

Integrating Search Interface to SAS[®] Content 3.6



Copyright Notice

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *Integrating Search Interface to SAS® Content 3.6*, Cary, NC: SAS Institute Inc., 2016.

Integrating Search Interface to SAS® Content 3.6

Copyright © 2016, SAS Institute Inc., Cary, NC, USA.

All rights reserved. Printed in the United States of America. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, by any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc. Limited permission is granted to store the copyrighted material in your system and display it on terminals, print only the number of copies required for use by those persons responsible for installing and supporting the SAS programming and licensed programs for which this material has been provided, and to modify the material to meet specific installation requirements. The SAS Institute copyright notice must appear on all printed versions of this material or extracts thereof and on the display medium when the material is displayed. Permission is not granted to reproduce or distribute the material except as stated above.

U.S. Government Restricted Rights Notice: Use, duplication, or disclosure of the software by the government is subject to restrictions as set forth in FAR 52.227-19 Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

May 2016

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.

® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

Table of Contents

| | |
|---|-----------|
| Chapter 1 — Introduction | 1 |
| Architecture | 1 |
| Data Feeds and Indexing..... | 2 |
| Installation and Prerequisites | 2 |
| Post-Installation Configuration to Enable SSL (Optional) | 3 |
| Integration with SAS Information Retrieval Studio | 3 |
| Integration with Apache Lucene | 4 |
| Apache Lucene Index..... | 4 |
| Verify the Installation | 5 |
| Verify the Integration with SAS Information Retrieval Studio | 5 |
| Verify the SAS Information Retrieval Studio Server Installation | 5 |
| Chapter 2 — Configuration Options | 7 |
| Step 1 - Verify the Registration Information in Metadata | 7 |
| Step 2 - Verify the Configuration of SAS Information Retrieval Studio | 9 |
| Rebuild the Search Index | 9 |
| Disable the CSRF Filter | 10 |
| Add Sites to the CSRF Whitelist | 10 |
| Whitelisting Methods | 11 |
| Chapter 3 — Logging | 12 |
| Log Settings and Location | 12 |
| Appendix — Troubleshooting (FAQ) | 13 |
| 1. I see no results when I submit a query string in Visual Analytics Hub. What can I do? | 13 |
| 2. How can I enable the Administration Console for SAS Information Retrieval Studio?.. | 13 |
| 3. How can I verify that SAS Information Retrieval Studio has been configured correctly?..... | 14 |
| 4. How can I verify the host and port of SAS Information Retrieval Studio?..... | 15 |
| 5. Some of the required fields are missing from the IR Studio schema, so its contents are not searchable. How can I add the missing fields? | 16 |
| 6. I have a 'Connection refused' message in my log file. What should I do?..... | 16 |
| 7. I am receiving email messages that report errors. What do they mean? | 16 |
| 8. What configuration changes are required for an external reverse proxy environment? | 17 |
| 9. How can I recover lost data for the Indexing Server after an indexing failure? | 18 |
| 10. Search does not work with SSL. What steps can I take in this situation? | 18 |
| 11. How can I change the frequency of indexing? | 18 |
| 12. What is the frequency of email notifications? | 19 |
| 13. Is IPv6 supported? | 19 |
| 14. How can I enable/disable the daily full feed? | 19 |
| 15. How can I change the default setting for the maximum number of records returned?..... | 19 |
| Technical Support | 20 |

Chapter 1 — Introduction

Search Interface to SAS Content is a web application that enables users to search the contents of SAS deployments. It is a user-interface component of the search facility that is installed on the middle tier of SAS Solutions. The search facility indexes and searches SAS content that is registered in metadata and in the Content Server. It returns only the results that are available to the requesting user, based on user account permissions.

Search Interface to SAS Content relies on a second component, an indexing provider that creates a content index. Two index providers are supported:

- SAS Information Retrieval Studio (enabled by default in IPv4 environments)
- Apache Lucene

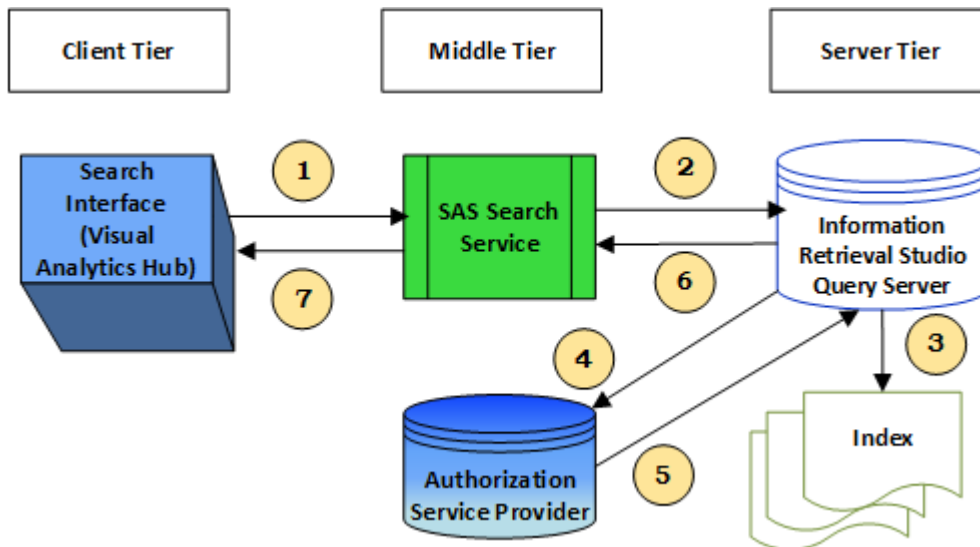
One of these providers is enabled during product installation.

Once the index has been built, Search Interface to SAS Content queries it in response to search criteria supplied by SAS Solution users.

A Search Interface to SAS Content plug-in is available in the Configuration Management plug-ins of the SAS Management Console to enable the administrator to perform configuration tasks.

Architecture

Search Interface to SAS Content is a component of the Platform Content Services within the SAS Web Infrastructure Platform. The Content Services provide common management of and access to different data types, often stored in multiple repositories. All of the services in the SAS Platform Content Services category contain methods that can interact with any content object type. The Search Service and the Index Data Service work together to provide the search functionality. The following diagram provides a high-level description of the architecture, focused on security requirements:



Note that the three “tiers” of the architecture can be running on a single machine, or on separate machines. The Metadata Server is part of the Server Tier, while the Content Server is part of the Middle Tier.

1. An end-user who has logged into a SAS user interface (such as the SAS Visual Analytics Hub, client software running on the user's PC) initiates a search. The client sends a query request to the SAS Search Service web application, which is configured on a separate server for the Middle Tier components. A connection is made to the SAS Search Service running in the application server using port 7980 by default, or an external port.
2. The SAS Search Service makes a connection to the Information Retrieval Studio Query Server on port 10731, and sends the search request.
3. Information Retrieval Studio searches its index for the query string. The index is stored in the following directory on the machine where your SAS solution server is running:
`<SASConfig>/LevN/Applications/SASInformationRetrievalStudioforSAS/work/index.`
4. Information Retrieval Studio sends an authorization request to the Search Service Authorization Provider running in the Middle Tier. The `query-server-db` file is used to find the authorization provider. The value of the `searchsas.auth.provider.url` parameter determines the REST endpoint to call for authorizing search results. For more information, see [“Step 1 - Verify the Registration Information in Metadata”](#) on page 7.
5. The Authorization Service Provider returns a filtered list of authorized objects.
6. The Information Retrieval Studio Query Server returns search results to the SAS Search Service.
7. The SAS Search Service returns the results to the user interface (in this example, the Hub).

The diagram shows a configuration with Information Retrieval Studio as the index server, but this service could also be provided by Apache Lucene.

Data Feeds and Indexing

Search Interface to SAS Content collects data from a variety of back-end systems, including the Metadata Server and the Object Type Service. It passes this information along to the Index Data Service. The data that is returned by the Index Data Service is accumulated for all types and is periodically pushed (“fed”) to the index server.

When the Index Data Service starts up for the first time, it takes all records from the Metadata Server and Content Server and adds them to the index. Then at a set interval (the default is every 15 minutes), it collects only new objects or objects with modifications since the previous feed and adds them to the index, registering them in metadata.

The frequency of indexing and the option to automatically push all of the available records to the index once each day are controlled by settings that you can change. The full data feed is disabled by default. For more information, see [“How can I change the frequency of indexing?”](#) on page 18 and [“How can I enable/disable the daily full feed?”](#) on page 19.

Installation and Prerequisites

Search Interface to SAS Content is installed by default with SAS Visual Analytics 7.2 and higher, as well as with other solutions that provide the SAS Visual Analytics Administration and Reporting component. You can configure it using the SAS Deployment Wizard. However, we recommend that you accept the default values during the initial configuration and make customizations after the initial deployment if required.

During the installation, you are prompted to configure a **SAS internal account for the Search Interface to SAS Content user**. The default Search Interface to SAS Content User is `sassearch@saspw`, which is required to enable access to content in SAS Information Retrieval Studio. This default user is one of the internal user accounts that are required by SAS Foundation and are authenticated on the SAS Metadata Server. For more information about these user accounts, see the *SAS® Visual Analytics Installation and Configuration Guide*.

Search Interface to SAS Content provides search results exclusively for the Search Interface to SAS Content User. The results are then authorized as a separate step for the individual user accounts that are logged in and performing searches.

SAS Information Retrieval Studio 1.53 requires Python version 2.6 through 2.7.x (not 3.0 or higher) for deployment. This third-party software is included in the Pre-Installation Checklist that accompanies your SAS Deployment Plan. It must be installed on the target machine before you can install SAS Information Retrieval Studio on the SAS Visual Analytics middle tier. You are prompted to specify the Python installation location during SAS Visual Analytics installation.

For more information about Python requirements and a link where you can download the software, go to <http://support.sas.com/resources/thirdpartysupport/v94/other.sw.html>.

Post-Installation Configuration to Enable SSL (Optional)

After you have used the SAS Deployment Wizard to install and configure SAS Visual Analytics, an extra set of steps may be required. The use of SSL for secure communications among SAS Visual Analytics components is supported, but is optional. To implement certificates, you must add the path to SAS Foundation to configure SAS Information Retrieval Studio.

On Linux, in addition to the path to SAS Foundation, you must add the path to the SAS Authentication Server certificate file.

The steps to add these paths are documented in the SAS Visual Analytics Installation and Configuration Guide.

Integration with SAS Information Retrieval Studio

SAS Information Retrieval Studio is enabled as the indexing provider by default. Typically, it is used as an integrated component of SAS Visual Analytics. SAS Information Retrieval Studio combines several technologies to provide a comprehensive solution for data collection, indexing, and searching. The version of SAS Information Retrieval Studio that is included with SAS Visual Analytics has been adapted for use with that solution and does not expose the full functionality. As a result, there is a 2-GB limit on the size of the search-content index due to the 32-bit integer size. After the index grows to exceed 2 GB, the processes can consume a large amount of memory.

By contrast, the version of SAS Information Retrieval Studio that is included with the SAS Text Analytics suite of products can handle larger volumes of content by means of distributed deployment and index segmentation. To perform large-scale searches, use one of the SAS Text Analytics solutions.

You can control the amount of data that is fed into the index by deploying search filters. The size of the index is about the same size as the documents being indexed. For example, if a 2-GB data set contains approximately 200,000 documents and the average document size is 10 KB, the index reaches the maximum size that Search Interface to SAS Content supports.

- On Windows, SAS Information Retrieval Studio runs as a local service (for example, SAS [Config-lev1] Information Retrieval Studio).

- On UNIX, the `IRStudio.sh` script runs in the SAS configuration directory under `/Applications/SASInformationRetrievalStudioforSAS`.

You can use the following commands to operate the server on UNIX:

```
IRStudio.sh start | stop | status | restart
```

Generated log files are in the `/logs` subdirectory.

When the SAS middle-tier search facility is first initialized, all available content is fully indexed. Subsequently, incremental changes to SAS content are regularly indexed. If a failure in generating or loading the index occurs, an email message is automatically sent to the address that is designated for administrative messages for your SAS deployment.

Refer to “[Chapter 2 – Configuration](#)” for the procedure to enable the index server of SAS Information Retrieval Studio to receive SAS contents.

Integration with Apache Lucene

If SAS Information Retrieval Studio is not available in the package, or if you want to use a different index server, you can instead configure Apache Lucene for the integration. Apache Lucene is required in IPv6 environments (where SAS Information Retrieval Studio is not supported).

Take the following steps to configure Apache Lucene:

1. Launch SAS Management Console and log in with administrator credentials.
2. Click the **Plug-ins** tab.
3. Navigate to SAS Management Console→Application Management and select Configuration Manager.
4. Right-click **Search Interface to SAS Content 3.6**, and select **Properties**. Then click the **Advanced** tab in the Properties dialog.
5. Verify that the property `searchsas.irstudio.is_available` is set to **false**.

Setting this property to “**false**” enables indexing by Apache Lucene.

Also check the default value of `searchsas.lucene.index.default.dir` to determine where index data will be stored. Change it if required.

To verify that content is being indexed, navigate to your equivalent of the following subdirectory of the SAS configuration directory:

```
Web\Applications\SearchInterfacetoSASContent\Index\Default\
```

If the folder is empty, content is not being indexed.

Apache Lucene Index

If Apache Lucene is enabled as the indexing server, you can use the SAS Management Console to find the location of the Lucene Index. The value is stored in the Advanced Properties.

Check the value of the following property to determine the location of the index:

```
searchsas.lucene.index.default.dir.
```

Typically, it is at your equivalent of the following subdirectory of the SAS configuration directory:

```
...Web\Applications\SearchInterfacetoSASContent\Index\Default\
```

Verify the Installation

You can verify that Search Interface to SAS Content has been installed properly by running a simple test of the API. When Search Interface to SAS Content has been installed and configured, access the following URL:

<http://host.com:port/SASSearchService/rest>

Log in using valid credentials defined in the SAS Metadata Server.

The response returned is a list of resources exposed by the service. Typically, the response is displayed in the browser and, depending on browser behavior, the XML response might be downloaded as a file. If required, open the file in a text editor to verify that a list of links to resources is returned.

Validation of your installation can also be performed by running a successful search in SAS Visual Analytics Hub.

Verify the Integration with SAS Information Retrieval Studio

Two conditions must be met in order for the integration between Search Interface to SAS Content and SAS Information Retrieval Studio to provide results in response to a search query:

- A. SAS Information Retrieval Studio is configured. Typically, it is configured automatically by Search Interface to SAS Content when the indexing process runs for the first time.
- B. The index of SAS Information Retrieval Studio must be populated with data from Search Interface to SAS Content. The data is fed into the index at scheduled intervals by the web application.

You can verify the configuration in the Administration Console of SAS Information Retrieval Studio. Be aware that the Administration Console is disabled by default for improved security. If it is not accessible, verify that the host server is running and that the Administration Console is enabled. You can enable it by taking the steps described in the following topic in the Appendix: “[How can I enable the Administration Console for SAS Information Retrieval Studio?](#)” on page 13.

Test the integration by accessing the following URL, including the appropriate host and port:

`http://<hostname>:<port>/SASSearchService/rest/content/search?query=*`

Note: You might be required to provide user credentials and log in.

Verify the SAS Information Retrieval Studio Server Installation

To verify that the SAS Information Retrieval Studio is installed correctly, access the Administration Console and check the server status.

Note that the Administration Console of SAS Information Retrieval Studio is disabled by default for improved security. You can enable it by taking the steps described in “[How can I enable the Administration Console for SAS Information Retrieval Studio?](#)” on page 13 in the Appendix.

Use the following URL with the appropriate host and port to call up the Administration Console:

`http://<hostname>:<port>/`

The default port for the SAS Information Retrieval Studio Administration Console is 10651. For example, type a URL like `http://windows-vm-005:10651/`

When you have successfully accessed the Administration Console, verify whether the following servers are listed as running:

- Proxy Server
- Pipeline Server
- Indexing Server
- Query Server

These servers are required to enable the Search applications to work. If any of them are missing from the administration page, you might have a licensing issue. Contact SAS Technical Support if you suspect a licensing issue.

If a server is not listed as running, start it by clicking the **Start** button on the top right side of the server Details page. Then wait for a minute or so to make sure it does not stop again. Troubleshooting has occasionally turned up a problem with this manual step.

If all of these servers are listed and are running, SAS Information Retrieval Studio has been successfully installed.

Chapter 2 — Configuration Options

Search Interface to SAS Content 3.6 supports feeding SAS contents to the Indexing Server of SAS Information Retrieval Studio 1.53. It applies user authentication and enables authorized users to search the SAS data. Users can view results based on their roles.

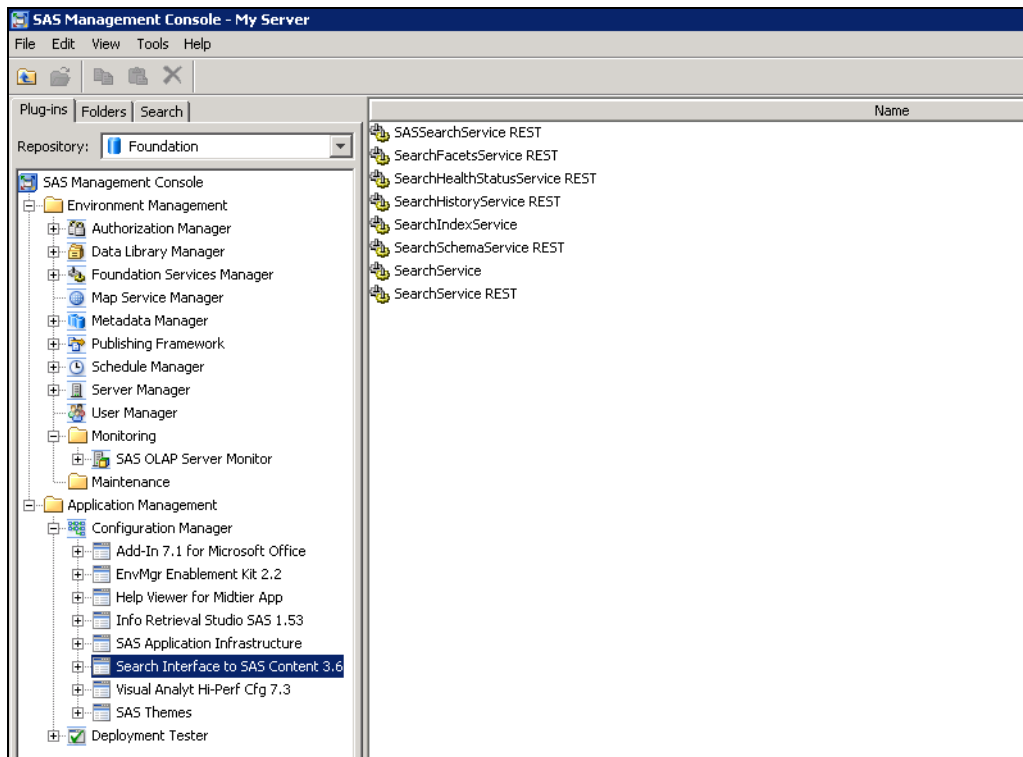
Configuration is automated through the SAS Deployment Wizard. If any component is not working properly following installation and deployment, take the steps below to verify the configuration.

Step 1 - Verify the Registration Information in Metadata

Make sure that both Search Interface to SAS Content and SAS Information Retrieval Studio are installed and running. Also verify that the correct information is used by Search Interface to SAS Content 3.6 to connect to SAS Information Retrieval Studio 1.53.

Take the following steps for verification:

1. Launch SAS Management Console and log in with administrator credentials.
2. Click the **Plug-ins** tab.
3. Navigate to **SAS Management Console**→**Application Management** and select **Configuration Manager**.



4. Right-click **Search Interface to SAS Content 3.6** and select **Properties**. Then click the **Advanced** tab in the Properties dialog box.

| Property Name | Property Value | Locked |
|--|--|--------|
| App.AcceptDirectCredentials | true | |
| sas.web.csrf.referers.performCheck | true | |
| searchsas.auth.provider.url | http://host:port/SASSearchService/?forward=authorize&* | |
| searchsas.excluded.object.types | LogicalServer.GenericBatch,User,Condition,Service,Channel,Spawner,Server.DataStepBatc... | |
| searchsas.feeder.scheduler.interval.minutes | 15 | |
| searchsas.feeder.scheduler.is_enabled | true | |
| searchsas.feeder.scheduler.notified.types.i... | 2 | |
| searchsas.history.records.limit | 20 | |
| searchsas.index.fullfeed.daily.enabled | true | |
| searchsas.irstudio.index.path | work\index\index | |
| searchsas.irstudio.is_available | true | |
| searchsas.irstudio.proxy.server.port | 10661 | |
| searchsas.irstudio.query.server.port | 10731 | |
| searchsas.irstudio.server.host | irstudiohost | |
| searchsas.irstudio.server.port | 10651 | |
| searchsas.lucene.index.default.dir | C:\SAS\Config\Lev1\Web\Applications\SearchInterfacetoSASContent\Index\Default | |
| searchsas.lucene.supported.locale | en_US | |
| searchsas.max.auth.records.count | 20000 | |
| searchsas.notification.email.is_active | true | |
| searchsas.notification.email.sender.address | abc@sas.com | |
| searchsas.notification.email.to.address | abc@sas.com | |

5. Verify that the following properties have the correct values. These properties were configured during the installation and configuration of SAS Information Retrieval Studio:
- **searchsas.irstudio.server.host**: the hostname of the server where SAS Information Retrieval Studio servers are running.
 - **searchsas.irstudio.server.port**: the port number of the server where SAS Information Retrieval Studio is running. The default (10651) should not be changed.
 - **searchsas.irstudio.query.server.port**: the port number of the server where the query server of SAS Information Retrieval Studio is running.
 - **searchsas.irstudio.proxy.server.port**: the port number of the server where the proxy server of SAS Information Retrieval Studio is listening.
 - **searchsas.auth.provider.url**: constructs the URL that is used to enable communications among SAS components. Search Interface to SAS Content uses this URL to contact the authorization service provider, which filters the query results that are returned based on user permissions.

If a load balancer or firewall is running in between client machines and the SAS Web Server, or if a proxy server is being used to mask URLs, you might need to change the default value on the **External Connection** tab in this Properties dialog box. By default, the **Use internal connection information** check box is selected. Clear that check box, and then enter the connection information for the proxy in situations where the proxy server must route connections.

Note: If you are unsure of the correct values, see the installation documentation for SAS Visual Analytics. For more information about configuration options in a proxy server or firewall environment, see “Specifying Connection Properties” in the SAS(R) 9.4 Intelligence Platform: Middle-Tier Administration Guide.

The following internal property should not be changed:

- **searchsas.irstudio.index.path:** Search Interface to SAS Content uses this path to initialize SAS Information Retrieval Studio index path.

Step 2 - Verify the Configuration of SAS Information Retrieval Studio

Search Interface to SAS Content 3.6 must configure SAS Information Retrieval Studio to support user authorization and to feed SAS contents into the index. The configuration is performed automatically when the content is pushed to the index for the first time.

To verify the configuration, open the Administration Console of Information Retrieval Studio and open the Indexing Server. The **Configuration** section should show the fields populated for the information to be indexed.

If the Information Retrieval Studio Administration Console is not running, refer to Question 2 on page 13 in the “Appendix – Troubleshooting (FAQ)” below.

The screenshot shows the SAS Information Retrieval Studio Administration Console. The left sidebar lists various components, with the Indexing Server selected. The main panel displays the Configuration section for the Indexing Server, which includes a table of field names and their functionalities. A green circle highlights this table.

| Field Name | Functionality |
|----------------------------|----------------------------------|
| relationship.depends.id | Custom (standard, info) |
| comments.content | Custom (standard, info) |
| modifieddate | Custom (standard, info) |
| relationship.includes.id | Custom (standard, info) |
| lastmodifiedby | Custom (standard, boolean, info) |
| relationship.synonymous.id | Custom (standard, info) |
| relationship.child.name | Custom (standard, info) |
| inputquerynames | Custom (standard, info) |
| sasauthordisplay | Custom (standard, boolean, info) |

Rebuild the Search Index

Typical usage of Search Interface to SAS Content does not require you to interact with the Index Server; all such interactions are performed by the Search Service and are transparent to you. However, if a failure occurs that affects the search index, you can manually delete and rebuild the index.

Note: The search index forms the basis for all searches using SAS Information Retrieval Studio or Apache Lucene.

To delete the index and rebuild it, take the following steps:

1. Open the Administration Console of SAS Information Retrieval Studio.
2. Click the **Indexing Server** option in the left pane.

3. Stop the indexing server.
4. Click the **Delete Index** button. You are asked to confirm the Delete operation.
5. Restart the indexing server.
6. Restart all SAS Web Application Server instances.

Search Interface to SAS Content will then index all data. The process can take 2 or 3 hours.

Disable the CSRF Filter

CSRF filtering prevents Cross-Site Request Forgery attacks. This filtering is provided because SAS Web Infrastructure Platform applications, including Search Interface to SAS Content, might otherwise be vulnerable to CSRF attacks.

With CSRF protection enabled, requests that do not originate from sites that belong to a whitelist are rejected. By default, the implications of this change are that the SAS installation is secure against browser-based CSRF attacks. In addition, applications that are external to the SAS installation cannot link to SAS web applications. If you link to a SAS application from a company intranet or a portal page that is not hosted in the SAS installation, you encounter `Access denied` messages when you click those links.

The Search Service property named `“sas.web.csrf.referers.performCheck”` determines whether the CSRF filter is deployed. The default value for this property is `true`, which indicates that CSRF checking is performed. You can set the value to `false` to disable the CSRF filter. When it is disabled, no checking is performed, regardless of the value of any other setting.

To change the value of the `performCheck` property, take the following steps:

1. Open the SAS Management Console and go to **Application Management** → **Configuration Manager**.
2. Right-click Search Interface to SAS Content, and click Properties.
3. Click the **Advanced** tab.
4. Disable the CSRF filter by setting the value of the `“sas.web.csrf.referers.performCheck”` property to **false**.
5. Click **OK**.
6. To enable the changes to take effect, restart all SAS Web Application Server instances.

This property is applied globally and is also available on the SAS Application Infrastructure Advanced properties tab. Individual SAS applications inherit SAS Application Infrastructure settings. If this property is locked within the Configuration Manager of the SAS Management Console, its value is not overridden when you change it in Search Interface to SAS Content.

Add Sites to the CSRF Whitelist

SAS Support recognizes many legitimate cases where it is necessary to link to SAS web applications. For such cases, it is much safer to add an individual site that contains a link to a SAS web application to the SAS CSRF whitelist than it is to disable CSRF filtering.

Take the following steps to add a site to the whitelist:

1. Open the SAS Management Console and go to **Application Management** → **Configuration Manager**.

2. Open the properties for SAS Application Infrastructure.
3. On the **Advanced** tab, add a property named `sas.web.csrf.referers.knownHosts`. Specify the value as a comma-delimited list of whitelisted sites.

For example, to add 'my-intranet.example.com' and 'test.my-intranet.example.com' to the whitelist, set the value of the property to the following:

```
http://my-intranet.example.com/,http://test.my-intranet.example.com/
```

4. Restart all SAS Web Application Server instances.

Notes: *Including a trailing slash at the end of each value is important because omitting the slash means that sites can use a prefix attack to bypass the protections.*

Be sure to include port numbers in the value if the whitelisted site uses ports other than the standard 80 or 443 for HTTP and HTTPS, respectively.

You can also restrict an allowed entry to an application on the whitelisted site. To restrict the whitelist entry to an application, include the path to the application in the value that you add to the `knownHosts` property. For example, use the following syntax:

```
http://my-intranet.example.com/my-application-path/
```

Whitelisting Methods

In addition to whitelisting sites, you can also whitelist certain HTTP methods. For example, you can choose to allow GET requests from any site. However, the result of this configuration is that SAS applications that use GET requests to invoke certain actions are now susceptible to CSRF attacks.

If you decide to whitelist certain methods, you can instead set the `sas.web.csrf.referers.skipMethods` property in the Advanced properties for the SAS Application Infrastructure. This value should be a comma-delimited list of HTTP methods (for example, GET, OPTIONS, TRACE) that are to be skipped.

To enable CSRF checking and enforcement, set the `sas.web.csrf.referers.performCheck` property to `true` in the Advanced properties for SAS Application Infrastructure.

After changing any of the properties that are discussed above, you must restart the SAS Web Application Server in order for the changes to take effect.

You also have the option to set CSRF properties on individual web applications in your SAS installation by using Configuration Manager. After you make any changes in Configuration Manager, you must restart all of the web application servers in your SAS middle tier in order for the settings to take effect.

Chapter 3 — Logging

Log Settings and Location

Logging for Search Interface to SAS Content is enabled by default. Also by default, the log4j root level is set to *error*, which can be changed to *warn* or *info* as required. The configuration file for Search Interface to SAS Content logs can be found at this location:

SAS-Configuration-Directory/LevN/Web/Common/LogConfig/SASSearchService-log4j.xml

Logs are generated at this location:

SAS-Configuration-Directory/LevN/Web/Logs

For “*LevN*,” substitute the location of the configuration file for the selected environment.

To change the default logging level from error to warn or info, take the following steps:

1. Use your preferred text editor to modify the following file:

SAS-Configuration-Directory/LevN/Web/Common/LogConfig/SASSearchService-log4j.xml

2. Restart all SAS Web Application Server instances for the changes to take effect.

Appendix — Troubleshooting (FAQ)

1. I see no results when I submit a query string in Visual Analytics Hub. What can I do?

First, verify that Search Interface to SAS Content and SAS Information Retrieval Studio have been correctly installed and configured. Chapter 1 contains more information.

If both components are installed and configured, the issue could be the query string, which may not have any matching record for the user (keeping the authorization constraints in mind). Try using ``*`` as the query string, which should return some results.

Another possibility is that a load balancer or firewall is preventing communications between client machines and the Middle-Tier components of your SAS solution software. The “authorization provider URL” property constructs the URL that is used to enable communications among SAS components. Search Interface to SAS Content uses this URL to contact the authorization server to authenticate the user who is performing the search before displaying results. If you suspect that your load balancer or firewall is causing this issue, or if a proxy server is being used to mask URLs in your network, you might need to change the default value of the `searchsas.auth.provider.url` property to an external URL. For more information, see [“What configuration changes are required for an external reverse proxy environment?”](#) on page 17.

2. How can I enable the Administration Console for SAS Information Retrieval Studio?

The SAS Information Retrieval Studio Administration Console is disabled by default so that the settings cannot be changed by an anonymous user.

You may need to enable the interface if SAS Information Retrieval Studio is running abnormally. You can use the administration interface to check server status and start/stop services.

To enable the SAS Information Retrieval Studio Administration Console, take the following steps:

1. On the machine that is hosting SAS Information Retrieval Studio, which is typically the SAS Application Server machine that hosts the SAS Visual Analytics server tier, make a backup copy of the following file:

```
SAS-Configuration-Directory/LevN/Applications
/SASInformationRetrievalStudioforSAS/work/information-retrieval-
studio.conf.
```

2. Use your preferred text editor to open `information-retrieval-studio.conf`.
3. Add the following line to the bottom of the file:
`enable-web-admin-interface=true`
4. Save the file and close it.
5. Restart the SAS Information Retrieval service:

```
SAS-Configuration-Directory/LevN/Applications
/SASInformationRetrievalStudioforSAS/IRStudio.sh restart
```

6. Access the SAS Information Retrieval Studio Administration Console:

```
http://host-name:port/
```


For the host name, substitute the host name of the machine that is hosting SAS Information Retrieval Studio. The default port is 10651.

To disable the SAS Information Retrieval Studio Administration Console, take the following steps:

1. Use your preferred text editor to open the following file:


```
SAS-Information-Retrieval-Studio-config-directory/work
/information-retrieval-studio.conf
```
2. Remove the following line:


```
enable-web-admin-interface=true
```
3. Save the file and close it.
4. Restart the SAS Information Retrieval service.

3. How can I verify that SAS Information Retrieval Studio has been configured correctly?

First, verify that SAS Information Retrieval Studio is installed correctly and that the search application has been installed correctly. Use the REST URL for search, which also silently configures the SAS Information Retrieval Studio for Search. The REST APIs can be accessed by using the context URL for the Search service: <http://host:port/SASSearchService/rest>

Next, open the Administration Console of SAS Information Retrieval Studio, and then open the Indexing Server. Verify the Information Retrieval Studio configuration by checking the following elements in the **Configuration** section of the Indexing Server:

- title
- description
- link
- sastype
- sasmetatype
- sasowner
- sasid
- keywords
- promptlabels

| | | |
|-----------------------------------|-------------------|----------------------------------|
| Proxy Server (running) | ▼ Configuration | |
| Pipeline Server (running) | Field Name | Functionality |
| <u>Indexing Server (running)</u> | title | Custom (standard, info) |
| Query Server (running) | description | Custom (standard, info) |
| Query Web Server (running) | link | Custom (standard, info, URL) |
| Query Statistics Server (running) | sastype | Custom (standard, boolean, info) |
| | sasmetatype | Custom (standard, info) |
| | sasowner | Custom (standard, boolean, info) |
| | sasid | Custom (standard, info) |
| | keywords | Custom (standard, boolean, info) |
| | promptlabels | Custom (standard, info) |

These elements must all be present in the **Configuration** section of the target setup. This ensures that the configuration of SAS Information Retrieval Studio has been invoked by Search Interface to SAS Content. Note that the page may include more field names than are shown above.

4. How can I verify the host and port of SAS Information Retrieval Studio?

The host and port number that SAS Information Retrieval Studio is using can be taken from the properties of the metadata software component of Search Interface to SAS Content.

Take the following steps:

1. Launch SAS Management Console and log in with administrator credentials.
2. Click the **Plug-ins** tab.
3. Navigate to SAS Management Console Application Management and select **Configuration Manager**.
4. Right-click **Search Interface to SAS Content 3.6** and select **Properties**. Then click the **Advanced** tab in the Properties dialog box.

| Property Name | Property Value | Locked |
|--|--|--------|
| App.AcceptDirectCredentials | true | |
| sas.web.csrf.referers.performCheck | true | |
| searchsas.auth.provider.url | http://host:port/SASSearchService/?forward=authorize&* | |
| searchsas.excluded.object.types | LogicalServer.GenericBatch,User,Condition,Service,Channel,Spawner,Server.DataStepBatc... | |
| searchsas.feeder.scheduler.interval.minutes | 15 | |
| searchsas.feeder.scheduler.is_enabled | true | |
| searchsas.feeder.scheduler.notified.types.i... | 2 | |
| searchsas.history.records.limit | 20 | |
| searchsas.index.fullfeed.daily.enabled | true | |
| searchsas.irstudio.index.path | work\index\index | |
| searchsas.irstudio.is_available | true | |
| searchsas.irstudio.proxy.server.port | 10661 | |
| searchsas.irstudio.query.server.port | 10731 | |
| searchsas.irstudio.server.host | irstudiohost | |
| searchsas.irstudio.server.port | 10651 | |
| searchsas.lucene.index.default.dir | C:\SAS\Config\Lev1\Web\Applications\SearchInterfacetoSASContent\Index\Default | |
| searchsas.lucene.supported.locale | en_US | |
| searchsas.max.auth.records.count | 20000 | |
| searchsas.notification.email.is_active | true | |
| searchsas.notification.email.sender.address | abc@sas.com | |
| searchsas.notification.email.to.address | abc@sas.com | |

searchsas.irstudio.server.host: The hostname of the server where SAS Information Retrieval Studio servers are running.

searchsas.irstudio.server.port: The port number of the server where the SAS Information Retrieval Studio Administration Console is running.

searchsas.irstudio.proxy.server.port: The port number where the feed is pushed in SAS Information Retrieval Studio.

searchsas.irstudio.query.server.port: The port number where queries to the Index are sent.

5. Some of the required fields are missing from the IR Studio schema, so its contents are not searchable. How can I add the missing fields?

If fields are missing from the configuration of the Information Retrieval Studio Indexing Server, the contents of those fields would neither be indexed nor available in responses. In this scenario, it is important to reconfigure the Indexing Server to make the missing fields available for searching.

Take the following steps to reconfigure the schema:

1. Stop all SAS Web Application Server instances.
2. Stop the Information Retrieval Studio server.
3. Go to the following location relative to the configuration directory:
`LevN\Applications\SASInformationRetrievalStudioforSAS\work`
4. Open the file `pipeline-server.db` and replace the text 'BIRD' with 'BIRD1'.
5. Indexed content is located in the 'index' folder inside the current directory of Information Retrieval Studio (the 'work' directory). Manually delete all of the files in this folder.
6. Save the file, and restart the Information Retrieval Studio server.
7. Start the Search Interface to SAS Content Web Application Server.

After these steps are completed, but before content is pushed to the index server, Information Retrieval Studio is reconfigured with the latest information available through the Index Data Service schema API.

6. I have a 'Connection refused' message in my log file. What should I do?

If queries are failing, check your log file for an error message similar to the following:

```
Please check that the host/port are correct and the Server is running
Message = Connection refused: status = SOCKET_PROTOCOL_ERROR
```

As the message indicates, either the host and port that are required to push the feed to SAS Information Retrieval Studio are not correctly configured, or the proxy server for SAS Information Retrieval Studio is down.

Refer to the diagram in "[Architecture](#)" on page 1. Step 2 is the relevant depiction of this data flow: In that step, the SAS Search Service makes a connection to the Information Retrieval Studio Query Server on port 10731 by default to send the search request. If you provided an incorrect port during the installation, correct it. You can verify the port that was configured by checking the metadata.

Or check the server status and start the server if necessary.

7. I am receiving email messages that report errors. What do they mean?

The email notifications that you receive from Search Interface to SAS Content typically report that "Search Interface to SAS Content encountered error while feeding the index to index server." The following email messages describe errors that you might see while running Search Interface to SAS Content and identify reasons for feed errors:

- 7 a) **Error Message:** `com.sas.svcs.search.client.IndexServiceException: com.sas.svcs.search.client.SearchServiceException: IR Studio query Server not reachable.`

You received this email message because Search Interface to SAS Content failed to connect to Information Retrieval Studio. Make sure Information Retrieval Studio is up and running. (Refer to Questions 2 and 3 [above](#).) Also verify that the correct values for the Information Retrieval Studio host and port are being used. (Refer to Question 4 [above](#).)

- 7 b) **Error Message:** `com.sas.svcs.search.client.IndexServiceException: Exception encountered related to Index Path of Lucene.`

You are receiving this email message because Search Interface to SAS Content could not access the Apache Lucene index directory. Verify that `searchsas.lucene.index.default.dir` has a valid directory path. (Refer to **Question 4** [above](#) and the image of the SAS Management Console.) You should also verify that the SAS install user has write access to the specified directory.

- 7 c) **Error Message:** `com.sas.svcs.search.client.IndexServiceException: An error occurred while retrieving data from Index Data Service.`

If you receive this email message, make sure that the Web Infrastructure Platform Data Server is running. If it is not running, start it, and also restart all Web application server instances. If the Web Infrastructure Platform Data Server is already running, check the SASSearchService log file for the exact cause of the failure.

- 7 d) **Error Message:** Search Interface to SAS Content encountered error while feeding the content to the Indexing Server.

This email message reports that the Search Interface to SAS Content “encountered [a] problem while retrieving content” for a specified list of data types. The message further instructs you to check web application server logs. The affected host is listed to help you parse log content.

The cause of this type of error can vary. It can be caused by an internal fault in the object type, or by corrupt data handling by the object type owner. When you check the web application server log, check for exceptions. The log information should indicate which team is responsible for troubleshooting the problem, based on the object types that are affected.

By default, the web application server logs are located in the following directory:
`Config/LevN/Web/Logs/SASServerN/SASSearchService.log`

You might also need a hot fix to resolve this issue. If your environment included a recent migration, some data object types can be unknown to the Metadata Server when indexing runs. The Indexing Service tries to find the lookup support service for those objects, and when it fails, it throws the exception.

If you believe that this description fits your situation, contact SAS Technical Support.

8. What configuration changes are required for an external reverse proxy environment?

If an external reverse proxy is used, ensure that the JVM parameter is set for the SAS Web Application Server as described below:

- For Windows, add
`wrapper.java.additional.XX=-Dsas.retry.internal.url=true`
to the following file:
`config\LevN\Web\WebAppServer\SASServer1_1\conf\wrapper.conf`
- For Linux (for x64), add
`-Dsas.retry.internal.url=true`
to the following file:
`config/LevN/Web/WebAppServer/SASServer1_1/bin/setenv.sh`

In an external reverse proxy environment, be sure to follow the steps provided in the *Middle Tier Administration Guide*. See the section titled “Configuring the Middle Tier to Use an Existing Customer Reverse Proxy.” You can find this guide at the following location:

<http://support.sas.com/documentation/cdl/en/bimtag/68217/HTML/default/viewer.htm>

9. How can I recover lost data for the Indexing Server after an indexing failure?

Search Interface to SAS Content does not lose data after an indexing failure. After the cause of the indexing failure has been found and corrected, all missing index fields will be populated automatically. Typical usage of Search Interface to SAS Content does not require you to interact with the Index Server; all such interactions are performed by the Search Service and are transparent to you. However, if a failure occurs that affects the search index, you can manually delete and rebuild the index. Any problem that involves corrupted Information Retrieval Studio indexes requires you to re-index all SAS data. Take the following steps:

1. Stop all SAS Web Application Server instances.
2. Stop Information Retrieval Studio. You can use the Administration Console to perform this step. For more information about the Administration Console, see [How can I enable the Administration Console for SAS Information Retrieval Studio?](#) on page 13.
3. Delete the Information Retrieval Studio index. You can use the Administration Console to perform this step.
4. Start all SAS Web Application Server instances.
5. Start Information Retrieval Studio.

Search Interface to SAS Content will then index all data. The process can take two to three hours.

10. Search does not work with SSL. What steps can I take in this situation?

SAS Information Retrieval Studio requires manual configuration to work in an environment where Secure Sockets Layer (SSL) encryption is used. Ensure that it has been configured properly. For details, refer to the section titled “Post-Deployment Step for SAS Information Retrieval Studio” in *SAS Visual Analytics Installation and Configuration Guide*, available at <http://support.sas.com/documentation/solutions/va/73/en/vaicg.pdf>

11. How can I change the frequency of indexing?

The scheduler by default uses an interval of 15 minutes for default indexing tasks (of all types) and an interval of 2 minutes for indexing tasks.

If you want to increase the frequency of indexing to make newly created or modified objects immediately available in the search results, or if you want to reduce the frequency of indexing to avoid unnecessary processing, take the following steps:

1. Open the SAS Management Console and go to Application Management → Configuration Manager.
2. Right-click the software component with the name “Search Interface to SAS Content X.X” and click **Properties**.
3. Click the **Advanced** tab.
4. The interval for the default indexing task can be customized by changing the value of the `searchsas.feeder.scheduler.interval.minutes` property.
5. The interval for the notified indexing task can be customized by changing the value of the `searchsas.feeder.scheduler.notified.types.interval.minutes` property.
6. Click **OK** to save the changes.

The changes that you make are reflected immediately. Restarting the Web Application Server is not required.

12. What is the frequency of email notifications?

Email notifications are now sent on the following intervals:

- **24 hours:** In cases where a single object or some specific types of objects failed to index, email notifications are sent once in 24 hours.
- **04 hours:** If indexing cannot be performed on any object of any type, an email notification is sent once every 4 hours. (The reason for this condition could be that the Index Data Service is down, or Information Retrieval Studio is down.)
- **01 hours:** If an unexpected exception occurs (most likely an internal exception), email notifications are sent hourly.

13. Is IPv6 supported?

Information Retrieval Studio does not support deployments on IPv6-only servers. As a result, the Search service is not available on IPv6-only servers. The service works on IPv4-only servers, and it also works on servers that support IPv4 and IPv6.

Configure Apache Lucene in an IPv6 environment to enable Search Service to SAS Contents. For more information, see “Integration with Apache Lucene” on page 4.

14. How can I enable/disable the daily full feed?

The full data feed that occurs at midnight actually runs between 00:01 hours and 00:15 hours if the default scheduler interval is 15 minutes. It pushes all the available records to the index. However, by default, the full feed is disabled.

Enable the full data feed to run daily at midnight by setting the property named `searchsas.index.fullfeed.daily.enabled` to `true`. If it is enabled, you can set the value to `false` to disable it.

15. How can I change the default setting for the maximum number of records returned?

Your search results are subject to a default setting that limits the number of records that can be returned from a single query. If your search results include a message indicating that “the matching record count has exceeded a threshold of *N* records,” you have experienced this limit.

If you have a very large database and do not want to see this message every time you do a search, you can change the maximum number of records that are returned from a search.

This limit is imposed in order to ensure optimal search performance. Each object that is returned from a query must be authorized for the user account that is performing the search. This series of checks and authorizations can easily become a bottleneck when searching a large number of database objects.

To change the maximum number of records that are returned per search, take the following steps:

1. Open the SAS Management Console and go to **Application Management → Configuration Manager**.
2. Right-click the software component with the name “Search Interface to SAS Content X.X” and click **Properties**.
3. Click the **Advanced** tab.
4. The number of records returned in a search can be customized by changing value of the `searchsas.max.auth.records.count` property. It is set to 20000 by default.
5. Click **OK** to save the changes.

Technical Support

If you need assistance with the software, we ask that only SAS support personnel call our Technical Support Division.

- For U.S. and Canadian customers, support is provided from our corporate headquarters in Cary, North Carolina. You may call (919) 677-8008, Monday through Friday.
- Customers outside of the U.S. can obtain local-language technical support through the local office in their countries. Customers in these locations should contact their local office for specific support hours. See <http://support.sas.com/techsup/contact/index.html> for contact information for local offices.

Before you call, we recommend exploring the SAS Support Web site at <http://support.sas.com/techsup/>

This site offers access to the SAS Knowledge Base, as well as discussion forums, Technical Support contact options, and other support materials that may answer your questions.



THE
POWER
TO KNOW.

SAS is the leader in business analytics software and services, and the largest independent vendor in the business intelligence market. Through innovative solutions delivered within an integrated framework, SAS helps customers at more than 50,000 sites improve performance and deliver value by making better decisions faster. Since 1976, SAS has been giving customers around the world THE POWER TO KNOW®.