

# IBM SPSS Statistics Server Administrator's Guide



*Note:* Before using this information and the product it supports, read the general information under Notices on p. 65.

This document contains proprietary information of SPSS Inc, an IBM Company. It is provided under a license agreement and is protected by copyright law. The information contained in this publication does not include any product warranties, and any statements provided in this manual should not be interpreted as such.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© **Copyright SPSS Inc. 1989, 2010.**

---

# Contents

<b>1</b>	<b>Overview</b>	<b>1</b>
	Products and Operating Systems . . . . .	1
	Architecture . . . . .	1
	Software Components . . . . .	3
	Using Distributed Mode . . . . .	4
	Administering Analytic Server Software . . . . .	4
	Using This Document . . . . .	5
<b>2</b>	<b>Installation</b>	<b>6</b>
	Installing the Analytic Server Software . . . . .	6
	Installing the Client Application. . . . .	6
	Installing Other Analytic Server Software . . . . .	7
<b>3</b>	<b>Data Access</b>	<b>8</b>
	View of the Data . . . . .	8
	Data Access Technology . . . . .	8
	Connect ODBC. . . . .	8
	Accessing Data . . . . .	9
	Referencing Data. . . . .	10
	Controlling Data Access . . . . .	10
	Data Sources. . . . .	11
	Configuring the UNIX Environment for Data Access. . . . .	11
<b>4</b>	<b>Configuring, Monitoring Usage, and Maintenance</b>	<b>14</b>
	Managing End-User Accounts and Files . . . . .	14
	Accounts. . . . .	14
	Data Access . . . . .	15
	Files. . . . .	15
	Profiles . . . . .	15

Configuring ODBC Data Sources . . . . .	15
ODBC Data Sources and IBM SPSS Data Access Pack . . . . .	15
Configuring OLE DB Data Sources . . . . .	16
Using a Third-Party Sort Engine . . . . .	16
Using the Administration Application . . . . .	17
Administered servers. . . . .	17
Adding new administered servers . . . . .	18
Viewing and editing administered server properties. . . . .	21
Connecting to administered servers. . . . .	22
Disconnecting administered servers . . . . .	22
Deleting administered servers . . . . .	23
IBM SPSS Statistics Server Connections . . . . .	23
Controlling the IBM SPSS Statistics Server . . . . .	23
IBM SPSS Statistics Server Configuration . . . . .	24
IBM SPSS Statistics Server User Profiles and Groups . . . . .	29
Monitoring IBM SPSS Statistics Server Users . . . . .	33
Controlling Service Startup. . . . .	34
Startup Script Command Line Parameters . . . . .	34
Other Maintenance. . . . .	35
Starting and Stopping the Analytic Server Software . . . . .	35
To Start the Service or Daemon . . . . .	35
To Stop the Service or Daemon . . . . .	35
Configuration for Improving Performance . . . . .	36

## **5 Supporting End Users 37**

Connecting to the Server Software . . . . .	37
Permissions. . . . .	38
Profiles . . . . .	38
Client and Server Versions. . . . .	39
Connecting Users through a Firewall . . . . .	39
Connecting Users with PPTP . . . . .	40
Using SSL to secure data transfer . . . . .	41
Other Authentication Options on UNIX. . . . .	45
Setting a Locale. . . . .	47
Accessing Data and Files . . . . .	49
Saving Data and Files . . . . .	49

**6 Analyzing and Improving Performance 50**

Obtaining Performance Information ..... 50  
Improving Disk Usage ..... 51  
Improving CPU Usage ..... 52  
Improving Memory Usage ..... 53  
Improving Network Usage ..... 53  
Using IBM SPSS Statistics Efficiently ..... 53

**Appendices**

**A Troubleshooting 54**

Analytic Server Software ..... 54  
Client Software ..... 54

**B The IBM SPSS Statistics Batch Facility 56**

What You Need to Know ..... 56

**C Windows Operating System Tasks 58**

File Properties ..... 58  
System Properties ..... 59  
User Manager ..... 59  
Services Control Panel ..... 59  
Task Manager ..... 60  
ODBC Administrator ..... 60  
    To Configure a System DSN ..... 60  
    To Configure a User DSN ..... 61

<b><i>D</i></b>	<b><i>UNIX Operating System Tasks</i></b>	<b>62</b>
	chmod .....	62
	env .....	62
	Scripts .....	63
	ps and kill .....	63
	odbc.ini .....	64
<b><i>E</i></b>	<b><i>Notices</i></b>	<b>65</b>
	<b><i>Glossary</i></b>	<b>67</b>
	<b><i>Index</i></b>	<b>69</b>

# Overview

The IBM® SPSS® Statistics analytic server technology is a **distributed architecture**, and coupled with key data management optimizations, it supports scalable analysis. The technology is client/server based. It distributes client requests for resource-intensive operations to powerful analytic server software. When the client and server work together like this, it is referred to as **distributed analysis mode**. Distributed analysis allows end users to perform analyses that their desktop computers cannot support.

For maximum flexibility, client applications that use the analytic server technology can also be configured to run solely on the end user's desktop computer—this is referred to as **local analysis mode**. End users can easily switch modes.

## ***Products and Operating Systems***

The analytic server technology supports the IBM® SPSS® Statistics client application, and the server software runs on several operating systems (see the installation instructions for specifics). You can install multiple versions of analytic server software at your site, on the same server computer, or on different server computers.

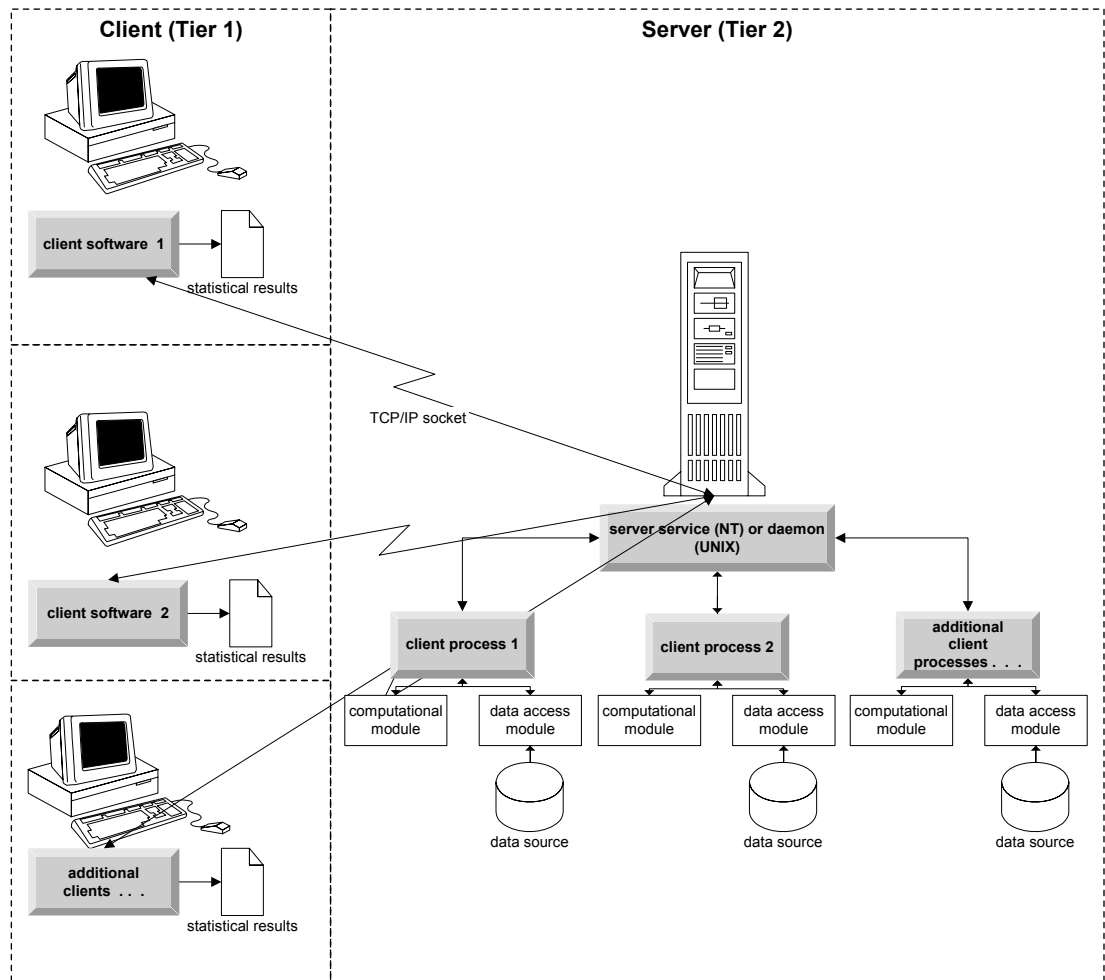
## ***Architecture***

The analytic server software has a two-tier, distributed architecture. It distributes software operations between the client and the server computers. Memory-intensive operations, such as accessing a large database or analyzing a large data file, are done on the server computer without downloading the data to the client computer.

**Tier1.** The **client** application. It is installed and runs on the end user's desktop computer. The client application provides the graphical user interface to data access and analysis. It presents the results of the end user's analyses.

**Tier2.** The **analytic server** software. It is installed and runs on a networked server computer. The server software provides the framework necessary to handle multiple clients, the algorithms used in statistical analysis, and data access.

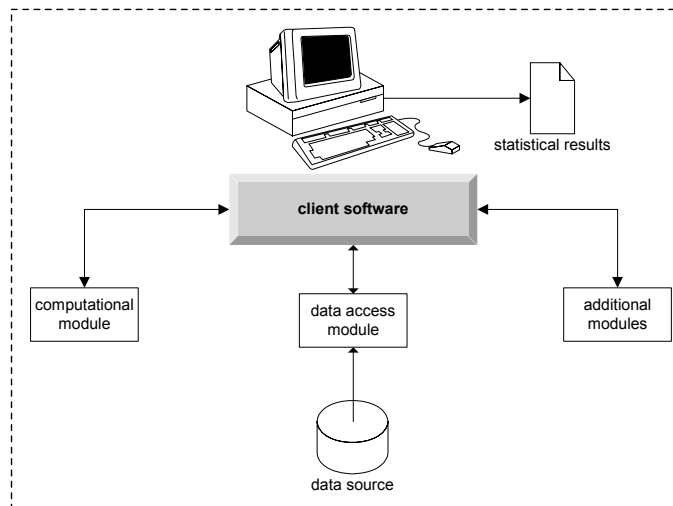
Figure 1-1  
Distributed analysis mode



For analyses that don't require intensive data access or numeric processing, the client software can be used as a standard standalone desktop application. When in local analysis mode, all data access and statistical processing are handled on the end user's desktop computer.



Figure 1-2  
Local analysis mode



## Software Components

As shown in the figure “Distributed analysis mode,” the components of the analytic server technology are the following. (See [Figure 1-1](#) on p. 2.)

**Client software.** The client application is a complete installation of the end user’s SPSS Inc. product (e.g., IBM® SPSS® Statistics). When it is used for distributed analysis, only the graphical user interface and the editing capabilities are used. When it is used for local analysis, its data access and statistical processing capabilities are also used.

**Analytic server software.** The analytic server software is made up of sub-components: a framework that manages client/server communication, client processes that manage client requests, and modules that access data and perform analyses.

- **Framework.** The framework of analytic server software is a service (on Windows) or daemon (on UNIX). It handles all communication between the client application and the modules. The framework runs continually on the server computer, waiting for client connections. When a client connects, the framework launches a process that handles requests for that client.
- **Client process.** A client process is effectively a session for the client. There is one process for each client. A process is launched when the client connects, and it is terminated when the client disconnects. The process manages its client’s requests for data and analyses. It loads the modules that are needed to access and analyze data. It unloads modules when they are no longer required.
- **Modules.** A module is an executable, DLL, or shared library that accesses data and runs analytic procedures. The analytical server software has several modules. Modules are loaded on demand. Modules may load other modules.

## **Using Distributed Mode**

The following steps occur when an end user runs a product in distributed analysis mode:

- ▶ **Launch the client application.** The end user launches the client software on his or her desktop computer. The client application presents a complete user interface.
- ▶ **Connect to the server.** The end user connects to the analytic server software by logging in from the client application. The server framework's service or daemon is always running, waiting for connection requests. When a connection is made, the server software launches a process to handle the end user.
- ▶ **Access data.** The end user accesses data as usual from the client application, except that his or her view of database drivers, data files, directories, and drives represents the remote server computer, not the desktop computer. The server process loads the appropriate data access modules and retrieves the data. A small segment of the data is sent to the client application so that the end user can refer to it when selecting an analysis. Most of the data remain on the server. You can also configure the analytic server software to prevent any data from being sent to the client application. You can use the administration application (IBM® SPSS® Statistics Administration Console) to prevent the data from being sent to all clients. For more information, see the topic [Users](#) in Chapter 4 on p. 25. You can also configure access for each user or group. For more information, see the topic [IBM SPSS Statistics Server User Profiles and Groups](#) in Chapter 4 on p. 29.
- ▶ **Analyze the data.** Using the client application's user interface, the end user selects the data and requests the type of analysis he or she wants. The request is sent to the server process, which loads the appropriate data analysis packages and processes the analysis. All data-related tasks, such as reading data, transforming data, computing new variables, and calculating statistics, are done on the server computer.
- ▶ **Review the results.** The server software sends the output from the client's request back to the client application. Only the results are sent, the data remain on the server. The end user can then use the client application to refine and edit the results.

## **Administering Analytic Server Software**

This guide is intended primarily for system administrators who are responsible for integrating analytic server technology into a networked environment in which client applications are run in distributed analysis mode. Administrative tasks include:

**Installation.** The analytic server software is designed to run continuously and respond to logins and requests from end-user desktop computers. Select an appropriate server computer for the analytic server software—one that has little downtime, is configured for end-user access, and is networked to the appropriate desktop computers. The more memory and processing power the server computer has, the faster client requests are handled. The client application must be installed on the end user desktop computers. Client installation can be done from a network location. [Chapter 2, Installation](#), provides an overview of how to install the analytic server software and the client application. Detailed installation instructions are included on the product DVD.

**Data access.** If you need to provide end users access to data on a remote server while they are working in distributed mode, the analytic server software needs to be able to access that data. SPSS Inc. products can access data from a variety of data file types, including databases. To make your job easier, SPSS Inc. products are distributed with DataDirect Connect ODBC for accessing data from a database. [Chapter 3, \*Data Access\*](#), introduces data access for SPSS Inc. products. Additional documentation is included on the product DVD.

**Configuration and maintenance.** Because the analytic server software is intended for continuous operation, it should be monitored at regular intervals by a system administrator. There are several configuration options that give you control over how the server software operates. [Chapter 4, \*Configuring, Monitoring Usage, and Maintenance\*](#), discusses configuring and monitoring the analytic server software.

**Supporting end users.** End users require information about server names, user accounts, and where to find data. You may also need to assist them in solving problems. [Chapter 5, \*Supporting End Users\*](#), discusses the kind of support that end users require.

**Performance.** [Chapter 6, \*Analyzing and Improving Performance\*](#), provides strategies for improving the performance of the analytic server software.

**Troubleshooting.** [Appendix A, \*Troubleshooting\*](#), provides troubleshooting tips.

**SPSS Statistics Batch Facility (SPSS Statistics Server only).** The IBM® SPSS® Statistics Server product includes SPSS Statistics Batch Facility, which is intended for automated production of statistical reports. If you are running SPSS Statistics Server at your site, read [Appendix B, \*The IBM SPSS Statistics Batch Facility\*](#), which describes the SPSS Statistics Batch Facility and the tasks that you may need to perform to support it.

## ***Using This Document***

This guide is intended primarily for system administrators who are responsible for installing and maintaining the analytic server software in a networked environment in which client applications are run in distributed analysis mode.

# ***Installation***

Products that use the analytic server technology are packaged on multiple media—one DVD for the analytic server software and one DVD for the client application.

To deploy the analytic server technology, you:

- Install the analytic server software on a networked server computer.
- Install, or supervise the installation of, the client application so that it is accessible from end-user desktop computers.

This chapter provides an overview of the installation process. Detailed installation instructions are available on your product DVD in the */Documentation/<language>/InstallationDocuments* directory.

Refer to [Products and Operating Systems](#) on p. 1 for a complete list of server products and their associated client applications.

## ***Installing the Analytic Server Software***

Install the analytic server software on a networked server computer. The server computer must be running the appropriate version of the operating system. If possible, use a server computer configured for, and dedicated to, rapid numeric processing and data access. Additional processing power and memory enhance the server software's performance. Detailed hardware and software requirements, including operating system requirements, appear in the installation instructions.

Installing the analytic server technology installs software that manages access to data and performs the computations required for statistical analysis. It also installs a service (on Windows) or daemon (on UNIX) that listens for incoming end-user login requests and launches a process to handle each end user.

To install the analytic server software, follow the instructions in the */Documentation/<language>/InstallationDocuments* directory on the server DVD.

## ***Installing the Client Application***

Installing the client application installs software that handles the user interface and the presentation of results. You must install, or supervise the installation of, the client application on the desktop computer of each end user. The desktop computer must be running Windows and must meet minimum hardware and operating system requirements. Detailed requirements appear in the installation instructions, which are in the */Documentation/<language>/InstallationDocuments* directory on the client DVD.

If you have a single user license or licenses, see:

- *Single User License Installation Instructions.pdf*

If you have a site license, see:

- *Site License Administrators Guide.pdf*
- *Site License Installation Instructions.pdf*

If you have a network license, see:

- *Network License Administrators Guide.pdf*
- *Network License Installation Instructions.pdf*

## ***Installing Other Analytic Server Software***

You can install multiple versions of the analytic server software at your site—for example, IBM® SPSS® Statistics Server and AnswerTree Server. They can be installed on the same server computer (in separate directories) or on different server computers. Detailed software requirements, including compatible versions of SPSS Inc. products, appear in the installation instructions.

---

# Data Access

If you want your end users to be able to access data on remote servers, including data from databases, you must plan, install, and configure data access. To do this, you need to understand how the application decides where to look for data. You also need to decide if you want to use the data access technology that offers DataDirect Connect ODBC. You can also use OLE DB data sources.

## View of the Data

Before you begin planning data access for the end users, it is important to understand how the application decides what data are available to the end user. The view of the data that is presented to end users depends on how they are running the program—locally or in distributed mode.

**Local analysis mode.** In local analysis mode, in which all data access and processing occur on the end user's desktop computer, the view of data files, ODBC data sources, directories, and drives is from the perspective of the desktop computer—that is, when the end user tries to open a data file, he or she sees the data files, directories, and network drives on his or her desktop computer.

**Distributed analysis mode.** In distributed analysis mode, in which data access and processing occur on a remote server, the view of data files, ODBC data sources, directories, and drives is from the perspective of the server computer—that is, when the end user tries to open a data file, he or she sees the data files, directories, and mounted drives on the server computer.

Your job is to configure data access in either local analysis mode or distributed analysis mode, as required by the end user.

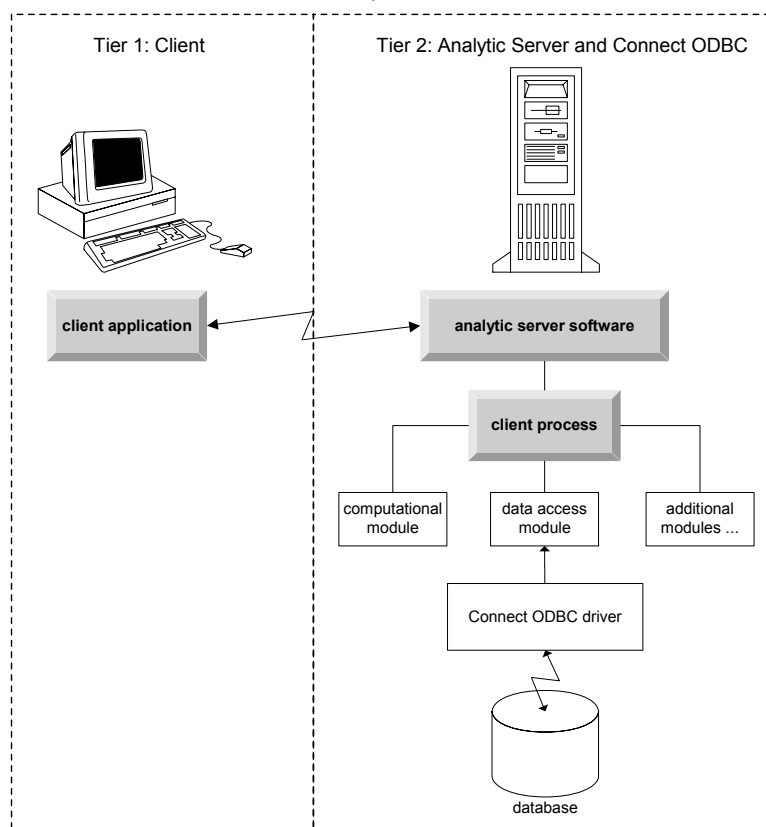
## Data Access Technology

A brief description of Connect ODBC follows. For more information about how the data access technology works with SPSS Inc. products and for links to detailed documentation for specific databases, see *IBM® SPSS® Data Access Pack Installation Instructions* in the */Documentation/<language>/InstallationDocuments* directory on the product DVD.

## Connect ODBC

Connect ODBC is a comprehensive set of individual, database-specific drivers that use ODBC to deliver connectivity to all major data stores, from relational databases to flat-file data.

Figure 3-1  
Connect ODBC in distributed analysis mode



## Accessing Data

As you set up data access, consider the following:

**Data access technology.** Decide if you want to use one of the data access technologies distributed with your SPSS Inc. product. For more information, see the topic [Data Access Technology](#) on p. 8. A more detailed discussion of how to choose a technology appears in the *IBM® SPSS® Data Access Pack Installation Instructions* (in /Documentation/<language>/InstallationDocuments on the product DVD).

**Analysis mode.** The end user's analysis mode determines what data he or she can access. For more information, see the topic [View of the Data](#) on p. 8.

**File system performance (Windows only).** If most of your data are in an proprietary format from SPSS Inc. (for example, .sav files) rather than in a database, we recommend that you store your data on a networked Windows NTFS drive for the best performance.

**File format.** The software handles opening and reading files in UNIX format automatically—you and your end users do not need to take any action to tell the software that a file is in UNIX format.

## Referencing Data

Some client software allows the end user to save references to data and other files. These references must be written from the perspective of the computer that will access the data. For example, if the end user is running in local analysis mode, a reference to `C:\mydata\mydata.sav` causes the software to try to access the file on the local *C* drive of the *desktop computer*. If the end user is running in distributed analysis mode, the same reference to `C:\mydata\mydata.sav` causes the software to try to access the file on the local *C* drive of the *server computer*, possibly resulting in an error.

**Windows.** If you are administering a Windows system, you may decide to store data on the same computer as the analytic server software. If you do, we recommend that users refer to the location of the data from the perspective of the server computer (for example, `C:\ServerData\mydata.sav`). Performance is faster because the network isn't used to locate the file. If your data are on another networked computer, we recommend that your users use UNC file references (for example, `\\mydataserver\ServerData\mydata.sav`). Note that UNC names can be used only when the referenced locations contain the name of a *shared resource* on the network. End users who frequently switch from distributed to local analysis mode are encouraged to use UNC file references because they work regardless of the mode.

**UNIX.** If you are administering a UNIX version of the analytic server software, you may decide to put files on a UNIX server. End users can reference files on a UNIX server—tell them to use the full file specification and forward slashes (for example, `/FILE = '/public/data/ourdata.txt'`.) Avoid using the backslash character in the UNIX directory and in filenames used with the analytic server software.

## Controlling Data Access

You can control access to data by using the operating system to set permissions by user IDs and groups. The end user connects to the server software by logging in from the client application. The server software uses the operating system to enforce the permissions for that user.

*Note:* Additional data security may be enforceable with your database software—the server software's data access modules prompt for IDs and passwords when the database requires them.

**Windows.** How you set access permissions on Windows depends on where the data are stored.

- If the files reside on a networked computer other than the computer running the analytic server software, assign permissions to shared resources.
- If the files reside on the server computer, and on an NTFS drive, use security settings. You cannot control file access for data on the server computer on a FAT drive.

See [File Properties](#) on p. 58 for information about setting sharing and security permissions on Windows.

**UNIX.** When the end user connects to the analytic server software by logging in from the client application, the server software passes the user's login ID and password to the operating system and launches a process for the user. The launched process has the file access rights of the end user's login account.



## Data Sources

### ODBC

The IBM® SPSS® Statistics analytic server software uses ODBC to access most data that are not in a proprietary format, including data that are stored in databases. ODBC requires an ODBC data source. An ODBC data source is the combination of:

- A descriptive name
- A specific driver
- A reference to a database or other type of data file

To access most data, you must configure, or help end users to configure, the ODBC data sources that they need.

The location of the configured data source is critical. It must be configured on the computer that is accessing and processing the data—so configure the ODBC data source on the *server computer* for distributed analysis and on the *desktop computer* for local analysis. For example, compare the location of the ODBC drivers in the figures shown in [Connect ODBC](#) on p. 8.

If you are just starting to use the data access technology (introduced in [Data Access Technology](#) on p. 8), you need to accomplish some additional tasks before you can configure a data source. Refer to the *IBM® SPSS® Data Access Pack Installation Instructions* (in /Documentation/<language>/InstallationDocuments on the product DVD). Data source configuration is discussed again in [Chapter 4, Configuring, Monitoring Usage, and Maintenance](#), of this guide.

### OLE DB

Refer to [Configuring OLE DB Data Sources](#) on p. 16 for information about setting up OLE DB data sources.

## Configuring the UNIX Environment for Data Access

For the data access technology to work on UNIX systems, the server software's startup script must be configured.

### Open the Startup Script

- ▶ Change to the */bin* subdirectory in the server software's installation directory. For example, at the UNIX prompt type:

```
cd /usr/local/serverproduct/bin
```

where */usr/local/serverproduct/bin* is the */bin* subdirectory of the directory in which the server software is installed.

- ▶ Open *statsenv.sh* with a text editor.

**Specify the DataDirect Script**

- ▶ Search for the first comment that contains the text:

```
MERANT_ENVIRONMENT_SCRIPT
```

- ▶ Find the line that defines the location of `odbc.sh`.
- ▶ Edit the line so it contains the correct path to your Connect ODBC *client installation*, and remove the comment character if it has one. For example, change:

```
# MERANT_ENVIRONMENT_SCRIPT=/usr/s1odbc50/5_01_00/.odbc.sh
```

to:

```
MERANT_ENVIRONMENT_SCRIPT=/usr/myDataAccess/s1odbc50/5_01_00/.odbc.sh
```

**Add `odbc.ini` Environment Variable**

- ▶ Add the following lines after the lines above to create an environment variable, `ODBCINI`, that allows IBM® SPSS® Statistics Server to find the `odbc.ini` file:

```
ODBCINI=ODBCDIR/odbc.ini
export ODBCINI
```

where `ODBCDIR` is replaced by the path to your Connect ODBC installation directory.

**Add Paths to Database Libraries**

- ▶ Add lines appropriate for your database, usually the *database home directory* and, if you are *not* using the Data Direct Wire Protocol drivers, a *path to the database libraries*. For example, if you are using Oracle on Solaris, add the following lines:

```
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/bigdisk/oracle/product/8.1.6/lib
export LD_LIBRARY_PATH
ORACLE_HOME=/bigdisk/oracle/product/8.1.6
export ORACLE_HOME
```

where `/bigdisk/oracle/product/8.1.6` is replaced by the path to your Oracle installation directory and `LD_LIBRARY_PATH` is the library path variable for your operating system.

Note that the Data Direct Wire Protocol drivers do not require the installation of database client libraries. However, other Data Direct drivers require these libraries.

**Save the Startup Script**

- ▶ Save `statsenv.sh`.

**Edit `odbc.ini`**

- ▶ Edit `odbc.ini`, the ODBC configuration file, so that ODBC data sources can be accessed from SPSS Statistics Server. See the appendix “The UNIX Environment” in DataDirect’s *DataDirect Connect ODBC Reference* (available if you installed the additional DataDirect documentation

when you installed Connect ODBC) and the chapters for specific drivers in *odbchelp.pdf* in the *doc* subdirectory of your Connect ODBC installation directory.

The change will take effect the next time you start the analytic server software.

*Note:* If you plan to use ODBC with the SPSS Statistics Batch Facility, you need to modify the SPSS Statistics Batch Facility startup script in the same manner.

# ***Configuring, Monitoring Usage, and Maintenance***

After you install the analytic server software, configure its environment by:

- Managing end-user accounts and files
- Configuring ODBC data sources
- Using the administration application (IBM® SPSS® Statistics Administration Console) to configure and monitor the server software
- Controlling service startup

These tasks are described in the following sections.

## ***Managing End-User Accounts and Files***

This section provides an overview of what you need to do to support end users at your site. For more information, see the topic [Supporting End Users](#) in Chapter 5 on p. 37.

### ***Accounts***

End users need accounts in order to log in to the analytic server software and access data. You can use existing login and database accounts, or you can create new accounts for use with the analytic server software. Use your usual system and database administration tools to create and manage end-user accounts (see [User Manager](#) on p. 59 for information about how to access the Windows User Manager). On UNIX, you can authenticate users via the Pluggable Authentication Module (PAM). For more information, see the topic [Configuring PAM](#) in Chapter 5 on p. 45. Also on UNIX, you can use internal authentication, which allows the analytic server software to run without root privileges. Note, however, that this option results in restrictions on client connections. For more information, see the topic [Configuring Internal Authentication](#) in Chapter 5 on p. 46.

On login, the analytic server software launches a process for the end user, passing the user's ID and password to the operating system. The launched process has the file access rights of the end user's account. For more information, see the topic [Permissions](#) in Chapter 5 on p. 38.

Database permissions are enforced by the database software. If the database is restricted, the server software's data access modules prompt the user to log in and pass that information back to the database for verification before accessing data.

## Data Access

By default, each end user can see all the data when opening a file while connected to the analytic server software. Displaying all the data can negatively impact performance and increase network traffic. You can choose to prevent the end user client computers from displaying the data by changing the global setting with the administration application (IBM® SPSS® Statistics Administration Console). For more information, see the topic [Users](#) on p. 25. You can also modify the user profile and groups settings to specify the data access for individual users or groups. For more information, see the topic [IBM SPSS Statistics Server User Profiles and Groups](#) on p. 29.

## Files

Most files that end users need to save should be saved on the desktop computer; however, you may want to allow users to save data files on a networked computer. When the end user logs in to the analytic server software for the first time, the default directory for opening and saving files is the server software's installation directory. Clearly this is not a location where you want users to write files, so set up a directory with write permission and distribute that location to the end users. Once they access that location from the user interface, the client application will store it, and it will become the default location for saved files.

## Profiles

The analytic server software also allows you to create profiles for users. A profile can specify the temporary directory, the UNIX umask setting, the CPU process priority, the client data access setting, and the maximum number of threads for each user or a group of users. For more information, see the topic [IBM SPSS Statistics Server User Profiles and Groups](#) on p. 29.

## Configuring ODBC Data Sources

If your end users access data from databases while they are working in distributed analysis mode, you must configure ODBC data sources on the computer on which the analytic server software is installed.

### ODBC Data Sources and IBM SPSS Data Access Pack

If you are using the SPSS Inc. data access technology, read [Chapter 3, Data Access](#), in this document. Read the appropriate *Installation Instructions* for an the overview of configuring database access and links to detailed documentation on data access for specific databases (the document is in */Documentation/<language>/InstallationDocuments* on the product DVD).

**Windows.** Configure ODBC data sources using the ODBC Administrator. For more information, see the topic [ODBC Administrator](#) in Appendix C on p. 60.

**UNIX.** Edit the startup environment script for the server software as described in [Configuring the UNIX Environment for Data Access](#) on p. 11 and configure ODBC data sources using the *odbc.ini* file. For more information, see the topic [odbc.ini](#) in Appendix D on p. 64.

## Configuring OLE DB Data Sources

If your end users need to access data from an OLE DB data source while they are working in distributed analysis mode, you must configure the OLE DB data sources on the computer on which the analytic server software is installed. The following must first be installed on the Windows computer on which the analytic server software is also installed:

- .NET Framework, version 1.1 or later.
- Dimension Data Model and OLE DB Access.

These are available on the Autoplay menu.

An end-user cannot create an OLE DB data source in distributed analysis mode. You must create the data source on the analytic server software computer or copy the UDL file created with the client application. The copied file should work if paths do not include mapped drives. Copy the file to `\Documents and Settings\All Users\Application Data\spssinc\UDL` on the analytic server software computer. If the directory doesn't exist, create it.

### To create the OLE DB data source:

- ▶ In `\Documents and Settings\All Users\Application Data\spssinc\UDL`, create a file with the extension `.udl`. You could do this by saving a blank document with a text editor or by creating a new text document in Windows Explorer.
- ▶ Double-click the file.
- ▶ Select the database by entering the directory location and database name or by clicking the button to browse to a database. (A user name and password may also be required.)
- ▶ Click OK after entering all necessary information. (You can make sure the specified database is available by clicking the Test Connection button.)

## Using a Third-Party Sort Engine

By default, the analytic server software tries to use an external, third-party engine for sorting. To use the correct sorting engine, complete the following steps:

- ▶ If the third-party sort engine is not installed on the server computer, install it. SPSS Inc. does not provide the engines. You must purchase and license an engine from a third-party vendor.
- ▶ Ensure the sort engine's library is on the system's execution or library path variable. On Windows, this is the PATH environment variable; on UNIX, this is LD\_LIBRARY\_PATH or LIBPATH, depending on the UNIX vendor. This step is required to allow the analytic server software to load the third-party sort library.
- ▶ Using the administration application (IBM® SPSS® Statistics Administration Console), set the Sort option to the appropriate third-party engine. For more information, see the topic [Users](#) on p. 25.

Any procedures that require sorting (e.g., `SORT`) will subsequently use the third-party sorting engine. Issuing the `SET SORT=INTERNAL` syntax command forces the analytic server software to use the internal algorithm for sorting. An end user can also explicitly specify third-party sorting by issuing the `SET SORT=EXTERNAL` command. However, this is not necessary because third-party sorting is the default.

### ***Checking the Current Sort Option***

To check which sorting option is being used, you can issue the `SET MESSAGES ON` syntax command and run a `SORT` job. You can also use the `SHOW SORT` syntax command.

## ***Using the Administration Application***

The IBM® SPSS® Statistics Administration Console provides a user interface to monitor and configure your IBM® SPSS® Statistics Server installations. The SPSS Statistics Administration Console uses the framework from the IBM® SPSS® Collaboration and Deployment Services Deployment Manager, so some of the documentation refers to the later product. The SPSS Statistics Administration Console can be installed only on Windows computers; however, it can administer SPSS Statistics Server installed on any supported platform.

### ***To Start IBM SPSS Statistics Administration Console***

- ▶ From the Windows Start menu, choose:  
[All] Programs > IBM SPSS Collaboration and Deployment Services > Deployment Manager > IBM SPSS Statistics Administration Console

## ***Administered servers***

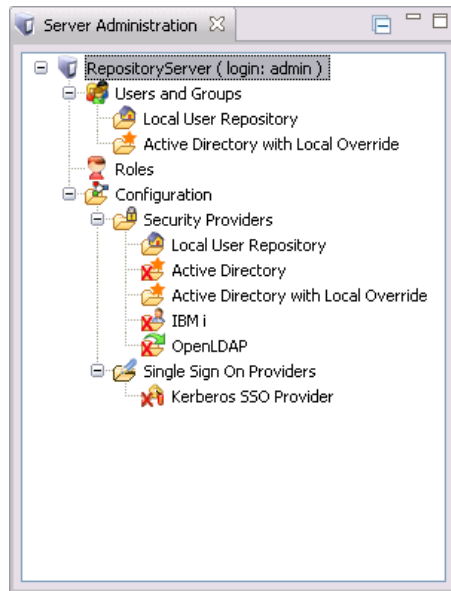
Server administration in IBM® SPSS® Collaboration and Deployment Services Deployment Manager involves:

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 4-1**  
*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 menu choose:

File > New > Administered Server Connection

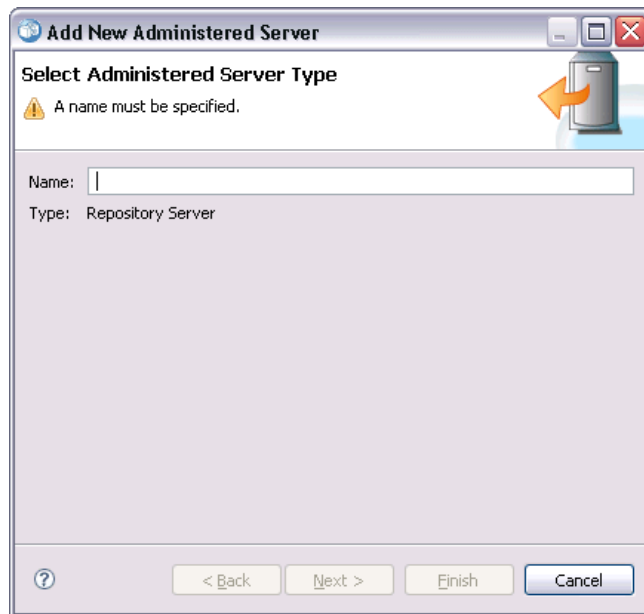
The Add New Administered Server dialog box opens. Adding a new connection requires the specification of the administered server type and the administered security server information.

### ***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 4-2  
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® Text Analytics Server

### **Selecting an Administered Server Type**

In the Select Administered Server Type dialog box:

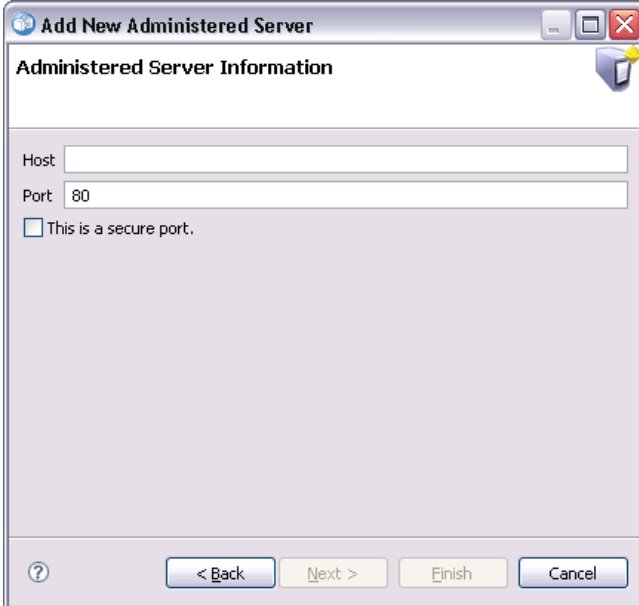
1. Enter a name for the server.

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 4-3  
Administered Security Server Information dialog box



The screenshot shows a dialog box titled "Add New Administered Server". Inside, there is a section titled "Administered Server Information". This section contains three input fields: a "Host" text box, a "Port" text box with the value "80", and a checkbox labeled "This is a secure port." which is currently unchecked. At the bottom of the dialog box, there are four buttons: "< Back", "Next >", "Finish", and "Cancel".

**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:

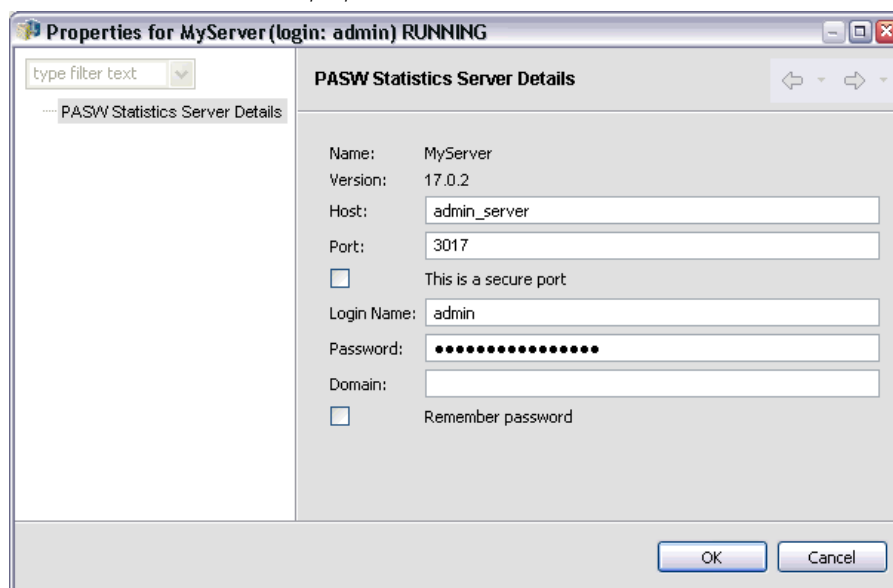
1. Enter the host name or IP address for the server being added.
2. Enter the port number on which the server being added is running.
3. Specify whether or not the server uses SSL, if applicable.
4. Click Finish.

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

### **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 4-4  
IBM SPSS Statistics Server properties



For SPSS Statistics 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.
- **This is a secure port.** If this option is selected, the server uses an SSL connection for communication.
- **Login Name.** The user logged in to the server, displayed in clear text.

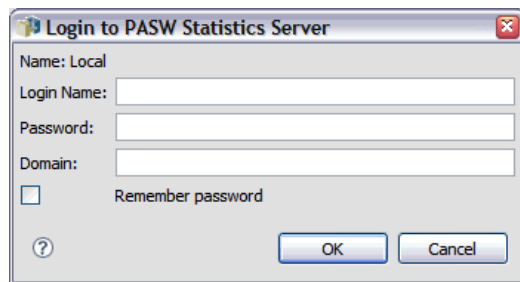
- **Password.** The string used to authenticate the user. For security, password text is displayed in a masked format.
- **Domain.** For Windows only, the domain name associated with the user account.
- **Remember password.** If selected, the password is remembered across sessions.

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.

Figure 4-5  
*Login dialog*



For SPSS Statistics servers, the login parameters include:

**Login Name.** The user to log in to the server, displayed in clear text. Use an account that has administrative access to the analytic server software.

**Password.** The string used to authenticate the user. For security, password text is displayed in a masked format.

**Domain.** For Windows only, the domain name associated with the user account. This is required only when the server and the computer running IBM® SPSS® Collaboration and Deployment Services Deployment Manager are in different domains.

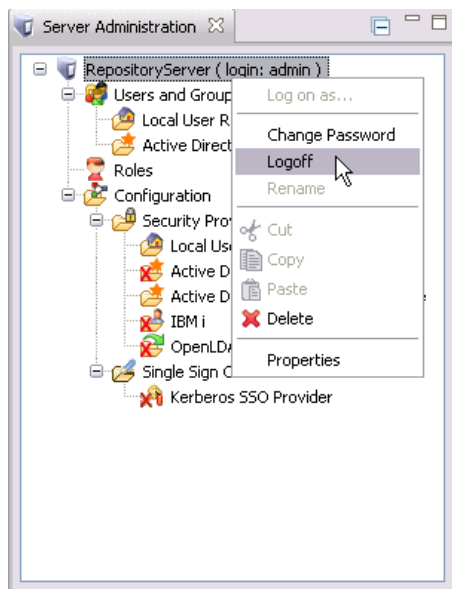
**Remember password.** Whether to remember the password across sessions.

### **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.

Figure 4-6  
Logging off from a server



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.

### ***IBM SPSS Statistics Server Connections***

You must specify a connection to each IBM® SPSS® Statistics Server on your network that you want to administer. You must then log in to each server that you want to configure or monitor.

After you log in to your SPSS Statistics Server, options are shown beneath the server name: [Configuration](#) and [Monitoring Current Users](#). Double-click one of these options to display the associated pane.

### ***Controlling the IBM SPSS Statistics Server***

The IBM® SPSS® Statistics Administration Console allows you to pause, shut down, and restart the IBM® SPSS® Statistics Server.

- ▶ In the Server Administration pane, select the SPSS Statistics Server.
- ▶ From the menus choose:  
Tools > SPSS Statistics Server
- ▶ Select an option:

**Pause.** Pause a server when you want to stabilize usage while you diagnose a problem, or when you want to temporarily limit the number of connections. Pausing the server prevents end users from connecting. It doesn't affect end users that are connected to the server when it is paused. Choose Resume when you want the server to accept connections again.

**Shutdown.** Shut down a server when you need to terminate the server software. Once you have shutdown, you must go to the computer where the server software is installed to restart it. Avoid shutting down servers when end users are connected. Shutting down will disconnect the users, and they may lose work.

**Restart.** Some configuration changes require that you restart the server before they take effect. These are indicated by an asterisk (\*) in the Configuration pane. Restarting a server temporarily stops the server software, and then restarts it. You must disconnect all end users before you can restart the server. Because end users can lose work when you disconnect them, please ask them to logoff from the client application. Restarting the server may take a few seconds.

## ***IBM SPSS Statistics Server Configuration***

The Configuration pane shows configuration options for IBM® SPSS® Statistics Server. To activate the pane, double-click the Configuration node beneath the desired server in the Server Administrator pane. To save any configuration changes, choose Save from the File menu. Some settings, indicated by an asterisk (\*), require the server to be restarted before taking effect.

### ***Connections***

**Host.** The host name—an alphanumeric name or an IP address—is the network designation for the computer where the server software is running. Change the host name when you move the server software to a different computer, when the network designation of the computer changes, or when the computer has more than one IP address and the server software is using the wrong one. End users must use the same name when they connect to the server software. If you make a change, distribute the new name.

**Port Number.** The port number is the port that the server software uses for communications. Change the port number only if another application on the server computer is using the same port. Enter an integer between 1025 and 32767. If you make a change, be sure to distribute the new port number to end users. The default port number is 30<version>, where <version> is the major version of IBM® SPSS® Statistics Server. For example, the default port number for version 17 is 3017.

**Maximum Number of Users.** End users connect to the analytic server software by logging in from a client application. When a connection is made, the server software launches a process to handle the end user. Increase or decrease the maximum number of users to adjust the load on the server computer. Enter an integer between 1 and 2000.

**Secure Sockets Layer.** If you use a Secure Sockets Layer (SSL) to encrypt the communications that occur when end users connect to the analytic server software, you can configure the server software to accept SSL connections from end users. If you select this option, you must specify the location of the public and private key files. For more information, see the topic [File Locations](#) on p. 25.

## **File Locations**

**Configuration File.** The analytic server software stores the information that it needs in a configuration file. The server software must locate this file on startup, and it needs to write to the file periodically. Since the file is writable, you may want to change its location.

**Server Temp Files Directory.** When the server software accesses and processes data, it often has to keep a temporary copy of that data on disk. The amount of disk space that will be used for temporary files depends on the size of the data file that the end user is analyzing and the type of analysis that he or she is doing. It varies from the size of the data file to three times the size of the data file. Because the temporary files are writable and can get quite large, you may want to change their location. If you do, the location that you define is a root directory. When an end user connects to the server software, it creates a unique, temporary subdirectory in the root location that you specify. You can specify multiple locations separated by commas. Using multiple temporary file locations can improve performance when the locations are on different spindles. For more information, see the topic [Improving Disk Usage](#) in Chapter 6 on p. 51. The temporary file location is a global setting for all users and can be overridden by the user profile or group setting. For more information, see the topic [IBM SPSS Statistics Server User Profiles and Groups](#) on p. 29.

**User Profile File.** The profile file defines the user profile and group settings. It is recommended that you move the profile file from the default location. This ensures that subsequent installations of the analytic server software do not overwrite your profile file. The settings in the profile file are modified within the administration application (IBM® SPSS® Statistics Administration Console). For more information, see the topic [IBM SPSS Statistics Server User Profiles and Groups](#) on p. 29.

**SSL Public Key File/SSL Private Key File.** If you selected Secure Sockets Layer in the Connections group, you must specify the full path to the SSL public and private key files. If the public and private keys are stored in the same file, specify the same file path in both fields.

## **Users**

**Admin Group.** When a user connects to the IBM® SPSS® Statistics Administration Console, the analytic server software checks if the user is a member of the administrator group. The administrator group corresponds to a system account group on the machine on which the analytic server software is installed. Any member of the administrator group has administrative-level rights to the analytic server software. This means that a member of this group can use the SPSS Statistics Administration Console to configure the analytic server. You can specify any system account group to be the administrator group. However, if the analytic server software cannot find the group, it will revert to the default group, which is the administrator group for the computer on which the analytic server software is installed (for example, the Administrators group on an English Windows system).

**Client Data Access.** When end users connect to the analytic server software and open a data file, they are able to see the actual data in the file. Always displaying the data in the client application can negatively affect performance and increase network traffic. Therefore, you may choose to prevent the data from displaying in the client application. The client will still display the data dictionary, allowing the end user to check which variables appear in the data. The client data access setting is a global setting for all users and can be overridden by the user profile or group setting. For more information, see the topic [IBM SPSS Statistics Server User Profiles and Groups](#) on p. 29.

**Sort.** When end users connect to the analytic server software and run a sort procedure, the analytic server software uses the specified sort engine. The engine must be installed and configured on the analytic server software computer. For more information, see the topic [Using a Third-Party Sort Engine](#) on p. 16.

**Umask.** You can adjust the umask that is used when users connect to the analytic server software running on a UNIX machine. This three-digit octal number is a global setting for all users and can be overridden by the user profile or group setting. For more information, see the topic [IBM SPSS Statistics Server User Profiles and Groups](#) on p. 29. Note that this setting is not available when the analytic server software is running under Windows.

**Maximum Threads.** You can specify a limit to the number of threads that the analytic server software can use. You may do this to limit the system resources available for multithreaded procedures. The maximum threads setting is a global setting for all users and can be overridden by the user profile or group setting. For more information, see the topic [IBM SPSS Statistics Server User Profiles and Groups](#) on p. 29.

**Cache Compression.** When end users explicitly issue a `CACHE` command or run a procedure that creates scratch files automatically, the analytic server software can compress the scratch (temporary) files using `zlib`. You should turn cache compression on if your users typically read large data files. A compressed scratch file reduces the amount of disk I/O compared to an uncompressed scratch file but requires more processing by the analytic server software. To minimize the effect of the extra CPU time on job elapsed time, compression and decompression are performed using separate threads. That means that one or more processors can be handling these chores while others are executing commands. Because the extra overhead is greater when writing a file compared to reading it, let your users know that they should use a higher value for `SET CACHE`. Doing so will reduce the frequency of writing scratch files and increase the number of times each scratch file is read. The cache compression setting determines the default behavior for cache compression. You can also set the compression override setting to allow users to override this default. The cache compression setting is a global setting for all users and can be overridden by the user profile or group setting. For more information, see the topic [IBM SPSS Statistics Server User Profiles and Groups](#) on p. 29.

**Compression Override.** You can specify whether end users can override the cache compression setting with `SET ZCOMPRESSION` syntax. The compression override setting is a global setting for all users and can be overridden by the user profile or group setting. For more information, see the topic [IBM SPSS Statistics Server User Profiles and Groups](#) on p. 29.

**Maximum JVM Memory.** When end users connect to the analytic server software, memory is allocated for the Java virtual machine (JVM) on the analytic server computer. Memory is allocated for each user. You can change the amount of allocated memory by specifying a number in megabytes (MB). You may need to this if users have jobs that require a large amount of memory,



for example when exporting very large custom tables. The JVM memory setting is a global setting for all users and can be overridden by the user profile or group setting. For more information, see the topic [IBM SPSS Statistics Server User Profiles and Groups](#) on p. 29.

*Note:* At this time, compiled transformations are supported only for IBM® SPSS® Statistics Server running on Windows. UNIX/Linux is not supported at this time.

**Compiled Transformations.** When a syntax job includes transformation commands, the analytic server software converts the transformations to bytecode when the syntax job is run. This bytecode is then “interpreted” while running the syntax job. The analytic server software can compile the transformations to machine code to improve performance when there are many cases in the data and several transformations in the syntax job. If Compiled Transformations is set to Yes and a user issues the `SET CMPTRANS=YES` command, the analytic server software will:

- Convert the transformations to bytecode as usual.
- Convert the bytecode to C++ code.
- Compile the C++ code to machine code.
- Execute the machine code.

**Compiler Path.** The full path to the C++ compiler that is used for compiling transformations. You must use MinGW, which is a Windows port of the GNU Compiler Collection (GCC). Be sure to include the executable name in the path specification, for example `C:\MinGW\bin\g++.exe`.

## ***COP Configuration***

If your site is running IBM® SPSS® Collaboration and Deployment Services 3.5 or later, you can configure one or more analytic servers to communicate with the Coordinator of Processes (COP). The COP registers the server and allows client computers to connect the server without knowing the server’s host name. In other words, the COP allows client computers to search for available analytic servers.

**COP Status.** The current status of the COP. You cannot change this value.

**COP Host.** The host name—an alphanumeric name or an IP address—is the network designation for the IBM SPSS Collaboration and Deployment Services server on which the COP is running.

**COP Port Number.** The port number is the port on the IBM SPSS Collaboration and Deployment Services server that the COP uses for communications.

**COP Login Name.** The login name for connecting to the IBM SPSS Collaboration and Deployment Services server on which the COP is running.

**COP Password.** The password for the login name.

**COP Provider.** The security provider against which to validate the user/password combination.

**COP Enabled.** Making an analytic server COP enabled allows it to communicate with the COP. When an end user searches for available servers, the COP will display the analytic server as one of the results.

**Server Name.** The COP displays this as the name of the analytic server.

**Description.** The COP displays this description for the analytic server.

**Update Interval.** The analytic server software sends an update to the COP after the specified interval (in minutes) has elapsed. The update lets the COP know whether the analytic server is running.

**Weight.** When an analytic server is part of a cluster, the COP uses the weight to determine how many client computers can connect to an individual server. For example, a weight of 5 indicates that 5 times as many client computers can connect to the analytic server compared to an analytic server with a weight of 1.

## **Logging**

The server software records session information in two types of log files:

- The active log file where session information is currently being recorded
- The backup log files

When the active log file reaches maximum size, it is moved to a backup log file, and the active log file starts over. The log file can also record performance information when the Log Sample Interval is set to a number greater than 0

**Active Log(s).** The name of the active log file.

**Number of Backup Log Files.** You can change the number of backup log files that the server keeps. Increase or decrease the number to accommodate the amount of information that you want to keep.

**Performance Log Interval.** By default, the log file does not record performance information. You can configure the log file to record this information if you are trying to determine problematic areas. You do this by setting the performance log interval to a number of seconds between 1 and 600. It is not recommended that you set the interval to a small number for a long period of time. Doing so may result in the log file being filled up quickly. It may also affect performance. To disable performance samples, set this to 0. Note that you can view performance log information directly in a grid. For more information, see the topic [Viewing Performance Log Information](#) on p. 29.

**Maximum Log File Size.** When the active log file reaches maximum size, it is moved to the backup log file, and the active log file starts over. Increase or decrease the size to accommodate the number sessions that you want to record.

## **Viewing the Log Files**

You can view the active log or the backup logs for a specific server.

### **To View a Log File**

- ▶ Select the IBM® SPSS® Statistics Server in the Server Administration pane.
- ▶ From the menus choose:  
Tools > SPSS Statistics Server > View Active Log
- ▶ In the View Log Selection for SPSS Statistics Server dialog box, select the log file that you want to view.
- ▶ Click OK.

The file opens in a new pane. If you are looking for specific information in the file, press Ctrl+F to open the Find/Replace dialog box.

### **Viewing Performance Log Information**

If the Performance Log Interval is set to a number greater than 0, performance log information is written to the log files. For more information, see the topic [Logging](#) on p. 28. You can see the performance log information by viewing the log files. However, the performance log information is interspersed with all the other log information and may be difficult to read. The IBM® SPSS® Statistics Administration Console offers an alternate interface for reading the performance log directly in a grid.

#### **To View the Performance Log Information Directly**

- ▶ Double-click the Performance Logs node beneath the desired IBM® SPSS® Statistics Server in the Server Administration pane.

Performance log information appears in the Performance Logs pane. Above the grid is the date/time range of the displayed performance log information. By default, performance log information from the last hour is displayed. You can change this date/time range. You can also subsequently export the displayed log information to a comma-separated values (CSV) file.

#### **To Change the Performance Log Date/Time Range**

- ▶ Click Change... next to the displayed date/time range.
- ▶ In the Select Date Range dialog box, enter a date and time for the start and end of the date/time range.
- ▶ Click OK.

#### **To Export the Performance Log Information to a CSV File**

- ▶ Click Export... next to the displayed date/time range.
- ▶ In the Select File to Export dialog box, browse to the location where you want to save the CSV file.
- ▶ Enter a name for the file. The file extension is added automatically, so you don't need to add it.
- ▶ Click Save.

## **IBM SPSS Statistics Server User Profiles and Groups**

The analytic server software provides the ability to create profiles of individual users and groups of users. The profile specifies the temporary files directory, the UNIX umask setting, the CPU process priority, client data access, and maximum threads for the user or group of users. These settings override the associated global, default settings. For information about creating user profiles and groups, see [Creating and Editing IBM SPSS Statistics Server User Profiles and Groups](#) on p. 30.

**Profile File**

User profile and group settings are stored in the profile file, *UserSettings.xml*. By default, this file is located in the IBM® SPSS® Statistics Server installation directory on Windows and the */bin* subdirectory of the installation directory on UNIX. Before making changes to the user profiles and groups, it is recommended that you move the default file and specify the new location in the administration application (IBM® SPSS® Statistics Administration Console). For more information, see the topic [File Locations](#) on p. 25.

**How the Server Applies Settings**

The analytic server follows these steps when determining the temporary files directory, the umask setting, the CPU priority, client data access, and maximum number of threads for a particular user.

- ▶ Look for a user profile for the user name and domain of the user connecting to the server. Select the first match found. On UNIX, the analytic server ignores the domain, and case matters for the user name. On Windows, the case of the user name or domain does not matter. Also, if a user logged onto the analytic server without specifying a domain, the analytic server looks for a user name match with a blank domain. If it can't find one, it uses a match on only the user name.
- ▶ If the user profile defines settings, apply these settings to the analytic server process.
- ▶ If the user profile does not define settings or does not define all settings, apply the user's group settings to the analytic server process.
- ▶ If some settings are still not defined or there was no matching user profile, apply the default umask, the temporary file directory, client data access setting, and maximum number of threads defined by the administration application (SPSS Statistics Administration Console). No default CPU priority processing is used. For information about the configuring the default settings, see [Users](#) on p. 25.

**Creating and Editing IBM SPSS Statistics Server User Profiles and Groups**

User profiles and groups are managed in the Manage Users and Groups pane.

**To Display the Manage Users and Groups Pane**

- ▶ Double-click the User Profiles and Groups node beneath the desired IBM® SPSS® Statistics Server in the Server Administration pane.

The Manage Users and Groups pane displays the currently defined user profiles and groups in the User Profiles and Groups grid. Select a user profile or groups from this grid. The grid on the right then displays the groups to which the selected user belongs or the users in the selected group, depending on what is selected. Both grids also display the current settings for the user profiles and groups.

**To Create a New User Profile**

- ▶ In the Manage Users and Groups pane, click New User Profile.

- ▶ In the Create New User Profile dialog box, enter the name of the user for whom you are creating the profile. If you want to limit the profile to a user with a specific domain, be sure to include the domain name (for example, domain\user). Note that you are not creating a user at this time. You are associating a profile to an *existing* user.
- ▶ If necessary, define any of the available settings. If you are only creating a user profile to assign a user to a group, you don't have to define any settings. For descriptions of all the available settings, see [Available User Profile and Group Settings](#) on p. 31.
- ▶ In the Manage Groups the User Belongs to area, identify the group(s) to which the user belongs. The grid on the left shows all defined groups, and the grid on the right shows the groups to which the user belongs. Select a defined group and click >>>> to add the user to the selected group. You can also click Add All to add the user to all groups. To remove a user from a group, select a group to which the user belongs and click <<<<. You can also click Remove All to remove the user from all groups.

#### ***To Create a New User Profile Group***

- ▶ In the Manage Users and Groups pane, click New Group.
- ▶ In the Create New Group dialog box, enter a name for the group. The name of the group is arbitrary and does not correspond to a group on your system.
- ▶ Define any of the available settings. For descriptions of all the available settings, see [Available User Profile and Group Settings](#) on p. 31.
- ▶ In the Manage Users in Group area, identify the user(s) that belong to the group. The grid on the left shows all users with defined user profiles, and the grid on the right shows the users that belong to the group. Select a user and click >>>> to add the selected user to the group. You can also click Add All to add all users with defined user profiles to the group. To remove a user from a group, select a user in the group and click <<<<. You can also click Remove All to remove all users from the group.

#### ***To Edit or Delete an Existing User Profile or Group***

- ▶ In the Manage Users and Groups pane, select a user profile or group. If deleting, you can select multiple user profiles or groups. Be careful because there is no option to undo the deletion.
- ▶ Click Edit to edit the user profile or group, or click Delete to delete the user profile or group.

### ***Available User Profile and Group Settings***

Following are the available settings that you can assign to user profiles and groups.

**Priority.** The CPU priority for the analytic server process. Use a negative or positive integer. On **UNIX**, this corresponds to the `nice` command, which has a range of -20 to 19, with -20 yielding more favorable scheduling and 19 yielding less favorable scheduling. `nice` settings can be viewed in the `NI` column of the `ps` command output. On **Windows**, a negative value yields a base priority of ABOVE\_NORMAL (more favorable scheduling), a positive value yields a base priority of

BELOW\_NORMAL (less favorable scheduling), and 0 yields NORMAL priority scheduling. You can view base priority settings in the Task Manager.

**Temp file location.** The directory to which the analytic server process writes temporary files. The user must have read/write access to this directory. To gain any performance benefit, you must specify a different physical drive for each user. You can use a different directory on the same drive for multiple users, but there will be no performance gain. You might do this so that users have temporary directories to which only they have access. You might also create different sized partitions and specify a partitioned drive for each user, thereby controlling the temporary files space allocated to each user. You can also specify multiple locations separated by commas. You can also use the administration application to specify this setting globally for all users. For more information, see the topic [Users](#) on p. 25.

**Client data access.** Access to the data from the client. Select an option to indicate whether the user sees the data in the client Data Editor. This option may increase client/server performance because data passes can be slow over the network. No Data Editor Data Access indicates that the user does see the data in the Data Editor. Although the user is able to see all the cases in the data, this option may degrade performance in some circumstances. You can also use the administration application to specify this setting globally for all users. For more information, see the topic [Users](#) on p. 25.

**Maximum threads.** The number of threads for the analytic server process. This setting limits the number of threads that the analytic server software can use for multithreaded procedures run by the user. You may do this to limit the system resources available for multithreaded procedures. You can also use the administration application to specify this setting globally for all users. For more information, see the topic [Users](#) on p. 25.

**Cache Compression.** When end users explicitly issue a `CACHE` command or run a procedure that creates scratch files automatically, the analytic server software can compress the scratch (temporary) files using `zlib`. You should turn cache compression on if an end user or group of users typically reads large data files. You can also use the administration application to specify this setting globally for all users. See [Users](#) on p. 25 for more information about cache compression and the global setting.

**Compression Override.** You can specify whether an end user or group of users can override the cache compression setting with `SET ZCOMPRESSION` syntax. You can also use the administration application to specify this setting globally for all users. For more information, see the topic [Users](#) on p. 25.

**Maximum JVM Memory.** When end users connect to the analytic server software, memory is allocated for the Java virtual machine (JVM) on the analytic server computer. Memory is allocated for each user. You can change the amount of allocated memory for a specific end user or group of users by specifying a number in megabytes (MB). You may need to this if users have jobs that require a large amount of memory, for example when exporting very large custom tables. You can also use the administration application to specify this setting globally for all users. For more information, see the topic [Users](#) on p. 25.

**Umask.** The umask for the analytic server process. Use an octal three-digit number. You can also use the administration application to specify this setting globally for all users. For more information, see the topic [Users](#) on p. 25.

## Monitoring IBM SPSS Statistics Server Users

The Monitor Current Users pane of the IBM® SPSS® Statistics Administration Console shows users that are connected to the IBM® SPSS® Statistics Server computer. To activate the pane, double-click the Monitor Current Users node beneath the desired server in the Server Administrator pane. This populates the pane with a list of connected users. The list refreshes at the rate shown. To refresh the list manually, from the menus choose:

View > Refresh

The monitoring pane displays the following information.

- **User.** The ID with which the user connected to the server software.
- **Client.** The name of the client application that the user is running.
- **Version.** The version of the client application that the user is running.
- **Connection ID.** An arbitrary number assigned to a user when the user connects to the server software. A lower number indicates that a user logged in earlier than another user with a higher number.
- **Authentication.** The access rights of the user.

*Note:* If the name or version is listed as <Unknown>, the user is connected with client software that the SPSS Statistics Administration Console cannot recognize.

You can perform user-specific actions in this pane. Right-click to see a list of choices:

- **Disconnect.** Disconnect one or more users from the analytic server software. For more information, see the topic [Disconnecting Users](#) on p. 33.
- **Broadcast.** Send a message to one or more users. For more information, see the topic [Broadcasting a Message to Users](#) on p. 33.

### Disconnecting Users

Disconnect users when you need to limit drain on the server computer, or when you notice a user using the system inappropriately. Avoid disconnecting all users since it is possible that they may lose work. Your ID also appears in the user list. The IBM® SPSS® Statistics Administration Console will prevent you from disconnecting yourself. If all users—including yourself—need to be disconnected from the server computer, disconnect all users and then disconnect from the server.

- ▶ On the Monitor Current User pane, click the Refresh button to ensure the user list is current.
- ▶ Select the user(s) you want to disconnect.
- ▶ Right-click and choose Disconnect User(s) from the menu.

### Broadcasting a Message to Users

You can broadcast a message to a specific user or all users who are connected to the analytic server software. You may do this if you need to restart the analytic server software and want to warn users before doing so.

- ▶ On the Monitor Current User pane, click the Refresh button to ensure the user list is current.
- ▶ Select the user(s) to whom you want to broadcast a message. If you want to broadcast the same message to all users, you don't need to select specific one.
- ▶ Right-click and choose Broadcast [All] User(s) from the menu.
- ▶ In the Broadcast dialog box, type the content of the message. You can add line breaks if needed.
- ▶ Click OK to broadcast the message.

## Controlling Service Startup

The analytic server software has a framework component that handles all communication between the client application and the modules. On Windows, the framework component is a service. On UNIX, the framework component is an application, usually run as a daemon.

**Windows.** By default, the service is configured for automatic startup, which means that it will restart automatically when the computer is rebooted. When started this way, the service runs unattended and the server computer can be logged off without affecting the service. You can use the Windows Services Control Panel to change the service startup parameters. For more information, see the topic [Services Control Panel](#) in Appendix C on p. 59.

**UNIX.** A startup script, `start_statistics_server`, is included in the `/bin` subdirectory of the installation directory. The script calls `statsenv.sh` to configure the environment for the analytic server software and then starts the application. You must start the server software with this script. The startup script must be run from the `/bin` subdirectory. To execute it, you must be logged in as **root** or, if using internal authentication, the user who owns the analytic server software daemon. For more information about internal authentication, refer to [Configuring Internal Authentication](#) on p. 46. The command:

```
./start_statistics_server -d
```

will start the server software as a daemon process, which is the recommended way to run the server software.

### Startup Script Command Line Parameters

The `start_statistics_server` script accepts the following command line parameters (in any order):

- **Daemon.** Run the server software as a daemon process by optionally specifying `-d`. If you omit the `-d`, the server will start as a foreground process. For example, to start the server software as a daemon, use the command:

```
./start_statistics_server -d
```

- **Port.** A port number can optionally be specified on the startup script command line. For example, to specify a port number, use the command:

```
start_statistics_server -p nnnn
```

where

`nnnn` is the desired port number.



Specify a port number only if you need to resolve a port number conflict. The default will work unless another application on the computer is using the same number. This command line setting overrides the value set in the administration application.

## Other Maintenance

**Remove unneeded files.** Periodically check the temporary file location and the log file location for unneeded files and remove them. The locations are defined with the administration application.

**Check running processes.** If you don't regularly reboot the server computer, periodically check the processes running on the computer and end any processes that are not in use. The process names are listed in [Starting and Stopping the Analytic Server Software](#) on p. 35.

## Starting and Stopping the Analytic Server Software

The administration application will restart the server software for you so that configuration changes can be committed; however, at times you may need to start or stop the server software from the operating system. Follow the steps below for your operating system, using the process name of your analytic server product. On Windows, the service name is *StatisticsNNService*, where *NN* is the major version number. On UNIX and Linux, the daemon name is *statisticsd*.

*Scheduling note:* Stopping the service or daemon disconnects end users and terminates their processes, so try to schedule configuration and maintenance tasks for a time when you expect few users to access the system (for example, early morning or late evening).

### To Start the Service or Daemon

**Windows.** Use the Windows Services Control Panel to start the service. For more information, see the topic [Services Control Panel](#) in Appendix C on p. 59.

**UNIX.** Start the Server with the startup script, *start\_statistics\_server*, which is included in the */bin* subdirectory of the installation directory. The startup script must be run from the */bin* subdirectory. To execute it, you must be logged in as **root** or, if using internal authentication, the user who owns the analytic server software daemon. For more information about internal authentication, refer to [Configuring Internal Authentication](#) on p. 46.

### To Stop the Service or Daemon

**Windows.** Use the Windows Services Control Panel to stop the service. For more information, see the topic [Services Control Panel](#) in Appendix C on p. 59.

**UNIX.** Kill the server process. (See [ps and kill](#) on p. 63 for an example.) The daemon automatically creates a file (*statisticsd.pid*) that contains the process ID for the daemon. You can use this file in conjunction with the kill command by running the following from the */bin* subdirectory of the installation directory:

```
kill -9 `cat statisticsd.pid`
```

**Platform independent.** Use the administration application (IBM® SPSS® Statistics Administration Console). For more information, see the topic [Controlling the IBM SPSS Statistics Server](#) on p. 23.

## ***Configuration for Improving Performance***

Refer to [Chapter 6, \*Analyzing and Improving Performance\*](#), for information about modifying the configuration of the analytic server software to improve performance.

# *Supporting End Users*

Supporting end users involves making sure that they have the information that they need to run their SPSS Inc. product in distributed analysis mode. To use the analytic server software, end users need to know:

- How to connect to the server software.
- How to access data and files.
- Where to save data and files.

## *Connecting to the Server Software*

The end user connects to the analytic server software by logging in from the client application. To login an end user needs the following information from you:

- **Computer name or IP address.** When ends users connect to the analytic server software, they log in from the client application. To do that, they need to correctly specify the name of the computer running the analytic server software. The server computer can be identified by an alphanumeric name (for example, myserver) or an IP address assigned to the server computer (for example, 202.123.456.78)—whichever you prefer. If you configure the server and client desktop computers to use Secure Sockets Layer (SSL), the end user must use a fully-qualified domain name (e.g., myserver.mycompany.com).
- **Port number.** End users need to correctly specify the port on which the server software is listening for connections. The port number is the default for the server, or whatever you specified when you configured the analytic server software.
- **Domain name (Windows only).** End users may also need to specify a domain name. A domain name is required only when the server computer is in a different domain than the end-user desktop computers.
- **User ID and password.** End users are required to log in to the server computer. To do this, the users need a valid account, with appropriate permissions, for the computer on which the server software is running.
- **Secure Socket Layer (SSL).** If you use SSL to encrypt the communications that occur when end users to connect to the analytic server software, tell the users to enable SSL when they set up the server connection.

## Permissions

**Windows.** A user connecting to analytic server software that is installed on an *NTFS* drive must login with an account that has the following permissions:

- Read and execute permissions to the server's installation directory and its subdirectories
- Read, execute, and write permissions to the directory location for temporary files

In Windows Server 2008 and later, you cannot assume that users have these permission. Be sure to explicitly set permissions as needed.

If the server software is installed on a *FAT* drive, you do not need to set permissions because all files allow users to have full control.

**UNIX.** If you are not using internal authentication, a user connecting to the analytic server software must login with an account that has the following permissions:

- Read and execute permissions to the server's installation directory and its subdirectories
- Read, execute, and write permissions to the directory location for temporary files.

To authenticate users, you can use the Pluggable Authentication Module (PAM). For more information, see the topic [Configuring PAM](#) on p. 45. You can also use internal authentication, which allows the analytic server software to run without root privileges. Note that this option results in restrictions on client connections. For more information, see the topic [Configuring Internal Authentication](#) on p. 46.

You can use the administration application (IBM® SPSS® Statistics Administration Console) to change the default location of the temporary files. For more information, see the topic [File Locations](#) in Chapter 4 on p. 25. You can also change the location for individual users or groups. For more information, see the topic [IBM SPSS Statistics Server User Profiles and Groups](#) in Chapter 4 on p. 29.

### Administrator-Level Permissions

By default, the administrator group for the analytic server software is the administrators group for the machine on which the analytic server software is running. You can change the administrator group for the analytic server software by specifying it in the Admin Group text box in the administration application (IBM® SPSS® Statistics Administration Console). For more information, see the topic [Users](#) in Chapter 4 on p. 25. If you are using internal authentication on UNIX, you can create administrators directly. For more information, see the topic [Configuring Internal Authentication](#) on p. 46.

## Profiles

The analytic server software provides the ability to create profiles of individual users and groups of users. These user profiles and groups allow you to define settings for specific users. For more information, see the topic [IBM SPSS Statistics Server User Profiles and Groups](#) in Chapter 4 on p. 29.

## Client and Server Versions

You can run multiple versions of the analytic server software on a server computer. Please keep in mind that the client software must connect to server software that is at the same release level as the client. So, for example, a version 10.1 client must connect to a version 10.1 server—it cannot connect to an 11.0 server.

When you distribute connection information to end users, keep in mind which version of the client software they are running and be sure that they have the connection information for a matching server version.

## Connecting Users through a Firewall

If you use a **firewall** to keep your network secure from intruders, you can configure your firewall and the analytic server software so that end users outside of the firewall can connect the client to the analytic server software. Your firewall can use **NAT** (Network Address Translation), but it isn't required.

The typical scenario for connecting end users through a firewall that uses NAT is as follows:

- ▶ The end user connects the client application to the analytic server software using the **masqueraded IP** (the IP address that NAT presents to the outside world) and the server's port number. For example, the end user connects with IP 10.10.10.2 and port number 3016.
- ▶ The firewall allows the connection because it has been configured to accept connections from the masqueraded IP.
- ▶ The firewall redirects the masqueraded IP to the analytic server's real internal IP. It allows the connection because the port (for example, 3016) is enabled on the firewall.
- ▶ The analytic server spawns a process for the end user's client connection and assigns it a port number from the list in the system environment variable `STATISTICS_CLIENT_PORTS`. For example, the process communicates through port 3287.
- ▶ The firewall allows the communication through that port (for example, 3287) because it is enabled on the firewall.

## Configuring Connections Through a Firewall

To configure the analytic server software and the firewall, follow these steps:

- ▶ Install the analytic server software as usual. You need to know IP address of the computer on which the server is installed and the port number which the server software uses for communications. For example, install the server on 202.123.456.78 at port 3016.
- ▶ Configure the system environment variable `STATISTICS_CLIENT_PORTS` by specifying at least one port number. If needed, you can specify a comma-delimited list and a range of ports (for example, 3287, 3289, 3291-3295).

**Windows.** Use the Windows System properties to create and configure the variable. See [System Properties](#) on p. 59 for instructions.

**UNIX.** Edit the server software's environment script, *statsenv.sh*, which is included in the */bin* subdirectory of the installation directory, and define the port that can be used by the client processes that the server launches. For example, add the following lines:

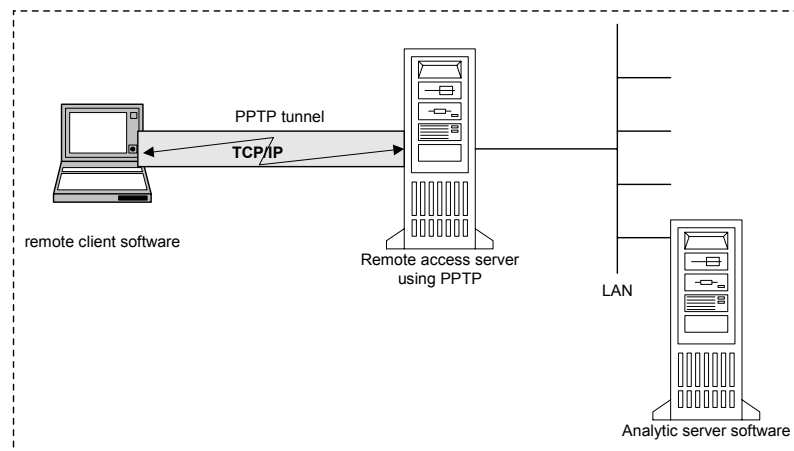
```
STATISTICS_CLIENT_PORTS=3287
export STATISTICS_CLIENT_PORTS
```

- ▶ If using Network Address Translation (NAT), create and map IPs. Using your firewall software, create a masqueraded IP for external use and map it to the server's internal IP. For example, create a masquerade IP 10.10.10.2 and map it to 202.123.456.78.
- ▶ Enable ports on the firewall. Using your firewall software, enable port numbers on the firewall, including:
  - The server's port number. For example, enable port 3016.
  - The port numbers that you specified in the `STATISTICS_CLIENT_PORTS` environment variable. For example, enable port 3287.
- ▶ Using your firewall software, make sure that *statisticsproc.exe* is allowed to accept network connections.
- ▶ Distribute connection information to the end users who will be connecting to the analytic server software from outside the firewall, including:
  - If used, the masqueraded IP of the computer on which the analytic server software is installed (do *not* distribute the server's internal IP). For example, distribute 10.10.10.2 as the server's IP.
  - Port number. Distribute the analytic server software's port number as usual. For example, distribute 3016 as the server's port number.

### ***Connecting Users with PPTP***

End users can connect a remote client computer to the analytic server software with **Point-to-Point Tunneling Protocol (PPTP)**. PPTP is a networking protocol that supports multiprotocol virtual private networks (VPNs). It enables remote end users to access your network securely across the Internet.

Figure 5-1  
PPTP Connections



**To use PPTP connections:**

- ▶ **Configure a remote access server for PPTP.** Be sure to create enough IP addresses for the clients because the server software supports multiple client connections. Each client connection requires its own IP address.
- ▶ **Configure the client desktop computer.** Use the Windows Network control panel to add a private network connection using PPTP. Enter an IP address that the remote access server will recognize as a PPTP connection.
- ▶ **Enable the PPTP connection on the client desktop computer.** When end users want to connect to the server software from a remote location, they enable the PPTP connection and then use the client software to connect to the server as usual.

## Using SSL to secure data transfer

Security Sockets Layer (SSL) is a protocol for encrypting data transferred between two computers. SSL ensures that communication between the computers is secure. SSL can encrypt the authentication of a username/password and the contents of an exchange between a server and client.

### How SSL works

SSL relies on the server's public and private keys, in addition to a public key certificate that binds the server's identity to its public key.

- ▶ When a client connects to a server, the client authenticates the server with the public key certificate.
- ▶ The client then generates a random number, encrypts the number with the server's public key, and sends the encrypted message back to the server.
- ▶ The server decrypts the random number with its private key.

- ▶ From the random number, both the server and client create the session keys used for encrypting and decrypting subsequent information.

The public key certificate is typically signed by a certificate authority. Certificate authorities, such as VeriSign and Thawte, are organizations that issue, authenticate, and manage security credentials contained in the public key certificates. Essentially, the certificate authority confirms the identity of the server. The certificate authority usually charges a monetary fee for a certificate, but self-signed certificates can also be generated.

### **Securing client-server and server-server communications with SSL**

The main steps in securing client-server and server-server communications with SSL are:

- ▶ Install OpenSSL on the server computer.
- ▶ Obtain and install the SSL certificate and keys.
- ▶ Enable and configure SSL in the server administration application (IBM® SPSS® Collaboration and Deployment Services Deployment Manager).
- ▶ If desired, install unlimited strength encryption on the client computers.
- ▶ If using a self-signed certificate, copy the certificate on the client computer.
- ▶ Instruct end users to enable SSL when connecting to the server.

*Note:* Occasionally a server product acts as a client. An example is IBM® SPSS® Statistics Server connecting to the IBM® SPSS® Collaboration and Deployment Services Repository. In this case, SPSS Statistics Server is the *client*.

### **Install OpenSSL**

If OpenSSL is not already available on the server, you must install it.

- ▶ Download OpenSSL from <http://www.openssl.org/>. Be sure to use the version of OpenSSL appropriate for the server version:

Table 5-1  
*OpenSSL versions*

<b>Server Product</b>	<b>Compatible OpenSSL Version</b>
IBM® SPSS® Statistics 17-19 ( <i>not</i> Linux® on System z®) IBM® SPSS® Modeler 14.1 ( <i>not</i> Linux® on System z®)	0.9.8 and its subversions (0.9.8a, 0.9.8b, and so on)
SPSS Statistics 19 (Linux® on System z®) SPSS Modeler 14.1 (Linux® on System z®)	1.0.0

- ▶ Follow the instructions for installing and configuring the software. If you build OpenSSL yourself, be aware of the following guidelines:

**Windows.** OpenSSL should be built with DLLs (which are multithreaded by default).



**UNIX.** OpenSSL should support multiple threads (which is not always by default) and shared libraries.

- ▶ Make sure the OpenSSL modules are included on the system path.

### ***Obtaining and installing SSL certificate and keys***

- ▶ Obtain an SSL certificate and key file. There are two ways you can do this:
  - Purchase them from a public certificate authority (such as VeriSign or Thawte). The public certificate authority signs the certificate to verify the server that uses it.
  - Generate the key and certificate files with an internal self-signed certificate authority. OpenSSL provides a certificate management tool for this purpose.
- ▶ Copy the certificate and key file or files to a local directory on the server. The public and private keys can be stored in one file. Ensure that access permissions deny casual browsing of this directory.
- ▶ Copy *root.pem* to the following location on the server computer. If you would like to use another location, create an `SSL_CERT_DIR` environment variable and set the value of the variable to the location.

**Windows.** *C:\Documents and Settings\All Users\Application Data\SPSSInc\certificates*

**Mac.** */Library/Application Support/SPSSInc/certificates*

**UNIX and Linux.** */opt/SPSSInc/certificates*

### ***Enable and Configure SSL in the IBM SPSS Collaboration and Deployment Services Deployment Manager***

- ▶ Start the server administration application (IBM® SPSS® Collaboration and Deployment Services Deployment Manager) and connect to the server.
- ▶ On the configuration page, set Secure Sockets Layer to Yes.
- ▶ In SSL Public Key File, specify the full path to the public key file.
- ▶ In SSL Private Key File, specify the full path to the private key file.

*Note:* If the public and private keys are stored in one file, specify the same file in SSL Public Key File and SSL Private Key File.

- ▶ From the menus choose:  
File > Save
- ▶ Restart the server service or daemon. When you restart, you will be prompted for the SSL password. On Windows, you can select Remember this password to store the password securely. This option eliminates the need to enter the password every time the server is started.

### **Installing unlimited strength encryption**

The Java Runtime Environment shipped with the product has US export-strength encryption enabled. For enhanced security of your data, we recommend that this is upgraded to unlimited-strength encryption.

- ▶ Download the Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files 5.0 from <http://java.sun.com/javase/downloads/index.jsp>.
- ▶ Unzip the downloaded file.
- ▶ Copy the two *.jar* files *local\_policy.jar* and *US\_export\_policy.jar* into *<installation folder>/jre/lib/security*, where *<installation folder>* is the folder in which you installed the product.

### **Copying the certificate file to client computers**

*Note:* Skip this step if you are using a certificate that is signed by a certificate authority.

If you are using a self-signed certificate, you need to copy the certificate to the *client* computers. Be aware that a server computer may also act as a client. An example is IBM® SPSS® Statistics Server connecting to the IBM® SPSS® Collaboration and Deployment Services Repository. In this case, SPSS Statistics Server is the *client*, and therefore you need to copy the certificate for the IBM SPSS Collaboration and Deployment Services Repository server to the SPSS Statistics Server.

- ▶ Copy *root.pem* to the following location on the client computers. By default, all SPSS Inc. client products look in this location for trusted self-signed certificate files. If you would like to use another location, create an `SSL_CERT_DIR` environment variable and set the value of the variable to the location.

**Windows.** *C:\Documents and Settings\All Users\Application Data\SPSSInc\certificates*

**Mac.** */Library/Application Support/SPSSInc/certificates*

**UNIX and Linux.** */opt/SPSSInc/certificates*

If you already copied a *root.pem* file to the client for another SPSS Inc. product, append the certificate information from the new server to the existing *root.pem* file. This file is a text file so you can copy and paste the certificate.

### **Adding the certificate to client keystore (for connections to the repository)**

*Note:* Skip this step if you are using a certificate that is signed by a certificate authority.

If you are using SSL to connect to the repository and you are using self-signed certificates, you need to add the certificate to the client's Java keystore. The following steps are completed on the client computer.

- ▶ Open a command prompt and change directories to the following location, where *<product install dir>* is the directory in which you installed the product:

```
<product install dir>/jre/bin
```

- ▶ Enter the following command:

```
keytool -import -alias <alias name> -file <path to cert> -keystore <path to key store>
```

Where *<alias name>* is an arbitrary alias for the certificate, *<path to cert>* is the full path to the certificate, and *<path to key store>* is the full path to the Java keystore, which may be *<product install dir>/lib/security/jssecacerts* or *<product install dir>/lib/security/cacerts*.

- ▶ When prompted, enter the keystore password, which is `changeit` by default.
- ▶ When prompted about trusting the certificate, enter `yes`.

### **Instructing end users to enable SSL**

When end users connect to the server through a client product, they need to enable SSL in the dialog box for connecting to the server. Be sure to tell your users to select the appropriate check box.

## **Other Authentication Options on UNIX**

On UNIX, there are additional options for authenticating users. Instead of using standard UNIX accounts, you can use the Pluggable Authentication Module (PAM) or internal authentication. Internal authentication allows the analytic server software to run without root privileges, but it results in restrictions on client connections.

### **Configuring PAM**

The analytic server software on UNIX can use the Pluggable Authentication Module (PAM) to authenticate users. You must first configure the analytic server software to use PAM. Then you configure PAM by following the instructions specific to your UNIX vendor. Steps follow for Linux and Solaris. These may vary, depending on version and vendor.

*Note:* If the analytic server software is running, you need to restart it after completing all the steps.

#### **Configure the Analytic Server Software to Use PAM**

- ▶ Log on to the UNIX machine as *root*.
- ▶ In the *bin* directory in the analytic server software installation directory, open the configuration file (e.g., *statisticsd.conf*) in a text editor.
- ▶ Find the *userauth* element and change the *value* parameter from `unix` to `pam`.
- ▶ Save the file.

**Configure PAM on Linux**

- ▶ Change to the PAM configuration directory (e.g., */etc/pam.d*).
- ▶ Use a text editor to create a file named *statisticsd*.
- ▶ Add the PAM configuration information that you want to use. For example:

```

auth      required /lib/security/pam_stack.so service=system-auth
auth      required /lib/security/pam_nologin.so
account   required /lib/security/pam_stack.so service=system-auth
password  required /lib/security/pam_stack.so service=system-auth
session   required /lib/security/pam_stack.so service=system-auth
session   required /lib/security/pam_limits.so
session   optional /lib/security/pam_console.so

```

*Note:* These lines may vary depending on your particular configuration. Consult the Linux documentation for more information.

- ▶ Save the file.

**Configure PAM on Solaris**

- ▶ Open */etc/pam.conf* in a text editor.
- ▶ Add configuration information to enable PAM for *statisticsd* (the IBM® SPSS® Statistics Server daemon process) and *statisticsproc* (user processes). For example, add the following lines before the first line that begins with *other auth*.

```

statisticsd  auth requisite pam_authtok_get.so.1
statisticsd  auth required pam_dhkeys.so.1
statisticsd  auth sufficient pam_unix_auth.so.1
statisticsd  auth required pam_krb5.so use_first_pass

```

*Note:* These lines may vary depending on your particular configuration. Consult the Solaris documentation for more information.

- ▶ Save the file.

**Configuring Internal Authentication**

Internal authentication allows the analytic server software to run without root privileges. However, it restricts client connections. Every user who connects to the analytic server software has the same disk-access security. Therefore, one user can delete another user's file.

To configure internal authentication, complete the following steps:

- ▶ Create a group for users who will connect to the analytic server software. We recommend naming this group **statistics**.
- ▶ A member of this group must install the analytic server software. This user will be the owner of the analytic server software daemon.
- ▶ Another member of this group (different from the daemon owner and typically the user who maintains the analytic server software users) creates an *statisticsusers* file in the *bin* directory in

the analytic server software installation directory. This file should have read/write access for the user who created it. It should have read access for the users group. No other users should be able to access it. If you don't create this file manually, it is automatically created the first time you run the `statisticsuser` command line tool (see the next step). The command line tool sets the appropriate permissions.

- ▶ In the *bin* directory, use the `statisticsuser` command line tool to add users. As the user who created the *statisticsusers* file, type `statisticsuser <username>` to create a regular user (e.g., `statisticsuser jdoe`). Use the `-a` option to create an admin user (e.g., `statisticsuser -a jdoe`). The `statisticsuser` command line tool prompts you for a password. An end user enters the user name and password to connect to the analytic server software. Be sure to distribute the user name and passwords appropriately. To delete a user, use the `-d` option (e.g., `statisticsuser -d jdoe`).
- ▶ Logged in as the owner of the analytic server software daemon, open the configuration file (e.g., *statisticsd.conf*) in a text editor.
- ▶ Find the *userauth* element and change the *value* parameter from `unix` to `internal`.
- ▶ Logged in as the owner of the analytic server software daemon, start the server.

## Setting a Locale

The analytic server software and the client that connects to it must run in the same character set, encoding, and locale. The analytic server software gets its locale from the client. By default, this is the client's *system* locale. However, the client can override the default to process data files in other locales. By overriding the default, the user instructs the analytic server software to run in a specified locale without changing the client's system locale.

### Syntax

The user overrides the default by using the `SET LOCALE` syntax command:

```
SET LOCALE="localeid"
```

*localeid* is a string that identifies the locale in which the analytic server software will run. `SET LOCALE` writes a registry entry on the client machine. This entry persists so that the next time IBM® SPSS® Statistics is started on the client machine, SPSS Statistics will run in that locale.

The naming convention for the locale ID can differ among platforms and vendors. Therefore, there is an XML file installed with the analytic server that maps client locales to server locales. This file, *loclmap.xml*, is located in the analytic server installation directory on Windows and the */bin* subdirectory on UNIX.

### loclmap.xml

The root element in *loclmap.xml* is the following. The root element also identifies the schema location.

```
<locale-map xmlns="http://xml.spss.com/spss/mls"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://xml.spss.com/spss/mls
```

```
http://xml.spss.com/spss/mls/locale-map-1.0.xsd">
```

The root element contains `<client-locale>` elements with a `name` attribute identifying the client locale. The `<client-locale>` elements contain one or more `<server-locale>` elements. Each `<server-locale>` element has a `name` attribute identifying a server locale that corresponds to the client locale. The analytic server software translates the client locale ID into one that can be used on the server machine. It checks each server locale in order, until it finds one that is valid on the server machine.

None of the default server locales in *loclmap.xml* are Windows locales. Windows system locales are generally not needed because the analytic server software first tries to use the same locale as the client's system locale. A Windows server should have the locale that matches the client locale. However, you can add Windows server locales to *loclmap.xml* if you need to substitute a different but similar Windows locale.

You can modify *loclmap.xml* as needed. Just be aware that your XML elements must validate against the schema.

### Example

Following is an example of the contents of *loclmap.xml*:

```
<client-locale name="Japanese_Japan.932">
  <server-locale name="ja_JP.sjis"></server-locale>
  <server-locale name="ja_JP.PCK"></server-locale>
  <server-locale name="ja_JP.SJIS"></server-locale>
  <server-locale name="Ja_JP.IBM-943"></server-locale>
</client-locale>
```

In this case, if the user issues `SET LOCALE="Japanese_Japan.932"` and the server machine is a Solaris machine, the analytic server software would check `ja_JP.PCK` first. This locale works on Linux but not Solaris, so the analytic server software checks the next locale. Because `ja_JP.PCK` is a locale ID recognized by Solaris, the analytic server software uses that locale.

### Using a Server Locale

If the user issues `SET LOCALE` using a server locale ID not recognized on the client machine, the client machine uses *loclmap.xml* to find client locale ID associated with a server locale ID. It writes this locale ID to the registry. For example, if the user issues `SET LOCALE="ja_JP.PCK"`, `Japanese_Japan.932` is written to the registry. To see which entry in *loclmap.xml* applies to the client, you can run the `SHOW LOCALE` command in local mode.

### Potential Issues

Be aware that using the `SET LOCALE` command can cause functional problems in some cases:

- The current variable names might not be legal in the new code page.
- Case-insensitive name matches might fail. The failure might occur because strings are converted to upper-case characters in case-insensitive name matches (for example, when comparing variable names). If the locale is incorrect, this conversion would change the character (for example, in the Central European code page, 1250).

- Some bytes could be interpreted incorrectly as lead bytes, and a problem might occur because of an unexpected trail byte.
- `SET LOCALE` does not change the client's system locale. Therefore, if the SPSS Statistics locale associated with `SET LOCALE` is different from the client's system locale, there will be display problems in various places. In this situation, a user is also unable to use an Input Method Editor (IME) to enter national characters.
- The SPSS Statistics locale, `OLANG` setting, and the encoding used for the data must be compatible. Otherwise, output might be unusable and unreadable.

## ***Accessing Data and Files***

When end users connect to the analytic sever software, their view of data sources and files is from the perspective of the server computer, not their desktop computers.

- **ODBC data sources.** If your end users need access to ODBC data sources defined on the server computer, distribute the names, descriptions, and login information for those data sources. See [Chapter 3, \*Data Access\*](#), for a discussion of database access from the analytic server software.
- **File access.** Distribute the names and locations of the files on the server computer that you want end users to access. For more information, see the topic [Referencing Data](#) in Chapter 3 on p. 10.

## ***Saving Data and Files***

When end users save files while they are connected to the analytic server software, the default location for the save is the directory from which the file was opened. In many cases, this is the local desktop computer; however for data files, it will often be a write-protected location on the server computer. Tell users where to save data files. Typically, the location is the user's home directory somewhere on your network.

*UNIX note:* Tell end users to use the full file specification and forward slashes when saving files (for example, `/public/myhome/myserverdata/data.sav`). Avoid using the backslash character in the UNIX directory and filenames used with the analytic server software.

# Analyzing and Improving Performance

If you need to improve the performance of the analytic server software, refer to this chapter for various strategies, ranging from configuration changes to hardware upgrades. Before making these changes, obtain performance information so you know which areas are problematic.

## Obtaining Performance Information

To check performance, compare usage in the following areas when the analytic server is not being used to when it is being heavily used.

- Disk usage
- CPU usage
- Memory usage
- Network usage

### Logging

The administration application (IBM® SPSS® Statistics Administration Console) allows you to configure the analytic server software to log performance information. Using the Performance Log Interval node, you can specify how often the analytic server software writes performance information to the log. For more information, see the topic [Logging](#) in Chapter 4 on p. 28. You can also get performance information directly from the operating system.

### Getting Performance Information on Windows

On Windows, you can obtain performance information by using the Performance Monitor.

Table 6-1  
*Windows Performance Information*

Area	Windows Performance Monitor Object	Useful Counter
Disk Usage	PhysicalDisk	% Idle Time
CPU Usage	Processor	% Processor Time
Memory Usage	Memory	Committed Bytes
Network Usage	Network Interface	Bytes Total/sec for each physical interface instance (not the MS TSP loopback interface)



### Getting Performance Information on UNIX

On UNIX, there are various commands for obtaining performance information, depending on the vendor.

Table 6-2  
UNIX Performance Information

Area	Vendor	Command	Note
Disk Usage	Linux	iostat -x	Check the %util column.
	Solaris	iostat -D	Check the util column.
	AIX	iostat -d	Check the % tm_actl column.
CPU Usage	Linux	top	Use the P interactive command to sort by CPU usage.
	Solaris	top	Use the o interactive command and then sort by cpu.
	AIX	ps aux	Pipe to the sort command to sort by the %CPU column.
	All	uptime	Check the load average.
Memory Usage	Linux	top	Use the M interactive command to sort by CPU usage.
	Solaris	top	Use the o interactive command and then sort by size.
	AIX	ps aux	Pipe to the sort command to sort by the RSS column.
Network Usage	All	netsat	Use the -i and -s switches for information.

#### Next Step

After gathering this information, you should be able to identify the area or areas that are problematic. The following sections describe possible solutions and recommendations for each area.

## Improving Disk Usage

Consider the following for improving disk usage.

**Space.** Allow enough disk space. Each user typically needs temporary disk space equal to twice the size of the data (SAV) file in use (the space needed ranges from 1 to 2.5 times). A user sorting a file might need temporary space more than three times the size of the file. For example, if six concurrent users are accessing a file and two are sorting at once, they might need as much as 17 times the size of the file. In practice, they will not be at peak usage simultaneously, so 12 times the size of the file would be sufficient.

**Hardware.** Use SCSI disks for fastest performance. Do not use IDE.

**System configuration.** Keep temporary files on a separate spindle. You can also define multiple temporary files locations using the administration application. For more information, see the topic [File Locations](#) in Chapter 4 on p. 25. Make sure each location is on a separate spindle. If you use RAID, use RAID0 for the temporary file spindle. The scratch files speed gained from

RAID0 is preferred over the redundancy gained from RAID1. If your CPU is not a problem and the analytic server computer runs Windows, you can also compress the data directory or data files on disk. Do not allocate more virtual memory.

**SPSS Statistics configuration.** If memory is not a problem but disk usage is, increase the workspace in IBM® SPSS® Statistics for faster performance. Try setting it by dividing the amount of RAM on the analytic server computer by the expected number of concurrent users. For example, if the analytic server computer has 1 GB of RAM, set the workspace to 0.25 GB RAM.

**Temporary files directory.** Modify the user profile or groups settings so that the temporary files directories for each user are located on different physical drives. For more information, see the topic [IBM SPSS Statistics Server User Profiles and Groups](#) in Chapter 4 on p. 29.

**Cache compression.** If your users are consistently working with large data files (especially if the size of the files is greater than half of the server's RAM), try enabling cache compression in the administration application. For more information, see the topic [Users](#) in Chapter 4 on p. 25.

## ***Improving CPU Usage***

Consider the following for improving CPU usage:

**Number.** Add more processors. If you want to approach the speed that a user would experience when running IBM® SPSS® Statistics locally, try to have one processor for every two concurrent users. Also use processors that are as fast as, or faster than, the processor on the desktop computer. For example, if you expect an average of four concurrent users, configure the analytic server computer with two fast processors.

**Hardware.** Use fast processors. Adding a few really fast processors is better than adding many slow ones. If CPU usage is still a problem with fast processors, consider adding more analytic server computers to your system.

**SAV file locations and access.** If certain files are used often by many concurrent users, consider moving the files across multiple servers to balance the user load. For example, if *TestScores.sav* and *GPA.sav* are both heavily used, put them on separate servers. Control access to the files with the operating system permissions (per group or per user) instead of controlling access through server accounts.

**CPU priority.** If certain users need a higher CPU priority than other users (for example, users who run quick jobs versus those who run long jobs), modify the user profile or group settings. For more information, see the topic [IBM SPSS Statistics Server User Profiles and Groups](#) in Chapter 4 on p. 29.

**Cache compression.** Cache compression has some CPU overhead for the compression and decompression of scratch files. If your users are not working with large data files, you may want to consider turning it off. For more information, see the topic [Users](#) in Chapter 4 on p. 25.

## ***Improving Memory Usage***

Consider the following for improving memory usage:

**Amount.** Add as much RAM as possible. Try to have 128 MB of RAM for each concurrent user. So, if there are four concurrent users, configure the server with 512 MB of RAM.

**SPSS Statistics configuration.** Decrease the workspace in IBM® SPSS® Statistics.

## ***Improving Network Usage***

Consider the following for improving network usage:

**System configuration.** Schedule network-intensive operations for times when the analytic server software is not in use (for example, run system backups overnight). If you identify a problem with network traffic on a computer on which the analytic server is running, SPSS Inc. will work with you to diagnose the problem further.

## ***Using IBM SPSS Statistics Efficiently***

In addition to focusing on specific problematic areas, you can also improve performance by adhering to the following guidelines for using IBM® SPSS® Statistics efficiently.

**Data management.** If you have large data files that require regular updating and are shared by users, consider doing the updates once and then releasing the files to the users for analysis. For example, if you regularly add monthly data to a file, sort it, and perform transformations, designate one person to run the job on the file. The other users can get the data they need without having to repeat the merge, sort, and transformations.

**Interactive vs. batch.** If you have regular, time-consuming operations that you do with SPSS Statistics, consider running them from the SPSS Statistics Batch Facility rather than from a client connected to the analytic server. Use the client for building the reports, and run them from the SPSS Statistics Batch Facility after the reports are ready.

---

# Troubleshooting

## **Analytic Server Software**

**Port number conflict.** If there is a port number conflict, the analytic server software may fail to start. Correct the problem by using the administration application (IBM® SPSS® Statistics Administration Console) to change the port number. For more information, see the topic [Connections](#) in Chapter 4 on p. 24. Be sure to distribute the new port number to the end users.

**Erratic behavior.** The analytic server software may behave erratically if its configuration file (for example, *statisticsd.conf*) is corrupted or missing. To correct the problem, restore the configuration file from your backup copy. Copy it to the location specified in the administration application or the configuration file environment variable and restart the server software. For information about restarting, see [Starting and Stopping the Analytic Server Software](#) in Chapter 4 on p. 35.

**Administration application doesn't work (UNIX only).** If you use the administration application to control or configure the server software and it doesn't work (for example, you cannot stop the server), it may be because you did not start the server software with the startup script provided by SPSS Inc.. Correct the problem by starting the server software with the *start\_statistics\_server* startup script. For more information, see the topic [To Stop the Service or Daemon](#) in Chapter 4 on p. 35. If you get an error message when you attempt to control or configure the server software, it may be because you have connected with an account that does not have administrator permissions. For more information, see the topic [Users](#) in Chapter 4 on p. 25.

**Can't change the location of temporary files (UNIX only).** If you use the administration application to change the location of temporary files and the change is not effective, it may be because the new location doesn't have sufficient file permissions for the end users. Choose a location that has **read**, **write**, and **execute** access for all of the users who will connect to the analytic server software.

**Server won't start (UNIX only).** If the server software will not start, it may be because you do not have the required operating system patches. To correct the problem, download and install the appropriate patch. The required patches are listed in the UNIX installation instructions for your server product.

## **Client Software**

**End user cannot connect to the analytic server.** The user may not have adequate permissions, or the firewall may be blocking the analytic server software. For information about user permissions, see [Permissions](#) on p. 38. For information about configuring the firewall, see [Configuring Connections Through a Firewall](#) on p. 39.

**End user login fails with the “specified remote server computer was not found” message.** The service or daemon may not be running. Confirm this by checking the status of the analytic server software. To correct the problem, restart the service or daemon. For more information, see the topic [Starting and Stopping the Analytic Server Software](#) in Chapter 4 on p. 35.

**End user login fails with the “error connecting to package” message.** The end user has specified the name or IP address of a server computer that isn’t on the network. To correct the problem, ask the end user to enter a valid server name.

**DataDirect ODBC data source fails with the “not licensed” message.** The DataDirect data access technology is distributed with SPSS Inc. products. It works only with newer SPSS Inc. products—it doesn’t work with earlier versions, nor does it work with non-SPSS Inc. applications. If end users attempt to use DataDirect data sources with an older or unlicensed product, they will get a message containing the text You are not licensed to use the DataDirect ODBC Driver. To correct the problem with the SPSS Inc. product, upgrade your users to a current version. To correct the problem with unlicensed products, upgrade your licensing with DataDirect or ask end users not to attempt to use the data sources that you have defined for SPSS Inc. products with unlicensed applications.

**End user can’t find a data file or ODBC data source.** When end users are running in distributed analysis mode, they will have access only to data files and ODBC data sources on the computer that is running the analytic server software. When end users are running in local analysis mode, they will have access only to data files and ODBC data sources on their desktop computers. To correct the problem, ask the end user to run the client application in the appropriate mode.

**End user can’t run a statistical procedure (SPSS Statistics Server only).** When end users are connected to the analytic server software, they have access only to the IBM® SPSS® Statistics options that were installed during the SPSS Statistics Server installation. To correct the problem, ask the end user to run the procedure while in local analysis mode or install the requested procedure on the server computer.

# ***The IBM SPSS Statistics Batch Facility***

*Note:* The IBM® SPSS® Statistics Batch Facility is a batch processing utility included with **SPSS Statistics Server**.

Typically the client for SPSS Statistics Server is SPSS Statistics running on a desktop computer. However, the SPSS Statistics Batch Facility is an alternative way to use the power of SPSS Statistics Server, and it runs on the server computer. The SPSS Statistics Batch Facility is intended for **automated production** of statistical reports. Automated production provides the ability to run analyses without user intervention. Automated production is advantageous if users at your site regularly require a set of time-consuming analyses, such as weekly reports.

The SPSS Statistics Batch Facility takes as its input a report request contained in a **command syntax** file. The SPSS Statistics Batch Facility then automatically produces the statistical reports specified by the syntax.

## ***What You Need to Know***

**Operating systems.** The IBM® SPSS® Statistics Batch Facility is currently available with all SPSS Statistics Servers, UNIX and Windows.

**Installation.** The SPSS Statistics Batch Facility is automatically installed in the SPSS Statistics Server installation directory on Windows and the */bin* subdirectory of the installation directory on UNIX.

**Invoking.** The SPSS Statistics Batch Facility is run from the command line using the *statisticsb* executable file. It runs independently of SPSS Statistics Server—SPSS Statistics Server does not have to be started for it to run. It can also be run concurrently with SPSS Statistics Server.

**Modes of operation.** Commands are submitted to the SPSS Statistics Batch Facility in either **batch** or **interactive mode**. In batch mode, the analyst or IT professional submits a command syntax file to the SPSS Statistics Batch Facility for execution—the commands in the file are read and acted upon as a batch, and output is directed to a file. The SPSS Statistics Batch Facility runs unattended and terminates after executing the last command. This is the typical way to use the SPSS Statistics Batch Facility. In interactive mode, the analyst types commands one at a time at a command prompt. The commands are executed immediately, and output is displayed in the window. The SPSS Statistics Batch Facility waits for the next command.

**Documentation.** The user's guide, written for the analysts and IT professionals at a site who will be using the SPSS Statistics Batch Facility, is on the SPSS Statistics Server DVD in */Documentation/<language>/Manuals*. The command syntax reference guide that analysts will need in order to create command syntax files for the SPSS Statistics Batch Facility is on the SPSS Statistics Server DVD in */Documentation/<language>/Manuals*. The SPSS Statistics

Batch Facility for UNIX is also distributed with a manual page, *statisticsb.1*, which is in the */bin* subdirectory of the SPSS Statistics Server installation directory. If you are administering a UNIX system, copy it to the location where you keep your manual pages.

**Additional documentation.** The SPSS Statistics Batch Facility user's guide contains sufficient information for an analyst who is experienced with the SPSS Statistics command syntax language to build command syntax files for the SPSS Statistics Batch Facility. If the analysts at your site are new to SPSS Statistics, they may require additional documentation. If they do, direct them to our Web site at <http://www.spss.com>, or ask them to contact your sales representative.

# Windows Operating System Tasks

You can do most administrative tasks with the administration application (see [Using the Administration Application](#) in Chapter 4 on p. 17); however a few tasks may need to be done with the Windows operating system. Use the following operating system features to administer analytic server software running on Windows:

- **File properties.** Used to set end-user access to the server software's installation directory, the temporary file location, and data files.
- **System properties.** Used to create environment variables.
- **User manager.** Used to create end-user accounts.
- **Services Control Panel.** Used to start, stop, and configure the service.
- **ODBC Administrator.** Used to configure data sources.

## File Properties

Use File Properties to set permissions on files. For data files, how you do this depends on where the data are stored. When you store data on the same computer as the analytic server software, you control access to the data directory by setting permissions on a directory on an NTFS drive.

### **On the server computer, logged on as an administrator:**

- ▶ Use the Windows Explorer to navigate to the data directory.
- ▶ Click the directory, right-click, and click Sharing on the context menu.
- ▶ Click the Security tab and configure the permissions.

*Note:* The Security tab is available only on NTFS drives. If you're not sure what type of file system your hardware uses, follow these steps:

- ▶ Use the Windows Explorer to navigate to the drive.
- ▶ Click the drive, right-click, and click Properties on the context menu.
- ▶ Click the General tab and look at the value for File System.

When you store data on a computer on your network, you can control access to the data directory by creating a shared resource and setting the permissions appropriately.

### **On the networked computer, logged on as an administrator:**

- ▶ Use the Windows Explorer to navigate to the data directory.



- ▶ Click the directory, right-click, and click Sharing on the context menu.
- ▶ Click the Sharing tab in the dialog box, click Shared As, enter a share name, and set the appropriate access.

## ***System Properties***

Use system properties to create environment variables.

### **On the server computer, logged on as an administrator:**

- ▶ On the Windows desktop, right-click on the icon for the computer. For example, right-click on My Computer.
- ▶ Select Properties from the menu.
- ▶ Click the Advanced tab and then click Environment Variables.
- ▶ Click New.
- ▶ Type the name of the new variable.
- ▶ Type the value for the new variable.

## ***User Manager***

Use the User Manager to create end-user accounts.

### **On the server computer, logged on as an administrator:**

- ▶ From the Windows Start menu choose:  
Programs > Administrative Tools  
  
Select Computer Management and then Local Users and Groups.
- ▶ Create the user accounts.

## ***Services Control Panel***

Use the Windows Services Control Panel to:

- Stop and start the service.
- Change service startup parameters.
- Check the server status.

### **To access and use the Services Control Panel:**

- ▶ From the Windows Start menu choose:  
Settings > Control Panel
- ▶ Select Administrative Tools and then Services.

- ▶ Select the service. You can now check its status, start or stop it, and edit startup parameters.

*Note:* You can start, stop, and check the status of the server software with the administration application. For more information, see the topic [Controlling the IBM SPSS Statistics Server](#) in Chapter 4 on p. 23.

## **Task Manager**

Use the Task Manager to see how many server-related processes are running.

- ▶ Open the Windows Task Manager by pressing Ctrl-Alt-Delete and choosing Task Manager.
- ▶ Click the Processes tab.
- ▶ Click Image Name to sort the processes alphabetically.
- ▶ Search for the filename of the server process (*statisticsrvr.exe*).
- ▶ Search for the filename of the client process (*statisticsproc.exe*). There is one process for each end user currently connected to the analytic server software.

*Note:* You can monitor server and client processes with the administration application. For more information, see the topic [Monitoring IBM SPSS Statistics Server Users](#) in Chapter 4 on p. 33.

## **ODBC Administrator**

Use the ODBC Administrator to configure system and user data sources for use with the analytic server software.

How the ODBC data source is created affects who can view and use it. Use *system* DSNs when you want to allow general access to the data source. Use *user* DSNs when you want to restrict access to sensitive information or when you want to tailor the DSN for a specific user.

### **To Configure a System DSN**

System DSNs can be used by anyone logged on to the computer on which they are defined. System DSNs are easier for you to configure and administer because you do it just once for all users.

**On the computer on which you want the data source to reside, logged on as an administrator:**

- ▶ From the Windows Start menu choose:  
Settings > Control Panel
- ▶ Select Administrative Tools and then Data Sources.
- ▶ Click the System DSN tab.
- ▶ Click Add.
- ▶ Select a driver from the list. If you are configuring a data source that uses the SPSS Inc. data access technology, the Connect ODBC driver names are labeled with the text SPSS Inc. OEM.
- ▶ Click Finish.

- ▶ Enter the appropriate information into the Driver Setup dialog box.
- ▶ Click OK.

### ***To Configure a User DSN***

User DSNs can be used only by the account of the user who created them. Configure user DSNs when you want to restrict access to sensitive information or when you want to tailor the DSN for a specific user.

Log in as the user and follow the steps for a system DSN, with this exception:

- Click the User DSN tab instead of the System DSN tab.

# *UNIX Operating System Tasks*

You can do most administrative tasks with the administration application (see [Using the Administration Application](#) in Chapter 4 on p. 17); however a few tasks may need to be done with the UNIX operating system. Use the following operating system features to administer analytic server software running on UNIX:

- **chmod**. Used to set end-user access to data files.
- **env**. Used to check values of environment variables.
- **scripts**. Used to start the server software and configure its environment.
- **ps** and **kill**. Used to check and stop server processes.
- **odbc.ini**. Used to configure ODBC data sources.

## *chmod*

Use the `chmod` (or `chown`) command to change or assign the permissions mode for directories and data files. For example, to set the `/usr/data` directory to read-only for everyone:

- ▶ Log in as a super-user or as the owner of the directory.
- ▶ At the UNIX prompt, type:  
`chmod a-w /usr/data`

## *env*

Use the `env` command to check the current values of environment variables. For example, to use `env` to check the current values of environment variables for the server software:

- ▶ Log in as the account that started the daemon, typically `root`.
- ▶ At the UNIX prompt, type:  
`env`
- ▶ Check the settings for the variable(s) of interest.

## Scripts

To change value of environment variables, edit the script that starts the server software. To edit the startup script:

- ▶ Use a text editor to open the *statsenv.sh* script, which is included in the */bin* subdirectory of the IBM® SPSS® Statistics Server installation directory. For example, open */usr/local/myserverproduct/bin/statsenv.sh*.
- ▶ If necessary, uncomment the line that defines the variable and then enter the new value for the variable.
- ▶ Save the file.

*statsenv.sh* is called by the *start\_statistics\_server* script. Environment variables set and exported in *statsenv.sh* affect only the processes started with the *start\_statistics\_server* script.

## ps and kill

Use the `ps` command to get information about what server processes are running and to report process status. For example:

- ▶ At the UNIX prompt, type:  
`ps -efl.`
- ▶ Search for the filename of the daemon process (for example, *statisticsd*). This process has the UID of the user who started the analytic server software daemon process (usually *root*).
- ▶ Search for the filename of the client process, *statisticsproc.exe*. There is one process for each end user currently connected to the analytic server software. The *UID* column displays the login ID of the end user who owns the client process.

Use the `kill` command to kill a process. For example:

- ▶ Log in as the user that started the daemon.
- ▶ At the UNIX prompt, type:

```
kill -9 pid
```

where *pid* it the process id of the process.

The analytic server software daemon also automatically creates a file that contains its process ID. Instead of manually finding the PID with the `ps` command, you can use this file in conjunction with the `kill` command directly to kill the daemon process directly:

```
kill -9 `cat statisticsd.pid`
```

*Note:* If you want to use the administration application to monitor and kill processes, you must start the server software with the startup script provided by SPSS Inc.. For more information, see the topic [Controlling Service Startup](#) in Chapter 4 on p. 34.

## ***odbc.ini***

You may need to configure ODBC data sources on the server computer if:

- You are using the SPSS Inc. Data Access Pack

and

- The analytic server software needs to access databases

No ODBC administrator exists on UNIX. To configure an ODBC data source on UNIX, you edit a system information text file, *odbc.ini*. *Odbc.ini* is installed when you install the data access pack for UNIX. Installation instructions appear in *SPSS Inc. Data Access Pack Installation Instructions for Unix.pdf* (the document is located in the */Documentation/<language>/InstallationDocuments* directory on the product DVD). Be sure to install the additional documentation so that you have access to the documents listed below.

**Connect ODBC.** For information about editing your *odbc.ini* file and setting important environment variables, see the “Configuring Drivers and Data Sources” section in the “Installation on UNIX” chapter of the *Connect ODBC Installation Instructions* for detailed instructions.

DataDirect’s product documentation for Connect ODBC is included, by default, as part of the IBM® SPSS® Data Access Pack installation. The installer creates the entry SPSS Inc. OEM Connect and ConnexXE for ODBC along with the entries for your other programs on the Start menu. The DataDirect product documentation is accessed from this menu item.

DataDirect’s product documentation for Connect ODBC can be found under the directory where you extracted the files. Online help is located in the *odbchelp* directory, and PDF versions are located in *books*.

*Note:* The documentation can also be accessed from the SPSS Inc. Web site at <http://www.spss.com/drivers/merant.htm>, or from the DataDirect home page at <http://www.datadirect.com/index.ssp>.

# Notices

Licensed Materials – Property of SPSS Inc., an IBM Company. © Copyright SPSS Inc. 1989, 2010.

Patent No. 7,023,453

**The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:** SPSS INC., AN IBM COMPANY, 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. SPSS Inc. 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-SPSS and 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 SPSS Inc. product and use of those Web sites is at your own risk.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

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

#### COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to SPSS Inc., for the purposes of developing,

using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. SPSS Inc., therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided “AS IS”, without warranty of any kind. SPSS Inc. shall not be liable for any damages arising out of your use of the sample programs.

### **Trademarks**

IBM, the IBM logo, and [ibm.com](http://www.ibm.com) 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>.

SPSS is a trademark of SPSS Inc., an IBM Company, registered in many jurisdictions worldwide.

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.

This product uses WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com>.

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

Adobe product screenshot(s) reprinted with permission from Adobe Systems Incorporated.

Microsoft product screenshot(s) reprinted with permission from Microsoft Corporation.





---

# *Glossary*

**analytical server.** A computer configured for, and dedicated to, rapidly executing commands from analytical applications.

**AnswerTree.** The AnswerTree (client) application that is installed on end users' desktop computers and handles user interface and presentation of results. AnswerTree can also run in local analysis mode without an analytic server software.

**AnswerTree Server.** The AnswerTree Server analytic server product that is installed on an analytical server; it manages requests from end users and handles data access and processing.

**client process.** The process that the analytic server software launches for each connected user.

**Connect ODBC.** A set of individual, database-specific drivers that use ODBC to access data.

**database server.** A computer system that processes database queries.

**daemon.** On UNIX, a program that runs continually as a background process that is not associated with a controlling terminal (tty).

**desktop application.** An application in which all data access and processing occur on the end user's desktop computer.

**distributed analysis mode.** A client application that is accessing and processing data on a remote server using the IBM® SPSS® Statistics analytic server technology.

**distributed architecture.** An architecture that distributes data access and processing to a remote server while running the user interface and presenting results on the end user's desktop computer.

**file server.** A computer and storage device dedicated to storing files.

**firewall.** A way of keeping a network secure from outside intruders while providing access to the Internet for internal users.

**local analysis mode.** A client application that is accessing and processing data on the end user's desktop computer.

**NT service.** A program that runs whenever the computer is running the operating system. It does not require a user to be logged in. Services are typically used to perform user-independent tasks.

**NTFS.** An advanced file system designed specifically for the Windows NT operating system. It offers many features, including better performance than other types of file systems.

**NAT.** Stands for Network Address Translation, a standard that allows an organization to present itself to the Internet with an external IP address that refers to many internal IP addresses. NAT is one of the techniques used for firewall protection.

**ODBC.** Stands for Open Database Connectivity, a specification for a low-level interface that defines a standard set of routines with which an application can access data in a data source.

**ODBC data source.** A named collection of information that consists of two essential pieces: the ODBC driver that will be used to access the data and the location of the database that you want to access.

**PPTP.** Stands for Point-to-Point Tunneling Protocol. PPTP is a networking protocol that supports multiprotocol virtual private networks (VPNs). It enables remote end users to access your network securely across the Internet.

**SPSS Statistics.**SPSS Statistics is a client application that is installed on end users' desktop computers and handles user interface and presentation of results. SPSS Statistics can also run in local analysis mode without an SPSS Statistics server.

**SPSS Statistics Batch Facility.** A batch processing utility included with the SPSS Statistics Server product. It is intended for automated production of statistical reports.

**SPSS Statistics Server.** A product that distributes client requests for resource-intensive operations to powerful analytic server software.

- accounts, 14
- active log file, 28
- adding
  - administered servers, 18
- administered servers
  - adding, 18
  - deleting, 23
  - logging in, 22
  - logging out, 22
  - properties, 21
  - server information, 20
  - types, 18
- administration, 17
- administration application, 17
  - connections, 23
- administrator group, 25
- administrator-level permissions, 38
- analytic server software
  - administrators, 38
  - architecture, 1
  - components, 1
  - configuration of, 14
  - configuring ODBC data sources, 15
  - configuring OLE DB data sources, 16
  - controlling startup, 34
  - defined, 1
  - installation of, 6
  - managing end-user accounts and files, 14
  - pause, 23
  - process names by product, 35
  - products, 1
  - restart, 23
  - routine maintenance of, 35
  - shutdown, 23
  - starting and stopping, 35
  - troubleshooting, 54
  - using the UNIX startup script, 34
- analytic server software products
  - AnswerTree Server defined, 67
  - IBM SPSS Statistics Server defined, 68
- analytical server
  - defined, 67
- authentication
  - internal, 46
  - PAM, 45
- automated production with IBM SPSS Statistics Server, 56
  
- broadcasting a message to users, 33
  
- cache compression, 26, 32
  
- client application
  - installation of, 6
  - troubleshooting, 54
- client data access, 26
- client desktop application products
  - AnswerTree defined, 67
- client process
  - defined, 67
- compiled transformations, 27
- computer name
  - what end users need to know, 37
- configuration file
  - location, 25
  - troubleshooting, 54
- configuration options, 24
  - connections, 24
  - COP, 27
  - data file access, 25
  - logging, 28
  - overview, 24
  - temp directory, 25
  - users, 25
- configuring the analytic server software, 14
- Connect ODBC
  - architecture, 8
  - defined, 67
  - overview of, 8
  - setting the UNIX environment for, 11
- Coordinator of Processes, 27
- COP configuration, 27
- CPU priority, 31
- CPU usage
  - improving, 52
  
- data access, 8, 15, 26
  - configuring ODBC data sources for, 15
  - Connect ODBC, 8
  - controlling, 9
  - factors to consider, 9
  - ODBC data sources for, 9
  - OLE DB, 16
  - on UNIX, 11
  - referencing data from client software, 9
- data access technology, 8
- Data Editor
  - hiding data in the client, 32
- data file access
  - what end users need to know, 49
- data view, 15, 32
- database server
  - defined, 67

- 
- deleting
    - administered servers, 23
  - desktop application
    - defined, 67
  - disconnecting users, 33
  - disk space, 51
  - disk usage
    - improving, 51
  - distributed analysis mode
    - defined, 1, 67
    - steps to use, 1
    - view of data, 8
  - distributed architecture, 1
    - defined, 67
  - domain name
    - what end users need to know, 37
  
  - editing
    - server properties, 21
  - encryption
    - SSL, 41
  - end users
    - computer name, 37
    - data file access, 49
    - domain name, 37
    - list of what they need to know, 37
    - ODBC data sources, 49
    - port number, 37
    - supporting, 37
    - user ID and password, 37
  
  - file server
    - defined, 67
  - firewall, 39
    - defined, 67
  
  - glossary, 67
  - groups
    - creating, 30
    - editing, 30
    - overview, 29
    - settings, 31
  
  - host, 24
  
  - IBM SPSS Statistics
    - defined, 68
  - IBM SPSS Statistics Administration Console, 17
    - connections, 23
  - IBM SPSS Statistics Batch Facility
    - introduction to, 56
    - what you need to know, 56
  - IBM SPSS Statistics servers
    - properties, 21
  - IBM SPSS Statistics workspace, 52–53
  
  - IBM SPSS StatisticsB
    - defined, 68
  - IDE, 51
  - improving performance, 50
  - installation, 6
    - analytic server software, 6
    - client application, 6
    - with other analytic server software products, 7
  
  - Java memory, 26, 32
  - JVM memory, 26, 32
  
  - legal notices, 65
  - local analysis mode
    - defined, 1, 67
    - view of data, 8
  - locale, 47
  - locations
    - configuration file, 25
    - profile file, 25
    - temporary files directory, 25
  - logs
    - configuration options, 28
    - viewing, 28
    - viewing performance information, 29
  
  - maximum JVM memory, 26, 32
  - maximum threads, 26, 32
  - memory, 26, 32
  - memory usage
    - improving, 53
  - message
    - broadcast to users, 33
  - monitoring users, 33
  
  - NAT, 39
    - defined, 67
  - network usage
    - improving, 53
  - NT service
    - defined, 67
  - NTFS
    - defined, 67
  
  - ODBC
    - defined, 67
  - ODBC data sources
    - and analytic server software, 11
    - configuring, 15
    - defined, 68
    - troubleshooting, 54
    - what end users need to know, 49
  - ODBC data sources, UNIX
    - defined in odbcc.ini, 64
  - ODBC data sources, Windows
    - system DSNs, 60

- user DSNs, 60
- OLE DB data sources
  - configuring, 16
- operating system tasks, UNIX
  - using `odbc.ini` to configure data sources, 64
  - using scripts to set environment variables, 63
  - using the `chmod` command to set file permissions, 62
  - using the `env` command to check environment variables, 62
  - using the `kill` command to stop server processes, 63
  - using the `ps` command to check server processes, 63
- operating system tasks, Windows
  - creating environment variables, 59
  - setting file properties, 58
  - using the ODBC Administrator, 60
  - using the Services control panel, 59
  - using the Task Manager, 60
  - using the User Manager, 59
- PAM, 45
- pause, 23
- performance
  - improving, 50
- performance information, 50
- performance log
  - interval, 28
  - viewing, 29
- permissions, 38
- Pluggable Authentication Module, 45
- point-to-point tunneling protocol, 40
- port number, 24
  - troubleshooting, 54
  - what end users need to know, 37
- port numbers, 21
- PPTP, 40
  - defined, 68
- process names by product, 35
- processors, 52
- products and operating systems, 1
- profile file, 25, 30
- profiles, 38
  - creating, 30
  - editing, 30
  - overview, 29
  - settings, 31
- RAID, 51
- RAM, 53
- restart, 23
- root privileges, 46
- running without root privileges, 46
- SCSI, 51
- Secure Sockets Layer, 41
- Secure Sockets Layer (SSL), 24
- security
  - SSL, 41
- server processes, 33
- shutdown, 23
- sorting, 16, 26
- SSL, 21, 24, 41
  - overview, 41
  - securing communications, 42
- `start_statistics_server`, 34
- `statisticsb`, 56
- SyncSort, 16
- system administrators
  - overview of administrative tasks, 4
  - what end users need to know, 37
- temp directory, 25
- temporary files directory, 32
- third-party sorting, 16, 26
- trademarks, 66
- transformations
  - compiling, 27
- troubleshooting
  - analytic server software, 54
  - client application, 54
  - client login, 54
  - configuration file, 54
  - ODBC data sources, 54
  - port number, 54
  - Solaris patches, 54
- umask, 26
- umask setting, 32
- UNC data file references, 49
- UNIX
  - checking environment variables, 62
  - checking server processes, 63
  - creating and configuring ODBC data sources, 64
  - setting environment variables, 63
  - setting file permissions, 62
  - stopping server processes, 63
- UNIX environment and data access, 11
- user groups
  - creating, 30
  - editing, 30
  - overview, 29
  - settings, 31
- user ID and password
  - what end users need to know, 37
- user profiles, 38
  - creating, 30
  - editing, 30
  - overview, 29
  - settings, 31
- users
  - broadcasting a message to, 33
  - disconnecting, 33

*Index*

monitoring, 33

versions, 39

view data, 15, 32

viewing

server properties, 21

**Windows**

changing service startup parameters, 59

checking server processes, 60

checking service status, 59

creating and configuring ODBC data sources, 60

creating end-user accounts, 59

creating environment variables, 59

setting file permissions, 58

starting and stopping services, 59

workspace, 52–53