



THE  
POWER  
TO KNOW.

# **SAS<sup>®</sup> Enterprise Content Categorization Servers 12.1 Administrator's Guide**



The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2012.  
*SAS® Enterprise Content Categorization Servers 12.1: Administrator's Guide*. Cary, NC:  
SAS Institute Inc.

## **SAS® Enterprise Content Categorization Servers 12.1: Administrator's Guide**

Copyright © 2012, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

**For a hard-copy book:** No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

**For a Web download or e-book:** Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

**U.S. Government Restricted Rights Notice:** Use, duplication, or disclosure of this software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227-19, Commercial Computer Software-Restricted Rights (June 1987).

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

1st electronic book, March 2012

SAS® Publishing provides a complete selection of books and electronic products to help customers use SAS software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit the SAS Publishing Web site at [support.sas.com/publishing](http://support.sas.com/publishing) or call 1-800-727-3228.

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 registered trademarks or trademarks of their respective companies.

---

# Contents

---

<b>About This Book .....</b>	<b>1</b>
Audience .....	1
Prerequisites .....	1
Conventions .....	2
<b>What's New in SAS Enterprise Content Categorization Servers 12.1 ..3</b>	
Documentation Changes for the 12.1 Release .....	3
<b>1 About SAS Enterprise Content Categorization Servers .....</b>	<b>7</b>
1.1 What is SAS Enterprise Content Categorization Servers? .....	7
1.2 Benefits of Using SAS Enterprise Content Categorization Servers .....	8
1.3 How Does SAS Enterprise Content Categorization Servers Work? .....	9
1.4 About the Architecture .....	9
<b>Part 1: Installing the Servers .....</b>	<b>11</b>
<b>2 Using the Installation and Related Processes .....</b>	<b>13</b>
2.1 Before You Install the Servers .....	13
2.1.1 Overview of Installation .....	13
2.1.2 Prerequisites .....	14
2.1.3 Using the SAS License File .....	15
2.2 Installing SAS Enterprise Content Categorization Servers .....	16
2.2.1 Install on Windows .....	16
2.2.1.A Before You Install on Windows .....	16
2.2.1.B Install on Windows .....	16
2.2.1.C The Folders That Appear after Installation .....	27
2.2.2 Installing on UNIX .....	27
2.2.2.A Install on UNIX .....	27
2.2.2.B See the Directory Structure .....	28
2.3 Starting and Stopping the Servers on a Windows Machine .....	29
2.4 Uninstall on Windows .....	30
2.5 Access the Servers .....	33
2.6 Processing Documents .....	34

---

<b>Part 2: SAS Document Conversion Server .....</b>	<b>37</b>
<b>3 Processing Various File Types into Plain Text Format .....</b>	<b>39</b>
3.1 Overview of SAS Document Conversion Server .....	39
3.2 The File Types That SAS Document Conversion Server Can Convert .....	40
<b>Part 3: SAS Content Categorization Server .....</b>	<b>41</b>
<b>4 Configuring and Running SAS Content Categorization Server .....</b>	<b>43</b>
4.1 Configuration Overview .....	43
4.2 SAS Content Categorization Server Configuration File .....	45
4.2.1 Windows Configuration File .....	45
4.2.2 UNIX Configuration File .....	45
4.2.3 The Directives .....	46
4.3 Specifying Project Users and Creators .....	50
4.4 Add a Project .....	51
4.5 Using the Models Directory .....	54
4.6 Using the Descriptors Directory .....	55
4.7 Using cat_log and concept_log Files .....	57
4.8 Specifying Multiple Project Files .....	58
4.9 Sending Documents to the Server .....	59
4.10 Run SAS Content Categorization Server .....	59
4.10.1 Verify That the Server Is Running on Windows .....	59
4.10.2 Run the Server on UNIX .....	60
4.11 Optimize Performance on a Client Windows Machine .....	61
4.11.1 Overview of Performance Optimization .....	61
4.11.2 Before and After You Optimize Performance .....	61
4.11.3 Adjust the TCP Time Wait State .....	61
4.11.4 Reset Ephemeral Ports .....	62
<b>5 SAS Content Categorization Server Web Administration .....</b>	<b>63</b>
5.1 Overview of SAS Content Categorization Server Web Administration .....	63
5.2 Access the Administration Web Interface .....	64
5.3 Using the Administration Web Page .....	65
5.3.1 Overview of the Administration Web Page .....	65
5.3.2 Use the Are you there? Page .....	65
5.3.3 Use the SAS Content Categorization Server Projects List Page .....	66
5.3.4 Use the SAS Content Categorization Server Categorization Statistics (Matches) Page .....	69

---

5.3.5 Use the SAS Content Categorization Server Categorization Statistics (Timing) Page .....	70
5.3.6 Use the SAS Content Categorization Server Concept Extraction Statistics Page .....	73
5.3.7 Use the SAS Enterprise Content Categorization Servers Concept Extraction Statistics (Timing) Page .....	75
5.3.8 Use the SAS Content Categorization Server Contextual Extraction Statistics (Matches) Page .....	77
5.3.9 Use the SAS Content Categorization Server Contextual Extraction Statistics (Timing) Page .....	79

## **Part 4: SAS Content Categorization Collaborative Server 81**

<b>6 Setting Up a Project on the Server .....</b>	<b>83</b>
6.1 Overview of Setting Up a Project on SAS Content Categorization Collaborative Server .....	83
6.2 Installing and Configuring a SQL-Based Server .....	84
6.2.1 Install the SQL-Based Server .....	84
6.2.2 Change the Size of the .cnf File for Linux .....	85
6.2.3 Change the Size of the Allowed Packet Size for Windows .....	85
6.3 Installing the _tk240 Utility .....	86
6.4 Setting Up ODBC Data Sources .....	87
6.4.1 Driver Types .....	87
6.4.2 Installing the ODBC Drivers .....	87
6.4.3 Configuring Data Sources .....	88
6.4.3.A Credentials .....	88
6.4.3.B Configuring ODBC Data Sources on UNIX .....	89
6.4.3.C Configure MySQL Data Sources .....	91
6.4.3.D Configure Microsoft SQL Server Data Source .....	94
6.4.3.E Download and Configure Oracle ODBC Drivers .....	101
6.5 Defining Users .....	105
6.6 Logging In .....	105
6.7 Quick Start Guide Using _tk240 .....	106

## **Part 5: Appendixes ..... 109**

<b>A Troubleshooting .....</b>	<b>111</b>
A.1 Installing SAS Enterprise Content Categorization Servers .....	111
A.1.1 The Installer Fails to Copy Files on Windows Vista or Server 2008	111

---

A.2 Tips and Guidelines for SAS Content Categorization Server .....	112
A.2.1 Trying to Upload a Project with a Duplicate Name .....	112
A.2.2 Noun Phrases Incorrectly Split by an Apostrophe (‘) .....	112
A.2.3 Using Synonym Lists .....	112
A.3 Using the Configuration File .....	113
A.3.1 Trailing Spaces in the Configuration File .....	113
A.3.2 Modifying the Configuration File .....	113
A.4 If SAS Content Categorization Server Does Not Appear to Be Running ..	114
A.4.1 Overview of When the Server Does Not Appear to Be Running ....	114
A.4.2 Checking and Debugging on a Windows Machine .....	114
A.4.3 Checking and Debugging on a UNIX Machine .....	115
A.5 Tips and Guidelines for SAS Content Categorization Collaborative Server	115
A.5.1 Working with the ODBC Drivers .....	115
A.5.1.A Receiving Error Messages on Windows .....	115
A.5.1.B Receiving Error Messages on UNIX .....	116
A.5.1.C Adding a User with _tk240 on Linux .....	116
A.5.1.D Connecting to the Database and SQL Syntax .....	116
<b>B Recommended Reading .....</b>	<b>117</b>
<b>C Glossary .....</b>	<b>119</b>
<b>Index .....</b>	<b>121</b>

---

# About This Book

---

## Audience

SAS Enterprise Content Categorization Servers is designed for the following types of administrators and users, depending on the server components that you install:

- Users who want to run SAS Document Conversion Server to preprocess files into plain text format.
- Administrators and users who want to collaborate on a single project, or work on one project as a single user using SAS Content Categorization Collaborative Server.
- Users who access the Administration Web Page to see the matching statistics for categories and concepts that are applied to input documents by SAS Content Categorization Server.
- Programmers who want to use the supplied APIs to write their own applications.

## Prerequisites

Here are the prerequisites for using SAS Enterprise Content Categorization Servers:

- Install Java 1.5 or higher and Python version 2.3 or higher if you plan to install the Java and Python APIs for SAS Content Categorization Server. If you plan to install SAS Document Conversion Server Java API, make sure that the Java 1.5, or higher, run-time environment is installed on the machine.
- Install one or more of the following SAS Enterprise Content Categorization Servers components:
  - SAS Document Conversion Server: Preprocess your documents into plain text format.

- SAS Content Categorization Collaborative Server: Enable multiple users to work collaboratively on the same SAS Content Categorization Studio project.
- SAS Content Categorization Server: Apply the categories or concepts that you uploaded to this server to documents.
- Configure the server, or servers, that you install.

## Conventions

This manual uses parts to define the chapters that are install or component specific. This manual also uses the following typographical conventions:

Convention	Description
TGM_ROOT	The root directory where SAS Enterprise Content Categorization Servers is installed, typically the following:  <b>Windows:</b> C:\Program Files\Teragram\SAS Enterprise Content Categorization Servers\ <b>UNIX:</b> /opt/sas_catcon_servers_linux64
<b>Next</b> button	The labels for user interface controls are shown in a bold, sans-serif font.
<a href="http://www.sas.com">www.sas.com</a>	The hypertext links are shown in a light blue, fixed-width font, and are underlined.



---

# What's New in SAS Enterprise Content Categorization Servers 12.1

---

The new features in SAS Enterprise Content Categorization Servers include the following:

- The addition of SAS Content Categorization Collaborative Server and SAS Document Conversion Server to the SAS Content Categorization Server package.
- A single installer for the package.

## Documentation Changes for the 12.1 Release

.See the following table to understand the documentation for the 12.1 release:

Documentation	12.1 Product	Tasks and 5.2 Product References
<i>SAS Content Categorization Studio: Installation Guide</i>	SAS Content Categorization Studio	Install the single user or the enterprise version of SAS Content Categorization Studio that you purchased. The enterprise version automatically installs support for collaborative features and LITI concepts.
	SAS Enterprise Content Categorization Studio	
<i>SAS Content Categorization Studio: User's Guide</i>	SAS Content Categorization Studio	Create a SAS Content Categorization Studio project, test, and upload the project to SAS Enterprise Content Categorization Servers.  This guide is written for a single user and is a companion book for <i>SAS Enterprise Content Categorization Studio: User's Guide</i> .

Documentation	12.1 Product	Tasks and 5.2 Product References
<i>SAS Enterprise Content Categorization Studio: Administrator's Guide</i>	SAS Enterprise Content Categorization Studio with collaborative operations.	Configure your server for collaborative operations. (In the 5.2 release, this book was <i>SAS Content Categorization Collaborative Server: Administrator's Guide</i> .)
<i>SAS Enterprise Content Categorization Studio: User's Guide</i>	SAS Enterprise Content Categorization Studio with collaborative operations and LITI concepts capabilities.	See the cell above and use this guide to understand how collaborative operations work. Use the second part of this guide to write LITI rules and to upload these rules to SAS Enterprise Content Categorization Servers. (In the 5.2 release, LITI rules were explained in <i>SAS Contextual Extraction Studio: User's Guide</i> .)
<i>SAS Enterprise Content Categorization Servers: Administrator's Guide</i>	<p>Download any, or all, of the following:</p> <ul style="list-style-type: none"> <li>- SAS Content Categorization Server</li> <li>- SAS Enterprise Content Categorization Studio</li> <li>- SAS Content Categorization Java API</li> <li>- SAS Content Categorization Python API</li> <li>- SAS Document Conversion Server Java API</li> </ul>	<p>Install, configure, and use SAS Enterprise Content Categorization Servers, SAS Enterprise Content Categorization Studio, and SAS Document Conversion Server. You can also upload .li files using this product.</p> <p>In the 5.2 release, the information in this book was found in the following manuals:</p> <ul style="list-style-type: none"> <li>- <i>SAS Content Categorization Server: Administrator's Guide</i></li> <li>- <i>SAS Content Categorization Collaborative Server: Administrator's Guide</i></li> <li>- <i>SAS Document Conversion: Developer's Guide</i></li> </ul>

Documentation	12.1 Product	Tasks and 5.2 Product References
<p><i>SAS Content Categorization Single User Servers: Administrator's Guide</i></p>	<p>Download any, or all, of the following:</p> <ul style="list-style-type: none"> <li>- SAS Content Categorization Server</li> <li>- SAS Content Categorization Java API</li> <li>- SAS Content Categorization Python API</li> <li>- SAS Document Conversion Server and Java API</li> </ul>	<p>Install, configure, and use SAS Enterprise Content Categorization Servers and SAS Document Conversion Server</p> <p>In the 5.2 release, the information in this book was found in the following manuals:</p> <ul style="list-style-type: none"> <li>- <i>SAS Content Categorization Server: Administrator's Guide</i></li> <li>- <i>SAS Document Conversion: Developer's Guide.</i></li> </ul>



---

## Chapter: 1

# About SAS Enterprise Content Categorization Servers

---

- *What is SAS Enterprise Content Categorization Servers?*
- *Benefits of Using SAS Enterprise Content Categorization Servers*
- *How Does SAS Enterprise Content Categorization Servers Work?*
- *About the Architecture*

## 1.1 What is SAS Enterprise Content Categorization Servers?

In most organizations it is necessary to obtain information about, and from, data that is created internally and externally. SAS Enterprise Content Categorization Servers installs several applications to perform these tasks. You can choose to install any, or all, of the servers included in SAS Enterprise Content Categorization Servers:

- Use SAS Document Conversion Server to preprocess your documents, converting them from various file types into plain text.
- Use SAS Content Categorization Server to automatically apply categorization and concept extraction to input documents. Categorization and concept extraction are applied using the category rules and concept definitions that your organization develops using SAS Content Categorization Studio.
- Use the collaborative functionality in SAS Content Categorization Collaborative Server that enables multiple users to work together on one project in SAS Content Categorization Studio.
- Write scripts to preprocess documents and apply categorization and concept extraction to input documents using Java and Python APIs.

---

The installation and configuration documentation for each server is found in this book. Some applications also have their own documentation that can be found in the **Start** menu.

#### Easy document preprocessing

Use SAS Document Conversion Server to preprocess your files into plain text format.

#### Easy collaboration

Use SAS Content Categorization Collaborative Server with SAS Content Categorization Studio to enable multiple users to work on the same project.

#### Easy configuration

Most of the configuration file for SAS Content Categorization Server is written for you.

#### Easy monitoring

You can use the Administration Web Page to see whether SAS Content Categorization Server is running. You can also see data on concept and category matches, input documents, and timing.

## 1.2 Benefits of Using SAS Enterprise Content Categorization Servers

SAS Enterprise Content Categorization Servers provides users with the following benefits:

#### Convert documents into plain text format

SAS Document Conversion Server automatically converts input documents, of various types, into plain text format.

#### Automatically locate matching documents

SAS Content Categorization Server automatically applies the category rules and concept definitions developed in SAS Content Categorization Studio to input documents.

---

Gain real-time knowledge of matches

The Administration Web Page enables you to see the information about matching concepts and categories in real time.

Save money on information retrieval and organization costs

All of the information created by, or within, your organization can automatically be classified and retrieved. You can find information that is related, whether you know the exact terms that you are seeking.

Write custom programs

Developers use the included APIs to write applications to perform document conversion, to upload files, and so on.

## 1.3 How Does SAS Enterprise Content Categorization Servers Work?

SAS Enterprise Content Categorization Servers consists of several servers. Administrators and other users use SAS Document Conversion Server to automatically convert documents to plain text format. These users upload models and input documents to SAS Content Categorization Server where automatic document categorization and concept extraction is performed. SAS Content Categorization Server applies the rules and definitions in the form of binary files. These rules and definitions are written in SAS Content Categorization Studio.

## 1.4 About the Architecture

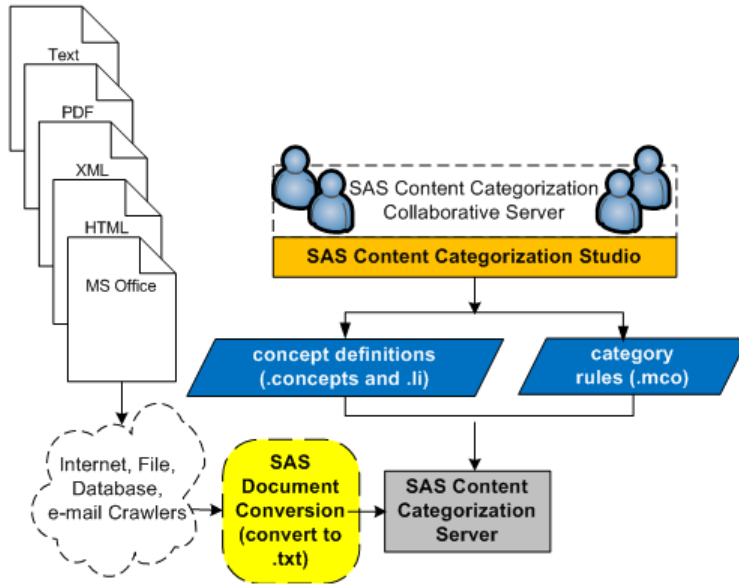
SAS Enterprise Content Categorization Servers provides rapid, run-time document conversion, categorization, and concept extraction for documents collected from your corporate intranet or the Internet. The SAS Document Conversion Server converts documents into plain text format that can be used by SAS Content Categorization Server. SAS Content Categorization Server automates the application of the compiled category rules and concept

---

definitions in the binary files that are uploaded from SAS Content Categorization Studio.

You can also use the Administration Web Page to see various types of reports on the matched categories and concepts.

*Figure 1-1 SAS Enterprise Content Categorization Servers Architecture*





---

# Part 1: Installing the Servers

---

- Chapter 2: *Using the Installation and Related Processes on page 13*



---

## Chapter: 2

# Using the Installation and Related Processes

---

- *Before You Install the Servers*
- *Installing SAS Enterprise Content Categorization Servers*
- *Starting and Stopping the Servers on a Windows Machine*
- *Uninstall on Windows*
- *Access the Servers*
- *Processing Documents*

## 2.1 Before You Install the Servers

### 2.1.1 Overview of Installation

The SAS Enterprise Content Categorization Servers installation kit for Windows (*SAS\_ConCat\_Collab\_Servers\_Setup.exe*) contains all of the components required to install (and uninstall) SAS Enterprise Content Categorization Servers. These servers are listed below:

- SAS Content Categorization Server
- SAS Content Categorization Collaborative Server
- SAS Document Conversion Server

---

**Note:** The necessary APIs are also included.

---

---

The installation is performed by a system administrator who is familiar with the operating system and who has sufficient system privileges to create directories and to define user permissions.

In some cases, you might want to install, or uninstall, one or more of the components of SAS Enterprise Content Categorization Servers. For example, you might want to remove a local copy of SAS Content Categorization Server from your machine because you are connecting to the main server.

To understand and use this example, complete these steps:

1. Install SAS Content Categorization Server and the SAS Content Categorization Java API when you install SAS Enterprise Content Categorization Servers.
2. Use SAS Content Categorization Server to verify that the Java API code is written correctly.
3. Use the Java code to connect to the main SAS Content Categorization Server.
4. Uninstall the local copy of SAS Content Categorization Server taking care not to uninstall SAS Content Categorization Java API. (For more information, see Section 2.4 *Uninstall on Windows* on page 30.)

## 2.1.2 Prerequisites

Configure the machine where you install SAS Enterprise Content Categorization Servers according to the recommended system configuration:

### **CPU**

x86 with 1 GHz or higher required. 2+ CPUs of 2 GHz or higher, each, are recommended

### **RAM**

1 GB or higher is recommended, but this base number depends on the number of binary files that you load

---

The table below lists the hardware requirements that are necessary to run SAS Enterprise Content Categorization Servers:

Table 2-1: Supported Operating Systems

Operating System	Platform
Linux, (Red Hat 7.x, 8, 9, Fedora 1-3, RHEL 2.1 and higher), SUSE	x86, x86-64
IBM AIX	PPC
HP-UX	PA-RISC
Sun Solaris (32-bit)	SPARC
Sun Solaris (64-bit)	UltraSPARC, x86-64
Windows	x86, x86-64

---

**Note:** Ensure that Java run-time environment 1.5 or later is installed on your machine or you cannot install SAS Document Conversion Server and Java API.

---

### 2.1.3 Using the SAS License File

This SAS license is the SAS installation data file (SID file) that is included in the Software Order E-mail (SOE) that you received. When you locate this file during installation, the folder location and path are specified in the `server.config` file located in the `Teragram CatCon Server/conf` folder. For more information, see Section 4.2 *SAS Content Categorization Server Configuration File* on page 45. If the path to this file does not appear in your server configuration file, see Section A.3.2 *Modifying the Configuration File* on page 113.

---

## 2.2 Installing SAS Enterprise Content Categorization Servers

### 2.2.1 Install on Windows

#### 2.2.1.A Before You Install on Windows

When you work through the SAS Enterprise Content Categorization Servers Setup wizard, there are some steps that pertain only to SAS Content Categorization Server. These pages specify SAS Content Categorization Server in the directions provided.

Install Java 1.5 or higher and Python version 2.3 or higher if you plan to install the Java and Python APIs for SAS Content Categorization Server. If you plan to install SAS Document Conversion Server and Java API, make sure that the Java 1.5, or higher, run-time environment is installed on the machine.

---

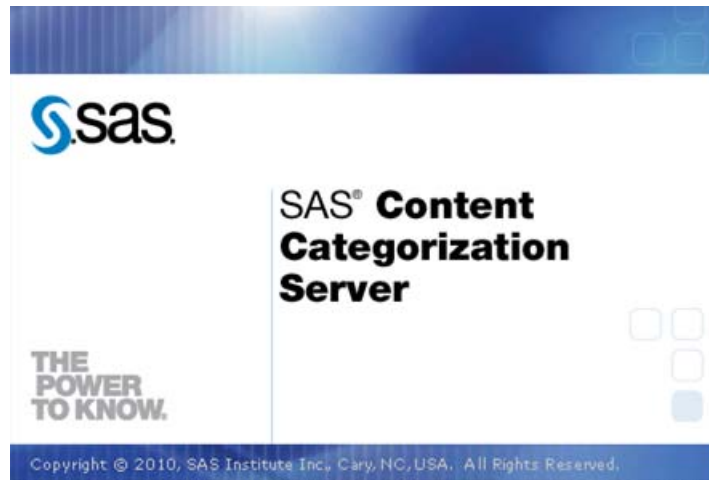
**Notes:** SAS Document Conversion Server and Java API cannot be installed unless the correct Java run-time environment is installed.

---

#### 2.2.1.B Install on Windows

To install the SAS Enterprise Content Categorization Servers software on a supported Microsoft Windows system, complete these steps:

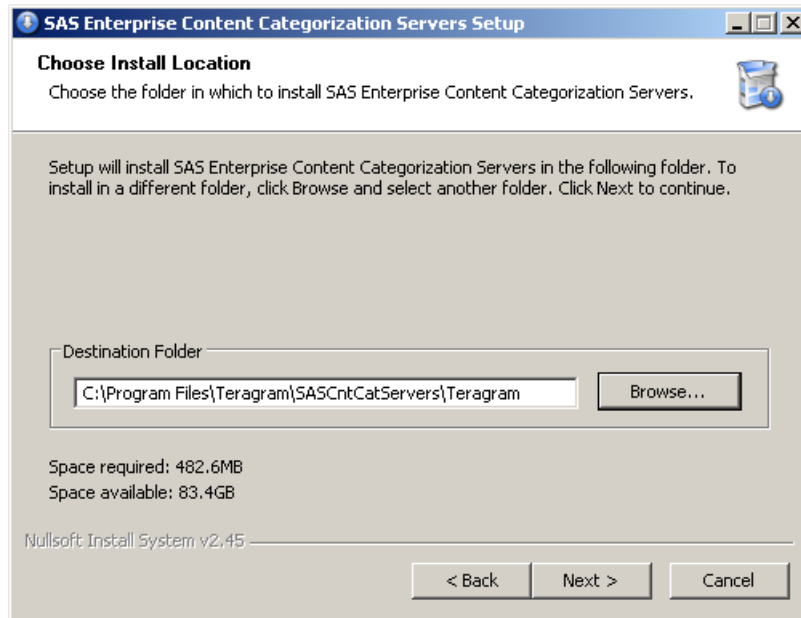
1. Double-click `SAS_ConCat_Server_Setup.exe` and the installation wizard appears.



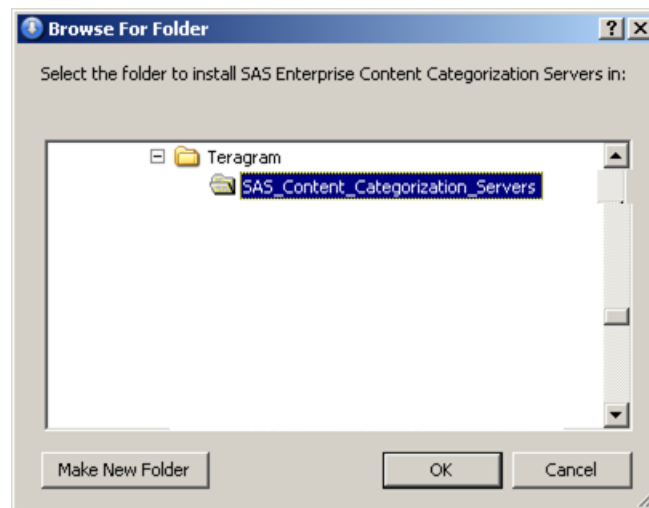
The Welcome page appears.



2. Click **Next** and the Choose Install Location page appears where you can enter the path to the installation folder.

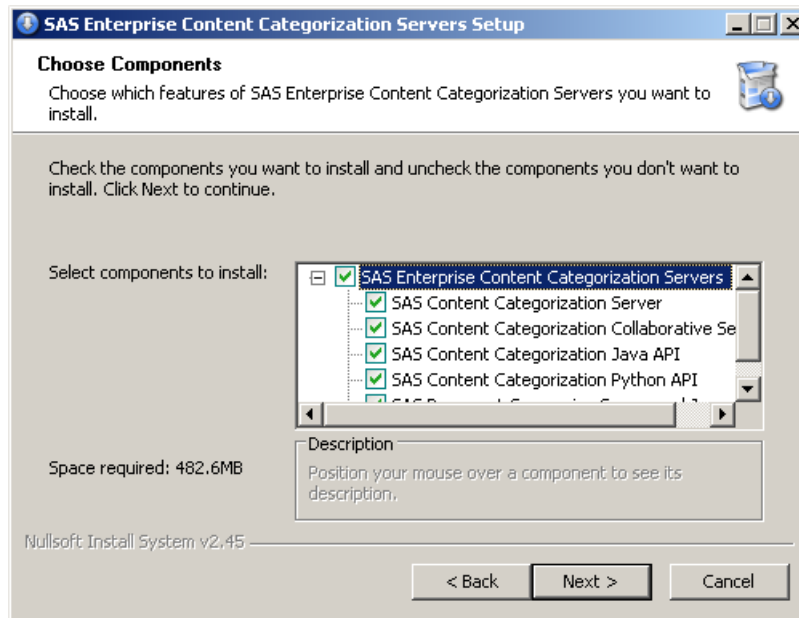



3. (Optional) Click **Browse** and the Browse For Folder window appears.



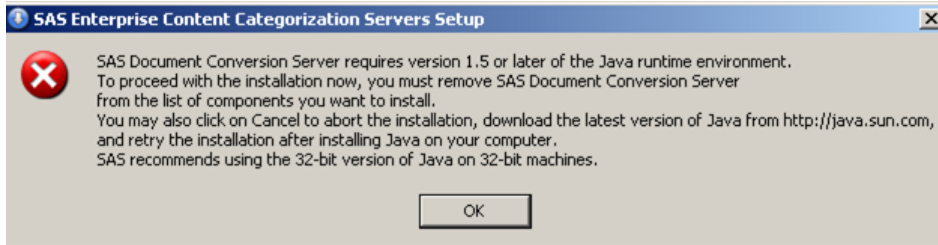


- a. (Optional) Select a different installation folder.
  - b. (Optional) Click **Make New Folder**.
  - c. Click **OK**.
4. Compare **Space required** with **Space available** in the Choose Install Location page to ensure that there is enough room on your hard drive for the application.
  5. Click **Next** and the Choose Components page appears.



6. (Optional) Click  to see all of the applications that you are installing and to deselect some, but not all, of these components:
  - SAS Content Categorization Server
  - SAS Content Categorization Collaborative Server
  - SAS Content Categorization Java API
  - SAS Content Categorization Python API

- 
- SAS Document Conversion Server and Java API
7. Click **Next** and one of the following SAS Content Categorization Servers Setup windows might appear. If none of these windows appears, see Step 8. on page 21.
- If you selected SAS Content Categorization Java API and you do not have version 1.5 or later loaded, the SAS Enterprise Content Categorization Servers Setup window appears. Note the URL and load Java after the installation of SAS Enterprise Content Categorization Servers is complete:



- If you selected SAS Content Categorization Python API and you do not have Python version 2.3 or higher loaded, the SAS Enterprise Content Categorization Servers Setup window appears. Note the URL and load the recommended Python version and type before you install SAS Content Categorization Python API.

---

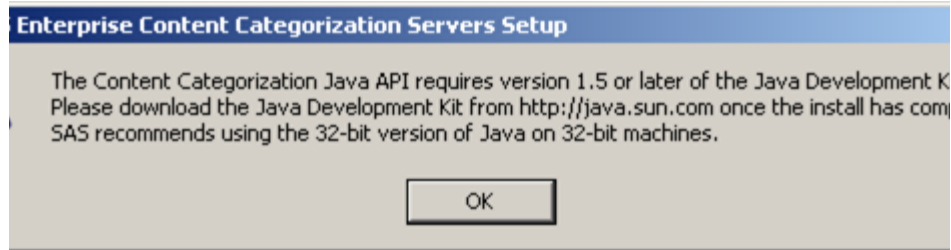
**Hint:** You can deselect this operation in order to continue the Install operation and install SAS Content Categorization Python API later.

---

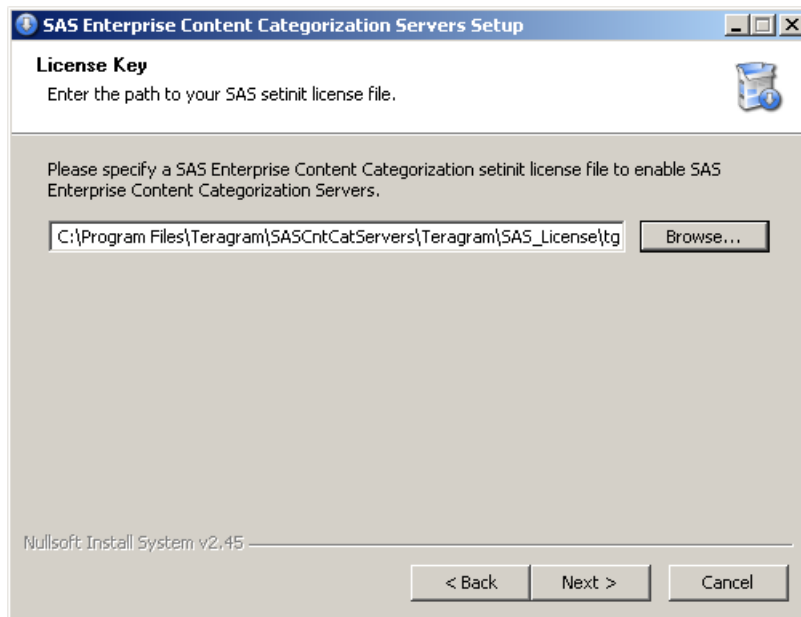


- (If you select SAS Document Conversion Server and you do not have version 1.5 of the Java run-time environment loaded, deselect

this component.) If, instead you select this component, the SAS Enterprise Content Categorization Servers Setup window appears. (If you install the 64-bit version, this window specifies 64-bit.) Note the URL and load Java before you try to install SAS Document Conversion Server:



8. Click **OK** in any of the SAS Enterprise Content Categorization Servers Setup windows that appear. See the License Key page that appears.



- 
9. Enter the path to the license key or click **Browse**. If you click **Browse**, see Step 3. on page 18. Specify this path on a local drive, not a network drive.

---

**Note:** This path should be on a local drive, not a network drive.

---

10. Click **Next** and the SAS Content Categorization Server Ports page appears.

SAS Content Categorization Servers Setup

**SAS Content Categorization Server Ports**

Enter SAS Content Categorization Server port numbers.

Please specify the port numbers that SAS Content Categorization Server will use. The server listens for content categorization requests on the query port, and displays a basic web page with administrative information on the admin port.

Enter the query port number:

Enter the admin port number:

Nullsoft Install System v2.45

< Back   Next >   Cancel

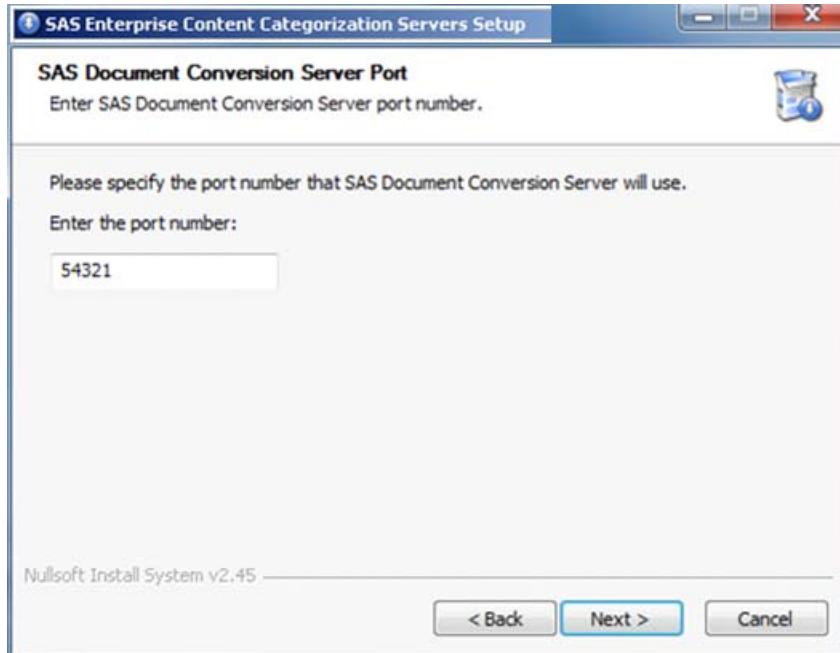
---

**Note:** This page is only for SAS Content Categorization Server.

---

11. (Optional) Use the **Enter the query port number** field to change the port number that appears by default.

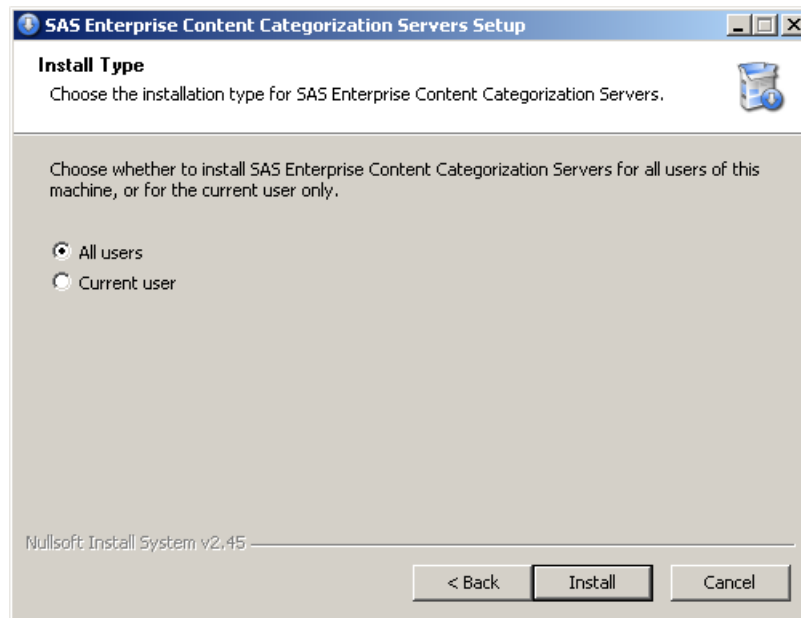
- 
12. (Optional) Use the **Enter the admin port number** field to change the port number that appears by default.
  13. Click **Next** and the SAS Document Conversion Server Port page appears.



14. (Optional) Use the **Enter the port number** field to change the port number that appears by default.

---

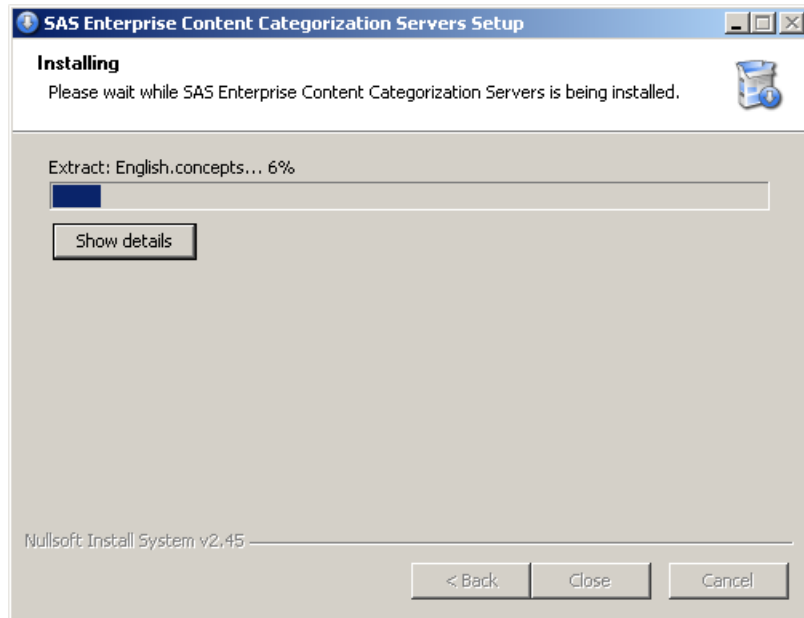
15. Click **Next**. The Install Type page appears:



16. Install the servers for either:

- **All users:** This operation enables all of the users on this machine to access the servers.
- **Current user:** This operation enables only the person who performs the Install operation to access the servers.

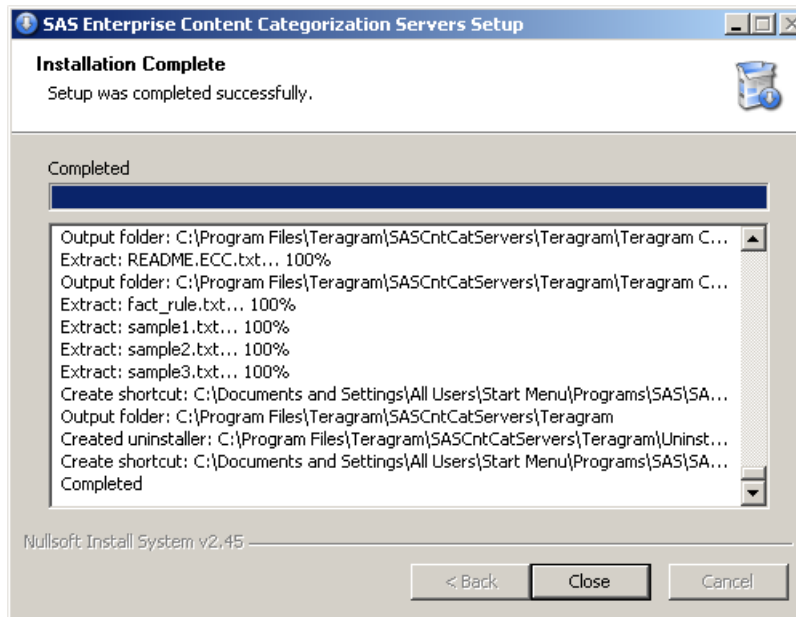
- 
17. Click **Install** and the Installing page appears where you can see the installation progress.



18. (Optional) Click **Show details** and see Step 5. on page 19 for an example of the pane that appears.

---

The Installation Complete page appears.



**19. Click Close.**

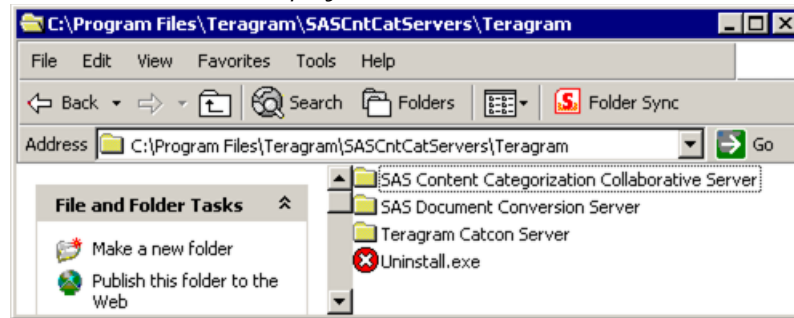


---

## 2.2.1.C The Folders That Appear after Installation

See the following example of the folders that might appear after installation:

*Display 2-1 Installed Folders*



SAS Content Categorization Collaborative Server

find the installer for SAS Content Categorization Studio

SAS Document Conversion Server

find the empty `client api` folder and a `README` file.

Teragram Catcon Server

find a number of files and folders, including the SAS Content Categorization Server configuration file in the `conf` directory.

`uninstall.exe`

uninstall SAS Content Categorization Servers.

## 2.2.2 Installing on UNIX

### 2.2.2.A Install on UNIX

SAS Enterprise Content Categorization Servers is distributed on UNIX systems as a `tar.gz` file. Install using the following UNIX command:

```
SAS_Enterprise_Content_Categorization_Server--<arch>--  
<date>.tar.gz
```

<arch>

is a string representing the architecture. For example, see `linux64`.

---

<date>

is the release date of the package.

To install the software, use the following UNIX commands:

```
SAS_Enterprise_Content_Categorization_Servers--<arch>--  
    <date>.tar.gz
```

```
SAS_Enterprise_Content_Categorization_Servers--<arch>--  
    <date>.tar
```

The switches on the `tar` command are used to extract the contents from the specified `tar` file and to preserve the file and directory permissions of the contents.

---

**Note:** The actual name of your `tar` file might vary from that shown in the example above.

---

Additional information about using the `gzip` and `tar` commands is available in the UNIX `man` pages.

### 2.2.2.B See the Directory Structure

When these files are extracted, the following directory structure is created. See the notes on what is in each directory in *italic font*.

```
+ sas_cc_collab_servers_<arch>  
|-- cc_server  
+  
|-- bin  
+  
|-- <arch> (SAS Content Categorization Server  
           executable)  
|-- client_api (README file for SAS Content  
               Categorization client APIs)  
+  
|-- doc (Web services API documentation)  
|-- java (Java client API)  
+  
|-- doc (Java client API technical documentation)  
|-- python (Python client API and Web services API)  
+  
|-- doc (Python client API technical documentation)
```

---

```

|-- test (Sample documents for testing the client
        APIs)
|-- conf (server.config file)
|-- descriptors (Descriptor files to load the included
        sample .mco, .concepts, and .li files)
|-- doc (SAS Enterprise Content Categorization Servers:
        Administrator's Guide)
|-- models (Sample .mco, .concepts, and .li files)
|-- collab_server
+
|-- bin
+
|-- <arch> (_tk240 executable)
|-- data (linguistic data files required to build SAS
        Content Categorization Studio projects in batch
        mode with _tk240)
|-- doc (SAS Content Categorization Collaborative
        Server: User's Guide)
|-- Shared Projects (placeholder directory for locally
        cached copies of SAS Content Categorization
        Collaborative Server projects)
|-- doc_conversion_server (Java JAR files for SAS
        Document Conversion Server and client)
+
|-- doc (README file for SAS Document Conversion Server
        and client API)
+
|-- javadoc (Java client API technical documentation)
|-- open-source (obligatory legal disclaimers for use
        of the open source library that SAS Document
        Conversion Server 12.1 )uses)

```

## 2.3 Starting and Stopping the Servers on a Windows Machine

After you install SAS Enterprise Content Categorization Servers, SAS Content Categorization Server automatically starts. The following steps are for SAS Content Categorization Server, but you can modify these steps for SAS Document Conversion Server.

To stop SAS Content Categorization Server, complete the following steps:

- 
1. Go to **Start --> Settings --> Control Panel --> Administrative Tools --> Services**.
  2. Select the SAS Content Categorization Server.
  3. Click **Stop** in the Services pane that appears.

To restart SAS Content Categorization Server, complete the following steps:

1. Complete Step 1. and Step 2. above.
2. Click **Start** in the Services pane that appears.

---

**Notes:** You can access only the SAS Content Categorization Server Administration Web Page from the **Start** menu when the server is running. For more information, see Chapter 5: *SAS Content Categorization Server Web Administration*.

---

## 2.4 Uninstall on Windows

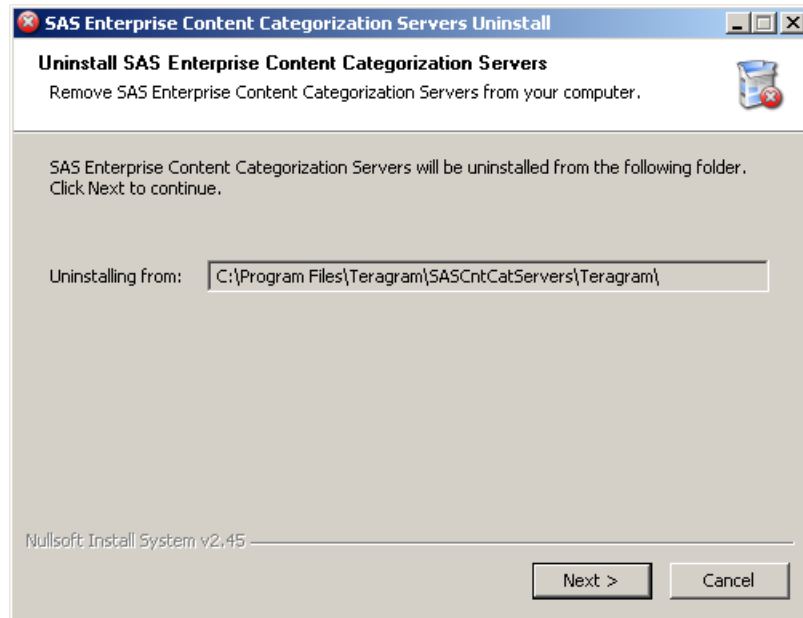
Administrators, or users with administrative permissions, are the only users that can uninstall SAS Enterprise Content Categorization Servers.

To uninstall SAS Enterprise Content Categorization Servers, complete these steps:

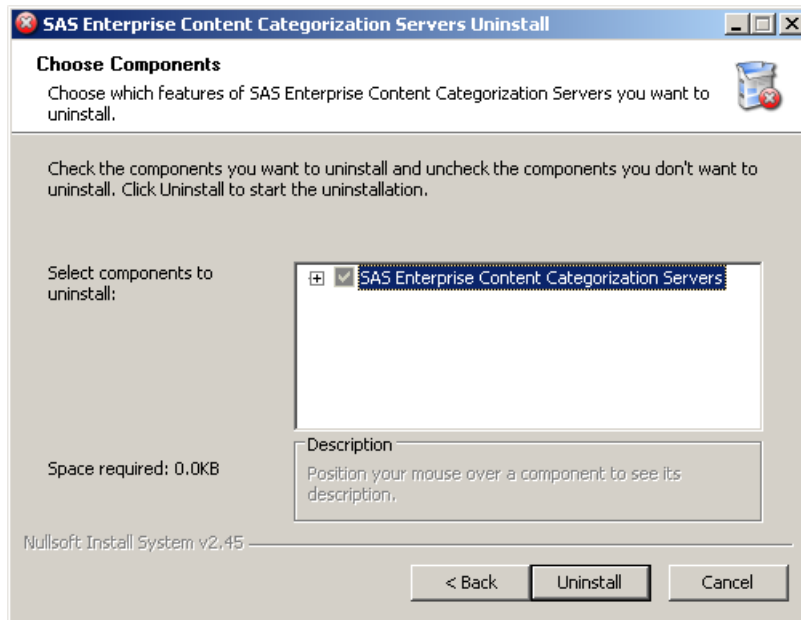
1. Go to **Start --> Programs --> SAS --> SAS Enterprise Content Categorization Servers --> Uninstall SAS Enterprise Content**

---

**Categorization Servers.** The Uninstall SAS Enterprise Content Categorization Servers page appears.

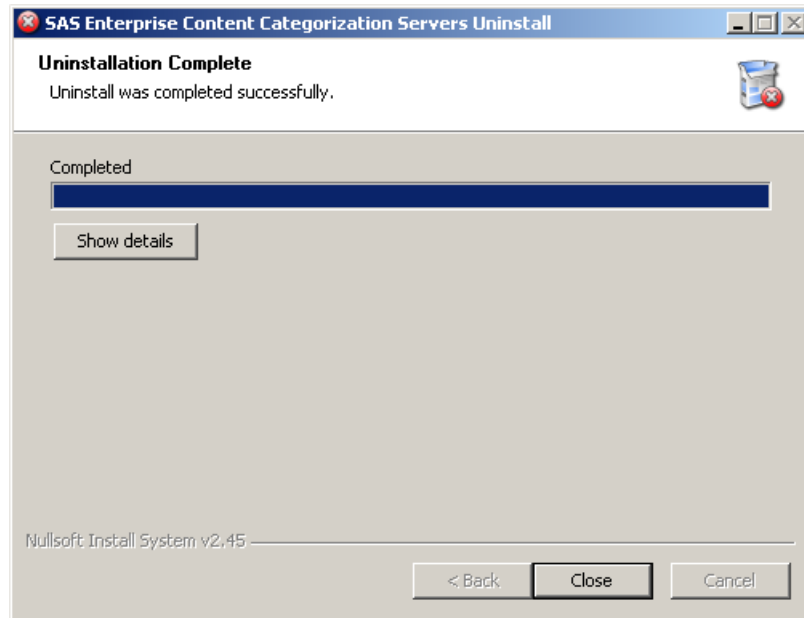


- 
2. Click **Next**. The Choose Components page appears:



3. (Optional) By default, all of the components are checked. Uncheck the components that you do not want to uninstall.

- 
4. Click **Uninstall**. The Uninstall page appears after the server stops running.

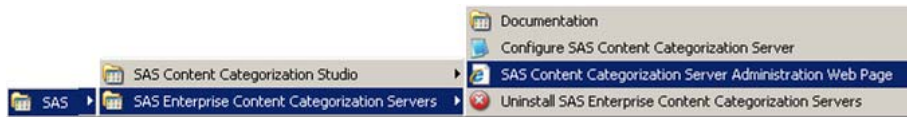


5. (Optional) Click **Show Details**. See Step 17. on page 25 for an example of the pane that appears.
6. Click **Close**.

## 2.5 Access the Servers

To use the Start menu to locate the SAS Enterprise Content Categorization Servers program, complete these steps:

- 
1. Go to **Start --> Programs --> SAS --> SAS Enterprise Content Categorization Servers**:



2. Select any of the following:
  - **Documentation:** See the documentation that is available for SAS Enterprise Content Categorization Servers.
  - **Configure SAS Content Categorization Server:** Access the configuration file, if you selected this component.
  - **SAS Content Categorization Server Administration Web Page:** Check SAS Content Categorization Server operations using this Web page.
  - **Uninstall SAS Enterprise Content Categorization Servers:** Remove the selected components from your machine.

---

**Hint:** If you did not install SAS Content Categorization Server, only the **Documentation** and **Uninstall SAS Enterprise Content Categorization Servers** selections are available.

---

## 2.6 Processing Documents

You can upload a project that uses the new synonym list feature in SAS Content Categorization Studio. In this case, make sure that the documents that you send to SAS Enterprise Content Categorization Servers do not exceed 1 MB in size.



---

**Note:** You cannot upload a synonym list to SAS Content Categorization Collaborative Server.

---



---

## Part 2: SAS Document Conversion Server

---

- Chapter 3: *Processing Various File Types into Plain Text Format on page 39*



---

## Chapter: 3

# Processing Various File Types into Plain Text Format

---

- *Overview of SAS Document Conversion Server*
- *The File Types That SAS Document Conversion Server Can Convert*

### 3.1 Overview of SAS Document Conversion Server

Use the SAS Document Conversion Server to preprocess different file types into plain text format. For an overview of why this process is often necessary, see Section 1.3 *How Does SAS Enterprise Content Categorization Servers Work?* on page 9. After you convert your files into plain text, this text can be sent to SAS Content Categorization Server. Use the `.mco`, `.concepts`, or `.li` files uploaded to SAS Content Categorization Server to apply categories and to locate concepts in the input text. For more information about this process, see Figure 1-1 on page 10.

---

## 3.2 The File Types That SAS Document Conversion Server Can Convert

SAS Document Conversion Server processes the document formats that are supported by *Apache Tika 1.0*, which is an open-source helper library. See the following table for a list of these file types and their description:

Table 3-1: Supported Document Types

Abbreviation	Format Type	Description
HTML	HyperText Markup Language	Web mark-up language that includes HTML and valid XHTML
XML and XML derived formats	Extensible Markup Language	The custom parsers for some widely used XML vocabularies like XHTML, OOXML, and ODF.
Microsoft Office	OLE 2 Compound Document and Office Open XML (OOXML) formats	The older OLE 2 format was introduced in Microsoft Office version 97 and was the default format until version 2007 and the XML-based OOXML format.
ODF	OpenDocument Format	The default format of the OpenOffice.org office suite and other applications.
PDF	Portable Document Format	A common document format used by Adobe Reader and Acrobat.
EPUB	Electronic Publication Format	The format used for many digital books.
RTF	rich text format	An older document interchange format.
	compression and packaging formats	The various compression and packaging formats such as .zip.
	text formats	The character encoding detection and normalization of plain text docs.
	mbox format	The e-mail messages from the mbox format that is used by many e-mail archives and Unix-style mailboxes.

---

## Part 3: SAS Content Categorization Server

---

- Chapter 4: *Configuring and Running SAS Content Categorization Server on page 43*
- Chapter 5: *SAS Content Categorization Server Web Administration on page 63*





---

## Chapter: 4

# Configuring and Running SAS Content Categorization Server

---

- *Configuration Overview*
- *SAS Content Categorization Server Configuration File*
- *Specifying Project Users and Creators*
- *Add a Project*
- *Using the Models Directory*
- *Using the Descriptors Directory*
- *Using cat\_log and concept\_log Files*
- *Specifying Multiple Project Files*
- *Sending Documents to the Server*
- *Run SAS Content Categorization Server*
- *Optimize Performance on a Client Windows Machine*

## 4.1 Configuration Overview

Configure SAS Content Categorization Server by specifying directives in the configuration file (`server.config`). These directives enable you to specify creators who are administrators who can upload projects and other users for this application. You also choose the types of connections for the server, directories, and other settings. The configuration file is automatically created when you install SAS Content Categorization Server and remains after the uninstall operation.

---

**Note:** If you are installing SAS Content Categorization Server for the first time, the configuration file is the default

---

---

file. For more information, see Section 4.2 *SAS Content Categorization Server Configuration File* on page 45. If you have previously installed SAS Content Categorization Server, the original config file is preserved. You can edit this file. For more information, see Appendix A *Troubleshooting*.

---

The SAS Content Categorization Server configuration file is a text file that contains key - value pair assignments. See the following form, where each pair appears on a single line. Any blank lines, as well as any comment lines that are preceded by the comment character (#), are ignored.

```
key=value
```

The `descriptor_dir` directive specifies the directory that contains descriptor files. For example, see the `descriptors` directory. The descriptor files determine the binary models that are loaded and the symbolic names for these models. When models are uploaded to SAS Content Categorization Studio, these files are automatically created.

See the following descriptor file examples:

*Example 4-1: IPTC.desc*

```
type=mcats
path=models\English.mco
name=IPTC
```

*Example 4-2: Entities.desc*

```
type=concepts
path=models\English.concepts
name=Entities
```

*Example 4-3: LITIDemo.desc*

```
type=liti
path=models\LITIDemo.li
name=LITIDemo
```

---

**Note:** The UNIX path contains a forward slash (/) instead of a backslash.

---

---

## 4.2 SAS Content Categorization Server Configuration File

### 4.2.1 Windows Configuration File

During installation, the configuration file is automatically created in the installation directory where you choose to install the program:

```
<INSTALL_DIR>\Teragram Catcon Server\conf
```

To see this file, select **Start --> Programs --> SAS --> SAS Enterprise Content Categorization Servers --> Configure SAS Content Categorization Server**. The configuration file that appears is similar to the example shown below:

*Example 4-4: Sample Configuration File for Windows*

```
basedir=/opt/sas/sas_cc_collab_servers/cc_server  
backupdir=backup  
setinit=/opt/sas/sas_cc_collab_servers/setinit.txt  
descriptor_dir=descriptors  
create_dir=server_data  
query_port=6500  
admin_port=6501  
skt_queue_size=10  
nb_threads=4  
persistent_connection=0  
timeout=60000000  
max_iterations_to_reinitialize=5000
```

### 4.2.2 UNIX Configuration File

On a UNIX system, the configuration file is located in the `conf` subdirectory. For example:

```
/opt/sas_catcon_server_linux64/conf/server.config
```

---

### 4.2.3 The Directives

Use the directives to modify the SAS Content Categorization Server configuration file. Directives such as `descriptor_dir` are treated as absolute paths. In other words, these paths are not relative to `basedir`. This is true if `basedir` is specified and these paths begin with `[A-Z, a-z]:\` on Windows or on UNIX. See the following example:

```
backupdir=c:\backups
```

The directives for the SAS Content Categorization Server configuration file are described in the table below:

Table 4-1: Configurable Directives

Directive	Description
<code>basedir</code>	Specifies the path to the project binaries, backup directory, and so on.
<code>backupdir</code>	Specifies the directory where the backup binaries are stored. When a categorization ( <code>.mco</code> ), concepts ( <code>.concepts</code> ), or contextual extraction ( <code>.li</code> ) binary file is opened by SAS Content Categorization Server at start-up and this directive is specified, a backup copy of the binary file is created. The backup copy is written to the specified directory. If a binary file cannot be subsequently opened, SAS Content Categorization Server attempts to use the backup version of the binary file. This directive enables SAS Content Categorization Server to keep running even if a binary file cannot be loaded.
<code>setinit</code>	Specifies the path to the SAS license file for SAS Content Categorization Server. This SAS installation data file (SID) is in the Software Order E-mail (SOE) that you received. For more information, see Section 2.1.3 <i>Using the SAS License File</i> on page 15.
<code>descriptor_dir</code>	Contains descriptor files that contain information about the projects that are loaded on SAS Content Categorization Server.  <b>Hint:</b> After you uninstall SAS Content Categorization Server, this directory and its files remain. This statement is true if you have uploaded one or more models to the server.

Table 4-1: Configurable Directives (Continued)

Directive	Description
create_dir	<p>Tells SAS Content Categorization Server where to store the binary files for the projects that are added to this server. Without this directive, the creator cannot upload new files.</p> <p><b>Hint:</b> After you uninstall SAS Content Categorization Server, this directory and its files remain. This statement is true if you have uploaded one or more models to the server.</p>
query_port	<p>Specifies the number of the TCP port where the categorization and concept extraction services are available. The clients connect to this port on the server host. This port number is used in the <b>Port</b> field of the Upload concepts to SAS Content Categorization Studio window. For more information, see Section 4.4 <i>Add a Project</i> on page 51.</p>
admin_port	<p>Specifies the number that corresponds to the TCP port where the server's Web-based administrative interface is available.</p>
skt_queue_size	<p>Specifies the number of simultaneous pending client connections that the server accepts, without dropping the connection. If all of the server threads are busy serving clients, this attribute specifies the maximum number of additional clients. This number of clients can connect to the server and wait for a thread to become available.</p>
nb_threads	<p>Specifies the number of parallel service threads to run. SAS Content Categorization Server is able to handle the specified number of clients, simultaneously.</p>
persistent_connection	<p>Specifies whether the server tries to maintain a continuous socket connection with the client, or not. The default value is zero (0). If this setting is set to one (1), persistent connections are enabled if the client also enables these connections.</p>
timeout	<p>Specifies the length of time (in microseconds) that the server waits. If no activity occurs during this period, the server forcibly drops the connection.</p>
max_iterations_to_reinitialize	<p>Tells the server to clear out its memory after the specified number of documents is reached.</p>
max_doc_size	<p>Specifies the largest size (in bytes) of documents that can be processed. Larger texts are truncated.</p>

Table 4-1: Configurable Directives (Continued)

Directive	Description
xml_weight_file	<p>Specifies the weights for structured-text fields that match MCAT rules. When rule terms match within these fields, the relevancy score for these terms is multiplied by the field weight. The syntax for the <code>xml_weight_file</code> is <code>field: weight</code> for each line in the file. You could specify the following:</p> <pre>title:3 body:1.5</pre> <p>In this example, if a match is located in the <code>body</code> field of an XML document, the match counts 1.5 times toward the relevancy score. However, a match for the <code>title</code> field is multiplied by three.</p>
user	<p>Specifies a user name and password. Users specify these entries to upload the binary files directly from SAS Content Categorization Studio to SAS Content Categorization Server. These users can upload only the new versions of projects that already exist on the server. For more information, see Section 4.3 <i>Specifying Project Users and Creators</i> on page 50.</p> <p><b>Tip:</b> This directive can be specified multiple times.</p>
creator	<p>Specifies a user name and password for the users who have the privileges necessary to upload new projects and refresh existing projects. For more information, see Section 4.3 <i>Specifying Project Users and Creators</i> on page 50.</p> <p><b>Tip:</b> This directive can be specified multiple times.</p>
io_log	<p>Generates a detailed input and output log file while performing the categorization operation. This log includes timestamps. The value is the name of the file where the logging information is written.</p> <p><b>Tip:</b> This directive only applies to categories.</p>
cat_log	<p>Generates a log of all of the category matches that are returned for the documents sent to the server. One entry is specified for each document that matches one or more categories in a category project. The value is the base for the category log file. For more information, see Section 4.7 <i>Using cat_log and concept_log Files</i> on page 57.</p>

Table 4-1: Configurable Directives (Continued)

Directive	Description
cat_log_max_entries	Specifies the maximum number of entries allowed in each cat_log file. For more information, see Section 4.7 <i>Using cat_log and concept_log Files</i> on page 57.
num_cat_logs	Specifies the maximum number of category log files to create. For more information, see Section 4.7 <i>Using cat_log and concept_log Files</i> on page 57.
do_cat_log_timing	Provides additional timing information in all of the category log files. No value is required.
concept_log	Generates a log of all of the concept matches that are returned for the documents sent to SAS Content Categorization Server. One entry is defined for each document that matches one or more concepts in a concepts project. The value is the base for the filename. For more information, see Section 4.7 <i>Using cat_log and concept_log Files</i> on page 57.
concept_log_max_entries	Specifies the maximum number of entries allowed in each concept_log file. For more information, see Section 4.7 <i>Using cat_log and concept_log Files</i> on page 57.
num_concept_logs	Specifies the maximum number of concept log files to create. For more information, see Section 4.7 <i>Using cat_log and concept_log Files</i> on page 57.
do_concept_log_timing	Obtains additional timing information in the concept log file. It is not necessary to specify any value for this attribute.
protocol_version	Enables SAS Content Categorization Server to emulate older versions of the client/server protocol that this application uses.

**Note:** After modifying and saving a configuration file, restart SAS Content Categorization Server in order to make the changes take effect. For more information, see Section 2.3 *Starting and Stopping the Servers on a Windows Machine* on page 29.  
If you have previously installed SAS Content Categorization Server, the original configuration file is

---

preserved. For more information about how to edit this file, see *Appendix A Troubleshooting*.

---

The directives in the following table continue to be supported for backwards compatibility purposes. Use the `descriptor_dir` directive when possible.

Table 4-2: Deprecated Directives

Directive	Description
<code>mcat</code>	Specifies either a relative, or an absolute, path to a categorization binary file ( <code>.mco</code> ). The symbolic name for the categories project is specified after a colon ( <code>:</code> ).
<code>concepts</code>	Specifies either a relative or an absolute path to a concepts binary file ( <code>.concepts</code> ). The symbolic name for the concepts project is specified after a colon ( <code>:</code> ).
<code>liti</code>	Specifies either a relative or an absolute path to a concept extraction binary file ( <code>.li</code> ). The symbolic name for the LITI project is specified after a colon ( <code>:</code> ).
<code>stat_cat</code>	Specifies either a relative or an absolute path to a statistical categorizer binary file ( <code>.st.cat</code> ). The symbolic name is specified after the colon ( <code>:</code> ).
<b>Tips:</b> These directives can be specified multiple times. If <code>basedir</code> is specified, these paths are relative to this directory. If this directory is not specified, these paths are absolute.	

## 4.3 Specifying Project Users and Creators

In order to upload a project file from SAS Content Categorization Studio to SAS Content Categorization Server, specify a `creator` in the server configuration file. You can also add users who have permissions to refresh existing project files to the server.

User

refresh a binary file for a project that is already uploaded to SAS Content Categorization Server. The user can reload these project files from SAS Content Categorization Studio to SAS Content Categorization Server.



---

## Creator

is an administrator who can upload a binary file for a new project to SAS Content Categorization Server. This class of administrators creates new projects.

Using either directive, specify the name and password for each user and creator. These specifications apply to any of the Upload operations in SAS Content Categorization Studio.

An example of the directives for the `user` and `creator` that you add to the configuration file are shown below:

*Example 4-5: User and Creator Directives*

```
user=user1:pw1
user=user2:pw2
creator=creator1:pw3
creator=creator2:pw4
```

All duplicate usernames are ignored after the first instance. See the following example:

```
creator=Joe:Joespassword
user=Joe:Joespassword
```

In this example, Joe has creator permissions.

## 4.4 Add a Project

Administrators who have creator privileges can upload projects built in SAS Content Categorization Studio to SAS Content Categorization Server. When a creator uploads a project, files for the project is automatically added to the data and descriptor directories. These folders are referenced by the configuration file for SAS Content Categorization Server. For this reason, it is not necessary to add new projects directly to the configuration file.

To add a `.concepts` project developed in SAS Content Categorization Studio to SAS Content Categorization Server, complete these steps. Make appropriate changes for other binary file types.

- 
1. In SAS Content Categorization Studio, select **Build --> Upload Concepts**. The Upload Concepts to SAS Content Categorization Studio window appears:



The screenshot shows a dialog box titled "Upload Concepts to SAS Content Categorization Server". It contains the following fields and values:

- Server Host Name: localhost
- Port: 6500
- Username: creator1
- Password: (masked with three dots)
- Server Project Name: MyConcepts

Buttons: OK, Cancel

By default, the **Server Host Name** and **Port** fields are automatically entered.

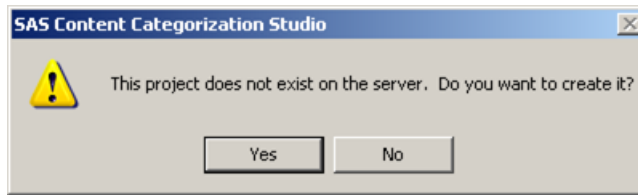
---

**Note:** The value for the **Port** field should be the value specified in the `query_port` directive of the `server.config` file, not the `admin_port` field. See Example 4-4 on page 45.

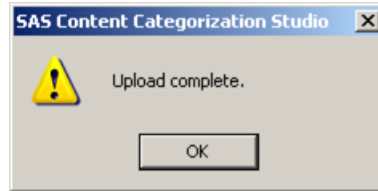
---

2. Enter the creator name that is specified in the configuration file into the **Username** field. For example, type `creator1`.
3. Enter your creator password that is specified in the configuration file into the **Password** field. For example, type `pw3`.
4. Enter the symbolic name for the file that you choose to upload into the **Server Project Name** field. For example, type `MyConcepts`.

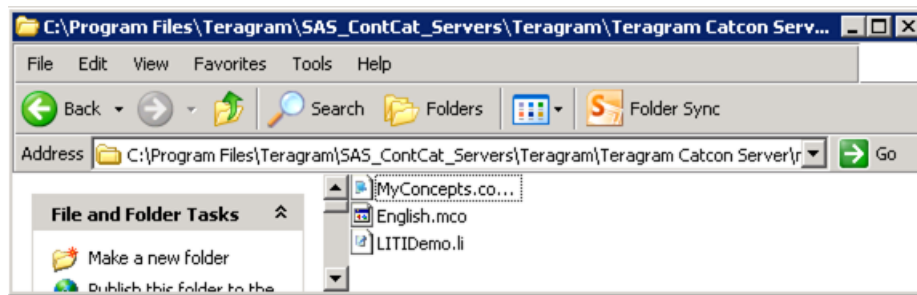
- 
5. Click **OK**. A SAS Content Categorization Studio status window appears.



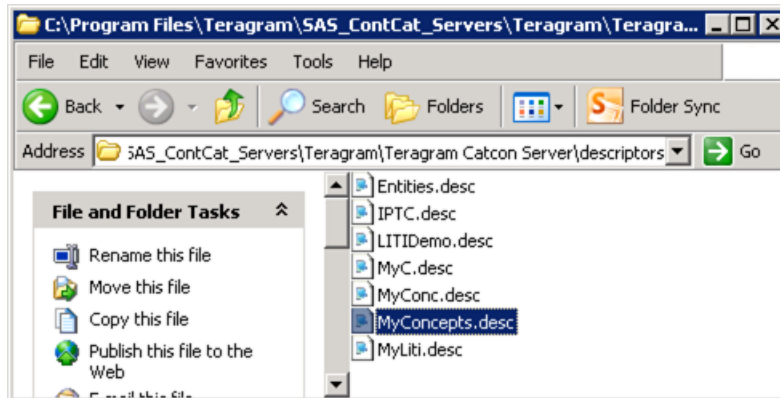
6. Click **Yes**. A second SAS Content Categorization Studio status window appears.



7. Click **OK** and the project is listed in the `models` folder. For more information, see Section 4.5 *Using the Models Directory* on page 54.



- 
8. The project is also listed in the descriptors folder as a <project name> .desc file. For example, see `MyConcepts.desc`:

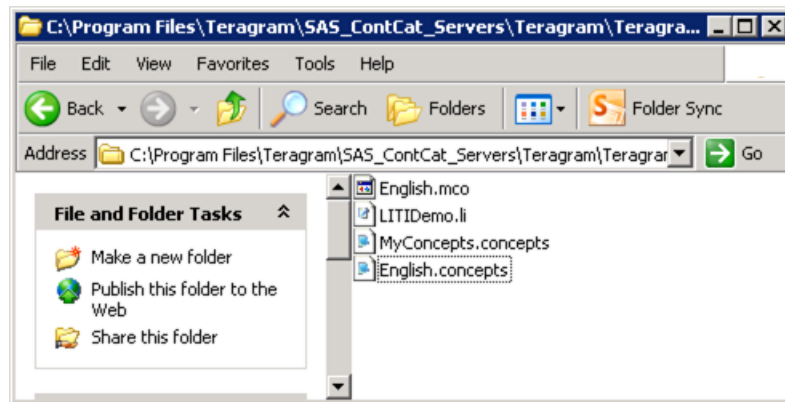


## 4.5 Using the Models Directory

The `models` directory displays each of the projects that are uploaded to SAS Content Categorization Server. These projects include any sample projects that are shipped with the application. If you uninstall and then re-install your application, the `.mco`, `.li`, and `.concepts` files that you uploaded remain in the `models` folder. For example, the names of the project files might be `English.concepts`, `LITIDemo.li`, and `English.mco`. See the following example:

---

Display 4-1 The Models Directory

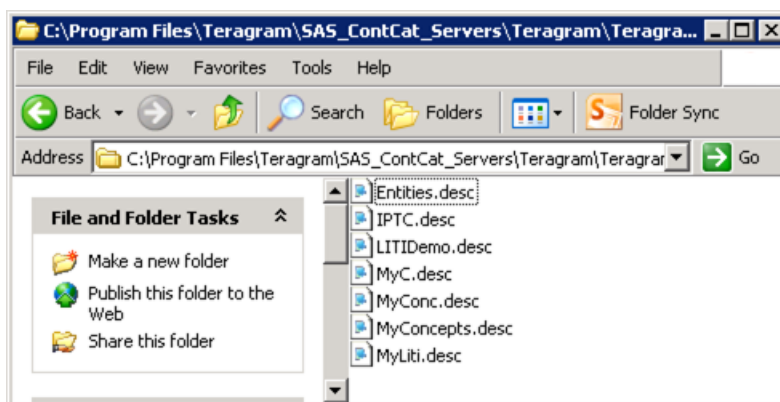


## 4.6 Using the Descriptors Directory

The descriptor directory (`descriptors`) contains information about each project that is uploaded to SAS Content Categorization Server. This information is in the form of binary files called descriptor files that follow the `<project name>.desc` naming convention. These files are automatically created for each project that is uploaded. Without the `descriptor` directory, no project files can be uploaded.

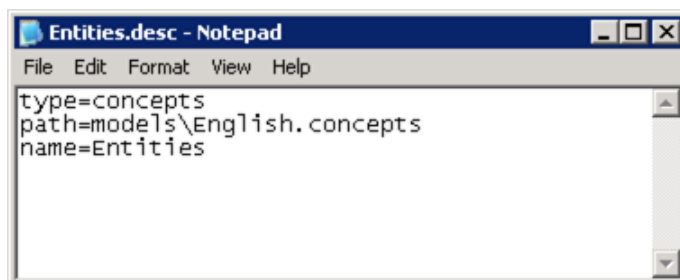
A plain text `.desc` file is created each time a creator uploads a project, unless a project with a duplicate name is uploaded. In this case, the existing file is overwritten. See the following example:

Display 4-2 The Descriptors Directory



Double-click the .desc file to access and read the file contents. The format for these files is displayed in the Entities.desc example that is shown below:

Display 4-3 Sample Descriptor File



See the explanations of the lines in this file that are shown below:

`type=[mcat|concepts|liti]`

specifies the type of project that is uploaded. You can upload a categories (mcat), concepts (concepts), or contextual extraction (li) file.

`path=<file_location>`

specifies the data path relative to the base, or installation, directory. If basedir is not specified in the configuration file, this path is treated as an absolute path.

---

`name=<symbolic_name>`

specifies the name that you assign to the project.

---

**Notes:** The file path is relative only to `basedir`. This path is not relative to `basedir` combined with `create_dir`. The deprecated directives can co-exist with the `descriptor_dir` directive. For more information, see Table 4-2 on page 50.

---

## 4.7 Using `cat_log` and `concept_log` Files

There are several ways that you can use the `cat_log` files. These examples also apply to `concept_log` files. The word *entry* is defined here in the context of the SAS Content Categorization Server configuration file. *Entry* represents a document that matches one or more categories or concepts within the project.

---

**Note:** If you want to use the features in SAS Enterprise Content Categorization Servers Administration Web Page, specify the appropriate log files in your `server.config` file. For more information, see Table 4-1 on page 46.

---

For example, if an input text matches one category in each of two projects, two entries are created. If however, another text matches five categories in one project and two in another project, two entries are also created. (For information about loading multiple project files into your SAS Content Categorization Server, see Section 4.8 *Specifying Multiple Project Files* on page 58.)

Your SAS Content Categorization Server configuration file might contain the following line:

```
cat_log=cat.log
```

---

In this example, the category log file is named `cat.log.0`. In this case, because there is no specification for `cat_log_max_entries`, the number of entries continue to grow. This growth continues until the log file is deleted or until there is no more disk space. However, you can specify a maximum number of entries for this file using the following example:

```
cat_log=cat.log
cat_log_max_entries=10000
```

In this example, the `cat.log.0` file is regenerated whenever more than 10,000 documents match at least one category, in at least one project, and during one session. When the `cat.log.0` file is regenerated, all existing data in the file is lost. For this reason, you can also configure SAS Content Categorization Server to create more than one log file. See the following example:

```
cat_log=cat.log
cat_log_max_entries=10000
num_cat_logs=10
```

To begin the regeneration process, SAS Content Categorization Server creates a file named `cat.log.0`. The server might attempt to exceed the specified number of entries. In this example 10,000 entries are specified. In this case, `cat.log.0` is copied to `cat.log.1`, and `cat.log.0` is regenerated to include the excess entries. This process can continue until the limit of `cat.log.9` is reached. In this case, excess data is copied and the first log file is destroyed when the maximum number of entries is reached.

## 4.8 Specifying Multiple Project Files

You can choose to load as many project files to SAS Content Categorization Server as your system can hold in memory. If you update your projects in SAS Content Categorization Studio, you can reload the updated projects using the same symbolic name specified for the original project.

The first project loaded on the server, or the first descriptor file that is read, for each binary file type is the default project. These project names are also used by the APIs for SAS Content Categorization Server.



---

## 4.9 Sending Documents to the Server

In order to send documents to SAS Content Categorization Server, you use the Java programs described in the `README` file. This file is located in the `client_api` directory that use the Java API. See the path below:

```
Teragram Catcon Server\client_api\java
```

---

**Note:** You can use SAS Content Categorization Studio to upload projects to SAS Content Categorization Collaborative Server. However, you cannot use SAS Content Categorization Studio to send documents to SAS Content Categorization Server.

---

## 4.10 Run SAS Content Categorization Server

### 4.10.1 Verify That the Server Is Running on Windows

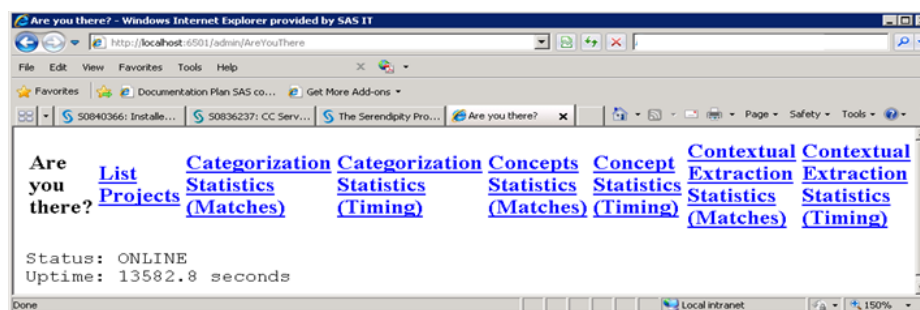
You can check to see whether SAS Content Categorization Server is running using the Administration Web Page or by checking the Services running on your machine.

To check whether the SAS Content Categorization Server service is running using the Administration Web page, complete these steps:

1. Go to **Start --> Programs --> SAS --> SAS Enterprise Content Categorization Servers --> SAS Content Categorization Server Administration Web Page.**

---

## Display 4-4 The Administrative Web Page



2. Click **Are you there?**
3. See the **Status** and **Uptime**. When the server is running the status is **ONLINE** and **Uptime** displays a number of seconds.

To stop or restart SAS Content Categorization Server, use the directions in Section 2.3 *Starting and Stopping the Servers on a Windows Machine* on page 29.

### 4.10.2 Run the Server on UNIX

To run SAS Content Categorization Server on a supported UNIX system, go to the installation root directory and enter the following command from the UNIX shell:

```
# ./bin/<arch>/-catcon_server.exe
```

For example:

```
# ./bin/linux64/_catcon_server
```

In this command line, `-server configfile` specifies the name and full path to the configuration file. The server program runs in the foreground. This means that it does not fork and writes its logging output to the terminal that initiated the program (`stdout`).

---

**Note:** Use the `-verbose` switch for debugging purposes.

---

---

## 4.11 Optimize Performance on a Client Windows Machine

### 4.11.1 Overview of Performance Optimization

The settings that are specified in the following sections should be applied if the client program that connects to SAS Content Categorization Studio is running on Windows. Otherwise, unexpected behaviors might occur when you process large amounts of documents.

### 4.11.2 Before and After You Optimize Performance

Before you use the following sections to optimize the performance of SAS Content Categorization Server, run the registry editor.

To run the registry editor, complete these steps:

1. Select **Start --> Run**.
2. Type `regedit` into the **Open** field of the Run window that appears.
3. Click **OK**.
4. After you use both Section 4.11.3 *Adjust the TCP Time Wait State* below and Section 4.11.4 *Reset Ephemeral Ports* on page 62, reboot your machine.

### 4.11.3 Adjust the TCP Time Wait State

Choose to lower the setting for the timed wait state in order to avoid depleting available ports on your servers. SAS recommends that you consider setting this selection in your system registry to 15 seconds.

To reset the `TCPTimedWaitDelay` setting, complete these steps:

1. Go to the registry subkey:  

```
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\  
Tcpip\Parameters
```
2. Select **Edit --> New --> DWORD value**. By default, the new value is named `New Value #1`.

- 
3. Rename the value by entering `TcpTimedWaitDelay`.
  4. Double-click on the new `TcpTimedWaitDelay` value.
  5. Select `Decimal` as the base, and enter 15 for the value data.

#### 4.11.4 Reset Ephemeral Ports

Ephemeral ports are short-lived ports that are used to create connections to the client computers from the server and between COM server objects. By default, these ports range from 1024 to 5000. Connection difficulties can occur if you run short of ports.

This section explains how to reset the parameter that controls the maximum port number that is used when the SAS Content Categorization Server program requests an available user port from the system.

Use the following steps to reset the valid range for ephemeral ports:

1. Go to the registry subkey:  

```
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\Services\  
    Tcpip\Parameters
```
2. Select **Edit --> New --> DWORD value**. By default, the new value is named `New Value #1`.
3. Rename the value by entering `MaxUserPort`.
4. Double-click on the new `MaxUserPort` value.
5. Select `Decimal` as the base, and enter 65534 for the value data.

---

## Chapter: 5

# SAS Content Categorization Server Web Administration

---

- *Overview of SAS Content Categorization Server Web Administration*
- *Access the Administration Web Interface*
- *Using the Administration Web Page*

## 5.1 Overview of SAS Content Categorization Server Web Administration

Use the SAS Content Categorization Server administration Web interface while SAS Content Categorization Server is running, to see information about category matching and concept extraction. This interface can be used by either the server or creator administrators, or by a regular user.

To use the features of the Web administration page, specify the appropriate log files in your `server.config` file. For more information, see Table 4-1 on page 46.

The Administration Web Page enables you to perform the following tasks:

- Check to see that SAS Content Categorization Server is running.
- See a list of all of the loaded category and concept extraction projects.
- Use tables of statistics to analyze the categorization and concept extraction results.

---

**Note:** The statistics generated for category matches and concept extraction do not appear by default. To see these results specify the `cat_log` and `concept_log` lines

---

in the SAS Content Categorization Server configuration file.

---

For more information about the files that are loaded using the server configuration file, see Chapter 4: *Configuring and Running SAS Content Categorization Server*.

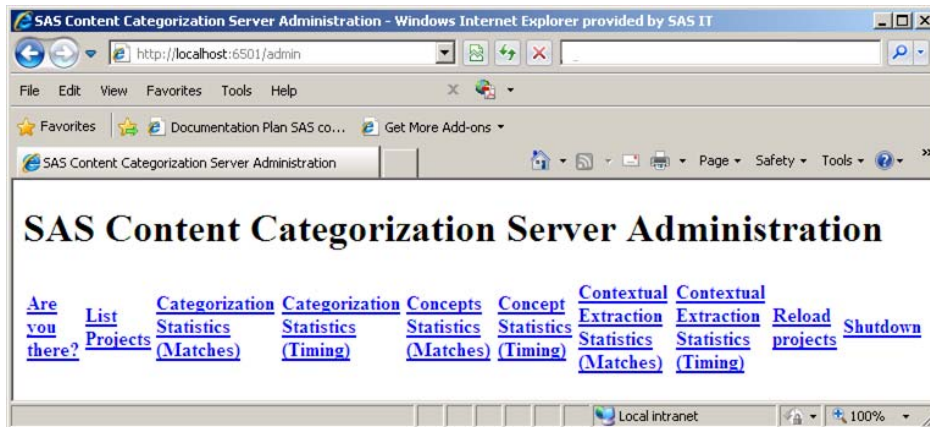
## 5.2 Access the Administration Web Interface

Use the links on the SAS Content Categorization Server Administration Page to gain information about the documents that are processed by this server.

The SAS Content Categorization Server Administration Page is displayed when you take the following step:

Go to **Start** —> **Programs** —> **SAS Enterprise Content Categorization Servers** —> **Start SAS Enterprise Content Categorization Servers Administration Web Page**.

The SAS Content Categorization Server - Administration page appears:



---

## 5.3 Using the Administration Web Page

### 5.3.1 Overview of the Administration Web Page

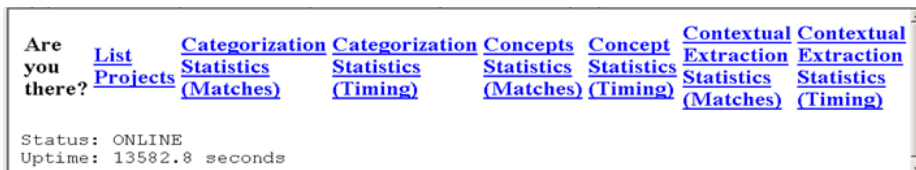
Use the administration Web page to see the operational data, in Web page format, as this information becomes available in SAS Content Categorization Server. Before completing the following steps, specify the `cat_log` or `concepts_log` directives in the server configuration file. For more information, see Section 4.7 *Using cat\_log and concept\_log Files* on page 57.

### 5.3.2 Use the Are you there? Page

The Are you there? page provides information about the status of the server and the length of time that the server was running.

To open and use the Are you there? page, complete these steps:

1. Click **Are you there?** and the Are you there? page appears.



2. See the following information about the server on this page:

#### Status

ONLINE, when running. Otherwise, there is no reply from the server. If the server is not started, see Section 2.3 *Starting and Stopping the Servers on a Windows Machine* on page 29.

---

**Uptime**

displays the length of time that SAS Content Categorization Server has been running, or when the process was restarted. In the example above, the time is 2.2 seconds.

### 5.3.3 Use the SAS Content Categorization Server Projects List Page

The SAS Content Categorization Server Projects List page displays the categorization and concepts projects that are loaded on the server. This page also displays relevant information about each project such as the name and type of the project.



To use the SAS Content Categorization Server Projects List page, complete these steps:

1. Click **List projects** and the SAS Content Categorization Server Projects List page appears.

## SAS Content Categorization Server Projects List

[Are you there?](#)
[List Projects](#)
[Categorization Statistics \(Matches\)](#)
[Categorization Statistics \(Timing\)](#)
[Concepts Statistics \(Matches\)](#)
[Concept Statistics \(Timing\)](#)
[Contextual Extraction Statistics \(Matches\)](#)
[Contextual Extraction Statistics \(Timing\)](#)

### Categorization Projects

	File Name	Symbolic Name	Type	Number of Rules	Default Category Bias	Default Relevancy Cutoff
0	models\English.mco	IPTC	Rule-based	1366	0	0.000000

### Concepts Projects

	File Name	Symbolic Name	Number of Rules
0	models\031912.concepts	031912	2

- Use the tables in the SAS Content Categorization Server Projects List page to analyze information about the projects running on your server.

Table 5-1: Projects List Page Information

Heading	Description
Categorization Projects	
File Name	Specifies the name and location of the project file that was exported from SAS Content Categorization Studio. For example, the <code>English.mco</code> file located in the <code>models</code> folder.
Symbolic Name	Specifies the name of the project that you enter into the Server <b>Project Name</b> field of the Upload Categories to SAS Content Categorization Studio window.
Type	Specifies the rule type, which is either rule-based or statistical.
Number of Rules	Specifies the number of categories in this project.
Default Category Bias	Specifies the number that is set as the relevancy bias in SAS Content Categorization Studio for your categories. By default, this setting is set to 0.
Default Relevancy Cutoff	Specifies that any matching documents with a score below this number is not considered a match for a category. By default, this setting is set to 0.000000.
<b>Note:</b> You can change the <b>Default Category Bias</b> and the <b>Default Relevancy Cutoff</b> settings in SAS Content Categorization Studio Project Settings - Category window.	
Concepts Projects	
File Name	Specifies the name of the project file that was exported from SAS Content Categorization Studio. For example, this name could be <code>English.concepts</code> .
Symbolic Name	Specifies the name of the project that you enter into the <b>Server Project Name</b> field of the Upload Concepts to SAS Content Categorization Studio window.
Number of Rules	Specifies the number of concepts in this project.
For contextual extraction Projects, see the information that is specified in the table cells beneath Concepts Projects above. An example of the filename is <code>Sample.li</code> .	

### 5.3.4 Use the SAS Content Categorization Server Categorization Statistics (Matches) Page

The SAS Content Categorization Server Categorization Statistics (Matches) page lists the names of categories in the projects loaded onto the server. This page also displays information about the matches for these categories.

To use the SAS Content Categorization Server Categorization Statistics (Matches) page, complete these steps:

1. Click **Categorization Statistics (Matches)** and the SAS Content Categorization Server Categorization Statistics (Matches) page appears.

**SAS Content Categorization Server Categorization Statistics (Matches)**

[Are you there?](#)
[List Projects](#)
[Categorization Statistics \(Matches\)](#)
[Categorization Statistics \(Timing\)](#)
[Concepts Statistics \(Matches\)](#)
[Concept Statistics \(Timing\)](#)
[Contextual Extraction Statistics \(Matches\)](#)
[Contextual Extraction Statistics \(Timing\)](#)

Total number of documents: 7

Total number of categories with at least one match: 28

Category Name	Percentage of Documents	Percentage (relative)
Top/04000000 - Economy, Business and Finance/04003000 - Computing and Information Technology/04003004 - Semiconductors and Active Components	14.285714%	
Top/04000000 - Economy, Business and Finance	14.285714%	
Top/13000000 - Science and Technology	14.285714%	
Top/04000000 - Economy, Business and Finance/04016000 - Company Information	14.285714%	
Top/13000000 - Science and Technology/13016000 - Electronics	14.285714%	

- 
2. Use the data that appears in the SAS Content Categorization Server Categorization Statistics (Matches) page to gain information about matching categories for the input texts:

Table 5-2: Categorization Statistics (Matches) Information

Information Type	Description
Total number of documents	This number represents the total number of documents that match one or more category rules. The example above specifies 7.
Total number of categories with at least one match	This number represents the total number of categories that have one or more matching documents. This example specifies 28 categories have matches.
Category Name	This number includes the full pathname of the category with one or more matching documents.
Percentage of Documents	This number represents the proportion of texts that matched the specified category.
Percentage (relative)	See the bar chart to visually compare the results shown in the <b>Percentage of Documents</b> column.

### 5.3.5 Use the SAS Content Categorization Server Categorization Statistics (Timing) Page

The SAS Content Categorization Server Categorization Statistics (Timing) page displays data about document processing and timing.

To use the SAS Content Categorization Server Categorization Statistics (Timing) page, complete these steps:

1. Select the **Categorization Statistics (Timing)** link and the SAS Content Categorization Server Categorization Statistics (Timing) page appears. The tables in this screen contain timing information relative to processing the input documents.

**SAS Content Categorization Server Categorization Statistics (Timing)**

[Are you there?](#)
[List Projects](#)
[Categorization Statistics \(Matches\)](#)
[Categorization Statistics \(Timing\)](#)
[Concepts Statistics \(Matches\)](#)
[Concept Statistics \(Timing\)](#)
[Contextual Extraction Statistics \(Matches\)](#)
[Contextual Extraction Statistics \(Timing\)](#)

Documents per second: 21.28

Time taken	Minimum (sec)	Maximum (sec)	Average (sec)
Idle	1299784688.321000	1299784688.477000	1299784688.391000
I/O	0.000000	0.016000	0.007750
Processing	0.016000	0.047000	0.039250

#	Name	Size (bytes)	Categorization project	Time (sec)	Time (relative)	Categories returned
1	..\test\fact_rule.txt	161	IPTC	0.016		0
2	..\test\sample1.txt	4708	IPTC	0.047		15
3	..\test\sample2.txt	3387	IPTC	0.047		5

2. See the Documents per second statistics to see the total number of documents processed by SAS Content Categorization Server. In the example shown above this number is 21.28.

- 
- Use the information in the first table for the time required to process the input documents:

Table 5-3: First Categorization Timing Table Information

Heading	Description
Time taken	The following types of timing occur with document processing:  Idle: The amount of time that SAS Content Categorization Server was not processing documents.  I/O: The amount of time that it took to input and output a single text.  Processing: The time required to process a document.
Minimum (sec)	The fewest number of seconds used to process any one document.
Maximum (sec)	The highest number of seconds used to process any single text.
Average (sec)	The number of seconds required to process all of the input documents divided by the total number of processed texts.

- Use the information in the second table to see the number of categories that match each input document:

Table 5-4: Second Categorization Timing Table Information

Heading	Description
#	Each input document incrementally increases by this number, beginning with 1.
Name	The name, if there is one, of the processed document.
Size (bytes)	The size of the processed documents in bytes.
Categorization project	The symbolic name of the categories project that is matched by these categories.

---

Table 5-4: Second Categorization Timing Table Information

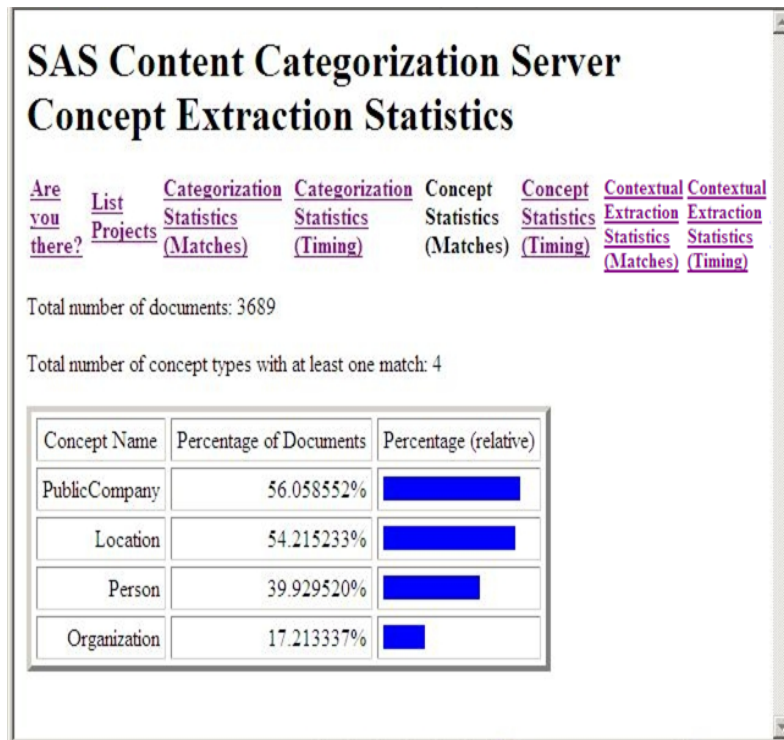
Heading	Description
Time (sec)	The number of seconds that it took to process the specified document.
Time (relative)	An overview of the preceding column <b>Time (sec)</b> . Use this bar chart for comparison purposes.
Categories returned	The number of category rules that this document matches.

### 5.3.6 Use the SAS Content Categorization Server Concept Extraction Statistics Page

The SAS Content Categorization Server Concept Extraction Statistics page displays data about the concepts that are extracted, or matched, in the input documents. This page displays information about the numbers of documents that are matched and the concepts that they match.

To use the SAS Content Categorization Server Concept Extraction Statistics page, complete these steps:

1. Select **Concept Statistics (Matches)** to see the statistical data compiled for matches in processed documents. The SAS Content Categorization Server Concept Extraction Statistics page appears.



2. Analyze the data that appears in the SAS Content Categorization Server Concept Extraction Statistics page:

**Total number of documents**

represents the total number of texts that have matched one or more concept definitions. For example, 3689 texts are matched.

**Total number of concept types with at least one match**

represents to the total number of concepts that have at least one matching document. For example, 4 concepts match at least one input document.

3. Evaluate the information that appears in the table:



---

**Concept Name**

see the name of the concept that one or more documents match.

**Percentage of Documents**

use these figures to see the concepts that have the highest percentage of matching documents.

**Percentage (relative)**

use this bar chart to visually compare the results shown in the **Percentage of Documents** column.

### 5.3.7 Use the SAS Enterprise Content Categorization Servers Concept Extraction Statistics (Timing) Page

The SAS Content Categorization Server Concept Extraction Statistics (Timing) page displays data about concepts that are extracted, or matched, in input documents. This page displays information about the numbers of documents that are matched and the concepts that they match.

To use the SAS Content Categorization Server Concept Extraction Statistics (Timing) page, complete these steps:

1. Select the **Concept Statistics (Timing)** link to see these statistics and the SAS Content Categorization Server Concept Extraction Statistics (Timing) page appears.

**SAS Content Categorization Server Concept Extraction Statistics (Timing)**

[Are you there?](#) [List Projects](#) [Categorization Statistics \(Matches\)](#) [Categorization Statistics \(Timing\)](#) [Concepts Statistics \(Matches\)](#) [Concept Statistics \(Timing\)](#) [Contextual Extraction Statistics \(Matches\)](#) [Contextual Extraction Statistics \(Timing\)](#)

Documents per second: 86.96

Time taken	Minimum (sec)	Maximum (sec)	Average (sec)
Idle	4.374000	4.484000	4.425000
I/O	0.000000	0.016000	0.011500
Processing	0.000000	0.000000	0.000000

#	Name	Size (bytes)	Concept extraction project	Time (sec)	Time (relative)	Concepts returned
1	..\test\fact_rule.txt	161	Entities	0.000		0
2	..\test\sample1.txt	4708	Entities	0.000		18

2. See processing speed for the documents to the right of **Documents per second**. For example, see 86.96.
3. Analyze the data that appears in the first table. For more information, see Table 5-3 on page 72.

- 
4. Analyze the data that appears in the second table:

Table 5-5: Second Contextual Extraction Timing Table Information

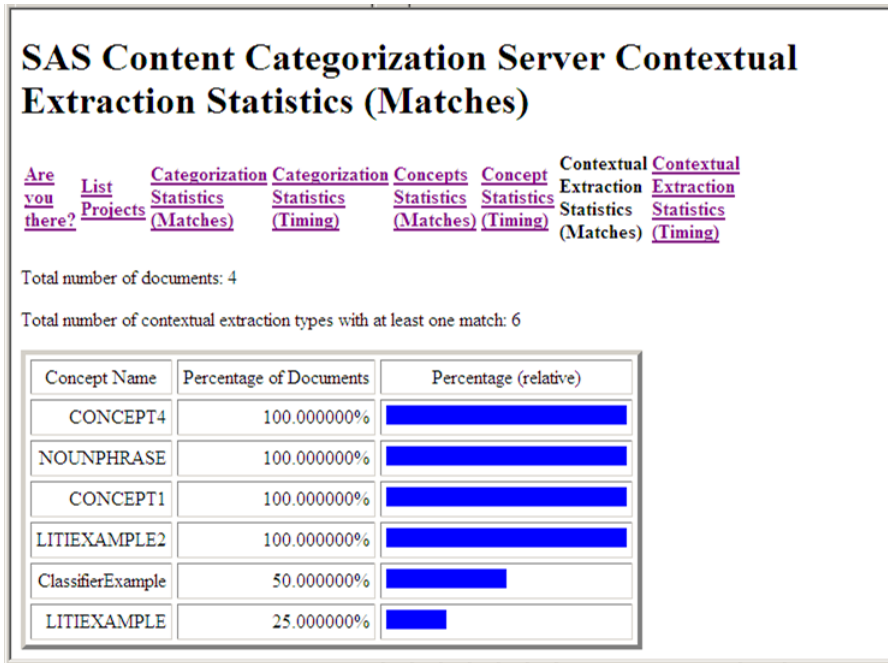
Heading	Description
#	Each input document incrementally increases by this number, beginning with 1.
Name	The name, if there is one, of the processed document.
Size (bytes)	The size of the processed documents in bytes.
Concept extraction project	When more than one concepts project is loaded, this column differentiates between the projects where concepts are extracted.
Time (sec)	The number of seconds that it took to process the specified document.
Time (relative)	An overview of the preceding column <b>Time (sec)</b> . Use this bar chart for comparison purposes.
Concepts returned	The number of concept definitions that this document matches.

### 5.3.8 Use the SAS Content Categorization Server Contextual Extraction Statistics (Matches) Page

The SAS Content Categorization Server Contextual Extraction Statistics (Matches) page displays data about the contextual extraction concepts matched in input documents. This page displays information about the numbers of documents that are matched and the concepts that they match.

To use the SAS Content Categorization Server Concept Extraction Statistics (Matches) page, complete these steps:

1. Select the **Contextual Extraction Statistics (Timing)** link to see these statistics and the SAS Content Categorization Server Contextual Extraction Statistics (Timing) page appears.



2. See the total number of documents processed to the right of **Total number of documents**. For example, see 4.
3. See the **Total number of contextual extraction types with at least one match**. For example, see 6.
4. Analyze the data that appear in the first table. For more information see Section 5.3.4 *Use the SAS Content Categorization Server Categorization Statistics (Matches) Page* on page 69.

### 5.3.9 Use the SAS Content Categorization Server Contextual Extraction Statistics (Timing) Page

The SAS Content Categorization Server Contextual Extraction Statistics (Timing) page displays data about the time that it takes to process the input documents. This page displays information about the numbers of documents that are matched and the concepts that they match.

To use the SAS Content Categorization Server Concept Extraction Statistics (Timing) page, complete these steps:

1. Select the **Contextual Extraction Statistics (Timing)** link to see these statistics and the SAS Content Categorization Server Contextual Extraction Statistics (Timing) page appears.

## SAS Content Categorization Server Contextual Extraction Statistics (Timing)

[Are you there?](#)
[List Projects](#)
[Categorization Statistics \(Matches\)](#)
[Categorization Statistics \(Timing\)](#)
[Concepts Statistics \(Matches\)](#)
[Concept Statistics \(Timing\)](#)
[Contextual Extraction Statistics \(Matches\)](#)
[Contextual Extraction Statistics \(Timing\)](#)

Documents per second: 15.09

Time taken	Minimum (sec)	Maximum (sec)	Average (sec)
Idle	1.453000	1.672000	1.542750
I/O	0.000000	0.016000	0.008000
Processing	0.000000	0.093000	0.058250

#	Name	Size (bytes)	Contextual extraction project	Time (sec)	Time (relative)	Concepts returned
1	..\test\fact_rule.txt	161	LITIDemo	0.000		12
2	..\test\sample1.txt	4708	LITIDemo	0.078	<div style="width: 100%; height: 10px; background-color: blue;"></div>	143
3	..\test\sample2.txt	3387	LITIDemo	0.062	<div style="width: 90%; height: 10px; background-color: blue;"></div>	109
4	..\test\sample3.txt	4776	LITIDemo	0.093	<div style="width: 100%; height: 10px; background-color: blue;"></div>	149

- 
2. See the total number of documents processed per second to the right of **Documents per second**. For example, see 15.09.
  3. Analyze the data that appear in the first table. For more information see Section 5.3.6 *Use the SAS Content Categorization Server Concept Extraction Statistics Page* on page 73.

---

# Part 4: SAS Content Categorization Collaborative Server

---

- Chapter 6: *Setting Up a Project on the Server on page 83*





---

# 6

## Setting Up a Project on the Server

---

- *Overview of Setting Up a Project on SAS Content Categorization Collaborative Server*
- *Installing and Configuring a SQL-Based Server*
- *Installing the \_tk240 Utility*
- *Setting Up ODBC Data Sources*
- *Defining Users*
- *Logging In*
- *Quick Start Guide Using \_tk240*

### 6.1 Overview of Setting Up a Project on SAS Content Categorization Collaborative Server

When your organization chooses to run SAS Content Categorization Studio in the collaborative mode, you as the database administrator, set up data sources on the server. (This mode is SAS Content Categorization Collaborative Server.) You also add users with varying levels of permissions to the database. This chapter is written primarily for database administrators, including those who might also function as a project administrator.

Sections of this chapter also contain information about the tasks that end users perform. These users can set up their own data sources that access the SAS Content Categorization Studio project running on the server.

---

**Note:** This chapter contains information for both administrators and regular users. The chapter assumes that administrators often set up data sources for users.

---

For more information about the tasks that apply to both types of users, see Section 6.4 *Setting Up ODBC Data Sources* on page 87.

## 6.2 Installing and Configuring a SQL-Based Server

### 6.2.1 Install the SQL-Based Server

Install a database server on your server machine in order to set up a repository for SAS Content Categorization Studio. Both the SAS Content Categorization Studio application and `_tk240` support MySQL, Microsoft SQL Server, and Oracle. For more information about `-tk240`, see Section 2.2.2.B *See the Directory Structure* on page 28.

---

**Note:** SAS has certified Microsoft SQL Server 2005, or newer, at this point in product development.

---

If you are running Microsoft SQL Server, the ODBC drivers are pre-installed on Windows. However, if you want to connect to a Microsoft SQL Server database from UNIX, a third-party solution is required.

The address to install MySQL is below:

<http://dev.mysql.com/downloads/mysql>

After installing MySQL, make the following changes to the MySQL server configuration file to ensure that SAS Content Categorization Studio runs smoothly. These changes are important if you plan to develop classifier concepts with 100,000, or more, entries.

---

## 6.2.2 Change the Size of the .cnf File for Linux

To change the size of data in the `my.cnf` file, complete these steps:

1. Copy the `my-xxx.cnf` file from `/usr/share/mysql` to `/etc` as `my.cnf`.

The size of this `.cnf` file can vary from small to large, depending on your requirements.

2. Locate the file. The location depends on the SQL installation. For example, see the following path:

```
cp /usr/share/sqlfilename /etc/my.cnf
```

or:

```
cp /usr/share/sqlfilename /etc/mysql/mysql.cnf
```

3. Change the following default setting in the `my.cnf` file:

```
max_allowed_packet = 1M
```

to:

```
max_allowed_packet = 32M
```

For classifier definitions of one million or more entries, the value is 96M.

4. Save the file.
5. Restart MySQL using the following command line:

```
/etc/init.d/mysql restart
```

## 6.2.3 Change the Size of the Allowed Packet Size for Windows

To increase the allowed packet size of data in the Microsoft SQL Server program directory, complete these steps:

1. Open the `my.ini` file that was either installed, or created when you configured the server. This file is located in the following, or in a similar, location:

```
C:\Program Files\MySQL
```

2. Change the following default setting under The SQL Server [mysqld] section:

---

```
max_allowed_packet = 1M  
to:  
max_allowed_packet = 32M
```

---

**Note:** Specify 96 MB for the `max_allowed_packet` if your concepts project is large. Large is defined as one million entries for a single classifier concept.

---

3. Save the file.
4. Restart Microsoft SQL Server. Select **Control Panel --> Administrative Tools --> Services --> SQL**.

## 6.3 Installing the `_tk240` Utility

Before you can configure your ODBC data sources, install `_tk240`. The `_tk240` utility is a command line program that enables you to set up the collaborative mode. This program is shipped in a package with a name that is similar to the example shown below:

```
sas_collab_server_<arch>.tar.gz (UNIX)
```

Download this `.zip` file onto your computer and extract its files into the location of your choice.

---

**Note:** To specify the path to the SAS license, see Section A.7 *Set the Path to the SAS License* on page 89.

---

---

## 6.4 Setting Up ODBC Data Sources

### 6.4.1 Driver Types

Install and configure the ODBC driver connectivity software on the server and client systems that are used to access the SAS Content Categorization Collaborative Server project repositories.

SAS Content Categorization Collaborative Server uses three types of ODBC data sources. The first two configurations are for the administrator's machine. The third configuration is for regular users.

- The first ODBC data source points to the default system database and is used by the database administrator to create the SAS Content Categorization Collaborative Server database on the server.
- The second ODBC data source points to the SAS Content Categorization Collaborative Server database. Database administrators use this driver to add users to this database.
- Regular users can create an ODBC data source on a local machine that points to the SAS Content Categorization Collaborative Server database.

For more information, see Section 6.4.3 *Configuring Data Sources* on page 88 and Section 6.7 *Quick Start Guide Using \_tk240* on page 106.

### 6.4.2 Installing the ODBC Drivers

SAS Content Categorization Studio and \_tk240 use ODBC drivers to connect to database repositories.

---

**Note:** To install the ODBC software, make sure that your privileges are set to perform an installation.

---

You can download the MySQL ODBC drivers for either Windows or UNIX, by going to the following location:

<http://dev.mysql.com/downloads/connector/odbc>.

Microsoft SQL Server, Windows: preinstalled

---

Microsoft SQL Server, UNIX: third-party solution required

Oracle: Complete these steps:

1. Go to <http://www.oracle.com/technetwork/database/features/instant-client/index-097480.html>.
2. Download the Database Instant Client for your architecture.

## 6.4.3 Configuring Data Sources

### 6.4.3.A Credentials

Use the following credentials to configure each of the three data sources that are required to run SAS Content Categorization Collaborative Server and \_tk240. For more information, see:

- Section 6.4.3.C *Configure MySQL Data Sources* on page 91
- Section 6.4.3.D *Configure Microsoft SQL Server Data Source* on page 94
- Section 6.4.3.E *Download and Configure Oracle ODBC Drivers* on page 101

There are several credentials that supply the information required to create the three different types of data sources:

Name

User

Password

Database: You can either specify a name or use the default entry. For more information, see Table 6-1 on page 90.

Examples of these credentials are provided for each of the following three data sources that you create:

- Administrators with database permissions should see Step 1 on page 106 to initialize a SAS Content Categorization Studio database on the server using these credentials.
- Administrators with database permissions should see Step 2 on page 106 to add users to the SAS Content Categorization Studio database.

- 
- Regular users use the example of credentials shown in Step 4 on page 107 to open a remote project after you create the third data source.

### 6.4.3.B Configuring ODBC Data Sources on UNIX

Administrators can use `_tk240` to set up database connectivity. If you, as an administrator, choose to configure through UNIX, all ODBC data sources are configured using an `odbc.ini` file. To locate this file, see the path specified by the `ODBCINI` environment variable or see your home directory. If the `odbc.ini` file is stored in `/etc`, it is not writeable. To create a local `.odbc.ini` file, create a file with this name in your home directory. If necessary, create this file by hand. For more information about the file format, see the following link:

<http://www.iodbc.org/index.php?page=docs/faq#iodbc237887083>

You can also use the graphical data source administrative tool if this tool is installed on your system.

Use the sample `odbc.ini` file displayed below to either create or to modify your `odbc.ini` file:

*Example 6-1: A Local odbc.ini File*

```
;
; odbc.ini
;
[ODBC Data Sources]
data_source_name1 = ODBC Driver for MySQL
data_source_name2 = ODBC Driver for MySQL

[data_source_name1]
Driver           = /usr/lib/odbc/libmyodbc.so
Server           = 127.0.0.1
Port             = 3306
Database         = database_name

[data_source_name2]
Driver           = /usr/lib/odbc/libmyodbc.so
Server           = server_name
Port             = 3306
Database         = database_name
```

---

**Note:** The example above uses the `MySQL` database. If you are connecting to Microsoft SQL Server or Oracle, specify the appropriate name.

---

Use the second example to create your third data source as well as your second data source.

As shown in Example 6-1 above, you create each data source on a separate line that appears below [ODBC Data Sources]. This section of the `odbc.ini` file lists all of the ODBC data sources. Below this section, you specify each data source name followed by separate lines detailing the `Driver`, `Server`, `Port`, and `Database` information.

Table 6-1: Data Source Parameters

Parameter	Description
Driver	Specify the name and location of the driver library that is installed as part of the ODBC driver. For example, the name of this file might be <code>libmyodbc.so</code> .
Server	Specify the name of the machine where MySQL Server is running. Use <code>127.0.0.1</code> instead of <code>localhost</code> if you are using MySQL Server running on <code>localhost</code> .
Port	Specify the number of the port that connects with MySQL Server. MySQL Server default: 3306 SQL Server default: 1433 Oracle default: 5560  This is true unless MySQL Server was configured to use another port.
Database	Specify the name of the database to access using this data source.

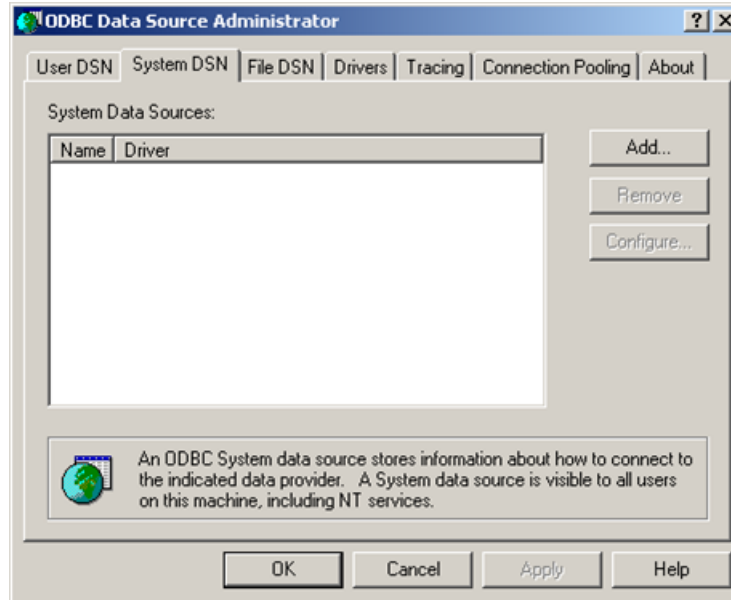


---

### 6.4.3.C Configure MySQL Data Sources

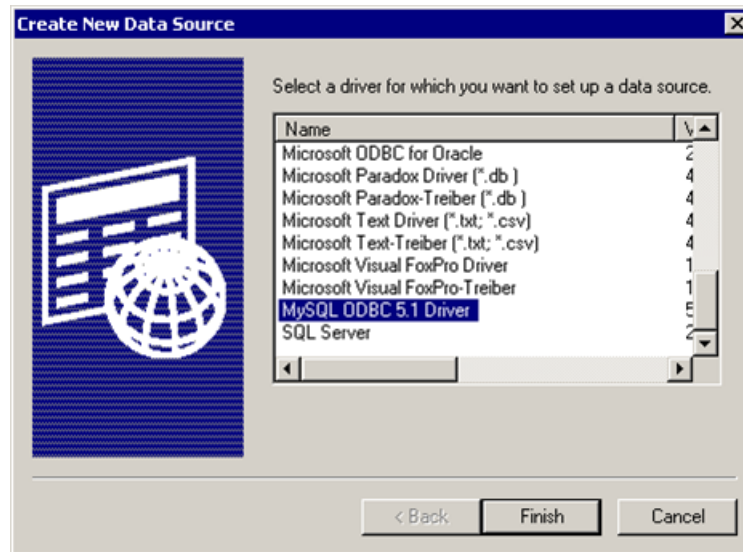
To configure data sources on a Windows machines, complete these steps:

1. Select **Control Panel --> Administrative Tools --> Data Sources (ODBC)** and the ODBC Data Source Administrator wizard appears.



2. (Administrators) Select the **System DSN** tab.  
(Regular users) Select the **User DSN** tab.

- 
3. Click **Add**. The Create New Data Source window appears.



4. Use the scroll bar to locate the appropriate driver. If you select MySQL, see Step 5 below.

If you are using Microsoft SQL Server, see Step 5 on page 95.

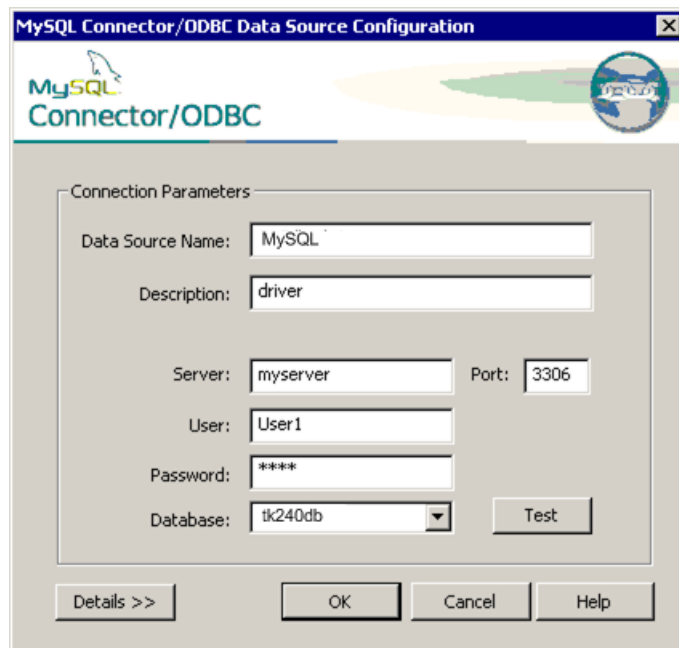
If you are using Oracle, see Step 5 on page 102.

---

**Note:** Version 5.1 of the MySQL ODBC driver is used for example purposes only.

---

5. Click **Finish** and the MySQL Connector/ODBC page appears. Use the MySQL Connector/ODBC wizard to point the ODBC driver to the server:



6. Enter the name of the data source into the **Data Source Name** field. For example, administrators might use `SASdbadmin` for the first database and `SASdb` for the second data source. Regular users can use the database name selected by the administrator. For example, they can use `SASdb`.
7. (Optional) Enter the description into the **Description** field. For example, type `driver`.
8. Enter the server name into the **Server** field. For example, type `myserver`.
9. (Optional) The default entry, shown in the right pane of this interface, is entered into the **Port** field.
10. Enter the name of the person making the connection, into the **User** field. Administrators enter their administrative user name as entered on the

---

server. Regular users specify the user name entered for them by the administrator who added users to the database.

---

**Note:** You cannot use a period (.) or an at sign (@) sign when you create a user name for MySQL Server. For example, you cannot use an e-mail address as your user name.

---

11. Enter your password into the **Password** field. The password for the administrator is the administrator's password on the server.
12. Enter the name of the database that you are connecting to into the **Database** field. The administrator specifies `mysql` the first time she or he uses this section. The second time, the administrator like all regular users who are using this section for the first time, specifies the database name. For example, you might type `tk240db`.
13. Click **Test**. If the connection is successful, the Connector/ODBC window appears with a message stating that the connection is successful.

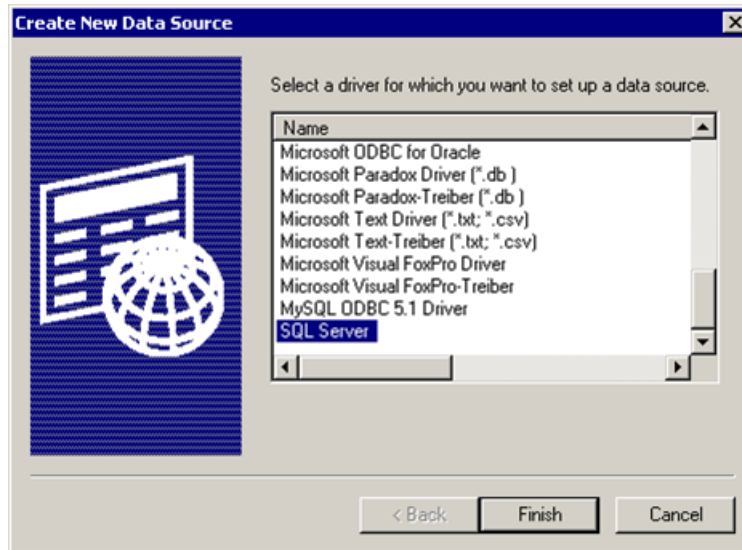


14. Click **OK**.
15. (Optional) Click **Details** in the MySQL Connector/ODBC window to set various types of flags.
8. Click **OK** in the MySQL Connector/ODBC window.

#### 6.4.3.D Configure Microsoft SQL Server Data Source

Use the following steps to configure the SAS Content Categorization Collaborative Server database using Microsoft SQL Server. These steps continue from Step 4 on page 92:

- 
5. (Optional) If you select `SQL Server` in the `Create New Data Source` page, click **Finish**.



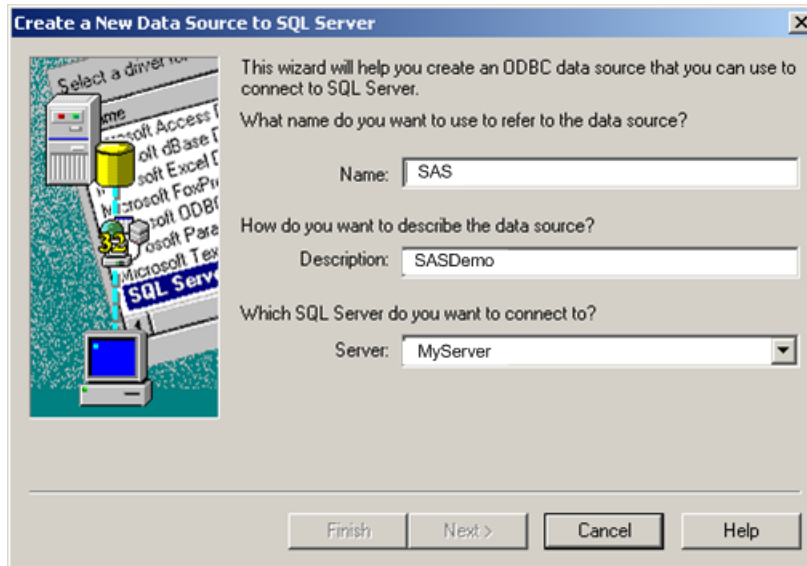
---

**Note:** SQL Native Client is also available on some machines. This SQL selection is also valid.

---

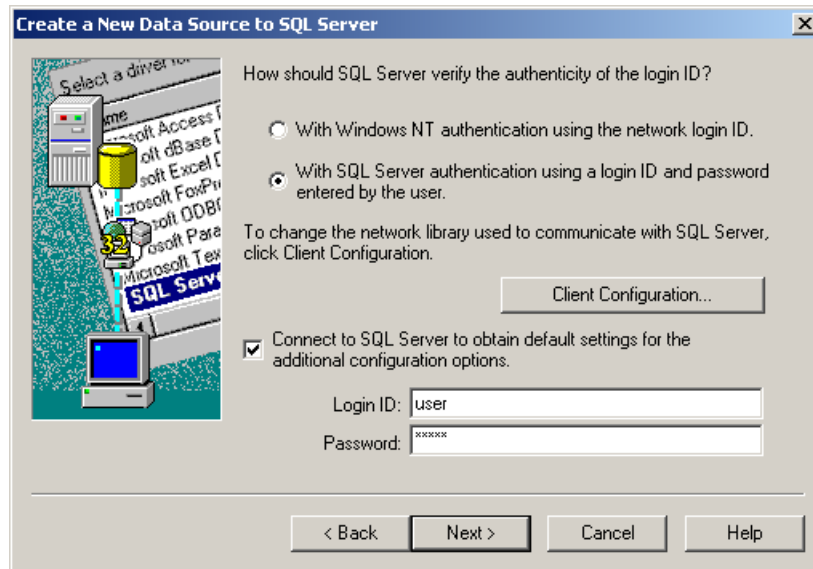
---

The first screen of the Create a New Data Source to SQL Server wizard appears.



6. Enter the name of your data source into the **Name** field. For example, type SAS.
7. (Optional) Enter the descriptive information for your database into the **Description** field. For example, type SASDemo.
8. Enter the name of the server into the **Server** field. For example, type MyServer.

- 
- Click **Next** and the next Create a New Data Source to SQL Server page appears.



- Select either **With Windows NT Authentication using the network login ID** or **With SQL Server authentication using a login ID and password entered by the user**. Your selection depends on your database server configuration.
- 

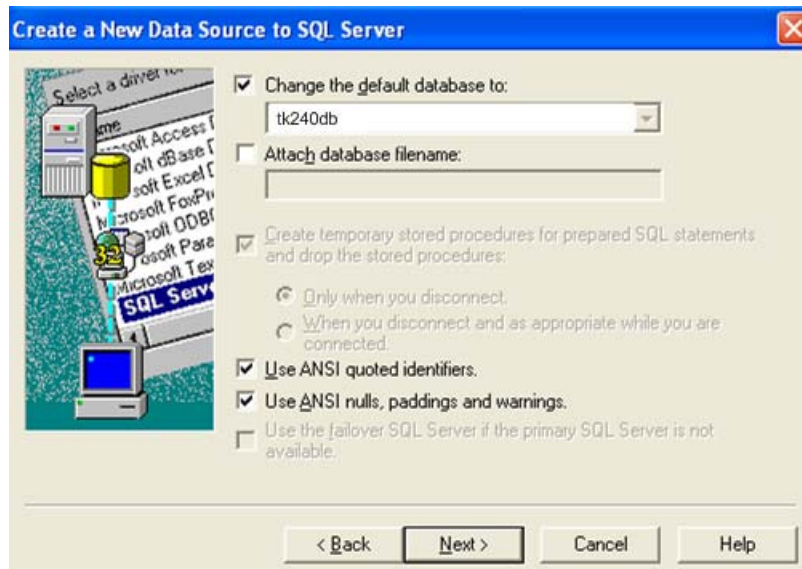
---

**Note:** Step 11 and Step 12 apply only if SQL Server authentication is selected.

---

- Enter your name into the **Login ID** field.
- Enter your password into the **Password** field.

- 
14. Click **Next** and the next Create a New Data Source to SQL Server page appears.

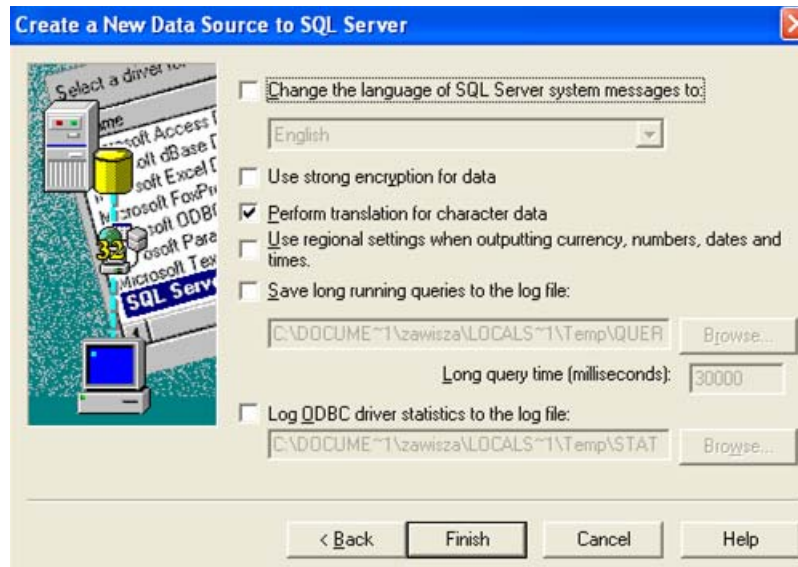


Any of the default settings are OK.

15. Click **Next**.

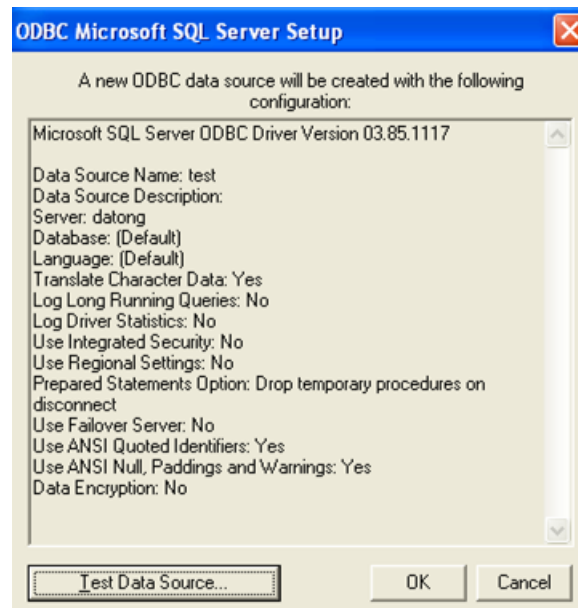


- 
16. Click **Next** and the next **Create a New Data Source to SQL Server** page appears.

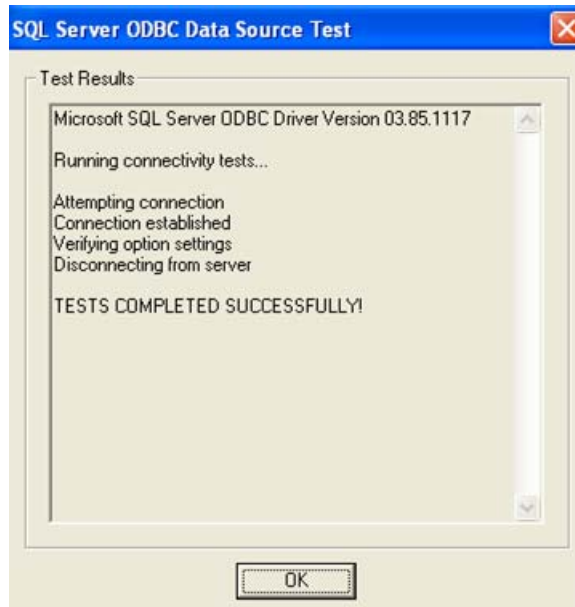


Any of the default settings are OK.

- 
17. Click **Finish**. The ODBC Microsoft SQL Server Setup page appears, displaying the configuration that you defined for the new ODBC data source.



- 
18. (Optional) Click **Test Data Source** and a window similar to the SQL Server ODBC Data Source Test window appears.



19. Click **OK**, if you see the SQL Server ODBC Data Source Test window.
20. Click **OK** in the ODBC Microsoft SQL Server Setup page.

#### 6.4.3.E Download and Configure Oracle ODBC Drivers

You can download Oracle ODBC drivers to your machine.

To download an Oracle ODBC driver, complete these steps.

1. Go to the Oracle Web site and download the Instant Client package (direct link: <http://www.oracle.com/technetwork/database/features/instant-client/index-097480.html>).
2. Choose the desired architecture in the Web page that opens from this link.
3. Click **Sign In/Register for Account** and log in.
4. Accept the license agreement.

- 
5. Download both the Instant Client – Basic and Instant Client – ODBC packages.

---

**Note:** Make sure that both client package files are located in the same directory.

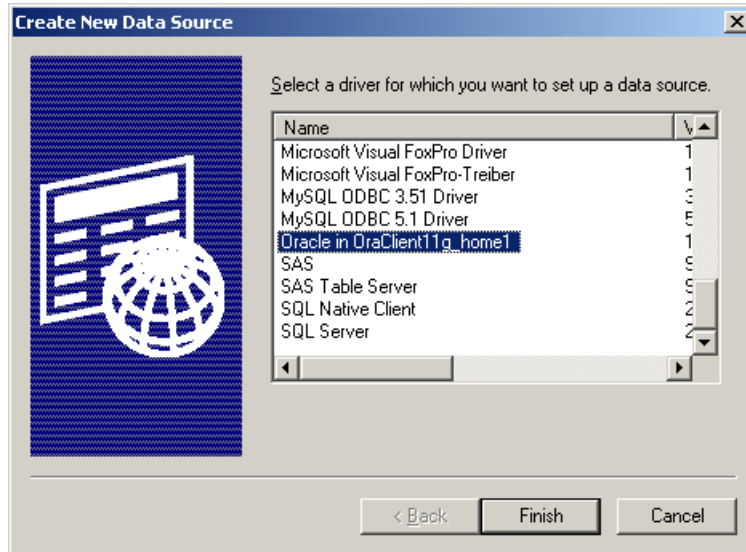
---

6. Extract the client packages.
7. Navigate to the subdirectory where the packages were extracted.
8. Run `odbc_install`. (On Windows, run `odbc_install.exe`.)

ODBC support for Oracle is configured after you run the install command.

To configure your Oracle ODBC drivers to set up your data source, complete these steps. These steps continue from Step 5 on page 85:

5. Select the ODBC driver that you installed. The exact name of the driver depends on the version of Oracle Instant Client that you install.



6. Click **Finish**.

---

The Oracle ODBC Driver Configuration window appears.

Oracle ODBC Driver Configuration

Data Source Name: SASdb

Description: SASDemo

TNS Service Name: MyOracleServer

User ID: SASUser

Application: Oracle | Workarounds | SQLServer Migration

Enable Result Sets  Enable Query Timeout  Read-Only Connection

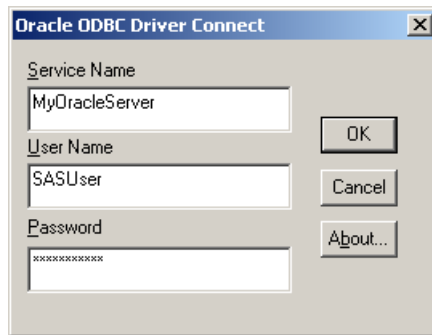
Enable Closing Cursors  Enable Thread Safety

Batch Autocommit Mode: Commit only if all statements succeed

Numeric Settings: Use Oracle NLS settings

7. Enter the name of the data source into the **Data Source Name** field. For example, type SASdb.
8. (Optional) Enter the information about your database into the **Description** field. For example, type SASDemo.
9. Enter the server name into the **TNS Service Name** field. For example, type MyOracleServer.
10. Enter the user name assigned by the database administrator into the **User ID** field. For example, type SASUser.

- 
11. Click **OK**. The Oracle ODBC Driver Connect window appears.



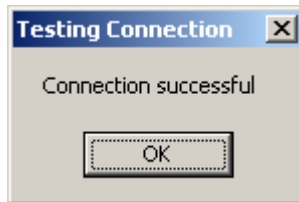
12. The fields in this window are automatically filled in for you with the exception of the **Password** field.

---

**Hints:** The **Server Name** field contains the same name as the entry in the **TNS Service Name** field.  
The **User Name** field contains the same name entered by the database administrator into the **User ID** field.  
See Step 10 on page 93.

---

13. Click **OK**. If the connection is successful, the Testing Connection window appears with this message.



14. Click **OK**.
15. Click **OK** to close the Oracle ODBC Driver Connect window.

---

## 6.5 Defining Users

After the SAS databases are configured, but before you add user accounts to a new project, define administrative and regular users. To make a new user a project administrator, assign them upload permissions for the project. For more information, see Section 6.5 *Defining Users* on page 105.

Only database administrators can perform some operations:

- Access every SAS Content Categorization Collaborative Server project on the server.
- Project administrators can also upload SAS Content Categorization Collaborative Server projects to the server.
- Add, change, and delete users in the SAS Content Categorization Collaborative Server project.

After regular users are added to a project by an administrator, their permission levels for the project are set using the SAS Content Categorization Collaborative Server interface. This permission level setting controls their access to the project. By default, a regular user has Read-Only permissions.

## 6.6 Logging In

After the connections to the remote server are set up from the local host, a user can log in from any host that they are set up to use. Three pieces of information are required to log in to the remote repository:

- the data source name
- your user name
- your password

---

## 6.7 Quick Start Guide Using \_tk240

Administrators use `_tk240` to enable the collaborative mode of SAS Content Categorization Studio.

To use `_tk240`, complete these steps:

1. You, as the database administrator, initialize the SAS Content Categorization Studio database using the following command line example:

```
_tk240 -initialize_database
      -driver <datasource_1>
      -user <sam>
      -pass <admin12>
      -verbose
```

2. You, as the database administrator, add one or more project administrators to the project:

```
_tk240 -add_user
      -driver <datasource_2>
      -user <sam>
      -pass <admin12>
      -username <pam>
      -password <pam34>
      -regular
      -can_upload
      -verbose
```

---

**Notes:** Project administrators are created by specifying both `-regular` and `-can_upload`. If you use SQL server with Windows Authentication, replace `-user` and `-pass` with `-windows_authentication` in this step and all of the following steps. Replace with `-login_name <DOMAIN\username>` in Step 2 above and Step 4 on page 107.

---

The user name of a new project administrator, who has upload privileges, is specified using the `<username>` parameter. `-password` specifies the password that the user enters in order to log onto the SAS Content Categorization Collaborative Server database server. For example, `pam` uses the password `pam34` to connect to the database.



- 
3. (Optional) Continue using Step 2 above, reiteratively, until all of the project administrators are added to your project.
  4. You, as the database administrator, add regular users to the project.

```
_tk240 -add_user -driver <datasource_2>  
-user <sam>  
-pass <admin12>  
-username <tom>  
-password <tom22>  
-regular  
-verbose
```

The user name of the new regular user is specified using the `<username>` parameter. In this example, `tom` enters the password `tom22` to connect to the SAS Content Categorization Collaborative Server database. In this example, `tom` uses this password `tom22` to access the project.

This user can access a project on the server as a regular user and download that project. This user cannot upload a project to the server, because `-upload` is not specified. If upload permissions are specified, this regular user becomes a project administrator.

5. (Optional) Continue using Step 4 above, reiteratively, until you have added all of the regular users to your project.



---

## Part 5: Appendixes

---

- Appendix A: *Troubleshooting on page 111*
- Appendix B: *Recommended Reading on page 117*
- Appendix C: *Glossary on page 119*



---

# Appendix: A

## Troubleshooting

---

- *Installing SAS Enterprise Content Categorization Servers*
- *Tips and Guidelines for SAS Content Categorization Server*
- *Using the Configuration File*
- *If SAS Content Categorization Server Does Not Appear to Be Running*
- *Tips and Guidelines for SAS Content Categorization Collaborative Server*

### A.1 Installing SAS Enterprise Content Categorization Servers

#### A.1.1 The Installer Fails to Copy Files on Windows Vista or Server 2008

If the SAS Enterprise Content Categorization Servers installers fails to copy files to the program directory on Windows Vista or Server 2008, the user account control might be enabled. Use the following steps even if you are signed in with administrative privileges.

In order to enable the SAS Enterprise Content Categorization Servers installer to function correctly, complete these steps:

1. Right-click on the `SAS_ConCat_Collab_Servers_<arch>_Setup.exe` file.
2. Select **Run as...**
3. Enter the administrator information in the Run As window that appears.
4. Click **OK** to close the Run As window.

---

## A.2 Tips and Guidelines for SAS Content Categorization Server

### A.2.1 Trying to Upload a Project with a Duplicate Name

If you try to upload a project to SAS Content Categorization Server with the same name as a project that has already been uploaded, the existing project is overwritten. This statement is true even when the type of project is different. For example, a `MyProj.desc` file for a `.mco` file can be overwritten by a `MyProj.desc` file for a `.concepts` file.

### A.2.2 Noun Phrases Incorrectly Split by an Apostrophe (')

Some noun phrases are incorrectly split by an apostrophe. This is a known issue.

### A.2.3 Using Synonym Lists

If you upload a project containing synonyms, the server cannot process documents that are larger than 1 MB in size.

Synonym lists cannot be used with SAS Content Categorization Collaborative Server.

---

## A.3 Using the Configuration File

### A.3.1 Trailing Spaces in the Configuration File

If you edit the server configuration file and there is a space at the end of a line that specifies the uploaded filename, the server might not deploy this model. Check all lines and remove all spaces in the configuration file if you find a trailing space.

### A.3.2 Modifying the Configuration File

If you installed SAS Content Categorization Server prior to the installation that you now have loaded on your machine, a backup version of the server configuration file named `server.config.bak` exists. This backup file overwrites the latest configuration file. For this reason, you might want to manually edit your configuration file. This is especially true if you have a new `setinit` file, because the old configuration file might refer to a previous version of the `setinit` file. For more information, see Chapter 4: *Configuring and Running SAS Content Categorization Server*.

If you specify `-verbose`, the server emits a warning for duplicate or conflicting entries.

---

## A.4 If SAS Content Categorization Server Does Not Appear to Be Running

### A.4.1 Overview of When the Server Does Not Appear to Be Running

If SAS Content Categorization Server does not appear to be running on your machine, access the SAS Content Categorization Server Administration Web Page. You can access this page using the Start menu, see Section 2.5 *Access the Servers* on page 33. Alternatively, open a Web browser and type:

```
http://<machine_name>:<admin_port>/admin
```

For example, if the server is running on your local machine, and you selected the default admin port of 6501 during the installation, type:

```
http://localhost:6501/admin
```

Click the [Are you there?](#) link in the SAS Content Categorization Server Administration Web Page to see whether the server is running.

### A.4.2 Checking and Debugging on a Windows Machine

On Windows, the server output is available when you go to **Start --> Control Panel --> Administrative Tools --> Windows Event Viewer**. SAS Content Categorization Server writes data to the Application log. If you can identify the issue in the server output, start a debugging instance of the server to obtain further information.

To start a debugging instance of the server, complete these steps:

1. Shut down the server using the Services window. For more information, see Section 2.3 *Starting and Stopping the Servers on a Windows Machine* on page 29.
2. Open a Windows command prompt window.
3. Enter the following command using the installation path:

```
cd <INSTALL_DIR>\Teragram Catcon Server\  
_catcon_server.exe -server conf\server.config -verbose
```



---

## A.4.3 Checking and Debugging on a UNIX Machine

You can shut down the server and start a debugging instance of the server.

To start a debugging instance, complete these steps:

1. Access a shell window.
2. Enter the following commands:

```
cd /path/to/sas_cc_servers/cc_server
./bin/<arch>/_catcon_server -server conf/server.config
                             -verbose
```

---

**Hint:** Replace <arch> with your system's architecture such as linux64.

---

## A.5 Tips and Guidelines for SAS Content Categorization Collaborative Server

### A.5.1 Working with the ODBC Drivers

#### A.5.1.A Receiving Error Messages on Windows

If you are getting ODBC errors, generate an ODBC trace log.

To generate an ODBC trace log, complete these steps:

1. Go to **Start --> Settings --> Control Panel --> Administrative Tools --> Data Sources (ODBC)**.
2. Click **Tracing**.
3. Click **Start Tracing Now**.
4. Click **OK**.

---

### A.5.1.B Receiving Error Messages on UNIX

There are several errors, or known issues, for ODBC drivers and UNIX machines:

- If you are getting ODBC errors, add the following lines to the ODBC data source name information in the `odbc.ini` file:

```
TraceFile = /tmp/odbc.trace
Trace = 1
```
- If you are getting an error message that states *Specified driver could not be loaded*, when using MySQL ODBC driver 5.1 on a Linux machine, use version 3.51 instead.
- If there is a `_tk240` command segmentation fault on a 64-bit UNIX machine with MySQL ODBC driver 5.1, use version 3.51 instead.

### A.5.1.C Adding a User with `_tk240` on Linux

If the command reports an error, the `odbc.ini` file might not be writable. Make this file writable and retry the operation.

### A.5.1.D Connecting to the Database and SQL Syntax

A status message that indicates an error with the SQL syntax and *NULL like tk\_lang\_proj* message might occur when you try to connect to the database. In this case, set the default database to your ODBC data source. For example, reset the database to `tk240db`.

---

# Appendix: B

## Recommended Reading

---

The following books are recommended:

- *SAS Content Categorization Single User Servers: Administrator's Guide*: Install, configure, and use SAS Enterprise Content Categorization Servers and SAS Document Conversion Server.
- *SAS Enterprise Content Categorization Studio: User's Guide*: Deploy the collaborative operations and develop complex LITI concepts and their rules for SAS Content Categorization Studio. Use this book after you read *SAS Content Categorization Studio: User's Guide*.
- *SAS Content Categorization Studio: Installation Guide*: Install SAS Content Categorization Studio.
- *SAS Content Categorization Studio: User's Guide*: Create a SAS Content Categorization Studio project, test, and upload the output to SAS Enterprise Content Categorization Servers.
- *SAS Enterprise Content Categorization Studio: Administrator's Guide*: Install and configure the server used for the collaborative operations available in SAS Enterprise Content Categorization Studio.

SAS offers instructor-led training and self-paced e-learning courses to help you get started with the SAS add-in, learn how the SAS add-in works with the other products in the SAS Enterprise Intelligence Platform, and learn how to run stored processes in the SAS add-in. For more information about the courses available, see [support.sas.com/training](http://support.sas.com/training).

For a complete list of SAS publications, see the current SAS Publishing Catalog. To order the most current publications or to receive a free copy of the catalog, contact a SAS representative at

SAS Publishing Sales  
SAS Campus Drive  
Cary, NC 27513  
Telephone: (800) 727-3228\*  
Fax: (919) 677-8166  
E-mail: [sasbook@sas.com](mailto:sasbook@sas.com)

---

Web address: [support.sas.com/pubs](http://support.sas.com/pubs)

\* For other SAS Institute business, call (919) 677-8000.

Customers outside the United States should contact their local SAS office.

---

# Appendix: C

## Glossary

---

**Administration Web Page**

moderate the progress and operations of SAS Content Categorization Server.

**categorization**

process of concisely defining the subject matter of a document, in other words, the main idea or subject of the document.

**definition**

defines a concept. Sometimes, this manual uses the word *rule* as a synonym for the word *definition*.

**document**

refers to an input text. Also see *Text*.

**plain text**

readable textual material that does not require much processing.

**rule**

defines a category. This term is also used, within this manual, to refer to a concept definition.

**string**

is a group of words or characters that you specify for a rule.

**text**

form that a written document, or a Web page takes, can be called a text. Also see *Document*.



---

# Index

---

#	
defined	72, 77
_tk240 utility	
defined	86
tk240_commandline.zip	86

## A

admin_port	
defined	47
Administration Web page	
view	64
administrative permission	
defined	105
Are you there?	
defined	65
Average (sec)	
defined	72

## B

backupdir	
defined	46
basedir	
defined	46

## C

cat_log	
defined	48
statistics	63
cat_log_max_entries	
defined	49
example	58
Categories returned	
defined	73

---

Categorization project	
defined .....	72
category	
rules .....	7
category matches	
log file .....	48
Category Name	
defined .....	70
comment character	
configuration file .....	44
concept	
definitions .....	7
concept extraction project	
defined .....	77
concept Name	
defined .....	75
concept_log	
defined .....	49
statistics .....	63
concept_log_max_entries	
defined .....	49
concepts	
defined .....	50
Concepts returned	
defined .....	77
configuration file	
comment character .....	44
restart program .....	49
UNIX .....	45
Windows .....	45
configure ODBC data sources	
ODBC.ini file .....	89
Connector/ODBC window	
usage .....	93
CPU .....	14
Create a New Data Source to SQL Server window	
open .....	96, 97, 98
create data sources	
credentials .....	88
Create New Data Source window	
open .....	92
usage .....	95



---

create_dir	
defined	47
creator	
defined	48
credentials	
create data sources	88

## D

data source name	
repository login	105
specify	93
Database parameter	
defined	90
Default Category Bias	
defined	68
Default Relevancy Cutoff	
defined	68
definitions	
concept	7
descriptor_dir	
defined	46
do_cat_log_timing	
defined	49
do_concept_log_timing	
defined	49
Driver parameter	
defined	90

## E

entries	
maximum exceeded	49
entry	
defined	57
ephemeral ports	
defined	62
exceed	
log file limits	58

---

## F

File Name	
defined .....	68

## I

I/O	
defined .....	72
Idle	
defined .....	72
installation	
wizard .....	17
installation guide	
UNIX .....	27
installation kit contents .....	13
installation root directory .....	60
io_log	
defined .....	48

## L

liti	
defined .....	50
log file	
exceed limits .....	58
regenerate .....	58
log in	
information .....	105

## M

max_doc_size	
defined .....	47
max_iterations_to_reinitialize	
defined .....	47
Maximum (sec)	
defined .....	72
mcat	
defined .....	50

---

memory	
project files .....	58
Microsoft SQL Server	
restart .....	86
Minimum (sec)	
defined .....	72
multiple project files	
specify .....	58
my.cnf file	
default setting .....	85
max packet size .....	85
my-huge.ini	
location .....	85
MySQL	
install .....	84

## N

Name	
defined .....	72, 77
nb_threads	
defined .....	47
num_cat_logs	
defined .....	49
num_concept_logs	
defined .....	49
Number of Rules	
defined .....	68

## O

ODBC Data Source Administrator window	
open .....	91
ODBC data sources	
3 types .....	87
ODBC drivers	
install .....	87
ODBC Microsoft SQL Server Setup window	
open .....	100, 101

---

odbc.ini file	
usage .....	89
operating systems	
supported .....	15

## P

password	
repository login .....	105
Percentage (relative)	
defined .....	70, 75
Percentage of Documents	
defined .....	70, 75
persistent_connection	
defined .....	47
Port parameter	
defined .....	90
Processing	
defined .....	72
protocol_version	
defined .....	49

## Q

query_port	
defined .....	47

## R

RAM .....	14
repository	
remote .....	105
restart program	
configuration file .....	49
rules	
category .....	7

---

## S

SAS Content Categorization Server	
running on Windows .....	59
UNIX shell command .....	60
-server configfile	
defined .....	60
server configuration	
change data size .....	85
Server parameter	
defined .....	90
setinit	
defined .....	46
Size (bytes)	
defined .....	72, 77
skt_queue_size	
defined .....	47
stat_cat	
defined .....	50
statistics	
cat_log .....	63
concept_log .....	63
generate .....	63
Status	
defined .....	65
Symbolic Name	
defined .....	68
system configuration	
specified .....	14

## T

Test Data Source button	
usage .....	101
Time (relative)	
defined .....	73, 77
Time (sec)	
defined .....	77
Time taken	
defined .....	72

---

timeout	
defined	47
tk240_commandline.zip	
_tk240 utility	86
Total number of categories with at least one match	
defined	70
Total number of concept types with at least one match	
defined	74
Total number of documents	
defined	70, 74
Type	
defined	68

## U

UNIX	
configuration file	45
installation guide	27
UNIX shell command	
running the server	60
upload	
project	105
Uptime	
defined	66
user	
defined	48
username	
repository login	105

## W

Web interface	
usage	63
Windows	
configuration file	45
installation guide	16