



# Exploring the Collections Dashboard

IBM Content Analytics with Enterprise Search Search Customizer | Analytics Customizer | Log Out | Help | About

Collections System Security

\* Create Collection

### Enterprise Search Collections

	Crawl	Parse and Index	Search	
English and Japanese collection	●	■	●	Actions ▼
Kiso collection	●	■	●	Actions ▼

### Content Analytics Collections

	Parse and Index	Analytics	
Text analytics collection		●	Actions ▼
Hadoop collection		●	Actions ▼

The **Collections** dashboard is where you will likely spend most of your time as an administrator. You use this dashboard to:

- Create collections
- Add content to collections
- Configure options for parsing, analyzing, and indexing content
- Specify how the content can be searched and explored by users, and
- Monitor the status and progress of all of these activities

# Exploring the Collections Dashboard

IBM Content Analytics with Enterprise Search Search Customizer | Analytics Customizer | Log Out | Help | About

Collections System Security

Create Collection

Before you add data that you want to analyze and search, you must create a collection.

### Enterprise Search

	Crawl	Parse and Index	Search	
English and Japanese collection	●	■	●	Actions ▼
Kiso collection	●	■	●	Actions ▼

### Content Analytics Collections

	Crawl	Parse and Index	Analytics	
Text analytics collection	●	■	●	Actions ▼
Hadoop collection	●	■	●	Actions ▼

## Exploring the Collections Dashboard

IBM Content Analytics with Enterprise Search Search Customizer | Analytics Customizer | Log Out | Help | About

Collections System Security

Collections > Create Collection

### Create a Collection

[Learn more](#)

A collection contains the content that you want to query. After you click **OK**, you return to the Collections view. On the Collections view, select your new collection and add content by adding a crawler or importing CSV files.

#### General settings

\* Collection name: My Collection

\* Collection type: [Select an item]  
[Select an item]  
Enterprise search collection  
Content analytics collection

environment from System > Server Configuration > Configure IBM InfoSpr...

[Advanced options](#)

OK Cancel

You must specify a unique name for the collection and select the collection type.

- Create an **enterprise search collection** if your goal is to search content to find relevant documents (that is, find what you know).

- Create a **content analytics collection** if your goal is to explore content to discover patterns, trends, connections, and anomalies in the data (that is, discover what you don't know).

the BigInsights

**IBM** IBM Content Analytics with Enterprise Search Search Customizer | Analytics Customizer | Log Out | Help | About

Collections System Security

**Collections > Create Collection**

### Create a Collection

[Learn more](#)

A collection contains the content that you want to query. After you click **OK**, you return to the Collections view. On the Collections view, select your new collection and add content by adding a crawler or importing CSV files.

#### General settings

\* Collection name:

If you previously added an IBM InfoSphere BigInsights server to your system topology, you can create the index for this new collection on the BigInsights server.

Use IBM InfoSphere BigInsights to help process large amounts of data. (This option is disabled. To enable this option, configure the BigInsights environment from System > Server Configuration > Configure IBM InfoSphere BigInsights Server.)

Document cache:  Enable the document cache (for crawling, indexing, and exporting documents):

Thumbnail generation:  Do not enable thumbnail generation

[Advanced options](#)

OK Cancel

You can accept the default settings for all other collection properties. Click **Advanced options** if you want to do things like:

- Enable security for the collection
- Detect duplicate documents in the collection
- Configure rules to assign documents to categories
- And much more.

## Exploring the Collections Dashboard

IBM Content Analytics with Enterprise Search Search Customizer | Analytics Customizer | Log Out | Help | About

Collections System Security

\* Create Collection

### Enterprise Search Collections

	Crawl	Parse and Index	Search	
English and Japanese collection	●	●	●	Actions ▼
Kiso collection			●	Actions ▼

### Content Analytics Collections

Text analytics collection				Analytics	●	Actions ▼
Hadoop collection	●	●	●			Actions ▼

After you create a collection, it appears on the **Collections** dashboard. The first thing you must do is add content to the collection. To do that, click to expand the collection.

# Exploring the Collections Dashboard

The screenshot displays the 'Content Analytics Collections' dashboard. At the top, there are three tabs: 'Crawl', 'Parse and Index', and 'Analytics'. Below these, two collections are listed: 'Text analytics collection' and 'Hadoop collection'. The 'Hadoop collection' is expanded to show three main stages: 'Crawl and Import', 'Parse and Index', and 'Search and Content Analytics'. Each stage has a progress bar and status indicator. A callout box with a purple background and white text is overlaid on the 'Crawl and Import' stage, providing instructions on how to add content to a collection. The 'Crawl and Import' stage shows a progress bar labeled 'crawling...' with a red square and a green play button. Below it, there are buttons for 'Crawlers' and 'Import'. The 'Parse and Index' stage shows a progress bar labeled 'waiting' with a red square. Below it, there are buttons for 'Export' and 'CSV'. The 'Search and Content Analytics' stage shows a progress bar labeled 'running' with a red square. Below it, there are buttons for 'XML', 'Request ID: 4', 'Description: To export data by ...', and 'Queued export requests'. The 'Export' stage shows a progress bar labeled 'waiting' with a red square. Below it, there are buttons for 'Database' and 'Exported documents: 3,924'. The 'CSV' stage shows a progress bar labeled 'waiting' with a red square. Below it, there are buttons for 'Exported documents: 3,924'. The 'XML' stage shows a progress bar labeled 'initializing...' with a red square. Below it, there are buttons for 'Request ID: 5', 'Description: description1', and 'Request ID: 6', 'Description: description2'. The 'Deep Inspection' stage shows a progress bar labeled 'running' with a red square. Below it, there are buttons for 'XML', 'Request ID: 4', 'Description: To export data by ...', and '30,020 documents'.

**Content Analytics Collections**

Text analytics collection    Crawl    Parse and Index    Analytics    Actions

Hadoop collection    Crawl    Parse and Index    Analytics    Actions

**Crawl and Import**    Parse and Index    Search and Content Analytics

3 running 3 stopped    380 documents    1 running

**Export**

Database    waiting    Exported documents: 3,924

**Export**

CSV    waiting    Exported documents: 3,924

**XML**

Request ID: 4    Description: To export data by ...    initializing...    4,420 documents

**Queued export requests**

Request ID: 5    Description: description1

Request ID: 6    Description: description2

**Deep Inspection**

XML    Request ID: 4    Description: To export data by ...    running    30,020 documents

In the first pane of the dashboard, you can add content to a collection by:

- Configuring crawlers to collect content from data and web sources
- Importing files that are in comma-separated value (CSV) format

## Exploring the Collections Dashboard

The screenshot displays the 'Content Analytics Collections' dashboard. At the top, there are three tabs: 'Crawl', 'Parse and Index', and 'Analytics'. Below these, two collections are visible: 'Text analytics collection' and 'Hadoop collection'. The 'Hadoop collection' is expanded to show three main stages: 'Crawl and Import', 'Parse and Index', and 'Search and Content Analytics'. Each stage has a progress bar and a status indicator. 'Crawl and Import' is 'crawling...', 'Parse and Index' is 'waiting', and 'Search and Content Analytics' is 'running'. A yellow callout box points to the 'Crawlers' section, providing instructions on creating a crawler and listing supported content types. Below the callout, there are sections for 'Export' (Database) and 'Deep Inspection' (XML).

### Content Analytics Collections

Text analytics collection    Crawl    Parse and Index    Analytics    Actions

Hadoop collection    Crawl    Parse and Index    Analytics    Actions

**Crawl and Import**    crawling...    1  
3 running 3 stopped  
1 running

**Parse and Index**    waiting    2  
380 documents

**Search and Content Analytics**    running    3

**Crawlers**    +

**Export**    Database    waiting  
Exported documents: 3,924

**Deep Inspection**    XML    running  
Request ID: 4  
Description: To export data by ...  
30,020 documents

**Callout Box:**  
When you create a crawler, you specify the type of content that you want to add to the collection. Before you create these types of crawlers, you must run scripts or deploy web services to enable the crawler server to communicate with the target server:

- Case Manager
- Content Integrator
- Content Manager
- DB2
- Exchange Server
- FileNet P8
- JDBC database
- Notes
- SharePoint



# Exploring the Collections Dashboard

The screenshot displays the 'Content Analytics Collections' interface. At the top, there are three tabs: 'Crawl', 'Parse and Index', and 'Analytics'. Below these, a 'Text analytics collection' and a 'Hadoop collection' are visible, each with a status indicator and an 'Actions' dropdown.

The main area is divided into three numbered sections:

- Crawl and Import:** Shows a progress bar labeled 'crawling' with a green arrow. Below it, it indicates '3 running 3 stopped' and '1 running'. A yellow callout box points to this section with the text: 'After you create a crawler, you can easily monitor how much content the crawler has added to the collection ...'. Below this is a 'Crawlers' list with items like 'Windows file system crawler 1' (63% done), 'Web crawler 1' (stopped), 'Notes crawler 1' (stopped), 'SharePoint crawler 1' (0% done), 'Exchange crawler 1' (100% done), and 'Seed list crawler 1' (stopped).
- Parse and Index:** Shows a progress bar labeled 'waiting' with a red square. Below it are sections for 'Analytic Resources', 'Global Processes', and 'Annotators'. An 'Export' window is open, showing 'CSV' export with 'Exported documents: 3,924'.
- Search and Content Analytics:** Shows a progress bar labeled 'running' with a red square. Below it are sections for 'Details', 'Export', and 'Deep Inspection'. The 'Export' window shows 'XML' export with '4,420 documents' and 'Queued export requests' (Request ID: 5, 6). The 'Deep Inspection' window shows 'XML' export with '30,020 documents'.

# Exploring the Collections Dashboard

The screenshot displays the 'Content Analytics Collections' interface. At the top, there are three tabs: 'Crawl', 'Parse and Index', and 'Analytics'. Below these, a 'Text analytics collection' and a 'Hadoop collection' are visible. The 'Hadoop collection' is expanded to show three main stages:

- Crawl and Import (1):** Shows a progress bar for 'crawling...'. A list of crawlers is shown below: 'Windows file system crawler 1' (63% done), 'Web crawler 1' (stopped), 'Notes crawler 1' (stopped), 'SharePoint crawler 1' (0% done), 'Exchange crawler 1' (100% done), and 'Seed list crawler 1' (stopped). A yellow callout bubble points to the 'Windows file system crawler 1' entry, containing the text: 'View detailed statistics about crawler activity and the specific sources crawled, including quick access to error logs ...'. Below the list are 'Import' and 'Export' buttons.
- Parse and Index (2):** Shows a progress bar for 'waiting' and '380 documents'. It includes a 'Details' panel with sub-sections for 'Analytic Resources', 'Global Processes', and 'Annotators'. Below this is an 'Export' window for 'CSV' showing 'Exported documents: 3,924'.
- Search and Content Analytics (3):** Shows a progress bar for 'running'. It includes a 'Details' panel and an 'Export' window for 'XML' showing '4,420 documents' and 'Queued export requests' with details for Request ID: 5 and Request ID: 6. Below this is a 'Deep Inspection' window for 'XML' showing '30,020 documents'.

# Exploring the Collections Dashboard

The screenshot displays the 'Content Analytics Collections' dashboard. At the top, there are three tabs: 'Crawl', 'Parse and Index', and 'Analytics'. Below these, a 'Text analytics collection' is shown with a progress bar and an 'Actions' dropdown. A 'Hadoop collection' is also visible, with its own progress bar and 'Actions' dropdown. The main area is divided into three columns: 'Crawl and Import' (1), 'Parse and Index' (2), and 'Search and Content Analytics' (3). A yellow callout box points to the 'Crawl and Import' section, containing the text: 'And edit the crawler's configuration, such as adding data sources or changing the crawler schedule.' The 'Crawl and Import' section lists several crawlers: 'Windows file system crawler' (63% done), 'Web crawler 1' (stopped), 'Notes crawler 1' (stopped), 'SharePoint crawler 1' (0% done), 'Exchange crawler 1' (100% done), and 'Seed list crawler 1' (stopped). The 'Parse and Index' section shows '380 documents' and an 'Export' window for 'CSV' (waiting) with 'Exported documents: 3,924'. The 'Search and Content Analytics' section shows '4,420 documents' and an 'Export' window for 'XML' (initializing...) with '4,420 documents' and 'Queued export requests' (Request ID: 5, 6). A 'Deep Inspection' window for 'XML' (running) shows '30,020 documents'.

# Exploring the Collections Dashboard

**Content Analytics Collections**

Text analytics collection    Crawl    Parse and Index    Analytics    Actions

Hadoop collection    Crawl    Parse and Index    Analytics    Actions

**1 Crawl and Import**  
3 running 3 stopped  
1 running

**2 Parse and Index**  
380 documents

**3 Search and Content Analytics**  
4,420 documents

**Crawlers**

- Windows file system crawler 1 (63% done)
- Web crawler 1 (stopped)
- Notes crawler 1 (stopped)
- SharePoint crawler 1 (0% done)
- Exchange crawler 1 (100% done)
- Seed list crawler 1 (stopped)

**Export**  
CSV  
Exported documents: 3,924

**Export**  
XML  
Request ID: 4  
Description: To export data by ...  
4,420 documents  
Queued export requests:  
Request ID: 5  
Description: description1  
Request ID: 6  
Description: description2

**Deep Inspection**  
XML  
Request ID: 4  
Description: To export data by ...  
30,020 documents

You can easily stop and start individual crawlers.

# Exploring the Collections Dashboard

The screenshot displays the 'Content Analytics Collections' dashboard. At the top, there are three tabs: 'Crawl', 'Parse and Index', and 'Analytics'. Below these, the 'Hadoop collection' is highlighted. The dashboard is divided into several sections:

- Crawl and Import:** A progress bar shows '3 running 3 stopped'. A tooltip points to a 'crawling...' indicator with the text: 'You can also select multiple crawlers to stop or start at the same time.' Below this, a list of crawlers is shown: 'windows file system crawler 1' (63% done), 'Web crawler 1' (stopped), 'Notes crawler 1' (stopped), 'SharePoint crawler 1' (0% done), 'Exchange crawler 1' (100% done), and 'Seed list crawler 1' (stopped).
- Parse and Index:** A progress bar shows 'stopped'.
- Search and Content Analytics:** A progress bar shows 'running'.

On the right side, there are several panels:

- Details:** A panel with a dropdown arrow, containing sub-sections for 'Analytic Resources', 'Global Processes', and 'Annotators'.
- Export:** A panel showing 'XML' with 'Request ID: 4' and 'Description: To export data by ...'. A progress bar indicates 'initializing...' with '4,420' documents. Below, 'Queued export requests' are listed: 'Request ID: 5 Description: description1' and 'Request ID: 6 Description: description2'.
- Deep Inspection:** A panel showing 'XML' with 'Request ID: 4' and 'Description: To export data by ...'. A progress bar indicates 'running' with '30,020 documents'.

At the bottom, there are 'Import' and 'Export' buttons.

# Exploring the Collections Dashboard

The screenshot displays the 'Content Analytics Collections' dashboard. At the top, there are three main stages: 'Crawl', 'Parse and Index', and 'Analytics'. Below these, a 'Text analytics collection' and a 'Hadoop collection' are visible. The 'Crawl and Import' section shows a progress bar and the status 'crawling...'. The 'Parse and Index' section shows a progress bar and the status 'waiting', with '380 documents' listed below. The 'Search and Content Analytics' section shows a progress bar and the status 'running'. A 'Crawlers' panel on the left lists several crawlers: 'Windows file system crawler 1' (63% done), 'Web crawler 1' (stopped), 'Notes crawler 1' (stopped), 'SharePoint crawler 1' (0% done), 'Exchange crawler 1' (100% done), and 'Seed list crawler 1' (stopped). A 'Select Crawlers' dialog box is open in the center, showing a list of crawlers with checkboxes: 'Web crawler 1', 'Notes crawler 1', and 'Seed list crawler 1'. The 'Start Selected Crawlers' button is highlighted. Other panels include 'Export' and 'Deep Inspection', both showing progress bars and document counts.

# Exploring the Collections Dashboard

## Content Analytics Collections

Crawl Parse and Index Analytics

Text analytics collection ● ■ ● Actions ▾

Hadoop collection ● ● ● Actions ▾

### Crawl and Import

■ ▶ crawling... 1

3 running 3 stopped  
1 running

Crawlers

Import +

Export

Database ■ waiting

Exported documents: 3,924

### Parse and Index

■ waiting 2

380 documents

Details

Analytic Resources

Export

CSV ■ waiting

Exported documents: 3,924

### Search and Content Analytics

■ running 3

Details

Export

Request ID: 4  
Description: To export data by ...

■ initializing...

4,420 documents

Queued export requests

Request ID: 5  
Description: description1

Request ID: 6  
Description: description2

Deep Inspection

XML

Request ID: 4  
Description: To export data by ...

■ running

30,020 documents

When you import files in CSV format, you tell the system how to read the content, such as what the column separator character is and whether the first line is a header line.

A wizard lets you browse to find the files that you want to add and preview the file content.

# Exploring the Collections Dashboard

The screenshot displays the 'Content Analytics Collections' dashboard. At the top, there are three tabs: 'Crawl', 'Parse and Index', and 'Analytics'. Below these, the dashboard is organized into several sections:

- Text analytics collection:** Shows a progress bar and an 'Actions' dropdown.
- Hadoop collection:** Shows a progress bar and an 'Actions' dropdown.
- Crawl and Import:** A progress bar labeled '1' with a 'crawling...' status. Below it, it says '3 running 3 stopped' and '1 running'.
- Parse and Index:** A progress bar labeled '2' with a 'waiting' status. Below it, it says '380 documents'.
- Search and Content Analytics:** A progress bar labeled '3' with a 'running' status.
- Crawlers:** A list of crawlers, including 'Import' which is 'running' and has '307,843 documents'. Below this is a 'Queued CSV imports' section with two entries.
- Export:** A section for 'Export' tasks, including 'CSV' (waiting, 3,924 documents) and 'Database' (waiting, 3,924 documents).
- Export (XML):** A section for 'Export' tasks, including 'XML' (initializing, 4,420 documents) with a list of 'Queued export requests'.
- Deep Inspection:** A section for 'Deep Inspection' tasks, including 'XML' (running, 30,020 documents).

A yellow callout box points to the 'Import' task in the 'Crawlers' section, containing the text: 'When the import task begins, you can easily monitor the progress or cancel the task.' A tooltip for the 'Import' task provides the following details: 'Import Space ID: C:\Users\t\_tejima\Desktop\csv\CIC\_SAFE\199707-199804.csv, Requested time: 3/6/11 11:53 PM'.



# Exploring the Collections Dashboard

**Content Analytics Collections**

Text analytics collection    Crawl    Parse and Index    Analytics    Actions

Hadoop collection    Crawl    Parse and Index    Analytics    Actions

**Crawl and Import** 1  
3 running 3 stopped  
1 running

**Parse and Index** 2  
380 documents

**Search and Content Analytics** 3

**Crawlers**

- Import**  
C:\Users\t\_tejima\Desktop\csv...  
307,843 documents
- Queued CSV imports  
C:\Users\t\_tejima\Desktop\csv...  
C:\Users\t\_tejima\Desktop\csv...

**Export**  
Database  
Exported documents: 3,924

**Export**  
CSV

**XML**  
Request ID: 4  
Description: To export data by ...  
4,420 documents

Queued export requests

- Request ID: 5  
Description: description1
- Request ID: 6  
Description: description2

**Deep Inspection**  
XML  
Request ID: 4  
Description: To export data by ...  
30,020 documents

You can export documents after they have been crawled and before they have been parsed or analyzed. For example, you might want to export the documents as XML files or to a relational database. A wizard helps you specify which documents you want to export and how they are to be exported.

After you configure the export wizard, you can easily change export settings and monitor progress of the task.

# Exploring the Collections Dashboard

**Content Analytics Collections**

Text analytics collection    Crawl    Parse and Index    Analytics    Actions ▾

Hadoop collection    Crawl    Parse and Index    Analytics    Actions ▾

**1** Crawl and Import  
3 running 3 stopped  
1 running  
Crawlers  
Import

**2** Parse and Index  
380 documents  
waiting

**3** Search and Content Analytics  
running

Export  
Database  
waiting  
Exported documents: 3,924

Export  
CSV  
waiting  
Exported documents: 3,924

Export  
XML  
Request ID: 4  
Description: To export data by ...  
initializing...  
4,420 documents  
Queued export requests  
Request ID: 5  
Description: description1  
Request ID: 6  
Description: description2

Deep Inspection  
XML  
Request ID: 4  
Description: To export data by ...  
running  
30,020 documents

**Callout Box:**  
In the second pane of the dashboard, you specify how you want content that was added to the collection to be parsed and analyzed before it is added to the index.  
You can easily stop and restart the parse and index services and monitor the status of index builds.

# Exploring the Collections Dashboard

The screenshot displays the 'Content Analytics Collections' dashboard. At the top, there are three tabs: 'Crawl', 'Parse and Index', and 'Analytics'. Below these, two collections are listed: 'Text analytics collection' and 'Hadoop collection'. The 'Hadoop collection' is expanded to show three main sections: 'Crawl and Import', 'Parse and Index', and 'Search and Content Analytics'.  
1. 'Crawl and Import' shows a progress bar labeled 'crawling...' with 3 running, 3 stopped, and 1 running crawler. Below it are 'Crawlers' and 'Import' sections.  
2. 'Parse and Index' shows a progress bar labeled 'waiting' with 380 documents. Below it is the 'Details' section, which is highlighted by a yellow callout box. The 'Details' section includes: 'Rebuild index' (not started), 'Rebuild optional facet index' (44% done), 'Index size (megabytes): 5 /', 'Number of documents: 380', 'Number of dropped documents: 12', 'Incoming document detection: Off', 'Index version: -', and 'Document processor nodes' (hostname2.ibm.com stopped, hostname.ibm.com running). Below 'Details' are 'Analytic Resources', 'Global Processes', and 'Annotators' sections.  
3. 'Search and Content Analytics' shows a progress bar labeled 'running'. Below it are 'Request ID: 4' (initializing... 4,420 documents) and 'Queued export requests' (Request ID: 5, description1; Request ID: 6, description2). At the bottom right is the 'Deep Inspection' section, showing 'XML' (Request ID: 4, running) with 30,020 documents and 'Queued inspection requests' (Request ID: 5).  
A yellow callout box points to the 'Details' section with the text: 'For example, expand the **Details** area to: - Reorganize or rebuild the index - Rebuild the facet index - See how many documents have been indexed - See how many documents were dropped and - Learn why documents could not be indexed, and - Stop and restart document processors.'

# Exploring the Collections Dashboard

**Content Analytics Collections**

Crawl Parse and Index Analytics

Text analytics collection [green dot] [grey square] [green dot] Actions ▾

Hadoop collection [green dot] [green dot] [green dot] Actions ▾

**Crawl and Import** 1  
[Progress bar] [red square] [green triangle] *crawling...*  
3 running 3 stopped  
1 running

**Parse and Index** 2  
[Progress bar] [red square] *waiting*  
380 documents

**Search and Content Analytics** 3  
[Progress bar] [red square] *running*

**Export**  
Database [Progress bar] *waiting*  
Exported documents: 3,924

**Export**  
CSV [Progress bar] *waiting*  
Exported documents: 3,924

**Export**  
XML  
Request ID: 4  
Description: To export data by ...  
[Progress bar] [red square] *initializing...*  
4,420 documents  
Queued export requests  
Request ID: 5  
Description: description1  
Request ID: 6  
Description: description2

**Deep Inspection**  
XML  
Request ID: 4  
Description: To export data by ...  
[Progress bar] [red square] *running*  
30,020 documents

**Details** [Hand cursor icon]

**Analytic Resources** [Pencil icon]

Global Processes [Pencil icon]

Annotators

Monitor analytic resources and configure facets, dictionaries, and custom rules for this collection.

For a content analytics collection, you can configure facets, dictionaries, and custom rule files, deploy these analytic resources, and then monitor the progress of the deployment.

# Exploring the Collections Dashboard

The screenshot displays the 'Content Analytics Collections' interface. At the top, there are three tabs: 'Crawl', 'Parse and Index', and 'Analytics'. Below these, two collection types are visible: 'Text analytics collection' and 'Hadoop collection'. The 'Hadoop collection' is the active view, showing three main sections: 'Crawl and Import' (1), 'Parse and Index' (2), and 'Search and Content Analytics' (3). Each section has a progress bar and status indicator. A yellow callout box on the left provides instructions on configuring global processes. The 'Global Processes' window is open, showing settings for 'Link analysis', 'Duplicate document detection', 'Rule-based categories', 'Thumbnail generation', 'Cluster deployment', and 'Custom'. The 'Export' and 'Deep Inspection' windows are also visible on the right side of the dashboard.

### Content Analytics Collections

Crawl Parse and Index Analytics

Text analytics collection [status icons] Actions

Hadoop collection [status icons] Actions

**1** Crawl and Import [progress bar] crawling...  
3 running 3 stopped  
1 running

**2** Parse and Index [progress bar] waiting  
380 documents

**3** Search and Content Analytics [progress bar] running

Crawlers

For a both types of collections, you can configure and monitor global processes, such as configuring categories and scheduling when the global processes run.

If you created the collection on an IBM InfoSphere BigInsights server, you can also configure, deploy, and monitor custom global processes.

Details

Analytic Resources

Global Processes

- Link analysis [progress bar] 100% waiting
- Duplicate document detection [progress bar] 100% waiting
- Rule-based categories [progress bar] 0% waiting
- Thumbnail generation [progress bar] stopped
- Cluster deployment [progress bar] stopped
- Custom [progress bar] 100% waiting

Annotators

Export

XML  
Request ID: 4  
Description: To export data by ...  
[progress bar] initializing...  
4,420 documents

Queued export requests

- Request ID: 5  
Description: description1
- Request ID: 6  
Description: description2

Deep Inspection

XML  
Request ID: 4  
Description: To export data by ...  
[progress bar] running  
30,020 documents

# Exploring the Collections Dashboard

## Content Analytics Collections

Crawl Parse and Index Analytics

Text analytics collection ● ■ ● Actions ▾

Hadoop collection ● ● ● Actions ▾

### Crawl and Import

■ ▶ crawling... 1

3 running 3 stopped  
1 running

Crawlers  
Import

### Parse and Index

■ waiting 2

380 documents

Details  
Analytic Resources  
Global Processes  
Annotators

### Search and Content Analytics

■ running 3

Details

### Database

■ waiting

Exported documents: 3,924

### Export

CSV ■ waiting

Exported documents: 3,924

### Export

XML ■ initializing...

4,420 documents

Queued export requests

- Request ID: 5  
Description: description1
- Request ID: 6  
Description: description2

### Deep Inspection

XML

Request ID: 4  
Description: To export data by ...

■ running

30,020 documents

In pane 2, you also enable and disable annotators for analyzing content in the collection. Not all annotators are available for both collection types.

Content Classification Annotator (disabled)

# Exploring the Collections Dashboard

**Content Analytics Collections**

Text analytics collection    Crawl    Parse and Index    Analytics    Actions

Hadoop collection    Crawl    Parse and Index    Analytics    Actions

**1** Crawl and Import  
3 running 3 stopped  
1 running  
Crawlers  
Import

**2** Parse and Index  
380 documents  
waiting

**3** Search and Content Analytics  
running

Details

Export  
Database  
waiting  
Exported documents: 3,924

Export  
CSV  
waiting  
Exported documents: 3,924

Export  
XML  
Request ID: 5  
Description: description1  
Request ID: 6  
Description: description2

Deep Inspection  
XML  
Request ID: 4  
Description: To export data by ...  
30,020 documents  
running

You can export documents or document metadata after content has been parsed or analyzed. For example, you might want to export analysis results to a relational database. A wizard helps you specify what you want to export and how it is to be exported.

After you configure the export wizard, you can easily change export settings and monitor the progress of export tasks.

# Exploring the Collections Dashboard

**Content Analytics Collections**

Crawl Parse and Index Analytics

Text analytics collection [●] [■] [●] Actions ▾

Hadoop collection [●] [●] [●] Actions ▾

**Crawl and Import** 1  
[Progress bar] [■] [▶] *crawling...*  
3 running 3 stopped  
1 running

Crawlers  
Import

**Parse and Index** 2  
[Progress bar] [■] *waiting*  
380 documents

Details  
Analytic Resources  
Global Processes  
Annotators

**Search and Content Analytics** 3  
[Progress bar] [■] *running*

**Export**  
Database [Progress bar] *waiting*  
Exported documents: 3,924

**Export**  
CSV [Progress bar] *waiting*  
Exported documents: 3,924

**Export**  
XML  
Request ID: 4  
Description: To export data by ...  
[Progress bar] [■] *initializing...*  
4,420 documents  
Queued export requests  
Request ID: 5  
Description: description1  
Request ID: 6  
Description: description2

**Deep Inspection**  
XML  
Request ID: 4  
Description: To export data by ...  
[Progress bar] [■] *running*  
30,020 documents

To specify your preferences for how content is to be parsed, analyzed, and indexed, click **Configure**.



# Exploring the Collections Dashboard

The screenshot displays the 'Content Analytics Collections' dashboard. At the top, there are three tabs: 'Crawl', 'Parse and Index', and 'Analytics'. Below these, two collection types are visible: 'Text analytics collection' and 'Hadoop collection'. The 'Hadoop collection' is expanded to show three stages: 'Crawl and Import' (with a progress bar and 'crawling...' status), 'Parse and Index' (with a progress bar and 'waiting' status), and 'Search and Content Analytics' (with a progress bar and 'running' status). A purple callout box explains that different actions are available for each collection type, with analytic resources and sentiment analysis being specific to content analytics collections. A 'Configure...' menu is open, listing options like 'Contextual views', 'Analytic resources', 'Sentiment analysis', and 'Terms of interest'. Other panels include 'Export' (with a 'Database' section and 'Exported documents: 3,924'), 'Queued export requests' (listing request IDs and descriptions), and 'Deep Inspection' (showing 'XML' and '30,020 documents').

# Exploring the Collections Dashboard

The screenshot displays the 'Content Analytics Collections' dashboard. At the top, there are three main stages: 'Crawl', 'Parse and Index', and 'Analytics'. Below these, the 'Hadoop collection' is shown with three sub-panels:

- Crawl and Import:** A progress bar is labeled 'crawling...' with a circled '1'. It shows '3 running 3 stopped' and '1 running'. Below are buttons for 'Crawlers' and 'Import'.
- Parse and Index:** A progress bar is labeled 'waiting' with a circled '2'. It shows '380 d'.
- Search and Content Analytics:** A progress bar is labeled 'running' with a circled '3'. It shows '4,420 documents' and 'Queued export requests'.

A purple callout box is positioned over the 'Parse and Index' section, containing the following text:

In the third pane of the dashboard, you specify how users can query the parsed and analyzed data in the index.  
You can specify preferences for improving search quality, such as associating custom dictionaries, and you can monitor query processing.

At the bottom, there are two 'Export' panels:

- Export (Database):** Shows 'waiting' and 'Exported documents: 3,924'.
- Export (CSV):** Shows 'waiting' and 'Exported documents: 3,924'.

On the right side, there are two more panels:

- Export (XML):** Shows 'initializing...' and '4,420 documents'. It lists 'Queued export requests' with 'Request ID: 5' (description1) and 'Request ID: 6' (description2).
- Deep Inspection (XML):** Shows 'running' and '30,020 documents'. It lists 'Queued inspection requests'.

# Exploring the Collections Dashboard

### Content Analytics Collections

	Crawl	Parse and Index	Analytics	Actions
Text analytics collection	●	■	●	Actions ▾
Hadoop collection	●	●	●	Actions ▾

  

#### Crawl and Import

1

3 running 3 stopped  
1 running

Export

Database *waiting*  
Exported documents: 3,924

#### Parse and Index

2

390 documents

Details

- Analytic Resources
- Global Processes
- Annotatrcs

Export

CSV *waiting*  
Exported documents: 3,924

#### Search and Content Analytics

3

hostname.ibm.com *running*

hostname2.ibm.com *stopped*

Export

XML  
Request ID: 4  
Description: To export data by N...  
4,420 documents

Queued export requests

- Request ID: 5  
Description: description1
- Request ID: 6  
Description: description2

Deep Inspection

You can easily start and stop one server or all search servers to apply configuration changes.

# Exploring the Collections Dashboard

## Content Analytics Collections

Crawl Parse and Index Analytics

Text analytics collection ● ■ ● Actions ▾

Hadoop collection ● ● ● Actions ▾

### Crawl and Import

▬ ■ ▶ *crawling...* **1**

3 running 3 stopped  
1 running

Crawlers  
Import

### Parse and Index

▬ ■ *waiting* **2**

380 documents

Details  
Analytic Resources  
Global Processes  
Annotators

### Search and Content Analytics

▬ ■ *running* **3**

You can view detailed **query statistics**, such as a history of the most recent or most popular queries, and you can save the statistics to a CSV file.

### Export

Database ▬ *waiting*

Exported documents: 3,924

### Export

CSV ▬ *waiting*

Exported documents: 3,924

### Export

XML  
Request ID: 4  
Description: To export data by ...  
▬ ■ *initializing...*

4,420 documents

Queued export requests  
Request ID: 5  
Description: description1  
Request ID: 6  
Description: description2

### Deep Inspection

XML  
Request ID: 4  
Description: To export data by ...  
▬ ■ *running*

30,020 documents

Queued inspection requests

# Exploring the Collections Dashboard

**Content Analytics Collections**

Text analytics collection    Crawl    Parse and Index    Analytics    Actions

Hadoop collection    Crawl    Parse and Index    Analytics    Actions

**Crawl and Import** 1  
3 running 3 stopped  
1 running

**Parse and Index** 2  
380 documents

**Search and Content Analytics** 3  
running

**Export**  
Database  
waiting  
Exported documents: 3,924

**Export**  
CSV  
waiting  
Exported documents: 3,924

**Export**  
XML  
Request ID: 4  
Description: To export data by ...  
initializing...  
4,420 documents  
Queued export requests  
Request ID: 5  
Description: description1  
Request ID: 6  
Description: description2

**Deep Inspection**  
XML  
Request ID: 4  
Description: To export data by ...  
running  
30,020 documents  
Queued inspection requests

Configure **deep inspections** for content analytics collections to explore large numbers of facet values that might be difficult to navigate interactively.

You can schedule when the deep inspections run, check the status of the process, and edit inspection settings.

Monitor deep inspection requests through the content analytics miner.

# Exploring the Collections Dashboard

### Enterprise Search Collections

	Crawl	Parse and Index	Search	
English and Japanese collection	●	■	●	Actions ▾
Kiso collection	●	■	●	Actions ▾

### Content Analytics Collections

	Crawl	Parse and Index	Analytics	
Text analytics collection	●	■	●	Actions ▾
Hadoop collection	●	●	●	Actions ▾

#### Crawl and Import

Progress bar: crawling... 1  
3 running 3 stopped  
1 running

Crawlers  
Import

#### Parse and Index

Progress bar: waiting 2  
380 documents

Details  
Analytic Resource  
Global Processes  
Annotators

#### Search and Content Analytics

Progress bar: running 3

Export  
Deep Inspection

#### Export

Database: waiting  
Exported documents: 3,924

#### Export

CSV: waiting  
Exported documents: 3,924

#### IBM Cognos BI Reports

Database: jdbc:db2://iies76.yamato.ibm.com:...  
IBM Cognos BI server URL: http://iies76.yamato.ibm.com:9300...

If you integrated with IBM Cognos Business Intelligence, you can enable analysts to **create reports** in the content analytics miner and view the reports with Cognos BI analytical tools.

You can schedule when the reports run and edit report generation settings.

### Enterprise Search Collections

	Crawl	Parse and Index	Search	
English and Japanese collection	<span style="color: green;">●</span>	<span style="color: gray;">■</span>	<span style="color: green;">●</span>	Actions ▾
Kiso collection	<span style="color: green;">●</span>	<span style="color: gray;">■</span>	<span style="color: green;">●</span>	Actions ▾

  

### Content Analytics Collections

	Crawl	Parse and Index	Analytics	
Text analytics collection	<span style="color: green;">●</span>	<span style="color: gray;">■</span>	<span style="color: green;">●</span>	Actions ▾
Hadoop collection	<span style="color: green;">●</span>	<span style="color: green;">●</span>	<span style="color: green;">●</span>	Actions ▾

**Crawl and Import** 1

▶ *crawling...*

3 running 3 stopped  
1 running

**Crawlers**

Import

**Parse and Index** 2

■ *waiting*

380 documents

**Search and Content Analytics** 3

■ *running*

**Export** 4

Database  ■ *waiting*

Exported documents: 3,924

**Export** 5

CSV  ■ *waiting*

Exported documents: 3,924

**Export** 6

XML  
Request ID: 4  
Description: To export data by ...

■ *initializing...*

4,420 documents

Queued export requests

Request ID: 5  
Description: description1

Request ID: 6  
Description: description2

**Details**

Analytic Resources

Global Processes

Annotators

**Export**

You can schedule when the export requests run, check the status of the process, and edit export settings.

**Export**

For both enterprise search and content analytics collections, you can enable users to **export documents** that match their query results.

# Exploring the Collections Dashboard

The screenshot displays the Enterprise Search Collections dashboard, divided into two main sections: Enterprise Search Collections and Content Analytics Collections.

**Enterprise Search Collections:**

Collection Name	Crawl	Parse and Index	Search	Actions
English and Japanese collection	●	■	●	Actions ▼
Kiso collection	●	■	●	Actions ▼

**Content Analytics Collections:**

Collection Name	Crawl	Parse and Index	Analytics	Actions
Text analytics collection	●	■	●	Actions ▼
Hadoop collection	●	●	●	Actions ▼

The Hadoop collection is expanded to show detailed process information:

- Crawl and Import (1):** Progress bar shows 'crawling...'. Status: 3 running, 3 stopped, 1 running. Includes sub-sections for Crawlers and Import.
- Parse and Index (2):** Progress bar shows 'waiting'. Status: 380 documents. Includes sub-sections for Details, Analytic Resources, Global Processes, and Annotators.
- Search and Content Analytics (3):** Progress bar shows 'running'. Includes an 'Export' window with details for XML, CSV, and Queued export requests.

A yellow callout box points to the 'Configure' menu icon in the Search and Content Analytics section, stating: "Open the **Configure** menu to specify options that help users search and explore collections."



# Exploring the Collections Dashboard

## Enterprise Search Collections

	Crawl	Parse and Index	Search	
English and Japanese collection	<span style="color: green;">●</span>	<span style="color: gray;">■</span>	<span style="color: green;">●</span>	Actions ▾
Kiso collection	<span style="color: green;">●</span>	<span style="color: gray;">■</span>	<span style="color: green;">●</span>	Actions ▾

  

## Content Analytics Collections

	Crawl	Parse and Index	Analytics	
Text analytics collection	<span style="color: green;">●</span>	<span style="color: gray;">■</span>	<span style="color: green;">●</span>	Actions ▾
Hadoop collection	<span style="color: green;">●</span>	<span style="color: green;">●</span>	<span style="color: green;">●</span>	Actions ▾

  

**Crawl and Import**

■ ▶ *crawling...* 1

3 running 3 stopped  
1 running

Different search actions are available for each collection type. For example, in an enterprise search collection, you can:

- Configure rules to dynamically expand queries and rank results when users search the collection.
- Enable users to export documents from an enterprise search application to a content analytics collection.

**Search and Content Analytics**

■ *running* 3

**Crawlers**

Import

**Configure...**

- Search server options
- Quick links
- Wildcard characters
- Scopes
- Type ahead options
- URI patterns for influencing scores
- Collapsed search results
- Document importance
- Document flags
- Memory for search servers
- Request ID: 5  
Description: description1
- Request ID: 6  
Description: description2

**Export**

Database waiting

Exported documents: 3,924

**Export**

CSV waiting

Exported documents: 3,924

# Exploring the Collections Dashboard

**Enterprise Search Collections**

	Crawl	Parse and Index	Search	
English and Japanese collection	●	■	●	Actions ▾
Kiso collection	●	■	●	Actions ▾

**Content Analytics Collections**

Text analytics collection

	Crawl	Parse and Index	Analytics	
Text analytics collection	●	■	●	Actions ▾

**Hadoop collection**

**Crawl and Import**  
3 running 3 stopped  
1 running

**Parse and Index**  
380 documents

**Export**  
Database  
Exported documents: 3,924

**Export**  
CSV  
Exported documents: 3,924

**Export**  
XML  
Request ID: 4  
Description: To export data by ...  
4,420 documents  
Queued export requests  
Request ID: 5  
Description: description1  
Request ID: 6  
Description: description2

Finally, for both types of collections, open the **Actions** menu to:

- Change collection settings, such as enabling or disabling the ability to create rule-based categories.
- View collection settings that cannot be changed.
- Specify options for logging messages and receiving alerts.
- Create a collection by cloning this one.
- Delete this collection.

## Exploring the Collections Dashboard

IBM Content Analytics with Enterprise Search Search Customizer | Analytics Customizer | Log Out | Help | About

Collections System Security

\* Create Collection

### Enterprise Search Collections

	Crawl	Parse and Index	Search	
English and Japanese collection	●	■	●	Actions ▼
Kiso collection	●	■	●	Actions ▼

### Content Analytics Collections

	Parse and Index	Analytics	
Text analytics collection	■	●	Actions ▼
Hadoop collection	●	■	Actions ▼

This concludes the tour of the Collections dashboard in the IBM Content Analytics with Enterprise Search administration console.