# IBM SPSS Text Analytics Administration Console 15 User's Guide

*Note*: Before using this information and the product it supports, read the general information under Notices on p. 13.

This edition applies to IBM® SPSS® Modeler Text Analytics 15 and to all subsequent releases and modifications until otherwise indicated in new editions.

# Preface

IBM® SPSS® Modeler Text Analytics offers powerful text analytic capabilities, which use advanced linguistic technologies and Natural Language Processing (NLP) to rapidly process a large variety of unstructured text data and, from this text, extract and organize the key concepts. Furthermore, SPSS Modeler Text Analytics can group these concepts into categories.

Around 80% of data held within an organization is in the form of text documents—for example, reports, Web pages, e-mails, and call center notes. Text is a key factor in enabling an organization to gain a better understanding of their customers' behavior. A system that incorporates NLP can intelligently extract concepts, including compound phrases. Moreover, knowledge of the underlying language allows classification of terms into related groups, such as products, organizations, or people, using meaning and context. As a result, you can quickly determine the relevance of the information to your needs. These extracted concepts and categories can be combined with existing structured data, such as demographics, and applied to modeling in IBM® SPSS® Modeler's full suite of data mining tools to yield better and more-focused decisions.

Linguistic systems are knowledge sensitive—the more information contained in their dictionaries, the higher the quality of the results. SPSS Modeler Text Analytics is delivered with a set of linguistic resources, such as dictionaries for terms and synonyms, libraries, and templates. This product further allows you to develop and refine these linguistic resources to your context. Fine-tuning of the linguistic resources is often an iterative process and is necessary for accurate concept retrieval and categorization. Custom templates, libraries, and dictionaries for specific domains, such as CRM and genomics, are also included.

## About IBM Business Analytics

IBM Business Analytics software delivers complete, consistent and accurate information that decision-makers trust to improve business performance. A comprehensive portfolio of business intelligence, predictive analytics, financial performance and strategy management, and analytic applications provides clear, immediate and actionable insights into current performance and the ability to predict future outcomes. Combined with rich industry solutions, proven practices and professional services, organizations of every size can drive the highest productivity, confidently automate decisions and deliver better results.

As part of this portfolio, IBM SPSS Predictive Analytics software helps organizations predict future events and proactively act upon that insight to drive better business outcomes. Commercial, government and academic customers worldwide rely on IBM SPSS technology as a competitive advantage in attracting, retaining and growing customers, while reducing fraud and mitigating risk. By incorporating IBM SPSS software into their daily operations, organizations become predictive enterprises – able to direct and automate decisions to meet business goals and achieve measurable competitive advantage. For further information or to reach a representative visit *http://www.ibm.com/spss*.

## Technical support

Technical support is available to maintenance customers. Customers may contact Technical Support for assistance in using IBM Corp. products or for installation help for one of the supported hardware environments. To reach Technical Support, see the IBM Corp. web site

at *http://www.ibm.com/support*. Be prepared to identify yourself, your organization, and your support agreement when requesting assistance.

Chapter
1

# Getting started

## Administered servers

Server administration in IBM® SPSS® Collaboration and Deployment Services Deployment Manager involves:

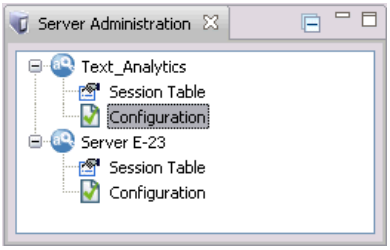
1. Adding the server to be administered to the system.
2. Logging in to the server being administered.
3. Performing administrative tasks for the server as needed.
4. Logging off from the server being administered.

The Server Administration tab offers access to this functionality. This tab lists the servers currently available to be administered. This list persists across Deployment Manager sessions, facilitating access to those servers.

From the menus choose:

Tools > Server Administration

Figure 1-1
*Administered server list*




The administered server list may include a variety of server types, including IBM® SPSS® Collaboration and Deployment Services Repository servers, IBM® SPSS® Modeler servers, and IBM® SPSS® Statistics servers. The actual administrative functionality available for a server depends on the server type. For example, security providers can be configured and enabled for repository servers but not for SPSS Modeler servers.

## Adding new administered servers

Before performing administrative tasks, a connection to the administered server must be established.

From the menus choose:

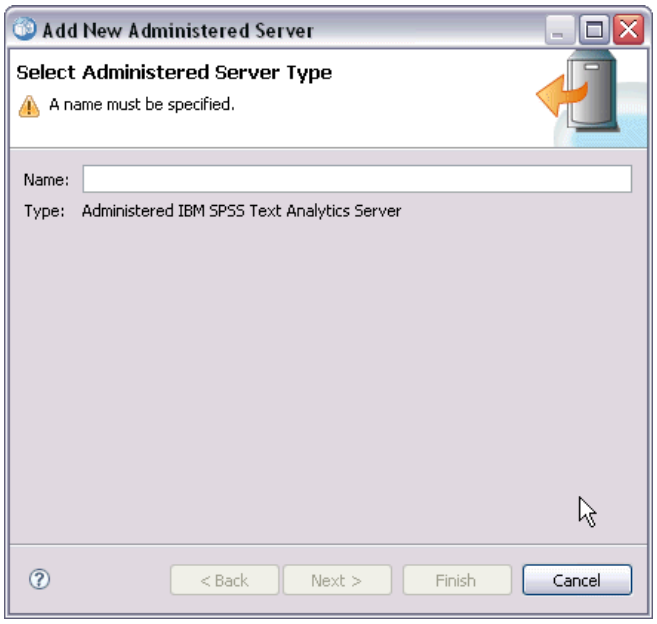
File > New > Administered Server Connection

The Add New Administered Server dialog box opens. Adding a new connection requires the specification of the administered server type and the administered security server information.

## Selecting the administered server name and type

The first step of adding a new administered server to the system involves the definition of two parameters for the server—the name and the type.

Figure 1-2
*Select Administered Server Type dialog box*




**Name.** A label used to identify the server on the Server Administration tab. Including the port number in the name, such as *my_server:8080*, may help to identify the server in the administered server list.

*Note*: Alphanumeric characters are recommended. The following symbols are prohibited:

- Quotation marks (single and double)
- Ampersands (&)
- Less-than (<) and greater-than (>) symbols
- Periods
- Commas
- Semicolons

**Type.** The type of server being added. The list of possible server types depends on the system configuration and may include:

- IBM® SPSS® Collaboration and Deployment Services Repository Server
- Administered IBM® SPSS® Modeler Server

- Administered IBM® SPSS® Statistics Server
- Administered IBM® SPSS® Modeler Text Analytics Server

#### *Selecting an Administered Server Type*

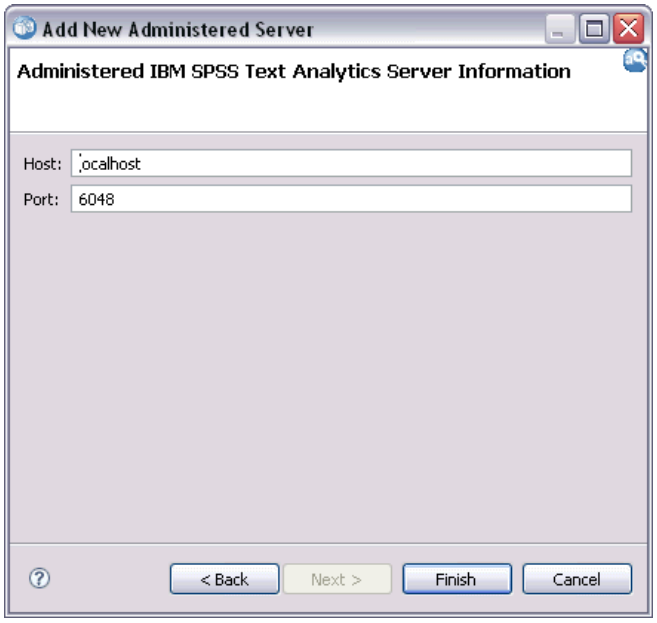
In the Select Administered Server Type dialog box:

1. Enter a name for the server.
2. Select the server type.
3. Click Next. The Administered Security Server Information dialog box opens.

## *Administered server information*

The second step of adding a new administered server to the system involves the definition of the server properties.

Figure 1-3
*Administered Security Server Information dialog box*




**Host.** The name or IP address of the server.

*Note*: Alphanumeric characters are recommended. The following symbols are prohibited:

- Quotation marks (single and double)
- Ampersands (&)
- Less-than (<) and greater-than (>) symbols
- Periods

- Commas
- Semicolons

**Port.** The port number used for the server connection.

**This is a secure port.** Enables or disables the use of a Secure Sockets Layer (SSL) for the server connection. This option is not offered for all types of administered servers.

#### *Specifying Administered Server Information*

In the Administered Security Server Information dialog box:

1. Enter the host name or IP address for the server being added.
2. Enter the port number on which the server being added is running.
3. Specify whether or not the server uses SSL, if applicable.
4. Click Finish.

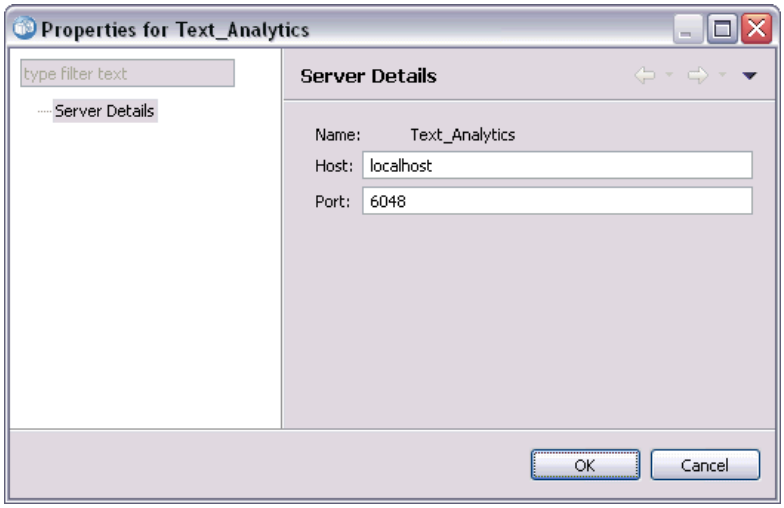
The server appears in the administered server list on the Server Administration tab.

*WARNING*: Validation of the SSL parameters modifies the axis2.xml file in the `<modeler_server_install_directory>/ext/bin/spss.TMWBServer/conf` directory on the server side. The file shows the password as readable plain text; this is a known defect in Axis2.

## Viewing and editing administered server properties

To view the properties of an existing administered server, right-click the server on the Server Administration tab and select Properties from the drop-down menu. The Properties dialog box opens. The displayed properties depend on the type of server selected.

Figure 1-4
*IBM SPSS Modeler Text Analytics Server properties*



For SPSS Modeler Text Analytics servers, the properties include:

- **Label.** The name associated with the server as it appears on the Server Administration tab.
- **Host.** The name or IP address of the server.
- **Port.** The port number used for the server connection.

Some connection properties can be modified. However, changing these properties does not change the setting on the server. The properties should be edited only if the server itself has changed in a way that requires new connection information. For example, if the server is running on a new port, the properties must be updated to successfully connect to the server.

## Connecting to administered servers

For most servers, you must connect to a server in the administered server list to perform administrative tasks. From the Server Administration tab, double-click the server to administer. The Login to Server dialog box opens.

## Disconnecting administered servers

After completing the desired administrative tasks, log off from the server.

1. On the Server Administration tab, right-click the server.
2. Select Logoff.

To administer the server, you must log in again.

## Deleting administered servers

A server appears in the administered server list until it is deleted from the list.

1. On the Server Administration tab, select the server to delete.
2. From the menus choose:
   Edit > Delete

   Alternatively, right-click the server and select Delete from the drop-down menu.

If further administrative tasks for the server are needed in the future, the server will need to be added to the system again.

Chapter
2

# Installing IBM SPSS Text Analytics Administration Console

## System Requirements for Installation

Before installing IBM® SPSS® Text Analytics Administration Console , you must first have installed IBM® SPSS® Modeler and IBM® SPSS® Modeler Text Analytics somewhere on your network or locally. Please verify the following system requirements before installing SPSS Text Analytics Administration Console .

- **Operating system.** Microsoft Windows 7 (Professional and Enterprise) x32 and x64 Editions; Microsoft Windows Vista (Business and Enterprise) with Service Pack 1 x32 and x64 Editions; Microsoft Windows XP Professional with Service Pack 3 x32 and x64 Editions.
- **Hardware.** Intel® Pentium® or Pentium-class processor or higher (for 32-bit Microsoft Windows); x64 (AMD 64 and EM64T) processor family (for 64-bit Microsoft Windows) running at 1GHz or faster. A monitor with 1024x768 resolution or higher. A disk drive is also required if you are installing from a disk.
- **Minimum free disk space.** 1 GB of available hard-disk space on Microsoft Windows machine running the SPSS Modeler/SPSS Modeler Text Analytics client.
- **Minimum RAM.** 512MB RAM minimum; 1GB recommended.
- **Software.** SPSS Modeler version 15 and SPSS Modeler Text Analytics version 15.
- **Virtual environment.** The following virtual environments support SPSS Text Analytics Administration Console :

  - Windows 2008® Terminal Services and R2
  - Windows 2003® Terminal Services and R2
  - Citrix XenApp 5 – Standard, Advanced, and Enterprise
  - Citrix Presentation Server 4.5 – Standard, Advanced, and Enterprise
  - VMWare ESX Server 3.5
  - VMWare vSphere 4.0

- **Additional requirements.** The SPSS Text Analytics Administration Console host computer must have network connectivity to the computer hosting IBM® SPSS® Modeler Text Analytics Server .

## Steps for Installing

***To install IBM SPSS Text Analytics Administration Console***

▶ From the installation autoplay menu, choose IBM® SPSS® Text Analytics Administration Console .

▶ Follow the instructions in the setup wizard to complete the installation.

***To use an existing administration console installation***

If you are already using an administrative console for other IBM Corp. server applications, you can add the IBM® SPSS® Modeler Text Analytics Server administration console to your existing installation.

- From the installation autoplay menu, choose SPSS Text Analytics Administration Console .
- In the setup wizard, choose to install in the same directory as your existing administration console installation. The setup wizard will add the new SPSS Text Analytics Administration Console files to the existing installation.

Chapter
3

# Using IBM SPSS Text Analytics Administration Console

The IBM® SPSS® Text Analytics Administration Console provides a user interface to monitor and configure your IBM® SPSS® Modeler Text Analytics Server installations. The SPSS Text Analytics Administration Console can be installed only on Microsoft Windows computers; however, it can administer SPSS Modeler Text Analytics Server s installed on any supported platform.

You can also use the SPSS Text Analytics Administration Console to manage your IBM® SPSS® Modeler Text Analytics sessions as well as to manage web server connections. Specifically, if you chose to "save and reuse" the Web Feed data or the Translate node translated text, you can delete this stored session data once you no longer need it.

If you use console for IBM® SPSS® Modeler, the session data for SPSS Modeler Text Analytics is managed in the same window. For more information on the console parameters specific to SPSS Modeler, please refer to *IBM® SPSS® Modeler Administration Console 15 User's Guide*.

## Starting IBM SPSS Text Analytics Administration Console

After you have installed and configured IBM® SPSS® Text Analytics Administration Console , the Server Administrator pane on the left displays a node for each IBM® SPSS® Modeler Text Analytics Server that you want to administer. The right-hand pane shows the configuration options for the selected server. When you first run the application, both panes are blank.

***To start IBM SPSS Text Analytics Administration Console :***

- From the Microsoft Windows Start menu, choose [All] Programs > IBM SPSS Modeler 15 > IBM® SPSS® Collaboration and Deployment Services Deployment Manager > SPSS Text Analytics Administration Console .

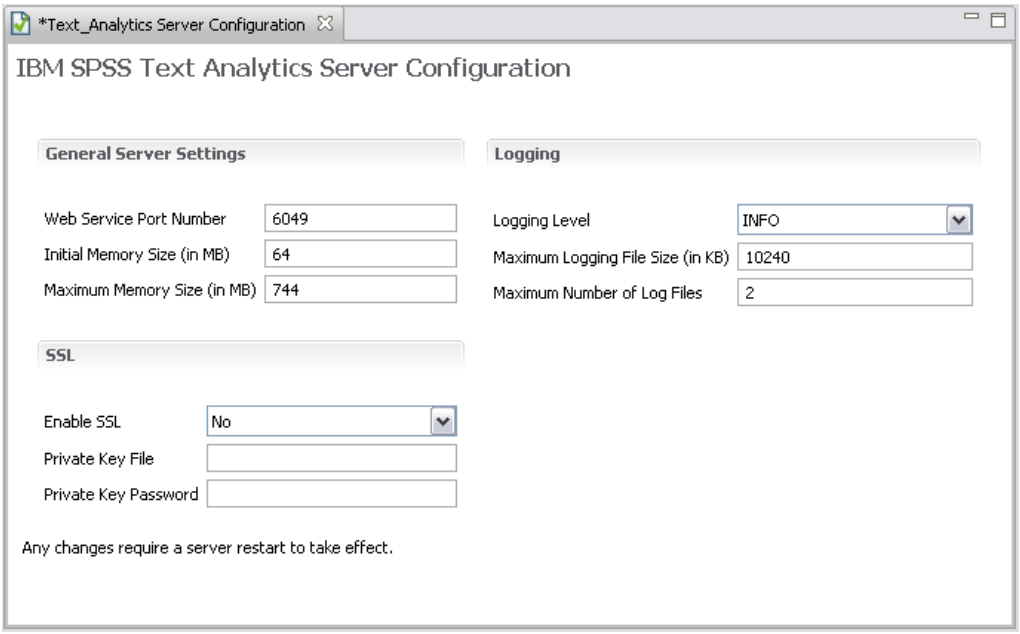
## Server Configuration Settings

The Configuration pane shows configuration options for IBM® SPSS® Modeler Text Analytics Server .

- Use this pane to change the options as desired and then choose File > Save from the menus to save the changes.

**Important!** In order for your changes to take effect, you must restart the server web service. For more information, see the topic Restarting Web Service on p. 11.

Figure 3-1
*Server Configuration*




### *General Server Settings*

| Setting | Description |
|---|---|
| Web Service Port Number | The port number for SPSS Modeler Text Analytics Server to listen on. Change if another application already uses the default. End users must know the port number in order to use SPSS Modeler Text Analytics Server . After changing and saving, you must now update the server properties with this new port number as well (Right-click the server on the Server Administration tab and select Properties from the drop-down menu.) For more information, see the topic Viewing and editing administered server properties in Chapter 1 on p. 4 . |
| Initial Memory Size (in MB) | The initial Java heap size to control the amount of memory that is allocated to JVM application. |
| Maximum Memory Size (in MB) | Maximum Java heap size. A value of –1 indicates no limit. |

### *SSL Parameters*

| Setting | Description |
|---|---|
| Enable SSL | Enables SSL encryption for connections between IBM® SPSS® Modeler Text Analytics and SPSS Modeler Text Analytics Server . |

| Setting | Description |
|---|---|
| Private key file | Enter the name of the PKCS-12 formatted keystore to be loaded when the server starts up. Enter a filename here. Make sure the file is located in the following subdirectory of the IBM® SPSS® Modeler Server installation directory:<br>■ On Microsoft Windows: *\ext\bin\spss.TMWBServer\lib*<br>■ On UNIX: */ext/bin/spss.TMWBServer/java*<br>To generate this keystore, run an openSSL command with the Certificate and Private Key files that were used to configure SSL for IBM® SPSS® Modeler. For example:<br>`openssl pkcs12 -export -in modeler_cert.pem -inkey modeler_key.pem -out pkcs-12-certificate-and-key-file`<br>■ `modeler_cert.pem` is the certificate file used with SPSS Modeler<br>■ `modeler_key.pem` is the private key file used with SPSS Modeler<br>■ `pkcs-12-certificate-and-key-file` is the generated PKCS-12 formatted keystore to be used with SPSS Modeler Text Analytics Server |
| Private key password | Password key for the specified certificate. Note this key must have been signed by at least one of the Certification Authorities listed in the configured certificate file. **Important!** If the private key file that you used with SPSS Modeler has a password, you must create the PKCS-12 keystore using the same password. However if the private key file for SPSS Modeler does not have a password, the PKCS-12 keystore for IBM® SPSS® Text Analytics Administration Console must still be created with a password. |

**Note for UNIX platforms**: The Java Cryptography Engine (JCE) shipped with some JVMs do not support all the cryptographic algorithms required by various specification used within the Identity Provider. In these cases, the BouncyCastle JCE is invoked to provide this support. To enable BouncyCastle JCE, you will need to edit the file `java.security` located in the `<modeler_server_install>/ext/bin/spss.TMWBServer/jre/lib/security` directory by adding the following line after the last `security.provider` entry: `security.provider.#=org.bouncycastle.jce.provider.BouncyCastleProvider` where `#` is replaced with a number one greater than the last provider in the list.

#### *Logging Parameters*

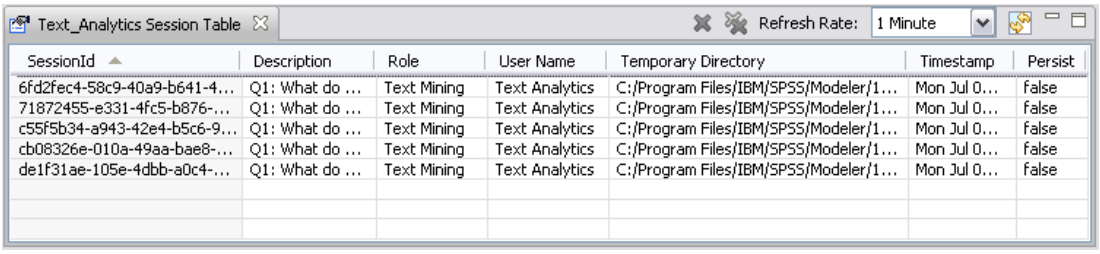
| Setting | Description |
|---|---|
| Logging Level | This indicates the level of logging that should be performed. Possible values are: NONE, DEBUG, INFO, WARN, ERROR, or FATAL. |
| Maximum Logging File Size (in KB) | This setting is the largest amount of space a log file can grow before the log file is rolled. The size is specified in kilobytes. This parameter accepts only numerical values. The default value of 0 disables log rolling. |
| Maximum Number of Log Files | This setting is the maximum number of log files to save before deleting the older ones. The default value of 0 means that there is no maximum number of logs. |

## *Server Session Table*

IBM® SPSS® Text Analytics Administration Console ’s session pane shows a snapshot of all sessions running on the IBM® SPSS® Modeler Text Analytics Server computer.

- To activate the session pane, double-click the Session node beneath the desired server in the Server Administrator pane. The data refreshes at the rate shown (one minute by default).
- To refresh the data manually, click the Refresh button.

Figure 3-2
*Session Table pane*




**SessionId.** This is the session's internal identifier.

**Description.** In the case where you have chosen to save and reuse web feeds or translation results, the label specified in the node is shown here. Otherwise a generic *Text_Mining* is shown.

**Role.** This column describes the role the session plays.

**User Name.** This is the username provided by the web server.

**Temporary Directory.** This is the directory path for the data generated by the session.

**Timestamp.** This is the date and time the session was started.

**Persist.** This indicates whether the session should be kept after a session has ended. When set to false, this indicates that the entry should be delete when a session terminates. When you have chosen to save and reuse web feeds or translation results, Persist is set to true since you want to keep this data until you choose to delete the session directly here.

# *Restarting Web Service*

Whenever you make changes to a IBM® SPSS® Modeler Text Analytics Server in the IBM® SPSS® Text Analytics Administration Console , you must restart the web service.

**To restart the Web Service on Microsoft Windows:**

- On the computer where you installed IBM® SPSS® Modeler Text Analytics , select Services from Administrative Tools on the Control Panel.
- Locate IBM SPSS Modeler Text Analytics Server 15 in the list and restart it.
- Click OK to close the dialog box.

**To restart the Web Service on UNIX:**

On UNIX, you must restart the IBM® SPSS® Modeler Server by running the modelersrv.sh script in the SPSS Modeler Server installation directory.

- Change to the SPSS Modeler Server installation directory. For example, at a UNIX command prompt, type:

  cd /usr/<modelersrv>, where `modelersrv` is the SPSS Modeler Server installation directory.

- To stop the server, at the command prompt, type

  ./modelersrv.sh stop

- To restart the server, at the command prompt, type

  ./modelersrv.sh start

Appendix

# A

# Notices

This information was developed for products and services offered worldwide.

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

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

*IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, U.S.A.*

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

*Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japan.*

**The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:** INTERNATIONAL BUSINESS MACHINES PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

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

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

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

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

*IBM Software Group, Attention: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.*

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

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

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

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

### *Trademarks*

IBM, the IBM logo, ibm.com, and SPSS are trademarks of IBM Corporation, registered in many jurisdictions worldwide. A current list of IBM trademarks is available on the Web at *http://www.ibm.com/legal/copytrade.shtml*.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

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

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

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