

IBM SPSS Modeler Server 15 管
理和性能指南



注意：使用本信息及其支持的产品之前，请阅读注意事项第 页码下的一般信息。

此版本适用于 IBM SPSS Modeler 15 及所有后续发布和修订，除非在新版本中另有说明。

Adobe 产品屏幕截图重印已获得 Adobe Systems Incorporated 的许可。

Microsoft 产品屏幕截图重印已获得 Microsoft Corporation 的许可。

受许可保护材料 - IBM 所有

Copyright IBM Corporation 1994, 2012.

美国政府用户受限权利 - 使用、复制或披露受与 IBM Corp. 签订的 GSA ADP Schedule Contract 的限制。

前言

IBM® SPSS® Modeler 是 IBM Corp. 企业级数据挖掘工作平台。SPSS Modeler通过深入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，这将生成更强有力的预测模型，缩减实现解决方案所需的时间。SPSS Modeler 提供了多种建模技术，例如预测、分类、细分和关联检测算法。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 软件为决策者提供可信赖的完整、一致和准确信息，以帮助其提升业务绩效。这一涵盖 [商务智能](#)、[预测分析](#)、[财务绩效与战略管理](#)以及[分析应用程序](#)的全面组合可提供有关当前业务表现的清晰、立即和切实可行的深入见解，并能够有效预测未来结果。其中整合了丰富的行业解决方案、经过验证的做法与专业服务，以帮助各种规模的组织提升生产效率、自动化决策并取得卓越成果。

作为该软件组合的一部分，IBM SPSS Predictive Analytics 软件能够帮助各类组织有效地预测未来事件，并针对所得到的深入见解提前采取行动，以取得更优秀的业务成果。全球企业、政府和学院客户依赖 IBM SPSS 技术作为吸引、留住和增加客户数量的竞争优势，并降低欺诈和转移风险。通过将 IBM SPSS 软件融入其日常运营中，这些组织将成为“预测型”企业，即能够指引并自动化决策，以实现业务目标和取得可衡量的竞争优势。有关详细信息，或联系我们的代表，请访问 <http://www.ibm.com/spss>。

技术支持

我们提供有技术支持服务以维护客户。客户可就 IBM Corp. 产品使用或某一受支持硬件环境的安装帮助寻求技术支持。要获得技术支持，请访问 IBM Corp. 网站 <http://www.ibm.com/support>。在请求帮助时，请做好准备，以便识别您自己、您的组织以及您的支持协议。

内容

1	关于 IBM SPSS Modeler	1
	IBM SPSS Modeler 产品	1
	IBM SPSS Modeler	1
	IBM SPSS Modeler Server	2
	IBM SPSS Modeler Administration Console	2
	IBM SPSS Modeler Batch	2
	IBM SPSS Modeler Solution Publisher	2
	IBM SPSS Modeler Server 适配器用于 IBM SPSS Collaboration and Deployment Services	2
	IBM SPSS Modeler 版本	3
	IBM SPSS Modeler 文档	3
	SPSS Modeler Professional 文档	4
	SPSS Modeler Premium 文档	4
	应用程序示例	5
	Demos 文件夹	6
2	基础架构和硬件建议	7
	IBM SPSS Modeler 基础架构	7
	基础架构描述	7
	硬件建议	9
	对临时磁盘空间和 RAM 的要求	10
	数据访问	11
	引用数据文件	13
	IBM SPSS Statistics 数据文件的导入	13
	安装说明	13
3	IBM SPSS Modeler 支持度	14
	连接到 IBM SPSS Modeler Server	14
	添加并编辑 IBM SPSS Modeler Server 连接	16
	搜索 IBM SPSS Collaboration and Deployment Services 中的服务器	17
	数据和文件系统	18
	用户验证	18
	文件创建	18
	不同的结果	19

4 IBM SPSS Modeler Server 管理 20

启动和停止 IBM SPSS Modeler Server	20
在 Windows 中启动、停止和检查状态	20
在 UNIX 中启动、停止和检查状态	20
处理不响应服务器进程 (UNIX 系统)	21
管理	22
使用 IBM SPSS Modeler Administration Console	22
启动 Modeler Administration Console	22
使用 Modeler Administration Console 配置访问权限	23
SPSS Modeler Server 连接	24
SPSS Modeler Server 配置	24
SPSS Modeler Server 监视	31
使用 options.cfg 文件	31
关闭不使用的数据库连接	32
IBM SPSS Statistics 许可证位置	32
使用 SSL 进行安全数据传输	33
SSL 如何运行	33
通过 SSL 确保客户端-服务器和服务器-客户端的通信安全	33
仅 URL 前缀配置	37
仅通过 SSL 实现安全 LDAP	38
服务器日志	39

5 性能概述 41

服务器性能和优化设置	41
客户端性能和优化设置	41
数据库使用情况和优化	43
SQL 优化	44

6 SQL 优化 45

SQL 生成如何运行	46
SQL 生成示例	46
配置 SQL 优化	48
预览生成的 SQL	49
查看模型块 SQL	50
最大化 SQL 生成的提示	51

支持 SQL 生成的节点	52
CLEM 支持 SQL 生成的表达式和运算符	55
使用 CLEM 表达式中的 SQL 函数	57
写入 SQL 查询	57

附录

A 为 UNIX 平台配置 Oracle 59

配置 Oracle 的 SQL 优化	59
------------------------------	----

B 配置 UNIX 启动脚本 61

简介	61
脚本	61
自动启动和停止 IBM SPSS Modeler Server	61
手动启动和停止 IBM SPSS Modeler Server	62
编辑脚本	62
控制文件创建权限	63
IBM SPSS Modeler Server 和数据访问包	63
ODBC 配置故障排除	65
库路径	68
为 Enterprise View 节点配置驱动程序	69

C 在 UNIX 上以非根进程方式配置和运行 SPSS Modeler Server 70

简介	70
不使用专有密码数据库以非根方式配置	70
使用专有密码数据库以非根方式配置	71
作为非根用户运行 SPSS Modeler Server	72
用户验证失败故障排除	73

D	使用服务器群集负载均衡	74
E	注意事项	76
	索引	78

关于 IBM SPSS Modeler

IBM® SPSS® Modeler 是一组数据挖掘工具，通过这些工具可以采用商业技术快速建立预测性模型，并将其应用于商业活动，从而改进决策过程。SPSS Modeler 参照行业标准 CRISP-DM 模型设计而成，可支持从数据到更优商业成果整个数据挖掘过程。

SPSS Modeler 提供了各种借助机器学习、人工智能和统计学的建模方法。通过建模选项板中的方法，您可以根据数据生成新的信息以及开发预测模型。每种方法各有所长，同时适用于解决特定类型的问题。

SPSS Modeler 可以作为独立产品购买，或作为客户端与 SPSS Modeler Server 一起使用。同时提供了大量其他选项，以下各节将对这些选项进行概述。有关详细信息，请参阅 <http://www.ibm.com/software/analytics/spss/products/modeler/>。

IBM SPSS Modeler 产品

IBM® SPSS® Modeler 系列产品及其相关软件包括如下成员：

- IBM SPSS Modeler
- IBM SPSS Modeler Server
- IBM SPSS Modeler Administration Console
- IBM SPSS Modeler Batch
- IBM SPSS Modeler Solution Publisher
- IBM SPSS Modeler Server 适配器用于 IBM SPSS Collaboration and Deployment Services

IBM SPSS Modeler

SPSS Modeler 是在个人电脑上安装并运行的完整功能版本产品。您可以在本地模式下将 SPSS Modeler 作为独立产品来运行；也可以在分布式模式下与 IBM® SPSS® Modeler Server 协同使用，从而提高了对大数据集的处理速度。

使用 SPSS Modeler，您可以快速直观地建构精确的预测模型，无需编程。使用其独特的可视界面，您可以轻松地将数据挖掘过程可视化。通过内嵌在该产品中的高级分析支持，您可以发现之前在您的数据中隐藏的模式与趋势。您可以建构输出模型，并理解其影响因子，让您可以更好地利用业务机会、降低风险。

SPSS Modeler 提供两个版本：SPSS Modeler Professional 和 SPSS Modeler Premium。有关详细信息，请参阅 [IBM SPSS Modeler 版本中的 IBM SPSS Modeler 15 用户指南](#)。

IBM SPSS Modeler Server

SPSS Modeler 使用客户端/服务器体系结构将资源集约型操作的请求分发给功能强大的服务器软件，因而使大数据集的传输速度大大加快。

SPSS Modeler Server 是一项独立许可的产品，可以协同一份或多份 IBM® SPSS® Modeler 安装，以分布式分析模式在服务器主机上持续运行。因此，SPSS Modeler Server 在处理大型数据集时具有卓越的性能，因为内存密集型的操作可在服务器完成而无需将数据下载到客户端计算机。IBM® SPSS® Modeler Server 还提供了对 SQL 优化的支持与数据库内建模能力，有助于进一步提高性能与自动化程度。

IBM SPSS Modeler Administration Console

Modeler Administration Console 是一个图形化的应用程序，可管理 SPSS Modeler Server 的多项配置选项，配置也可通过编辑一个选项文件来进行。该应用程序提供了一个控制台用户界面，用以监控和配置所安装的 SPSS Modeler Server，SPSS Modeler Server 的现用户可以免费使用该应用程序。应用程序只能安装在 Windows 计算机上；但是它可以管理安装在任何受支持平台上的服务器。

IBM SPSS Modeler Batch

虽然数据挖掘通常是交互式的过程，但是也可以通过命令行来运行 SPSS Modeler 而无需图形用户界面。例如，您可能有些任务需长期或重复性地运行而无用户干预。SPSS Modeler Batch 是该产品一个特殊版本，无需通过常规的用户界面即可完整地实现 SPSS Modeler 的分析功能。使用 SPSS Modeler Batch 需要具备 SPSS Modeler Server 许可。

IBM SPSS Modeler Solution Publisher

SPSS Modeler Solution Publisher 是一个工具，让您可以创建 SPSS Modeler 流的打包版本，由外部运行时引擎运行，或嵌入一个外部应用程序。以此方式，您可以发布与部署完整的 SPSS Modeler 流，在未安装 SPSS Modeler 的环境下也能使用。SPSS Modeler Solution Publisher 是作为 IBM SPSS Collaboration and Deployment Services - Scoring 服务的一部分来发行的，需另行购买许可。获得许可后，您会收到 SPSS Modeler Solution Publisher Runtime，让您可以执行已发布的流。

IBM SPSS Modeler Server 适配器用于 IBM SPSS Collaboration and Deployment Services

有若干种 IBM® SPSS® Collaboration and Deployment Services 适配器可以让您通过 SPSS Modeler 和 SPSS Modeler Server 与 IBM SPSS Collaboration and Deployment Services 存储库交互。以此方式，部署在存储库中的 SPSS Modeler 流即可实现多用户共享，或通过瘦客户端应用程序 IBM SPSS Modeler Advantage 访问。适配器须安装在存储库的主机系统中。

IBM SPSS Modeler 版本

SPSS Modeler 提供下列版本。

SPSS Modeler Professional

SPSS Modeler Professional 提供您在处理大多数类型的结构化数据（例如 CRM 系统中跟踪的行为或交互活动、人口统计、购买行为与销售数据）时所需的所有工具。

SPSS Modeler Premium

SPSS Modeler Premium 是一项单独许可的产品，它扩展了 SPSS Modeler Professional 的功能，使其可以处理如实体分析或社会网络专门数据，以及非结构化的文本数据。SPSS Modeler Premium 包含下列组件。

IBM® SPSS® Modeler Entity Analytics 在 IBM® SPSS® Modeler 预测分析的基础上添加了全新的维度。预测分析会尝试根据过去数据预测未来行为，而实体分析侧重于通过解析记录自身的身份冲突，提高当前数据的连贯性和一致性。身份可以指个人、组织、对象或可能不确定的任何其他实体的身份。身份解析在许多领域中都非常重要，包括客户关系管理、检测、反洗钱以及国家与国际安全。

IBM SPSS Modeler Social Network Analysis 将关于关系的信息转换为字段，这些字段可描述个人和组社交行为的特征。使用介绍社交网络之下关系的数据，IBM® SPSS® Modeler Social Network Analysis 可识别影响网络中他人行为的社交领导。此外，可确定受其他网络参与者影响最大的人员。通过结合这些结果和其他测量，您可创建个人的综合配置文件，作为预测模型的基础。包括此社交信息的模型比不包括的模型执行效果更好。

Text Analytics for IBM® SPSS® Modeler 采用了先进语言技术和 Natural Language Processing (NLP)，以快速处理大量无结构文本数据，抽取和组织关键概念，以及将这些概念分为各种类别。抽取的概念和类别可以和现有结构化数据中进行组合（例如人口统计学），并且可用于借助 SPSS Modeler 的一整套数据挖掘工具来进行建模，以此实现更好更集中的决策。

IBM SPSS Modeler 文档

可以从 SPSS Modeler 的帮助菜单中获取在线帮助格式的文档。此文档包括 SPSS Modeler、SPSS Modeler Server 和 SPSS Modeler Solution Publisher 的文档以及《应用程序指南》和其他支持材料。

每个产品的完整文档（包括安装说明）也在每个产品 DVD 的 \Documentation 文件夹下以 PDF 格式提供。安装文档也可从以下网页中下载：
<http://www-01.ibm.com/support/docview.wss?uid=swg27023172>。

两种格式的文档均可从 SPSS Modeler 信息中心获取，其网址如下：
<http://publib.boulder.ibm.com/infocenter/spssmodl/v15r0m0/>。

SPSS Modeler Professional 文档

SPSS Modeler Professional 的文档套件（不含安装说明）如下：

- **IBM SPSS Modeler 用户指南。** 使用 SPSS Modeler 的一般使用介绍，包括如何构建数据流、处理缺失值、生成 CLEM 表达式、处理项目和报告以及将用于部署的流打包为 IBM SPSS Collaboration and Deployment Services、Predictive Application 或 IBM SPSS Modeler Advantage。
- **IBM SPSS Modeler 源、处理和输出节点。** 介绍用于以不同的格式读取、处理和输出数据的所有节点。实际上这表示所有节点而非建模节点。
- **IBM SPSS Modeler 建模节点。** 有关用于创建数据挖掘模型的所有节点的描述。IBM® SPSS® Modeler 可提供各种借助机器学习、人工智能和统计学的建模方法。 [有关详细信息，请参阅第 3 章中的建模节点概述中的 IBM SPSS Modeler 15 建模节点。](#)
- **IBM SPSS Modeler 算法指南。** 介绍 SPSS Modeler 中所用建模方法的数学基础。此指南仅提供 PDF 版。
- **IBM SPSS Modeler 应用程序指南。** 本指南中的示例旨在为具体的建模方法和技术提供具有针对性的简介。还可以在“帮助”菜单中查阅本指南的在线版本。 [有关详细信息，请参阅应用程序示例中的 IBM SPSS Modeler 15 用户指南。](#)
- **IBM SPSS Modeler 脚本编写与自动化。** 通过编写脚本实现系统自动化的相关信息，包括用于操作节点和流的属性信息。
- **IBM SPSS Modeler 部署指南。** 有关在 IBM® SPSS® Collaboration and Deployment Services Deployment Manager 中以处理作业的步骤形式运行 SPSS Modeler 流和方案的信息。
- **IBM SPSS Modeler CLEF 开发人员指南** CLEF 提供了将第三方程序（例如，数据处理例程或建模算法）作为节点集成到 SPSS Modeler 的功能。
- **IBM SPSS Modeler 数据库内数据挖掘指南。** 有关如何利用数据库的功能通过第三方法来改进性能并增强分析功能的信息。
- **IBM SPSS Modeler Server 管理和性能指南。** 有关如何配置和管理 IBM® SPSS® Modeler Server 的信息。
- **IBM SPSS Modeler Administration Console 用户指南。** 有关安装和使用控制台用户界面以监视和配置 SPSS Modeler Server 的信息。控制台实现为 Deployment Manager 应用程序的插件。
- **IBM SPSS Modeler Solution Publisher 指南。** SPSS Modeler Solution Publisher 是一个附加式组件，通过它组织可发布在标准 SPSS Modeler 环境之外使用的流。
- **IBM SPSS Modeler CRISP-DM 指南。** 借助 CRISP-DM 方法进行 SPSS Modeler 数据挖掘的分步指南。
- **IBM SPSS Modeler Batch 用户指南。** 在批处理模式下使用 IBM SPSS Modeler 的完整指南，包括批处理模式的执行与命令行参数的详细信息。此指南仅提供 PDF 版。

SPSS Modeler Premium 文档

SPSS Modeler Premium 的文档套件（不含安装说明）如下：

- **IBM SPSS Modeler Entity Analytics 用户指南。** 关于通过 SPSS Modeler 使用实体分析的信息，涵盖存储库的安装与配置、实体分析节点以及管理任务。

- **IBM SPSS Modeler Social Network Analysis 用户指南。** 通过 SPSS Modeler 进行社会网络分析的指南，包括群组分析与传播分析。
- **Text Analytics for SPSS Modeler 用户指南。** 关于通过 SPSS Modeler 使用文本分析的信息，涵盖文本发掘节点、交互式工作台、模板及其他资源。
- **Text Analytics for IBM SPSS Modeler Administration Console 用户指南。** 有关安装和使用控制台用户界面监视和配置 IBM® SPSS® Modeler Server 用于 Text Analytics for SPSS Modeler 的信息。控制台实现为 Deployment Manager 应用程序的插件。

应用程序示例

SPSS Modeler 中的数据挖掘工具可以帮助解决很多业务和组织问题，应用程序示例将提供有关特定建模方法和技术的简明的针对性说明。此处使用的数据集比某些数据挖掘器管理的大量数据存储要小得多，但涉及的概念和方法应可扩展到实际的应用程序。

可以通过在 SPSS Modeler 中的“帮助”菜单中单击[应用程序示例](#)来访问示例。数据文件和样本流安装在产品安装目录下的 Demos 文件夹中。[有关详细信息，请参阅Demos 文件夹中的IBM SPSS Modeler 15 用户指南。](#)

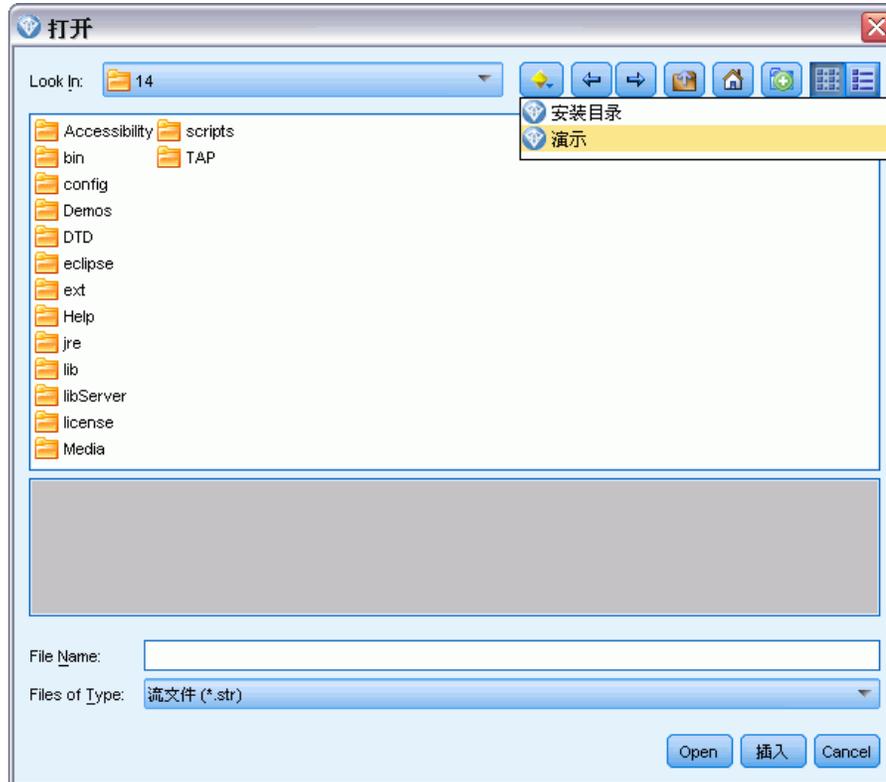
数据库建模示例。 请参阅 IBM SPSS Modeler 数据库内挖掘指南 中的示例。

编写示例脚本。 请参阅 IBM SPSS Modeler 脚本编写和自动化指南 中的示例。

Demos 文件夹

与应用程序示例一起使用的数据文件和样本流安装在产品安装目录下的 Demos 文件夹中。可从 Windows 的“开始”菜单中 IBM SPSS Modeler 15 程序组访问该文件夹，也可以在“文件打开”对话框中最近目录的列表中单击 Demos。

图片 1-1
在最近使用的目录列表中选择 Demos 文件夹



基础架构和硬件建议

IBM SPSS Modeler 基础架构

本节说明了 IBM® SPSS® Modeler Server 的基础架构，其中包括服务器软件、客户端软件和数据库。本节说明了 SPSS Modeler Server 如何实现性能优化，并且建议通过选择适当规模的硬件来尽可能地提高性能。数据访问章节中进行了相应的总结，此节描述了如何设置必要的 ODBC 驱动程序。

基础架构描述

IBM® SPSS® Modeler Server 采用三层分布式基础架构。服务器和客户端计算机共享软件操作项。安装和使用 SPSS Modeler Server（相对于独立的 IBM® SPSS® Modeler）有诸多优势，尤其是在处理大型数据集合时：

- SPSS Modeler Server 除 Windows 系统以外，可以在 UNIX 上运行，这样在选择安装平台时更具灵活性。在任何平台上，您均可指定快速、大型的服务器计算机来进行数据挖掘。
- 对 SPSS Modeler Server 进行优化，以提高运行速度。当无法将操作项推送到数据库中时，SPSS Modeler Server 会将中间结果作为临时文件存储到磁盘而不是 RAM 中。因为服务器上的磁盘可用空间通常较大，所以 SPSS Modeler Server 可以对大型的数据集合执行排序、合并和聚合操作。
- 通过客户-服务器的基础架构，您可以将组织内的数据挖掘过程集合在一起进行。集中化有助于规范业务流程中数据挖掘的作用。
- 使用管理员工具，例如，IBM® SPSS® Modeler Administration Console（SPSS Modeler Server 附带）和 IBM® SPSS® Collaboration and Deployment Services（单独出售），您可以监视数据挖掘过程，同时确保有足够的可用计算资源。通过 IBM SPSS Collaboration and Deployment Services，您可以让某个数据挖掘任务自动化，管理对数据模型的访问，并且可以在组织内共享数据挖掘成果。

IBM® SPSS® Modeler 分布式基础架构的组件将显示在“IBM SPSS Modeler Server 基础架构”图形第 8 页码中。

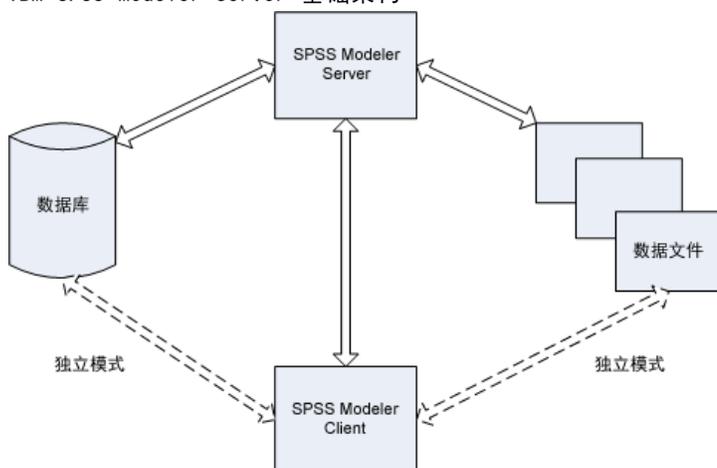
- **SPSS Modeler。**客户端软件将安装在最终用户的计算机上。客户端软件将提供用户界面并显示数据挖掘结果。此客户端是一个完整的 SPSS Modeler 软件安装程序，但当其与 SPSS Modeler Server 连接来进行分布式分析时，此客户端的引擎处于非激活状态。此 SPSS Modeler 仅能在 Windows 操作系统上运行。

- **SPSS Modeler Server**。此服务器软件安装在服务器计算机上，并通过网络与 SPSS Modeler(s) 和数据库相连接。SPSS Modeler Server 将作为服务（在 Windows 中）或作为守护进程（在 UNIX 中）运行，同时等待客户端连接。它将执行使用 SPSS Modeler 创建的流和脚本。
- **数据库服务器**。数据库服务器可以是某个现有的数据集市（例如，基于大型 UNIX 服务器的 Oracle 或基于或者为了降低对其他操作系统的影响，数据库服务器还可以是本地或部门服务器中的数据集市（例如，基于 Windows 的 SQL Server）。

IBM SPSS Modeler Server 基础架构

图片 2-1

IBM SPSS Modeler Server 基础架构



借助分布式基础架构，多数处理将在服务器计算机上进行。最终用户执行某个流时，SPSS Modeler 会将此流的描述发送给服务器。服务器将判断哪些操作可以在 SQL 中执行并创建相应查询。这些查询可以在数据库中执行，执行后得出的数据将传递给服务器来进行无法用 SQL 语言描述的处理。处理完成后，仅将相关结果回传给客户端。

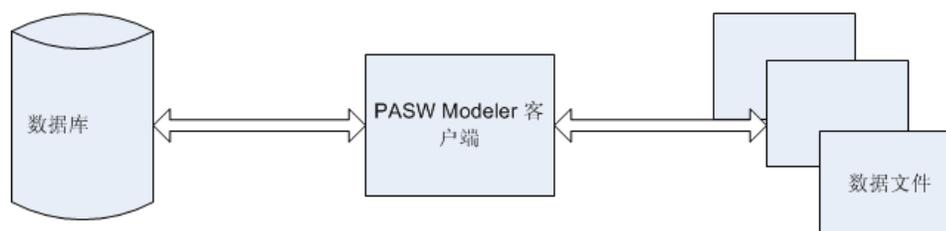
必要时，SPSS Modeler Server 可以在数据库之外执行所有 SPSS Modeler 操作。Clementine Server 会自动平衡对 RAM 和磁盘的使用以执行数据操作。此过程将使 SPSS Modeler Server 完全兼容平面文件。

使用服务器群集进行处理时还可以使用负载均衡功能。聚类通过进程协调器插件可以在 IBM SPSS Collaboration and Deployment Services 3.5 中启动。有关详细信息，请参阅第 74 页码附录 D 中的使用服务器群集负载均衡。通过 SPSS Modeler 的服务器登录对话框可以直接连接到在进程协调器中受管的服务器或群集。有关详细信息，请参阅第 3 章中的连接到 IBM SPSS Modeler Server 中的 IBM SPSS Modeler 15 用户指南。

独立客户端

SPSS Modeler 还可以通过配置来作为自包含桌面应用程序运行，如下面的“IBM SPSS Modeler 单机”图形所示。有关详细信息，请参阅第 14 页码第 3 章中的 IBM SPSS Modeler 支持度。

图片 2-2
IBM SPSS Modeler 单机



硬件建议

当您打算安装 IBM® SPSS® Modeler Server 时，应考虑要使用的硬件。尽管 SPSS Modeler Server 本身就是高速设计，您仍可通过在数据挖掘任务中使用适当规模的硬件来尽可能地提高性能。硬件升级通常是提高整体性能最简单且最经济的方式。

专用服务器。在专用服务器计算机上安装 SPSS Modeler Server 程序，专用服务器上的该程序将不与其他应用程序（包括 SPSS Modeler Server 可能连接到的数据库）争用资源。特别是模型构建操作，作为资源密集型操作，模型构建操作可以在不与其他应用程序争用资源的情况下更好地执行。

注意：虽然在作为数据库的同一计算机上安装 SPSS Modeler Server 可以通过避免网络开支来减少数据库和服务器间的数据传输时间，但多数情况下，最佳配置是使服务器和数据库分别位于独立的计算机上，以避免相互之间对资源的争用。通过提供服务器和数据库之间的快速连接来尽可能地降低数据传输成本。

处理器。计算机上处理器的数量不应小于您希望定期运行的并行任务（同时执行的流）的数量。通常来讲，处理器数量越多越好。

- SPSS Modeler Server 的单个实例将接受来自多个客户端（用户）的连接，每个客户端连接均可对多个流执行进行初始化。服务器因此可以在任一时间执行若干个任务。
- 基本原则是：一个处理器最多可由两个用户使用，两个处理器最多可由四个用户使用，四个处理器最多可由八个用户使用。此外，请为每两到四个用户增加一个处理器，具体情况取决于工作环境。
- 某些时候，一些处理可以通过 SQL 优化来推送回数据库，此时两个或多个用户可以在性能损耗最小化的同时共享 CPU 资源。
- 多线程功能使得单个任务可以获取多处理器的优势，就是说增加处理器可以提高性能，即便一次仅有一个任务运行。多线程通常用于 C5.0 模型构建和某些数据准备操作（排序、聚合和合并）。

64 位平台。如果您想在大量数据的基础上处理或构建模型，请使用 64 位 Solaris、Windows 或 Linux 作为 SPSS Modeler Server 平台，同时请将计算机内存最大化。对于大型数据集，服务器可能很快会耗尽 32 位平台所分配的单处理内存限制，这样将导致数据溢出到磁盘并大大增加运行时间。采用 64 位服务器可获取辅助 RAM（建议最小为 8 GB）的优势。64 位支持适用于 Solaris、Windows 和 Linux 平台。

未来需求。如果可行，请确保内存和 CPU 等服务器硬件可以扩展，可以同时容纳未来增加的使用量（例如，增加的同步用户数量或增加的现有用户处理需求）和增加的 SPSS Modeler Server 多线程功能。

对临时磁盘空间和 RAM 的要求

IBM® SPSS® Modeler Server 采用临时磁盘空间来处理大量数据。所需的临时磁盘空间量取决于您所处理数据的量和类型以及所执行操作的类型。数据量与行数和列数成正比。处理的行和列数越多，所需磁盘空间越大。

本节描述了需要临时磁盘空间和辅助 RAM 的情形，以及如何估算所需空间量。请注意，本节不讨论在数据库中进行数据处理时的临时磁盘空间需求，因为此类需求会因各种数据库而有所不同。

需要临时磁盘空间的情形

由于 IBM® SPSS® Modeler Server 的强大 SQL 优化功能，只要可以实现，处理都将在数据库（而不是服务器）中进行。然而，当处于以下情形时，则不能使用 SQL 优化功能：

- 要处理的数据保存在纯文本文件而不是数据库中。
- SQL 优化功能处于关闭状态。
- 不能使用 SQL 来优化处理操作。

无法使用 SQL 优化功能时，以下数据操控类节点和 CLEM 功能将创建部分或所有数据的临时磁盘备份。如果您的站点所用的流包含以下处理命令或功能，您可能需要对服务器上的辅助磁盘空间进行设置。

- 聚合节点
- 差异节点
- 分级节点
- 采用“按关键字合并”选项时合并节点
- 所有建模节点
- 排序节点
- 表输出节点
- @OFFSET 函数，其中的查询条件使用 @THIS。
- 任何 @ 函数（例如 @MIN、@MAX 和 @AVE），其中将计算偏移参数。

计算临时磁盘空间量

通常，IBM® SPSS® Modeler Server 需要能够写临时文件，临时文件大小至少是原始数据集合的三倍。例如，如果数据文件为 2GB 且未使用 SQL 生成，SPSS Modeler Server 将需要 6GB 磁盘空间来处理数据。由于每个并行用户帐户均会创建其自有临时文件，所以您需要相应地增加每个并行用户的磁盘空间。

如果您发现自己的站点经常使用大型临时文件，则请考虑对 IBM® SPSS® Modeler 临时文件使用独立文件系统（在某个单一磁盘上创建）。为了获得最佳效果，可以采用 RAID 0 或跨越多个物理磁盘的分区数据集合来加速磁盘操作，理想状态下，可以是每个磁盘位于一个独立磁盘控制器上的分区文件系统中。

RAM 的要求

对于多数无法在数据库中完成的操作，IBM® SPSS® Modeler Server 会把中间结果作为临时文件存储到磁盘而不是内存（RAM）中。但对于建模节点，可能的话还将采用 RAM。神经网络、Kohonen 和 K-Means 节点需要大量 RAM。如果您的站点经常使用这些节点，请考虑为服务器安装更多 RAM。

通常，所需 RAM 字节数可通过下面的公式计算：

$$(\text{number_of_records} * \text{number_of_cells_per_record}) * \text{number_of_bytes_per_cell}$$

当存在名义字段时，`number_of_cells_per_record` 的值可以变得非常大。

有关对 RAM 的当前建议，请参阅服务器安装指南的系统要求部分。如果有四个甚至更多同步用户，建议安装更大的 RAM。并行任务必须共享内存，所以应相应增加内存。通常，增加内存似乎是提高整体性能的最有效的方式之一。

数据访问

要读取或写入到数据库中，您必须为相关数据库安装并配置 ODBC 数据源，并根据需要配置读取或写入权限。IBM® SPSS® Data Access Pack 包括一组用于此用途的 ODBC 驱动程序，在 IBM SPSS Data Access Pack DVD 或从下载站点可找到这些驱动程序。如果您有关于创建或设置 ODBC 数据源权限的问题，请联系您的数据库管理员。

在 IBM® SPSS® Modeler 中数据库支持分为三层，分别代表对 SQL 回送和优化的不同支持级别，具体取决于数据库供应商。不同的支持级别采用一定的系统设置来实现，并作为服务合约的组成部分加以自定义。

数据库支持的三个层包括：

表 2-1
数据库支持层

支持层	描述
第 1 层	所有可能的 SQL 回送都可用，并具有数据库特定的 SQL 优化。
第 2 层	多数 SQL 回送可用，具有非数据库特定的 SQL 优化。
第 3 层	没有 SQL 回送或优化，只能向数据库读取和写入数据。

支持的 ODBC 驱动程序

有关使用 SPSS Modeler 15 支持和测试的数据库和 ODBC 驱动程序的最新信息，请参阅公司支持站点上的产品兼容性矩阵 (<http://www.ibm.com/support>)。

在哪里安装驱动程序

注意，必须在每台可能进行处理的计算机上安装并配置 ODBC 驱动程序。

- 如果您以本地（独立）模式运行 IBM® SPSS® Modeler，必须在本地计算机上安装驱动程序。
- 如果您以分布模式针对远程 IBM® SPSS® Modeler Server 运行 SPSS Modeler，需要在安装 SPSS Modeler Server 的计算机上安装 ODBC 驱动程序。对于 UNIX 系统中的 SPSS Modeler Server，也可参阅本节稍后说明的“在 UNIX 系统中配置 ODBC 驱动程序”。
- 如果您需要从 SPSS Modeler 和 SPSS Modeler Server 中访问相同数据源，必须在两个计算机上都安装 ODBC 驱动程序。
- 如果您通过终端服务运行 SPSS Modeler，需要在安装 SPSS Modeler 的终端服务服务器上安装 ODBC 驱动程序。
- 如果您使用 IBM® SPSS® Modeler Solution Publisher Runtime 在单独的计算机上运行发布的流，您也需要在该计算机上安装并配置 ODBC 驱动程序。

注意：如果您在 UNIX 上使用 SPSS Modeler Server 访问 Teradata 数据库，必须使用与 Teradata ODBC 驱动程序一起安装的 ODBC 驱动程序管理器。为了对 SPSS Modeler Server 进行此更改，请在靠近 modelersrv.sh 脚本的顶部、注释所指示的地方为 ODBC_DRIVER_MANAGER_PATH 指定一个值。此环境变量需要设置为 Teradata ODBC 驱动程序自带的 ODBC 驱动程序管理器的位置（Teradata ODBC 驱动程序默认安装中的 /usr/odbc/lib）。您必须重新启动 SPSS Modeler Server 以使所做更改生效。有关为 Teradata 访问提供支持的 SPSS Modeler Server 平台以及支持的 Teradata ODBC 驱动程序版本的详细信息，请访问公司支持站点 <http://www.ibm.com/support>。

在 UNIX 系统中配置 ODBC 驱动程序

默认情况下，DataDirect 驱动程序管理器尚未配置 SPSS Modeler Server 在 UNIX 中的使用。要配置 UNIX 载入 DataDirect 驱动程序管理器，输入如下命令：

```
cdmodeler_server_install_directory/bin
rm -f libspssodbc.so
ln -s libspssodbc_datadirect.so libspssodbc.so
```

此命令可删除默认链接并新建至 DataDirect 驱动程序管理器的链接。

注意：上述规则特定于数据库中的数据访问。其他文件操作类型，例如流、项目、模型、节点、PMML、输出和脚本文件的打开与保存，始终在客户端进行，并且始终在客户端计算机的文件系统条件中进行指定。此外，SPSS Modeler 中的“设置目录”命令为本地客户端对象（例如流）设置工作目录时不会影响服务器工作目录。

UNIX。关于如何配置基于 UNIX 的 SPSS Modeler Server 以采用 IBM® SPSS® Statistics 数据访问技术，请参阅[配置 UNIX 启动脚本第 61 页码](#)。

引用数据文件

Windows。如果在作为 IBM® SPSS® Modeler Server 的同一计算机上存储数据，我们建议您给出服务器计算机的数据路径（例如，C:\ServerData\Sales 1998.csv）。不使用网络查找文件时，速度更快。

如果数据存储在不同的主机，我们建议使用 UNC 文件引用（例如，\\mydataserver\ServerData\Sales 1998.csv）。请注意，仅当路径中包含共享网络资源的名称时，UNC 名称有效。引用计算机必须具有指定文件的读取权限。如果您经常在分布式和本地分析方式之间切换，请采用 UNC 文件引用，因为此种引用不考虑模式。

UNIX。要引用驻留在 UNIX 服务器中的数据文件，请采用完整文件路径和正斜杠（例如，/public/data/ServerData/Sales 1998.csv）。请避免在 SPSS Modeler Server 所用数据的文件名和 UNIX 目录中使用反斜杠字符。文本文件可以随意采用 UNIX 或 DOS 格式，因为系统将自动处理这两种格式。

IBM SPSS Statistics 数据文件的导入

如果您同时在站点上运行 IBM® SPSS® Statistics Server，则用户可能想以分布式方式导入或导出 SPSS Statistics 数据。请记住，当 IBM® SPSS® Modeler 以分布式方式运行时，此客户端提供的是服务器文件系统。SPSS Statistics 客户端以相同方式运行。如果要进行两个应用程序间的导入和导出，则客户端操作方式必须相同。如果客户端操作方式不同，则它们的文件系统视图也将不同，并且将无法共享文件。IBM® SPSS® Modeler 的 SPSS Statistics 节点可以自动启动 SPSS Statistics 客户端，但用户必须首先保证 SPSS Statistics 客户端与 SPSS Modeler 的操作方式相同。

安装说明

有关 IBM® SPSS® Modeler Server 的安装信息，请参阅 Server CD 中 \documentation\installation\ 文件夹内的说明信息。有适用于 Windows 和 UNIX 的独立文档。

有关安装和使用 IBM® SPSS® Modeler 的完整信息，请参阅客户端 CD。

IBM SPSS Modeler 支持度

此节供为 IBM® SPSS® Modeler 用户提供支持的管理人员和客户支持人员参考。本节的主题包括：

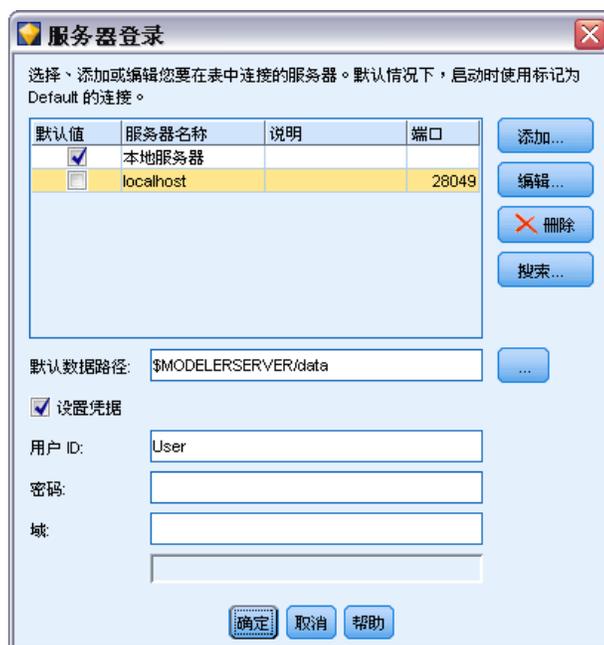
- 如何登录到 IBM® SPSS® Modeler Server（或从服务器断开，单机运行）。
- 用户可能需要的数据和文件系统
- SPSS Modeler Server 的用户帐户和文件使用权限
- 用户在 SPSS Modeler Server 和 IBM® SPSS® Modeler 之间切换时可以看到的结果差异。

连接到 IBM SPSS Modeler Server

IBM® SPSS® Modeler 可作为独立的应用程序运行，或作为直接连接到 IBM® SPSS® Modeler Server 的客户端运行，或者作为通过进程协调器（COP）插件从 IBM® SPSS® Collaboration and Deployment Services 连接到 SPSS Modeler Server 或服务器群集的客户端运行。当前连接状态显示在 SPSS Modeler 窗口的左下角。

无论何时想连接到服务器，都请手动输入想要连接的服务器的名称或选择之前已定义的名称。但是，如果您拥有 IBM SPSS Collaboration and Deployment Services，则可以从“服务器登录”对话框搜索服务器列表或服务器群集列表。可以通过进程协调器执行浏览网络上运行的 Statistics 服务的功能。[有关详细信息，请参阅第 74 页码附录 D 中的使用服务器群集负载均衡。](#)

图片 3-1
“服务器登录”对话框



连接到服务器

- ▶ 在“工具”菜单上，单击服务器登录。将打开“服务器登录”对话框。或者，双击 SPSS Modeler 窗口的连接状态区域。
- ▶ 使用该对话框指定要连接到本地服务器计算机的选项或从表中选择连接。
 - 单击添加或编辑以添加或编辑连接。有关详细信息，请参阅[添加并编辑 IBM SPSS Modeler Server 连接中的 IBM SPSS Modeler 15 用户指南](#)。
 - 单击搜索以访问进程协调器中的服务器或服务器群集。有关详细信息，请参阅[搜索 IBM SPSS Collaboration and Deployment Services 中的服务器中的 IBM SPSS Modeler 15 用户指南](#)。

服务器表。该表包含已定义的服务器连接集。该表显示默认连接、服务器名称、说明和端口号。可以手动添加新的连接，以及选择或搜索现有连接。要将特定的服务器设置为默认连接，请在表中“默认”列中为此连接选择复选框。

默认数据路径。指定用于服务器计算机上的数据的路径。单击省略号按钮 (...)，以浏览至所需要的位置。

设置凭证。不选中此复选框可启用单点登录功能，该功能尝试使您使用本地计算机用户名和密码详细信息登录服务器。如果无法使用单点登录，或您选中此复选框以禁用单点登录（例如，登录管理员帐户），则启用以下字段让您输入您的凭证。

用户 ID。输入用于登录到服务器的用户名。

密码。输入与指定用户名关联的密码。

域。指定用于登录到服务器的域。只有服务器计算机与客户计算机处于不同的 Windows 域时，才需要域名。

- ▶ 单击**确定**以完成此连接。

断开与服务器的连接

- ▶ 在“工具”菜单上，单击**服务器登录**。将打开“服务器登录”对话框。或者，双击 SPSS Modeler 窗口的连接状态区域。
- ▶ 在此对话框中，选择“本地服务器”，然后单击**确定**。

添加并编辑 IBM SPSS Modeler Server 连接

可以在“服务器登录”对话框中手动编辑或添加服务器连接。单击“添加”可以访问空的“添加/编辑服务器”对话框，在此对话框中可以输入服务器连接的详细信息。在“服务器登录”对话框中选择现有连接并单击“编辑”，将打开“添加/编辑服务器”对话框，其中包含所选连接的详细信息，以便可以进行任何更改。

注意：不能编辑从 IBM® SPSS® Collaboration and Deployment Services 中添加的服务器连接，因为名称、端口及其他详细信息已在 IBM SPSS Collaboration and Deployment Services 中做过定义。

图片 3-2
“服务器登录：添加/编辑服务器”对话框



添加服务器连接

- ▶ 在“工具”菜单上，单击**服务器登录**。将打开“服务器登录”对话框。
- ▶ 在此对话框中，单击**添加**。将打开“服务器登录：添加/编辑服务器”对话框。
- ▶ 输入服务器连接的详细信息，然后单击**确定**保存此连接并返回“服务器登录”对话框。
 - **服务器。**指定可用服务器或从列表选择一个服务器。服务器计算机的名称可以使用字母数字（例如 myserver）或指派给服务器计算机的 IP 地址（例如，202.123.456.78）。
 - **端口。**指定服务器正在侦听的端口号。如果默认设置不可用，请向系统管理员索取正确的端口号。

- **说明。** 输入此服务器连接的说明（可选）。
- **确保安全连接（使用 SSL）。** 指定是否应使用 SSL（安全套接层）连接。SSL 是常用于确保网络发送数据的安全的协议。要使用此功能，必须在承载 IBM® SPSS® Modeler Server 的服务器中启用 SSL。必要时请联系本地管理员，以了解详细信息。

编辑服务器连接

- ▶ 在“工具”菜单上，单击**服务器登录**。将打开“服务器登录”对话框。
- ▶ 在此对话框中，选择希望编辑的连接，然后单击**编辑**。将打开“服务器登录：添加/编辑服务器”对话框。
- ▶ 更改服务器连接详细信息，然后单击**确认**保存更改内容并返回至“服务器登录”对话框。

搜索 IBM SPSS Collaboration and Deployment Services 中的服务器

在 IBM® SPSS® Collaboration and Deployment Services 中，可以使用进程协调器选择网络上可用的服务器或服务器群集，从而代替手动输入服务器连接。服务器群集是一组服务器，进程协调器从这组服务器中确定最适合对处理要求作出响应的服务器。 [有关详细信息，请参阅第 74 页码附录 D 中的使用服务器群集负载均衡。](#)

尽管可在“服务器登录”对话框中手动添加服务器，但通过搜索可用的服务器，可在无需知道正确服务器名称和端口号的情况下连接到服务器。此信息是自动提供的。但仍需输入正确的登录信息，如用户名、域和密码。

注意：如果您没有访问进程协调器功能的权限，仍然可以手动输入希望连接的服务器名称或选择之前已定义的服务器名称。 [有关详细信息，请参阅添加并编辑 IBM SPSS Modeler Server 连接中的 IBM SPSS Modeler 15 用户指南。](#)

图片 3-3
“搜索服务器”对话框



搜索服务器和服务器群集

- ▶ 在“工具”菜单上，单击**服务器登录**。将打开“服务器登录”对话框。
- ▶ 在此对话框中，单击**搜索**打开“搜索服务器”对话框。如果在尝试浏览进程协调器时未登录到 IBM SPSS Collaboration and Deployment Services，则系统会提示您执行此

项操作。有关详细信息，请参阅第 9 章中的[连接至存储库中的 IBM SPSS Modeler 15 用户指南](#)。

- ▶ 从列表中选择服务器或服务器群集。
- ▶ 单击确定以关闭对话框，然后将此连接添加到“服务器登录”对话框的表中。

数据和文件系统

使用 IBM® SPSS® Modeler Server 的用户可能需要访问数据文件及网络上的其他数据源，并在网络上保存文件。用户可能需要如下信息：

- **ODBC 数据源信息。**用户访问服务器计算机上定义的 ODBC 数据源时，需要有数据源的名称、描述和登录信息（包括数据库登录 ID 和密码）。
- **数据文件访问权。**用户访问服务器计算机或网络上其他位置的数据文件时，需要有数据文件的名称和位置。
- **保存文件的位置。**用户在连接到 SPSS Modeler Server 的情况下保存数据时，可能试图在服务器计算机上保存文件。但是，服务器计算机通常是一个写保护位置。如果遇到这种情况，请告知用户应保存数据文件的位置。（一般来说，该位置为用户的主目录。）

用户验证

IBM® SPSS® Modeler Server 使用服务器计算机上的操作系统来验证连接到服务器的用户。用户连接到 SPSS Modeler Server 时，代表用户执行的所有操作均在用户的安全环境下执行。对数据库表格的访问权取决于用户和/或密码本身对数据库的使用权限。

Windows。在 Windows 上，持有主机网络有效帐户的任何用户都可以登录。使用默认验证时，用户必须具有 <modeler_server_install>\Tmp 目录的修改访问权限。如果没有这些权限，则用户使用 Windows 默认验证将无法从客户端登录 SPSS Modeler Server。

UNIX。默认情况下，假定 SPSS Modeler Server 在 UNIX 上以 root 身份运行。这允许任何在主机网络上具有有效帐户的用户登录，并将用户的文件访问权限限制为可以访问他们自己的文件和目录。但是，用户可以配置 SPSS Modeler Server，在无 root 权限的情况下运行。如果这样，则用户必须创建用于验证的专有密码数据库，且所有 IBM® SPSS® Modeler 用户共享唯一一个 UNIX 用户帐户（因而也共享数据文件的访问权）。有关详细信息，请参阅第 71 页码附录 C 中的[使用专有密码数据库以非根方式配置](#)。

在 Solaris、HP-UX、Linux 和 AIX 平台上，SPSS Modeler Server 使用 PAM 进行验证。用户需要时可使用服务名称 `modelerserver` 来配置 SPSS Modeler Server 的 PAM 模块。

文件创建

IBM® SPSS® Modeler Server 访问和处理数据时，通常必须在磁盘上保存这些数据的临时副本。临时文件使用的磁盘空间大小取决于最终用户分析的数据文件的大小和用户所执行分析的类型。有关详细信息，请参阅第 10 页码第 2 章中的[对临时磁盘空间和 RAM 的要求](#)。

UNIX。UNIX 版本的 SPSS Modeler Server 使用 UNIX `umask` 命令为临时文件设置文件权限。用户可以覆盖服务器的默认权限。有关详细信息，请参阅第 63 页码附录 B 中的控制文件创建权限。

不同的结果

在两种模式下运行分析的用户可以看到 IBM® SPSS® Modeler 和 IBM® SPSS® Modeler Server 的结果稍微有所不同。该差异的发生通常是因为记录排序和取整的不同。

记录排序。除非有流明确地分类记录和排列记录顺序，否则本地执行的流和服务器上执行的流在记录表示顺序上可能有所不同。数据库中运行操作的顺序与 SPSS Modeler Server 中运行操作的顺序可能会有差异。产生这些差异的原因是每个系统用来执行功能的不同算法可能重新排列记录顺序，比如汇总。而且请注意，在没有明确的排序操作的情况下，SQL 不会指定从数据库返回的记录顺序。

取整差异。在本机模式中运行的 IBM® SPSS® Modeler 在保存浮点值时使用与 SPSS Modeler Server 不同的内部格式。由于取整差异，各个版本之间的结果可能稍微有所不同。

IBM SPSS Modeler Server 管理

本章包含了有关如何启动和停止 IBM® SPSS® Modeler Server，如何配置各种服务器选项，以及如何解释日志文件的信息。本章还描述了如何使用 IBM® SPSS® Modeler Administration Console，该应用程序可简化服务器的配置和监视。有关此组件的安装说明，请产品附带的 SPSS Modeler Server 安装说明。

启动和停止 IBM SPSS Modeler Server

IBM® SPSS® Modeler Server 在 Windows 中作为服务运行，或在 UNIX

计划注意事项：停止 SPSS Modeler Server 会断开最终用户，终止他们的会话，因此请尽量将服务器重新启动安排在使用率较低的时段。如果做不到这点，则请务必在停止服务器之前通知用户。

在 Windows 中启动、停止和检查状态

在 Windows 上，使用 Windows 控制面板中的“服务”对话框控制 IBM® SPSS® Modeler Server。

- ▶ **Windows XP。** 打开 Windows “开始”菜单。选择设置，然后选择控制面板。双击管理工具，然后双击服务。
- ▶ **Windows 2003 或 2008。** 打开 Windows “开始”菜单。选择控制面板、管理工具，然后选择服务。
- ▶ 选择 IBM SPSS Modeler Server <nn.n> 服务。现在您可以检查其状态、启动或停止该服务，以及编辑启动参数（如果需要的话）。

默认情况下，该服务配置为自动启动，这就表示，如果您停止了该服务，则当计算机重新引导时它会自动重启。如果按照上述方式启动了该服务，则该服务以无人参与模式运行，注销服务器计算机时不会影响该服务。

在 UNIX 中启动、停止和检查状态

在 UNIX 上，可通过运行 IBM® SPSS® Modeler Server 安装目录中的 modelersrv.sh 脚本启动或停止 SPSS Modeler Server。

- ▶ 切换至 SPSS Modeler Server 安装目录。例如，在 UNIX 命令提示符下，键入

```
cd /usr/modelersrv
```

其中 modelersrv 是 SPSS Modeler Server 安装目录。

- ▶ 要启动服务器，请在命令提示符下键入

```
./modelersrv.sh start
```

- ▶ 要停止服务器，请在命令提示符下键入

```
./modelersrv.sh stop
```

- ▶ 要检查 SPSS Modeler Server 的状态，请在 UNIX 命令提示符下键入

```
./modelersrv.sh list
```

然后查看输出，该输出与 UNIX `ps` 命令生成的结果相似。列表中的第一个进程是 SPSS Modeler Server 守护进程，其他进程是 IBM® SPSS® Modeler 会话。

SPSS Modeler Server 安装程序包括一个脚本 (`auto.sh`)，该脚本会将系统配置为在引导时自动启动该服务器守护进程。如果您运行了该脚本然后停止了服务器，则该服务器守护进程会在计算机重新引导时自动重启。有关详细信息，请参阅第 61 页码附录 B 中的自动启动和停止 IBM SPSS Modeler Server。

UNIX 内核函数限制

您必须确保系统上的内核函数限制对于 SPSS Modeler Server 的操作足够。数据、内存和文件限制尤其重要，应在 SPSS Modeler Server 环境内设为无限。为此：

- ▶ 添加以下命令到 `modelersrv.sh`：

```
ulimit -d unlimited
```

```
ulimit -m unlimited
```

```
ulimit -f unlimited
```

- ▶ 重新启动 SPSS Modeler Server。

处理不响应服务器进程（UNIX 系统）

IBM® SPSS® Modeler Server 进程可能由于多种原因而不响应，包括其系统调用或 ODBC 驱动程序调用被堵塞（调用不返回或返回用时很长）的情况。UNIX 进程出现这种情况时，可使用 UNIX `kill` 命令将其清理（由 IBM® SPSS® Modeler 客户端发起的中止，或 SPSS Modeler 客户端的关闭都无效）。`kill` 命令是正常 `stop` 命令的替代，使管理员能够使用 `modelersrv.sh` 方便地发起适当的 `kill` 命令。

在容易累计无用的（“死”）服务器进程的系统上，我们建议使用以下命令序列定期停止和重启 SPSS Modeler Server：

```
cdmodeler_server_install_directory  
./modelersrv.sh stop  
./modelersrv.sh kill
```

这些使用 `modelersrv.sh kill` 命令终止的 SPSS Modeler 进程会留下临时文件（在临时目录中），需要手动删除这些文件。在某些其他情况下也会留下临时文件，包括应用程序由于资源耗尽、用户中止、系统崩溃或其他原因而崩溃。因此我们建议，作为定期重新启动 SPSS Modeler Server 过程的一部分，应当从 SPSS Modeler 临时目录中删除保留下来的所有文件。

在关闭了所有服务器进程并删除了临时文件后，就可以安全地重新启动 SPSS Modeler Server。

管理

IBM® SPSS® Modeler Server 具有若干控制其行为的配置选项。您可以通过两种方式设置这些选项：

- 使用 IBM® SPSS® Modeler Administration Console 应用程序，当前使用 IBM® SPSS® Modeler 的客户可免费使用此应用程序。 [有关详细信息，请参阅第 22 页码使用 IBM SPSS Modeler Administration Console。](#)
- 使用 options.cfg 文本文件，该文件位于 [服务器安装路径]/config 目录。 [有关详细信息，请参阅第 31 页码使用 options.cfg 文件。](#)

我们建议您安装 Modeler Administration Console 并将其用作管理工具，而不要编辑 options.cfg 文件。编辑该文件需要访问 SPSS Modeler Server 文件系统，而通过 Modeler Administration Console 则可以授予任何具有用户帐户的人员调整这些选项的权限。另外，Modeler Administration Console 还提供了服务器进程的其他信息，从而使您能够监视使用情况和性能。与编辑配置文件不同，大多数配置选项都可在不重新启动 SPSS Modeler Server 的情况下进行更改。

有关使用 Modeler Administration Console 和 options.cfg 文件的详细信息，请参阅下面的 章节。

使用 IBM SPSS Modeler Administration Console

当前的 SPSS Modeler Server 用户可以免费使用 Modeler Administration Console 应用程序，该程序可提供一个用以监控和配置 SPSS Modeler Server 安装程序的控制台用户界面。应用程序只能安装在 Windows 计算机上；但是它可以管理安装在任何受支持平台上的服务器。

Modeler Administration Console 中的许多可用选项也可在 options.cfg 文件中进行指定，此文件位于 /config 下的 SPSS Modeler Server 安装目录中。然而，Modeler Administration Console 应用程序提供了一个共享图形界面，通过此界面可以连接、配置和监视多个服务器。

启动 Modeler Administration Console

- ▶ 请依次选择 Windows “开始” 菜单上的 [所有] 程序、IBM SPSS Collaboration and Deployment Services、部署管理器、IBM SPSS Modeler Administration Console <nn.n>。

首次运行应用程序时，您将看到空的“服务器管理”和“属性”窗格（除非您已安装 Deployment Manager，并且设置了 IBM SPSS Collaboration and Deployment Services 服务器连接）。在配置 Modeler Administration Console 之后，左侧的“服务器管理员”窗格会显示您要管理的每个 SPSS Modeler Server 的节点。右边窗格则显示选定服务器的配置选项。用户必须首先为要管理的每台服务器都[创建连接](#)。

SPSS Modeler Server 连接

用户必须为网络上希望管理的每台 SPSS Modeler Server 指定连接。然后必须登录每台服务器。虽然整个 Modeler Administration Console 会话过程中服务器连接设置都将被记住，但不会记住登录凭证。每次启动 Modeler Administration Console 时都必须登录。

创建服务器连接

- ▶ 确保 IBM® SPSS® Modeler Server 服务已启动。
- ▶ 选择“文件”菜单上的新建，然后选择管理的服务器连接。
- ▶ 在向导首页，输入服务器连接的名称。名称供用户自己使用，应该是描述性质的名称，例如，生产服务器。确保已将“类型”设为管理的 IBM SPSS Modeler Server，然后单击下一步。
- ▶ 在第 2 页，输入服务器的主机名和 IP 地址。如果更改了默认的端口，则请输入端口号。单击完成。新服务器连接即显示在“服务器管理员”窗格中。

要执行管理任务，现在必须登录。

登录到 Server

- ▶ 在“服务器管理员”窗格中双击以选择您要登录的服务器。
- ▶ 在“登录”对话框中输入用户凭证。（使用服务器主机的用户帐号。）单击确定。

如果登录失败并显示无法获取服务器上的管理员权限的消息，最有可能的原因是未正确配置管理员访问权限。[有关详细信息，请参阅第 23 页码使用 Modeler Administration Console 配置访问权限。](#)

如果登录失败，并显示消息无法连接到服务器‘<服务器>’，请确保用户 ID 和密码正确，而且 SPSS Modeler Server 服务正在运行。例如，在 Windows 上，前往控制面板 > 管理工具 > 服务，然后选中 IBM SPSS Modeler Server 条目。如果“状态”列未显示已启动，则在屏幕上选择此行，单击启动，然后重试登录。

登录到 SPSS Modeler Server 后，可以看到服务器名称下方显示的两个选项，即[配置](#)和[监视](#)。双击其中一个选项。

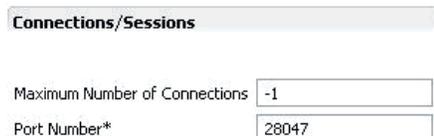
SPSS Modeler Server 配置

配置窗格显示 SPSS Modeler Server 的配置选项。请根据需要更改选项。单击工具栏上的保存，以保存更改。（注意：更改端口号需要重启服务器才能生效。）

以下说明这些配置选项并在括号中给出每个选项在 options.cfg 中对应的行。本节的结尾处描述了仅在 options.cfg 中可见的选项。

连接/会话

图片 4-2
连接/会话设置



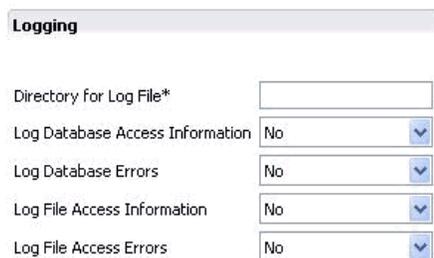
Connections/Sessions	
Maximum Number of Connections	-1
Port Number*	28047

最大连接数。 (max_sessions) 同时连接用户的最大数量。值为 -1 则表示无限制。

端口号。 (port_number) SPSS Modeler Server 侦听的端口号。如果默认端口 (28047) 已被其他应用程序使用，则请更改端口号。最终用户必须知晓端口号，才能使用 SPSS Modeler Server。

日志记录

图片 4-3
日志设置



Logging	
Directory for Log File*	
Log Database Access Information	No
Log Database Errors	No
Log File Access Information	No
Log File Access Errors	No

日志文件的目录。 (log_directory) messages.log 文件的位置。空字符串表示[服务器安装路径]/log。注意：必须使用正斜线。要了解关于此日志文件的内容，请参阅 [服务器日志第 39 页码](#)。

在日志中记录数据库访问信息。 (log_database_access) 将数据库访问信息添加到日志文件。

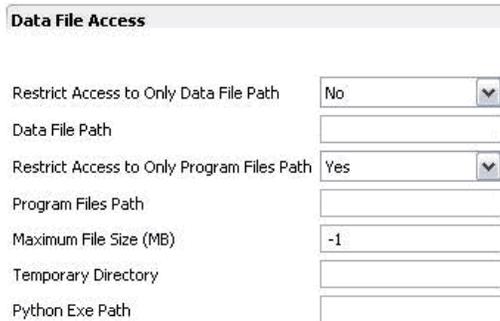
在日志中记录数据库错误。 (log_database_errors) 将数据库错误添加到日志文件。

在日志中记录文件访问信息。 (log_file_access) 将文件访问信息添加到日志文件。

在日志中记录文件访问错误。 (log_file_errors) 将文件访问错误添加到日志文件。

数据文件访问权限

图片 4-4
数据文件访问权设置



Data File Access	
Restrict Access to Only Data File Path	No
Data File Path	
Restrict Access to Only Program Files Path	Yes
Program Files Path	
Maximum File Size (MB)	-1
Temporary Directory	
Python Exe Path	

限制仅能访问数据文件路径。 (`data_files_restricted`) 设置为是时，此选项将限制只能访问标准数据路径和以下数据文件路径中列出的数据文件。

默认数据路径。 (`data_file_path`) 允许用户读和写数据文件的其他目录的列表。除非启用限制仅能访问数据文件路径选择，否则此选项将被忽略。注意：所有路径名中应使用正斜线。指定多个目录时请使用半角分号隔开（例如，`[server install path]/data;c:/data;c:/temp`）

限制仅能访问程序文件路径。 (`program_files_restricted`) 设置为是时，此选项将限制只能访问标准 `bin` 路径和以下程序文件路径中列出的程序文件。

程序文件路径。 (`program_file_path`) 允许用户执行程序的其他目录的列表。除非启用限制仅能访问程序文件路径选择，否则此选项将被忽略。注意：所有路径名中应使用正斜线。指定多个目录时请使用半角分号隔开。

最大文件大小 (MB)。 (`max_file_size`) 流执行期间创建的临时和导出数据文件的最大大小（不适用于 SAS 和 SPSS Statistics 数据文件）。值为 `-1` 则表示无限制。

临时目录。 (`temp_directory`) 该目录用于存储临时数据文件（缓存文件）。在理想情况下，该目录应当在独立的高速驱动器或控制器上，因为对该目录的访问速度对性能可能有显著影响。可以指定多个临时目录，并分别用逗号隔开。这些目录应当位于不同的磁盘，其中第一个目录使用最频繁，而其他目录则在执行过程中特定数据准备操作（例如排序）平行应用时，用于存储临时工作文件。允许每个执行线程使用独立磁盘进行临时存储能够提高性能。在所有路径指定中均应使用正斜线。注意：在 SPSS Modeler Server 启动期间会在此目录中生成临时文件。确保您对该目录拥有所需的访问权限（例如，如果临时目录为共享网络文件夹），否则 SPSS Modeler Server 启动将失败。

Python 执行路径。 (`python_exe_path`) Python 可执行程序的完整路径（包括可执行程序的名称）。注意，根据 Python 安装的位置，可能需要将限制仅能访问程序文件路径设置为否。

性能/优化

图片 4-5
性能/优化设置

Performance/Optimization	
Stream Rewriting	Yes
Parallelism	-1
Buffer Size	32768
Cache Compression	1
Memory Usage Multiplier	100
Modelling Memory Limit Percentage	25
Allow Modelling Memory Override	Yes
Minimum Server Port	0
Maximum Server Port	0
Array Fetch Optimization	1

流重写。 (`stream_rewriting_enabled`) 允许服务器通过重写流来实现对流的优化。例如，服务器可能会在接近源节点的位置进行数据削减操作以尽早实现数据集规模的最小化。通常仅在此优化导致错误或其他不应有的结果时，才建议禁用此选项。此设置将覆盖相应的客户端的优化设置。

并行性。 (`max_parallelism`) 说明运行流时，允许 SPSS Modeler 使用的并行工作程序的线程数量。如果设置为 0 或负数，则 IBM® SPSS® Modeler 将认定允许使用的线程数为计算机上可用处理器的数量；此选项的默认值为 -1。要关闭（针对多处理器的计算机）并行处理，请将此选项设置为 1。要限制并行处理线程数，请将此选项设置为比计算机上的处理器数量小的数字。注意：超线程或双核处理器将被当作两个处理器。

缓冲区大小（字节）。 (`io_buffer_size`) 从服务器传输到客户机的数据文件通过该字节数的缓冲区进行传递。

缓存压缩。 (`cache_compression`) 一个 0 到 9 的整数值，它控制缓存和服务临时目录中其他文件的压缩。压缩在磁盘空间有限时很重要，因为压缩可以降低磁盘空间的使用量，而且压缩还能通过减少创建和读取缓存所需要的磁盘活动量，改善性能。虽然压缩会延长处理器时间，但因为缩短了磁盘访问时间，因而得到几乎等量的补偿。注意：只有某些被顺序访问的缓存才能被压缩。此选项不适用于随机访问缓存，比如网络训练算法使用的缓存。值为 0 将完全禁用压缩。值从 1 增大，压缩程度会随之增加，但相应的访问时间也会增加。默认值为 1；除非磁盘空间实在有限，否则通常不需要比默认值更大的值。

内存使用倍增器。 (`memory_usage`) 控制为排序和其他内存缓存分配的物理内存的比例。默认值为 100，这相当于大约 10% 的物理内存。有可用的空闲内存时，增加该值可提高排序性能，但要注意，如果该值增加得太高则有可能导致过度分页。

建模内存限制百分比。 (`modelling_memory_limit_percentage`) 控制为训练神经网络、Kohonen 和 k-means 模型分配的物理内存的比例。默认值为 25%。有可用的空闲内存时，增加该值可提高训练性能，但要注意，如果该值增加得太高则有可能在数据溢出到磁盘时导致过度分页。

允许覆盖建模内存。 (`allow_modelling_memory_override`) 在特定建模节点中启用或禁用优化速度选项。默认为已启用。此选项使建模算法能够要求获得所有可用的内存，而忽略百分比限制选项。如果需要在服务器计算机上共享内存资源，则可能希望禁用此选项。

最大和最小服务器端口。 (`max_server_port` 和 `min_server_port`) 指定端口号范围，该端口号用于互动模型和流执行所需要的客户机和服务器之间的其他套接字连接。这些套接字连接要求服务器侦听其他端口；未限制范围将对系统带有防火墙的用户造成使用问题。两者的默认值均为 -1，表示“无限制”。例如，要通过设置使服务器侦听端口 8000 或数值更大的端口，必须将 `min_server_port` 设置为 8000，将 `max_server_port` 设置为 -1。

注意，您必须打开主服务器端口上的附加端口，以便打开或执行流，如果要打开或执行并行的流，还应打开对应的更多端口。这需要用于捕获流执行的反馈。

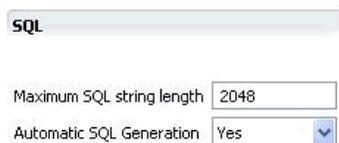
默认情况下，SPSS Modeler 将使用任何一个可用的打开端口；如果它未找到端口（例如，所有的端口都被防火墙关闭），那么执行流时就会显示错误信息。要配置端口的范围，除主服务器端口外，SPSS Modeler 还需要两个每个并行的流都能使用的打开端口，以及为每个具有 ODBC 连接的连接客户端打开一个附加的端口。如果要执行更多并行的流，您就可以增大端口号的范围，然后相应地在防火墙上打开更多的端口。

注意：如果您更改了这些参数，就需要重新启动 SPSS Modeler Server 以使所做更改生效。

数组获取优化。 (`sql_row_array_size`) 控制 SPSS Modeler Server 从 ODBC 数据源获取数据的方法。默认值为 1，表示一次获取一行。增大此值可使服务器读取更大数据块的信息，即获取指定数量的行并放入数组。对于某些操作系统/数据库组合，这样做可以提高 SELECT 语句的性能。

SQL

图片 4-6
SQL 设置



最大 SQL 字符串长度。 (`max_sql_string_length`) 对于用 SQL 从数据库导入的字符串，保证成功传递的最大字符数。根据具体的操作系统，大于此长度的字符串可能会在没有警告的情况下从右截断。有效范围在 1 和 65,535 个字符之间。

自动 SQL 生成。 (`sql_generation_enabled`) 允许为流自动生成 SQL，这可以极大提高性能。默认为已启用。仅在数据库不能支持由 SPSS Modeler Server 提交的查询时建议禁用此选项。注意，此设置将覆盖相应的客户端的优化设置；同时注意，因为评分的目的，必须分别为每个建模节点启用 SQL 生成，而忽略此设置。

SSL

图片 4-7
SSL 设置

启用 SSL。 (`ssl_enabled`) 为 SPSS Modeler 和 SPSS Modeler Server 之间的连接启用 SSL 加密。

证书文件。 (`ssl_certificate_file`) 服务器启动时要加载的公共 SSL 证书文件。

私有密钥文件。 (`ssl_private_key_file`) 服务器启动时要加载的私有 SSL 密钥文件。

私有密钥密码。 (`ssl_private_key_password`) 指定证书的密码密钥。注意，此密钥必须经过至少一个证书授权机构（列于配置的证书文件中）签署。

进程协调器配置

图片 4-8
进程协调器配置设置

Items marked (*) require a server restart to take effect.

主机。 (`cop_host`) 进程协调器服务的主机名或 IP 地址。默认的“spsscop”是一个无效名称，管理员可以选择添加它作为 DNS 中 IBM SPSS Collaboration and Deployment Services 主机的别名。

端口号。 (`cop_port_number`) 进程协调器服务的端口号。默认为 8080，也是 IBM SPSS Collaboration and Deployment Services 的默认值。

登录名。 (`cop_user_name`) 对进程协调器服务进行身份验证的用户名。它是 IBM SPSS Collaboration and Deployment Services 登录名，所以可包括一个安全提供者的前缀（例如：ad/jsmith）。

密码。 (`cop_password`) 用于对进程协调器服务进行身份验证的加密密码。

要创建密码文件，需要使用密码实用程序 `pwutil`，该程序位于 SPSS Modeler Server 安装程序的分级目录中。

已启用。 (`cop_enabled`) 确定服务器是否应尝试注册进程协调器。默认为不注册，这是因为管理员应选择通过进程协调器公开哪些服务。

服务器名称。 (`cop_service_name`) 此 SPSS Modeler Server 实例的名称；默认为主机名。

描述。 (`cop_service_description`) 此实例的说明。

更新的时间间隔（分钟）。 (`cop_update_interval`) 保持活动消息的分钟数；默认值为 2。

权重。 (`cop_service_weight`) 该实例的权重可指定为 1 和 10 之间的整数。权重越高，所吸引的连接越多。默认值为 1。

服务主机。 (`cop_service_host`) IBM® SPSS® Modeler Server 主机的完全限定主机名。默认的主机名可自动获取，管理员可以覆盖多宿主机的 hostname。

仅在 `options.cfg` 中可见的选项。

administrators。 指定您要授予管理员访问权限的那些用户的用户名。 [有关详细信息，请参阅第 23 页码使用 Modeler Administration Console 配置访问权限。](#)

allow_config_custom_overrides。 除非得到技术支持代表的指示，否则不要修改。

db_udf_enabled。 如设为 Y（默认值），则导致 SQL 生成选项产生用户定义函数（UDF）SQL 而不是纯 SPSS Modeler SQL。UDF SQL 的性能通常优于纯 SQL。

default_sql_string_length。 指定将在数据库缓存表中创建的字符串列的默认宽度。在数据库缓存表中创建字符串字段时，如果没有上游类型信息，则采用 255 的默认宽度。如果在数据中有更宽的值，则可以使用这些值实例化上游类型节点，或将此参数设置为能够适应这些字符串值的足够大的值。

max_transfer_size。 仅供内部使用；不得修改。

shell。（仅 UNIX 服务器）覆盖 UNIX shell 的默认设置，例如 `shell`，`"/usr/bin/ksh"`。默认情况下，IBM® SPSS® Modeler 使用在当前连接到 IBM® SPSS® Modeler Server 的用户的配置文件中定义的 shell。

start_process_as_login_user。 如果使用单点登录，并且连接到采用不同于 SPSS Modeler Server 主机的验证方法的 IBM® SPSS® Collaboration and Deployment Services 服务器，则将此设为 Y。如果您在 UNIX 上以非根进程方式运行 SPSS Modeler Server，且未使用专有密码数据库，也可将此设置为 Y。

supported_sql_type_logging。 仅用于调试目的；除非得到技术支持代表的指示，否则不要修改。

time_code_execution_log。如果您希望 message.log 文件（在 SPSS Modeler Server 安装目录的 log 文件夹中）包含流中每个节点各个执行时间，可将此设置为 Y。

SPSS Modeler Server 监视

Modeler Administration Console 的监视窗格用一个快照来显示在 SPSS Modeler Server 计算机上运行的所有进程，就跟 Windows 任务管理器一样。要激活监视窗格，请双击“服务器管理员”窗格中所需服务器下方的“监视”节点。这样就向窗格填充了来自服务器的当前数据快照。数据将按所显示的速率（默认为一分钟一次）进行刷新。要手动刷新数据，请单击刷新。要在此列表中仅显示 SPSS Modeler Server 进程，请单击过滤掉非 SPSS Modeler 进程按钮。

图片 4-9

IBM SPSS Modeler Administration Console 监控窗格

Executable	Memory	CPU	User Name
FrameworkService	13,036 K	00	SYSTEM
LocalSch	1,640 K	00	SYSTEM
Mcshield	23,364 K	00	SYSTEM
SoftMon	1,632 K	00	SYSTEM
System	176 K	00	SYSTEM
UpdaterUI	2,156 K	00	cstrahs
VsTskMgr	464 K	00	SYSTEM
clemserv	10,452 K	00	SYSTEM
clemserv	22,312 K	00	SYSTEM

使用 options.cfg 文件

options.cfg 文件位于 [服务器安装路径]/config 目录下。每个设置都由一个逗号分隔的名称-值对表示，其中名称是该选项的名称，值是该选项的值。井字符（#）表示注解。

注意：大多数配置选项均可使用 IBM® SPSS® Modeler Administration Console 来进行更改，而不必使用此配置文件，但存在一些例外。有关详细信息，请参阅第 30 页码仅在 options.cfg 中可见的选项。。

通过使用 Modeler Administration Console，不用重新启动服务器即可使得所有选项生效，但服务器端口除外。有关详细信息，请参阅第 22 页码使用 IBM SPSS Modeler Administration Console。

可添加到默认文件的配置选项

默认情况下，数据库内高速缓存通过 IBM® SPSS® Modeler Server 启用。要禁用此功能，可将以下行添加到 options.cfg 文件中。

```
enable_database_caching, N
```

这将导致在服务器上而不是数据库中创建临时文件。

查看或更改 IBM SPSS Modeler Server 配置选项

- ▶ 使用文本编辑器打开 options.cfg 文件。
- ▶ 找到相关的选项。有关选项的完整列表，请参阅[SPSS Modeler Server 配置第 24 页码](#)。
- ▶ 根据需要编辑这些值。请注意，所有路径名值都必须使用正斜线 (/) 作为路径名分隔符，而不要使用反斜线。
- ▶ 保存该文件。
- ▶ 停止并重新启动 SPSS Modeler Server 以使得这些更改生效。 [有关详细信息，请参阅第 20 页码启动和停止 IBM SPSS Modeler Server。](#)

关闭不使用的数据库连接

默认情况下，在访问一个连接时，IBM® SPSS® Modeler 至少会将该连接缓存入数据库。即使不执行需要数据库访问的流，数据库会话也会保持打开。

缓存数据库连接可使每次执行流时不再需要 SPSS Modeler 重新连接到数据库，可增加执行次数。然而，在某些环境中，应用程序应尽可能快地释放数据库资源。如果过多的 SPSS Modeler 会话保持到数据库的连接且不再被使用，可能会耗尽数据库资源。

您可以在自定义数据库配置文件中通过关闭 SPSS Modeler 选项 `cache_connection` 来避免这种情况。这也使得 SPSS Modeler 能更快修复因 SPSS Modeler 会话长期使用连接而导致的数据库连接故障（如超时）。

要关闭不使用的数据库连接：

- ▶ 找到 [server install path]/config 目录。
- ▶ 添加以下文件（如已存在，则打开该文件）：
odbc-custom-properties.cfg
- ▶ 添加以下行到文件：
`cache_connection, N`
- ▶ 保存并关闭该文件。
- ▶ 重新启动 IBM® SPSS® Modeler Server 使更改生效。

IBM SPSS Statistics 许可证位置

要启用 IBM® SPSS® Modeler 以使用 Statistics 变换、Statistics 模型和 Statistics 输出节点，必须在执行流的计算机上拥有 IBM® SPSS® Statistics 安装和许可的一个副本。依靠服务器运行时，必须在该台服务器上完成许可证配置。

Windows

- ▶ 在命令提示符下，切换至 IBM® SPSS® Modeler Server bin 目录，然后运行：

```
statisticsutility -location =<path to IBM SPSS Statistics Server license file>
```

其中 <path to SPSS Statistics Server license file> 是经许可的 SPSS Statistics 服务器的安装目录。

UNIX

- ▶ 在命令提示符下，切换至 SPSS Modeler Server bin 目录，然后运行：

```
./statisticsutility -location =<path to IBM SPSS Statistics Server license file>
```

其中 <path to SPSS Statistics Server license file> 是经许可的 SPSS Statistics 服务器的安装目录。

使用 SSL 进行安全数据传输

安全套接字层（SSL）是对两台计算机之间传输的数据进行加密的协议。SSL 可以确保计算机之间的通信安全。SSL 可以对用户名/密码的身份验证进行加密，还可以对服务器与客户端之间的交换内容进行加密。

SSL 如何运行

SSL 依赖于服务器的公钥和私钥，并且公钥证书将服务器身份绑定在其公钥上。

- ▶ 当客户端连接到服务器时，客户端使用公钥证书对服务器进行身份验证。
- ▶ 然后客户端生成随机数，使用服务器的公钥对该数字进行加密，并将加密后的消息发回服务器。
- ▶ 服务器使用其私钥对随机数进行解密。
- ▶ 服务器和客户端根据此随机数创建会话密钥，用以加密和解密后续传输的信息。

公钥证书通常由证书授权机构签署。证书授权机构，如 VeriSign 和 Thawte 等组织，专门负责签发、验证和管理公钥证书中所包含的安全凭证。实际上，证书授权机构确认了服务器的身份。证书授权机构通常会收取证书费用，但用户也可以自行生成自签署证书。

通过 SSL 确保客户端-服务器和服务器-客户端的通信安全

通过 SSL 确保客户端-服务器和服务器-客户端通信安全的主要步骤包括：

- ▶ 在服务器计算机上安装 OpenSSL。
- ▶ 获取并安装 SSL 证书和密钥。
- ▶ 在服务器管理应用程序（IBM® SPSS® Collaboration and Deployment Services Deployment Manager）中启用和配置 SSL。
- ▶ 如果需要，在客户机上安装无限强度的加密。
- ▶ 如果使用自签署证书，则将此证书复制到客户端计算机上。

- ▶ 将证书添加到客户端的密钥库中。
- ▶ 指导最终用户在连接到服务器时启用 SSL。

注意：有时服务器产品会作为客户端运行。例如当 IBM® SPSS® Statistics Server 连接到 IBM® SPSS® Collaboration and Deployment Services Repository 时。这时，SPSS Statistics Server 是作为客户端。

安装 OpenSSL

如果服务器上尚未安装 OpenSSL，则必须进行安装。

- ▶ 请从 <http://www.openssl.org/> 下载 OpenSSL。确保针对服务器版本使用适当的 OpenSSL 版本。

表 4-1
OpenSSL 版本

服务器产品	兼容的 OpenSSL 版本
IBM® SPSS® Statistics 20	1.0.0
SPSS Statistics 17-19（不是 System z® 上的 Linux®） IBM® SPSS® Modeler 15（不是 System z® 上的 Linux®）	0.9.8 及其后续版本（0.9.8a、0.9.8b 等等）
SPSS Statistics 19（System z® 上的 Linux®） SPSS Modeler 15（System z® 上的 Linux®）	1.0.0

注意：在 IBM® SPSS® Modeler Server for UNIX 被配置为使用 SSL 时，它会寻求链接具有以下名称的两个库（.so 可能与平台相关）：

- libssl.so.0.9.8
- libcrypto.so.0.9.8

如果使用 SSL 的子版本（0.9.8a、0.9.8b 等），则需要使用 SPSS Modeler Server 要求的 xxx.0.9.8 版本来创建这些库的符号链接。

- ▶ 按照说明安装和配置软件。如果自行构建 OpenSSL，请注意遵循以下原则：

Windows。 OpenSSL 应随 DLL 一起构建（默认为多线程）。

UNIX。 OpenSSL 应当支持多线程（并不总是默认选项）和共享库。

- ▶ 确保系统可以找到 SSL 库：

Windows。 确保 OpenSSL 模块包含在系统路径中。

UNIX。 将 SSL 库追加到在 modelersrv.sh 中定义的加载库路径末尾。例如，在 Solaris 上，可在服务器启动脚本中添加此行：

```
LD_LIBRARY_PATH_64=$LD_LIBRARY_PATH_64:/usr/local/openssl; export LD_LIBRARY_PATH_64
```

获取并安装 SSL 证书和密钥。

- ▶ 获取 SSL 证书和密钥文件。有两种方法可以解决这一问题：
 - 从公共证书授权机构（如 Verisign 或 Thwaite）购买。公共证书授权机构签署证书以验证使用它的服务器。
 - 使用内部自签署证书颁发程序生成密钥和证书文件。OpenSSL 为此提供了证书管理工具，或者您可搜索 Internet 以了解有关创建自签署 SSL 证书的说明。
- ▶ 将证书和密钥文件复制到服务器上的本地目录。公钥和私钥可以存储在同一文件中。确保将访问权限设置为拒绝随意浏览此目录。
- ▶ 在应用服务器上安装 SSL 证书和密钥。有关密钥和证书与特定应用服务器之间的互操作信息，请参阅原始提供商的文档。注意，您可能需要将证书和密钥添加到 Java 密钥库中。

在以下中启用和配置 SSL： IBM SPSS Collaboration and Deployment Services Deployment Manager

- ▶ 启动服务器管理应用程序（IBM® SPSS® Collaboration and Deployment Services Deployment Manager）并连接到服务器。
 - ▶ 在配置页上，将安全套接字层设为是。
 - ▶ 在 SSL 公钥文件中，指定公钥文件的完整路径。
 - ▶ 在 SSL 私钥文件中，指定私钥文件的完整路径。
- 注意：如果公钥和私钥存储在同一文件中，请在 SSL 公钥文件和 SSL 私钥文件中指定相同的文件。
- ▶ 从菜单中选择：
 - 文件 > 保存
 - ▶ 重新启动服务器服务或守护进程。在重新启动时，将提示您输入 SSL 密码。在 Windows 上，可以选择记住该密码以安全地存储密码。选择此选项，无需每次启动服务器时都输入密码。

在 IBM SPSS Modeler Server 中设置 SSL

如果需要密码来访问 SSL 私钥文件，则密码的提供方式取决于具体的操作系统。

Windows 系统

在 Windows 系统上，您可以通过以下方式之一来提供密码：

- 通过使用 `ssl_private_key_password` 选项在 `options.cfg` 文件中指定。 [有关详细信息，请参阅第 29 页码 SSL。](#)
- 通过提示窗口来指定密码。在 `options.cfg` 中忽略 `ssl_private_key_password` 选项，将导致在启动时显示提示窗口。

UNIX 系统

在 UNIX 系统上，您可以通过以下方式之一来提供密码：

- 通过使用 `ssl_private_key_password` 选项在 `options.cfg` 文件中指定。 [有关详细信息，请参阅第 29 页码SSL。](#)
- 添加开关到 `modelersrv.sh` 文件以在开始时提示输入密码。

如果使用后一种选项，在 `modelersrv.sh` 文件中找到以下行：

```
if "$INSTALLEDPATH/$SCLEMDNAME" -server $ARGS; then
```

按以下添加 `-request_ssl_password` 开关：

```
if "$INSTALLEDPATH/$SCLEMDNAME" -request_ssl_password -server $ARGS; then
```

安装无限强度的加密

随产品发售的 Java Runtime Environment 启用了 US 出口强度的加密。为使您的数据获得增强的安全性，我们建议升级到无限强度的加密。

IBM J9

- ▶ 从 IBM.com Web 站点下载与您的 SDK 版本对应的 Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files。
- ▶ 提取打包在压缩文件中的无政策限制权限文件。压缩文件包含一个 `US_export_policy.jar` 文件和一个 `local_policy.jar` 文件。在您的 WebSphere Application Server 安装目录中，前往 `$JAVA_HOME/jre/lib/security` 目录，然后备份您的 `US_export_policy.jar` 和 `local_policy.jar` 文件。
- ▶ 用您下载的两个文件替换此处的 `US_export_policy.jar` 和 `local_policy.jar` 文件。
- ▶ 启用 WebSphere Application Server 管理控制台中的安全性。请确保单元中的所有节点代理都已预先启动。有关详细信息，请参阅 WebSphere 文档。请注意，您必须从安全性 > 安全管理、应用程序和基础结构中的列表选择一个可用的领域定义，然后单击设为当前，即可在服务器重启时启用安全性。
- ▶ 注销管理控制台。
- ▶ 停止服务器。
- ▶ 重启服务器。

Sun Java

- ▶ 从 Sun Java Web 站点下载与您的 SDK 版本对应的 Java Cryptography Extension (JCE) Unlimited Strength Jurisdiction Policy Files。
- ▶ 对下载的文件解压缩。
- ▶ 将两个 `.jar` 文件 `local_policy.jar` 和 `US_export_policy.jar` 复制到 `<installation folder>/jre/lib/security`，其中的 `<installation folder>` 为产品安装目录。

将证书添加到客户端密钥库（适合连接到存储库 或 IBM SPSS Modeler Server）

注意：如果使用由证书授权机构签署的证书，请跳过此步骤。

如果使用 SSL 连接到 IBM® SPSS® Collaboration and Deployment Services 存储库 或 IBM® SPSS® Modeler Server，并且使用了自签署证书，则需要将证书添加到客户端的 Java 密钥库中。在客户机上完成下列步骤。

注意，服务器计算机也可能会作为客户端。例如当 IBM® SPSS® Statistics Server 连接到 IBM® SPSS® Collaboration and Deployment Services Repository 时。在此情况下，SPSS Statistics Server 为客户端，因此需要将 IBM SPSS Collaboration and Deployment Services Repository 服务器的证书复制到 SPSS Statistics Server 上。

- ▶ 从 SSL 服务器将证书 (.cer) 文件复制到客户机上的 <install_dir>\jre\lib\security 目录中。
- ▶ 在该目录，右键单击 cacerts 文件并选择属性。
- ▶ 取消选中只读复选框。
- ▶ 打开命令提示符，将目录切换至以下位置，其中 <product install dir> 是您的产品安装目录：

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

- ▶ 输入下面的命令：

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

其中 <alias name> 是证书的专有别名，<path to cert> 是证书的完整路径，<path to key store> 是 Java 密钥库的完整路径（可能为 <product install dir>/lib/security/jssecacerts 或 <product install dir>/lib/security/cacerts）。

- ▶ 提示时，输入密钥库密码，默认为 **changeit**。
- ▶ 当提示是否信任证书时，输入 **是**。

指导最终用户启用 SSL

当最终用户通过客户端产品连接到服务器时，他们需要在服务器连接对话框中启用 SSL。确保告诉他们选择相应的复选框。例如，在 IBM® SPSS® Modeler 中，用户应选中存储库上的确保安全连接复选框：从“工具”菜单上的“存储卡”选项访问“服务器”对话框。

仅 URL 前缀配置

如果设置 IBM® SPSS® Collaboration and Deployment Services Repository 以使用 SSL 访问，则必须修改“URL 前缀”配置设置的值如下：

1. 使用基于浏览器的控制台登录到存储库。
2. 打开 URL 前缀配置选项。

配置 > 设置 > URL 前缀

3. 将前缀值设置为 `https` 以取代 `http`，并将端口值设置为 SSL 端口号。例如：

```
[default]
http://<hostname>:<port>
[SSL-enabled]
https://<hostname>:<SSLport>
```

仅通过 SSL 实现安全 LDAP

轻量目录访问协议 (LDAP) 是一项有关包含任意级别信息的网络目录和数据库间信息交换的互联网工程任务编组 (IETF) 标准。对于需要附加安全性的系统，LDAP 提供程序（如 Microsoft 的 Active Directory）可以运行在安全套接字层 (SSL) 上，前提是 Web 或应用服务器支持 SSL 上的 LDAP。将 SSL 与 LDAP 配合使用，可以确保登录密码、应用程序信息和其他敏感数据不会遭受黑客攻击、威胁或窃取。

下例示例说明了如何使用 Microsoft 的 Active Directory 作为安全提供程序来启用 LDAP。有关任何步骤的更多信息，或打算了解安全提供程序的特定版本细节，请参阅原始提供商的文档。

1. 验证 Active Directory 和 Enterprise Certificate Authority 是否已安装且正常工作。
2. 使用证书颁发程序生成证书，并将证书导入 IBM® SPSS® Collaboration and Deployment Services Deployment Manager 安装的证书库中。这样允许在 IBM® SPSS® Collaboration and Deployment Services Repository 与 Active Directory 服务器之间建立 LDAPS 连接。

要配置 Deployment Manager 进行安全的 Active Directory 连接，请检查确保存在到存储库的连接。

3. 启动 IBM® SPSS® Collaboration and Deployment Services Deployment Manager。
4. 从“工具”菜单选择服务器管理。
5. 登录到管理的服务器（经预先定义）。
6. 双击服务器的配置图标以展开层次结构。
7. 双击安全提供程序图标以展开层次结构。
8. 双击 Active Directory 安全提供程序。
9. 为已安装安全证书的 Active Directory 实例输入配置值。
10. 选择使用 SSL 复选框。
11. 注意“域用户”字段中的名称。使用 Active Directory 的后续登录将通过 SSL 进行身份验证。

有关在特定应用服务器上安装、配置和实施 LDAPS 的更多信息，请参阅原始提供商的文档。

服务器日志

IBM® SPSS® Modeler Server 在一个名为 messages.log 的日志文件中保存着其重要操作的记录。默认情况下，此文件位于 [服务器安装路径]/log，但该位置可以配置。启动 SPSS Modeler Server 后，该日志文件将进行更新，以表明启动的时间和它正在侦听的端口号。从此时开始，每当客户端尝试连接服务器时该日志文件都会不断更新连接结果。

默认情况下服务器向日志文件写入下列事件：

- 服务器启动
- 服务器停止
- 连接已接受
- 连接已拒绝
- 会话启动
- 会话停止
- 系统错误
- 登录成功
- 登录失败

如果您希望记录数据库访问事件的相关信息，则请将 SPSS Modeler Server 配置中的日志数据库访问信息和日志数据库错误设置为是。将记录下列数据库访问事件：

- 数据库登录成功
- 数据库登录失败
- 数据库注销

如果您希望记录文件访问事件的相关信息，则请将 SPSS Modeler Server 配置中的日志文件访问信息和日志文件访问错误设置为是。将记录下列文件访问事件：

- 文件打开/创建
- 文件关闭

下面是一个日志文件示例（添加了行号）：

```
1. 2006/06/14 08:41:23 [167]: IBM SPSS Modeler Server Version 0.1.0 started
2. 2006/06/14 08:41:23 [167]:Listening on port 28028
3. 2006/06/14 08:49:28 [167]:Accepted connection from localhost:1827
4. 2006/06/14 08:49:28 [167]:Session 105 (localhost:1827) started
5. 2006/06/14 08:49:29 [167-105]:Login failed for user:SPSS\awilson
6. 2006/06/14 09:27:27 [167]:Session 105 ended
7. 2006/06/14 09:53:39 [167]:Accepted connection from localhost:1837
8. 2006/06/14 09:53:39 [167]:Session 63 (localhost:1837) started
9. 2006/06/14 09:53:40 [167-63]:Login succeeded for user:SPSS\awilson
10. 2006/06/14 10:27:31 [167]:Session 63 ended
```

日志文件中的所有项均包含时间戳。启动 SPSS Modeler Server 时日志文件中添加了第 1 行和第 2 行。当计算机 localhost 上的用户尝试在端口号 1827 连接服务器时，添加了第 3 行。第 4 行表示此用户启动了一个服务器会话。第 5 行显示该用户未能正确登

录。一段时间之后，该用户终止了该会话。半小时后，尝试了另一个来自 localhost 的连接，并启动了一个会话。这一次登录成功。又过了半小时后，该会话终止，连接关闭。

性能概述

分析数据时的真正性能受若干因素的影响，这些因素包括服务器和数据库配置，以及流中单个节点的顺序等。一般来说，您可以通过执行下列操作获取最佳的性能：

- 在 DBMS 中存储数据，尽可能多地使用 SQL 生成和优化。
- 使用满足或超过 [第 2 章](#) 中所给出推荐条件的硬件。
- 确保正确配置了客户端和服务器的性能和优化设置。注意，当 SPSS Modeler 连接到 SPSS Modeler Server 安装时，服务器性能和优化设置将覆盖客户端的相应设置。
- 设计流，以获得最大的性能。有关详细信息，请参阅 [第 13 章](#) 中的流和节点的性能注意事项中的 IBM SPSS Modeler 15 用户指南。

有关每个性能因素的详细信息，请参阅下面的章节。

服务器性能和优化设置

某些 IBM® SPSS® Modeler Server 设置可以进行配置，以优化性能。您可以使用 SPSS Modeler Server 随附的 IBM® SPSS® Modeler Administration Console 应用程序调整这些设置。有关详细信息，请参阅 [第 22 页码第 4 章](#) 中的使用 IBM SPSS Modeler Administration Console。

这些设置组合在 Modeler Administration Console 配置窗口的性能和优化标题下。对于大多数安装，都针对最佳性能预配置了这些设置。但是，您可能需要根据您具体的硬件、数据集的大小以及流的内容调整这些设置。有关详细信息，请参阅 [第 27 页码第 4 章](#) 中的性能/优化。

客户端性能和优化设置

从“流属性”对话框的“选项”选项卡上可以访问客户端性能和优化设置。要显示这些选项，请从客户端菜单上选择以下项目。

工具 > 流属性 > 选项 > 优化

您可以使用“优化”设置来优化流性能。请注意，IBM® SPSS® Modeler Server（如果使用的话）上的性能和优化设置会覆盖客户端上的任何相应设置。

注意：数据库建模和 SQL 优化需要在 IBM® SPSS® Modeler 计算机上启用 SPSS Modeler Server 连接。通过启用此设置，您可以访问数据库算法，直接从 SPSS Modeler 回送 SQL 以及访问 SPSS Modeler Server。要验证当前许可证的状态，请从 SPSS Modeler 菜单中选择以下项目。

帮助 > 关于 > 其他详细信息

如果启用了连接，您可以在“许可证状态”选项卡中看到选项服务器启用。

有关详细信息，请参阅 [第 3 章](#) 中的连接到 IBM SPSS Modeler Server 中的 IBM SPSS Modeler 15 用户指南。

图片 5-1
设置流优化选项



注意：是否支持 SQL 回送和优化，取决于使用的数据库类型。有关使用 IBM® SPSS® Modeler 15 支持和测试的数据库和 ODBC 驱动程序的最新信息，请参阅公司支持站点 <http://www.ibm.com/support>。

启用流重写。选择此选项可启用 SPSS Modeler 中的流重写。提供的重写类型有两种，您可以选择其中一种也可以两种都选。流重写会对流中的节点在后台进行重新排序，使其在不更改流语义的情况下更高效的操作。

- **优化 SQL 生成。**利用此选项可对流中的节点进行重新排序，以便可以使用 SQL 生成回送更多的操作，使其在数据库中执行。当它发现某个节点无法呈现到 SQL 中时，该优化选项则会向下查看，看是否存在任何下游节点可呈现到 SQL 中，并且可以在不影响流语义的情况下安全移动到该问题节点的前面。数据库执行操作不但可以比 SPSS Modeler 效率更高，而且这样的回送操作可以减小返回到 SPSS Modeler 进行处理的数据集的大小。而这又可以进一步减小网络通信量，加快流操作的速度。请注意，必须选中生成 SQL 复选框后 SQL 优化才能生效。有关详细信息，请参阅第 45 页码第 6 章中的 SQL 优化。

- **优化语法执行。**这种流重写方法会提高合并多个包含 IBM® SPSS® Statistics 语法节点的操作的效率。通过将语法命令合并到单个操作中，而不是作为单独的操作分别运行，从而实现优化。
- **优化其他执行。**这种流重写方法会提高无法委派给数据库的操作的效率。优化是通过尽可能早地减少流中的数据量来获得的。在保持数据完整性的同时，该流会进行重写，以使操作距离数据源更接近，因而减少了具有较高消耗的操作（如连接）的下游数据。

启用并行处理。当在一台带有多处理器的计算机上运行时，此选项使得系统能够平衡这些处理器之间的负载，从而可能产生更佳的性能。使用多个节点或者使用下列单个节点可能会由于并行处理受益：C5.0、合并（按键进行）、排序、间隔（秩和分位法），以及聚合（使用或多个键字段）。

生成 SQL。选择此选项可启用 SQL 生成，从而可以通过使用 SQL 代码使得流操作回送到数据库以生成执行处理，因此可能能够改善性能。为了进一步改善性能，还可以选择优化 SQL 生成，从而实现将回送到数据库的操作数最大化。节点的操作回送到数据库之后，该节点将在流运行时以紫色突出显示。

- **数据库高速缓存。**对于生成 SQL 以便在数据库中执行的流，数据可以在中游时缓存在数据库而不是文件系统临时表中。如果与 SQL 优化相组合，则此操作将使性能得到显著提高。例如，可以对合并多个表以创建数据挖掘视图的流的输出进行缓存并在需要时重新使用。启用数据库缓存后，请右键单击非终端节点以在此位置缓存数据，下次运行流时，将直接在数据库中自动创建缓存。如此一来，为下游节点生成 SQL 将成为可能，从而可进一步提高性能。此外，必要时可禁用该选项，例如当策略或权限排除正被写入数据库的数据时。如果未启用数据库缓存或 SQL 优化，则缓存将被转而写入文件系统。有关详细信息，请参阅第 5 章中的节点的高速缓存选项中的 IBM SPSS Modeler 15 用户指南。
- **使用宽松转换。**利用此选项可以将数据从字符串转换为数字，或从数字转换为字符串（如果以适当的格式存储）。例如，如果数据在数据库中以字符串的形式保存，但实际上包含有意义的数字，则在出现回送时转换数据供使用。

注意：由于 SQL 实现的微小差别，在数据库中运行的流返回的结果可能会与在 SPSS Modeler 中运行时返回的结果有着轻微的不同。也是因为上述原因，这些差别还可能根据数据库提供商的不同而有所区别。

数据库使用情况和优化

数据库服务器。如果可能的话，请创建一个专门用于数据挖掘的数据库实例，以便生产服务器不会受到 IBM® SPSS® Modeler 查询的影响。SPSS Modeler 生成的 SQL 语句可能要求会很高，即，IBM® SPSS® Modeler Server 计算机上的多个任务能够在同一个数据库中执行 SQL。

数据库内挖掘。很多数据库提供商都提供了用于其产品的数据挖掘扩展。这些扩展允许数据挖掘活动（如模型构建或评分）在数据库服务器内运行，或在单独的专用服务器内运行。SPSS Modeler 的数据库内挖掘功能补充并扩展了它的 SQL 生成功能，提供了一种驱动提供商特定的数据库扩展的方式。在某些情况下，采用此方式避免了 SPSS Modeler 和数据库之间用于数据传输的潜在的昂贵开销。数据库缓存可增加更多好处。有关详细信息，请参阅文件 DatabaseMiningGuide.pdf，该文件位于 IBM® SPSS® Modeler 安装盘上 \Documentation 文件夹下。

SQL 优化

为了获得最佳的性能，您总是应该尝试尽可能扩大所生成 SQL 的数量，以开发数据库的性能和可伸缩性。只有不可编译到 SQL 的流的部分应该在 IBM® SPSS® Modeler Server 中执行。有关详细信息，请参阅 [第 6 章](#)。

上载基于文件的数据

未存储在数据库中的数据不能从 SQL 优化中获益。如果您要分析的数据尚未位于数据库中，则可以使用数据库输出节点上载这些数据。您还可以使用此节点存储来自数据准备的中间数据集以及部署的结果。有关详细信息，请参阅 [第 7 章中的数据库导出高级选项中的 IBM SPSS Modeler 15 源、过程和输出节点](#)。

IBM® SPSS® Modeler 可与很多常见数据库系统的外部载入程序进行接口。该软件还附带了几个脚本，它们及其文档位于 SPSS Modeler 安装文件夹下的 /scripts 子目录中。

下表显示了批量载入的潜在性能优点。这些数据显示了向 Oracle 数据库导出 250,000 个记录和 21 个字段所花的时间。外部载入程序为 Oracle 的 `sqlldr` 实用程序。

表 5-1
批量载入的性能优点

导出选项	时间（秒）
默认 (ODBC)	409
通过 ODBC 的批量载入	52
通过外部载入程序的批量载入	33

SQL 优化

IBM® SPSS® Modeler 最强大的功能之一是直接在数据库中执行很多数据准备和挖掘操作的功能。通过生成可回送到数据库进行执行的 SQL 代码，很多操作（如抽样、排序、导出新字段以及某些类型的绘图）可以在数据库中执行，而不必在 IBM® SPSS® Modeler 或 IBM® SPSS® Modeler Server 计算机上执行。当您对较大数据集进行操作时，这些回送可在以下几个方面显著地增强性能：

- 减小从 DBMS 传输到 SPSS Modeler 的结果集的大小。通过 ODBC 驱动程序读取大型结果集时，可能会导致网络 I/O 或驱动程序的低效率。因此，由于 SQL 优化而受益最多的操作是行和列的选择和聚合（选择节点、抽样节点、汇总节点），这样通常会减小要传输的数据集的大小。数据还可以在流中的某些重要点（例如，合并或选择节点之后）高速缓存到数据库的一个临时表中，从而进一步改善性能。
- 利用数据库的性能和可伸缩性。因为 DBMS 通常可以利用并行处理、功能更强大的硬件、更为复杂的磁盘存储管理，以及由于索引的存在，所以效率会得到提高。

因为存在这些优势，所以 SPSS Modeler 的设计会最大化每个流生成的 SQL 量，因而只有那些无法编译为 SQL 的操作才会由 SPSS Modeler Server 执行。但由于可以使用标准 SQL (SQL-92) 表达的内容有限制，某些操作可能不受支持。 [有关详细信息，请参阅第 51 页码最大化 SQL 生成的提示。](#)

注意：由于 SQL 实现之间的微小差别，在数据库中执行的流与在 SPSS Modeler 中执行的流返回的结果可能略有不同。类似的原因，这些差别还可能根据数据库提供商的不同而有所区别。例如，根据字符串比较和字符串排序规则中有关区分大小写的数据库配置，使用 SQL 回送执行的 SPSS Modeler 流可能会与不使用 SQL 回送执行的流产生不同的结果。有关配置数据库的建议，请与数据库管理员联系。为尽可能提高与 SPSS Modeler 的兼容性，数据库字符串比较应区分大小写。

注意：数据库建模和 SQL 优化需要在 SPSS Modeler 计算机上启用 SPSS Modeler Server 连接。通过启用此设置，您可以访问数据库算法，直接从 SPSS Modeler 回送 SQL 以及访问 SPSS Modeler Server。要验证当前许可证的状态，请从 SPSS Modeler 菜单中选择以下项目。

帮助 > 关于 > 其他详细信息

如果启用了连接，您可以在“许可证状态”选项卡中看到选项服务器启用。

[有关详细信息，请参阅第 3 章中的连接到 IBM SPSS Modeler Server 中的 IBM SPSS Modeler 15 用户指南。](#)

注意：当在数据库中执行流时，将从该数据库中提取日期和时间详细信息。这可能与您的本地或 SPSS Modeler Server 日期和时间不同，例如，如果数据库所在的计算机位于不同的国家或时区。

数据库要求

有关使用 SPSS Modeler 支持和测试的数据库和 ODBC 驱动程序的最新信息，请参阅公司支持站点上的产品兼容性矩阵 (<http://www.ibm.com/support>)。

请注意，通过使用数据库建模还可以获得进一步的性能改善。有关详细信息，请参阅第 2 章中的数据库建模概述中的 IBM SPSS Modeler 15 数据库内数据挖掘指南。

SQL 生成如何运行

来自数据库源节点的流的初始片断是 SQL 生成的主要目标。当节点遇到无法编译为 SQL 的内容时，数据则会从数据库提取出来，由 IBM® SPSS® Modeler Server 完成后续处理。

在流准备期间以及执行之前，SQL 的生成过程按照下列步骤进行：

- 服务器对流重新排序，将下游节点移动到“SQL 区域”（如果已证实这样做不会产生问题）。（可以在服务器上禁用此功能。）
- 从源节点向终端节点不断执行操作，将 SQL 表达式逐渐构建起来。当节点遇到无法转换为 SQL 的内容或者终端节点（例如表节点或图形节点）转换为 SQL 时，此阶段停止。在此阶段的最后，每个节点都会带有一个 SQL 语句标签（如果节点及其前面的内容具有对等 SQL 的话）。
- 再从具有最复杂对等 SQL 的节点向源节点反方向不断执行操作，检查 SQL 的有效性。成功验证的 SQL 将被选择用于执行。
- 其所有操作均生成了 SQL 的节点在流工作区中突出显示为紫色。基于这些结果，您可能希望在合适时进一步重新组织您的流，以充分利用数据库执行。有关详细信息，请参阅第 51 页码最大化 SQL 生成的提示。

改善发生在何处

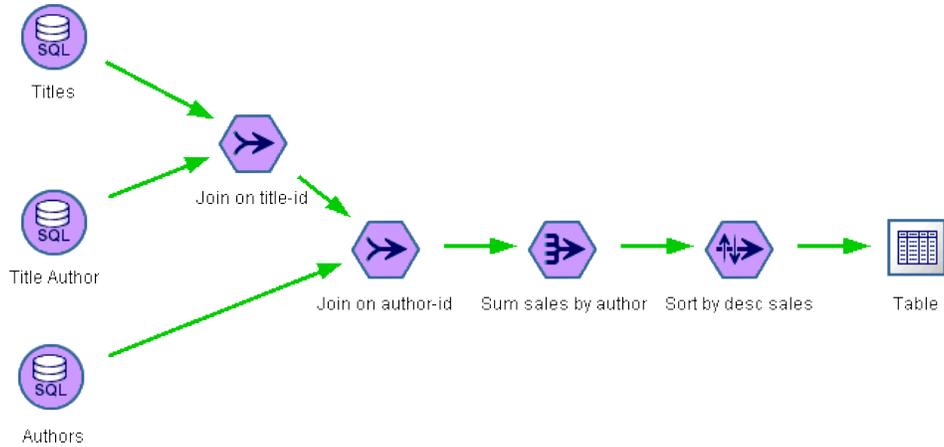
SQL 优化改善了一些数据操作中的性能：

- **连接（使用关键字合并）**。连接操作可以增强数据库内的优化。
- **汇总**。汇总节点、条形图节点和网络节点全部使用汇总生成其结果。汇总后的数据使用的带宽比原始数据要小很多。
- **选择**。基于某些标准选择记录会减少记录的数量。
- **排序**。对记录排序是一个资源密集型活动，在数据库中执行的效率会更高。
- **字段派生**。在数据库中生成新字段时效率更高。
- **字段投射**。SPSS Modeler Server 仅从数据库中提取后续处理所必需的字段，这样会最小化带宽和内存要求。对于平面文件中的多余字段也是上述情况：尽管服务器必须读取多余的字段，但它不会为其分配任何存储。
- **评分**。SQL 可从决策树、结果集、线性回归和因子生成的模型生成。

SQL 生成示例

下面的流会通过关键字操作连接三个数据库表，然后执行一次汇总和一次排序。

图片 6-1
带有表明 SQL 回送（在数据库中执行的操作）的紫色节点的优化流



生成的 SQL

对于此流生成的 SQL 如下所示：

```

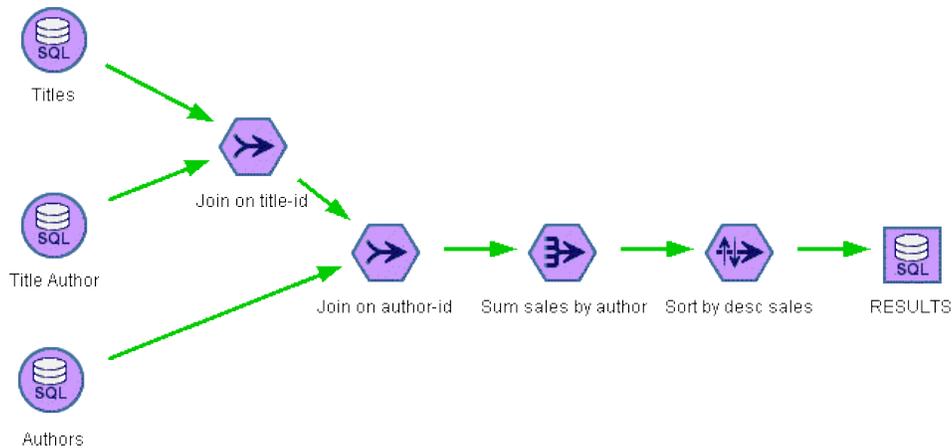
SELECT
    T2. au_lname AS C0,
    T2. au_fname AS C1,
    SUM({fn CONVERT(T0. ytd_sales ,SQL_BIGINT)}) AS C2
FROM
    dbo . titles T0,
    dbo . titleauthor T1,
    dbo . authors T2
WHERE
    (T0. title_id = T1. title_id )
    AND (T1. au_id = T2. au_id )
GROUP BY T2. au_lname ,T2. au_fname
ORDER BY 3 DESC

```

执行流

如果流以数据库导出节点终止，则可以在数据库中执行整个流。

图片 6-2
在数据库中执行的整个流



配置 SQL 优化

- ▶ 安装 ODBC 驱动程序，为要使用的数据库配置数据源。有关详细信息，请参阅第 11 页码第 2 章中的数据访问。
- ▶ 创建一个使用源节点从该数据库拉出数据的流。
- ▶ 检查以确保在客户端和服务器上均启用了 SQL 生成（如果可应用的话）。默认情况下客户端和服务器上都是启用的。

在客户端上启用 SQL 优化

- ▶ 在“工具”菜单中，选择用户选项。
- ▶ 单击“优化”选项卡。选择生成 SQL 启用 SQL 优化。您还可以选择其他设置以改善性能。有关详细信息，请参阅第 41 页码第 5 章中的客户端性能和优化设置。

在服务器上启用 SQL 优化

因为服务器设置会覆盖在客户端上指定的任何内容，所以服务器配置设置流重写和自动 SQL 生成必须都打开。有关如何更改 IBM® SPSS® Modeler Server 设置的详细信息，请参阅性能/优化中的第 4 章一节。

对模型评分时启用优化

为了进行评分，必须为每个建模节点分别启用 SQL 生成，而不管任何服务器或客户端级的设置如何。这样做是因为有些模型生成非常复杂的 SQL 表达式，在数据库中可能无法有效评估。由于 SQL 的大小或复杂性，在尝试执行生成的 SQL 时数据库可能报告错误。

可能需要一定数量的试错以确定 SQL 生成是否改进给定模型的性能。此过程是在生成的模型添加到流之后在“设置”选项卡上完成的。有关详细信息，请参阅第 3 章中的使用流中的模型块中的 IBM SPSS Modeler 15 建模节点。

预览生成的 SQL

在数据库中执行生成的 SQL 之前，您可以在消息日志中进行预览。这可能有助于调试程序，通过预览还可以导出生成的 SQL 以便将来在数据库中进行编辑或运行。预览还会表明哪些节点将回送到数据库，这样可能有助于您确定是否可以对流重新排序以改善性能。

- ▶ 确保在“用户选项”对话框中选择了执行流期间在消息日志中显示 SQL 以及流准备期间在消息日志中显示关于 SQL 生成的详细信息。有关详细信息，请参阅第 41 页码第 5 章中的客户端性能和优化设置。
- ▶ 在流工作区中，选择要预览的节点或流。
- ▶ 单击工具栏上的预览 SQL 按钮。

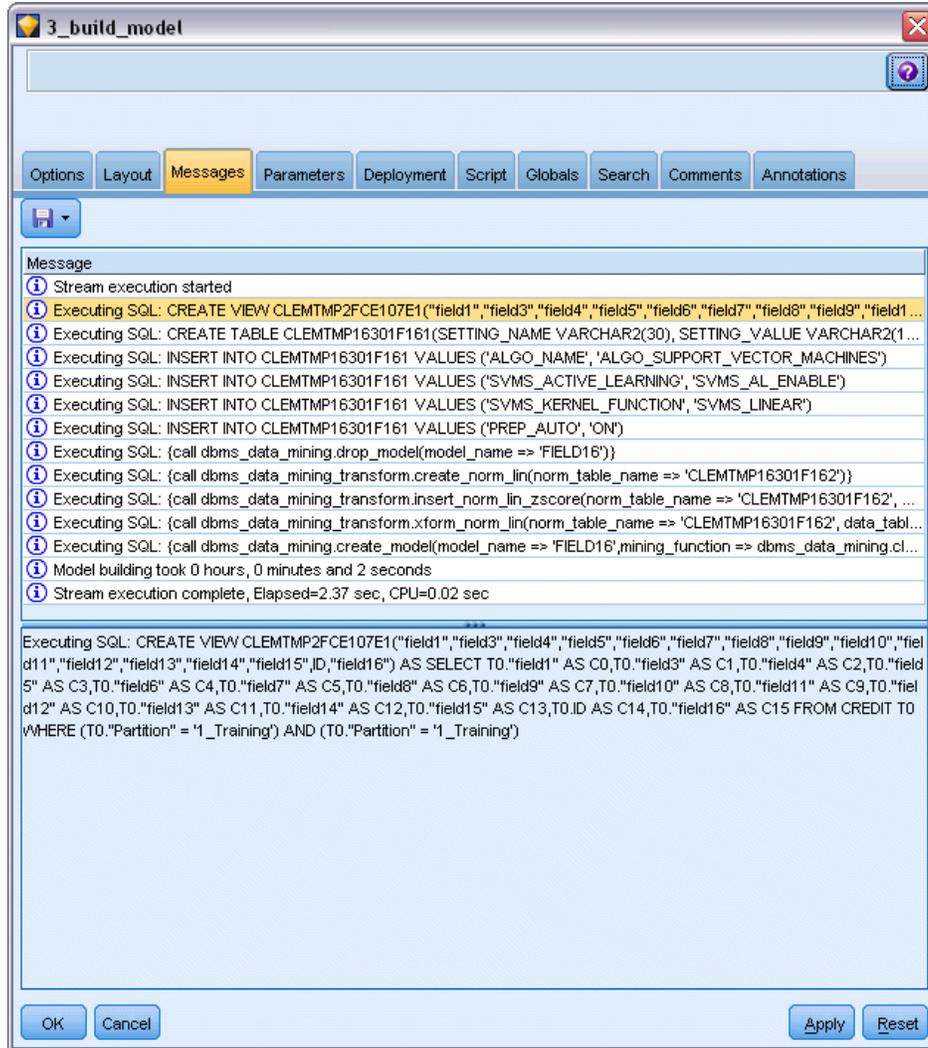
为其生成 SQL 的所有节点（以及执行流时将回送到数据库的节点）在流工作区中显示为紫色。

图片 6-3
预览 SQL 按钮



- ▶ 要预览生成的 SQL，请从菜单中选择：
工具 > 流属性 > 消息...

图片 6-4
显示在消息日志中的已生成 SQL



查看模型块 SQL

对于某些模型，可以生成模型块 SQL，将模型评分阶段回送到数据库。此功能的主要用途不是提高性能，而是允许包含这些块的流进行完整的 SQL 回送。有关详细信息，请参阅第 52 页码支持 SQL 生成的节点。

要查看支持 SQL 生成的模型块 SQL：

- ▶ 选择模型块上的“设置”选项卡。
- ▶ 根据情况选择一个选项，在缺失值支持的情况下生成或为此模型生成 SQL。
- ▶ 在模型块菜单中，选择：
文件 > 导出 SQL

- ▶ 保存该文件。
- ▶ 打开文件以查看 SQL。

最大化 SQL 生成的提示

要从 SQL 优化获取最佳的性能改善，请注意以下几点。

流顺序。因为 IBM® SPSS® Modeler’ 的数据挖掘功能比标准 SQL 支持的传统数据处理操作更丰富，所以当节点功能与 SQL 没有对等语义时，SQL 生成可能会停止。如果发生这种情况，任何下游节点的 SQL 生成也都会受到抑制。因此，通过对节点重新排序，以使停止 SQL 的操作放置在流中尽可能靠下的位置，您或许能够显著地改善性能。SQL 优化器可以自动执行一定数量的重新排序（确保启用了流重写即可），但还可以进行进一步的改善。选择节点即是一个很好的备选，该节点通常可以向前推进一些。[有关详细信息，请参阅第 52 页码支持 SQL 生成的节点。](#)

CLEM 表达式。如果流无法重新排序，您或许可以更改节点选项或 CLEM 表达式，或者重新设计操作的执行模式，以便它不再阻止 SQL 生成。导出、选择以及类似的节点通常可以呈现为 SQL，条件是所有 CLEM 表达式运算符都具有对等 SQL。大多数运算符都可以呈现，但有一些运算符会阻止 SQL 生成（具体来说，为序列函数 [“@functions”]）。有时生成会因为生成的查询变得太复杂，以至于数据库无法处理而停止。[有关详细信息，请参阅第 55 页码 CLEM 支持 SQL 生成的表达式和运算符。](#)

多个源节点。对于流具有多个数据库源节点的情况，SQL 生成会独立应用于每个输入分支。如果生成在某个分支上停止，它可以在另一个分支上继续。对于两个分支合并的情况（并且合并之前两个分支都可以表达为 SQL），合并本身通常会被替代为数据库连接，生成可以在下游继续。

数据库算法。模型估计总是在 IBM® SPSS® Modeler Server 上执行，而不在数据库中执行，但使用 Microsoft、IBM 或 Oracle 提供的数据库自有算法时除外。[有关详细信息，请参阅第 2 章中的数据库建模概述中的 IBM SPSS Modeler 15 数据库内数据挖掘指南。](#)

对模型评分。通过将生成的模型呈现为 SQL，可支持某些模型在数据库内部评分。但某些模型会生成极其复杂的 SQL 表达式，而这些表达式总是无法在数据库内进行有效计算。因此，必须为每个模型节点单独启用 SQL 生成。如果您发现某个模型节点正在阻止 SQL 生成，则请转至该节点对话框上的“设置”选项卡，然后选择生成此模型的 SQL（对于某些模型，可能还有一些其他选项可控制生成）。运行检验，确认该选项是否有益于您的应用。[有关详细信息，请参阅第 52 页码支持 SQL 生成的节点。](#)

当测试建模节点以查看模型的 SQL 生成是否有效工作，我们建议首先保持 IBM® SPSS® Modeler 的所有流。一些数据库系统可能尝试处理（潜在复杂）生成的 SQL 时可能中止，这需从 Windows 任务管理器关闭 SPSS Modeler。

数据库高速缓存。如果您要使用节点高速缓存在流中的某些重要点（例如合并或汇总节点之后）保存数据，则请确保启用了数据库高速缓存和 SQL 优化。这样将使得数据在大多数情况下都能够高速缓存到数据库（而不是文件系统中）的一个临时表中。[有关详细信息，请参阅第 48 页码配置 SQL 优化。](#)另请参阅 [节点的高速缓存选项](#) 了解详细信息。

供应商特定的 SQL。大多数生成的 SQL 都符合标准 (SQL-92)，但在实际中会采用一些非标准的、供应商特定的功能。根据数据库源的不同，SQL 优化的程度可能有所差别。

支持 SQL 生成的节点

下表显示了表示支持 SQL 生成的数据挖掘操作的节点。如果节点（数据库建模节点除外）不出现在这些表格中，则它不支持 SQL 生成。

您可以在执行前预览生成的 SQL。[有关详细信息，请参阅第 49 页码预览生成的 SQL。](#)

源

支持 SQL 生成的节点	附注
数据库	此节点用于指定要用于进一步分析的表和视图。此节点实现了 SQL 查询的输入。请避免带有重复列名的结果集。 有关详细信息，请参阅第 57 页码写入 SQL 查询。

记录操作

支持 SQL 生成的节点	附注
选择	只有当支持选定表达式本身的 SQL 生成（请参阅下面的表达式）时，才支持生成。如果任何字段值为空，则 SQL 生成所给出的丢弃结果与本地 IBM® SPSS® Modeler 中给出的结果不同。
样本	简单抽样支持 SQL 生成。请参阅第 52 页码中的表 6-1。对于 Netezza 数据库，复杂抽样支持 SQL 生成。
聚合	聚合的 SQL 生成支持取决于数据存储类型。请参阅第 53 页码中的表 6-2。
RFM 汇总	支持生成，除非保存了第二个或第三个最近交易的日期，或仅仅包含最近的几次交易。但是，如果回送了 <code>datetime_date(YEAR, MONTH, DAY)</code> 函数，则包括最近几次交易时，也支持生成。
排序	
合并	按顺序合并不会生成任何 SQL。仅当数据库/驱动程序支持使用具有完整或部分外部连接的关键字进行合并时才支持此功能。可通过过滤器节点或源节点的“过滤器”选项卡重命名非匹配的输入字段。支持按条件合并的 SQL 生成。对于所有类型的合并，如果输入源自不同的数据库，则不支持 <code>SQL_SP_EXISTS</code> 。
追加	如果未排序输入，则支持生成。
区分	

表 6-1
在样本节点中支持 SQL 生成以用于简单抽样

模式	样本	大小最大 大值	种子 数	DB2 OS/Z	DB2 OS/400	DB2 Windows/ UNIX	Netezza	Oracle	SQL Server	Teradata
Include	First	n/a		Y	Y	Y	Y	Y	Y	Y
	n 中取 1	关闭		Y	Y	Y	Y	Y		Y
		max		Y	Y	Y	Y	Y		Y
	随机 %	关闭	关闭			Y	Y	Y		Y

模式	样本	大小最大值	种子数	DB2 OS/Z	DB2 OS/400	DB2 Windows/UNIX	Netezza	Oracle	SQL Server	Teradata	
			在			Y		Y			
		max	关闭			Y	Y	Y		Y	
			在			Y		Y			
Discard	First	关闭					Y	Y			
		max					Y	Y			
	n 中取 1	关闭		Y	Y	Y	Y	Y		Y	
		max		Y	Y	Y	Y	Y		Y	
	随机 %	关闭	关闭				Y	Y	Y		Y
			在				Y		Y		
		max	关闭				Y	Y	Y		Y
			在				Y		Y		

表 6-2
在汇总节点中支持 SQL 生成

存储	和	均值	最小	最大	SDev	中位数	Count	方差	百分位数
整数	Y	Y	Y	Y	Y		Y	Y	
实数	Y	Y	Y	Y	Y		Y	Y	
日期			Y	Y			Y		
时间			Y	Y			Y		
时间戳			Y	Y			Y		
字符串			Y	Y			Y		

字段操作

支持 SQL 生成的节点	附注
类型	如果类型节点已经实例化，且没有指定 ABORT 或 WARN 类型检查，则支持 SQL 生成。
筛选器	
导出	如果支持为派生表达式生成的 SQL，则支持 SQL 生成（请参阅下面的表达式）。
整体	支持连续目标的 SQL 生成。对于其他目标，仅当使用“赢得最高置信度”整体方法时，才支持生成。
填充	如果支持为派生表达式生成的 SQL，则支持 SQL 生成（请参阅下面的表达式）。
匿名化	支持连续目标的 SQL 生成，以及名义和标志目标的部分 SQL 生成。
重新分类	
离散化	如果使用“分位数（同等计数）”分箱方法，并且选中“如果可用，从分级值选项卡读取”选项，则支持 SQL 生成，
RFM 分析	如果选中“如果可用，从分级值选项卡读取”选项，则支持 SQL 生成，但下游节点不支持。

支持 SQL 生成的节点	附注
分区	支持 SQL 生成以分配记录到分区。
设为标志	
重建	

图形

支持 SQL 生成的节点	附注
图形板	以下图形类型支持 SQL 生成：面积图、3-D 面积图、条形图、3-D 条形图、计数条、Heat 图、饼图、3-D 饼图以及计数饼图。对于直方图，只有分类数据支持 SQL 生成。
分布	
Web	
评估	

模型块

对于某些模型，可以生成模型块 SQL，将模型评分阶段回送到数据库。此功能的主要用途不是提高性能，而是允许包含这些块的流进行完整的 SQL 回送。[有关详细信息，请参阅第 50 页码查看模型块 SQL。](#)

支持 SQL 生成的模型块	附注
C&R 树	支持单个树选项的 SQL 生成，但不支持推进、bagging 或大型数据集选项的 SQL 生成。
QUEST	
CHAID	
C5.0	
决策表	
线性	支持标准模型选项的 SQL 生成，但不支持推进、bagging 或大型数据集选项的 SQL 生成。
神经网络	支持标准模型选项（仅多层感知器）的 SQL 生成，但不支持推进、bagging 或大型数据集选项的 SQL 生成。
主成分分析/因子	
Logistic	支持多项（而非二项）过程的 SQL 生成。对于多项过程，选中置信度时，不支持生成，除非目标类型为标志。
生成的结果集	

输出

支持 SQL 生成的节点	附注
表	如果对于突出显示表达式（请参阅下面的表达式）支持 SQL 生成，则支持生成。
矩阵	除选择了字段选项的“所有数字”外，都支持生成。
分析	支持生成，具体取决于选择的选项。
转换	

支持 SQL 生成的节点	附注
Statistics	如果不使用“相关”选项，则支持生成。
报告	
设置全局量	

导出

支持 SQL 生成的节点	附注
数据库	
发布者	发布的流将包含生成的 SQL。

CLEM 支持 SQL 生成的表达式和运算符

下表显示了支持 SQL 生成，并在数据挖掘中经常使用的数学运算和表达式。未在此表中显示的运算在当前版本中不支持 SQL 生成。

运算符

支持 SQL 生成的运算	附注
+	
-	
/	
*	
><	用于连接字符串。

关系运算符

支持 SQL 生成的运算	附注
=	
/=	用于指定“不等于”。
>	
>=	
<	
<=	

函数

支持 SQL 生成的运算	附注
abs	
allbutfirst	
allbutlast	
and	
arccos	
arcsin	
arctan	

支持 SQL 生成的运算	附注
arctanh	
cos	
div	
exp	
fracof	
hasstartstring	
hassubstring	
integer	
intof	
isaplhacode	
islowercode	
isnumbercode	
isstartstring	
issubstring	
isuppercode	
last	
length	
locchar	
log	
log10	
lowertoupper	
max	
member	
min	
negate	
not	
number	
or	
pi	
real	
rem	
round	
sign	
sin	
sqrt	
string	
strmember	
subscrs	
substring	
substring_between	
uppertolower	
to_string	

特殊函数

支持 SQL 生成的运算	附注
@NULL	
@GLOBAL_AVE	该特殊全局函数用于检索设置全局变量节点计算的全局值。
@GLOBAL_SUM	
@GLOBAL_MAX	
@GLOBAL_MEAN	
@GLOBAL_MIN	
@GLOBALSDEV	

汇总函数

支持 SQL 生成的运算	附注
Sum	
Mean	
Min	
Max	
Count	
SDev	

使用 CLEM 表达式中的 SQL 函数

@SQLFN函数可用于在 CLEM 表达式中添加指定的 SQL 函数（仅用于数据库执行）。对于需要采用专有 SQL 或其他供应商特定自定义的特殊情况，此函数十分有用。

此函数的使用不在标准 IBM® SPSS® Modeler 支持协议的涉及范围内，因为执行依靠的外部数据库组件超出 IBM Corp. 的控制范围，但在某些特殊情况下可能被采纳（通常包含在服务合约中）。如有必要，请访问 <http://www.ibm.com/software/analytics/spss/services/> 以获得详细信息。

写入 SQL 查询

使用数据库节点时，您应该对会产生带有重复列名的数据集的任何 SQL 查询特别注意。这些重复列名通常会阻止任何下游节点的 SQL 优化。

IBM® SPSS® Modeler 使用嵌套的 **SELECT** 语句为在数据库源节点中使用 SQL 查询的流回送 SQL。换句话说，该流会将数据库源节点中指定的查询嵌套在一个或多个在下游节点优化期间生成的 **SELECT** 语句中。因此，如果查询的结果集包含重复列名，RDBMS 则无法嵌套该语句。嵌套困难最常发生在表连接过程中，因为具有相同名称的列已在多个连接表中选中。例如，假设这种查询情况发生在源节点中：

```
SELECT e. ID, e. LAST_NAME, d.*
FROM EMP e RIGHT OUTER JOIN
DEPT d ON e. ID = d. ID;
```

该查询会阻止后面的 SQL 优化，因为此 **SELECT** 语句会生成带有两个名为 **ID** 的列的数据集。

为了允许完整 SQL 优化，编写 SQL 查询时您应该更加明确，出现重复列名的情况时，您应该指定列别名。下面的语句会说明一个更加明确的查询：

```
SELECT e. ID AS ID1, e. LAST_NAME, d.*  
FROM EMP e RIGHT OUTER JOIN  
DEPT d ON e. ID = d. ID;
```

为 UNIX 平台配置 Oracle

配置 Oracle 的 SQL 优化

当在 UNIX 平台上运行 IBM® SPSS® Modeler Server 并从 Oracle 数据库读取时，请参考下面的提示，以确保在数据库中彻底地优化生成的 SQL。

适当的环境规范

当在除连接 ODBC 驱动程序随附环境以外的环境中运行 SPSS Modeler Server 时，应当重新配置计算机以增强 SQL 优化。仅将随附的 ODBC 驱动程序与 en_US 环境文件连接。结果，如果在不同的环境中运行 SPSS Modeler Server 计算机，或在启动了 SPSS Modeler Server 的命令解释程序未完全定义环境时，则在 Oracle 中，生成的 SQL 可能不会完全优化。原因如下：

- SPSS Modeler Server 使用与其运行的环境对应的 ODBC 环境文件，以将从数据库返回的代码转换为文本字符串。然后使用这些文本字符串确定实际连接到哪个数据库。
- 如果环境（使用系统 `$LANG` 查询返回到 SPSS Modeler Server）不是 en_US，则 IBM® SPSS® Modeler 无法将它从 ODBC 驱动程序接收的代码转换为文本。换言之，在启动数据库连接时，返回到 SPSS Modeler Server 的是未转换的代码，而不是字符串 Oracle。这意味着 SPSS Modeler 无法优化 Oracle 的流。

检查并重设环境规范：

- ▶ 在 UNIX 命令解释程序中，运行：

```
#locale
```

这将返回命令解释程序的环境信息。例如：

```
$ locale
LANG=en_US. ISO8859-15
LC_CTYPE="en_US. ISO8859-15"
LC_NUMERIC="en_US. ISO8859-15"
LC_TIME="en_US. ISO8859-15"
LC_COLLATE="en_US. ISO8859-15"
LC_MONETARY="en_US. ISO8859-15"
LC_MESSAGES="en_US. ISO8859-15"
LC_ALL=en_US. ISO8859-15
```

- ▶ 更改为连接 ODBC/环境目录。（您将在此处看到一个单独的目录，en_US。）
- ▶ 创建指向此 en_US 目录的软链接，指定命令解释程序中环境设置的名称。示例如下：

```
#ln -s en_US en_US. ISO8859-15
```

对于非英语环境，如 fr_FR. ISO8859-1，应该使用如下命令创建软链接：

```
#ln -s en_US fr_FR. ISO8859-1
```

- ▶ 创建了此链接后，请从同一命令解释程序中重新启动 SPSS Modeler Server。（SPSS Modeler Server 将接收到来自它启动时所在命令解释程序的环境信息。

备注

在优化 SQL 回送到 Oracle 的 UNIX 计算机时，请考虑以下提示：

- 必须指定整个环境。在上面的示例中，必须以 `language_territory.code-page` 形式创建链接。现有的 `en_US` 环境目录并不充分。
- 要完全优化数据库内挖掘，必须在用于启动 SPSS Modeler Server 的命令解释程序中定义 `LANG` 和 `LC_ALL`。在重新启动 SPSS Modeler Server 之前，可以在命令解释程序中将 `LANG` 定义为所需的任何其他环境变量。例如，请参见下面的定义：

```
#LANG=en_US.ISO8859-15; export LANG
```

- 每次启动 SPSS Modeler Server 时，需要检查是否完全定义了命令解释程序环境信息，以及 ODBC/环境目录中是否存在相应的软链接。

配置 UNIX 启动脚本

简介

此附录描述了某些与 UNIX 版本的 IBM® SPSS® Modeler Server 一起提供的脚本，并且说明了如何配置这些脚本。这些脚本可用于：

- 将 SPSS Modeler Server 配置成服务器计算机重新启动时自动启动。
- 手动停止并重新启动 SPSS Modeler Server。
- 更改 SPSS Modeler Server 所创建的文件权限。
- 配置 SPSS Modeler Server 以和 SPSS Modeler Server 提供的 ODBC Connect 驱动程序一起使用。 [有关详细信息，请参阅第 63 页码 IBM SPSS Modeler Server 和数据访问包。](#)
- 配置 SPSS Modeler Server 以在用 Enterprise View 节点读取数据时使用 IBM® SPSS® Collaboration and Deployment Services Enterprise View Driver。 [有关详细信息，请参阅第 69 页码为 Enterprise View 节点配置驱动程序。](#)

脚本

IBM® SPSS® Modeler Server 使用多个脚本，包括：

- **modelersrv.sh**。SPSS Modeler Server 的手动启动脚本位于 SPSS Modeler Server 的安装目录中。当手动启动服务器守护进程时，手动启动脚本将配置服务器环境。您可以通过运行此脚本实现手动启动和关闭服务器。当需要更改手动启动配置时，您可以编辑此脚本。
- **auto.sh**。此脚本用于对您的系统进行配置，以便在引导时自动启动服务器守护进程。运行此脚本即可将系统配置成自动启动。您无需对此脚本进行编辑。此脚本位于 SPSS Modeler Server 的安装目录中。
- **rc.modeler**。当运行 auto.sh 时，系统将在某个位置（取决于您服务器的操作系统）创建此脚本。当自动启动时，此脚本会配置服务器环境。当需要更改自动启动配置时，您可以编辑此脚本。

表 B-1
不同操作系统中 rc.modeler 的位置

操作系统	位置
AIX	/etc/rc.modeler
HP-UX	/sbin/init.d/rc.modeler
Solaris	/etc/init.d/rc.modeler

自动启动和停止 IBM SPSS Modeler Server

IBM® SPSS® Modeler Server 必须作为守护进程启动。安装程序中包括您想运行用来配置系统以自动停止和重新启动 SPSS Modeler Server 的脚本 (auto.sh)。

将系统配置为自动启动和关闭

- ▶ 以 root 身份登录。
- ▶ 切换至 SPSS Modeler Server 安装目录。
- ▶ 运行此脚本。在 UNIX 提示符处键入：

```
./auto.sh
```

在上表中所显示的位置将创建自动启动脚本 (rc.modeler)。每当服务器计算机被重新引导时，操作系统将使用 rc.modeler 来启动 SPSS Modeler Server 守护进程。每当系统关闭时，操作系统还将采用 rc.modeler 来停止守护进程。

手动启动和停止 IBM SPSS Modeler Server

可通过运行 modelersrv.sh 脚本手动启动和停止 IBM® SPSS® Modeler Server。

手动启动和停止 IBM SPSS Modeler Server

- ▶ 切换至 SPSS Modeler Server 安装目录。
- ▶ 要启动此服务器，请在 UNIX 命令提示符下输入：

```
./modelersrv.sh start
```

- ▶ 要停止此服务器，请在 UNIX 命令提示符下输入：

```
./modelersrv.sh stop
```

编辑脚本

如果您同时采用了手动和自动启动，则请同时对 modelersrv.sh 和 rc.modeler 脚本进行修改。如果您仅采用了手动启动，则请修改 modelersrv.sh。如果您仅采用了自动启动，则请修改 rc.modeler。

编辑脚本

- ▶ 停止 IBM® SPSS® Modeler Server。（有关详细信息，请参阅第 62 页码[手动启动和停止 IBM SPSS Modeler Server](#)。）
- ▶ 找到相应脚本。（有关详细信息，请参阅第 61 页码[脚本](#)。）
- ▶ 用文本编辑器打开脚本，修改并保存文件。
- ▶ 自动（通过重新启动服务器计算机）或手动启动 SPSS Modeler Server。

控制文件创建权限

IBM® SPSS® Modeler Server 为所有用户创建了具有读、写和执行权限的临时文件。您可以通过编辑启动脚本（modelersrv.sh 和/或 rc.modeler）中的 **UMASK** 设置来覆盖此默认值。（更多信息，请参阅上述[编辑脚本](#)。）我们建议采用 **077** 作为最具约束力的 **UMASK** 设置。约束力较强的设置可能会导致 SPSS Modeler Server 发生权限问题。

IBM SPSS Modeler Server 和数据访问包

如果想在 IBM® SPSS® Modeler Server 中使用 ODBC 驱动程序，则启动 SPSS Modeler Server 进程时必须通过 `odbc.sh` 配置 ODBC 环境。通过编辑 `modelersrv.sh` 和/或 `rc.modeler` 中相应的 IBM® SPSS® Modeler 启动脚本，可以实现上述配置。（[有关详细信息，请参阅第 62 页码编辑脚本](#)。）

有关详细信息，请访问技术支持 Web 站点 <http://www.ibm.com/support>。如果您有关于创建或设置 ODBC 数据源权限的问题，请联系您的数据库管理员。

配置 ODBC 以随 IBM SPSS Modeler Server 启动

- ▶ 如果 SPSS Modeler Server 主机正在运行，请将其停止。
- ▶ 为您安装 SPSS Modeler Server 的平台下载相关的压缩 tar 存档文件。确保下载对应于您所安装的 SPSS Modeler Server 版本的驱动程序。将文件复制到您要安装 ODBC 驱动程序的位置（例如，`/usr/spss/odbc`）。
- ▶ 提取存档文件。
- ▶ 运行从存档文件中提取的 `setodbcpath.sh` 脚本。
- ▶ 编辑脚本 `odbc.sh`，在该脚本底部添加 ODBCINI 定义，并将其导出，例如：

```
ODBCINI=/usr/spss/odbc/odbc.ini; export ODBCINI
```

ODBCINI 必须指向 `odbc.ini` 文件的完整路径名，SPSS Modeler 将从该文件中读取您所定义的 ODBC 数据源列表（默认的 `odbc.ini` 随驱动程序一起安装）。

- ▶ 保存 `odbc.sh`。
- ▶ （仅适合 64 位 SPSS Modeler Server 安装；对于其他安装，从下一步继续）在 `odbc.sh` 中定义并导出 `LD_LIBRARY_PATH_64`：

```
if [ "$LD_LIBRARY_PATH_64" = "" ]; then
    LD_LIBRARY_PATH_64=<library_path>
else
    LD_LIBRARY_PATH_64=<library_path>:$LD_LIBRARY_PATH_64
fi
export LD_LIBRARY_PATH_64
```

其中 `library_path` 与在脚本中存在的 `LD_LIBRARY_PATH` 定义相同，该脚本已在您的安装路径中初始化（例如 `/usr/spss/odbc/lib`）。最简便的方法是在 `odbc.sh` 文件中复制 `LD_LIBRARY_PATH` 的 `if` 和 `export` 语句，并将其附加到文件末尾，然后使用“`LD_LIBRARY_PATH_64`”替换新附加的 `if` 和 `export` 语句中的“`LD_LIBRARY_PATH`”。

因此，在 64 位 SPSS Modeler Server 安装上的最终 `odbc.sh` 文件应如下所示：

```
if [ "$LD_LIBRARY_PATH" = "" ]; then
    LD_LIBRARY_PATH=/usr/spss/odbc/lib
else
    LD_LIBRARY_PATH=/usr/spss/odbc/lib:$LD_LIBRARY_PATH
fi
export LD_LIBRARY_PATH
if [ "$LD_LIBRARY_PATH_64" = "" ]; then
    LD_LIBRARY_PATH_64=/usr/spss/odbc/lib
else
    LD_LIBRARY_PATH_64=/usr/spss/odbc/lib:$LD_LIBRARY_PATH_64
fi
export LD_LIBRARY_PATH_64
ODBCINI=/usr/spss/odbc/odbc.ini; export ODBCINI
```

记住导出 `LD_LIBRARY_PATH_64`，并使用 `if` 循环来定义它。

- ▶ 编辑您之前使用 `$ODBCINI` 定义的 `odbc.ini` 文件。定义所需的数据源名称（取决于您当前访问的数据库）。
- ▶ 保存 `odbc.ini` 文件。
- ▶ 配置 SPSS Modeler Server 以使用这些驱动程序。为此，应编辑 `modelersrv.sh` 并紧接在 `SCLEMDNAME` 定义行下面添加以下行：

```
. <odbc.sh_path>
```

其中 `odbc.sh_path` 是您在此过程开始时编辑过的 `odbc.sh` 文件的完整路径，例如：

```
./usr/spss/odbc/odbc.sh
```

注意：此处的语法很重要；请在第一个句点和文件路径之间保留一个空格。

- ▶ 保存 `modelersrv.sh`。

测试连接

- ▶ 重新启动 SPSS Modeler Server。
- ▶ 从客户端连接到 SPSS Modeler Server。
- ▶ 在客户端上，将数据库源节点添加到工作区中。
- ▶ 打开节点，确认您可以看到之前配置过程中在 `odbc.ini` 文件中定义的数据源名称。

如果看不到期望结果，或者在您尝试连接到所定义的数据源时发生错误，请遵照故障排除步骤执行检查。有关详细信息，请参阅第 65 页码 [ODBC 配置故障排除](#)。

配置 ODBC 以随 IBM SPSS Modeler Solution Publisher Runtime 启动

在成功地从 SPSS Modeler Server 连接到数据库后，您可以从 IBM® SPSS® Modeler Solution Publisher Runtime 的启动脚本中引用相同的 `odbc.sh` 脚本，以便在同一服务器上配置 SPSS Modeler Solution Publisher Runtime 安装。

- ▶ 编辑 SPSS Modeler Solution Publisher Runtime 中的 `modelerrun` 脚本，并紧邻脚本末行上方添加以下行：

```
. <odbc.sh_path>
```

其中 `odbc.sh_path` 是您在此过程开始时编辑过的 `odbc.sh` 文件的完整路径，例如：

```
./usr/spss/odbc/odbc.sh
```

注意：此处的语法很重要；请在第一个句点和文件路径之间保留一个空格。

- ▶ 保存 `modelerrun` 脚本文件。

配置 ODBC 以随 IBM SPSS Modeler Batch 启动

对于 ODBC，无需配置 IBM® SPSS® Modeler Batch 脚本。这是因为您是从 SPSS Modeler Batch 连接到 SPSS Modeler Server 来运行流。确保已按照本节前面的说明完成了 SPSS Modeler Server ODBC 配置，并且能够正常工作。

添加或编辑数据源名称

- ▶ 编辑 `odbc.ini` 文件以包含新的或更改后的名称。
- ▶ 按照本节前面的说明测试连接。

如果与 SPSS Modeler Server 的连接工作正常，则新的或更改后的数据源也应与 SPSS Modeler Solution Publisher Runtime 和 SPSS Modeler Batch 正常工作。

ODBC 配置故障排除

未列出数据源，或显示随机文本

如果您打开数据库源节点，可用数据源列表为空或包含异常条目，则可能是由于启动脚本存在问题。

- ▶ 检查在 `modelersrv.sh` 中是否定义了 `$ODBCINI`，它要么在该脚本中明确定义，要么在 `modelersrv.sh` 引用的 `odbc.sh` 脚本中定义。
- ▶ 对于后一种情况，应确保 `ODBCINI` 指向您用于定义 ODBC 数据源的 `odbc.ini` 文件的完整路径。
- ▶ 如果在 `ODBCINI` 中正确指定了路径，则应通过回显 `modelersrv.sh` 中的 `$ODBCINI` 值，以检查在 IBM® SPSS® Modeler Server 环境中使用的变量值。为此，应在 `modelersrv.sh` 中的 `ODBCINI` 定义位置添加以下行：

```
echo $ODBCINI
```

- ▶ 保存并执行 `modelersrv.sh`。在 SPSS Modeler Server 环境中设置的 `$ODBCINI` 值将被写入 `stdout` 以供验证。
- ▶ 如果根本没有值返回到 `stdout`，而且您是在 `odbc.sh` 脚本中定义 `$ODBCINI`，并从 `modelersrv.sh` 中进行引用，则应检查引用语法是否正确。它应为：

```
. <odbc.sh_path>
```

其中 `odbc.sh_path` 是您在此过程开始时编辑过的 `odbc.sh` 文件的完整路径，例如：

```
. /usr/spss/odbc/odbc.sh
```

注意：此处的语法很重要；请在第一个句点和文件路径之间保留一个空格。

如果在运行 `modelersrv.sh` 时 `stdout` 回显了正确值，则在您重启 SPSS Modeler Server 并从客户端进行连接后，应当在数据源节点中看到数据源名称。

在“数据库连接”对话框中单击“连接”时 IBM SPSS Modeler 客户端挂起

这种情况可能是因为您的库路径未正确设置以包含 ODBC 库路径。该库路径由 `$LD_LIBRARY_PATH` 定义（在 64 位版本上为 `$LD_LIBRARY_PATH_64`）。

要在 SPSS Modeler Server 守护进程环境中查看该库路径的值，则应从 `modelersrv.sh` 中回显适当的环境变量值（位于将 ODBC 库路径附加到库路径的所在行之后），并执行脚本。在您下次执行脚本时，库路径值将回显到终端。

如果您从 `modelersrv.sh` 中引用 `odbc.sh` 来设置您的 SPSS Modeler Server ODBC 环境，则应从引用 `odbc.sh` 脚本的所在行之后回显库路径值。要回显该值，应在脚本中添加以下行，然后保存并执行脚本文件：

```
echo ${library_path_variable}
```

其中 `<library_path_variable>` 是您的服务器操作系统的相应库路径变量。

库路径的返回值必须包含指向 ODBC 安装的 `lib` 子目录的路径。否则，应将此位置附加到该文件。

如果您在运行 64 版本的 SPSS Modeler Server，则应 `$LD_LIBRARY_PATH_64` 将覆盖 `$LD_LIBRARY_PATH`（如设置）。如果您在某个 64 位平台上遇到此问题，则可从 `modelersrv.sh` 中回显 `LD_LIBRARY_PATH_64` 和 `$LD_LIBRARY_PATH`，根据需要，设置 `$LD_LIBRARY_PATH_64` 以包含您的 ODBC 安装的 `lib` 子目录。

找不到数据源名称，且未指定默认驱动程序

如果您在“数据库连接”对话框中单击“连接”时看到此错误，通常表明您的 `odbc.ini` 文件定义不正确。检查该文件顶部的 [ODBC Data Sources] 部分中定义的数据源名称 (DSN)，它应与 `odbc.ini` 的较下方部分中定义 DSN 的方括号之间指定的字符串相符。如果二者不同，则您在 IBM® SPSS® Modeler 中使用该 DSN 进行连接时将看到此错误。以下示例显示了错误的指定：

```
[ODBC Data Sources]
Oracle=Oracle Wire Protocol

...
...
[Oracle Driver]
Driver=/usr/ODBC/lib/XEora22. so
Description=SPSS 5.2 Oracle Wire Protocol
AlternateServers=
...
```

您需要更改两个黑体字符串之一，以确保它们完全相同。这样即可解决此错误。

无法加载指定的驱动程序

该错误也表明 `odbc.ini` 文件定义不正确。一种可能是在驱动程序部分中的驱动程序参数设置不正确，例如：

```
[ODBC Data Sources]
Oracle=Oracle Wire Protocol

...
...
[Oracle]
Driver=/nosuchpath/ODBC/lib/XEora22. so
Description=SPSS 5.2 Oracle Wire Protocol
AlternateServers=
```

- ▶ 检查驱动程序参数指定的共享对象是否存在。
- ▶ 如果不正确，则纠正共享对象的路径。
- ▶ 如果其中以此格式指定驱动程序参数：

```
Driver=ODBCHOME/lib/XEora22. so
```

这表明您尚未初始化 ODBC 相关脚本。运行随驱动程序安装的 `setodbcpath.sh` 脚本。[有关详细信息，请参阅第 63 页码 IBM SPSS Modeler Server 和数据访问包。](#) 在您运行此脚本后，应能看到字符串 “ODBCHOME” 已被替换为您的 ODBC 安装的路径。这应当可以解决该问题。

另一种原因可能是驱动程序库存在问题。请使用 ODBC 提供的 `ivtestlib` 工具来确认是否无法加载驱动程序。对于 `Connect64`，请使用 `ddtestlib` 工具。通过在启动脚本中设置库路径变量排除上述故障。

例如，如果对于 32 位安装无法加载 Oracle 驱动程序，请执行以下步骤：

- ▶ 使用 `ivtestlib` 工具来确认是否无法加载驱动程序。例如，在 UNIX 提示符处输入：

```
sh
cd ODBC DIR
. odbc.sh
```

```
./bin/ivtestlib MFor815
```

在此将用 ODBC 安装目录的路径来替换 ODBC DIR。

- ▶ 查阅消息，看是否存在错误。例如，消息为：

```
MFor815. so 加载失败:ld. so. 1:bin/ivtestlib:严重错误:libclntsh. so:无法打开:文件或目录不存在
上述消息指出: Oracle 客户端库 (libclntsh. so) 丢失或其库路径错误 (例如, 在 Solaris 中, 客户端库的路径不是 LD_LIBRARY_PATH)。
```

- ▶ 确认库的存在。如果库不存在，则请重新安装 Oracle 客户端。如果库存在，则请在 UNIX 命令提示符处输入以下命令序列：

```
LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/bigdisk/oracle/product/8.1.6/lib
export LD_LIBRARY_PATH
./bin/ivtestlib Mfor815
```

在此将用 libclntsh.so 的路径来替换 /bigdisk/oracle/product/8.1.6/lib, LD_LIBRARY_PATH 是针对操作系统的库路径变量。

请注意，如果正在 Linux 或 Solaris 上运行 64 位 SPSS Modeler，则库路径变量包含后缀 _64。因此，上一示例中的前两行将变为：

```
LD_LIBRARY_PATH_64=$LD_LIBRARY_PATH_64:/bigdisk/oracle/product/8.1.6/lib
export LD_LIBRARY_PATH_64
```

- ▶ 查阅消息以确认现在是否可以加载驱动程序。例如，消息为：

```
MFor815. so 加载成功, qehandle 为 0xFF3A1BE4
此消息说明可以加载 Oracle 客户端库。
```

- ▶ 更正 SPSS Modeler 启动脚本中的库路径。
- ▶ 使用编辑过的启动脚本 (modelersrv.sh 或 rc.modeler) 重新启动 SPSS Modeler Server。

库路径

库路径变量的名称取决于您所用的操作系统。当对系统进行配置或故障排除时，下表将作为指南来指导您做出相应的替换。

表 B-2
不同操作系统的库路径

操作系统	库路径变量名称
AIX	LIBPATH
HP-UX	SHLIB_PATH
Solaris 或 Linux 32 位	LD_LIBRARY_PATH
Solaris 或 Linux 64 位	LD_LIBRARY_PATH_64

为 Enterprise View 节点配置驱动程序

使用 Enterprise View 节点，可以在共享的 IBM® SPSS® Collaboration and Deployment Services Repository 中创建并维护 IBM® SPSS® Modeler 会话和 IBM SPSS Collaboration and Deployment Services Enterprise View 间的连接。[有关详细信息，请参阅第 9 章中的存储和部署存储库对象中的 IBM SPSS Modeler 15 用户指南。](#)

要将 Enterprise View 节点用于 IBM® SPSS® Modeler Server，必须安装并配置 IBM® SPSS® Collaboration and Deployment Services Enterprise View Driver。有关安装此驱动程序的详细信息，请与本地管理员联系。安装后，必须在启动脚本 中添加对 pev.sh 的引用（添加到 modelersrv.sh、rc.modeler 或同时添加到这两者中），格式如下：

```
. /usr/odbc/pev.sh
```

在 UNIX 上以非根进程方式配置和运行 SPSS Modeler Server

简介

这些说明提供了在 UNIX 系统上以非根进程方式运行 IBM® SPSS® Modeler Server 的相关信息。

以根进程方式运行。 SPSS Modeler Server 的默认安装程序会假设服务器守护进程以根进程方式运行。以根进程方式运行允许 IBM® SPSS® Modeler 对每位用户的登录进行安全验证，并且会在对应的 UNIX 用户帐户上启动用户会话。这样可以确保用户只能访问他们各自的文件和目录。但是，可能会按照下列说明在非根帐户上运行守护进程。

以非根进程方式运行。 以非根进程方式运行 SPSS Modeler Server 表示具有为所选帐户设置的服务器守护进程的真实有效的用户 ID。所有由 SPSS Modeler Server 启动的用户会话将使用相同的 UNIX 帐户，这意味着任何由 SPSS Modeler 读取和写入的文件数据将由所有 SPSS Modeler 用户共享。由于用户必须分别在所用的每个数据库数据源上验证自己的身份，因此对数据库进行的访问不会受到影响。SPSS Modeler 不需根特权即可按下列两种方式之一进行操作：

- **不使用专有密码数据库。** 用这种方法，SPSS Modeler 将使用 UNIX 系统上常用于用户验证的现有 UNIX 密码数据库、NIS 或 LDAP 服务器。 [有关详细信息，请参阅第 70 页码不使用专有密码数据库以非根方式配置。](#)
- **使用专有密码数据库。** 用这种方法，SPSS Modeler 根据专有密码库验证用户，这不同于 UNIX 上常用于验证的 UNIX 密码数据库、NIS 或 LDAP 服务器。 [有关详细信息，请参阅第 71 页码使用专有密码数据库以非根方式配置。](#)

不使用专有密码数据库以非根方式配置

要配置 IBM® SPSS® Modeler Server 在非根帐户上运行且无需专有密码数据库，请执行下列步骤：

1. 打开 SPSS Modeler Server options.cfg 文件进行编辑。
2. 将选项 `start_process_as_login_user` 设置为 Y。
3. 保存并关闭 options.cfg 文件。

警告： 不要启用 `start_process_as_login_user` 设置，然后将 SPSS Modeler Server 作为 root 启动。这意味着，对于所有连接到服务器的用户而言，其服务器进程将运行作为 root；此举将造成安全风险。

使用专有密码数据库以非根方式配置

如果您选择使用专有密码数据库的方式验证用户，所有用户会话将在同一个非根用户帐户上启动。但您必须仍然以 `root` 或 `su` 为 `root` 身份进行登录，以启动和停止服务器，然后执行常规管理。

要用此方法配置 IBM® SPSS® Modeler Server 在非根帐户上运行，请执行下列步骤：

1. 创建包含所有用户的组。可以为该组命名任意名称，例如，称它为 `modelerusers`。
2. 创建在其上运行 SPSS Modeler Server 的用户帐户。该帐户仅用于 SPSS Modeler Server 守护进程。换句话说，您无法使用该帐户启动和停止服务器，或执行常规 IBM® SPSS® Modeler 管理。例如，称它为 `modelerserv`。

创建帐户时，请注意下列事项：

- 帐户无需登录功能。
 - 主要的组应为先前创建的 `<modelerusers>` 组。
 - 主目录可以是 SPSS Modeler 的安装目录或任何其他适当的默认目录（如果需要通过帐户来继续进行升级，请考虑使用除安装目录之外的目录）。
3. 然后，配置启动脚本，以使用新建帐户启动 SPSS Modeler Server。找到相应的启动脚本并在文本编辑器中将其打开。有关详细信息，请参阅第 61 页码附录 B 中的脚本。
 - ▶ 在相应的脚本中，更改设置 `CLEMUSER` 变量的行以引用所建用户帐户的名称，从而运行 SPSS Modeler Server。例如：

```
CLEMUSER=modelerserv
```

- ▶ 更改 `umask` 设置，以允许在下列创建的文件上至少可以进行组读取访问：

```
umask 027
```

4. 编辑服务器选项文件 `config/options.cfg`，以追加下列行通过专有密码数据库指定验证：

```
authentication_methods, "pasw_modeler"
```

5. 然后，需要创建存储在文件 `config/passwords.cfg` 中的专有密码数据库。密码文件定义允许登录到 SPSS Modeler 的用户名/密码组合。注意：这些用户名和密码仅供 SPSS Modeler 使用，与用于登录到 UNIX 的用户名和密码无关。为方便起见，可以使用相同的用户名，但不能使用相同的密码。

要创建密码文件，需要使用密码实用程序 `pwutil`，该程序位于 SPSS Modeler Server 安装程序的分级目录中。该程序的概要如下所示：

```
pwutil [ username [ password ] ]
```

该程序使用用户名和纯文本密码，并且采用适合于包含在密码文件中的格式将用户名和加密密码写入标准输出。例如，要定义其密码为“data mining”的用户 `modeler`，您可以键入以下内容：

```
bin/pwutil modeler "data mining" > config/passwords.cfg
```

大多数情况下，定义一个用户名就足够了，这样所有用户都使用同一用户名和密码进行登录。但是，也可以创建其他用户，方法是：通过使用 >> 运算符在该文件中追加每个用户，例如：

```
bin/pwutil modeler "data miner2" >> config/passwords.cfg
```

注意：如果使用一个 >，则每次会覆盖 passwords.cfg 的内容，从而会替换先前设置的所有用户。记住，无论创建多少用户，所有用户都会共享同一 UNIX 用户帐户。

6. 递归式地将 SPSS Modeler 安装目录的所有权以及其内容更改为用户 <modelerserv> 和组 <modelerusers>，其中所引用的名称为先前所创建的名称。例如：

```
chown -R -h modelerserv:modelerusers .
```

7. 考虑为 SPSS Modeler 用户在数据目录中创建子目录，以便他们有足够的空间在互不影响的情况下存储工作数据。这些目录应按 <modelerusers> 组进行分组，并具备组读写和组搜索的权限。例如，为用户 bob 创建工作目录：

```
mkdir data/bob
chown bob:modelerusers data/bob
chmod ug=rwx,o= data/bob
```

另外，可以在目录上设置设置组 ID 位，以便任何复制到目录的数据文件自动按 <modelerusers> 进行分组：

```
chmod g+s data/bob
```

作为非根用户运行 SPSS Modeler Server

要作为非根用户运行 SPSS Modeler Server，请执行下列步骤：

1. 使用先前创建的非根用户帐户登录。
2. 如果您运行时启用了配置文件选项 `start_process_as_login_user`，您可以启动和停止 SPSS Modeler Server 以及检查其状态。[有关详细信息，请参阅第 20 页码第 4 章中的在 UNIX 中启动、停止和检查状态。](#)

最终用户可通过从客户端软件中登录而连接到 SPSS Modeler Server。必须将用户连接所需的信息给予最终用户，这些信息包括 IP 地址或服务器计算机的主机名。

在 AIX 上作为非根用户运行 IBM SPSS Modeler Server

当在 AIX 上作为非根用户启动 IBM® SPSS® Modeler Server 时，错误 `lsattr not found` 可能将被输出到控制台。这是因为 `lsattr` 位于 `/usr/sbin` 目录中，但是当 SPSS Modeler Server 登录一个新用户（以作为非根运行）时会重置路径，只包括 `/usr/bin` 和 `/bin`。

如果出现此错误，解决方案是在 `/usr/bin` 下创建以下链接：

```
su
cd /usr/bin
ln -s /usr/sbin/lsattr lsattr
```

创建链接后，请重新启动 SPSS Modeler Server。

用户验证失败故障排除

根据配置操作系统进行验证的方式，当运行于非根配置时您可能会遇到登录 SPSS Modeler Server 失败的情况。例如，如果您的操作系统配置为（使用 `/etc/nsswitch.conf` 或类似的文件）检查本地 shadow 密码文件，而非使用 NIS 或 LDAP，将会出现以上情况。发生这种情况是因为 SPSS Modeler Server 要求对于执行验证的文件具有访问权限，包括 `/etc/shadow` 文件或同类文件（例如，AIX 上的 `/etc/security/passwd`），其中存储安全用户帐户信息。但是，通常会设置操作系统文件权限，因此 `/etc/shadow` 仅能由根用户访问。在这些情况下，非根进程无法读取 `/etc/shadow` 以验证用户密码，即导致验证错误。

有几个方法可以解决此问题：

- 请要求您的系统管理员将操作系统配置为使用 NIS 或 LDAP 验证。
- 更改受保护文件的文件权限，例如可授予 `/etc/shadow` 文件读取权限，如此一来，运行 SPSS Modeler Server 的本地用户帐户即可访问该文件。虽然此解决方式在生产环境中不合宜，但仍可在测试环境中暂时应用此方式，以检验验证失败是否与操作系统配置有关。
- 为 `/etc/shadow` 文件指定访问控制列表（ACL）（可参阅本节稍后说明的示例）。
- 以根进程方式运行 SPSS Modeler Server，可使服务器进程读取 `/etc/shadow` 文件。警告：在这种情况下，请确保 SPSS Modeler Server 的 `options.cfg` 文件包含选项 `start_process_as_login_user, N`，以避免之前提到的安全问题。

访问控制列表（ACL）示例

在 Solaris 操作系统上，您可以使用 `setfacl` 命令创建访问控制列表（请与您的管理员联系以确保此命令在您的系统上可用）。以下示例显示如何为用户 `modserv` 创建此类 ACL：

```
# vi /tmp/shadow.acl
user::r--
user:modserv:r--
group:---
mask:r--
other:---

# setfacl -f /tmp/shadow.acl /etc/shadow

# getfacl /etc/shadow

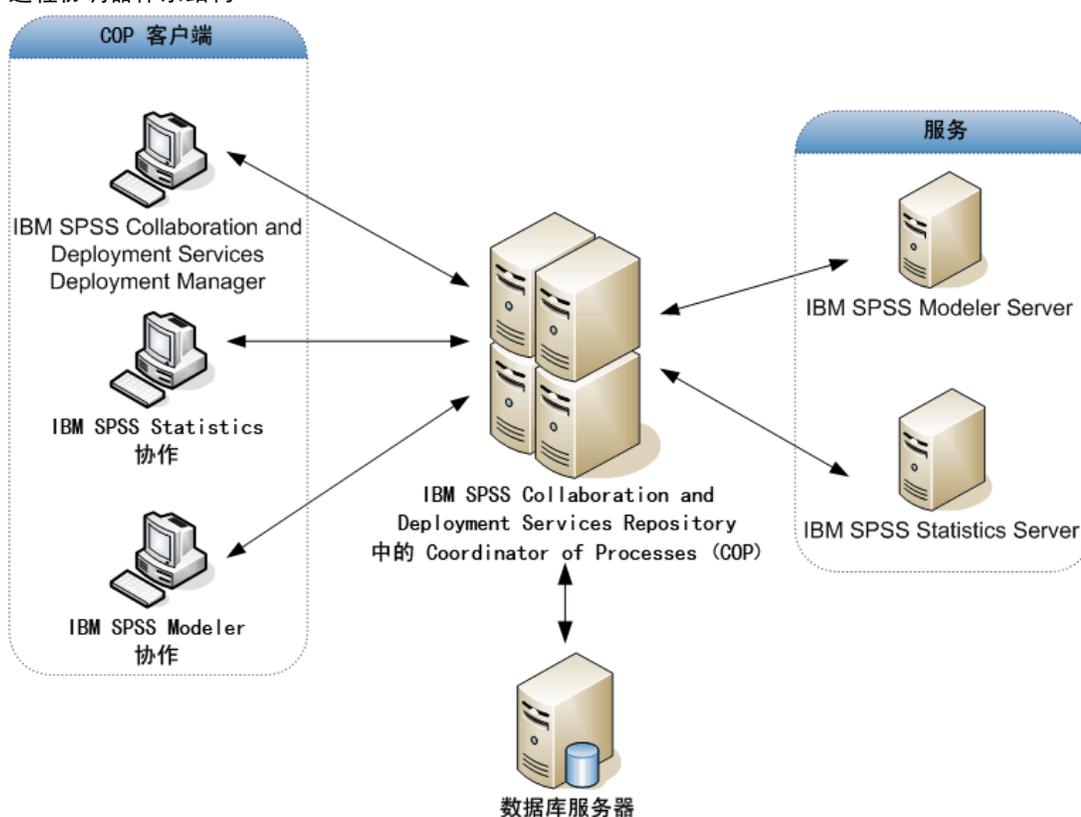
# file: /etc/shadow
# owner: root
# group: sys
user::r--
user:modserv:r--      #effective:r--
group:---             #effective:---
mask:r--
other:---
```

使用服务器群集负载均衡

对于 IBM® SPSS® Collaboration and Deployment Services, 可使用名为进程协调器的插件来管理网络上的服务。进程协调器提供了服务器管理功能, 旨在优化客户端和服务端之间的通信和处理。

要管理的服务, 如 IBM® SPSS® Statistics Server 或 IBM® SPSS® Modeler Server, 在启动时使用进程协调器注册并定期发送更新的状态消息。服务还可在 IBM® SPSS® Collaboration and Deployment Services Repository 中存储任何必要的配置文件, 并在初始化时检索这些文件。

图片 D-1
进程协调器体系结构



在服务器上执行 IBM® SPSS® Modeler 流可以增强性能。在某些情况下, 只能选择一个或两个服务器。在另一些情况下, 由于各个服务器之间存在本质差异, 如服务器所有者、访问权限、服务器数据、测试及制作服务器等, 所以可选择的服务器很多。此外, 如果在网络上安装了进程协调器, 则会向您提供一个服务器群集。

服务器群集是一组可以根据配置和资源互相交换的服务器。进程协调器使用根据多个标准（包括服务器加权、用户属性和当前处理量）平衡负载的算法，确定最适合响应处理要求的服务器。有关详细信息，请参阅《进程协调器服务开发人员指南》，可从 IBM Corp. 索取。

无论何时连接到 SPSS Modeler 中的服务器或服务器群集，都可以手动输入服务器名称或使用进程协调器搜索服务器或服务器群集。[有关详细信息，请参阅第 3 章中的连接到 IBM SPSS Modeler Server 中的 IBM SPSS Modeler 15 用户指南。](#)

注意事项

这些信息开发用于在全球提供的产品和服务。

IBM 可能在其他国家/地区中不提供在本文档中讨论的产品、服务或功能。请咨询您当地的 IBM 代表以了解有关您所在地区当前可用产品和服务的信息。任何对 IBM 产品、程序或服务的引用，并不意味着仅可使用这些 IBM 产品、程序或服务。作为替代，可以使用任何功能相当的产品、程序或服务，前提是不侵犯任何 IBM 知识产权。不过，用户应自行负责评估和验证任何非 IBM 产品、程序或服务的工作情况。

在本文档中介绍的主题可能涉及 IBM 的专利或申请中的专利。您获得该文档并不意味着授予您任何这些专利许可。您可以将书面许可查询函件发送至：

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

要查询双字节字符集 (DBCS) 相关许可证信息，请联系所在国家/地区中的 IBM 知识产权部门，或者以书面形式将查询函件发送至：

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

以下段落不适用于英国或此类条款与当地法律不符的其他国家/地区： INTERNATIONAL BUSINESS MACHINES 公司“按原样”提供本出版物，不保证任何明示或暗示，包括但不限于对非侵权性、适销性或对特定用途适用性的暗示担保。有些州/省不允许针对某些交易的明示或暗示免责条款，因此本声明可能不适用于您。

本信息可能含有技术误差或排版错误。此处的信息会定期进行更改；这些更改会体现在本出版物的新版本中。IBM 可能随时对本出版物中所述的产品和/或程序进行改进和/或更改，恕不另行通知。

在本信息材料中对任何非 IBM 网站的引用仅为了方便用户，并不以任何方式表明对这些网站的认可。这些网站上的材料并非本 IBM 产品材料的一部分，您对这些网站的使用需自担风险。

IBM 可以自认为适当并且不会对您构成任何约束的任何方式使用或分发您提供的任何信息。

如果本程序的受许可方试图了解有关程序的信息以启用：(i) 在独立创建的程序和其他程序（包括本程序）之间交换信息；(ii) 相互使用交换的信息，则应联系：

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

此类信息的提供应遵照相关条款和条件，其中包括在某些情况下支付适当费用。

在本文档中介绍的受许可保护程序，及其所有受许可保护材料由 IBM 在双方签署的“IBM 客户协议”、“IBM 国际程序许可证协议”或任何其他等同协议下提供。

此处所含的性能数据均在受控环境下决定。因此，在其他操作环境中获得的结果可能差异较大。有些测量可能在开发级的系统中进行，不保证这些测量结果与常用系统上的测量结果相同。此外，有些测量结果可能通过推断来估计得出。实际结果可能有所差异。此文档的用户应针对其具体环境验证适用的数据。

非 IBM 产品的相关信息来自这些产品的供应商，及其发布的公告或其他公开来源。IBM 并未测试过这些产品，无法确认有关非 IBM 产品的性能准确性、兼容性或任何其他声明。有关非 IBM 产品功能的问题应由这些产品的供应商负责。

有关 IBM 未来方向或意向的所有声明均可能未经通知即变更或撤销，并且仅代表目标和目的。

本信息包含日常业务运营中使用的数据和报告的示例。为了尽可能详尽地对其进行说明，示例中包含了人员的姓名、公司、品牌和产品的名称。所有这些名称均为虚构，与真实商业企业使用的名称和地址的任何雷同纯属巧合。

如果您正在查阅此信息的软拷贝，照片和彩色插图可能不会显示。

商标

IBM、IBM 徽标、ibm.com 和 SPSS 是 IBM Corporation 的商标，在全球许多司法辖区注册。有关最新的 IBM 商标列表，请访问网页 <http://www.ibm.com/legal/copytrade.shtml>。

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 或其子公司在美国和其他国家或地区的商标或注册商标。

Linux 是 Linus Torvalds 在美国、其他国家或地区或两者的注册商标。

Microsoft、Windows、Windows NT 和 Windows 徽标是 Microsoft Corporation 在美国、其他国家或地区或两者的商标。

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 Sun Microsystems, Inc. 在美国、其他国家或地区或两者的商标。

其他产品和服务名称可能是 IBM 或其他公司的商标。



- 64 位操作系统, 9
- allow_modelling_memory_override
 - options.cfg 文件, 27
- auto.sh (UNIX)
 - 的位置, 61
- cache_compression
 - options.cfg 文件, 27
- cache_connection 选项, 32
- chemsrv.sh (UNIX)
 - 的位置, 61
- CLEM表达式
 - SQL 生成, 55
- COP, 17
 - 服务器群集, 74
 - 负载均衡, 74
- COP 配置
 - 适用于 IBM SPSS Modeler Server, 29
- cop_enabled
 - options.cfg 文件, 29
- cop_host
 - options.cfg 文件, 29
- cop_password
 - options.cfg 文件, 29
- cop_port_number
 - options.cfg 文件, 29
- cop_service_description
 - options.cfg 文件, 29
- cop_service_host
 - options.cfg 文件, 29
- cop_service_name
 - options.cfg 文件, 29
- cop_service_weight
 - options.cfg 文件, 29
- cop_update_interval
 - options.cfg 文件, 29
- cop_user_name
 - options.cfg 文件, 29
- data_file_path
 - options.cfg 文件, 26
- data_files_restricted
 - options.cfg 文件, 26
- DB2
 - SQL 优化, 45 - 46
- IBM SPSS Modeler, 1
 - 文档, 3
- IBM SPSS Modeler Administration Console, 22
 - 管理员访问权限, 23
- IBM SPSS Modeler Server
 - COP 配置, 29
 - password, 14
 - temp 目录, 26
 - 不响应进程, 21
 - 与客户端的不同结果, 19
 - 主机名, 14, 16
 - 供最终用户阅读的信息, 18
 - 域名 (Windows), 14
 - 文件创建, 18
 - 日志文件, 25
 - 服务器进程, 31
 - 用户 ID, 14
 - 用户帐户, 18
 - 用户验证, 18
 - 监视使用情况, 31
 - 端口号, 14, 16, 25
 - 管理, 22
 - 管理员访问权限, 23
 - 管理选项, 22
 - 进程协调器配置, 29
 - 配置选项, 24
- IBM SPSS Modeler Server 的
 - 硬件建议, 9
- IBM SPSS Statistics 数据文件
 - 导入和导出, 13
- IBM SPSS Statistics 数据访问技术, 11
- IBM SPSS Statistics 许可证位置, 32
- io_buffer_size
 - options.cfg 文件, 27
- LDAP, 38
 - 安全, 38
- log_database_access
 - options.cfg 文件, 25
- log_database_errors
 - options.cfg 文件, 25
- log_directory
 - options.cfg 文件, 25
- log_file_access
 - options.cfg 文件, 25
- log_file_errors
 - options.cfg 文件, 25
- max_file_size
 - options.cfg 文件, 26
- max_login_attempts
 - options.cfg 文件, 25
- max_parallelism
 - options.cfg 文件, 27
- max_sessions
 - options.cfg 文件, 25
- max_sql_string_length
 - options.cfg 文件, 28
- memory_usage
 - options.cfg 文件, 27
- messages.log 文件, 39
- Microsoft SQL Server
 - SQL 优化, 45 - 46
- modelling_memory_limit_percentage
 - options.cfg 文件, 27
- ODBC
 - 基于 UNIX 进行配置, 63
- ODBC 数据源
 - ODBC 和 UNIX 脚本, 63

- 和 UNIX, 63
- options.cfg 文件, 31
- Oracle
 - SQL 优化, 45 - 46, 59
- password
 - IBM SPSS Modeler Server, 14
- port_number
 - options.cfg 文件, 25
- program_file_path
 - options.cfg 文件, 26
- program_files_restricted
 - options.cfg 文件, 26
- RAM, 11
- rc.modeler (UNIX)
 - 的位置, 61
- SPSS Modeler Server, 2
- SQL
 - 优化 Oracle, 59
 - 查看模型块, 50
 - 查询, 57
 - 重复列名, 57
 - 预览生成的, 49
- SQL Server
 - SQL 优化, 45 - 46
- SQL 回送。请参阅 SQL 生成, 45
- SQL 生成, 41, 45 - 46, 52
 - CLEM表达式, 51, 55
 - 启用, 48
 - 对 IBM SPSS Modeler Server 启用, 28
 - 提示, 51
 - 日志记录, 49
 - 查看模型块, 50
 - 流重写, 51
 - 预览, 49
- sql_generation_enabled
 - options.cfg 文件, 28
- @SQLFN 函数, 57
- SSL, 33
 - 概述, 33
 - 确保通信安全, 33
- SSL 数据加密
 - 对 IBM SPSS Modeler Server 启用, 29
- ssl_certificate_file
 - options.cfg 文件, 29
- ssl_enabled
 - options.cfg 文件, 29
- ssl_private_key_file
 - options.cfg 文件, 29
- ssl_private_key_password
 - options.cfg 文件, 29
- stream_rewriting_enabled
 - options.cfg 文件, 27
- temp 目录
 - 适用于 IBM SPSS Modeler Server, 26
- temp_directory
 - options.cfg 文件, 26
- UNC 文件名, 13
- UNIX
 - 库路径, 68
 - 用户验证, 18
 - 配置文件权限, 63
 - UNIX 上的内核函数限制, 21
 - UNIX 内核函数限制, 21
 - UNIX 脚本
 - auto.sh, 61
 - modelersrv.sh, 61
 - rc.modeler, 61
 - 编辑, 62
 - URL 前缀, 37
- 临时文件, 10
 - 权限 (IBM SPSS Modeler Server), 18
- 主机名
 - IBM SPSS Modeler Server, 14, 16
- 优化, 41
 - SQL 生成, 45 - 46, 48
- 体系结构
 - 组件, 7
- 内存, 11
- 内存管理
 - 管理选项, 27
- 加密
 - SSL, 33
- 单点登录, 15
- 商标, 77
- 回送, 41, 45 - 46, 48, 52
 - CLEM表达式, 55
- 在 UNIX 上
 - IBM SPSS Modeler Server 的状态, 20
 - 停止 IBM SPSS Modeler Server, 20
 - 启动 IBM SPSS Modeler Server, 20
- 在 Windows 上
 - IBM SPSS Modeler Server 的状态, 20
 - 停止 IBM SPSS Modeler Server, 20
 - 启动 IBM SPSS Modeler Server, 20
- 域名 (Windows)
 - IBM SPSS Modeler Server, 14
- 处理器, 9
 - 多个, 27
- 多个流执行, 28

索引

- 安全
 - LDAP, 38
 - SSL, 33
 - 文件创建, 18
 - 配置 UNIX 中的文件创建, 63
- 安全套接字层, 33
- 层, 数据库支持, 11
- 并行处理
 - 启用, 41
 - 控制, 27
- 应用程序示例, 3
- 建模
 - 内存管理, 27
- 性能
 - IBM SPSS Modeler Server, 41
- 执行流时出现错误, 28
- 操作系统
 - 64 位, 9
- 数据库
 - 支持层, 11
 - 访问, 11
 - 数据库内挖掘, 43
 - 数据库内高速缓存, 31
 - 数据库服务器, 43
 - 数据库连接
 - 关闭, 32
 - 数据库高速缓存
 - SQL 生成, 51
 - 从 options.cfg 中控制, 31
 - 数据文件
 - IBM SPSS Statistics, 13
 - 导入和导出, 13
 - 数据访问, 11
 - 数据访问包
 - ODBC (基于 UNIX 进行配置), 63
 - UNIX 中的 ODBC 故障排除, 65
 - 和 UNIX 库路径, 68
 - 针对下列对象配置 UNIX, 63
- 文件使用权限
 - 在 IBM SPSS Modeler Server 上, 18
 - 基于 UNIX 进行配置, 63
- 文件名
 - UNIX, 13
 - Windows, 13
- 文档, 3
- 日志文件
 - 显示生成的 SQL, 49
 - 适用于 IBM SPSS Modeler Server, 25, 39
- 服务器
 - 添加连接, 16
 - 登录, 14
 - 通过 COP 搜索服务器, 17
 - 服务器的自动启动
 - 基于 UNIX 进行配置, 61
 - 服务器端口设置
 - options.cfg 文件, 27
- 模型块
 - 查看 SQL, 50
- 死进程, IBM SPSS Modeler Server, 21
- 法律注意事项, 76
- 流重写, 51
 - 启用, 41
- 消息
 - 显示生成的 SQL, 49
- 添加 IBM SPSS Modeler Server 连接, 16 - 17
- 用户 ID
 - IBM SPSS Modeler Server, 14
- 用户帐户
 - IBM SPSS Modeler Server, 18
- 用户验证, 18
- 登录到 IBM SPSS Modeler Server, 14
- 硬盘, 11
- 磁盘空间, 10
 - 计算, 10
- 示例
 - 应用程序指南, 3
 - 概述, 5
- 端口号
 - IBM SPSS Modeler Server, 14, 16, 25
- 端口设置
 - options.cfg 文件, 27
- 管理
 - IBM SPSS Modeler Server, 22
 - IBM SPSS Statistics 许可证位置, 32
 - 管理员访问权限
 - 适用于 IBM SPSS Modeler Server, 23

- 紫色节点, 41
 - SQL 优化, 46
- 结果
 - Client 和 Server 之间的差异, 19
 - 取整, 19
 - 记录顺序, 19
 - 缓存压缩, 27
- 节点
 - 支持 SQL 生成, 52
- 节点缓存
 - SQL 生成, 51
 - 写入数据库, 51
- 路径, 13
- 运算符
 - SQL 生成, 55
- 进程, 不响应, 21
- 进程协调器, 17
 - 服务器群集, 74
 - 负载均衡, 74
- 进程协调器配置
 - 适用于 IBM SPSS Modeler Server, 29
- 连接
 - 服务器群集, 17
 - 至 IBM SPSS Modeler Server, 14, 16 - 17
 - 通过 COP 搜索连接, 17
- 配置选项
 - COP, 29
 - IBM SPSS Modeler Server, 22
 - SQL 字符串长度, 28
 - SSL 数据加密, 29
 - temp 目录, 26
 - 内存管理, 27
 - 并行处理, 27
 - 性能和优化, 27
 - 数据文件访问, 26
 - 日志记录, 25
 - 概述, 24
 - 流重写, 27
 - 登录尝试, 25
 - 端口号, 25
 - 自动 SQL 生成, 28
 - 进程协调器, 29
 - 连接和会话, 25
- 防火墙设置
 - options.cfg 文件, 28
- 预览
 - SQL 生成, 49
- 验证, 18
- 高速缓存, 数据库内, 31