IBM SPSS Modeler 14.2 Deployment Guide

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

This edition applies to IBM SPSS Modeler 14 and to all subsequent releases and modifications until otherwise indicated in new editions.

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

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

Licensed Materials - Property of IBM

Preface

IBM® SPSS® Modeler is the IBM Corp. enterprise-strength data mining workbench. SPSS Modeler helps organizations to improve customer and citizen relationships through an in-depth understanding of data. Organizations use the insight gained from SPSS Modeler to retain profitable customers, identify cross-selling opportunities, attract new customers, detect fraud, reduce risk, and improve government service delivery.

SPSS Modeler's visual interface invites users to apply their specific business expertise, which leads to more powerful predictive models and shortens time-to-solution. SPSS Modeler offers many modeling techniques, such as prediction, classification, segmentation, and association detection algorithms. Once models are created, IBM® SPSS® Modeler Solution Publisher enables their delivery enterprise-wide to decision makers or to a database.

About IBM Business Analytics

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

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

Technical support

Technical support is available to maintenance customers. Customers may contact Technical Support for assistance in using IBM Corp. products or for installation help for one of the supported hardware environments. To reach Technical Support, see the IBM Corp. web site at *http://www.ibm.com/support*. Be prepared to identify yourself, your organization, and your support agreement when requesting assistance.

Contents

4　*Scenario Job Steps*　24

5　*Champion Challenger Job Steps*　34

Appendix

Overview

IBM SPSS Collaboration and Deployment Services

IBM® SPSS® Collaboration and Deployment Services is an enterprise-level application that enables widespread use and deployment of predictive analytics. IBM SPSS Collaboration and Deployment Services provides centralized, secure, and auditable storage of analytical assets and advanced capabilities for management and control of predictive analytic processes, as well as sophisticated mechanisms for delivering the results of analytical processing to the end users. The benefits of IBM SPSS Collaboration and Deployment Services include:

- safeguarding the value of analytical assets
- ensuring compliance with regulatory requirements
- improving the productivity of analysts
- minimizing the IT costs of managing analytics

IBM SPSS Collaboration and Deployment Services allows you to securely manage diverse analytical assets and fosters greater collaboration among those developing and using them. Furthermore, the deployment facilities ensure that the right people get the information they need to take timely, appropriate action.

Collaboration

Collaboration refers to the ability to share and reuse analytic assets efficiently, and is the key to developing and implementing analytics across an enterprise. Analysts need a location in which to place files that should be made available to other analysts or business users. That location needs a version control implementation for the files to manage the evolution of the analysis. Security is required to control access to and modification of the files. Finally, a backup and restore mechanism is needed to protect the business from losing these crucial assets.

To address these needs, IBM® SPSS® Collaboration and Deployment Services provides a repository for storing assets using a folder hierarchy similar to most file systems for organization. Files stored in the IBM® SPSS® Collaboration and Deployment Services Repository are available to users throughout the enterprise, provided those users have the appropriate permissions for access. To assist users in finding assets, the IBM SPSS Collaboration and Deployment Services Repository offers a search facility.

Analysts can work with files in the IBM SPSS Collaboration and Deployment Services Repository from client applications that leverage the service interface of IBM SPSS Collaboration and Deployment Services. Products such as IBM® SPSS® Statistics and IBM® SPSS® Modeler allow direct interaction with files in the repository. An analyst can store a version of a file in development, retrieve that version at a later time, and continue to modify it until it is finalized and ready to be moved into a production process. These files can include custom interfaces that run analytical processes allowing business users to take advantage of an analyst's work.

The use of the IBM SPSS Collaboration and Deployment Services Repository protects the business by providing a central location for analytical assets that can be easily backed-up and restored. In addition, permissions at the user, file, and version label levels control access to individual assets. Version control and object version labels ensure the right versions of assets are being used in production processes. Finally, logging features provide the ability to track file and system modifications.

Deployment

To realize the full benefit of predictive analytics, the analytic assets need to provide input for business decisions. Deployment bridges the gap between analytics and action by delivering results to people and processes on a schedule or in real time.

In IBM® SPSS® Collaboration and Deployment Services, individual files stored in the IBM® SPSS® Collaboration and Deployment Services Repository can be included in processing **jobs** that define an execution sequence for the files. The execution results can be stored in the repository, on a file system, or delivered to specified recipients. Results stored in the repository can be accessed by any user with sufficient permissions using the IBM® SPSS® Collaboration and Deployment Services Deployment Portal interface. The jobs themselves can be triggered according to a defined schedule or in response to system events.

In addition, the scoring service of IBM SPSS Collaboration and Deployment Services allows analytical results from deployed models to be delivered in real time when interacting with a customer. An analytical model configured for scoring can combine data collected from a current customer interaction with historical data to produce a score that determines the course of the interaction. The service itself can be leveraged by any client application, allowing the creation of custom interfaces for defining the process.

The deployment facilities of IBM SPSS Collaboration and Deployment Services are designed to easily integrate with your enterprise infrastructure. Single sign-on reduces the need to manually provide credentials at various stages of the process. Moreover, the system can be configured to be compliant with Federal Information Processing Standard Publication 140-2.

System Architecture

In general, IBM® SPSS® Collaboration and Deployment Services consists of a single, centralized IBM® SPSS® Collaboration and Deployment Services Repository that serves a variety of clients, using execution servers to process analytical assets.

Figure 1-1
IBM SPSS Collaboration and Deployment Services Architecture
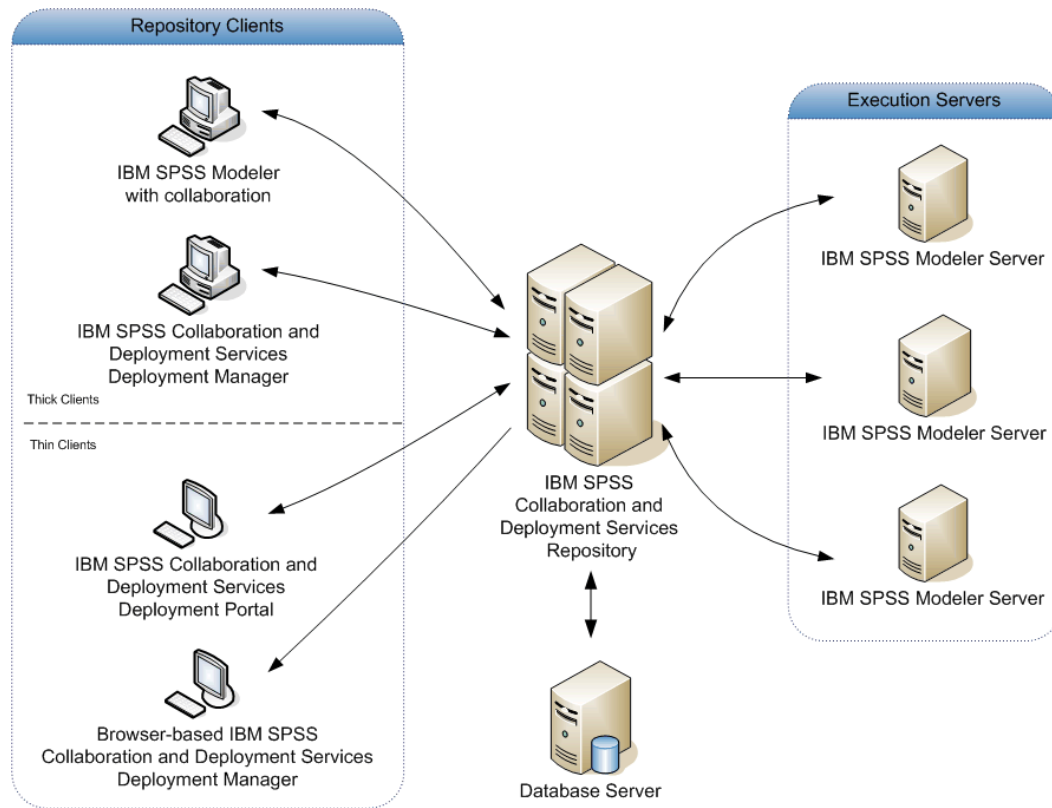
IBM SPSS Collaboration and Deployment Services consists of the following components:

- IBM SPSS Collaboration and Deployment Services Repository for analytical artifacts
- IBM® SPSS® Modeler with Collaboration
- IBM® SPSS® Collaboration and Deployment Services Deployment Manager
- IBM® SPSS® Collaboration and Deployment Services Deployment Portal
- browser-based IBM® SPSS® Collaboration and Deployment Services Deployment Manager
- IBM® SPSS® Collaboration and Deployment Services Enterprise View
- BIRT Report Designer for IBM® SPSS®

IBM SPSS Collaboration and Deployment Services Repository

The IBM® SPSS® Collaboration and Deployment Services Repository provides a centralized location for storing analytical assets, such as models and data. The IBM SPSS Collaboration and Deployment Services Repository includes facilities for:

- Security
- Version control

- Searching
- Auditing

The IBM SPSS Collaboration and Deployment Services Repository requires an installation of a relational database, such as Oracle, IBM DB2 UDB, or Microsoft SQL Server.

Configuration options for the IBM SPSS Collaboration and Deployment Services Repository are defined using the IBM® SPSS® Collaboration and Deployment Services Deployment Manager or the browser-based IBM® SPSS® Collaboration and Deployment Services Deployment Manager. The contents of the repository are managed with the Deployment Manager and accessed with IBM® SPSS® Collaboration and Deployment Services Deployment Portal.

IBM SPSS Modeler with Collaboration

IBM® SPSS® Modeler with collaboration allows interaction with the IBM® SPSS® Collaboration and Deployment Services Repository from within the SPSS Modeler interface. Files can be stored and retrieved directly from SPSS Modeler.

In addition, SPSS Modeler streams and scenarios stored in the repository can be executed as steps within jobs. A job can contain any number of steps, with each step corresponding to a separate file. Relationships defined between the steps determine the processing flow. The job can be scheduled to execute at a specific time, according to a recurrence pattern, or in response to a defined event. Moreover, notifications can be sent to specified recipients to report on individual step and overall job execution status.

Collaboration between IBM® SPSS® Collaboration and Deployment Services and SPSS Modeler is enabled through the use of adapters. These adapters are installed into the IBM SPSS Collaboration and Deployment Services environment to add the product-specific features. For more information, consult the SPSS Modeler documentation.

IBM SPSS Collaboration and Deployment Services Deployment Manager

IBM® SPSS® Collaboration and Deployment Services Deployment Manager is a client application that allows users to schedule, automate, and execute analytical tasks, such as updating models or scores, using the IBM® SPSS® Collaboration and Deployment Services Repository. Deployment Manager allows a user to:

- View any existing files within the system, including IBM® SPSS® Modeler streams, scenarios, and data files.
- Import files into the IBM SPSS Collaboration and Deployment Services Repository.
- Schedule jobs to be executed repeatedly using a specified recurrence pattern, such as quarterly or hourly.
- Modify existing job properties in a user-friendly interface.
- Determine the status of a job.
- Specify e-mail notification of job status.

In addition, Deployment Manager allows users to perform administrative tasks for IBM® SPSS® Collaboration and Deployment Services, including:

- user management
- security provider configuration
- role and action assignment

IBM SPSS Collaboration and Deployment Services Deployment Portal

IBM® SPSS® Collaboration and Deployment Services Deployment Portal is a thin-client interface for accessing the IBM® SPSS® Collaboration and Deployment Services Repository. Unlike the browser-based IBM® SPSS® Collaboration and Deployment Services Deployment Manager, which is intended for IBM® SPSS® Collaboration and Deployment Services administrators, Deployment Portal is a web portal serving a variety of users. Deployment Portal includes the following functionality:

- Browsing the IBM SPSS Collaboration and Deployment Services Repository content by folder
- Opening published content
- Running jobs
- Generating scores using models stored in the IBM SPSS Collaboration and Deployment Services Repository
- Searching IBM SPSS Collaboration and Deployment Services Repository content.
- Viewing content properties.
- Accessing individual user preferences, such as e-mail address and password, general options, subscriptions, and options for output file formats.

Browser-based IBM SPSS Collaboration and Deployment Services Deployment Manager

The browser-based IBM® SPSS® Collaboration and Deployment Services Deployment Manager is a thin-client interface for performing setup and system management tasks, including:

- Configuring the system.
- Configuring security providers.
- Managing MIME types.

Non-administrative users can perform any of these tasks provided they have the appropriate actions associated with their login credentials. The actions are assigned by an administrator.

IBM SPSS Collaboration and Deployment Services Enterprise View

IBM® SPSS® Collaboration and Deployment Services Enterprise View provides a single, consistent view of enterprise data. IBM SPSS Collaboration and Deployment Services Enterprise View allows users to define and maintain a common view of warehoused and transaction data needed to perform analytics, optimization, deployment, and reporting. Underlying data may

come from a variety of sources, including a data warehouse, an operational data store, and an online transaction database. IBM SPSS Collaboration and Deployment Services Enterprise View ensures a consistent use of enterprise data and hides the complexities of stored data structures from the end user. IBM SPSS Collaboration and Deployment Services Enterprise View is the data backbone for the predictive enterprise.

Data discovery requires a major investment of resources from the organizations deploying predictive analytics. The process is labor intensive—it can involve representatives from departments across the organization and often entails resolving differences in data structure and semantics across organizational boundaries. IBM SPSS Collaboration and Deployment Services Enterprise View provides a mechanism for recording the outcomes of the data discovery process, versioning and securing the resulting schema, and tracking changes over time.

IBM SPSS Collaboration and Deployment Services Enterprise View includes the IBM® SPSS® Collaboration and Deployment Services Enterprise View Driver component designed to provide other applications access to IBM SPSS Collaboration and Deployment Services Enterprise View objects stored in the IBM® SPSS® Collaboration and Deployment Services Repository. The driver operates similarly to ODBC drivers with the exception that it does not directly query a physical data source but rather references IBM SPSS Collaboration and Deployment Services Enterprise View data provider definitions and application views. Note that while IBM SPSS Collaboration and Deployment Services Enterprise View is installed as part of IBM® SPSS® Collaboration and Deployment Services Deployment Manager, IBM SPSS Collaboration and Deployment Services Enterprise View Driver must be installed separately. For more information, see the installation instructions.

Execution Servers

Execution servers provide the ability to execute resources stored within the IBM® SPSS® Collaboration and Deployment Services Repository. When a resource is included in a job for execution, the job step definition includes the specification of the execution server used for processing the step. The execution server type depends on the resource.

The IBM® SPSS® Modeler execution server is IBM® SPSS® Modeler Server, which permits distributed analysis for data mining and model building. This execution server requires the specification of user credentials under which the processing occurs. For more information, see the topic Server definitions in Chapter 2 on p. 8.

To allow load balancing, two or more execution servers can be grouped together in a server cluster. When a job step uses a cluster for execution, IBM® SPSS® Collaboration and Deployment Services determines which managed server in the cluster is best suited to handle processing requests at that time. For more information, consult the IBM® SPSS® Collaboration and Deployment Services Deployment Manager documentation.

During job creation, assign an execution server or server cluster to each step included in the job. When the job executes, the IBM SPSS Collaboration and Deployment Services Repository uses the specified execution servers to perform the corresponding analyses.

BIRT Report Designer for IBM SPSS

The reporting functionality of IBM® SPSS® Collaboration and Deployment Services is enabled by BIRT (Business Intelligence and Reporting Tools), an open-source package distributed by Eclipse Foundation under the Eclipse Public License. BIRT provides core reporting features, such as report layout, data access, and scripting. For more information about BIRT, see the BIRT project page (*http://www.eclipse.org/birt*).

The IBM SPSS Collaboration and Deployment Services installation includes the BIRT reporting engine server components, which enable the execution of BIRT report syntax files as part of the IBM SPSS Collaboration and Deployment Services reporting job steps. BIRT Report Designer for IBM® SPSS® is a standalone application that can be used in conjunction with IBM SPSS Collaboration and Deployment Services. It provides a rich user interface with a number of advanced features for creating reports and must be installed separately.

If a BIRT Report Designer for IBM SPSS report requires a JDBC-based database connection, a corresponding JDBC driver must be installed with the IBM® SPSS® Collaboration and Deployment Services Repository. For application server-specific information on the location of the JDBC drivers, see the corresponding section of the IBM SPSS Collaboration and Deployment Services Repository installation instructions.

To start BIRT Report Designer for IBM SPSS, execute the file *BIRT.exe* in the installation directory. For information on using BIRT Report Designer for IBM SPSS, see the documentation installed with the application.

Working With IBM SPSS Modeler Files

In IBM® SPSS® Collaboration and Deployment Services Deployment Manager, the general process for working with IBM® SPSS® Modeler files involves:

1. Define an execution server for processing the SPSS Modeler stream or scenario if an appropriate definition does not already exist.

2. Add the SPSS Modeler stream or scenario to a job as a job step.

3. Specify the job properties, including any desired schedules for execution.

Server definitions

Executing an IBM® SPSS® Collaboration and Deployment Services Repository resource as a job step requires the specification of an appropriate corresponding server to process the instructions contained in the job step. The connection information for such a server is specified within a **server definition**. Server definitions can be classified as either execution servers or repository servers.

■ Execution servers process the contents of an IBM SPSS Collaboration and Deployment Services Repository resource. The execution server type must correspond to the resource type being processed. An IBM® SPSS® Modeler job step requires an SPSS Modeler server definition.

■ A repository server corresponds to an IBM® SPSS® Collaboration and Deployment Services repository installation. A server of this type is typically used by job steps that need to return result artifacts to a repository.

Server definitions are contained in the *Resource Definitions* folder of the Content Explorer. Specifically, they are defined in the *Servers* subfolder.

Adding new server definitions

To add a new server:

1. In the Content Explorer, open the *Resource Definitions* folder.

2. Click on the *Servers* folder.

3. From the File menu, choose:
   New > Server Definition

   The Add New Server Definition wizard opens. Alternatively, the new server definition dialog box can be accessed by clicking New next to a server field on the General tab for some steps. The process for defining new servers consists of:

1. Naming the server definition and specifying its type.

2. Selecting a location in the *Servers* folder for the definition.

3. Specifying parameters for the server that define connection or execution information. The parameter set depends on the server type.

IBM SPSS Modeler server parameters

An IBM® SPSS® Modeler Server Definition specifies the connection parameters for SPSS Modeler servers used to process job steps. SPSS Modeler streams are executed on the SPSS Modeler server.

Figure 2-6
*Defining a new IBM SPSS Modeler execution server*

1. In the Host field, enter the hostname where the server resides. For example, if you are creating an SPSS Modeler server definition, the host would be the machine that contains your SPSS Modeler server.

2. In the Port field, enter the port number to be used to connect to the host.

3. In the Default Data Path field, enter the path on which you want to place data files.

4. If Secure Socket Layer (SSL) is to be used for the server connection, select This is a secure port.

5. Click Finish. The new definition appears in the *Servers* folder.

IBM ShowCase Enterprise Server definition

After you have specified the name and selected IBM® ShowCase® Enterprise Server as the type, the Server Definition dialog box opens.

Figure 2-9
Defining a new IBM ShowCase Enterprise Server
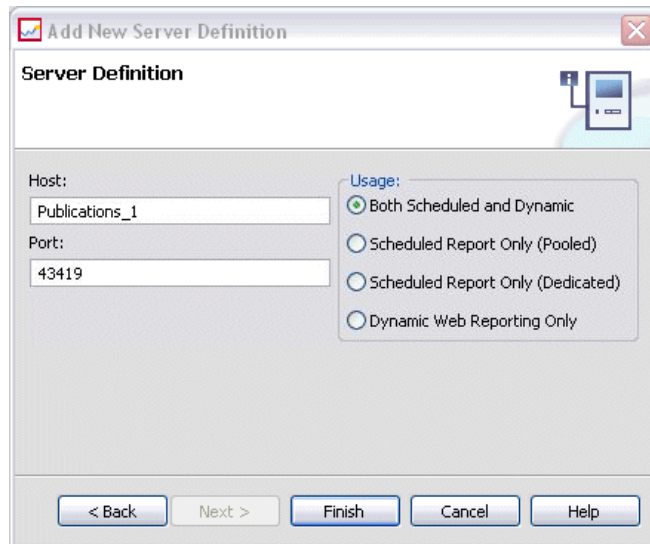
1. In the Host field, enter the machine name where the server resides.

2. In the Port field, enter the port number to be used to connect to the host (43419 is the default).

3. In the Usage section, select from the available options:

   ■ **Both Scheduled and Dynamic:** Select this option if you will use the server for both scheduled reporting and dynamic reporting via IBM® SPSS® Collaboration and Deployment Services Deployment Portal.

   ■ **Scheduled Report Only (Pooled):** Select this option if you will configure multiple servers for scheduled reporting. A pooled server handles incoming scheduled report requests. Pooled servers logically act as one server, and the server with the lightest load processes incoming requests.

   ■ **Scheduled Report Only (Dedicated):** Select this option if you will use the server only for scheduled reporting. A dedicated server handles specific incoming scheduled report requests.

   ■ **Dynamic Web Reporting Only:** Select this option if you will use the server only for dynamic reporting via Deployment Portal.

4. Click Finish. The new definition appears in the *Servers* folder.

Modifying server definitions

To modify a server definition:

1. In the Content Explorer, open the *Resource Definitions* folder.

2. Open the *Servers* folder.

3. Double-click the server to be modified. The Edit Server Definition dialog opens.

4. Modify the server definition parameters as necessary.

5. Click Finish to save your changes.

IBM SPSS Modeler Job Steps

Working with IBM SPSS Modeler Streams

This section describes IBM® SPSS® Modeler streams within the context of IBM® SPSS® Collaboration and Deployment Services Deployment Manager. SPSS Modeler streams are brought into the Deployment Manager fully formed. For more detailed information on how to create and work with streams, refer to the SPSS Modeler documentation.

You can work with SPSS Modeler streams in Deployment Manager. Like any other step, an SPSS Modeler stream has to be added to a job before you can execute it using Deployment Manager. Specifically, you can perform the following tasks:

- Importing streams
- Modifying stream parameters
- Executing streams

IBM SPSS Modeler Server Configuration

Before you begin working with IBM® SPSS® Modeler streams in the IBM® SPSS® Collaboration and Deployment Services Deployment Manager, you need to perform the following configuration tasks:

- Create an SPSS Modeler server definition.
- Define server credentials.

Viewing IBM SPSS Modeler Job Properties

When you click on an IBM® SPSS® Modeler stream within a job, the following job properties appear:

- General
- Data Files
- ODBC Data Sources
- Parameters
- Results
- Cognos Import
- Cognos Export
- Notifications

IBM SPSS Modeler Job Properties - General

By default, the General properties tab appears when you click on an IBM® SPSS® Modeler stream in the job canvas.

Figure 3-1
IBM SPSS Modeler job general properties

The General properties tab contains the following information:

Job step name. The name of the job step. Typically, the name of the job step is the name of the SPSS Modeler stream, appended with the suffix _step. However, you can modify the job step name.

To modify the name of the job step, type the revised name in the Job step name field. Your changes are reflected in the job canvas.

SPSS Modeler Stream. The name and path of the original SPSS Modeler stream. The source stream is not modifiable.

Object version. The labeled version of the SPSS Modeler stream to use. From the Object version drop-down list, select the labeled version that you want to use.

Iterative Variable List. If the step acts as an iterative consumer, identify the variable from the iterative producer providing the values for the step. An iterative producer step must appear immediately before the step in the job for the step to act as an iterative consumer.

SPSS Modeler server. The SPSS Modeler server or server cluster on which the stream will be executed. The list contains all servers and server clusters currently configured to execute SPSS Modeler steps. To change the server, select from the SPSS Modeler Server drop-down list. To create a new server definition, click New to launch the server definition wizard.

SPSS Modeler login. The credential information used to access the SPSS Modeler server or server cluster. To change the credentials, select a credential definition from the SPSS Modeler Login drop-down list. To define new credentials, click New to launch the credentials definition wizard.

Content Repository Server. The content repository server allows a job to save files to an IBM® SPSS® Collaboration and Deployment Services Repository. Typically, the content repository server is specified when refreshing models using SPSS Modeler. To specify a content repository server, select a server from the Content Repository Server drop-down list. To create a new server

definition, click New to launch the server definition wizard. To generate a content repository server definition based on the current server information, click Generate. A server definition is created and automatically populated in the *Content Repository Server* field.

Content Repository Login. The login information for the content repository server. To specify a content repository login, select a credential from the Content Repository Login drop-down list. To create a new login, click New to launch the content repository login wizard. If single sign-on is not used to connect to the IBM SPSS Collaboration and Deployment Services Repository, click Generate to generate a content repository server login based on existing security settings. A content repository login is created and automatically populated in the *Content Repository Login* field. Login generation is not available when using single sign-on.

Warning expression. Define warnings for job steps connected by a Conditional connector. The warning expression (e.g., `completion_code`, `warning`, or `success`) must be lowercase.

To use warning expressions:

1. Connect two job steps with a Conditional connector. In the Expression field for the conditional connector, type `warning==true`.

2. Navigate to the General tab of the parent job step.

3. In the Warning expression field, specify a warning code—for example, `completion_code==18`. This expression overrides the default warning code, if any.

When the job is run, the system will execute the parent job step. Then the system will evaluate the condition for `warning==true`. If true, the system will look at the warning expression specified and determine if the condition was met. If the condition specified in the warning expression was met, the system proceeds to the next job step.

Guidelines for Providing IBM SPSS Modeler and Content Repository Information

For all SPSS Modeler job steps, both SPSS Modeler server and SPSS Modeler login information are required. The additional server and login information required vary by the type of stream used in the SPSS Modeler job step.

- **SPSS Modeler stream without an Enterprise View node.** For streams that do not use an Enterprise View node, server and login information for the Content Repository is optional. However, if information is provided for the Content Repository, then both server and login information must be specified. Partial server and login information generates an error.

- **SPSS Modeler stream with an Enterprise View node.** For streams that use an Enterprise View node, both Content Repository server and Content Repository login information are required. Incomplete information generates an error.

IBM SPSS Modeler Job Properties - Data Files

The Data Files table contains the following information:

Node Name. The name of the input node that contains the data used by the stream. The name is prefixed by the names of any supernodes containing the node separated by slashes. For example, if the node *MyNode* is within a supernode named *Supernode1*, the name appears as */Supernode1/MyNode*. The node name is not modifiable.

Node Type. The node type as defined in the stream. The node type is not modifiable.

File Name. The name of the input data file. To change the name, click in the File Name cell and change the name.

Format. The format of the output file—for example, a comma delimited file. To modify the file format type, click in the Format cell. A drop-down arrow appears. Select the format type.

Location. The location of the input data files. To modify location, click in the column and then click the resulting ellipsis button. The Input File Location dialog box opens. Change the location as necessary.

Nodes within locked supernodes are not accessible. They cannot be viewed or modified.
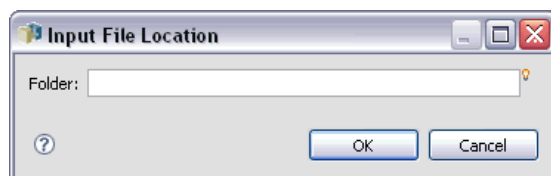
Figure 3-2
IBM SPSS Modeler job data files

Specifying the Input File Location

When working with data files in an IBM® SPSS® Modeler job step, an input file location may need to be specified in the Input File Location dialog.

Figure 3-3
Specifying the input file location

To change the location:

1. Type the new path for the data file in the Folder field.

2. Click OK.

IBM SPSS Modeler Job Properties - Enterprise View

The Enterprise View tab specifies the Enterprise View settings for accessing data. This tab lists only the node associated with the default branch of the stream.

Figure 3-4
IBM SPSS Modeler job Enterprise View settings

Enterprise View Node. Name of the Enterprise View node in the stream or scenario.

Data Provider Definition. Data provider definition associated with the Enterprise View node. To change the data provider used for a node, select the cell containing the data provider and click the resulting ellipsis (...) button. For more information, see the topic Select Data Provider Definition in Chapter 4 on p. 28.

Label. Label identifying the version of the data provider used.

IBM SPSS Modeler Job Properties - ODBC Data Sources

The ODBC Data Sources table contains the following information:

Node Name. The name of the input node that contains the data used by the stream. The name is prefixed by the names of any supernodes containing the node separated by slashes. For example, if the node *MyNode* is within a supernode named *Supernode1*, the name appears as */Supernode1/MyNode*.

Node Type. The node type as defined in the stream.

ODBC DSN. The ODBC data source name.

User Name. The user name that corresponds to the database.

Password. The password that corresponds to the user name.

Database Table. The database table that corresponds to the node.

Nodes within locked supernodes are not accessible. They cannot be viewed or modified.

IBM SPSS Modeler Job Properties - Parameters

You can modify parameters for IBM® SPSS® Modeler streams using the IBM® SPSS® Collaboration and Deployment Services Deployment Manager. The Parameters table contains the following information:

Name. The parameter name.

Storage. Describes how the parameter is stored—for example, as a string.

Value. The value of the parameter.
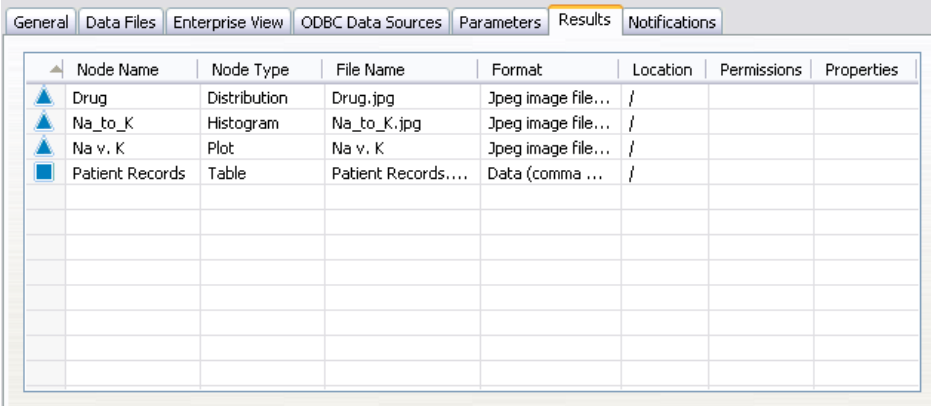
Type. The parameter type.

Parameters within locked supernodes are not accessible. They cannot be viewed or modified.

IBM SPSS Modeler Job Properties - Results

To display output settings for an IBM® SPSS® Modeler job step, click the Results tab.

Figure 3-5
IBM SPSS Modeler job step results

	Node Name	Node Type	File Name	Format	Location	Permissions	Properties
▲	Drug	Distribution	Drug.jpg	Jpeg image file…	/		
▲	Na_to_K	Histogram	Na_to_K.jpg	Jpeg image file…	/		
▲	Na v. K	Plot	Na v. K	Jpeg image file…	/		
■	Patient Records	Table	Patient Records….	Data (comma …	/		

Node Name. The name of the node that contains the output of the stream processing. The name is prefixed by the names of any supernodes containing the node separated by slashes. For example, if the node *MyNode* is within a supernode named *Supernode1*, the name appears as */Supernode1/MyNode*.The node name is not modifiable.

Node Type. The node type as defined in the stream. The node type is not modifiable.

File Name. The name of the corresponding file. To modify the name, click in the File Name column and type the new name.

Note: The file extension is determined by the selected file format, which is automatically appended to the file name.

Format. The format of the output file. To modify the file format type, click in the Format column. A drop-down arrow appears. Select the format type.

Location. The location of the file. To modify location, open the Results Location dialog box by clicking in the column and then clicking the resulting ellipsis button.

Permissions. Access permissions for the file if saved to the repository. To modify permissions, open the Output Permissions dialog box by clicking in the Permissions column and then clicking the resulting ellipsis button.

Properties. The properties (metadata) of the file. To define properties, open the Output Properties dialog box by clicking in the Properties column and then clicking the resulting ellipsis button.

Nodes within locked supernodes are not accessible. They cannot be viewed or modified.

Viewing Output Results

To view the output results of your stream:

1. In the Results column of the Job Step History table, double-click on the results you want to view. The results are opened in a separate window.

   For example, suppose one of your output results is a histogram. A sample histogram might look like this.

   Figure 3-6
   Sample histogram
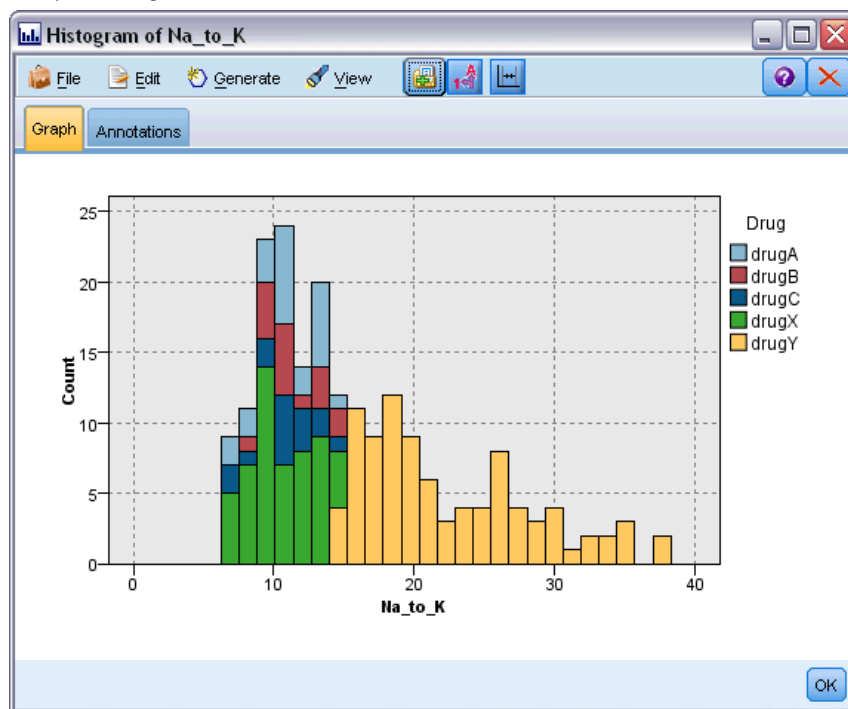
IBM SPSS Modeler Job Properties - Cognos Import

If the stream contains an IBM Cognos BI source node, the Cognos connection details are displayed here.

Figure 3-7
IBM SPSS Modeler job Cognos source node details

Node Name. The name of the Cognos source node.

Connection URL. The URL of the Cognos server to which the connection is made.

Package Name. The name of the Cognos package from which the metadata is imported.

Anonymous. Contains Anonymous if anonymous login is used for the Cognos server connection, or Credential if a specific Cognos username and password are used.

User Name. The username (if required) on the Cognos server.

Password. The password for this username.

IBM SPSS Modeler Job Properties - Cognos Export

If the stream contains an IBM Cognos BI Export node, the Cognos and ODBC connection details are displayed here.

Figure 3-8
IBM SPSS Modeler job Cognos export node details

Node Name. The name of the Cognos export node.

Connection URL. The URL of the Cognos server to which the connection is made.

Package Name. The name of the Cognos package used to export the metadata.

Datasource. The name of the Cognos database used to export the data.

Folder. The path and name of the folder on the Cognos server where the export package is created.

Anonymous. Contains Anonymous if anonymous login is used for the Cognos server connection, or Credential if a specific Cognos username and password are used.

User Name. The username (if required) on the Cognos server.

Password. The password for this username.

DSN. The database source name (DSN) of the Cognos database.

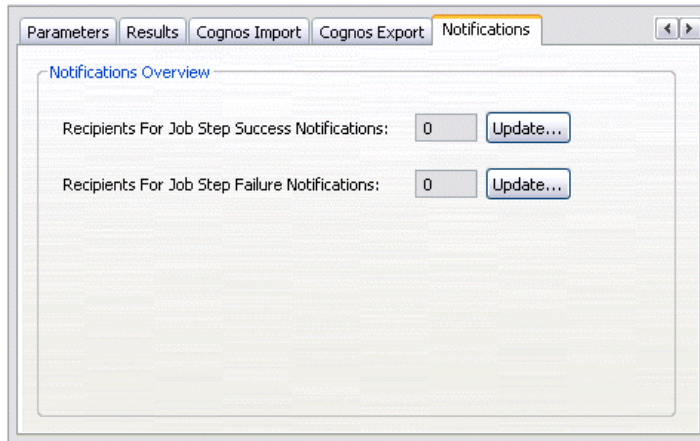**Table Name.** The name of the database table used for the export.

User Name. The username used for connection to the database used for the export.

Password. The password for this username.

IBM SPSS Modeler Job Properties - Notifications

Allows you to specify e-mail notifications for job step failure and success.

Figure 3-9
IBM SPSS Modeler job step notifications

Click the Update button in each case to add or delete notification recipients.

Viewing Streams in IBM SPSS Modeler

You can launch the IBM® SPSS® Modeler application directly from IBM® SPSS® Collaboration and Deployment Services Deployment Manager.

To view the stream in the SPSS Modeler application:

► In the Content Explorer, double-click on the stream. The system launches the SPSS Modeler application and displays the stream in SPSS Modeler.

It is important to note that if any changes are made to the files in a job—for example, an SPSS Modeler stream (*.str*) or a scenario file (*.scn*), any job that contains the file is affected. When changes are made to the file, a new version of the file is saved to the repository. However, the job that contains the file is not automatically updated with the modified file. To incorporate the file updates into the affected job:

1. Reopen the job. When the job is reopened, an asterisk appears beside the job name in the job canvas, indicating that the job contains unsaved changes.

2. Resave the job.

IBM SPSS Modeler Completion Codes

The completion codes for IBM® SPSS® Modeler jobs are described in the following table. Use these completion codes for any conditional relationships that involve SPSS Modeler streams.

Table 3-1
Completion codes for IBM SPSS Modeler jobs

Code	Description
0	success
1	stream execution error
2	publishing error
8	unknown error

IBM SPSS Modeler Stream Limitations

When you work with streams in the IBM® SPSS® Collaboration and Deployment Services Deployment Manager, the system has the following constraints:

- **Naming.** If the node name, label, and type are the same, you cannot schedule a job for execution because a conflict arises.

- **Scripting.** If a stream script contains overrides for a specific node—for example, the script sets the output location for a graph—the script supersedes any conflicting user-specified values defined in Deployment Manager.

- **Supernodes.** Execution of an IBM® SPSS® Modeler job step corresponds to the processing and execution of all top-level terminal nodes in the stream. If any terminal node is a terminal supernode, the terminal nodes within that supernode execute recursively. In contrast, if source or process supernodes, which are by definition non-terminal, contain terminal nodes, those nodes do not execute. The terminal nodes in the non-terminal supernodes do appear in Deployment Manager but are not processed during execution of the step.

- **Parameters.** Parameters defined for supernodes cannot have the same name as parameters defined for a stream. The names must be unique.

Node Types

When you open an IBM® SPSS® Modeler stream from IBM® SPSS® Collaboration and Deployment Services Deployment Manager, you see that the stream nodes are represented by different shaped icons. Circle icons represent source data nodes, while hexagonal nodes represent processing operations on data records and fields. Triangles indicate graphical output, while pentagons represent modeling nodes. Output (other than graphical) and export operations are indicated by rectangular nodes.

The various nodes are fully described in the *SPSS Modeler Source, Process and Output Nodes* and the *SPSS Modeler Modeling Nodes* guides.

Script Example: Comparing Champion vs. Challenger Models

The following excerpt of an IBM® SPSS® Modeler script demonstrates a champion/challenger comparison. The champion model is compared to the challenger model. If the challenger model is better than the champion model, the champion model is overwritten. Thus, the challenger model becomes the new champion model.

```
# -----------------------------------
# Before rebuilding model, examine performance of existing model
# -----------------------------------
#--- LOAD Model
set working_file = ^SPSS_SMS_INT_PERSIST_PATH >< ^slash >< "ResolutionStatus.gm"
load model ^working_file

#--- First do the evaluation of the current champion model
delete 'ResolutionStatusModel'
insert model 'ResolutionStatus' at 558 652
rename 'ResolutionStatus':applyc50node as 'ResolutionStatusModel'
connect 'ValidationSet' to 'ResolutionStatusModel'
connect 'ResolutionStatusModel' to 'PredictQuality'

set 'ModelEval'.full_filename = ^SPSS_SMS_INT_TEMP_PATH >< ^slash >< "ModelEvalChamp.txt"
execute 'ModelEval':tablenode
var missRateChamp
set missRateChamp = value 'ModelEval':tablenode.output at 1 6
delete 'ResolutionStatusModel'
clear generated palette

#--- Execute c50node 'ResStatusModelBuild'
if ^useBoost = 1
   set 'ResStatusModelBuild'.use_boost = T
   set 'ResStatusModelBuild'.boost_num_trials = ^numBoostTrials
else
   set 'ResStatusModelBuild'.use_boost = F
endif
if ^useMisCosts = 1
   set 'ResStatusModelBuild'.use_costs = T
   set 'ResStatusModelBuild'.costs = [{'Not fraud' 'Not fraud' 0}{'Not fraud' 'Fraud' ^falsePosCost}
   {'Fraud' 'Not fraud' ^falseNegCost}{'Fraud' 'Fraud' 0}]
endif
execute 'ResStatusModelBuild'

#--- Now evaluate the new model
insert model 'ResolutionStatus' at 558 652
rename 'ResolutionStatus':applyc50node as 'ResolutionStatusModel'
connect 'ValidationSet' to 'ResolutionStatusModel'
connect 'ResolutionStatusModel' to 'PredictQuality'

set 'ModelEval'.full_filename = ^SPSS_SMS_INT_TEMP_PATH >< ^slash >< "ModelEvalChallenge.txt"
execute 'ModelEval':tablenode
var missRateChallenge
set missRateChallenge = value 'ModelEval':tablenode.output at 1 6

#--- Determine which one should be the champion
#--- Overwrite existing champion if challenger is better

if ^missRateChamp > ^missRateChallenge
   #--- SAVE Model
   set working_file = ^SPSS_SMS_INT_PERSIST_PATH >< ^slash >< "ResolutionStatus.gm"
```

```
      save model 'ResolutionStatus':applyc50node as ^working_file

      #--- EXPORT Model
      set working_file = ^SPSS_SMS_INT_PERSIST_PATH >< ^slash >< "ResolutionStatus.xml"
      export model 'ResolutionStatus' as ^working_file format pmml

      #--- EXPORT Model
      set working_file = ^SPSS_SMS_INT_PERSIST_PATH >< ^slash >< "ResolutionStatus_summary.html"
      export model 'ResolutionStatus' summary  as ^working_file format html
endif
```

Scenario Job Steps

Scenario Job Step Overview

A scenario is a container for models and the metadata describing the models. The key component of a scenario (*.scn*) file is an IBM® SPSS® Modeler stream (*.str*), which is responsible for performing the data analysis. Scenario files, which are created in SPSS Modeler, are not modifiable in IBM® SPSS® Collaboration and Deployment Services Deployment Manager. For additional information on how to create scenario files and store them to the repository, refer to the SPSS Modeler documentation.

What is the difference between a scenario and an SPSS Modeler stream? A scenario refers to the container that houses the analysis file (in this case, an SPSS Modeler stream) and any auxiliary files. Thus, a stream is a component of a scenario. In fact, the stream is the heart of the scenario because it defines the relationship among a set of fields within the data and analyzes that data.

Scenario Types

The type of scenario determines the input parameters and output available for analysis. Scenarios types include:

■ Refresh

■ Score

■ Evaluation

Refresh occurs when an existing model is rebuilt using newer data. The stream itself does not change. For example, the algorithm type and stream-specific settings remain the same. If the scenario includes several models, you can select any number of the models to be refreshed.

Scenarios of the type **Score** use production data to generate predicted values. When scoring the stream within a scenario, the following conditions are assumed:

■ When a scenario is scored, the Enterprise View node may not be modified in the job step. The system assumes that the Enterprise View node in the IBM® SPSS® Modeler stream contains the proper settings.

■ The Application View and the Enterprise View must be identical.

■ Scoring results are written to the format (e.g., a table) specified in the SPSS Modeler stream. Therefore, there are no output options for scoring on the Results tab.

Finally, the **Evaluation** type involves passing a sample of analytic data and scoring it against the stream. The predicted outcome is then compared against the actual outcome to determine model effectiveness.

Process Overview and Order Dependency

Unlike other types of job steps, the tabs in the scenario job step are order dependent. For example, the information that appears in the following tabs is contingent upon the options selected in the General tab:

- Enterprise View
- Data Files
- ODBC Data Sources
- Parameters
- Results

The process for working with scenario job steps consists of the following tasks:

1. Specifying general job information, including the scenario type. For more information, see the topic General Information on p. 26.

2. Viewing and specifying Enterprise View settings. For more information, see the topic Enterprise View on p. 27.

3. Viewing data file settings. For more information, see the topic Data Files on p. 29.

4. Viewing ODBC data source information. For more information, see the topic ODBC Data Sources on p. 30.

5. Viewing parameter information. For more information, see the topic Parameters on p. 30.

6. Viewing result settings. For more information, see the topic Results on p. 31.

7. Viewing Cognos import settings (if relevant). For more information, see the topic Cognos Import on p. 32.

8. Viewing Cognos export settings (if relevant). For more information, see the topic Cognos Export on p. 32.

9. Specifying notifications. For more information, see the topic IBM SPSS Modeler Job Properties - Notifications in Chapter 3 on p. 20.

Before a scenario can be added to a job, the scenario file must be added to the repository. After the scenario file is accessible from the repository, the scenario file can be added to a job, just like any other file.

It is important to note that if any changes are made to the files in a job—for example, an IBM® SPSS® Modeler stream (*.str*) or a scenario file (*.scn*), any job that contains the file is affected. When changes are made to the file, a new version of the file is saved to the repository. However, the job that contains the file is not automatically updated with the modified file. To incorporate the file updates into the affected job:

1. Reopen the job. When the job is reopened, an asterisk appears beside the job name in the job canvas, indicating that the job contains unsaved changes.

2. Resave the job.

General Information

The General tab contains information that pertains to the overall job step. IBM® SPSS® Modeler server and login information is required to execute the scenario job step. Content repository server and login information is required only if saving the results to an IBM® SPSS® Collaboration and Deployment Services Repository.

Figure 4-1
Scenario job step—general tab

Job step name. The name of job step. By default, the system fills in the job step name as `<scenario name>.scn_step`. The name, which will appear in the job history table after the job is run, is modifiable.

Scenario. The directory containing the scenario file. This field is not modifiable. To use an alternate scenario, the new scenario file must be added to the existing job separately.

Object version. The version of the scenario file in use.

SPSS Modeler server. The SPSS Modeler server or server cluster on which the stream will be executed. The list contains all servers and server clusters currently configured to execute SPSS Modeler steps. To change the server, select from the SPSS Modeler Server drop-down list. To create a new server definition, click New to launch the server definition wizard.

SPSS Modeler login. The credential information used to access the SPSS Modeler server or server cluster. To change the credentials, select a credential definition from the SPSS Modeler Login drop-down list. To define new credentials, click New to launch the credentials definition wizard.

Content Repository Server. The content repository server allows a job to save files to an IBM SPSS Collaboration and Deployment Services Repository. Typically, the content repository server is specified when refreshing models using SPSS Modeler. To specify a content repository server, select a server from the Content Repository Server drop-down list. To create a new server definition, click New to launch the server definition wizard. To generate a content repository server definition based on the current server information, click Generate. A server definition is created and automatically populated in the *Content Repository Server* field.

Content Repository Login. The login information for the content repository server. To specify a content repository login, select a credential from the Content Repository Login drop-down list. To create a new login, click New to launch the content repository login wizard. If single sign-on is

not used to connect to the IBM SPSS Collaboration and Deployment Services Repository, click Generate to generate a content repository server login based on existing security settings. A content repository login is created and automatically populated in the *Content Repository Login* field. Login generation is not available when using single sign-on.

Type. The scenario type. Valid scenario types include *Refresh*, *Score*, and *Evaluation*.

Evaluation Options

Performance. Specify the threshold percentages for the following categories:

■ Bad

■ Good

A percentage for *Better* is not user-defined. The system subtracts the *Good* percentage from 100 to arrive at the *Better* threshold. Threshold values are mandatory for evaluation scenarios.

Metric. The measurement criteria by which the scenario's effectiveness is assessed. Valid values include *accreditation*, *accuracy*, and *gains*. If *gains* is selected, then a percentile needs to be specified as well. If *accreditation* is selected, then the accreditation step needs to be specified. Valid values include *collect statistics* and *run evaluation*.
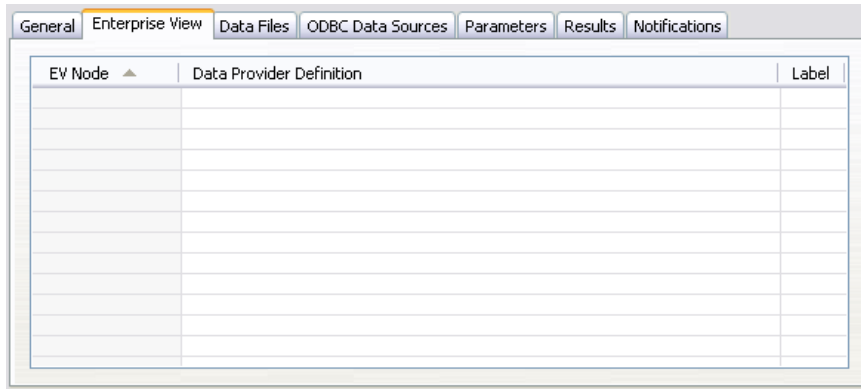
Refresh Options

Refresh Relationships. Lists all model-nugget relationships in the scenario that can be refreshed using the node names defined in the scenario. Select a specific relationship to be refreshed by checking the box before the relationship. Uncheck a relationship to prevent it from being refreshed when the step executes.

To select all relationships for refresh, click the Check All button. To unselect all relationships, click the Uncheck All button.

Enterprise View

The Enterprise View tab specifies the Enterprise View settings for accessing data. For model refresh, this tab lists all Enterprise View nodes. For the score and evaluation scenario types, the list contains only the node associated with the default branch of the stream in the scenario. The data provider for any of the Enterprise View nodes can be changed to any Data Provider Definition in the system that is compatible with the node.

Figure 4-2
Scenario job steps—Enterprise View

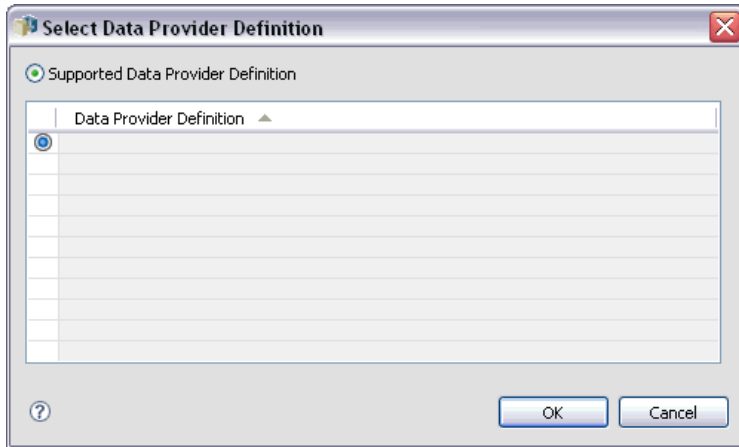**Enterprise View Node.** Name of the Enterprise View node in the stream or scenario.

Data Provider Definition. Data provider definition associated with the Enterprise View node. To change the data provider used for a node, select the cell containing the data provider and click the resulting ellipsis (...) button. For more information, see the topic Select Data Provider Definition on p. 28.

Label. Label identifying the version of the data provider used.

Select Data Provider Definition

The Select Data Provider Definition dialog box specifies the data provider definition used for an Enterprise View node. Select a definition from the list of supported definitions or specify a custom definition path.

Figure 4-3
Select Data Provider Definition

Supported Data Provider Definitions. Lists all data provider definitions in the system that are compatible with the Enterprise View. Select a particular definition by clicking the circle in the first column for that definition.

Specify Data Provider Definition. Defines a custom data provider definition path. Use this option to include job variables in the path, allowing the data provider definition to be dependent on parameter values specified when the job runs. Click the Browse button to select an IBM® SPSS® Collaboration and Deployment Services Repository path. To include a job variable, place the cursor at the desired insertion point in the field and type a $ symbol. Select the variable from the resulting list of available job variables. Example custom paths include:

${my_dpd}

and

\refresh_files\dpd\${my_dpd}

Data Files

Any changes made in this tab apply to the scenario job step only. Modifications made to the data file information are not propagated back to the scenario saved in the IBM® SPSS® Collaboration and Deployment Services Repository. The Data Files table contains the following information:

Figure 4-4
Scenario job steps—data files

Node Name. The name of the input node that contains the data used by the stream. The name is prefixed by the names of any supernodes containing the node separated by slashes. For example, if the node *MyNode* is within a supernode named *Supernode1*, the name appears as */Supernode1/MyNode*. The node name is not modifiable.

Node Type. The node type as defined in the stream. The node type is not modifiable.

File Name. The name of the input data file. To change the name, click in the File Name cell and change the name.

Format. The format of the output file—for example, a comma delimited file. To modify the file format type, click in the Format cell. A drop-down arrow appears. Select the format type.

Location. The location of the input data files. To modify location, click in the column and then click the resulting ellipsis button. The Input File Location dialog box opens. Change the location as necessary.

Nodes within locked supernodes are not accessible. They cannot be viewed or modified.

ODBC Data Sources

Any changes made in this tab apply to the scenario job step only. Modifications made to ODBC data sources information are not propagated back to the scenario saved in the IBM® SPSS® Collaboration and Deployment Services Repository. The ODBC Data Sources table contains the following information:

Figure 4-5
Scenario job steps—ODBC data sources

Node Name. The name of the input node that contains the data used by the stream. The name is prefixed by the names of any supernodes containing the node separated by slashes. For example, if the node *MyNode* is within a supernode named *Supernode1*, the name appears as */Supernode1/MyNode*.

Node Type. The node type as defined in the stream.

ODBC DSN. The ODBC data source name.

User Name. The user name that corresponds to the database.

Password. The password that corresponds to the user name.

Database Table. The database table that corresponds to the node.

Nodes within locked supernodes are not accessible. They cannot be viewed or modified.

Parameters

Any changes made in this tab apply to the scenario job step only. Modifications made to parameter information are not propagated back to the scenario saved in the IBM® SPSS® Collaboration and Deployment Services Repository. The Parameters table contains the following information.

Figure 4-6
Scenario job steps—parameters

Name. The parameter name.

Storage. Describes how the parameter is stored—for example, as a string.

Value. The value of the parameter.

Type. The parameter type.

Results

The Results table contains the following information:

Figure 4-7
Scenario job steps—results

Name. The name of the target variable upon which the results are based.

File name. The name of the output file to which results are written.

Format. The format of the output file. To modify the file format type, click in the Format column. A drop-down arrow appears. Select the format type.

Location. The location of the file. To modify location, open the Results Location dialog box by clicking in the column and then clicking the resulting ellipsis button.

Permissions. Access permissions for the file if saved to the repository. To modify permissions, open the Output Permissions dialog box by clicking in the Permissions column and then clicking the resulting ellipsis button.

Properties. The properties (metadata) of the file. To define properties, open the Output Properties dialog box by clicking in the Properties column and then clicking the resulting ellipsis button.

Results for Scoring Scenarios

If *Score* was selected as the type in the General tab, the results are not written to a file. Scoring results are written to the format (e.g., a table) specified in the terminal node in the scenario file scoring branch. The scoring format is established in the IBM® SPSS® Modeler application when the scenario file is created and is not modifiable in IBM® SPSS® Collaboration and Deployment Services Deployment Manager.

Cognos Import

If the scenario contains an IBM Cognos BI source node, the Cognos connection details are displayed here.

Figure 4-8
IBM SPSS Modeler scenario Cognos source node details

Node Name. The name of the Cognos source node.

Connection URL. The URL of the Cognos server to which the connection is made.

Package Name. The name of the Cognos package from which the metadata is imported.

Anonymous. Contains Anonymous if anonymous login is used for the Cognos server connection, or Credential if a specific Cognos username and password are used.

User Name. The username (if required) on the Cognos server.

Password. The password for this username.

Cognos Export

If the scenario contains an IBM Cognos BI Export node, the Cognos and ODBC connection details are displayed here.

Figure 4-9
IBM SPSS Modeler scenario Cognos export node details

Node Name. The name of the Cognos export node.

Connection URL. The URL of the Cognos server to which the connection is made.

Package Name. The name of the Cognos package used to export the metadata.

Datasource. The name of the Cognos database used to export the data.

Folder. The path and name of the folder on the Cognos server where the export package is created.

Anonymous. Contains Anonymous if anonymous login is used for the Cognos server connection, or Credential if a specific Cognos username and password are used.

User Name. The username (if required) on the Cognos server.

Password. The password for this username.

DSN. The database source name (DSN) of the Cognos database.

Table Name. The name of the database table used for the export.

User Name. The username used for connection to the database used for the export.

Password. The password for this username.

Champion Challenger Job Steps

Champion Challenger Overview

Using IBM® SPSS® Collaboration and Deployment Services Deployment Manager, it is possible to compare scenarios generated by IBM® SPSS® Modeler to determine which scenario contains the most effective predictive model. The champion challenger job step evaluates a scenario (the first challenger) and compares that scenario to one or more scenarios (challengers). After the system compares the results, the best scenario becomes the new champion.

Champion. The scenario that is determined to contain the most effective model. For the initial execution of the champion challenger job step, there is no champion—only the first challenger and the corresponding list of challengers. For subsequent executions of the job step, the system determines the champion scenario.

Challenger. Challenger scenarios are compared against each other. The scenario that generates the best results then becomes the new champion scenario.

Champion Selection Process

The champion challenger comparison process consists of the following tasks:

▶ Scoring each of the competing scenarios.

▶ Evaluating the resulting scores.

▶ Comparing the evaluation results and determining which scenario is the champion.

▶ Saving the new champion to the repository (optional).

Adding Champion Challenger Work to a Job

To add a champion challenger job step to a job, select the Champion Challenger Work tool from the job palette and click on the job canvas.

Figure 5-1
Champion challenger work on the job canvas
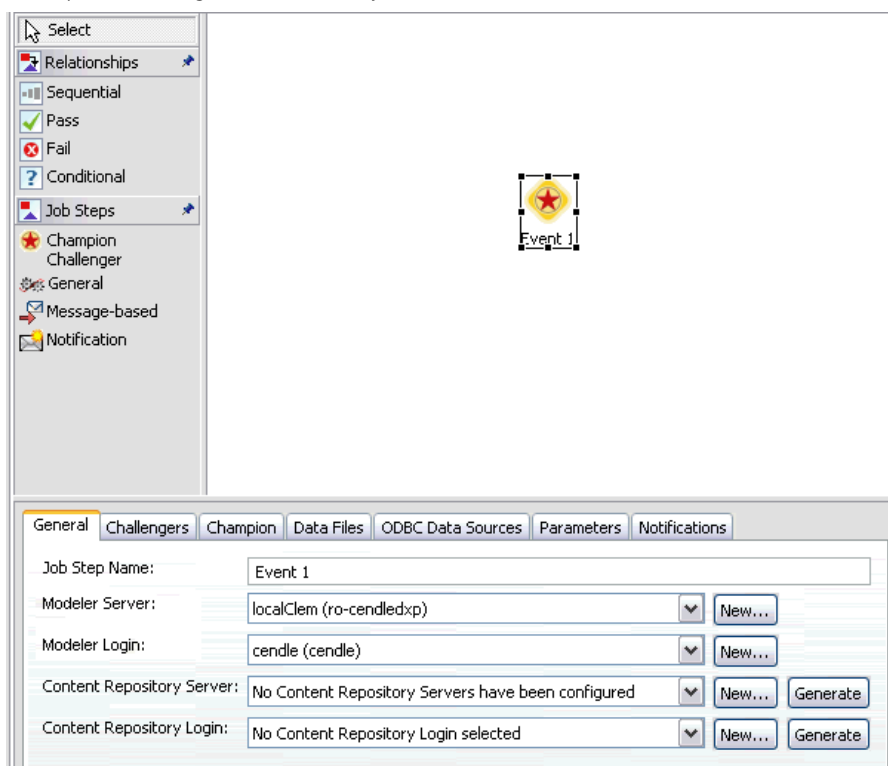
Order Dependency

Unlike other types of job steps, the tabs in the champion challenger job step are order dependent. For example, a challenger must be selected in the Challengers tab before information can be modified in the Champion tab. In addition, the information that appears in the following tabs is contingent upon the challenger scenarios selected in the Challengers table:

■ Enterprise View

■ Parameters

The information in the preceding tabs is automatically updated when the Challengers table is updated. For example, if additional challengers are selected in the Challengers table, the information in these tabs is updated accordingly.

The process for executing a champion challenger comparison consists of the following steps:

1. Providing general job information.

2. Identifying challenger scenarios.

3. Specifying champion information.

4. Viewing Enterprise View information and specifying Data Provider Definition assignments.

5. Viewing parameter information.

6. Specifying notifications.

General Information

The General Information tab contains information that pertains to the overall champion challenger job step. IBM® SPSS® Modeler server and login information is required to execute a champion challenger job step. Content repository server and login information is required to execute the job and save new champion scenarios to the IBM® SPSS® Collaboration and Deployment Services Repository. (Content repository server and login information is required, even if the user discards the results of the analysis.) Credentials are based upon the user currently logged into the system.

Figure 5-2
Champion challenger general information tab

Job step name. The name of the job step. By default, the name of the first job step is *Event 1*. Subsequent job steps are named *Event 2*, *Event 3*, and so on. The name specified here appears in the job history table after the job step is executed.

SPSS Modeler server. The SPSS Modeler server or server cluster on which the stream will be executed. The list contains all servers and server clusters currently configured to execute SPSS Modeler steps. To change the server, select from the SPSS Modeler Server drop-down list. To create a new server definition, click New to launch the server definition wizard.

SPSS Modeler login. The credential information used to access the SPSS Modeler server or server cluster. To change the credentials, select a credential definition from the SPSS Modeler Login drop-down list. To define new credentials, click New to launch the credentials definition wizard.

Content Repository Server. The content repository server allows a job to save files to an IBM SPSS Collaboration and Deployment Services Repository. Typically, the content repository server is specified when refreshing models using SPSS Modeler. To specify a content repository server, select a server from the Content Repository Server drop-down list. To create a new server definition, click New to launch the server definition wizard. To generate a content repository server definition based on the current server information, click Generate. A server definition is created and automatically populated in the *Content Repository Server* field.

Content Repository Login. The login information for the content repository server. To specify a content repository login, select a credential from the Content Repository Login drop-down list. To create a new login, click New to launch the content repository login wizard. If single sign-on is

not used to connect to the IBM SPSS Collaboration and Deployment Services Repository, click Generate to generate a content repository server login based on existing security settings. A content repository login is created and automatically populated in the *Content Repository Login* field. Login generation is not available when using single sign-on.

Challengers

At least one first challenger scenario must be selected to execute a champion challenger job step. It is important to note that the first challenger selected does not imply an order of comparison or any priority in the evaluation process. The first challenger is simply the baseline. The Application View and labels used to determine subsequent challengers are established by the first challenger. After the first challenger is selected or updated, the remaining fields in this tab are updated with information that corresponds to the first challenger.

Figure 5-3
Challengers tab

First challenger. Name of the first challenger scenario. To browse through the repository, click Browse.

First challenger label. Label associated with the first challenger scenario. To select an alternate label, click on the drop-down arrow and select from the list of available labels.

Metric. The measurement criteria by which the challenger scenarios are compared. Valid values include *accuracy* and *gains*. If *gains* is selected, then a percentile needs to be specified as well.

Challengers Table

The Challengers table contains the challengers that match the Application View and the label associated with the first challenger. Only the challengers selected from the table will be compared to the first challenger when the job step is executed. Selecting (or deselecting) a challenger from the list will cause the system to update the corresponding information on the following tabs:

■ Enterprise View
■ Parameters

Each time a job that contains a previously saved champion challenger step is opened, the list of challengers is automatically updated. New challengers may be added to the list if they match the Application View and label criteria of the first challenger. Conversely, challengers that no longer meet those criteria may be removed from the list of challengers. If a selected challenger has been removed from the repository, the system will generate a message indicating that the challenger is no longer available.

Although challenger scenarios can be selected and deselected for comparison, information in the challengers table is not modifiable. Specifically, the Challengers table contains the following information.

Name. Name of the challenger scenario.

Label. Label associated with the challenger scenario.

Description. A description of the scenario.

Modifications to the First Challenger

Changes made to the first challenger after the job is saved may affect the champion challenger analysis. For example, suppose that the first challenger is removed from the repository or the label associated with the first challenger is removed. When the Challengers tab is accessed, the system will generate a message, indicating that the first challenger is no longer available for use. In this case, a new first challenger needs to be specified.

Invalid Challengers

To be compared to one another, two scenarios must have a scoring branch that shares the same Application View. In addition, when a scenario is added to the repository, the Application View used by the scenario must also exist.

If the system cannot find challengers comparable to the first challenger selected, the Invalid Challenger dialog appears. To select a new challenger:

1. Click OK to return to the Challengers tab.

2. Select a new challenger.

Champion

Before information can be specified for a champion scenario, at least one challenger scenario must be selected. If the Champion tab is accessed before a challenger has been selected, the First Challenger Not Selected dialog appears, indicating that a challenger must be selected.

In the Champion tab, the *File name* field is text field, and a name needs to be specified by the user. To modify location, permission, and metadata information, click the corresponding Browse buttons.

Figure 5-4
Champion tab

File name. The name of the champion file.

Location. The location in which the new version of the champion file is stored.

Permission. The permissions associated with the new version of the champion scenario.

Metadata. Properties associated with the new version of the champion scenario. Specifying metadata for champion scenario output is the same as specifying metadata for other job output.

Using the Champion Scenario in Other Jobs

After the champion challenger job has been run, the resulting champion scenario can be used in other jobs. To include the champion scenario in another job, the following information is required:

■ The name of the champion scenario.
■ The location of champion scenario.

When the champion scenario is used in another job, the *LATEST* label is applied. This label cannot be modified.

Testing the Champion Scenario

By default, the system creates a new version of the champion scenario each time the champion challenger job step is run, stores the new version to the output location specified, and writes the results to the job history log. However, there may be instances in which saving a new version of the champion scenario is not desired. For example, suppose that you simply want to test the champion challenger job step.

To disable the creation of a new version, select the Do not create a new version of the champion scenario check box. If this check box is selected, the remaining options in the tab are disabled. The system will use the same information applied to the current champion scenario.

The system will execute the champion challenger job step and determine a new champion. However, a new version of the champion scenario will not be created or saved to the repository. Instead, the results will only be written to the job history log, indicating which scenario would have been chosen as the champion.

For example, suppose a champion challenger job is run and the creator of the job elected not to create a new version of the champion scenario, the resulting job history log might look like this:

```
Stream execution started
500 500
1000 1000
1500 1500
2000 2000
2500 2500
Stream execution complete, Elapsed=26.22 sec, CPU=18.97 sec
Stream execution started
1000 0
2000 0
Field 'Correct_Sum' has only one value
Field 'Count' has only one value
Field 'Traffic Light Result' has only one value
2855 145
2855 1145
2855 2145
Field 'campaign' has only one value
Field 'gold_card' has only one value
Field 'response' has only one value
Stream execution complete, Elapsed=0.39 sec, CPU=0.2 sec
Stream execution started
500 500
1000 1000
1500 1500
2000 2000
2500 2500
Stream execution complete, Elapsed=26.06 sec, CPU=17.75 sec
Stream execution started
1000 0
2000 0
Field 'Correct_Sum' has only one value
Field 'Count' has only one value
Field 'Traffic Light Result' has only one value
2855 145
2855 1145
2855 2145
Field 'campaign' has only one value
Field 'gold_card' has only one value
Field 'response' has only one value
Stream execution complete, Elapsed=0.48 sec, CPU=0.19 sec
Stream execution started
500 500
1000 1000
1500 1500
2000 2000
```

2500 2500
Stream execution complete, Elapsed=21.48 sec, CPU=17.34 sec
Stream execution started
1000 0
2000 0
Field 'Correct_Sum' has only one value
Field 'Count' has only one value
Field 'Traffic Light Result' has only one value
2855 145
2855 1145
2855 2145
Field 'campaign' has only one value
Field 'gold_card' has only one value
Field 'response' has only one value
Stream execution complete, Elapsed=0.39 sec, CPU=0.17 sec
The result for challenger cc_cartresponse.scn is 98.809.
The result for challenger cc_neuralnetresponse.scn is 98.844.
The result for challenger cc_c51response.scn is 98.809.
The declared Champion is cc_neuralnetresponse.scn.

Note the last line in the log file:

The declared Champion is cc_neuralnetresponse.scn.

This line indicates that the *cc_neuralnetresponse* scenario would have been the champion scenario.
However, a copy of this scenario was not saved to the repository because the system did not create
a new version of the scenario. If the scenario had been saved to the repository, the log would
contain an additional line, indicating that the scenario was saved to the repository—for example:

Adding artifact spsscr:/PMDemo/ModelManagement/cc_neuralnetresponse.scn.

Enterprise View

The Enterprise View tab specifies the Enterprise View settings for accessing data. This tab lists
only the node associated with the default branch of the streams in the scenarios.

Figure 5-5
Enterprise View

Enterprise View Node. Name of the Enterprise View node in the stream or scenario.

Data Provider Definition. Data provider definition associated with the Enterprise View node. To change the data provider used for a node, select the cell containing the data provider and click the resulting ellipsis (...) button. For more information, see the topic Select Data Provider Definition in Chapter 4 on p. 28.

Label. Label identifying the version of the data provider used.

Parameters

Parameters information appears for the challengers that were selected from the list of scenarios in the Challengers tab. Any changes made in this tab apply to the champion challenger job step only. Modifications made to parameter information are not propagated back to the challenger scenario saved in the IBM® SPSS® Collaboration and Deployment Services Repository. The Parameters table contains the following information.

Figure 5-6
Parameters tab

Challenger name. Name of the selected challenger scenario. The scenarios that appear in this column are the scenarios that were selected in the Challengers table.

Challenger label. Label applied to the challenger scenario. The labels that appear in this column correspond to the scenarios that were selected in the Challengers table.

Name. The parameter name.

Storage. Describes how the parameter is stored—for example, as a string.

Value. The value of the parameter.

Type. The parameter type.

Notices

This information was developed for products and services offered worldwide.

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

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

IBM Director of LicensingIBM CorporationNorth Castle DriveArmonk, NY 10504-1785U.S.A.

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

Intellectual Property LicensingLegal and Intellectual Property LawIBM Japan Ltd.1623-14, Shimotsuruma, Yamato-shiKanagawa 242-8502 Japan

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

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

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

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

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

IBM Software GroupAttention: Licensing233 S. Wacker Dr.Chicago, IL 60606USA

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

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

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

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

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

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

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

Trademarks

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

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.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.

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.

ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.

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

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

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

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

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

Index