status of the index area is automatically changed to full. The status is changed when the number of full-text indexes equals or exceeds this threshold. A value of -1 indicates that no threshold applies.
- Index maximum size: A threshold that governs when the status of the index area is automatically changed to full. The status is changed when the number of indexed objects that

are assigned to the index equals or exceeds this threshold. A value of -1 indicates that no threshold applies.

- Resource status: Indicates the status of the index area.

You can set the status directly if the index area is not full.

- Affinity group
- Standby activation priority

IBM Training

Index jobs

- An index job represents one or more index requests for Content Platform Engine objects.
- You create an index job on the document class that you want to index.
  - The document class must first be CBR-enabled.
- The index job is queued and then runs.
  - Indexing occurs automatically after the index job is queued.
- Schedule to run index jobs during non-peak hours:
  - P8Domain > Text Search Subsystem tab > Schedule
- Use the Index Jobs Manager to check index progress
- Index errors are logged in the P8_Server_ error.log files.

Create content-based indexes

Figure 4-12. Index jobs

Help path

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Indexing object text>Maintaining indexes for the CBR text configuration>Managing index jobs and index requests>Running index jobs

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_indexjob_running.htm

To create an index job in Administration Console for Content Platform Engine, open the class definition and then select Action > Index Class for Content Search (include subclasses).

Scheduling index jobs

Index jobs can be time and resource expensive. To avoid inconvenience to users, schedule the index job during non-peak usage hours.

Use the Index Jobs Manager to perform the following actions:

View job status.

Pause or resume jobs.

Cancel jobs.

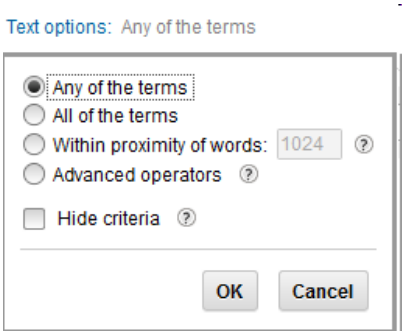
Delete jobs.

View index requests.

IBM Training

How to run a CBR search

- From IBM Content Navigator
  - Create a search on the Search page.
  - Enter words into the *Find items with the following terms* field.
  - Select text options.
- From Administration Console for Content Platform Engine
  - Open the object store page.
  - Select the Search link
  - Open the SQL Query tab.
  - Enter the SQL query



Text options in IBM Content Navigator CBR search.

Figure 4-13. How to run a CBR search

IBM Content Navigator

IBM Content Navigator changes your CBR search. For example, if you enter the text "Lion tiger" either "Lion AND tiger," or "Lion OR tiger," is sent to the database, depending on the text option that you selected.

The graphic shows the text options that are available in IBM Content Navigator. Selecting "Any of the terms" inserts an OR between search terms. Selecting All of the terms inserts an AND between search terms. With the Proximity option you can search for terms that must be in proximity to one another. You set the number of words that can separate the terms. Use Advanced operators to include more complex queries with mixed expressions.

Users can use IBM Content Navigator to perform routine searches for their content. Administrators typically use Administration Console for Content Platform Engine to find documents in order to perform administration tasks, such as changing metadata or security, or to perform bulk operations.

IBM Training

CBR SQL query format

- You might have to create a CBR query in SQL query format.
- Use this syntax:
  - ```
    SELECT d.This
    FROM Document d
    INNER JOIN ContentSearch c ON d.This = c.QueriedObject
    WHERE CONTAINS(d.*,'lion AND tiger')
    ```
- Substitute your search terms for lion AND tiger
- You can use AND, OR, or NOT in the CONTAINS statement.

Create content-based indexes

Figure 4-14. CBR SQL query format

Help path

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Querying for object text>CBR query syntax introduction

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_cbrquery_introduction.htm

The CBR SQL query format is the standard method for making a CBR query. This query statement is useful if you must use a content search to find documents for administration tasks, such as bulk operations.

Note

You can save the search and then reuse it by changing the highlighted text.
    ```

IBM Training

CBR enabled properties

- You can enable properties for CBR.
  - Property values are included in the full-text index.
- Why enable properties for CBR?
  - A CBR search can provide many query syntax capabilities that a DB search cannot:
  - Examples: Phrase, wildcard, synonyms, language processing, grouping.
  - A property can be combined with other CBR criteria for a more efficient query.
  - Adding a database index might be more expensive than an entry to the full text index.
- Configuration:
  - Class > Property Definition > CBR Enabled [True or False].

Figure 4-15. CBR enabled properties

Help path

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Indexing object text>Setting the CBR-enabled status for a class

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_class_cbrenabling.htm

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Indexing object text>Enabling CBR for a property

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_property_cbrenabling.htm

Guidelines

- When you use IBM Content Navigator to do queries based on document contents, be aware that:
  - The text that is entered into the content field in IBM Content Navigator is augmented before being sent to the database.
  - Boolean expressions, such as AND and OR are added to the query.
- Index only the document subclasses that are going to be searched.
- Schedule indexing jobs for non-peak hours.
- Reindex
  - When an index failure occurs.
  - If you create a partition.

Figure 4-16. Guidelines

4.2. Optimize CBR queries

IBM Training

Optimize CBR queries

Figure 4-17. Optimize CBR queries

IBM Training

Topics

- Create content-based indexes
- Optimize CBR queries

Figure 4-18. Topics

IBM Training

Combined CBR ad database searches

- A search can include both content and relational database criteria:
  - Example: Find all documents with document title LIKE "Test" and that contain the word "Copyright."
- CBR searches can sometimes take too long in these circumstances:
  - Too many full-text search hits.
  - Too few relational database hits.
  - Example: All of the documents contain the word "Copyright," but only one document has "Test" in the document title.
- This search is slow because it retrieves many full-text hits that are not needed.

Figure 4-19. Combined CBR ad database searches

IBM Training

How the mixed query works

- Query statement
  - ```
    Select D.This FROM Document D INNER JOIN ContentSearch
    CS
    ON D.This = CS.QueriedObject
    WHERE D.DocumentTitle LIKE 'Test%'
    AND CONTAINS (*, 'copyright')
    ```
- What happens:
  1. The full text portion of the search retrieves a batch.
  2. The batch of full text hits is inserted into a DB Temp table.
  3. A JOIN query is issued between the Temp and DocVersion tables.
  4. Steps 1-3 are repeated to fill the page.
- Limitation
  - Works poorly if many full text hits and many iterations of the join query are needed to fill the page.

Figure 4-20. How the mixed query works

Help path

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Querying for object text>CBR query syntax introduction

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_cbrquery_introduction.htm
    ```

IBM Training

CBR Query Optimization

- What is Dynamic Switching?
  - For every CBR search, an estimate is first made of the full text hit count based on the full text search criteria.
  - If the estimate is larger than a configured threshold, the relational database portion of the search is run first (DB First).
  - Otherwise, the full-text search is issued first (CBR First).
  - The database results are then searched against Content Search Services.
- Benefits
  - Much faster searches in the case of few relational results.
  - All of the results can be found quickly in the database.

Figure 4-21. CBR Query Optimization

Help path

FileNet P8 Platform 5.2.1>Developing FileNet P8 applications>Content Engine Development>Content Engine .NET API Reference>FileNet.Api.Core Namespace>IObjectStore Interface>IObjectStore Properties>CBRQueryOptimization Property

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.dev.net.doc/P_FileNet_Api_Core_IObjectStore_CBRQueryOptimization.htm

IBM Training

Configuration

- You set the CBR Query Optimization property on each Object Store
- Default value is Null (not set)
  - Searches are done Content First
- Set to value of 1 (Dynamic switching)
  - Searches dynamically switch
- Set CBR Query Dynamic Threshold property
  - For example: 10000
  - Only used if CBR Query Optimization property is set to Dynamic switching.

Figure 4-22. Configuration

Help path

FileNet P8 Platform 5.2.1>Administering>Administering Content Platform Engine>Retrieving documents>Finding objects with content-based retrieval>Querying for object text>Optimizing CBR query performance

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.admin.tasks.doc/cbr/csscbr_queryperformance_optimizing.htm

IBM Training

Handling requests for ranked results

- Users can request search results to be ordered by rank.
  - Rank order can be calculated only when the search is CBR-first.
  - You specify how ranked results requests are handled.
- Grant always:
  - The server always runs CBR-first searches if rank order is requested.
- Grant conditionally:
  - CBR-first searches are ordered by rank.
  - DB-first searches are not ordered by rank.
- Grant never:
  - Searches are not ordered by rank.
  - This option can boost performance.

Figure 4-23. Handling requests for ranked results

Help path

FileNet P8 Platform 5.2.1>Developing FileNet P8 applications>Content Engine Development>Content Engine .NET API Reference>FileNet.Api.Core Namespace>IObjectStore Interface>IObjectStore Properties>CBRQueryRankOverride Property

http://www.ibm.com/support/knowledgecenter/SSNW2F_5.2.1/com.ibm.p8.ce.dev.net.doc/P_FileNet_Api_Core_IObjectStore_CBRQueryRankOverride.htm

Search results are ordered by rank when you use the ORDER BY clause in a CBR query. Objects are returned in order of search relevance based on term instance frequency. IBM Content Search Services calculates the value of the Rank property for each returned object.

IBM Training

Configure CBR query optimization options

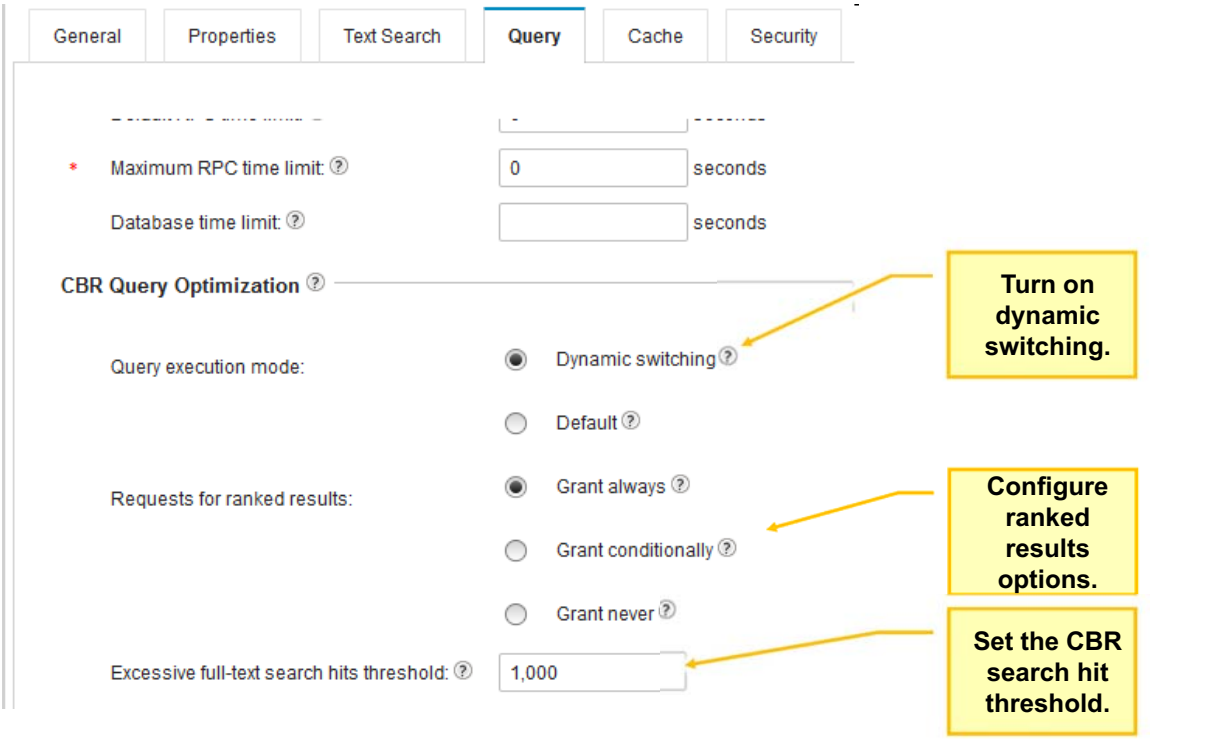


Figure 4-24. Configure CBR query optimization options

IBM Training

Instructor demonstration

- Create an index area
- Create an index job

Figure 4-25. Instructor demonstration

Create an index area

1. Create a folder on C:\ named Indexes. Give P8Admin full control over this folder.
2. In Administration Console, go to Sales > Administrative > Indexes.
3. Create an index area.
4. Configure the index area by specifying the root directory (C:\Indexes).

Create an index job (setup)

1. Setup
   a. Enable IBM Content Search Services on the object store.
      i. Open the Text Search tab of the Sales object store.
      ii. Select a language.
      iii. Check the option to Enable IBM Content Search Services.
      iv. Save.
   b. Enable CBR on a document class.
      i. Select a document subclass, such as Product.

ii. On the General tab, check CBR Enabled.

iii. Save.

2. Create an index job.
   a. Open the document subclass that is CBR-enabled.
   b. Click Actions > Index Class or Content Search (include subclasses).
   c. Click OK.
3. View index job progress (optional)
   a. Go to Sales > Administrative > Index Jobs Manager.
   b. Review the information. Click Refresh if nothing is displayed.

IBM Training

Unit summary

- Configure CBR.
- Configure an index area.
- Check indexing logs.
- Reindex.
- Optimize CBR queries.

Figure 4-26. Unit summary

IBM Training

Exercise: Create content based indexes

Use your student system and the Course Exercises guide to complete the exercise.

Create content-based indexes

Figure 4-27. Exercise: Create content based indexes

IBM Training

Exercise introduction

- Configure CBR
  - Create an index area
  - Enable IBM Content Search Services on the object store
  - Enable CBR on a document class
  - Create an index job
- Configure an index area (optional challenge)
- Check indexing logs
  - Find index entries in the log file
- Reindex (optional challenge)
- Configure CBR query optimization options

Figure 4-28. Exercise introduction

IBM Training