

IBM SPSS Modeler Server

18.1 管理和性能指南

The IBM logo, consisting of the letters "IBM" in a bold, black, sans-serif font. Each letter is composed of eight horizontal bars of varying lengths, creating a striped effect. The logo is centered on the page.

IBM

注释

在使用本资料及其支持的产品之前，请阅读第 85 页的『声明』中的信息。

产品信息

此版本适用于 IBM SPSS Modeler V18.0.0，及所有后续发行版和修订版，除非在新版本中另有说明。

目录

前言	v	IBM SPSS Modeler Server 管理	30
第 1 章 关于 IBM SPSS Modeler	1	启动 Modeler Administration Console	31
IBM SPSS Modeler 产品	1	重新启动 Web Service	31
IBM SPSS Modeler	1	使用 Modeler Administration Console 来配置访问权	31
IBM SPSS Modeler Server	1	SPSS Modeler Server 连接	32
IBM SPSS Modeler Administration Console	2	SPSS Modeler Server 配置	32
IBM SPSS Modeler Batch	2	SPSS Modeler Server 监视	39
IBM SPSS Modeler Solution Publisher	2	使用 options.cfg 文件	39
用于 IBM SPSS Collaboration and Deployment Services 的 IBM SPSS Modeler Server 适配器	2	关闭未使用的数据库连接	39
IBM SPSS Modeler 版本	2	使用 SSL 进行安全数据传输	40
IBM SPSS Modeler 文档	3	SSL 如何运行	40
SPSS Modeler Professional 文档	3	通过 SSL 确保客户端/服务器和服务器/服务器通信的安全	40
SPSS Modeler Premium 文档	3	Cognos SSL 连接	44
应用程序示例	4	Cognos TM1 SSL 连接	45
Demos 文件夹	4	配置组	45
许可证跟踪	4	服务器日志	47
第 2 章 基础架构和硬件建议	5	第 5 章 性能概述	49
IBM SPSS Modeler 体系结构	5	服务器性能和优化设置	49
基础架构描述	5	客户端性能和优化设置	49
硬件建议	6	数据库使用情况和优化	50
对临时磁盘空间和 RAM 的要求	7	SQL 优化	51
数据访问	8	第 6 章 SQL 优化	53
引用数据文件	9	SQL 生成如何运行	54
IBM SPSS Statistics 数据文件的导入	10	SQL 生成示例	54
安装说明	10	配置 SQL 优化	56
第 3 章 IBM SPSS Modeler 支持	11	预览生成的 SQL	56
正在连接到 IBM SPSS Modeler Server	11	查看模型块 SQL	57
配置单点登录	12	最大化 SQL 生成的提示	57
添加并编辑 IBM SPSS Modeler Server 连接	17	支持 SQL 生成的节点	58
搜索 IBM SPSS Collaboration and Deployment Services 中的服务器	17	支持 SQL 生成的 CLEM 表达式和运算符	61
数据和文件系统	18	使用 CLEM 表达式中的 SQL 函数	63
用户认证	18	写入 SQL 查询	63
权限	19	Scoring Adapter for Teradata - 重复行	64
文件创建	19	附录 A. 为 UNIX 平台配置 Oracle	65
不同的结果	19	配置 Oracle 的 SQL 优化	65
第 4 章 IBM SPSS Modeler 管理	21	附录 B. 配置 UNIX 启动脚本	67
启动和停止 IBM SPSS Modeler Server	21	介绍	67
在 Windows 中启动、停止和检查状态	21	脚本	67
在 UNIX 中启动、停止和检查状态	21	自动启动和停止 IBM SPSS Modeler Server	67
处理无响应的服务器进程 (UNIX 系统)	22	手动启动和停止 IBM SPSS Modeler Server	68
配置服务器概要文件	22	编辑脚本	68
使用服务器概要文件	23	控制文件创建权限	68
概要文件结构	24	IBM SPSS Modeler Server 和数据访问包	68
概要文件脚本	26	ODBC 配置故障诊断	71
管理	30	库路径	73

附录 C. 在 UNIX 上配置 SPSS Modeler Server 并使其作为非 root 用户进程运行 . 75

介绍 75
在不使用专用密码数据库的情况下配置为非 root 用户进程 75
在使用专用密码数据库的情况下配置为非 root 用户进程 76
以非 root 用户身份运行 SPSS Modeler Server 77
对用户认证故障进行故障诊断 77

附录 D. 在 Windows 上使用专用密码文件来配置和运行 SPSS Modeler Server . 79

介绍 79

配置专用密码数据库 79

附录 E. 使用服务器集群负载均衡 81

附录 F. LDAP 认证. 83

声明 85

商标 86

产品文档的条款和条件 86

索引 89

前言

IBM® SPSS® Modeler 是 IBM Corp. 企业级数据挖掘工作平台。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 软件提供了完整、一致且正确的信息，决策人可以放心地根据此信息提高业务绩效。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，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得更好的结果。

作为此产品服务组合的组成部分，IBM SPSS Predictive Analytics 软件可以帮助组织预测未来事件，并在此洞察的基础上提前行动以实现更好的业务结果。全世界的商业、政府和学校客户都依靠 IBM SPSS 技术作为吸引、保留和增加客户的竞争优势，同时减少欺诈并降低风险。通过在其日常运营中融入 IBM SPSS 软件，组织将成为前瞻型企业 - 能够指引并实现决策自动化，以满足业务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 Corp. 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。要与技术支持联系，请访问 IBM Corp. Web 站点 (<http://www.ibm.com/support>)。寻求帮助时，请准备好标识您的身份、所在组织以及支持协议。

第 1 章 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。IBM 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的整个数据挖掘过程。

IBM SPSS Modeler 提供了各种来源于机器学习、人工智能和统计学的建模方法。"建模"选用板中提供的方法使您可以根据数据派生新信息，并开发预测模型。每种方法各有所长，并且最适合于解决特定类型的问题。

SPSS Modeler 可以作为独立产品购买，也可以作为客户机与 SPSS Modeler Server 配合使用。另外，还提供了很多其他选项，以下各节概述了这些选项。有关更多信息，请参阅<https://www.ibm.com/analytics/us/en/technology/spss/>。

IBM SPSS Modeler 产品

IBM SPSS Modeler 系列产品及关联的软件包括以下各项。

- IBM SPSS Modeler
- IBM SPSS Modeler Server
- IBM SPSS Modeler Administration Console (包含在 IBM SPSS Deployment Manager 中)
- IBM SPSS Modeler Batch
- IBM SPSS Modeler Solution Publisher
- IBM SPSS Collaboration and Deployment Services 的 IBM SPSS Modeler Server 适配器

IBM SPSS Modeler

SPSS Modeler 是产品的完整功能版本，您可以在个人计算机上安装并运行此版本。可以在本地方式下将 SPSS Modeler 作为独立产品运行，也可以在分布式方式下将其与 IBM SPSS Modeler Server 配合使用，以提高大型数据集的性能。

借助 SPSS Modeler，您可以快速直观地构建准确的预测模型而无需进行编程。通过使用独特的可视界面，可以轻松实现数据挖掘过程的可视化。在本产品中提供的高级分析的支持下，您可以发现数据中先前隐藏的模式和趋势。可以对结果进行建模并了解影响结果的因素，这使您可以利用业务机会并降低风险。

现已推出两个版本的 SPSS Modeler：SPSS Modeler Professional 和 SPSS Modeler Premium。请参阅主题第 2 页的『IBM SPSS Modeler 版本』以获取更多信息。

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 是需要单独许可证的产品，在分布式分析方式下它在服务器主机上与一个或多个 IBM SPSS Modeler 安装一起连续运行。通过此方式，SPSS Modeler Server 极大地提高了大型数据集的性能，因为可以在服务器上完成内存密集型操作，而无需将数据下载到客户端计算机。IBM SPSS Modeler Server 还提供对 SQL 优化和数据库内建模功能的支持，从而在性能和自动化方面提供更多优势。

IBM SPSS Modeler Administration Console

Modeler Administration Console 是一个图形用户界面，用于管理多个 SPSS Modeler Server 配置选项，这些选项还可以通过选项文件进行配置。控制台包含在 IBM SPSS Deployment Manager，可以用于监视和配置 SPSS Modeler Server 安装，并且可供当前 SPSS Modeler Server 客户免费使用。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

数据挖掘通常是交互过程，因此，还可以从命令行运行 SPSS Modeler 而不需要图形用户界面。例如，您可能具有长时间运行或重复任务，并且希望在用户不进行干预的情况下执行这些任务。SPSS Modeler Batch 是该产品的一个特殊版本，可提供对 SPSS Modeler 完整分析性能的支持，而无需访问常规的用户界面。要使用 SPSS Modeler Batch，需要 SPSS Modeler Server。

IBM SPSS Modeler Solution Publisher

SPSS Modeler Solution Publisher 是一个工具，它使您能够创建 SPSS Modeler 流的打包版本，该版本的流可以由外部运行时引擎运行或者可以嵌入在外部应用程序中。通过此方式，您可以发布和部署完整的 SPSS Modeler 流，以便在未安装 SPSS Modeler 的环境中进行使用。SPSS Modeler Solution Publisher 作为 IBM SPSS Collaboration and Deployment Services - Scoring 服务的组成部分进行分发，需要单独的许可证。通过此许可证，您将接收 SPSS Modeler Solution Publisher Runtime，它使您可以执行已发布的流。

有关 SPSS Modeler Solution Publisher 的更多信息，请参阅 IBM SPSS Collaboration and Deployment Services 文档。IBM SPSS Collaboration and Deployment Services Knowledge Center 包含名为"IBM SPSS Modeler Solution Publisher"和"IBM SPSS Analytics Toolkit"的部分。

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

提供了一些用于 IBM SPSS Collaboration and Deployment Services 的适配器，这些适配器使 SPSS Modeler 和 SPSS Modeler Server 可以与 IBM SPSS Collaboration and Deployment Services 存储库进行交互。通过这种方式，部署到该存储库的 SPSS Modeler 流可以由多个用户共享，也可以从瘦客户机应用程序 IBM SPSS Modeler Advantage 进行访问。将在适配器安装在主管该存储库的系统上。

IBM SPSS Modeler 版本

SPSS Modeler 推出了下列版本。

SPSS Modeler Professional

SPSS Modeler Professional 提供处理大多数类型的结构化数据所需要的所有工具，例如 CRM 系统中跟踪的行为和交互、人口统计信息、采购行为和销售数据。

SPSS Modeler Premium

SPSS Modeler Premium 是一个单独授权的产品，它对 SPSS Modeler Professional 进行了扩展，以便后者能够处理专门的数据和非结构化文本数据。SPSS Modeler Premium 包含 IBM SPSS Modeler Text Analytics：

IBM SPSS Modeler Text Analytics 使用先进的语言技术和自然语言处理 (NLP) 来快速处理各种各样的无结构文本数据、抽取和组织关键概念，以及将这些概念分组为类别。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 IBM 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

文档可以从 SPSS Modeler 中的“帮助”菜单获取。这样会打开可在产品外部公开访问的 SPSS Modeler Knowledge Center。

作为产品下载的一部分，还会在单独的压缩文件夹中以 PDF 格式提供每个产品的完整文档（包括安装指示信息）。也可以从 Web 下载 PDF 文档，地址为：<http://www.ibm.com/support/docview.wss?uid=swg27046871>。

SPSS Modeler Professional 文档

SPSS Modeler Professional 文档套件（安装指示信息除外）如下。

- **IBM SPSS Modeler 用户指南。**对于使用 SPSS Modeler 的一般简介，包括如何构建数据流、处理缺失值、构建 CLEM 表达式、处理项目和报告以及将用于部署的流打包到 IBM SPSS Collaboration and Deployment Services 或 IBM SPSS Modeler Advantage。
- **IBM SPSS Modeler Source、Process 和 Output 节点。**描述用于以不同格式读取、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 Modeling 节点。**描述所有用于创建数据挖掘模型的节点。IBM 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。
- **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。请参阅主题第 4 页的『应用程序示例』，了解更多信息。
- **IBM SPSS Modeler Python 脚本编制和自动化。**通过编写 Python 脚本实现系统自动化的相关信息，其中包括可以用于处理节点和流的属性的信息。
- **IBM SPSS Modeler 部署指南。**有关在 IBM SPSS Deployment Manager 下以处理作业的步骤形式运行 IBM SPSS Modeler 流的信息。
- **IBM SPSS Modeler CLEF 开发者指南。**CLEF 提供了将第三程序（例如，数据处理例程或建模算法）作为节点集成到 IBM SPSS Modeler 的功能。
- **IBM SPSS Modeler 数据库内挖掘指南。**有关如何利用数据库的功能通过第三方算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。**提供有关如何配置和管理 IBM SPSS Modeler Server 的信息。
- **IBM SPSS Deployment Manager 用户指南。**有关使用 Deployment Manager 应用程序中包含的管理控制台用户界面来监视和配置 IBM SPSS Modeler Server 的信息。
- **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。**提供在批处理方式下使用 IBM SPSS Modeler 的完整指导，包括批处理方式执行和命令行自变量的详细信息。本指南仅以 PDF 格式提供。

SPSS Modeler Premium 文档

SPSS Modeler Premium 文档套件（安装指示信息除外）如下。

- **SPSS Modeler Text Analytics 用户指南。**提供有关通过 SPSS Modeler 使用文本分析的信息，涵盖文本挖掘节点、交互式工作台、模板和其他资源。

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要的针对性说明。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小得多，但涉及的概念和方法可扩展到实际应用程序。

要访问示例，请在 SPSS Modeler 中单击“帮助”菜单中的**应用程序示例**。

数据文件和样本流安装在产品安装目录下的 Demos 文件夹中。有关更多信息，请参阅『Demos 文件夹』。

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。

脚本编制示例。请参阅 *IBM SPSS Modeler 脚本编写和自动化指南* 中的示例。

Demos 文件夹

在产品安装目录下的 Demos 文件夹中安装用于应用程序示例的数据文件和样本流（例如：C:\Program Files\IBM\SPSS\Modeler\

许可证跟踪

在使用 SPSS Modeler 时，将定期跟踪并记录许可证使用情况。记录的许可证度量为 *AUTHORIZED_USER* 和 *CONCURRENT_USER*，并且记录的度量类型取决于 SPSS Modeler 的许可证类型。

IBM License Metric Tool 可处理生成的日志文件，您可以从中生成许可证使用情况报告。

在记录 SPSS Modeler 客户机日志文件的相同目录中创建许可证日志文件（缺省情况下，%ALLUSERSPROFILE%/IBM/SPSS/Modeler/<version>/log）。

第 2 章 基础架构和硬件建议

IBM SPSS Modeler 体系结构

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 IBM 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 采用三层分布式体系结构。服务器计算机与客户端计算机共享软件操作。安装和使用 IBM 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：

- IBM SPSS Modeler Server 不仅可以在 Windows 系统上运行，还可以在 UNIX 上运行，这使您在确定安装位置时可以更加灵活。在任何平台上，都可以指定速度更快、更大型的服务器计算机来执行数据挖掘过程。
- 为了实现最佳性能，IBM SPSS Modeler Server 进行了优化。当无法将操作项推送到数据库中时，IBM 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 IBM SPSS Modeler Server 可以对大型的数据集执行排序、合并和汇总操作。
- 通过使用客户端/服务器体系结构，您可以将组织内的数据挖掘过程集中到一起执行。集中化有助于规范数据挖掘在业务流程中的角色。
- 使用管理员工具，例如，IBM SPSS Modeler Administration Console（IBM SPSS Deployment Manager 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 Server 体系结构"图显示了 IBM SPSS Modeler 分布式体系结构的组件。

- **IBM SPSS Modeler.** 客户端软件安装在最终用户的计算机上。客户端软件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 IBM SPSS Modeler 软件安装程序，但当其与 IBM 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 IBM SPSS Modeler 仅能在 Windows 操作系统上运行。
- **IBM SPSS Modeler Server.** 服务器软件安装在服务器计算机上，并通过网络与 IBM SPSS Modeler 和数据库相连接。IBM SPSS Modeler Server 作为服务（在 Windows 中）或守护进程（在 UNIX 中）运行，并等待客户端进行连接。它将执行使用 IBM SPSS Modeler 创建的流和脚本。
- **数据库服务器。** 数据库服务器可以是现有的数据仓库（例如，大型 UNIX 服务器上的 Oracle），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/部门服务器上的数据集市（例如 Windows 上的 SQL Server）。

IBM SPSS Modeler Server 体系结构

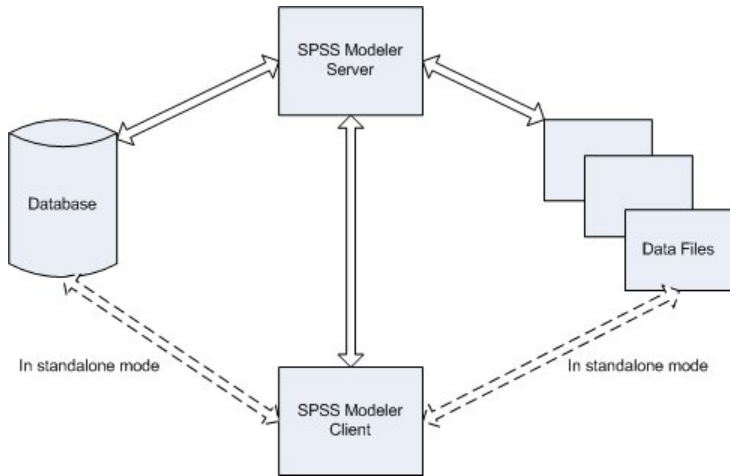


图 1. IBM SPSS Modeler Server 体系结构

借助分布式基础架构，大部分处理将在服务器计算机上进行。最终用户执行某个流时，IBM 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。这些查询在数据库中执行，执行后得到的数据将传递到服务器，以执行任何无法以 SQL 语言表达的处理。处理完成后，仅将相关结果回传给客户端。

必要时，IBM SPSS Modeler Server 可以在数据库之外执行所有 IBM SPSS Modeler 操作。它会自动平衡对用于存放所要处理的数据的 RAM 和磁盘存储器的使用。此过程将使 IBM SPSS Modeler Server 完全兼容平面文件。

使用服务器集群进行处理也可以实现负载均衡。通过进程协调程序插件，可以在 IBM SPSS Collaboration and Deployment Services 3.5 中启动集群。有关更多信息，请参阅第 81 页的附录 E，『使用服务器集群负载均衡』主题。通过 IBM SPSS Modeler 的“服务器登录”对话框，您可以直接连接到进程协调程序中管理的服务器或集群。有关更多信息，请参阅第 11 页的『正在连接到 IBM SPSS Modeler Server』主题。

独立客户端

IBM SPSS Modeler 还可以配置成作为独立的桌面应用程序运行，如下图所示。有关更多信息，请参阅第 11 页的第 3 章，『IBM SPSS Modeler 支持』。



图 2. 独立 IBM SPSS Modeler

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 IBM 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的最简单最经济方式。

专用服务器。在专用服务器上安装 IBM SPSS Modeler Server，专用服务器上的 IBM SPSS Modeler Server 不会与其他应用程序（包括它可能连接到的任何数据库）争用资源。模型构建操作尤其如此，作为耗费资源的操作，模型构建操作在不与其他应用程序争用资源的情况下可以更好地执行。

注：虽然在数据库所在的计算机上安装 IBM SPSS Modeler Server 可以通过避免网络开销来缩短数据库与服务器之间的数据传输时间，但在大多数情况下，最佳配置是将服务器和数据库安装在不同的机器上，以避免相互争用资源。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。机器上的处理器数目不应小于您希望经常运行的并发任务（同时执行的流）的数目。通常，处理器数目越多越好。

- IBM 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。因此，一个服务器在任意时间可以执行多项任务。
- 根据经验，一个处理器可以由一个或两个用户使用，两个处理器可以由多达四个用户使用，四个处理器可以由多达八个用户使用。除此之外，每增加两到四个用户另外添加一个处理器，具体情况取决于工作环境。
- 某些时候，一些处理可能通过 SQL 优化回送到数据库，此时两个或两个以上用户可以共享 CPU，而对性能的影响最小。
- 多线程功能使单个任务可以利用多个处理器，即，增加处理器可以提高性能，即便一次只有一个任务在运行也是如此。多线程技术通常用于 C5.0 模型构建和某些数据准备操作（排序、汇总以及合并）。对于在 IBM SPSS Analytic Server 中运行的所有节点（例如，GLE、线性 AS、随机森林、LSVM、树 AS、时间序列、TCM、关联规则和 STP），还支持多线程。

64 位平台。如果您计划在大量数据的基础上处理或构建模型，请使用 64 位机器作为 IBM SPSS Modeler Server 平台，同时尽可能增加机器上的 RAM 容量。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。64 位服务器实现可以利用更多的 RAM；建议的最小容量为 8 GB。

未来的需求。如果可行，请确保内存和 CPU 等服务器硬件可以扩展，以便同时容纳未来增加的使用量（例如，并发用户数增加或者现有用户处理需求增加）以及增加的 IBM SPSS Modeler Server 多线程处理能力。

对临时磁盘空间和 RAM 的要求

IBM SPSS Modeler Server 使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处理的数据量和数据类型以及执行的操作类型。数据量与行数和列数成正比。处理的行和列越多，需要的磁盘空间就越多。

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

IBM SPSS Modeler Server 的强大的 SQL 优化功能意味着处理能够尽可能在数据库中（而不是在服务器上）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。

- 汇总节点

- 差异节点
- 分级节点
- 采用“按键合并”选项时合并节点
- 所有建模节点
- 排序节点
- 表输出节点
- 其中的查询条件使用了 @THIS 的 @OFFSET 函数。
- 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，IBM 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 节点需要大量的 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。

要了解最新的 RAM 建议，请参阅服务器安装指南的系统需求部分。如果存在四个或四个以上的并发用户，建议安装更多的 RAM。并发任务必须共享内存，所以应相应地增加内存。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

要对数据库执行读写操作，您必须为相关数据库安装并配置 ODBC 数据源，并根据需要分配读或写权限。IBM SPSS Data Access Pack 包括一组用于此用途的 ODBC 驱动程序，您可以从下载站点获取这些驱动程序。如果您有关于创建或设置 ODBC 数据源权限方面的疑问，请与数据库管理员联系。

支持的 ODBC 驱动程序

有关使用 IBM 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>)。

在哪里安装驱动程序

注：必须在每台可能要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。

- 如果您以分布式模式针对远程 IBM SPSS Modeler Server 运行 IBM SPSS Modeler，需要在安装 IBM SPSS Modeler Server 的计算机上安装 ODBC 驱动程序。对于 UNIX 系统中的 IBM SPSS Modeler Server，另请参阅本节后面部分中的“在 UNIX 系统中配置 ODBC 驱动程序”。
- 如果您需要从 IBM SPSS Modeler 和 IBM SPSS Modeler Server 中访问相同数据源，必须在两台计算机上都安装 ODBC 驱动程序。
- 如果您通过终端服务运行 IBM SPSS Modeler，需要在安装 IBM SPSS Modeler 的终端服务服务器上安装 ODBC 驱动程序。

在 UNIX 系统中配置 ODBC 驱动程序

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

```
cd <modeler_server_install_directory>/bin
rm -f libspssodbc.so
ln -s libspssodbc_datadirect.so libspssodbc.so
```

此命令可删除缺省链接并新建至 DataDirect 驱动程序管理器的链接。

注：对于某些数据库，使用 SAP HANA 或 IBM DB2 CLI 驱动程序需要 UTF16 驱动程序包装程序。DashDB 需要 IBM DB2 CLI 驱动程序。要为 UTF16 驱动程序包装器创建链接，请改为输入以下命令：

```
rm -f libspssodbc.so
ln -s libspssodbc_datadirect_utf16.so libspssodbc.so
```

要配置 SPSS Modeler Server，请执行以下操作：

1. 通过将以下行添加到 modelersrv.sh，配置 SPSS Modeler Server 启动脚本 modelersrv.sh 以寻源 IBM SPSS Data Access Pack odbc.sh 环境文件：

```
. /<pathtoSDAPinstall>/odbc.sh
```

其中，<pathtoSDAPinstall> 是 IBM SPSS Data Access Pack 安装的完整路径。

2. 重新启动 SPSS Modeler Server。

此外，仅对于 SAP HANA 和 IBM DB2，在 odbc.ini 文件中，向 DSN 添加以下参数定义，以避免连接期间发生缓冲区溢出：

```
DriverUnicodeType=1
```

注：libspssodbc_datadirect_utf16.so 包装程序还兼容其他 SPSS Modeler Server 支持的 ODBC 驱动程序。

注：上述规则仅适用于数据库中的数据访问。其他类型的文件操作（例如流、工程、模型、节点、PMML、输出和脚本文件的打开与保存）始终在客户端执行，并且始终根据客户端计算机的文件系统进行指定。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 和 SPSS Statistics

有关如何在 UNIX 上配置 SPSS Modeler Server 以便使用 IBM SPSS Statistics 数据访问技术的信息，请参阅第 67 页的附录 B，『配置 UNIX 启动脚本』。

引用数据文件

Windows。如果在 IBM SPSS Modeler Server 所在的计算机上存储数据，我们建议您从服务器计算机的角度给出数据路径（例如 C:\ServerData\Sales 1998.csv）。不使用网络来查找文件时，性能较好。

如果数据存储在另一主机上，我们建议您使用 UNC 文件引用（例如 `\\mydataserver\ServerData\Sales 1998.csv`）。请注意，仅当路径包含共享网络资源的名称时，UNC 名称才有效。进行引用的计算机必须有权读取指定的文件。如果您经常在分布式和本地分析模式之间切换，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 服务器上的数据文件，请指定完整文件路径，并使用正斜杠（例如 `/public/data/ServerData/Sales 1998.csv`）。请避免在 IBM 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 - 系统将自动进行处理。

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模式导入或导出 IBM SPSS Statistics 数据。请记住，当 IBM SPSS Modeler 以分布式模式运行时，此客户端提供的是服务器文件系统。IBM SPSS Statistics 客户端以相同方式运行。如果要在两个应用程序之间进行导入和导出，那么两个客户端必须以同一方式操作。如果客户端操作方式不同，那么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 中的 IBM SPSS Statistics 节点可以自动启动 IBM SPSS Statistics 客户机，但是用户必须先确保 IBM SPSS Statistics 客户机是以与 IBM SPSS Modeler 相同的方式运行。

安装说明

有关安装 IBM SPSS Modeler Server 的信息，请参阅作为部分产品下载提供的 PDF 文件格式的安装指示信息。对于 Windows 和 UNIX，提供了不同的文档。

有关安装和使用 IBM SPSS Modeler 客户机的完整信息，请参阅作为部分产品下载提供的 PDF 文件。根据您的拥有的许可证类型，也可能提供独立安装文档。

第 3 章 IBM SPSS Modeler 支持

此节供为 IBM SPSS Modeler 用户提供支持的管理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
- IBM SPSS Modeler Server 的用户帐户和文件使用权限
- 用户在 IBM SPSS Modeler Server 和 IBM SPSS Modeler 之间切换时可以看到的差异。

正在连接到 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程序 (COP) 插件从 IBM SPSS Collaboration and Deployment Services 连接到 IBM SPSS Modeler Server 或服务器集群的客户端运行。当前连接状态显示在 IBM 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集群列表。可以通过进程协调程序执行浏览网络上运行的 Statistics 服务的功能。

连接到服务器

1. 在“工具”菜单上，单击**服务器登录**。这将打开“服务器登录”对话框。或者，双击 IBM SPSS Modeler 窗口的连接状态区域。
2. 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。请参阅主题第 17 页的『添加并编辑 IBM SPSS Modeler Server 连接』以获取更多信息。
 - 单击**搜索**以访问进程协调程序中的服务器或服务器集群。请参阅主题第 17 页的『搜索 IBM SPSS Collaboration and Deployment Services 中的服务器』以获取更多信息。

服务器表。此表包含已定义的服务器连接集。该表显示缺省连接、服务器名称、描述和端口号。您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缺省连接，请在表中“缺省”列中为此连接选择复选框。

缺省数据路径。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。

用户标识。输入用于登录服务器的用户名。

密码。输入与指定用户名相关联的密码。

域。指定用于登录服务器的域。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。

3. 单击 **确定** 以完成此连接。

断开与服务器的连接

1. 在"工具"菜单上, 单击**服务器登录**。这将打开"服务器登录"对话框。或者, 双击 IBM SPSS Modeler 窗口的连接状态区域。
2. 在此对话框中, 选择"本地服务器", 然后单击**确定**。

配置单点登录

可以使用单点登录连接到正在任何受支持的平台上运行的 IBM SPSS Modeler Server。要使用单点登录进行连接, 必须先配置 IBM SPSS Modeler 服务器和客户端机器。

如果您使用单点登录连接到 IBM SPSS Modeler Server 和 IBM SPSS Collaboration and Deployment Services, 那么在连接到 IBM SPSS Modeler 之前, 必须先连接到 IBM SPSS Collaboration and Deployment Services。

IBM SPSS Modeler Server 使用 Kerberos 进行单点登录。

Kerberos 是 Windows Active Directory 的核心组件, 在 Active Directory 基础结构中假定以下信息: 具体如下:

- 客户端计算机为连接到 Active Directory 域的 Windows 计算机
- 客户机用户已使用域帐户登录计算机。用于登录的机制不重要, 可采用智能卡、指纹等。
- IBM SPSS Modeler Server 可通过引用 Active Directory 域控制器验证客户机用户凭证

此文档描述如何配置 Windows 和 UNIX 服务器来以这种方式进行认证。可能有其他配置, 但未经过测试。

要与最现代、安全的 Active Directory 安装进行互操作, 必须为 Java 安装高强度加密包, 因为缺省情况下不支持需要的加密算法。必须为客户机和服务器安装此包。当服务器连接由于此包未安装而失败时, 客户机上将显示错误消息, 例如 Illegal key size。请参阅第 44 页的『安装无限强度的加密』。

服务主体名称

每个服务器实例都必须注册唯一的服务主体名称 (SPN) 以标识其自身, 并且客户机在连接到服务器时必须指定同一 SPN。

SPSS Modeler Server 实例的 SPN 具有以下格式:

```
modelerserver/<host>:<port>
```

例如:

```
modelerserver/jdoemachine.spss.com:28054
```

请注意, 必须使用主机名的 DNS 域 (在本示例中, 这是 spss.com) 对主机名进行限定, 并且这个域必须映射到 Kerberos 领域。

主机名与端口号的组合使 SPN 具有唯一性 (这是因为, 给定主机上的每个实例都必须使用不同的端口号进行侦听)。并且, 客户机和服务器都已具有主机名和端口号, 因此可以为实例构造适当的 SPN。需要执行的附加配置步骤是在 Kerberos 数据库中注册该 SPN。

在 Windows 上注册 SPN

如果要使用 Active Directory 作为 Kerberos 实现, 请使用 setspn 命令来注册 SPN。要运行此命令, 必须满足下列条件:

- 您必须登录到域控制器

- 您必须运行提升了特权的命令提示符（以管理员身份运行）
- 您必须是 Domain Admins 组的成员（或者域管理员已授予您适当的许可权）

有关更多信息，请参阅下列文章：

- [Setspn Command-Line Reference](#)
- [Delegating Authority to Modify SPNs](#)

对于缺省实例，即，使用标准端口（例如，28054）进行侦听并使用"本地系统"帐户运行，您必须针对服务器计算机名称注册 SPN。例如：

```
setspn -s modelerserver/jdoemachine.spss.com:28054 jdoemachine
```

对于每个后续（概要文件）实例，使用定制端口（例如：29000）进行侦听并使用任何用户帐户（例如：jdoe）运行且选项 `start_process_as_login_user` 设置为 Y，您必须针对服务用户帐户名称注册 SPN：

```
setspn -s modelerserver/jdoemachine.spss.com:29000 jdoe
```

请注意，在这种情况下，服务帐户并非"本地系统"，因此注册 SPN 并不足以使客户机能够建立连接。下一节描述了附加的配置步骤。

要查看向 jdoe 帐户注册了哪些 SPN，请运行以下命令：

```
setspn -l jdoe
```

在 UNIX 上注册 SPN

如果要使用 Active Directory 作为 Kerberos 实现，可以按上面 Windows 一节中描述的方式使用 `setspn` 命令；这假定您已在目录中创建计算机或用户帐户。或者您可以使用 `ktpass`，如第 14 页的『在 UNIX 和 Linux 上配置 IBM SPSS Modeler Server』中所示。

如果要使用另外某种 Kerberos 实现，请使用 Kerberos 管理工具将服务主体添加到 Kerberos 数据库中。要将 SPN 转换为 Kerberos 主体，必须追加 Kerberos 领域的名称。例如：

```
modelerserver/jdoemachine.spss.com:28054@MODELERSSO.COM
```

请将此主体和密码添加到服务器的密钥表中。对于主机上运行的每个实例，该密钥表都必须包含一个相应的条目。

在 Windows 上配置 IBM SPSS Modeler Server

在缺省方案中，SPSS Modeler Server 服务在"本地系统"帐户下运行，它将使用本机 Windows API 认证用户凭证且不需要在服务器上执行其他配置。

在替代方案中，SPSS Modeler Server 服务在专用用户帐户下运行且 `start_process_as_login_user` 设置为 Y，那么它将使用 Java API 认证用户帐户且不需要在服务器上执行其他配置。

首先，验证缺省方案是否可行。客户机应该能够使用 SSO 来连接到使用"本地系统"帐户运行的缺省实例。这将对未更改的客户端配置进行验证。您需要按上文所述为缺省实例注册 SPN。

然后，执行下列步骤：

1. 创建 `<MODELERSERVER>\config\sso` 目录。
2. 在步骤 1 中创建的 `sso` 文件夹中创建名为 `krb5.conf` 的文件。有关如何创建此文件的指示信息，请参阅第 15 页的『配置 IBM SPSS Modeler 客户机』下的步骤 3。此文件在服务器和客户机上必须相同。
3. 使用以下命令在服务器 SSO 目录中创建 `krb5.keytab` 文件：

```
<MODELERSERVER>\jre\bin\ktab -a <spn>@<realm> -k krb5.keytab
```

例如：

```
"..\jre\bin\ktab.exe" -a modelerserver/jdoemachine.spss.com:29000@SPSS.COM  
-k krb5.keytab
```

这将提示您输入密码。您输入的密码必须是服务帐户的密码。因此，如果服务帐户为 jdoe，那么必须输入 jdoe 用户的密码。

在密钥表中未指定服务帐户本身，但您之前已使用 setspn 对该帐户注册了 SPN。这意味着服务主体的密码与服务帐户的密码完全相同。

对于您创建的每个新实例（概要文件），必须使用 setspn（请参阅第 22 页的『配置服务器概要文件』和 第 12 页的『服务主体名称』）为该实例注册 SPN，并使用 jre\bin\ktab 向密钥表中添加一个条目。只有一个密钥表文件，对于每个未以“本地系统”身份运行的实例，密钥表文件都必须包含一个相应条目。缺省实例或者任何其他以“本地系统”身份运行的实例使用 Windows API 进行认证，因此无需在密钥表中指定。Windows API 不使用密钥表。

要验证某个实例是否包括在密钥表中：

```
ktab.exe -l -e -k krb5.keytab
```

对于每个主体，您可能会看到多个具有不同加密类型的条目，但这是正常情况。

在 UNIX 和 Linux 上配置 IBM SPSS Modeler Server

先决条件

IBM SPSS Modeler Server 依靠 Windows Active Directory (AD) 启用单点登录，以下先决条件对其至关重要：

- SPSS Modeler Client (Windows) 计算机是 Active Directory (AD) 域的成员。
- 客户机用户使用 AD 域帐户登录到计算机。
- SPSS Modeler Server (UNIX) 计算机通过植根于 AD DNS 域中的标准域名进行识别。例如，如果 DNS 域为 modelersso.com，那么服务器主机名可能为 myserver.modelersso.com。
- AD DNS 域支持 SPSS Modeler Server 主机名的正向和反向查找。

如果 SPSS Modeler Server 机器不是 AD 域的成员，那么必须创建域用户帐户以表示目录中的服务。例如，可以创建名为 ModelerServer 的域帐户。

在 UNIX 或 Linux 上配置 SPSS Modeler Server

1. 在 SPSS Modeler Serverconfig 文件夹中，创建名为 sso 的子文件夹。
2. 在 sso 文件夹中，创建 keytab 文件。可以在 AD 端生成 keytab 文件；但是，根据 SPSS Modeler Server 机器是否为 AD 域的成员，需求不同：
 - 如果 SPSS Modeler Server 机器是 AD 域的成员，请使用计算机帐户名称作为服务用户名：

```
ktpass -princ <spn>@<realm> -mapUser <domain>\<computer account> -pass <password> -out  
<output file> -ptype KRB5_NT_PRINCIPAL
```

例如：

```
ktpass -princ modelerserver/myserver.modelersso.com:28054@MODELERSO.COM -mapUser  
MODELERSO\myserver$ -pass Pass1234 -out c:\myserver.keytab -ptype KRB5_NT_PRINCIPAL
```

- 如果 SPSS Modeler Server 机器不是 AD 域的成员，请将作为先决条件创建的域用户帐户指定为服务用户：

```
ktpass -princ <spn>@<realm> -mapUser <domain>\ <user account> -mapOp set -pass <password>
-out <output file> -ptype KRB5_NT_PRINCIPAL
```

例如：

```
ktpass -princ modelerserver/myserver.modelersso.com:28054@MODELERSO.COM -mapUser MODELERSO\
ModelerServer -mapOp
set -pass Pass1234 -out c:\myserver.keytab -ptype KRB5_NT_PRINCIPAL
```

有关更多信息，请参阅Ktpass Command-Line Reference。

3. 将 sso 文件夹中的 keytab 文件重命名为 krb5.keytab。

注：如果将服务器计算机重新连接到域，那么将生成新密钥表文件。

4. 在步骤 1 中创建的 sso 文件夹中创建名为 krb5.conf 的文件。有关如何创建此文件的指示信息，请参阅『配置 IBM SPSS Modeler 客户机』下的步骤 3。此文件在服务器和客户机上必须相同。

配置 IBM SPSS Modeler 客户机

1. 使 Java 能够访问 TGT 会话密钥：
 - a. 从开始菜单中，单击运行。
 - b. 输入 regedit 并单击确定以打开“注册表编辑器”。
 - c. 浏览到本地计算机的操作系统相应的注册表位置：
 - 在 Windows XP 上：My Computer\HKEY_LOCAL_MACHINE\System\CurrentControlSet\Control\Lsa\Kerberos
 - 在 Windows Vista 或 Windows 7 上：My Computer\HKEY_LOCAL_MACHINE\System\CurrentControlSet\Control\Lsa\Kerberos\Parameters
 - d. 右键单击该文件夹并选择新建 > **DWORD** 值。新值的名称应该为 allowtgtsessionkey。
 - e. 将 allowtgtsessionkey 的值设置为十六进制值 1，即，0x0000001。
 - f. 关闭“注册表编辑器”。
 - g. 注：当用户帐户为本地管理员组的成员且用户帐户控制 (UAC) 启用时，存在已知问题。在这种情况下，检索服务凭单中的会话密钥为空，这将导致 SSO 认证失败。要避免此问题，请执行其中一个操作：
 - 以管理员身份运行应用程序
 - 禁用用户帐户控制
 - 使用非管理员帐户的帐户
2. 在 IBM SPSS Modeler 安装位置的 config 文件夹中，创建名为 sso 的文件夹。
3. 在 sso 文件夹中，创建 krb5.conf 文件。可以在 http://web.mit.edu/kerberos/krb5-current/doc/admin/conf_files/krb5_conf.html 中找到有关如何创建 krb5.conf 文件的指示信息。下面提供了 krb5.conf 文件的示例：

```
[libdefaults]
    default_realm = MODELERSO.COM
    dns_lookup_kdc = true
    dns_lookup_realm = true

[realms]
    MODELERSO.COM = {
        kdc = ad.modelersso.com:88
        admin_server = ad.modelersso.com:749
        default_domain = modelersso.com
    }
```

```
[domain_realm]
.modelersso.com = MODELERSSO.COM    modelersso.com = MODELERSSO.COM
```

4. 重新启动本地计算机和服务计算机。

获取 SSO 用户的组成员资格

当用户使用 SSO 登录到 SPSS Modeler Server 并且服务器以非 root 用户身份运行时，已认证的用户名称与操作系统用户帐户无关。服务器无法获取该用户的操作系统组成员资格。那么，在这种情况下如何执行组配置？

我们假定用户在 LDAP 目录（这可以是 Active Directory）中注册，这样，我们就可以请求从 LDAP 服务器中获取组成员资格。SPSS Modeler Server 可以查询 IBM SPSS Collaboration and Deployment Services 中的 LDAP 提供程序以获取组成员资格。

在 SPSS Modeler Server 上的 options.cfg 中，有两个属性用于控制服务器对 IBM SPSS Collaboration and Deployment Services 存储库的访问：

```
repository_enabled, N
repository_url, ""
```

要启用组查找，必须同时设置这两个属性。例如：

```
repository_enabled, Y
repository_url, "http://jdoemachine.spss.ibm.com:9083"
```

存储库连接仅用于 SSO 组查找，因此除非您需要使用此功能，否则无需更改这些属性设置。

为了使组查找功能正常工作，您必须先配置存储库以添加 LDAP 或 Active Directory 提供程序，然后使用该提供程序来启用 SSO：

1. 启动 IBM SPSS Deployment Manager 客户机并选择文件 > 新建 > 受管服务器连接...，以便为存储库创建受管服务器连接（如果尚未存在此连接）。
2. 登录到受管服务器连接，并展开配置文件夹。
3. 右键单击安全提供程序，选择新建 > 安全提供程序定义...，并输入适当的值。单击对话框中的帮助可以获得更多信息。
4. 展开单点登录提供程序文件夹，右键单击 **Kerberos SSO 提供程序**，并选择打开。
5. 单击启用，选择安全提供程序，然后单击保存。在这里，除非您希望使用 SSO，否则不必填写任何其他详细信息（仅仅启用提供程序即可允许执行组查找）。

要点：为了使组查找功能正常工作，此处配置的 Kerberos 提供程序必须与您为 SPSS Modeler Server 配置的提供程序相同。特别是，它们必须在同一个 Kerberos 领域中工作。因此，如果用户使用 SSO 登录到 SPSS Modeler Server 并将自身标识为 jdoe@SPSS.COM（其中 SPSS.COM 是领域），那么 IBM SPSS Collaboration and Deployment Services 中的安全提供程序应该能够识别用户主体名称并从 LDAP 目录中返回相应的组成员资格。

数据源单点登录

可以使用单点登录功能从 IBM SPSS Modeler 连接到数据库。如果要使用单点登录功能来创建数据库连接，那么必须先使用 ODBC 管理软件正确配置数据源和单点登录令牌。然后，在 IBM SPSS Modeler 中连接到数据库时，IBM SPSS Modeler 将使用该单点登录令牌，并且不会提示用户登录到该数据源。

但是，如果没有正确地对该数据源进行单点登录配置，那么 IBM SPSS Modeler 将提示用户登录到该数据源。该用户在提供有效凭证后仍能访问该数据源。

有关在启用了单点登录的系统上如何配置 ODBC 数据源的完整详细信息，请参阅数据库供应商的文档。以下是可能涉及的一般步骤的示例：

1. 配置数据库，以使其能够支持 Kerberos 单点登录。
2. 在 IBM SPSS Modeler Server 机器上，请创建 ODBC 数据源并对其进行测试。DSN 连接不应需要用户标识和密码。
3. 使用单点登录功能连接到 IBM SPSS Modeler Server，并开始使用创建的 ODBC 数据源，然后在步骤 2 中进行验证。

添加并编辑 IBM SPSS Modeler Server 连接

您可以在"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：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。最佳实践指出，应该使用相同的端口与 IBM SPSS Collaboration and Deployment Services 和 SPSS Modeler Client 进行通信。这些端口可以设置为 options.cfg 文件中的 max_server_port 和 min_server_port。

添加服务器连接

1. 在"工具"菜单上，单击**服务器登录**。这将打开"服务器登录"对话框。
2. 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。
3. 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**。指定服务器要侦听的端口号。如果缺省设置不可用，请向系统管理员索取正确的端口号。
 - **描述**。输入此服务器连接的可选描述。
 - **确保安全连接（使用 SSL）**。指定是否应该使用 SSL（安全套接字层）连接。SSL 是一个常用协议，用于确保通过网络发送的数据的安全。要使用此功能，必须在承载 IBM SPSS Modeler Server 的服务器中启用 SSL。必要时请与本地管理员联系，以了解详细信息。

编辑服务器连接

1. 在"工具"菜单上，单击**服务器登录**。这将打开"服务器登录"对话框。
2. 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。
3. 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，可以使用进程协调程序选择网络上可用的服务器或服务器集群，从而代替手动输入服务器连接。服务器集群是一组服务器，进程协调程序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。

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：如果您无权访问进程协调程序功能，那么仍然可以手动输入要连接的服务器名称或选择先前已定义的名称。请参阅主题『添加并编辑 IBM SPSS Modeler Server 连接』以获取更多信息。

搜索服务器和服务器集群

1. 在"工具"菜单上, 单击**服务器登录**。这将打开"服务器登录"对话框。
2. 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程序时未登录到 IBM SPSS Collaboration and Deployment Services, 则系统会提示您执行此项操作。
3. 从列表中选择服务器或服务器集群。
4. 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**。如果用户需要访问服务器计算机上定义的 ODBC 数据源, 那么将需要数据源的名称、描述和登录信息 (包括数据库登录标识和密码)。
- **数据文件访问**。如果用户需要访问位于服务器计算机或者网络中其他位置的数据文件, 那么将需要数据文件的名称和位置。
- **保存的文件的位置**。用户在连接到 IBM SPSS Modeler Server 的情况下保存数据时, 可能会尝试在服务器计算机上保存文件。但是, 服务器通常是写保护位置。如果遇到这种情况, 请告知用户应该将数据文件保存到何处。(一般来说, 该位置为用户的主目录。)

用户认证

IBM SPSS Modeler Server 使用服务器上的操作系统对连接到服务器的用户进行认证。当用户连接到 SPSS Modeler Server 时, 代表该用户执行的所有操作均在该用户的安全上下文中执行。对数据库表的访问权取决于数据库本身中的用户和/或密码特权。

Windows。在 Windows 上, 任何在主机网络中拥有有效帐户的用户都可以登录。使用缺省认证时, 用户必须具有 `<modeler_server_install>\Tmp` 目录的修改访问权。如果没有这些权限, 则用户使用 Windows 缺省认证将无法从客户端登录 SPSS Modeler Server。

UNIX。缺省情况下, 在 UNIX 上, 假定 SPSS Modeler Server 作为 root 用户进程运行。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认证的专有密码数据库, 且所有 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)。有关更多信息, 请参阅第 76 页的『在使用专用密码数据库的情况下配置为非 root 用户进程』。

配置 PAM

在 Linux 和 AIX 平台上, SPSS Modeler Server 使用可插入认证模块 (PAM) 进行认证。

要使用 PAM 认证, 必须在主机系统上正确配置适当的 PAM 模块; 例如, 要让 PAM 与 LDAP 进行交互, 主机操作系统上必须存在 PAM LDAP 模块, 并且必须正确地对其进行配置。请参阅操作系统文档以获取更多信息。这是 SPSS Modeler Server 能够使用 PAM 的先决条件。

要将 SPSS Modeler Server 配置为使用 PAM, 请编辑 SPSS Modeler Server 的"options.cfg"文件, 并添加或编辑下面这一行: **authentication_methods, pam**。

有需要时, 可以使用服务名称 `modelerserver` 为 SPSS Modeler Server 提供特定的 PAM 配置。例如, 下列步骤说明如何为 Red Hat Linux 进行配置:

1. 切换到 PAM 配置目录。例如：`/etc/pam.d`。
2. 使用文本编辑器创建名为"modelerserver"的新文件。
3. 添加要使用的 PAM 配置信息。例如：

```
auth      include      system-auth
account   include      system-auth
password  required     pam_deny.so
session   required     pam_deny.so
```

注：根据您的特定配置不同，这些行可能有所变化。有关更多信息，请参阅 Linux 文档。

4. 保存文件，然后重新启动 Modeler 服务。

权限

Windows。 连接到安装在 NTFS 驱动器上的服务器软件的用户必须使用具有下列权限的帐户进行登录。

- 对服务器的安装目录及其子目录的读权限和执行权限。
- 对临时文件的目录位置的读、执行和写权限。

在 Windows Server 2008 及更高版本中，不得假定用户具有这些权限。请确保根据需要显式地设置权限。

如果服务器软件安装在 FAT 驱动器上，那么无需设置权限，这是因为所有文件都允许用户具有完全控制。

UNIX。 如果未使用内部认证功能，那么连接到服务器软件的用户必须使用具有下列权限的帐户进行登录：

- 对服务器的安装目录及其子目录的读权限和执行权限。
- 对临时文件的目录位置的读、执行和写权限。

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量取决于最终用户要分析的数据文件的大小以及用户要执行的分析类型。有关更多信息，请参阅第 7 页的『对临时磁盘空间和 RAM 的要求』主题。

UNIX。 UNIX 版本的 IBM 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缺省权限。有关更多信息，请参阅第 68 页的『控制文件创建权限』主题。

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。 除非流明确地对记录进行排序，否则本地执行的流与服务器上执行的流在提供记录的顺序方面可能有所差异。数据库中运行操作的顺序与 IBM SPSS Modeler Server 中运行操作的顺序也可能有所不同。产生这些差异的原因是，各个系统使用不同的算法来实现有可能对记录进行重新排序的功能（例如汇总）。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的顺序。

取整差异。 以本地方式运行的 IBM SPSS Modeler 在存储浮点值时，使用的内部格式与 IBM SPSS Modeler Server 使用的格式不同。由于取整差异，各个版本之间的结果可能稍微有所不同。

第 4 章 IBM SPSS Modeler 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，如何配置组，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 IBM SPSS Modeler Server 安装说明。

启动和停止 IBM SPSS Modeler Server

IBM SPSS Modeler Server 在 Windows 上作为服务运行，而在 UNIX 上作为守护进程运行。

调度注意事项：停止 IBM SPSS Modeler Server 将断开最终用户的连接并终止他们的会话，因此请尽量将服务器重新启动安排在使用率较低的时段进行。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。

1. **Windows XP**。打开 Windows 的“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 的“开始”菜单。依次选择控制面板、管理工具和服务。

2. 选择 **IBM SPSS Modeler Server <nn.n>** 服务。现在，您可以检查该服务的状态，启动或停止该服务，以及编辑启动参数（如果适当）。

缺省情况下，该服务配置为自动启动，这表示如果您将其停止，那么它将在计算机重新引导时自动重新启动。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 IBM SPSS Modeler Server。

1. 切换至 IBM 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 IBM SPSS Modeler Server 安装目录。

2. 要启动服务器，请在命令提示符下键入

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

3. 要停止服务器，请在命令提示符下键入

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

4. 要检查 IBM 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 IBM SPSS Modeler Server 守护进程，其他进程是 IBM SPSS Modeler 会话。

IBM 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，然后停止服务器，那么服务器守护进程将在计算机重新引导时自动地重新启动。有关更多信息，请参阅第 67 页的『自动启动和停止 IBM SPSS Modeler Server』主题。

UNIX 内核限制

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

1. 添加以下命令到 *modelersrv.sh*：

```
ulimit -d unlimited
ulimit -m unlimited
ulimit -f unlimited
```

此外，将堆栈限制设置为系统允许的最大值 (*ulimit -s XXXX*)，例如：

```
ulimit -s 64000
```

2. 重新启动 IBM SPSS Modeler Server。

处理无响应的服务器进程 (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 IBM 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。

在容易累积不可使用的服务器进程（僵尸进程）的系统上，我们建议使用以下命令序列定期停止并重新启动 IBM SPSS Modeler Server：

```
cd modeler_server_install_directory
./modelersrv.sh stop./modelersrv.sh kill
```

这些使用 *modelersrv.sh kill* 命令终止的 IBM 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。在其他一些情况下，也可能会留下临时文件，这包括应用程序因资源耗尽而崩溃、用户中断、系统崩溃或其他原因。因此我们建议，作为定期重新启动 IBM SPSS Modeler Server 过程的一部分，应当从 IBM 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 IBM SPSS Modeler Server。

配置服务器概要文件

服务器概要文件使您能够从单个安装中运行 SPSS Modeler Server 的多个独立的实例。对于客户机而言，这些实例作为单独的服务器出现，这些服务器位于同一台主机上，但对不同的端口号进行侦听。多个实例共享一个安装可以简化维护，从而使管理员受益。第一个实例之后的后续实例可以比完全安装和卸载更快地进行创建和删除，而修订包只需应用一次。

在同一台主机上运行多个服务器实例的原因是，这样能够单独地配置每个实例。如果所有实例完全相同，那么就不会有什么益处。特别是，如果这些实例以非 *root* 用户身份运行（以使所有会话共享同一个用户帐户），那么每个实例都可以使用不同的用户帐户，从而在用户组之间实现数据隔离。例如，对于登录到实例 A 的用户，将向其分配由某个特定 User-A 拥有的会话，这将使其仅有权访问该用户的文件和文件夹，而登录到实例 B 的用户将看到可供 User-B 访问的另一组文件和文件夹。这可以与组配置结合使用，以便仅允许特定的组登录到特定实例，这意味着最终用户只能登录到一个或多个与其角色相对应的实例。请参阅第 45 页的『配置组』。

在标准 SPSS Modeler Server 安装中，文件夹 config、data 和 tmp 特定于服务器实例。config 文件夹的用途是使实例具有专用配置，而 data 和 tmp 文件夹支持数据隔离。每个实例都具有这些文件夹的专用副本，所有其他内容均共享。

请注意，许多服务器配置（例如数据库设置）仍可以为公共配置，因此概要文件配置将覆盖公共配置。服务器将先检查概要文件配置，然后再检查缺省配置。对于概要文件，最有可能更改的文件是 options、groups 和 passwords。

有关更多信息，请参阅第 24 页的『概要文件结构』。

有关如何配置概要文件以使用 SSO 的信息，请参阅第 12 页的『配置单点登录』。这要求您注册服务主体名称 (SPN)，执行某些配置（如果 Windows 服务帐户非本地），并在某些情况下启用组查找。

使用服务器概要文件

服务器概要文件的一些常见用例如下所示。其中的一些用法通过使用脚本受支持（请参阅第 26 页的『概要文件脚本』），并可能需要管理/root 用户权限。

创建服务器概要文件

SPSS Modeler Server 管理员 Jane 使用脚本来创建新概要文件：

- Jane 必须对该概要文件指定唯一的名称（不能是现有的概要文件名）。如果 profiles 目录尚未存在，那么系统将为 Jane 创建该目录。然后，将在 profiles 目录中创建新的子目录并对其指定名称 Jane，该子目录中包含 config、data、log 和 tmp 目录。
- Jane 还可以选择指定要用作模板的现有概要文件的名称，在这种情况下，现有概要文件中 config 文件夹的内容将复制到新概要文件中。如果未指定模板，或者现有概要文件未包含 options 文件（尽管应该包含此文件），那么将在新概要文件中创建空的 options 文件。
- Jane 还可以选择对该概要文件指定端口号，在这种情况下，该端口号将作为 port_number 属性的值写入该概要文件的 options 文件。如果未指定端口号，那么系统将选择一个值并将其写入 options 文件。
- Jane 还可以选择指定对该概要文件具有独占访问权的操作系统组的名称，在这种情况下，将在 options 文件中启用组配置。在这种情况下，将创建一个 groups 文件，用于阻止登录到除指定组以外所有的组。

配置服务器概要文件

服务器管理员 Jane 通过手动编辑概要文件配置文件来配置概要文件，或者通过使用 IBM SPSS Deployment Manager 中的 IBM SPSS Modeler Administration Console 连接到概要文件服务来完成此配置。

为服务器概要文件创建 Windows 服务

在 Windows 上，管理员使用脚本为指定的概要文件创建服务：

- Jane 必须指定现有概要文件的名称，系统将为该概要文件创建服务实例。该服务的命令行将包含 profile 参数。该服务的名称将遵循标准模式，即，包括概要文件名称。
- 以后，如果 Jane 需要更改该服务的用户名和密码以便以非 root 用户身份运行该服务，那么可能需要使用服务管理控制台并编辑服务属性。

在 UNIX 上，也有办法创建在系统引导时自动启动的“服务”。管理员可能希望使用这些机制来创建概要文件服务，但请注意，IBM SPSS Modeler 未提供对这些服务的正式支持。

为服务器概要文件管理 Windows 服务

管理员可以使用脚本来执行下列任务：

- 查看哪些服务器概要文件服务处于运行状态
- 启动特定的服务
- 启动所有服务
- 停止特定服务
- 停止所有服务

启动或停止所有服务时，将通过搜索 `profiles` 目录的子目录来获取概要文件列表。

删除服务器概要文件的 Windows 服务

在 Windows 上，管理员可以使用脚本来删除所指定概要文件的服务（如果存在该概要文件的服务）。必须指定该概要文件的名称。

除去服务器概要文件

将概要文件的服务停止后，管理员可以通过从 `profiles` 目录中删除概要文件的文件夹来除去该概要文件。

更新SPSS Modeler Server

将修订包应用于 SPSS Modeler Server 时，该修订包将应用于所有服务器概要文件。在 Windows 上，所有概要文件服务都将自动停止和启动。在 UNIX 上，您必须将其手动停止并重新启动。

卸载 SPSS Modeler Server

卸载 SPSS Modeler Server 时，将卸载所有的服务器概要文件。请注意，不会自动除去 `profiles` 目录及其包含的任何概要文件。必须将其手动删除。在 Windows 上，所有概要文件服务都将自动卸载。在 UNIX 上，您必须将其手动除去。

安装新版本的 SPSS Modeler Server

安装新版本的 SPSS Modeler Server 时，不会自动迁移任何现有的服务器概要文件。管理员必须以手动方式将概要文件从一个安装复制到下一个安装（必要时编辑配置），以重新创建服务。

概要文件结构

`profiles` 目录

服务器概要文件存储在服务器管理员选中的目录位置中。缺省目录是 SPSS Modeler Server 上 `[server install path]\config\` 中名为 `profiles` 的目录，但是由于以下原因，建议使用不同的目录用于概要文件存储：

- 可以在集群中的节点之间共享概要文件
- 升级时可以保留概要文件
- 不需要将 SPSS Modeler Server 安装目录的写权限授予管理员和其他配置概要文件的用户

执行 SPSS Modeler Server 的全新安装后，不存在 `profiles` 目录。此目录在您创建第一个概要文件时创建。

对于每个概要文件，`profiles` 目录都包含一个子目录，并且子目录名称与概要文件名匹配。由于目录名与概要文件名相同，因此概要文件名不得包含在文件名中无效的字符。另外，概要文件名也不应包含空格，否则有可能导致脚本中出现错误。并且，请记住，概要文件名在单个安装中必须是唯一的。

识别某个安装的所有概要文件的唯一方法是识别 profiles 目录的子目录。系统未在任何其他位置维护单独的概要文件列表。对于一个安装，除非超过主机系统的支持能力，否则对可以创建的概要文件数目没有限制。

在 profiles 目录中，任何给定概要文件的子目录都必须至少包含一个名为 config 的目录，在该目录中，必须至少存在一个名为 options.cfg 的文件（用于定义概要文件配置）。此文件包含标准 SPSS Modeler Server options.cfg 文件（位于 [server install path]/config）中的部分设置，具体包含的设置数目视概要文件的需求而定。在概要文件配置中不存在的设置必须在安装 config 目录中的公共 options 文件中设置。概要文件配置必须至少包含 port_number 设置，这是因为，每个概要文件服务必须使用不同的端口号进行侦听。

概要文件配置可以包含其他通常位于安装 config 目录中的 *.cfg 文件，在这种情况下，将读取这些文件而不读取标准文件（只有 options 文件具有累积效应）。最有可能包括在概要文件配置中的其他文件是 groups 和 passwords。在概要文件配置中忽略的文件包括所有概要文件之间共享的 JVM 和 SSO 配置文件。

profile 目录还可以包含 data 和 tmp 目录，除非概要文件配置中指定了替代位置，否则这些目录将覆盖公共的 data 和 tmp 文件位置。

如果您使用概要文件来实现数据隔离，请确保对相关目录正确地设置许可权。

概要文件配置文件

profiles 目录的位置在名为 [server install path]\config\profiles.cfg 的新配置文件中指定。此文件与同一目录中的其他配置文件具有相同的格式，用于设置 profiles 目录的键是 profiles_directory。例如：

```
profiles_directory, "C:\\SPSS\\Modeler\\profiles"
```

将一个单独的文件用于概要文件配置（而不是向标准 options 文件中添加设置）的原因有两个：

- 概要文件配置确定如何读取 options 文件，因此在本质上难以在一个 options 文件中定义另一个 options 文件
- 概要文件配置文件设计成通过脚本自动管理，因此在简单情况下，用户完全不需要关心此文件（但是，仍可以安全地对其进行手动编辑以支持更复杂的方案）

除 profiles 目录的位置以外，profiles.cfg 中存在的唯一其他条目是端口号。例如：

```
profile_port, 28501
```

这是要创建的下一个概要文件的缺省端口号，每次使用脚本创建概要文件时，此端口号都将自动递增。profiles.cfg 文件只有在需要时才会创建，因此在全新安装中不存在。

启动概要文件

服务可执行文件 (modelerserver.exe) 接受一个附加参数 profile，此参数用于为服务指定概要文件：

```
modelerserver -server profile=<profile-name>
```

如果各个服务使用不同的概要文件，那么可以从同一个安装中运行多个服务。如果省略了 profile 参数，那么服务将使用公共安装缺省值，而不进行任何概要文件覆盖。

在指定了 profile 参数的情况下进行调用时，此服务将执行下列操作：

- 读取 [server install path]\config\profiles.cfg 以获取 profiles 目录的位置
- 读取 [profiles directory]\[profile name]\config\options.cfg 以获取概要文件配置（尤其是端口号）

如果任意一个步骤因任何原因而失败，那么服务将向日志输出一条错误消息并停止。如果在指定了概要文件的情况下调用服务，但服务无法装入该概要文件，那么服务将无法运行。

环境变量

服务定义了一些附加的环境变量，以便可以表达路径名等内容，而不必了解当前概要文件：

表 1. 环境变量

变量	值
PROFILE_NAME	当前概要文件的名称，如果未指定概要文件，则为空字符串。
MODELERPROFILE	当前概要文件的目录的完整路径（例如， <code>\$MODELERSERVER\profiles\ \$PROFILE_NAME</code> ）。如果未指定概要文件，那么值与 <code>\$MODELERSERVER</code> 相同。
MODELERDATA	当前概要文件的 <code>data</code> 目录的完整路径（例如， <code>\$MODELERPROFILE\data</code> ）。如果未指定概要文件，那么值将指向标准 <code>data</code> 目录 <code>\$MODELERSERVER\data</code> 。

这些环境变量由服务进程设置，因此仅在该进程及其创建的所有子进程中可见。如果您在服务进程外部设置这些变量，那么系统会将其忽略，并在该进程中按描述的方式重新定义这些变量。

日志记录

每个概要文件服务都需要一个单独的专用文件夹，用于存放其日志文件。对于每个概要文件，都存在 `server_logging.log` 等文件的一个副本。

安装的 `config` 目录中的缺省 `log4cxx.properties` 配置使用 `PROFILE_NAME` 环境变量来指示服务的日志目录：

```
log4j.appender.LoggingAppender.File=${ALLUSERSPROFILE}/IBM/SPSS/Modeler
Server/17/log/${PROFILE_NAME}/server_logging.log
```

通过更改上面这一行并包括两个特定于概要文件的环境变量中的一个（`PROFILE_NAME` 或 `MODELERPROFILE`），可以更改所有概要文件的日志位置。例如，要重新定位 `profile` 目录中的 `log` 目录：

```
log4j.appender.LoggingAppender.File=${MODELERPROFILE}/log
/server_logging.log
```

或者，您可以通过在概要文件配置中创建 `log4cxx.properties` 文件的副本并对其进行编辑，来更改特定概要文件的日志位置。

概要文件脚本

提供本节中描述的脚本是为了帮助您创建和管理 SPSS Modeler Server 概要文件。所有脚本都包括在 SPSS Modeler Server 安装目录的 `scripts/profiles` 目录（例如 `C:\Program Files\IBM\SPSS\ModelerServer\18\scripts\profiles`）中。

公共脚本（适用于所有平台）

以下脚本帮助创建和管理概要文件。产品提供了此脚本的多个变体，这些变体具有不同的扩展名，并适用于不同的平台（`.bat` 用于 Windows，`.sh` 用于 UNIX）。在每种情况下，操作均相同。

创建概要文件

create_profile [options] <profile-name>

创建具有指定名称的新概要文件。概要文件名必须适合于用作服务器主机上的目录名（这是因为，此脚本将创建具有该名称的目录），并且不应包含空格。此名称不得与任何现有的概要文件名相同。

选项：

-d, --profiles-directory<profiles-directory>

指定应该在其中创建此概要文件以及所有后续概要文件的 profiles 目录。仅对于第一个概要文件才必须指定此目录，但最好每次创建概要文件时都指定此目录。如果在第一次创建概要文件时省略了此目录，那么系统将选择缺省位置。如果您在后续调用中更改了 profiles 目录，那么将在新位置中创建新概要文件，但将忽略所有的现有概要文件，除非将其单独移到新位置中。

-t, --template <profile-name>

指定要用作模板的现有概要文件的名称。现有概要文件中的概要文件配置将复制到新概要文件中，并且仅更改端口号。

-p, --port-number <port-number>

指定概要文件服务的端口号。端口号对于此概要文件而言必须是唯一的。如果省略了端口号，那么系统将选择缺省值。

-g, --group-name <group-name>

指定对此概要文件具有独占访问权的操作系统组的名称。此概要文件将配置为仅允许这个组的成员进行登录访问。

未更改文件系统许可权，因此您必须单独执行该操作。

示例：

```
scripts\profiles\create_profile.bat -d C:\Modeler\Profiles comet
```

在 C:\Modeler\Profiles 目录中创建名为 comet 的新概要文件。此概要文件将使用缺省端口号进行侦听。要确定端口号，请打开为此概要文件生成的 options.cfg 文件（在本示例中，这是 C:\Modeler\Profiles\comet\config\options.cfg）。

```
scripts\profiles\create_profile.bat --template comet --group-name "Meteor Users" --port-number 28510 meteor
```

在 C:\Modeler\Profiles 目录（名称取自上一个示例）中创建名为 meteor 的新概要文件。此概要文件将使用端口号 28510 进行侦听，并且仅允许 Meteor Users 组的成员进行登录访问。所有其他配置选项都将从现有的概要文件 comet 中复制。

Windows 脚本

这些脚本帮助为 SPSS Modeler Server 概要文件创建和管理 Windows 服务。这些脚本使用"Windows 服务控制"程序 (SC.EXE) 来执行所请求的操作，除非另有说明，否则脚本输出来自 SC.EXE。您在本地计算机上必须具有管理员特权才能执行其中的大部分任务。

有关更多信息，请参阅有关 SC.EXE 的 Microsoft TechNet 文档。

为概要文件创建 Windows 服务

create_windows_service [options] <profile-name>

为指定的概要文件创建 Windows 服务。您必须具有管理员特权才能创建服务。创建服务后，请使用服务管理控制台为该服务设置其他属性（例如，设置用于服务登录的帐户详细信息）。

选项：

-u, --service-user <account-name>

指定用于服务登录的帐户 (passim)。这可以是本地用户帐户、域用户帐户或本地计算机名称（代表本地系统帐户）。缺省值为本地系统帐户。如果您指定了除本地系统帐户以外的帐户，那么必须先转到服务管理控制台并设置该帐户的密码，然后该服务才能启动。

-s, --register-spn

为服务注册服务主体名称 (SPN)，以使客户机能够使用 Kerberos SSO 进行连接。在这种情况下，您必须指定服务登录帐户 (-u)，以便可以对该帐户注册 SPN。您必须具有域管理员特权（或者已被授予注册 SPN 的权限）才能使用此选项。

-H, --service-host <host-name>

指定用于构造 SPN 的主机名。这必须是客户机用于进行连接的主机名，并且必须使用映射到 Kerberos 领域的域名进行限定（在简单 Active Directory 配置中，域名就是 Kerberos 领域）。

示例：

```
scripts\profiles\create_windows_service.bat comet
```

为 comet 概要文件创建 Windows 服务。此服务由本地系统帐户拥有，客户机应该使用用户名和密码进行登录。

```
scripts\profiles\create_windows_service.bat -s -H modelerserver.mycompany.com -u MYCOMPANY\ProjectMeteor meteor
```

为 meteor 概要文件创建 Windows 服务。此服务由 ProjectMeteor 域帐户拥有，并且客户机可以使用 SSO 进行登录。直到您转到服务管理控制台并为 ProjectMeteor 帐户设置密码之后，此服务才会启动。系统将自动授权此帐户作为服务登录。

删除概要文件的 Windows 服务

delete_windows_service [options] <profile-names...>

删除所指定概要文件的 Windows 服务。您必须具有管理员特权才能删除服务。

选项：

-s, --summary

列出已删除的服务的名称。不会列出不存在或者无法删除的服务。如果未指定此选项，那么将列出所有指定服务的删除状态。

-a, --all

删除所有概要文件的服务。

示例：

```
scripts\profiles\delete_windows_service.bat comet
```

删除 comet 概要文件的 Windows 服务。

```
scripts\profiles\delete_windows_service.bat --all
```

删除所有概要文件的 Windows 服务。

启动概要文件的 Windows 服务

start_windows_service [options] <profile-names...>

启动所指定概要文件的 Windows 服务。您必须具有管理员特权才能启动服务。

选项：

-s, --summary

列出已启动的服务的名称。不会列出已处于运行状态或者无法启动的服务。如果未指定此选项，那么将列出所有列出服务的状态。

-a, --all

启动所有概要文件的服务。

示例：

```
scripts\profiles\start_windows_service.bat -s comet meteor
```

尝试启动 comet 和 meteor 概要文件的 Windows 服务，并列出生成启动的服务的名称。

停止概要文件的 Windows 服务

```
stop_windows_service [options] <profile-names...>
```

停止所指定概要文件的 Windows 服务。您必须具有管理员特权才能停止服务。

选项：

```
-s, --summary
```

列出已停止的服务的名称。不会列出已停止或者无法停止的服务。如果未指定此选项，那么将列出所有列出服务的状态。

```
-a, --all
```

停止所有概要文件的服务。

示例：

```
scripts\profiles\stop_windows_service.bat -a -s
```

尝试停止所有概要文件的 Windows 服务，并输出成功停止的服务的名称。由所有概要文件组成的集合从 profiles 目录中获取。

查询概要文件的 Windows 服务的状态

```
query_windows_service [options] <profile-names...>
```

显示所指定概要文件的 Windows 服务的状态。您无需具有管理员特权即可查询服务。

选项：

```
-s, --summary
```

仅列出服务的名称及其当前状态（RUNNING 和 STOPPED 等等）。如果由于任何原因而无法查询某个服务（例如，该服务不存在），那么状态将报告为 UNKNOWN。如果未指定此选项，那么将列出所有列出服务的全面状态。

```
-a, --all
```

查询所有概要文件的服务状态。

示例：

```
scripts\profiles\query_windows_service.bat -a
```

报告所有概要文件的全面服务状态。

UNIX 脚本

现在，用于管理 SPSS Modeler Server 服务的现有 UNIX 脚本接受一个附加的 profile 参数，因此可以独立地管理 SPSS Modeler Server 概要文件服务。

```
modelersrv.sh [options] {start|stop|kill|list}
```

管理 SPSS Modeler Server 主服务。有关更多信息，请参阅第 21 页的第 4 章，『IBM SPSS Modeler 管理』。

选项：

```
-p, --profile <profile-name>
```

管理所指定概要文件的服务实例。如果使用了此参数，那么指定的命令将仅应用于所指定概要文件的实例。如果未使用此参数，那么 `start` 命令将仅启动缺省实例（没有概要文件的服务），但是 `stop`、`kill` 和 `list` 命令将应用于所有活动实例。

示例：

```
./modelersrv.sh --profile comet start
```

启动 comet 概要文件的服务。

```
./modelersrv.sh --profile meteor start
```

启动 meteor 概要文件的服务。

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

列出所有处于活动状态的服务的进程。

```
./modelersrv.sh --profile comet stop
```

停止 comet 概要文件的服务。

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

停止所有处于活动状态的服务。

在 UNIX 上，目前没有用于自动启动 SPSS Modeler Server 概要文件服务的受支持方法。提供了标准的 `auto.sh` 脚本，用于将系统配置为随操作系统一起启动和停止 SPSS Modeler Server 主服务，但此脚本仅应用于缺省服务，而不会应用于任何概要文件服务。

管理

IBM SPSS Modeler Server 具有许多用于控制其行为的可配置选项。您可以通过两种方式设置这些选项：

- 使用 IBM SPSS Modeler Administration Console，当前使用 IBM SPSS Modeler 的客户可免费使用此应用程序。有关更多信息，请参阅『IBM SPSS Modeler Server 管理』主题。
- 使用 `options.cfg` 文本文件，该文件位于 [服务器安装路径]/`config` 目录。请参阅主题第 39 页的『使用 `options.cfg` 文件』，了解更多信息。

我们建议您安装 IBM SPSS Deployment Manager 并使用其 IBM SPSS Modeler Administration Console 作为管理工具，而不要编辑 `options.cfg` 文件。编辑该文件需要访问 IBM SPSS Modeler Server 文件系统，而通过 IBM SPSS 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，IBM SPSS 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 IBM SPSS Modeler Server 的情况下进行更改。

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

IBM SPSS Modeler Server 管理

当前的 SPSS Modeler Server 用户可以免费使用 IBM SPSS Deployment Manager 中的 Modeler Administration Console，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 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

从 Windows"开始"菜单中，依次选择 [所有] 程序、IBM SPSS Collaboration and Deployment Services、Deployment Manager。

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。

重新启动 Web Service

在管理控制台中，对 IBM SPSS Modeler Server 或 IBM SPSS Modeler Text Analytics Server 做出更改时，必须重新启动 Web Service。

要重新启动 Microsoft Windows 上的 Web Service：

1. 在安装了 IBM SPSS Modeler 的计算机上，从"控制面板"上的"管理工具"中选择服务。
2. 在列表中找到 **IBM SPSS Modeler Server** 并重新启动。
3. 单击**确定**并关闭对话框。

要重新启动 UNIX 上的 Web Service：

在 UNIX 上，必须通过运行 IBM SPSS Modeler Server 安装目录中的 **modelersrv.sh** 脚本来重新启动 IBM SPSS Modeler Server。

1. 切换至 IBM SPSS Modeler Server 安装目录。例如，在 UNIX 命令提示符处，输入：

```
cd /usr/<modelersrv>, 其中 modelersrv 是 IBM SPSS Modeler Server 安装目录。
```

2. 要停止服务器，请在命令提示符下键入

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

3. 要重新启动服务器，请在命令提示符下键入

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

使用 Modeler Administration Console 来配置访问权

使用 IBM SPSS Deployment Manager 中包含的 Modeler Administration Console 访问 SPSS Modeler Server 的管理员访问权由 options.cfg 文件中的 administrators 行进行控制，该文件位于 /config 下的 SPSS Modeler Server 安装目录中。在缺省情况下，此行被注释掉，因此您必须编辑此行以便将访问权授予特定人员，或者使用 * 将访问权授予所有用户，如下列示例所示：

```
administrators, "*"
administrators, "jsmith,mjones,achavez"
```

- 各行必须以 administrators, 开头，并且条目必须括在引号内。条目区分大小写。
- 各个用户标识之间以逗号分隔。
- 对于 Windows 帐户，请勿使用域名。

- 谨慎地使用星号。星号表示具有 IBM SPSS Modeler Server 的有效用户帐号的任何人（大多数情况下指服务器网络上的所有人）均可以登录和更改配置选项。

使用"用户访问控制"来配置访问权

要使用 Modeler Administration Console 来更新启用了"用户访问控制"(UAC) 的 Windows 机器上安装的 SPSS Modeler Server 配置，必须对 *config* 目录和 *options.cfg* 文件定义读、写和执行许可权。这些 (NTFS) 权限必须在特定用户级别进行定义，而不得在组级别进行定义，这是由 UAC 权限与 NTFS 权限的交互方式决定的。

Modeler Administration Console 包含在 IBM SPSS Deployment Manager 中。

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，您必须登录到每台服务器。虽然在 IBM SPSS Deployment Manager 中的所有 Modeler Administration Console 会话中都会记住服务器连接设置，但不会记住登录凭证。每次启动 IBM SPSS Deployment Manager 时都必须登录。

设置服务器连接

1. 确保 IBM SPSS Modeler Server 服务已启动。
2. 选择"文件"菜单上的新建，然后选择管理的服务器连接。
3. 在向导的第一页，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。
4. 在第二页，输入服务器的主机名和 IP 地址。如果更改了缺省的端口，则请输入端口号。单击**完成**。新服务器连接将显示在"服务器管理员"窗格中。

要执行管理任务，现在必须登录。

登录服务器

1. 在"服务器管理员"窗格中，双击以选择您要登录的服务器。
2. 在"登录"对话框中，输入凭证。（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。请参阅主题第 31 页的『使用 Modeler Administration Console 来配置访问权』，了解更多信息。

如果登录失败，并显示消息无法连接到服务器 '<服务器>'，请确保用户标识和密码正确，而且 IBM 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 IBM 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。请使用此窗格根据需要更改选项。单击工具栏上的 **保存**，以保存更改。请注意，更改标有星号 (*) 的任何选项需要重新启动服务器才能生效。

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。

注：如果非 root 用户希望更改这些选项，那么需要 SPSS Modeler Server **config** 目录的写许可权。

连接/会话

最大连接数。 (max_sessions) 同时连接的服务器会话的最大数目。值为 -1 表示无限制。

端口号。 (port_number) SPSS Modeler Server 要侦听的端口号。如果缺省端口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。

Analytic Server 连接

启用分析服务器 SSL (as_ssl_enabled): 要对 Analytic Server 与 SPSS Modeler 之间的通信进行加密，请指定 Y；否则，请指定 N。

主机 (as_host): Analytic Server 的 IP 地址。

端口号 (as_port): Analytic Server 端口号。

上下文根 (as_context_root): Analytic Server 的上下文根。

使用者 (as_tenant): SPSS Modeler Server 安装所属的使用者。

提示输入密码 (as_prompt_for_password): 如果 SPSS Modeler Server 配置为使用 Analytic Server 上使用的相同用户及密码认证系统，请指定 N，例如当使用 Kerberos 认证时；否则，请指定 Y。

注：如果想要使用 Kerberos SSO，您必须在 options.cfg 文件中设置额外的选项。有关更多信息，请参阅本章随后的主题"options.cfg 中可见的选项"。

数据文件访问

限制为只能访问数据文件路径 (data_files_restricted) 设置为 **yes** 时，此选项会将数据文件限制为标准数据目录中的数据文件以及以下**数据文件路径**中列出的数据文件。

数据文件路径 (data_file_path) 这是允许客户端在其中读写数据文件的附加目录的列表。除非启用**限制为只能访问数据文件路径**选择，否则此选项将被忽略。请注意，在所有路径名中，都应使用正斜杠。在 Windows 上，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）。在 Linux 和 UNIX 上，请使用冒号 (:) 代替分号。数据文件路径必须包含下面描述的 temp_directory 参数所指定的任何路径。

限制为只能访问程序文件路径 (program_files_restricted) 设置为 **yes** 时，此选项限制只能访问标准 bin 目录中的程序文件以及以下**程序文件路径**中列出的程序文件。在 R17 中，唯一限制访问的程序文件是 Python 可执行文件（请参阅下文中的 **Python 可执行文件路径**）。

程序文件路径 (program_file_path) 这是允许客户端从中执行程序程序的附加目录的列表。除非启用**限制为只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杠。指定多个目录时请使用半角分号隔开。

最大文件大小 (max_file_size) 流执行期间创建的临时数据文件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) 用于存储临时数据文件（缓存文件）的目录。在理想情况下，此目录应该位于独立的高速驱动器或控制器上，这是因为对此目录的访问速度可能会显著影响性能。您可以指定多个临时目录，并使用逗号分隔各个目录。这些目录应该位于不同的磁盘上；第一个目录使用最频繁，而其他目录则在执行期间特定数据准备操作（例如排序）并行工作时，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。在指定的所有路径中，都应使用正斜杠。

注：

- 在 SPSS Modeler Server 启动期间会在此目录中生成临时文件。请确保您对此目录具有所需的访问权（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。
- 当通过 IBM SPSS Collaboration and Deployment Services 作业运行评估流时，不适用 temp_directory 设置。当您运行此类作业时，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。

Python 可执行文件路径 (python_exe_path) Python 可执行文件的完整路径（包括可执行文件的名称）。如果限制访问程序文件，那么您必须将包含 Python 可执行文件的目录添加到程序文件路径（请参阅上文中的限制为只能访问程序文件路径）。

性能/优化

流重写。 (stream_rewriting_enabled) 允许服务器通过重写流对流进行优化。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，仅当此优化导致错误或其他意外结果时，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。如果在服务器中禁用了此设置，那么客户机无法将其启用。但是，如果在服务器中启用了此设置，那么客户机可以选择将其禁用。

并行性。 (max_parallelism) 描述运行流时允许 SPSS Modeler 使用的并行工作程序线程数。如果设置为 0 或任何负数，那么 IBM SPSS Modeler 将认为允许使用的线程数是计算机上的可用处理器数；此选项的缺省值为-1。要对多处理器计算机关闭并行处理，请将此选项设置为 1。要限制并行处理，请将此选项设置为小于计算机上的处理器数的数字。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端的数据文件通过此字节数的缓冲区进行传递。

缓存压缩。 (cache_compression) 这是 0 到 9 范围内的整数值，用于控制服务器临时目录中缓存文件和其他文件的压缩。压缩可以减少磁盘空间使用量，这在磁盘空间有限时至关重要。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。请注意，只有按顺序访问的特定缓存才可压缩。此选项不适用于随机访问缓存，例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，压缩程度会随之提高，但相应的访问时间也会延长。缺省值为 1；磁盘空间有限时可能需要更大的值。

内存使用乘数。 (memory_usage) 控制分配给排序和其他此类内存内缓存的物理内存比例。缺省值为 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) 控制为了训练 Kohonen 和 K-Means 模型而分配的物理内存比例。缺省值为 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。缺省值为 enabled。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，这些端口号可以用于客户端与服务器之间交互模型和流执行所需的其他套接字连接。这些套接字连接要求服务器侦听其他端口；未限制范围将对在带防火墙的系统上工作的用户造成问题。两者的缺省值均为 -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。

请注意，您必须打开主服务器端口上方的其他端口才能打开或执行流，如果要打开或执行并发流，那么还需要打开更多的相应端口。必须执行此操作才能捕获执行流所产生的反馈。

缺省情况下，IBM 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。要配置端口范围，IBM SPSS Modeler 除了需要主服务器端口以外，对于每个并发的流还需要两个打开的可用端口，并且对于来自任何相连接客户机的每个 ODBC 连接另外需要 3 个端口（2 个端口在该 ODBC 连接的持续期间用于该 ODBC 连接，另一个临时端口用于执行认证）。

注：ODBC 连接是数据库连接列表中的一个条目，并可以在指定了同一个数据库连接的多个数据库节点之间共享。

注：如果各个连接在不同的时间建立，那么可以共享认证端口。

注：最佳实践指出，应该使用相同的端口与 IBM SPSS Collaboration and Deployment Services 和 SPSS Modeler Client 进行通信。这些端口可以设置为 `max_server_port` 和 `min_server_port`。

注：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。

数组访存优化。 (`sql_row_array_size`) 控制 SPSS Modeler Server 从 ODBC 数据源访存数据的方法。缺省值为 1，表示一次访存一行。增大此值将使服务器读取更大数据块的信息，即，访存指定数目的行并将其放入数组。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

最大 SQL 字符串长度。 (`max_sql_string_length`) 对于使用 SQL 从数据库中导入的字符串，保证成功传递的最大字符数。根据操作系统不同，超出此长度的字符串值可能会在右端截断，并且不会发出警告。有效范围在 1 和 65,535 个字符之间。此属性也适用于“数据库导出”节点。

注：此参数的缺省值为 2048。如果要分析的文本的长度超过 2048 个字符（例如，使用“SPSS Modeler Text Analytics Web 订阅源”节点时，可能会发生这种情况），我们建议您增大此值（如果在本机方式下有效，否则结果可能会截断）。如果您正在使用数据库和用户定义的函数 (UDF)，那么不存在此限制；这是由于本机方式与 UDF 方式之间结果存在差异所致。

自动 SQL 生成。 (`sql_generation_enabled`) 允许自动为流生成 SQL，这可以显著提高性能。缺省值为 `enabled`。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优化设置；另请注意，要进行评分，必须为每个建模节点分别启用 SQL 生成，而与此设置无关。如果在服务器中禁用了此设置，那么客户机无法将其启用。但是，如果在服务器中启用了此设置，那么客户机可以选择将其禁用。

缺省 SQL 字符串长度。 (`default_sql_string_length`)。指定要在数据库缓存表中创建的字符串列的缺省宽度。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 的缺省宽度。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。

启用数据库 UDF。 (`db_udf_enabled`)。如果设置为 Y (缺省值)，那么将导致 SQL 生成选项生成用户定义的函数 (UDF) SQL，而不是生成纯 SPSS Modeler SQL。UDF SQL 通常优于纯 SQL。

SSL

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

密钥库。 (`ssl_keystore`) 服务器启动时要装入的 SSL 密钥数据库文件 (SPSS Modeler 安装目录的完整路径或相对路径)。

密钥库存储文件。 (ssl_keystore_stash_file) 服务器启动时要装入的密钥数据库密码存储文件的名称（完整路径或者相对于 SPSS Modeler 安装目录的路径）。如果要将此设置保留为空白并在启动 SPSS Modeler Server 时提示输入密码，请参阅以下指示信息：

- 在 Windows 上：
 1. 确保 options.cfg 中的 ssl_keystore_stash_file 设置没有值。
 2. 重新启动 SPSS Modeler Server。将提示您输入密码。输入正确的密码，单击**确定**，然后服务器将启动。
- 在 Linux/UNIX 上：
 1. 确保 options.cfg 中的 ssl_keystore_stash_file 设置没有值。
 2. 在 modelersrv.sh 文件中找到以下行：

```
if "$INSTALLEDPATH/$SCLEMDNAME" -server $ARGS; then
```
 3. 按以下添加 -request_ssl_password 开关：

```
if "$INSTALLEDPATH/$SCLEMDNAME" -request_ssl_password -server $ARGS; then
```
 4. 重新启动 SPSS Modeler Server。将提示您输入密码。输入正确的密码，单击**确定**，然后服务器将启动。

密钥库标签。 (ssl_keystore_label) 所指定证书的标签。

注：要将管理控制台与 SSL 服务器设置配合使用，必须将 SPSS Modeler Server 所需的所有证书导入到 ../jre/lib/security 之下的 Deployment Manager 信任库中。

注：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。

进程协调程序配置

主机。 (cop_host) 进程协调器服务的主机名或 IP 地址。缺省的"spsscop"是一个虚名称，管理员可以选择添加此名称作为 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。缺省为 8080，也是 IBM SPSS Collaboration and Deployment Services 的缺省值。

上下文根。 (cop_context_root) 进程协调器服务的 URL。

登录名。 (cop_user_name) 向进程协调器服务进行认证的用户名。这是 IBM SPSS Collaboration and Deployment Services 登录名，因此可以包含安全提供程序前缀（例如：ad/jsmith）

密码。 (cop_password) 用于向进程协调器服务进行认证的密码。

注：如果您手动更新 options.cfg 文件而不是使用 IBM SPSS Deployment Manager 中的 Modeler Administration Console，那么必须对此文件中指定的 cop_password 值进行手动编码。纯文本密码无效，并且将导致向进程协调程序进行的注册失败。

要以手动方式对密码进行编码，请完成下列步骤：

1. 打开命令提示符并浏览到 SPSS Modeler ./bin 目录。
2. 接收到请求时，请输入用户名（您在 options.cfg 中指定的 cop_user_name）并按 Enter 键。
3. 接收到请求时，请输入该用户的密码。

经过编码的密码将作为返回的字符串的组成部分显示在命令行上，并括在双引号内。例如：

```
C:\Program Files\IBM\SPSS\Modeler\18\bin>pwutil
User name: copuser
Password: Pass1234
copuser, "0Tqb4n.ob0wrs"
```

4. 请复制经过编码的密码（不带双引号）并将其粘贴到 options.cfg 文件中的 cop_password 值中已存在的双引号之间。

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

已启用 SSL。 (cop_ssl_enabled) 确定是否使用了 SSL 来连接到“协调程序”或“进程”服务器。如果使用了此选项，那么必须将 SSL 证书文件导入到 SPSS Modeler Server JRE 中。要执行此操作，必须获取 SSL 证书文件及其别名和密码。然后，在 SPSS Modeler Server 上运行以下命令：

```
$JAVA_HOME/bin/keytool -import -trustcacerts -alias $ALIAS_NAME -file $CERTIFICATE_FILE_PATH -keystore
$ModelerServer_Install_Path/jre/lib/security/cacerts
```

服务器名称。 (cop_service_name) 此 SPSS Modeler Server 实例的名称；缺省值为主机名。

描述。 (cop_service_description) 此实例的描述。

更新时间间隔（分钟）。 (cop_update_interval) 保持活动消息的间隔分钟数；缺省值为 2。

权重。 (cop_service_weight) 此实例的权重，指定为 1.0 与 10.0 之间的整数。较高的权重可以吸引更多的连接。缺省值为 1。

服务主机。 (cop_service_host) IBM SPSS Modeler Server 主机的标准主机名。缺省的主机名可自动获取，管理员可以覆盖多宿主机的主机名。

缺省数据路径。 (cop_service_default_data_path) 进程协调器注册的 IBM SPSS Modeler Server 安装的缺省数据路径。

options.cfg 中的可见选项

大部分配置选项都可以使用 IBM SPSS Deployment Manager 包含的 IBM SPSS Modeler Administration Console 进行更改。但是，存在一些例外情况，例如本节所述的情况。本节中的选项只能通过编辑 options.cfg 文件进行更改。有关更多信息，请参阅第 30 页的『IBM SPSS Modeler Server 管理』和第 39 页的『使用 options.cfg 文件』。请注意，options.cfg 中可能有此处未列出的其他设置。

注：此信息仅适用于远程服务器（例如，IBM SPSS Modeler Server）。

管理员。请指定要向其授予管理员访问权的用户的用户名。请参阅主题第 31 页的『使用 Modeler Administration Console 来配置访问权』，了解更多信息。

allow_config_custom_overrides。除非技术支持代表指示您进行修改，否则请不要这样做。

fips_encryption。启用符合 FIPS 标准的加密。缺省值为 N。

group_configuration。如果启用此选项，那么 IBM SPSS Modeler Server 将检查 groups.cfg 文件，该文件确定哪些用户能够登录到服务器。有关详细信息，请参阅第 45 页的『配置组』。

max_transfer_size。仅供系统内部使用。**请勿修改。**

shell。（仅限于 UNIX 服务器）覆盖 UNIX Shell 的缺省设置，例如 shell, "/usr/bin/ksh"。缺省情况下，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。如果要使用专用密码数据库来运行 SPSS Modeler Server，从而使用非 root 帐户来启动服务器服务，请将此选项设置为 Y。

use_bigint_for_count。要统计的记录数超过常规整数 ($2^{31}-1$) 范围时，请将此选项设置为 Y。此选项设置为 Y 并且有流连接到 DB2、SQL Server 或者 Teradata、Oracle 或 Netezza 数据库时，将在需要进行计数时使用函数（例如，由“汇总”节点生成的 **Record_Count** 字段）。

启用此选项后，如果使用的是 DB2 或 SQL Server，那么 SPSS Modeler 会将 COUNT_BIG() 用于记录计数。如果使用的是 Teradata、Oracle 或 Netezza，那么 SPSS Modeler 将使用 COUNT()。对于所有其他数据库，函数没有 SQL 回送。差别在于，启用 use_bigint_for_count 时，与禁用这些选项时的正常整数（32 位带符号整数， $2^{31}-1$ 到最大值）相比，所有记录计数将保存为 BIG INT（或 LONG）类型（64 位带符号整数， $2^{63}-1$ 到最大值）。

cop_ssl_enabled。如果您使用 SSL 连接到进程协调程序服务，请将此选项设置为 Y。如果使用了此选项，那么必须将 SSL 证书文件导入到 SPSS Modeler Server JRE 中。要执行此操作，必须获取 SSL 证书文件及其别名和密码。然后，在 SPSS Modeler Server 上运行以下命令：

```
$JAVA_HOME/bin/keytool -import -trustcacerts -alias $ALIAS_NAME -file $CERTIFICATE_FILE_PATH -keystore $ModelerServer_Install_Path/jre/lib/security/cacerts
```

cop_service_default_data_path。您可以使用此选项为进程协调器注册的 IBM SPSS Modeler Server 安装设置缺省数据路径。

用户可以通过 SPSS Modeler 中的工具 > **Analytic Server** 连接来创建自己的 Analytic Server 连接。管理员还可以使用以下属性定义缺省 Analytic Server 连接：

as_ssl_enabled。Y 或 N。

as_host。指定 Analytic Server 主机名或 IP 地址。

as_port。指定 Analytic Server 端口号。

as_context_root。指定 Analytic Server 上下文根。

as_tenant。指定 IBM SPSS Modeler Server 所属的租户的名称

as_prompt_for_password。Y 或 N。

缺省情况下，不启用使用 Kerberos 方法的 Analytic Server 认证。要启用 Kerberos 认证，请使用下列三个属性：

as_kerberos_auth_mode。要启用 Kerberos 认证，请将此选项设置为 Y。

as_kerberos_krb5_conf。指定 Analytic Server 应使用的 Kerberos 配置文件的路径；例如：c:\windows\krb5.conf。

as_kerberos_krb5_spn。指定 Analytic Server Kerberos SPN；例如：HTTP/ashost.mydomain.com@MYDOMAIN.COM。

SPSS Modeler Server 监视

IBM SPSS Deployment Manager 中的 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”窗格中需要服务器下方的“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（缺省为一分钟一次）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。

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。# 符表示注释。

注：大多数配置选项均可使用 IBM SPSS Deployment Manager 中的 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。请参阅主题第 37 页的『options.cfg 中的可见选项』，了解更多信息。

通过使用 IBM SPSS 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。有关更多信息，请参阅第 30 页的『IBM SPSS Modeler Server 管理』主题。

注：此信息仅适用于远程服务器（例如，IBM SPSS Modeler Server）。

可以添加到缺省文件中的配置选项

缺省情况下，数据库内缓存通过 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
```

执行此操作将在服务器上创建临时文件，而不是在数据库中创建这些文件。

要查看或更改 IBM SPSS Modeler Server 配置选项：

1. 使用文本编辑器打开 options.cfg 文件。
2. 找到感兴趣的选项。要获取完整的选项列表，请参阅第 32 页的『SPSS Modeler Server 配置』。
3. 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杠 (/) 作为路径名分隔符，而不得使用反斜杠。
4. 保存该文件。
5. 停止并重新启动 IBM SPSS Modeler Server 以使得这些更改生效。请参阅主题第 21 页的『启动和停止 IBM SPSS Modeler Server』，了解更多信息。

关闭未使用的数据库连接

缺省情况下，在访问一个连接时，IBM SPSS Modeler 至少会将该连接缓存入数据库。数据库会话将保持打开，即使未执行要求访问数据库的流也是如此。

缓存数据库连接可使每次执行流时不再需要 IBM SPSS Modeler 重新连接到数据库，可增加执行次数。但是，在某些环境中，应用程序尽快释放数据库资源至关重要。如果过多的 IBM 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。

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

要关闭未使用的数据库连接，请执行以下操作：

1. 找到 `[server install path]/config` 目录。
2. 添加以下文件（如果此文件已存在，请将其打开）：

```
odbc-custom-properties.cfg
```

3. 在此文件中添加下面这一行：

```
cache_connection, N
```

4. 保存并关闭该文件。
5. 重新启动 IBM SPSS Modeler Server。

注：

根据，每个数据库的实施，数据库内缓存在数据库中保存为常规表或临时表。例如，将临时表用于 DB2、HP Neoview、Oracle、Amazon Redshift、Sybase 和 Teradata。对于这些数据库，将 `cache_connection` 设置为 N 不会如期望一样工作，因为临时表仅在会话中有效（在关闭数据库连接时数据库将自动进行清除）。

因此在针对将 `cache_connection` 设置为 N 的这些数据库之一运行 SPSS Modeler 流时，可能生成诸如无法为数据库内缓存创建表。使用文件缓存代替。的错误。这指示 SPSS Modeler 无法创建数据库内缓存。另外，在某些情况下，对于 SPSS Modeler 生成的 SQL 查询，将使用临时表，但是表为空。

要解决此问题，您可以选择针对数据库内缓存使用常规数据库表。要执行此操作，请创建包含以下行的定制数据库属性配置文件：

```
table_create_temp_sql, 'CREATE TABLE <table-name> <(table-columns)>'
```

这强制将常规数据库表用于数据库内缓存，并且在关闭与数据库的所有连接或关闭工作流时，将删除表。

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/密码的认证进行加密，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 依赖于服务器的公用密钥和专用密钥，并且公用密钥证书将服务器身份绑定在其公用密钥上。

1. 当客户端连接到服务器时，客户端使用公用密钥证书对服务器进行认证。
2. 然后客户端生成随机数，使用服务器的公用密钥对该数字进行加密，并将加密后的消息发回服务器。
3. 服务器使用其专用密钥对随机数进行解密。
4. 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。

公用密钥证书通常由认证中心签署。认证中心，如 VeriSign 和 Thawte 等组织，专门负责签发、认证和管理公用密钥证书中所包含的安全凭证。实际上，认证中心确认了服务器的身份。认证中心通常会收取证书费用，但用户也可以自行生成自签名证书。

通过 SSL 确保客户端/服务器和服务器/服务器通信的安全

通过 SSL 确保客户端/服务器和服务器/服务器通信安全的主要步骤包括：

1. 获取并安装 SSL 证书和密钥。
2. 在服务器管理应用程序 (IBM SPSS Deployment Manager) 中启用和配置 SSL。

3. 如果使用强度超过 2048 位的加密证书, 请在 客户端计算机上安装无限强度的加密。
4. 指示用户在启用 SSL 的情况下连接到服务器。

注: 有时, 服务器产品可以充当客户端。例如当 IBM SPSS Statistics Server 连接到 IBM SPSS Collaboration and Deployment Services Repository 时。这时, IBM SPSS Statistics Server 是作为客户端。

获取并安装 SSL 证书和密钥。

要配置 SSL 支持, 必须先完成下列步骤:

1. 获取 SSL 证书和密钥文件。可以通过三种方法完成此任务:
 - 向公共认证中心 (例如 VeriSign、Thawte 或 Entrust) 购买。公共认证中心 (CA) 签发证书, 以便对该证书的服务器进行验证。
 - 通过第三方认证中心生成密钥和证书文件。如果采用这种方法, 那么必须将第三方 CA 的根证书导入到客户机和服务器密钥库文件中。有关更多信息, 请参阅第 43 页的『导入第三方根 CA 证书』主题。
 - 使用内部自签名认证中心来生成密钥和证书文件。完成此操作的步骤如下所示:
 - a. 准备密钥数据库。有关更多信息, 请参阅第 42 页的『创建 SSL 密钥数据库』主题。
 - b. 创建自签名证书。有关更多信息, 请参阅第 42 页的『创建自签名 SSL 证书』主题。
2. 将第 1 步中创建的 .kdb 和 .sth 文件复制到 IBM SPSS Modeler Server 有权访问的目录中, 并在 *options.cfg* 文件中指定该目录的路径。

注: 使用正斜杠作为目录路径中的分隔符。

3. 在 *options.cfg* 文件中设置下列参数:
 - 将 *ssl_enabled* 设置为 Y
 - 将 *ssl_keystore* 设置为 "<filename>.kdb", 其中 <filename> 是密钥数据库的名称。
 - 将 *ssl_keystore_stash_file* 设置为 "<filename>.sth", 其中 <filename> 是密钥数据库密码存储文件的名称。
 - 将 *ssl_keystore_label* 设置为 <label>, 其中 <label> 是证书的标签。
4. 对于自签名证书或第三方证书, 请在客户机系统上安装该证书。对于购买的公用 CA 证书, 不必执行此步骤。请确保将访问权设置为拒绝随意浏览包含该证书的目录。有关更多信息, 请参阅第 42 页的『安装自签名 SSL 证书』主题。

配置环境以运行 GSKit: GSKCapiCmd 是并非基于 Java 的命令行工具, 您无需在系统上安装 Java? 即可使用此工具; 此工具位于 <Modeler installation directory>/bin 文件夹中。根据使用的平台不同, 配置环境以运行 IBM Global Security Kit (GSKit) 的过程也有所变化。

要为 Linux/Unix 进行配置, 请将共享库目录 <Modeler installation directory>/lib 添加到环境中:

```
$export <Shared library path environment variable>=<modeler_server_install_path>/bin
$export PATH=$PATH:<modeler_server_install_path>/bin
```

根据平台不同, 共享库路径变量名称也有所变化:

- AIX 使用以下变量名称: LIBPATH
- HP-UX 使用以下变量名称: SHLIB_PATH
- Linux 使用以下变量名称: LD_LIBRARY_PATH

例如, 要在 Linux 上设置环境, 请使用:

```
$export LD_LIBRARY_PATH=/path/to/gskit/bin
$export PATH=$PATH:/path/to/gskit/bin
```

帐户对文件的访问权

请确保为将访问 SSL 文件的帐户授予正确的权限：

1. 对于 SPSS Modeler 用于连接的所有帐户，请授予对 SSL 文件的读访问权。

注：这也适用于 SPSS Modeler Server 服务中定义的登录身份用户。在 UNIX 或 Linux 上，适用于您启动 SPSS Modeler Server 时所用的用户。

2. 对于 Windows，帐户位于管理员组中且在启用用户访问控制 (UAC) 时向该管理员组提供权限还不够。此外，您必须执行下列其中一个操作：
 - 单独提供帐户权限。
 - 创建新组，向新组添加帐户，然后提供访问 SSL 文件的组权限。
 - 禁用 UAC。

创建 SSL 密钥数据库： 使用 GSKCapiCmd 工具可以创建密钥数据库。在使用此工具之前，必须配置环境；有关更多信息，请参阅第 41 页的『配置环境以运行 GSKit』主题。

要创建密钥数据库，请运行 GSKit 并输入以下命令：

```
gsk<ver>capiCmd[_64] -keydb -create -populate -db <filename>.kdb -pw <password> -stash
```

其中，<ver> 是 GSKit 版本号，<filename> 是要用于密钥数据库文件的名称，而 <password> 是密钥数据库的密码。

-stash 选项将在密钥数据库所在的路径中创建一个存储文件，其文件扩展名为 .sth。GSKit 使用该存储文件来获取密钥数据库的密码，从而使您不必每次都在命令行中输入该密码。

注：应该对 .sth 文件实施强文件系统保护。

创建自签名 SSL 证书： 要生成自签名证书并将其存储在密钥数据库中，请使用以下命令：

```
gsk<ver>capiCmd[_64] -cert -create -db <filename>.kdb -stashed -dn "CN=myserver,OU=mynetwork,O=mycompany,C=mycountry" -label <label> -expire <Number of days certificate is valid> -default_cert yes
```

其中，<ver> 是 GSKit 版本号，<filename> 是密钥数据库文件的名称，<Number of days certificate is valid> 是证书实际生效天数，而 <label> 是描述性标签，用于帮助您识别该文件（例如，您可以使用 *myselfsigned* 之类的标签）。

安装自签名 SSL 证书： 对于使用 SSL 来连接到服务器的客户端机器，您必须将证书的公用部分分发到客户机，以使其可以存储在这些客户机的密钥数据库中。为此：

1. 使用以下命令将公用部分抽取到一个文件中：

```
gsk<ver>capiCmd[_64] -cert -extract -db <filename>.kdb -stashed -label <label> -format ascii -target mycert.arm
```

2. 将 mycert.arm 分发给客户机。应该将其复制到客户机的 jre/bin 目录中。

3. 使用以下命令将新证书添加到客户机的密钥数据库中：

```
keytool -import -alias <label> -keystore ..\lib\security\cacerts -file mycert.arm
```

如果系统提示您输入密码，请使用 changeit。keytool 位于 <Modeler installation directory>\jre\bin 目录中（或者在 Mac 上，位于 <Modeler installation directory>/SPSSModeler.app/Contents/PlugIns/jre/Contents/Home/bin 目录中）。

导入第三方根 CA 证书： 您可以使用第三方认证中心 (CA) 对服务器证书进行签名，以代替向熟知认证中心购买证书或者创建自签名证书。客户机和服务器必须有权访问第三方 CA 的根证书，以便对第三方 CA 所签发的服务器证书进行验证。为此：

1. 获取第三方 CA 根证书。根据第三方 CA 的过程不同，此操作的流程也有所变化。第三方 CA 通常提供了他们的根证书供您下载。

2. 使用以下命令将该证书添加到服务器的密钥数据库中：

```
gsk<ver>capicmd[_64] -cert -add -db <filename>.kdb -stashed -label <label> -file <ca_certificate>.crt  
-format binary -trust enable
```

3. 使用以下命令将该证书添加到客户机的密钥数据库中：

在 Windows 上：

```
C:> cd <Modeler Client installation path>\jre\bin  
C:> keytool -import -keystore ..\lib\security\cacerts -file <ca_certificate>.crt -alias <label>
```

在 Mac 上：

```
C:> cd <Modeler Client installation path>/SPSSModeler.app/Contents/PlugIns/jre/Contents/Home/bin  
C:> keytool -import -keystore ..\lib\security\cacerts -file <ca_certificate>.crt -alias <label>
```

如果系统提示您输入密码，请使用 *changeit*。keytool 位于 <Modeler installation directory>\jre\bin 目录中（或者在 Mac 上，位于 <Modeler installation directory>/SPSSModeler.app/Contents/PlugIns/jre/Contents/Home/bin 目录中）。

4. 使用以下命令，对服务器的包含根 CA 证书的密钥数据库进行验证：

```
gsk<ver>capicmd[_64] -cert -validate -db <filename>.kdb -stashed -label <label>
```

返回以下消息表示验证成功：OK。

注：上述命令使用二进制格式的第三方 CA 根证书。如果证书采用 ASCII 格式，请使用 *-format ascii* 选项。

-db 参数指定密钥数据库的名称，第三方 CA 根证书将导入到此数据库中。

-label 参数指定密钥数据库文件中用于此第三方 CA 根证书的标签。此处使用的标签可以是任意内容，这是因为，它与 IBM SPSS Modeler options.cfg 文件中使用的标签无关。

-file 参数指定包含第三方 CA 根证书的文件。

在 IBM SPSS Deployment Manager 中启用和配置 SSL

1. 如果要安装自签名 SSL 证书，请将您创建的 cacerts 文件复制到 <Deployment Manager installation directory>\jre\lib\security 目录中。请参阅主题第 42 页的『安装自签名 SSL 证书』，了解更多信息。
2. 启动服务器管理应用程序（作为 IBM SPSS Deployment Manager 一部分安装的 IBM SPSS Deployment Manager）并连接到服务器。
3. 在配置页上，将安全套接字层设为是。
4. 在 **SSL 公用密钥文件**中，指定公用密钥文件的完整路径。
5. 在 **SSL 专用密钥文件**中，指定专用密钥文件的完整路径。

注：如果公用密钥和专用密钥存储在同一文件中，请在 **SSL 公用密钥文件**和 **SSL 专用密钥文件**中指定同一个文件。

6. 从菜单中选择：

文件 > 保存

7. 重新启动服务器服务或守护进程。重新启动时，系统将提示您输入 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。

安装无限强度的加密

产品附带的 Java 运行时环境启用了美国出口强度的加密。为使您的数据获得增强的安全性，建议升级到无限强度的加密。必须对客户机和服务器安装重复此过程。

安装无限强度的加密

1. 从 IBM.com (选择适用于 Java 7 的文件) 下载无限制 SDK JCE 策略文件。
2. 将打包在压缩文件中的不限加密强度司法协议文件解压缩。该压缩文件包含 `US_export_policy.jar` 文件和 `local_policy.jar` 文件。
3. 从目录 `jre/lib/security` 备份 `US_export_policy.jar` 和 `local_policy.jar` 的现有副本。
4. 将 `US_export_policy.jar` 和 `local_policy.jar` 文件的现有副本替换为下载和解压缩的两个文件。
5. 根据需要重新启动 IBM SPSS Modeler 客户机或服务器。

指示用户启用 SSL

当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。

Cognos SSL 连接

要使用 HTTPS 和 SSL 安全端口连接到 Cognos Analytics 服务器，必须先更改 Cognos 内部和外部分派器的某些设置。有关如何进行需要的更改的详细信息，请参阅《Cognos Server Configuration and Administration Guide》。

更改分派器设置后，按照下列步骤将您在 Cognos 中创建的 SSL 证书导入 SPSS Modeler JRE 中：

1. 在 Cognos 配置中，定义 IBM Cognos 密钥存储库的密码：
 - a. 在"资源管理器"窗口中，单击**密码术 > Cognos**。
 - b. 在"属性"窗口中，在**加密密钥设置**下设置加密密钥存储库密码。
 - c. 在"文件"菜单中，选择"保存"。
 - d. 在"操作"菜单中，选择"重新启动"。
2. 从命令行转至 `c10_location\bin` 目录。
3. 将 `JAVA_HOME` 环境变量设置为运行 Cognos 的应用程序服务器使用的 Java™ 运行时环境位置。例如：

```
set JAVA_HOME=c11_location\bin\jre<version>
```
4. 在命令行中，运行证书工具。例如：

```
ThirdPartyCertificateTool.bat -E -T -r ca.cer -k ..\configuration\encryptkeypair\jEncKeystore -p <password>
```
5. 将 `ca.cer` 文件复制到 SPSS Modeler Server 位置。
6. 打开命令行并切换至 `<ModelerInstallationLocation>\jre\bin` 文件夹。
7. 运行命令以导入证书。例如：

```
.\keytool -import -alias ca -file <Directory where ca.cer is located>\ca.cer -keystore "<ModelerInstallationLocation>\jre\lib\security\cacerts"
```

然后，您可以使用 HTTPS 和 SSL 安全分派器连接到 Cognos。例如：

`https://9.119.83.37:9343/p2pd/servlet/dispatch`

Cognos TM1 SSL 连接

要使用 HTTPS 和 SSL 安全端口连接到 Cognos TM1，请完成下列步骤：

1. 配置 Tomcat SSL。（例如，有关更多信息，请参阅：<http://tomcat.apache.org/tomcat-7.0-doc/ssl-howto.html>）。

a. 从命令行，转至 C:\Program Files\ibm\cognos\tm1_64\bin64\jre\7.0\bin 目录（这是缺省安装路径）并运行以下命令以在主文件夹中生成名为 .keystore 的文件：

```
keytool -genkey -alias tomcat -keyalg RSA
```

b. 在 C:\Program Files\ibm\cognos\tm1_64\tomcat\conf 文件夹中，将以下连接器设置添加到 server.xml 文件：

```
<Connector SSLEnabled="true" acceptCount="100" clientAuth="false"
  disableUploadTimeout="true" enableLookups="false" maxThreads="25"
  port="8443" keystoreFile="/Users/loiane/.keystore" keystorePass="password"
  protocol="org.apache.coyote.http11.Http11NioProtocol" scheme="https"
  secure="true" sslProtocol="TLS" />
```

c. 重新启动 IBM Cognos TM1 应用程序服务器服务。

2. 从命令行，使用以下命令以导出新创建的密钥库的证书文件：

```
keytool -export -alias tomcat -file certfile.cer -keystore C:\Users\Administrator\.keystore
```

3. 从命令行，使用以下命令以将证书文件导入到 SPSS Modeler Server 使用的 JRE：

```
keytool -import -alias as -file C:\Leon\Temp\certfile.cer -keystore
"c:\Program Files\IBM\SPSS\Modeler\<version>\jre\lib\security\cacerts"
```

然后，您可以使用 HTTPS 和 SSL 安全端口号连接到 Cognos TM1。

配置组

已认证的用户通常属于一个或多个安全组，而且对 SPSS Modeler Server 启用基于组的配置后，可以使用这些组来允许或拒绝登录到服务器，或者对用户会话的选项设置进行定制。

在以下方案中支持组配置：

- 在缺省安装中，SPSS Modeler Server 服务在本地系统或 root 用户帐户下运行，而且用户使用显式凭证或单点登录 (SSO) 进行登录：在这种情况下，组为用于控制文件访问等的用户的操作系统安全组。
- 在非 root 用户安装中，SPSS Modeler Server 服务在非特权帐户下运行，而且用户使用 SSO 登录：在这种情况下，组为与 SSO 主体关联的 LDAP 组。这些组从 IBM SPSS Collaboration and Deployment Services 中的 LDAP 安全提供程序获得，因此需要执行一些其他配置来启用此方案。有关更多信息，请参阅第 16 页的『获取 SSO 用户的组成员资格』。

如果这两种方案都不适用，该用户的组将不可用，而且组配置不受支持。特别是，在非 root 用户安装中，SPSS Modeler Server 服务在非特权帐户下运行，而且用户使用用户名和密码进行登录，那么操作系统组将对服务器不可用且组配置不受支持。

基于组的配置的原理是，应用于用户会话的选项设置可以根据该用户的组成员资格不同而有所变化。这些选项设置通常是从 SPSS Modeler Server options.cfg 文件中读取的服务器端设置（对于所有会话，此文件包含相同的设置）。options.cfg 文件提供所有会话的缺省值，但可以存在特定于组的配置文件，这些文件将覆盖特定会话的部分设置。

组配置允许控制各种设置，例如：

- 控制文件和 DSN 访问
- 控制资源使用情况

在 options.cfg 中启用 group_configuration 选项时，IBM SPSS Modeler Server 将检查 groups.cfg 文件，该文件控制哪些用户可登录到服务器。缺省值为 N。以下是一个 groups.cfg 示例，此示例拒绝 Test 组访问服务器，并允许 Fraud 组使用指定的配置进行访问。星号允许所有其他组使用缺省配置进行访问。

```
Test, DENY
Fraud, "groups/fraud.cfg"
*,
```

特定组配置（例如，上面的 Fraud）可能限制访问特定数据源或更改资源设置（相对于 SQL 回推、内存使用、多线程等）以提高该组成员的性能。

按组控制 DSN 访问权

多因子认证 (MFA) 要求可以根据用户的组成员资格在允许其访问的一组 ODBC 数据源名称 (DSN) 方面对其进行限制。

用于实现此目标的方案与文件访问权的现有方案类似。在 options.cfg 中提供了两项配置设置：

```
sql_data_sources_restricted, N
sql_data_source_path, ""
```

如果 sql_data_sources_restricted 设置为 Y，那么将仅允许该用户访问相关路径中列出的 DSN。在 Windows 上，各个 DSN 之间以标准路径分隔符 ; (分号) 分隔，在 UNIX 上以 : (冒号) 分隔。例如，在 Windows 上：

```
sql_data_sources_restricted, Y
sql_data_source_path, "Fraud - Analytic;Fraud - Operational"
```

启用此限制后，结果如下所示：

- 当用户进行浏览以查找数据源（例如，从 ODBC 连接对话框中进行浏览，或者使用 PSAPI 会话 getServerDataSourceNames API 进行浏览）时，产品并不会显示服务器系统上定义的所有 DSN，用户将仅看到 options.cfg 路径中定义的 DSN 子集。请注意，此路径可能包含未在服务器上定义的 DSN，这些 DSN 将被忽略 - 用户不会看到这些名称。
- 如果用户构造使用了脚本或 PSAPI 的 ODBC 节点（或者任何使用了 ODBC 连接的节点），并指定了未包括在 options.cfg 路径中的 DSN，那么该节点将无法运行，而系统将显示类似于以下的错误：**对数据源 <X> 的访问被拒绝。**

数据源路径可以包含其他有关文件路径的内容中描述的 PATH、GROUP 和 USER 插入项。使用基于组的配置时，PATH 插入项使您能够根据用户的组成员资格以递增方式构造路径。并且，可能还存在最好根据拥有 DSN 的组对该 DSN 进行命名的情况。

以前一个示例为基础，如果仅允许“内部欺诈分析人员”组的成员访问 Fraud 数据源，那么站点可以启用组配置，并创建特定于“内部欺诈分析人员”的配置，其中至少包含下面这一行：

```
sql_data_source_path, "${PATH};Fraud - Analytic;Fraud - Operational"
```

在此示例中，添加 PATH 前缀可以确保内部欺诈分析人员仍能访问其他允许所有人或者他们所属的其他组访问的数据源。

服务器日志

IBM SPSS Modeler Server 在名为 `server_logging.log` 的日志文件中保存其重要操作的记录。在 UNIX 上，此文件位于安装目录中的 `log` 文件夹内；在 Windows 上，此文件位于 `%ALLUSERSPROFILE%/IBM/SPSS/Modeler Server/<version>/log` 中。

用于控制如何在安装中执行日志记录的设置包含在 `log4cxx.properties` 文件中。

更改日志文件的大小

日志文件的缺省位置在 `log4cxx.properties` 文件中设置为：

```
log4j.appender.MainLog.File=${app_log_location}/${PROFILE_NAME}/${app_type}logging.log
```

要更改日志文件位置，请编辑此条目。

启用跟踪

在有些情况下，可能需要比信息的基本列表（显示主要操作）更精细的详细信息；例如，支持人员可能要求提供此详细信息以帮助识别问题。在这些情况下，您可以修改日志以提供更详细的跟踪信息。

要启用跟踪，请在 `log4cxx.properties` 文件中禁用行 `log4j.rootLogger=INFO, MainLog, ConsoleLog`，然后在其位置中启用以下行：`log4j.rootLogger=TRACE, MainLog, TraceLog`

要更改跟踪日志的位置，请编辑以下条目：

```
log4j.appender.TraceLog.File=${app_log_location}/${PROFILE_NAME}/${app_type}tracing_${PROCESS_ID}.log
```

修改日志记录选项

`log4cxx.properties` 文件包含用于控制如何记录各种事件的控件。这些控件通常设置为 **INFO**（记录日志文件中的操作）或 **WARN**（通知用户潜在问题）。如果是使用日志文件识别潜在问题，那么也可以将某些控件设置为 **TRACE**。

控制日志文件的大小

缺省情况下，每次您使用 SPSS Modeler Server 时，日志文件的大小都继续增长。要防止日志变得过大，可以将其设置为每天从头开始启动，或者为其定义大小限制。

要将日志设置为每天作为新日志启动，请在 `log4cxx.properties` 文件中使用以下条目：

```
log4j.appender.MainLog=org.apache.log4j.DailyRollingFileAppender
```

```
log4j.appender.MainLog.DatePattern='.'yyyy-MM-dd
```

或者，要为日志定义大小限制（例如 8 Mb），请在 `log4cxx.properties` 文件中使用以下条目：

```
log4j.appender.MainLog=org.apache.log4j.RollingFileAppender
```

```
log4j.appender.MainLog.MaxFileSize=8MB
```

第 5 章 性能概述

分析数据时的实际性能受众多因素影响，其中包括服务器和数据库配置，以及流中各个节点的顺序等等。一般来说，您可以通过执行下列操作获取最佳的性能：

- 将数据存储在 DBMS 中，并且尽量使用 SQL 生成和优化功能。
- 使用满足或超出第 5 页的第 2 章，『基础架构和硬件建议』中给出的建议的硬件。
- 确保正确配置客户端和服务器的性能及优化设置。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。
- 对流进行设计，实现最佳性能。

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 IBM SPSS Deployment Manager 中包含的 IBM SPSS Modeler Administration Console 界面调整这些设置。有关更多信息，请参阅第 30 页的『IBM SPSS Modeler Server 管理』主题。

这些设置组合在 IBM SPSS Modeler Administration Console 配置窗口的**性能和优化**标题下。在大多数安装中，都以最佳性能为目标对这些设置进行了预先配置。但是，您可能需要根据您的特定硬件、数据集大小以及流内容调整这些设置。有关更多信息，请参阅第 34 页的『性能/优化』主题。

客户端性能和优化设置

"流属性"对话框的"选项"选项卡提供了客户端性能和优化设置。要显示这些选项，请从客户端菜单中进行如下选择。

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。如果在服务器中禁用了这些设置，那么客户机无法将其启用。但是，如果在服务器中启用了这些设置，那么客户机可以选择将其禁用。

注：数据库建模和 SQL 优化要求在 IBM SPSS Modeler 计算机上启用 IBM 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 IBM SPSS Modeler 回送 SQL 以及访问 IBM SPSS Modeler Server。要验证当前许可证的状态，请从 IBM 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**。

有关更多信息，请参阅第 11 页的『正在连接到 IBM SPSS Modeler Server』主题。

注：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 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 IBM SPSS Modeler 中的流重写。提供了四种重写类型，您可以选择其中的一种或多种类型。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。在发现某个节点无法以 SQL 表达时，优化器将会向前检查，确定是否有任何下游节点能够以 SQL 表达并可以在不影响流语义的情况下安全地移到问题节点前面。数据库执行操作不但可以比 IBM SPSS Modeler 效率更高，而且这样的回送操作可以减小返回到 IBM 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。
- **优化 CLEM 表达式。**此选项启用优化器来搜索能够在流运行之前被预处理的 CLEM 表达式，从而提高处理速度。举个简单的示例，如果您有一个像 $\log(\text{salary})$ 这样的表达式，优化器将计算实际工资值并通过传递来对其进行处理。这可以用来提高 SQL 后推和 IBM SPSS Modeler Server 性能。
- **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而受益：C5.0、合并（按键进行）、排序、分箱（秩法和分位法）以及汇总（使用一个或多个关键字段）。

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。

- **数据库缓存。**对于生成了要在数据库中执行的 SQL 的流，可以在中游将数据缓存到数据库中的临时表，而不是缓存到文件系统。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。如果已启用数据库缓存，则只需用右键单击任意非终端节点就可以缓存该点上的数据，并且在下次运行流时，可以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，则会转而将缓存写入文件系统中。
- **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，如果数据作为字符串保存在数据库中，但实际包含有意义的数字，那么可以转换数据以便在回送时使用。

注：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 IBM 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供应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。如果有可能，请创建一个专门用于数据挖掘的数据库实例，以使生产服务器不受 IBM SPSS Modeler 查询影响。IBM SPSS Modeler 生成的 SQL 语句可能要求很高 - IBM SPSS Modeler Server 计算机上的多个任务可以在同一个数据库中执行 SQL。

数据库内挖掘。很多数据库供应商都提供了用于其产品的数据挖掘扩展。这些扩展使得数据挖掘活动（例如模型构建或评分）能够在数据库服务器内或者在独立的专用服务器内运行。IBM SPSS Modeler 的数据库内挖掘功能补充并扩展了它的 SQL 生成功能，从而提供了一种驱动供应商特有的数据库扩展的方式。在某些情况下，

采用此方式避免了 IBM SPSS Modeler 和数据库之间用于数据传输的潜在的昂贵开销。数据库缓存可以进一步增加益处。有关详细信息，请参阅作为部分下载 elmage 提供的 *DatabaseMiningGuide.pdf* 文件。

SQL 优化

为了实现最佳性能，您始终应该尝试尽可能增加所生成的 SQL 数量，以便利用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。有关更多信息，请参阅第 53 页的第 6 章，『SQL 优化』。

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。如果要分析的数据尚未存在于数据库中，您可以使用“数据库输出”节点上载这些数据。您还可以使用此节点来存储数据准备工作所产生的中间数据集以及部署结果。

IBM SPSS Modeler 可以与许多常用数据库系统的外部加载程序进行交互。该软件还附带了几个脚本，它们及其文档位于 IBM SPSS Modeler 安装文件夹下的 */scripts* 子目录中。

下表显示了批量加载的潜在性能优势。这些数据显示向 Oracle 数据库导出 250,000 个记录和 21 个字段所耗用的时间。外部加载程序为 Oracle 的 *sqlldr* 实用程序。

表 2. 批量加载的性能优点

导出选项	时间 (秒)
缺省 (ODBC)	409
通过 ODBC 的批量加载	52
通过外部加载程序的批量加载	33

第 6 章 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 IBM 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 或驱动程序的效率较低。因此，由于 SQL 优化而受益最多的操作是行和列的选择和汇总（“选择”、“采样”和“汇总”节点），这样通常会减小所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）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 IBM 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 IBM SPSS Modeler Server 执行。但是，由于可以使用标准 SQL (SQL-92) 表示的内容有限制，可能不支持某些操作。有关更多信息，请参阅第 57 页的『最大化 SQL 生成的提示』主题。

注：请谨记以下有关 SQL 的信息：

- 由于 SQL 实现之间存在细微差别，因此在数据库中执行的流所返回的结果可能与其在 IBM SPSS Modeler 中执行时返回的结果略有不同。由于类似的原因，这些差别还可能根据数据库供应商不同而有所变化。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 IBM 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 IBM SPSS Modeler 的兼容性，数据库字符串比较应区分大小写。
- 数据库建模和 SQL 优化需要在 IBM SPSS Modeler 计算机上启用 IBM 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 IBM SPSS Modeler 回送 SQL 以及访问 IBM SPSS Modeler Server。要验证当前许可证的状态，请从 IBM SPSS Modeler 菜单中选择以下项目。
- 使用 IBM SPSS Modeler 生成 SQL 时，在某些平台（例如 Linux/zLinux）上，使用 SQL 后推的结果可能与 IBM SPSS Modeler 本机不一致。原因是浮点在不同平台上以不同方式进行处理。

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。

有关更多信息，请参阅第 11 页的『正在连接到 IBM SPSS Modeler Server』主题。

注：在 Netezza 数据库中执行流时，将从该数据库中获取日期和时间详细信息。例如，如果数据库所在的机器位于另一国家或地区或者时区，那么日期和时间可能与您的本地或 IBM SPSS Modeler Server 日期和时间不同。

数据库需求

有关使用 IBM 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，使用数据库建模功能还可以进一步提高性能。

ODBC 驱动程序设置

在 32 位 Windows 8 系统上使用 SQL 2012 时，为了确保正确处理时间详细信息（例如 HH:MM:SS），在设置 ODBC SQL Server Wire Protocol Driver 时，应同时选中启用括在引号内的标识和以时间形式访存 **TWFS** 选项。

SQL 生成如何运行

来自"数据库源"节点的流的初始片段是 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。有关更多信息，请参阅第 57 页的『最大化 SQL 生成的提示』主题。

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（按键合并）。**连接操作可以增强数据库内的优化。
- **汇总。**"汇总"、"分布"和"Web"节点全都使用汇总功能来生成其结果。汇总后的数据使用的带宽比原始数据要小很多。
- **选择。**根据特定条件选择记录可以减少记录的数量。
- **排序。**对记录进行排序是一项耗费资源的活动，在数据库中执行会更有效率。
- **字段派生。**在数据库中生成新字段效率更高。
- **字段投影。**IBM SPSS Modeler Server 仅从数据库中抽取后续处理所需的字段，这样可以最大限度降低带宽和内存需求。对于平面文件中的多余字段也是如此：尽管服务器必须读取多余的字段，但不会为其分配任何存储器。
- **评分。**可以根据决策树、结果集、线性回归以及因子生成的模型生成 SQL。

SQL 生成示例

下面的流会按键操作连接三个数据库表，然后执行一次汇总和一次排序。

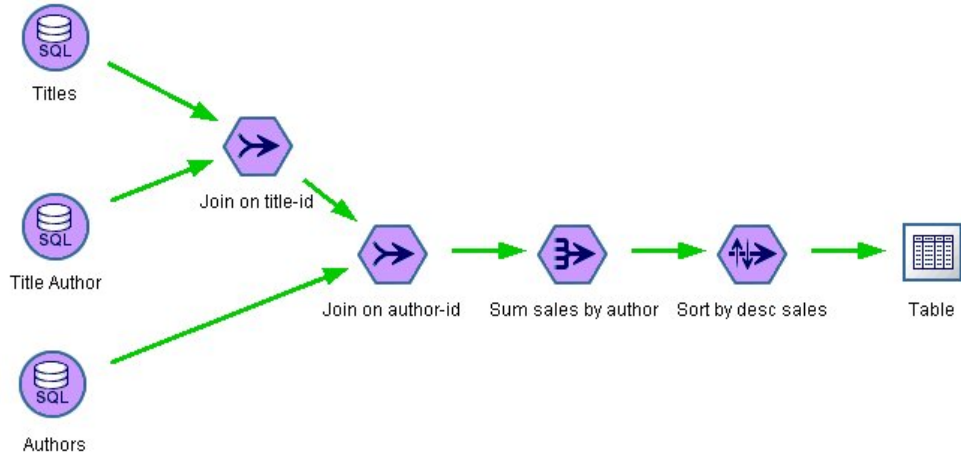


图 3. 经过优化的流 - 紫色节点指示 SQL 回送（在数据库中执行的操作）

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。

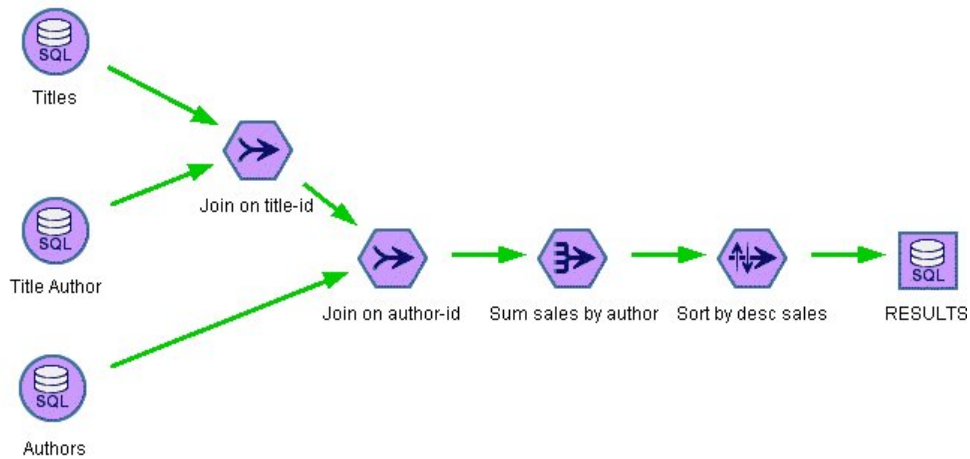


图 4. 在数据库中执行的整个流

配置 SQL 优化

1. 安装 ODBC 驱动程序，并为所要使用的数据库配置数据源。有关更多信息，请参阅第 8 页的『数据访问』主题。
2. 创建一个使用源节点从该数据库中拉取数据的流。
3. 进行检查，确保在客户端和服务器上启用了 SQL 生成（如果适用）。缺省情况下客户端和服务上都是启用的。

在客户端启用 SQL 优化

1. 在"工具"菜单中，选择流属性 > 选项。
2. 单击"优化"选项卡。选择 **生成 SQL 启用 SQL 优化**。（可选）您可以选择其他设置以提高性能。有关更多信息，请参阅第 49 页的『客户端性能和优化设置』主题。

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 设置的更多信息，请参阅第 34 页的『性能/优化』一节。请注意，如果在服务器中禁用了这些设置，那么客户机无法将其启用。但是，如果在服务器中启用了这些设置，那么客户机可以选择将其禁用。

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。这样做是因为一些模型会生成极其复杂的 SQL 表达式，这些表达式在数据库中可能无法有效地进行求值。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。

您可能需要进行特定数量的试错，以确定 SQL 生成能否针对给定模型提高性能。此任务在生成的模型添加到流之后通过"设置"选项卡完成。

预览生成的 SQL

在数据库中执行生成的 SQL 之前，可以先在消息日志中进行预览。这可能有助于您执行调试，并且，您还可以将生成的 SQL 导出，以便将来进行编辑或者在数据库中运行。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提高性能。

1. 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**。有关更多信息，请参阅第 49 页的『客户端性能和优化设置』主题。
2. 在流工作区中，选择要预览的节点或流。
3. 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。

4. 要预览生成的 SQL，请从菜单中选择：

工具 > 流属性 > 消息...

查看模型块 SQL

对于某些模型，可以生成模型块的 SQL，从而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将其完整的 SQL 回送。有关更多信息，请参阅第 58 页的『支持 SQL 生成的节点』主题。

要查看支持 SQL 生成的模型块 SQL：

1. 选择模型块上的“设置”选项卡。
2. 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。
3. 在模型块菜单中，选择：

文件 > 导出 SQL

4. 保存该文件。
5. 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 所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。发生这种情况时，任何下游节点的 SQL 生成也都会受抑制。因此，通过对节点重新排序，使导致 SQL 停止的操作尽可能放在下流，可能能够显著改善性能。SQL 优化器可以自动执行特定数量的重新排序（仅仅确保启用流重写），但还可以进行进一步的改进。“选择”节点就是一个很好的备选，此节点通常可以前移。有关更多信息，请参阅第 58 页的『支持 SQL 生成的节点』主题。

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方式，以使其不再阻止 SQL 生成。导出、选择以及类似的节点通常可以用 SQL 表达，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，但有许多运算符会阻止 SQL 生成（尤其是序列函数["@ 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。有关更多信息，请参阅第 61 页的『支持 SQL 生成的 CLEM 表达式和运算符』主题。

多个源节点。对于流具有多个“数据库源”节点的情况，SQL 生成将独立应用于每个输入分支。即使生成在某个分支中停止，也可以在另一个分支中继续。对于两个分支合并的情况（并且两个分支在合并前都可以表示为 SQL），合并本身通常可以替换为数据库连接，生成可以在下游继续。

数据库算法。模型估算总是在 IBM SPSS Modeler Server 上执行，而不是在数据库中执行，但使用 Microsoft、IBM 或 Oracle 提供的数据库自有算法时除外。

对模型评分。通过将生成的模型呈现为 SQL，支持对某些模型进行数据库内评分。但是，一些模型会生成极其复杂的 SQL 表达式，而这些表达式在数据库中并非始终都能有效地进行求值。因此，必须对各个模型节点分别启用 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）。运行测试，确认此选项是否有益于您的应用。有关更多信息，请参阅第 58 页的『支持 SQL 生成的节点』主题。

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，这需要从 Windows 任务管理器关闭 IBM SPSS Modeler。

数据库缓存。如果要使用节点缓存在流中的某些重要点（例如，在"合并"或"汇总"节点之后）保存数据，请确保启用数据库缓存和 SQL 优化。这样将使数据在大多数情况下都能够缓存到数据库中的临时表（而不是缓存到文件系统）。有关更多信息，请参阅第 56 页的『配置 SQL 优化』主题。

特定于供应商的 SQL。生成的大部分 SQL 都符合标准 (SQL-92)，但在实际应用中，会利用一些特定于供应商的非标准功能。根据数据库源不同，SQL 优化程度可能有所变化。

支持 SQL 生成的节点

下列各表显示了一些节点，这些节点表示支持 SQL 生成功能的数据挖掘操作。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。有关更多信息，请参阅第 56 页的『预览生成的 SQL』主题。

表 3. 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。此节点允许输入 SQL 查询。请避免包含重复列名的结果集。有关更多信息，请参阅第 63 页的『写入 SQL 查询』主题。

表 4. 记录操作

支持 SQL 生成的节点	附注
选择	仅当支持 SELECT 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	根据数据库不同，简单采样在各种程度上支持 SQL 生成。请参阅表 5。
汇总	对汇总的 SQL 生成支持依赖于数据存储类型。请参阅第 59 页的表 6。
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 datetime_date(YEAR,MONTH,DAY) 函数，则包括最近几次交易时，也支持生成。
排序	
合并	对于按顺序合并，不会生成任何 SQL。 仅当数据库/驱动程序支持具有完全或部分外连接的按键合并时，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 SQL_SP_EXISTS。
追加	如果输入未经排序，那么支持生成。 注：仅当您的输入拥有相同列数时，才有可能实现 SQL 优化。
区分	选择了（缺省）方式为每个组创建组合记录的"区分"节点不支持 SQL 优化。

表 5. 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 值	种子	DB 2 for z/OS	DB2 for OS/400	DB2 for Win/ UNIX	Netezza	Oracle	SQL Server	Teradata
包括	第一个 (F)	n/a		Y	Y	Y	Y	Y	Y	Y

表 5. 在样本节点中支持 SQL 生成以用于简单抽样 (续).

模式	样本	大小最大值	种子	DB2 for z/OS	DB2 for OS/400	DB2 for Win/UNIX	Netezza	Oracle	SQL Server	Teradata
	n 中取 1	关闭		Y	Y	Y	Y	Y		Y
		最大		Y	Y	Y	Y	Y		Y
	随机 %	关闭	关闭	Y		Y	Y	Y		Y
			在	Y		Y		Y		
		最大	关闭	Y		Y	Y	Y		Y
			在	Y		Y		Y		
放弃	第一个 (F)	关闭					Y	Y		
		最大					Y	Y		
	n 中取 1	关闭		Y	Y	Y	Y	Y		Y
		最大		Y	Y	Y	Y	Y		Y
	随机 %	关闭	关闭	Y		Y	Y	Y		Y
			在	Y		Y		Y		
		最大	关闭	Y		Y	Y	Y		Y
			在	Y		Y		Y		

表 6. 在汇总节点中支持 SQL 生成.

存储	合计	平均值	最小值	最大值	标准差	中位数	计数	方差	百分位
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		
时间戳记			Y	Y			Y		
字符串			Y	Y			Y		

表 7. 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。
过滤器	
派生	如果支持为派生表达式生成的 SQL, 那么支持 SQL 生成 (参见下面的表达式)。
整体	对于连续目标, 支持 SQL 生成。对于其他目标, 仅当使用"最高置信度获胜"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	

表 7. 字段操作 (续)

支持 SQL 生成的节点	附注
分箱	如果使用"分位数 (同等计数)"分箱方法, 并且选中了"从'分箱值'选项卡读取 (如果可用)"选项, 那么支持 SQL 生成。 注: 由于分箱边界的计算方式存在差异 (这是由分箱字段中的数据分布性质所造成), 因此在比较一般流执行结果和 SQL 后推结果时, 分箱输出可能会有不同。要避免此情况, 请使用记录计数平铺方法, 或者添加到下一个分箱或在当前分箱中保留平铺, 以获取两种流执行方法之间最接近的匹配。
RFM 分析	如果选中"从'分箱值'选项卡读取 (如果可用)"选项, 则支持 SQL 生成, 但下游节点不支持。
分区	支持 SQL 生成以分配记录到分区。
设为标志	
重建	

表 8. 图形

支持 SQL 生成的节点	附注
图形板	下列图形类型支持 SQL 生成: 面积图、3-D 面积图、条形图、3-D 条形图、计数条、热图、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。图形板中的动画不支持 SQL 生成。
分布	
Web	
评估(E)	

对于某些模型, 可以生成模型块的 SQL, 从而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将其完整的 SQL 回送。有关更多信息, 请参阅第 57 页的『查看模型块 SQL』主题。

表 9. 模型块

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、组装或大型数据集选项的 SQL 生成。
QUEST	
CHAID	
C5.0	
决策列表	
线性	支持标准模型选项的 SQL 生成, 但不支持推进、组装或大型数据集选项的 SQL 生成。
类神经网络	支持标准模型选项 (仅多层感知器) 的 SQL 生成, 但不支持推进、组装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	
自动分类器	如果启用了用户定义的函数 (UDF) 评分适配器, 那么这些块支持 SQL 回送。并且, 如果为连续目标生成 SQL, 或者使用了"最高置信度获胜"整体方法, 那么这些块支持进一步向下游回送。
自动数值	

表 10. 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(M)	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。
变换(T)	
图形	如果在图形板节点中启用动画，那么不支持 SQL 回送。
Statistics	如果不使用"相关"选项，则支持生成。
报告	
设置全局量	

表 11. 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。

支持 SQL 生成的 CLEM 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。

表 12. 运算符

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

表 13. 关系运算符

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

表 14. 函数

支持 SQL 生成的运算	附注
abs	
allbutfirst	
allbutlast	

表 14. 函数 (续)

支持 SQL 生成的运算	附注
and	
arccos	
arcsin	
arctan	
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	

表 14. 函数 (续)

支持 SQL 生成的运算	附注
strmember	
subscrs	
substring	
substring_between	
uppertolower	
to_string	

表 15. 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量节点计算的全局值。
@GLOBAL_SUM	
@GLOBAL_MAX	
@GLOBAL_MEAN	
@GLOBAL_MIN	
@GLOBALSDEV	

表 16. 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 或其他供应商特有定制的特殊情况，此函数非常有用。

此函数的使用不在标准 IBM SPSS Modeler 支持协议的涉及范围之内，这是因为执行所依赖的外部数据库组件超出 IBM Corp. 的控制范围，但在特殊情况下可以进行部署（通常作为服务合约的组成部分）。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，您应该对所有将会产生包含重复列名的数据集的 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 语句为那些在"数据库源"节点中使用了 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，如果查询结果集包含重复的列名，那么 RDBMS 将无法嵌套该语句。嵌套问题通常在多个相连接的表中选择了同名的列的表连接期间发生。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 JOINDEPT d ON e.ID = d.ID;
```

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

为了允许进行完全 SQL 优化，您应该更加明确地编写 SQL 查询，出现列名重复的情况时，应指定列别名。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 JOINDEPT d ON e.ID = d.ID;
```

Scoring Adapter for Teradata - 重复行

IBM SPSS Modeler Server Scoring Adapter for Teradata 期望输入数据中无相同的行。Teradata 不允许表中存在两个相同的行。但是，如果连接表，或者在用户仅使用表字段的一部分作为输入时可能出现重复行。这些重复行将导致在笛卡尔连接后记录数不正确。

附录 A. 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，请考虑下面的提示，以确保在数据库中彻底地优化生成的 SQL。

指定正确的语言环境

当在除连接 ODBC 驱动程序随附环境以外的语言环境中运行 IBM 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* 语言环境文件连接。结果，如果在不同的语言环境中运行 IBM SPSS Modeler Server 计算机，或在启动了 IBM 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：

- IBM SPSS Modeler Server 使用与其运行语言环境相对应的 ODBC 语言环境文件，以便将那些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的数据库。
- 如果语言环境（使用系统 `$LANG` 查询返回到 IBM 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 IBM SPSS Modeler Server 的是未转换的代码，而不是字符串 *Oracle*。这意味着 IBM SPSS Modeler 无法优化 Oracle 的流。

检查并重设语言环境规范：

1. 在 UNIX 命令解释程序中，运行：

```
#locale
```

这将返回 Shell 的语言环境信息。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
```

2. 切换到 `Connect ODBC/locale` 目录。（您将在此处看到一个单独的目录，*en_US*。）
3. 创建指向此 *en_US* 目录的软链接，指定命令解释程序中语言环境的名称。示例如下：

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

对于非英语语言环境，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
```

4. 创建了此链接后，请从同一命令解释程序中重新启动 IBM SPSS Modeler Server。（IBM SPSS Modeler Server 将接收到来自它从中启动的 Shell 的语言环境信息。）

注意

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

- 必须指定完整的语言环境。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 语言环境目录并不充分。
- 要完全优化数据库内挖掘，必须在用于启动 IBM SPSS Modeler Server 的命令解释程序中定义 LANG 和 LC_ALL。在重新启动 IBM SPSS Modeler Server 之前，可以在 Shell 中定义 LANG，就像定义任何其他环境变量一样。例如，请参见下面的定义：

```
#LANG=en_US.IS08859-15; export LANG
```

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

附录 B. 配置 UNIX 启动脚本

介绍

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 IBM 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 IBM SPSS Modeler Server。
- 更改 IBM SPSS Modeler Server 所创建的文件权限。
- 配置 IBM SPSS Modeler Server 以和 IBM SPSS Modeler Server 提供的 ODBC Connect 驱动程序一起使用。有关更多信息，请参阅第 68 页的『IBM SPSS Modeler Server 和数据访问包』主题。

脚本

IBM SPSS Modeler Server 使用多个脚本，其中包括：

- **modelersrv.sh**。IBM SPSS Modeler Server 的手动启动脚本位于 IBM 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停止服务器。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。运行一次此脚本即可将系统配置为自动启动。您无需对此脚本进行编辑。此脚本位于 IBM 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（取决于服务器的操作系统）创建此脚本。当服务器自动启动时，此脚本为其配置环境。如果您需要更改自动启动配置，请编辑此脚本。

表 17. 不同操作系统中 *rc.modeler* 的位置。

操作系统	位置
AIX	<i>/etc/rc.modeler</i>

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 IBM SPSS Modeler Server 的脚本 (*auto.sh*)。

将系统配置为自动启动和关闭

1. 以 root 用户身份登录。
2. 切换至 IBM SPSS Modeler Server 安装目录。
3. 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*)。每当服务器计算机重新引导时，操作系统将使用 *rc.modeler* 来启动 IBM 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

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

```
./modelersrv.sh start
```
3. 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

1. 停止 IBM SPSS Modeler Server。（请参阅主题『手动启动和停止 IBM SPSS Modeler Server』以获取更多信息。）
2. 找到相应脚本。（请参阅主题第 67 页的『脚本』以获取更多信息。）
3. 用文本编辑器打开脚本，修改并保存文件。
4. 自动（通过重新启动服务器计算机）或手动启动 IBM SPSS Modeler Server。

控制文件创建权限

IBM SPSS Modeler Server 将创建临时文件，并将其读、写和执行权限授予每个人。您可通过编辑启动脚本（*modelersrv.sh* 和/或 *rc.modeler*）中的 *UMASK* 设置来覆盖此缺省值。（有关更多信息，请参阅前面的『编辑脚本』。）我们建议您使用 *077* 作为限制性最强的 *UMASK* 设置。限制性较强的设置可能会导致 IBM SPSS Modeler Server 发生权限问题。

IBM SPSS Modeler Server 和数据访问包

如果要将 ODBC 驱动程序与 IBM SPSS Modeler Server 配合使用，那么在 IBM SPSS Modeler Server 进程启动时，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。（请参阅『编辑脚本』以获取更多信息。）

有关更多信息，请参阅技术支持 Web 站点 (<http://www.ibm.com/support>)。如果您有关于创建或设置 ODBC 数据源权限方面的疑问，请与数据库管理员联系。

配置 ODBC 以开始使用 IBM SPSS Modeler Server

1. 如果 IBM SPSS Modeler Server 主机处于运行状态，请将其停止。
2. 下载适用于 IBM SPSS Modeler Server 安装平台的压缩 TAR 归档。确保下载适用于您安装的 IBM SPSS Modeler Server 版本的正确驱动程序。将文件复制到您要安装 ODBC 驱动程序的位置（例如，*/usr/spss/odbc*）。
3. 使用 **tar -xvof** 提取 TAR 归档文件。
4. 运行从存档文件中提取的 *setodbcpath.sh* 脚本。
5. 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, IBM SPSS Modeler 将从该文件中读取您所定义的 ODBC 数据源列表 (缺省的 odbc.ini 随驱动程序一起安装)。

6. 保存 odbc.sh。
7. (仅适合 64 位 IBM 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)。完成此任务的最简单方法是, 在 odbc.sh 文件中复制 LD_LIBRARY_PATH 的 if 和 export 语句, 并将其追加到文件末尾, 然后将新追加的 if 和 export 语句中的"LD_LIBRARY_PATH"字符串替换为"LD_LIBRARY_PATH_64"。

例如, 在 64 位 IBM 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
```

请不要忘记导出 LD_LIBRARY_PATH_64, 并使用 if 循环对其进行定义。

8. 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 (这些名称依赖于您要访问的数据库)。
9. 保存 odbc.ini 文件。
10. 配置 IBM 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: 此处的语法十分重要; 请确保在第一个句点与文件路径之间保留一个空格。

11. 保存 modelersrv.sh。

要点: 为使 SDAP 7.1 驱动程序在 DB2 on z/OS 上工作, 您必须授予对 SYSIBM.SYSPACKSTMT 的访问权。

测试连接

1. 重新启动 IBM SPSS Modeler Server。
2. 从客户端连接到 IBM SPSS Modeler Server。
3. 在客户端上，将“数据库源”节点添加到工作区中。
4. 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。

如果在此处没有看到您期望的内容，或者尝试连接到定义的数据源时发生错误，请执行故障诊断过程。有关更多信息，请参阅第 71 页的『ODBC 配置故障诊断』主题。

配置 ODBC 以开始使用 IBM SPSS Modeler Solution Publisher Runtime

成功地从 IBM SPSS Modeler Server 连接到数据库后，您可以通过从 IBM SPSS Modeler Solution Publisher Runtime 的启动脚本中引用同一个 `odbc.sh` 脚本，在同一服务器上配置 IBM SPSS Modeler Solution Publisher Runtime 安装。

1. 编辑 IBM 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：此处的语法十分重要。请确保在第一个句点与文件路径之间保留一个空格。

2. 保存 `modelerrun` 脚本文件。
3. 在 UNIX 系统上，缺省情况下，未配置 DataDirect 驱动程序管理器，因此 IBM SPSS Modeler Solution Publisher Runtime 无法使用 ODBC。要配置 UNIX 以加载 DataDirect 驱动程序管理器，请输入以下命令（其中 `sp_install_dir` 是 Solution Publisher Runtime 的安装目录）：

```
cd sp_install_dir
rm -f libspssodbc.so
ln -s libspssodbc_datadirect.so libspssodbc.so
```

配置 ODBC 以开始使用 IBM SPSS Modeler Batch

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

添加或编辑数据源名称

1. 编辑 `odbc.ini` 文件以包含新的或更改后的名称。
2. 测试连接，如本节前面的内容所述。

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

SQL Server 对 Data Access Pack 驱动程序的支持

SQL Server 的 ODBC 配置必须将启用括在引号内的标识 ODBC 连接属性设置为是（该驱动程序的缺省值是否）。在 UNIX 上，该属性在系统信息文件 (`odbc.ini`) 中使用 `QuotedId` 选项进行配置。

ODBC 配置故障诊断

未列出数据源，或者显示了随机文本

打开"数据库源"节点后，如果可用数据源列表为空或包含意外的条目，那么可能是由于启动脚本存在问题所致。

1. 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。
2. 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。
3. 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
```

4. 保存并执行 *modelersrv.sh*。在 IBM SPSS Modeler Server 环境中设置的 \$ODBCINI 值将被写入 *stdout* 以供验证。
5. 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。这应该是：

```
. <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：此处的语法十分重要；请确保在第一个句点与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，则在您重新启动 IBM 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

此行为可能是由于未将库路径正确设置为包括 ODBC 库路径所致。库路径由 \$LD_LIBRARY_PATH（以及 64 位版本中的 \$LD_LIBRARY_PATH_64）定义。

要在 IBM 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 库路径追加到库路径的所在行之后），并执行脚本。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 IBM 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，请在脚本中添加下面这一行，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，请将此位置追加到文件中。

如果您在运行 64 版本的 IBM 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，并且未指定缺省驱动程序

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。请检查文件开头处 [ODBC Data Sources] 部分中定义的数据源名称 (DSN)，它应与 *odbc.ini* 中用于定义 DSN 的较下方部分中方括号之间指定的字符串匹配。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=
....
```

您需要更改两个粗体字符串中的一个，以使它们完全匹配。这样做应该能够更正错误。

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。一种可能的原因是驱动程序节中的 *Driver* 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=
```

1. 请检查 *Driver* 参数所指定的共享对象是否存在。
2. 如果共享对象的路径不正确，请将其更正。
3. 如果以如下格式指定 *Driver* 参数：

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

那么这表明尚未初始化与 ODBC 相关的脚本。运行随驱动程序安装的 *setodbcpath.sh* 脚本。有关更多信息，请参阅第 68 页的『IBM SPSS Modeler Server 和数据访问包』主题。运行此脚本后，您应该会看到字符串“ODBCHOME”已替换为 ODBC 安装路径。这应当可以解决该问题。

另一原因可能是驱动程序库存在问题。请使用 ODBC 随附的 *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：

1. 使用 *ivtestlib* 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 ODBCDir
. odbc.sh
./bin/ivtestlib MFor815
```

在此，请将 *ODBCDir* 替换为 ODBC 安装目录的路径。

2. 阅读消息，确定是否存在错误。例如，消息 *Load of MFor815.so failed: ld.so.1: bin/ivtestlib: fatal: libclntsh.so: open failed: No such file or directory* 指示缺少 Oracle 客户机库 *libclntsh.so*，或它不在库路径上。

3. 确认库是否存在。如果不存在，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，请将 */bigdisk/oracle/product/8.1.6/lib* 替换为 *libclntsh.so* 的路径，LD_LIBRARY_PATH 是操作系统的库路径变量。

请注意，如果正在 Linux 上运行 64 位 IBM 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
```

4. 查阅消息以确认现在是否可以加载驱动程序。例如，消息 *Load of MFor815.so successful, qehandle is 0xFF3A1BE4* 指示可以加载 Oracle 客户机库。
5. 更正 IBM SPSS Modeler 启动脚本中的库路径。
6. 使用编辑过的启动脚本 (*modelersrv.sh* 或 *rc.modeler*) 重新启动 IBM SPSS Modeler Server。

库路径

库路径变量的名称随操作系统不同而有所变化。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 18. 不同操作系统的库路径.

操作系统	库路径变量名称
AIX	LIBPATH
Linux 64 位	LD_LIBRARY_PATH_64

附录 C. 在 UNIX 上配置 SPSS Modeler Server 并使其作为非 root 用户进程运行

介绍

这些指示信息提供有关在 UNIX 系统上使 IBM SPSS Modeler Server 作为非 root 用户进程运行的信息。

作为 root 用户进程运行。IBM SPSS Modeler Server 的缺省安装假定服务器守护进程作为 root 用户进程运行。作为 root 用户进程运行允许 IBM SPSS Modeler 可靠地对每个用户的登录进行认证，并且将使用相应的 UNIX 用户帐户来启动每个用户会话。这样可以确保用户只能访问他们自己的文件和目录。

作为非 root 用户进程运行。使 IBM SPSS Modeler Server 作为非 root 用户进程运行表示将服务器守护进程的真实有效用户标识设置为您选择的帐户。由 SPSS Modeler Server 启动的所有用户会话将使用同一个 UNIX 帐户，这表示 SPSS Modeler 读或写的所有文件数据都由所有 SPSS Modeler 用户共享。由于用户必须向他们使用的每个数据库数据源独立认证自己的身份，因此对数据库数据的访问权不受影响。在不具有 root 用户特权的情况下，IBM SPSS Modeler 将按以下两种方式之一操作：

- **不使用专用密码数据库。**如果采用此方法，那么 SPSS Modeler 将使用 UNIX 系统上通常用于执行用户认证的现有 UNIX 密码数据库、NIS 或 LDAP 服务器。有关更多信息，请参阅『在不使用专用密码数据库的情况下配置为非 root 用户进程』主题。
- **使用专用密码数据库。**如果采用此方法，那么 SPSS Modeler 将根据有别于 UNIX 上通常用于执行认证的 UNIX 密码数据库、NIS 或 LDAP 服务器的专用密码数据库来认证用户。有关更多信息，请参阅第 76 页的『在使用专用密码数据库的情况下配置为非 root 用户进程』主题。

在不使用专用密码数据库的情况下配置为非 root 用户进程

要在不需要专用密码数据库的情况下将 IBM SPSS Modeler Server 配置为使用非 root 用户帐户运行，请完成下列步骤：

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

缺省情况下，SPSS Modeler Server 将尝试每一种认证方法，直至找到一种可行的方法为止。但是，如果您愿意，可以使用 *options.cfg* 中的 `authentication_methods` 选项将服务器配置为仅尝试一种特定的认证方法。此选项的可能值为 `pasw_modeler`、`gss`、`pam`、`sspi`、`unix` 或 `windows`。

请注意，以非 root 用户身份运行可能需要一些配置更新。有关更多信息，请参阅第 77 页的『对用户认证故障进行故障诊断』主题。

注意：

请勿先启用 `start_process_as_login_user` 设置，然后以 root 身份启动 IBM SPSS Modeler Server。这样做意味着，对于所有连接到服务器的用户，他们的服务器进程将以 root 用户身份运行；这存在安全性风险。请注意，如果您尝试这样做，那么服务器可能会自动停止。

在使用专用密码数据库的情况下配置为非 root 用户进程

如果您选择使用专用密码数据库来认证用户，那么将使用同一个非 root 用户帐户来启动所有用户会话。

要以此方式将 IBM SPSS Modeler Server 配置为使用非 root 用户帐户运行，请完成下列步骤：

1. 创建一个包含所有用户的组。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 IBM SPSS Modeler Server 的用户帐户。该帐户仅用于 IBM SPSS Modeler Server 守护进程。例如，称它为 *modelerserv*。

创建帐户时，请注意下列事项：

- 主要的组应为先前创建的 *<modelerusers>* 组。
 - 主目录可以是 IBM SPSS Modeler 的安装目录或任何其他适当的缺省目录（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 IBM SPSS Modeler Server。找到相应的启动脚本，并在文本编辑器中将其打开。有关更多信息，请参阅第 67 页的『脚本』主题。
 - a. 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认证：

```
authentication_methods, "pasw_modeler"
```

5. 将选项 *start_process_as_login_user* 设置为 Y。

6. 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 IBM SPSS Modeler 的用户名/密码组合。注：这些用户名和密码仅供 IBM 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*，此程序位于 IBM SPSS Modeler Server 安装位置下的 *bin* 目录中。该程序的概要如下所示：

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

此程序接收用户名和纯文本密码，并以适合于包括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，要定义密码为"data mining"的用户 *modeler*，您可以输入以下命令：

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

在大多数情况下，定义单个用户名就已足够，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：如果使用单个 *>*，那么每次都会覆盖 *passwords.cfg* 的内容，从而替换先前设置的任何用户。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。

注：如果在 SPSS Modeler Server 处于运行状态时将新用户添加到专用密码数据库，那么需要重新启动 SPSS Modeler Server 以使其能够识别新定义的用户。在执行此操作之前，在 SPSS Modeler Server 上次重新启动后通过 *pwutil* 添加的所有新用户的登录都将失败。

7. 递归式地将 IBM 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 .
```

8. 考虑为 IBM 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/bobchown bob:modelerusers data/bobchmod ug=rwx,o= data/bob
```

另外，可以在目录上设置设置组标识位，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
```

以非 root 用户身份运行 SPSS Modeler Server

要以非 root 用户身份运行 SPSS Modeler Server，请完成下列步骤：

1. 使用先前创建的非 root 用户帐户登录。
2. 如果您是在启用了配置文件选项 `start_process_as_login_user` 的情况下运行，那么可以启动和停止 SPSS Modeler Server 以及检查其状态。有关更多信息，请参阅第 21 页的『在 UNIX 中启动、停止和检查状态』主题。

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 地址或服务器的主机名。

在 AIX 上以非 root 用户身份运行 IBM SPSS Modeler Server

在 AIX 上以非 root 用户身份启动 IBM SPSS Modeler Server 时，可能会将错误找不到 `lsattr` 输出到控制台。这是因为 `lsattr` 位于 `/usr/sbin` 目录中，但是当 IBM SPSS Modeler Server 以新用户身份登录以便以非 root 用户身份运行时，会将路径重置为只包括 `/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：

```
sudo /usr/binln -s /usr/sbin/lsattr lsattr
```

创建链接后，请重新启动 IBM SPSS Modeler Server。

对用户认证故障进行故障诊断

根据对操作系统配置的认证执行方式不同，在使用非 root 用户配置运行时，登录 SPSS Modeler Server 可能会遇到故障。例如，如果使用 `/etc/nsswitch.conf` 文件或类似文件将操作系统配置为检查本地影子密码文件，而不是使用 NIS 或 LDAP，那么可能发生这种情况。这是因为 SPSS Modeler Server 必须对用来执行认证的文件（包括 `/etc/shadow` 文件或其等效文件，例如 AIX 上的 `/etc/security/passwd`，其中存储了安全用户帐户信息）具有读访问权。但是，操作系统文件权限通常设置为仅允许 root 用户访问 `/etc/shadow`。在这些情况下，非 root 用户进程无法读取 `/etc/shadow` 以验证用户密码，从而引起认证错误。

可以通过以下几种方法解决此问题：

- 请系统管理员将操作系统配置为使用 NIS 或 LDAP 进行认证。
- 更改受保护文件的文件权限，例如，通过授予 `/etc/shadow` 文件的读访问权，使得用来运行 SPSS Modeler Server 的本地用户帐户能够访问该文件。在生产环境中，此变通方法可能不适用，此方法只是临时应用于测试环境以验证授权故障是否与操作系统配置相关。
- 为 `/etc/shadow` 文件指定访问控制表 (ACL)。
- 以 root 用户身份运行 SPSS Modeler Server，以使服务器进程能够读取 `/etc/shadow` 文件。

注意：

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`，以避免上文说明的安全性问题。

附录 D. 在 Windows 上使用专用密码文件来配置和运行 SPSS Modeler Server

介绍

这些指示信息提供了有关在 Windows 系统上使用专用密码文件运行 IBM SPSS Modeler Server 的信息。通过此方法，IBM SPSS Modeler 根据专用密码数据库对用户进行认证，不同于 Windows 上的系统认证。

配置专用密码数据库

如果您选择使用专用密码数据库来认证用户，那么将使用同一个用户帐户来启动所有用户会话。

要以此方式配置 SPSS Modeler Server，请完成以下步骤：

1. 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。您必须在 SPSS Modeler Server 18.1 服务的“登录”选项卡中以该用户帐户的身份启动守护进程。例如，称它为 *modelerserv*。
2. 编辑服务器选项文件 (`config/options.cfg`)，以便将 `start_process_as_login_user` 选项设置为 Y，并通过追加下面这一行指定根据专用密码数据库进行认证：

```
authentication_methods, "pasw_modeler"
```

3. 然后，需要创建存储在文件 `config/passwords.cfg` 中的专用密码数据库。密码文件定义允许登录到 SPSS Modeler 的用户名/密码组合。请注意，这些组合仅供 SPSS Modeler 使用，与用于登录 Windows 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`，此程序位于 SPSS Modeler Server 安装位置下的 `bin` 目录中。该程序的概要如下所示：

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

此程序接收用户名和纯文本密码，并以适合于包括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，要定义为 `modeler`、密码为 `data mining` 的用户，需要使用 DOS 提示导航至 SPSS Modeler Server 安装目录，然后输入：

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

注：确保在文件中对于每个用户只有 1 个实例；重复项会阻止 SPSS Modeler Server 启动

在大多数情况下，定义单个用户名就已足够，这样所有用户都使用同一用户名和密码进行登录。但是，也可以创建更多用户，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：

如果使用单个 `>`，那么每次都会覆盖 `passwords.cfg` 的内容，从而替换先前设置的任何用户。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。

如果在 SPSS Modeler Server 处于运行状态时将新用户添加到专用密码数据库，那么需要重新启动 SPSS Modeler Server 以使其能够识别新定义的用户。在执行此操作之前，在 SPSS Modeler Server 上次重新启动后通过 `pwutil` 添加的所有新用户的登录都将失败。

4. 为步骤 1 中创建的用户提供对服务器选项文件 `config\options.cfg` 以及 `%ALLUSERSPROFILE%\IBM\SPSS` 目录的完全控制权。
5. 在系统服务中，停止 **IBM SPSS Modeler Server** 服务并将登录从本地系统帐户更改为步骤 1 中创建的用户帐户。然后重新启动该服务。

附录 E. 使用服务器集群负载均衡

对于 IBM SPSS Collaboration and Deployment Services, 可使用名为进程协调程序的插件来管理网络上的服务。进程协调程序提供了服务器管理功能, 旨在优化客户端与服务器之间的通信和处理。

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程序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。

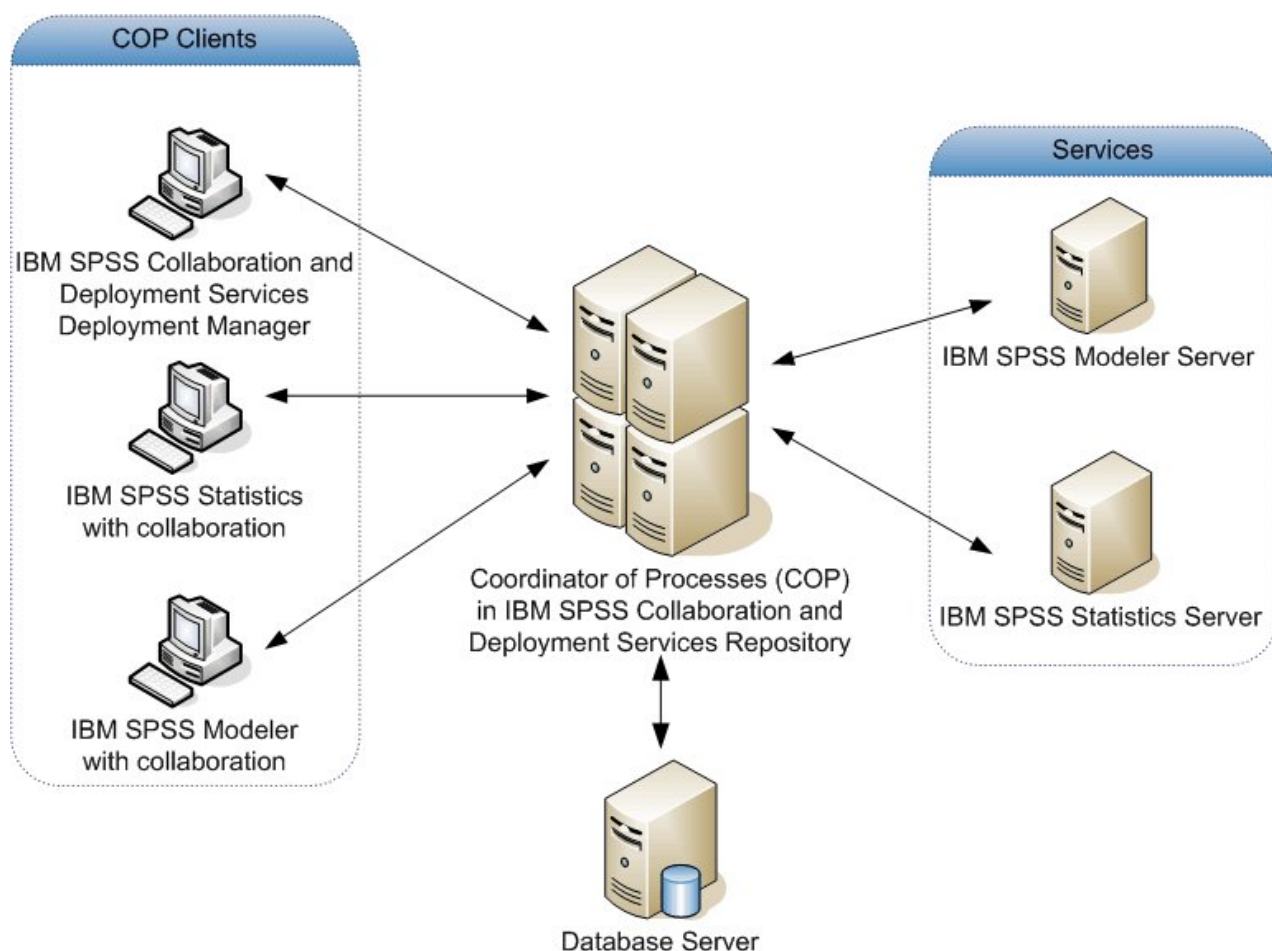


图 5. 进程协调程序体系结构

在服务器上执行 IBM SPSS Modeler 流有助于提高性能。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实质差异 (例如所有者、访问权、服务器数据、测试服务器与生产服务器等方面的差异), 所以可供选择的服务器更多。此外, 如果在网络上安装了进程协调程序, 则会向您提供一个服务器集群。

服务器集群是一组在配置和资源方面可以互换的服务器。进程协调程序使用一种根据多个条件 (包括服务器权重、用户优先级和当前处理负载) 平衡负载的算法, 确定最适合响应处理请求的服务器。有关更多信息, 请参阅 IBM SPSS Collaboration and Deployment Services 文档套件中提供的 *Coordinator of Processes Service Developer's Guide*。

无论何时连接到 IBM SPSS Modeler 中的服务器或服务器集群，都可以手动输入服务器名称或使用进程协调程序搜索服务器或服务器集群。有关更多信息，请参阅第 11 页的『正在连接到 IBM SPSS Modeler Server』主题。

附录 F. LDAP 认证

下列指示信息提供基本的准则，旨在说明在 UNIX 上如何配置 SPSS Modeler Server 以使用 LDAP 认证，其中，要登录到服务器的用户的身份存储在 LDAP 目录中。

注：作为先决条件，必须在主机操作系统上正确配置 LDAP 客户机软件。有关更多信息，请参阅原始供应商文档。

通常，无需执行其他配置工作，LDAP 的使用对服务器而言并不明显。不需要进行其他更改的情况示例包括：

- LDAP 客户机和服务器软件根据 RFC 2307 进行配置。
- 对 passwd（以及 shadow，如果适用）数据库进行的访问重定向到 LDAP，例如在 nsswitch.conf 中进行重定向。
- SPSS Modeler Server 的每个有效用户都有一个存储在 LDAP 目录中的 passwd（和 shadow）条目。
- SPSS Modeler Server 服务使用 root 用户帐户启动。

在两组情况下，可能有必要专门为 LDAP 配置 SPSS Modeler Server：

- 使用除 root 以外的帐户启动服务时，该服务可能无权使用缺省方法进行认证。通常，这是因为对影子数据库的访问权受限。
- 当用户没有存储在目录中的 passwd（或 shadow）条目时；即，他们不具有登录到主机系统所需的有效用户身份。

LDAP 认证过程使用 PAM 子系统，而且要求存在 PAM LDAP 模块并为主机操作系统正确配置该模块。有关更多信息，请参阅原始供应商文档。

请完成下列步骤，以便将 SPSS Modeler Server 配置为仅使用 LDAP 认证。

注：这些步骤提供应该能正常工作的最基本配置。根据您的操作系统以及本地安全策略不同，可能需要更多选项或替代设置。有关更多信息，请参阅原始操作文档。

1. 编辑服务配置文件 (options.cfg) 并添加（或编辑）下面这一行：authentication_methods, pam。此行指示服务器优先使用 PAM 认证，以代替缺省认证。
2. 为 SPSS Modeler Server 服务提供 PAM 配置；该服务通常需要 root 用户特权。该服务由名称 modelerserver 标识。
3. 在 Linux/Unix 类型的系统上（使用 /etc/pam.d），在该目录中创建名称为 modelerserver 的文件，并添加类似于以下示例的内容：

```
# IBM SPSS Modeler Server
auth required pam_ldap.so
account required pam_ldap.so
password required pam_deny.so
session required pam_deny.so
```

4. 对于使用 /etc/pam.conf 的较旧系统，请编辑该文件并添加类似于以下示例的行：

```
# IBM SPSS Modeler Server
modelerserver auth required pam_ldap.so
modelerserver account required pam_ldap.so
modelerserver password required pam_deny.so
modelerserver session required pam_deny.so
```

5. 所引用的 PAM 模块的名称随操作系统不同而有所变化；请确认主机操作系统所需的模块。

注：第 3 步和第 4 步中的行指定 SPSS Modeler Server 必须引用 PAM LDAP 模块以进行认证和帐户管理。但是，不支持更改密码和会话管理，因此不允许执行这些操作。如果不需要或者不适合进行帐户管理，请更改相应的行以允许所有请求，如以下示例所示：

```
# IBM SPSS Modeler Server
auth required pam_ldap.so
account required pam_permit.so
password required pam_deny.so
session required pam_deny.so
```

声明

本信息是为在美国提供的产品和服务编写的。IBM 可能以其他语言提供本材料。然而，您可能需要拥有产品或产品版本的该语言副本才能进行访问。

IBM 可能在其他国家或地区不提供本文档中讨论的产品、服务或功能特性。有关您当前所在区域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 公司可能已拥有或正在申请与本文档内容有关的各项专利。提供本文档并未授予用户使用这些专利的任何许可。您可以用书面方式将许可查询寄往：

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，或用书面方式将查询寄往：

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 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) 使其能够在独立创建的程序和其它程序（包括本程序）之间进行信息交换，以及 (ii) 使其能够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。

本资料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 国际软件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 产品的信息可从这些产品的供应商、其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 是 International Business Machines Corp. 在全球许多行政管辖地区的商标或注册商标。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。

Adobe、Adobe 徽标、PostScript 和 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 或其子公司在美国和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

只有遵守以下条款和条件才会授予使用这些出版物的许可权。

适用性

除用于 IBM Web 站点的任何条款外，还需要遵守这些条款和条件。

个人使用

您可以为了个人使用而非商业性使用复制这些出版物，但前提是保留所有专有权声明。没有 IBM 的明确同意，您不能对这些出版物或者其中的任何部分进行分发、展示或从中派生内容。

商业使用

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 的明确许可，您不得制作这些出版物的演绎作品，也不得在贵公司外部复制、分发或显示这些出版物或其部分出版物。

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 将有权撤销此处授予的许可权。

您不得下载、出口或再出口这些信息，除非完全遵守所有适用的法律和法规（包括所有的美国出口法律和法规）。

IBM 对这些出版物的内容不作任何保证。本出版物以"按现状"的基础提供，不附有任何形式的（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、非侵权以及适用于某特定用途的保证。

索引

[A]

- 安全套接字层 40
- 安全性
 - 配置 UNIX 中的文件创建 68
 - 文件创建 19
 - SSL 40

[B]

- 并行处理
 - 控制 34

[C]

- 操作系统
 - 64 位 6
- 重新启动 Web Service 31
- 处理器 6
 - 多个 34
- 磁盘空间 7
 - 计算 8

[D]

- 单点登录 11
- 登录 IBM SPSS Modeler Server 11
- 端口号
 - IBM SPSS Modeler Server 11, 17, 33
- 端口设置
 - options.cfg 文件 34
 - 多个流执行 34

[F]

- 防火墙设置
 - options.cfg 文件 34
- 服务器
 - 单点登录 12, 14
 - 登录 11
 - 添加连接 17
 - 通过 COP 搜索服务器 17
- 服务器的自动启动
 - 基于 UNIX 进行配置 67
- 服务器端口设置
 - options.cfg 文件 34

[G]

- 管理
 - IBM SPSS Modeler Server 的 30
- 管理员访问权
 - 对于 IBM SPSS Modeler Server 31
 - 使用"用户访问控制"(UAC) 32

[H]

- 缓存压缩 34
- 缓存, 数据库内 39
- 回送 53, 54, 56, 58, 64
 - CLEM 表达式 61

[J]

- 加密
 - FIPS 37
 - SSL 40
- 建模
 - 内存管理 34
- 节点
 - 支持 SQL 生成 58, 64
- 节点缓存
 - 写入数据库 57
 - SQL 生成 57
- 结果
 - 记录顺序 19
 - 取整 19
 - Client 和 Server 之间的差异 19
- 进程协调程序 17
 - 服务器集群 81
 - 负载均衡 81
- 进程协调器配置
 - 对于 IBM SPSS Modeler Server 36
- 进程, 无响应 22

[K]

- 客户机
 - 单点登录 15

[L]

- 连接
 - 到 IBM SPSS Modeler Server 11, 17
 - 服务器集群 17
- 临时文件 7

临时文件 (续)

- IBM SPSS Modeler Server 的权限 19
- 流重写 57
- 路径 9

[M]

- 密码
 - IBM SPSS Modeler Server 11
- 模型块
 - 查看 SQL 57

[N]

- 内存 8
- 内存管理
 - 管理选项 34

[P]

- 配置选项
 - 并行处理 34
 - 登录尝试 33
 - 端口号 33
 - 概述 32, 33
 - 进程协调器 36
 - 连接和会话 33
 - 流重写 34
 - 内存管理 34
 - 数据文件访问 33
 - 性能和优化 34
 - 自动 SQL 生成 35
 - COP 36
 - IBM SPSS Modeler Server 的 30
 - SQL 字符串长度 35
 - SSL 数据加密 35
 - temp 目录 33

[Q]

- 启动 IBM SPSS Modeler Server
 - 在 UNIX 中 21
 - 在 Windows 中 21
- 权限 19

[R]

- 认证 18

日志文件

对于 IBM SPSS Modeler Server 47

显示生成的 SQL 56

[S]

示例

概述 4

应用程序指南 3

数据访问 8

数据访问包

和 UNIX 库路径 73

针对下列对象配置 UNIX 68

ODBC, 在 UNIX 上配置 68

UNIX 上的 ODBC 故障诊断 71

数据库

访问 8

数据库服务器 50

数据库缓存

从 options.cfg 中控制 39

SQL 生成 57

数据库连接

关闭 39

数据库内缓存 39

数据库内挖掘 50

数据文件

导入和导出 10

IBM SPSS Statistics 10

数据源

单点登录 16

[T]

体系结构

组件 5

添加 IBM SPSS Modeler Server 连接

17

停止 IBM SPSS Modeler Server

在 UNIX 中 21

在 Windows 中 21

通过 COP 搜索连接 17

[W]

文档 3

文件名

UNIX 9

Windows 9

文件权限

基于 UNIX 进行配置 68

在 IBM SPSS Modeler Server 上 19

[X]

消息

显示生成的 SQL 56

性能

IBM SPSS Modeler Server 的 49

[Y]

应用程序示例 3

硬件建议

对于 IBM SPSS Modeler Server 6

硬盘 8

用户标识

IBM SPSS Modeler Server 11

用户认证 18

用户帐户

权限 19

IBM SPSS Modeler Server 18

优化

SQL 生成 53, 54, 56

预览

SQL 生成 56

域名 (Windows)

IBM SPSS Modeler Server 11

运算符

SQL 生成 61

[Z]

执行流时出现错误 34

主机名

IBM SPSS Modeler Server 11, 17

状态

UNIX 上 IBM SPSS Modeler Server

的 21

Windows 上的 IBM SPSS Modeler

Server 的 21

紫色节点

SQL 优化 54

[数字]

64 位操作系统 6

A

allow_modelling_memory_override

options.cfg 文件 34

auto.sh (UNIX)

的位置 67

C

cache_compression

options.cfg 文件 34

cache_connection 选项 39

chemsrv.sh (UNIX)

的位置 67

CLEM 表达式

SQL 生成 61

Cognos SSL 连接 44

Cognos TM1 SSL 连接 45

COP 17

服务器集群 81

负载均衡 81

COP 配置

对于 IBM SPSS Modeler Server 36

cop_enabled

options.cfg 文件 36

cop_host

options.cfg 文件 36

cop_password

options.cfg 文件 36

cop_port_number

options.cfg 文件 36

cop_service_description

options.cfg 文件 36

cop_service_host

options.cfg 文件 36

cop_service_name

options.cfg 文件 36

cop_service_weight

options.cfg 文件 36

cop_update_interval

options.cfg 文件 36

cop_user_name

options.cfg 文件 36

D

data_files_restricted

options.cfg 文件 33

data_file_path

options.cfg 文件 33

DB2

SQL 优化 53, 54

F

FIPS 加密 37

G

group_configuration 37

I

- IBM SPSS Analytic Server
 - 配置选项 33
- IBM SPSS Modeler 1
 - 文档 3
- IBM SPSS Modeler Administration
 - Console 30
 - 管理员访问权 31
 - "用户访问控制"访问权 32
- IBM SPSS Modeler Client
 - 单点登录 15
- IBM SPSS Modeler Server 1
 - 不响应进程 22
 - 单点登录 12, 14
 - 端口号 11, 17, 33
 - 服务器进程 39
 - 供最终用户阅读的信息 18
 - 管理 30
 - 管理选项 30
 - 管理员访问权 31
 - 监视使用情况 39
 - 进程协调器配置 36
 - 密码 11
 - 配置选项 32
 - 权限 19
 - 数据源单点登录 16
 - 文件创建 19
 - 用户标识 11
 - 用户认证 18
 - 用户帐户 18
 - 与客户端的不同结果 19
 - 域名 (Windows) 11
 - 主机名 11, 17
 - COP 配置 36
 - temp 目录 33
 - "用户访问控制"访问权 32
- IBM SPSS Statistics 数据访问技术 8
- IBM SPSS Statistics 数据文件
 - 导入和导出 10
- io_buffer_size
 - options.cfg 文件 34

K

- Kerberos 37

L

- LDAP
 - 认证 83
- Linux
 - 单点登录 14

M

- max_file_size
 - options.cfg 文件 33
- max_login_attempts
 - options.cfg 文件 33
- max_parallelism
 - options.cfg 文件 34
- max_sessions
 - options.cfg 文件 33
- max_sql_string_length
 - options.cfg 文件 35
- memory_usage
 - options.cfg 文件 34
- Microsoft SQL Server
 - SQL 优化 53, 54
- modelling_memory_limit_percentage
 - options.cfg 文件 34

O

- ODBC
 - 基于 UNIX 进行配置 68
- ODBC 驱动程序设置 53
- ODBC 数据源
 - 和 UNIX 68
 - ODBC 和 UNIX 脚本 68
- options.cfg 37
- options.cfg 文件 39
- Oracle
 - SQL 优化 53, 54, 65

P

- PAM
 - 认证 83
- port_number
 - options.cfg 文件 33
- program_files_restricted
 - options.cfg 文件 33
- program_file_path
 - options.cfg 文件 33

R

- RAM 8
- rc.modeler (UNIX)
 - 的位置 67

S

- server_logging.log 47
- SQL
 - 查看模型块 57
 - 查询 63

SQL (续)

- 重复列名 63
- 优化 Oracle 65
- 预览生成的 56
- SQL 回送。另请参阅 SQL 生成 53
- SQL 生成 53, 54, 58, 64
 - 查看模型块 57
 - 对 IBM SPSS Modeler Server 启用 35
 - 流重写 57
 - 启用 56
 - 日志记录 56
 - 提示 57
 - 预览 56
 - CLEM 表达式 57, 61
- SQL Server
 - SQL 优化 53, 54
- sql_generation_enabled
 - options.cfg 文件 35
- SSL 40
 - 概述 40
 - 确保通信安全 40
 - Cognos 连接 44
 - Cognos TM1 连接 45
- SSL 数据加密
 - 对 IBM SPSS Modeler Server 启用 35
- ssl_certificate_file
 - options.cfg 文件 35
- ssl_enabled
 - options.cfg 文件 35
- ssl_private_key_file
 - options.cfg 文件 35
- ssl_private_key_password
 - options.cfg 文件 35
- stream_rewriting_enabled
 - options.cfg 文件 34

T

- temp 目录
 - 对于 IBM SPSS Modeler Server 33
- temp_directory
 - options.cfg 文件 33

U

- UNC 文件名 9
- UNIX
 - 重新启动 Web Service 31
 - 单点登录 14
 - 库路径 73
 - 配置文件权限 68
 - 权限 19
 - 用户认证 18

UNIX 脚本
 编辑 68
 auto.sh 67
 modelersrv.sh 67
 rc.modeler 67
UNIX 内核限制 21
UNIX 上的内核限制 21
UNIX shell 37

W

Web Service - 重新启动 31
Windows
 重新启动 Web Service 31

Z

zombie 进程, IBM SPSS Modeler
 Server 22

[特别字符]

@SQLFN 函数 63



Printed in China