

*IBM SPSS Modeler 18.5* 部署指南



## 注

在使用本资料及其支持的产品之前，请阅读第 25 页的『[注意事项](#)』中的信息。

## 产品信息

本版本适用于的版本 18、发行版 4、IBM® SPSS Modeler 的修订 0 以及所有后续版本和修改，除非在新版本中另有说明

© Copyright International Business Machines Corporation .

# 内容

前言.....	V
<b>第 1 章 概况.....</b>	<b>1</b>
IBM SPSS 协作和部署服务.....	1
协作.....	1
部署.....	1
系统体系结构.....	2
IBM SPSS Collaboration and Deployment Services Repository.....	2
IBM SPSS Modeler with collaboration.....	3
IBM SPSS Deployment Manager.....	3
IBM SPSS Collaboration and Deployment Services Deployment Portal.....	4
基于浏览器的 IBM SPSS Deployment Manager.....	4
执行服务器.....	4
使用 IBM SPSS Modeler 文件.....	4
<b>第 2 章 服务器定义.....</b>	<b>5</b>
添加新服务器定义.....	5
IBM SPSS Modeler 服务器参数.....	5
修改服务器定义.....	5
<b>第 3 章 IBM SPSS Modeler 作业步骤.....</b>	<b>7</b>
处理 IBM SPSS Modeler 流.....	7
IBM SPSS Modeler Server 配置.....	7
查看 IBM SPSS Modeler 作业属性.....	7
IBM SPSS Modeler 作业属性 - 常规.....	7
IBM SPSS Modeler 作业属性 - 数据文件.....	9
IBM SPSS Modeler 作业属性 - 数据视图.....	9
IBM SPSS Modeler 作业属性 - ODBC 数据源.....	9
IBM SPSS Modeler 作业属性 - 地理空间.....	10
IBM SPSS Modeler 作业属性 - 参数.....	10
IBM SPSS Modeler 作业属性 - 结果.....	11
IBM SPSS Modeler 作业属性 - Cognos 导入.....	11
IBM SPSS Modeler 作业属性 - Cognos 导出.....	12
IBM SPSS Modeler 作业属性 - 遗存 TM1 导入.....	12
IBM SPSS Modeler 作业属性 - 遗存 TM1 导出.....	12
IBM SPSS Modeler 作业属性 - TM1 导入.....	12
IBM SPSS Modeler 作业属性 - TM1 导出.....	13
IBM SPSS Modeler 作业属性 - Analytic Server 导入.....	13
IBM SPSS Modeler 作业属性 - Analytic Server 导出.....	13
IBM SPSS Modeler 作业属性 - 通知.....	14
在 IBM SPSS Modeler 中查看流.....	14
IBM SPSS Modeler 完成代码.....	14
IBM SPSS Modeler 流限制.....	14
节点类型.....	15
脚本示例：比较优胜者与参选者模型.....	15
<b>第 4 章 评分服务.....</b>	<b>17</b>
IBM SPSS Modeler 流限制.....	17
<b>第 5 章 优胜者参选者作业步骤.....</b>	<b>19</b>

优胜者参选者概述.....	19
模型评估度量.....	19
顺序相关性.....	19
常规信息.....	20
参选者.....	20
选择参选者.....	21
无效参选者.....	21
选择参选者数据源.....	21
优胜者.....	21
测试优胜者.....	22
数据文件.....	23
数据视图.....	23
ODBC 数据源.....	24
Cognos 导入.....	24
<b>注意事项.....</b>	<b>25</b>
商标.....	26
产品文档的条款和条件.....	26
<b>索引.....</b>	<b>27</b>

# 前言

---

IBM SPSS Modeler 是 IBM 企业级的数据挖掘工作台。SPSS Modeler 通过深度的数据分析帮助组织改进与客户和市民的关系。组织通过借助源自 SPSS Modeler 的洞察力可以留住优质客户，识别交叉销售机遇，吸引新客户，检测欺诈，降低风险，促进政府服务交付。

SPSS Modeler 的可视化界面让用户可以应用他们自己的业务专长，这将生成更加强有力的预测模型，缩减实现解决方案所需时间。SPSS Modeler 提供了多种建模技术，例如预测、分类、分割和关联检测算法。模型创建成功后，通过 IBM SPSS Modeler Solution Publisher，在广泛的企业内交付给决策者，或通过数据库交付。

## 关于 IBM Business Analytics

IBM Business Analytics 软件提供完整、一致和正确的信息，决策人依据此信息来提高业务性能。企业智能、预测分析、财务业绩和战略管理的完整产品组合，和分析应用程序一起提供对当前业绩的清晰、直接和实用的洞察力，以及预测未来结果的能力。结合丰富的行业解决方案，久经证明的实践和专业服务以及各种规模的组织都能够实现最高生产力、确信地自动作出决策以及获取更好的结果。

作为此产品服务组合的组成部分，IBM SPSS Predictive Analytics 软件可帮助组织预测未来事件，并在该洞察的基础上提前行动以实现更好的业务结果。减少欺诈和降低风险时，世界范围的商业、政府和学术客户都依赖 IBM SPSS 技术作为吸引、保留和增加客户的竞争优势。通过在日常活动中融入 IBM SPSS 软件，成为预测企业的组织可指引并实现决策的自动化，以满足企业目标并实现可衡量的竞争优势。有关详细信息或要联系一位代表，请访问 <http://www.ibm.com/spss>。

## 技术支持

技术支持可供维护客户使用。客户可就 IBM 产品使用问题或某一受支持硬件环境的安装帮助寻求技术支持。要寻求技术支持，请访问 IBM Web 站点：<http://www.ibm.com/support>。请求帮助时，请准备好标识您自身、组织和支持协议。



# 第 1 章 概况

## IBM SPSS 协作和部署服务

IBM SPSS 协作和部署服务 是企业级应用程序，支持广泛使用和部署预测性分析。

IBM SPSS 协作和部署服务 以集中、安全、可审核的方式存储分析资产，具备管理和控制预测性分析过程的高级功能，并以成熟的机制将分析处理结果交付给用户。IBM SPSS 协作和部署服务 的优点包括：

- 保护分析资产的价值
- 确保符合法规要求
- 确保分析人员的工作效率
- 将管理分析的 IT 成本最小化

IBM SPSS 协作和部署服务 允许您安全地管理各类分析数据，并促进开发者与用户之间的更有效协作。此外，该部署工具还确保人员可及时获取所需的信息，从而采取相应的措施。

### 协作

协作是指有效率地共享和复用分析资产的能力，这是在整个企业内开发和实现分析的关键。

分析员需要一个位置来存放应该提供给其他分析员或业务用户使用的文件。该位置需要实施文件版本控制，用来管理分析的演进。需要安全性来控制文件的访问和修改。最后，需要备份和恢复机制以保护这些重要的业务数据，避免丢失。

为了应对这些需求，IBM SPSS 协作和部署服务 提供了用于存储资产的存储库，该存储库使用与大多数文件系统类似的文件夹层次结构。存储在 IBM SPSS Collaboration and Deployment Services Repository 中的文件可供整个企业内的用户使用，前提是这些用户具有适当的访问许可权。该存储库提供了搜索工具，可帮助用户查找数据。

分析员可以从利用 IBM SPSS 协作和部署服务 服务接口的客户机应用程序中处理存储库中的文件。诸如 IBM SPSS Statistics 与 IBM SPSS Modeler 等产品则允许直接同存储库中的文件进行交互。分析员可以在开发过程中存储文件的一个版本，稍后检索该版本并继续修改，直到它最终完成并准备好移入生产过程为止。这些文件可以包括运行分析过程的定制接口，以使业务用户能够利用分析员的工作。

使用存储库将提供一个可轻松备份和恢复的集中分析资产存储位置，从而保护业务。此外，用户、文件和版本标签级别的许可权控制对各个资产的访问。版本控制和对象版本标签可确保生产过程中使用正确的资产版本。最后，日志记录功能支持跟踪文件和系统修改。

### 部署

要实现预测性分析的所有有点，那么分析资产需要提供业务决策的输入。部署将分析与传递结果给人员的操作之间的差异连接起来，并按计划或实时处理它。

在 IBM SPSS 协作和部署服务 中，存储在存储库中的各个文件可以包括在处理**作业**中。作业定义了分析工件的执行顺序，可通过 IBM SPSS Deployment Manager 创建。执行结果可以存储在存储库中、存储在文件系统中，或者交付给指定的接收方。任何具有相应权限的用户可以使用 IBM SPSS Collaboration and Deployment Services Deployment Portal 界面来访问在存储库中存储的结果。作业本身可以根据已定义的调度来触发，或者为响应系统事件而触发。

此外，IBM SPSS 协作和部署服务 的评分服务允许与客户交互时，实时地交付已部署的模型所生成的分析结果。配置来进行评分的分析模型可以将当前客户交互收集到的数据与历史数据组合起来，从而产生一个决定交互过程的分值。该服务本身可以由任何客户机应用程序利用，以便创建定制接口来定义流程。

IBM SPSS 协作和部署服务 的部署工具专门适合方便地与您的企业基础设施进行集成。单点登录将减少手动提供不同过程阶段的凭证的需要。此外，该系统可配置为符合联邦信息处理标准出版物 140-2。

注：如果 SPSS Modeler 流包含使用列表类型的节点，那么包含此节点的分支将不支持评分服务。

## 系统体系结构

通常，IBM SPSS 协作和部署服务 由单个集中式 IBM SPSS Collaboration and Deployment Services Repository 组成，它使用执行服务器处理分析资产来服务于各种客户机。

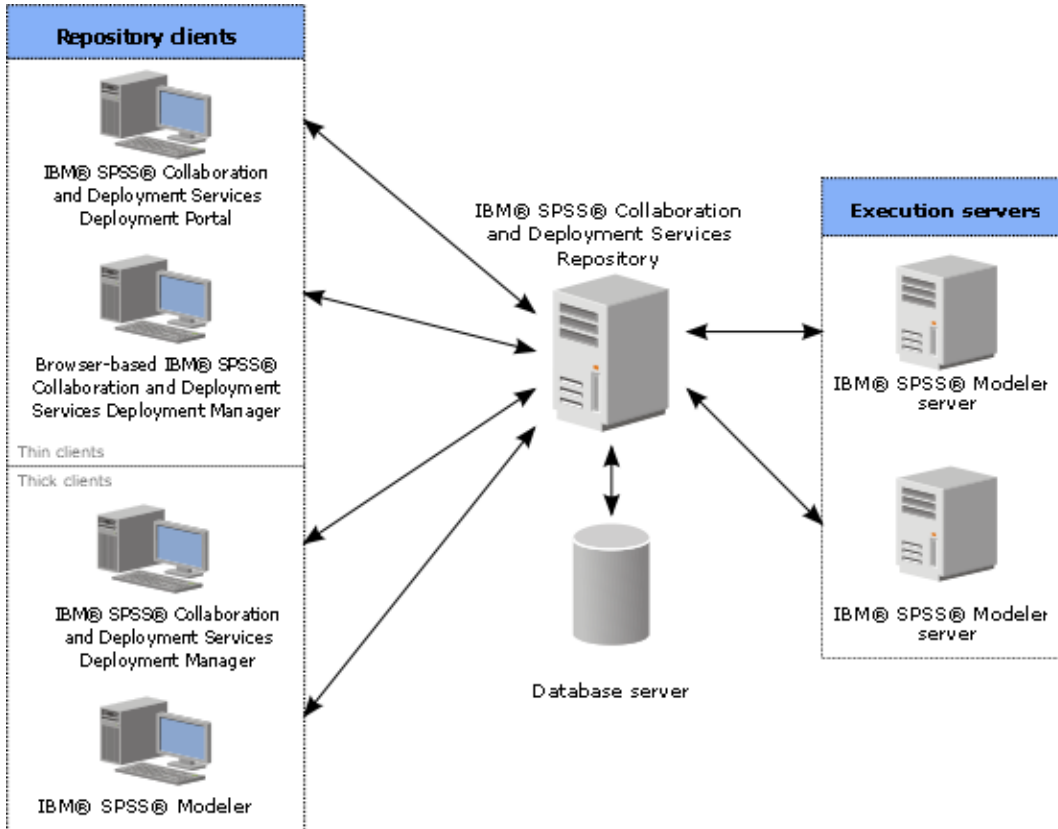


图 1: IBM SPSS 协作和部署服务体系结构

IBM SPSS 协作和部署服务 由下列组件组成：

- 分析构件的 IBM SPSS Collaboration and Deployment Services Repository
- IBM SPSS Modeler
- IBM SPSS Deployment Manager
- IBM SPSS Collaboration and Deployment Services Deployment Portal
- 基于浏览器的 IBM SPSS Deployment Manager

## IBM SPSS Collaboration and Deployment Services Repository

存储库提供一个用于存储分析资产（例如模型和数据）的集中位置。存储库需要安装关系数据库，例如 IBM Db2、Microsoft SQL Server 或 Oracle。

存储库包含以下方面的工具：

- 安全性
- 版本控制
- 搜索
- 审核

存储库的配置选项可使用 IBM SPSS Deployment Manager 或基于浏览器的 IBM SPSS Deployment Manager 来定义。存储库的内容通过 Deployment Manager 进行管理，并通过 IBM SPSS Collaboration and Deployment Services Deployment Portal 进行访问。



## IBM SPSS Modeler with collaboration

IBM SPSS Modeler 允许从 IBM SPSS Modeler 界面与 IBM SPSS Collaboration and Deployment Services Repository 进行交互。可以直接从 IBM SPSS Modeler 中存储和检索文件。

此外，存储在存储库中的 IBM SPSS Modeler 流可以在作业中作为步骤执行。作业可以包含任意数量的步骤，每个步骤对应于单独的文件。步骤之间定义的关系确定了过程流。可调度作业以在特定时间、根据重复模式或这响应定义的事件执行。此外，可以向指定的收件人发送通知以报告单个步骤和整体作业执行状态。

IBM SPSS 协作和部署服务 和 IBM SPSS Modeler 之间的协作通过使用适配器来启用。这些适配器安装到 IBM SPSS 协作和部署服务 环境以添加特定于产品的功能部件。有关更多信息，请参阅 IBM SPSS Modeler 文档。

**注：**如果有的作业中包含使用 Analytic Server 源节点的 SPSS Modeler 流，那么必须允许在 IBM SPSS 协作和部署服务 Server 与 Analytic Server 之间直接连接。否则，如果防火墙阻止这两个服务器之间的连接，作业将失败。

## IBM SPSS Deployment Manager

IBM SPSS Deployment Manager 是 IBM SPSS Collaboration and Deployment Services Repository 的客户端应用程序，它使用户能够安排、自动化以及执行分析任务，例如更新模型或生成分数。

通过该客户端应用程序，用户可以执行以下任务：

- 查看系统中的任何现有文件，包括 IBM SPSS Modeler 流、方案以及数据文件
- 将文件导入存储库
- 将作业调度为使用指定的重复模式（例如每季度或每小时）重复执行
- 修改现有的作业属性
- 确定作业的状态
- 指定作业状态的电子邮件通知

另外，该客户端应用程序还允许用户执行 IBM SPSS 协作和部署服务 管理任务，其中包括：

- 管理用户
- 配置安全提供程序
- 指定角色和操作

### 基于浏览器的 IBM SPSS Deployment Manager

基于浏览器的 IBM SPSS Deployment Manager 为精简客户端界面，用于执行设置与系统管理任务，包括：

- 设置系统配置选项
- 配置安全提供程序
- 管理 MIME 类型

如果非管理用户具有与登录凭证关联的相应操作，那么他们可以执行所有这些任务。这些操作由管理员进行指定。

通常，您需要访问基于浏览器的 IBM SPSS Deployment Manager，它位于以下 URL：

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

**注：**必须用方括号将 IPv6 地址括起来，例如 [3ffe:2a00:100:7031::1]。

如果您的环境配置为使用定制上下文路径进行服务器连接，请将该路径包括在 URL 中。

```
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 是用于访问存储库的瘦客户机界面。与旨在供管理员使用的基于浏览器的 IBM SPSS Deployment Manager 不同，IBM SPSS Collaboration and Deployment Services Deployment Portal 是服务于各种用户的 Web 门户网站。

该 Web 门户网站包含以下功能：

- 按文件夹浏览存储库内容
- 打开发布的内容
- 运行作业
- 使用存储库中存储的模型生成分数
- 搜索存储库内容
- 查看内容属性
- 访问各个用户首选项，例如电子邮件地址和密码、一般选项、预订，以及输出文件格式选项

通过在以下 URL 处访问主页：

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

注：必须用方括号将 IPv6 地址括起来，例如 [3ffe:2a00:100:7031::1]。

如果您的环境配置为使用定制上下文路径进行服务器连接，请将该路径包括在 URL 中。

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

## 基于浏览器的 IBM SPSS Deployment Manager

基于浏览器的 IBM SPSS Deployment Manager 为精简客户端界面，用于执行设置与系统管理任务，包括：

- 配置系统。
- 配置安全提供程序。
- 管理 MIME 类型。

如果非管理用户具有与登录凭证关联的相应操作，那么他们可以执行所有这些任务。这些操作由管理员分配。

## 执行服务器

通过执行服务器可执行存储器中存储的资源。当资源包括在作业中以待执行时，作业步骤定义就会包括用于处理该步骤的执行服务器的规范。执行服务器类型取决于资源。

IBM SPSS Modeler 执行服务器为 IBM SPSS Modeler Server，它允许执行有关数据挖掘与模型构建的分布式分析。该执行服务器要求指定进行处理所依据的用户凭证。

要允许负载均衡，可以将两个或更多个执行服务器聚集在一个服务器集群中。当作业步骤使用集群来执行时，IBM SPSS 协作和部署服务 会确定该集群中的哪个受管服务器当时最适合对处理请求进行处理。有关更多信息，请参阅 IBM SPSS Deployment Manager 文档。

在作业创建期间，为作业包含的每个步骤指定执行服务器或服务器集群。当作业执行时，存储库将使用指定的执行服务器来执行相应的分析。

## 使用 IBM SPSS Modeler 文件

在 IBM SPSS Deployment Manager 中，使用 IBM SPSS Modeler 文件的一般过程涉及：

1. 定义执行服务器以处理 IBM SPSS Modeler 流或方案（如果相应的定义尚不存在）。
2. 将 IBM SPSS Modeler 流或方案作为作业步骤添加到作业。
3. 指定作业属性，包括任何执行调度。

---

## 第 2 章 服务器定义

要将 IBM SPSS Collaboration and Deployment Services Repository 资源 作为作业步骤加以执行，需要指定相应的服务器来处理该作业步骤中包含的指令。此类服务器的连接信息在服务器定义中指定。

服务器定义可以分为执行服务器或存储库服务器。

- 执行服务器处理 IBM SPSS Collaboration and Deployment Services Repository 资源的内容。执行服务器类型必须与要处理的资源类型相对应。例如，IBM SPSS Modeler 作业步骤需要 IBM SPSS Modeler 服务器定义。
- 存储库服务器对应于 IBM SPSS 协作和部署服务 存储库安装。该类型的服务器通常由需要返回结果误差给存储库的作业步骤使用。

服务器定义包含在内容资源管理器的 *Resource Definitions* 文件夹中。具体来说，它们在 *Servers* 子文件夹中定义。

---

### 添加新服务器定义

要添加新服务器，请执行下列操作：

1. 在“内容资源管理器”中，打开资源定义文件夹。
2. 单击服务器文件夹。
3. 在“文件”菜单中，选择：

**新建 > 服务器定义**

这将打开“添加新服务器定义”向导。此外，也可通过单击某些步骤的“常规”选项卡上的服务器字段旁边的**新建**来访问“新建服务器定义”对话框。定义新服务器的过程包含下列步骤：

1. 命名服务器定义并指定其类型。请注意，可用的服务器类型取决于安装到存储库中的产品适配器。
2. 在 *Servers* 文件夹中选择定义的位置。
3. 指定用于定义连接或执行信息的服务器参数。参数集取决于服务器类型。

### IBM SPSS Modeler 服务器参数

IBM SPSS Modeler 服务器定义指定用于处理作业步骤的 IBM SPSS Modeler 服务器的连接参数。IBM SPSS Modeler 流在 IBM SPSS Modeler 服务器上执行。

1. 在“主机”字段中，输入服务器所驻留主机的名称。例如，如果创建一个 IBM SPSS Modeler 服务器定义，则主机应当是包含您的 IBM SPSS Modeler 服务器的计算机。
2. 在“端口”字段中，输入用来连接该主机的端口号。
3. 在“缺省数据路径”字段中，输入要用来放置数据文件的路径。
4. 如果服务器连接使用安全套接字层 (SSL)，请选择**这是一个安全端口**。
5. 单击**完成**。新定义将显示在服务器文件夹中。

**注：**在“评估”流中运行 IBM SPSS 协作和部署服务 作业时，将创建临时文件。缺省情况下，此文件保存在 IBM SPSS Modeler 服务器安装目录中。当您在 IBM SPSS Modeler 中创建 IBM SPSS Modeler 服务器连接时，可以更改用于保存临时文件的缺省数据文件夹。

---

### 修改服务器定义

要修改服务器定义，请执行下列操作：

1. 在“内容资源管理器”中，打开资源定义文件夹。
2. 打开 *Servers* 文件夹。
3. 双击要修改的服务器。这将打开“编辑服务器定义”对话框。

4. 根据需要修改服务器定义参数。
5. 单击“完成”以保存更改。

---

## 第 3 章 IBM SPSS Modeler 作业步骤

### 处理 IBM SPSS Modeler 流

---

本部分在 IBM SPSS Deployment Manager 的上下文内描述 IBM SPSS Modeler 流。IBM SPSS Modeler 流以完整格式引入到 Deployment Manager 中。有关如何创建与使用流的更多详细信息，请参阅 IBM SPSS Modeler 文档说明。

您可以在 Deployment Manager 中使用 IBM SPSS Modeler 流。与任何其他步骤一样，必须先将 IBM SPSS Modeler 流添加到作业，然后才能使用 Deployment Manager 执行该流。特别地，可以执行以下任务：

- 导入流
- 修改流参数
- 执行流

**注：**如果有的作业中包含使用 Analytic Server 源节点的 SPSS Modeler 流，那么必须允许在 IBM SPSS 协作和部署服务 Server 与 Analytic Server 之间直接连接。否则，如果防火墙阻止这两个服务器之间的连接，作业将失败。

### IBM SPSS Modeler Server 配置

在 Deployment Manager 中开始使用 IBM SPSS Modeler 流之前，需要执行以下配置任务：

- 创建 IBM SPSS Modeler 服务器定义。有关更多信息，请参阅主题 [第 5 页的『IBM SPSS Modeler 服务器参数』](#)。
- 定义服务器凭证。

### 查看 IBM SPSS Modeler 作业属性

---

单击作业内的 IBM SPSS Modeler 流时，将显示以下作业属性：

- 常规
- 数据文件
- ODBC 数据源
- 参数
- 结果
- Cognos 导入
- Cognos 导出
- 通知

### IBM SPSS Modeler 作业属性 - 常规

缺省情况下，在单击作业工作区中的 IBM SPSS Modeler 流时，会显示“常规”属性选项卡。

“常规属性”选项卡包含以下信息：

**作业步骤名称。** 作业步骤的名称。通常，作业步骤的名称是 IBM SPSS Modeler 流的名称，附加后缀 `_step`。但是，可以修改作业步骤名称。

要修改作业步骤名称，在**作业步骤名称**字段中键入修改后的名称。更改会反映在作业工作区中。

**IBM SPSS Modeler 流。** 原始 IBM SPSS Modeler 流的名称与路径。源流不可修改。

**覆盖类型** 使用此选项可以指定某个作业步骤还是某个流中节点内的设置在执行期间优先执行。

- **作业覆盖流** 如果您选择此选项，那么执行作业时，节点属性将使用您在作业步骤中设置的值。

- **流覆盖作业** 选择此选项将禁用作业步骤中的节点属性控件；唯一的例外是某些凭证控件。如果更改了流中的节点属性，那么执行作业步骤时，将使用更改后的节点属性。

**覆盖类型**选项将通过下列方式影响凭证设置：

- 如果您选择**作业覆盖流**，那么可以编辑凭证选择控件，并且所选凭证将在运行时生效。
- 如果您选择**流覆盖作业**，那么有两种可能的情况：
  - 在 SPSS Modeler Client 中创建使用凭证的节点时，如果认证方式为**存储凭证**，并且 IBM SPSS 协作和部署服务中存在凭证名称，那么凭证选择控件不可用。但是，如果 IBM SPSS 协作和部署服务中不存在凭证名称，请使用凭证选择控件；所选凭证将在运行时生效。
  - 在 SPSS Modeler Client 中创建使用凭证的节点时，如果认证方式为**用户名和密码**，那么您可以使用凭证选择控件；所选凭证将在运行时生效。

**对象版本**。要使用的 IBM SPSS Modeler 流的标注版本。从**对象版本**下拉列表中，选择您要使用的标注版本。

**迭代变量列表**。如果此步骤作为迭代使用者，则标识来自迭代生成者的变量，以便为步骤提供值。迭代生产者步骤必须为作业中作为迭代使用者的步骤之前的那个步骤。

**IBM SPSS Modeler 服务器**。执行流的 IBM SPSS Modeler 服务器或服务器集群。包含目前配置用于执行 IBM SPSS Modeler 步骤的所有服务器和服务器集群的列表。要更改服务器，请从**IBM SPSS Modeler Server**下拉列表中进行选择。要创建新的服务器定义，单击**新建**以启动服务器定义向导。

**IBM SPSS Modeler 登录信息**。用于访问 IBM SPSS Modeler 服务器或服务器集群的凭证信息。要更改凭证，请从**IBM SPSS Modeler 登录信息**下拉列表中选择凭证定义。要定义新凭证，单击**新建**以启动凭证定义向导。

**内容存储库服务器**。内容存储库服务器允许作业将文件保存到 IBM SPSS Collaboration and Deployment Services Repository。通常在使用 IBM SPSS Modeler 刷新模型时指定内容存储库服务器。要指定内容存储库服务器，从**内容存储库服务器**下拉列表中选择服务器。要创建新的服务器定义，单击**新建**以启动服务器定义向导。要生成基于当前服务器信息的内容存储库服务器，单击**生成**。这将创建服务器定义，并在内容存储库服务器字段中自动填写。

**内容存储库登录**。内容存储库服务器的登录信息。要指定内容存储库登录信息，从**内容存储库登录**下拉列表中选择凭证。要创建新的登录，单击**新建**以启动内容存储库登录向导。如果不使用单点登录连接到 IBM SPSS Collaboration and Deployment Services Repository，单击**生成**以生成基于现有安全设置的内容存储库服务器登录。这将创建内容存储库登录，并在内容存储库登录字段中自动填写。当使用单点登录时，登录生成功能不可用。

**警告表达式**。定义通过条件连接器连接的作业步骤的警告。警告表达式（例如，`completion_code`、**警告**或**成功**）必须为小写字母。

要使用警告表达式，请执行以下操作：

1. 使用条件连接器连接两个作业步骤。在条件连接器的**表达式**字段中，键入 `warning==true`。
2. 浏览至父作业步骤的“常规”选项卡。
3. 在**警告表达式**字段中，指定警告代码，例如 `completion_code==18`。此表达式会覆盖缺省警告代码（如果有）。

运行作业时，系统将执行父作业步骤。然后，系统将评估 `warning==true` 条件。如果为 `true`，那么系统将查看所指定的警告表达式并确定是否满足条件。如果满足警告表达式中指定的条件，那么系统将执行下一个作业步骤。

**类型**。运行流的方式。

- **运行流**。如果使用此选项，您可以运行完整的流，或者如果流包含分支，可选择运行一个或多个分支。在后一种情况中，会显示**运行选项**列表。缺省情况下会选中所有分支，这意味着运行流时将执行所有分支。取消选中所有不想执行的分支，这样做可以提高性能。

**注：**如果该流包含脚本且配置为执行时“运行脚本”，那么不会有任何执行分支显示在 IBM SPSS 协作和部署服务 作业编辑器中的列表中。

- **模型管理**。如果您希望流使用模型管理功能（如评估、刷新和评分），则选择此选项。从显示的“类型列表”中选择功能。**评估选项**

**性能。** 指定以下类别的阈值百分比：“差”或“优”。“更好”百分比不是用户定义的。系统将 100 减去“好”百分比即得到“更好”阈值。阈值对评估流来说是强制的。

**度量。** 评估流有效性的衡量标准。有效值包括“授信”、“准确性”和“收益”。如果选择了**收益**，则需要指定相应的百分数。如果选择了**授信**，则需要指定授信步骤。有效值包括“收集统计信息”和“运行评估”。

#### 刷新选项

列出在流中可以使用在流中定义的节点名称刷新的所有模型块关系。选中关系前面的复选框，以指定要刷新的特定关系。取消选中某关系可在执行步骤时防止关系被刷新。要选择所有要刷新的关系，请单击“选中所有”按钮。要取消选中所有关系，请单击“取消选中所有”按钮。

#### 分数选项

选择要指定为评分分支的分支。

### 提供 IBM SPSS Modeler 和内容存储库信息的准则

对于所有 IBM SPSS Modeler 作业步骤，需要 IBM SPSS Modeler 服务器与 IBM SPSS Modeler 登录信息。

## IBM SPSS Modeler 作业属性 - 数据文件

“数据文件”表包含以下信息：

**节点名。** 包含流所使用数据的输入节点名称。名称的前缀为包含了以斜杠分隔的节点的任何超节点的名称。例如，如果节点 *MyNode* 位于名为 *Supernode1* 的超节点中，则名称显示为 */Supernode1/MyNode*。节点名称不能修改。

**节点类型。** 在流中定义的节点类型。节点类型不能修改。

**文件名。** 输入数据文件的名称。要更改名称，请单击“文件名”单元格并更改名称。

**格式。** 输出文件的格式，例如以逗号分隔的文件。要修改文件格式类型，请单击“格式”单元格。将显示下拉箭头。选择格式类型。

**位置。** 输入数据文件的位置。要修改位置，单击列，然后单击显示的省略号按钮。此时将打开“输入文件位置”对话框。必要时更改位置。

无法访问已锁定超节点中的节点。也无法查看或修改这些节点。

## IBM SPSS Modeler 作业属性 - 数据视图

如果流包含“地理空间”节点，将在此处显示连接详细信息。

**节点名。** “数据视图”节点的名称。

**分析数据视图。** 使用的分析数据视图。

**标签。** 使用的标签。

**表名称。** 使用的数据库表的名称。

**数据访问方案。** 请从分析数据视图中选择一个数据访问方案。数据访问方案会将分析数据视图中的数据模型表与物理数据源进行关联。分析数据视图通常包含多个数据访问方案。更改使用中的数据访问方案时，您可以更改流所使用的数据。例如，如果分析数据视图包含用于训练模型的数据访问方案以及用于检验模型的方案，那么您可以通过更改使用中的数据访问方案从训练数据切换至检验数据。

## IBM SPSS Modeler 作业属性 - ODBC 数据源

ODBC 数据源表包含以下信息。

**节点名。** 包含流所使用数据的输入节点名称。名称的前缀为包含了以斜杠分隔的节点的任何超节点的名称。例如，如果节点 *MyNode* 位于名为 *Supernode1* 的超节点中，则名称显示为 */Supernode1/MyNode*。

**节点类型。** 在流中定义的节点类型。

**ODBC 数据源。** 当前 ODBC 数据源名称 (DSN)。要切换至其他 ODBC 数据源，请单击包含当前数据源名称的单元格，然后单击所显示的“...”按钮。执行上述步骤后，将会显示可在其中选择现有 DSN 或创建新 DSN

的对话框。请注意，作业步骤设置始终会覆盖流；因此，如果您将流修改为使用其他数据源，那么还必须编辑作业以使用同一源，否则该作业可能无法运行。

**凭证。** 要在更改 ODBC 数据源时更改数据库用户名和密码，请单击包含当前凭证的单元格，然后单击 "..." 按钮。这将显示一个对话框，可以在该对话框中选择现有凭证定义或创建新的凭证定义。

**数据库表。** 对应于节点的数据库表。

无法访问已锁定超节点中的节点。也无法查看或修改这些节点。

## 更改 ODBC 连接

可以将 ODBC 连接更改为已经定义的连接，或者也可以在此创建连接。

**使用现有 ODBC DSN。** 单击**浏览**以显示可供选择的现有 ODBC DSN 列表。

**创建新的 ODBC DSN。** 选择此选项，然后单击**新建**以定义新的 ODBC 连接来用于流作业步骤。

## 浏览 ODBC 连接

这是已经在此主机上定义的并具有访问权限的所有 ODBC 连接的列表。

从列表中选择一项，并单击**确定**。

## 更改数据库凭证

更改连接时，您可以切换为在此处使用相应的数据库用户名和密码。可以使用存储库中现有的凭证定义，或者可以创建一个新的定义。

**从现有用户名和密码生成存储库凭证。**（仅当导入来自 IBM SPSS 协作和部署服务 版本 4.2.1 的作业时显示）如果在导入的作业中设置了用户名和密码，则单击**生成**按钮可从现有用户名和密码生成存储库凭证。例如，如果用户名 sa 已存在，那么重复单击**生成**将创建用户名 sa1、sa2，依此类推。

**使用现有凭证。** 单击**浏览**按钮可显示现有凭证定义列表以便从中选择。

**创建新凭证。** 选择此选项，然后单击**新建**来创建新凭证定义，以用于当前连接。

## 浏览凭证定义

这是已经在此主机上创建的并具有访问权限的所有凭证定义的列表。

从列表中选择一项，并单击**确定**。

## IBM SPSS Modeler 作业属性 - 地理空间

如果流包含“地理空间”节点，将在此处显示连接详细信息。

**源类型。** 数据源类型。

**文件名。** 如果使用形状文件，请指定文件名。

**地图服务 URL。** 如果使用地图服务，请指定该服务的 URL。

**地图标识。** 指定地图标识。

## IBM SPSS Modeler 作业属性 - 参数

您可以使用 IBM SPSS Deployment Manager 来修改 IBM SPSS Modeler 流参数。“参数”表包含以下信息：

**名称。** 参数名。

**存储器。** 描述如何存储参数 - 例如，作为字符串。

**值。** 参数的值。

**类型。** 参数类型。

无法访问已锁定超节点中的参数。也无法查看或修改这些节点。



## IBM SPSS Modeler 作业属性 - 结果

要显示 IBM SPSS Modeler 作业步骤的输出设置，单击“结果”选项卡。

**节点名称。** 包含流处理输出的节点的名称。名称的前缀为包含了以斜杠分隔的节点的任何超节点的名称。例如，如果节点 *MyNode* 位于名为 *Supernode1* 的超节点中，则名称显示为 */Supernode1/MyNode*。节点名称不能修改。

**节点类型。** 流中定义的节点类型。节点类型不能修改。

**文件名。** 对应文件的名称。要修改名称，请在“文件名”列中单击，然后输入新的名称。

注：文件扩展名由所选的文件格式决定，将自动附加在文件名之后。

**格式。** 输出文件的格式。要修改文件格式类型，单击“格式”列。将显示下拉箭头。选择格式类型。

**位置。** 文件位置。要修改位置，单击列打开“结果位置”对话框，然后单击显示的省略号按钮。

**许可权。** 保存到存储库时的文件访问权限。要修改权限，单击“权限”列打开“输出权限”对话框，然后单击显示的省略号按钮。

**属性。** 文件的属性（元数据）。要定义属性，单击“属性”列打开“输出属性”对话框，然后单击显示的省略号按钮。

无法访问已锁定超节点中的节点。也无法查看或修改这些节点。

## 查看输出结果

要查看流的输出结果，请在“作业步骤历史记录”表的“结果”列中单击要查看的结果。会在一个单独的窗口中打开结果。

例如，假设某个输出结果是直方图。那么样本直方图如图所示。

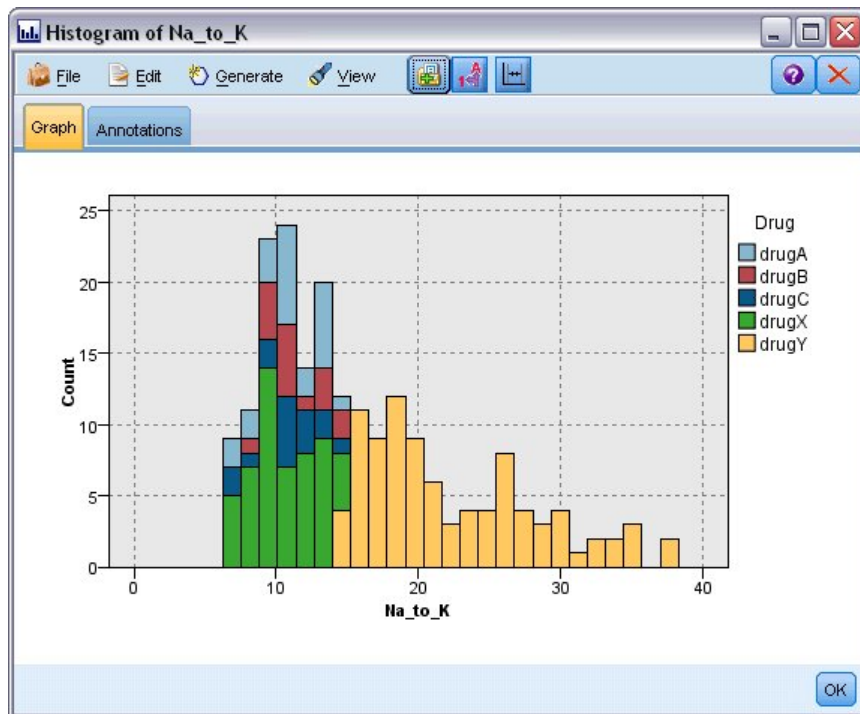


图 2: 样本直方图

## IBM SPSS Modeler 作业属性 - Cognos 导入

如果流包含 IBM Cognos 源节点，那么将在此显示 Cognos 连接详细信息。

**节点名。** Cognos 源节点的名称。

**连接 URL。** 进行连接的 Cognos 服务器的 URL。

**包名称。** 导入元数据的 Cognos 数据包名称。

**匿名。** 如果 Cognos 服务器连接使用了匿名登录，包含**匿名**，或者在使用了特定的 Cognos 用户名和密码时，包含**凭证**。

**凭证。** Cognos 服务器上的用户名和密码（如果需要）。

**注：**必须在代表 Cognos 名称空间标识的域中创建 Cognos 凭证。

## IBM SPSS Modeler 作业属性 - Cognos 导出

如果流包含 IBM Cognos 导出节点，那么将在此显示 Cognos 连接详细信息。

**节点名。** Cognos 导出节点的名称。

**连接 URL。** 进行连接的 Cognos 服务器的 URL。

**包名称。** 用于导出元数据的 Cognos 数据包名称。

**数据源。** 用于导出数据的 Cognos 数据库名称。

**文件夹。** 要在其上创建导出数据包的 Cognos 服务器的路径和文件夹名称。

**匿名。** 如果 Cognos 服务器连接使用了匿名登录，包含**匿名**，或者在使用了特定的 Cognos 用户名和密码时，包含**凭证**。

**凭证。** Cognos 服务器上的用户名和密码（如果需要）。

**DSN。** Cognos 数据库的数据库源名称 (DSN)。

**表名称。** 用于导出的数据库表名称。

**凭证。** 连接到用于导出的数据库的用户名和密码。

**注：**必须在代表 Cognos 名称空间标识的域中创建 Cognos 凭证。

## IBM SPSS Modeler 作业属性 - 遗存 TM1 导入

如果流包含“遗存 TM1 导入”节点，将在此处显示连接详细信息。

**节点名。** “遗存 TM1 导入”节点的名称。

**TM1 服务器。** Cognos TM1 服务器名称。

**多维数据集。** 要从中导入数据的 TM1 多维数据集。

**视图。** 要从 TM1 多维数据集导入的视图。

**凭证。** Cognos TM1 服务器的凭证。

## IBM SPSS Modeler 作业属性 - 遗存 TM1 导出

如果流包含“遗存 TM1 导出”节点，将在此处显示连接详细信息。

**节点名。** “遗存 TM1 导出”节点的名称。

**TM1 服务器。** Cognos TM1 服务器名称。

**多维数据集。** 要将数据导出到的 TM1 多维数据集。

**凭证。** Cognos TM1 服务器的凭证。

## IBM SPSS Modeler 作业属性 - TM1 导入

如果流包含“TM1 导入”节点，将在此处显示连接详细信息。

**节点名。** “TM1 导入”节点的名称。

**管理主机。** Cognos TM1 管理主机。

**TM1 服务器。** Cognos TM1 服务器名称。

**多维数据集。** 要从中导入数据的 TM1 多维数据集。

**视图。** 要从 TM1 多维数据集导入的视图。可以导入公共或私有视图。

**行。** 要从 TM1 多维数据集导入的行。

**列。** 要从 TM1 多维数据集导入的列。

**凭证。** Cognos TM1 服务器的凭证。

## IBM SPSS Modeler 作业属性 - TM1 导出

如果流包含“TM1 导出”节点，将在此处显示连接详细信息。

**节点名。** “TM1 导出”节点的名称。

**管理主机。** Cognos TM1 管理主机。

**TM1 服务器。** Cognos TM1 服务器名称。

**多维数据集。** 要将数据导出到的 TM1 多维数据集。

**测量。** 将导出的测量。

**凭证。** Cognos TM1 服务器的凭证。

## IBM SPSS Modeler 作业属性 - Analytic Server 导入

如果流包含 Analytic Server 导入节点，那么此处将显示连接新选项。

**使用缺省 Analytic Server。** **True** 或 **False** 指定将使用管理员在 `options.cfg` 中定义的缺省 Analytic Server 连接 (**True**) 还是使用另一个 Analytic Server，即使用在流/作业中指定的服务器 (**False**)。

**URL。** 这是格式为 `https://hostname:port/contextroot` 的 Analytic Server 的 URL，其中 `hostname` 是 Analytic Server 的 IP 地址或主机名，`port` 是其端口号，而 `contextroot` 是 Analytic Server 的上下文根。

**租户。** 这是 SPSS Modeler Server 所属的租户的名称。

**凭证。** 这是用于登录 Analytic Server 的凭证。

**服务主体名称。** 这是 Kerberos 服务主体名称。

**配置文件路径。** 这是 Kerberos 服务配置文件路径。

**数据源。** 这是 Analytic Server 上的数据源名称。

## IBM SPSS Modeler 作业属性 - Analytic Server 导出

如果流包含 Analytic Server 导出节点，那么此处将显示连接新选项。

**使用缺省 Analytic Server。** **True** 或 **False** 指定将使用管理员在 `options.cfg` 中定义的缺省 Analytic Server 连接 (**True**) 还是使用另一个 Analytic Server，即使用在流/作业中指定的服务器 (**False**)。

**URL。** 这是格式为 `https://hostname:port/contextroot` 的 Analytic Server 的 URL，其中 `hostname` 是 Analytic Server 的 IP 地址或主机名，`port` 是其端口号，而 `contextroot` 是 Analytic Server 的上下文根。

**租户。** 这是 SPSS Modeler Server 所属的租户的名称。

**凭证。** 这是用于登录 Analytic Server 的凭证。

**服务主体名称。** 这是 Kerberos 服务主体名称。

**配置文件路径。** 这是 Kerberos 服务配置文件路径。

**数据源。** 这是 Analytic Server 上的数据源名称。

## IBM SPSS Modeler 作业属性 - 通知

使您能够指定关于作业步骤已失败和已成功的电子邮件通知。

单击每个观测值的**更新**按钮以添加或删除通知接收方。

## 在 IBM SPSS Modeler 中查看流

您可以直接从 IBM SPSS Deployment Manager 中启动 IBM SPSS Modeler 应用程序。

要在 IBM SPSS Modeler 应用程序中查看流，请在内容资源管理器中双击该流。系统将启动 IBM SPSS Modeler 应用程序，并在 IBM SPSS Modeler 中显示流。

请务必注意，如果对作业中的文件进行了任何更改，例如 IBM SPSS Modeler 流 (.str)，那么包含此文件的任何作业都会受影响。在对文件做出更改后，文件的新版本将保存到存储库。但是，包含此文件的作业不会随修改后的文件自动更新。要合并文件更新到受影响的作业，请执行下列操作：

1. 重新打开作业。重新打开作业后，在作业工作区中的作业名称旁边将出现星号，这表示此作业包含未保存的更改。
2. 重新保存作业。

## IBM SPSS Modeler 完成代码

IBM SPSS Modeler 作业的完成代码在下表中进行了描述。请为任何涉及 IBM SPSS Modeler 流的条件关系使用这些完成代码。

代码	描述
0	成功
1	流执行错误
2	发布错误
8	未知错误

## IBM SPSS Modeler 流限制

当您在 IBM SPSS Deployment Manager 中使用流时，系统存在以下限制：

- **命名。** 如果节点名称、标签以及类型均相同，则无法计划作业执行，因为会引起冲突。
- **脚本编制。** 如果流脚本包含对特定节点的覆盖（例如，脚本设置了图的输出位置），那么此脚本将取代在 Deployment Manager 中定义的、存在冲突的所有用户指定值。
- **超节点。** IBM SPSS Modeler 作业步骤的执行对应于流中所有顶层终端节点的处理与执行。如果某个终端节点是终端超节点，那么将重复执行该超节点中的终端节点。然后，如果源或过程超节点（定义为非终端节点）包含终端节点，那么将不执行这些节点。非终端超节点中的终端节点会在 Deployment Manager 中出现，但不会在步骤执行期间进行处理。
- **参数。** 为超节点定义参数不能与为流定义的参数同名。名称必须是唯一的。
- **“文本分析”节点限制。** 您不能在 IBM SPSS 协作和部署服务-评分 配置内将 SPSS Modeler 文本分析“文件列表”或“Web 订阅源”节点用于评分。

除了以上所列各项，还有以下具体约束。

- 在 IBM SPSS 协作和部署服务 存储库中存储流时，没有可用的解锁复选框。解锁是存储流时的缺省值。要锁定或解锁对象，选择工具 > 存储库 > 探索，导航到对象，右键单击其名称以显示上下文菜单。
- 在 Deployment Manager 中，当您运行某个作业来产生“增益”图形，且该作业包含设置了“评估”节点的流时，如果系统正使用 Oracle JRockit JRE 在 Oracle Weblogic 11g 下运行，该图形输出可能不完整。为避免此问题，请使用 IBM JRE。

## 节点类型

当从 IBM SPSS Deployment Manager 中打开 IBM SPSS Modeler 流时，您可以看到流节点由不同形状的图标来表示。圆圈图标表示源数据节点，而六角形节点表示数据记录和字段上的处理操作。三角形表示图形输出，五边形表示建模节点。输出（除了图形外）和导出操作由矩形节点表示。

各种节点在 *IBM SPSS Modeler* 源、过程、输出节点和 *IBM SPSS Modeler* 建模节点指南中进行了详细描述。

## 脚本示例：比较优胜者与参选者模型

以下这段 IBM SPSS Modeler 脚本摘录展示了优胜者/参选者比较。优胜者模型将与参选者模型进行比较。如果参选者模型优于优胜者模型，那么优胜者模型将被覆盖。这样，参选者模型将变为新的优胜者模型。

```
# -----
# 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':tablnode
var missRateChamp
set missRateChamp = value 'ModelEval':tablnode.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':tablnode
var missRateChallenge
set missRateChallenge = value 'ModelEval':tablnode.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 pmm1

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



---

## 第 4 章 评分服务

评分服务允许客户端应用程序采用源自 IBM SPSS Modeler 中开发的预测模型的实时评分。服务将访存指定模型，加载此模型，调用正确的评分实现并将结果返回至客户机。

评分是通过向预测模型提供输入数据生成实时值的过程。评分模型是可用于根据给定输入数据生成输出值的任何工件。通常，要使用用于生成评分的模型，请执行以下操作：

1. 从 IBM SPSS Collaboration and Deployment Services Repository 中选择要用于评分的模型。
2. 定义此模型的评分配置。
3. 向已配置模型提供数据并生成评分。

有关评分和评分服务的更多信息，请参阅 IBM SPSS 协作和部署服务文档。

---

### IBM SPSS Modeler 流限制

使用 IBM SPSS Deployment Manager 中的流时，系统对评分服务施加了以下约束：

- **超节点。** 不支持超节点内的源节点，并且不支持超节点内的终端节点。
- **地理空间节点。** 不支持地理空间节点。
- **模型构建器节点。** 不支持作为终端节点的模型构建器节点。
- **数据库内挖掘。** 不支持包含数据库内挖掘块的流。
- **UDF、UDA 和 WUDA。** 不支持使用数据库函数 (UDF)、数据库汇总 (UDA) 或数据库窗口汇总 (WUDA) 的流。
- **源节点。** 不支持带有包含列表类型的输出数据模型的源节点。
- **终端节点。** 不支持带有包含列表类型的输入数据模型的终端节点。
- **Text Analytics 节点限制。** 不能使用 SPSS Modeler 文本分析文件列表或 Web 订阅源节点在 IBM SPSS 协作和部署服务-评分配置中评分。
- **Analytic Server 源节点。** 如果有的作业中包含使用 Analytic Server 源节点的 SPSS Modeler 流，那么必须允许在 IBM SPSS 协作和部署服务 Server 与 Analytic Server 之间直接连接。否则，如果防火墙阻止这两个服务器之间的连接，作业将失败。





# 第 5 章 优胜者参选者作业步骤

## 优胜者参选者概述

使用 IBM SPSS Deployment Manager 可比较 IBM SPSS Modeler 生成的模型文件以确定哪个文件包含最有效的预测模型。优胜者参选者作业步骤会评估模型并将其与一个或多个参选者比较。

在系统比较结果之后，最佳模型将成为新的优胜者。

**优胜者。** 优胜者对应于最有效的模型。在优胜者参选者作业步骤的最初执行中，没有优胜者 - 只有第一个参选者和相对应的其他参选者的列表。在随后的作业步骤执行中，系统将确定优胜者。

**参选者。** 将对各个参选者进行相互比较。然后，生成最佳结果的参选者将成为新的优胜者。

### 优胜者选择过程

优胜者参选者比较过程由以下任务组成：

1. 对每个参选模型进行评分。
2. 评估由此得出的评分。
3. 比较评估结果并确定哪个参选者是优胜者。
4. 保存新的优胜者到存储库（可选）。

### 添加“优胜者参选者工作”到作业

要向作业添加优胜者参选者作业步骤，请从作业选用板中选择“优胜者参选者”工具，然后单击作业工作区。

## 模型评估度量

模型评估和比较可侧重于准确性、收益或授信。

- **准确性。** 模型的准确性反映正确预测的目标响应的百分比。相对于正确预测变量百分比较低的模型，优先考虑使用百分比较高的模型。
- **收益。** 收益统计信息是模型的性能指标。此测量方式将使用模型的结果与不使用模型的结果相比较。使用模型时在结果中取得的改进称为收益。将两个模型相比较，首先考虑使用在指定的百分位数上收益值较高的模型。
- **授信。** 模型授信反映了模型的可信度。此方法对新数据与模型所基于的培训数据之间的相似性进行检查。授信值在 0 到 1 之间变化，值较大表示两个数据集中的预测变量之间的相似度较高。比较两个模型时，授信值较高的模型基于与新数据更相似的培训数据，此类模型更可信并值得优先选择。

## 顺序相关性

与其他作业步骤类型不同，优胜者参选者作业步骤中的选项卡存在顺序相关性。

例如，在“优胜者”选项卡上修改信息之前必须在“参选者”选项卡上选择参选者。此外，在某些选项卡上显示的信息取决于“参选者”表中所选的参选者。

执行优胜者参选者比较的过程包含以下步骤：

1. 提供常规作业信息。
2. 指示参选者。
3. 指定优胜者信息。
4. 查看参数信息。
5. 指定通知。

## 常规信息

“常规”选项卡包含与整个优胜者参选者作业步骤相关的信息。

需要 IBM SPSS Modeler 服务器和登录信息才能执行优胜者参选者作业步骤。执行作业并保存新优胜者到 IBM SPSS Collaboration and Deployment Services Repository 需要内容存储库服务器和登录信息。（需要内容存储库服务器和登录信息，即使您放弃分析结果。）凭证基于当前登录系统的用户。

**作业步骤名称。** 作业步骤的名称。缺省情况下，第一个作业步骤的名称为事件 1。后续作业步骤的名称为事件 2、事件 3 等。在执行完作业步骤后，这里指定的名称将出现在作业历史记录表中。

**IBM SPSS Modeler 服务器。** 执行流的 IBM SPSS Modeler 服务器或服务器集群。包含目前配置用于执行 IBM SPSS Modeler 步骤的所有服务器和服务器集群的列表。要更改服务器，请从 **IBM SPSS Modeler Server** 下拉列表中进行选择。要创建新的服务器定义，单击**新建**以启动服务器定义向导。

**IBM SPSS Modeler 登录信息。** 用于访问 IBM SPSS Modeler 服务器或服务器集群的凭证信息。要更改凭证，请从 **IBM SPSS Modeler 登录信息** 下拉列表中选择凭证定义。要定义新凭证，单击**新建**以启动凭证定义向导。

**内容存储库服务器。** 内容存储库服务器允许作业将文件保存到 IBM SPSS Collaboration and Deployment Services Repository。通常在使用 IBM SPSS Modeler 刷新模型时指定内容存储库服务器。要指定内容存储库服务器，从**内容存储库服务器**下拉列表中选择服务器。要创建新的服务器定义，单击**新建**以启动服务器定义向导。要生成基于当前服务器信息的内容存储库服务器，单击**生成**。这将创建服务器定义，并在内容存储库服务器字段中自动填写。

**内容存储库登录。** 内容存储库服务器的登录信息。要指定内容存储库登录信息，从**内容存储库登录**下拉列表中选择凭证。要创建新的登录，单击**新建**以启动内容存储库登录向导。如果不使用单点登录连接到 IBM SPSS Collaboration and Deployment Services Repository，单击**生成**以生成基于现有安全设置的内容存储库服务器登录。这将创建内容存储库登录，并在内容存储库登录字段中自动填写。当使用单点登录时，登录生成功能不可用。

## 参选者

必须选择至少一个参选者才能执行优胜者参选者作业步骤。很重要的是，所选的第一个参选者不意味着评估过程中的比较顺序或任何优先程度。第一个参选者只是底线。

用于确定后续参选者的数据源和标签由第一个参选者确立。在选择或更新第一个参选者后，将使用对应于第一个参选者的信息更新该选项卡中的剩余字段。

**第一个参选者。** 第一个参选者的名称。要浏览存储库，单击**浏览**。

**第一个参选者的标签。** 与包含第一个参选者的模型文件关联的标签。在选择第一个参选者时，请指定此值。

**数据源参选者。** 提供用于该作业步骤的数据源的参选者。请单击**浏览**以从“参选者”表中选择的条目列表中选择此参选者。

**度量。** 比较参选者的衡量标准。有效值包括授信、准确性和收益。如果选择了收益，则需要指定相应的百分数。有关更多信息，请参阅主题 [第 19 页的『模型评估度量』](#)。

### “参选者”表格

“参选者”表会列出与数据源匹配的参选者的缺省分数分支以及与第一个参选者关联的标签。当执行作业步骤时，只将从该表格中选择的参选者与第一个参选者进行比较。从列表中选择（或取消选择）参选者将使系统相应地更新其他作业步骤选项卡上的对应信息。

每次在打开包含之前保存的优胜者参选者步骤的作业时，将自动更新参选者列表。如果新的参选者与第一个参选者的数据源和标签条件匹配，那么可以向列表添加这些新参选者。相反，不再满足这些标准的参选者可能会从参选者列表中删除。如果从存储库删除了选定的参选者，系统将显示信息表示参选者不再可用。

虽然可选择和取消选择用于比较的参选者，但是“参选者”表中的信息无法修改。具体来说，“参选者”表格包含以下信息：

**名称。** 参选者的名称。

**标签。** 与参选者关联的标签。

**描述。** 参选者的描述。

## 修改第一个参选者

在保存作业后对第一个参选者所进行的更改可能影响优胜者参选者分析。例如，假定从存储库中删除了第一个参选者，或删除了与第一个参选者关联的标签。当访问“参选者”选项卡时，系统将显示信息表示第一个参选者不再可用。在这种情况下，需指定新的第一个参选者。

## 选择参选者

要选择将包含在优胜者参选者分析中的参选者模型，请执行以下步骤：

1. 在优胜者参选者步骤中的“参选者”选项卡中，单击**浏览**以选择第一个参选者。如果您要手动添加参选者，请单击**添加**以打开“参选者”表。
2. 单击**浏览**以选择模型文件。模型文件是 IBM SPSS Modeler 流，其中包含具有有效模型块的缺省分数分支。
3. 选择标签以指示要使用的选定模型文件版本。
4. 在“参选者”表中，选择要使用的分数分支。
5. 单击**确定**。

## 无效参选者

要进行比较，模型文件必须具有使用公共数据体系结构的评分分支。

所有参选者都必须符合的数据特征包括以下几项：

- 数据源必须具有相同数量的字段。
- 字段名必须在各个数据源中都必须一致。
- 字段测量级别在各个数据源中都必须一致。

如果系统无法找到可与所选的第一个参选者进行比较的参选者，将显示“无效参选者”对话框。要选择新的参选者：

1. 单击**确定**返回到“参选者”选项卡。
2. 选择新的参选者。

## 选择参选者数据源

要在优胜者参选者分析中使用的数据源，请执行以下步骤：

1. 在优胜者参选者步骤中的“参选者”选项卡中，单击**浏览**以找到数据源参选者。
2. 从分析中包含的参选者列表中，选择包含要使用的数据源的分数分支。
3. 单击**确定**。

## 优胜者

为优胜者指定信息之前，必须选择至少一个参选者。如果在选择参选者之前访问了“参选者”选项卡，将显示“第一个参选者未选定”对话框，表示必须选择一个参选者。

**不创建优胜者的新版本。** 选中此选项可防止创建优胜者的新版本。在这种情况下，将会修改选择的优胜者标签版本。取消选中此选项可创建优胜者的新版本，而不是修改标签版本。

**文件名。** 要对指定为优胜者的参选者副本使用的名称。

**位置。** 优胜者副本文件的存储位置。

**许可权。** 与优胜者副本关联的许可权。

**元数据。** 与优胜者副本关联的属性。指定优胜者输出的元数据与指定其他作业输出的元数据相同。

## 在其他作业中使用优胜者

运行优胜者参选者作业后，可在其他作业中使用生成的优胜者。要在其他作业中包含优胜者，需要以下信息：

- 优胜者的名称。
- 优胜者的位置。

在其他作业中使用该优胜者时，将应用最新标签。无法修改该标签。

## 测试优胜者

缺省情况下，每次运行优胜者参选者作业步骤时，系统都会创建优胜者的新副本，将该副本存储在指定的输出位置，并将结果写入作业历史记录日志中。

但是，您可能不希望保存优胜者的副本。例如，假定您只希望测试优胜者参选者作业步骤。

要禁用副本创建，请选中**不创建优胜者的新版本**复选框。如果选中该复选框，将禁用选项卡中的剩余选项。系统将使用已应用于当前优胜者的信息。

系统将执行优胜者参选者作业步骤并确定新的优胜者。但是，将不会创建优胜者的新版本并将其保存至存储库。而只会将结果写入作业历史记录日志（指示哪个参选者应选为优胜者）。

例如，假设运行了某个优胜者参选者作业，并且作业的创建者已选择不创建优胜者的新版本。生成的作业历史记录日志可能如下：

```
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.
```

请注意日志文件的最后一行:

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

这一行表示 *cc\_neuralnetresponse* 流应该是优胜者。然而, 该流的副本将不会保存到存储库中, 因为系统未创建该流的新版本。如果流已经保存到存储库, 日志将还会多包含一行, 表示该流已经保存到存储库 - 例如:

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

## 数据文件

会为已从“参选者”选项卡的参选者列表中选择参选者显示数据文件信息。

在此选项卡中所进行的任何更改只适用于优胜者参选者作业步骤。对数据文件信息所作的修改不会逆向应用于在 IBM SPSS Collaboration and Deployment Services Repository 中保存的参选者。数据文件表包含以下信息。

**节点名。** 包含流所使用数据的输入节点名称。节点名称不能修改。

**节点类型。** 在流中定义的节点类型。节点类型不能修改。

**文件名。** 输入数据文件的名称。要更改名称, 请单击“文件名”单元格并更改名称。

**格式。** 输出文件的格式, 例如以逗号分隔的文件。要修改文件格式类型, 单击“格式”单元格。将显示下拉箭头。选择格式类型。

**位置。** 输入数据文件的位置。要修改位置, 单击列, 然后单击显示的省略号按钮。此时将打开“输入文件位置”对话框。必要时更改位置。

## 数据视图

对于从“参选者”选项卡的参选者列表中选择参选者, 将会显示分析数据视图信息。

在此选项卡中所进行的任何更改只适用于优胜者参选者作业步骤。对数据视图信息所作的修改不会传播回 IBM SPSS Collaboration and Deployment Services Repository 中保存的参选者。数据视图表包含以下信息。

**节点名。** 这是包含流所用数据的数据视图节点的名称。节点名称不能修改。

**分析数据视图。** 这是由数据视图节点引用的分析数据视图。

**标签。** 这是标识所用分析数据视图的版本的标签。

**表名称。** 这是包含输入数据字段的表。

**数据访问方案。** 这是为输入数据字段提供数据记录的方案。要更改对节点使用的数据访问方案，请选择包含该访问方案的单元格，然后单击显示的省略号 (...) 按钮。

## ODBC 数据源

---

会为已从“参选者”选项卡的参选者列表中选择参选者显示 ODBC 数据源信息。

在此选项卡中所进行的任何更改只适用于优胜者参选者作业步骤。对 ODBC 数据源信息所作的修改不会逆向应用于在 IBM SPSS Collaboration and Deployment Services Repository 中保存的参选者。ODBC 数据源表包含以下信息。

**节点名。** 包含流所使用数据的输入节点名称。名称的前缀为包含了以斜杠分隔的节点的任何超节点的名称。例如，如果节点 *MyNode* 位于名为 *Supernode1* 的超节点中，则名称显示为 */Supernode1/MyNode*。

**节点类型。** 在流中定义节点类型。

**ODBC 数据源。** 当前 ODBC 数据源名称 (DSN)。要切换至其他 ODBC 数据源，请单击包含当前数据源名称的单元格，然后单击所显示的 "..." 按钮。执行上述步骤后，将会显示可在其中选择现有 DSN 或创建新 DSN 的对话框。请注意，作业步骤设置始终会覆盖流；因此，如果您将流修改为使用其他数据源，那么还必须编辑作业以使用同一源，否则该作业可能无法运行。

**凭证。** 要在更改 ODBC 数据源时更改数据库用户名和密码，请单击包含当前凭证的单元格，然后单击 "..." 按钮。这将显示一个对话框，可以在该对话框中选择现有凭证定义或创建新的凭证定义。

**数据库表。** 对应于节点的数据库表。

无法访问已锁定超节点中的节点。也无法查看或修改这些节点。

## Cognos 导入

---

如果模型文件包含任何 IBM Cognos BI 源节点，那么此处将显示 Cognos 连接详细信息。

**节点名。** Cognos 源节点的名称。

**连接 URL。** 进行连接的 Cognos 服务器的 URL。

**包名称。** 导入元数据的 Cognos 数据包名称。

**匿名。** 如果 Cognos 服务器连接使用了匿名登录，包含**匿名**，或者在使用了特定的 Cognos 用户名和密码时，包含**凭证**。

**凭证。** Cognos 服务器上的用户名和密码（如果需要）。

**注：**必须在代表 Cognos 名称空间标识的域中创建 Cognos 凭证。

## 注意事项

---

本信息是为在美国提供的产品和服务编写的。IBM 可能会提供其他语言形式的本资料。但是，您可能需要拥有该语言的产品副本或产品版本，才能对其进行访问。

IBM 可能在其他国家或地区不提供本文档中讨论的产品、服务或功能。有关您所在区域当前可获得的产品和服务的信息，请向您当地的 IBM 代表咨询。任何对 IBM 产品、程序或服务的引用并非意在明示或暗示只能使用 IBM 的产品、程序或服务。只要不侵犯 IBM 的知识产权，任何同等功能的产品、程序或服务，都可以代替 IBM 产品、程序或服务。但是，评估和验证任何非 IBM 产品、程序或服务，则由用户自行负责。

IBM 可能已拥有或正在申请与本文档内容有关的各项专利。提供本文档并不意味着授予用户使用这些专利的任何许可。您可以以书面形式将许可查询寄往：

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

有关双字节 (DBCS) 信息的许可查询，请与您所在国家或地区的 IBM 知识产权部门联系，或以书面形式将查询寄往：

*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“按现状”提供本出版物，不附有任何种类的（无论是明示的还是暗含的）保证，包括但不限于暗含的有关非侵权、适销和适用于某种特定用途的保证。某些管辖区域在某些交易中不允许免除明示或暗含的保证。因此本条款可能不适用于您。

本信息中可能包含技术方面不够准确的地方或印刷错误。此处的信息将定期更改；这些更改将编入本资料的新版本中。IBM 可以随时对本出版物中描述的产品和/或程序进行改进和/或更改，而不另行通知。

本信息中对非 IBM Web 站点的任何引用都只是为了方便起见才提供的，不以任何方式充当对那些 Web 站点的保证。那些 Web 站点中的资料不是本 IBM 产品资料的一部分，使用那些 Web 站点带来的风险将由您自行承担。

IBM 可以按它认为适当的任何方式使用或分发您所提供的任何信息而无须对您承担任何责任。

本程序的被许可方如果要了解有关程序的信息以达到如下目的：(i) 允许在独立创建的程序和其他程序（包括本程序）之间进行信息交换，以及 (ii) 允许对已经交换的信息进行相互使用，请与下列地址联系：

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

只要遵守适当的条件和条款，包括某些情形下的一定数量的付费，都可获得这方面的信息。

本文档中描述的许可程序及其所有可用的许可资料均由 IBM 依据 IBM 客户协议、IBM 国际程序许可协议或任何同等协议中的条款提供。

所引用的性能数据和客户示例仅作参考用途。实际的性能结果可能会因特定的配置和运营条件而异。

涉及非 IBM 产品的信息是从这些产品的供应商、已出版说明或其他可公开获得的资料中获取。IBM 没有对这些产品进行测试，也无法确认其性能的精确性、兼容性或任何其他关于非 IBM 产品的声明。有关非 IBM 产品性能的问题应当向这些产品的供应商提出。

关于 IBM 未来方向或意向的声明都可随时更改或收回，而不另行通知，它们仅仅表示了目标和意愿而已。

本信息包含在日常业务操作中使用的数据和报告的示例。为了尽可能完整地说明这些示例，示例中可能会包括个人、公司、品牌和产品的名称。所有这些名字都是虚构的，若与实际个人或业务企业相似，纯属巧合。

## 商标

---

IBM、IBM 徽标和 [ibm.com](http://ibm.com) 是 International Business Machines Corp.，在全球许多管辖区域注册的商标或注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。IBM 商标的最新列表可从 Web 上的“Copyright and trademark information”处获得，网址为：[www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)。

Adobe、Adobe 徽标、PostScript 以及 PostScript 徽标是 Adobe Systems Incorporated 在美国和/或其他国家或地区的注册商标或商标。

Intel、Intel 徽标、Intel Inside、Intel Inside 徽标、Intel Centrino、Intel Centrino 徽标、Celeron、Intel Xeon、Intel SpeedStep、Itanium 和 Pentium 是 Intel Corporation 或其子公司在美国或其他国家或地区的商标或注册商标。

Linux 是 Linus Torvalds 在美国和/或其他国家或地区的注册商标。

Microsoft、Windows、Windows NT 和 Windows 徽标是 Microsoft Corporation 在美国和/或其他国家或地区的商标。

UNIX 是 The Open Group 在美国和其他国家或地区的注册商标。

Java 和所有基于 Java 的商标和徽标是 Oracle 和/或其子公司的商标或注册商标。

## 产品文档的条款和条件

---

根据以下条款和条件授予这些出版物的使用许可权。

### 适用性

这些条款和条件是对 IBM Web 站点的任何使用条款的补充。

### 个人使用

您可以复制这些出版物供个人非商业性使用，但前提是保留所有专有权声明。未经 IBM 明确同意，您不可以分发、展示或制作这些出版物或其中任何部分的演绎作品。

### 商业性使用

您仅可在贵公司内部复制、分发和显示这些出版物，但前提是保留所有专有权声明。未经 IBM 明确同意，您不可以制作这些出版物的演绎作品，或者在您的企业外部复制、分发或展示这些出版物或其中的任何部分。

### 权利

除非本许可权中明确授予，否则不得授予对这些出版物或其中包含的任何信息、数据、软件或其他知识产权的任何许可权、许可证或权利，无论明示的还是暗含的。

只要 IBM 认为这些出版物的使用会损害其利益或者 IBM 判定未正确遵守上述指示信息，IBM 将有权撤销本文授予的许可权。

只有您完全遵循所有适用的法律和法规，包括所有的美国出口法律和法规，您才可以下载、出口或再出口该信息。

IBM 对这些出版物的内容不作任何保证。这些出版物“按现状”提供，不附有任何种类的（无论是明示的还是暗含的）保证，包括但不限于暗含的有关适销性、非侵权和适用于某种特定用途的保证。



# 索引

## Special Characters

比较  
    模型 [15, 19](#)

步骤  
    IBM SPSS Modeler [7, 9–11, 14, 17](#)

参数 [14](#)

参选者模型  
    数据源 [21](#)

超节点 [14](#)

分析数据视图  
    优胜者参选者 [23](#)

服务器定义  
    添加 [5](#)  
    修改 [5](#)  
    IBM SPSS Modeler [5](#)

服务器集群 [2, 4](#)

脚本编制  
    优胜者参选者 [19](#)  
    优胜者与参选者 [15](#)  
    IBM SPSS Modeler [15, 19](#)

流  
    IBM SPSS Modeler [7, 9–11, 14, 17](#)

模型  
    比较 [15, 19](#)  
    监视 [15, 19](#)  
    评估 [15, 19](#)  
    优胜者参选者 [19](#)  
    优胜者与参选者 [15](#)  
    IBM SPSS Modeler [15, 19](#)

评分服务 [17](#)

数据文件  
    优胜者参选者 [23](#)

文件  
    IBM SPSS Modeler [7, 9–11, 14, 17](#)

文件夹  
    资源定义 [5](#)

协作 [1](#)

用户首选项 [4](#)

优胜者参选者 [19–24](#)

优胜者模型 [15, 19, 21, 22](#)

优胜者与参选者 [15](#)

执行服务器  
    IBM SPSS Modeler [2, 4](#)

资源定义  
    服务器定义 [5](#)

作业  
    IBM SPSS Modeler [7, 9–11, 14, 17](#)

作业步骤名称  
    优胜者参选者 [20](#)

## C

Cognos  
    优胜者参选者 [24](#)

## D

deployment [1](#)

## I

IBM SPSS Collaboration and Deployment Services  
    Deployment Manager [2, 3](#)  
IBM SPSS Collaboration and Deployment Services  
    Deployment Portal [2, 4](#)  
IBM SPSS Collaboration and Deployment Services  
    Repository [2](#)  
IBM SPSS Modeler  
    服务器定义 [5](#)  
    流 [7, 9–11, 14, 17](#)  
    执行服务器 [2, 4, 5](#)

## O

ODBC 数据源  
    优胜者参选者 [24](#)





