IBM SPSS Modeler 18.4 Deployment Guide

Note

Before you use this information and the product it supports, read the information in "Notices" on page 31.

Contents

Preface

IBM SPSS Modeler is the IBM 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 products or for installation help for one of the supported hardware environments. To reach Technical Support, see the IBM web site at http://www.ibm.com/support. Be prepared to identify yourself, your organization, and your support agreement when requesting assistance.

Chapter 1. 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 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 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. 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 repository offers a search facility.

Analysts can work with files in the 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 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 correct 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 repository can be included in processing **jobs**. Jobs define an execution sequence for analytical artifacts and can be created with IBM SPSS Deployment Manager. 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.

Note: If a SPSS Modeler stream contains a node that uses a list type, the branch containing that node does not support the scoring service.

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.
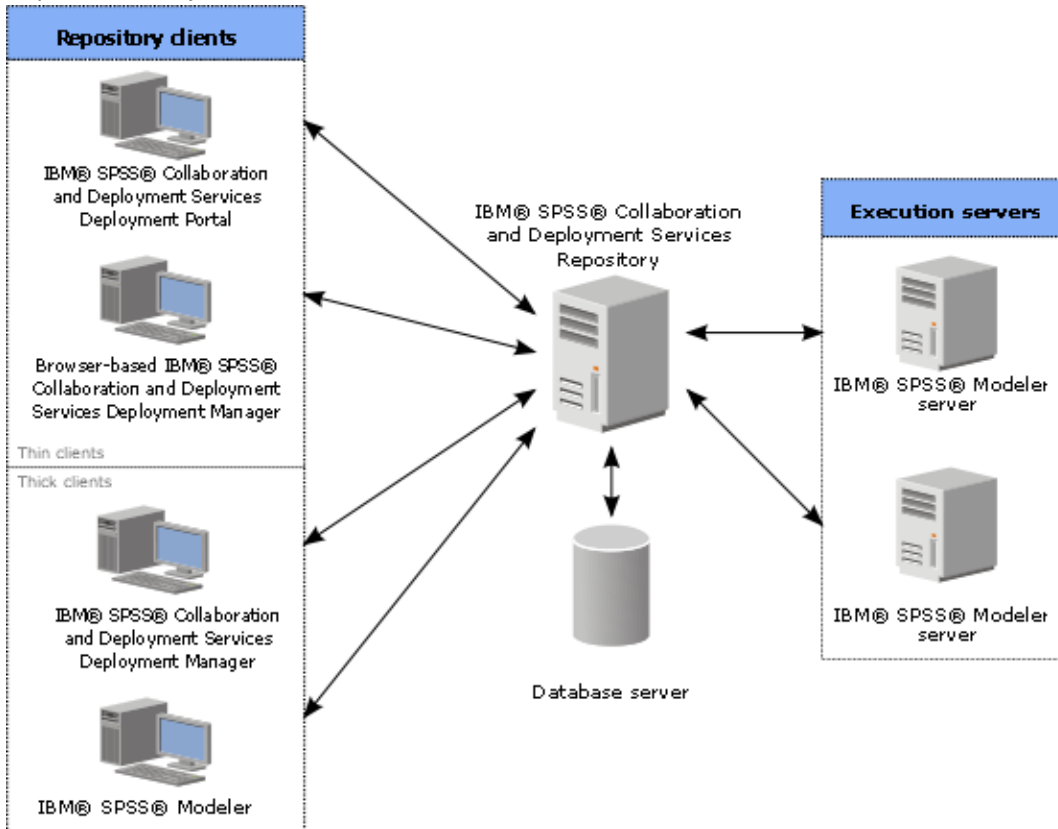
Figure 1. IBM SPSS Collaboration and Deployment Services Architecture

IBM SPSS Collaboration and Deployment Services consists of the following components:

- IBM SPSS Collaboration and Deployment Services Repository for analytical artifacts
- IBM SPSS Modeler
- IBM SPSS Deployment Manager
- IBM SPSS Collaboration and Deployment Services Deployment Portal
- Browser-based IBM SPSS Deployment Manager

IBM SPSS Collaboration and Deployment Services Repository

The repository provides a centralized location for storing analytical assets, such as models and data. The repository requires an installation of a relational database, such as IBM Db2, Microsoft SQL Server, or Oracle.

The repository includes facilities for:

- Security
- Version control
- Searching
- Auditing

Configuration options for the repository are defined using the IBM SPSS Deployment Manager or the browser-based IBM SPSS Deployment Manager. The contents of the repository are managed with

the Deployment Manager and accessed with the 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 IBM SPSS Modeler interface. Files can be stored and retrieved directly from IBM SPSS Modeler .

In addition, IBM SPSS Modeler streams 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 IBM 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 IBM SPSS Modeler documentation.

Note: If you have a job containing an SPSS Modeler stream that uses an Analytic Server source node, you must allow a direct connection between the IBM SPSS Collaboration and Deployment Services Server and Analytic Server. Otherwise, the job will fail if the firewall blocks the connection between the two servers.

IBM SPSS Deployment Manager

IBM SPSS Deployment Manager is a client application for IBM SPSS Collaboration and Deployment Services Repository that enables users to schedule, automate, and execute analytical tasks, such as updating models or generating scores.

The client application allows a user to perform the following tasks:

- View any existing files within the system, including IBM SPSS Modeler streams, scenarios, and data files
- Import files into the repository
- Schedule jobs to be executed repeatedly using a specified recurrence pattern, such as quarterly or hourly
- Modify existing job properties
- Determine the status of a job
- Specify email notification of job status

In addition, the client application allows users to perform administrative tasks for IBM SPSS Collaboration and Deployment Services, including:

- Manage users
- Configure security providers
- Assign roles and actions

Browser-based IBM SPSS Deployment Manager

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

- Setting system configuration options
- 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.

You typically access the browser-based IBM SPSS Deployment Manager at the following URL:

```
http://<host IP address>:<port>/security/login
```

Note: An IPv6 address must be enclosed in square brackets, such as [3ffe:2a00:100:7031::1].

If your environment is configured to use a custom context path for server connections, include that path in the URL.

```
http://<host IP address>:<port>/<context path>/security/login
```

IBM SPSS Collaboration and Deployment Services Deployment Portal

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

The web portal includes the following functionality:

- Browsing the repository content by folder
- Opening published content
- Running jobs
- Generating scores using models stored in the repository
- Searching repository content
- Viewing content properties
- Accessing individual user preferences, such as email address and password, general options, subscriptions, and options for output file formats

You typically access the home page at the following URL:

```
http://<host IP address>:<port>/peb
```

Note: An IPv6 address must be enclosed in square brackets, such as [3ffe:2a00:100:7031::1].

If your environment is configured to use a custom context path for server connections, include that path in the URL.

```
http://<host IP address>:<port>/<context path>/peb
```

Browser-based IBM SPSS Deployment Manager

The browser-based IBM SPSS 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.

Execution servers

Execution servers provide the ability to execute resources stored within the 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.

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 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 repository uses the specified execution servers to perform the corresponding analyses.

Working with IBM SPSS Modeler files

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

1. Define an execution server for processing the IBM SPSS Modeler stream or scenario if an appropriate definition does not already exist.
2. Add the IBM SPSS Modeler stream or scenario to a job as a job step.
3. Specify the job properties, including any schedules for execution.

Chapter 2. 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 IBM 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 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. Note that the server types available depend on which product adapters are installed to the repository.
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 IBM SPSS Modeler servers used to process job steps. IBM SPSS Modeler streams are executed on the IBM SPSS Modeler server.

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

Note: When you run a IBM SPSS Collaboration and Deployment Services job in an Evaluation stream, a temporary file is created. By default, the file is saved to the IBM SPSS Modeler Server installation directory. You can change the default data folder that the temp files are saved to when you create the IBM SPSS Modeler Server connection in IBM SPSS Modeler.

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.

Chapter 3. IBM SPSS Modeler Job Steps

Working with IBM SPSS Modeler streams

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

You can work with IBM SPSS Modeler streams in Deployment Manager. Like any other step, an IBM SPSS Modeler stream must 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

Note: If you have a job containing an SPSS Modeler stream that uses an Analytic Server source node, you must allow a direct connection between the IBM SPSS Collaboration and Deployment Services Server and Analytic Server. Otherwise, the job will fail if the firewall blocks the connection between the two servers.

IBM SPSS Modeler Server configuration

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

- Create an IBM SPSS Modeler server definition. See the topic " IBM SPSS Modeler server parameters" on page 7 for more information.
- Define server credentials.

Viewing IBM SPSS Modeler job properties

When you click 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 an IBM SPSS Modeler stream in the job canvas.

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 IBM 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.

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

Override Type Use this option to specify whether the settings in either a job step or the node in a stream take priority during execution.

- **Job overrides stream** If you select this option, when you execute the job, the node properties use the values that you set in the job step.
- **Stream overrides job** Selecting this option disables the node property controls in the job step; the only exceptions are some of the credential controls. If you change the node properties in the stream, when you execute the job step, the changed node properties are used.

The **Override Type** options affect the credential settings in the following ways:

- If you choose **Job overrides stream**, you can edit the credential selection controls, and your selected credentials take effect at run-time.
- If you choose **Stream overrides job**, there are two possible cases:
  - When you create nodes in SPSS Modeler Client that use credentials, and if the authentication mode is **Stored credential**, and the credential name exists in IBM SPSS Collaboration and Deployment Services, the credential selection control is unavailable. However, if the credential name does not exist in IBM SPSS Collaboration and Deployment Services, you can use the credential selection control; your selected credentials take effect at run-time.
  - When you create nodes in SPSS Modeler Client that use credentials, if the authentication mode is **Username and password**, you can use the credential selection control; your selected credentials take effect at run-time.

Object version. The labeled version of the IBM 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.

IBM SPSS Modeler server. The IBM 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 IBM SPSS Modeler steps. To change the server, select from the **IBM SPSS Modeler Server** drop-down list. To create a new server definition, click **New** to launch the server definition wizard.

IBM SPSS Modeler login. The credential information used to access the IBM SPSS Modeler server or server cluster. To change the credentials, select a credential definition from the **IBM 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 IBM 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.

Type. The way you want to run the stream.

- **Run Stream.** With this option, you can run the complete stream or, if the stream contains branches, choose one or more branches to run. In the latter case, the **Run Options** list is displayed. All branches are checked by default, meaning that all branches will execute when the stream runs. Uncheck any branches that you do not want to execute; doing so can improve performance.

  Note: If the stream contains a script and is configured to "Run Script" on execution, no execution branches will be displayed in the list in the IBM SPSS Collaboration and Deployment Services job editor.

- **Model Management.** Choose this option if you want the stream to make use of model management features, such as evaluation, refresh and score. Choose the feature from the Types list that is displayed. **Evaluation Options**

  Performance. Specify the threshold percentages for the following categories: Bad or 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 streams.

  Metrics. The measurement criteria by which the stream'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

  Lists all model-nugget relationships in the stream that can be refreshed using the node names defined in the stream. 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.

  Score Options

  Choose the branch or branches you want to designate as scoring branches.

Guidelines for Providing IBM SPSS Modeler and Content Repository Information

For all IBM SPSS Modeler job steps, both IBM SPSS Modeler server and IBM SPSS Modeler login information are required.

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 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 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.

IBM SPSS Modeler Job Properties - Data View

If the stream contains a Geo Spatial node, the connection details are displayed here.

Node Name. The name of the Data View node.

Analytic Data View. The analytic data view used.

Label. The label used.

Table Name. The name of the database table used.

Data Access Plan. Select a data access plan from the analytic data view. A data access plan associates the data model tables in an analytic data view with physical data sources. An analytic data view typically contains multiple data access plans. When you change the data access plan in use, you change the data used by your stream. For example, if the analytic data view contains a data access plan for training a model and a data access plan for testing a model, you can switch from training data to testing data by changing the data access plan in use.

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 Data Sources. The current ODBC data source name (DSN). To change to a different ODBC data source, click the cell containing current data source name, then click the "..." button that is displayed. Doing so displays a dialog box from where you can choose an existing DSN or create a new one. Note that the job step settings always override the stream; therefore, if you modify the stream to use a different data source you must also edit the job to use the same source, otherwise the job may fail to run.

Credentials. To change the database username and password when changing the ODBC data source, click the cell containing the current credentials, then click the "..." button that is displayed. Doing so displays a dialog box from where you can choose an existing credential definition or create a new one.

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

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

Changing the ODBC connection

You can change the ODBC connection either to one that is already defined, or to one that you create here.

Use Existing ODBC DSN. Click **Browse** to display a list of existing ODBC DSNs from which you can choose.

Create New ODBC DSN. Select this option and click **New** to define a new ODBC connection for use with stream job steps.

Browsing for an ODBC connection

This is a list of all the ODBC connections that have been defined on this host and to which you have access.

Choose one from the list and click **OK**.

Changing the database credentials

When changing the connection, you can switch to use the corresponding database username and password here. You can use an existing credential definition from the repository, or you can create a new definition.

Generate Repository credentials from existing Username and Password. (displayed only if importing a job from Release 4.2.1 of IBM SPSS Collaboration and Deployment Services) Click the **Generate** button to generate repository credentials from an existing username and password if these have been set in the imported job. For example, if the username sa already exists, repeatedly clicking **Generate** will create the usernames sa1, sa2, and so on.

Use Existing Credential. Click the **Browse** button to display a list of existing credential definitions from which you can choose.

Create New Credential. Select this option and click **New** to create a new credential definition for use with the current connection.

Browsing for a credential definition

This is a list of all the credential definitions that have been created on this host and to which you have access.

Choose one from the list and click **OK**.

IBM SPSS Modeler Job Properties - Geo Spatial

If the stream contains a Geo Spatial node, the connection details are displayed here.

Source Type. The datasource type.

File Name. If using a Shape file, specify the file name.

Map Service URL. If using a Map service, specify the URL to the service.

Map ID. Specify the map ID.

IBM SPSS Modeler Job Properties - Parameters

You can modify parameters for IBM SPSS Modeler streams using the IBM SPSS 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.

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, double-click on the results that you want to view in the Results column of the Job Step History table. 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.

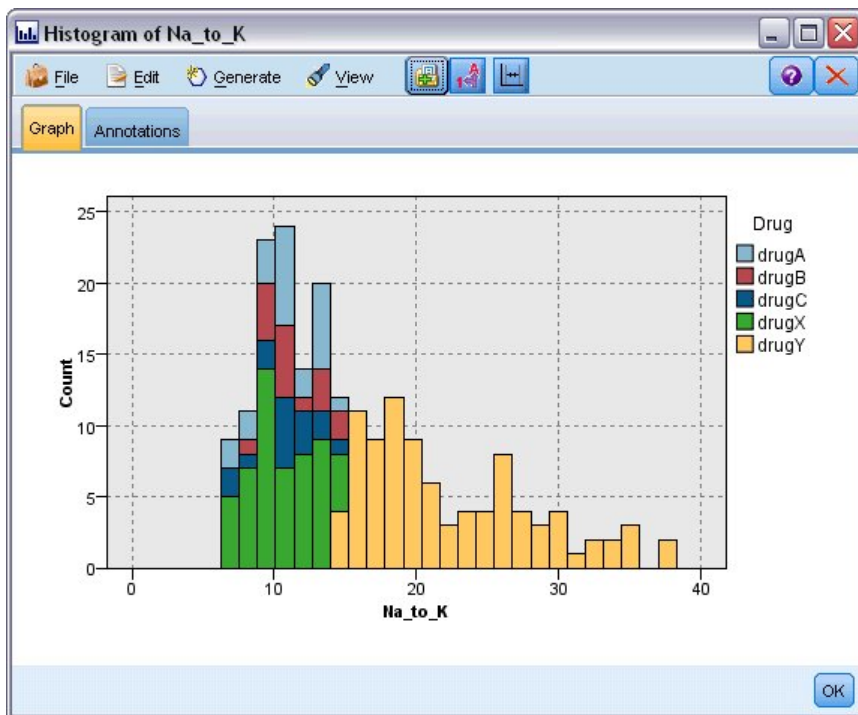*Figure 2. Sample histogram*

IBM SPSS Modeler Job properties - Cognos Import

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

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.

Credentials. The username and password (if required) on the Cognos server.

Note: Cognos credentials must be created in a domain, which represents the Cognos namespace ID.

IBM SPSS Modeler Job properties - Cognos Export

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

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.

Credentials. The username and password (if required) on the Cognos server.

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

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

Credentials. The username and password used for connection to the database used for the export.

Note: Cognos credentials must be created in a domain, which represents the Cognos namespace ID.

IBM SPSS Modeler Job Properties - Legacy TM1 Import

If the stream contains a Legacy TM1 Import node, the connection details are displayed here.

Node Name. The name of the Legacy TM1 Import node.

TM1 Server. The Cognos TM1 Server name.

Cube. The TM1 cube from which the data will be imported.

View. The view to import from the TM1 cube.

Credential. Credential for the Cognos TM1 Server.

IBM SPSS Modeler Job Properties - Legacy TM1 Export

If the stream contains a Legacy TM1 Export node, the connection details are displayed here.

Node Name. The name of the Legacy TM1 Export node.

TM1 Server. The Cognos TM1 Server name.

Cube. The TM1 cube into which the data will be exported.

Credential. Credential for the Cognos TM1 Server.

IBM SPSS Modeler Job Properties - TM1 Import

If the stream contains a TM1 Import node, the connection details are displayed here.

Node Name. The name of the TM1 Import node.

Admin Host. The Cognos TM1 administration host.

TM1 Server. The Cognos TM1 Server name.

Cube. The TM1 cube from which the data will be imported.

View. The view to import from the TM1 cube. You can import a public or a private view.

Rows. The rows to import from the TM1 cube.

Columns. The columns to import from the TM1 cube.

Credential. Credential for the Cognos TM1 Server.

IBM SPSS Modeler Job Properties - TM1 Export

If the stream contains a TM1 Export node, the connection details are displayed here.

Node Name. The name of the TM1 Export node.

Admin Host. The Cognos TM1 administration host.

TM1 Server. The Cognos TM1 Server name.

Cube. The TM1 cube into which the data will be exported.

Measure. The measure that will be exported.

Credential. Credential for the Cognos TM1 Server.

IBM SPSS Modeler Job Properties - Analytic Server Import

If the stream contains an Analytic Server Source node, the connection details are displayed here.

Use Default Analytic Server. True or **False** specifies whether the default Analytic Server connection defined by an administrator in `options.cfg` is used (**True**), or if a different Analytic Server is used – the one specified in the stream/job (**False**).

URL. The URL for the Analytic Server in the format `https://hostname:port/contextroot`, where hostname is the IP address or host name of the Analytic Server, `port` is its port number, and `contextroot` is the context root of the Analytic Server.

Tenant. The name of the tenant that the SPSS Modeler Server is a member of.

Credential. The credential for logging on to Analytic Server.

Service Principal Name. The Kerberos service principal name.

Config File Path. The Kerberos service config file path.

Data Source. The data source name on Analytic Server.

IBM SPSS Modeler Job Properties - Analytic Server Export

If the stream contains an Analytic Server Export node, the connection details are displayed here.

Use Default Analytic Server. True or **False** specifies whether the default Analytic Server connection defined by an administrator in `options.cfg` is used (**True**), or if a different Analytic Server is used – the one specified in the stream/job (**False**).

URL. The URL for the Analytic Server in the format `https://hostname:port/contextroot`, where hostname is the IP address or host name of the Analytic Server, `port` is its port number, and `contextroot` is the context root of the Analytic Server.

Tenant. The name of the tenant that the SPSS Modeler Server is a member of.

Credential. The credential for logging on to Analytic Server.

Service Principal Name. The Kerberos service principal name.

Config File Path. The Kerberos service config file path.

Data Source. The data source name on Analytic Server.

IBM SPSS Modeler Job Properties - Notifications

Allows you to specify email notifications for job step failure and success.

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 Deployment Manager.

To view the stream in the IBM SPSS Modeler application, double-click on the stream in the Content Explorer. The system launches the IBM SPSS Modeler application and displays the stream in IBM SPSS Modeler.

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)--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 with 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 IBM SPSS Modeler streams.

Table 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 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.
- **Text Analytics node restrictions.** You cannot use the SPSS Modeler Text Analytics File List or Web Feed nodes for scoring within an IBM SPSS Collaboration and Deployment Services - Scoring configuration.

In addition to the items listed above, there are the following specific constraints.

- No unlock checkbox is available when you store a stream in the IBM SPSS Collaboration and Deployment Services repository. Unlock is the default when storing streams. To lock or unlock an object, choose **Tools** > **Repository** > **Explore**, navigate to the object, and right-click on its name to display the context menu.
- In Deployment Manager, when you run a job that contains a stream with an Evaluation node set to produce a Gains graph, the graph output may be incomplete if the system is running under Oracle Weblogic 11g using the Oracle JRockit JRE. To avoid this problem, use the IBM JRE.

Node Types

When you open an IBM SPSS Modeler stream from IBM SPSS 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 *IBM SPSS Modeler Source, Process and Output Nodes* and the *IBM 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
```

Chapter 4. Scoring Service

The Scoring Service allows client applications to employ real-time scores derived from predictive models developed in IBM SPSS Modeler. The service fetches the specified model, loads it, invokes the correct scoring implementation, and returns the result to the client.

Scoring is the process of generating real-time values by supplying predictive models with input data. A scoring model is any artifact that can be used to produce output values given input data. In general, to use a model for generating scores:

1. Select a model to use for scoring from the IBM SPSS Collaboration and Deployment Services Repository.
2. Define a scoring configuration for the model.
3. Supply the configured model with data and generate scores.

For more information about scoring and the Scoring Service, see the IBM SPSS Collaboration and Deployment Services documentation.

IBM SPSS Modeler stream limitations

When you work with streams in IBM SPSS Deployment Manager, the system has the following constraints for the Scoring Service:

- **Supernodes.** Source nodes inside supernodes are not supported, and terminal nodes inside supernodes are not supported

- **Geospatial nodes.** Geospatial nodes are not supported.

- **Model builder nodes.** Model builder node as the terminal node is not supported.

- **In-database mining.** Streams containing in-database mining nuggets are not supported.

- **UDF, UDA, and WUDA.** Streams using database functions (UDF), database aggregations (UDA), or database window aggregates (WUDA) are not supported.

- **Source nodes** Source nodes with an output data model containing a list type are not supported.

- **Terminal nodes.** Terminal nodes with an input data model containing a list type are not supported.

- **Text Analytics node restrictions.** You cannot use the SPSS Modeler Text Analytics File List or Web Feed nodes for scoring within an IBM SPSS Collaboration and Deployment Services - Scoring configuration.

- **Analytic Server source nodes.** If you have a job containing an SPSS Modeler stream that uses an Analytic Server source node, you must allow a direct connection between the IBM SPSS Collaboration and Deployment Services Server and Analytic Server. Otherwise, the job will fail if the firewall blocks the connection between the two servers.

Chapter 5. Champion challenger job steps

Champion challenger overview

By using IBM SPSS Deployment Manager, it is possible to compare model files generated by IBM SPSS Modeler to determine which file contains the most effective predictive model. The champion challenger job step evaluates a model and compares it to one or more challengers.

After the system compares the results, the best model becomes the new champion.

Champion. The champion corresponds to 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.

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

Champion Selection Process

The champion challenger comparison process consists of the following tasks:

1. Scoring each of the competing models.
2. Evaluating the resulting scores.
3. Comparing the evaluation results and determining which challenger is the champion.
4. 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 tool from the job palette and click the job canvas.

Model evaluation metrics

Model evaluation and comparison can focus on accuracy, gains, or accreditation.

- **Accuracy.** The accuracy of a model reflects the percentage of target responses that are predicted correctly. Models having a high percentage of correct predictions are preferred to those having a low percentage.
- **Gains.** The gains statistic is an indicator of the performance of a model. This measure compares the results from a model to the results obtained without using a model. The improvement in the results when using the model is referred to as the gains. When comparing two models, the model having the higher gains value at a specified percentile is preferred.
- **Accreditation.** Model accreditation reflects the credibility of a model. This approach examines the similarity between new data and the training data on which a model is based. Accreditation values vary from 0 to 1, with high values indicating greater similarity between the predictors in the two data sets. When comparing two models, the model having the higher accreditation value is based on training data that is more similar to the new data, making it more credible and preferred.

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 on some tabs is contingent upon the challengers selected in the Challengers table.

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

1. Providing general job information.
2. Identifying challengers.
3. Specifying champion information.
4. Viewing parameter information.
5. Specifying notifications.

General information

The General 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 champions to the IBM SPSS Collaboration and Deployment Services Repository. (Content repository server and login information is required, even if you discard the results of the analysis.) Credentials are based upon the user currently logged in to the system.

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.

IBM SPSS Modeler server. The IBM 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 IBM SPSS Modeler steps. To change the server, select from the **IBM SPSS Modeler Server** drop-down list. To create a new server definition, click **New** to launch the server definition wizard.

IBM SPSS Modeler login. The credential information used to access the IBM SPSS Modeler server or server cluster. To change the credentials, select a credential definition from the **IBM 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 IBM 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 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 data source 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.

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

First challenger label. Label associated with the model file containing the first challenger. Specify this value when selecting the first challenger.

Data Source Challenger. The challenger supplying the data source node used for the job step. Click **Browse** to choose this challenger from the list of entries selected in the Challengers table.

Metric. The measurement criteria by which the challengers are compared. Valid values include *accreditation*, *accuracy* and *gains*. If *gains* is selected, then a percentile needs to be specified as well. See the topic "Model evaluation metrics" on page 23 for more information.

Challengers Table

The Challengers table lists the default score branches for challengers that match the data source 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 clearing) a challenger from the list will cause the system to update the corresponding information on the other job step tabs accordingly.

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 data source 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 challengers can be selected and cleared for comparison, information in the challengers table is not modifiable. Specifically, the Challengers table contains the following information.

Name. Name of the challenger.

Label. Label associated with the challenger.

Description. A description of the challenger.

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.

Selecting challengers

To select challenger models for inclusion in champion challenger analyses, perform the following steps:

1. On the Challenger tab of a Champion Challenger step, click **Browse** for the First Challenger. If you are adding challengers manually, click **Add** for the Challengers table.
2. Select the model file by clicking **Browse**. The model file is a IBM SPSS Modeler stream that contains a default score branch with a valid model nugget.
3. Select the label designating the version of the selected model file to use.
4. In the Challengers table, select the score branch to use.
5. Click **OK**.

Invalid challengers

To be compared, model files must have scoring branches that use a common data structure.

Data characteristics that must match across challengers include the following items:

- Data sources must have the same number of fields.
- The field names must be identical across data sources.
- The field measurement levels must be identical across data sources.

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.

Selecting challenger data sources

To select the data source used in the champion challenger analyses, perform the following steps:

1. On the Challenger tab of a Champion Challenger step, click **Browse** for the data source challenger.
2. From the list of challengers included in the analysis, select the score branch that includes the data source to use.
3. Click **OK**.

Champion

Before information can be specified for a champion, at least one challenger 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.

Do not create a new version of the champion. Select this option to prevent the creation of a new version of the champion. In this case, the selected labeled version of the champion will be modified. Clear this option to create a new version of the champion instead of modifying the labeled version.

File name. The name to use for the copy of the challenger identified as the champion.

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

Permission. The permissions associated with the copy of the champion.

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

Using the champion in other jobs

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

- The name of the champion.
- The location of champion.

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

Testing the champion

By default, the system creates a new copy of the champion each time the champion challenger job step is run, stores the copy to the output location specified, and writes the results to the job history log.

However, there may be instances in which saving a copy of the champion is not wanted. For example, suppose that you simply want to test the champion challenger job step.

To disable the creation of a copy, select the **Do not create a new version of the champion** 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.

The system will execute the champion challenger job step and determine a new champion. However, a new version of the champion will not be created or saved to the repository. Instead, the results will only be written to the job history log, indicating which challenger 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. 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.str is 98.809.
The result for challenger cc_neuralnetresponse.str is 98.844.
The result for challenger cc_c51response.str is 98.809.
The declared Champion is cc_neuralnetresponse.str.
```

Note the last line in the log file:

```
The declared Champion is cc_neuralnetresponse.str.
```

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

```
Adding artifact spsstr:/PMDemo/ModelManagement/cc_neuralnetresponse.str.
```

Data files

Data files information appears for the challengers that were selected from the list of challengers in the Challengers tab.

Any changes made in this tab apply to the champion challenger job step only. Modifications made to the data file information are not propagated back to the challenger saved in the IBM SPSS Collaboration and Deployment Services Repository. 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 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 the location, click in the column and then click the resulting ellipsis button. The Input File Location dialog box opens. Change the location as necessary.

Data view

Analytic data view information appears for the challengers that were selected from the list of challengers in the Challengers tab.

Any changes made in this tab apply to the champion challenger job step only. Modifications made to the data view information are not propagated back to the challenger saved in the IBM SPSS Collaboration and Deployment Services Repository. The data view table contains the following information.

Node Name. The name of the data view node that contains the data used by the stream. The node name is not modifiable.

Analytic Data View. The analytic data view referenced by the data view node.

Label. Label identifying the version of the analytic data view used.

Table Name. The table containing the input data fields.

Data Access Plan. The plan supplying the data records for the input data fields. To change the data access plan used for a node, select the cell containing the access plan and click the resulting ellipsis (...) button.

ODBC data sources

ODBC data source information appears for the challengers that were selected from the list of challengers in the Challengers tab.

Any changes made in this tab apply to the champion challenger job step only. Modifications made to the ODBC data source information are not propagated back to the challenger saved in the IBM SPSS

Collaboration and Deployment Services Repository. 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 Data Sources. The current ODBC data source name (DSN). To change to a different ODBC data source, click the cell containing current data source name, then click the "..." button that is displayed. Doing so displays a dialog box from where you can choose an existing DSN or create a new one. Note that the job step settings always override the stream; therefore, if you modify the stream to use a different data source you must also edit the job to use the same source, otherwise the job may fail to run.

Credentials. To change the database username and password when changing the ODBC data source, click the cell containing the current credentials, then click the "..." button that is displayed. Doing so displays a dialog box from where you can choose an existing credential definition or create a new one.

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

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

Cognos import

If the model files contain any IBM Cognos BI source nodes, the Cognos connection details are displayed here.

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.

Credentials. The username and password (if required) on the Cognos server.

Note: Cognos credentials must be created in a domain, which represents the Cognos namespace ID.

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

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

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

IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US

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

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

INTERNATIONAL BUSINESS MACHINES CORPORATION 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 jurisdictions 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 websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide 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 Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785
US

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.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

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-IBMproducts. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

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 actual people or business enterprises is entirely coincidental.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

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

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

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

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

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

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Terms and conditions for product documentation

Permissions for the use of these publications are granted subject to the following terms and conditions.

Applicability

These terms and conditions are in addition to any terms of use for the IBM website.

Personal use

You may reproduce these publications for your personal, noncommercial use provided that all proprietary notices are preserved. You may not distribute, display or make derivative work of these publications, or any portion thereof, without the express consent of IBM.

Commercial use

You may reproduce, distribute and display these publications solely within your enterprise provided that all proprietary notices are preserved. You may not make derivative works of these publications, or reproduce, distribute or display these publications or any portion thereof outside your enterprise, without the express consent of IBM.

Rights

Except as expressly granted in this permission, no other permissions, licenses or rights are granted, either express or implied, to the publications or any information, data, software or other intellectual property contained therein.

IBM reserves the right to withdraw the permissions granted herein whenever, in its discretion, the use of the publications is detrimental to its interest or, as determined by IBM, the above instructions are not being properly followed.

You may not download, export or re-export this information except in full compliance with all applicable laws and regulations, including all United States export laws and regulations.

IBM MAKES NO GUARANTEE ABOUT THE CONTENT OF THESE PUBLICATIONS. THE PUBLICATIONS ARE PROVIDED "AS-IS" AND WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, AND FITNESS FOR A PARTICULAR PURPOSE.

Index

A

analytic data view
    champion challenger 28

C

challenger models
    data sources 26
champion challenger 23–26, 28, 29
champion models 18, 23, 26
champion vs. challenger 18
Cognos
    champion challenger 29
collaboration 1
comparing
    models 18, 23

D

data files
    champion challenger 28
deployment 2

E

execution servers
    IBM SPSS Modeler 3, 6

F

files
    IBM SPSS Modeler 9, 11–14, 17, 21
folder
    resource definitions 7

I

IBM SPSS Collaboration and Deployment Services
Deployment Manager 3, 4
IBM SPSS Collaboration and Deployment Services
Deployment Portal 3, 5
IBM SPSS Collaboration and Deployment Services
Repository 3
IBM SPSS Modeler
    execution server 7
    execution servers 3, 6
    server definition 7
    streams 9, 11–14, 17, 21

J

job step name
    champion challenger 24
jobs
    IBM SPSS Modeler 9, 11–14, 17, 21

M

models
    champion challenger 23
    champion vs. challenger 18
    comparing 18, 23
    evaluating 18, 23
    IBM SPSS Modeler 18, 23
    monitoring 18, 23

O

ODBC data sources
    champion challenger 28

P

parameters 17

R

resource definitions
    server definitions 7

S

scoring service 21
scripting
    champion challenger 23
    champion vs. challenger 18
    IBM SPSS Modeler 18, 23
server clusters 3, 6
server definitions
    adding 7
    IBM SPSS Modeler 7
    modifying 8
steps
    IBM SPSS Modeler 9, 11–14, 17, 21
streams
    IBM SPSS Modeler 9, 11–14, 17, 21
supernodes 17

U

user preferences 5